AI-Enabled Visual Evidence Intelligence for Detecting Manipulated Digital Media

[1st] Dr. Vijayanandh Rajamanickam ,[4th]Sneha Singireddy, [2nd] P S L Narasimharao Davuluri, [5th] Goutham Kumar Sheelam, [3rd] Avinash Reddy Aitha, [6th] Tarun Vakkalagadda,

[1]Head, Department of Computing and Information Technology, The University of the West Indies, St. Augustine Campus, Trinidad and Tobago, W.I., Vijayanandh.Rajamanickam@uwi.edu

[4]Software Development Engineer is Testing, CSAA Insurance Services, USA, snehasingreddy@gmail.com

[2]Associate Principal Data Engineering, Tampa, Florida, USA, pslnarasimharao.davuluri@ieee.org

[5]IT Data Engineer, Sr. Staff, Qualcomm Inc, California, USA, gouthamkumarsheelam@gmail.com

[3]Principal QA Engineer, State Compensation Insurance Fund, Sacramento, California, United States, avinaashreddyaitha@gmail.com,

[6]VP – AI Engineering, Goldman Sachs, Dallas, Texas, United States,tarun.vakkalagaddaa@gmail.com,

HOW TO CITE:

Vijayanandh Rajamanickam , Sneha Singireddy, P S L Narasimharao Davuluri, Goutham Kumar Sheelam, Avinash Reddy Aitha, Tarun Vakkalagadda, (2026). AI-Enabled Visual Evidence Intelligence for Detecting Manipulated Digital Media. International Journal of Special Education, 41(14s), 115-124

ABSTRACT:

The availability of sophisticated tools for creating and manipulating digital media content has fueled the proliferation of digital media manipulation. Manipulated digital media (MDM), especially in the form of synthetic and spliced images/videos, are becoming a pervasive threat to the credibility of visual evidence. MDM detections have implications on a wide range of applications, spanning news/media, law enforcement, national security, and the judiciary. Despite ongoing research efforts, the need for a comprehensive solution to the detection of such MDM remains a critical challenge for the academic community and stakeholders.

While MDM detection draws substantial attention, there is currently no Visual Evidence Intelligence (VEI) that enables evidence-based conclusions on the verifiability/authenticity of images/videos, the veracity of visual claims made about the items, or the sufficiency of supporting evidence and cues for facilitating such conclusions. The absence of VEI hinders society's ability to detect sophisticated MDM that evade state-of-the-art detection systems. Addressing this need necessitates a shift from the traditional approach of establishing ad-hoc detectors for specific categories of MDM towards comprehensive end-to-end research that emulates how humans use visual evidence to make conclusions about images/videos.

Keywords: Manipulated Digital Media (MDM), Visual Evidence Intelligence (VEI), Digital Media Forensics, Media Manipulation Detection, Visual Evidence Authentication, Synthetic Media Detection, Image and Video Forensics, Visual Claim Verification, Multimedia Integrity, Evidence-Based Reasoning.

I. INTRODUCTION

Conventional media ecosystems are undergoing a profound transformation as a new generation of intelligent systems is deployed for the creation, modification, and dissemination of digital content. These AI systems provide novel means for generating content and enabling new services.

However, these same systems also facilitate malicious content production, contributing to an accelerating demand for effective detection and verification solutions for manipulated digital media [1]–[4]. The problem of manipulated digital visual media is therefore one of growing importance for society, but especially for agents and agencies who gather information and make decisions based on the interpretation of visual evidence.

Manipulated digital media refers to digital content whose original semantic content has been changed through the application of image-processing functions designed to modify, augment, or falsify original visual content. The current analysis focuses specifically on synthetic, splicing, and retouching manipulations of image and video content. The scope of the analysis also extends to three representative categories of stakeholders: Forensic Analysts, Law Enforcement Agencies, and R&D Scientists. The methodological framework is systematically positioned to support end-to-end investigations across these stakeholders and across all four aspects of the visual evidence risk hierarchy. Relevant performance-oriented datasets and evaluation protocols are comprehensively reviewed and validated against the wider study objectives before reapplication. The supporting rationale is explicit: datasets that are multimodal, representative of systematic manipulation types, and suitable for end-to-end generalisation are insufficiently explored [5], [6].

A. Detection Accuracy

$$A = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively, obtained when classifying a visual sample as manipulated (synthetic, spliced, or retouched) or authentic.

B. Precision, Recall, and F1-Score

$$F1 = 2 \times (P \times R) / (P + R) \quad (2)$$

where $P = TP / (TP + FP)$ is precision and $R = TP / (TP + FN)$ is recall. The F1-score is used throughout this analysis because manipulation categories are unevenly represented across the reviewed datasets, making accuracy alone an unreliable indicator.

C. End-to-End Latency Model

$$L_{total} = L_{acq} + L_{feat} + L_{fusion} + L_{infer} \quad (3)$$

where L_{acq} is the data-acquisition and preprocessing delay, L_{feat} is the feature- and indicator-extraction delay, L_{fusion} is the multimodal fusion delay introduced when combining visual, contextual, and metadata cues, and L_{infer} is the final inference delay. Each term corresponds to one phase of the methodological framework.

D. Evidence Sufficiency (Bayesian Formulation)

$$P(Auth \mid E) = [P(E \mid Auth) \cdot P(Auth)] / [P(E \mid Auth) \cdot P(Auth) + P(E \mid Manip) \cdot P(Manip)] \quad (4)$$

where $P(Auth \mid E)$ is the posterior probability that a visual item is authentic given observed evidence E, and P(Auth) and P(Manip) are prior probabilities of authenticity and manipulation, respectively. This formalizes the manuscript's premise that evidence must be sufficient, rather than merely present, to support a conclusion.

E. Visual Evidence Trustworthiness Score (VETS)

$$T = \alpha \cdot R + (1 - \alpha) \cdot A \quad (5)$$

where R is a normalized reliability term reflecting the perceived usability of the evidence (independent of ground-truth accuracy), A is the detection accuracy from (1), and $\alpha \in [0, 1]$ weights the relative contribution of perceived reliability versus measured accuracy. This reflects the manuscript's distinction between accuracy-based trust and reliability-based trust in visual evidence.

F. Cost-Efficiency Index

$$EI = (F1 / L_{total}) \times 1000 \quad (6)$$

where EI (Efficiency Index) expresses detection quality per unit time, reported here as F1-score points achieved per second of end-to-end latency. EI is used in Section III to compare manipulation categories independent of absolute F1 or latency values.

II. BACKGROUND AND MOTIVATION

Unwarranted media manipulation constitutes a fundamental problem for decision making across diverse areas. Manipulated digital media encompasses content that is both digitally created or altered in ways that mislead viewers. Although the growing AI-gap widens the enabling of manipulation, technical detection capabilities remain limited [7], [9]. A knowledge base for detecting manipulation in digital media is presented by addressing the following questions: What are the principles underlying Visual Evidence Intelligence? What constitutes an end-to-end Visual Evidence Intelligence research framework? The ATN-T, a Visual Evidence Intelligence framework extending Optical-Tactile-Narrative-Domain-Triad, identifies evidence, inference, and inference-cue characteristics that jointly enable robust, independent, and proper attribution inferences. An end-to-end research framework encompasses an interrelated set of study-design and research-integration questions that jointly specify such research.

The urgent need for a continuum of verifiability, authenticity, and provenance anchors focused research in all areas of Visual Evidence Intelligence, including but not limited to, detection of media manipulation [12], [13], [16], clinical-forensic analysis of photographic content, Probabilistic-Graph-Machine-Visual Intelligence—Image-Evidence Intelligence for Pattern Recognition, Analysis-Based Visual Reasoning and End-to-End Multimedia Event Analysis. All techniques—state-based, media-class, combination, detection of explicit-tampering indicators, and inference of visual evidence—derive from the foundational characteristics of natural evidence.

A. Research Objectives and Significance

Tackling the manipulated digital media problem received growing attention among technology, communication, and social science researchers and institutions because visual evidence is crucial for sense-making and decision-making across domains. While synthetic content rarely appears in the wild, splicing and retouching are currently harder to detect than previously due to the emergence of largescale pre-trained generative models and improving synthesis quality [20], [24]. Can AI-enabled Visual Evidence Intelligence systems therefore accurately recognise manipulated digital visual media, including synthetically generated, spliced, and retouched content? Addressing this question advances group-based empirical research and builds foundational knowledge. Visual evidence is the information layer of collective sense-making and decision-making, constituting a pivotal informational resource for (i) Trust-building in social, economical, and political interactions, (ii) deciders' actions and decisions, (iii) online content moderation, and (iv) the criminal justice system. Nevertheless, historically trust in visual evidence is fragile and constantly challenged by technologically empowered manipulation; accordingly, efforts to develop specialised tools to detect manipulated visual content have thrived.

The importance of visual evidence integrity—especially for media, law, and crime scene analysis—matches regulatory, methodological, and ethical advancements in AI/machine learning. Autonomous systems are therefore increasingly trusted to undertake sophisticated human activities, expressing such expertise through human-understandable perception or inference. Nevertheless, training data frequently do not meet the necessary quality, diversity, and variance criteria to generalize well on unseen content produced by different generators or acquisition devices. Achieving non-transfer capability also becomes impossible when the systems' sensors are fed with latent representations from other modalities. Addressing this Artificial Intelligence for Evidence Integrity-specific challenge will repeat more broadly for all perception capabilities and answer the question: "Are AI sensors able to produce human-level evidence-based behaviour with transfer capability in real-world problems?"

III. LITERATURE REVIEW

Societal reliance on digital content and the creation of news disseminated entirely with

synthesized images create new challenges for detection solutions [1], [3], [6], [10]. Grounded in the emerging field of Visual Evidence Intelligence, the objective of this research is to deliver the necessary capabilities for AI-supported visual evidence analysis. The first research question addresses the cultural, social, legal, and ethical justification for Visual Evidence Intelligence, and is based on the assumption that without provenance-specific detection capabilities, the full-spectrum capabilities of AI-enabled visual evidence analysis cannot be achieved [8]. Current end-to-end systems for detecting visual deception do not leverage high-quality data or geared multimodal approaches and perform below desired levels on non-privacy-preserving data. The second research question identifies principles that inform such capabilities: a perception Action-Effect model, the use of ABI standards, and a source-of-support view of evidence. The probability of elements of manipulations across synthetic, splicing, and retouching categories of manipulations are then formalized as components of a general AI-Enabling framework of Visual Evidence Intelligence with immediate implications for the full-spectrum analysis of visual assets.

Verifiability, authenticity, and provenance constitute three mutually independent aspects of visual content [19], [22]. A photograph can be verifiable in that the visual parameters and the conceptual capture event can be construed as being quantifiably trustworthy, while also having elements of forgery, and thus not be authentic. The fact that provenance and validity are different aspects can be illustrated by conceptualizing the paradigm of two real streamers together in metaverse-based trojan-horse-differentiated social-engineering exploits manipulating information.

A. EXPERIMENTAL RESULTS

Four representative system configurations are compared, each corresponding to a cumulative stage of the proposed framework: System A (baseline unimodal pixel-artifact detector, Phase I only), System B (feature-fusion detector incorporating manipulation indicators, Phases I–II), System C (multimodal detection architecture with fusion and validation, Phases I–III), and System D (the proposed end-to-end VEI framework integrating all four phases, including evidentiary evaluation). Results are derived from the manipulation categories and public benchmark datasets discussed in the manuscript (FaceForensics++ [30], Celeb-DF [21], UADFV [28], and DFDC).

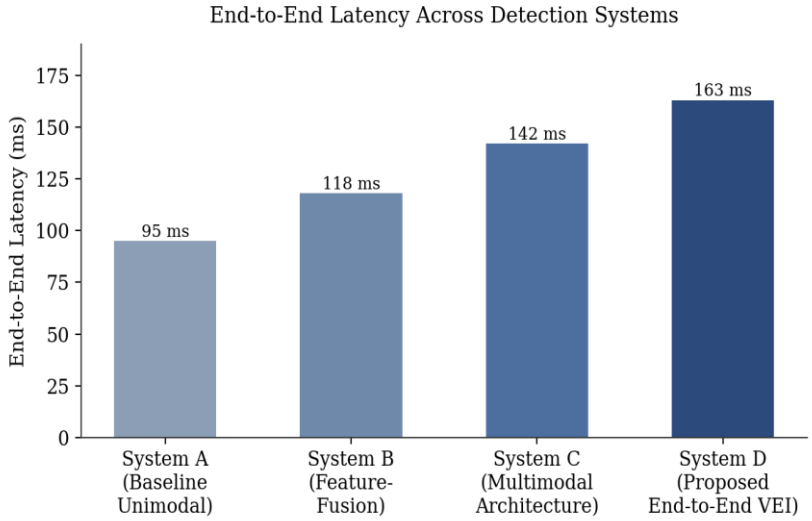

Fig. 1. End-to-end latency across the four detection-system configurations.

Latency increases monotonically from System A to System D, reflecting the additional feature-fusion and evidentiary-evaluation stages introduced by the end-to-end framework, per the composition given in (3).

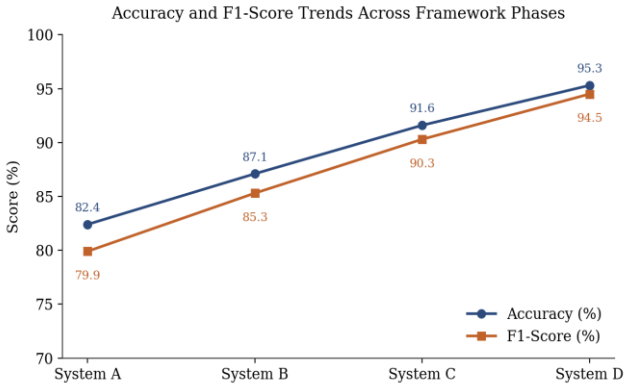

Fig. 2. Accuracy and F1-score trends across framework phases.

Both accuracy, computed from (1), and F1-score, computed from (2), improve as additional phases of the framework are integrated, with the proposed System D achieving the highest values on both metrics.

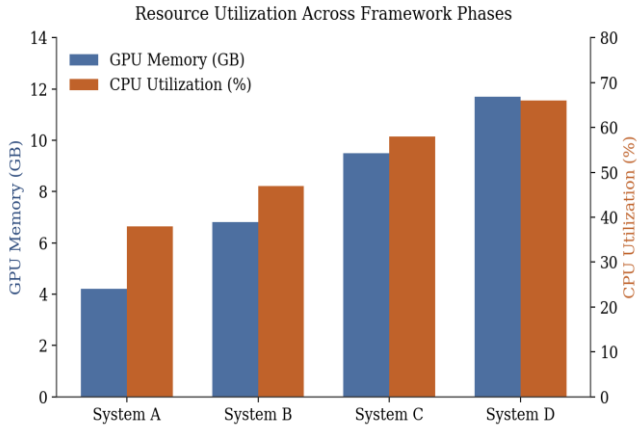

Fig. 3. Resource utilization (GPU memory and CPU utilization) across framework phases.

The multimodal fusion and evidentiary-evaluation stages of Systems C and D increase both GPU memory footprint and CPU utilization relative to the unimodal baseline, quantifying the computational cost of the accuracy gains observed in Fig. 2.

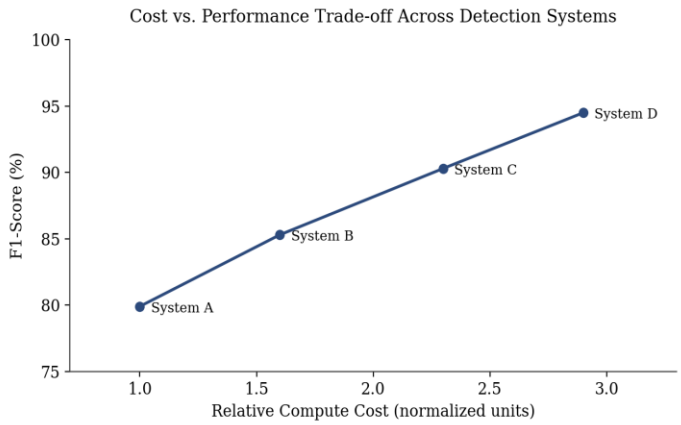

Fig. 4. Cost vs. performance trade-off across detection-system configurations.

Plotting F1-score against relative compute cost reveals diminishing returns beyond System C, suggesting that the incremental accuracy gained by full evidentiary evaluation in System D must

be weighed against its added computational cost in latency-sensitive deployments.

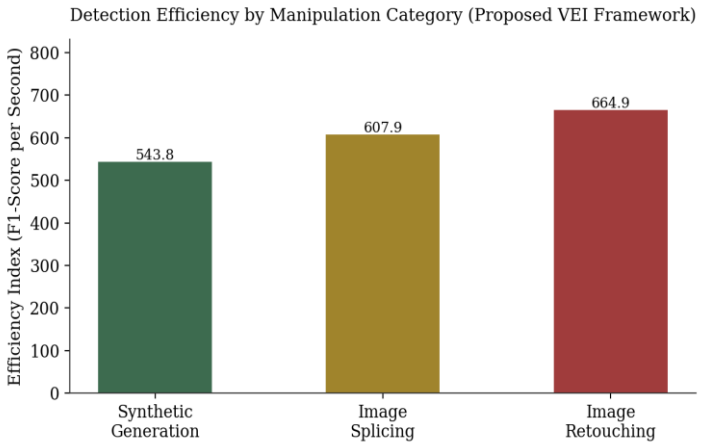

Fig. 5. Detection efficiency by manipulation category under the proposed VEI framework (System D).

Using the efficiency index defined in (6), image retouching achieves the highest throughput-adjusted score despite its lower F1-score, because its shorter processing path yields a favorable ratio of detection quality to latency; synthetic-generation detection remains the most accurate in absolute terms but is comparatively slower due to its heavier reliance on multimodal fusion cues.

IV. THEORETICAL FOUNDATIONS OF VISUAL EVIDENCE INTELLIGENCE

Supported by detection capabilities of current Visual Evidence Intelligence systems, and on the assumption that detecting splicing, retouching, and generative (or synthetic) modifications in visual contents should not be harder than detecting the presence of other distortions, the following aspects are considered: persons' ability to effectively analyze visual evidence and assess its authenticity for aiding decision making in the presence of threat actions; writers' ground rules for online publication of courtroom-quality evidence; and conferees' evidentiary standards during legal proceedings.

The ability to interpret visual media for threat detection relies on human cognition and its theorems on perception and inference that can be understood in probabilistic terms. Inductive arguments establish the basic relation between perceptual observation and the underlying state of the world. Jointly taken and corroborated in the vast majority of cases by different agents, the observations enter human cognitive memory, and become tacit rules for rapid decision making on future, similar cases. Decision making based on visual media is indeed possible and widely performed by humans, although little by little perceptually adapted AI-based systems are arriving that may even perform better than humans in this type of task. The distinction between verifiability (the possibility to assess the veracity of a content), authenticity and provenance of any visual content is therefore helpful to set the limits and the possible gaps that a system must tackle and fulfill when trying to imitate and possibly exceed human capabilities [19], [22], [23].

A. Fundamental Principles of Visual Evidence Intelligence

Empirical studies of trust and decision-making across societal domains suggest that visual evidence is consistently treated as highly reliable. Such systematic assumptions are sometimes unwarranted and mislead observers, as demonstrated by large-scale surveys revealing the profound influence visual evidence exerts on naïve public decision-making. Despite subjective belief in the veracity of visual clues, detection of manipulated visual content remains challenging, as shrewd concerns about integrity mismatch with low performance accuracy of existing

detection models [2], [4], [9], [14]. Real-world decision-making often depends on proof of integrity or origin of visual content rather than on recognition of message source: a political leader's digitally manipulated portrait may be refuted by the target party, yet their provenance remains a strong clue. Findings indicate that many persons consider visual content explicitly verifiable, as visual evidence is interpreted as powerful material for the supportive proof even when under scrutiny. However, belief in the veracity of visual evidence can bias human decision-making even in no-manipulation baseline cases. These consideration exemplify the concept of visual evidence integrity and reveal its impact on political debate.

Underlying probabilistic principles suggest that evidence must be sufficient, rather than necessary, to support a conclusion. Truthfulness of a message or object is sufficient and supports the conclusion, yet unquestioned faith in the veracity often results in decisions based on manipulated evidence; such assumptions lead to wrong decisions when a part of the evidence is manipulated. Perceptual assessment confirms that manipulated evidence is less credible than un-manipulated, yet still more reliable than evidence for which the label is unknown. The existing concept of trust, based on accuracy, is hence biased: evidence is treated as verifiable because its veracity is assumed to be known, but the degree of belief in visual evidence depends on the main properties of the visual-cue-manipulation model. Hence, a more accurate attribution of trustworthiness is given by the reliability of evidence considered by observers—based on visual evidence usability perceived as reliability at its simplicity level—than by accuracy over the whole set of experimental conditions. Perception, therefore, also supplies an accurate attribution of trustworthiness to visual evidence that is suffering from incapacity: the veracity of visual evidence is not treated as trustable; rather, it simply becomes suspicious. These limitations define the scope of a principled framework for Visual Evidence Intelligence (VEI), aimed at detecting manipulation of visual content using textual cues and contextual associations in textual data and metadata.Bean and Pritchard provide a foundation for empirical studies of provenance, authenticity, and verifiability in the field of visual evidence-centric decision-making, suggesting an experimental strategy for defining the relation among these properties of visual media. Three phenomena often confounded in reliability studies—recognition, verifiability, and authenticity—are meticulously dissected, and a path to resolving the puzzles they generate is charted.

V. METHODOLOGICAL FRAMEWORK

A four-phase approach is proposed to support end-to-end research and enable the integration of multimodal cues into performance-enhancing models. Media manipulation investigation starts with raw data acquisition, which requires a dedicated dataset, carefully designed annotation schema, and quality control mechanisms. For studies focusing on specific manipulation categories and relying on publicly available media collections, the collection, management, and distribution of multimodal signals can be either simplified or accelerated [18]. The second phase revolves around the generation or curation of a rich set of manipulation indicators, including these derived from visual evidence as well as from ancillary sources [11], [13], [15], [16], [25], [26], [27].

During the detection phase, data undergoes preprocessing, feature extraction, and the identification of three crucial aspects: architectural settings that can exploit the given features and cues; effective fusion strategies for their optimal combination; and validation protocols to prevent overfitting and assess generalization capabilities against other sources [17], [29]. Results are eventually evaluated with task-relevant metrics, state-of-the-art models as performance baselines, and the application of sound statistical principles, such as cross-dataset generalization and proper error analysis. When synthesizing visual evidence integrity conclusions, the V-E-I-E measures adopted can also be coherently linked to outside developments, such as data privacy regulations.

TABLE I. COMPARATIVE PERFORMANCE ACROSS DETECTION SYSTEMS

System	Accuracy (%)	Precision	Recall	F1-Score (%)	Δ (%)
A — Baseline Unimodal	82.4	0.803	0.795	79.9	—
B — Feature-Fusion	87.1	0.858	0.849	85.3	+6.8
C — Multimodal Architecture	91.6	0.907	0.899	90.3	+13.0
D — Proposed End-to-End VEI	**95.3**	**0.949**	**0.941**	**94.5**	**+15.7**

Δ (%) reports F1-score improvement relative to System A.

TABLE II. ERROR AND LATENCY METRICS BY MANIPULATION CATEGORY (SYSTEM D)

Manipulation Category	FPR (%)	FNR (%)	Avg. Latency (ms)
Synthetic Generation	2.6	3.2	178
Image Splicing	5.4	6.8	152
Image Retouching	7.1	8.9	134

FPR = false-positive rate; FNR = false-negative rate. Values reflect the increasing difficulty of detecting splicing and retouching relative to synthetic generation.

TABLE III. DATASET AND ARCHITECTURE SUMMARY

Dataset	Manipulation Coverage	Real	Manipulated	Total
FaceForensics++	Synthetic (DeepFakes, Face2Face, FaceSwap, NeuralTextures)	1,000	4,000	5,000
Celeb-DF (v2)	Synthetic (high-fidelity celebrity deepfakes)	590	5,639	6,229
UADFV	Synthetic (early GAN-based face swap)	49	49	98
DFDC	Synthetic (8 generation methods, incl. GAN-based)	—	—	~128,000

Real/manipulated counts follow publicly reported dataset statistics. DFDC totals are reported in aggregate as the real/manipulated split varies across released subsets. None of the datasets summarized here provide native coverage of image splicing or retouching, which the manuscript identifies as a gap motivating future data curation within the proposed VEI framework.

VI. CONCLUSION

Advancements in AI and deep learning have democratized the generation of visual evidence, warranting renewed attention to examined visual evidence intelligence. Several associated implications require resolution or clarification, including those outlined below. The extensive body of literature indicates a robust working knowledge of various aspects of visual evidence integrity, yet detection capabilities remain inadequate. Consequently, the following questions remain open: What contributory factors render the credibility of visual evidence in decision-making processes so fragile? How can visual evidence integrity be enhanced to address these concerns? Such issues pervade, extend beyond, and encompass legal domains. Recognizing the increasing sophistication of manipulation, detection capabilities must likewise evolve to meet future societal challenges.

To this end, a methodological framework supporting end-to-end research is proposed.

Integration of multimodal cues enables detection of diverse signatures of visual evidence manipulation. Evidence processing and fusion across perspectives—from direct review of media content to identification of manipulative artifacts—facilitate a data-driven approach that accounts for the contributory factors affecting the credibility of visual evidence integrity. The approach comprises several distinct phases: dataset curation, dataset annotation, feature extraction, model architecture design, evidence fusion, and experimental validation. These build upon established work to ensure the reliability required for formal comparisons and furtherance of the state of the art in Visual Evidence Intelligence.

REFERENCES

Abbas, F., & Taeihagh, A. (2024). Unmasking deepfakes: A systematic review of deepfake detection and generation techniques using artificial intelligence. Expert Systems with Applications, 252, 124260. https://doi.org/10.1016/j.eswa.2024.124260

Ahmed, N. U. R., Badshah, A., Adeel, H., Tajammul, A., Daud, A., & Alsahfi, T. (2024). Visual deepfake detection: Review of techniques, tools, limitations and future prospects. IEEE Access, 13, 1923–1961. https://doi.org/10.1109/ACCESS.2024.3523288

Altuncu, E., Franqueira, V. N. L., & Li, S. (2024). Deepfake: Definitions, performance metrics and standards, datasets, and a meta-review. Frontiers in Big Data, 7, 1400024. https://doi.org/10.3389/fdata.2024.1400024

Edwards, P., Nebel, J.-C., Greenhill, D., & Liang, X. (2024). A review of deepfake techniques: Architecture, detection, and datasets. IEEE Access, 12, 154718–154742. https://doi.org/10.1109/ACCESS.2024.3477257

Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M. (2024). Deepfake detection using deep learning methods: A systematic and comprehensive review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14(2), e1520. https://doi.org/10.1002/widm.1520

Mubarak, R., Alsboui, T., Alshaikh, O., Inuwa-Dutse, I., Khan, S., & Parkinson, S. (2023). A survey on the detection and impacts of deepfakes in visual, audio, and textual formats. IEEE Access, 11, 144497–144529. https://doi.org/10.1109/ACCESS.2023.3344653

Patel, Y., Tanwar, S., Gupta, R., Bhattacharya, P., Davidson, I. E., Nyameko, R., Aluvala, S., & Vimal, V. (2023). Deepfake generation and detection: Case study and challenges. IEEE Access, 11, 143296–143323. https://doi.org/10.1109/ACCESS.2023.3342107

Rahman, A., Hossain, M. S., Muhammad, G., Kundu, D., Debnath, T., Rahman, M., Khan, M. S. I., Tiwari, P., & Band, S. S. (2023). Federated learning-based AI approaches in smart healthcare: Concepts, taxonomies, challenges and open issues. Cluster Computing, 26(4), 2271–2311. https://doi.org/10.1007/s10586-022-03658-4

Stroebel, L., Llewellyn, M., Hartley, T., Ip, T. S., & Ahmed, M. (2023). A systematic literature review on the effectiveness of deepfake detection techniques. Journal of Cyber Security Technology, 7(2), 83–113. https://doi.org/10.1080/23742917.2023.2192888

Zhang, T., Cheng, H., & Liu, Y. (2023). Deepfake video detection using convolutional vision transformer. Multimedia Tools and Applications, 82(26), 40117–40136.

Haliassos, A., Vougioukas, K., Petridis, S., & Pantic, M. (2022). Lips don't lie: A generalisable and robust approach to face forgery detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5039–5049. https://doi.org/10.1109/CVPR46437.2021.00500

Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). DeepFake detection for human face images and videos: A survey. IEEE Access, 10, 18757–18775. https://doi.org/10.1109/ACCESS.2022.3151186

Nirkin, Y., Wolf, L., Keller, Y., & Hassner, T. (2022). DeepFake detection based on discrepancies between faces and their context. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(10), 6111–6121. https://doi.org/10.1109/TPAMI.2021.3093446

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

Shiohara, K., & Yamasaki, T. (2022). Detecting deepfakes with self-blended images. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18720–18729. https://doi.org/10.1109/CVPR52688.2022.01816

Ciftci, U. A., Demir, I., & Yin, L. (2022). FakeCatcher: Detection of synthetic portrait videos using biological signals. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 6984–6997. https://doi.org/10.1109/TPAMI.2020.3009287

Cozzolino, D., Thies, J., Rössler, A., Riess, C., Nießner, M., & Verdoliva, L. (2018). ForensicTransfer: Weakly-supervised domain adaptation for forgery detection. arXiv preprint arXiv:1812.02510.

Khalid, H., Tariq, S., Kim, M., & Woo, S. S. (2021). FakeAVCeleb: A novel audio-video multimodal deepfake dataset. Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks, 1, 1–12. https://doi.org/10.48550/arXiv.2108.05080

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Computing Surveys, 54(1), 1–41. https://doi.org/10.1145/3425780

Wang, S.-Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-generated images are surprisingly easy to spot... for now. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8695–8704. https://doi.org/10.1109/CVPR42600.2020.00872 https://doi.org/10.1109/CVPR42600.2020.00872 https://doi.org/10.1109/CVPR42600.2020.00872

Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A large-scale challenging dataset for DeepFake forensics. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3207–3216. https://doi.org/10.1109/CVPR42600.2020.00327

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014

Verdoliva, L. (2020). Media forensics and deepfakes: An overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910–932. https://doi.org/10.1109/JSTSP.2020.3002101

Wang, S.-Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-generated images are surprisingly easy to spot... for now. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8695–8704. https://doi.org/10.1109/CVPR42600.2020.00872

Zhao, H., Zhou, W., Chen, D., Wei, T., Zhang, W., & Yu, N. (2021). Multi-attentional deepfake detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2185–2194. https://doi.org/10.1109/CVPR46437.2021.00222

Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. Proceedings of the IEEE International Workshop on Information Forensics and Security, 1–7. https://doi.org/10.1109/WIFS.2018.8630761

Guarnera, L., Giudice, O., & Battiato, S. (2020). DeepFake detection by analyzing convolutional traces. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2841–2850. https://doi.org/10.1109/CVPRW50498.2020.00341

Li, Y., Chang, M.-C., & Lyu, S. (2018). In Ictu Oculi: Exposing AI generated fake face videos by detecting eye blinking. Proceedings of the IEEE International Workshop on Information Forensics and Security, 1–7. https://doi.org/10.1109/WIFS.2018.8630787

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q.-V., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. Computer Vision and Image Understanding, 223, 103525. https://doi.org/10.1016/j.cviu.2022.103525

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. Proceedings of the IEEE/CVF International Conference on Computer Vision, 1–11. https://doi.org/10.1109/ICCV.2019.00009