Volume 01, Issue 02, December 2025 : 47-53 Received: Nopember 03, 2025. Revised: December 26, 2025.

IJEIE International Journal of Electrical and Intelligent Engineering

Deepfake Image Detection Using Transfer Learning Method

Tsalatsatun Nur Rohmah[1] Dewi Purnamasari[2]* Kurniawati[3] Didin Herlinudinkhaji[4]
Yunifa Miftachul Arif [5] Santiago Criollo-C[6]

[1,2,3] *Department of System and Information Technology, Universitas Ivet, Semarang, Indonesia*
[4]*Digital Bussiness, Universitas Ivet, Semarang, Indonesia*
[5]*Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia*
[6]*Carrera de Ingeniería en Ciberseguridad, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de Las Américas, Quito 170125, Ecuador*
* Corresponding author's Email: dewi.poernamasari.09@gmail.ac.id

Abstract: The development of Artificial Intelligence (AI) technologies, particularly deep learning has led to the emergence of innovative applications such as deepfake technology, which enables the realistic manipulation of digital images and videos. While this technology offers positive applications in fields such as entertainment and education, it also poses significant risks of misuse, particularly in the dissemination of false information and violations of privacy. Therefore, deepfake detection has become a crucial aspect in preserving the authenticity of digital content. This study aims to analyze the effectiveness of transfer learning methods in detecting deepfake images using VGG16, VGG19, and ResNet50 architectures. The research employs a dataset of deepfake and real images sourced from Kaggle, comprising 10,826 facial images with a resolution of 256 × 256 pixels, evenly balanced between authentic and manipulated content. The data are split in an 80:20 ratio for training and testing purposes. Each model is trained using identical parameter configurations. The performance evaluation of the models was conducted using confusion matrix metrics, including accuracy, precision, recall, and F1-score. The results indicate that the VGG16 model achieved the best performance, with an accuracy of 76%, followed by VGG19 at 72%, and ResNet50 at 58%. VGG16 also outperformed the other models in terms of precision, recall, and F1-score, demonstrating more effective performance in identifying visual manipulation patterns. In contrast, ResNet50 exhibited the lowest performance, which may be attributed to its architectural complexity not being optimally aligned with the characteristics of the dataset. It can be concluded that the transfer learning approach using the VGG16 model is more effective in detecting deepfake images on this dataset. This study also highlights the importance of selecting appropriate architectures and fine-tuning models to the characteristics of the data.

Keywords: Artificial Intelligence, ResNet50, Transfer Learning, VGG16, VGG19

1. Introduction

Indonesia is currently entering the era of Industry 5.0, characterized by the integration of advanced digital technologies into various sectors. This demonstrates the rapid development of various revolutions and innovations. Rapid technological advancements have blurred the boundaries between physical and digital media. In the era of Society 5.0, it makes it easier for us to exchange information and data. These technological advancements have generated both positive and negative impacts across multiple domains[1][2][3]. One of the technologies is Artificial Intelligence (AI).

The development of Artificial Intelligence (AI) technology, particularly deep learning has led to new innovations in the field of digital image processing, one of which is deepfake technology [4]. Deepfake refers to a technique that utilizes artificial neural network algorithms to manipulate images or videos by modifying elements such as faces, backgrounds, facial expressions, and voices. This technology generates highly realistic but manipulated content, making it challenging to differentiate from authentic media. [5] Currently, deepfake technology has not only advanced technically but has also emerged as a widely discussed social phenomenon, particularly within the digital realm, where it attracts public attention while simultaneously raising concerns. [6]

The use of deepfake technology for digital image manipulation has become increasingly widespread, encompassing both positive and negative aspects. This technology holds positive potential in fields such as entertainment, education, advertising, and the film industry. However, on the other hand deepfake technology also raises serious concerns related to the dissemination of false information, privacy violations, and defamation, particularly when it is used unethically to manipulate human faces in the digital real image [5][7]. Therefore, deepfake image detection has become a crucial component in efforts to mitigate the risks associated with the misuse of this technology, particularly in distinguishing between authentic visual content and that manipulated using deepfake techniques. Research and development of accurate and efficient detection models are essential to maintain public trust in the authenticity of digital content and to prevent broader negative impacts on individuals and society.

Alongside the advancement of artificial intelligence (AI) based forgery technologies, researchers and technology experts continuously strive to develop methods for detecting manipulated media [8][9]. Fake content generated by deepfake technology is increasingly difficult to distinguish from authentic content due to the growing sophistication and complexity of deepfake methods, which result in highly realistic media. Therefore, various approaches continue to be tested and refined to effectively identify signs of manipulation, such as unnatural facial expressions or inconsistencies in lighting and motion [10]. Artificial intelligence (AI) based approaches, particularly deep learning offer effective and accurate solutions to address these challenges. Deep learning is capable of extracting and learning complex features from visual data, making it highly effective in distinguishing authentic images from manipulated ones [9].

Deep learning methods such as Convolutional Neural Networks (CNN) and autoencoder based architectures have been widely applied in deepfake detection models and have proven effective in identifying visual manipulation patterns generated by deepfake techniques [11]. Moreover, the transfer learning approach which utilizes pre-trained models on large scale datasets and fine-tunes them for specific tasks, can also be effectively applied to deepfake detection models [12]. In this study, the transfer learning approach is utilized, as it has been shown to be more efficient than building a model from scratch. By leveraging pre-trained models on large-scale datasets, the training process can be accelerated and computational resource requirements significantly reduced. This approach is particularly advantageous when training data is limited. Pre-trained models typically possess the ability to recognize general visual patterns, requiring only fine-tuning for specific tasks such as detecting facial manipulations generated by deepfake techniques [13][14][15][16].

2. Literature Review

This study Iqbal et al research the transfer learning method using Inception V3, VGG19, and VGG16. dataset using deepfake and real image. The transfer learning method using VGG16 achieved the highest accuracy, reaching 90% [17].

The study proposes a deepfake detection approach using a transfer learning-based Xception model fine-tuned on datasets such as FaceForensics++ and Celeb-DF. Experimental results show that the method achieves high accuracy (over 90%) by effectively capturing subtle deepfake artifacts, though its performance decreases on unseen datasets, indicating the need for improved generalization through more robust training strategies [14].

Singh et al the VGG16 using transfer learning method achieved an accuracy of 96% on the testing dataset. The dataset comprises both authentic and deepfake media, although the specific dataset name is not explicitly stated [16]. Matei et al research used the VGG16 transfer learning method. The datasets used were Diffusion DB and the Open Images dataset. The results showed that transfer learning with a pre-trained VGG16 model achieved 97% accuracy. Research used transfer learning with a pre-trained VGG16 model. The dataset used included real and fake images. The results showed an accuracy of 93.53% [18]. Ivancakova et al research methodology used the Cross Industry Standard Process for Data Mining (CRISP-DM). The research data used was the Wisconsin Breast Cancer Database (WDBC).[19]

The state of the art in this research is that previous evaluations only calculated accuracy. However, in this study, model performance was evaluated using several evaluation metrics, namely confusion matrix, accuracy, precision, recall, and F1-score. The testing dataset used in this study consisted of 2,164 images, comprising 1,082 images labeled as fake (0) and 1,082 images labeled as real (1). The dataset was evenly balanced to ensure fair evaluation and reliable confusion matrix interpretation. This research was conducted using the VGG16, VGG19, and ResNet50 approaches for deepfake image detection. The dataset used was a collection of deepfake and real images sourced from Kaggle. This dataset contains a collection of human facial images divided into two categories: fake and real. The total data set used in this study was 10,826 images, consisting of 5,413 fake images and 5,413 real images.

3. Methods

This study uses the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, which is widely applicable in data science, particularly in areas such as data mining, Artificial Intelligence (AI), machine learning, deep learning, big data, and data analytics. The advantage of this methodology lies in its simplicity and structured approach, as all its phases are clearly identified, well organized, and not overly complex, making it easy to understand. CRISP-DM consists of six main phases which include business understanding, data understanding, data preparation, modeling, evaluation, and deployment [20]. In this study, the research process was conducted only up to the evaluation phase and did not proceed to the deployment or implementation of the model in an application. The research stages are shown in Fig. 1 below.

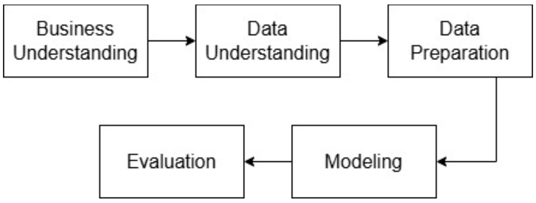

Figure 1. Research Stages

Fig. 1 shown research stages consists:

3.1 Business Understanding

This stage defines the research objectives and understands the business needs that underlie the research. Analysis is conducted to identify key issues, such as the high prevalence of deepfake content, and the role of technology in addressing these issues.

3.2 Data Understanding

At this stage, the collected dataset is explored. Analysis is performed on the dataset, including identifying the number of data points, categories (real and fake), data distribution, and data types.

3.3 Data Preparation

At this stage, data preprocessing is performed to ensure the dataset is ready for application to the model. This process includes image resizing and data augmentation.

3.4 Modeling

At this stage, a transfer learning model is created and trained using the processed dataset. The transfer learning model is built using pre-trained models to improve training accuracy and efficiency.

3.5 Evaluation

In this phase, the model is evaluated using metrics such as the confusion matrix, accuracy, precision, recall, and F1-score. The evaluation is performed on a testing set of 2,164 images, comprising 1,082 images labeled as fake (0) and 1,082 images labeled as real.

4. Result and Discussion

4.1 Research Dataset

This study was conducted using VGG16, VGG19, and ResNet50 approaches for deepfake image detection. The dataset used in this research was obtained from Kaggle and consists of deepfake and real images. The dataset consists of 10,826 facial images with a resolution of 256 × 256 pixels format jpg. It was accessed in January 2025. The dataset is publicly available [21]. It contains a collection of human facial images categorized into two classes, fake and real. The total number of images used in this study is 10,826, comprising 5,413 fake images and 5,413 real images. The fake images were generated using deepfake techniques, while the real images are authentic. Several example images from the dataset used in this study are presented in Fig. 2.

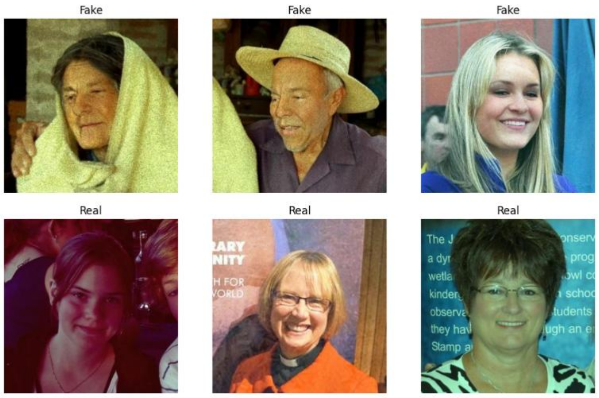

Figure 2. Dataset Images

Volume 01, Issue 02, December 2025 : 47-53 Received: Nopember 03, 2025. Revised: December 26, 2025.

IJEIE International Journal of Electrical and Intelligent Engineering

In this Fig. 2, the data were divided into two subsets, training data and testing data with a ratio of 80:20, which means 80% was allocated for training data and 20% for testing data. The total number of training samples was 8,662, while the total number of testing samples was 2,164.

4.2 Parameter Value Settings

The selection and tuning of training parameters are critical factors that influence the ultimate performance of the model during the training process. Essential parameters, including the number of epochs, batch size, activation functions, and optimization algorithms, must be carefully adjusted to optimize the training procedure.

All experiments were conducted using identical training configurations to ensure a fair comparison among models. The models were trained for 25 epochs with a batch size of 16 using the Adam optimizer with a learning rate of 0.0001. The binary cross-entropy loss function was used due to the binary classification nature of the task.

A validation split of 10% from the training set was applied to monitor model generalization during training. Early stopping was implemented with a patience of 5 epochs to prevent overfitting. No learning rate scheduler was applied.

The dataset was divided using a stratified random splitting strategy with a fixed random seed of 42 to ensure reproducibility and balanced class distribution across training and testing sets.

In the transfer learning configuration, the convolutional base layers of VGG16, VGG19, and ResNet50 were fully frozen during training, and only the newly added fully connected layers were trained. No additional fine-tuning of deeper layers was performed. The configurations of these parameters for the model architectures utilized in this study are summarized in Table 1.

Table 1. Configuration of Model Parameter

Model	Number of Epochs	Batch Size	Activation	Optimizer
VGG16	25	16	ReLU	Adam
VGG19	25	16	ReLU	Adam
ResNet50	25	16	ReLU	Adam

As presented in Table 1, all three models utilized consistent parameter settings to ensure uniformity during the training process, thereby facilitating a fair and objective comparison of their performance. The research stages: The research methodology consists of several stages, including business understanding, data understanding, data preparation, modeling, and evaluation. Each stage is described in detail in this section.

4.2.1 Business Understanding

This phase focuses on understanding the business objectives and the problems to be solved. It involves analyzing the business requirements of the study, including identifying needs and defining the research objectives

4.2.2 Data Understanding

In this phase, relevant data are collected and evaluated to ensure their quality. An initial analysis is also conducted to understand patterns and extract key information from the data. In this study, the data were collected from Kaggle, specifically the Deepfake and Real Images dataset, which consists of human face images in jpg format with a resolution of 256 × 256 pixels per image. The total number of images used is 10,826, comprising 5,413 fake images generated through deepfake manipulation techniques and 5,413 real images.

4.2.3 Data Preparation

In this phase, the data are processed and prepared for model development through data preprocessing. The preprocessing steps include resizing, normalization, data augmentation, labeling, and data splitting, augmentation.

a. Resizing

Resizing is the process of adjusting image dimensions to a specified size in order to ensure uniformity and compatibility with the model's requirements. In this study, images were resized to 224 × 224 pixels to match the input size expected by the pre-trained model architecture. This process aims to ensure data compatibility and consistency, as well as to optimize model performance during training.

b. Normalization

The next step is normalization, which involves scaling the pixel values of image data to a specific. Typically, image pixels have values ranging from 0 to 255. In this normalization process, the pixel values are rescaled to a range between 0 and 1 to accelerate the model training process. The images before and after normalization are shown in the Fig. 2.

Figure 3. Differences between Images Before and After Normalization

Fig. 2 shows that normalization changes pixel values from the original 0–255 scale to a 0–1 range. Before normalization, color and brightness differences appear more distinct, while after normalization the image looks smoother and slightly less contrasty. This process makes pixel values more uniform, helping machine learning models process data more consistently, reduce sensitivity to lighting variations, and improve training efficiency.

c. Data Augmentation

Data augmentation is a technique used to increase the size of the training dataset by applying transformations such as image shifting and flipping. This technique is implemented to enhance the diversity of the dataset, allowing the model to learn from a broader range of variations without the need to manually collect additional data. An example of the image results that have gone through the data augmentation process can be seen in Fig. 3.

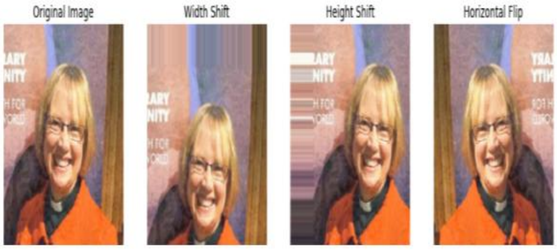

Figure 3. Data Augmentation Process Results

Fig.3 above is an example of the results of the data augmentation process carried out using the width shift, height shift, and horizontal flip techniques.

d. Labeling

Labeling is the process of assigning labels or categories to data so that it can be used to train a machine learning model. In the context of image classification, labeling involves indicating the class to which each image belongs. In this study, the images were labeled with 0 for fake data and 1 for real data.

e. Data Split

The next step is data splitting, which is the process of dividing the dataset into subsets such as training data and testing data. This step aims to objectively evaluate the model's performance. The training data are used to build and train the model so that it can learn patterns within the data, while the testing data are used to assess the model’s ability to generalize to unseen data.

4.2.4 Modeling

In this phase, models for deepfake image detection are developed using a transfer learning approach. The transfer learning models used in this study are VGG16, VGG19, and ResNet50. All three models were trained on the dataset for 25 epochs.

a. VGG16 Model

VGG16 is a convolutional neural network (CNN) architecture used for image recognition, consisting of16 layers. The VGG16 model includes convolutional layers that have been pre-trained to recognize basic image patterns such as edges and textures. The pre-trained VGG16 model is used as the initial part of the transfer learning architecture by retaining its convolutional and pooling layers. The features extracted by VGG16 are then passed to fully connected layers to adapt the model for the deepfake image detection task. Finally, an output layer provides the classification result indicating whether an image is a deepfake or not.

This model is built using a transfer learning approach with the VGG16 architecture pre-trained on the ImageNet dataset. The fully connected layers of VGG16 are removed, and the pre-trained weights are frozen to preserve the learned visual features. Additional layers include a flatten layer, a dense layer with ReLU activation, dropout to prevent overfitting, and an output layer with sigmoid activation for binary classification. The input images have a size of 224 × 224 pixels with 3 color channels (RGB).

b. VGG19 Model

The architecture of the VGG19 model is similar to that of the VGG16 model. In the VGG19 architecture, the fully connected layers are removed, and all pre-trained weights are frozen to preserve the general feature extraction capabilities learned from the ImageNet dataset. After feature extraction through VGG19, the image data pass through a flatten layer, followed by a dense layer to recognize important patterns before being sent to the output layer. A dropout layer is also included to reduce the risk of overfitting.

c. ResNet50 Model

This model features an input layer that accepts images of size 224 × 224 pixels with 3 color channels (RGB). The main component of the model is ResNet50, a deep convolutional neural network consisting of 50 layers, pre-trained on the ImageNet dataset. In this model, ResNet50's weights are frozen, meaning they remain unchanged during fine-tuning. The output from ResNet50 is aggregated using global average pooling, then passed through a dense layer with sigmoid activation for classification.

4.2.5 Evaluation

The next step is model evaluation, which involves measuring the performance of the trained model using previously unseen data, specifically the testing dataset.

The evaluation of the model was conducted to assess its performance on previously unseen data, specifically the testing dataset. This evaluation utilized several metrics, including the confusion matrix, accuracy, precision, recall, and F1-score. The testing dataset consisted of 2,164 images, comprising 1,082 fake images (label 0) and 1,082 real images (label 1). The balanced class distribution ensures that accuracy, precision, recall, and F1-score are not biased toward a dominant class. The evaluation results are presented in Tables 2 and 3.

Table 2. Model Performance Evaluation Based on Confusion Matrix

Model	**TP**	**FP**	**FN**	**TN**
VGG16	**882**	**200**	**325**	**757**
VGG19	839	243	356	726
ResNet50	822	260	640	442

Table 3. Model Performance Evaluation Based on Accuracy, Precision, Recall, and F1-score

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
VGG16	**0.76**	**0.82**	**0.73**	**0.77**
VGG19	0.72	0.78	0.70	0.74
ResNet50	0.58	0.76	0.56	0.65

The confusion matrix results presented in Table 2 indicate that the VGG16 model achieved relatively high counts of True Positives (TP) and True Negatives (TN) compared to the other two models, demonstrating superior capability in detecting deepfake images. The VGG19 model exhibited slightly lower performance, with increased counts of False Positives (FP) and False Negatives (FN). In contrast, ResNet50, despite being a deeper and more complex architecture, showed significantly decreased performance. Its high FN and FP values suggest frequent failures in detecting deepfake images. The elevated FN rate for ResNet50 implies difficulty in recognizing manipulation patterns within deepfake images. This issue may be attributed to overfitting during training or suboptimal adaptation of the pretrained architecture to the characteristics of the dataset used in this study.

As shown in Table 3, the VGG16 model demonstrated the best performance compared to VGG19 and ResNet50, achieving the highest values in accuracy, precision, recall, and F1-score. VGG16 effectively identified both authentic and deepfake images, as evidenced by its high counts of True Positives and True Negatives. Conversely, ResNet50 exhibited the lowest performance, with the lowest accuracy and recall values, alongside a notably high number of False Negatives. This suggests that ResNet50 is less effective in detecting deepfake images, potentially due to its complex architecture not being fully compatible with the data characteristics or insufficient fine-tuning during the training process.

In general, these results indicate that VGG16 is the most effective model for deepfake image classification on the dataset used in this study. The advantage of VGG16 may be attributed to the balance between model complexity and generalization capability. Its relatively simpler network architecture likely provides an advantage in handling images with distinctive textures and patterns, such as manipulated facial images. Conversely, the poor performance of ResNet50 suggests that model success depends not only on architectural complexity but also on how well the model is adapted to the characteristics of the data. Overly deep models without comprehensive fine-tuning or appropriate regularization may experience a decline in performance.

Although previous studies reported accuracy above 90% when using VGG16 for deepfake detection, the best accuracy achieved in this study was 0.76. Several factors may explain this difference.

First, the Kaggle dataset used in this study contain more challenging deepfake samples with higher visual realism, making manipulation patterns harder to detect. Second, the dataset splitting strategy was performed randomly but strictly separated training and testing samples, which may reduce the possibility of data leakage and thus result in lower but more realistic performance.

Third, the convolutional base layers were fully frozen during training. While this approach reduces computational cost, it may limit the model's ability to adapt to subtle deepfake-specific artifacts. Previous studies that achieved higher accuracy often applied partial fine-tuning on deeper convolutional layers, allowing better adaptation to dataset characteristics.

Additionally, differences in preprocessing strategies and augmentation techniques may significantly affect performance. The relatively moderate accuracy obtained in this study suggests that deeper architectures such as ResNet50 require more extensive fine-tuning and hyperparameter optimization to perform effectively on this dataset.

5. Conclusions

Based on the evaluation results of the three models, namely VGG16, VGG19, and ResNet50, it can be concluded that the VGG16 model demonstrated the best performance in classifying authentic and manipulated deepfake facial images. This is evidenced by its higher evaluation metric scores compared to the other two models. VGG16 exhibited strong generalization capabilities and effectively recognized distinctive visual patterns in deepfake-manipulated images. Conversely, ResNet50 showed the lowest performance, indicating that despite being a complex and deep model, its performance is suboptimal without adequate fine-tuning tailored to the specific characteristics of the dataset. This finding suggests that architectural complexity does not necessarily correlate directly with accuracy.

For future research, it is recommended that more extensive fine-tuning be applied to complex models such as ResNet50 in order to better align them with the specific characteristics of the data. Exploring alternative neural network architectures may yield improved performance. Enhancing the diversity of the training dataset is essential to enable the model to recognize a wider variety of patterns. Moreover, the application of advanced data augmentation techniques together with additional regularization strategies could help mitigate overfitting and further improve model accuracy.

References

[1] D. Purnamasari, D. Herlinudinkhaji, and Z. Mauludin, "Analisis Kualitas Citra Foveal Avascular Zone (FAZ) Dengan Teknik Kombinasi Pengacakan Piksel Jurnal Pepadun," vol. 4, no. 3, pp. 325–333, 2023.

[2] D.Purnamasari, D.Herlinudinkhaji, and A.K.Dewi, "Foveal Avascular Zone (FAZ) Image Encryption Using Pixel Scrambling Combination Technique for Medical Image Security," *Infotel*, vol. 16, no. 1, 2024, [Online]. Available: https://ejournal.ittelkom-pwt.ac.id/index.php/infotel/article/view/1029

[3] D. Purnamasari and N. Erwanti, "Enkripsi Citra Fovea Avascular Zone (FAZ) Menggunakan Kriptografi Vigener Cipher," vol. 9, no. September, pp. 114–121, 2022.

[4] M. R. Shoaib, Z. Wang, M. T. Ahvanooey, and J. Zhao, "Deepfakes , Misinformation , and Disinformation in the Era of Frontier AI , Generative AI , and Large AI Models," pp. 1–8, 2023.

[5] Regina Angelika Septi Rahayu;Handri Santoso, "ANALISIS GAMBAR WAJAH PALSU : MENDETEKSI KEASLIAN GAMBAR YANG DIMANIPULASI MENGGUNAKAN METODE VARIATIONAL AUTOENCODER DAN FORENSICS DEEP NEURAL NETWORK ANALYSIS OF FAKE FACE IMAGES : DETECTING THE AUTHENTICITY OF MANIPULATED IMAGES USING VARIATIONAL AUTOE," vol. 2, no. 9, pp. 2701–2726, 2023.

[6] C. Gilbert and M. A. Gilbert, "Navigating the Dual Nature of Deepfakes : Ethical , Legal ,

and Technological Perspectives on Generative Artificial Intelligence AI) Technology," vol. 3, no. 10, 2024.

[7] N. Misirlis and H. Bin Munawar, "FROM DEEPFAKE TO DEEP-USEFUL : RISKS AND OPPORTUNITIES THROUGH A SYSTEMATIC LITERATURE REVIEW," pp. 26–32, 2022.

[8] M. Indra, I. Nurtanio, and A. Achmad, "Deepfake detection in videos using Long Short-Term Memory and CNN ResNext," vol. 14, no. 3, pp. 178–185, 2022.

[9] A. Heidari, N. J. Navimipour, and M. Unal, "Deepfake detection using deep learning methods : A systematic and comprehensive review," no. August 2022, pp. 1–45, 2024, doi: 10.1002/widm.1520.

[10] S. M. Qureshi, A. Saeed, S. H. Almotiri, F. Ahmad, and M. A. Al Ghamdi, "Deepfake forensics : a survey of digital forensic methods for multimodal deepfake identification on social media," pp. 1–40, 2024, doi: 10.7717/peerj-cs.2037.

[11] A. Hatem, S. Omnia, S. Hala, T. R. Ragab, and S. Mohsen, "Deepfake detection using convolutional vision transformers and convolutional neural networks," *Neural Comput. Appl.*, vol. 36, no. 31, pp. 19759–19775, 2024, doi: 10.1007/s00521-024-10181-7.

[12] S. Suratkar and F. Kazi, "Deep Fake Video Detection Using Transfer Learning Approach," *Arab. J. Sci. Eng.*, 2022, doi: 10.1007/s13369-022-07321-3.

[13] L. Boongasame, J. Boonpluk, S. Soponmanee, J. Muangprathub, and K. Thammarak, "Design and Implement Deepfake Video Detection Using VGG-16 and Long Short-Term Memory," vol. 2024, 2024, doi: 10.1155/2024/8729440.

[14] V. Rajakumareswaran, S. Raguvaran, V. Chandrasekar, S. Rajkumar, and V. Arun, "Erode Sengunthar Engineering College , Tamil Nadu , Thuduppathi , India Sona College of Technology , Salem , India DEEPFAKE DETECTION USING TRANSFER LEARNING-BASED XCEPTION MODEL," pp. 89–98, 2024.

[15] M. Kaur, S. Singh, and M. Kaur, "Computational Image Encryption Techniques: A Comprehensive Review," *Math. Probl. Eng.*, vol. 2021, no. i, 2021, doi: 10.1155/2021/5012496.

[16] A. V. Singh, D. Moghe, and K. Meenakshi, "Deepfake detection using fine-tuned VGG16 model : A transfer learning approach," pp. 825–829, 2025, doi: 10.1201/9781003559085-141.

[17] F. Iqbal and U. A. Emirates, "Data Augmentation-based Novel Deep Learning Method for Deepfaked Images Detection Data Augmentation-based Novel Deep Learning Method for," vol. 20, no. 11, 2026, doi: 10.1145/3592615.

[18] D. S. M. V, H. J. Vidyarani, G. Krishna, and S. Mohan, "Deep Fake Detection," vol. 12, no. 9, pp. 298–305, 2024.

[19] B. Halima *et al.*, "CRISP-MED-DM a Methodology of Diagnosing Breast Cancer," vol. 9, no. 1, pp. 709–721, 2024.

[20] A. Rianti, N. Wachid, A. Majid, and A. Fauzi, "CRISP-DM : Metodologi Proyek Data Science," pp. 107–114, 2023.

[21] M. Karki, "deepfake and real images dataset," Kaggle. Accessed: Jan. 08, 2025. [Online]. Available: https://www.kaggle.com/datasets/manjilkarki/deepfake-and-real-images