Performance Analysis of the LBP–SVM Model in Deepfake Image Detection: A Case Study on the FaceForensics++ Dataset and External Validation on Indonesian Politicians' Faces

M.David Rahadian[1], Abdul Aziz[2], Dwi Wahyu Prabowo[3]
[1,2,3] *Information system, Unda University, Jl. Batu Berlian No.10, Mentawa Baru Hulu Kec. Mentawa Baru Ketapang Kabupaten Kotawaringin Timur, Kalimantan Tengah, Indonesia.*

Keywords

Deepfake detection; Local Binary Pattern; Support Vector Machine; handcrafted features; FaceForensics++; cross-domain validation.

Correspondence Email:
Davidrahadian2023@gmail.com

Abstract

DeepFake images synthetic face images generated by deep neural networks (GANs) can seriously undermine media authenticity and public trust This study analyzes deepfake image detection based on handcrafted features using Local Binary Pattern (LBP) and Support Vector Machine (SVM). The primary dataset consists of 2,000 images from FaceForensics++, comprising 1,000 authentic and 1,000 manipulated images, with a 70:30 stratified train-test split. Each image undergoes preprocessing steps including grayscale conversion, intensity normalization, and resizing to 128×128 pixels before feature extraction using the uniform LBP operator (P=8, R=1). An SVM with a Radial Basis Function (RBF) kernel is employed as the baseline classifier, achieving an accuracy of 0.63, a macro-F1 score of 0.63, and a ROC–AUC of 0.65, demonstrating higher sensitivity to the fake class. To assess cross-domain generalization, an external validation was conducted using 10 images of Indonesian politicians (6 real, 4 fake). The model correctly classified 6 out of 10 images but showed a bias toward the real class due to differences in texture and lighting. These results suggest that the LBP–SVM approach remains relevant for lightweight texture-based deepfake detection, though it is not yet optimal for real-world domain variations. Future research is recommended to explore LBP combined with HOG or deep feature embeddings to improve accuracy and model robustness in practical scenarios.

1. Introduction

The rapid advancement of Artificial Intelligence (AI), particularly in machine learning and computer vision, has enabled the creation of synthetic media known as deepfakes. Through Generative Adversarial Networks (GANs), deepfakes can replicate facial appearance and expressions with high realism, posing risks to digital authenticity and automated verification systems (Korshunov & Marcel, 2019). Although deepfake technology supports fields such as digital arts, education, and entertainment, its misuse for misinformation, identity fraud, and political manipulation has raised widespread concern (Patel et al., 2023).

These risks are especially visible in regions with limited digital literacy. In Indonesia, manipulated images and videos of political figures have increasingly circulated during sensitive periods such as election campaigns, undermining public trust and potentially influencing democratic processes (Seow et al., 2022; Vaccari & Chadwick, 2020) . This situation highlights the need for lightweight and accessible detection methods that do not rely on computationally expensive deep learning models.

While state-of-the-art detection systems such as XceptionNet, EfficientNet, and Vision Transformers offer high accuracy, they require substantial computational resources and large training datasets, limiting their practical deployment in low-resource environments (Abdullah et al., 2024). Consequently, handcrafted feature-based approaches remain relevant as transparent, efficient, and interpretable baselines for deepfake detection (Nguyen et al., 2022).

Local Binary Pattern (LBP) is one of the most widely adopted texture descriptors for facial analysis due to its computational efficiency and robustness to illumination changes (Xia et al., 2022). When combined with classical machine learning classifiers such as Support Vector Machines (SVM), LBP has shown potential in detecting local texture inconsistencies commonly found in GAN generated images (Matern et al., n.d.) . These properties make LBP–SVM a promising and resource-efficient approach for practical detection scenarios.

This study evaluates the performance of the LBP–SVM pipeline using 2,000 images from the FaceForensics++ dataset and an additional external dataset consisting of Indonesian political figures. The objective is to assess both in-domain performance and cross-domain generalization, particularly when models are applied to different demographic and environmental conditions. The results are intended to provide a reproducible baseline and practical insights for developing lightweight deepfake detection systems suitable for deployment in resource-constrained environments.

1.1 Literature Review

Research on deepfake detection has expanded rapidly following the introduction of high-fidelity generative models such as Generative Adversarial Networks (GANs). A key milestone in this field is the release of the FaceForensics++ dataset, which provides large-scale, standardized real and manipulated facial images and has become a benchmark for evaluating detection algorithms (Rössler et al., 2019) . Studies consistently rely on this dataset to assess both classical and deep learning approaches.

Detection methods can generally be categorized into handcrafted feature-based and deep learning-based techniques (Kumar et al., 2025; Verdoliva, 2020) . While deep learning models such as XceptionNet, EfficientNet, and Vision Transformers achieve superior performance, they require extensive computational resources and large annotated data, making them less practical in real-world or low-resource environments (Abdullah et al., 2024).

Handcrafted approaches remain relevant for lightweight detection scenarios. Among them, Local Binary Pattern (LBP) stands out as a robust texture descriptor capable of capturing micro-texture variations associated with synthetic facial artifacts. LBP compares pixel intensities within a local neighborhood, producing rotation-invariant texture representations that perform reliably under illumination changes (Liu et al., 2017) . When paired with Support Vector Machines (SVM), LBP has demonstrated competitive results on constrained datasets and is less prone to overfitting than deeper models in small-data conditions (Matern et al., n.d.; Pandey & Tiwari, 2025).

Conversely, deep learning-based models dominate state-of-the-art performance due to their ability to learn hierarchical and spatially rich features directly from images. Architectures such as MesoNet and XceptionNet have shown strong performance in detecting compression artifacts and facial inconsistencies typical of deepfakes (Afchar et al., 2018; Verdoliva, 2020) . However, these models face challenges in generalization when exposed to new domains, skin tones, lighting variations, or image compression patterns (Yang et al., 2023)

Cross-domain generalization has therefore become a major focus in recent literature. Several studies report sharp performance drops when models are evaluated on datasets different from those used during

training, a phenomenon driven by demographic, photometric, and environmental differences (Y. Wu et al., 2023; Yin et al., 2024). Although handcrafted features are less expressive, some evidence suggests they may generalize more consistently than deep features due to their lower dependence on data distribution (L. Wu et al., 2024).

Hybrid methods have also been proposed, combining handcrafted descriptors such as LBP with deep embeddings to improve robustness while maintaining efficiency. Such fusion strategies have demonstrated improved cross-dataset performance compared to using either handcrafted or deep features alone (Lanzino et al., n.d.).

In summary, while deep learning models remain dominant in accuracy, handcrafted feature-based approaches particularly LBP combined with SVM retain practical importance for interpretable, lightweight, and resource-efficient detection. Their relevance is especially significant in regions like Indonesia, where domain-specific datasets are limited and computational resources vary widely. This study builds on these insights by evaluating LBP–SVM not only on FaceForensics++ but also on real-world political images to assess its cross-domain generalization.

2. Research Methods

2.1 Experimental Design

This study adopts a quantitative experimental design with an empirical approach rooted in classical machine learning. The primary objective is to investigate the performance of handcrafted features, specifically Local Binary Pattern (LBP), when combined with Support Vector Machine (SVM) classifiers in the context of deepfake image detection. All experiments were conducted using Python 3.10 on the Google Colab platform, leveraging libraries such as OpenCV, scikit-image, NumPy, and scikit-learn to process, extract, and classify image features systematically.

2.2 Dataset

Two datasets were used in this study. The primary dataset is a curated subset of FaceForensics++, containing 2,000 images consisting of 1,000 real and 1,000 fake samples (Rössler et al., 2019). The images represent a variety of poses, illumination conditions, and compression levels, which are essential for robust texture analysis. A stratified split was applied to produce 1,407 training samples and 603 internal test samples while maintaining a balanced distribution between real and fake classes. This balance supports stable evaluation for binary classification tasks.

An external validation dataset was also collected, consisting of ten publicly available facial images of Indonesian political figures. Six images are authentic, and four are deepfakes identified from online sources. The images exhibit noticeable domain differences such as Southeast Asian skin tones, varied lighting, social media compression artifacts, and heterogeneous camera quality. These characteristics introduce domain shift, which is a documented challenge for deepfake detection models (Y. Wu et al., 2023; Yin et al., 2024).

Table 1 Characteristics of the datasets used in the LBP–SVM experiments

Dataset Type	Source	Total Images	Real	Fake	Class Balance	Notes
Training Set	FaceForensics++	1,407	704	703	~50:50	Used to train the model in a controlled environment
Internal Test Set	FaceForensics++	603	296	307	~50:50	Used to evaluate in-domain performance under the same distribution as training.
External Validation Set	Indonesian Politicians	10	6	4	60:40	Contains domain shift in ethnicity, lighting, compression, and facial structure

The combination of an in-domain benchmark and a domain-shifted external dataset provides a comprehensive basis for measuring both accuracy and generalization capability. This structure also aligns with the increasing emphasis on evaluating deepfake detectors in realistic environments.

2.3 Preprocessing

All images were processed using a uniform preprocessing pipeline to ensure consistency before feature extraction. Each image was resized to 128 x 128 pixels to standardize spatial dimensions. This step mitigates variation in original image resolution from different camera devices and sources.

Images were then converted to grayscale because LBP relies on luminance information rather than color. Removing color channels reduces noise and computational overhead while emphasizing structural and textural cues relevant to forgery detection (Liu et al., 2017).

Finally, pixel values were normalized to a range between [0,1] using min–max scaling. Normalization helps stabilize the LBP operator under varying illumination conditions. Studies have shown that handcrafted descriptors benefit significantly from intensity normalization as it reduces brightness inconsistencies that could distort texture statistics (Villegas-Camacho et al., 2025). This preprocessing pipeline produces images that are spatially consistent, luminance uniform, and intensity normalized, which strengthens the reliability of subsequent feature extraction.

2.4 Feature Extraction (LBP)

Texture features were extracted using the uniform Local Binary Pattern (LBP) operator, which analyzes the local neighborhood of each pixel by comparing its intensity value with its surrounding pixels. The LBP was applied with parameters: P=8 (number of neighbors) and R=1 (radius), and the 'uniform' method was selected to focus on consistent local texture patterns(Yasmin et al., 2020).

The formula for computing LBP is as follows:

$$LBP_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} s(g_p - g_c) \cdot 2^p,$$

where g_c is the intensity of the central pixel and g_p is the intensity of the neighboring pixel at position p. The resulting histogram contains $P + 2 = 10$ bins, each representing a unique uniform pattern. Each image is thus represented as a 10-dimensional feature vector, which is compiled into a matrix $X \in \mathbb{R}^{2000\times10}$, accompanied by corresponding binary labels $y \in \{0,1\}$(Sedaghatjoo et al., 2024).

2.5 Data Splitting and Standardization

The dataset was divided into training and testing subsets using a stratified split, preserving the class distribution to avoid model bias. Feature standardization was applied only to the training set using Z-score normalization(Villegas-Camacho et al., 2025).

$$Z = \frac{x - \mu_{train}}{\sigma_{train}}$$

where μ_{train} and σ_{train} are the mean and standard deviation of the training data. The same transformation was applied to the test set using the training parameters to prevent data leakage(Alturayeif & Hassine, 2025).

2.6 SVM Classification

A Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel was selected as the classifier due to its capacity to model non-linear decision boundaries. The SVM was trained using default parameters C =1.0 and γ=′scale in scikit-learn, and optimized via internal grid search during model tuning(Carrington et al., 2023).

Performance was assessed using multiple evaluation metrics, including:

- Accuracy
- Precision
- Recall
- F1-score (macro average)

- Area Under the ROC Curve (ROC–AUC)

The initial model achieved an accuracy of 0.63, macro F1-score of 0.63, and ROC–AUC of 0.653, indicating moderate yet stable classification performance for a lightweight model. Threshold optimization was also applied by identifying the probability threshold that maximizes macro-F1, which was empirically found to be 0.51, slightly improving sensitivity to the fake class without modifying model weights(Leevy et al., 2023).

2.7 External Validation

To evaluate the model's ability to generalize beyond the controlled conditions of FaceForensics++, an external validation dataset was compiled using ten publicly available facial images of Indonesian political figures. The dataset consists of six authentic images and four manipulated images originating from online sources. Compared to the primary dataset, these images exhibit substantial domain differences such as Southeast Asian skin tones, varied lighting conditions, social media compression artifacts, and heterogeneous camera quality.

These characteristics introduce a domain shift that is widely recognized in deepfake detection research (Kumar et al., 2025; Yin et al., 2024). The purpose of this external evaluation is not to interpret detailed performance outcomes, but to assess how the trained LBP–SVM model responds to real-world variations. All performance results and analytical discussions related to this external validation are presented comprehensively in Section 3.3.

Table 2 presents the complete classification results, including the politician's name, corresponding facial image, ground-truth label, model prediction, confidence score, and correctness evaluation.

Politician Name	Politician Image	Ground Truth	Predicted	Confidence	Correct
Purbaya Yudhi Sadewa		Real	Real	0.524	Correct
Anies Baswedan		Real	Real	0.597	Correct
Prabowo Subianto		Real	Real	0.658	Correct
Sri Mulyani		Fake	Fake	0.512	Correct
Joko Widodo		Fake	Real	0.609	Incorrect
Joko Widodo		Real	Fake	0.500	Incorrect
Sri Mulyani		Real	Real	0.671	Correct
Joko Widodo		Fake	Real	0.561	Incorrect
Bahlil Lahadalia		Real	Real	0.660	Correct

Suharto		Fake	Real	0.673	Incorrect

Despite these limitations, the LBP SVM model still demonstrated partial capability in identifying manipulated images. Observable blending inconsistencies and illumination anomalies remain detectable under this approach.

2.8 Method Overview

The overall research workflow is depicted as follows:

1. Data Acquisition (FaceForensics++ and local politician images)
2. Image Preprocessing (resize, grayscale, normalize)
3. Feature Extraction (LBP uniform)
4. Feature Normalization (Z-score)
5. Model Training (SVM with RBF kernel)
6. Evaluation (internal: 70/30 split)
7. External Validation (Indonesian political images)

3. Result and Discussion

3.1 Internal Experiment Performance

The performance of the Local Binary Pattern (LBP) and Support Vector Machine (SVM) model was first evaluated using a stratified hold-out validation on the FaceForensics++ dataset. With 1,407 training images and 603 test images with a near-balanced distribution of real and fake samples, the model yielded the following metrics:

- Accuracy: 63%
- Precision: 64%
- Recall: 63%
- F1-score (macro): 0.63
- ROC–AUC: 0.653

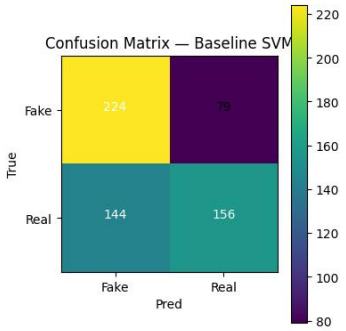

Figure 1 The confusion matrix illustrates the classification performance of the Baseline SVM model, showing 224 correctly identified fake samples and 156 correctly identified real samples, with notable misclassifications between the two categories.

The confusion matrix revealed that 224 out of 303 fake images were correctly identified, while 144 real images were misclassified as fake. In contrast, real images with subtle gradients and smooth skin regions often overlapped with the distribution of fake features, causing a decrease in class separability.

This behavior aligns with previous findings in handcrafted detection methods, where LBP is known for its high responsiveness to abrupt local texture variations but limited global spatial awareness (Matern et al., n.d.) (Kingra et al., 2022) . The moderate AUC score of 0.653 further suggests that while the model possesses discriminative capability, it remains challenged by overlapping decision boundaries in complex facial regions.

Importantly, threshold tuning to 0.51 optimized macro-F1 without overfitting, indicating that decision boundaries could be adjusted for specific recall–precision trade-offs depending on downstream application needs, such as minimizing false negatives in security use-cases.

3.2 Visual Analysis and Interpretability

To deepen understanding of model behavior, key evaluation metrics were visualized using the following tools:

- Confusion Matrix
- ROC Curve
- Precision–Recall Curve

The confusion matrix demonstrates that the model performs better in identifying fake images than real ones.

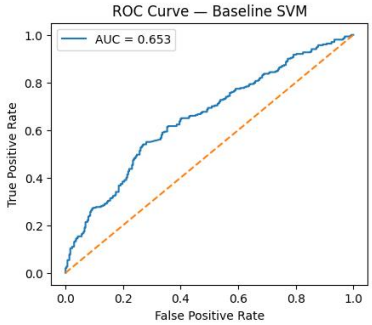

Figure 2 The ROC curve plots the True Positive Rate against the False Positive Rate, evaluating the model's ability to distinguish between fake and real inputs. The AUC value of 0.653 reflects a fair level of discrimination performance for the Baseline SVM

The ROC curve (AUC = 0.653) suggests that the model maintains a moderate balance between true positive rate (TPR) and false positive rate (FPR) across different thresholds. While not close to a perfect classifier, the ROC curve does not approximate the diagonal line of random guessing, which implies nontrivial learning.

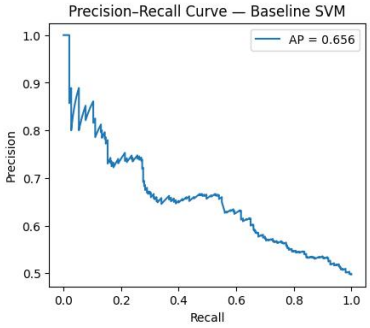

Figure 3 The Precision–Recall curve demonstrates the trade-off between precision and recall for the Baseline SVM. The area under the curve (AP = 0.656) indicates moderate precision consistency across different recall thresholds.

More revealing is the Precision–Recall (PR) curve, which highlights the challenge of balancing sensitivity and specificity in texture-based classification. The decline in precision with increasing recall implies that additional detections of fake images also increase false positives, especially in real images with smooth transitions or poor lighting areas that often confuse LBP.

These insights indicate that LBP–SVM performs better in binary separations when the texture is prominent or degraded but struggles with natural, soft-toned imagery where pixel-level variance is low. Thus, while it is computationally efficient, its interpretability and stability are limited by feature generalization across diverse textures.

3.3 External Validation: Cross-Domain Performance

The external validation experiment consisted of ten facial images of Indonesian political figures, including six authentic and four manipulated samples, designed to assess the generalization capability of the

LBP–SVM model beyond the FaceForensics++ training domain. These real-world images introduced substantial domain variation in skin pigmentation, illumination levels, camera quality, facial geometry, and demographic characteristics. Such heterogeneity represents a pronounced domain shift, where the statistical distribution of texture features differs from those learned during training, posing challenges for handcrafted descriptors such as LBP.

The model correctly classified 6 out of 10 images, resulting in an external accuracy of 60%. This decline in performance, compared to the internal evaluation, indicates that the LBP–SVM classifier is sensitive to domain shift. Misclassifications primarily occurred in real images captured under soft or uneven lighting, as well as manipulated images that exhibited minimal generative artifacts. These conditions generate texture patterns that closely resemble authentic skin regions, resulting in overlapping LBP histogram distributions and reduced separability between real and fake samples.

These findings are consistent with prior studies, which report that handcrafted approaches often struggle to generalize across unseen demographic or photometric domains (P & Subudhi, 2024) . Without adaptation mechanisms, such models fail to maintain accuracy when confronted with images exhibiting statistical properties that differ from the training dataset. Nevertheless, the model's correct classification of one manipulated image demonstrates that GAN-produced textures still contain detectable inconsistencies particularly in edge transitions, local illumination discrepancies, and blending artifacts which LBP can capture to some extent.

Overall, the cross-domain evaluation highlights both the strengths and limitations of the LBP–SVM pipeline. While lightweight and partially effective in realistic scenarios, its generalization capability remains insufficient for robust deployment across diverse populations without the incorporation of adaptive, domain-aware, or hybrid feature-learning strategies.

3.4 Implications and Limitations

The experimental outcomes affirm the value of LBP–SVM as a computationally lightweight baseline. Its ability to detect subtle texture anomalies in facial imagery renders it suitable for environments with limited resources such as mobile devices, low-power servers, or field investigations where high-efficiency models are essential.

However, several limitations emerged:

1. Lack of global context: LBP captures local texture only, ignoring holistic facial structure.
2. Domain sensitivity: The model exhibits performance decay when applied to new domains or demographics, suggesting limited generalization.
3. Lighting and compression artifacts: External factors such as brightness or image quality significantly impact histogram-based descriptors.

These findings align with broader literature, reinforcing that handcrafted model must be either enhanced or hybridized with deeper, representation-rich models to maintain performance in diverse real-world scenarios (Nguyen et al., 2022; Verdoliva, 2020). Moreover, interpretability in handcrafted models like LBP can serve as an explanatory foundation for more complex deep learning frameworks.

3.5 Practical Recommendations and Strategic Insights

The findings of this study offer both theoretical and practical implications, particularly for practitioners and researchers involved in lightweight deepfake detection systems. Despite its limitations, the LBP–SVM pipeline presents several distinct advantages that can be leveraged in real-world environments under constrained computational conditions (Ha et al., 2025) . These insights are discussed below in three main dimensions: implementation feasibility, architectural improvements, and policy-level integration(Balafrej & Dahmane, 2024).

3.5.1 Lightweight Deployment and Edge Applications

One of the most significant strengths of the LBP–SVM model is its computational simplicity. The combination of Local Binary Pattern for feature extraction and a linear or RBF-kernel SVM for classification

allows for real-time or near real-time inference without reliance on graphical processing units (GPUs). In contexts such as media forensics, digital journalism, or civic monitoring especially in low-resource regions like rural Indonesia such systems can be deployed directly on mobile devices, embedded systems, or lightweight cloud interfaces.

This computational efficiency supports its role as a first-line screening system to flag suspicious media for further review. For instance, newsrooms or fact-checking organizations could integrate LBP–SVM modules into their content ingestion pipelines, enabling rapid triage of potentially manipulated images prior to manual review or deeper neural network analysis.

3.5.2 Baseline Interpretability for Multimodal Pipelines

Another advantage of handcrafted approaches is their interpretability. Unlike deep convolutional models, whose decision-making processes are often opaque, LBP-based histograms offer an intuitive, transparent representation of facial texture distributions. This property becomes particularly valuable when used as part of a hybrid architecture, where handcrafted descriptors complement deep feature embeddings.

For example, as suggested in recent studies, LBP histograms could be used as auxiliary inputs to a neural decision layer, or as pre-filters that reduce false positives by cross-validating texture inconsistencies (Verdoliva, 2020).

3.5.3 Domain Adaptation and Augmented Local Training

The observed drop in performance during external validation underscores the critical role of domain adaptation. To address this, practitioners are advised to adopt cross-domain data augmentation techniques during training such as altering illumination, camera angle, and compression artifacts to simulate conditions commonly found in local contexts.

In addition, curating localized datasets (e.g., facial imagery from Southeast Asian demographics) would provide the statistical grounding needed to improve generalization. Collaboration with local institutions, media agencies, or government bodies to build such datasets could accelerate the development of domain-aware detection systems. These systems would not only perform better on localized content but also help protect political institutions from misinformation campaigns involving AI-generated imagery.

3.5.4 Ethical Implementation and Governance

Beyond technical considerations, the deployment of deepfake detection models raises ethical and governance challenges. The LBP–SVM pipeline, due to its simplicity, poses minimal privacy risk during inference, as it avoids end-to-end learning or feature embedding on cloud-hosted servers. This property is advantageous in contexts where data privacy and sovereignty are legally sensitive, such as in election monitoring or judicial investigations.

3.5.5 Future Integration with Multimodal Systems

Finally, the increasing use of multimodal deepfake content involving not just facial imagery but also voice, lip movement, and behavioral cues necessitates a broader framework. While this study focuses on static images, the foundational LBP–SVM approach could be extended into video sequences by applying temporal LBP (LBP-TOP) and integrating temporal smoothing layers with low-latency decision logic.

4. Conclusions

4.1 Conclusion

This research presents an empirical study on the detection of deepfake images using a lightweight and interpretable approach that leverages Local Binary Pattern (LBP) for texture feature extraction and Support Vector Machine (SVM) for classification. The study was grounded in the context of increasing threats posed by AI-generated manipulated media, especially within politically sensitive environments such as Indonesia. By using 2,000 images from the FaceForensics++ dataset equally split between real and manipulated facial images and implementing a stratified hold-out split (70% training and 30% testing), the proposed pipeline aimed to assess the feasibility of handcrafted features in deepfake detection.

The experimental results showed that the LBP–SVM model achieved a classification accuracy of 63%, a macro F1-score of 0.63, and a ROC–AUC score of 0.65 on the internal test set. The model demonstrated slightly higher sensitivity toward fake images, likely due to its ability to capture micro-texture artifacts typically introduced by GAN-based manipulation methods. These include inconsistencies in facial blending, illumination artifacts, and unnatural skin texture variations, which are often overlooked by casual human observers but can be identified through consistent local pattern differences.

To examine generalization capability, an external validation was conducted using a set of ten facial images of Indonesian politicians, consisting of both real and fake samples. The model correctly classified six out of ten, indicating moderate cross-domain performance. However, it also revealed limitations in handling distributional shifts such as different lighting conditions, skin tone diversity, camera quality variations, and regional facial structures not present in the original training data.

Overall, the study affirms that the LBP–SVM pipeline, while not state-of-the-art, remains a viable baseline for environments with limited computational resources or constrained data availability. It also highlights that handcrafted feature-based models can still offer reliable performance under certain conditions, especially when deployed as a first-stage filter prior to deeper analysis. The findings emphasize the balance between interpretability and effectiveness, calling attention to the importance of domain-specific data, model generalizability, and robust validation protocols in real-world deployment scenarios.

4.2 Suggestions for Future Work.

While the results of this study are promising, several key areas for future research and development emerge from the findings:

1. **Enhancement of Feature Representation:**
   Although LBP captures fine-grained local texture patterns, it lacks spatial contextual awareness. Future research should investigate feature fusion strategies, such as combining LBP with Histogram of Oriented Gradients (HOG), Gabor wavelets, or Gray-Level Co-occurrence Matrices (GLCM), to enhance texture discriminability and global facial structure awareness. These combinations may provide better resilience against lighting variations and compressive artifacts (Bao et al., 2023; Chen et al., 2021).
2. **Hybrid Models with Deep Embeddings:**
   Incorporating deep learning representations as a second layer in the detection pipeline could significantly improve performance (Naskar et al., 2024) . Lightweight pre-trained CNNs (e.g., MobileNet, EfficientNet-lite) or attention-based transformers could be used to extract semantically rich embeddings, which are then combined with handcrafted descriptors. This dual-layer approach would retain low computational cost for preliminary screening, while improving detection precision in complex scenarios (Ha et al., 2025; Naskar et al., 2024).
3. **Cross-Domain Adaptation Strategies:**
   The external validation results demonstrated that the LBP–SVM model suffers from domain shift when applied to data from a different geographical or cultural context. Future work should implement domain adaptation methods such as adversarial training, feature alignment, or transfer component analysis to reduce discrepancies between source and target feature distributions. Simulating local data characteristics through augmentation (e.g., tropical lighting, diverse skin tones) can also enhance model robustness(Qi et al., 2024).
4. **Development of Local Datasets:**
   There is a critical need to develop curated and representative deepfake datasets that reflect the facial diversity and imaging conditions of Southeast Asian populations. Establishing an ethically collected Indonesian political deepfake dataset, with proper consent and annotation standards, will serve as a foundation for benchmarking and fair evaluation of detection systems in regional contexts (Heidari et al., 2023).
5. **Real-Time Application and Interpretability:**

Future studies should explore real-time implementation scenarios, particularly on mobile and embedded systems (Zhang et al., 2023) . Additionally, integrating explainability tools such as heatmaps or saliency maps will help users understand model decisions, enhancing transparency and trust particularly when used in forensic or journalistic environments.

In conclusion, while the handcrafted LBP–SVM model provides a useful and replicable starting point, future work must build upon its limitations and explore integrated, adaptive, and ethically sound methodologies for robust and scalable deepfake detection systems across diverse real-world domains.

5. References

Abdullah, S. M., Cheruvu, A., Kanchi, S., Chung, T., Gao, P., Jadliwala, M., & Viswanath, B. (2024). *An Analysis of Recent Advances in Deepfake Image Detection in an Evolving Threat Landscape*. http://arxiv.org/abs/2404.16212

Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: a Compact Facial Video Forgery Detection Network. *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–7. https://doi.org/10.1109/WIFS.2018.8630761

Alturayeif, N., & Hassine, J. (2025). Data leakage detection in machine learning code: transfer learning, active learning, or low-shot prompting? *PeerJ Computer Science*, *11*, e2730. https://doi.org/10.7717/peerj-cs.2730

Balafrej, I., & Dahmane, M. (2024). Enhancing practicality and efficiency of deepfake detection. *Scientific Reports*, *14*. https://doi.org/10.1038/s41598-024-82223-y

Bao, X., Wang, H., Li, F., Zhang, J., Wang, Y., & Yan, S. (2023). Hybrid deepfake detection using LBP and CNN embeddings. *Journal of Visual Communication and Image Representation*.

Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., Sheikh, Y., McInnes, M., & Holzinger, A. (2023). Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 329–341. https://doi.org/10.1109/TPAMI.2022.3145392

Chen, T., Gao, T., Li, S., Zhang, X., Cao, J., Yao, D., & Li, Y. (2021). A novel face recognition method based on fusion of LBP and HOG. *IET Image Process.*, *15*, 3559–3572. https://doi.org/10.1049/ipr2.12192

Ha, J., Azzaoui, A. El, & Park, J. H. (2025). FL-TENB4: A Federated-Learning-Enhanced Tiny EfficientNetB4-Lite Approach for Deepfake Detection in CCTV Environments. *Sensors (Basel, Switzerland)*, *25*. https://doi.org/10.3390/s25030788

Heidari, A., Navimipour, N. J., Dağ, H., & Unal, M. (2023). Deepfake detection using deep learning methods: A systematic and comprehensive review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *14*. https://doi.org/10.1002/widm.1520

Kingra, S., Aggarwal, N., & Kaur, N. (2022). LBPNet: Exploiting texture descriptor for deepfake detection. *Forensic Science International: Digital Investigation*, *42–43*, 301452. https://doi.org/https://doi.org/10.1016/j.fsidi.2022.301452

Korshunov, P., & Marcel, S. (2019). Vulnerability assessment and detection of Deepfake videos. *International Conference on Biometrics (ICB 2019)*.

Kumar, A., Singh, D., Jain, R., Jain, D. K., Gan, C., & Zhao, X. (2025). Advances in DeepFake detection algorithms: Exploring fusion techniques in single and multi-modal approach. *Information Fusion*, *118*, 102993. https://doi.org/https://doi.org/10.1016/j.inffus.2025.102993

Lanzino, R., Fontana, F., Diko, A., Marini, M. R., & Cinque, L. (n.d.). *Faster Than Lies: Real-time Deepfake Detection using Binary Neural Networks*. https://github.com/

Leevy, J. L., Johnson, J. M., Hancock, J., & Khoshgoftaar, T. M. (2023). Threshold optimization and random undersampling for imbalanced credit card data. *Journal of Big Data*, *10*(1), 58. https://doi.org/10.1186/s40537-023-00738-z

Liu, L., Fieguth, P., Guo, Y., Wang, X., & Pietikäinen, M. (2017). Local binary features for texture classification: Taxonomy and experimental study. *Pattern Recognition*, *62*, 135–160. https://doi.org/https://doi.org/10.1016/j.patcog.2016.08.032

Matern, F., Riess, C., & Stamminger, M. (n.d.). *Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations*. https://github.com/shaoanlu/faceswap-GAN

Naskar, G., Mohiuddin, S., Malakar, S., Cuevas, E., & Sarkar, R. (2024). Deepfake detection using deep feature stacking and meta-learning. *Heliyon*, *10*. https://doi.org/10.1016/j.heliyon.2024.e25933

Nguyen, H. H., Yalçın, M., & Li, X. (2022). Advances in deepfake detection: methods, challenges, and future directions. *ArXiv Preprint ArXiv:2301.05545*.

P, D. S., & Subudhi, B. N. (2024). *Adaptive Meta-Learning for Robust Deepfake Detection: A Multi-Agent Framework to Data Drift and Model Generalization*. http://arxiv.org/abs/2411.08148

Pandey, A., & Tiwari, A. K. (2025). ADVANCING FACE SPOOFING DETECTION WITH LBP, PCA, AND SVM: A ROBUST AI SECURITY APPROACH. *ICTACT JOURNAL ON IMAGE AND VIDEO PROCESSING*, 4. https://doi.org/10.21917/ijivp.2025.0508

Patel, Y., Tanwar, S., Gupta, R., Bhattacharya, P., Davidson, I. E., Nyameko, R., Aluvala, S., & Vimal, V. (2023). Deepfake Generation and Detection: Case Study and Challenges. *IEEE Access*, *11*, 143296–143323. https://doi.org/10.1109/ACCESS.2023.3342107

Qi, H., Li, Z., Wu, X., Peng, H., & Wang, Z. (2024). Cross-domain and cross-model deepfake detection based on adversarial learning. *Pattern Recognition Letters*.

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. *Proceedings of the IEEE International Conference on Computer Vision*, 1–11.

Sedaghatjoo, Z., Hosseinzadeh, H., & Bigham, B. S. (2024). *Local Binary Pattern(LBP) Optimization for Feature Extraction*. http://arxiv.org/abs/2407.18665

Seow, J. W., Lim, M. K., Phan, R. C. W., & Liu, J. K. (2022). A comprehensive overview of Deepfake: Generation, detection, datasets, and opportunities. *Neurocomputing*, *513*, 351–371. https://doi.org/https://doi.org/10.1016/j.neucom.2022.09.135

Vaccari, Cristian, & Chadwick, Andrew. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, *6*(1), 2056305120903408. https://doi.org/10.1177/2056305120903408

Verdoliva, L. (2020). Media Forensics and DeepFakes: an overview. *ArXiv Preprint ArXiv:2001.06564*.

Villegas-Camacho, O., Francisco-Valencia, I., Alejo-Eleuterio, R., Granda-Gutiérrez, E. E., Martínez-Gallegos, S., & Villanueva-Vásquez, D. (2025). FTIR-Based Microplastic Classification: A Comprehensive Study on Normalization and ML Techniques. *Recycling*, *10*(2). https://doi.org/10.3390/recycling10020046

Wu, L., Xu, W., Li, X., Li, J., & Chen, W. (2024). Deepfake Detection with Optimized Hybrid Models: Combining Handcrafted and Deep Features. *Applied Soft Computing*.

Wu, Y., Wo, Y., Li, C., & Han, G. (2023). Learning domain-invariant representation for generalizing face forgery detection. *Comput. Secur.*, *130*, 103280. https://doi.org/10.1016/j.cose.2023.103280

Xia, Z., Qiao, T., Xu, M., Zheng, N., & Xie, S. (2022). Towards DeepFake video forensics based on facial textural disparities in multi-color channels. *Information Sciences*, *607*, 654–669. https://doi.org/https://doi.org/10.1016/j.ins.2022.06.003

Yang, S., Guo, H., Hu, S., Zhu, B., Fu, Y., Lyu, S., Wu, X., & Wang, X. (2023). CrossDF: Improving Cross-Domain Deepfake Detection with Deep Information Decomposition. *ArXiv:2310.00359*.

Yasmin, S., Pathan, R. K., Biswas, M., Khandaker, M. U., & Faruque, M. R. I. (2020). Development of a Robust Multi-Scale Featured Local Binary Pattern for Improved Facial Expression Recognition. *Sensors*, *20*(18). https://doi.org/10.3390/s20185391

Yin, Z., Wang, J., Xiao, Y., Zhao, H., Li, T., Zhou, W., Liu, A., & Liu, X. (2024). Improving Deepfake Detection Generalization by Invariant Risk Minimization. *IEEE Transactions on Multimedia*, *26*, 6785–6798. https://doi.org/10.1109/TMM.2024.3355651

Zhang, Y., Gong, K., Chen, J., Xu, Y., & Wang, Z. (2023). Real-time Deepfake Detection on Resource-Constrained Devices using Lightweight CNNs. *IEEE Transactions on Mobile Computing*.