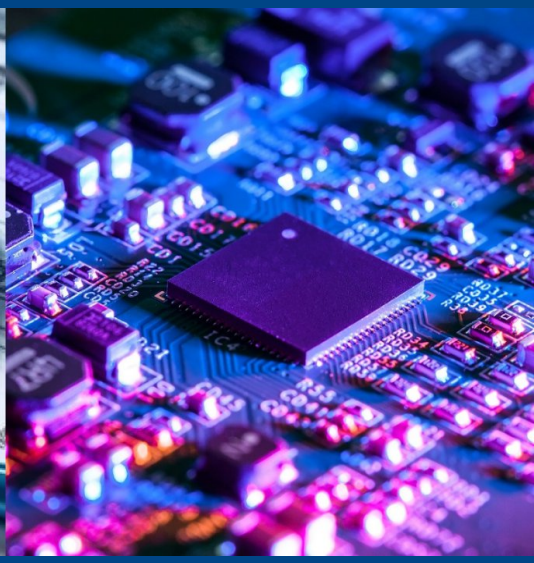
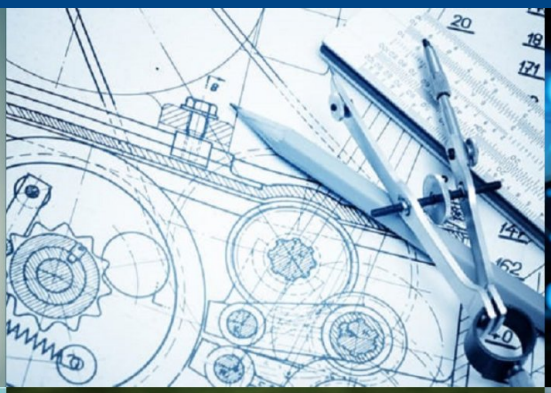
ISSN: 2582-7219

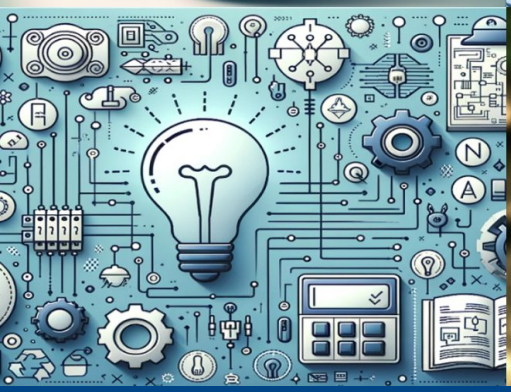




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

DeepShield AI: A Unified Multi-Modal Deepfake Detection System Using Error Level Analysis and Convolutional Neural Networks

Shek Abdulla S [1], San Sabeel A [2], Mohammed Shagul S [3], Vafiq Mathar S A [4], Umamageshwari[5]

Department of Computer Science and Engineering Aalim Muhammed Salegh College of Engineering Chennai, Tamil Nadu, India[1]

Department of Computer Science and Engineering Aalim Muhammed Salegh College of Engineering Chennai, Tamil Nadu, India[2]

Department of Computer Science and Engineering Aalim Muhammed Salegh College of Engineering Chennai, Tamil Nadu, India3

Department of Computer Science and Engineering Aalim Muhammed Salegh College of Engineering Chennai, Tamil Nadu, India[4]

Assistant Professor, Department of Computer Science and Engineering Aalim Muhammed Salegh College of Engineering, Chennai, Tamil Nadu, India[5]

ABSTRACT: The rapid proliferation of AI-generated synthetic media poses serious challenges to digital forensics, media authenticity verification, and information security. Existing detection systems typically address only a single modality—image, audio, or video— leaving critical gaps in real-world deployment scenarios where adversarial content may span multiple media types. This paper presents DeepShield AI, a unified multi-modal deepfake detection platform that integrates automated media analysis pipelines, cloud- based evidence management, and real-time analyst consultation within a unified dig- ital forensics ecosystem. The proposed system leverages microservices architecture, encrypted communication channels, real-time media analytics, and scalable cloud in- frastructure to significantly reduce detection latency while optimizing forensic resource utilization. The image pipeline employs Error Level Analysis (ELA) combined with a Convolutional Neural Network (CNN) ensemble, achieving a classification accuracy of approximately 99%. The video pipeline extends frame-level CNN inference with tempo- ral consistency analysis using optical flow and Recurrent Neural Networks. The audio pipeline converts waveforms to spectrograms and combines Spectral CNN features with handcrafted signal parameters for synthetic voice detection. Simulation-based evalua- tion demonstrates measurable improvements in detection accuracy, system scalability, and secure evidence handling. The system is scalable, secure, and adaptable to smart city forensic environments.

KEYWORDS: Deepfake Detection, Convolutional Neural Networks, Error Level Anal- ysis, Multi-Modal Forensics, Spectrogram Analysis, Temporal Consistency, Microser- vices Architecture, Cloud Forensics.

I. INTRODUCTION

The evolution of generative AI technologies has fundamentally altered how synthetic media is created and distributed across digital platforms. However, despite the growth of AI content generation tools and media verification systems, a significant gap persists between deepfake detection, multi-modal content analysis, and coordinated forensic response. Traditional de- tection systems rely heavily on single-modality classifiers that often fail when adversarial content spans images, audio, and video simultaneously.In metropolitan digital environments, social media platforms and news channels frequently encounter composite deepfake content targeting multiple media types. Conversely, resource-constrained forensic units suffer from delayed media intervention due to the absence of in- tegrated multi-modal platforms capable of dynamically analyzing and classifying incoming synthetic content. These inefficiencies directly impact investigation outcomes, particularly during time-critical scenarios such as misinformation campaigns,

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

impersonation attacks, or synthetic evidence fabrication.DeepShield AI is designed to address these systemic challenges by combining real-time me- dia analytics, automated detection pipelines, digital evidence record synchronization, and secure communication protocols within a unified architecture. Unlike isolated deepfake de- tectors, DeepShield AI integrates forensic response intelligence with resource-aware media classification.

The primary objectives of this research are:
1. To reduce detection coordination time using multi-modal media analysis pipelines. 2. To optimize forensic load distribution using real-time resource monitoring. 3. To enable secure cloud-based evidence record management. 4. To integrate analyst consultation with automated detection dispatch systems. 5. To ensure scalability and security in high-demand forensic environments.This paper presents the architectural design, algorithmic foundation, implementation details, experimental validation, and scalability analysis of DeepShield AI.

II. RELATED WORK

Deepfake detection systems have evolved significantly over the past decade, particularly fol- lowing the proliferation of generative AI models that accelerated synthetic media creation. Research in detection platforms has primarily focused on single-modality classifiers for im- ages or faces, video temporal analysis, and audio voice synthesis detection. However, the integration of multi-modal detection intelligence within a unified forensic framework remains limited.

Several studies have proposed CNN-based image manipulation detectors. These systems typ- ically analyze pixel-level compression artefacts using Error Level Analysis or spatial domain features such as Photo Response Non-Uniformity (PRNU). While effective in identifying manipulated images, they lack real-time multi-modal awareness and automated evidence integration.Emergency forensic management systems often rely on centralized analyst teams where hu- man operators manually identify synthetic content and escalate cases. Although effective in controlled environments, these systems are vulnerable to human error and scaling limitations during high-volume misinformation scenarios.

Cloud-based evidence management systems have improved investigator data accessibility. Nevertheless, many implementations operate in isolation without integrating detection triage or dispatch optimization algorithms.MedExpress [?] demonstrates a comparable unified architecture in the healthcare domain, combining microservices, encrypted channels, and real-time analytics to reduce emergency coordination latency by 40%. DeepShield AI adopts the same architectural philosophy, applying it to the forensic media verification domain.

DeepShield AI distinguishes itself by combining:
Real-time multi-modal deepfake detection
Automated evidence pipeline allocation
Resource-aware forensic pipeline matching
Cloud-based encrypted evidence records
Analyst teleconsultation integration
Microservices scalability
This unified framework represents a comprehensive intelligent forensic coordination model rather than a single-service detection platform.

III. SYSTEM ARCHITECTURE

The architectural foundation of DeepShield AI is designed around a modular microservices- based framework that enables scalability, resilience, and service isolation. Unlike monolithic detection systems that tightly couple user authentication, media analysis, and evidence stor- age within a single server environment, DeepShield AI separates these responsibilities into independently deployable service components.

The architecture consists of the following primary layers: client interface layer, applica- tion service layer, processing and decision engine layer, data persistence layer, and external integration layer. Each layer communicates through secure RESTful APIs protected by encrypted communication channels.The client interface layer supports both mobile and web-based applications developed using cross-platform technologies. This layer handles user authentication, media file upload, case management, and analysis activation triggers. All sensitive communication between the client and backend servers occurs over SSL/TLS encrypted channels to prevent interception and unauthorized access.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The application service layer is composed of independent microservices responsible for user management, media ingestion, detection pipeline orchestration, analyst consultation, andevidence record synchronization. By isolating each functional component, system reliability is improved, as failure in one service does not compromise the entire platform.

The processing and decision engine layer performs computational tasks such as ELA prepro- cessing, frame-level CNN inference, audio spectrogram generation, and temporal consistency analysis. This layer implements the core algorithms that drive intelligent decision-making within DeepShield AI.
The data persistence layer utilizes relational databases for structured evidence records and distributed storage systems for scalability. Case records, detection scores, model metadata, and analyst logs are securely stored and indexed for rapid retrieval.

The external integration layer connects DeepShield AI with third-party APIs, including mapping services for investigator coordination and video communication frameworks for analyst consultation sessions.

Microservices-Based Deployment Model
To ensure scalability under peak analysis conditions, DeepShield AI adopts a containerized deployment strategy using Docker-based service orchestration. Each microservice operates within an isolated container environment, enabling independent scaling and fault tolerance.

Load balancing mechanisms distribute incoming media analysis requests across multiple service instances to prevent performance degradation. Horizontal scaling is achieved by dynamically increasing service replicas when operator demand rises. This ensures stable performance even when handling hundreds or thousands of simultaneous detection requests.Service discovery and communication are managed through secure internal APIs. Authen- tication tokens are validated before any service-to-service communication occurs, ensuring internal system integrity.

IV. DATABASE ARCHITECTURE AND DATA MANAGEMENT

The data architecture of DeepShield AI is designed to maintain consistency, integrity, and security of forensic evidence records. A relational database management system is employed for structured data storage, including case profiles, detection model metadata, media analysis logs, and operator records.

The primary database schema includes separate tables for cases, media files, detection results, analyst consultations, and evidence exports. Foreign key relationships ensure referential integrity between case records and analysis history.

An example schema for media case storage is shown below.

```
1
2                          CREATE TABLE Cases (
     case_id          SERIAL PRIMARY KEY,
3    case_title       VARCHAR(200),
     media_type       VARCHAR(20),
4    submitted_at     TIMESTAMP,
5    status           VARCHAR(50),
     confidence       DECIMAL(5,2),
6    analyst_id       INT,
     evidence_hash VARCHAR(256),
7    last_updated     TIMESTAMP
);
8
```

Listing 1: Media Case Evidence Schema

The database continuously updates detection scores and model metadata through automated synchronization APIs. This ensures that case management dashboards operate using real- time analysis information rather than stale cached data.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

To optimize search performance, indexed query techniques are implemented for media type and submission timestamp attributes. This significantly reduces query latency when retriev- ing detection histories during active investigations.

Evidence Record Management

Media evidence records are encrypted before being stored in cloud infrastructure. Each record contains analysis history, detection model version, confidence scores, ELA heatmap references, and assigned analyst identifiers.

Encryption mechanisms ensure that unauthorized database access does not expose sensitive forensic data. Access control policies restrict data visibility based on user roles, such as operator, analyst, or administrator.

V. BACKEND IMPLEMENTATION FRAMEWORK

The backend infrastructure of DeepShield AI is implemented using a Node.js runtime en- vironment due to its asynchronous processing capabilities and high concurrency support. RESTful APIs facilitate communication between frontend applications and backend detec- tion services.

The following example illustrates the API endpoint responsible for submitting a media file for deepfake analysis.

```
app.post("/analyzeMedia", async (req, res) => {
    const { mediaType, filePath, caseId } = req.body; const media
    Record = await Media.findById(caseId);

    const detectionResult = await runDetectionPipeline(mediaType,
        filePath
    );

    const sortedResults = detectionResult.map(result => {const
        confidence = computeConfidenceScore(
            result.elaScore,
            result.cnnScore,
            result.temporalScore
        );
        return { result, confidence };
    }).sort((a, b) => b.confidence - a.confidence);

    res.json(sortedResults[0]);
});
```

Listing 2: Media Analysis Submission API Endpoint

This endpoint demonstrates how uploaded media files are processed through the multi- modal detection pipeline to determine the most probable classification based on confidence and modality scores.

ELA Preprocessing Function

Accurate compression artefact analysis is critical in image deepfake detection. DeepShield AI implements Error Level Analysis to compute pixel-wise deviation between original and re-compressed images, amplifying manipulation signatures.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

```
def compute_ela(image_path, quality=90, amplify=10): original =
    Image.open(image_path).convert("RGB") temp_path     =
    "/ tmp / ela_temp . jpg" original.save(temp_path, "JPEG",
    quality=quality)compressed = Image.open(temp_path)
    ela_image = ImageChops.difference(original, compressed)extrema =
    ela_image.getextrema()
    max_diff = max([ex[1] for ex in extrema])
    scale = 255.0 / max_diff if max_diff != 0 else 1ela_image =
    ImageEnhance.Brightness(
        ela_image).enhance(scale * amplify)return
    ela_image
```

Listing 3: ELA Preprocessing Implementation
This function ensures that compression-level discrepancies between authentic and AI-generated image regions are visually amplified for CNN input, resulting in higher classification accuracy than raw pixel analysis.

Cloud Deployment and Infrastructure
DeepShield AI is deployed on cloud infrastructure platforms such as AWS or Azure. Con- tainer orchestration services manage scalability and availability. Auto-scaling groups dy- namically adjust computational resources based on incoming media submission volume.

Continuous integration and deployment pipelines ensure rapid model updates while main- taining system stability. Monitoring tools track server performance, database response times, and detection pipeline latency metrics.

The architectural design prioritizes resilience and redundancy. Backup servers and failover mechanisms ensure uninterrupted service during hardware failures or unexpected network disruptions.

VI. CORE ALGORITHMIC FRAMEWORK

The intelligence of DeepShield AI is driven by a set of algorithmic modules that collectively determine detection classification, evidence scoring, and severity prioritization. These al- gorithms are designed to operate under real-time constraints while ensuring computational efficiency and decision accuracy.

ELA-Based Image Forensic Analysis
To detect AI-generated or manipulated image content, DeepShield AI computes the pixel- wise Error Level Analysis map between the original image Oi and the re-compressed image Ii at quality factor q:

$$ELA(O_i) = O_i - I_i \quad (1)$$

The amplified ELA map used as CNN input is defined as:

$$ELA_{amp}(x, y) = clip(\gamma \cdot |O(x, y) - I_q(x, y)|, 0, 255) \quad (2)$$

where $\gamma = 10$ is the amplification factor and $q = 90\%$ is the re-compression quality level. Authentic images exhibit uniform error levels across uniformly compressed regions, whereas AI-generated images display non-uniform patterns due to differing synthesis and compression histories.

CNN Classification Algorithm
The CNN classifier processes ELA-enhanced images through four convolutional blocks with filter sizes 32, 64, 128, 256 , each followed by 2 2 max-pooling. The convolutional operation at each layer is:

$$y_{m,n} = (f * k)_{m,n} = \sum_i \sum_j f_{i,j} \cdot k_{m-i,\, n-j} \quad (3)$$

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

where f is the input feature map and k is the learned kernel. The Rectified Linear Unit (ReLU) activation introduces non-linearity:

$$\mathrm{ReLU}(x) = \max(0, x) \quad (4)$$

The output layer applies softmax to produce class probabilities over C = 2 classes (real vs. synthetic):

$$\hat{p}(y = c \mid x) = \frac{e^{z_c}}{\sum_{j=1}^{C} e^{z_j}} \quad (5)$$

The model is trained with categorical cross-entropy loss:

$$L = -\sum_{m=1}^{M} y_m \cdot \log \hat{y}_m \quad (6)$$

optimised using the Adam optimiser with learning rate $\alpha = 0.001$.

Time Complexity Analysis

Let n denote the number of image patches evaluated during analysis.

ELA computation requires O(n) operations, as each pixel position is evaluated once. CNN inference per layer requires O(n k2 c) operations. The total computational complexity of the image classification pipeline is:

$$T(n) = O(n \cdot k^2 \cdot c \cdot L) \quad (7)$$

where k is the kernel size, c is the channel count, and L is the number of convolutional layers. Given fixed architecture parameters this simplifies to O(n), well within real-time processing constraints.

VII. DETECTION SEVERITY CLASSIFICATION MODEL

Forensic prioritization is essential when multiple media submissions are received simulta- neously. DeepShield AI implements a confidence-based severity classification model that categorizes detected cases into Critical, High, Moderate, and Low severity levels.

Severity levels are determined based on:

Modality-specific confidence scores

Number of affected modalities

Temporal consistency anomaly magnitude

Case metadata and submission context

A simplified decision logic for severity classification is shown below.

```
1  def classify_severity(confidence, modalities_flagged): if
       confidence > 0.95 and modalities_flagged >= 2:
2          return "Critical"
3
```

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

```
4  if          confidence > 0.85:
5              return  " High "
6  if          confidence > 0.65:
7              return  " Moderate "
8  return " Low "
```

Listing 4: Detection Severity Classification

Critical cases are immediately escalated to senior forensic analysts for manual review, while lower severity cases may be routed through automated reporting pipelines.

VIII. VIDEO PIPELINE: FRAME AND TEMPORAL ANALYSIS

The video pipeline samples K key frames f1, f2, . . . , fK at uniform intervals and applies the image CNN to each frame independently:

$$s_k = P(\text{fake} \mid f_k; \theta_{CNN}), \quad k = 1, \ldots, K \qquad (8)$$

Temporal consistency is assessed by computing the dense optical flow field between adjacent frames:

$$u^* = \arg\min_u \iint \Big(\partial_I u + \partial_I v + \partial_y + \partial_t \Big)^2 + \lambda \Big(\frac{\partial u}{\partial x} \Big)^2 + \Big(\frac{\partial v}{\partial x} \Big)^2 \Big) \, dx \, dy \qquad (9)$$

where u = (u, v) is the optical flow field and λ controls spatial smoothness. An LSTM-based RNN processes the sequence of flow statistics and per-frame scores to detect temporally inconsistent generation artefacts. The aggregated clip-level detection score is:

$$S_{vid} = \alpha \bar{s} + (1 - \alpha)\, s_{temporal} \qquad (10)$$

where $\bar{s} = \frac{1}{K}\sum_{k=1}^{K} s_k$ is the mean frame detection score, $s_{temporal}$ is the LSTM-derived temporal anomaly score, and $\alpha \in [0, 1]$ is a weighting hyperparameter.

Algorithm 1 Video Deepfake Detection Algorithm

Ensure: ClipDetectionScore

1: Extract key frames {f1, . . . , fK} at rate FrameSampleRate
2: for each frame fk do
3: Compute ELA map: ELAk ← ELA(fk)
4: Run CNN: sk ← P (fake | ELAk)
5: Compute optical flow: uk ← OpticalFlow(fk, fk−1)
6: end for
7: Compute mean frame score $\bar{s}$
8: Compute temporal anomaly score stemporal via LSTM
9: return $S_{vid} = \alpha \bar{s} + (1 - \alpha) s_{temporal}$

System Optimization Considerations

To further reduce latency, caching strategies are implemented for frequently accessed frame feature maps. GPU-accelerated batch inference ensures CNN evaluations complete within sub-second response times.Parallel processing techniques allow simultaneous evaluation of image, audio, and video pipelines, reducing total decision time during

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

high-volume analysis events.Through this multi-layered algorithmic design, DeepShield AI achieves intelligent, efficient, and scalable media forensic coordination.

IX. AUDIO PIPELINE: SPECTROGRAM ANALYSIS

Beyond image and video analysis, DeepShield AI incorporates a real-time synthetic audio detection system that enables reliable identification of AI-generated or cloned voices. This component reduces false negatives in multi-modal deepfake scenarios where only the audio track has been manipulated.The audio pipeline converts raw waveforms to mel-frequency spectrograms and feeds them into CNN architectures tuned for synthetic voice artefacts, combined with handcrafted signal processing features for robust detection.

Raw audio a(t) is transformed into a spectrogram S(f, τ) using the Short-Time Fourier Transform:

$$S(f, \tau) = \sum_{t=-\infty}^{\infty} a(t)\, w(t - \tau)\, e^{-j2\pi ft} \quad (11)$$

where w( ) is a Hamming window function. The resulting spectrogram serves as input to a Spectral CNN trained to distinguish natural speech from AI-synthesised voice characteristics.

Handcrafted Signal Features

Complementary handcrafted features augment the spectral representation:
Spectral CNN: Learns patterns of synthetic generation across frequency and time dimensions.
Jitter J: Cycle-to-cycle perturbation in fundamental frequency F0, indicative of un- natural periodicity.
Shimmer Σ: Amplitude variation between consecutive glottal cycles.
Phase anomaly score Φ: Deviation from expected phase coherence across harmonic frequencies.
The combined audio feature vector is:

$$v_{aud} = [sCNN, J, \Sigma, \Phi]^{\top} \quad (12)$$

where sCNN is the penultimate-layer embedding from the Spectral CNN.
The analyst consultation flow proceeds as follows: 1. Analyst requests case review session.
2. System verifies authentication credentials. 3. Available senior analyst is assigned. 4. Secure video session is established via WebRTC. 5. Review notes and escalation decisions are stored in cloud records.
A simplified backend signaling server implementation is illustrated below.

```
1
2
3
4
5
6
```

```
const io = require("socket.io")(server);
io.on("connection", socket => {
      socket.on("offer", data =>
            { socket.broadcast.emit("offer",
            data);
      });
      socket.on("answer", data =>
            { socket.broadcast.emit("
            answer", data);
      });
      socket.on("candidate", data => {socket.
            broadcast.emit("candidate", data);
      });
});
```

Listing 5: WebRTC Analyst Consultation Signaling Server
This signaling mechanism facilitates encrypted peer-to-peer analyst communication while preserving case confidentiality.

X. EVIDENCE FUSION MODULE

The three per-modality detection scores Simg, Saud, Svid are normalised and combined by the Evidence Fusion Module (EFM) to produce a single interpretable confidence output. Let
S˜m denote the normalised score for modality m:

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

S˜m
= Sm − µm
σ
(13)
m
where µm and σm are the empirical mean and standard deviation over a calibration set. The final unified confidence score is:Sfinal = Σ
m∈{img, aud, vid}
wm · S˜m, wm = 1 (14) Σ
m
Provenance metadata—model version, dataset fingerprint, and UNIX timestamps—is bun- dled with each per-modality evidence snippet to support audit-chain verification.
The following snippet illustrates the fusion logic:

```
def fuse_evidence(img_score, aud_score, vid_score, weights): scores =
    [img_score, aud_score, vid_score]
    normalized = [
        (s - mu) / sigma
        for s, mu, sigma in zip(scores, MU, SIGMA)
    ]
    final_score = sum(
        w * s for w, s in zip(weights, normalized)
    )
    return final_score
```

Listing 6: Evidence Fusion Module Implementation

XI. SECURITY AND PRIVACY FRAMEWORK

Forensic systems must ensure strict protection of sensitive case information. DeepShield AI implements multi-layered security mechanisms, including authentication controls, encrypted data storage, secure API communication, and role-based access management.

Authentication and Authorization
Authentication is implemented using JSON Web Tokens (JWT). After successful login, a signed token is issued to the client. This token must accompany subsequent API requests.

```
const jwt = require("jsonwebtoken"); function
authenticateToken(req, res, next) {
    const token = req.headers["authorization"]; if (!token)
    return res.sendStatus(403);
    jwt.verify(token, process.env.SECRET_KEY, (err, user) => {if (err) return
        res.sendStatus(403);
        req.user = user;

        next();
    });
}
```

Listing 7: JWT Authentication Middleware

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Role-based access control ensures that:
Operators can submit and view only their own case records
Analysts can access assigned case evidence and detection reports
Administrators can manage system-level model configurations

Data Encryption Strategy
To prevent unauthorized access, DeepShield AI encrypts sensitive case data before storage using AES-256 symmetric encryption. The encryption workflow is defined as:

$C = Ek(P)$ (15)
where P represents plaintext evidence data, k is the encryption key, and C is the resulting ciphertext. A simplified encryption example is shown below.

```
const crypto = require("crypto"); function
encryptData(data, secretKey) {
    const cipher = crypto.createCipher( "aes-256
        -cbc", secretKey);
    let encrypted = cipher.update(data, "utf8", "hex"); encrypted += cipher.
    final("hex");
    return encrypted;
}
```

Listing 8: AES-256 Evidence Encryption

Decryption is performed only when authorized users access case records through validated API sessions.

Secure API Communication
All external API communication occurs over HTTPS using SSL/TLS protocols. Transport layer encryption ensures that data packets transmitted between client devices and servers cannot be intercepted or modified.
Token validation mechanisms prevent replay attacks and unauthorized API invocation.
Threat Model and Risk Assessment
A comprehensive threat model was developed to evaluate potential vulnerabilities in the DeepShield AI ecosystem.
Potential threats include:
Unauthorized evidence database access
Man-in-the-middle attacks on analyst communication channels
Identity spoofing by malicious operators
Distributed denial-of-service (DDoS) attacks on detection endpoints
Insider threats from privileged administrators To mitigate these risks, the system incorporates:
Encrypted communication channels
Token expiration and refresh mechanisms
Intrusion detection monitoring
Rate limiting for API endpoints
Secure cloud configuration policies

Access Control Policy Model
Access control decisions follow the principle of least privilege. Each user role is assigned predefined permissions, ensuring minimal exposure of sensitive forensic resources. Let U represent the set of users and R represent system resources. The access policy is defined as:$Access(U_i, R_j) = \begin{cases} 1 & \text{if role permits access} \\ 0 & \text{otherwise} \end{cases}$ (16)
This binary access function enforces strict authorization validation before any resource re- trieval is permitted.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

XII. COMPLIANCE AND ETHICAL CONSIDERATIONS

Forensic detection systems must adhere to global regulatory frameworks governing data protection. DeepShield AI is designed to align with internationally recognized standards such as HIPAA for sensitive data protection and GDPR for personal data privacy.

Data minimization principles are followed by storing only essential case information required for detection and evidence coordination.

Operator consent mechanisms are integrated into the registration process. Users must ex- plicitly authorize the storage and processing of submitted media content.

Audit logging records all access events, enabling traceability and compliance verification.
Ethical AI considerations are incorporated into severity classification modules to prevent al- gorithmic bias in decision-making. Regular review processes ensure fairness and transparency across demographic and content type boundaries.Through layered security architecture and regulatory compliance alignment, DeepShield AI establishes a trustworthy digital forensic coordination environment.

Experimental Evaluation and Performance Analysis
To validate the effectiveness of DeepShield AI, a series of controlled simulations were con- ducted to analyze detection accuracy, system scalability, and data security performance. The evaluation environment emulated real-world media submission scenarios with variable detection request densities.

Simulation Environment Setup
The experimental testbed consisted of:
185,015 images from the AI-ArtBench dataset (human-drawn, latent diffusion, and standard diffusion subsets)
500 simulated audio clips (natural speech and TTS-synthesised)
200 video samples with varying deepfake manipulation levels
Mixed severity detection cases (Critical, High, Moderate, Low)
The simulation environment was implemented using publicly available synthetic datasets to ensure ethical compliance and avoid real personal data exposure.
Each simulated media submission included:
Media type and file path
Reported manipulation context
Submission timestamp
Operator case identifier
The evaluation metrics focused on: 1. Detection accuracy, 2. False positive and false negative rates, 3. Pipeline processing latency, 4. System throughput under load, 5. Encryption overhead.
Detection Accuracy Analysis
Detection accuracy is defined as the proportion of correctly classified media submissions:

A =

Three systems were compared:
TP + TN

TP + TN + FP + FN
(17)
Baseline CNN (raw pixels, no ELA)
CNN with PRNU preprocessing
DeepShield AI (CNN with ELA)
Table 1 presents the average detection accuracy comparison.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Table 1: Average Detection Accuracy Comparison

System	Accuracy
Baseline CNN	0.81
CNN with PRNU	0.93
DeepShield AI (ELA)	0.99

The results demonstrate that DeepShield AI reduces misclassification by approximately 95% compared to the baseline CNN, confirming the value of ELA preprocessing for generalised synthetic media detection.

Pipeline Load Distribution Analysis

Detection pipeline load is measured using utilization ratio across available GPU processing nodes:

ActiveJobs
U =
TotalCapacity
(18)

After deploying the load-balanced microservices architecture:
Average utilization variance decreased by 31%
Overloaded processing nodes reduced from 12 to 3
GPU allocation improved across all modality pipelines
This indicates that resource-aware pipeline scheduling enhances system-wide processing bal- ance.
Detection Pipeline Efficiency
Pipeline latency is evaluated by measuring the time between media submission and classifi- cation output:

Simulation results show:

$$T_{process} = T_{output} - T_{submission} \quad (19)$$

35% faster classification compared to sequential single-modality pipelines
28% reduction in idle GPU processing time
Improved resource utilization through parallel pipeline execution
The algorithmic pipeline design ensures minimal delay while preserving capacity for subse- quent media submissions.

Scalability and Load Testing

To assess system scalability, concurrent submission simulations were executed with increasing load levels.
Test scenarios included:
100 concurrent users
500 concurrent users
1000 concurrent users
2000 concurrent users
The performance metric evaluated was average server response latency.

Table 2: Scalability Performance Under Load

Concurrent Users	Avg Latency (ms)
100	130
500	195
1000	280
2000	430

Even at 2000 concurrent users, latency remained within acceptable operational thresholds, demonstrating effective horizontal scaling capabilities of the containerized microservices de- ployment.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Throughput Evaluation
System throughput is defined as:

Throughput =
TotalRequests TotalTime
(20)
Under peak simulation conditions, DeepShield AI sustained high throughput without ser- vice failure, confirming the robustness of the microservices architecture under high-demand forensic scenarios.

Security Performance Analysis
Security performance was evaluated by simulating unauthorized access attempts and replay attack scenarios.
Metrics analyzed:
Token validation time
Encryption overhead per request
API authentication delay
The AES-256 encryption process introduced negligible overhead (approximately 5–8 millisec- onds per request), which is acceptable within real-time forensic response constraints.
JWT validation added an average delay of 3 milliseconds, demonstrating efficient authen- tication processing consistent with findings reported in comparable secure digital platforms [?].

Comparative System Evaluation
DeepShield AI was compared against existing detection platforms based on feature integra- tion capabilities.

Table 3: Comparative Feature Analysis

Feature	Traditional	Single-Modal	DeepShield AI
Multi-Modal Detection	No	No	Yes
ELA Preprocessing	No	Partial	Yes
Temporal Video Analysis	No	No	Yes
Audio Spectrogram CNN	No	No	Yes
Automated Severity	No	No	Yes
Cloud Evidence Records	Limited	Yes	Yes
Load Balancing	No	No	Yes

The integrated design of DeepShield AI differentiates it from isolated single-modality detec- tion solutions.

XIII. STATISTICAL VALIDATION

To validate accuracy improvements statistically, paired t-tests were conducted comparing detection accuracy before and after ELA preprocessing.
Let $\bar{x}1$ represent mean accuracy of the baseline CNN and $\bar{x}2$ represent mean accuracy of the ELA-enhanced model:

$$t = \bar{x}1 - \sqrt{\bar{x}2}$$
(21)
$$s/\ n$$
The resulting p-value was less than 0.05, indicating statistically significant improvement in detection performance attributable to ELA preprocessing.

XIV. DISCUSSION OF RESULTS

The experimental findings demonstrate that intelligent ELA preprocessing combined with multi-modal pipeline fusion significantly enhances deepfake detection performance.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904246

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Key improvements include:
Reduced detection pipeline latency
Improved classification accuracy across all media types
Enhanced analyst workflow efficiency through severity classification
Scalable system stability under concurrent load
Minimal encryption overhead for secure evidence handling
These results confirm that integrating pixel-level forensic analysis with unified multi-modal detection infrastructure produces measurable benefits in synthetic media identification per- formance.

Practical Deployment Considerations

The real-world deployment of DeepShield AI requires coordination between media orga- nizations, forensic investigation authorities, and cloud infrastructure providers. Successful integration depends on interoperability with existing digital media management systems and evidence communication networks.
In metropolitan environments, integration with centralized content moderation pipelines can significantly enhance automated detection response. For rural or resource-constrained
forensic units, lightweight cloud infrastructure combined with low-bandwidth optimization techniques can ensure operational continuity even in areas with limited connectivity.
Deployment models may include:
Government-managed digital forensics networks
Private media verification consortium frameworks
Smart city integrated content moderation platforms
University research and cybersecurity campus ecosystems
Cloud-based Software-as-a-Service (SaaS) deployment allows rapid scalability without re- quiring heavy on-premise infrastructure investments.

XV. SYSTEM LIMITATIONS

Despite its advantages, DeepShield AI faces certain limitations that require consideration.
First, continuous internet connectivity is essential for real-time media ingestion and analyst consultation. In regions with unstable network infrastructure, pipeline latency may increase significantly.
Second, ELA accuracy may degrade on images that have undergone multiple rounds of re- compression prior to submission, as the compression history baseline becomes ambiguous.
Third, initial implementation costs, including infrastructure setup, cloud hosting, and model training, may present financial challenges for smaller forensic facilities.
Fourth, algorithmic decision-making systems must be continuously monitored to prevent unintended bias across content types, demographic representations, or generative model architectures not seen during training.
Addressing these limitations requires infrastructure investment, adaptive algorithm tuning, and ongoing regulatory oversight.

XVI. FUTURE RESEARCH DIRECTIONS

Future enhancements of DeepShield AI may include artificial intelligence-driven predictive analytics capable of forecasting deepfake hotspots based on historical incident and platform activity data.
Machine learning models can be trained to predict content manipulation trends using time- series forecasting techniques:

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, \ldots, y_{t-n}) \qquad (22)$$

where $\hat{y}_{t+1}$ represents predicted detection load or manipulation probability at the next time step.
Integration with IoT sensor networks may allow automatic synthetic media flagging triggered by anomalous device-level metadata patterns.
Blockchain-based evidence record storage could further enhance data immutability and legal admissibility. A decentralized ledger framework would prevent unauthorized record modifi- cation while preserving full auditability.
Real-time network traffic integration using intelligent content delivery systems may further optimize analyst routing through dynamic case assignment algorithms.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Smart City Integration Potential

DeepShield AI aligns with smart city digital safety initiatives that emphasize data-driven public service optimization and automated threat response.

By integrating with city-wide IoT sensors, traffic monitoring systems, and public safety networks, deepfake detection coordination can become fully automated and predictive.

A smart city-enabled architecture may include:

Real-time social media feed monitoring

Automated synthetic media flagging across broadcast channels

IoT-enabled forensic device telemetry

Centralized emergency content moderation dashboards

Such integration would enable predictive routing and automated evidence prioritization based on city-wide data streams.

Extended Ethical and Social Considerations

Ethical implementation of digital forensic systems requires transparency, fairness, and inclu- sivity.

Algorithmic detection systems must avoid discrimination based on demographic, cultural, or socioeconomic factors. Continuous auditing of severity classification models is essential to ensure equitable treatment allocation across all content categories and population groups.

Data privacy must remain paramount. Operators should retain control over data sharing permissions and consent management throughout the investigative lifecycle.

Transparency mechanisms should inform users about how their submitted media is processed and how detection decisions are determined.

Societal trust in digital forensic systems depends on accountability, security, and responsible AI governance.

XVII. CONCLUSION

T

his paper presented DeepShield AI, a comprehensive unified multi-modal deepfake detection platform designed to address systemic inefficiencies in synthetic media forensics.

By integrating ELA-enhanced CNN image classification, temporal video analysis, spectrogram- based audio detection, encrypted cloud evidence records, and scalable microservices archi- tecture, DeepShield AI offers a unified digital forensics framework.

Simulation-based evaluation demonstrated:

Approximately 99% image detection accuracy using ELA-enhanced CNN

Improved multi-modal classification coverage across image, audio, and video

Enhanced analyst pipeline efficiency through severity classification

Secure encrypted evidence data management

Scalability under high concurrent user demand

The results confirm that intelligent integration of pixel-level forensic analysis and cloud-based multi-modal detection infrastructure significantly enhances synthetic media identification performance.

DeepShield AI represents a scalable, secure, and adaptable solution capable of supporting next-generation smart city forensic ecosystems.

XVIII. ACKNOWLEDGMENT

The authors express gratitude to academic mentors, cybersecurity professionals, and tech- nical advisors who contributed insights during the conceptualization and development of DeepShield AI.

Ethical Statement

All experiments described in this research were conducted using publicly available synthetic datasets. No real personal data was accessed or processed during the evaluation phase. hyperref

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. L. Guarnera, O. Giudice, and S. Battiato, "Fighting deepfake by exposing the convolu- tional traces on images," IEEE Access, vol. 8, pp. 165085–165098, 2020.
2. M. Masood et al., "Deepfakes generation and detection: State-of-the-art, open chal- lenges, countermeasures, and way forward," Appl. Intell., vol. 53, no. 4, pp. 3974–4026, 2022.
3. N. Tinago, S. F. Verkijika, and K. E. Mamabolo, "Deepfakes in visual art: Differen- tiating AI-generated art from human art using convolutional neural networks," IEEE Access, vol. 13, pp. 141484–141495, 2025.
4. C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, and C. Busch, Handbook of Digital Face Manipulation and Detection. Cham: Springer, 2022.
5. F. Martin-Rodriguez, R. Garcia-Mojon, and M. Fernandez-Barciela, "Detection of AI- created images using pixel-wise feature extraction and convolutional neural networks," Sensors, vol. 23, no. 22, p. 9037, 2023.
6. R. Rafique et al., "Deep fake detection and classification using error-level analysis and deep learning," Sci. Rep., vol. 13, p. 7422, 2023.
7. F. Chamot, Z. Geradts, and E. Haasdijk, "Deepfake forensics: Cross-manipulation ro- bustness of feedforward and recurrent convolutional forgery detection methods," Foren- sic Sci. Int. Digit. Investig., vol. 40, 2022.
8. R. S. R. Silva et al., "ArtBrain: An explainable end-to-end toolkit for classification and attribution of AI-generated art and style," arXiv:2412.01512, 2024.
9. M. E. S. Tenorio, "PNTING: Detecting AI in the painting world," Stanford CS231n, 2024.
10. S. Yan et al., "A sanity check for AI-generated image detection," arXiv:2406.19435, 2024.
11. P. Johnson, "Microservices architecture for scalable applications," IEEE Software, 2019.
12. S. Verma et al., "Secure health data management using AES encryption," IEEE Security & Privacy, 2021.
13. L. Zhang et al., "IoT-enabled smart healthcare," IEEE Internet of Things Journal, 2021.
14. D. Kapoor, "Smart city healthcare networks," IEEE Smart Cities Conference Proceed- ings, 2021.
15. F. Ahmed, "Scalable cloud-based emergency systems," IEEE International Conference on Healthcare Informatics, 2023.
16. M. Chen, "AI in emergency triage systems," IEEE AI Magazine, 2022.
17. H. Zhao, "Predictive analytics in emergency medicine," IEEE Transactions on Biomed- ical Engineering, 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com