Journal, Advances in Mathematical & Computational Sciences
Vol. 14 No. 1, March 2026 Series
www.isteams.net/mathematics-computationaljournal

Journal of Advances in Mathematical & Computational Sciences
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA

Available online at https://www.isteams.net/ mathematics-computationaljournal.info
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

Deepfake Image Detection Using ResNet-50: A Systematic Literature Review

[1]Ojeniyi J. A., [2]Yisa R. N., [3]Subairu S. O., [4]Abdulhamid S. M., [5]Alhassan, J.K., & [6]Noel, M.D
[1-2.,3,4 & 6] Department of Cyber Security, Federal University of Technology, Minna, Nigeria
[5] Department of Computer Science, Federal University of Technology, Minna, Nigeria
E-mails; [1]ojeniyija@futminna.edu.ng, [2]rhodannabayisa@gmail.com, [3]islam4life@futminna.edu.ng, [4]shafii.abdulhamid@futminna.edu.ng, [5]jkalhassan@futminna.edu.ng, [6]moses.noel@futminna.edu.ng.

ABSTRACT

With the rapid proliferation of deepfake technologies, the generation and propagation of manipulated images pose significant threats to digital security, privacy, and public trust. Convolutional Neural Networks (CNNs), particularly ResNet-50, have gained attention for their robust feature extraction capabilities in deepfake detection tasks. This systematic literature review aims to synthesize recent research efforts that leverage ResNet-50 for the detection of deepfake images. Following PRISMA guidelines, we analyzed peer-reviewed articles published between 2020 and 2025 across major scientific databases including IEEE Xplore, Science Direct, SpringerLink, and ACM Digital Library. Our review explores how ResNet-50 is employed either as a standalone classifier, a transfer learning backbone, or part of hybrid architectures and evaluates its performance across various datasets such as, Flickr / FFHQ FaceForensics++, Celeb-DF, and DFDC. We identify key limitations in current approaches, particularly in terms of generalization to unseen manipulations, dataset biases, and adversarial robustness. The findings indicate that while ResNet-50 demonstrates competitive accuracy and computational efficiency, challenges remain in ensuring model reliability in real-world applications. This review provides a foundational reference for future work aimed at enhancing deepfake image detection systems using deep learning.

Keywords: Deepfake image Detection, Machine Learning, Deep Learning, ResNet-50

Ojeniyi J. A., Yisa R. N., Subairu S. O., Abdulhamid S. M., Alhassan, J.K., & Noel, M.D (2026): Deepfake Image Detection Using ResNet-50: A Systematic Literature Review. Journal of Advances in Mathematical & Computational Science. Vol. 14, No. 1. Pp 113-146. Available online at www.isteams.net/mathematics-computationaljournal. dx.doi.org/10.22624/AIMS/MATHS/V14N1P8

1. INTRODUCTION

Over the past few years, the introduction and rapid advancement of deepfake technology have raised significant concerns across media, security, and social sectors globally, characterized with an annual growth rate of 67% (Sandotra and Arora.,2023; Hanush et al., 2024) . Deepfake is the process of capturing a person's lip, eye, and facial movements and superimposing them on a different background to simulate that person in a fake situation (Samal et al., 2024).

Deepfakes are digitally generated media where one person's image is swapped with another's. While this technology has valid uses, like producing convincing visual effects for entertainment, it has also been employed to generate false news and spread misleading information (Patel and Jain, 2024). These manipulated images and videos can realistically change faces, expressions, or even create entirely false identities. This poses serious threats to credibility, personal privacy, security, political stability, and public trust. Because these sophisticated forgeries are becoming easier to accessible and more convincing, it is now critically important to develop good techniques/ strategies to detect them. (Abdullah et al., 2024).

Among various detection strategies, deep learning approaches especially convolutional neural networks (CNNs) have shown remarkable promises to learn fine-grained visual features, detect inconsistencies typical in synthetic images and scale well with data and generalize across manipulation types (Kroiß & Reschke, 2025). One of the most frequently used CNN architectures in image analysis tasks is ResNet-50, a 50-layer residual network known for its balance of depth and computational efficiency. ResNet-50 has been employed across numerous studies for the detection of manipulated media, leveraging its ability to learn variations between real and fake images (Kroiß & Reschke, 2025).

However, despite growing research in this area, there is a lack of consolidated understanding regarding the effectiveness of ResNet-50 in the context of deepfake image detection. Studies differ in terms of datasets used, preprocessing techniques, evaluation metrics, and types of manipulations considered. These differences make it difficult to draw clear conclusions about the model's generalizability and real-world applicability.To address this gap, this study presents a systematic literature review (SLR) focused on deepfake image detection using ResNet-50. The objective is to evaluate and synthesize existing research to identify trends, challenges, and best practices.

By systematically analyzing peer-reviewed studies, our contributions are to provide a comprehensive overview of:

- Deepfake generated methods
- Deepfake image detection methods
- CNN models for deepfake image detection
- The performance of ResNet-50 in various deepfake image detection scenarios,
- The types of datasets, evaluation metrics used and performance compared to other model,
- Limitations, open issues, and future directions.

This review serves as a valuable resource for researchers, practitioners, and policymakers interested in advancing the development of robust, interpretable, and scalable deepfake detection systems.
The remainder of the paper is organized as follows: Section II presents the review procedure by defining interest research questions. In Section III, we thoroughly discuss the findings from different studies. Finally, Section VI concludes the paper.

2. PROCESS OF SLR

We utilize their methods in our systematic literature review (SLR) and organize the review procedure into three key phases, depicted in Figure 1. This is done to identify, evaluate, and understand various research related to particular research questions.

- **Planning the Review:** This initial phase aims to (a) determine the necessity for the review, (b) establish the standards and methods for conducting it, and (c) assess the suitability of those standards and methods for this specific systematic literature review.
- **Conducting the Review:** Following the guiding principles outlined in (Rana et al., 2022), this phase comprises six essential steps.

i. Research Questions (RQs): The purpose of defining research questions is to pinpoint the relevant studies that should be included in the current review. We establish a series of research questions (detailed later) specifically focusing on the field of Deepfakes.
ii. Search strategy (SS): A planned search approach is used to identify a broad range of relevant studies related to our research questions. We aim to develop an objective and comprehensive search strategy to capture as much pertinent literature as possible.
iii. Study Selection Criteria (SSC): Selecting relevant literature can be challenging, due to potential biases such as the language of publication, familiarity with the authors, institutions, journals, or publication year (Rana et al., 2022). Before defining selection criteria, we carefully consider factors to ensure objectivity in choosing primary studies that offer meaningful evidence related to our research questions.
iv. Quality Assessment Criteria (QAC): Assessing the quality of each selected study is essential to ensure the findings are reliable and impartial. We develop a set of quality criteria for evaluating the individual studies.
v. Data extraction and monitoring (DEM): We carefully plan how to collect the necessary data from the chosen studies and document the corresponding evidence.
vi. Data Synthesis (DS): Data synthesis involves organizing and summarizing the findings from the selected studies. We use a structured set of methods to effectively integrate the information.

Reporting the Review. After completing the review of all the studies, we present the results in an appropriate format for distribution and the intended audience.

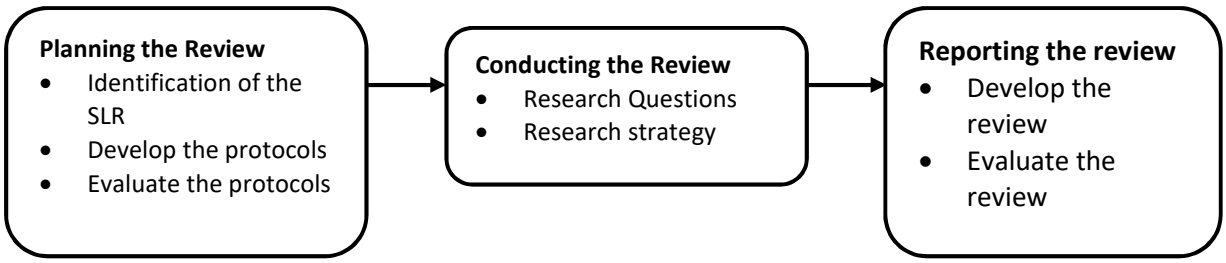

Fig. 1: The process of SLR

Research Questions (RQs)

Choosing research questions (RQs) is a crucial initial step in outlining a study's primary goal and anticipated results. We formulate our RQs to ensure they are valuable to investigators; as well-chosen questions can foster greater certainty in a specific area. To understand recent work in Deepfake image identification utilizing ResNet-50, we've developed five key questions (RQ 1-5) detailed in Table 1. First (RQ-1), we examine how ResNet-50 is applied to identify deepfake images. Second (RQ-2), we explore the datasets used alongside ResNet-50 for deepfake image detection. Third (RQ-3), we consider the performance metrics employed to evaluate the effectiveness of deepfake image detection models using ResNet-50. Fourth (RQ-4), we further examine the performance of ResNet-50 in comparison to alternative CNN architectures. Finally, (RQ-5), we assess the shortcomings and difficulties associated with using ResNet-50.

TABLE 1. Define research questions for the SLR.

QID	Research Questions	Objectives
RQ1	How is ResNet-50 applied for deepfake image detection?	To utilize the ResNet-50 deep CNN for detection and classification of deepfake image
RQ2	What datasets are commonly used with ResNet-50 deepfake image detection?	To identify and utilize benchmark for datasets commonly paired with ResNet-50 for training and evaluating deepfake image detection models.
RQ 3	What evaluation metric are used to computing the performance of deepfake image detection model performance using ResNet-50 utilizing the ResNet-50 architecture	To identify the evaluation metrics commonlyused to assess the of deepfake image detection models
RQ4	What is the performance compared to other models?	To evaluate the performance of ResNet-50 in deepfake image detection by comparing its accuracy, precision, recall, F1-score, and AUC with other state-of-the-art models such as VGG-19, DenseNet, MesoNet, and MobileNet.
RQ5	What are the limitations and challenges in using ResNet-50 for this task?	To analyze the limitations and challenges associated with using ResNet-50 for deepfake image detection

Search Strategy (SS)

To conduct a thorough investigation, we consulted multiple databases rather than limiting ourselves to a few sources. While numerous online repositories offer research articles, we prioritized prominent platforms based on their relevance and accessibility, as detailed below:

- Web of Science
- IEEE Xplore
- Digital Library ACM

- Digital Library, Science Direct (ELSEVIER)
- SpringerLink
- Google Scholar
- Semantic Scholar
- ResaerchGate
- Arxiv
- Wiley online Library

The repositories include journals, conferences, and archives. We limit our search duration from January 2020 to December 2025.

Search Keywords and Terms

The search keywords and terms for this SLR are as follows: deepfake/ image forgery, detection/ classification and ResNet-50/ deep learning.

- Deepfake / Image Forgery: "deepfake" "image forgery" "digital image manipulation" "face manipulation" "image tampering" "fake image detection" "synthetic face detection" "AI-generated images" "face swapping"
- Detection / Classification: "detection" "classification" "identification" "recognition" "fake image detection" "forgery detection" "image analysis"
- ResNet-50 / Deep Learning: "ResNet-50" "residual neural network" "ResNet" "deep convolutional neural network" "CNN" "transfer learning" "deep learning" "pretrained CNN" "fine-tuned ResNet-50"

Boolean search

We intended to collect as many works as possible that are relevant to our research questions. During collecting Deepfake image detection using ResNet-50 studies, we tried to include all the combinations of related search phrases or keywords to avoid any bias. The key idea of using Boolean terminology for combining those searching terms with AND or OR. The search words can be outlined primarily (“Image Forgery” OR “DeepFake detection” OR “Deep fakes” OR “Deepfakes” OR “Deepfake Images” OR “Faceswap” OR (“ai generated image” OR “manipulated image” OR “generative image” OR “fabricated image” OR “digital image manipulation" OR "face manipulation" OR "image tampering" OR "fake image detection" OR "synthetic face detection" "AI-generated images" OR "face swapping" AND"detection" OR "classification" OR "identification" AND "recognition" OR "fake image detection" OR "forgery detection" OR "image analysis" OR "ResNet-50" OR "residual neural network" OR "ResNet" OR "deep convolutional neural network" OR "CNN" OR "transfer learning" OR "deep learning" OR "pretrained CNN" OR "fine-tuned ResNet-50").

Fig. 2: Word Clouds analysis regarding Deepfake Image Detection Using ResNet-50

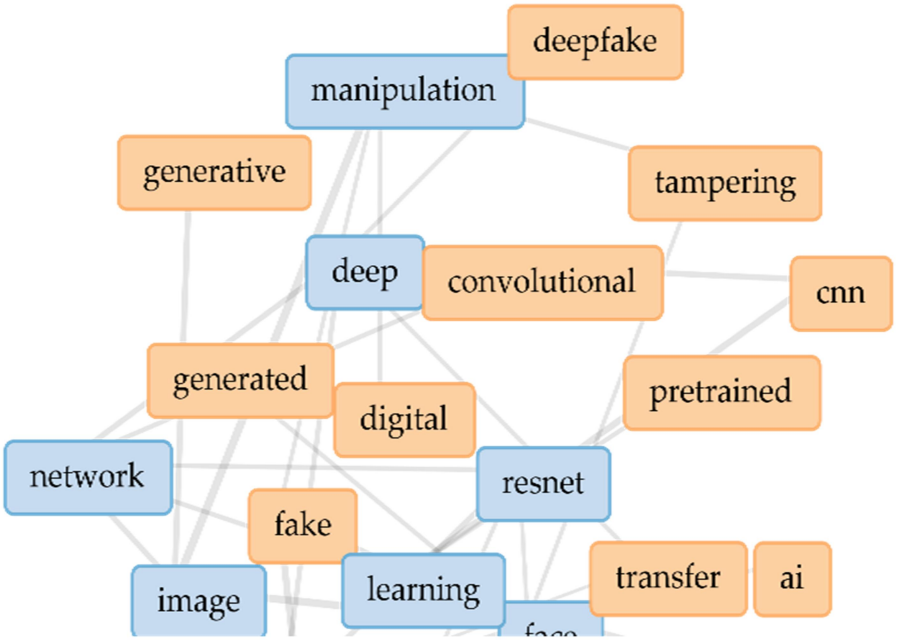

Fig. 3 Text analysis visualization of search string generated using the VOYANT tools

Study Selection Process

The evidence gathering process is documented in a PRISMA flowchart, as depicted in Figure 4, which outlines the data selection and evaluation stages.

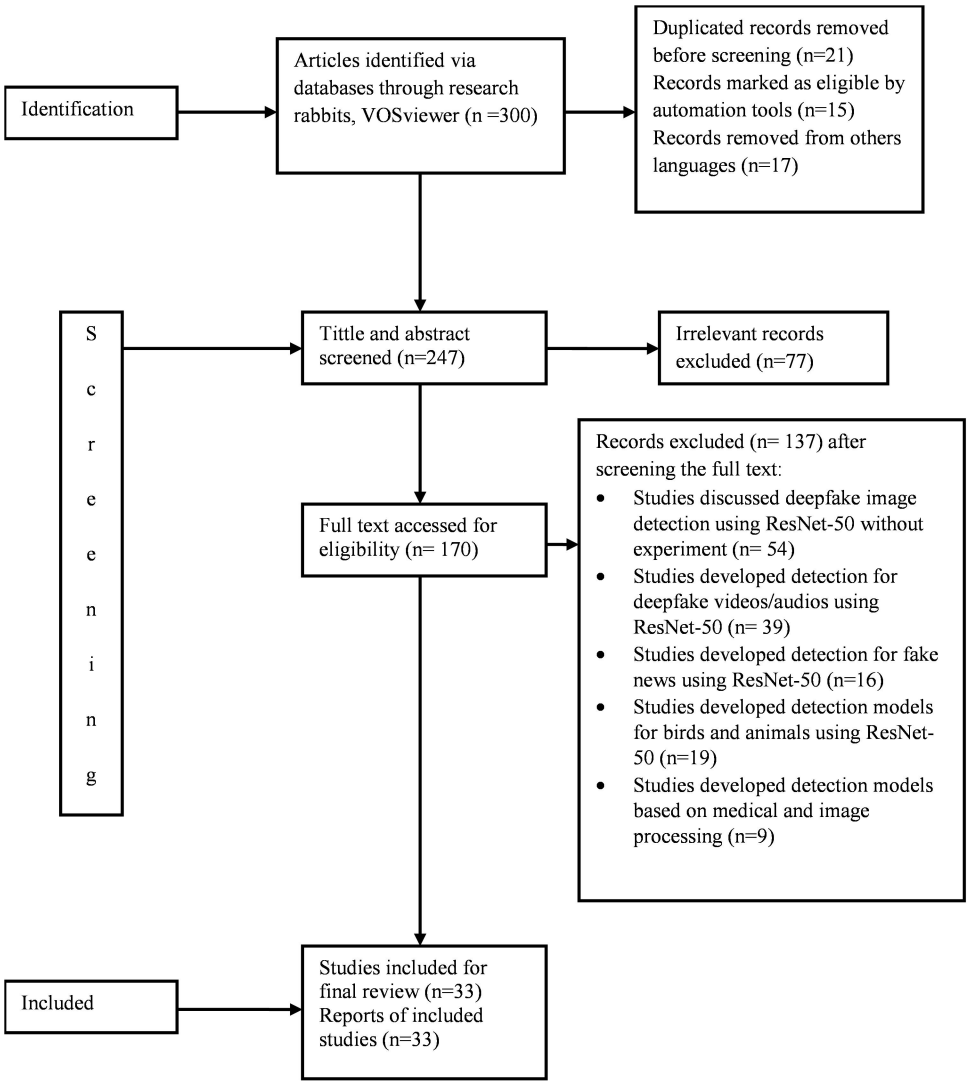

Fig. 4. PRISMA flow diagram reporting the data selection and screening process.

Our search, conducted in July 2025, initially yielded 300 records from the specified databases. The PICO framework (Pico, 2022) was used during the selection process to refine the research questions and strategies for locating pertinent studies. After removing redundancies and non-English publications, 247 records remained. Subsequent screening of abstracts, titles, and full texts resulted in the exclusion of 170 records deemed irrelevant. The final evaluation identified 45 studies that met the defined criteria and were included in the final literature analysis.

Inclusion and exclusion criteria

We considered different inclusion and exclusion criteria to gather relevant studies for our systematic review. The criteria used to select articles were as follows: **Inclusioncriteria** (i) published articles/ journals/conference papers between 2020 to 2025 (ii) articles that proposed and developed deepfake image detection CNN (ResNet-50), methods, effectiveness of ResNet-50, modified ResNet-50, dataset and metrics to identify the deepfake image detection; and (iii) articles related to research questions and studies that focus on deepfakes image detection using ResNet-50 only. **Exclusion criteria**articles were as follows: (i) duplicate articles found via different databases; (ii) articles written in other languages; (iii) articles that do not consider deepfakes image detection using ResNet-50 (iv) irrelevant studies (video only or audio only deepfake studies) and (v) survey articles. Using this strategy, we ensured that all related studies were included and that inapplicable studies were excluded from the system article literature review.

Quality Evaluation

Assessing the strength of the evidence in a systematic review is just as crucial as analyzing the information itself. Problems in research design can skew the results of weaker studies, and this should be acknowledged carefully. Studies with flawed methods or limitations that could affect the accuracy of the results should be identified in, or removed from, literature reviews. Choosing the right criteria to assess the trustworthiness of information and possible problems in each study is also essential. Following established standards like those in the PICO Portal (Pico, 2022), we used such criteria to verify the chosen studies and to examine them for relevant information in this systematic review.

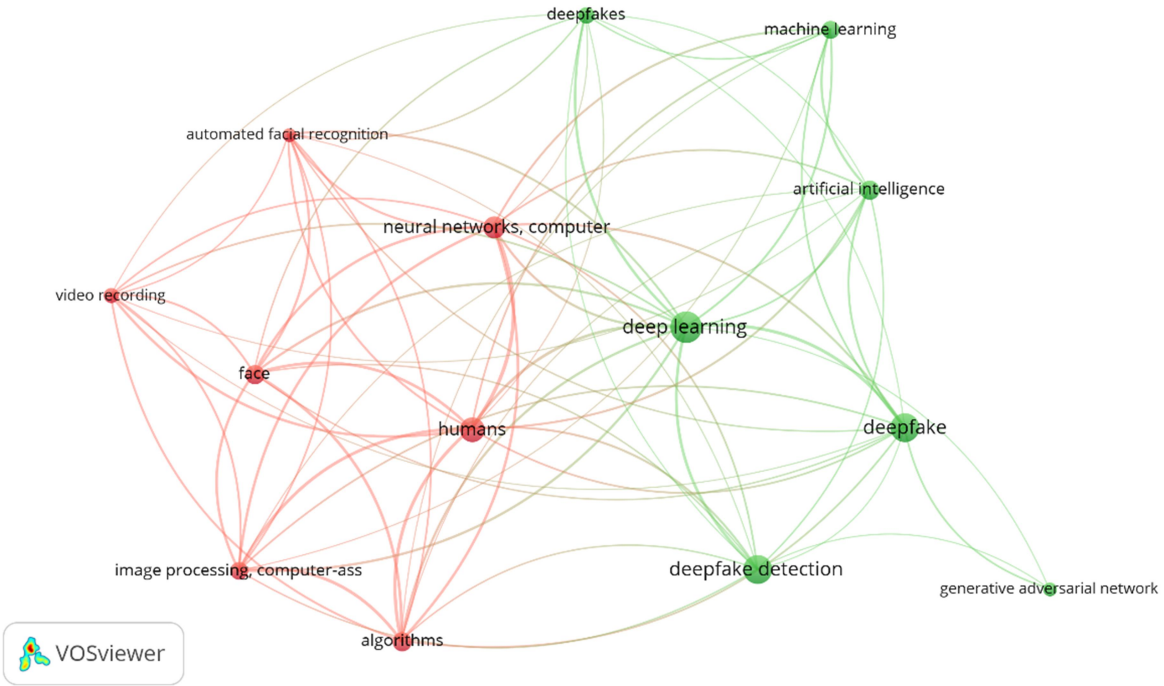

Fig. 5. Keyword co-occurrences visualization analysis of eligible studies using the VOSviewer tool.

Moreover, we employed a confirmation process to judge the articles included, ensuring consistency across different sets of information. After evaluating the quality, we found 45 relevant studies discussing deepfake image detection using ResNet-50. Figure 3 shows the analysis of how frequently different keywords appear together in the relevant studies for our systematic literature review.

3. DEEPFAKE GENERATION METHODS

Deep learning generation methods have dramatically transformed the deepfake landscape, enabling the creation of highly convincing fabricated media, techniques like Autoencoders, VAE, GAN, Diffusion models (Heo et al., 2021). Fig. 2 illustrates taxonomy of deepfake generation.

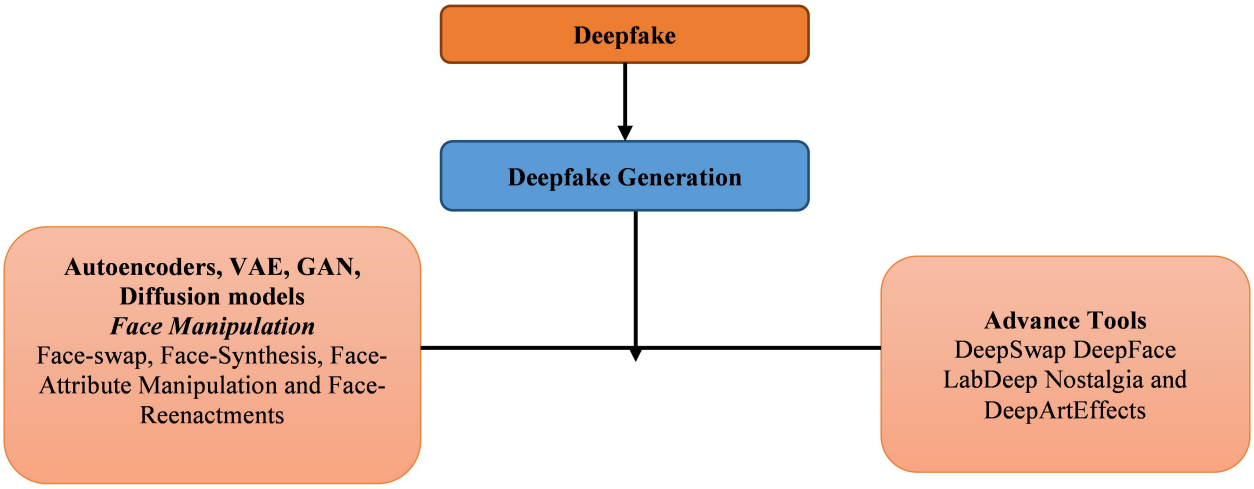

Fig. 6. Taxonomy of Deepfake Generation Methods

Autoencoders: It consists of two main components: an encoder and a decoder. The encoder analyzes digital media to identify key characteristics and compress them into a simplified representation (a code that removes irrelevant details). The decoder then uses this representation to reconstruct the original media. By training the encoder and decoder on a varied dataset of authentic and manipulated videos, the autoencoder learns to apply the features of one image to another. This makes it useful for face replacement. However, a significant limitation is that the generated image may be easily detected by humans as artificial. (Kaushik et al., 2025).

Generative Adversarial Networks: They tools are made of two parts– the Generator (G) and Discriminator (D). The generator is used to produce novel data samples (fake data), and the discriminator evaluates them against the existing data to draw comparison and conclude whether the content is fake or real. Generative Adversarial Network (GAN) techniques for generating paraphrases are often centered around facial imagery.

These methods generally fall into two main types. First is “Entire face synthesis” involves constructing a face entirely from scratch and second is “Attribute Manipulation” modifies existing faces by changing features like hair color or adding accessories. These methods generate very convincing fake images, but they need substantial amounts of training data and tuning. However, they are limited to a single image domain and do not support multimodal outputs (Kaushik et al., 2025).

- **Variational Autoencoders (VAE):** Variational autoencoders (VAEs) build upon conventional autoencoders by representing the compressed information as a probability distribution, typically Gaussian. Like their predecessors, they use encoding and decoding processes and leverage neural networks to capture intricate, non-linear connections and probabilistic models. The strength and adaptability of VAEs stem from their capacity to integrate the advantages of both approaches. While they may produce less realistic outputs, they excel at generating a wider range of diverse results compared to typical autoencoders. GANs (Abdullah et al., 2024).
- **Diffusion Models:** These models use a step-by-step noise reduction method. They repeatedly refine an initial, noisy version of the input, gradually removing the noise to better resemble the desired output. This process allows them to generate believable images with finer details and more distinctive characteristics (Ho et al., 2020). While they surpass GANs and VAEs in some performance measures, their training is comparatively more complex. (Aghasanli et al., 2023).
- **Face Manipulation Techniques**
- Face Manipulation are different types in images and videos generation. Face manipulation types are categorized into following types: Face-swap, Face-Synthesis, Face-Attribute Manipulation and Face-Reenactments (Irfan et al., 2025).
- **Face-swap:** In this face-swapping technique, which utilizes an auto-encoder framework, a face's underlying characteristics are retrieved from an image using an encoder. Subsequently, a decoder rebuilds the face with a more typical or conventional appearance. The first Face-Swap used FaceForensics + + dataset for the generation of deepfake (Rössler et al., 2019).
- **Face-Synthesis:** Face synthesis is the creation of realistic images of human faces, which can be based on existing people or entirely fabricated.
- **Face-Attribute Manipulation:** This involves modifying a specific facial characteristic, while preserving other unrelated features, to alter the appearance of a face.
- **Face-Reenactment:** This method changes how someone looks by overlaying the facial movements, such as how they move their eyes or head, from one person onto a picture or video of another.

DEEPFAKE IMAGE DETECTION METHODS

In this Section, we first determine the taxonomy of Deepfake image detection methods.

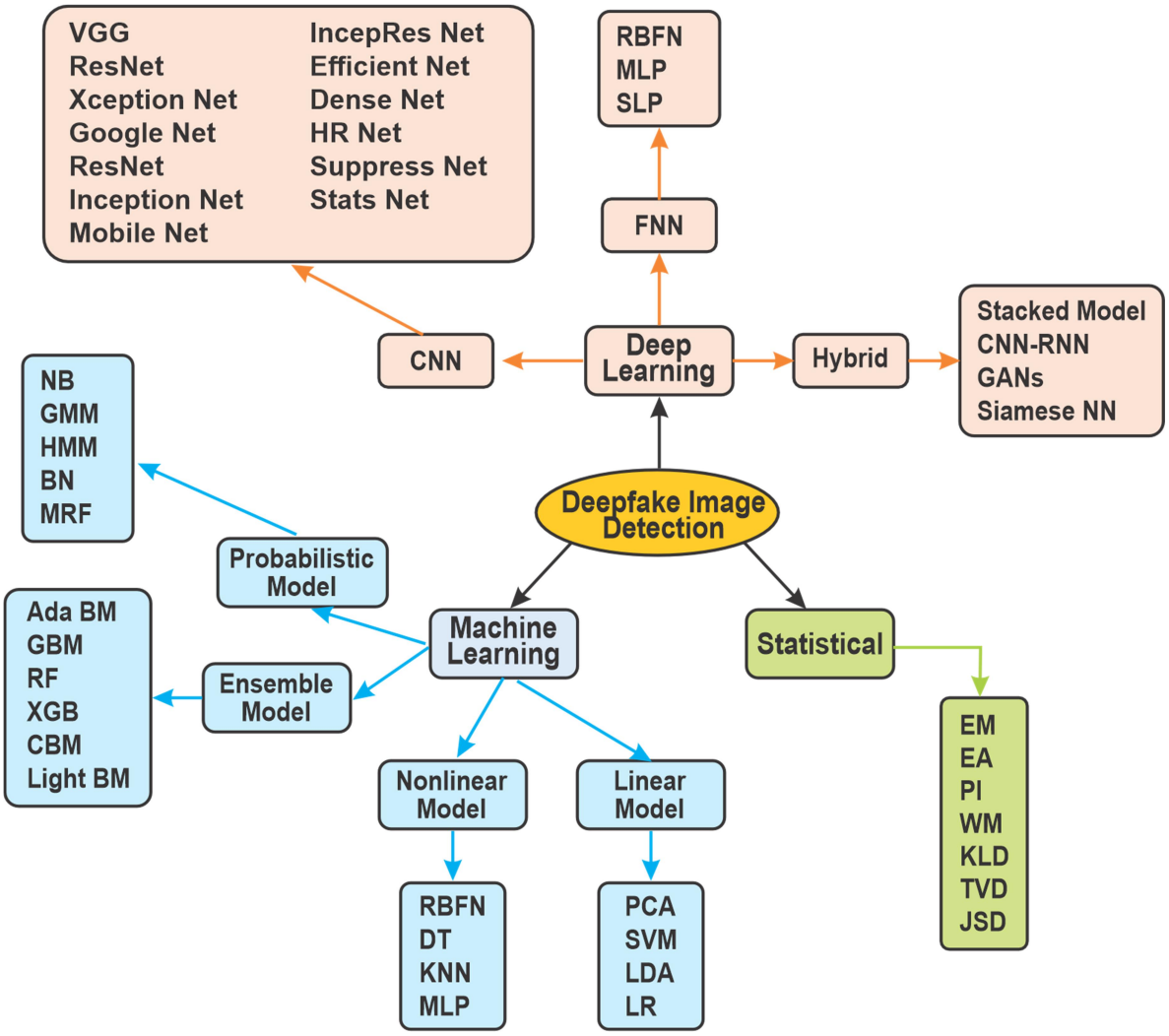

Fig. 7. Taxonomy of major Deepfake image detection methods

Machine Learning Method

Traditional machine learning methods are beneficial for interpreting the reasoning behind any outcome that can be expressed in understandable terms. Because machine learning approaches address many current challenges, they are well-suited for tasks like deepfake detection, offering improved data and process understanding. Additionally, adjusting settings and modifying model designs is much simpler. Generative Adversarial Networks (GANs), on the other hand, are self-training generative models that handle unsupervised problems as supervised, producing highly realistic fake faces in images or videos. (Agarwal et al., 2021) explores deepfake detection using Support Vector Machines (SVM).

The approach identifies subtle, artificial features imperceptible to the human eye. It employs frequency domain analysis to extract these features, followed by a classifier to distinguish between authentic and manipulated images. The method's effectiveness is assessed using a dataset of deepfake celebrity images gathered from various online sources. The findings indicate significant potential for accurately identifying deepfakes, achieving an accuracy of 99.76%. (Hanady Sabah Abdul Kareem et al., 2023) proposed a new technique for spotting AI-generated fake faces, using a Support Vector Machine (SVM) classifier, demonstrates improved reliability compared to existing approaches. The initial steps involve a preprocessing phase, which begins by converting images from RGB to YCbCr color space, followed by gamma correction. Crucially, the results indicate that edges identified using the Canny filter are valuable for differentiating between real and synthetic faces. Subsequently, two different detection strategies are employed: one using an SVM with Principal Component Analysis (PCA), and another using an SVM without PCA. The study reveals that the highest level of precision, reaching 96.8%, is achieved with the SVM-PCA combination, while the standalone SVM achieves a precision of 72.2%.

Deep Learning Methods

Many studies explore Deepfake detection in images, employing deep learning techniques to identify telltale signs left by the Deepfake creation process. Deep learning methods, including Convolutional Neural Networks (CNNs), Feedforward Neural Networks (FNNs), and combined approaches, are widely used for automatically identifying Deepfake characteristics. These methods often outperform traditional machine learning models in detecting Deepfakes (Kroiß & Reschke, 2025). Deep learning-based detection methods offer the possibility of enhanced accuracy. One study analyzing the racial makeup of Deepfakes discovered effective training signals in unusual faces resulting from face swapping (Trinh & Liu, 2021). This research introduces key elements for Deepfake detection: a system for separating and classifying facial expressions. While achieving high precision (0.94% - 0.99%), it exposes vulnerabilities to adversarial attacks (Cao & Gong, 2021). Another study (Ismail et al., 2021) introduces a new Deepfake detection method (YOLO-CNN-XGBoost) that functions as a high-level CNN Network and achieves an accuracy of 90.73%.

Amrithat (2025) presented an enhanced CNN architecture,DeepGuardNet, a deepfake detection model that is simple and effective for determining whether images are real or fake. DeepGuardNet is a straightforward, sequential, and robust network designed for deepfake recognition and detection. Additionally, the network has an enhanced, ability to detect tampered content with fewer parameters due to the use of separable convolution. They utilize depth wise separable convolution to efficiently extract deepfake features. The DeepGuardNet architecture effectively captures deep fake image features in both the spatial and depth dimensions. Celeb-DFdataset demonstrated the competence of the proposed method with an accuracy of 91% when compared with conventional methods. The proposed DeepGuardNet architecture is productive in terms of the better feature extraction and reduced computational complexity.

(Yadav et al., 2025) presents a new technique for spotting forged images by blending several methods. It uses CNNs at different resolutions, analyzes image frequencies, incorporates transformer networks that focus on important areas, and combines the predictions of multiple models for improved accuracy. The method was evaluated on a large dataset of 140,000 images (half authentic, half fake), achieving 93% accuracy during training and 88% accuracy during testing.

By using techniques to make the model resistant to adversarial attacks and extracting comprehensive features, the approach effectively identifies subtle traces left behind during image tampering. The findings demonstrate that the model can be applied to various types of image manipulation and different datasets, making it a practical and dependable solution for real-world use. The study highlights the value of combining different detection methods to handle the intricate nature of fake media analysis and emphasizes the need to integrate features from multiple domains to tackle the increasing challenges posed by deepfakes. (Gura et al., 2024) proposed a novel Convolutional Neural Network (CNN) approach for identifying deepfake videos. The method begins by extracting key facial features from video frames using facial landmark detection. This structured facial data is then fed into a specialized CNN. To improve the model's performance, a data augmentation technique is employed within the CNN itself, creating synthetic "fake data" or "fake images" to train on.

The system was developed using Python and associated libraries. The researchers used footage from the Deep Fake Detection Challenge dataset, specifically 242 videos (199 fake, 53 real) and an additional 76 (40 fake, 36 real), each segment being 10 seconds long. The proposed CNN achieved a high testing accuracy of 91.47%, with a loss of 0.342 and an AUC score of 0.92. These results surpassed the performance of two comparative methods: a standard CNN and an MLP-CNN. Notably, this new method also demonstrated superior accuracy compared to existing state-of-the-art models, including XceptionNet, Meso-4, EfficientNet-BO, MesoInception-4, VGG-16, and DST-Net. The core innovation of this work lies in creating a new CNN-based learning model specifically designed for precise detection of manipulated facial images in deepfakes. (Kosarkar et al 2023) identify deepfake images within a video dataset, we employed a specifically designed convolutional neural network (CNN) algorithm. We then compared its performance against two alternative approaches to ascertain the optimal method. Our model was trained and tested using a dataset from Kaggle.

This research leverages CNNs to differentiate between genuine and manipulated images by training three separate CNN models. Furthermore, we created and implemented a refined CNN model, incorporating additional components such as a fully connected layer, maximum pooling, and a dropout layer. The methodology involves extracting frames, identifying facial features, preparing the data, and finally classifying the images as either authentic or fabricated, aligning with the research aims. Performance was evaluated using metrics like accuracy, loss, and the area under the ROC curve. Our enhanced CNN model demonstrated superior results, achieving 91.4% accuracy, a lower loss value of 0.342, and an AUC of 0.92. In addition, the standard CNN model reached 85.2% accuracy on the test set, while the MLP-CNN model achieved 95.5% testing accuracy.

Statistical Measurements Methods

Statistical methods have become important resources for identifying deepfakes. These methods use complex statistical analysis to find trends and differences in data, helping to spot altered content. By studying various statistical characteristics, such as image quality measures, noise variations, facial points, and timing problems, these approaches aim to distinguish between real and fake media. (Hou et al., 2023) proposed a statistically-driven attack, termed Stat Attack, has been introduced to challenge DeepFake detectors. This approach operates in two stages. Initially, a collection of naturally occurring distortions (specifically, exposure, blur, and noise) are deliberately introduced to the synthesized images to undermine the detector.

Subsequently, it has been observed that discrepancies in statistical properties between real and DeepFake images correlate with shifts in the data distributions distinguishing the two categories. (Rana and Sung, 2024) explores a machine learning-based method for spotting manipulated videos, commonly known as Deepfakes. It includes a statistical analysis and compares the performance of different machine learning models. The findings show exceptionally good results, achieving 99.84% accuracy on the FF++, 99.38% on the DFDC, 99.66% on the VDFD, and 99.43% on the Celeb-DF datasets. The statistical examination confirms the reliability of these strong accuracy scores across a range of datasets. Therefore, the research questions the widely held belief that deep learning methods are the only efficient way to identify Deepfakes. It demonstrates that smart application of machine learning techniques can provide highly precise, affordable, and more environmentally friendly solutions for Deepfake detection.

CNN MODELS USE FOR DEEPFAKE IMAGE DETECTION

This segment describes various CNN models that are used for detecting Deepfake. In Deepfake image detection studies, we found the following deep learning-based models have been used: convolutional neural network (CNN) model (e.g., XceptionNet, GoogleNet, VGG, ResNet, EfficentNet, HRNet, InceptionResNetV2, MobileNet, InceptionV3, DenseNet, SuppressNet, StatsNet) (Rana et al., 2022) (Meraj et al., 2025) explored the use of various EfficientNet architectures (including older versions like B3 and B5, as well as newer ones like V2S and V2M) as backbones for deepfake face detection models. The goal was to assess their performance in identifying face deepfakes. A key contribution is the inclusion of images generated by the latest StyleGAN3 technique, which, to the best of our knowledge, has not been previously used in deepfake detection research. This makes the study a valuable starting point for understanding how pretrained models like EfficientNet handle StyleGAN3-generated images.

The findings revealed that EfficientNet B3 achieved the highest accuracy (92.77%) on a test dataset *excluding* StyleGAN3 images, even outperforming newer EfficientNet versions. However, when evaluated against a separate, pre-processed test set of StyleGAN3 images, the accuracy dropped significantly. The best performance in this scenario was only 64.29%, achieved by EfficientNet V2M, which had a lower overall accuracy of 88.60% on the dataset excluding StyleGAN3. These results indicate that, despite achieving high accuracy on datasets containing a variety of generated images (including previous StyleGAN versions and some "real-world" scenarios), pretrained backbone-based detectors struggle to maintain their performance when specifically challenged with pre-processed StyleGAN3 images.

(Kumari et al., 2025) presents a method for identifying Deepfake images and videos using the XceptionNet model, a deep learning technique. The system employs deep learning and image analysis methods to distinguish between authentic and manipulated content. The process involves extracting frames from videos and images, preparing them for analysis, and then using a trained XceptionNet model to categorize them as either genuine or fabricated. Video classification is determined by the most frequent prediction across all analyzed frames. A user-friendly application, developed with Streamlit, allows users to upload and analyze media instantly. (Praditha et al., 2023) proposed convolutional neural networks, particularly those based on the VGG face architecture, are a promising approach for identifying synthetic or altered images, including those generated by GANs.

While these models demonstrate considerable capability in detecting diverse image manipulation methods. The model shown accuracy 89.4%, ROC 0.961, F1-Score 0.894, Recall 0.899 using Flickr dataset (140k real and fake faces). (Doegar et al., 2020) this research introduces an automated method for identifying and pinpointing manipulated regions within images. To achieve reliable forgery detection, a sufficient set of distinguishing characteristics is necessary. This is accomplished using a deep learning model, specifically circumventing the need for manual feature design or handcrafted techniques. This study utilizes the GoogleNet deep learning architecture to automatically extract relevant image features. Subsequently, the Random Forest machine learning algorithm is employed to classify images as either authentic or forged. The proposed method is tested on the widely used MICC-F220 dataset, employing k-fold cross-validation to partition the data into training and testing sets. The performance of the developed approach is also benchmarked against existing cutting-edge techniques.

(Nagagopiraju et al., 2024) employed the InceptionNet architecture to detect counterfeit faces. Their method leverages various shifts and alterations present in authentic images, alongside test parameters including the count of significant points within images, comparative evaluation, and the time needed for each technique to perform. The findings of this research indicate a total precision of 93% for the DFD dataset. This research is capable of categorizing deepfake videos originating from multiple sources, utilizing varied convolutional layers.

ResNet-50 Architecture

A residual neural network (ResNet-50) is a type of convolution neural network (CNN) that stacks residual blocks on top of each other to form a deep network (Sharma et al., 2022). ResNet-50 whole design is seen in Figure 2.5. The ResNet-50 architecture is structured into five distinct stages, each featuring both identity and convolutional blocks. ResNet-50's size is progressively decreased due to the incorporation of pooling layers at each stage. Convolutional blocks consist of three convolutional layers each, while identity blocks also contain three layers. ResNet-50 has approximately 23 million adjustable parameters. Its strong performance has led to widespread adoption in numerous image-based detection and classification tasks.

RQ1: How is ResNet-50 applied for deepfake image detection

(Ho et al., 2024) this study introduces a defense mechanism against adversarial images using a ResNet-based framework. The proposed method utilizes the ResNet-50 architecture and incorporates adversarial training techniques to enhance resilience against adversarial examples. The model's performance is evaluated on a mixed CIFAR-10 dataset, intended to boost both robustness and precision through the inclusion of GAN-synthesized images. Experimental results demonstrate that the suggested model attains an accuracy exceeding 95%, surpassing the performance of three established models – VGG19_bn, ShuffleNetV2, and RepVGG_a2.

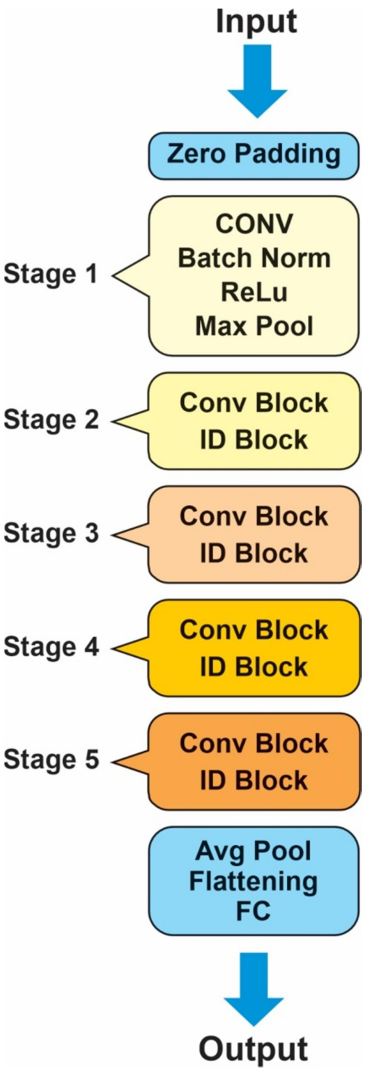

Fig. 8 Taxonomy of ResNet-50 Architecture

(Singh & Ramachandra, 2022) presents a a Deep Fake Detection system (DFD) powered by deep learning. The system utilizes enhanced convolutional layers before feeding into a Resnet-50 architecture. DFD is trained from start to finish using low-quality images from the FaceForensics++ dataset. Around 1.68 million images are used for training, 315,000 for validation, and 340,000 for testing. The DFD is evaluated in three settings: a mix of various image manipulations (achieving 96.07% accuracy, significantly better than the existing best methods at 85.14%), individual manipulation techniques (achieving perfect 100% accuracy on neural textures), and manipulations involving different images (achieving 94.28% accuracy on the previously unseen faceswap category, outperforming the state-of-the-art).

(Li and Lima, 2021) a facial expression recognition system leveraging a ResNet-50 architecture. The proposed approach utilizes the deep residual network ResNet-50 for feature extraction, integrating it with a convolutional neural network for facial emotion recognition. Experimental results on a defined dataset demonstrate the model's improved performance in facial emotion detection compared to existing state-of-the-art models. The ResNet-50 method achieved a peak accuracy of 95.39%. (Kroiß & Reschke, 2025) developed a deepfake detection system for face images using a Convolutional Neural Network. They leveraged a pre-trained ResNet-50 model because of its high accuracy in image classification. They adapted this model to classify images as real or fake by adding a fully connected output layer with a single neuron to represent image authenticity. Fine-tuning and transfer learning techniques were then applied to optimize the model and improve its performance. For training, they used the 'Diverse Face Fake Dataset,' which contains a variety of image manipulation techniques and a diverse range of faces. The resulting model achieved impressive performance, with a precision of 0.98, a recall of 0.96, an F1-Score of 0.97, and an area-under-curve of 0.99.

(Metin & Ozkaynak, 2025) presents a combined deep learning model for identifying artificial intelligence-created fabricated images. The method uses ResNet-50 and Vision Transformer (ViT) models to obtain key characteristics. These characteristics are then merged into a unified characteristic vector, which is handled by the XGBoost algorithm for categorization. The model was educated and assessed using the CIFAKE dataset, comprised of 60,000 authentic and 60,000 fabricated images. Experimental outcomes reveal that our technique achieves a substantial accuracy of 98.77% in identifying fabricated images. These results emphasize the capability and dependability of our approach, outperforming current methods in effectiveness.

(Lipianina-Honcharenko et al., 2024) this study compares the performance of three deep learning networks—ResNet 50, EfficientNet, and Xception—for spotting manipulated images. The main purpose was to determine the best model for categorizing deceptive images and videos. The researchers also looked into how adding tools like Long Short Term Memory (LSTM) and attention mechanisms could improve the models' correctness. A collection of genuine and fabricated images and videos was used to test each model, and their performance was measured using metrics like correctness, precision, recall, and F1-score. The findings revealed that Xception achieved the best correctness (87.7%), while EfficientNet was also highly efficient, especially when resources are limited. ResNet showed consistent performance but had difficulty with less common categories.

(Ritu, 2025) explores a method for identifying fake face images and videos using a combined approach. It examines the performance of two advanced image processing models, ResNet50 and EfficientNetB7, in detecting manipulated faces. The models were trained and tested using standard datasets like FaceForensics++ and DFDC, which represent both controlled and real-world scenarios. ResNet50, known for its ability to identify subtle image discrepancies, performs well in detecting small alterations at the frame level. EfficientNetB7, on the other hand, demonstrates stronger adaptability and precision in more complex situations due to its advanced scaling techniques and detailed feature analysis. While tools like Adobe Photoshop require skill to manipulate images, recent advancements in artificial intelligence allow for the automatic creation of fake images and videos, including face swapping, without specialized knowledge. These AI-generated fakes pose a significant challenge due to their potential for misuse.

(Sanghvi,2023) a new approach to identifying manipulated images using CNNs and the ResNet50 architecture. Initial findings after two training cycles showed a significant difference in performance. The CNN exhibited a higher accuracy of 0.9067 compared to ResNet50's accuracy of 0.7066 on a dataset comprising labelled real and fake instances. (Rajagukguk et al., 2024) carry out image classification by employing a CNN model based on the ResNet50 architecture, a well-established framework for this purpose. ResNet50 is recognized for addressing the vanishing gradient issue, allowing for efficient training with a reduced number of layers while maintaining high accuracy. The research utilizes a dataset consisting of 589 authentic human face images and 700 synthesized human face images. Prior to training, image augmentation techniques are applied, and the data is split into training (80%) and validation (20%) sets. The model achieved an accuracy of 76.07% on the training dataset. However, the model's performance on the validation dataset yielded an accuracy of 53%.

(Gunukula et al., 2025) resents a new deep learning approach for identifying artificially created images. The method combines the ResNet-50 structure with Squeeze-and-Excitation (SE) attention mechanisms. This enhanced model, called SE-ResNet50, improves the adjustment of feature importance across different channels and the understandability of the model's decisions. By incorporating SE blocks into the ResNet-50 framework, the model can dynamically focus on subtle indicators of artificial generation, like odd textures and logical errors, leading to more accurate classification. Testing on the CIFAKE dataset shows that the model performs well, reaching a test accuracy of 96.12%, a precision of 97.04%, a recall of 88.94%, an F1-score of 92.82%, and an AUC score of 0.9862.

(Latif and Khalifa, 2023) carried out three Deep Transfer Learning (DLT) algorithms to distinguish between healthy images, pneumonia cases, and COVID-19 infections. The findings indicate that ResNet50 achieved superior training accuracy, reaching 94.72%. Moreover, the ResNet50 model demonstrated satisfactory testing performance, with an accuracy of 80.66%. (Alzamily et al., 2022) perform image categorization for enhanced data security, employing a Convolutional Neural Network (CNN) algorithm with a ResNet50 framework. By training the model on ciphered images, the research demonstrates the ability to classify encoded images directly, bypassing the need for decryption. The image classification model in this research achieved a high level of performance, exhibiting an accuracy of 99.75%, a Recall of 94.12%, a Precision of 94.23%, and an F1-Score of 94.70%.

(Vedhapriya et al., 2025) presents a straightforward yet powerful deepfake recognition system for still images of faces. Leveraging transfer learning with the ResNet50 convolutional neural network, the system differentiates authentic images from altered ones by spotting minor variations in facial characteristics. A simple interface was developed using Streamlit and made available through Hugging Face Spaces, enabling immediate categorization. Although the system achieves fair precision (roughly 61%), it highlights the promise of readily available and deployable AI instruments for use in digital investigation. (Wangsadidjaja, 2023) developed a deepfake detection tool using a Convolutional Neural Network, specifically the ResNet-50 architecture, trained on fake images generated by a ProGAN model. This method reached 85% accuracy, perfect precision, and 65% recall when identifying these manipulated images.

However, the model was less successful at detecting deepfakes produced by the original StyleGAN and BigGAN techniques. (Zahra Nazemi Ashani, 2024) they took a look at the potential of artificial intelligence for identifying manipulated images (deepfakes). The researchers examined the performance of three types of convolutional neural networks (CNNs): ResNet, VGG16, and VGG19. These models were tested on a collection of 1,200 images, comprising both authentic and fabricated examples. The deepfake images were generated using FaceApp, a widely used application for modifying images. The results indicated that VGG19 performed better than VGG16 and ResNet50, achieving a 98% accuracy rate.

(Hasan Abir et al., 2023) examines the widespread use of deep learning for creating generative media, like deepfakes. It focuses on how advanced algorithms are used to identify these deepfakes and how their effectiveness is assessed using Local Interpretable Model-Agnostic Explanations (LIME). The study involved transferring styles from two sets of images: 70,000 genuine images from the Flickr dataset (originally used by Nvidia) and 70,000 artificial images generated using StyleGAN at a 256-pixel resolution. To maximize precision, different Convolutional Neural Network (CNN) architectures (specifically, InceptionResNetV2, DenseNet201, InceptionV3, and ResNet152V2) were implemented. The LIME method was used to clarify which parts of the image influence the model's classification outcomes. InceptionResNetV2 achieved the highest accuracy at 99.87%, outperforming DenseNet201 (99.81%), InceptionV3 (99.68%), and ResNet152V2 (99.19%). The LIME technique supported these results, improving the clarity of the models for the purpose of explainable artificial intelligence (XAI).

(Wang et al., 2023) proposed a method for identifying manipulated facial images (deepfakes) using convolutional neural networks. The effectiveness of the approach is tested using three established collections of real and fake face images: 140k Real and Fake Faces, Real and Fake Face Detection, and Fake Faces. To enhance the performance of the proposed model, it's integrated with pre-existing VGG16 and ResNet50 models to create deep learning ensembles. This combined system significantly boosts the accuracy, achieving rates of 98.79%, 75.79%, and 95.52% on the respective datasets. (Bartos and Akyol, 2023) explores the use of two well-known deep learning techniques, specifically Residual Networks (ResNet-50) and Variational Autoencoders (VAEs), for identifying artificial intelligence-created synthetic images. These methods tackle the problem in distinct ways: ResNets strive to differentiate genuine images from fabricated ones by examining advanced characteristics and organizational structures within the images. VAEs, on the other hand, treat this as an unusual occurrence identification task.

The effectiveness of these techniques is evaluated using the CIFAKE image dataset, which includes 120,000 images (60,000 artificial images and 60,000 authentic images). The experiments showed that ResNet-50 performed better than VAEs at accurately categorizing real and AI-generated images with accuracy of 0.94. (Mansoor and Iliev, 2024) this study uses two different CNN structures for categorization, highlighting the growing need for DeepFake identification in today's world. The research creates DeepFake detection methods by constructing two CNN models using ResNet50 and DenseNet121 designs. The two models are compared based on accuracy and F1-score performance metrics. The data shows the CNN models used in the work reach an accuracy of 86% and an F1-score of 0.87

Table 2: ResNet-50 Models used for deepfake image detection

Authors	Models	Results	Datasets	Limitation
(Mansoor and Iliev, 2024)	ResNet-50, DenseNet121	Accuracy 86% F1-Score 0.87	Flickr dataset (140 real and fake Faces)	Provides poor results and lacked generalizability
Bartos and Akyol, 2023)	ResNet-50, VAEs	Accuracy	CIFAKE	Model need to extend the scope of fine-tuning the algorithms to improve the performance
(Sharma et al., 2022)	ResNet-50, VGG16 and Ensemble	Accuracy 95.85%, 75.79% and 98.79%	Flickr dataset (140k real and fake faces) FFD and FF++	Like ensemble approach and improvement by incorporating the concept of occlusion invariant
(Ashani et al., 2024)	VGG16, VGG19, and ResNet-50	Accuracy 98%	Flickr dataset (140k real and fake faces)	exploring advanced AI techniques are necessary to enhance the robustness and reliability of CNN architectures for deepfake detection in real-world scenarios.
(Wangsadidjaja, 2023)	ResNet-50	Accuracy:85% Precision: 100% Recall: 65%	Flickr dataset (140k real and fake faces)	Provides poor results and lacked generalizability
(Abir et al., 2023)	InceptionV3 ResNet152 V2 DenseNet201 InceptionRes NetV2	99.68%, 99.19%, 99.81%, 99.87%	Flickr dataset (140k real and fake faces)	project can be improved by examining the model with updated transfer learning methods from its new versions.
(Vedhapriya et al., 2025)	ResNet-50	Accuracy above Above 61 %	Flickr dataset (140k real and fake faces)	Enhancing model accuracy through fine-tuning and t raining on more diverse datasets.

(Latif and Khalifa, 2023)	ResNet-50	Accuracy of 94.72%	COVID-19 dataset	Like generation across datasets
(Yousef et al., 2022)	ResNet-50	Accuracy (99.75%), Recall (94.12%), recision (94.23%) and F1-score (94.70%)	Designed Dataset	Like different feature techniques for selection and analysis
Rajagukguk et al., 2024	ResNet-50	Accuracy of 76.07%	Flickr dataset (140k real and Fake faces)	Need to carry out classification using architectures, activation functions, and optimizers that are different from the research
(Gunukula, et al., 2025)	ResNet-50 with Squeeze-and-Excitation (SE)	Accuracy 96.12%, precision 97.04%, recall of 88.94%, F1-score 92.82%, AUC score 0.9862.	CIFAKE dataset	Like generalization across datasets and domains, Adversarial Robustness
(Sanghvi et al., 2023)	ResNet-50	Accuracy 0.9067	FaceForensics++:	Like Incomplete Model Training, Potential Overfitting and Dataset Specificity
(Ritu et al., 2025)	ResNet50 and EfficientNetB7	Accuracy around 49% to 54%	FaceForensics++, DFDC, CNN, GAN	No improvements include newmethods like transformer networks and using time-based models
(Zulfiye and Ozkaynak, 2025)	ResNet-50 and Vision Transformer (ViT)	Accuracy 98.77%	CIFAKE dataset	Like generalization across datasets and domains
(Borade et al., 2024)	ResNet-50	Accuracy 97%	CelebDF and FaceForensics++	The proposed model need further enhancement and refinement and to improve computational time
(Kroi and Reschke,	ResNet-50	precision = 0.98 recall 0.96,	CelebA, FF++ and FFHQ	The hypothesis was not analyzed in detail

2025)		F1-Score = 0.97 and AUC= 0.99		and an interpretation of the model could be interesting and revealing as well
(Li and Lima,2021)	ResNet-50	Accuracy 95.39	Designed dataset	poor generalization ability of network model and low robustness of recognition system
(Singh and Ramachandra, 2022)	Resnet-50	Accuracy 96.07%	FF++	The proposed model need to be modified to improve cross image manipulation results
(Ho et al., 2024)	Resnet-50	Accuracy of over 95%	CIFAR-10	Like generalization across datasets and the proposed model need more optimal than the ResNet architecture
(Lipianina-Honcharenko et al., 2024)	ResNet 50, EfficientNet, and Xception	Accuracy 87.7%	DFDC	Model need to improve results in real-world conditions by developing effective ensemble approaches

RQ2: What datasets are commonly used with ResNet-50 for deepfake detection?

We found several manipulated image datasets that have been used widely in different research projects for both training and evaluation. These resources have facilitated significant improvements in the identification of manipulated images. These dataset includes; CIFAKE dataset, Flickr / FFHQ dataset,CelebA dataset, Open Forensics dataset, FaceForensics++(dataset), Celeb-DeepFake (Dataset), DFDC/DFD dataset, AttGAN dataset, STARGAN dataset, StyleGAN dataset, StyleGAN2 dataset.

The datasets listed below are frequently used for the task of detecting deepfake images:

1. CelebA. A large collection of celebrity images (over 200,000) with associated attributes and identity information. (CelebA Dataset, 2015)
2. FF++: A dataset built from 1000 authentic videos that have been manipulated using common face alteration techniques like FaceSwap, Face2Face, Deepfakes, and NeuralTextures. (Onagri FaceForensics++, 2023)
3. HFF: A significant dataset composed of over 155,000 artificially created facial images (Guo et al., 2021).

4. Flickr/FFHQ: A high-quality image dataset with a large number of unique identities, used as a benchmark for image-related tasks. (Yin et al., 2024).
5. AttGAN: A dataset containing a substantial number of images, exceeding 30,000. (Lin, 2023)
6. StyleGAN: A dataset of over 7000 artificially generated images. (NVlabs/stylegan, 2023)
7. StyleGAN2: A larger dataset focused on fake face images, containing approximately 100,000 samples. (NVlabs/stylegan2, 2023)
8. OpenForensics: A dataset for detecting and segmenting forgeries in images containing multiple faces in unconstrained environments, comprising of real and fake images (over 190,000) (Le, 2021).
9. DFDC/DFD: The Deepfake Detection Challenge (DFDC) dataset is a comprehensive collection for deepfake forensics, comprising 590 real videos and 5,639 DeepFake videos (Brian et al., 2024).
10. CIFAKE/ CIFAR-10 dataset: A dataset of 120,000 images labeled as either real or fake, using real images from the CIFAR-10 dataset. (Bird and Lotfi, 2024)

(Abdal Mashkoor, 2025) this work explores distinguishing authentic images from those created by artificial intelligence using a customized ResNet50 model. The model was trained and evaluated using a dataset containing 140,000 images of both genuine and AI-generated faces sourced from Flickr. The results show the method's capability in categorizing images. Initially, the feature extraction model achieved an accuracy of 89%.
Further optimization through fine-tuning over 5 training cycles improved the accuracy to 98% (Lamichhane et., al 2022) proposed a method for identifying image-level forgeries, tested on both the complete and a subset of the DeepFake Detection Challenge (DFDC) dataset. The research evaluates the proposed model's effectiveness using established pre-trained architectures like VGG-19, Xception, and Inception-ResNet-v2. Notably, the experiments explore various image sizes while preserving aspect ratios of 1:1 and 9:16, a relatively unexplored area within DFDC dataset research. This allows the model to handle diverse image resolutions and aspect ratios, making it adaptable for general image classification tasks.

Unlike much existing research that focuses solely on DFDC samples or previews, this work extends the evaluation to the full DFDC dataset. The comprehensive results and resource analysis across different scenarios suggest that this deepfake detection approach can offer valuable insights and direction for future research using the DFDC dataset. Performance was evaluated using common metrics like accuracy, recall, precision, ROC area under the curve, and F1-score.

(Naskar et al., 2025) this paper introduces a new deep learning approach for spotting fabricated facial images, combining the strengths of Generative Adversarial Networks (GANs) and RESNET-50 architecture. The performance of this combined model was compared against established pre-trained models like VGG16 and RESNET 50. The results indicate that the new model is significantly better at correctly identifying fake faces, representing a substantial improvement in facial image recognition and authentication technologies. Tests on a standard dataset reveal that the proposed model achieves impressive results, including a precision of 0.79, recall of 0.88, F1-score of 0.83, accuracy of 0.83, and ROC AUC Score of 0.825.

The study concludes that this hybrid model demonstrates strong capabilities in identifying fake faces, particularly in terms of correctness, precision, and efficient memory use with Celeb-DF and FF++ datasets. (Chouhan et al., 2024) this study investigates deepfake detection using the ResNet-50 architecture. The researchers conducted multiple tests, training ResNet-50 with the Celeb-DF dataset while varying parameters and employing different fine-tuning approaches. They assessed the resulting models using standard metrics to understand and quantify their effectiveness. Across twelve distinct experiments, some configurations of the model demonstrated remarkable ability, reaching an AUC score of 98.11%. The results of this research aim to foster trust within society by addressing the increasing danger presented by deepfakes

(Naskar et al., 2025) presented a deepfake detection method leveraging deep feature integration and meta-learning. Initially, a stacking ensemble method is employed, merging characteristics extracted from two well-known deep learning architectures, Xception and EfficientNet-B7. Subsequently, a near-optimal feature selection process, based on ranking, identifies a reduced set of features. The final classification, distinguishing between authentic and manipulated videos, is carried out using a meta-learner, specifically a multi-layer perceptron. The method attained a precision of 96.33% on the Celeb-DF (V2) dataset and 98.00% on the FaceForensics++ dataset with the meta-learning model, surpassing the performance of the individual base models. (Berrahal et al., 2023) Investigating how well advanced machine learning techniques work for identifying deepfakes.

The models were trained using a substantial collection of authentic images from CelebA-HQ and synthetic images created by a generative adversarial network (GAN). All the tested models were highly successful at detecting the fabricated images, with the residual neural network (ResNet50) demonstrating the best performance at 99.43% accuracy. (Jasim and Atia, 2023) developed an evolutionary learning approach was created to automatically devise convolutional neural network (CNN) architectures for identifying deepfakes. A genetic algorithm (GA), utilizing residual networks (ResNet) and densely connected convolutional networks (DenseNet) as fundamental components for feature extraction, in conjunction with multilayer perceptrons (MLP), random forests (RF), and support vector machines (SVM) as classification methods, produces diverse CNN designs.

A localized search mutation strategy was introduced to fine-tune three levels: batch normalization, activation function, and regularization techniques. This technique offers the benefit of functioning across various datasets without requiring preliminary data preparation. Results obtained from the HHF dataset and a collection of 140,000 manipulated images demonstrate the effectiveness of the presented method, where the resulting models surpass existing techniques by improving accuracy by 1%. This underscores that automated design is a promising path toward enhanced generalization.

(Samal et al., 2024) explores a method for identifying and categorizing manipulated images (deepfakes) using Conv2D Neural Networks. Unlike many existing techniques that rely on smaller datasets or pre-trained models, this approach trains a Conv2D model from scratch on a substantial dataset of 140,000 images, equally divided between authentic and fake examples. To improve the model's training effectiveness, sparse-categorical cross entropy is used as a loss function, and the Adam optimizer is employed. The resulting model demonstrates a strong performance, achieving 94.54% accuracy on the challenging OpenForensics dataset, a large collection of real-world forged images with multiple faces.

(Patel et al., 2023) introduces a new and enhanced deep convolutional neural network (D-CNN) design for identifying deepfakes, achieving good precision and broad applicability. The model is trained using images from diverse origins, which boosts its ability to generalize. These images are resized and then processed by the D-CNN. Binary cross-entropy and the Adam optimizer are employed to accelerate the D-CNN's learning process. The research uses seven different datasets from the reconstruction challenge, consisting of 5000 deepfake images and 10000 authentic images. The suggested model achieves the following accuracy rates: 98.33% on AttGAN, 99.33% on GDWCT, 95.33% on StyleGAN, 94.67% on StyleGAN2, and 99.17% on StarGAN datasets, discerning between genuine and fabricated images.

(Bird & Lotfi, 2024) proposed the use of a Convolutional Neural Network (CNN) was employed to distinguish between authentic and artificial images. After optimizing settings and training 36 different network structures, the best model achieved a classification accuracy of 92.98%. To understand the model's decision-making process, explainable AI techniques, specifically Gradient Class Activation Mapping, were utilized to identify important image features. Interestingly, the analysis revealed that the central subject of the images was not a key factor for classification. Instead, the model relied on subtle flaws in the image background. The dataset created for this research, named CIFAKE, has been released for public use by other researchers.

Table 2, Highlighting the author, dataset, types, metrics and common use with ResNet-50 in Deepfake image Detection.

Table 3. Publicly accessible datasets for Deepfake image Detection using ResNet-50

Authors	Dataset s	Types	Metrics	Common use with ResNet-50
(Bird et al., 2024)	CIFAKE	Image	Accuracy	
(Aabid and Mashkoor, 2025)	Flickr dataset (140k real and Fake faces)	Image	Accuracy	Image level deepfake detection (binary classification, fine turned on face level image
(Jasim and Atia, 2023)	HHF	Image	Accuracy	Deepfake/face f orgery detection, fine-turned as binary classifier to detect fake and real face
(Lamichhane., 2022)	DFDC/ DFD	Image/Video	Accuracy, recall, precision, ROC, F1	Deepfake video detection, frame - based ResNet-50 detection or part of ensemble
(Naskar et al., 2025)	FFHQ	Image	Accuracy, recall,	High quality facial face image modeling

			precision, ROC, F1-score AUC	(face analysis, synthesis, face attribute and classification
(Berrahal et al., 2023)	CelebA-HQ	Image	AUC	face attributes/identity, deepfake and spoof detection using ResNet-50 as binary classifier
(Chouhan et al., 2024)	CELEB-DF	Image/Video	Accuracy	Deepfake video detection/facial forgery classification used as binary classifier (real or fake)
(Safwat et al., 2024)	FF++	Image/Video	Accuracy	Deepfake image detection, train ResNet-50 for fake or real face classification
(Samal et al., 2024)	Open Forensics	Image	Accuracy	Deepfake/image manipulation detection, evaluating generalization to unseen manipulation
(Patel et al. 2023)	AttGAN, Image STAR GAN, StyleGAN and StyleGAN2		Accuracy	face attribute manipulation, deepfake images editing classify manipulated or original faces

RQ 3: What evaluation metric are used to computing the performance of deepfake image detection model using ResNet-50?

This section outlines several evaluation metrics used to assess how well models perform in identifying manipulated images and videos. A confusion matrix summarizes the results of the classification, showing both correct and incorrect predictions. The effectiveness of the methods is evaluated and validated using data from this matrix. Specifically, True Positives (TP) represent the number of Deepfakes correctly identified as Deepfakes, while True Negatives (TN) indicate the

number of genuine images/videos correctly identified as real. Conversely, False Positives (FP) represent the number of real images/videos incorrectly classified as Deepfakes, and False Negatives (FN) represent the number of Deepfakes incorrectly classified as real. These values are used to calculate various performance metrics, which are frequently reported in research. The most commonly used metrics include Accuracy (AC), the Receiver Operating Characteristic (ROC) curve and its Area Under the Curve (AUC), Recall, Error Rate (ER), Precision (P), and the F1-score.

Table 4: Confusion matrix

		ACTUAL	
		POSITIVE	NEGATIVE
PREDICTION	POSITIVE	TRUE POSITIVE (TP)	FALSE POSITIVE (FP)
	NEGATIVE	FALSE NEGATIVE (FN)	TRUE NEGATIVE (TN)

Table 5: Common Evaluation metric used inResnet-50 for deepfake image detection

Metrics	Definition
Accuracy	AC = (TN+TP)/(TN+FP+TP)
Area Under Curve (AUC)	AUC provides an aggregate measure of performance across all possible classification thresholds
Receiver Operating Characteristic (ROC)	ROC is the plot with recall value on the y-axis and the specificity values on the x-axis
Error Rate(ER)	Error rate = (FP+FN)/(TN+FN+FP+TP)
Recall/ True Positive Rate (TPR)	Recall = TP/(FP+TN)
F-measure (F1)	F1=2*Recall*Precision/ (Recall + Precision)
Precision(P)	Precision = TP/(TP+FN)
False Positive Rate (FPR)	FPR = FP/(FP=TN)

RQ4: What is the performance compared to other models?

(Akramet al., 2025) proposed a deep learning strategy was developed to tackle the problem, contrasting a tailored Convolutional Neural Network (CNN) with the MobileNetV2 model. The assessment utilized the "CIFAKE: Real and AI-Generated Synthetic Images" dataset, which included 120,000 images equally divided between real and synthetic content, with 50,000 images per category for training and 10,000 for testing. The study revealed that the designed CNN architecture excelled in this particular task. It produced precise outcomes, achieving 95.48% accuracy, 0.94 precision, 0.97 recall, and a 0.95 F1-score. The MobileNetV2 model also performed well, attaining 95% accuracy, 0.96 precision, 0.94 recall, and a 0.95 F1-score.

The findings show the utility of CNN and MobileNetV2 for identifying manipulated images. These outcomes suggest that deep learning methods can significantly improve the precision and correctness of systems designed for Deepfake image detection. (Alzahrani, 2024) proposed Digital image forensics is advancing with the use of deep learning, but previous methods haven't adequately addressed the issue of vanishing gradients. To overcome the limitations of traditional CNNs, like vanishing gradients and excessive layers, a DenseNet model was developed. This DenseNet architecture, built with densely connected layers, aims to accurately distinguish between authentic and manipulated images. The model was tested against other leading deep learning techniques using a forged image dataset, and the results demonstrated its superior performance. The suggested improved DenseNet model can identify altered images with a high accuracy of 92.32%. (Killi et al., 2024) investigates the identification of manipulated images using a convolutional neural network (CNN) structure, which has demonstrated effectiveness in numerous image categorization applications. The suggested VGG19 model surpasses current approaches, achieving a 96% correctness rate.

(Widiyasono, 2025) developed a system for identifying manipulated images (deepfakes) was developed using the MesoNet algorithm. MesoNet, a type of Convolutional Neural Network (CNN), employs a shallow neural network that focuses on intermediate-level image characteristics, encompassing both basic and complex features. This deepfake detection model, built using the MesoNet algorithm, demonstrated strong performance, achieving an accuracy of 90.93%, precision of 87.78%, recall of 95.09%, and an F1-score of 91.28%. (Kroiß & Reschke, 2025) developed a deepfake detection system for face images using a Convolutional Neural Network. They leveraged a pre-trained ResNet-50 model because of its high accuracy in image classification. They adapted this model to classify images as real or fake by adding a fully connected output layer with a single neuron to represent image authenticity. Fine-tuning and transfer learning techniques were then applied to optimize the model and improve its performance. For training, they used the 'Diverse Face Fake Dataset,' which contains a variety of image manipulation techniques and a diverse range of faces. The resulting model achieved impressive performance, with accuracy of 96.06%, a precision of 0.98, a recall of 0.96, an F1-Score of 0.97, and an area-under-curve of 0.99.

Table 6: Comparison performance compared to (DensNet, VGG 19, MesoNet and MobileNet models)

Author	Model	Performance
(Alzahrani, 2024)	DenseNet model	Accuracy of 92.32%.
(Killi et al., 2024)	VGG19 model	Accuracy 96%
(Widiyasono, 2025)	MesoNet model	Accuracy 90.93%, precision 87.78%, recall 95.09% and f1-score 91.28%
(Akram et al., 2025)	MobileNetV2 model	Accuracy 95%, precision 0.96, recall 0.94 and F1-score 0.95
(Kroiß & Reschke, 2025))	ResNet50 model	Accuracy of 96.06%.

RQ 5: What are the limitations and challenges in ResNet-50 for this task?

This section will discuss some limitations and challenges that we observed during the preparation of this SLR.

Model-Centric Limitation

ResNet-50's architectural constraints (fixed receptive field, depth limitations) may cap detection performance. The SLR's focus on this single architecture narrows the field and may overlook more robust or lightweight alternatives.

Methodological Rigor and Transparency

Although the review identified 32+ studies using ResNet-50 for deepfake image detection, the lack of a registered protocol and limited coverage of non-english studies may have led to publication bias or missed key contributions

Dataset and Evaluation Inconsistency

Most studies used FaceForensics++ or Celeb-DF, which are high-quality but controlled datasets. This creates a dataset bias, underrepresenting real-world content diversity like lighting, occlusion, or compression artifacts.

Generalization and Robustness

The SLR highlights that ResNet-50 achieves 97% accuracy on FF++, but accuracy drops sharply on datasets like WildDeepfake or DFDC, pointing to poor cross-domain generalization.

4. CONCLUSION

This systematic literature review (SLR) examines current approaches for detecting Deepfake images using ResNet-50, based on 33 studies published between early 2020 and mid-2025. This paper focuses on Deepfake image detection using ResNet-50 and summarizes the overall findings as follows:

- The SLR indicates that ResNet-50, when used for Deepfake image detection, performs effectively, especially when combined with other models.
- Experimental results show that the FF++, Flickr/FFHQ, CelebA datasets yield better performance compared to other Deepfake image datasets.
- The SLR reveals that detection accuracy is the most commonly used metric for evaluating performance.
- The SLR suggests that ResNet-50 outperforms several selected CNN models in this application.

Overall, the findings suggest that ResNet-50 can effectively detect Deepfake images. However, given the rapid advancements in multimedia technology, deepfake image detection still encounters significant challenges. This SLR is intended to be a helpful resource for researchers in developing robust detection techniques and protective measures.

REFRENCES

1. Abdal Mashkoor, K. A. (2025). Classifying Real and AI-Generated Images Using Fine-Tuned ResNet50. Journal of Advanced Artificial Intelligence, Engineering and Technology, 01(02). https://doi.org/10.56147/aaiet.1.2.7
2. Abdullah, S. M., Cheruvu, A., Kanchi, S., Chung, T., Gao, P., Jadliwala, M., & Viswanath, B. (2024). An Analysis of Recent Advances in Deepfake Image Detection in an Evolving Threat Landscape (No. arXiv:2404.16212). arXiv. https://doi.org/10.48550/arXiv.2404.16212
3. Agarwal, S., Girdhar, N., & Raghav, H. (2021, January 28). A Novel Neural Model based Framework for Detection of GAN Generated Fake Images. 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India. https://doi.org/10.1109/confluence51648.2021.9377150
4. Aghasanli, A., Kangin, D., & Angelov, P. (2023). Interpretable-through-prototypes deepfake detection for diffusion models. 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 467–474. https://doi.org/10.1109/iccvw60793.2023.00053
5. Akram, O., Mohamed, A., Magdy, H., Abdellatif, M.M., Abdelghafar, S. (2025). Comparative Analysis of Custom CNN Architecture and MobileNet for Deepfake Image Detection. Proceedings of the 11th International Conference on Advanced Intelligent Systems and vol 238. https://doi.org/10.1007/978-3-031-81308-5_6
6. Alzahrani, A. (2024). Digital Image Forensics: An Improved DenseNet Architecture for Forged Image Detection. Engineering, Technology & Applied Science Research, 14(2), 13671–13680. https://doi.org/10.48084/etasr.7029.
7. Alzamily, J. Y. I., Ariffin, S. B., & Naser, S. S. A. (2022). CLASSIFICATION OF ENCRYPTED IMAGES USING DEEP LEARNING –RESNET50. . . Vol., 21.
8. Amrithat, D.V & Philimon, S., (2025): Deep GuardNet: A Novel CNN Architecture for Deep Fake Image Detection. International Conference on Machine Learning and Data Engineering Procedia Computer Science. 258, 811–818.
9. Bartos, G. E., & Akyol, S. (2024). Deep Learning for Image Authentication: A Comparative Study on Real and AI-Generated Image Classification.
10. Borade, S., Jain, N., Patel, B., Kumar, V., Godhrawala, M., Kolaskar, S., Nagare, Y., Shah, P., & Shah, J. (2024). ResNet50 DeepFake Detector: Unmasking Reality. Indian Journal of Science and Technology, 17(13), 1263–1271. https://doi.org/10.17485/ijst/v17i13.285
11. Berrahal, M., Boukabous, M., Yandouzi, M., Grari, M., & Idrissi, I. (2023). Investigating the effectiveness of deep learning approaches for deep fake detection. Bulletin of Electrical Engineering and Informatics, 12(6), 3853–3860. https://doi.org/10.11591/eei.v12i6.6221
12. Bird, J. J., & Lotfi, A. (2024). CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images. IEEE Access, 12, 15642–15650. https://doi.org/10.1109/access.2024.3356122
13. Brian, D., Joanna, B., Ben, P., Jikuo, Lu., Russ, H., Menglin, W., & Cristian, C. F. (2024): Dataset: Deepfake Detection Challenge (DFDC). https://doi.org/10.57702/rdr5mk24
14. Cao, X., & Gong, N. Z. (2021). Understanding the Security of Deepfake Detection (No. arXiv:2107.02045). arXiv. https://doi.org/10.48550/arXiv.2107.02045

15. CelebFaces Attributes (CelebA) Dataset, 2015. URL: https://www.kaggle.com/datasets/jessicali9530/celeba-dataset.
16. Chouhan, Y.A., Chudasama, C., Verma, D.K. (2025). Deepfake Detection Using ResNet50: Performance Analysis on Celeb-DF Dataset. In Artificial Intelligence Based Smart and Secured Applications. vol 2426. Springer, Cham. https://doi.org/10.1007/978-3-031-86296-0_28
17. Rajagukguk, I. P. E. N. Kencana, & I. G. N. L. W. Kusuma, (2025). Classification of Original and Fake Images Using Deep Learning- Resnet50. Advances in Computer Science Research (pp. 51–61). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-413-6_6
18. Doegar, A., Dutta, M., & Kumar, G. (n.d.). 1Department of Computer Science and Engineering, NITTTR, Chandigarh, India 2Magma Research and Consultancy Pvt. Ltd., Ambala, India.
19. Guo, Z., Yang, G., Chen, J. and Sun, X. (2021): Fake face detection via adaptive manipulation traces extraction network, Computer Vision and Image Understanding 204(2021). doi:10.1016/j.cviu.2021. 103170.
20. Gunukula, A. R., Das Gupta, H., & Sheng, V. S. (2025). Detecting AI-Generated Images Using a Hybrid ResNet-SE Attention Model. Applied Sciences, 15(13), 7421. https://doi.org/10.3390/app15137421
21. Gura, D., Dong, B., Mehiar, D., & Said, N. A. (2024). Customized Convolutional Neural Network for Accurate Detection of Deep Fake Images in Video Collections. Computers, Materials & Continua, 79(2), 1995–2014. https://doi.org/10.32604/cmc.2024.048238.
22. Hanady Sabah Abdul kareem & Mohammed Sahib Mahdi Altaei, (2023): Detection of Deep Fake in Face Images Based Machine Learning. I-Salam Journal for Engineering and Technology. 2790-4822, DOI: https://doi.org/10.55145/ajest.2023.02.02.001.
23. Hasan Abir, W., Rahman Khanam, F., Nabiul Alam, K., Hadjouni, M., Elmannai, H., Bourouis, S., Dey, R., & Monirujjaman Khan, M. (2023). Detecting Deepfake Images Using Deep Learning Techniques and Explainable AI Methods. Intelligent Automation & Soft Computing, 35(2), 2151–2169. https://doi.org/10.32604/iasc.2023.029653
24. Heo, Y.-J., Choi, Y.-J, Kim, B.-G. & Lee, Y.-W. (2021). Deepfake Detection Scheme Based on Vision Transformer and Distillation. arXiv preprint arXiv:2104.01353VI[cs.CV].
25. Ho, T. P., Pham, D. T., Dang, V. H., & Nguyen, N. H. (2024). Proposed method for against adversarial images based on ResNet architecture. Journal of Science and Technology on Information security. vol. 6, no. 1, p. 74.
26. Hou, Y., Guo, Q., Huang, Y., Xie, X., Ma, L., & Zhao, J. (2023). Evading DeepFake Detectors via Adversarial Statistical Consistency. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 12271–12280. https://doi.org/10.1109/cvpr52729.2023.01181.
27. Irfan,T., Arorab, B., Sandotra, N. & Raza, A. A, (2025). On Machine Learning and Deep Learning based Deepfake Generation and Detection. International Conference on Futuristic Trends in Networks and Computing Technologies. Procedia Computer Science 259 (2025) 1927–1936.
28. Ismail, A., Elpeltagy, M., Zaki, M. S., & Eldahshan, K. (2021). A New Deep Learning-Based Methodology for Video Deepfake. MDPI, Basel, Switzerland, 1–15.

29. Jasim, R. M., and Atia, T. S. (2023). An evolutionary-convolutional neural network for fake image detection. Indonesian Journal of Electrical Engineering and Computer Science. Vol. 29, No. 3, pp. 1657-1667, DOI: 10.11591/ijeecs.v29.i3.pp1657-1667.
30. Kaushik, A., Kaurb,P. Gupta, S. (2025). Deepfake Overview: Generation, Detection, Risks and Opportunities. Artificial Intelligence Review. https://link.springer.com/article/10.1007/s10462-024 10810-6
31. Killi, Rao, C.B, Narayanan, B. and Rao, C.S. (2024). Deep Fake Image Classification Using VGG-19 Model. Academic Journal. 28(2), p. 509, doi. 10.18280/isi.280228
32. Kosarkar, U., Sarkarkar, G., & Gedam, S. (2023). ScienceDirect ScienceDirect Revealing and Classification of Deepfakes Video ' s Images using a Customize Convolution Neural Network Revealing and Usha Classification of Deepfakes Video ' s using a Gedam Images Customize Convolution Neural Network Model. Procedia Computer Science, 218, 2636–2652. https://doi.org/10.1016/j.procs.2023.01.237.
33. Kroiß, L., & Reschke, J. (2025). Deepfake Detection of Face Images based on a Convolutional Neural Network (No. arXiv:2503.11389). arXiv. https://doi.org/10.48550/arXiv.2503.11389
34. Kumari, K., Priyanka, M., Deepthi, R. M., Gnana, D. D.and Swetha B. (2025). Deepfake Detection in Images & Videos Using XceptionNet: A Deep Learning Approach. International Journal of Advanced Research in Computer and Communication Engineering. Vol. 14, Issue 3, 411- 22.
35. Lamichhane, B., Thapa, K. and Yang, S-H. (2022). Detection of Image Level Forgery with Various Constraints Using DFDCFull and Sample Datasets. Sensors 2022, 22, 9121. https://doi.org/10.3390/ s22239121.
36. Latif, E.I.A. and Khalifa, N.E. (2023). COVID-19 digital forgery classification model using deep learning International Journal of Artificial Intelligent.12(4), 1821 - 1827.
37. Le, T.-N., Nguyen, H. HYamagishi, J. Echizen, I. (2021): Openforensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild, in: International Conference on Computer Vision, 2021.
38. Li, Y. Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics, 2023. URL: https://github.com/yuezunli/celeb-deepfakeforensics, original-date: 2019-10-02T00:31:06Z.
39. Lin, E. Y.-J. AttGAN-PyTorch, 2023. URL: https://github.com/elvisyjlin/AttGAN-PyTorch, original-date: 2018-11-28T08:56:52Z.
40. Li, B. and Lima, D. (2021). Facial expression recognition via ResNet-50. International Journal of Cognitive Computing in Engineering. 2(2021),57 -64.
41. Lipianina-Honcharenko, K., Telka, M., & Melnyk, N. (2024). Comparison of ResNet, EfficientNet, and Xception architectures for deepfake detection. International Workshop on Advanced Applied Information Technologies.
42. Meraj, F., Shahra, E.Q., Basurra, S. (2025). EfficientNET-Based Deepfake Detection: Performance Analysis Against StyleGAN3 and In-the-Wild Scenarios. Advances on Intelligent Computing and Data Science vol 254. https://doi.org/10.1007/978-3-031-91351-8_11

43. Mansoor, N., Iliev, A. (2024). DeepFake Detection Using Deep Learning. In: Arai, K. (eds) Intelligent Computing. SAI 2024. Lecture Notes in Networks and Systems, vol 1018. Springer, Cham. https://doi.org/10.1007/978-3-031-62269-4_14
44. Metin, Z. B., & Ozkaynak, F. (2025). Deep Learning-Based Hybrid Model for Detecting AI-Generated Fake Images. International Conference I Academic Research in Science, Technology and Engineering.
45. Nagagopiraju, V., Kancharla, A., Pallabothula, A., Jillalamudi, S. and Kakumanu, T.T. (2024). A EFFCIET DEEP FAKE FACE DETECTION USING DEEP INCEPTION NET LEARNING ALGORITHM. Journal of Computer and Mathematics Education. Vol.15 No.1(2024),138-141
46. NVlabs/stylegan, 2023. URL: https://github.com/NVlabs/stylegan, original-date: 2019-02 04T15:33:58Z.
47. NVlabs/stylegan2, 2023. URL: https://github.com/NVlabs/stylegan2, original-date: 2019 11-26T20:52:23Z.
48. Ondyari, FaceForensics++: Learning to Detect Manipulated Facial Images, 2023. URL: https://github.com/ondyari/FaceForensics, original-date: 2018-04-13T12:47:46Z.
49. Patel, Y., & Jain, M. (2024). Deepfake Image Detection Using Machine Learning and Deep Learning. 30(5), 14987–14993. https://doi.org/10.53555/kuey.v30i5.7980.
50. Patel, Y., Tanwar, S., Bhattacharya, P., Gupta, R., Alsuwian, T., Davidson, I. E., & Mazibuko, T. F. (2023). An Improved Dense CNN Architecture for Deepfake Image Detection. IEEE Access, 11, 22081–22095. https://doi.org/10.1109/access.2023.3251417
51. Pico. (2022). Pico Portal. PICO Portal. https://picoportal.org/legal-statement/
52. Potrimba, P. (2024) “What is resnet-50?” 13.03.2024. [Online]. Available: https://blog.roboflow.com/what-is-resnet-50/;visited:15.09.
53. Praditha, B., Mahajan, R.S., Vaishnavi, V. R., Vibha, T. K. and Pattan, D. (2023). Deepfake Image Detection Using Convolutional Neural Networks. International Research Journal. 10(7), 76 - 81.
54. Rana, S. M and Sung, A. H. (2024). Advanced Deepfake Detection using Machine Learning Algorithms: A Statistical Analysis and Performance Comparison. International Conference on Information and Computer Technologies. DOI 10.1109/ICICT62343.2024.00019
55. Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake Detection: A Systematic Literature Review. IEEE Access, 10, 25494–25513. https://doi.org/10.1109/access.2022.3154404
56. Ritu, D. and NEEtu, I.B. (2025). Deepfake Detection Using Hybrid Model. International Journal of Scientific Research and Engineering Developmen.- 8 (3), 3977 - 3988 https://doi.org/10.5281/ZENODO.15735942.
57. Rössler, A., Davide, C., Luisa, V., Christian R., Justus, T., and Matthias N. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. IEEE Trans actions on Information Forensics and Security, 8(7).
58. Samal, D., Agrawal, P. and Madaan,V. (2024). Deepfake Image Detection & Classification using Conv2D Neural Networks. Workshop on Advances in Computational Intelligence. 113 - 128.

59. Sandotra, N. and Arora, B. (2023). Deep Learning Model for Deepfake Image Detection: An Analytical Approach. International Conference on innovation for Industry Application. 1018 - 1026.
60. Sanghvi, D., Solanki, V., Sonarikar, A. and Jaiswal, R. (2023). Novel solution for Deepfake Image Detection using CNN and ResNet50 Architecture. International Journal of Computer Applications. (0975 – 8887).
61. Singh, J., & Ramachandra, R. (2022). DLDFD: Recurrence Free 2D Convolution Approach for Deep Fake Detection. Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 568–574. https://doi.org/10.5220/0010880500003124
62. Singh, D., Jain, T., Gupta, N., Tolani, B., Seeja, K.R. (2023): Fake Image Detection Using Ensemble Learning. Proceedings on International Conference on Data Analytics and Computing. vol 175. https://doi.org/10.1007/978-981-99-3432-4_30
63. Taeb M. Maryam R. and Hongmei C. (2022): Comparison of Deepfake Detection Techniques through Deep Learning." Journal of Cybersecurity and Privacy. 2(1).
64. Trinh, L., & Liu, Y. (2021). An Examination of Fairness of AI Models for Deepfake Detection. IJCAI International Joint Conference on Artificial Intelligence, 567–574. https://doi.org/10.24963/ijcai.2021/79
65. Vedhapriya, P., Vyshali, M., Yashoda, N., Lavanya, P., Vyshali, M., Yashoda, N., Lavanya, V., Kusuma. J. M. and Kusuma J. M. (2025). Deepfake Detection System using ResNet50 (based lightweight for static images) – Deep Dect. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering Impact Factor. 13, (7), 2321-5526.
66. Wang, B., Wu, X., Tang, Y., Ma, Y., & Shan, Z. (2023). Frequency Domain Filtered Residual Network for Deepfake Detection. 1–13.
67. Wang, Y., Vahid, Z. & Suxia, C. (2021): Fake Face Detection Using Local Binary Pattern and Ensemble Modeling." International Conference on Image Processing, 3917-3921.
68. Wangsadidjaja, A. A. (2023). Application of Deep Learning for Image Deepfake Detector Using Convolutional Neural Network Algorithm. Bit-Tech, 6 (2), 199.
69. Yadav, U., Dasarwar, P., Bondre, S., & Kalamkar, S. (2025). A Hybrid Approach for Robust Deep Fake Image Detection using Spatial and Frequency Domain Features. Engineering, Technology & Applied Science Research, 15(3), 22786–22791. https://doi.org/10.48084/etasr.10458
70. Yin, C., Guandao, Y., Andreas, V., Xun, H. & Serge, B. (2024): Dataset: Flickr 8k Dataset. https://doi.org/10.57702/8w6sv6a0Zahra Nazemi Ashani. (2024). Comparative Analysis of Deepfake Image Detection Method Using VGG16, VGG19 and ResNet50. Journal of Advanced Research in Applied Sciences and Engineering Technology, 47(1), 16–28. https://doi.org/10.37934/araset.47.1.1628