International Journal Software Engineering and Computer Science (IJSECS)
6 (2), 2026, 452-464
Published Online August 2026 in IJSECS (http://www.journal.lembagakita.com/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v6i2.7244.

RESEARCH ARTICLE Open Access

Digital Forensic Analysis of Signature Images Using Error Level Analysis, Image Hashing, and Support Vector Machine Within the DFRWS Framework

Amelia Yahya [1*], **Taswanda Taryo** [2], **Kahfi Heryandi Suradiradja** [3]

[1*] Master of Informatics Engineering, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.
[2,3] Department of Informatics Engineering, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.

**Corresponding author*: ameliayahya.ay@gmail.com.

Received: May 17, 2026; Accepted: June 1, 2026; Published: August 1, 2026.

Abstract: The increasing use of digital documents in administrative and legal activities has expanded the use of image-based signatures for authentication and verification. However, signature images are vulnerable to manipulation using image-editing software, potentially resulting in document forgery and disputes over authenticity. This study examined the use of Error Level Analysis (ELA), perceptual hashing (pHash), and the Gray Level Co-occurrence Matrix (GLCM) to detect manipulation in signature images. It also evaluated the performance of a Support Vector Machine (SVM) in classifying genuine and forged signatures within the Digital Forensic Research Workshop (DFRWS) framework. The dataset comprised 720 signature images obtained from the Starter Handwritten Signatures Dataset. The research process involved image preprocessing, feature extraction, model training, and performance evaluation using a confusion matrix, accuracy, precision, recall, and F1-score. The model achieved an accuracy of 80.56% on previously unseen test data. The developed system also produced visual analysis outputs and generated digital investigation reports based on the DFRWS framework. These results indicate that the combination of ELA, pHash, GLCM, and SVM can support a structured digital forensic process for distinguishing between genuine and forged signature images.

Keywords: Digital Forensics; Signature Image Manipulation; Error Level Analysis; Perceptual Hashing; Support Vector Machine.

1. Introduction

The rapid development of information technology has increased the use of digital documents in administrative activities, electronic transactions, and document validation. Image-based handwritten signatures are often embedded in these documents for authentication and verification purposes. Their digital format offers practical advantages, including easier storage, distribution, retrieval, and document management. Consequently, signature images are increasingly used in documents with administrative and legal significance (Alzahrani & Alotaibi, 2021). However, unlike cryptographic digital signatures, image-based signatures do not inherently guarantee the identity of the signer or the integrity of a document. Their evidentiary reliability therefore depends partly on whether the images can be verified as genuine and free from unauthorized modification.

The availability of image-editing software has made digital signature images easier to copy, modify, and reproduce. Manipulation techniques such as copy-move, splicing, recompression, cropping, and visual alteration can produce forged images that closely resemble genuine signatures (Tolosana *et al.*, 2021). These manipulations may contribute to document forgery, identity misuse, and disputes concerning document authenticity. Detection becomes particularly difficult when manipulated and genuine signatures share similar visual characteristics or when the image has undergone resizing, scanning, repeated compression, or changes in resolution. For this reason, the examination of signature images has become an important aspect of digital forensics, particularly in investigations involving the authenticity and integrity of electronic documents.

Verification is commonly performed through visual inspection by comparing a questioned signature with one or more reference signatures. Although this approach can identify visible differences, its reliability may be affected by subjective judgment, examiner experience, image quality, and the complexity of the forgery. Genuine signatures may also vary naturally because of writing conditions, physical movement, writing instruments, and other individual factors. Conversely, a skilled forgery may reproduce many of the visual characteristics found in an authentic signature. Hafemann *et al.* (2021) noted that these variations make offline handwritten signature verification a challenging pattern-recognition problem. As a result, forged signatures may be accepted as genuine, while authentic signatures may be incorrectly classified as forged. Computational analysis can assist the verification process by extracting image characteristics that may not be assessed consistently through visual examination alone.

Several image-processing and machine-learning methods have been used to address signature verification and image manipulation detection. Huda (2024) applied a Convolutional Neural Network (CNN) to signature verification and reported an accuracy of approximately 86–90%. CNN models can automatically learn visual patterns from images, but they generally require sufficient training data and considerable computational resources. Fahmi (2022) used Error Level Analysis (ELA) for compression-based manipulation detection and reported an accuracy range of approximately 87–91%. ELA can reveal differences in image-compression characteristics, although its effectiveness depends on the image format, compression history, and type of manipulation. Li *et al.* (2023) examined the use of a Support Vector Machine (SVM) for image classification and reported favorable performance on high-dimensional datasets. SVM is suitable for datasets of limited size, but its classification performance depends on the relevance and quality of the extracted features. Because these studies used different datasets and evaluation procedures, their reported results should be interpreted within their respective experimental settings rather than compared directly.

Previous approaches have often focused on a single category of image features, even though signature manipulation may affect several image characteristics. Compression artifacts alone may not adequately represent texture or visual similarity, while texture features may not capture inconsistencies resulting from editing and recompression. This study therefore combines Error Level Analysis (ELA), perceptual hashing (pHash), and the Gray Level Co-occurrence Matrix (GLCM) to represent different characteristics of signature images. ELA is used to identify compression inconsistencies that may be associated with image alteration. Perceptual hashing captures visual characteristics that remain relatively stable under minor modifications, whereas GLCM describes texture through contrast, correlation, energy, and homogeneity. The resulting features are used as inputs to an SVM classifier to distinguish genuine signature images from forged ones. This combination is intended to provide multiple forms of information for classification, although its performance must be assessed through experimental testing.

The examination process is organized using the Digital Forensic Research Workshop (DFRWS) framework to ensure that the handling and analysis of signature images follow structured investigation stages. These stages cover the identification, collection, examination, analysis, and reporting of digital evidence. The analytical methods are also implemented in a web-based system that supports dataset management, image preprocessing, feature extraction, model training, classification, result visualization, and forensic report generation. The system is intended to assist the analysis process rather than replace expert judgment, particularly because classification results may be influenced by image quality, dataset characteristics, and similarities between genuine and forged signatures.

Based on the identified problem and research gap, this study aims to analyze the use of ELA, pHash, and GLCM in extracting characteristics relevant to signature-image manipulation; evaluate the performance of SVM in classifying genuine and forged signature images; apply the DFRWS framework to structure the digital forensic analysis process; and develop a web-based system for signature-image verification and forensic reporting. The study focuses on the classification performance obtained from the combined features and the implementation of these methods within a structured digital forensic workflow.

2. Related Work

This section reviews studies related to handwritten signature verification, digital image manipulation detection, image forensics, and machine learning-based image classification. It also outlines the technical foundations of Error Level Analysis (ELA), perceptual hashing (pHash), the Gray Level Co-occurrence Matrix (GLCM), Support Vector Machine (SVM), and the Digital Forensic Research Workshop (DFRWS) framework. The review identifies the strengths and limitations of previous approaches and establishes the position of the present study.

2.1 State-of-the-Art Research

Research on handwritten signature verification and image manipulation detection has progressed through image-forensic, machine learning, and deep learning approaches. Deep learning methods automatically learn visual representations from image data, whereas conventional machine learning methods rely on features extracted before classification. Image-forensic methods, meanwhile, examine characteristics such as compression inconsistencies, texture patterns, and visual similarities that may indicate image alteration. Huda (2024) developed a signature verification system using a Convolutional Neural Network (CNN) and reported an accuracy of approximately 86–90%. The CNN learned visual patterns directly from signature images, reducing the need for manually designed features. However, the effectiveness of a CNN generally depends on the quantity and diversity of the training data, model configuration, and available computational resources. Saputra and Nurhaida (2024) applied EfficientNetV2M to signature verification and obtained an accuracy of 82.07%. Their findings demonstrated the applicability of a deep learning architecture to signature classification, although its implementation involved relatively high computational requirements and model complexity.

In image forensics, Fahmi (2022) used Error Level Analysis (ELA) to detect digital image manipulation and reported an accuracy of approximately 87–91%. ELA identifies differences in compression-error levels that may occur when parts of an image have different compression histories. Nevertheless, ELA primarily represents compression-related characteristics, and its effectiveness can be affected by image format, recompression history, and the type of manipulation. It may therefore be insufficient as the sole source of information for detecting all forms of signature forgery. Li *et al.* (2023) examined the use of Support Vector Machine (SVM) for image classification and reported performance ranging from approximately 90% to 95% on high-dimensional image datasets. SVM is particularly suitable for classification tasks involving a limited number of samples and can construct decision boundaries in high-dimensional feature spaces. Its performance, however, depends strongly on feature quality, feature scaling, kernel selection, and parameter configuration. These studies demonstrate that deep learning, image forensics, and conventional machine learning can support signature verification or image manipulation detection. However, each method represents different image characteristics and has limitations when used independently. Compression analysis does not necessarily capture texture variations, while texture analysis alone may not identify differences in compression history or perceptual similarity. Combining complementary feature types may therefore provide a broader representation of signature images for classification. Because the previous studies used different datasets, class distributions, and evaluation protocols, their reported accuracy values should be interpreted within their respective experimental settings rather than compared directly.

2.2 Comparison with Previous Studies

Table 1 summarizes the methods, datasets, reported results, advantages, and limitations of the studies reviewed. The comparison is intended to describe the position of the present study rather than establish a direct performance ranking because the studies were conducted under different experimental conditions.

Table 1. Comparison of Previous Studies and the Present Study

Study	Method	Dataset	Reported result	Main advantage	Limitation
Huda (2024)	CNN	Digital signature dataset	86–90% accuracy	Automatically learns visual image features	Requires sufficient training data and computational resources
Fahmi (2022)	ELA	Digital image dataset	87–91% accuracy	Detects inconsistencies in image compression	Primarily limited to compression-related characteristics
Li *et al.* (2023)	SVM	Image dataset	90–95% accuracy	Suitable for high-dimensional feature spaces	Performance depends on feature

					quality and model configuration
Saputra and Nurhaida (2024)	EfficientNetV2M	Digital signature dataset	82.07% accuracy	Learns complex visual representations	Involves relatively high computational requirements
Present study	ELA, pHash, GLCM, SVM, and DFRWS	720 signature images	80.56% accuracy	Combines compression, perceptual, and texture features within a structured forensic process	Uses a limited dataset and has not been evaluated across multiple datasets

As shown in Table 1, previous studies applied CNN, EfficientNetV2M, ELA, or SVM to different image-analysis tasks. Deep learning methods can learn visual representations automatically but may require more data and computational resources. ELA offers interpretable information about compression inconsistencies, although its scope is mainly related to the compression history of an image. SVM is suitable for high-dimensional classification but relies on the quality of the features supplied to the model. The present study differs by combining ELA, pHash, and GLCM features before classification with SVM. This combination represents compression characteristics, perceptual image information, and texture patterns. The reported accuracy of 80.56% reflects performance on the dataset and testing configuration used in this study. It should not be interpreted as directly higher or lower than the results of previous studies because their datasets and evaluation protocols were not identical.

2.3 Positioning of the Present Study

The present study is positioned at the intersection of image forensics, machine learning, and digital forensic process management. Rather than relying on a single feature-extraction method, it combines ELA, pHash, and GLCM to represent different characteristics of signature images. ELA examines compression inconsistencies that may be associated with image alteration, pHash represents perceptual image characteristics, and GLCM describes texture patterns through statistical relationships between grayscale pixel values. These features are combined as inputs to an SVM classifier for distinguishing genuine and forged signature images. The study also applies the DFRWS framework to organize the forensic process into identification, collection, examination, analysis, and reporting stages. This implementation connects the image-classification process with a structured forensic workflow instead of treating classification as an isolated computational task. The methods are implemented in a web-based system that supports dataset management, preprocessing, feature extraction, model training, classification, visualization, and report generation. The contribution of this study therefore lies in the combined use of several image-feature categories and their implementation within a DFRWS-based analysis process. The study does not assume that combining these methods necessarily produces better performance than every individual method. Instead, it evaluates the classification performance obtained from the selected feature combination and examines its practical implementation in a web-based forensic analysis system.

2.4 Technologies, Frameworks, and Algorithms

Digital forensics concerns the identification, collection, preservation, examination, analysis, and reporting of digital evidence. Its purpose is to ensure that digital evidence is handled through procedures that maintain its integrity and allow the analytical findings to be scientifically explained. In the present study, the objects of examination are digital images containing handwritten signatures. The forensic process includes organizing the image dataset, examining image characteristics, classifying the images, recording the results, and generating investigation reports. Digital forensic analysis requires a clear distinction between evidence-integrity verification and perceptual image comparison. A cryptographic hash may be used to verify whether a file has changed, whereas a perceptual hash is intended to identify visual similarity between images. These two forms of hashing serve different purposes and should not be treated as interchangeable.

Error Level Analysis is an image-forensic technique used to examine variations in compression-error levels within an image. The process generally involves recompressing an image at a specified quality level and comparing the recompressed image with the original. Regions with different compression histories may produce different error patterns and appear more prominent in the resulting ELA image. ELA can assist in identifying potential inconsistencies associated with image editing or recompression. Its interpretation, however, depends on factors such as file format, image quality, previous compression, and editing history. A prominent ELA region does not independently prove that manipulation has occurred. In this study, ELA is treated as a source of image characteristics used by the classification model rather than as conclusive evidence of forgery.

Perceptual hashing, or pHash, generates a compact representation based on the visual characteristics of an image. Unlike cryptographic hashing, in which a minor file-level change produces a substantially different hash, perceptual hashing is designed so that visually similar images can produce similar hash representations. Similarity between two perceptual hashes can be assessed using an appropriate comparison measure, such as Hamming distance. The pHash process commonly involves image resizing, grayscale conversion, application of the Discrete Cosine Transform (DCT), selection of low-frequency components, and binary hash generation. This method can retain similarity information following minor operations such as resizing or recompression. Its effectiveness nevertheless depends on the type and extent of the image modification. In this study, pHash provides perceptual information that complements the compression features obtained through ELA and the texture features extracted using GLCM.

The Gray Level Co-occurrence Matrix is a statistical texture-analysis method that describes the spatial relationships between grayscale pixel values. A GLCM records how frequently particular pixel-intensity pairs occur at a specified distance and direction. Statistical measures derived from the matrix can represent regularity, variation, and local texture patterns within a signature image. The GLCM features used in this study are:

1) Contrast, which represents the degree of local variation between neighboring grayscale values;
2) Correlation, which measures the statistical relationship between neighboring pixel values;
3) Energy, which represents the uniformity or repeated patterns within the texture; and
4) Homogeneity, which indicates the proximity of the distribution to the main diagonal of the GLCM.

These features provide texture information that cannot be obtained solely from compression analysis or perceptual hashing. Their values may be influenced by image resolution, grayscale quantization, pixel distance, and the orientation used to construct the matrix.

Support Vector Machine is a supervised machine learning algorithm used for classification and regression. In binary classification, SVM identifies a decision boundary that maximizes the margin between samples from two classes. When the data cannot be separated linearly, a kernel function can map the features into a higher-dimensional space in which a more suitable decision boundary can be constructed. In this study, SVM is used to classify signature images into genuine and forged classes based on the combined ELA, pHash, and GLCM features. SVM is suitable for this task because it can process high-dimensional feature vectors and can be applied to datasets of limited size. Its performance depends on feature scaling, kernel selection, parameter configuration, class distribution, and the quality of the extracted features.

The DFRWS framework provides a structured approach to digital forensic investigation. In this study, the framework is applied through five operational stages:

1) Identification, which determines the digital evidence and the scope of the examination;
2) Collection, which involves acquiring and organizing the signature-image dataset;
3) Examination, which includes image preprocessing and the preparation of data for feature extraction;
4) Analysis, which applies ELA, pHash, GLCM, and SVM to examine and classify the images; and
5) Reporting, which documents the procedures, classification results, performance metrics, and relevant visual outputs.

The framework is used to organize the analytical workflow and document the procedures applied to the digital evidence. Its implementation supports process traceability by connecting dataset acquisition, image examination, classification, evaluation, and report generation within the developed system.

3. Methodology

This study developed a web-based system for the digital forensic analysis of signature images by combining image-processing techniques, machine learning, and the Digital Forensic Research Workshop (DFRWS) framework. The methodology covered system architecture, dataset preparation, image preprocessing, feature extraction, SVM-based classification, performance evaluation, and forensic report generation. The system was designed to classify signature images into genuine and forged categories and to document the analysis process in a structured digital forensic workflow.

3.1 System Architecture and DFRWS Implementation

The system was developed as a web-based platform using CodeIgniter 3 with a Model-View-Controller (MVC) architecture. Image-processing and machine-learning operations were performed using Python, while MySQL was used to store image information, extracted features, prediction results, and analysis records. The web interface managed user interaction and data input, whereas the Python-based analysis engine handled preprocessing, feature extraction, model training, and classification. The analysis workflow began with the

acquisition and organization of the signature-image dataset. Each image was preprocessed and then analyzed using Error Level Analysis (ELA), perceptual hashing (pHash), and the Gray Level Co-occurrence Matrix (GLCM). The extracted features were used as inputs to a Support Vector Machine (SVM) classifier. During the testing stage, the trained model classified each image as genuine or forged. The system then displayed the prediction results, performance visualizations, and digital forensic investigation reports.

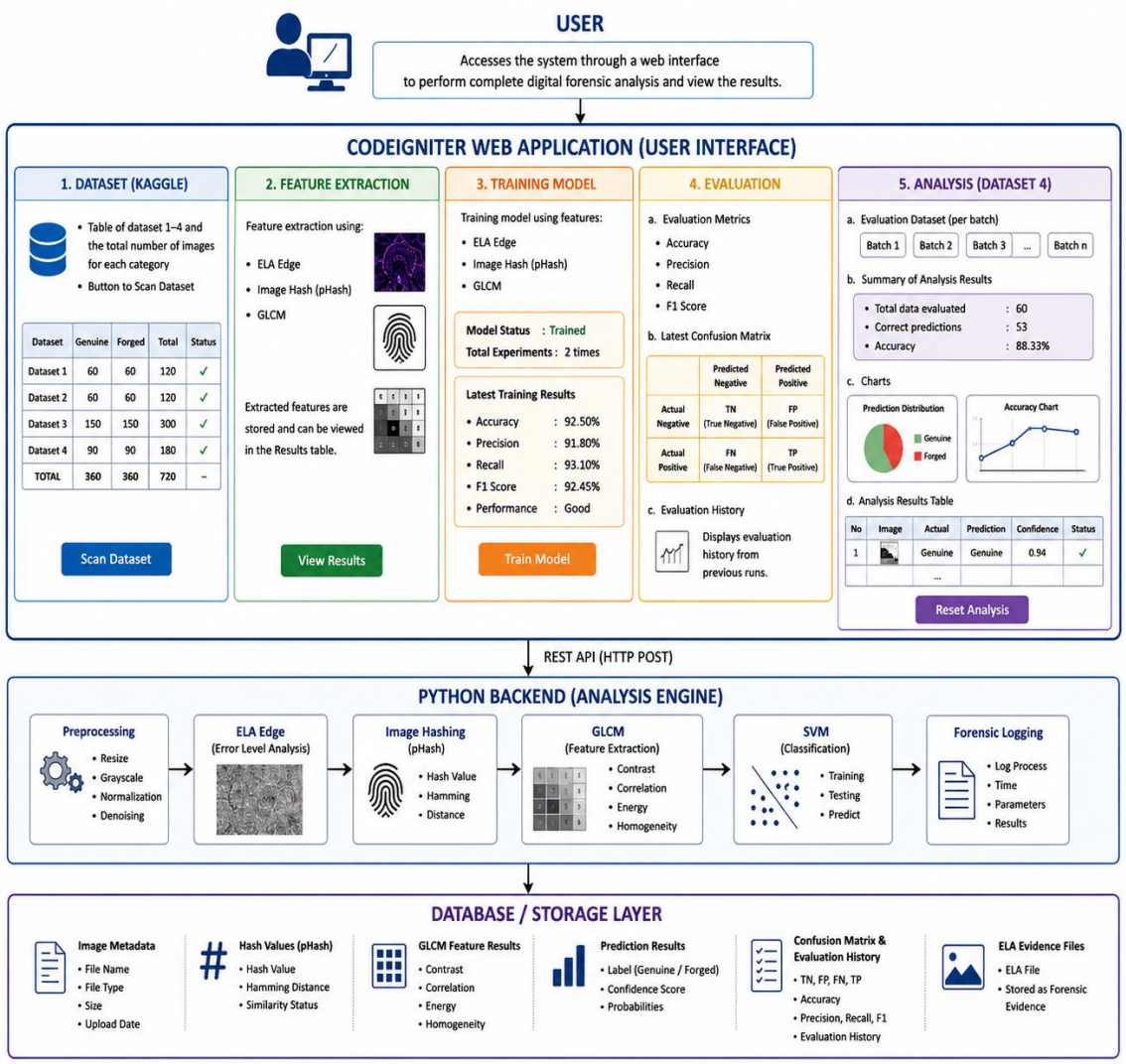

Figure 1. Proposed System Architecture

As illustrated in Figure 1, the system consisted of three main components: a CodeIgniter-based web interface, a Python-based analysis engine, and a database layer. These components supported the complete process from dataset acquisition and preprocessing to feature extraction, model training, testing, and report generation. The DFRWS framework was applied to organize the investigation process into identification, collection, examination, analysis, and reporting stages. During identification, the investigation objects were defined as handwritten signature images obtained from a public dataset. The collection stage involved acquiring and organizing the Starter Handwritten Signatures Dataset available on Kaggle. During examination, the images were resized, converted to grayscale, normalized, and prepared for feature extraction. The analysis stage applied ELA, pHash, and GLCM to obtain image features, followed by SVM-based classification. The reporting stage documented the classification results, confusion matrix, prediction distribution, performance metrics, and relevant visual outputs in PDF-based digital forensic investigation reports.

3.2 Dataset Preparation and Feature Extraction

The dataset used in this study was obtained from the Starter Handwritten Signatures Dataset available on Kaggle. It consisted of 720 signature images divided into two classes: genuine and forged signatures. Dataset 1–3 were used for model training, while Dataset 4 was reserved for testing. This partitioning allowed the model to be evaluated using images that were not included in the training process. Before feature extraction, all images underwent preprocessing to standardize their characteristics. The process included image acquisition, resizing, grayscale conversion, pixel normalization, and storage of the processed images.

These steps reduced differences in image dimensions and pixel-value ranges, allowing the feature-extraction methods to process the images in a consistent format. Three methods were then applied to extract different image characteristics. ELA was used to examine inconsistencies in image compression by recompressing an image and comparing the recompressed version with the original. Regions with different compression histories may exhibit different error patterns. In this study, ELA-derived information was used as one component of the feature set rather than as independent proof that an image had been manipulated.

Perceptual hashing was used to produce a compact representation of the visual characteristics of each image. The process involved resizing the image, applying the Discrete Cosine Transform (DCT), selecting low-frequency information, and generating the perceptual hash. Unlike a cryptographic hash, pHash represents perceptual similarity and is therefore intended for image comparison rather than file-integrity verification. GLCM was used to extract texture characteristics based on the spatial relationships between grayscale pixel values. The features derived from GLCM were contrast, correlation, energy, and homogeneity. Contrast represented local intensity variation, correlation described the statistical relationship between neighboring pixels, energy measured texture uniformity, and homogeneity represented the closeness of the GLCM distribution to its main diagonal. The ELA, pHash, and GLCM outputs were subsequently combined as inputs to the classification model.

3.3 Classification and Experimental Setup

Classification was performed using SVM, a supervised machine learning algorithm that identifies a decision boundary between classes. In this study, the model was trained to distinguish genuine signature images from forged signature images based on the combined features extracted through ELA, pHash, and GLCM. The classification process involved preparing the extracted feature vectors, training the SVM model using Dataset 1–3, storing the trained model, predicting the classes of images in Dataset 4, and generating the classification results. The experimental configuration used Dataset 1–3 as the training set and Dataset 4 as the previously unseen test set. All images followed the same preprocessing and feature-extraction procedures. The resulting feature vectors were supplied to the SVM classifier, and its performance was evaluated using a confusion matrix, accuracy, precision, recall, and F1-score. The confusion matrix recorded correct and incorrect predictions for the genuine and forged classes, while the remaining metrics provided complementary measures of classification performance. Accuracy measured the proportion of correctly classified images among all test images. Precision measured the proportion of positive predictions that were classified correctly, while recall represented the proportion of actual positive samples detected by the model. The F1-score was used to assess the balance between precision and recall. Together, these metrics described the model's ability to distinguish genuine signature images from forged ones.

3.4 Tools and Research Workflow

The software and technologies used to develop and evaluate the system are listed in Table 2.

Table 2. Tools and Technologies

Component	Technology
Web framework	CodeIgniter 3
Programming languages	PHP and Python
Image processing	OpenCV
Machine learning	scikit-learn
Database	MySQL
Operating system	Windows 11
Integrated development environment	Visual Studio Code
Dataset source	Kaggle

PHP and CodeIgniter 3 were used to develop the web interface and manage interactions among system components. Python was used for image processing and machine learning because it supports the required analytical libraries. OpenCV handled image-processing operations, while scikit-learn supported SVM model development and performance evaluation. MySQL stored the relevant dataset records, extracted features, and prediction results. The complete research workflow is shown in Figure 2. It began with dataset collection and partitioning, followed by image preprocessing and feature extraction using ELA, pHash, and GLCM. The extracted features from Dataset 1–3 were used to train the SVM model, while Dataset 4 was used to evaluate its performance. The final stages included classification analysis, performance evaluation, visualization of the results, and generation of a digital forensic investigation report.

Figure 2. Research Workflow

Figure 2 summarizes the sequence of activities performed in this study, from dataset acquisition to forensic report generation. The workflow connects the technical classification process with the DFRWS stages so that image examination, analysis, and reporting are conducted through a consistent procedure.

4. Result and Discussion

4.1 Results

The developed system was evaluated based on its functional implementation and performance in classifying genuine and forged signature images. Dataset 1–3 were used for model training, while Dataset 4 was reserved as previously unseen test data. The evaluation covered the system interfaces, training and testing results, confusion matrix, accuracy, precision, recall, and F1-score. The digital forensic analysis system was implemented as a web-based platform using CodeIgniter connected to a Python-based analysis engine. Its main functions included dataset management, image preprocessing, automatic feature extraction, SVM model training, signature-image classification, result visualization, and DFRWS-based forensic report generation. The system also displayed prediction distributions, confusion matrices, and extracted feature information to support the interpretation and documentation of the analysis.

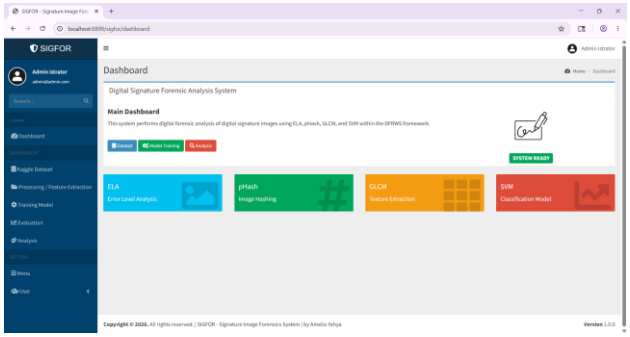

Figure 3. Main Interface of the Developed System

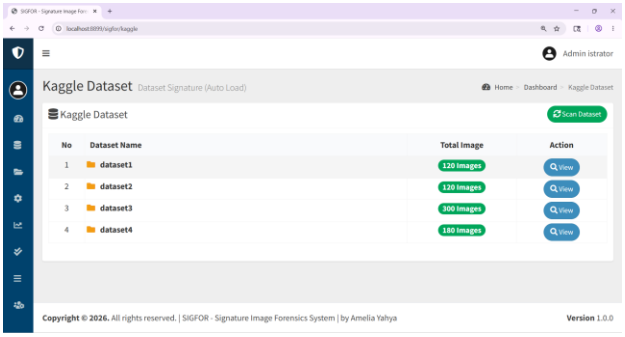

Figure 4. Dataset Management Interface

Figure 3 shows the main dashboard, which provides access to dataset processing, model training, and forensic analysis functions. The interface organizes the main functions within a single platform. Communication between the CodeIgniter web application and the Python-based image-processing modules enables the exchange of data and analysis results between system components. Figure 4 shows the dataset management interface used to organize genuine and forged signature images from the Starter Handwritten Signatures Dataset. Through this interface, images were grouped according to their classes and assigned to the training or testing set before preprocessing and feature extraction.

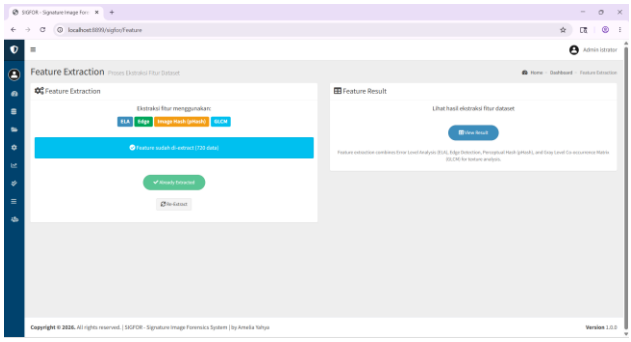

Figure 5. Feature Extraction Results Interface

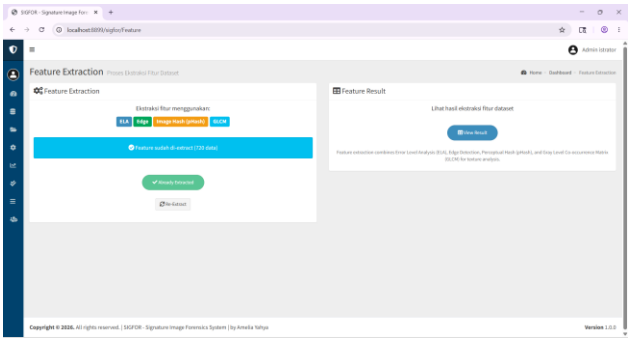

Figure 6. Model Training Interface

Figure 5 displays the feature extraction results generated using Error Level Analysis (ELA), perceptual hashing (pHash), and the Gray Level Co-occurrence Matrix (GLCM). ELA provided information related to image-compression inconsistencies, pHash represented perceptual image characteristics, and GLCM extracted texture features consisting of contrast, correlation, energy, and homogeneity. The resulting features were combined and used as inputs to the classification model. Figure 6 presents the SVM model training interface. The interface displays the training process based on features extracted from Dataset 1–3 and records the resulting performance metrics. After training, the model was stored and subsequently used to classify images from Dataset 4.

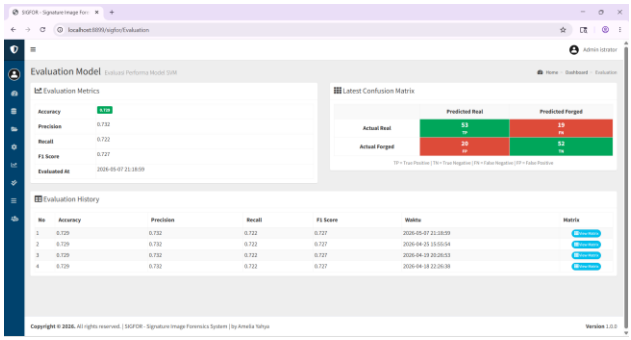

Figure 7. Confusion Matrix of the SVM Model

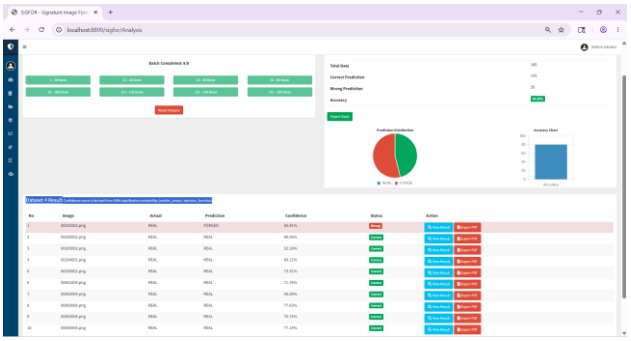

Figure 8. Dataset 4 Testing Results and Prediction Performance

Figure 7 shows the confusion matrix generated during testing. It summarizes correct and incorrect classifications through True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. These values provide the basis for calculating accuracy, precision, recall, and F1-score. Figure 8 presents the classification results for Dataset 4, which was not included during model training. The interface displays the predicted classes, prediction distribution, accuracy, and SVM model scores for the tested images. These outputs allow users to review individual predictions and examine the general behavior of the model. The displayed scores should be interpreted according to the output method implemented in the SVM model, either probability estimates generated through predict_proba() or decision scores generated through decision_function(). Before classification, all images underwent the same preprocessing and feature-extraction procedures. The SVM model was trained using Dataset 1–3, with the features obtained through ELA, pHash, and GLCM serving as model inputs. The performance recorded during model development is reported in Table 3.

Table 3. Model Training Performance

Metric	Value
Accuracy	72.90%
Precision	73.20%
Recall	72.20%
F1-score	72.70%

As shown in Table 3, the model recorded an accuracy of 72.90% during the training stage. Its precision, recall, and F1-score were 73.20%, 72.20%, and 72.70%, respectively. The relatively small differences among these values indicate that the reported precision and recall were reasonably balanced during model development. The trained model was then evaluated using Dataset 4. This test set consisted of 180 images that were not used during model training. Table 4 shows that the model correctly classified 145 images and incorrectly classified 35 images.

Table 4. Dataset 4 Testing Results

Parameter	Value
Total test images	180
Correct predictions	145
Incorrect predictions	35
Accuracy	80.56%

The testing accuracy was calculated as follows:

$$\text{Accuracy} = \frac{145}{180} \times 100\% = 80.56\%$$

The result indicates that the model correctly classified approximately four-fifths of the images in Dataset 4. However, 35 images were misclassified, showing that the extracted features did not completely separate genuine and forged samples. The complete classification metrics obtained from Dataset 4 are reported in Table 5. Accuracy represents the proportion of correctly classified images among all test images. Precision measures the proportion of positive predictions that were correct, whereas recall measures the proportion of actual positive samples detected by the model. The F1-score represents the balance between precision and recall.

Table 5. Classification Performance on Dataset 4

Metric	Value
Accuracy	80.56%
Precision	81.00%
Recall	80.00%
F1-score	80.50%

The model achieved a precision of 81.00%, a recall of 80.00%, and an F1-score of 80.50%. The small differences among these metrics indicate that precision and recall were relatively balanced under the evaluation procedure used. These aggregate values should nevertheless be interpreted together with the confusion matrix and class distribution because they do not independently show how classification errors were distributed between genuine and forged images.

4.2 Discussion

The testing accuracy of 80.56% indicates that the combination of ELA, pHash, GLCM, and SVM distinguished most genuine and forged signature images in Dataset 4. Each feature-extraction method represented a different aspect of the images. ELA captured information associated with compression errors, pHash represented perceptual image characteristics, and GLCM described texture through contrast, correlation, energy, and homogeneity. The combination supplied the SVM classifier with compression-related, perceptual, and texture information within a single feature set. The result should not be interpreted as evidence that the combined features performed better than each feature-extraction method used separately. The study did not conduct baseline or ablation experiments that compared ELA, pHash, and GLCM individually under identical testing conditions. Therefore, the findings show the performance obtained from the selected combination but do not establish the contribution of each feature category or demonstrate a measurable improvement over single-feature approaches. Such comparisons would require separate models trained and tested using the same dataset partition and evaluation protocol. The result also cannot be directly ranked against the accuracy values reported in previous studies. The studies reviewed in Section 2 used different datasets, image characteristics, class distributions, preprocessing procedures, model configurations, and

evaluation protocols. Their results provide contextual information about the application of CNN, EfficientNetV2M, ELA, and SVM but do not constitute direct benchmarks for the present study. A valid performance comparison would require all methods to be evaluated using the same dataset and testing procedure.

The model recorded a training-stage accuracy of 72.90% and a testing accuracy of 80.56%. A higher testing score can occur when the test samples are less difficult to classify or when their class distribution and image characteristics differ from those of the training data. This difference does not independently demonstrate stronger generalization. The interpretation also depends on whether the value reported in Table 3 was calculated directly from the training set, obtained from a validation set, or averaged through cross-validation. The specific evaluation procedure used to generate the training-stage metrics should therefore be stated clearly. The 35 incorrect predictions indicate that some genuine and forged signature images remained difficult to distinguish. Similar visual and texture patterns may cause their extracted features to overlap. Image quality, resizing, grayscale conversion, compression history, and variations in signature structure may also affect the values generated by ELA, pHash, and GLCM. False-positive and false-negative cases should be examined separately to determine whether the model was more likely to classify genuine signatures as forged or forged signatures as genuine. This distinction is important because the two types of errors may have different consequences in document verification.

The web-based implementation shows that dataset management, image preprocessing, feature extraction, model training, classification, visualization, and report generation can be conducted within one system. Organizing these functions according to the DFRWS stages also supports a structured analysis process from evidence identification and collection to examination, analysis, and reporting. Nevertheless, the system should be regarded as an analytical support tool rather than a replacement for forensic examination or expert judgment. Its predictions remain dependent on the characteristics of the dataset, the feature-extraction configuration, the SVM parameters, and the quality of the input images. The study was limited to 720 signature images obtained from one dataset and evaluated through one predefined training–testing partition. The reported results therefore represent performance under the conditions used in this experiment and should not be generalized to all signature images or document-forgery scenarios. Evaluation using larger and more diverse datasets, repeated experiments, writer-independent partitioning, and comparisons among individual and combined feature sets would provide stronger evidence regarding model robustness. Alternative classifiers, including deep learning models, may also be examined, but they should be assessed using the same dataset and evaluation protocol to ensure a valid comparison.

5. Conclusion and Recommendations

This study developed a web-based forensic analysis system for classifying genuine and forged signature images. The system combined Error Level Analysis (ELA), perceptual hashing (pHash), and the Gray Level Co-occurrence Matrix (GLCM) with a Support Vector Machine (SVM) classifier. CodeIgniter was used to develop the web application, while Python supported the image-processing and machine-learning operations. The analysis process was organized according to the Digital Forensic Research Workshop (DFRWS) stages, covering identification, collection, examination, analysis, and reporting. The model achieved an accuracy of 80.56% on Dataset 4, correctly classifying 145 of the 180 test images. Its precision, recall, and F1-score were 81.00%, 80.00%, and 80.50%, respectively. These results show that the selected combination of ELA, pHash, GLCM, and SVM distinguished most genuine and forged signature images under the experimental conditions used. The developed system also supported dataset management, image preprocessing, feature extraction, model training, classification, result visualization, and digital forensic report generation.

The contribution of this study lies in combining compression-related, perceptual, and texture features within a structured forensic analysis process. ELA represented image-compression characteristics, pHash provided perceptual image information, and GLCM described texture through contrast, correlation, energy, and homogeneity. However, because the study did not include baseline or ablation experiments, the results do not establish that the combined feature set performed better than each method used separately. The reported performance should therefore be interpreted as the result of the complete configuration evaluated in this study. Several limitations must also be considered. The evaluation used 720 signature images from a single dataset and one predefined training–testing partition, which limits the generalizability of the findings. Thirty-five test images were incorrectly classified, particularly among samples with similar visual characteristics or lower image quality. Model performance may also have been affected by image resolution, compression history, preprocessing procedures, feature configuration, and similarities between genuine and forged signatures. The system should consequently be treated as an analytical support tool rather than a replacement for expert forensic judgment. Future studies should evaluate the method using larger and more diverse datasets, repeated experiments, and writer-independent data partitioning. Separate tests using ELA, pHash,

and GLCM would help determine the contribution of each feature category, while comparisons using identical datasets and evaluation protocols would allow a more reliable assessment of alternative methods. CNN, ResNet, and EfficientNet may also be examined as alternative classifiers. Further development should include detailed error analysis, systematic parameter optimization, validation across different image formats and compression conditions, and stronger mechanisms for preserving and documenting the integrity of digital evidence.

References

Akinbi, A., & Ojie, D. (2021). Digital forensic framework for investigation process. *Forensic Science International: Digital Investigation, 38*, Article 301221. https://doi.org/10.1016/j.fsidi.2021.301221

Alzahrani, A., & Alotaibi, B. (2021). Digital signature security in e-government: A review. *IEEE Access, 9*, 35645–35655. https://doi.org/10.1109/ACCESS.2021.3052376

Barkur, R., & Thomas, P. (2024). Image forgery detection based on ELA and deep learning. In *Lecture notes in electrical engineering*. Springer. https://doi.org/10.1007/978-981-95-5835-3_17

Fatihia, W. M., Fariza, A., & Karlita, T. (2023). CNN with batch normalization adjustment for offline handwritten signature verification. *International Journal of Intelligent Engineering and Systems, 16*(2), 123–134. https://doi.org/10.22266/ijies2023.0430.12

Gorle, S., & Guttavelli, R. (2025). Enhanced image tampering detection using error level analysis and CNN. *Engineering, Technology & Applied Science Research, 15*(1), 9593–9600. https://doi.org/10.48084/etasr.9593

Hafemann, L. G., Sabourin, R., & Oliveira, L. S. (2021). Offline handwritten signature verification: Literature review. *Pattern Recognition, 119*, Article 107972. https://doi.org/10.1016/j.patcog.2021.107972

Herman, R., Santoso, B., & Pratama, A. (2022). Digital forensics investigation framework: A systematic review based on DFRWS. *Procedia Computer Science, 197*, 202–209. https://doi.org/10.1016/j.procs.2022.12.133

Kaggle. (2022). *Handwritten signature dataset* [Data set]. https://www.kaggle.com/datasets/divyanshrai/handwritten-signatures

Kaur, G., & Jindal, R. (2021). Image forgery detection using error level analysis. *Materials Today: Proceedings, 51*, 1881–1886. https://doi.org/10.1016/j.matpr.2021.03.467

Kaur, N., & Goel, S. (2022). Digital image forensics: A comprehensive survey. *Journal of Information Security and Applications, 64*, Article 103060. https://doi.org/10.1016/j.jisa.2022.103060

Li, J., Wang, H., & Chen, X. (2023). Image classification using support vector machine: A review. *Pattern Recognition Letters, 167*, 12–22. https://doi.org/10.1016/j.patrec.2023.01.005

Sari, W. P., & Fahmi, H. (2021). The effect of error level analysis on image forgery detection using deep learning. *Journal of Physics: Conference Series, 1898*, Article 012012. https://doi.org/10.1088/1742-6596/1898/1/012012

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *Journal of King Saud University—Computer and Information Sciences, 34*(10), 8888–8903. https://doi.org/10.1016/j.jksuci.2021.05.017

Sethy, P. K., & Behera, S. K. (2021). Texture analysis using GLCM for image classification. *Procedia Computer Science, 167*, 236–245. https://doi.org/10.1016/j.procs.2021.01.012

Tang, Z. (2022). Perceptual hashing for image authentication: A survey. *IEEE Transactions on Information Forensics and Security, 17*, 1234–1248. https://doi.org/10.1109/TIFS.2022.3145678

Tolosana, R., Vera-Rodriguez, R., & Fierrez, J. (2021). Deep learning for signature verification: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(9), 5666–5687. https://doi.org/10.1109/TPAMI.2021.3051230

Zhang, W., & Zhao, Y. (2022). Image forgery detection based on error level analysis and deep features. *IEEE Access, 10*, 45872–45883. https://doi.org/10.1109/ACCESS.2022.3167890.