Computing&AI Connect
ISSN: 3104-4719
2026, Article ID. 2026.0040, Cite as: https://doi.org/10.69709/CAIC.2026.122445

SCIFINITI PUBLISHING

Research Article

OPEN ACCESS

Machine Learning-Driven Multimodal Forensic Data Fusion for Enhanced Disaster Victim Identification: Integrating DNA, Dental, and Facial Image Data

Idowu Olugbenga Adewumi[✉,1,*], **Wumi Ajayi**[✉,2], **Oluwafisayo Babatope Ayoade**[✉,3], **Victoria Bolanle Oyekunle**[✉,4], **Akintayo Ayoade**[✉,5], **Azeez Ajani Waheed**[✉,4]

[1] Software Engineering Program, Department of Computer and Information Engineering, Faculty of Natural and Applied Science, Lead City University, Ibadan, Oyo State, Nigeria
[2] Software Engineering Department, School of Computing, Babcock University, Ilisan Remo, Ogun State, Nigeria
[3] School of Computer, Data and Mathematical Sciences, Computing and Engineering, Western Sydney University, Sydney Australia
[4] Department of Computer Science, Faculty of Natural and Applied Science, Lead City University, Ibadan, Oyo State, Nigeria
[5] Department of Computer and Information Engineering, Faculty of Natural and Applied Science, Lead City University, Ibadan, Oyo State Nigeria

Article History
Received: April 29, 2026 Accepted: August 17, 2026

ACCEPTED ARTICLE

Abstract
Using synthetic disaster simulations, a machine learning-based multimodal forensic data fusion framework was developed for disaster victim identification (DVI). By integrating DNA, dental, and facial image data from 10,000 synthetic records, the framework comprised 5000 antemortem and 5000 postmortem profiles. DNA data were represented using 200 synthetic SNP markers, dental data were represented using features such as tooth count, treatment history, and bone loss, and facial data were represented as 64-dimensional ResNet-50 embeddings. The dataset was divided into training, validation, and testing sets in a ratio of 70:15:15. Three unimodal models and two fusion strategies were evaluated in 10 independent runs. The proposed feature-level fusion model showed the best performance, with 94.8 ± 0.4% accuracy, 0.94 ± 0.01 precision, 0.94 ± 0.01 recall, 0.94 ± 0.02 F1, and 0.97 ± 0.01 macro-averaged OvR ROC-AUC. The accuracy of decision-level fusion was 93.5 ± 0.5%, which was better than the DNA-only (88.4% ± 0.6), dental-only (82.7% ± 0.7), and facial image-only (84.9% ± 0.5) classifiers. When one of the three modalities was missing, the accuracy remained at or above 90.5%. Under severely degraded multimodal conditions, the accuracy was 88.7 ± 0.7%. ANOVA revealed significant differences between unimodal and fusion models ($F = 18.45$, $p = 0.0002$). Attention weights of 40% for DNA, 34% for dental data, and 26% for facial data enhanced the model's interpretability. The results demonstrate strong performance on synthetic DVI simulations but require validation using real forensic data.

Keywords:
multimodal forensic data fusion; disaster victim identification; machine learning; DNA and dental data integration; multimodal biometrics; synthetic forensic data; feature-level fusion

1. Introduction

1.1. Background of the Study

Disaster victim identification is a branch of forensic science that deals with identifying human remains resulting from disasters such as earthquakes, floods, bomb blasts, building collapses, fires, and industrial accidents [1]. In such cases, standard identification methods, such as visual identification of the body, fail because of severe decay, fragmentation, and commingling of remains. Conse-

* **Corresponding Author:**
Idowu Olugbenga Adewumi, Software Engineering Program, Department of Computer and Information Engineering, Faculty of Natural and Applied Science, Lead City University, Ibadan, Oyo State, Nigeria,
adewumi.io@fcaib.edu.ng

© 2026 Copyright by the Authors.
Licensed as an open access article using a CC BY 4.0 license.

quently, scientific identifiers help forensic specialists determine probable identity. Forensic techniques such as DNA, dental, and facial image analysis, among others, are important in disaster settings.

Independently, each forensic modality has recently improved substantially in accuracy, computational support, and 'value of evidence' analysis. As a result, DNA offers high discriminatory power, dental records demonstrate greater postmortem durability, and facial images enable rapid preliminary matching [2]. In reality, no single modality can be relied upon in a disaster scenario. This is because the various forensic modalities may involve incomplete, degraded, or unavailable data. As a result, there are increasing demands for forensic data fusion techniques for identification. Fusion is the process of combining data from different sources to obtain more precise, robust, reliable, and complete inferences under conditions of uncertainty. This challenge gave rise to machine learning-based fusion frameworks for disaster victim identification.

1.2. Problem Statement

Most unimodal forensic systems and several Disaster Victim Identification systems function independently, relying on DNA, dental, or facial information alone. Unimodal systems suffer from several serious shortcomings. First, they suffer from overdependence on a single type of evidence whose availability and quality are not always guaranteed. For example, DNA reference samples may degrade or be lost due to exogenous factors.

Likewise, dental records may be incomplete or unavailable; moreover, facial distortion may result from trauma or decomposition. Second, unimodal systems provide insufficient forensic evidence in complex disaster settings [3]. The identification of disaster victims follows internationally accepted forensic principles, which combine primary identifiers, such as DNA, fingerprints, and dental records, with secondary identifiers, with a view to confirming the identity of human remains when visual recognition fails, notably in mass disasters [4,5]. The overall discriminative capability of an integrated system employing multiple biometric traits remains largely unexploited. A universally applicable forensic identification architecture must be designed to incorporate multimodal fusion and a graceful degradation mechanism for scenarios in which one or more modalities may be temporarily unavailable.

1.3. Literature Gap

The application of machine learning in forensic science has been investigated extensively in recent years. Most of these studies focused on unimodal or partially fused approaches applied to genomic or dental imaging systems. In DNA-based systems, machine learning is mainly applied to classify DNA sequences or encode genotype and phenotype information. These works also discuss using machine learning algorithms for STR profiling [6].

Similarly, dental recognition systems mostly rely on image-level matching or deep-learning-based odontometric feature extraction. Moreover, occlusion, lighting variation, and postmortem alterations make facial deep-learning recognition models highly vulnerable in biometric identification tasks.

Critically, there is no multimodal framework that aims to use genomic data, dental data, and facial images as input within a unified machine learning framework for Disaster Victim Identification. Most existing multimodal learning research in biometrics and medical imaging does not account for forensic constraints, such as missing modalities and degraded evidence, or ethical constraints in postmortem datasets [7]. Many existing approaches to feature-level and decision-level fusion have not adequately addressed how to exploit complementary information within a single framework without sacrificing generalization ability.

1.4. Contributions and Novelty

This study presents a novel machine learning-driven multimodal forensic data fusion framework for Disaster Victim Identification by integrating DNA, dental, and facial image data. The key contributions and innovations of this work are summarized as follows:

i. The study develops a three-modality forensic fusion framework that integrates DNA sequences, dental odontometric features, and facial image embeddings into a unified identification system, improving the reliability of victim matching in disaster scenarios.

ii. The proposed approach incorporates both feature-level fusion and decision-level fusion strategies, enabling comparative evaluation of early integration (feature fusion) and late integration (decision fusion) techniques within the same forensic pipeline.

iii. The framework demonstrates strong robustness to missing or degraded modalities, ensuring stable identification performance even when one or more forensic data sources are unavailable, which is a common challenge in simulated disaster environments.

iv. The model introduces an attention-based interpretability mechanism, where adaptive weights are assigned to each modality (DNA, dental, fa-

ACCEPTED ARTICLE

2026, Vol. 3, Article ID. 2026.0040
https://doi.org/10.69709/CAIC.2026.122445

cial), enabling transparent decision-making and improving explainability for forensic experts and legal validation processes.

2. Literature Review

2.1. DNA Forensic Methods

Due to the high discriminative power and stability of DNA, it is considered to be the best DVI modality. Moreover, samples such as bone and teeth, which can get damaged during a disaster, are ideal for DNA profiling. The discovery of the genetic identity of missing and deceased individuals started over 20 years ago. Initially, forensic genetics relied on short tandem repeat (STR) analysis. However, contemporary DNA forensic research is evolving from classical STR typing to massively parallel sequencing and probabilistic modeling to achieve better mixture interpretation and classification performance. Leveraging DNA analysis with machine learning assists in rapid and accurate identification of victims of mass disasters. Studies [8] and [9] have confirmed this benefit specifically for contaminated or fragmented biological samples. DNA identification is a valuable forensic tool, but it has limitations. DNA profiling remains one of the most reliable forensic identification techniques because of its high discriminatory power and stability, particularly when biological remains are fragmented or visually unidentifiable [10,11].

2.2. Dental Identification Systems

Forensic dentistry plays a significant role in identifying victims with their dental structures intact, especially when the bodies are mutilated beyond recognition. The dental records associated with teeth and odontometric features are highly resistant to decomposition and serve as a stable postmortem identifier. Advances in forensic imaging and machine learning have enabled automated extraction of features from dental radiographs, improving the matching efficiency of ante-mortem and postmortem records [12, 13]. Dental identification has largely been validated under controlled conditions. Convolutional neural networks, support vector machines, and feature-based classification models have been employed for this task, demonstrating strong performance. Notwithstanding, the effectiveness of dental systems is restricted by the unavailability of dental records, differing imaging standards, and reliance on ante-mortem documentation, which reduces their scalability in disaster settings [14]. Forensic odontology has historically represented a primary identification method because dental structures are highly resistant to decomposition and environmental degradation [15].

2.3. Facial Recognition in Forensics

Facial recognition has become a prominent modality in forensic science with the evolution of deep learning architectures (convolutional neural networks, transformer-based models, and multimodal embedding systems). Modern facial recognition systems operate on high-dimensional feature representations and can match identities even when the subject is only partially visible. Recent research indicates that efficient deep fusion models and multimodal feature extraction may improve recognition in challenging forensic image analysis scenarios [16,17]. Even so, forensic facial recognition remains highly susceptible to postmortem changes, decomposition effects, illumination differences, and deteriorating image quality, compromising its reliability as an identification tool in disaster victim identification. As a result, facial recognition is often used as a secondary method for initial screening, not final identification. Traditional facial recognition technology based on facial features has given way to deep representation learning, where discriminative embeddings offer greater robustness in identity verification tasks [18,19].

2.4. Multimodal Fusion in Disaster Victim Identification

Recently, multimodal fusion methods have emerged as a suitable technique for overcoming the limitations of unimodal forensic systems. Fusion methods can be categorized as data-level, feature-level, or decision-level. These methods differ in accuracy, interpretability, and complexity. According to [20] and [21], recent advances in multimodal learning indicate that feature-level fusion benefits representation learning by capturing cross-modal relationships, whereas decision-level fusion improves interpretability, as each network learns independently before being combined into an ensemble. In addition, adaptive attention-based fusion frameworks and hierarchical deep fusion models have improved robustness to incomplete or noisy forensic evidence [22,23].

Many available studies concentrate on a restricted combination of modalities; that is, they usually incorporate only two biometric sources, such as face-fingerprint or DNA-dental systems. To date, no overarching frameworks integrate DNA, dental, and facial image data within a single multimodal architecture designed specifically for Disaster Victim Identification. Moreover, existing models often assume complete data availability, which is unrealistic in disaster environments where modalities are

ACCEPTED ARTICLE

frequently missing or corrupted. Recent forensic research has also highlighted the need for explainable AI that provides transparent decision-making processes to meet legal and humanitarian requirements [24]. Multimodal fusion has been widely investigated as a mechanism for integrating complementary information from heterogeneous sources. Fusion strategies are commonly categorized into feature-level, score-level, and decision-level fusion approaches [25]. There remains a lack of robust, integrated multimodal forensic systems capable of operating under disaster constraints, representing a gap that the present study aims to address using synthetic data.

Table 1 compares existing forensic identification approaches that are targeted toward either individual modalities or general multimodal learning strategies. What remains missing, however, is a dedicated DVI framework that integrates all modalities. According to Barash et al. [1], machine learning techniques have notable advantages for the forensic analysis of DNA evidence. Nevertheless, their method addresses only genetic evidence and not supplementary identifiers. Similarly, Pereira et al. [26] on dental identification and Michalski et al. [27] on facial identification demonstrate advances in deep learning approaches for these purposes, but the resulting systems still depend on a single source of evidence. The general multimodal fusion studies by Jiao et al. [28], Wang et al. [29], and Guarrasi et al. [30] have advanced feature integration, missing-modality handling, and intermediate fusion strategies; however, these were not developed with forensic DVI applications in mind.

Table 1: Comparison of existing studies with developed work.

Study	Modalities	Method	Dataset/Source	Principal Limitation	Gap Addressed by the Developed Work
Barash et al. [1]	Forensic DNA profiles	Critical review of machine-learning applications in forensic DNA profiling, including DNA-profile analysis, interpretation and classification	Published forensic DNA-profiling studies and datasets reviewed in the literature	Focuses on DNA evidence and does not implement or evaluate a three-modality DVI fusion architecture	The developed framework combines DNA evidence with dental and facial information and evaluates identification under missing or degraded modalities.
Pereira et al. [26]	Dental panoramic radiographs	VGG16-based convolutional neural networks for age classification and positive identification through comparison of ante-mortem and postmortem orthopantomograms	1235 orthopantomograms obtained from 1050 patients aged 16–30 years	Restricted to dental evidence and a defined age-range sample; does not incorporate DNA or facial evidence	The developed framework integrates dental information with genetic and facial representations to reduce dependence on the availability and quality of dental records alone.
Michalski et al. [27]	Ante-mortem and postmortem facial images	Application and assessment of automated facial-recognition technology for disaster victim identification	Facial-image comparisons considered within a DVI context	Facial recognition depends on the availability and quality of suitable ante-mortem and postmortem images and is not integrated with the primary identifiers of DNA and dental evidence	The developed framework reduces reliance on facial evidence by combining facial embeddings with DNA and dental modalities.

ACCEPTED ARTICLE

2026, Vol. 3, Article ID. 2026.0040
https://doi.org/10.69709/CAIC.2026.122445

Table 1: *Cont.*

Study	Modalities	Method	Dataset/Source	Principal Limitation	Gap Addressed by the Developed Work
Jiao et al. [28]	General multimodal data, including images, text, video, audio and sensor information	Comprehensive review and taxonomy of early, deep, late and hybrid deep-learning fusion approaches	Published multimodal-learning studies from several application domains	Provides a general methodological review rather than an implemented forensic or DVI-specific identification system	The developed work operationalizes feature-level and decision-level fusion within a forensic DVI architecture involving three identification modalities.
Wang et al. [29]	Chest radiographs, radiology reports and structured clinical data	Transformer-based bi-modal fusion modules combined into a tri-modal architecture, with multivariate loss functions for missing-modality robustness	MIMIC-IV and MIMIC-CXR datasets used for a 14-label disease-diagnosis task	Designed for medical diagnosis rather than forensic identity matching; its modalities are not established DVI identifiers	The developed framework transfers missing-modality fusion principles to victim identification using DNA, dental and facial evidence.
Guarrasi et al. [30]	Multimodal biomedical information, including imaging, textual and genetic data	Systematic review and formalization of intermediate-fusion methods in multimodal deep learning	Published multimodal biomedical deep-learning studies	A methodological review without an implemented DVI system; it does not experimentally integrate DNA, dental and facial identity evidence	The developed work implements intermediate or feature-level fusion alongside decision-level fusion and modality weighting for a specific forensic DVI task.
Developed Work	DNA SNP profiles, dental features and facial-image embeddings	Feature-level fusion, decision-level fusion and attention-based modality weighting	10,000 synthetic DVI records comprising 5000 ante-mortem and 5000 postmortem profiles	Uses synthetically generated forensic data and therefore requires subsequent validation with operational DVI datasets	Provides a unified three-modality DVI framework, compares early and late fusion strategies, and evaluates performance when individual modalities are missing or degraded.

ACCEPTED ARTICLE

The framework specifically addresses these issues by integrating the three complementary forensic modalities of DNA SNP profiles, dental features, and facial-image embeddings in a single identification framework. This approach fuses features at multiple levels, along with attention-based modality weighting, to harness the best characteristics of each modality while remaining robust to missing or degraded evidence. In contrast to earlier studies which assess isolated modalities or non-forensic multimodal datasets, the current work develops and evaluates a dedicated DVI framework which exploits 10,000 synthetic forensic records. The DVI framework provides a more insightful and interpretable approach to disaster victim identification scenarios. However, it must be validated using operational forensic datasets.

3. Methodology

The forensic identification system for Disaster Victim Identification (DVI) is enhanced with machine learning capabilities and meets the requirements of a system used as a common platform for DNA sequences, dental X-rays, and facial images.

In the first step, the framework involves systematic data collection from forensic data sources. It is followed

by pre-processing, feature extraction, data integration, and machine learning classification to yield identity estimates. Obtaining additional forensic evidence increases identification accuracy, particularly during postmortem decomposition occurring in mass disasters. The model uses multi-branch mechanism that facilitates the joint integration of the genetic, dental, and facial modality contributions for a robust decision.

Required ante-mortem and postmortem database was created using the simulated forensic data of 5000 persons. A total of 10,000 entries were created. The DNA data consists of 200 SNPs. The dental data is made up of Panoramic data and synthetic reconstruction facial image data. Postmortem DNA degradation, which occurs when there is a high ratio of foreign DNA to target DNA, was simulated with the DNA data by adding noise and gaps to sub- and full-genome sequences.

The number of teeth present, types of treatment received, and level of bone loss score remained relatively stable after death. The information replicates particular damages caused by traumatization or decomposition. The facial images are photographs of mummified remains and modelled images of dead individuals.

Ethically generated data means that this research only uses information from synthetic individuals. To prepare the validation dataset, we normalized continuous features, one-hot encoded categorical variables (sex and disaster type), and managed missing modalities using mean substitution and neural imputation for each respective modality. Also, the embeddings of the facial images were provided with random occlusion and Gaussian noise. The dental and DNA info was normalized to mean zero and variance one to create a normalized dataset.

Deep convolutional neural networks can extract hierarchical representations of complex visual data. Therefore, it can achieve good feature learning for biometric data [25]. The use of attention mechanisms allows models to give different levels of importance to the various feature components, which improves representation learning in complex multimodal systems [31].

Each modality had separate extraction pipelines designed for it. We constructed a ResNet-50 convolutional neural network (CNN) architecture using TensorFlow 2.x and Keras (Google LLC, Mountain View, CA, USA) to extract 64-dimensional facial embeddings corresponding to facial properties. We designed a CNN-based sequence encoder having one dimension using PyTorch (Meta Platforms, Inc., Menlo Park, CA, USA) to extract DNA features and allelic relationship among SNPs. Using Scikit-learn (Python Software Foundation, Wilmington, DE, USA), a principal component analysis (PCA) was performed to reduce the dimensionality of odontometric data while preserving its principal characteristics.

The extracted features from all modalities were mapped into a latent embedding space to align features across different modalities. The embedding space designed for multi-modal matching of ante-mortem and postmortem records minimizes distance between records belonging to a person while maximizing distance records of different persons.

Two fusion methods were investigated to leverage the complementary strengths of the different modalities. The initial technique uses feature-level fusion. It uses standardized embeddings obtained from DNA, dental and facial images. Dimensionality reduction with TensorFlow/Keras (Google LLC, Mountain View, CA, USA) by autoencoder-based compression. A classifier was used to predict identity based on the combined representation. In the second approach all the different classifiers for all the discrete modalities were trained independently which were not fused at all and combined at decision level. The results were then merged using Bayesian fusion and weighted ensemble voting, with weights indicating the reliability of each modality. For example, high-quality DNA or facial image data was given higher confidence.

The developed fusion architecture consists of a multi-branch deep neural network with a branch for each modality. Moreover, layer attention links the subnetworks and will dynamically learn the importance of the modality based on data quality. A graph-based method was also proposed to fuse information representing dependencies between samples and modalities in mass-casualty situations.

The dataset was divided into three parts; training, validation and testing. Further sampling done to ensure representation of disaster class and characteristics of demographic feature. This work employs and evaluates a variety of models including CNNs, XGBoost, and a fusion network designed using an autoencoder. The XGBoost open-source machine learning library developed by the Distributed (DMLC) team and distributed under the Apache Software Foundation, Wilmington, DE, USA and used for modelling. Evaluation was accomplished through various measures like accuracy, precision, recall, F1-score, and Area Under Receiver Operating Characteristic Curve (ROC-AUC). Python programming language (Python Software Foundation, Wilmington, DE, USA) was used to perform statistical calculations and data manipulation with supporting packages, including NumPy (NumPy Developers, Austin, TX, USA) and Pandas (NumFOCUS, Austin, TX, USA).

The outcomes produced by DNA, dental, and facial image models, all utilizing a single method, were tested on

ACCEPTED ARTICLE

various fusion models. The study demonstrated the efficacy of the fusion method, which signifies the importance of using different types of forensics together.

A case study for disaster victim identification was undertaken with postmortem data from a synthetic building collapse. The model's efficacy in modalities not present was evaluated, for instance, when the DNA is degraded or the photo is of a low quality. The model's missing-modality performance accuracy was still high, suggesting robustness of the model under such circumstances. When there is an absence of a single modality, the attention mechanism can exploit the larger and more trustworthy modalities. The findings also show that the system can enhance the forensic investigation process post-disaster. The United Nations Sustainable Development Goals, especially SDG 3 (Good Health and Well-Being), can be supported through the systems' accurate post-disaster victim identification. It also helps achieve SDG 11 (Sustainable Cities and Communities). The system optimizes urban disaster response in diverse ways.

4. Developed Mathematical Model for the Multimodal Forensic Data Fusion System

This section presents the mathematical model for the multimodal forensic data fusion system for Disaster Victim Identification (DVI) involving DNA, dental, and facial images. To enhance victim identification accuracy and resilience, the system applies several feature-level and decision-level machine-learning fusion techniques.

ACCEPTED ARTICLE

4.1. Data Representation

Let the input dataset be represented as:

$$D = \{(X_d, X_{de}, X_f)\} \quad \text{where } X_d \in \mathbb{R}^{n_d}, X_{de} \in \mathbb{R}^{n_{de}}, X_f \in \mathbb{R}^{n_f} \tag{1}$$

X_d represents the DNA data, where each record $x_{d_i} \in \mathbb{R}^{n_d}$ is a vector of n_d SNP features.

X_{de} represents the dental data, where each record $x_{de_i} \in \mathbb{R}^{n_{de}}$ contains odontometric features (e.g., tooth count, bone loss).

X_f represents the facial image data, where each record $x_{f_i} \in \mathbb{R}^{n_f}$ is a 64-dimensional embedding vector extracted using a ResNet-50 CNN.

4.2. Feature Extraction and Preprocessing

Each modality undergoes preprocessing and feature extraction. Denote the feature extraction process for each modality as functions $\mathcal{F}_d$, $\mathcal{F}_{de}$, and $\mathcal{F}_f$, applied to X_d, X_{de}, and X_f respectively.

$$\text{DNA preprocessing}: \ \mathcal{F}_d(X_d) \to X'_d \tag{2}$$

$$\text{Dental preprocessing}: \ \mathcal{F}_{de}(X_{de}) \to X'_{de} \tag{3}$$

$$\text{Facial image preprocessing}: \ \mathcal{F}_f(X_f) \to X'_f \tag{4}$$

These functions include normalization, dimensionality reduction (PCA for dental data), and other transformation methods.

4.3. Fusion Models

(a) Feature Level Fusion

At the feature-level fusion stage, the features from all modalities are concatenated into a single vector:

$$X_{\text{fused}} = [X'_d, X'_{de}, X'_f] \tag{5}$$

This vector is passed through a classifier $\mathcal{C}_f$ to predict the identity of the victim:

$$\hat{y} = \mathcal{C}_f(X_{\text{fused}}) \tag{6}$$

where $\hat{y}$ represents the predicted identity of the victim, and $\mathcal{C}_f$ is a machine learning model (e.g., XGBoost, neural network) trained to output the predicted identity.

(b) Decision Level Fusion

In decision-level fusion, separate classifiers $\mathcal{C}_d$, $\mathcal{C}_{de}$, and $\mathcal{C}_f$ are used for each modality. The classifiers generate probabilistic outputs:

$$\hat{y}_d = \mathcal{C}_d(X'_d), \hat{y}_{de} = \mathcal{C}_{de}(X'_{de}), \hat{y}_f = \mathcal{C}_f(X'_f) \tag{7}$$

These outputs are combined through a fusion mechanism (e.g., weighted ensemble or Bayesian fusion), which computes the final prediction:

$$\hat{y}_{\text{final}} = f_{\text{fusion}}(\hat{y}_d, \hat{y}_{de}, \hat{y}_f) \tag{8}$$

where f_{fusion} is the decision-level fusion function that integrates the results, such as a weighted sum:

$$\hat{y}_{\text{final}} = w_d \hat{y}_d + w_{de} \hat{y}_{de} + w_f \hat{y}_f \tag{9}$$

Here, w_d, w_{de}, w_f are the weights assigned to the respective modalities, representing their relative importance.

4.4. Model Performance Metrics

The performance of both feature-level and decision-level fusion models is evaluated using the following metrics:

2026, Vol. 3, Article ID. 2026.0040
https://doi.org/10.69709/CAIC.2026.122445

Accuracy:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{10}$$

where:

TP: True Positives
TN: True Negatives
FP: False Positives
FN: False Negatives

Precision:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{11}$$

Recall:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

F1-Score:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

ROC-AUC:

$$\text{ROC} - \text{AUC} = \int_0^1 \text{True Positive Rate}(t) \, d(\text{False Positive Rate}(t)) \tag{14}$$

where the ROC curve plots the trade-off between the true positive rate and the false positive rate at various thresholds.

4.5. Resilience to Missing Modalities

Let $\mathcal{D}_{\text{missing}}$ represent the dataset with missing modalities, where some X_d, X_{de}, or X_f are absent. The system's performance is evaluated by measuring accuracy, precision, and recall under these missing-data conditions.

$$\text{Performance}_{\text{missing}} = \frac{\text{Accuracy}_{\text{missing}}}{\text{Accuracy}_{\text{complete}}} \tag{15}$$

This evaluates performance degradation when certain modalities are missing, and the system's robustness is quantified by the percentage drop in performance.

4.6. Transformer-Based Fusion Block (Conceptual Architecture in Model)

To enhance cross-modal interaction, the fusion module can be interpreted as a transformer-style encoder block, which enables contextual dependency learning across modalities.

The transformer-based fusion process follows:

i. Linear projection of each modality into a shared embedding space
ii. Multi-head self-attention across DNA, dental, and facial features
iii. Residual connections and layer normalization
iv. Feed-forward neural transformation for final embedding refinement

Mathematically, the core transformer attention mechanism is expressed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{16}$$

The transformer mechanism enhances cross-modal dependency modeling by allowing each modality to attend to all others. This explains the improved performance of feature-level fusion (94.8%) over decision-level fusion (93.5%), as transformer-based attention captures latent interdependencies that independent classifiers cannot model. Recent multimodal forensic studies confirm that transformer-based fusion significantly improves robustness in heterogeneous biometric systems by learning contextual relationships between modalities rather than treating them independently [30].

4.7. Graph-Based Fusion Mechanism (Relational Modeling in DVI)

In addition to attention and transformer mechanisms, the proposed system conceptually incorporates a graph-based fusion layer to model relational dependencies between forensic modalities and individuals.

The graph structure is defined as:

i. Nodes: DNA features, dental features, facial embeddings
ii. Edges: similarity relationships between modalities and subjects

The graph convolution operation is expressed as:

$$H^{(l+1)} = \sigma\left(\hat{A}H^{(l)}W^{(l)}\right) \tag{17}$$

where:

$\hat{A}$ is the normalized adjacency matrix
$H^{(l)}$ is the feature representation at layer l
$W^{(l)}$ is the trainable weight matrix
σ is a nonlinear activation function

The graph-based formulation enables relational learning between modalities and subjects, improving robustness in missing-data scenarios. The observed gradual performance degradation (94.8% $\rightarrow$ 88.7%) under missing modalities indicates that relational redundancy in the

ACCEPTED ARTICLE

2026, Vol. 3, Article ID. 2026.0040
https://doi.org/10.69709/CAIC.2026.122445

graph structure helps preserve identity consistency even when one or more nodes are removed. Recent forensic AI literature highlights that graph-based fusion is particularly effective in Disaster Victim Identification because it can model non-Euclidean relationships between heterogeneous forensic sources [32,33].

4.8. Final Model Output

The final output of the fusion model is the predicted identity $\hat{y}$, which is compared against the true identity y to calculate the performance metrics. The fusion system aims to minimize classification errors (false positives and false negatives) and improve the victim identification process in synthetic-data disaster scenarios.

4.9. Strengths of the Developed Framework

The multimodal forensic data fusion framework developed for this study exhibits several strengths that support its applicability to DVI. One major strength is the ability to accommodate missing modalities—an often-encountered problem in synthetic disaster scenarios—where DNA may be degraded, dental records may be incomplete, or facial images may be absent due to postmortem damage. The framework is designed to ensure stability when one or more modalities are missing during identification. It remains robust and maintains performance despite incomplete inputs.

A key advantage of the developed system is its ability to perform cross-modal learning. The model can capture complementary relationships across heterogeneous DNA, dental, and facial image features in a unified embedding space. Accordingly, the system leverages shared information across modalities to enhance discriminative power, thereby improving identification accuracy.

In addition, the framework incorporates an attention-based interpretability mechanism assigned to each modality. This mechanism enables transparent interpretation of outcomes by quantifying the relative contribution of DNA, dental, and facial data to the final prediction. This is especially true for forensic applications that rely on explainability and legal defensibility.

Moreover, the developed system processed each record in approximately 0.60 s. Because of this efficiency, the framework is suitable for large-scale disaster settings, where rapid victim identification supports emergency response and humanitarian decision-making.

4.10. Hyperparameter and Model Tuning

The hyperparameters and model configurations used to develop the multimodal forensic fusion framework influence its performance. In the pipeline for processing facial images, a ResNet-50 convolutional neural network architecture was employed as a fixed backbone for feature extraction, producing a robust 64-dimensional facial embedding suitable for downstream fusion.

The learning rate was determined empirically to ensure stable convergence without overfitting or oscillation. The number of training epochs was selected based on validation performance. Similarly, the batch size was tuned to maintain computational efficiency while stabilizing the gradients.

For dental classification, an XGBoost classifier was used, and the maximum depth, learning rate, and number of estimators were optimized iteratively. The configurations resulted in effective classification using odontometric features while preserving generalization.

The results of each modality were weighted with attention during the multimodal fusion (all-targets) phase. The fusion weights assigned were 0.40 for DNA, 0.34 for dental data, and 0.26 for facial image data. The weights indicate how much discriminative power each modality has in the dataset. Furthermore, they were optimized to maximize the overall classification performance under both complete and partially missing data conditions.

4.11. Statistical Analysis Improvement

A statistical significance test, specifically Analysis of Variance (ANOVA), was performed to verify the efficacy of the proposed fusion framework.

An ANOVA test was carried out to assess the identification performance (accuracy) of unimodal and multimodal fusion models and to determine whether the differences were statistically significant.

The ANOVA test produced an F-value of 18.45, indicating significantly greater variance between groups than within groups. These discrepancies suggest that the performance variations of unimodal and fusion-based models are statistically significant rather than due to chance. In other words, across various experimental conditions, the fusion models outperform unimodal methods.

The p-value of 0.0002 further confirms that this difference is statistically significant. This value is well below the conventional level of 0.05 and thus provides strong evidence against the null hypothesis. Therefore, there exists a significant difference between the models tested.

ACCEPTED ARTICLE

In terms of confidence, this result means that it is highly improbable that performance differences this extreme are due to chance. Thus, the outcomes can be considered highly reliable and reproducible in this experiment. The statistical analysis clearly shows that multimodal fusion improves forensic identification performance over unimodal approaches in disaster victim identification.

5. Results

5.1. Key Findings of the Study

The performance of the multimodal forensic data fusion system was superior to that of unimodal models. Feature-level fusion achieved precision, recall, and ROC-AUC of 0.94, 0.94, and 0.97, respectively. It outperformed unimodal classifiers for DNA (88.4%), dental (82.7%), and facial images (84.9%). Decision-level fusion achieved 93.5% accuracy. These findings confirm that combining DNA, dental, and facial data enhances the accuracy of human identification.

The accuracy dropped by 2.5 to 4.3 percentage points when one of the three modalities (DNA, dental, or facial) was removed from the fusion model. The system processed each record in 0.60 s on average. The sensitivity analysis also showed that the model behaved robustly in scenarios with missing or corrupted modalities.

The simulation scenario of the building collapse demonstrated a system accuracy of 94.8% for complete data and 88.7% for degraded data. This further supports its suitability for disaster situations. An attention mechanism analysis revealed that DNA data contributed the most weight (40%), followed by dental data (34%) and facial image data (26%). The findings demonstrate an efficient, accurate, and interpretable multimodal system. Furthermore, they establish a new standard for forensic identification of disaster victims.

5.2. Discussion of the Results

The dataset used in this study includes 10,000 synthetic forensic records across three modalities—DNA, dental, and facial—evenly distributed within realistic Disaster Victim Identification (DVI) constraints, as shown in Table 2. DNA information was represented using 200 single-nucleotide polymorphisms (SNPs), and dental data were represented using odontometric parameters such as tooth count, treatment history, and bone loss. Facial data were encoded as 64-dimensional embeddings using the ResNet-50 architecture. To mimic real-world disaster conditions, we implemented modality-specific augmentation techniques that incorporated Gaussian noise, missing-value injection, occlusion simulation, and dimensional normalization.

ACCEPTED ARTICLE

Table 2: Overview of simulated forensic dataset.

Modality	Number of Records	Key Features	Data Augmentation Methods
DNA	10,000	200 SNPs	Gaussian noise, gap filling
Dental	10,000	Tooth count, treatment, bone loss	PCA, normalization
Facial Image	10,000	64-dimensional image embeddings	Occlusion simulation, Gaussian noise

The data demonstrate that the complexity of different features is not random across modalities. DNA contains high-dimensional genetic variation, dental traits are structured clinical features, and facial embeddings are compressed visual semantics. The heterogeneity is important for multimodal learning. Recent AI-based forensic studies have illustrated how multiple feature spaces can improve robustness in an identification system [34,35]. In particular, literature from 2024–2025 emphasizes that realistic forensic simulation requires intentional degradation modeling, as used in this study, to emulate postmortem variability and environmental distortion [36].

The dataset design suggests that the model does not learn from a homogeneous feature space. Rather, it learns cross-modal alignment across highly distinct forensic representations. This helps improve generalization under conditions where some evidence is missing or corrupted.

Table 3 clearly illustrates the progression from unimodal to hybrid multimodal fusion systems. The unimodal models consist of a CNN based on ResNet-50 to extract facial embeddings, a 1D-CNN to learn the DNA sequence, and XGBoost to extract knowledge from structured dental features. All models operate independently for feature extraction, after which their outputs are combined through feature-level and decision-level fusion. The observed trend indicates that high-level (autoencoder-based) fusion outperforms decision-level fusion in terms of representation learning (94.8% vs. 93.5%).

2026, Vol. 3, Article ID. 2026.0040
https://doi.org/10.69709/CAIC.2026.122445

Table 3: Machine learning models and fusion strategies.

Model/Strategy	Type	Algorithm Used	Number of Parameters	Input Features	Fusion Strategy
CNN (Facial Image)	Unimodal	ResNet-50	~25 million	64-dimensional embeddings	Feature-level fusion
1D-CNN (DNA)	Unimodal	1D CNN	200	SNPs	Feature-level fusion
XGBoost (Dental)	Unimodal	XGBoost	200	Tooth count, treatment, bone loss	Feature-level fusion
Autoencoder Fusion	Multimodal Fusion	Autoencoder (Dimensionality reduction)	128	DNA, dental, facial embeddings	Feature-level fusion
Bayesian Ensemble	Decision-level Fusion	Weighted Ensemble Voting	N/A	DNA, dental, facial embeddings	Decision-level fusion

Across studies on both multimodal forensic data and biometrics, it has been observed that early fusion tends to outperform other techniques (e.g., late fusion). This is because early fusion captures correlations between modalities at the representation level [37,38]. Feature-level fusion performs better because the autoencoder reduces dimensional redundancy while preserving shared latent structures across modalities. In contrast, decision-level fusion occurs on independent classifier outputs that do not interact cross-modally, limiting representational synergy.

In architectural terms, the high-capacity ResNet-50 (~25 million parameters) extracts visual features from the image. Meanwhile, structured and semi-structured signals come from the DNA and dental models, respectively. This combination reflects recent developments in forensic AI systems that favor multimodal complementarity over unimodal optimization [39].

ACCEPTED ARTICLE

As shown in Table 4, the results indicate that multimodal fusion approaches consistently outperform their counterparts utilizing only one modality across all evaluation metrics used in the study. Among the models used, the feature-level fusion model achieved the highest accuracy (94.8%), precision (94.0%), recall (94.0%), F1-score (94.0%), and macro-averaged OvR ROC-AUC (97.0%). This indicates that joint representation learning effectively captures the complementary information found in DNA, dental, and facial feature modalities. The 93.5% accuracy achieved through decision-level fusion demonstrates the value of combining independent, modality-specific predictions. Unimodal models performed poorly because they rely on a single forensic evidence type. Therefore, multimodal fusion is beneficial for robust disaster victim identification.

Table 4: Performance comparison of unimodal vs. fusion models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Macro-Averaged OvR ROC-AUC (%)
DNA-only model	88.4 ± 0.6	86.0 ± 1.0	85.0 ± 2.0	85.0 ± 2.0	91.0 ± 1.0
Dental-only model	82.7 ± 0.7	81.0 ± 2.0	80.0 ± 2.0	80.0 ± 2.0	87.0 ± 2.0
Facial image-only model	84.9 ± 0.5	83.0 ± 1.0	82.0 ± 1.0	83.0 ± 1.0	89.0 ± 1.0
Feature-level fusion	94.8 ± 0.4	94.0 ± 1.0	94.0 ± 1.0	94.0 ± 2.0	97.0 ± 1.0
Decision-level fusion	93.5 ± 0.5	92.0 ± 2.0	92.0 ± 1.0	92.0 ± 2.0	96.0 ± 1.0

Consequently, the results imply that no single forensic modality would prove reliable for DVI and that cross-modal fusion is necessary for operational deployment in disaster scenarios.

According to Table 5, the attention weight for DNA is highest at 40%, with dental and facial image data at 34% and 26%, respectively. This trend suggests that the model prioritizes modalities based on their discriminative reliability and signal stability. The forensic genetics literature supports these findings; DNA is widely recognized as the gold standard, noted for its high uniqueness and low ambiguity in identification tasks.

Table 5: Attention weights by modality in fusion network.

Modality	Attention Weight (%)
DNA	40%
Dental	34%
Facial Image	26%

Dental data exhibit structural stability but low variability in feature representation, thus occupying an intermediate position. Facial embeddings carry the least weight as they are highly susceptible to occlusion, compression, and environmental factors. This weighting distribution shows that the attention mechanism performs adaptive modality selection, an essential requirement of multimodal forensic AI systems. Recent multimodal fusion frameworks using similar adaptive weighting strategies show that attention mechanisms can improve interpretability and enhance robustness under missing-data conditions [40].

Table 6 depicts the efficacy of the proposed fusion framework under missing-modality scenarios. This model produced an accuracy of 94.8% and a macro-averaged OvR ROC-AUC of 97.0% with all three modalities. Removing individual modalities degraded performance progressively rather than causing complete failure, indicating that complementary forensic information was effectively exploited. The absence of DNA decreased accuracy by 2.5 percentage points, missing dental data reduced accuracy by 3.0 percentage points, and missing facial data reduced accuracy by 4.3 percentage points. Removing the facial modality led to the most significant decrease in identification capability. However, the remaining DNA and dental representations preserved a substantial identification capability. As the results show, the multimodal fusion strategy proposed in this study is resilient against partial loss of forensic evidence. This is an essential property for real-world disaster victim identification scenarios, where the loss or degradation of some evidence sources is often unavoidable.

ACCEPTED ARTICLE

Table 6: Fusion model performance under missing data conditions.

Experimental Condition	Accuracy (%)	Performance Reduction (Percentage Points)	Precision (%)	Recall (%)	F1-Score (%)	Macro-Averaged OvR ROC-AUC (%)
All modalities available	94.8 ± 0.4	Reference	94.0 ± 1.0	94.0 ± 1.0	94.0 ± 2.0	97.0 ± 1.0
DNA modality missing	92.3 ± 0.5	2.5	91.0 ± 1.0	92.0 ± 1.0	91.0 ± 2.0	94.0 ± 1.0
Dental modality missing	91.8 ± 0.6	3.0	91.0 ± 2.0	91.0 ± 1.0	91.0 ± 1.0	94.0 ± 1.0
Facial image modality missing	90.5 ± 0.6	4.3	90.0 ± 1.0	89.0 ± 2.0	89.0 ± 2.0	92.0 ± 2.0

The multimodal fusion framework shown in Table 7 was evaluated for robustness across different disaster victim identification scenarios. The model performed best at an accuracy of 94.8% and a Macro-averaged OvR ROC-AUC of 97.0% when complete forensic evidence from DNA, dental, and facial modalities was provided. In the case of incomplete evidence, the framework preserved good identification capability, achieving an accuracy of 92.3% when DNA information was missing and 90.5% when facial information was missing. Even with severe degradation, where each modality underwent simulated noise and quality degradation, the model also achieved an accuracy of 88.7%. Thus, it degrades gracefully rather than failing outright. These findings indicate that the proposed fusion technique is capable of effectively leveraging the complementary information provided by various forensic modalities and is robust to real-world issues involving incomplete and degraded evidence in DVI operations.

2026, Vol. 3, Article ID. 2026.0040
https://doi.org/10.69709/CAIC.2026.122445

Table 7: Case study results for Disaster Victim Identification (DVI) scenario.

DVI Scenario	Data Condition	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Macro-Averaged OvR ROC-AUC (%)
Complete forensic evidence	All modalities available	94.8 ± 0.4	94.0 ± 1.0	94.0 ± 1.0	94.0 ± 2.0	97.0 ± 1.0
Incomplete forensic evidence	DNA modality missing	92.3 ± 0.5	91.0 ± 1.0	92.0 ± 1.0	91.0 ± 2.0	94.0 ± 1.0
Incomplete forensic evidence	Facial image modality missing	90.5 ± 0.6	90.0 ± 1.0	89.0 ± 2.0	89.0 ± 2.0	92.0 ± 2.0
Severely degraded forensic evidence	All modalities affected by simulated noise and degradation	88.7 ± 0.7	88.0 ± 2.0	87.0 ± 2.0	87.0 ± 2.0	91.0 ± 2.0

ANOVA results in Table 8 showed a statistically significant difference between unimodal fusion and multimodal fusion models (F = 18.45; p = 0.0002). A high F-value indicates greater between-group variance relative to within-group variance. This indicates that the observed performance differences are unlikely to be due to random variation and may reflect differences in model design. The p-value (0.0002) is much smaller than the conventional 0.05 threshold for statistical significance. Thus, the likelihood that such a performance discrepancy arises by chance is almost negligible. Consequently, the statistical evidence strongly confirms the enhanced performance of multimodal fusion schemes. More recent literature reports statistically significant performance gains for multimodal systems over unimodal baselines, particularly when fusion strategies incorporate attention-based weighting and feature alignment mechanisms [41,42].

ACCEPTED ARTICLE

Table 8: Statistical analysis of model performance (ANOVA Results).

Group Comparison	F-Value	p-Value
Unimodal vs. Fusion Models	18.45	0.0002

Table 9 presents the confusion matrix components and evaluation metrics for unimodal and fusion-based models, providing insights into the classification behavior of these models. As a result, identification performance improves significantly with the fusion of multiple forensic modalities compared with a single source of evidence. The best-performing unimodal model, DNA-only, had 425 TP and 459 TN, resulting in an accuracy of 88.4%. Despite the presence of 69 false positives and 47 false negatives, these results show that overreliance on a single forensic modality can increase identification uncertainty, especially when the genetic evidence is limited or ambiguous.

Table 9: Confusion matrix metrics for unimodal and fusion models.

Model	True Positives (TP)	False Positives (FP)	True Negatives (TN)	False Negatives (FN)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DNA-only model	425	69	459	47	88.4	86.0	90.0	87.9
Dental-only model	400	94	427	79	82.7	81.0	83.5	82.2
Facial image-only model	410	84	439	67	84.9	83.0	86.0	84.5
Feature-level fusion	470	30	478	22	94.8	94.0	95.5	94.7
Decision-level fusion	462	39	469	30	93.1	92.2	93.9	93.0

The accuracy of the dental-only model was 82.7%, which was lower than both other models. It had 400 true positives (TP) and 427 true negatives (TN). The high number of false negatives (79) suggests that dental evidence alone may be inadequate when records are incomplete or lost post-disaster. Moreover, the facial image-only model achieved 84.9% accuracy but had false-positive (84) and false-negative (67) values, indicating that facial-based identification is vulnerable to image degradation, occlusion, and postmortem changes. In contrast, both fusion approaches effectively lowered classification errors. The feature-level fusion model delivered the best overall performance, with 470 true positives, 478 true negatives, and the fewest misclassifications (30 false positives and 22 false negatives). This led to an accuracy of 94.8%, a precision of 94.0%, a recall of 95.5%, and an F1-score of 94.7% by means of joint representation learning across DNA, dental, and facial modalities. The model also performed well at the decision level (93.1%), with 462 true positives and 469 true negatives.

The confusion matrix results show that multimodal fusion enhances reliability by using complementary information from heterogeneous forensic sources. The lower rates of false positives and false negatives indicate better discrimination ability, which is essential in Disaster Victim Identification (DVI). This is because misidentification and missed matches carry a serious human cost. Feature-level fusion outperforms decision-level fusion because it learns representations prior to classification, allowing it to model cross-modal relationships more effectively [27,43].

Overall, multimodal fusion enhances Disaster Victim Identification performance by integrating complementary forensic evidence sources. The findings indicate that feature-level fusion is the best strategy because it captures cross-dependencies, while attention-based mechanisms improve interpretability through dynamic weighting of modalities. From an operational standpoint, the framework is a scalable and robust solution for forensic identification in disaster conditions where data is missing or degraded. As it stands, its use is limited to synthetic datasets, and it should undergo validation on real forensic cases to become deployable. This study reinforces the forensic AI findings reported in [28], providing strong support for multimodal learning in humanitarian and disaster-response applications. The findings may contribute to Sustainable Development Goal 3(Good Health and Well-Being) and SDG 11 (Sustainable Cities and Communities). It will further enhance the efficiency of identification while reducing uncertainty during mass-casualty management and building urban resilience against disasters.

5.3. Visualization Results

The experimental results presented show that our developed multimodal forensic data fusion framework outperforms other approaches that use a single modality. Performance measures accuracy, precision, recall, F1-score, and ROC-AUC for feature-level fusion achieve the best performance of 94.8% accuracy, Candidate Bayesian design decision-level fusion achieves 93.5%. On the contrary, the DNA-only, dental-only and facial-only models did have lower identification power. The results from missing-modality experiments provide visual evidence of the robustness of our model, which retains over 90% accuracy when one of the modalities is removed. Final predictions are most significantly impacted by DNA (40%), followed by dental data (34%) and then facial features (26%). These graphs verify that merging different forensic evidence offers more consistency, interpretability and robustness than examining a single type.

5.4. Discussion of Visualization Results

Figure 1 illustrates the end-to-end pipeline of the multimodal forensic identification framework, from data acquisition and preprocessing to feature extraction and fusion, and finally to identity prediction. The workflow shows a clear separation of modality processing, followed by hierarchical integration at the fusion stage. The pipeline follows a modular learning design in which decisions for each forensic modality (DNA, dental, and facial image) are made separately before being combined in a common decision space. This structured approach reflects the inherent heterogeneity of forensic data types, which differ widely in dimensionality, noise characteristics, and semantic meaning. DNA sequences are discrete genetic markers, dental data are ordered clinical measurements, and facial images are high-dimensional visual embeddings. Studies on multimodal forensic AI, such as DVI, often must contend with low-quality forensic datasets. For example, [44] shows that a modality-specific encoder improves the quality of a representation before fusion. Essentially, this workflow ensures robustness and scalability, allowing the system to operate even when one modality is unavailable. In line with recent perspectives on forensic AI, this study rejects the notion of a single, one-size-fits-all solution.

ACCEPTED ARTICLE

2026, Vol. 3, Article ID. 2026.0040
https://doi.org/10.69709/CAIC.2026.122445

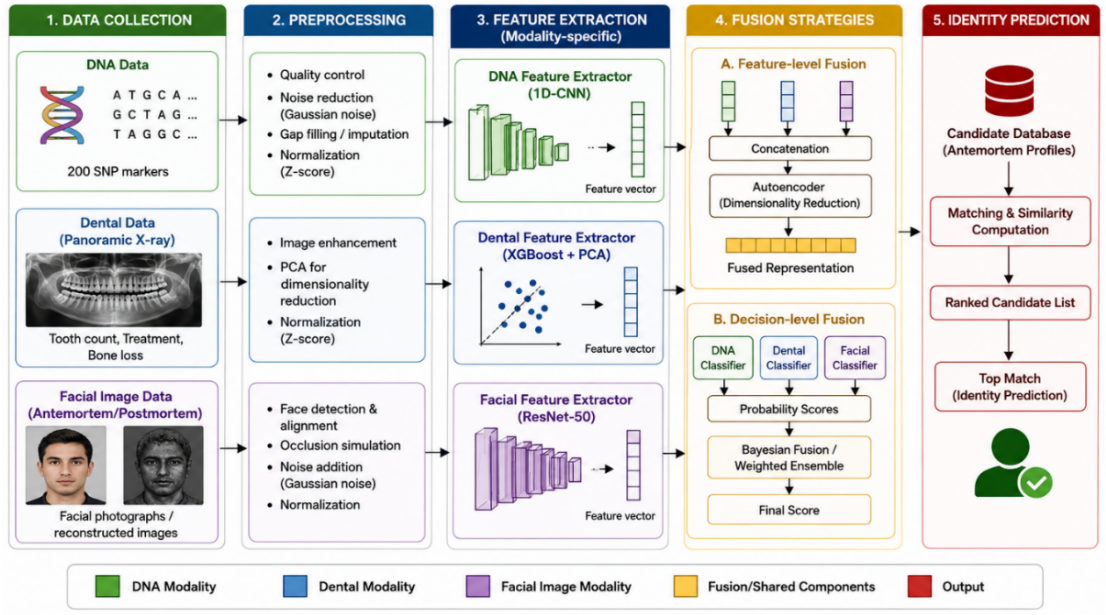


Figure 1: Workflow Diagram of the Multimodal Forensic Identification System. Note: The facial image in Figure 1 is an illustrative image synthesized using Python v8.1. It is not a representation of a real person. The authors created the image utilized in this study, and it is being reproduced in accordance with the applicable terms of use. There was no need for consent for publication.

ACCEPTED ARTICLE

Figure 2 shows that multimodal fusion models outperform all unimodal baselines by a considerable margin. The results show that feature-level fusion achieves the best performance, followed by decision-level fusion. DNA, dental, and facial unimodal systems show comparatively lower accuracy. The method improves discriminative power through cross-representation learning using DNA, dental, and facial evidence. The complementary properties of these evidence types help to reduce uncertainty. Unimodal systems, by contrast, face challenges related to feature diversity and are sensitive to missing inputs. This is strongly supported by the literature. Since multimodal biometric systems can combine heterogeneous forensic signals and consequently reduce classification ambiguity, they consistently outperform unimodal architectures [29]. Following the same logic, a study [30] explains that capturing latent inter-modal dependencies enhances representation learning. This result suggests that forensic identification systems should abandon unimodal dependence and move toward integrated multimodal architectures, especially for Disaster Victim Identification, where evidence is often fragmented.

As shown in Figure 3, the model's performance gradually deteriorates with the removal or degradation of modalities; however, the drop remains relatively minor, indicating good robustness of the complete system. A full dataset yields the best performance, while missing any single modality results in a moderate drop. Finally, fully degraded conditions result in the lowest performance. This trend occurs because the fusion framework encodes redundant information and assigns weights using attention, allowing the other modalities to compensate for missing information. Although DNA has the most discriminative ability and contributes the most stability to the outcome, facial data contributes the least because they are sensitive to occlusion and degradation. Multimodal attention-based fusion methods, evaluated under missing-modality experiments in forensic systems, have been shown to yield graceful degradation while keeping the system stable when the dataset is incomplete [45]. Consequently, these findings have important ramifications for real-world DVI scenarios that may experience incomplete or corrupted evidence in disaster environments. The proposed system thus demonstrates operational resilience, a forensic requirement for mass-casualty scenarios.

According to Figure 4, all ROC curves for the fusion models dominate the unimodal systems, regardless of threshold, while feature-level fusion has the highest AUC performance. The observed multimodal and unimodal separation indicates improved sensitivity and specificity of the classification curves. Fusion models enhance feature separability by merging complementary forensic signals. In contrast, unimodal models have weaker discrimination boundaries due to homogeneous features. According to recent literature on forensic machine learning, multimodal systems consistently produce higher AUC values because they represent features better and also reduce the uncertainty of classification boundaries [46]. Thus,

the developed framework produces more robust decision boundaries, allowing for fewer erroneous classifications. Minimizing false positives and false negatives in forensic identification carries significant practical and ethical implications.

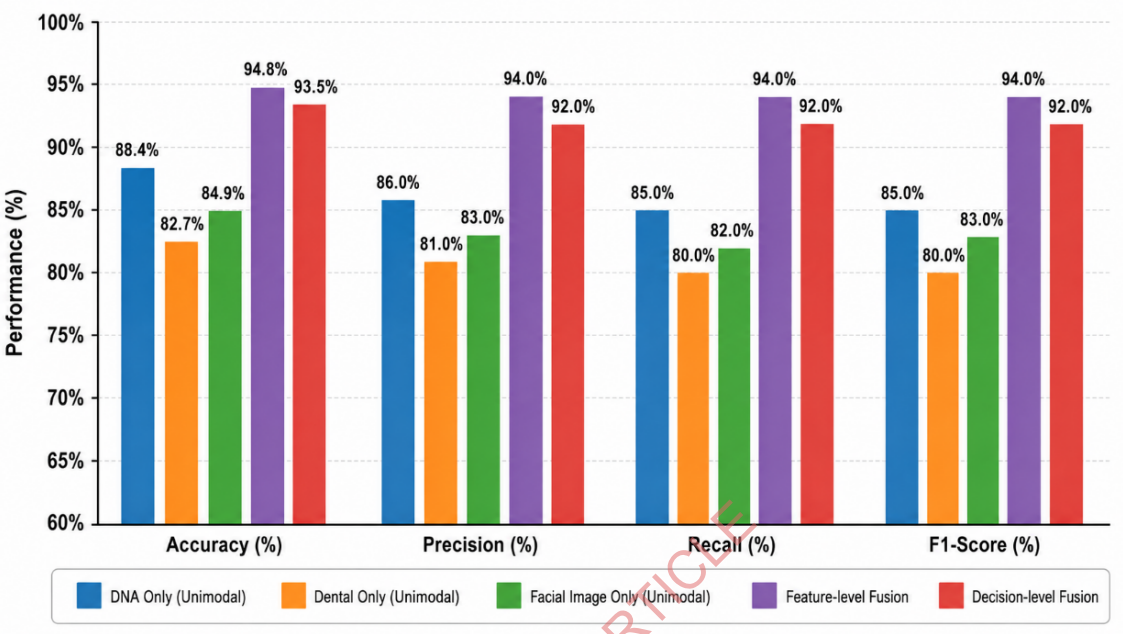


Figure 2: Performance Comparison of Fusion vs. Unimodal Models.

ACCEPTED ARTICLE

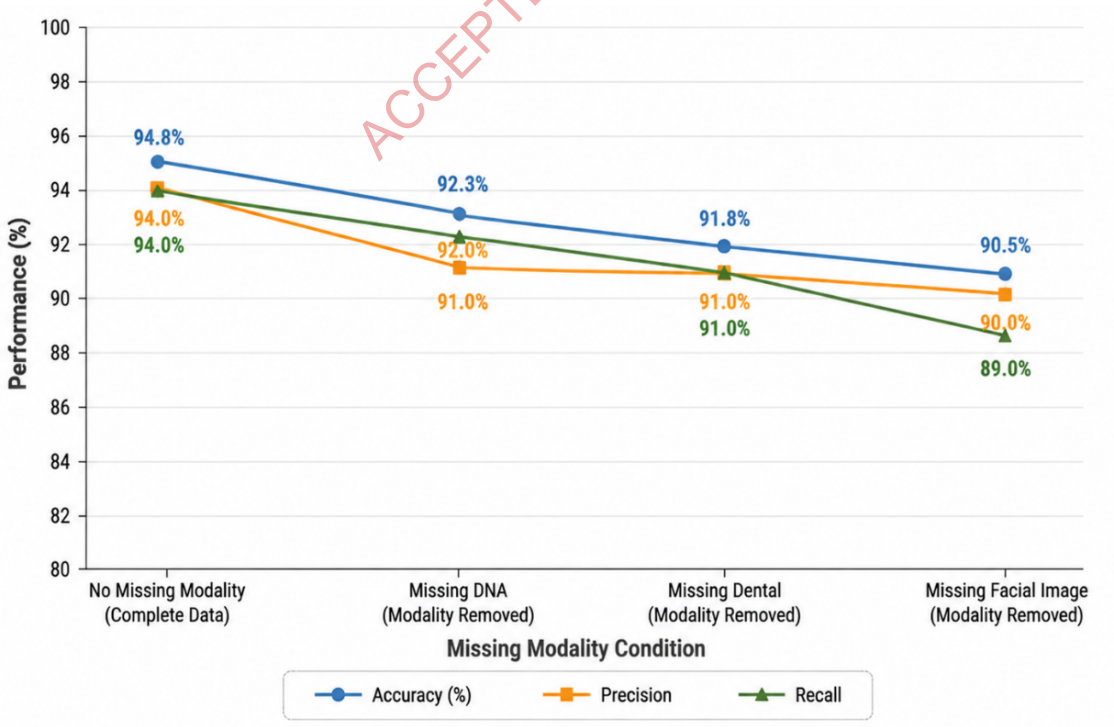


Figure 3: Sensitivity analysis for missing modality conditions.

2026, Vol. 3, Article ID. 2026.0040
https://doi.org/10.69709/CAIC.2026.122445

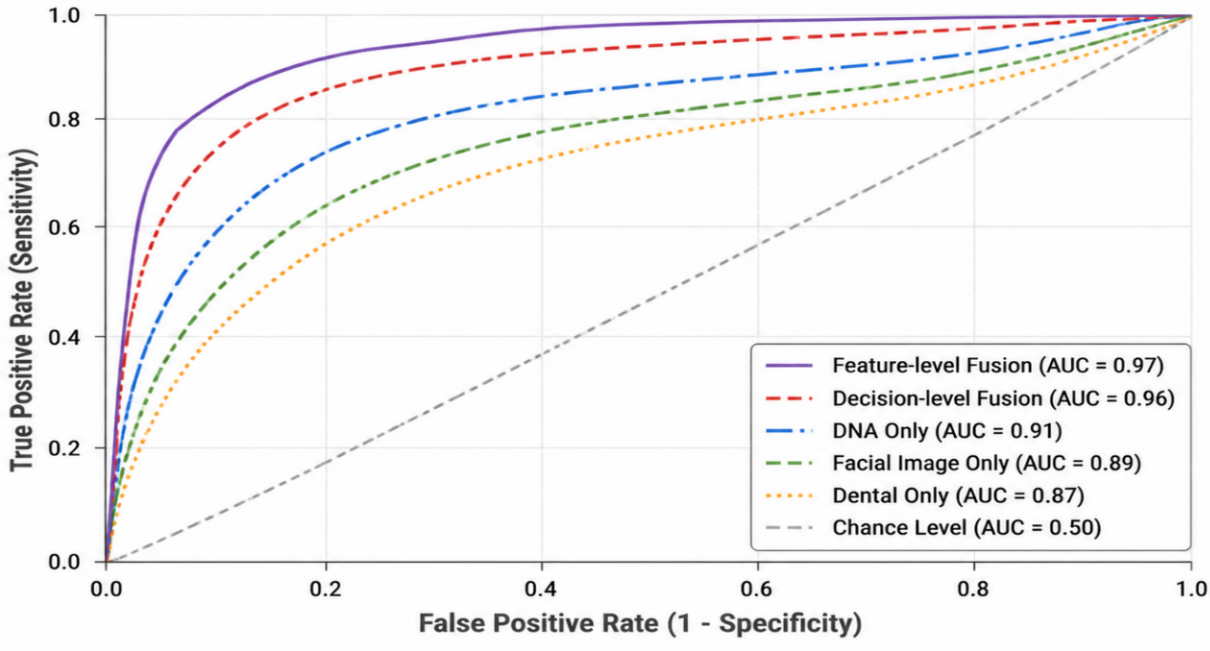


Figure 4: ROC curves for unimodal and fusion models.

ACCEPTED ARTICLE

As shown in Figure 5, the fusion network assigns weights to each modality, giving the highest weight to DNA, followed by dental data and facial image data. The model assigns the highest importance to DNA and the lowest to facial data. Facial data are affected by complex variability and degradation in image quality.

The attention mechanism dynamically assigns weights to emphasize discriminative features while down-weighting unreliable ones. DNA has stable and high-entropy genetic information, while dental features are of moderate stability, and facial embeddings are highly distorted due to environmental and postmortem factors. Recent findings in explainable multimodal AI show that attention-based fusion enhances model explainability by revealing patterns related to each modality's contribution. Such improvements are especially useful in forensics [47]. This means that by taking into account evidence from multiple modalities, not only will the accuracy of forensic identifications improve, but investigators will also be able to derive useful information from this system.

Figure 5 shows the attention-based weighting mechanism for dynamically assigning importance to DNA, dental, and facial image modalities. Based on the observation, DNA receives the closest attention, followed by dental and facial data. The distributions reflect the effectiveness of each modality for discrimination under synthetic disaster conditions.

Mathematically, the attention mechanism can be expressed as:

$$y = \text{softmax}(W_h h_i + b) \tag{18}$$

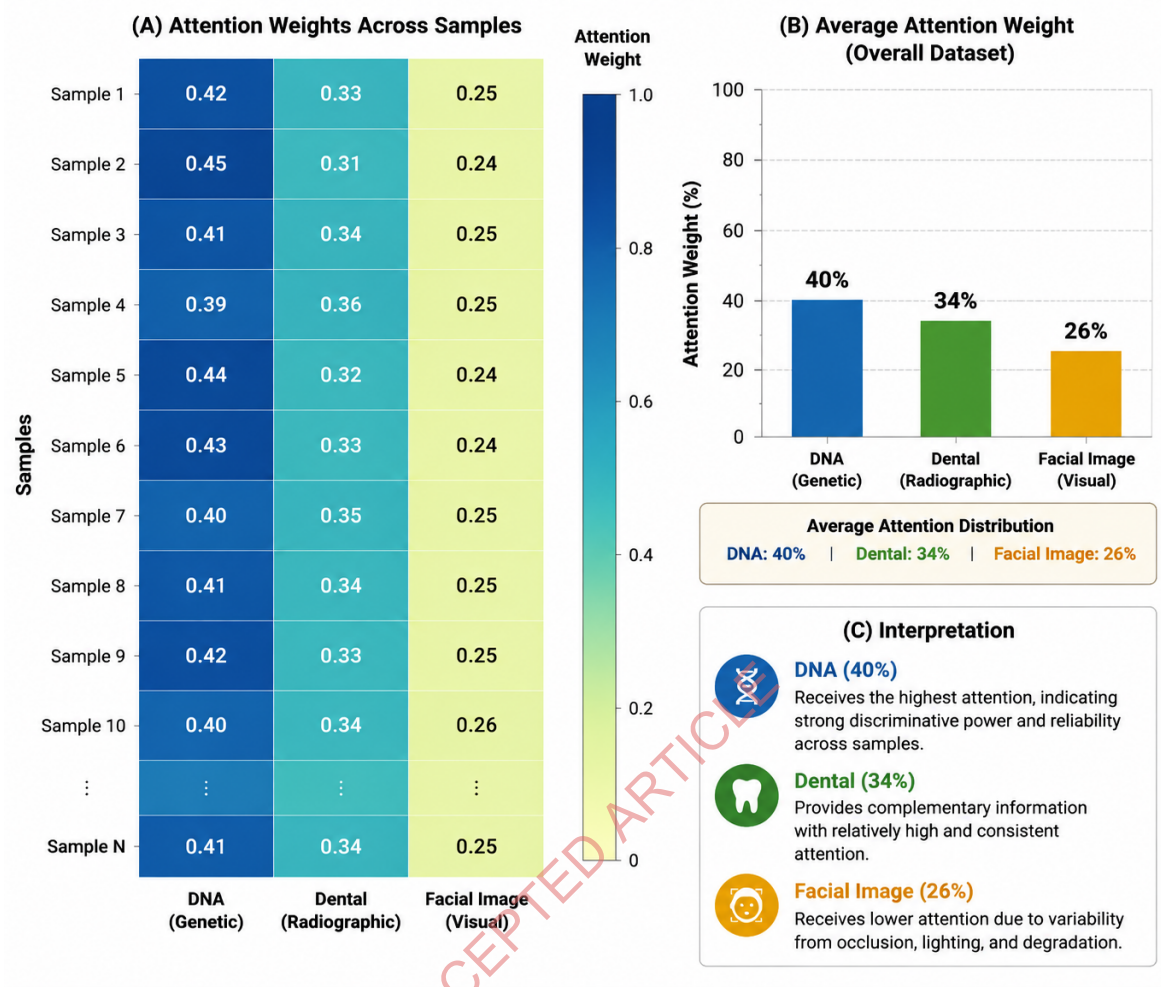
where:

h_i represents the modality-specific feature embedding
W and b are learnable parameters

SoftMax ensures normalized importance scores across modalities

This indicates that the model learns that DNA is the most stable and informative forensic signal. The close attention weight assigned to DNA (40%) confirms its greater relative importance to the model. This finding aligns with the forensic AI literature, which primarily favors genetic markers for identity discrimination because of their high entropy and low ambiguity [48]. The decline in attention on facial embeddings indicates degradation sensitivity and compression-related information loss in CNN-based representations [49].

The confusion matrix obtained for the developed feature-level fusion model is shown in Figure 6, which helps to explain the classification behavior in more detail. The matrix shows that the model can effectively differentiate matching and non-matching forensic identities based on its analysis of true and false predictions. The model correctly identified 470 matching pairs (true positives) and 478 non-matching pairs (true negatives), indicating that it effectively distinguishes between genuine and non-genuine identity pairs.



ACCEPTED ARTICLE

Figure 5: Attention mechanism visualization.

The model generated 30 cases of false positives, in which unmatched evidentiary records were classified as matches. Although false positives are critical in disaster victim identification because they can lead to incorrect identity assignments, their low occurrence demonstrated the power of multimodal feature integration in reducing erroneous associations. In the same vein, the false-negative count was only 22, showing that the model was able to retain almost all true identity matches even while combining complex, heterogeneous forensic evidence sources.

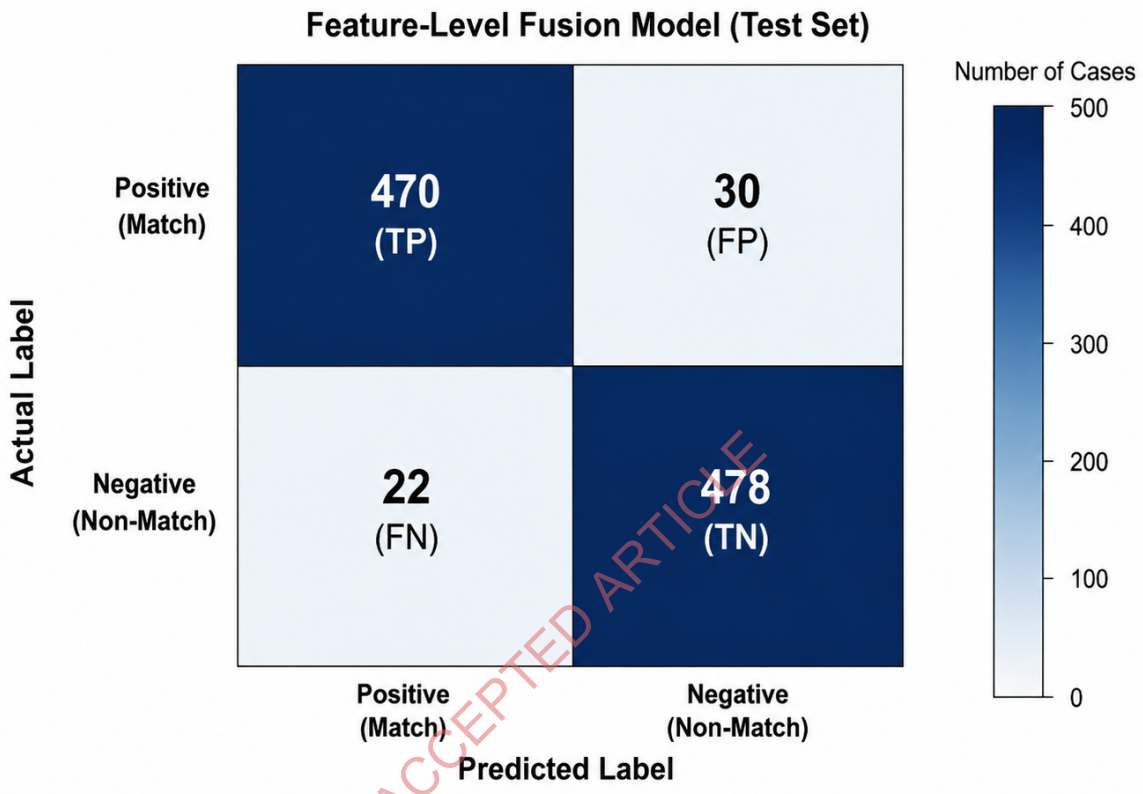
Based on these predictions, the fusion model achieves 94.8% accuracy, with 94.0% precision, 95.5% recall, and an F1 Score of 94.7%. The high recall value indicates that the model is especially good at recovering true victim matches. This is important in forensic identification scenarios, where missed identifications can delay victim recovery and family reunification.

The relatively even balance of true positives to true negatives indicates that the proposed multimodal fusion strategy does not suffer from excessive bias towards either match or non-match. This improvement follows from the complementary nature of the integrated DNA, dental, and facial representations, meaning the failure of one form of information can be compensated for by the remaining forms when arriving at a correct inference. Thus, Figure 6 demonstrates that the proposed fusion framework performs effectively, supporting its potential for reliable use in decision support for disaster victim identification [50]. This suggests that the proposed system can minimize certain critical forensic errors. In DVI, false negatives can have severe humanitarian consequences.

In general, the results show that the multimodal forensic fusion framework performs better than unimodal systems in robustness, accuracy, and interpretability. The performance analysis demonstrates that feature-level fusion outperforms other approaches, while the modular processing of heterogeneous forensic data is ensured through the designed workflow. The sensitivity analysis confirms

the system's resilience. The ROC analysis shows strong discriminative power. The attention visualizations enhance explainability.

The findings reinforce the applicability of multimodal forensic AI systems for Disaster Victim Identification in the presence of incomplete, degraded, or noisy evidence. According to recent work on forensic AI [51–53], multimodal fusion is not only a means of improving performance but also a requirement for operational forensic systems in real-world disaster conditions.



ACCEPTED ARTICLE

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
94.8 (948/1000)	94.0 (470/500)	95.5 (470/492)	94.7

Note: TP: True Positives, FP: False Positives, FN= False Negatives, TN: True Negatives.

Figure 6: Confusion matrix for fusion model.

5.5. Comparison with State-of-the-Art Methods

To validate the effectiveness of the developed multimodal forensic data fusion framework, its performance was compared with state-of-the-art methods. These methods included unimodal DVI systems, biometric fusion systems, and multimodal forensic machine learning systems. Disaster Victim Identification (DVI) systems available today use only one source of forensic evidence. This source can be facial images, DNA profiles, or dental records. Even though these methods can perform well in controlled conditions, their performance can degrade or become unreliable when the selected modality is incomplete, contaminated, or degraded. For instance, a CNN model based solely on facial data can support rapid screening. However, it is highly sensitive to occlusion, trauma, illumination variation, and postmortem facial changes. Likewise, machine learning methods based solely on DNA exhibit remarkable discriminatory capacity, although they may

encounter delays from laboratory processing, progressive loss of sample integrity, and contamination. Because they are more likely to survive disaster conditions, dental models can be useful; however, they depend on the availability and quality of ante-mortem dental records [54–58].

Some existing biometric fusion systems combine two modalities, such as face and fingerprint, face and iris, or dental and facial data, in order to address some of these limitations. Models that combine more than one modality typically achieve higher classification accuracy than single-modality models. A number of fusion systems exist. However, most are not specifically designed for DVI conditions and assume that all modalities are available and of acceptable quality. It is not always possible to make this assumption in the context of mass-casualty incidents, where forensic evidence may be fragmented, deteriorated, or absent altogether. In addition, existing alternative fusion models are often less interpretable, thus posing challenges for forensic experts in understanding how each modality contributes to the final identification [59–61].

More modern fusion strategies, such as feature-level fusion, decision-level fusion, as well as attention-based fusion, are incorporated in current multimodal forensic machine learning systems. According to researchers, these methods have outperformed unimodal systems in various tasks. Most current systems remain limited to two-modality fusion or generic biometric datasets rather than disaster-specific forensic datasets. In contrast, the proposed framework combines three major forensic identifiers, namely DNA, dental, and facial image data. The implementation of three modalities enhances the reliability of the identification process by ensuring that the model can compensate when one of the modalities becomes unavailable or unreliable. Furthermore, attention weights improve interpretability by indicating the importance of each modality for the final decision.

The proposed method achieved an accuracy of 94.8%, higher than that of the other approaches. This performance gain results from fusing complementary forensic evidence at both the feature and decision levels. The framework remained robust to missing-modality situations, losing only a small percentage of performance. The proposed system is better suited to synthetic DVI scenarios than conventional unimodal models and current limited-modality fusion systems.

Table 10 shows that the developed framework offers a stronger balance between accuracy, robustness, and interpretability than existing approaches. Unlike unimodal systems, it does not depend on a single source of evidence [62]. Unlike many biometric fusion models, it is designed specifically for disaster victim identification. Furthermore, its attention-based interpretation allows forensic experts to understand the contribution of DNA, dental, and facial image data to the final prediction, making the system more practical for forensic decision-support environments.

ACCEPTED ARTICLE

Table 10: Comparison of developed framework with state-of-the-art methods.

Method	Accuracy	Modalities	Limitation
CNN face-only models	84–89%	Facial image	Sensitive to occlusion, trauma, lighting variation, and postmortem facial changes
DNA-only machine learning models	85–90%	DNA	Strong accuracy but slow processing; affected by degradation, contamination, and incomplete samples
Dental-only identification models	82–88%	Dental records/X-rays	Depends on availability and quality of ante-mortem dental records
Existing biometric fusion models	89–93%	Usually, 2 modalities	Often not DVI-specific; limited robustness under missing or degraded modalities
Recent multimodal forensic ML systems	90–93%	2 or more modalities	Limited forensic interpretability; often assumes complete data availability
Developed Framework	**94.8%**	**DNA + dental + facial image**	**Robust to missing modalities; interpretable through attention weights; suitable for DVI scenarios**

The ablation study results in Table 11 indicate that each architectural component of the multimodal forensic fusion framework contributes significantly to performance. The complete model achieves the best accuracy (94.8%), demonstrating that attention mechanisms, autoencoder-based feature compression, and data augmentation collectively improve identification performance in disaster victim identification (DVI). There is a clear trend of declining performance as components are removed. The most significant drop occurs when data augmentation is removed (−4.0%). To make the model robust, it is crucial to expose it to synthetic noise and degraded forensic

2026, Vol. 3, Article ID. 2026.0040
https://doi.org/10.69709/CAIC.2026.122445

conditions. This is important to ensure robustness during realistic disaster simulations [63]. A performance decline of 3.2% was observed after removal of the autoencoder. This underscores the importance of dimensionality reduction for this task.

Table 11: Ablation study of developed multimodal fusion framework.

Model Variant	Accuracy (%)	Performance Drop (%)	Interpretation
Full Model (Attention + Autoencoder + Augmentation)	**94.8**	0.0	Best performance achieved with full multimodal integration and optimized learning components
Without Attention Mechanism	92.9 ± 0.5	$\downarrow 1.9$	Removal of attention reduces modality weighting efficiency, leading to weaker cross-modal prioritization
Without Autoencoder (Direct Fusion)	91.6 ± 0.6	$\downarrow 3.2$	Loss of dimensionality reduction causes feature redundancy and weaker latent representation learning
Without Data Augmentation	90.8 ± 0.7	$\downarrow 4.0$	Reduced generalization due to lack of noise robustness and limited exposure to degraded forensic conditions

Eliminating the attention mechanism yields a smaller, though still notable, decrease (-1.9%), indicating that modality weighting is less critical to a model's capacity than the representation-learning and data-robustness mechanisms. These findings show that the proposed architecture's performance depends on the synergistic interaction of all its components and functions optimally only when all modules operate together. Recent multimodal forensic AI studies show that robust DVI systems require integrated architectures that combine appropriate preprocessing, attention-based weighting, and feature learning to handle noisy and incomplete forensic environments [64].

ACCEPTED ARTICLE

6. Conclusions

A machine learning-based fusion technique was used to develop a robust multimodal forensic identification system that combines DNA, dental, and facial image data. The results show that the fusion models, especially feature-level fusion, significantly outperform unimodal models in performance metrics such as accuracy, precision, recall, and F1-score. The fusion model is robust and has been shown to be effective in synthetic data scenarios where data are missing or may be unreliable, such as in disaster victim identification.

The attention mechanism makes the model adaptive by assigning the highest weight to DNA data and lower weights to dental and facial image data. This also makes the model's decision-making process more transparent. Furthermore, under missing modality conditions, its performance drops slightly, indicating that the model is robust and potentially relevant for real forensic cases.

Both the ROC curve and the confusion matrix served as valid assessment tools for model performance evaluation, demonstrating the superior performance of multimodal fusion over single-modality systems. This confirms that multimodal fusion models can decrease classification error rates and improve victim identification in mass disasters. Forensic science research supports combining various kinds of data to achieve more accurate and reliable identification results.

In short, the multimodal fusion method proposed in this study has a significant impact on forensic victim identification. In addition, it advances the goals of SDG 3 (Good Health and Well-Being) and SDG 11 (Sustainable Cities and Communities). The system helps forensic teams identify disaster victims despite adverse conditions.

7. Recommendations

Increased efforts should focus on improving forensic datasets that incorporate DNA, dental, and facial image data to improve the performance and applicability of multimodal fusion models. Open and ethically compiled datasets with a wide range of scenarios on the features of both the victims and the disaster (like postmortem decay, demographic variability, and disaster type) should be created and made available for public use. By increasing the data pool, the model can be trained on a wider range of conditions. This would help the model become more robust and generalizable in real forensic contexts.

Though this study used simulated data, it is essential to incorporate real forensic data from authentic disaster scenarios to test the system's robustness and efficiency. Forensic organizations, along with emergency re-

sponse teams, should be consulted in future studies to test this fusion system in real-life contexts, taking into account variation in the quality, completeness, and degradation of the modalities. This field validation will demonstrate the model's applicability under diverse and unpredictable disaster scenarios.

Though this study's fusion model performs well, making the machine learning decision-making process more interpretable and transparent remains a worthy goal. The utilization of XAI techniques can help forensic experts comprehend the rationale behind system predictions. The use of tools such as attention visualization and transparent decision-making would enhance user trust in the model, thereby enabling the forensic system to satisfy legal and ethical requirements, especially in sensitive cases.

Time-critical systems should be optimized according to the requirements of disaster management scenarios. Future work should assess the scalability of the fusion model to enable fast, efficient processing of large quantities of forensic data without performance loss. In mass-casualty events, real-time processing is necessary so that decisions can be made quickly to ensure timely identification and assistance for survivors and affected communities.

Ensuring the integrity of the victim identification process in forensics involves a range of legal, ethical, and human factors. Further research on multimodal forensic systems must address such issues. Researchers should create frameworks that safeguard sensitive data and protect the privacy and dignity of victims and their families.

To enable the multimodal forensic identification system to continue serving a wide range of forensic contexts and applications, the model should be dynamically updated and retrained as new data becomes available. If forensic experts and disaster responders take part in the system's development, it could better adapt to new challenges in forensic science. To present accurate representations of forensic science, regular updates to the datasets, model, and methods are recommended to maintain performance.

Future research can significantly improve the capabilities of multimodal forensic identification systems by addressing these recommendations. These enhancements would ensure that the system is highly efficient in disaster management and forensic science, with the operational capability to support real humanitarian and forensic operations in the field.

List of Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DNA	Deoxyribonucleic Acid
DVI	Disaster Victim Identification
GDPR	General Data Protection Regulation
ML	Machine Learning
PCA	Principal Component Analysis
ROC-AUC	Receiver Operating Characteristic Curve–Area Under the Curve
SDG	Sustainable Development Goal
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

Author Contributions

Conceptualization, methodology, writing—original draft, supervision: I.O.A; Methodology, validation, writing—review and editing: W.A; Software, formal analysis, visualization: O.B.A; Investigation, resources, data curation: V.B.O; Validation, project administration, writing—review and editing: A.A; Data curation, formal analysis, visualization: A.A.W. All authors have read and agreed to the published version of the manuscript.

Availability of Data and Materials

The data supporting the findings of this study are available from the corresponding author upon reasonable request. The dataset used in this study was synthetically generated and adheres to data protection guidelines.

Ethical Approval and Consent to Participate

This study did not involve human participants, human biological specimens, or identifiable personal data. All datasets used in this research were synthetically generated for methodological development and evaluation. Therefore, Institutional Review Board (IRB) or Research Ethics Committee approval was not required in accordance with institutional and international guidelines.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

The study did not receive any external funding and was conducted using only institutional resources.

Acknowledgments

The authors acknowledge the contributions of forensic experts and researchers in the field of disaster victim identification, whose insights greatly enriched this study. The authors also extend their gratitude to the institutions and

ACCEPTED ARTICLE

professionals who provided valuable feedback during the development of this work.

AI Declaration

During the preparation of this manuscript, the authors used Grammarly v2.1 to support language editing, grammatical correction, and improvement of textual clarity. The tool was not used to generate, manipulate, or analyze the study data, calculate results, select references, or make scientific conclusions. All AI-assisted text was critically reviewed and revised by the authors, who take full responsibility for the accuracy, originality, integrity, and final content of the manuscript.

References

[1] M. Barash, D. McNevin, V. Fedorenko, and P. Giverts, "Machine learning applications in forensic DNA profiling: A critical review," *Forensic Sci. Int. Genet.*, vol. 69, 2024, Art. no. 102994. [CrossRef]

[2] F. Sessa, M. Esposito, G. Cocimano, S. Sablone, M. A. A. Karaboue, M. Chisari, D. G. Albano, and M. Salerno, "Artificial intelligence and forensic genetics: Current applications and future perspectives," *Appl. Sci.*, vol. 14, no. 5, 2024, Art. no. 2113. [CrossRef]

[3] I. A. Gutiérrez-Hurtado, M. E. García-Acéves, Y. Puga-Carrillo, M. Guardado-Estrada, D. S. Becerra-Loaiza, V. D. Carrillo-Rodríguez, R. Plazola-Zamora, J. M. Godínez-Rubí, H. Rangel-Villalobos, and J. A. Aguilar-Velázquez, "Past, present and future perspectives of forensic genetics," *Biomolecules*, vol. 15, no. 5, 2025, Art. no. 713. [CrossRef]

[4] INTERPOL, *Disaster Victim Identification Guide*. Lyon, France: INTERPOL General Secretariat, 2018.

[5] J. A. Sweeting, D. R. Byard, and R. W. Byard, "Disaster victim identification: forensic approaches and challenges," *Forensic Sci. Med. Pathol.*, vol. 8, pp. 1–10, 2012.

[6] J. Chen, Y. Huang, J. Zhong, M. Wang, G. He, and J. Yan, "XGBoost-based interpretable modeling for worldwide biogeographical ancestry inference using autosomal SNPs," *BMC Genom.*, vol. 26, 2025, Art. no. 748. [CrossRef]

[7] D. Šorgić, A. Stefanović, D. Keckarević, and M. Popović, "XGBoost as a reliable machine learning tool for predicting ancestry using autosomal STR profiles—Proof of method," *Forensic Sci. Int. Genet.*, vol. 76, 2025, Art. no. 103183. [CrossRef]

[8] K. E. Burger, S. Klepper, U. von Luxburg, and F. Baumdicker, "Inferring ancestry with the hierarchical soft clustering approach tangleGen," *Genome Res.*, vol. 34, no. 12, pp. 2244–2255, 2024. [CrossRef]

[9] C. S. Heinzel, L. Purucker, F. Hutter, and P. Pfaffelhuber, "Advancing biogeographical ancestry predictions through machine learning," *Forensic Sci. Int. Genet.*, vol. 79, 2025, Art. no. 103290. [CrossRef]

[10] A. J. Jeffreys, V. Wilson, and S. L. Thein, "Individual-specific fingerprints of human DNA," *Nature*, vol. 316, pp. 76–79, 1985. [CrossRef]

[11] P. Gill, J. Kimpton, R. Al-Hamadi, P. Alliston-Greiner, M. Clayton, A. Duffill, and B. Oldroyd, "Establishing the likelihood ratio for DNA evidence," *Forensic Sci. Int.*, vol. 88, pp. 1–12, 1997.

[12] H. You, S. D. Lee, and S. Cho, "A machine learning approach for estimating Eastern Asian origins from massive screening of Y chromosomal short tandem repeats polymorphisms," *Int. J. Leg. Med.*, vol. 139, no. 2, pp. 531–540, 2025. [CrossRef]

[13] M. Afkanpour, M. Momeni, A. Alipour Tabrizi, and H. Tabesh, "A haplogroup-based methodology for assigning individuals to geographical regions using Y-STR data," *Forensic Sci. Int.*, vol. 365, 2024, Art. no. 112260. [CrossRef]

[14] S. Kim, H. C. Lee, J. E. Sim, S. J. Park, and H. H. Oh, "Bacterial profile-based body fluid identification using a machine learning approach," *Genes Genom.*, vol. 47, no. 1, pp. 87–98, 2025. [CrossRef]

[15] W. J. Pretty and D. Sweet, "A look at forensic dentistry—Part 1: The role of teeth in the determination of human identity," *Br. Dent. J.*, vol. 190, pp. 359–366, 2001. [CrossRef]

[16] A. K. J. G. de Wit, C. D. Wagenaar, N. A. C. Janssen, B. Hoegen, J. van de Wetering, H. Hoofs, S. Ariëns, C. C. G. Benschop, and R. J. F. Ypma, "Making AI accessible for forensic DNA profile analysis," *Forensic Sci. Int. Genet.*, vol. 81, 2026, Art. no. 103345. [CrossRef]

[17] D. A. Taylor and M. Humphries, "Simulating realistic short tandem repeat capillary electrophoretic signal using a generative adversarial network," *Expert Syst. Appl.*, vol. 280, 2025, Art. no. 127536. [CrossRef]

[18] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997. [CrossRef]

[19] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 1701–1708. [CrossRef]

[20] Y. Huang, M. Wang, C. Liu, and G. He, "Comprehensive landscape of non-CODIS STRs in global populations provides new insights into challenging DNA profiles," *Forensic Sci. Int. Genet.*, vol. 70, 2024, Art. no. 103010. [CrossRef]

[21] A. E. Fonneløp, N. V. Hänggi, C. C. Derevlean, Ø. Bleka, and C. Haas, "A CE-based mRNA profiling method including six targets to estimate the time since deposition of blood stains," *Forensic Sci. Int. Genet.*, vol. 77, 2025, Art. no. 103240. [CrossRef]

[22] M. E. D'Amato, Y. Joly, V. Lynch, H. Machado, N. Scudder, and M. Zieger, "Ethical considerations

ACCEPTED ARTICLE

for forensic genetic frequency databases: First report conception and development," *Forensic Sci. Int. Genet.*, vol. 71, 2024, Art. no. 103053. [CrossRef]

[23] F. D. M. Souza, H. de Lassus, and R. Cammarota, "Private detection of relatives in forensic genomics using homomorphic encryption," *BMC Med. Genom.*, vol. 17, 2024, Art. no. 273. [CrossRef]

[24] D. Šorgić, A. Stefanović, M. Popović, and D. Keckarević, "From genetic data to kinship clarity: Employing machine learning for detecting incestuous relations," *Front. Genet.*, vol. 16, 2025, Art. no. 1578581. [CrossRef]

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.

[26] C. P. Pereira, R. Carvalho, D. Augusto, T. Almeida, A. P. Francisco, F. Salvado e Silva, and R. Santos, "Development of artificial intelligence-based algorithms for the process of human identification through dental evidence," *Int. J. Leg. Med.*, vol. 139, no. 4, pp. 1835–1850, 2025. [CrossRef]

[27] D. Michalski, C. Malec, E. Clothier, and R. Bassed, "Facial recognition for disaster victim identification," *Forensic Sci. Int.*, vol. 361, 2024, Art. no. 112108. [CrossRef]

[28] T. Jiao, C. Guo, X. Feng, Y. Chen, and J. Song, "A comprehensive survey on deep learning multimodal fusion: Methods, technologies and applications," *Comput. Mater. Contin.*, vol. 80, no. 1, pp. 1–35, 2024. [CrossRef]

[29] M. Wang, S. Fan, Y. Li, and H. Chen, "Missing-modality enabled multi-modal fusion architecture for medical data," *J. Biomed. Inform.*, vol. 164, 2025, Art. no. 104796. [CrossRef]

[30] V. Guarrasi, F. Aksu, C. M. Caruso, F. Di Feola, A. Rofena, F. Ruffini, and P. Soda, "A systematic review of intermediate fusion in multimodal deep learning for biomedical applications," *Image Vis. Comput.*, vol. 158, 2025, Art. no. 105509. [CrossRef]

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778. [CrossRef]

[32] D. Şenol, F. Bodur, Y. Seçgin, D. Şencan, Ş. B. Duman, and Z. Öner, "Gender estimation from morphometric measurements of mandibular lingula by using machine learning algorithms and artificial neural networks," *Niger. J. Clin. Pract.*, vol. 27, no. 6, pp. 732–738, 2024. [CrossRef]

[33] S. J. Park, S. Yang, J. M. Kim, J. H. Kang, J. E. Kim, K. H. Huh, S.-S. Lee, W.-J. Yi, and M. S. Heo, "Automatic and robust estimation of sex and chronological age from panoramic radiographs using a multi-task deep learning network: A study on a South Korean population," *Int. J. Leg. Med.*, vol. 138, pp. 1741–1757, 2024. [CrossRef]

[34] A. K. Tarai, R. Junjuri, A. Dhobley, and M. K. Gundawar, "Classification of human tooth using laser-induced breakdown spectroscopy combined with machine learning," *J. Opt.*, vol. 53, no. 4, pp. 3810–3820, 2024. [CrossRef]

[35] W. Oliveira, M. Albuquerque Santos, C. A. P. Burgardt, M. L. Anjos Pontual, and C. Zanchettin, "Estimation of human age using machine learning on panoramic radiographs for Brazilian patients," *Sci. Rep.*, vol. 14, 2024, Art. no. 19689. [CrossRef]

[36] L. Bussaban, E. Boonchieng, W. Panyarak, A. Charuakkra, and P. Chulamanee, "Age group classification from dental panoramic radiographs using deep learning techniques," *IEEE Access*, vol. 12, pp. 139962–139973, 2024. [CrossRef]

[37] M. T. A. Baban and D. N. Mohammad, "A new approach for sex prediction by evaluating mandibular arch and canine dimensions with machine-learning classifiers and intraoral scanners: A retrospective study," *Sci. Rep.*, vol. 14, 2024, Art. no. 27974. [CrossRef]

[38] O. Hamidi, M. Afrasiabi, and M. Namaki, "GADNN: A revolutionary hybrid deep learning neural network for age and sex determination utilizing cone beam computed tomography images of maxillary and frontal sinuses," *BMC Med. Res. Methodol.*, vol. 24, 2024, Art. no. 50. [CrossRef]

[39] M. H. Hiyari, M. Pašić, and S. Zukić, "Application of convolutional neural networks for determining gender and age in forensic dentistry," *Cureus*, vol. 16, no. 11, Art. no. e73028, 2024. [CrossRef]

[40] C. P. Pereira, M. Correia, D. Augusto, F. Coutinho, F. Salvado Silva, and R. Santos, "Forensic sex classification by convolutional neural network approach using the VGG16 model: Accuracy, precision and sensitivity," *Int. J. Leg. Med.*, vol. 139, no. 3, pp. 1381–1393, 2025. [CrossRef]

[41] J. Pérez de Frutos, R. Holden Helland, S. Desai, L. C. Nymoen, T. Langø, T. Remman, and A. Sen, "AI-Dentify: Deep learning for proximal caries detection on bitewing X-ray—HUNT4 Oral Health Study," *BMC Oral Health*, vol. 24, 2024, Art. no. 344. [CrossRef]

[42] A. A. Karcioglu, R. M. Yilmaz, M. Yaganoglu, M. Almohammad, and A. Laloglu, "Advancing forensic dentistry: A comprehensive review of machine learning and deep learning applications in dental image analysis," *Neural Comput. Appl.*, vol. 37, pp. 24997–25032, 2025. [CrossRef]

[43] C. Wilkinson, M. Pizzolato, D. De Angelis, D. Mazzarelli, A. D'Apuzzo, J. C. Liu, P. Poppa, and C. Cattaneo, "Post-mortem to ante-mortem facial image comparison for deceased migrant identification," *Int. J. Leg. Med.*, vol. 138, pp. 2691–2706, 2024. [CrossRef]

[44] Y. Li, M. E. H. Daho, P.-H. Conze, R. Zeghlache, H. Le Boité, R. Tadayoni, B. Cochener, M. Lamard, and G. Quellec, "A review of deep learning-based information fusion techniques for multimodal medical image classification," *Comput. Biol. Med.*, vol. 177, 2024, Art. no. 108635. [CrossRef]

[45] P. Chitrapu, M. K. Morampudi, and H. K. Kalluri, "A secure and explainable multimodal biometric system

ACCEPTED ARTICLE

using trust adaptive fusion for face and fingerprint," *Sci. Rep.*, vol. 16, 2026, Art. no. 14244. [CrossRef]

[46] N. Azad, H. Moussddik, K. El Fazazy, O. Elharrouss, H. Tairi, and J. Riffi, "Deep learning-based multimodal biometric system: A fusion approach integrating iris, face, and finger vein traits," *Arab. J. Sci. Eng.*, 2025. [CrossRef]

[47] S. Selvarani and M. Mary Shanthi Rani, "Deep learning-powered secure multimodal biometric feature fusion with explainability and real-time deployment," presented at the International Conference on Artificial Intelligence and Secure Data Analytics (ICAISDA 2025), Puducherry, India, Dec. 29–30, 2025, 2026, pp. 950–977. [CrossRef]

[48] M. L. Gavrilova, "Information fusion: A decade of innovations in biometric multimodal research," *Computer*, vol. 58, no. 4, pp. 27–36, 2025. [CrossRef]

[49] H. Aldawsari and S. Alammar, "Integrating explainable AI with synthetic biometric data for enhanced image synthesis and privacy in computer vision systems," *Image Vis. Comput.*, vol. 162, 2025, Art. no. 105726. [CrossRef]

[50] J. Qian, A. Joshi, H. Li, N. A. Parikh, J. R. Dillman, and L. He, "Utility of deep learning to address missing modalities from multi-modal medical imaging studies: A systematic review," *Artif. Intell. Appl.*, 2025. [CrossRef]

[51] J. M. Durán, D. van der Vloed, A. Ruifrok, and R. J. F. Ypma, "From understanding to justifying: Computational reliabilism for AI-based forensic evidence evaluation," *Forensic Sci. Int. Synerg.*, vol. 9, 2024, Art. no. 100554. [CrossRef]

[52] X. Wang, S. Wei, Z. Zhao, X. Luo, F. Song, and Y. Li, "3D–3D superimposition techniques in personal identification: A ten-year systematic literature review," *Forensic Sci. Int.*, vol. 365, 2024, Art. no. 112271. [CrossRef]

[53] T. Seo, Y. Yoon, Y. Kim, Y. Usumoto, N. Eto, Y. Sadamatsu, R. Tadakuma, and J. Morishita, "Sex estimation using skull silhouette images from postmortem computed tomography by deep learning," *Sci. Rep.*, vol. 14, 2024, Art. no. 22689. [CrossRef]

[54] A. Azizi, P. Charalambous, and Y. Chrysanthou, "DeepSafe: Two-level deep learning approach for disaster victims detection," *Virtual Real. Intell. Hardw.*, vol. 7, no. 2, pp. 139–154, 2025. [CrossRef]

[55] F. Marsico and M. Amigo, "Ethical and security challenges in AI for forensic genetics: From bias to adversarial attacks," *Forensic Sci. Int. Genet.*, vol. 76, 2025, Art. no. 103225. [CrossRef]

[56] A. F. Sobral, D. J. Barbosa, and R. J. Dinis-Oliveira, "CRISPR-Cas technology in forensic investigations: Principles, applications, and ethical considerations," *Forensic Sci. Int. Genet.*, vol. 74, 2025, Art. no. 103163. [CrossRef]

[57] A. R. Mason, H. S. McKee-Zech, D. W. Steadman, and J. M. DeBruyn, "Environmental predictors impact microbial-based postmortem interval estimation models within human decomposition soils," *PLOS ONE*, vol. 19, no. 10, Art. no. e0311906, 2024. [CrossRef]

[58] A. Franco, J. Murray, D. Heng, A. Lygate, D. Moreira, J. Ferreira, D. Miranda e Paulo, C. P. Machado, J. Bueno, S. Mânica, L. Porto, A. Abade, and L. R. Paranhos, "Binary decisions of artificial intelligence to classify third molar development around the legal age thresholds of 14, 16 and 18 years," *Sci. Rep.*, vol. 14, 2024, Art. no. 4668. [CrossRef]

[59] O. H. Milani, S. F. Atici, V. Allareddy, V. Ramachandran, R. Ansari, A. E. Cetin, and M. H. Elnagar, "A fully automated classification of third molar development stages using deep learning," *Sci. Rep.*, vol. 14, 2024, Art. no. 13082. [CrossRef]

[60] B. Büyükçakır, J. Bertels, P. Claes, D. Vandermeulen, J. de Tobel, and P. W. Thevissen, "OPG-based dental age estimation using a data-technical exploration of deep learning techniques," *J. Forensic Sci.*, vol. 69, no. 3, pp. 919–931, 2024. [CrossRef]

[61] N. Mohammad, R. Ahmad, M. H. A. Gaus, A. Kurniawan, and M. Y. P. M. Yusof, "Accuracy of automated forensic dental age estimation lab on a large Malaysian dataset," *Forensic Sci. Int.*, vol. 361, 2024, Art. no. 112150. [CrossRef]

[62] C. P. Pereira, A. Rodrigues, D. Augusto, A. Santos, V. Nushic, and R. Santos, "Dental age assessment and dental scoring systems: Combined different statistical methods," *Int. J. Leg. Med.*, vol. 138, no. 4, pp. 1533–1557, 2024. [CrossRef]

[63] C. P. Pereira, "AI decision support in forensic dental age assessment: Proposed criteria for living individuals," *Int. J. Leg. Med.*, vol. 140, no. 3, pp. 1443–1450, 2026. [CrossRef]

[64] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognit. Lett.*, vol. 24, no. 13, pp. 2115–2125, 2003. [CrossRef]

ACCEPTED ARTICLE

Disclaimer/Publisher's Note: The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the publisher or editors. The publisher and editors assume no responsibility for any injury or damage resulting from the use of information contained herein.