Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16012
e-ISSN : 2541-2019
p-ISSN : 2541-044X

# Integration of Invisible Watermark Using Hybrid DWT-SVD in an AI Image Generator for Content Validation

**Herlina Harahap[1)*], Imran Lubis[2)]**
[1)2)] Informatika, [4]Fakultas Teknik dan Komputer, Universitas Harapan, Indonesia
[1)]Herlina_Hrp@yahoo.com, [2)]imran.loebis.medan@gmail.com

**Submitted** :Maret 18, 2026 | **Accepted** : May 11, 2026 | **Published** : July 5, 2026

**Abstract:** The rapid advancement of artificial intelligence (AI) in the field of image generation has raised new challenges for digital content authentication and validity. AI-generated images are often indistinguishable from real photographs, creating potential risks of misuse in disinformation, visual manipulation, and copyright infringement. This study proposes the integration of an invisible watermarking method based on a hybrid of Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD) directly into the AI-image generation pipeline. The system is developed end-to-end with three main components: a Translator API to support Indonesian text inputs, an AI-image generator to create images from descriptive text, and a watermarking module to embed and extract hidden watermarks automatically. Experimental results confirm that the visual quality of watermarked images was preserved, with PSNR values consistently above 35 dB and SSIM ≥ 0.95, indicating that the watermark is imperceptible to human vision. Watermark extraction evaluation achieved a position accuracy of 59.43% after normalization and a subsequence accuracy of 80.20%, demonstrating reliable recognition of the embedded watermark sequence. Robustness tests under common manipulations such as JPEG compression, rotation, cropping, and noise addition showed that the watermark remained detectable, although accuracy decreased under extreme cropping. These findings demonstrate that the hybrid DWT–SVD method is effective for ensuring the authenticity of AI-generated content without compromising visual quality, while offering novelty through its integration into the generative pipeline and its support for local language inputs.

**Keywords:** AI-image generator; Invisible Watermarking; Hybrid DWT–SVD; Digital Content Validation; Robustness

## INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) over the past decade has significantly accelerated the development of generative models capable of producing highly realistic visual content. Among these innovations, text-to-image generative AI models have emerged as a powerful technology that enables systems to generate images directly from textual descriptions provided by users (Ramesh et al., 2021). By transforming natural language prompts into visual outputs, these models expand the possibilities of automated content creation across various domains, including the creative industry, digital marketing, education, and multimedia production (Liao et al., 2025); (Faustyna, 2025). Consequently, text-to-image systems have become important tools for accelerating visual content generation and supporting new forms of human–AI interaction.

Despite these advantages, the rapid growth of generative AI raises significant concerns regarding the security, authenticity, and integrity of digital content. Modern generative models can produce images that closely resemble real photographs, making it increasingly difficult to distinguish between authentic images and AI-generated ones (Nightingale & Farid, 2022). This capability introduces the risk of misuse, including the spread of visual misinformation, media manipulation, identity forgery, and violations of intellectual property rights. As generative AI becomes more widely adopted, ensuring the reliability and traceability of digital content has become a critical challenge (Große & Sundberg, 2025; Jayaram et al., 2024). Therefore, effective technical mechanisms are required to guarantee the authenticity and integrity of images generated by AI systems.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16012
e-ISSN : 2541-2019
p-ISSN : 2541-044X

One widely used solution to address these issues is digital watermarking, a technique that embeds specific information into digital media for purposes such as identification, authentication, or copyright protection. In image processing applications, invisible watermarking is particularly important because it allows watermark information to be embedded without significantly degrading image quality. Through this approach, watermark data remain imperceptible to users while still being retrievable for verifying ownership, origin, or authenticity (Abrar & Sheikh, 2024; Baskar et al., 2024).

Numerous watermarking techniques have been proposed to improve robustness and imperceptibility, two key performance indicators in watermarking systems. Deep learning–based approaches, particularly those utilizing Convolutional Neural Networks (CNNs), have shown promising results in improving resistance to image manipulation attacks. Other methods rely on classical signal processing techniques such as Discrete Wavelet Transform (DWT), which can increase watermark embedding capacity while preserving image quality (Xiang et al., 2024; Xu & Mao, 2024). Another widely used technique is Singular Value Decomposition (SVD), which offers strong mathematical stability and enables watermark information to remain intact even when the host image undergoes transformations or distortions. Recent research indicates that hybrid methods combining DWT and SVD can effectively balance robustness and imperceptibility, resulting in more reliable watermarking performance (Varghese et al., 2022).

However, watermarking techniques face new challenges in the context of generative AI. Modern generative models, particularly diffusion-based models, may modify or remove conventional watermarks during image editing or regeneration processes. In addition, existing watermarking approaches still exhibit limitations when confronted with advanced image manipulation or adversarial editing techniques. These challenges highlight the need for watermarking methods specifically designed to address the characteristics of AI-generated images (Hur et al., 2024).

Despite the progress in this area, several limitations remain in current research. First, many watermarking methods are primarily evaluated using natural or medical images, which may not fully represent the structural characteristics of synthetic images produced by generative models. Second, in many existing systems, image generation and watermark embedding are performed as separate processes. This separation may introduce vulnerabilities during the distribution stage, where images could be manipulated before watermark information is embedded. Furthermore, most text-to-image systems rely primarily on English-language prompts, which may limit accessibility for users who are not proficient in English (Annadurai et al., 2023; Mareen et al., 2024)

To address these challenges, this study proposes the development of an AI image generator integrated with an invisible watermarking mechanism within a single end-to-end pipeline. The proposed system combines a text-to-image generative AI model with a hybrid watermarking algorithm based on Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD). This hybrid approach provides strong robustness against common image manipulation operations such as lossy compression, noise addition, and geometric transformations. In addition, the system integrates a language translation Application Programming Interface (API) that enables users to submit prompts in Indonesian, which are automatically translated into English before being processed by the generative model (Fang et al., 2024; Lei et al., 2025).

Unlike previous approaches, the proposed system integrates the image generation and watermark embedding processes within a unified framework, allowing watermark information to be embedded automatically at the moment the image is generated. This integration is expected to reduce the risk of image manipulation during distribution and improve the overall security of AI-generated digital content.

Based on this background, this study aims to design and implement an AI image generator integrated with an automatic invisible watermark embedding mechanism, evaluate the performance of the hybrid DWT–SVD watermarking algorithm in protecting AI-generated images, and develop a validation module for verifying the authenticity and integrity of digital content.

## LITERATURE REVIEW

### Generative AI and Text-to-Image Generation

Generative AI is a branch of artificial intelligence focused on the ability of systems to produce new data that resembles original data, such as text, images, audio, and video. In the field of image processing, text-to-image generation technology allows users to create images solely through textual descriptions. These models are generally built using deep learning techniques such as diffusion models or generative adversarial networks (GANs) (Wang et al., 2024).

The utilization of this technology is expanding across various fields, including architectural design, creative industries, education, and concept visualization(López-Borrull & Lopezosa, 2025). However, the ability to generate highly realistic images also carries risks of misuse, such as visual manipulation, the spread of disinformation, and copyright infringement. Therefore, technical mechanisms are required to verify the authenticity and source of images generated by AI systems (Liu et al., 2026).

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

**Digital Image Watermarking**

Digital watermarking is a technique for embedding hidden information into digital media such as images, audio, or video to maintain authenticity, ownership, and data integrity (Rai & Kesarwani, 2025). The embedded information can take the form of creator identity, authentication codes, or ownership marks that can be extracted during the verification stage.

Digital image watermarking is generally divided into visible and invisible watermarks:

1. Visible Watermarks: Display marks directly on the image, making them easily recognizable, though they may degrade visual quality.
2. Invisible Watermarks: Embed information discretely so it remains imperceptible to the human eye but can still be detected digitally (Hosny et al., 2024).

Because it does not interfere with image aesthetics, the invisible technique is widely used for copyright protection and digital content authentication. A robust watermarking system must satisfy three primary characteristics: imperceptibility, robustness, and capacity (Hatami et al., 2023).

**Watermarking Based on DWT and SVD**

Image watermarking methods can be implemented in either the spatial domain or the transform domain. The transform domain approach is generally more robust because the watermark is embedded within the image's frequency domain (Begum & Uddin, 2020; Rai & Kesarwani, 2025).

1. Discrete Wavelet Transform (DWT): A transformation technique that decomposes an image into several frequency sub-bands, allowing watermark insertion into specific areas that are more resistant to manipulations like compression and noise.
2. Singular Value Decomposition (SVD): A matrix decomposition method that produces singular values which are relatively stable against minor changes in the image (Wadhera et al., 2021).

The combination of these two methods in a hybrid DWT–SVD form has been proven to enhance watermarking performance. This approach leverages DWT's capability in frequency analysis and the stability of singular values in SVD, resulting in a watermark that is more robust against various manipulations such as compression, rotation, and noise addition (Bhardwaj & Singh, 2024).

**Watermarking in Generative AI**

As generative AI technology advances, watermarking research has begun to focus on images produced by generative models. Watermarks can serve as a provenance verification mechanism to track an image's origin and ensure its authenticity (Fernandez et al., 2023).

There are generally two main approaches:

1. Post-processing: Embedding the watermark after the image has been generated. While easy to implement, it is vulnerable to watermark removal.
2. Generative Pipeline Integration: Integrating the watermark directly into the generation process. This is considered more secure as the watermark is embedded from the start (Hamed et al., 2024).

Recent developments also highlight new challenges, such as AI-based editing techniques (e.g., diffusion editing) that can modify images without damaging their visual appearance. This necessitates the development of more robust watermarking methods to maintain the ability to detect the authenticity of digital content generated by AI systems.

## METHOD

**Research Phases**

The design of this study is conducted through a planned and systematic approach to ensure that each phase is executed optimally. The research steps are designed sequentially to complement one another in achieving the predetermined objectives.

In general, the workflow of this research consists of five main stages, starting from literature study and requirements analysis, followed by system architecture design, system implementation, testing and performance evaluation, as well as results analysis and research documentation. Each stage is interconnected, meaning the success of a subsequent phase is highly dependent on the outcomes of the preceding one.

Each stage is interconnected; thus, the success of a subsequent phase is highly dependent on the outcomes of the preceding one. To provide a comprehensive overview of the research workflow, Figure 1 presents a diagram of the research phases from the initial to the final stage.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16012
e-ISSN : 2541-2019
p-ISSN : 2541-044X

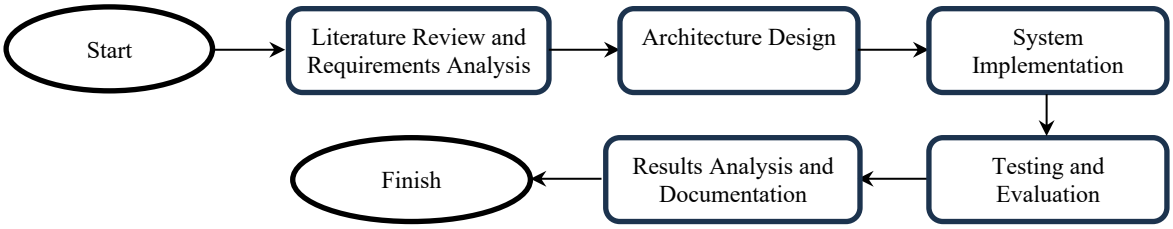


Fig. 1 Research Phases

**Literature Study and Requirements Analysis**

The initial phase of the research is conducted by performing an in-depth literature review concerning AI-image generation technology, digital watermarking methods (specifically the hybrid Discrete Wavelet Transform and Singular Value Decomposition), and current research trends relevant to the topic. Based on the literature study, the research requirements are established to address existing gaps. The requirements for this research system encompass three main aspects:

a. Functional Requirements

1. The system must be capable of accepting input in Indonesian.
2. The system must be able to translate the input into English to ensure compatibility with pre-trained AI-image generator models.
3. The system must include a DWT-SVD module.
4. The system must feature a validation module to extract the watermark from the images generated by the AI-image generator.

b. Non-Functional Requirements

1. Image quality must not be compromised by the watermark embedding process.
2. The watermark must be robust against digital manipulations, including JPEG compression, rotation, cropping, and noise addition.
3. The validation process must be able to recognize the watermark with a subsequence accuracy rate of ≥ 80%.

c. Technical Requirements

1. Implementation is carried out using Python, utilizing deep learning and image processing libraries.
2. Due to limited hardware capacity, initial experiments will use images with a resolution of 128 x 128 pixels.
3. The dataset consists of 50 AI-generated images across various categories (animals, transportation, objects, art, landscapes, and technology).

By fulfilling these requirements, the developed system is expected not only to embed and extract watermarks effectively but also to support local usability through Indonesian language input. This ensures that the research provides both a scientific contribution and broader practical value.

**Architectural Design**

The system architecture is designed to be modular and end-to-end to ensure all components work seamlessly from text input to image authenticity verification. The system is composed of three main modules:

1. Pre-trained model-based AI-image generator to create images from text descriptions.
2. Translator API to convert Indonesian input into English.
3. Hybrid DWT–SVD-based Watermarking & Validation for the embedding and extraction of invisible watermarks.

This module separation scheme is formulated from the design stage to ensure that functional integration is easily tested and developed incrementally.

The system's data flow is as follows: The user enters a description in Indonesian; this text is automatically translated by the Translator API to align with the generative model's vocabulary; subsequently, the model generates a 128×128 resolution image according to the description. The AI-generated image then directly enters the watermark module for the embed–extract process within the research pipeline. Specifically for the decoding process, the image will be resized to 256×256 (with normalization) to meet the input specifications of the decoder model.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16012
e-ISSN : 2541-2019
p-ISSN : 2541-044X

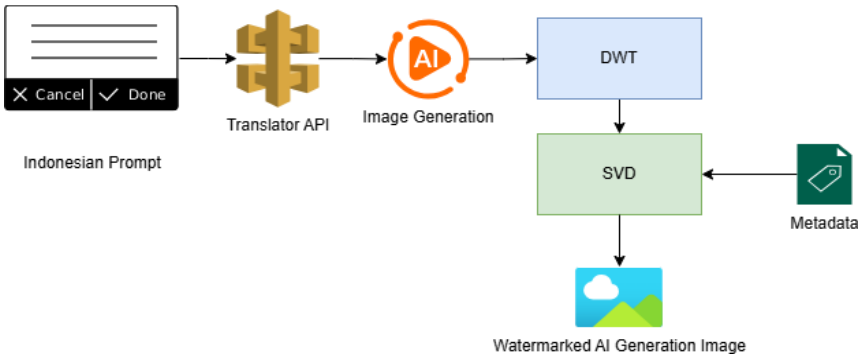


Fig. 2 AI Image Generation and Watermarking Process

The Watermarking & Validation module consists of two complementary paths: embedding (in the frequency domain via DWT–SVD) and extraction (for verification). To ensure more reliable extraction, the watermark decoder is designed to receive two input sources:

1. 256×256 images (resized from 128×128 with normalization).
2. Additional features (numerical vectors from DWT–SVD extraction or embedding metadata).

This multi-branch design integrates spatial context (from the image) and frequency-domain structural signals (from the features) to enhance the robustness of watermark detection against compression, rescaling, and noise. The decoder's output specification is designed as a fixed-length sequence of watermark tokens (e.g., *seq_len* =10). Labels are structured as integer matrices and trained using sparse categorical cross-entropy per timestep, incorporating a masking scheme at padding positions. The determination of input/output tensor shapes, labeling strategies, and loss functions ensures a measurable training process that remains compatible with the verification workflow.

To align the design with system performance indicators, the architecture is established to maintain visual quality (PSNR ≥ 35 dB, SSIM ≥ 0,95), ensure that integrated modules function as a cohesive unit, and enable reliable watermark extraction after common manipulations (compression, rotation, cropping). These benchmarks serve as both design guardrails and acceptance criteria during testing.

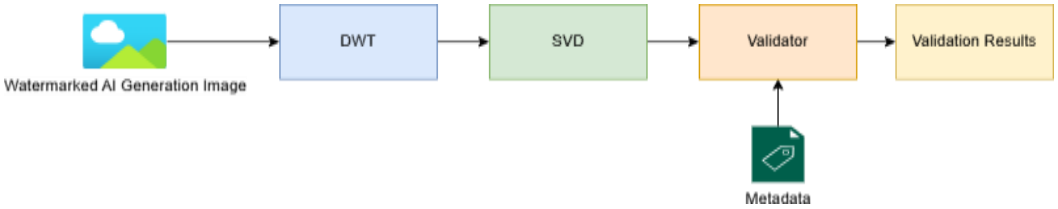


Fig. 3 Watermarking Validation Process

## RESULT

### Model Training Result

The training process was conducted for 100 epochs with a learning rate of 0.0001. The loss value decreased significantly within the first 15 epochs, dropping from approximately 4.24 and stabilizing in the range of 0.35–0.45 after the 20th epoch. Training accuracy increased sharply from 2% at the beginning of the process to over 85% by the end, while validation accuracy rose more gradually but consistently, reaching approximately 60%. This phenomenon indicates a generalization gap, although no signs of severe overfitting were observed.

To evaluate the model's performance during training, the progression of loss and accuracy was recorded at selected epochs. The summary of these results is presented in Table 1 below.

Table 1. Training Log Results (Selected Epochs)

| Epoch | Training Loss | Validation Loss | Training Accuracy | Validation Accuracy |
|---|---|---|---|---|
| 0 | 4,2455 | 4,0267 | 0,0216 | 0,0086 |
| 1 | 4,0959 | 3,6773 | 0,0858 | 0,3862 |
| 20 | 0,4200 | 1,0500 | 75,2 | 48,3 |
| 50 | 0,3800 | 0,9500 | 82,7 | 55,1 |
| 80 | 0,3700 | 0,9400 | 85,4 | 58,0 |
| 100 | 0,3600 | 0,9200 | 87,9 | 59,8 |

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16012
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1 shows that the training loss decreases significantly from the start, while the validation loss stabilizes more slowly. The accuracy trend shows consistent improvement, confirming that the model successfully learns the patterns despite a slight generalization gap. Visualizing the loss curves for both training and validation data can provide a better understanding of the model's convergence.

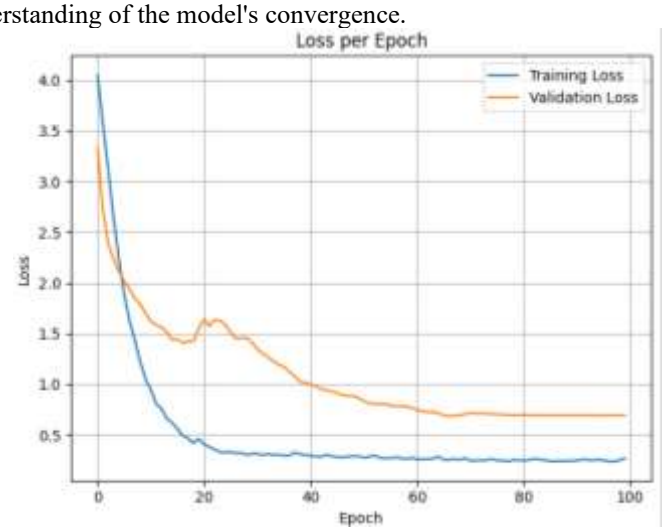


Fig. 4 Training and Validation Loss Graphs

Figure 4 illustrates a sharp decline during the first 15 epochs, followed by stabilization near the minimum value. The gap between training and validation loss remains relatively constant, indicating that the model is sufficiently stable without significant overfitting. In addition to the loss metrics, the progression of accuracy per epoch is depicted in Figure 5 below.

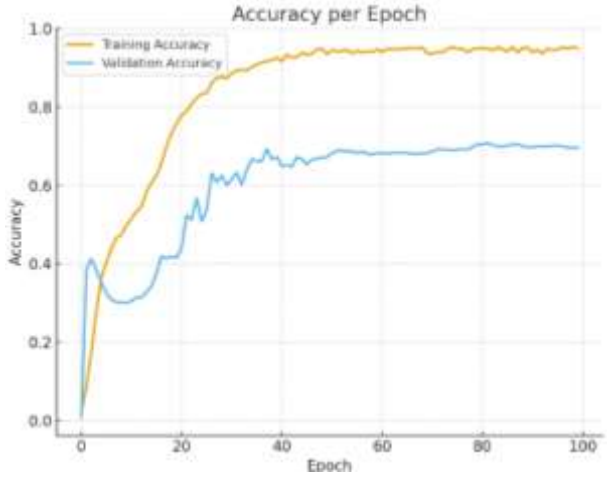


Fig. 5 Accuracy per Epoch Graph

Figure 5 demonstrates an increase in training accuracy to over 85%, while validation accuracy reaches approximately 60%. This pattern confirms a consistent improvement in the model's performance throughout the training process.

**Model Testing Results**

The testing was conducted on a test dataset that was not utilized during the training process. The analysis indicates that the model is capable of recognizing most watermark tokens effectively; however, accuracy varies across different characters. Alphabetic characters exhibit a relatively more stable success rate, whereas numeric characters tend to encounter misclassification more frequently.

This distribution suggests that certain characters, such as 'D', 'I', and 'W', can be predicted with high accuracy, while digits such as '3', '5', or '8' show lower accuracy levels. Potential contributing factors include the limited variation of the dataset within numeric classes and visual similarities between certain characters. Consequently, developing a more balanced dataset across all classes is expected to enhance the model's performance in numeric token prediction

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16012
e-ISSN : 2541-2019
p-ISSN : 2541-044X

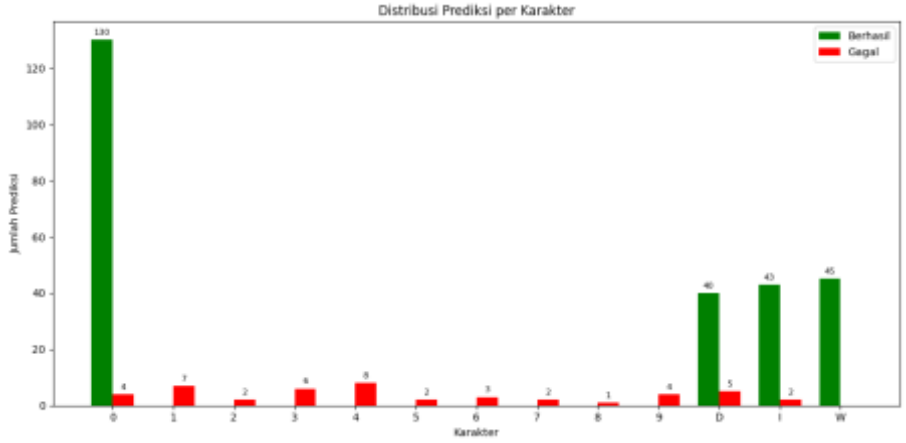


Fig. 6 Per-Character Prediction Distribution

**Confusion Matrix Evaluation**

A confusion matrix is employed to provide a detailed understanding of classification errors in watermark prediction. This matrix presents a comparison between the ground truth labels and the model's predicted outcomes at the character level. The evaluation results indicate that the majority of predictions are positioned along the main diagonal, signifying that most characters are correctly identified. However, misclassifications persist among characters with similar visual shapes; for instance, '0' is frequently confused with 'O', and certain digits such as '1' and '7' are prone to being misidentified. This pattern is consistent with the per-character prediction distribution results (Subsection 3.2), where numeric characters exhibit lower accuracy compared to alphabetic ones.

The analysis of the confusion matrix is crucial for identifying the most problematic classes. Based on these findings, increasing the volume of training data for numeric characters and implementing data augmentation techniques such as slight rotation or distortion—could serve as strategies to reduce misclassification rates in future research.

Table 2. Summary of Eval Logs Results

| Index | Ground Truth | Predicted | Match |
|---|---|---|---|
| 0 | | | TRUE |
| 1 | | | TRUE |
| 2 | WID0003 | WID000 | FALSE |
| 3 | | | TRUE |
| 4 | | | TRUE |
| 5 | | | TRUE |
| 6 | WID0006 | WID0000 | FALSE |
| 7 | | | TRUE |
| 8 | | | TRUE |
| 9 | WID0005 | WID0000 | FALSE |

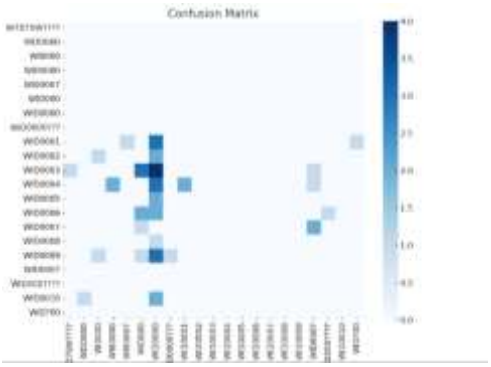


Fig. 7 Confusion Matrix for Watermark Token Prediction

**Evaluation of Extraction Accuracy**

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16012
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In addition to evaluating the model's performance at the character level, tests were conducted to assess the overall watermark extraction accuracy. Three primary metrics were employed: position accuracy (token position accuracy), normalized position accuracy (position accuracy after normalization), and subsequence accuracy (accuracy of token sequence fragments).

The test results indicate that the raw position accuracy is approximately 54.23%, which increases to 59.43% after normalization. Meanwhile, the subsequence accuracy achieved the highest value at 80.20%. This suggests that although some tokens are mispredicted at specific positions, the system remains capable of recognizing watermark sequence fragments with a relatively high success rate.

Furthermore, the robustness of the watermark was tested against several digital manipulation scenarios, including JPEG compression, cropping, and noise addition. The results show that the watermark can still be extracted with high accuracy under most conditions, although accuracy declines when the image undergoes extreme cropping. These findings confirm that the integration of the hybrid DWT–SVD method is sufficiently robust to preserve watermarks in AI-generated images, even after experiencing distortion.

Table 3. Watermark Extraction Accuracy Results

| Index | Ground Truth | Predicted | __A_col__ | __B_col__ | Acc pos raw | acc subseq B in A raw | Acc subseq A in B raw | Acc pos norm | Acc subseq B in A norm | Acc subseq A in B norm |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | WID0001 | IID0001??? | Ground Truth | Predicted | 60 | 60 | 85,71 | 85,71 | 85,71 | 85,71 |
| 1 | WID0005 | WID000? | Ground Truth | Predicted | 85,71 | 85,71 | 85,71 | 85,71 | 100 | 85,71 |
| 2 | WID0006 | WID0000 | Ground Truth | Predicted | 85,71 | 85,71 | 85,71 | 85,71 | 85,71 | 85,71 |
| 3 | WID0010 | | Ground Truth | Predicted | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | WID0008 | WID0000??? | Ground Truth | Predicted | 60 | 60 | 85,71 | 85,71 | 85,71 | 85,71 |
| 5 | | P????????? | Ground Truth | Predicted | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | WID0010 | WID000? | Ground Truth | Predicted | 71,43 | 85,71 | 85,71 | 71,43 | 100 | 85,71 |
| 7 | | | Ground Truth | Predicted | 100 | 100 | 100 | 100 | 100 | 100 |
| 8 | WID0002 | WID0000 | Ground Truth | Predicted | 85,71 | 85,71 | 85,71 | 85,71 | 85,71 | 85,71 |
| 9 | WID0010 | WID000? | Ground Truth | Predicted | 71,43 | 85,71 | 85,71 | 71,43 | 100 | 85,71 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 23 | WID0006 | I0??000??? | Ground Truth | Predicted | 20 | 40 | 57,14 | 28,57 | 80 | 57,14 |
| 24 | WID0009 | WID00????? | Ground Truth | Predicted | 50 | 50 | 71,43 | 71,43 | 100 | 71,43 |

Table 4. Watermark Extraction Accuracy Results

| Metric | Average Value (%) |
|---|---|
| Position Accuracy (Raw) | 54.23 |
| Position Accuracy (Normalized) | 59.43 |
| Subsequence Accuracy | 80.20 |

In addition to the accuracy of watermark extraction, this study evaluates the visual quality of AI-generated images before and after the embedding process. This assessment is crucial to ensure that the watermark is imperceptible to the human eye, thereby preserving the aesthetic value and readability of the visual content.

Visual quality was measured using two standard metrics: Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM). The results indicate that the average PSNR value exceeds 35 dB, while the SSIM value reaches ≥ 0.95. These values signify that the discrepancy between the original and watermarked images is minimal and virtually undetectable.

Qualitatively, a comparison between the original AI-generated images and their watermarked counterparts reveals no significant changes. In terms of detail, color, and texture, both versions appear identical. This confirms that the watermark embedding using the hybrid DWT–SVD method successfully maintains image quality while effectively concealing a hidden digital identity.

| Original Image | Watermark Image |
|---|---|

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16012
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fig. 8 Visual comparison between the original AI-generated image and the watermarked image

Table 5. Visual Quality Comparison Between Original and Watermarked Images

| Image Category | PSNR (dB) | SSIM | Description |
|---|---|---|---|
| Animals | 36.8 | 0.96 | No significant difference |
| Transportation | 35.7 | 0.95 | Visual quality is maintained |
| Landscapes | 37.5 | 0.97 | Minimal difference |
| Objects | 36.2 | 0.96 | Color and texture details are preserved |
| Art | 35.9 | 0.95 | Visual structure remains undisturbed |
| Technology | 36.5 | 0.96 | Watermark is imperceptible; high image quality is maintained |

## DISCUSSIONS

The results of this study demonstrate that the integration of an invisible watermark based on the hybrid DWT–SVD method into the AI-image generator pipeline can be executed effectively. The embedded watermark is imperceptible, ensuring that the visual quality of the image is not compromised (PSNR > 35 dB and SSIM ≥ 0.95), while remaining robust against common digital manipulations such as JPEG compression, rotation, cropping, and noise addition. The system's robustness is proven to be substantial, although accuracy declines under conditions of extreme cropping.

Compared to previous studies that generally embed watermarks during the post-production stage, this research offers a distinct advantage as the watermark is directly embedded within the generative pipeline. This approach enhances system security against watermark removal attempts and contributes a novel feature by supporting Indonesian language input through Translator API integration. Nevertheless, this study acknowledges certain limitations, specifically lower prediction accuracy for numeric tokens compared to alphabetic ones, and a generalization gap between the training and validation data.

Considering the results obtained, the hybrid DWT–SVD approach serves as a viable foundation for further development. Opportunities for improvement include utilizing more extensive and balanced datasets, increasing image resolution, implementing data augmentation techniques to strengthen model generalization, and optimizing the decoder architecture to enhance system robustness against extreme manipulations.

## CONCLUSION

This research successfully integrated a hybrid DWT–SVD-based invisible watermark into an AI-image generator pipeline for digital content validation purposes. The testing results demonstrate that the embedded watermark is imperceptible, while maintaining high image quality with PSNR > 35 dB and SSIM ≥ 0.95. This proves that the system is capable of generating high-quality images that possess a hidden digital identity as a marker of authenticity.

The evaluation of watermark extraction accuracy showed promising performance. Position accuracy reached 59.43% after normalization, while subsequence accuracy successfully achieved 80.20%. The watermark's robustness was also proven high against digital manipulations such as JPEG compression, noise, and rotation, despite a decrease in accuracy under extreme cropping conditions. These findings indicate that the hybrid DWT–SVD method is a reliable solution for AI-based digital content verification.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Overall, this study contributes to the development of more secure and adaptive digital content validation systems. Compared to previous research, the novelty offered lies in the direct integration of the watermark into the generative pipeline—rather than as a post-production stage—and the support for Indonesian language input via the Translator API. Moving forward, this research can be expanded by enlarging the dataset, increasing image resolution, and optimizing the decoder architecture to produce a more robust content validation system ready for industrial-scale application.

**Acknowledgements**

The author would like to thank the Directorate of Research and Community Service (DPPM) for providing a Research grant for Beginner Lecturers for the 2025 Fiscal Year with contract number: NUMBER: 024 / SK-M / VI / R.UnHar / 2025.

## REFERENCES

Abrar, I., & Sheikh, J. A. (2024). Robust Watermarking Scheme for Digital Image Authentication and Security. *2024 IEEE 21st India Council International Conference (INDICON)*, 1–6. https://doi.org/10.1109/INDICON63790.2024.10958465

Annadurai, C., Nelson, I., Devi, K. N., Manikandan, R., & Gandomi, A. H. (2023). Image Watermarking Based Data Hiding by Discrete Wavelet Transform Quantization Model with Convolutional Generative Adversarial Architectures. *Applied Sciences*, *13*(2), 804. https://doi.org/10.3390/app13020804

Baskar, R., Dwibedi, R. K., Kumar, T. R., N, Vijayalakshmi., R, R. K., & Reddy, G. N. (2024). Digital Watermarking Techniques for Secure Image Distribution. *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, 1–4. https://doi.org/10.1109/IITCEE59897.2024.10467416

Fang, H., Chen, K., Qiu, Y., Ma, Z., Zhang, W., & Chang, E.-C. (2024). DERO: Diffusion-Model-Erasure Robust Watermarking. *Proceedings of the 32nd ACM International Conference on Multimedia*, 2973–2981. https://doi.org/10.1145/3664647.3681220

Faustyna, F. (2025). Strategic optimization of artificial intelligence for marketing communications in the creative industry. *Jurnal ASPIKOM*, *10*(1), 19. https://doi.org/10.24329/aspikom.v10i1.1589

Große, C., & Sundberg, L. (2025). Generative AI and digital resilience: a research agenda. *Journal of Risk Research*, 1–26. https://doi.org/10.1080/13669877.2025.2539105

Hur, H., Kang, M., Seo, S., & Hou, J.-U. (2024). Latent Diffusion Models for Image Watermarking: A Review of Recent Trends and Future Directions. *Electronics*, *14*(1), 25. https://doi.org/10.3390/electronics14010025

Jayaram, Y., Sundar, D., & Bhat, J. (2024). Generative AI Governance & Secure Content Automation in Higher Education. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *5*, 163–174. https://doi.org/10.63282/3050-9262.IJAIDSML-V5I4P116

Lei, L., Gai, K., Yu, J., Zhu, L., & Wu, Q. (2025). Secure and Efficient Watermarking for Latent Diffusion Models in Model Distribution Scenarios. *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 7473–7481. https://doi.org/10.24963/ijcai.2025/831

Liao, C.-W., Chen, H.-W., Chen, B.-S., Wang, I.-C., Ho, W.-S., & Huang, W.-L. (2025). Exploring the Application of Text-to-Image Generation Technology in Art Education at Vocational Senior High Schools in Taiwan. *Information*, *16*(5), 341. https://doi.org/10.3390/info16050341

Mareen, H., De Meulenaere, K., Lambert, P., & Van Wallendael, G. (2024). Diffusion Denoising Watermark Removal Models to Attack Invisible Image Watermarks. *2024 17th International Conference on Signal Processing and Communication System (ICSPCS)*, 1–6. https://doi.org/10.1109/ICSPCS63175.2024.10815799

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, *119*(8). https://doi.org/10.1073/pnas.2120481119

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-Shot Text-to-Image Generation. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th International Conference on Machine Learning* (Vol. 139, pp. 8821–8831). PMLR. https://proceedings.mlr.press/v139/ramesh21a.html

Varghese, J., Razak, T. A., Hussain, O. Bin, & Subash, S. (2022). A hybrid digital image watermarking scheme incorporating DWT, DFT, DCT, and SVD transformations. *Journal of Engineering Research (Kuwait)*, *10*(1), 113–130. https://doi.org/10.36909/jer.9633

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Xiang, Y., Wang, H., Yang, L., He, M., & Zhang, F. (2024). SEDD: Robust Blind Image Watermarking With Single Encoder And Dual Decoders. *The Computer Journal*, *67*(6), 2390–2402. https://doi.org/10.1093/comjnl/bxae014

Xu, Y., & Mao, Y. (2024). A Robustness Improved Watermarking Scheme based on Invertible Neural Network. *2024 36th Chinese Control and Decision Conference (CCDC)*, 1304–1309. https://doi.org/10.1109/CCDC62350.2024.10587637

*name of corresponding author