International Journal of Research and Development in Applied Science and Engineering (IJRDASE)
ISSN: 2454-6844

Unmasking Digital Deception: A Review of SVM and CNN Techniques for Deepfake Detection in Images and Videos

Anuj Dwivedi[1], Shiwangi Choudhary[2]
Dept. of Computer Science and Engineering,
Rameshwaram Institute of Technology & Management, (AKTU), Lucknow, India

Abstract— **The rapid advancement of deep learning technologies has led to the proliferation of deepfake media, posing significant threats to privacy, trust, and security in the digital realm. Deepfakes, which involve the manipulation of images and videos to create realistic yet fraudulent content, challenge traditional detection methods due to their high visual fidelity. This review paper, titled "Unmasking Digital Deception: A Review of SVM and CNN Techniques for Deepfake Detection in Images and Videos," explores the latest developments in Support Vector Machine (SVM) and Convolutional Neural Network (CNN)-based approaches for identifying and combating deepfake content. We analyze key architectural variations, feature extraction mechanisms, training strategies, and evaluation metrics employed across state-of-the-art models. Furthermore, the paper highlights the strengths and limitations of each method, the role of hybrid systems combining SVM and CNN, and the implications for real-world deployment in social media monitoring, digital forensics, and cybersecurity. By consolidating recent research efforts, this review aims to support the development of more robust and generalizable deepfake detection frameworks.**

Keywords— **Deepfake Detection, Support Vector Machine (SVM), Convolutional Neural Network (CNN), Image Forensics, Video Manipulation, Digital Deception, Hybrid Models, Machine Learning, Deep Learning, Media Authentication.**

I. INTRODUCTION

The digital era has witnessed an unprecedented surge in the creation and distribution of manipulated multimedia content, largely driven by the advancement of deep learning techniques. Among these, deepfakes—synthetic media where a person in an image or video is replaced with someone else's likeness—have emerged as a significant concern due to their potential misuse in political misinformation, cybercrime, identity theft, and social manipulation (Chesney & Citron, 2019). The core technologies enabling deepfakes, such as Generative Adversarial Networks (GANs) and autoencoders, have made it increasingly difficult to distinguish between authentic and forged content using the human eye or conventional digital forensic techniques (Korshunov & Marcel, 2018).

To address this challenge, researchers have turned to machine learning, particularly Support Vector Machines (SVM) and Convolutional Neural Networks (CNN), for automated deepfake detection. SVM, a classical supervised learning method, has shown promise in binary classification tasks using handcrafted features such as facial landmarks, eye blinking patterns, and head movements (Afchar et al., 2018). On the other hand, CNNs, a deep learning architecture known for their efficacy in image and video analysis, can automatically extract hierarchical features from raw data, making them especially suitable for detecting subtle visual artifacts introduced by generative models (Nguyen et al., 2019).

While CNNs offer robust performance in detecting frame-level and spatial inconsistencies, SVMs complement them by performing well on reduced, interpretable features and in hybrid pipelines. Several studies have explored combining these two approaches to leverage their respective strengths (Li et al., 2020). Additionally, benchmark datasets such as FaceForensics++, DFDC, and Celeb-DF have facilitated standardized evaluation and comparison of proposed models (Rossler et al., 2019).

Despite promising results, deepfake detection remains an evolving field plagued by generalization issues, adversarial robustness challenges, and real-time implementation constraints. The continued development of deepfake generation techniques necessitates the parallel evolution of detection systems that are adaptive, scalable, and resilient. This review aims to consolidate current research on SVM and CNN-based deepfake detection techniques, assess their performance across various datasets, and explore future directions for more secure and reliable digital content authentication.

II. LITERATURE SURVEY

The growing prevalence of deepfakes has prompted extensive research into automated detection methods, particularly those leveraging machine learning (ML) and deep learning (DL) techniques. Among these, Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) have emerged as prominent tools due to their capacity to detect subtle inconsistencies introduced during the generation of synthetic content.

Afchar et al. (2018) proposed MesoNet, a CNN-based architecture with low computational complexity designed for real-time detection of facial forgeries. The model performed effectively on compressed videos by focusing on mesoscopic features, which are neither too fine (like pixel-level) nor too global (like entire face context). Similarly, Rössler et al. (2019) introduced FaceForensics++, a benchmark dataset comprising manipulated videos and evaluated multiple CNN architectures, including XceptionNet, for deepfake detection. Their work demonstrated that deeper networks can capture the intricate spatial and temporal inconsistencies introduced during manipulation.

On the other hand, SVM-based approaches have shown promising results, especially when used with handcrafted features. Li et al. (2018) utilized SVMs to classify deepfakes based on physiological signals such as eye blinking patterns, which are often missing or inconsistent in synthetically generated videos. Their system used spatio-temporal features extracted from facial regions, demonstrating that even basic physiological cues can aid in detection.

Khalid and Woo (2021) proposed a hybrid approach combining CNN for feature extraction and SVM for classification. Their model outperformed traditional CNN classifiers in terms of generalization, particularly when trained on one dataset and tested on another, highlighting the potential of hybrid systems in handling cross-domain deepfake detection.

Nguyen et al. (2019) conducted a comprehensive survey on DL-based deepfake generation and detection methods. They concluded that while CNNs are effective in detecting manipulated content, their performance suffers when evaluated on datasets not seen during training. This observation was further echoed by Verdoliva (2020), who emphasized the generalization problem in deepfake detectors and advocated for anomaly detection and forensic-inspired techniques to enhance model robustness.

Recent research has also explored temporal features. Sabir et al. (2019) introduced a recurrent neural network (RNN) architecture on top of CNN to capture temporal dependencies in deepfake videos. Though not purely CNN or SVM-based, their work indicated that incorporating motion consistency across frames can improve detection accuracy.

Additionally, Zhang et al. (2020) highlighted that CNN-based detectors are vulnerable to adversarial attacks and proposed adversarial training to improve robustness. The study emphasized the importance of incorporating defensive mechanisms in detection systems to withstand manipulation by evolving deepfake generation algorithms.

In terms of datasets, Li et al. (2020) leveraged the Celeb-DF dataset to benchmark various CNN and SVM-based models, finding that models trained on FaceForensics++ often struggled with the higher visual quality and diversity of Celeb-DF. This further underscored the importance of diverse training data for building generalized detection systems.

Despite the progress, challenges remain. Most detectors suffer performance degradation under video compression, noise, and low resolution. Furthermore, real-time implementation and explainability of detection remain critical areas for further exploration (Mirsky & Lee, 2021).

Table 1: Literature Review Table Based On Previous Year Research Paper Key Findings

S. No.	Title of the Paper	Author(s)	Year	Techniques Used	Dataset Used	Key Findings
1	MesoNet: a Compact Facial Video Forgery Detection Network	Afchar et al.	2018	CNN (Meso-4, MesoInception-4)	Deepfake, FaceForensics	Shallow CNNs effective for compressed deepfake videos
2	FaceForensics++	Rössler et al.	2019	CNN (XceptionNet)	FaceForensics++	XceptionNet achieves state-of-the-art detection accuracy
3	Exposing DeepFake Videos by Detecting Eye Blinking	Li et al.	2018	SVM + Eye Blink Detection	CelebA	Eye blink inconsistencies indicate deepfake presence
4	Deep Learning for Deepfakes Creation and Detection: A Survey	Nguyen et al.	2019	CNN, GAN	Multiple datasets	Comprehensive review of generation and detection methods
5	Hybrid CNN-SVM Model for DeepFake Detection	Khalid & Woo	2021	CNN + SVM	FaceForensics++	Hybrid improves generalization across datasets
6	In Ictu Oculi: Eye Blinking for Fake Face Detection	Li et al.	2018	SVM + Temporal Pattern	YouTube, Custom	Temporal patterns help detect generative anomalies
7	DeepFakes and Beyond: A Survey	Tolosana et al.	2020	CNN, SVM	FaceForensics++, DFDC	Overview of face manipulation

	of Face Manipulation Detection					detection techniques
8	Detecting Deep-Fake Videos from Biological Signals	Ciftci et al.	2020	SVM on rPPG	Celeb-DF	Physiological signals like pulse used to detect fakes
9	Detecting AI-synthesized Speech using SVM	Wu et al.	2020	SVM	ASVspoof	SVM can classify audio-based deepfakes
10	Celeb-DF: A New Dataset for DeepFake Forensics	Li et al.	2020	CNN (XceptionNet)	Celeb-DF	Harder dataset to benchmark generalization performance
11	Fighting Deepfake with Adversarial Examples	Zhang et al.	2020	CNN + Adversarial Training	FaceForensics++	Improved model robustness with adversarial training
12	A CNN+LSTM Approach for DeepFake Video Detection	Sabir et al.	2019	CNN + LSTM	FaceForensics++	LSTM captures temporal inconsistencies
13	Deep Learning Techniques for Deepfake Detection	Tariq et al.	2021	CNN	DFDC, Deepfake TIMIT	Ensemble CNNs provide high detection accuracy
14	On the Generalization of Deepfake Detection	Agarwal et al.	2020	CNN	Celeb-DF, DFDC	Generalization remains a major challenge
15	Autoencoder-based Feature Extraction with SVM Classifier	Korshunov & Marcel	2018	Autoencoder + SVM	Deepfake TIMIT	Hybrid methods offer interpretable features
16	Deep Forensics: A CNN-based Deepfake Detector	Amerini et al.	2019	CNN	FaceForensics++	Pixel-level inconsistencies exploited by CNN
17	Detecting Deepfakes via Facial Regions Segmentation	Dang et al.	2020	Region-based CNN	FaceForensics++, Celeb-DF	Segment-wise analysis boosts accuracy
18	SVM-based Temporal Pattern Recognition for Deepfake Detection	Sun et al.	2020	SVM	YouTube	Good for short-form and face-focused deepfakes
19	FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals	Ciftci et al.	2020	CNN + SVM	Custom	Uses heart rate estimation for detection
20	Evaluation of Deepfake Detection using Xception and ResNet	Gupta et al.	2021	CNN (Xception, ResNet)	FaceForensics++	Deep CNNs yield best results with high-quality inputs
21	A Survey on SVM for Face Forgery Detection	Bose et al.	2021	SVM	Multiple	SVMs still relevant when paired with meaning

						ful features
22	Video-based CNN Model for Deepfake Detection	Nguyen et al.	2020	3D-CNN	DFDC	Spatio-temporal CNNs outperform frame-wise models
23	Deepfake Detection via Eye Movement Consistency	Hussain et al.	2022	SVM + Eye Tracking	Custom	Eye motion cues reveal forgeries
24	Detection of Deepfake Faces with Fused CNN and SVM Features	Varol et al.	2021	CNN + SVM	Celeb-DF, FaceForensics++	Feature-level fusion improves accuracy
25	Adversarial Robustness in Deepfake Detection	Mehra et al.	2022	CNN + Adversarial Defense	DFDC	Adversarial perturbations reduce model reliability

III. ALGORITHM

In the context of deepfake detection, two primary algorithmic paradigms are prominently utilized: Support Vector Machine (SVM) and Convolutional Neural Network (CNN). Each follows distinct algorithmic workflows designed to detect spatial, temporal, or physiological anomalies introduced by synthetic content generation methods.

A. Support Vector Machine (SVM) Algorithm for Deepfake Detection

Step-by-Step Workflow:

Preprocessing:

- Convert video to frame sequences.
- Detect and crop facial regions using face detection methods (e.g., Haar Cascade, MTCNN).

Extract handcrafted features such as:

- Blink frequency
- Head movement
- Local Binary Patterns (LBP)
- Histogram of Oriented Gradients (HOG)
- Physiological signals (e.g., rPPG)

Feature Vector Construction:

- Normalize and standardize the features.
- Convert into feature vectors suitable for classification.
- Training Phase:
- Use labeled dataset (real vs. fake).
- Train an SVM classifier with a kernel (linear, RBF, or polynomial).
- Optimize hyperparameters using cross-validation.

Testing/Inference:

- Apply the same preprocessing and feature extraction.
- Use trained SVM to classify as real or fake.
- Output:
- Binary classification (0 = Real, 1 = Fake).
- Confidence score (optional).

B. Convolutional Neural Network (CNN) Algorithm for Deepfake Detection

Step-by-Step Workflow:

Input Preparation:

- Extract video frames or use single image inputs.
- Resize images to standard dimensions (e.g., 224x224).
- Normalize pixel values (e.g., between 0 and 1).
- CNN Architecture (e.g., XceptionNet, ResNet, MesoNet):
- Convolution Layers: Learn spatial features via kernel filters.
- Activation Functions: Typically ReLU to introduce non-linearity.
- Pooling Layers: Reduce spatial dimensionality and retain dominant features.

- Batch Normalization: Improve convergence speed and stability.
- Dropout Layers: Prevent overfitting.

Feature Extraction:

- Output from final convolutional block is flattened into a feature vector.
- Classification Layer:
- Fully connected (dense) layers.
- Final softmax or sigmoid layer for binary classification.

Training:

- Loss Function: Binary Cross-Entropy.
- Optimizer: Adam or SGD.
- Epochs: Multiple iterations over training data with backpropagation.

Prediction:

- Classifies each input as either real or deepfake.
- Can be applied frame-wise or averaged over video for final decision.

C. Hybrid CNN + SVM Approach

Workflow:

- CNN is used as a feature extractor (without final classification layer).
- Extracted deep features are fed into an SVM classifier.
- The SVM performs the final classification based on CNN-learned representations.
- This hybrid approach leverages:
- CNN's powerful feature learning
- SVM's ability to generalize better in some cases, especially with smaller datasets

IV. CONCLUSION

The rapid advancement of deepfake generation technologies poses a significant threat to digital media authenticity and public trust. This review has examined the effectiveness of Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) as core techniques in the detection of deepfake images and videos. While SVM offers a lightweight and interpretable approach when combined with well-engineered features such as blink frequency, facial landmarks, or physiological signals, CNN-based methods excel in automatic feature learning, especially in high-dimensional visual spaces.

Our analysis of the literature highlights that CNN models, particularly those based on deep architectures like XceptionNet and ResNet, have consistently achieved high accuracy in benchmark datasets such as FaceForensics++ and Celeb-DF. Hybrid approaches that combine CNN feature extraction with SVM classification have also shown promise in improving detection robustness and cross-dataset generalization. However, challenges remain in ensuring the scalability, generalizability, and real-time deployment of these models, especially in adversarial settings where deepfakes are crafted to evade detection.

Future research should focus on multimodal detection strategies, adversarial defense mechanisms, and the integration of explainability frameworks to enhance the trustworthiness and transparency of detection systems. As the arms race between deepfake creation and detection continues, the fusion of classical machine learning techniques like SVM with deep learning frameworks like CNN presents a balanced and effective strategy in the fight against digital deception.

REFERENCES

[1] Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: a compact facial video forgery detection network. In IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–7.

[2] Chesney, R., & Citron, D. K. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107(6), 1753–1820.

[3] Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? Assessment and detection. arXiv preprint arXiv:1812.08685.

[4] Li, Y., Chang, M. C., & Lyu, S. (2020). Exposing DeepFake Videos by Detecting Face Warping Artifacts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

[5] Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., & Nahavandi, S. (2019). Deep learning for deepfakes creation and detection: A survey. arXiv preprint arXiv:1909.11573.

[6] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1–11.

[7] Khalid, M., & Woo, W. L. (2021). Deepfake Detection: Hybrid SVM-CNN Approach with Enhanced Generalization. IEEE Access, 9, 123–135.

[8] Verdoliva, L. (2020). Media Forensics and DeepFakes: An Overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910–932.

[9] Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., & Natarajan, P. (2019). Recurrent Convolutional Strategies for Face Manipulation Detection in Videos. arXiv:1905.00582.

[10] Zhang, R., Duan, Y., & Li, Y. (2020). Detecting adversarial deepfakes with adversarial training. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), pp. 1586–1590.

[11] Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A New Dataset for DeepFake Forensics. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3207–3216.

[12] Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. ACM Computing Surveys (CSUR), 54(1), 1–41.

[13] H. T. Nguyen, J. Yamagishi, and I. Echizen, "Deep Learning for Deepfakes Creation and Detection: A Survey," arXiv preprint arXiv:1909.11573, 2019.

[14] J. Khalid and W. Woo, "Hybrid CNN-SVM Model for DeepFake Detection," in Proc. IEEE Int. Conf. on AICS, 2021, pp. 102–107.

[15] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," Information Fusion, vol. 64, pp. 131–148, 2020.

[16] U. A. Ciftci, I. Demir, and L. Yin, "FakeCatcher: Detection of Synthetic Portrait Videos Using Biological Signals," IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 3, pp. 1148–1161, 2022.

[17] Y. Wu, Y. Xia, and M. Li, "Detecting AI-synthesized Speech Using SVM and Prosodic Features," in Proc. Interspeech, 2020, pp. 1111–1115.

[18] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," in Proc. IEEE CVPR, 2020, pp. 3207–3216.

[19] R. Zhang, Y. Liu, and Y. Zhang, "Fighting Deepfakes with Adversarial Examples," in Proc. IEEE ICASSP, 2020, pp. 2927–2931.

[20] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, and P. Natarajan, "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos," arXiv preprint arXiv:1905.00582, 2019.

[21] R. Tariq et al., "Deepfake Detection Using CNN-based Feature Extraction and Classification," in Proc. Int. Conf. on ITT, 2021, pp. 1–6.

[22] A. Agarwal, R. Singh, M. Vatsa, and N. Ratha, "On the Effectiveness of Facial Regions for Detecting Deepfakes," IEEE Trans. Inf. Forensics Secur., vol. 16, pp. 3966–3978, 2021.

[23] P. Korshunov and S. Marcel, "Deepfakes: A New Threat to Face Recognition? Assessment and Detection," arXiv preprint arXiv:1812.08685, 2018.

[24] I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, "Deep Video Forgery Detection with Recurrent Neural Networks," in Proc. AVSS, 2019, pp. 1–6.

[25] H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. Jain, "On the Detection of Digital Face Manipulation," in Proc. IEEE CVPR, 2020, pp. 5781–5790.

[26] B. Sun, Y. Yang, and T. Guo, "Temporal Pattern Recognition for Video Deepfake Detection Using SVM," in Proc. IEEE CISP, 2020, pp. 1–5.

[27] P. Gupta, S. Kumar, and S. S. Oberoi, "Evaluation of Deepfake Detection Using Xception and ResNet Models," in Proc. ICICC, 2021, pp. 233–243.

[28] R. Varol and M. Baykara, "Detection of Deepfake Faces Using Fused CNN and SVM Features," Turkish J. Elect. Eng. Comput. Sci., vol. 29, no. 5, pp. 2835–2848, 2021.

[29] S. Mehra, M. K. Singh, and A. Saini, "Adversarial Robustness in Deepfake Detection Using CNN-Based Architectures," in Proc. ICVGIP, 2022, pp. 81–89.