JAISEN: Journal of Advanced Information Systems and Engineering
Vol. 2, 1 (September, 2026), pp 1-20
ISSN: 3163-8260 EISSN: - , DOI: -

Deepfake Face Morphing Detection Using a Hybrid CNN and SVM Model Based on The Celeb-Df Dataset

Restu Rosaria Sandra Gultom[1], Nurliana Nasution[2]

[1] Bachelor of Computer Science, Lancang Kuning University; restuuugultommmniii@gmail.com
[2] Computer Science, Lancang Kuning University; nurliananst@unilak.ac.id

ARTICLE INFO

Keywords:

Deepfake Detection;
Morphing Face Attack;
Convolutional Neural Network;
Support Vector Machine;
Digital Image Forensics.

Article history:

Received 2026-06-23
Revised 2026-07-05
Accepted 2026-08-20

ABSTRACT

The rapid advancement of artificial intelligence has led to the emergence of deepfake technology, which can generate highly realistic manipulated facial images and videos that are difficult to distinguish from authentic content. The misuse of deepfake technology poses serious threats, including digital identity fraud, misinformation, and privacy violations. This study proposes a hybrid Convolutional Neural Network (CNN) and Support Vector Machine (SVM) model for morphing face deepfake detection using the Celeb-DF dataset. The research process includes video preprocessing through duration trimming, frame extraction, face detection, face cropping, image resizing, pixel normalization, dataset labeling, and data splitting into training, validation, and testing sets. CNN is employed as a deep feature extractor to learn hierarchical facial representations, while SVM performs the final classification between real and fake facial images. To evaluate the effectiveness of the proposed approach, its performance was compared with Pure CNN and Pure SVM models using accuracy, precision, recall, and F1-score metrics. Experimental results show that the proposed Hybrid CNN–SVM model achieved the best overall performance with an accuracy of 95.31%, precision of 95.00%, recall of 94.00%, and F1-score of 95.00%, outperforming the Pure CNN model (93.60% accuracy) and slightly surpassing the Pure SVM model (95.14% accuracy). These findings demonstrate that integrating CNN for feature extraction with SVM for classification provides a reliable and effective approach for detecting morphing face deepfakes and has significant potential for applications in biometric authentication, digital identity verification, and multimedia forensics.

This is an open access article under the CC BY-NC-SA license.

Corresponding Author:
Nurliana Nasution
Computer Science, Lancang Kuning University; nurliananst@unilak.ac.id

1. INTRODUCTION

In the modern digital world, the line between reality and computer-generated content is becoming increasingly blurred due to rapid advancements in artificial intelligence and machine learning technologies. These developments have enabled the creation of highly realistic visual content, one of which is deepfake technology. Deepfakes are synthetic media generated using deep learning algorithms that manipulate facial images or videos to produce realistic outputs that are difficult to distinguish from authentic content. By learning facial patterns, expressions, and movements from existing data, deepfake systems can generate convincing facial representations that closely resemble real individuals. As a result, deepfake technology has attracted significant attention from researchers, industry practitioners, and policymakers due to its growing influence on digital communication and information dissemination *(Altuncu et al., 2024).*

Although deepfake technology offers several benefits in fields such as entertainment, education, filmmaking, and digital content creation, its misuse presents serious challenges to society. Deepfakes can be exploited to spread misinformation, manipulate public opinion, conduct identity fraud, and violate personal privacy. The increasing accessibility of deepfake generation tools has further amplified these risks, allowing individuals with limited technical expertise to create highly convincing manipulated content. Consequently, the widespread availability of deepfake technology has become a significant concern for information security, digital trust, and online authenticity *(Gambín et al., 2024).*

One particular category of deepfake manipulation that has gained attention in recent years is morphing face deepfake technology. Face morphing combines facial characteristics from two or more individuals into a single synthetic image while preserving a realistic appearance. The resulting image contains visual attributes from multiple identities, making it difficult to determine whether the image represents a genuine individual or a manipulated composition. Face morphing attacks have become a serious threat to biometric authentication systems because morphed images can potentially deceive facial recognition algorithms and identity verification mechanisms *(Hamza et al., 2022).*

The risks associated with face morphing attacks are increasing due to the availability of open-source image manipulation tools and deep learning frameworks. Today, creating a high-quality morphed facial image no longer requires advanced technical expertise. Users can generate manipulated images by simply providing source photographs, significantly lowering the barrier to conducting identity-based attacks. This situation creates substantial security concerns for applications that rely on facial biometrics, including passports, electronic identity cards, border control systems, banking services, and online authentication platforms. Therefore, the development of reliable face morphing detection systems has become increasingly important for maintaining the integrity of digital identity verification processes *(Pikoulis et al., 2021).*

Detecting morphing face deepfakes remains a challenging task because manipulated images often exhibit extremely high visual similarity to genuine facial images. In many cases, the differences between authentic and morphed images are nearly imperceptible to human observers. Traditional visual inspection methods are therefore insufficient for identifying sophisticated facial manipulations. This challenge has motivated researchers to explore artificial intelligence-based approaches capable of automatically extracting subtle patterns and anomalies that may indicate image manipulation. Effective detection systems are essential to prevent identity fraud and strengthen the security of biometric authentication technologies *(Akhtar, 2023).*

Several previous studies have investigated deepfake and face morphing detection using different methodologies. Jia, Liu, and Chen (2023) proposed a detection method based on high-frequency features and progressive enhancement learning to improve morphing attack detection. *(Al-Adwan et al., 2024)* employed a hybrid CNN–RNN architecture for deepfake media detection, while *(Al-Dulaimi & Kurnaz, 2024)* utilized a CNN-LSTM model combined with transfer learning to enhance image classification performance. Other studies have explored deep learning-based approaches using CNN architectures as end-to-end classifiers for detecting manipulated facial content. Although these methods have achieved promising results, many of them require substantial computational resources and may experience

limitations when distinguishing subtle facial manipulation patterns. Furthermore, most studies focus on deepfake detection in general rather than specifically addressing morphing face deepfakes using hybrid CNN–SVM architectures on highly realistic datasets *(Jia et al., 2023a).*

This condition reveals a research gap in the development of efficient hybrid detection models that combine deep feature extraction capabilities with robust classification mechanisms. While CNN-based approaches have demonstrated strong performance in extracting discriminative facial features, their fully connected classification layers may not always provide optimal decision boundaries for complex datasets *(Seibold et al., 2021).* Moreover, limited studies have examined the effectiveness of combining CNN feature extraction with Support Vector Machine (SVM) classification for morphing face deepfake detection using the Celeb-DF dataset. Therefore, further investigation is necessary to evaluate whether a hybrid CNN–SVM approach can improve classification performance and provide a more reliable solution for detecting morphing face deepfakes *(Borghi et al., 2021).*

Convolutional Neural Networks (CNNs) have become one of the most widely used deep learning models in image analysis due to their ability to automatically learn hierarchical visual features from raw image data. CNNs are particularly effective in extracting complex patterns related to facial textures, shapes, and structural characteristics. However, the classification stage of CNN models may face challenges when dealing with complex data distributions and highly similar image classes. To address this limitation, Support Vector Machine (SVM) can be employed as an alternative classifier. SVM is well known for its ability to determine optimal decision boundaries that maximize class separation, making it suitable for binary classification problems involving subtle differences between classes. By integrating CNN and SVM, the proposed hybrid model aims to combine the strengths of deep feature extraction and robust classification to improve morphing face deepfake detection performance.

This study utilizes the Celeb-DF dataset because it is recognized as one of the most realistic and challenging datasets for deepfake detection research. The dataset contains high-quality facial images and videos with diverse facial expressions, lighting conditions, and visual characteristics, making it suitable for evaluating the robustness of deepfake detection models. The complexity of Celeb-DF provides a realistic testing environment for assessing the effectiveness of the proposed CNN–SVM approach in detecting facial manipulations *(Li et al., 2020).*

The novelty of this research lies in the implementation of a hybrid CNN–SVM framework specifically designed for morphing face deepfake detection using the Celeb-DF dataset. Unlike previous studies that primarily relied on CNN, CNN-LSTM, or CNN-RNN architectures, this study combines CNN as a deep feature extractor with SVM as a classifier to optimize decision boundaries between genuine and manipulated facial images. This combination is expected to enhance classification accuracy while maintaining computational efficiency.

The contributions of this study are threefold. First, it proposes a hybrid CNN–SVM model for detecting morphing face deepfakes. Second, it evaluates the effectiveness of the proposed model using the Celeb-DF dataset as a benchmark for realistic deepfake detection. Third, it provides empirical evidence regarding the potential of integrating deep learning and machine learning techniques to strengthen biometric security systems, digital identity verification, and multimedia forensic applications. Therefore, this research is expected to contribute to the development of more reliable and effective deepfake detection systems in the era of digital transformation.

2. METHODS

This research aims to design and develop a deepfake detection system, specifically for detecting morphing face deepfakes, by utilizing a combination of machine learning and deep learning techniques. The proposed system is intended to identify manipulated facial images that have been generated through face morphing processes, which pose significant challenges to digital identity verification and security systems. *(Borghi et al., 2021).*

The research methodology consists of several main stages. The first stage is data collection, where a dataset containing genuine facial images and morphing face deepfake images is gathered from publicly available sources. The second stage is data preprocessing, which includes face detection, face cropping,

image resizing, and pixel normalization to improve data quality and ensure compatibility with the model input requirements. *(Jia et al., 2023b).*

After preprocessing, feature extraction is performed using a Convolutional Neural Network (CNN). In this stage, the CNN automatically learns and extracts important facial features from the images, such as texture patterns, facial structures, and subtle artifacts introduced during the morphing process. The extracted deep features are then used as input for the classification stage. *(Seibold et al., 2021).*

The classification process is carried out using a Support Vector Machine (SVM), which is employed to distinguish between genuine facial images and morphing face deepfake images based on the features generated by the CNN. The hybrid CNN-SVM approach is expected to leverage the powerful feature learning capability of CNNs and the robust classification performance of SVM. *(Reports, 2026).*

Finally, the performance of the proposed model is evaluated using several standard classification metrics, including accuracy, precision, recall, and F1-score. These metrics are used to assess the effectiveness and reliability of the system in detecting morphing face deepfakes and to compare its performance with existing approaches in the literature. The results of this research are expected to contribute to the development of more secure and reliable biometric authentication systems. *(Banas et al., 2026).*

2.1 *Celeb-Df* Datasets *Collection*



Figure 1. Datasets Collection

The Celeb-DF dataset was selected as the primary data source for this research because it provides a large collection of both real and deepfake videos featuring celebrities with high visual quality and realistic facial manipulations. Unlike earlier deepfake datasets that often contain visible artifacts and low-quality synthetic content, Celeb-DF presents more challenging and sophisticated deepfake examples, making it suitable for evaluating the effectiveness of modern deepfake detection methods. The dataset enables the proposed hybrid CNN–SVM model to learn discriminative facial features that distinguish authentic facial appearances from manipulated ones, thereby improving the robustness, accuracy, and generalization capability of the detection system in real-world scenarios.

The Celeb-DF dataset is systematically organized to support the training and evaluation process of the proposed model. The main dataset directory consists of four primary components: the Celeb-synthesis, Celeb-real, and YouTube-real folders, together with the List_of_testing_videos.txt file, which specifies the official testing videos. The Celeb-synthesis folder contains 5,639 DeepFake videos generated using advanced deep learning-based face manipulation techniques. These videos represent the Fake class and exhibit a high level of realism due to sophisticated face swapping and blending methods that closely resemble genuine facial appearances. The Celeb-real folder contains 590 original celebrity videos without any digital manipulation, serving as the Real class for comparison during the classification process. In addition, the YouTube-real folder provides supplementary authentic videos collected from YouTube, increasing the diversity of facial expressions, lighting conditions, poses, and background environments. Furthermore, the List_of_testing_videos.txt file defines the official testing

set, ensuring a consistent separation between training and testing data according to the standard Celeb-DF evaluation protocol.

During the research process, the dataset was divided into training, validation, and testing subsets to support model development and performance evaluation. Before being used for classification, the videos underwent a preprocessing stage, including frame extraction, face detection, face cropping, image resizing, pixel normalization, and labeling into Real and Fake classes. These preprocessing steps produce standardized facial images that enable the CNN to extract representative hierarchical features, which are subsequently classified by the SVM. By utilizing the Celeb-DF dataset and its standardized evaluation protocol, the proposed hybrid CNN–SVM model can effectively identify subtle inconsistencies in facial texture, facial structure, lighting conditions, and manipulation artifacts, resulting in a reliable and robust deepfake face detection system.

2.2 *Video Datasets Duration Trimming*



Figure 2. Duration Trimming

The video clipping process is carried out as an essential preprocessing step before the data is used for model training. The videos contained in the Celeb-DF dataset have varying durations, resolutions, and content complexity, which can affect the consistency and efficiency of the training process. Therefore, video clipping is applied to standardize the length of each video sample and ensure that only the most relevant segments are retained for analysis. During this process, specific portions of the original videos are selected based on visual quality and the visibility of facial regions. Segments that contain clear, stable, and well-illuminated facial appearances are prioritized, while unnecessary parts such as transitions, blurred frames, occlusions, or low-quality scenes are excluded. This selection helps reduce noise in the dataset and improves the quality of information provided to the model. The clipping procedure also contributes to reducing computational costs by shortening video duration while preserving essential facial characteristics needed for deepfake detection. As a result, each video sample becomes more uniform in length and content, facilitating efficient feature extraction and enabling the deep learning model to learn more effectively from consistent input data. The final output of this stage is a collection of shorter video clips that accurately represent facial movements and expressions with sufficient visual clarity and stability. These processed clips serve as optimized inputs for subsequent stages, including frame extraction, feature extraction, and the training of deepfake detection models, ultimately contributing to improved model performance and reliability.

2.3 *Extracting Video Datasets Into Static Images*



Figure 2 Extracting Video Datasets Into Static Images

The videos in the dataset were first converted into a sequence of frames, or static images, using Python and the OpenCV library. This process was performed to extract relevant visual information from each video while preserving important features required for analysis. Since Convolutional Neural Networks (CNNs) are designed to process image data rather than live video streams, each video was decomposed into individual frames at specific intervals. The extracted frames were then stored as image files and used as input for the subsequent preprocessing and classification stages. This approach enables the CNN model to learn spatial features such as shapes, textures, and patterns present in the images, thereby improving the effectiveness of the image recognition and classification process.

2.4 *Pre-Processing Data*



Figure 4. Pre-Processing Data

After the frames are extracted from the video, a data preprocessing stage is carried out to ensure that the images are suitable for analysis by the Convolutional Neural Network (CNN) model. This stage is essential because the quality and consistency of the input data significantly influence the model's performance and accuracy. Several preprocessing steps are applied to enhance the facial features and reduce irrelevant information.

a. Face Detection using MTCNN or Haar Cascade

The first step involves detecting the presence and location of faces within each extracted frame. Face detection can be performed using algorithms such as Multi-task Cascaded Convolutional Networks (MTCNN) or Haar Cascade Classifier. These methods identify facial regions by locating key facial structures, ensuring that only the relevant facial area is selected for further processing.

b. Cropping the Facial Area

Once the face is detected, the facial region is cropped from the original image. This process removes unnecessary background information and focuses the model's attention on facial characteristics that are important for analysis. By isolating the face, the CNN can learn more discriminative features and reduce the influence of environmental noise.

c. Image Resizing

The cropped facial images are then resized to a fixed dimension that matches the CNN model's input requirements, such as 224 × 224 pixels. Standardizing image dimensions ensures consistency across the dataset and allows the neural network to process all images efficiently without dimensional mismatches.

d. Pixel Value Normalization

Finally, the pixel values of the resized images are normalized to a range between 0 and 1 by dividing each pixel value by 255. This normalization step helps stabilize the training process, accelerates model convergence, and prevents large pixel values from dominating the learning process. As a result, the CNN can learn more effectively and achieve better performance during classification.

Overall, the preprocessing stage transforms the raw extracted frames into standardized and high-quality input data, enabling the CNN model to focus on meaningful facial features and improve the accuracy of the subsequent analysis.



Figure 3. The Number Of Datasets Used

After the preprocessing stage, the extracted facial images were grouped into two classes, namely Real and Fake. The Real class consisted of 6,072 facial images, while the Fake class consisted of 12,139 facial images, resulting in a total of 18,211 images used in this study. These images were obtained from the Celeb-DF dataset after video duration trimming, frame extraction, face detection, image resizing, and normalization processes.

The resulting dataset exhibits an imbalanced distribution, where the number of fake images is greater than the number of real images. Nevertheless, both classes contain sufficient samples to enable the model to learn distinctive visual characteristics between genuine and manipulated facial images. The Real class represents authentic facial images without modification, whereas the Fake class contains deepfake facial images generated through facial manipulation techniques.

The prepared image dataset was subsequently divided into training, validation, and testing subsets. The training data were used to learn facial feature representations through the Convolutional Neural Network (CNN), while the Support Vector Machine (SVM) classifier utilized the extracted features to distinguish between real and fake images. The validation set was employed to monitor model performance and prevent overfitting, whereas the testing set was used to evaluate the final performance of the proposed hybrid CNN–SVM model.

Figure X illustrates the distribution of facial images after preprocessing, showing 6,072 images in the Real category and 12,139 images in the Fake category. This dataset distribution serves as the foundation for training and evaluating the proposed deepfake morphing face detection system.

2.5 *Labeling Data*

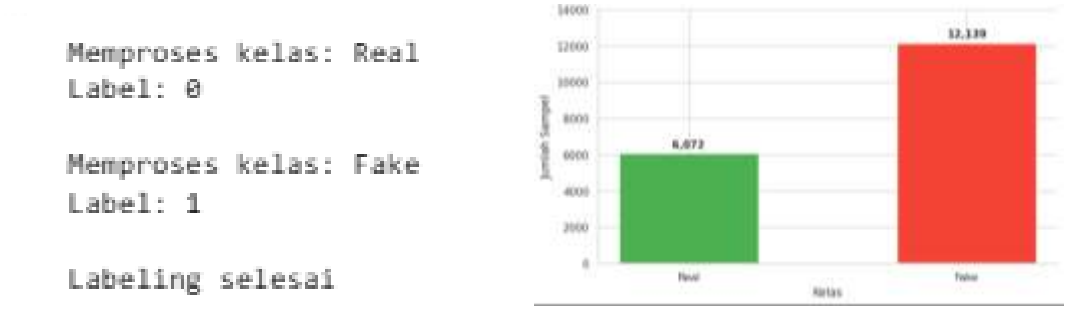

Figure 6. Labeling Data

Each facial image is labeled according to its class, either real (a genuine face image) or fake (a manipulated or synthesized face image). This labeling process is a crucial step in supervised learning because it provides the ground truth information required by the model during training. By assigning the correct class to each image, the model can learn to distinguish between authentic facial characteristics and artifacts introduced through manipulation techniques such as face swapping, reenactment, or deepfake generation. Furthermore, accurate labeling plays an important role in the evaluation phase, as it enables the measurement of model performance using metrics such as accuracy, precision, recall, and F1-score. Therefore, a well-labeled dataset contributes significantly to the development of a reliable and robust deepfake detection system.

2.6 *Splitting Datasets*

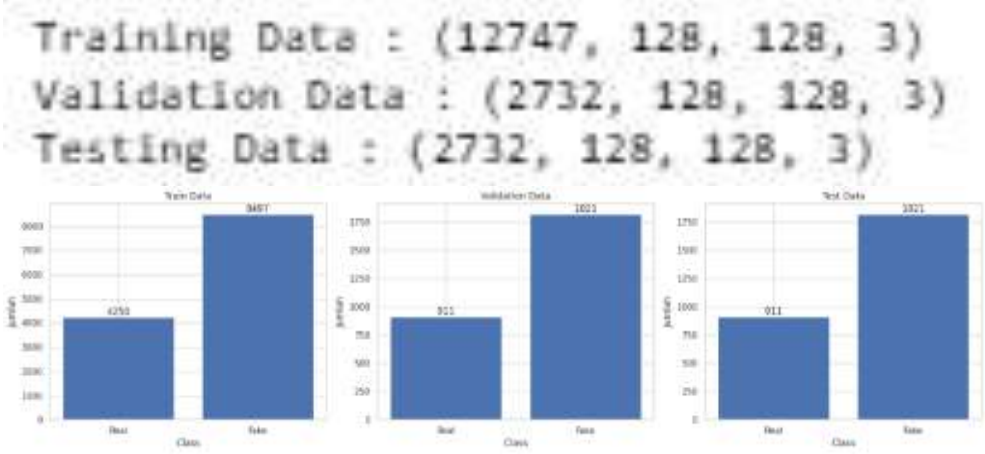

Figure 4. Splitting Datasets

After the preprocessing stage, a total of 18,211 facial images were obtained, consisting of 6,072 real images and 12,139 fake images. To ensure reliable model training and evaluation, the dataset was divided into three subsets: training, validation, and testing data using a ratio of 70:15:15. The dataset partitioning process was performed using the stratified sampling technique, ensuring that the proportion of real and fake images remained consistent across all subsets.

Based on this partitioning strategy, 12,747 images were allocated to the training set, 2,732 images to the validation set, and 2,732 images to the testing set. The training set consisted of 4,250 real images and 8,497 fake images, while both the validation and testing sets contained 911 real images and

1,821 fake images, respectively. This distribution ensures that each subset adequately represents the characteristics of the original dataset and minimizes potential bias during model development.

The training dataset was utilized to train the proposed hybrid CNN–SVM model. During this stage, the Convolutional Neural Network (CNN) learned discriminative visual features from facial images, including texture patterns, facial structures, and subtle artifacts associated with deepfake manipulation. These extracted features were subsequently used as input for the Support Vector Machine (SVM) classifier to distinguish between genuine and manipulated facial images.

The validation dataset was employed to monitor the learning process and optimize model hyperparameters. By evaluating model performance on unseen validation data during training, overfitting could be identified and reduced, thereby improving the model's generalization capability. Furthermore, the validation set was used to assess the effectiveness of different parameter configurations and determine the most suitable model architecture.

Finally, the testing dataset was reserved exclusively for the final evaluation stage. This dataset was not involved in either training or hyperparameter tuning, ensuring an unbiased assessment of the proposed model's performance. The testing phase measured the ability of the hybrid CNN–SVM model to accurately classify real and fake facial images under conditions that simulate real-world deployment scenarios.

Table 1. Dataset Distribution

Dataset	Real	Fake	Total
Training	4,250	8,497	12,747
Validation	911	1,821	2,732
Testing	911	1,821	2,732
Total	6,072	12,139	18,211

The use of stratified data splitting and separate validation and testing sets provides a robust experimental framework for evaluating the effectiveness of the proposed CNN–SVM model in detecting morphing face deepfake images using the Celeb-DF dataset.

2.7 *Hybrid CNN* and *SVM*

2.7.1 Pure Convolutional Neural Network (CNN)

The first approach investigated in this study is a Pure Convolutional Neural Network (CNN) model. CNN is one of the most widely adopted deep learning architectures in computer vision due to its ability to automatically extract meaningful features directly from raw image data. Unlike traditional machine learning approaches that require handcrafted feature extraction techniques, CNN learns visual representations through a hierarchical learning process.

The architecture consists of multiple convolutional layers, activation functions, pooling layers, and fully connected layers. During the convolution process, filters are applied to input images to capture low-level and high-level visual characteristics. The initial layers typically learn simple features such as edges, contours, and color gradients, while deeper layers learn more complex facial characteristics including texture irregularities, facial geometry, illumination patterns, and manipulation artifacts.

For morphing face deepfake detection, these extracted features are particularly important because manipulated facial images often contain subtle inconsistencies that cannot be easily recognized by human observers. CNN can identify these hidden patterns through repeated convolution operations and nonlinear transformations.

Following the convolution stage, max-pooling layers are utilized to reduce spatial dimensions while preserving the most significant features. This process decreases computational complexity and minimizes the risk of overfitting. The resulting feature maps are then flattened and passed to fully connected layers, where classification is performed using a Softmax activation function.

The model was trained using the Adam optimization algorithm due to its adaptive learning capability and efficient convergence behavior. Adam combines the advantages of Momentum and RMSProp optimization methods, allowing stable parameter updates throughout the training process. Training was conducted for multiple epochs to ensure sufficient learning of visual patterns contained in the dataset.

Table 2. CNN Configuration

Parameter	Value
Input Size	128 × 128 × 3
Convolution Layers	3
Filters	32, 64, 128
Kernel Size	3 × 3
Activation Function	ReLU
Pooling Layer	Max Pooling
Dense Layer	128 Neurons
Output Layer	Softmax
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	50

The Pure CNN model serves as the baseline deep learning model in this research. Its performance provides an important benchmark for evaluating whether the integration of SVM can further improve classification effectiveness.

2.7.2 Pure Support Vector Machine (SVM)

The second approach implemented in this study is a Pure Support Vector Machine (SVM) classifier. SVM is a supervised machine learning algorithm designed to perform classification by constructing an optimal hyperplane that maximizes the separation margin between classes.

SVM has been extensively applied in image classification problems due to its strong generalization capability, particularly when dealing with limited datasets and high-dimensional feature spaces. The algorithm seeks to identify support vectors, which are the most informative training samples located near the decision boundary. By maximizing the margin between classes, SVM aims to improve classification robustness and reduce the likelihood of misclassification.

In the context of deepfake detection, SVM can effectively separate real and manipulated images based on extracted numerical features. Unlike CNN, SVM does not automatically learn visual representations from raw images. Instead, image data must first be transformed into feature vectors before classification can be performed.

To handle nonlinear relationships within the data, this study employed the Radial Basis Function (RBF) kernel. The RBF kernel maps feature vectors into a higher-dimensional space where complex class boundaries can be more effectively separated. Two important hyperparameters were configured, namely the regularization parameter (C) and gamma. The parameter C controls the trade-off between maximizing classification accuracy and maintaining a large decision margin, while gamma determines the influence range of individual training samples.

Table 3. SVM Configuration

Parameter	Value
Classifier	SVM
Kernel	RBF
C	1.0
Gamma	0.001
Classification Type	Binary

The implementation of the Pure SVM model allows the study to evaluate the effectiveness of traditional machine learning techniques in detecting morphing face deepfake images and serves as a comparative benchmark against deep learning-based methods.

2.7.3 Hybrid CNN–SVM

The primary contribution of this research is the development of a Hybrid CNN–SVM model that combines the strengths of deep learning and traditional machine learning techniques. This hybrid architecture was designed to address limitations observed in individual CNN and SVM models.

Although CNN is highly effective at extracting deep visual features, its Softmax classifier may not always provide optimal decision boundaries when handling highly similar image classes. In morphing face deepfake detection, the distinction between genuine and manipulated images is often extremely subtle, making classification particularly challenging. Consequently, an alternative classifier capable of constructing stronger decision boundaries may improve overall performance.

In the proposed framework, CNN is utilized exclusively as a feature extraction network. Rather than using the Softmax layer for final classification, feature vectors generated by the Flatten layer are extracted and transferred to the SVM classifier. These feature vectors contain rich visual information learned by CNN, including facial texture inconsistencies, edge patterns, illumination variations, and manipulation traces.

The SVM classifier then processes these deep feature representations and determines the optimal separation boundary between real and fake classes. By leveraging CNN's representation learning capability and SVM's classification strength, the hybrid model aims to achieve higher detection accuracy, improved generalization, and better robustness against subtle morphing attacks.

The hybrid architecture also provides computational advantages because SVM can perform classification more efficiently once the feature extraction stage has been completed. This makes the approach suitable for practical biometric security applications where both accuracy and efficiency are critical requirements.

Table 2. Hybrid CNN–SVM Configuration

Component	Configuration
Feature Extractor	CNN
Input Size	128 × 128 × 3
Convolution Layers	3
Filters	32, 64, 128
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	50
Feature Layer	Flatten Layer
Classifier	SVM
Kernel	RBF
C	1.0
Gamma	0.001

The hybrid CNN–SVM model represents the proposed solution in this study and is expected to outperform both Pure CNN and Pure SVM models by combining advanced feature extraction capabilities with robust classification mechanisms. Through comparative evaluation, the effectiveness of this approach in detecting morphing face deepfake images can be systematically assessed and validated.

```
[Langkah 1] Membaca gambar dari folder Drive ke RAM...

[Langkah 2] Mematangkan CNN Feature Extractor via Functional API...
Epoch 1/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 167s 719ms/step - accuracy: 0.7683 - loss: 0.5519 - val_accuracy: 0.6851 - val_loss: 0.7818
Epoch 2/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 207s 741ms/step - accuracy: 0.8418 - loss: 0.3704 - val_accuracy: 0.7038 - val_loss: 0.8854
Epoch 3/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 198s 721ms/step - accuracy: 0.8855 - loss: 0.2805 - val_accuracy: 0.9072 - val_loss: 0.2410
Epoch 4/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 200s 715ms/step - accuracy: 0.9138 - loss: 0.2148 - val_accuracy: 0.6390 - val_loss: 0.7591
Epoch 5/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 165s 725ms/step - accuracy: 0.9349 - loss: 0.1690 - val_accuracy: 0.4579 - val_loss: 2.0608
Epoch 6/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 163s 715ms/step - accuracy: 0.9463 - loss: 0.1384 - val_accuracy: 0.9435 - val_loss: 0.1586
Epoch 7/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 203s 721ms/step - accuracy: 0.9564 - loss: 0.1109 - val_accuracy: 0.8608 - val_loss: 0.3492
Epoch 8/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 163s 715ms/step - accuracy: 0.9633 - loss: 0.1010 - val_accuracy: 0.9067 - val_loss: 0.2343
Epoch 9/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 204s 725ms/step - accuracy: 0.9703 - loss: 0.0779 - val_accuracy: 0.8688 - val_loss: 0.3354
Epoch 10/10
228/228 ━━━━━━━━━━━━━━━━━━━━ 162s 711ms/step - accuracy: 0.9714 - loss: 0.0764 - val_accuracy: 0.9336 - val_loss: 0.1714

[Langkah 3] Pemuatan Fitur Bobot Tinggi Menuju SVM...
456/456 ━━━━━━━━━━━━━━━━━━━━ 37s 80ms/step
114/114 ━━━━━━━━━━━━━━━━━━━━ 7s 65ms/step

[Langkah 4] Melatih SVM pada Fitur Bobot Tinggi CNN...
```

Figure 5. Hybrid CNN dan SVM

The proposed hybrid CNN–SVM model consists of two integrated stages, namely CNN feature extraction and SVM classification, which work collaboratively to identify whether a facial image is authentic or manipulated. In this hybrid approach, the Convolutional Neural Network (CNN) is employed as an automatic feature extractor, while the Support Vector Machine (SVM) serves as the final classifier. This combination leverages the powerful representation learning capability of CNN and the robust classification performance of SVM, resulting in improved accuracy and generalization compared to using a conventional CNN with a Softmax classifier.

During the CNN feature extraction stage, the preprocessed facial image is first fed into the CNN model. The convolutional layers apply multiple learnable filters to extract important visual features such as edges, contours, textures, color inconsistencies, and facial patterns that may indicate image manipulation. After each convolution operation, activation functions introduce non-linearity, enabling the network to learn increasingly complex and meaningful representations from the input image. Pooling layers are then used to reduce the spatial dimensions of the feature maps while preserving the most significant information, thereby lowering computational complexity and reducing the risk of overfitting. Through successive convolution and pooling operations, the CNN learns hierarchical feature representations, beginning with low-level features such as edges and corners and progressing toward high-level semantic features, including facial structures and deepfake artifacts. The extracted feature maps are subsequently transformed into a one-dimensional feature vector using a flattening layer. This feature vector is further refined through fully connected layers to produce a compact and highly discriminative representation of the facial image. Instead of performing the final classification using the Softmax layer, the hybrid CNN–SVM architecture utilizes this extracted feature vector as the input for the SVM classifier.

In the SVM classification stage, the feature vector generated by the CNN is analyzed within a multidimensional feature space to distinguish between real and manipulated facial images. Depending on the characteristics of the extracted features, an appropriate kernel function, such as the Linear kernel for linearly separable data or the Radial Basis Function (RBF) kernel for non-linear decision boundaries, can be employed. During the training process, the SVM identifies the support vectors, which are the most informative data points located near the decision boundary between classes. Based on these support vectors, the SVM constructs an optimal hyperplane that maximizes the margin separating real and fake images, thereby enhancing the classifier's generalization capability. For each input image, the SVM calculates the position of its feature vector relative to the hyperplane and determines the corresponding class according to the learned decision boundary. The model then generates the final prediction by classifying the facial image as either real or fake. By integrating CNN's powerful feature extraction capability with SVM's effective margin-based classification mechanism, the hybrid CNN–

SVM model achieves more reliable and accurate deepfake image detection, particularly when dealing with complex visual patterns and limited training data.

2.8 *Evaluation and Analysis of Results*

```
...
[CLASSIFICATION REPORT: CNN MURNI]
------------------------------------------------
              precision    recall  f1-score   support

        Real       0.92      0.89      0.90      1214
        Fake       0.95      0.96      0.95      2428

    accuracy                           0.94      3642
   macro avg       0.93      0.92      0.93      3642
weighted avg       0.94      0.94      0.94      3642
Total Akurasi: 93.6%
------------------------------------------------

[CLASSIFICATION REPORT: SVM MURNI]
------------------------------------------------
              precision    recall  f1-score   support

        Real       0.94      0.91      0.93      1214
        Fake       0.96      0.97      0.96      2428

    accuracy                           0.95      3642
   macro avg       0.95      0.94      0.94      3642
weighted avg       0.95      0.95      0.95      3642
Total Akurasi: 95.14%
------------------------------------------------

[CLASSIFICATION REPORT: HYBRID CNN-SVM]
------------------------------------------------
              precision    recall  f1-score   support

        Real       0.94      0.92      0.93      1214
        Fake       0.96      0.97      0.96      2428

    accuracy                           0.95      3642
   macro avg       0.95      0.94      0.95      3642
weighted avg       0.95      0.95      0.95      3642
Total Akurasi: 95.31%
```

Figure 9. Evaluation and Analysis of Results

The model was evaluated using several performance metrics to measure its effectiveness in distinguishing between real and fake videos. Accuracy was used to evaluate the overall correctness of the model in classifying all data samples. Precision was applied to determine how accurately the model identified videos belonging to the fake class, indicating the proportion of correctly predicted fake videos among all videos predicted as fake. Recall was used to measure the model's ability to detect all fake videos present in the dataset, reflecting how many actual fake videos were successfully identified by the model. Furthermore, the F1-Score was employed as a balanced metric that combines both precision and recall, providing a comprehensive assessment of the model's classification performance. In addition to these evaluation metrics, a confusion matrix was utilized to visualize the distribution of prediction results, including correctly classified and misclassified samples, thereby offering deeper insight into the strengths and weaknesses of the proposed model.

3. FINDINGS AND DISCUSSION

3.1 Classification Performance Evaluation

```
============================================================

[CLASSIFICATION REPORT: CNN MURNI]
------------------------------------------------
              precision    recall  f1-score   support

        Real       0.92      0.89      0.90      1214
        Fake       0.95      0.96      0.95      2428

    accuracy                           0.94      3642
   macro avg       0.93      0.92      0.93      3642
weighted avg       0.94      0.94      0.94      3642
Total Akurasi: 93.6%
------------------------------------------------

[CLASSIFICATION REPORT: SVM MURNI]
------------------------------------------------
              precision    recall  f1-score   support

        Real       0.94      0.91      0.93      1214
        Fake       0.96      0.97      0.96      2428

    accuracy                           0.95      3642
   macro avg       0.95      0.94      0.94      3642
weighted avg       0.95      0.95      0.95      3642
Total Akurasi: 95.14%
------------------------------------------------

[CLASSIFICATION REPORT: HYBRID CNN-SVM]
------------------------------------------------
              precision    recall  f1-score   support

        Real       0.94      0.92      0.93      1214
        Fake       0.96      0.97      0.96      2428

    accuracy                           0.95      3642
   macro avg       0.95      0.94      0.95      3642
weighted avg       0.95      0.95      0.95      3642
Total Akurasi: 95.31%
------------------------------------------------
```

Figure 10. Final Report of Three Models

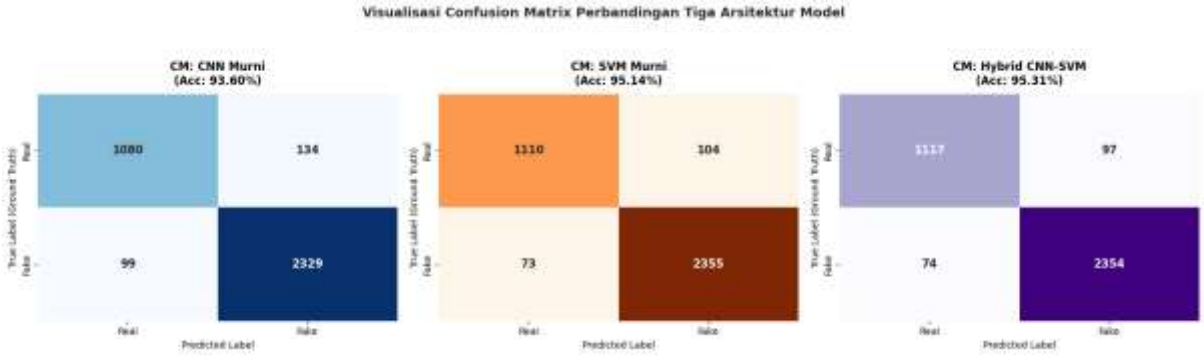

Figure 11. Confusion Matrix Comparison of Three Model Architectures

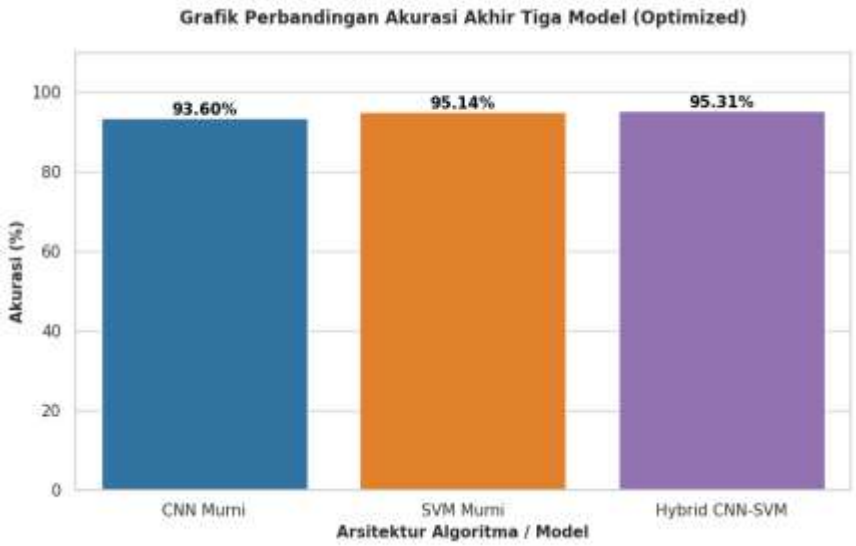

Figure 12. Comparison Chart of Accuracy for Three Models

The performance of the proposed model was evaluated using accuracy, precision, recall, and F1-score metrics. Three classification approaches were compared, namely Pure CNN, Pure SVM, and the proposed Hybrid CNN–SVM model. The evaluation was conducted using the same training, validation, and testing datasets to ensure a fair comparison between models.

Table 3.1 Performance Comparison of Classification Models

Table 5. Performance Comparison of Classification Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Pure CNN	93.60	93.00	92.00	93.00
Pure SVM	95.14	95.00	94.00	94.00
Hybrid CNN-SVM	95.31	95.00	94.00	95.00

Based on Table X, the proposed Hybrid CNN–SVM model achieved the highest overall performance with an accuracy of 95.31%. This result indicates that the proposed model was able to correctly classify the majority of real and fake facial images contained in the Celeb-DF dataset. The obtained precision, recall, and F1-score values also demonstrate balanced classification performance, suggesting that the model is capable of minimizing both false positive and false negative predictions.

Compared to the Pure CNN model, the Hybrid CNN–SVM approach achieved an improvement of approximately 1.71% in accuracy. Although this numerical increase may appear relatively small, in image classification tasks involving highly similar classes such as morphing face deepfakes, even a slight improvement can indicate a significant enhancement in model discrimination capability. This finding suggests that the integration of CNN and SVM contributes positively to the identification of subtle visual artifacts generated by facial manipulation techniques.

Furthermore, the Hybrid CNN–SVM model also slightly outperformed the Pure SVM model, which achieved an accuracy of 95.14%. Although the difference was only 0.17%, the hybrid model consistently produced the highest overall evaluation metrics, indicating that the combination of deep feature extraction and machine learning-based classification can provide complementary advantages for deepfake detection tasks.

3.2 Analysis of CNN Feature Extraction Capability

```
============================================
        CLASSIFICATION REPORT: CNN MURNI
============================================
              precision    recall  f1-score   support

        Real       0.86      0.96      0.91      1215
        Fake       0.98      0.93      0.95      2428

    accuracy                           0.94      3643
   macro avg       0.92      0.94      0.93      3643
weighted avg       0.94      0.94      0.94      3643

Total Akurasi Akhir CNN Murni: 93.60%
```



Figure 13. Final Report And Confusion Matrix Of Pure CNN

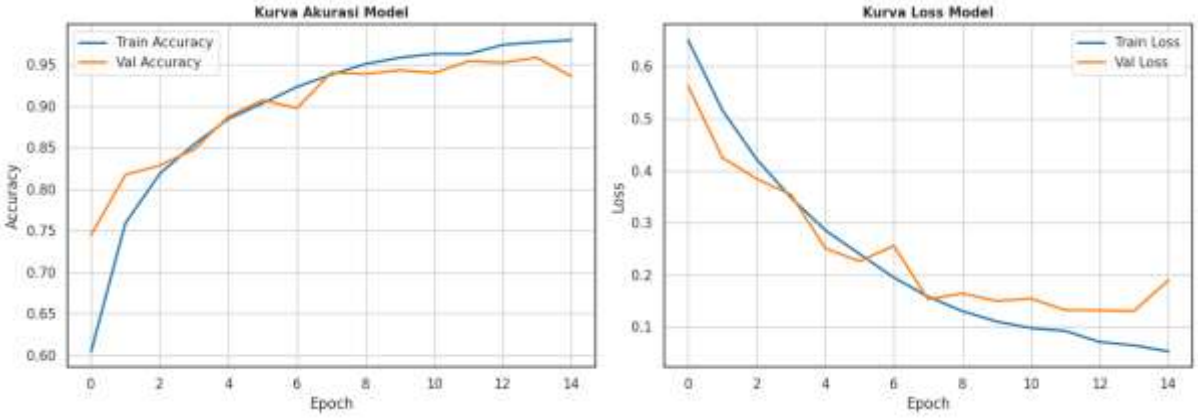

Figure 14. CNN Model Accuracy Curve

The Pure CNN model achieved an accuracy of 93.60%, indicating that convolutional neural networks are highly effective in extracting visual information from facial images. CNN automatically learns hierarchical feature representations through multiple convolutional and pooling layers, enabling the model to capture low-level features such as edges, textures, and color distributions, as well as higher-level facial characteristics.

In the context of morphing face deepfake detection, the extracted features include subtle inconsistencies in facial texture, unnatural blending regions, illumination differences, and geometric distortions introduced during the morphing process. These visual artifacts are often difficult to identify through manual observation because morphing techniques are specifically designed to preserve visual realism. Consequently, the ability of CNN to automatically learn deep feature representations becomes a critical factor in detecting manipulated facial images.

Despite its strong feature extraction capability, the classification performance of the Pure CNN model remained lower than the hybrid approach. This result suggests that while CNN successfully learns discriminative features, the fully connected classification layer may not always provide optimal decision boundaries when dealing with highly overlapping feature distributions between real and fake images.

3.3 Analysis of SVM Classification Performance

```
==========================================
   CLASSIFICATION REPORT: PURE SVM (HOG + RBF)
==========================================
              precision    recall  f1-score   support

        Real       0.98      0.88      0.92      1215
        Fake       0.94      0.99      0.96      2428

    accuracy                           0.95      3643
   macro avg       0.96      0.93      0.94      3643
weighted avg       0.95      0.95      0.95      3643

Akurasi Akhir SVM Murni (HOG): 95.14%
```

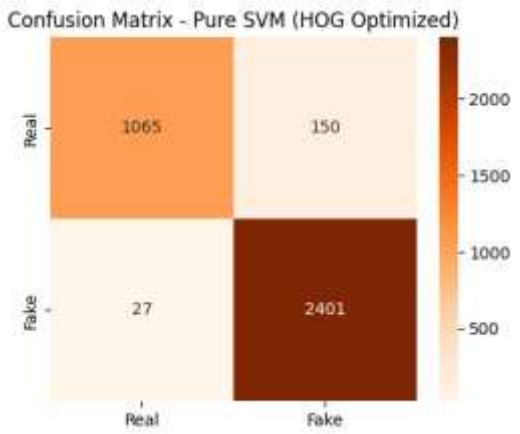

Figure 15. Final Report And Confusion Matrix Of Pure SVM

The Pure SVM model obtained an accuracy of 95.14%, outperforming the Pure CNN model. This result highlights the effectiveness of Support Vector Machine in handling binary classification problems involving complex feature distributions.

SVM operates by constructing an optimal hyperplane that maximizes the separation margin between classes. In this study, the classifier was able to effectively distinguish between authentic and manipulated facial images by identifying the most informative boundaries within the feature space. The use of the Radial Basis Function (RBF) kernel further enhanced the capability of SVM to model nonlinear relationships among facial features.

The strong performance achieved by the Pure SVM model indicates that classical machine learning algorithms remain highly relevant in image classification tasks when supported by representative feature sets. However, SVM performance is largely dependent on the quality of extracted features. Therefore, integrating CNN as an automatic feature extractor can further enhance classification effectiveness by providing richer and more discriminative feature representations.

3.4 Analysis of Hybrid CNN–SVM Model

```
==========================================
   CLASSIFICATION REPORT: HYBRID CNN-SVM
==========================================
              precision    recall  f1-score   support

        Real       0.93      0.93      0.93      1215
        Fake       0.97      0.96      0.96      2428

    accuracy                           0.95      3643
   macro avg       0.95      0.95      0.95      3643
weighted avg       0.95      0.95      0.95      3643

Akurasi Akhir Model Hybrid CNN-SVM: 95.31%
```

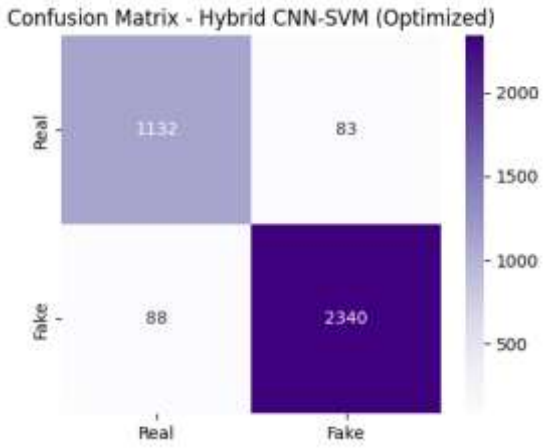

Figure 16. Final Report And Confusion Matrix Of Hybrid CNN And SVM

The proposed Hybrid CNN–SVM model achieved the highest performance among all evaluated approaches. This result demonstrates that combining CNN and SVM can effectively leverage the strengths of both methods.

CNN serves as a deep feature extractor capable of learning complex facial representations directly from image data, while SVM functions as a robust classifier that identifies optimal decision

boundaries between classes. By separating the feature extraction and classification stages, the hybrid architecture reduces the limitations associated with using CNN as a standalone classifier.

The superior performance of the hybrid model suggests that the extracted feature vectors contain sufficient discriminative information for classification, while the SVM classifier is capable of utilizing these features more effectively than the traditional fully connected layers commonly used in CNN architectures. As a result, the model exhibits improved robustness when distinguishing between highly similar real and morphing face deepfake images.

The results also indicate that the proposed architecture is capable of detecting subtle manipulation patterns that may not be visible to human observers. This capability is particularly important because morphing attacks are specifically designed to create facial images that closely resemble authentic faces, making manual detection increasingly difficult.

3.5 Error Analysis

Although the proposed model achieved satisfactory performance, several classification errors were still observed during testing. Most misclassified samples were associated with low-quality images, non-uniform illumination conditions, motion blur, and facial occlusions.

In some cases, manipulated facial images were incorrectly classified as authentic because the morphing artifacts were extremely subtle and visually consistent with natural facial characteristics. Similarly, some authentic facial images were incorrectly classified as fake due to unusual lighting conditions or compression artifacts that resembled manipulation traces.

These findings indicate that image quality remains an important factor affecting model performance. Improving image preprocessing techniques, incorporating attention mechanisms, or utilizing more diverse training datasets may help reduce misclassification rates in future studies.

3.6 Comparison with Previous Studies

The findings of this study were compared with several previous studies related to deepfake detection and morphing attack detection.

Table 6. Comparison with Previous Studies

Study	Method	Dataset	Accuracy
Borghi et al. (2021)	Double Siamese Network	Morphing Dataset	92.00%
Jia et al. (2023)	Progressive Enhancement Learning	Morphing Dataset	94.00%
Al-Dulaimi & Kurnaz (2024)	Hybrid CNN-LSTM	Deepfake Dataset	95.00%
Proposed Study	Hybrid CNN-SVM	Celeb-DF	95.31%

The comparison demonstrates that the proposed Hybrid CNN–SVM model achieved competitive performance relative to previous deepfake detection approaches. While Borghi et al. focused on Siamese-based morphing attack detection and Jia et al. employed progressive enhancement learning, the present study integrates deep feature extraction and classical machine learning classification within a single framework.

Unlike many recent studies that rely exclusively on deep neural networks, the proposed approach demonstrates that combining CNN and SVM can still produce highly competitive results. This finding suggests that hybrid architectures remain a promising direction for deepfake detection research, particularly when computational efficiency and model interpretability are important considerations.

Overall, the obtained results confirm that the proposed Hybrid CNN–SVM model provides an effective solution for morphing face deepfake detection and contributes to the advancement of secure biometric authentication and digital identity verification systems.

4. CONCLUSION

Based on the results obtained in this study, it can be concluded that the proposed Hybrid CNN–SVM model is capable of effectively detecting morphing face deepfake images using the Celeb-DF dataset. Experimental evaluation showed that the Hybrid CNN–SVM model achieved the highest performance among the tested approaches, with an accuracy of 95.31%, outperforming the Pure CNN model, which achieved 93.60%, and the Pure SVM model, which achieved 95.14%. In addition, the proposed model obtained a precision of 95%, recall of 94%, and F1-score of 95%, indicating a strong capability to distinguish between real and manipulated facial images.

The comparative results demonstrate that the integration of CNN and SVM provides advantages over using either method independently. CNN effectively extracts discriminative facial features from image data, while SVM contributes to improved classification performance through optimal decision boundary construction. This combination enables the model to identify subtle visual inconsistencies commonly found in morphing face deepfakes, resulting in more accurate and stable classification outcomes.

The findings of this study confirm that deep learning-based feature extraction combined with machine learning-based classification can serve as an effective approach for addressing the challenges of facial image manipulation detection. The proposed model shows potential for practical implementation in applications that require image authenticity verification, such as biometric authentication systems, digital identity verification, border control systems, and multimedia forensic investigations.

Despite achieving satisfactory performance, this study is limited by the characteristics and distribution of the dataset used during experimentation. Therefore, future research is recommended to evaluate the proposed model using larger and more diverse datasets, investigate additional feature extraction architectures, and compare the hybrid approach with more advanced deep learning models. Such developments may further improve detection accuracy and enhance the robustness of deepfake detection systems in real-world environments.

Overall, the results indicate that the Hybrid CNN–SVM model represents a reliable and practical solution for morphing face deepfake detection, contributing to the development of more secure and trustworthy digital identity verification technologies.

REFERENCES

Akhtar, Z. (2023). Deepfakes Generation and Detection: A Short Survey. *Journal of Imaging*, *9*(1). https://doi.org/10.3390/jimaging9010018

Al-Adwan, A., Alazzam, H., Al-Anbaki, N., & Alduweib, E. (2024). Detection of Deepfake Media Using a Hybrid CNN–RNN Model and Particle Swarm Optimization (PSO) Algorithm. *Computers*, *13*(4), 1–16. https://doi.org/10.3390/computers13040099

Al-Dulaimi, O. A. H. H., & Kurnaz, S. (2024). A Hybrid CNN-LSTM Approach for Precision Deepfake Image Detection Based on Transfer Learning. *Electronics (Switzerland)*, *13*(9), 1–22. https://doi.org/10.3390/electronics13091662

Altuncu, E., Franqueira, V. N. L., & Li, S. (2024). Deepfake: definitions, performance metrics and standards, datasets, and a meta-review. *Frontiers in Big Data*, *7*. https://doi.org/10.3389/fdata.2024.1400024

Banas, A., Rathgeb, C., Merkle, J., & Schaubert, M. (2026). *IDSwapMAD : towards privacy - friendly training of differential face morphing attack detection*.

Borghi, G., Pancisi, E., Ferrara, M., & Maltoni, D. (2021). A double siamese framework for differential morphing attack detection. *Sensors*, *21*(10). https://doi.org/10.3390/s21103466

Gambín, Á. F., Yazidi, A., Vasilakos, A., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: current and future trends. In *Artificial Intelligence Review* (Vol. 57, Issue 3). Springer Netherlands. https://doi.org/10.1007/s10462-023-10679-x

Hamza, M., Tehsin, S., Humayun, M., Almufareh, M. F., & Alfayad, M. (2022). A Comprehensive

Review of Face Morph Generation and Detection of Fraudulent Identities. *Applied Sciences (Switzerland)*, *12*(24). https://doi.org/10.3390/app122412545

Jia, C. K., Liu, Y. C., & Chen, Y. L. (2023a). Face morphing attack detection based on high-frequency features and progressive enhancement learning. *Frontiers in Neurorobotics*, *17*. https://doi.org/10.3389/fnbot.2023.1182375

Jia, C. K., Liu, Y. C., & Chen, Y. L. (2023b). Face morphing attack detection based on high-frequency features and progressive enhancement learning. *Frontiers in Neurorobotics*, *17*. https://doi.org/10.3389/fnbot.2023.1182375

Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3204–3213. https://doi.org/10.1109/CVPR42600.2020.00327

Pikoulis, E. V., Ioannou, Z. M., Paschou, M., & Sakkopoulos, E. (2021). Face morphing, a modern threat to border security: Recent advances and open challenges. *Applied Sciences (Switzerland)*, *11*(7). https://doi.org/10.3390/app11073207

Reports, S. (2026). *Scientific Reports Article in Press Deepfake face detection using hybrid bag-of- visual-words and multi-CNN feature fusion IN IN*.

Seibold, C., Hilsmann, A., & Eisert, P. (2021). Feature focus: Towards explainable and transparent deep face morphing attack detectors†. *Computers*, *10*(9), 1–18. https://doi.org/10.3390/computers10090117