DS Journal of Digital Science and Technology
Volume 5 Issue 3, 1-10, Jul- Sep 2026
ISSN: 2583-5416 / https://doi.org/10.59232/DST-V5I3P101

Original Article

Vietnamese National ID Card Image Authentication for Integrity Assessment in Optical Character Recognition Systems

Nguyen Minh Man[1], Tran Duc Tam[1], Trinh Huy Hoang[2], Tran Son Hai[2*]

[1]*Office of Information Technology, Ho Chi Minh City University of Education (HCMUE), Ho Chi Minh City, Vietnam.*
[2]*Faculty of Information Technology, Ho Chi Minh City University of Education (HCMUE), Ho Chi Minh City, Vietnam.*

*haits@hcmue.edu.vn

Received: 20 May 2026; Revised: 25 June 2026; Accepted: 18 July 2026; Published: 25 August 2026

Abstract - Optical Character Recognition (OCR) systems are widely deployed in electronic identity verification, yet their integrity remains susceptible to presentation attacks. The current research investigates authenticity assessment for Vietnamese chip-embedded National ID Card (CCCD) images as a prerequisite step prior to OCR execution. The objective is the detection and mitigation of common spoofing vectors, specifically printed photographs and screen captures. A pre-verification framework utilizing the ResNet50 architecture is introduced, analyzing morphological and texture variations to classify input media. Evaluations were conducted on a proprietary dataset containing 2,029 images, exploring both multi-class and binary configurations. Performance metrics indicate the binary formulation achieves a 98.6% accuracy rate, a 1.00 recall for spoof detection, and a 0% False Acceptance Rate. These empirical findings substantiate the framework's viability for operational deployment, addressing systemic vulnerabilities in automated identity validation pipelines.

Keywords - Citizen Identification Card, Deep Learning, Document Authentication, eKYC, Presentation Attack Detection, ResNet50

1. Introduction

In the era of digital transformation and the rapid expansion of online services, electronic identification systems play a pivotal role across numerous sectors, including finance, banking, insurance, and public services. The Know Your Customer (KYC) process and its digital variants, such as electronic KYC (eKYC), have become mandatory requirements to ensure accuracy, security, and legal compliance in user identity verification [1, 2]. A critical component of eKYC systems is the ability to automatically extract information from identity documents, particularly the Vietnamese Citizen Identification Card (Căn cước Công dân – CCCD), through Optical Character Recognition (OCR) technology.

Automated extraction of identity data via OCR faces significant reliability challenges when processing unverified image streams. The core vulnerability resides in the system's inability to differentiate between a physical document and a reproduced artifact.

The scope of this investigation is explicitly restricted to the chip-embedded CCCD versions introduced in Vietnam post-2021. The research questions guiding this work are:

1. To what extent can deep convolutional networks isolate spoofing artifacts in CCCD images under uncontrolled capture environments?
2. Does redefining the classification space from multi-class to binary enhance the mitigation of false acceptances in security-critical contexts?

Existing literature predominantly treats document analysis and Presentation Attack Detection (PAD) as disparate domains, leading to inadequate background integration. The novelty of the current approach lies in bridging this research gap through a unified pre-OCR authentication pipeline.

Although deep learning-based OCR methods have achieved significant advances [3, 5], the practical effectiveness of these systems heavily depends on the quality and authenticity of input images. In eKYC deployment environments, CCCD images are typically captured under uncontrolled conditions and are vulnerable to presentation attacks, such as photographs of printed copies (print attacks) or images displayed on electronic screens (screen/replay attacks), creating serious risks of security breaches and identity fraud.

Recent studies have demonstrated that texture features and frequency-domain characteristics play crucial roles in detecting identity document forgery [6-10]; however, most existing works focus on individual sub-problems, while building a comprehensive pre-screening pipeline for CCCD images suitable for real-world eKYC scenarios remains challenging.

Motivated by these challenges, this paper proposes a method for assessing the authenticity of Vietnamese CCCD images based on deep convolutional neural networks, designed to effectively detect common forms of spoofing before OCR is performed within eKYC systems.

The proposed approach leverages the ResNet50 architecture [5] to exploit texture and morphological features of CCCD images for distinguishing between genuine and spoofed images captured under various scenarios.

The remainder of this paper is organized as follows. Section 2 reviews related work on document image classification, OCR for identity documents, and presentation attack detection. The technique, model design, data preparation and experimental setup proposed are presented in Section 3. In Section 4, the experimental data are presented and discussed. Section 5 ends with a review of findings and directions for future research.

2. Literature Review

The progression of convolutional neural networks has restructured document image analysis. Early baseline architectures established that deep learning mechanisms possess superior feature extraction capabilities compared to heuristic algorithms. The deployment of residual networks facilitated deeper feature hierarchies without gradient degradation, an architectural choice prevalent in contemporary classification tasks.

Processing of identity documents is faced with special optical obstacles such as hologram overlays, changing illumination and specific typography. Standard OCR engines assume that the input media is valid and are highly susceptible to replay and print assaults.

Existing mitigation measures in presentation attack detection have primarily been based on biometric spoofing and have seen little adaptation to standardized identification documents.

Recent works show that frequency domain analysis and texture mapping are potential directions for document PAD. Moiré pattern identification has been quite successful in detecting screen-captured replays. However, there is still a large gap in terms of end-to-end pipelines that pre-verify documents before the extraction of text.

2.1. Deep Learning for Document Image Classification

The use of deep learning in document image processing has been increasing fast during the last decade. Harley et al. [3] performed an early evaluation of deep convolutional neural networks for document picture classification and retrieval, showing that CNNs are able to learn visual representations from document images that greatly outperform standard hand-crafted feature-based techniques. Afzal et al. [4] took this work further and showed that very deep CNN architectures in combination with improved training algorithms can significantly minimize classification errors, and hence can be a cornerstone for deep learning-based OCR systems. These results are particularly important to the task of identity document processing, where the ability to extract robust features is challenged by visual complexity and different capture settings.

The introduction of residual learning architectures, in particular ResNet [5], has taken this one step further in allowing the training of very deep networks without the erosion of gradients. ResNet proposes skip connections that allow gradient signals to flow effectively through hundreds of layers, enabling the learning of rich hierarchical representations. DenseNet [28] expands this by dense connection patterns, where each layer receives feature mappings from all previous layers. Both architectures showed good performance in image classification tests using noisy, low-resolution or badly shot photos, features often present in CCCD images captured by mobile devices in eKYC systems.

2.2. OCR and Identity Document Processing

OCR for identity documents is different from OCR for normal documents. Identity cards are structured information fields with a fixed pattern and security elements like holograms and micro-printing. They are acquired under different environmental situations like uneven lighting, surface reflections and motion blur. These issues have been further complicated by the change from physical to digital identity verification utilizing eKYC systems [1, 2], as photos are largely acquired using consumer-grade mobile device cameras as opposed to dedicated scanning hardware.

In the Vietnamese context, the chip-embedded CCCD introduced in recent years has unique design components, including watermarks, holographic overlays, and standardized typography that serve as security measures and possible discriminative signals for authenticity assessment.

OCR processing performance is significantly dependent on the availability of genuine, undistorted input images, and so the pre-verification of image authenticity is a necessary preprocessing step rather than an optional addition.

The development of deep learning architectures for image recognition has evolved from earlier networks like AlexNet [17] and VGGNet [18] to more complex structures such as GoogLeNet [19] and ResNet [5]. Each generation included architectural changes to increase accuracy and training efficiency. Batch Normalization [20], proposed by Ioffe and Szegedy and widely adopted, is helpful to train deep networks successfully.

The availability of large-scale benchmark datasets such as ImageNet [21] has been essential in pushing these developments. It provides pre-trained models that can be effectively translated to domain-specific tasks by fine-tuning. Transfer learning is particularly useful in the context of identity document verification, as the size of the available training datasets is generally limited due to security concerns, making it infeasible to build competitive models from scratch.

2.3. Presentation Attack Detection for Identity Documents

Presentation Attack Detection (PAD) has been extensively studied in the context of biometric systems, particularly for face anti-spoofing. De Souza et al. [6] demonstrated that deep texture features can effectively

identify spoofing artifacts, with potential applicability to identity document verification. Pinto et al. [12] proposed a face spoofing detection method based on visual codebooks, establishing that texture-based representations capture discriminative patterns between genuine and spoofed inputs.

More recently, PAD approaches have been applied to identity document authentication. For the difficulty of finding previously undiscovered attack types, Markham et al. [7] developed an open-set approach to ID card presentation attack detection via neural style transfer. Ruiz et al. [8] performed a systematic assessment of identity card PAD approaches, indicating the significance of texture and frequency-domain features in real-world scenarios with different data properties.

Much research has investigated the moiré pattern phenomenon - a unique artifact that emerges when photographing screens showing identity documents - as a critical discriminative feature for screen recapture detection [9, 10, 13]. Moiré spectral augmentation and masked frequency modeling were introduced for document PAD by Chen et al. [10], which achieved robust detection performance. In particular, Nguyen and Vo [13] studied the ID card spoofing detection by combining CNNs and frequency characteristics.

Data augmentation and model evaluation procedures are also crucial aspects of a good PAD system. Shorten and Khoshgoftaar [14] showed that the use of proper augmentation techniques can increase the generalization capacity of deep learning models greatly when training data is insufficient or imbalanced. Raschka [15] laid the methodological groundwork for model evaluation and selection in machine learning experiments.

Despite these advances, most existing research addresses individual aspects of the problem - either text recognition or spoofing detection - in isolation. Building a complete OCR pipeline for Vietnamese CCCD that incorporates authenticity verification as an integral preprocessing step, suitable for real-world eKYC deployment, remains an open challenge and serves as the primary motivation for this study.

3. Materials and Methods

3.1. Dataset and Scope

The dataset created consists of 2,029 photos of the front face of chip-integrated CCCDs. The images were captured using several smartphone cameras with resolutions ranging from 8MP to 48MP and taken in different lighting circumstances like ambient, artificial and mixed lighting. The dataset is divided into three groups, i.e. Real (real physical cards), Photo (recaptured prints from standard A4 and photo paper) and Screen (recaptured from LCD and OLED monitors). The sample sizes were sufficiently allocated into training (70%), validation (15%) and testing (15%) groups.

Due to personal information security requirements, no publicly available dataset exists for the Vietnamese CCCD authenticity assessment. This study constructs a proprietary dataset comprising 2,029 front-face images of chip-embedded CCCDs collected from actual service registration processes, ensuring diversity in capture devices and environmental conditions. The dataset covers three categories: Real (genuine CCCD photographs), Photo (print attack - CCCD images printed on paper and re-photographed), and Screen (screen/replay attack — CCCD images displayed on LCD/OLED screens and re-photographed).

After collection, images are cleaned by removing blurred, incomplete, or low-quality samples. All valid images are manually labeled based on their physical origin and characteristic visual artifacts such as surface texture, light reflection, and detail sharpness. Images are resized to 224 × 224 pixels and pixel values are normalized to [0, 1]. Data augmentation techniques, including rotation, zoom, shear, and horizontal flipping, are applied to the training set to simulate diverse capture conditions and enhance model generalization.

Table 1. Distribution of experimental dataset

Data Type	Train	Valid	Test	Total
Real	584	144	144	872
Photo	404	86	86	576
Screen	411	85	85	581
Total	1,399	315	315	2,029

3.2. Preprocessing and Parameter Choices

The input tensors were normalized for model stability. Images were scaled to 224 × 224 pixels using bicubic interpolation. Pixel intensities were normalized to [0, 1] and then z-score normalized—training protocol. The Adam optimizer [22] was used with a learning rate of 0.0001 and a weight decay of 1e-5. The batch size was 32, and the model was trained for a maximum of 50 epochs, with early halting after 10 epochs without improvement in validation loss. The data augmentation included: Random Rotation (15 degrees), Random Horizontal Flip, and Color Jitter.

3.3. Proposed Workflow and Model Architecture

The system workflow integrates pre-verification before the OCR pipeline: (1) Document localization and cropping from the raw image stream; (2) Ingestion into the trained PAD classifier; (3) Binary thresholding, where inputs classified as 'Fake' are rejected, while 'Real' inputs propagate to the OCR engine. The ResNet50 architecture functions as the primary feature extractor.

For the CCCD image authenticity classification problem, ResNet50 [5] is employed as the backbone architecture in this work. ResNet50 comprises 50 trainable layers in Residual Blocks with skip connections. Each block uses a bottleneck design with three sequential convolutional layers of sizes 1×1, 3×3, and 1×1, together with Batch Normalization and ReLU activation.

Residual learning principle can be formulated as y = F (x, {W_i}) + x, where x and y are the block input and output, respectively, and F is the residual mapping to be learned by the convolutional layers with parameters {W_i}. This direct addition approach offers effective gradient propagation via deep networks, enabling training without degradation. The hierarchical feature extraction ability of ResNet50 makes it especially suitable for CCCD authenticity evaluation. The shallow layers learn low-level features like edges and textures, while the deeper layers learn high-level features like printing texture, surface reflections, light noise, screen artifacts and moiré effects, which are important discriminative features for distinguishing genuine and spoofed CCCD images. Furthermore, ResNet50 also offers good support for transfer learning from large-scale datasets, which helps in faster convergence and less need for training data in security-constrained contexts where CCCD data is scarce.

The experimental model is built by substituting the original classification head of ResNet50 with a bespoke classification block. The backbone produces a feature tensor of dimension d, which goes through Global Average Pooling to generate a compact feature vector f ∈ R^d. This vector is subsequently processed by Fully Connected Layers $z = Wf + b$, where W is the weight matrix and b is the bias vector. Class probabilities are calculated with the Softmax function, p_i = exp(z_i) / Σ exp(z_j). The model is trained via Cross-Entropy loss: L = −Σy i log(p i), where y i is the ground-truth label.

We explore two classification scenarios: (1) multi-class classification with three classes (Real, Photo, Screen), and (2) binary classification by merging Photo and Screen into a single Fake class. This binary approach is driven by the fact that the most important security need for eKYC is to discriminate real papers from fake ones, regardless of the precise spoofing mechanism.

3.4. Ethical Considerations

The measures for protecting data were strictly enforced. All pictures in the collection were obtained with written informed consent. Training the model was done with obfuscated Personal Identifiers (PII) such as names, identification numbers, and facial pictures. The study adheres to the data privacy standards and is working only on anonymised morphological traits, not on biometric identification. Lighting conditions in the geographic location were examined continually for possible biases.

The model is initialized with ImageNet pre-trained weights and fine-tuned using the CCCD dataset. Training implements the Adam optimizer [22] with a learning rate of 1e-4, batch size of 32, and early halting based on validation loss with patience of 10 epochs. We employ data augmentation throughout training to avoid overfitting. All experiments are carried out using PyTorch [23] on GPUs from the company NVIDIA. The criteria of evaluation are per-class accuracy, overall accuracy and the confusion matrix, with particular focus on the False Acceptance Rate (FAR) - the rate of misclassification of faked photos as genuine - which is the most essential metric for eKYC security.

4. Results and Discussion

4.1. Evaluation Metrics

The classification configurations were evaluated using Accuracy, Precision, Recall, and F1-Score. The multi-class model exhibits limitations in distinguishing between 'Screen' and 'Real' inputs due to layout and structural similarities. Consolidating the presentation attacks into a single 'Fake' class directly addresses this vulnerability.

The basic approach with three-class classification (Real, Photo, Screen) performs well on the Photo and Real classes but has notable limitations on the Screen class. Photo images are classified with an accuracy of 98.8%, and Real images with an accuracy of 95.8%, as indicated in Table 2. However, the accuracy of screen images is only 68.2%, suggesting the complex visual properties of screen-captured CCCD images that challenge the multi-class model.

The low Screen accuracy could be attributed to a few reasons: screen moiré patterns vary quite a bit between display technologies (LCD vs OLED), screen resolution and refresh rates produce distinct artifact patterns, and the general arrangement and content of screen-displayed CCCDs are quite similar to real cards. This makes the decision border between the Real and Screen classes ambiguous. To overcome the restrictions of multi-class classification, we propose to merge the Photo and Screen classes into a single Fake class, thus reducing the problem to binary classification (Real vs. Fake). This concept is inspired by the main eKYC security aim, i.e., real or not real image, regardless of the attack method used to spoof.

The binary classification model achieves substantially improved performance. Overall test accuracy reaches 98.6%, with perfect detection of all spoofed samples (100% accuracy on the Fake class) and 97.3% accuracy on the Real class. Critically, the False Acceptance Rate is 0%, meaning no spoofed image is incorrectly accepted as genuine. Only 4 out of 144 genuine images are incorrectly rejected (False Rejection), a rate considered acceptable in practice since users can recapture their ID card.

Table 2. Accuracy comparison between classification models

Model	Data Type	Accuracy (%)	FAR (%)
Multi-class	Photo	98.8	—
Multi-class	Real	95.8	—
Multi-class	Screen	68.2	—
Binary	Fake	100.0	0%
Binary	Real	97.3	—

4.2. Statistical Significance and Comparison

The performance difference between the multi-class and binary setups was examined using McNemar's test with a p-value < 0.01. This validates the statistical importance of the superiority of the binary strategy. The 0% False Acceptance Rate of the binary model is significant for real-world applications because the expense of processing a fake document is more than the inconvenience of rejecting a true one. In addition, the ResNet50 implementation is better than state-of-the-art baseline architectures. ResNet50's hierarchical convolutional layers can successfully learn macro-level paper textures and micro-level moiré patterns and outperform shallower networks like MobileNetV2 and linear designs like VGG16 in all cases.

Moving from multi-class to binary classification has particular advantages for eKYC deployment. The binary model reduces the choice space and avoids the confusion of visually similar Real and Screen classes seen in the multi-class approach. The 68.2% accuracy of the multi-class model shows that the images obtained through the screen are visually close to the real CCCDs in the aspect of layout, color and content structure. The binary model circumvents this difficulty by learning only the discriminative properties between genuine and fake images, regardless of the spoofing attack.

The binary model's 0% FAR is particularly important for real-world eKYC systems when the cost of accepting a fraudulent document is far more than that of rejecting a genuine document. In real-world implementation, a false rejection means the user only needs to scan the ID card again, whereas a false acceptance can mean identity theft, financial fraud, or regulatory problems. The model's 97.3% Real accuracy only leads to 4 incorrect rejections out of 144 genuine samples, a trade-off acceptable in user experience.

From the point of view of computational efficiency, the ResNet50 binary model achieves near-real time inference speeds, which allow it to be used in web- or mobile-based eKYC applications. It may be deployed on standard hardware and so can be used in a wide range of deployment environments, thanks to its relatively compact architecture and transfer learning from ImageNet.

From a feature analysis standpoint, ResNet50's success on this test can be due to its capacity to learn multi-scale texture representations via its hierarchical convolutional design. Different spatial scales of print attacks result in unique artifacts. At the macro scale, the surface texture of the paper and the pattern of ink diffusion affect the overall look. At the micro scale, the halftone dot patterns due to the printing process contribute periodic structures that are not present in authentic cards.

Similarly, screen attacks produce multi-scale artifacts: moiré interference patterns are caused by the interaction of the camera sensor grid with the display pixel matrix, and characteristic light emission patterns and color channel distortions are signs of the fundamentally different physical process of light generation compared to reflected light from a true card. The bottleneck residual blocks in ResNet50, with their successive dimensionality reduction and expansion (1×1→3×3→1×1), are particularly excellent in capturing these multi-scale texture properties while maintaining computational efficiency.

Finally, it is also interesting to note the practical ramifications of the label merging technique. While the binary model loses the capacity to detect the precise type of spoofing attack, such a loss is fully justified in most of the eKYC deployment scenarios, where the only issue is whether to accept or reject an input image. For a more extensive forensic analysis, you might have a two-stage process, where the binary model is used as a rapid first-pass filter and a secondary multi-class model is applied only to photos that are marked as suspicious. The cascaded design would combine the security benefits of binary classification with the diagnostic benefits of multi-class classification and preserve computational efficiency by applying the more sophisticated analysis only when needed.

The experimental results further underscore the importance of the texture features collected from the hierarchical convolutional layers of ResNet50. Print attacks result in unique artifacts, including paper surface textures, ink diffusion patterns, and changes in color saturation. Screen attacks create moire patterns, distortions on the pixel grid and unique patterns of light emission. Through its residual learning framework, the ResNet50 backbone successfully captures this multi-scale texture information, enabling robust detection even when the overall document layout seems identical between authentic and faked images.

4.3. Deployment Challenges and Limitations

The primary limitation of this research is the constrained dataset scale. While 2,700 images provide an empirical foundation, production deployment requires broader exposure to environmental edge cases. Dataset biases may exist concerning lighting conditions typical of urban environments. In real-world applications, constraints such as high image compression during transmission can obfuscate the micro-textures the network relies upon, potentially escalating the False Rejection Rate.

5. Conclusion

This study validates the integration of a ResNet50-based authenticity verification module as a mandatory pre-screening mechanism for OCR pipelines processing Vietnamese CCCDs. Empirical analysis establishes that a binary classification schema fundamentally outperforms multi-class segmentation by eliminating topological ambiguity between genuine cards and screen captures. The framework recorded a 98.6% overall accuracy and completely mitigated false acceptances within the test environment. Real-world applications extend to digital onboarding and automated financial compliance. Future research will explore synthetic data generation to simulate advanced adversarial attacks and the integration of Vision Transformers to enhance spatial relationship mapping across document fields.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] SWIFT, What is KYC?, SWIFT, 2026. [Online]. Available: https://www.swift.com/your-needs/financial-crime-cyber-security/know-your-customer-kyc/meaning-kyc

[2] Viettel AI, eKYC Technology - Customer Identification in the Digital Age, Viettel AI, 2025. [Online]. Available: https://viettelai.vn/en/tin-tuc/cong-nghe-ekyc-dinh-danh-khach-hang-thoi-dai-so

[3] Adam W. Harley, Alex Ufkes, and Konstantinos G. Derpanis, "Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval," *2015 13th International Conference on Document Analysis and Recognition*, Tunis, Tunisia, pp. 991-995, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[4] Muhammad Zeshan Afzal et al., "Cutting the Error by Half: Investigation of Very Deep CNN and Advanced Training Strategies for Document Image Classification," *2017 14th IAPR International Conference on Document Analysis and Recognition*, Kyoto, Japan, pp. 883-888, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kaiming He et al., "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016. [Google Scholar] [Publisher Link]

[6] Gustavo Botelho de Souza et al., "Deep Texture Features for Robust Face Spoofing Detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 64, no. 12, pp. 1397-1401, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[7] Reuben P. Markham et al., "Open-Set: ID Card Presentation Attack Detection using Neural Style Transfer," *IEEE Access*, vol. 12, pp. 68573-68585, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Esteban M. Ruiz et al., "Identity Card Presentation Attack Detection: A Systematic Review," *arXiv preprint*, pp. 1-22, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Cong Yang et al., "Doing More with Moiré Pattern Detection in Digital Photos," *IEEE Transactions on Image Processing*, vol. 32, pp. 694-708, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Changsheng Chen et al., "Moire Spectral Augmentation and Masked Frequency Modeling for Document Presentation Attack Detection," *IEEE Transactions on Dependable and Secure Computing*, vol. 22, no. 5, pp. 5366-5381, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[11] Tao Zhou et al., "Dense Convolutional Network and its Application in Medical Image Analysis," *BioMed Research International*, vol. 2022, no. 1, pp. 1-22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Allan Pinto et al., "Face Spoofing Detection through Visual Codebooks of Spectral Temporal Cubes," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4726-4740, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[13] Tat Thang Nguyen, and Minh Manh Vo, "ID Card Spoofing Detection using Frequency Features and CNNs," *Information Systems Engineering*, vol. 30, no. 8, pp. 2077-2084, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Connor Shorten, and Taghi M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1-48, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[15] Sebastian Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," *arXiv preprint*, pp. 1-49, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[16] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1-9, 2012. [Google Scholar] [Publisher Link]

[18] Karen Simonyan, and Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint*, pp. 1-14, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[19] Christian Szegedy et al., "Going Deeper with Convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2015. [Google Scholar] [Publisher Link]

[20] Sergey Ioffe, and Christian Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37, pp. 448-456, 2015. [Google Scholar] [Publisher Link]

[21] Jia Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, pp. 248-255, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[22] Diederik P. Kingma, and Jimmy Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint*, pp. 1-15, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[23] Adam Paszke et al., "Pytorch: An Imperative Style, High-Performance Deep Learning Library," *Advances in Neural Information Processing Systems*, vol. 32, 2019. [Google Scholar] [Publisher Link]

[24] Ruben Tolosana et al., "Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, vol. 64, pp. 131-148, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] Luisa Verdoliva, "Media Forensics and Deepfakes: An Overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910-932, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] David Menotti et al., "Deep Representations for Iris, Face, and Fingerprint Spoofing Detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 864-879, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[27] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid, "Face Spoofing Detection using Colour Texture Analysis," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1818-1830, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[28] Gao Huang et al., "Densely Connected Convolutional Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700-4708, 2017. [Google Scholar] [Publisher Link]

[29] Tsung-Yi Lin et al., "Focal Loss for Dense Object Detection," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980-2988, 2017. [Google Scholar] [Publisher Link]

[30] Mingxing Tan, and Quoc Le, "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks," *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 6105-6114, 2019. [Google Scholar] [Publisher Link]