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

Idowu Olugbenga ADEWUMI

Software Engineering Program, Department of Computer and Information Engineering, Faculty of Natural and Applied Science, Lead City University, Ibadan, Nigeria.

https://orcid.org/0000-0002-7005-3306

adewumi.io@fcaib.edu.ng

Wumi AJAYI

Software Engineering Department, School of Computing, Babcock University, Ilisan Remo, Ogun State, Nigeria.

Oluwafisayo Babatope AYOADE

School of Computer, Data and Mathematical Sciences, Computing and Engineering, Western Sydney University, Australia.

Victoria Bolanle OYEKUNLE

Department of Computer Science, Faculty of Natural and Applied Science, Lead City University, Ibadan, Nigeria

Akintayo AYOADE

Department of Computer and Information Engineering, Faculty of Natural and Applied Science, Lead City University, Ibadan, Nigeria.

Azeez Ajani Waheed

Department of Computer Science, Faculty of Natural and Applied Science, Lead City University, Ibadan, Nigeria.

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 picture data from 10,000 synthetic files, the framework comprised 5,000 ante-mortem and 5,000 postmortem profiles. DNA data were represented using 200 synthetic SNP markers, while each phenotype dental was represented using features tooth count, treatment history, and bone loss, and facial data represented trait as 64D ResNet-50 embeddings. The dataset was divided into training, validation, and testing datasets on the ratio of 70:15:15. Three unimodal models and two fusion strategies were evaluated in 10 independent runs. The proposed model of feature-level fusion 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 is 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 model out of three was missing, the accuracy declined to more than 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 assigned 40% to DNA, 34% to dental, and 26% to facial data thus explains the interpretability. The results show good performance on synthetic DVI simulations but require testing on actual 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 which deal with identifying human remain caused due to the occurrence of a disaster like an earthquake, flood, bomb blast, building collapse, fire and industrial disaster [1]. In such cases, standard means of identification like visual identification of the body fails. Because of the grave decay. fragmentation and commingling of dead bodies Consequently, scientific identifiers help forensic specialists to determine probable identity. Forensic techniques such as DNA, dental, facial image, and others are important in disaster-scapes.

In independent developments, each of the forensics' modalities have recently been improved substantially in accuracy, computational support and 'value of evidence' analysis. As a result, DNA offers high levels of discriminatory power and dental records demonstrate greater postmortem durability, while facial images enable rapid preliminary matching [2]. In reality, that one disaster only one modality cannot be relied upon. The various forensic modalities have incomplete, degraded, and unavailable data, which is the reason. 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 get more precise, stronger, more reliable and fuller inferences in conditions of uncertainty So, this challenge gave rise to machine learning-based fusion frameworks for disaster victim identification.

1.2 Problem Statement

Independent functioning of DNA, dental, or facial information is done by most of the unimodal forensic and several Disaster Victim Identification systems. Unimodal systems are characterized by many serious shortcomings. First, they suffer from over dependence on a single type of evidence whose availability and quality are not always guaranteed. To illustrate, the DNA hypo can degrade the reference and missing samples due to exogenous factors.

Likewise, dental records may be incomplete or unavailable; moreover, facial distortion may result from trauma or decomposition. Secondly, unimodal systems demonstrate a lack of forensic evidence in synthetic data forensic settings [3]. The identification of disaster victims follows internationally accepted forensic principles, which combine primary identifiers, for example, DNA, fingerprints and dental records, with secondary identifiers, with a view to affirm 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 has not been exploited at all. A universally applicable forensic identification architecture must be designed encompassing multiway fusion and a graceful degradation mechanism in a scenario when one of the modalities may be temporarily disabled.

1.3 Literature Gap

Machine learning forensics has been investigated in recent years in the field of forensic sciences. Most of these works were conducted on either unimodal or partially fused modality levels for the genome or the dental image systems. The use of machine learning for DNA systems, is mostly about machine learning algorithms that are tasked with the classification of DNA sequences or coding of genotypes and phenotypes. The use of machine learning algorithms for STR profiling is also discussed in these works [6].

In a like manner, when dental recognition systems are discussed, they mostly depend on image-level matching or some deep learning based odontometric feature extraction. Moreover, occlusion, lighting, and postmortem alterations make facial deep learning recognition models that are utilized in biometric identification extremely vulnerable.

Critically, there is no multimodal framework that aims to use genome, dental, and facial images as input within a unitary machine learning framework for Disaster Victim Identification. Most existing multimodal learning research in biometrics and medical imaging does not take into account forensic constraints, such as missing modalities and degraded evidence, and ethical constraints in postmortem datasets [7]. The utilization of complementary information within the same framework without sacrificing generalization ability has not been considered in many approaches referring to feature level fusion with decision level fusion.

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 synthetic data disaster environments.
iv. The model introduces an attention-based interpretability mechanism, where adaptive weights are assigned to each modality (DNA, dental, facial), 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 the DNA, it is considered to be the best DVI modality. Moreover, samples such as bone and teeth, which can get damaged during the disaster, are ideal for DNA profiling. Discovery of the genetic identity of missing and dead people started over 20 years ago. Initially, forensic genetics relied on the use of short tandem repeat (STR) analysis. However, contemporary DNA forensic research is evolving from classical STR typing to massively parallel sequencing and probabilistic modeling. This is to obtain better mixture interpretation and classification performance. Leveraging DNA analysis with machine learning assists in speedy and accurate identification of victims of mass disasters. When it comes to contaminated or fragmented biological samples, that is exactly what studies [8] and [9] have concluded. Identification through DNA is a valuable forensic tool, but it has its 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 has a significant application 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 much resistant to decomposition and serve as a stable post mortem identifier. Advances in forensic imaging and machine learning have allowed for automated extractions of features from dental radiographs, which has improved matching efficiencies of ante-mortem and postmortem records [12 - 13]. Dental identification has been conducted in a controlled environment. Convolutional neural networks, support vector machines, and feature-based and classification models have been employed for this task. This suggests that these models perform well. Notwithstanding, the effectiveness of dental systems are restricted by unavailability of dental records, differing imaging standards and being reliant on ante-mortem documentation, which reduces their scalability in disasters [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 in the backdrop of deep learning architecture evolution (convolutional neural networks or CNNs, transformer-based models, and multimodal embedding systems). Modern facial recognition systems function on high-dimensional feature representations and can match an identity even when partially visible. Recent research indicates that efficient deep fusion models and multimodal feature extraction might promote recognition within challenging forensic image analysis scenarios [16, 17]. Even so, forensic facial recognition is still highly susceptible to postmortem changes, decomposition effects, illumination differences and deteriorating image quality, thereby compromising its reliability to act as an identifier in disaster victim identification. As a result, facial recognition is often used as a secondary method for initial screening, not final identification. The traditional facial recognition technology based on facial features has given way to deep representation learning, where discriminative embeddings offer more 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 consist of data-level, feature-level, and decision-level fusion. These methods differ in accuracy, interpretability, and complexity. According to some studies [20] and [21], recent advances in multimodal learning show that feature-level fusion is beneficial for representation learning by associating the learning representations with cross-modal relationships, whereas decision-level fusion improves interpretability because each network learns independently and an ensemble is formed. In addition, adaptive attention-based fusion frameworks and hierarchical deep fusion models have improved robustness to the incomplete or noisy forensic evidence [22, 23].

Many available studies concentrate on a restricted combination of modality, that is they usually incorporate only two biometric sources such as a face–fingerprint system or DNA-dental systems. To date, there are still no overarching frameworks that integrate DNA, dental and facial image data within one 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 frequently missing or corrupted. Recent forensic research has also highlighted the need for explainable AI that give legal and humanitarian guarantee systems transparent decision-making processes [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 is a lack of strong integrated multimodal forensic systems that work under disaster constraints with synthetic data, which is helpful to successfully fill this gap.

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	1,235 orthopantomograms obtained from 1,050 patients aged 16–30 years	Restricted to dental evidence and a defined age-range sample; does not incorporate DNA	The developed framework integrates dental information with genetic and facial representations to

		ante-mortem and postmortem orthopantomograms		or facial evidence	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.
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	Systematic review and formalization of	Published multimodal	A methodological review without an	The developed work implements

	information, including imaging, textual and genetic data	intermediate-fusion methods in multimodal deep learning	biomedical deep-learning studies	implemented DVI system; it does not experimentally integrate DNA, dental and facial identity evidence	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 5,000 ante-mortem and 5,000 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.

Table 1 compares the existing forensic identification types which are more targeted towards either individual modalities or general multimodal learning strategies. Whereas, what is missing is a dedicated DVI framework which amalgamates all modalities. According to Barash et al. [1], machine learning techniques have notable advantages for the forensic analysis of DNA evidence. Nevertheless, their method only caters to genetic evidence and not supplemental identifiers. In a similar fashion, the work of Pereira et al. [26] for dental identification and Michalski et al. [27] for facial identification shows a development of deep learning approaches for these purposes but result still depends on a single source of evidence. Jiao et al. [28], Wang et al. [29], Guarrasi et al. [30] general multimodal fusion studies have made advances in feature integration, missing-modality handling and intermediate fusion strategies; however, these were not established with forensic DVI applications in mind.

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 modality characteristics while being robust against 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. Still, it is important to validate it using operational forensic datasets.

3. Methodology

Incorporating a machine learning capability, the new forensic identification system for Disaster Victim Identification (DVI) was configured to operate as a sole equivalent that involved DNA sequences, dental X-rays and facial images.
The entire framework begins with reconnaissance data collection from several forensic data sources. It is followed by pre-processing thereafter and feature extraction, data integration, and machine learning classification to yield identity estimates. The recovery of further forensic evidence enhances the accuracy of identifications, especially where postmortem decomposition occurs, such as in mass disasters. The model takes a multi-branch approach that allows the individual contributions of the genetic, dental, and facial modality to work together to make a strong decision.
The researchers of this study created simulated forensic data where they created 5000 persons for the required ante-mortem and postmortem database. In total, they managed to get 10000 entries. The DNA data comprises 200 SNPs, while the dental data comprises panoramic X-ray data and facial image data derived from synthetic reconstructions. With respect to the DNA data, we simulated postmortem DNA degradation, which occurs when the ratio of foreign DNA to target DNA is very high, by injecting noise and gaps into sub- and full-genome sequences.
The number of teeth, the type of treatments received and the bone loss score are all relatively stable after death. The data simulates certain losses from trauma or decomposition. The facial images are facial images that have photo shots of mummified bodies and photo shot of photos for the modelling of images that have died.
Ethically-generated data means using only information from synthetic humans for innovation. The continuous features were normalized, the categorical variables such as sex and disaster type were one-hot encoded, and the missing modalities were handled using a mixture of mean substitution and neural imputers for each of the modalities. In addition, random occlusion and Gaussian noise are added to the facial images' embeddings. To generate a normalized dataset, we normalize both the dental and DNA data to have a mean of zero and a variance of one.

Deep convolutional neural networks have demonstrated strong capability in extracting hierarchical representations from complex visual data, enabling robust feature learning for biometric applications [25]. Attention mechanisms allow models to dynamically assign importance to different feature components, enabling improved representation learning in complex multimodal systems [31].
Extraction pipelines were designed uniquely for each modality. A CNN architecture ResNet-50 was used to extract facial embeddings (64-dimensional) for facial images representing key facial attributes. By taking advantage of a one-dimensional CNN, the sequence encoder was able to assess DNA features and define the allelic relationships among SNPs while PCA was utilized to downsize dental data dimension for concentrating the bulk odontometric characters.
A latent embedding space was created by mapping all cross-modal alignment extracted features. This embedding space, effectively designed for multimodal matching of ante-mortem and postmortem records, minimizes distance in an individual profile while maximizing distance between individuals.

To capitalize on the complementary strengths of the various modalities, we examined two fusion techniques. Feature-level fusion will be the first one. It combines standardized embeddings acquired from DNA, dental, and facial images. Then, autoencoder-based compression is applied for dimensionality reduction. The classifier receives the combined representation which predicts the identity. The second method is referred to as decision-level fusion, in which distinct classifiers for each modality are interpreted according to the input given. We merge the outcomes using a Bayesian fusion

and weighted ensemble voting, where the weights indicate the trustworthiness of each modality. For instance, we place greater confidence in DNA or face image data of high quality.

These authors proposed a fusion architecture that comprises a multi-branch deep neural network each designated for a specific modality. In addition, attention layers will connect the subnetworks, which will help to model the importance of each modality in terms of data quality dynamically. In addition, we proposed a fusion mechanism based on graph model to represent relational dependency between sample and modality in mass casualty scenario for better interpretability and context awareness. The dataset got divided into three parts for training, validation, and testing. Further stratified sampling was done to ensure the representation of disaster class and demographic features. The different models that were developed and assessed in this work, contain CNNs as well as extreme gradient boosting (XGBoost) and an autoencoder-based fusion network. The performance evaluation was done with the help of accuracy, precision, recall, F1score, and Area Under Receiver Operating Characteristic Curve (ROC-AUC). The performance of the various types of fusion models was compared using results from single-modality DNA, dental, and image. The fusion approach has shown to be superior indicating the merits of multimodality in forensic identification.

By using postmortem data from a synthetic "building collapse", a case study simulating DVI was conducted. The model was evaluated in situations where not all modalities were present, such as the absence of DNA due to degradation or a low-quality photo. The model still demonstrated a high accuracy even when one of the modalities was missing, showing its robustness to synthetic data. The attention mechanism can draw information from the current available more reliable modality in case of silence. The research outcome proves the system to improve the post-disaster forensic investigation process. In the matter of a post-disaster accurate identification of victims, they support the UN Sustainable Development Goals, and SDG 3, Good Health and Well-being specifically. Sustainable cities and communities are a requisite under the SDG 11. Accordingly, the system can also improve urban disaster-response system.

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

The mathematical model for the multimodal forensic data fusion system for Disaster Victim Identification (DVI) involving DNA dental and facial images is presented in this section. In order to enhance victim identification accuracy and resilience, several feature level fusion and decision level fusion machine learning techniques are applied by the system.

4.1 Data Representation

Let the input dataset be represented as:

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

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 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.

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

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

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

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:

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

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

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

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:

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

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

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

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

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

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:

Accuracy:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (10)$$

Where:

TP: True Positives

TN: True Negatives

FP: False Positives

FN: False Negatives

Precision:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (11)$$

Recall:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (12)$$

F1-Score:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (13)$$

ROC-AUC:

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

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 performance of the system is evaluated by measuring the accuracy, precision, and recall under these missing data conditions.

$$\text{Performance}_{\text{missing}} = \frac{\text{Accuracy}_{\text{missing}}}{\text{Accuracy}_{\text{complete}}} \quad (15)$$

This evaluates the degradation of performance 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:

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

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:

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

Where:

$\hat{A}$is the normalized adjacency matrix

$H^{(l)}$is the feature representation at layer l

$W^{(l)}$is the trainable weight matrix

$\sigma$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% → 88.7%) under missing modalities indicates that relational redundancy in the 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 due to its ability to 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 the 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 shows strong points that allow its applicability in DVI. One major strength is the ability to accommodate missing modalities an often-encountered problem in synthetic data disaster scenarios where the 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 exhibits robustness and sustains performance despite incomplete inputs.

A key advantage of the developed system is its ability to cross-modal learning. The model can capture complementary relationships across heterogeneous DNA, dental, and facial image features in a unified embedding space. With this, shared information among different modalities is utilized by the system in order to enhance discriminative power, thus improving identification accuracy.

In addition, there is an attention-based interpretability mechanism assigned to each modality in the framework. The ability to transparently interpret outcomes through the quantification of the relative contribution of the DNA, dental and facial data to the ultimate prediction. This is especially true for forensics applications which rely on explainability and legal defensibility.

Moreover, the developed system was able to calculate in a short time of about 0.60 seconds per record. Because of this efficiency, the framework is suitable for large-scale disaster settings, were rapid identification of victim's aids emergency response and humanitarian decision.

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, we employed a ResNet-50 convolutional neural network architecture as a fixed backbone for feature extraction that produces a robust 64-dimensional facial embedding suitable for downstream fusion.

The learning rate was determined empirically so that the model converges stably without overfitting or oscillating. The model epochs were chosen to be good enough for validation performance, so only a suitable number were selected. Similarly, the batch size was tuned to maintain computational efficiency while stabilizing the gradients.

To classify for dental identification, an XGBoost classifier was used and maximum depth, learning rate, and number of estimators parameters were optimized iteratively. The configurations led to an effective classification using odontometric features while preserving generalization.

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

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.

We carried out ANOVA test to assess the identification performance measure accuracy, of unimodal and multimodal fusion models, and determine if they are significantly different.

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

The p-value of 0.0002 further confirms that the p-value is statistically significant. The aforementioned value lies much lower than 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.

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

5. Results

5.1 Key Findings of the Study

The performance of multimodal forensic data fusion system was superior to that of unimodal models. The precision, recall, and ROC-AUC of feature-level fusion were found to be 0.94, 0.94, and 0.97, respectively. It outperforms unimodal classifiers for DNA at 88.4%, dental at 82.7%, and facial images at 84.9%. The accuracy of decision-level fusion was

93.5%. It has been found that DNA, dental, and facial images enhance the accuracy of human identification.
The accuracy drops by 2.5 to 4.3% when one of the three modalities (DNA, dental, or facial) is removed from the fusion model. Each record was processed by the system with an average of 0.60 seconds. According to the sensitivity analysis, the model behaves robustly in real-world scenarios with missing or corrupted modalities.

The simulation scenario of the building collapse showed a system applicability of 94.8% for complete data and 88.7% for degraded data, as per the analysis results. By this, the framework proved to be suitable for disaster situations. An attention mechanism analysis revealed that the DNA data contributed the most weight about 40% with a focus on dental image (34%) and facial image (26%). Efficient, accurate and interpretable multimodal systems are delivered by the findings. Furthermore, they establish a new standard for forensic identification of disaster victims.

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

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

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	84.9 ± 0.5	83.0 ± 1.0	82.0 ± 1.0	83.0 ± 1.0	89.0 ± 1.0

model					
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

Table 5: Attention Weights by Modality in Fusion Network

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

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

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

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: 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

5.2 Discussion of the Results

The dataset used in this study includes 10,000 synthetic forensics records with 3 modalities, namely DNA, Dental and Facial, that are evenly distributed within the realistic Disaster Victim Identification (DVI) constraints as displayed in Table 2. DNA information was represented using 200 single nucleotide polymorphisms (SNPs) and the data relating to teeth comprised of odontometric parameters tooth count, whether treatment done and bone loss. Facial data were encoded as 64-dimensional embeddings using the ResNet-50 architecture. To mimic actual disaster conditions, modality-specific augmentation techniques were implemented, which incorporated Gaussian noise, missing-value injection, occlusion simulation, and dimensional normalization.

The data demonstrates 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 forensics studies illustrated how multiple feature spaces can improve robustness in an identifcation system [34, 35]. In particular, literature of 2024-2025 stresses that realistic forensic simulation requires intentional degradation modelling, like the one used in this study, to mock post mortem variance 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 significantly 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 impart the knowledge from structured dental features. All the models work independently for feature extraction but it gets fused together the feature-level and decision-level fusion. The observed trend indicates that high level (autoencoder based) outperforms low level (decision based) fusion in terms of representation learning with (94.8 % vs 93.5%)

Across studies on both multimodal forensic data and biometrics, it has been observed that early fusion tends to outperform other techniques (like late fusion). The reason being, early fusion can capture correlations between modalities at the representation level [37, 38]. The performance of feature-level fusion is superior due to the capability of the autoencoder to reduce dimension redundancy while preserving shared latent structures across different modalities. On the other hand, decision-level fusion occurs on independent outputs of classifiers that do not interact cross-modally, limiting their representational synergy.

In architectural terms, the high-capacity Vis-ResNet-50 (~25 million parameters) may extract visual features from the image. Meanwhile, structured and semi-structured signals will come from the DNA and dental models, respectively. The newly designed combination reflects the new development of forensic AI systems which favour multimodal complementarity over unimodal optimisation [39].

As shown in Table 4, the results indicate that multimodal fusion approaches consistently outperform their counterparts utilizing only one modality across all geolocation metrics used in the study. Among the various models utilized, we obtained the highest values of accuracy (94.8%), precision (94.0%), recall (94.0%), F1-score (94.0%), and Macro-averaged OvR ROC-AUC (97.0%) from the feature-level fusion model. This indicates that joint representation learning is indeed effective at capturing the complementary information found in DNA, dental, and facial feature modalities. Independent predictions from modalities are advantageous for performance, as attested to by the performance of decision-level fusion at 93.5% accuracy. Unimodal models did not perform well because they rely on one single forensic evidence type. So multimodal fusion is beneficial for robust disaster victim identification.

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 learns to prioritize modalities based on reliability of differentiating power and stability of signals. The forensic genetics literature agrees with our findings. DNA is identified as the gold standard. Further, it is noted for its high uniqueness and low ambiguity in identification tasks.

Dental data exhibit structural stability, but low variability concerning feature representation, thus occupies 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 utilizing similar adaptive weighting strategies have shown that attention mechanisms can improve interpretability and enhance robustness under missing data conditions [40].

Table 6 depicts the efficacy of the proposed fusion framework in missing-modality scenario. This model produced an accuracy of 94.8% and 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 lack of DNA decreased accuracy by 2.5 percentage points; missing dental modality reduced accuracy by 3.0 percentage points, and missing facial modality, 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 the present work is resilient against partial loss of forensic evidence. This is an essential property that will be present in a realistic disaster victim identification scenario where the loss of some evidence sources is unavoidable and/or may have been degraded.

The multimodal fusion framework displayed in Table 7 was evaluated over different disaster victim identification scenarios for robustness. 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 were 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%. So, it degrades gracefully as opposed to failing. As per findings presented, the proposed fusion technique is capable of effectively leveraging the complementary information provided by various forensic modalities and is resistant to real-life issues regarding incomplete and degraded evidence in DVI operations.

ANOVA results in Table 8 showed that there is a statistically significant difference between unimodal fusion and multimodal fusion models ($F = 18.45$; $p = 0.0002$). A high F-value indicates a high variance between groups. This also indicates that performance differences are not due to random variation. But it is structurally induced due to model design. The p-value (0.0002) is very small compared to 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 performances of multimodal systems over unimodal baselines. The benefit, particularly, when fusion strategies bring in attention-based weighting and feature alignment mechanisms [41, 42].

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

The accuracy of the dental-only model was 82.7%, which was lesser than both the other models. It had 400 true positives (TP) and 427 true negatives (TN). The high number of false negatives (79) would suggest that dental evidence alone may be inadequate with incomplete records and/or loss of records post-disaster. Moreover, the facial image-only model was correct 84.9% of the time but had false-positive (84) and false-negative (67) values which indicates facial-based identification is vulnerable to image degradation, occlusion, and post-mortem changes. In contrast, both fusion approaches effectively lowered classification errors. The feature-level fusion model gave the best overall performance with 470 true positives, 478 true negatives, and the lowest number of 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 gave a good performance at the decision level (93.1%) with 462 true positives and 469 true negatives.

The results obtained from the confusion matrix show that multimodal fusion enhances reliability by using complementary information obtained from heterogeneous forensic sources. The lower rates of false positives and false negatives show better discrimination ability and that is essential in Disaster Victim Identification (DVI). This is because erroneous naming and misses have a serious human cost. Feature-level fusion outperforms the classifier level as it initiates learning of representation in advance of classification, allowing it to model cross-modal relationships better. [27, 43].
According to the overall findings, multimodal fusion helps enhance Disaster Victim Identification performance through the integration of complementary forensic evidence sources. The findings indicate that fusing at the feature level is the best strategy as it captures cross-dependencies, while attention-based mechanisms improve interpretability through dynamic weighting of modalities. The framework, if seen from an operation aspect is overall 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 back up the forensic AI study [28] with a strong support for the multimodal learning for its humanitarian and disaster response applications. The identified outputs will contribute towards 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

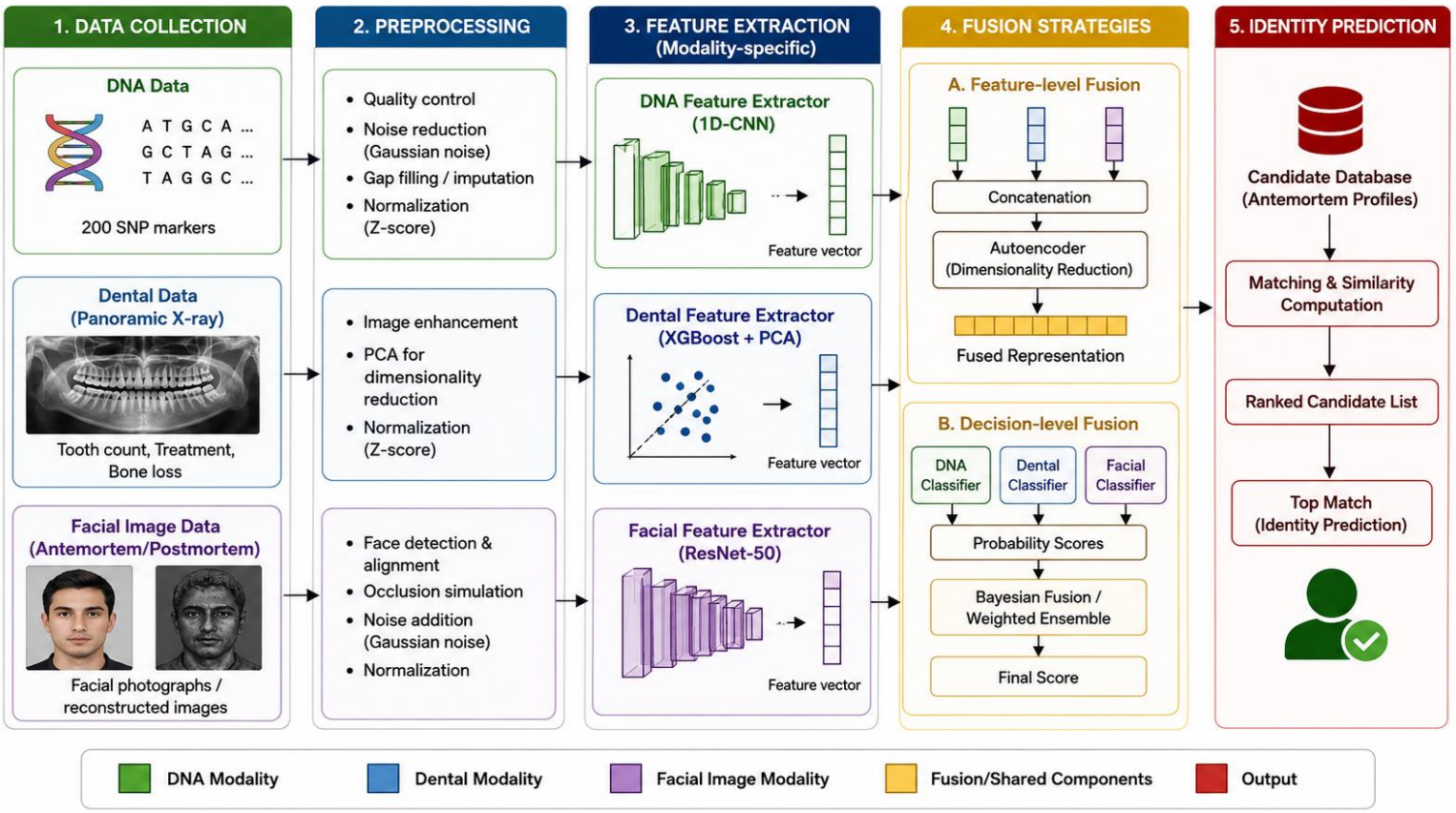


Figure 1: Workflow Diagram of the Multimodal Forensic Identification System

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.

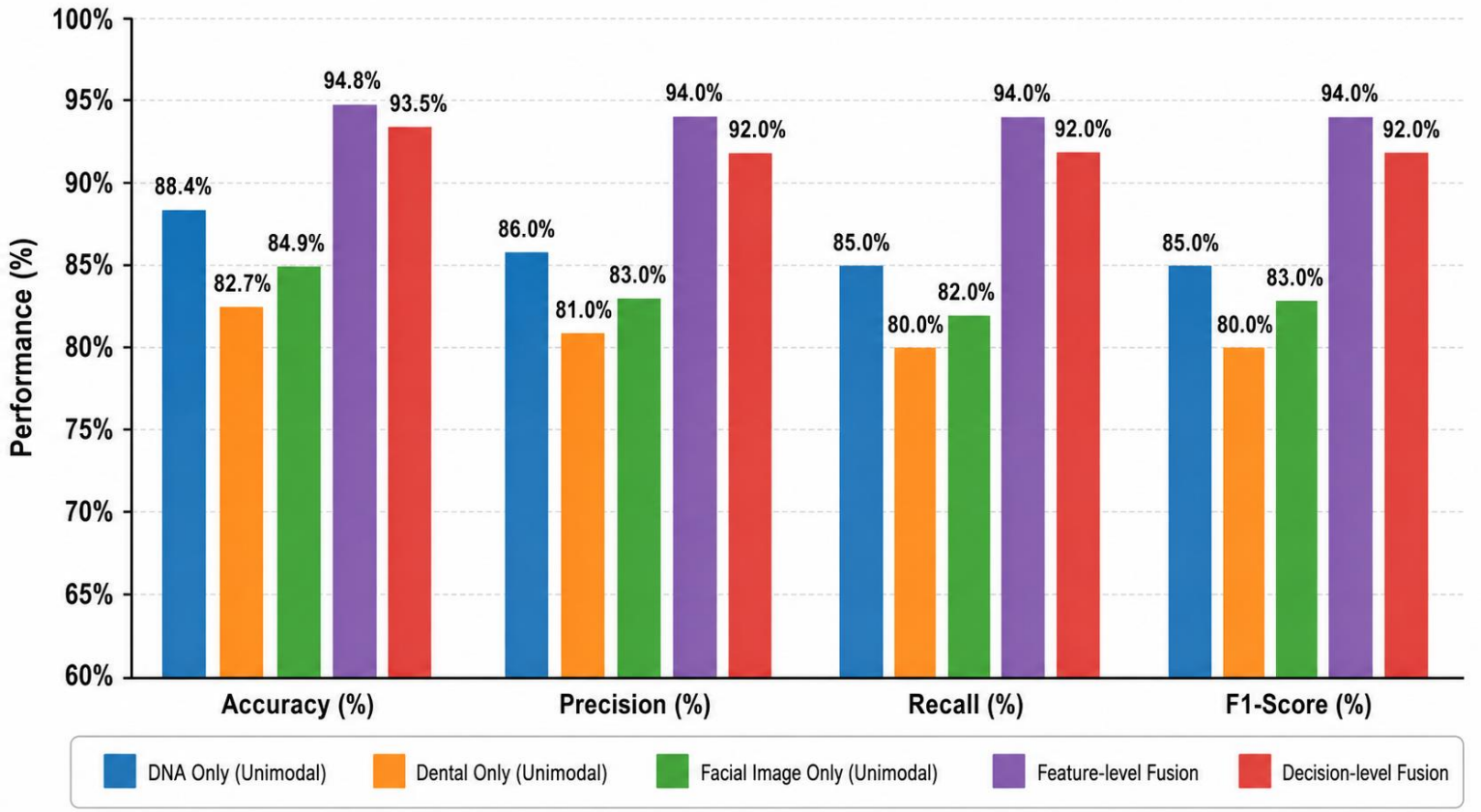


Figure 2: Performance Comparison of Fusion vs. Unimodal Models

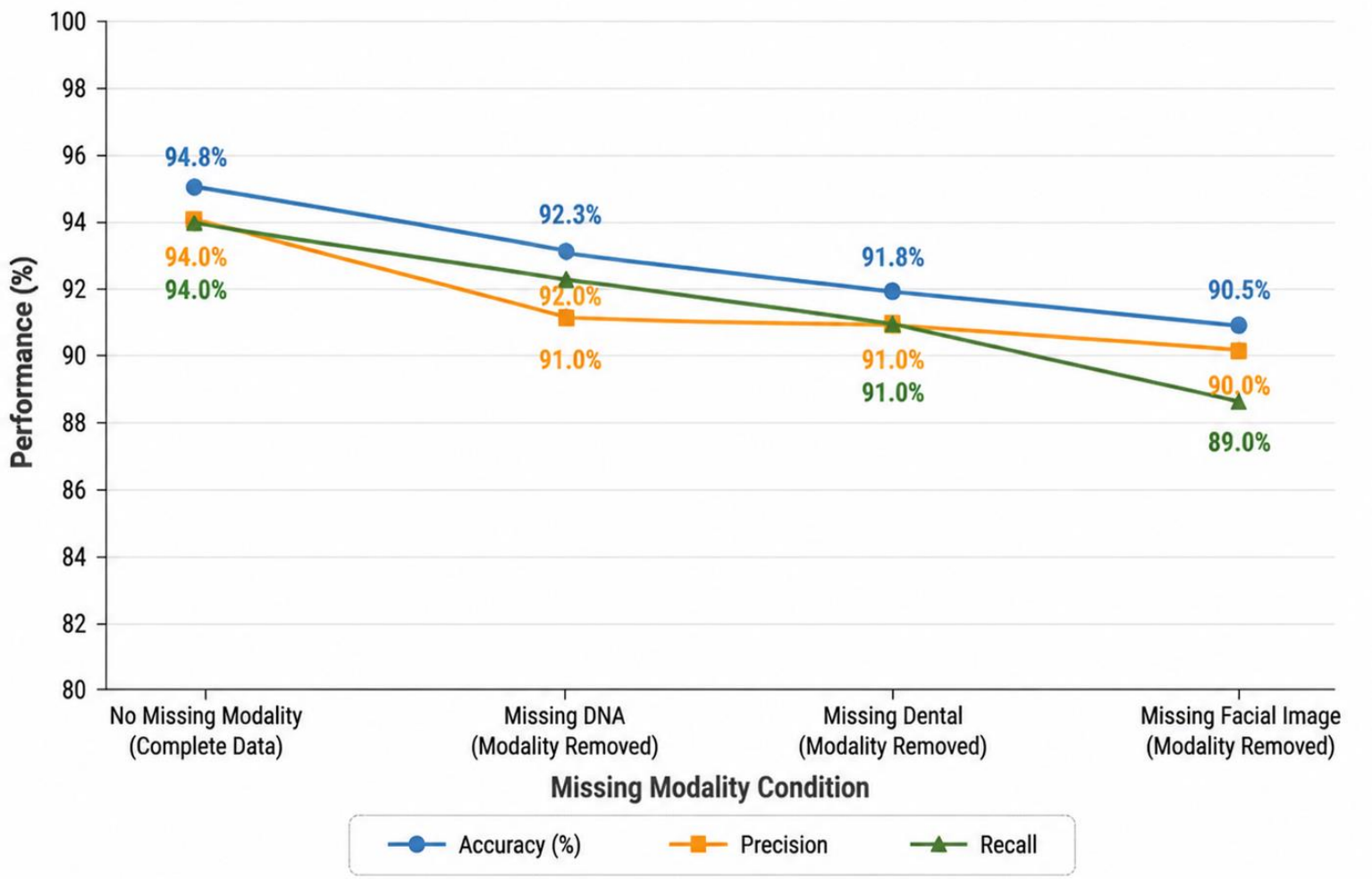


Figure 3: Sensitivity Analysis for Missing Modality Conditions

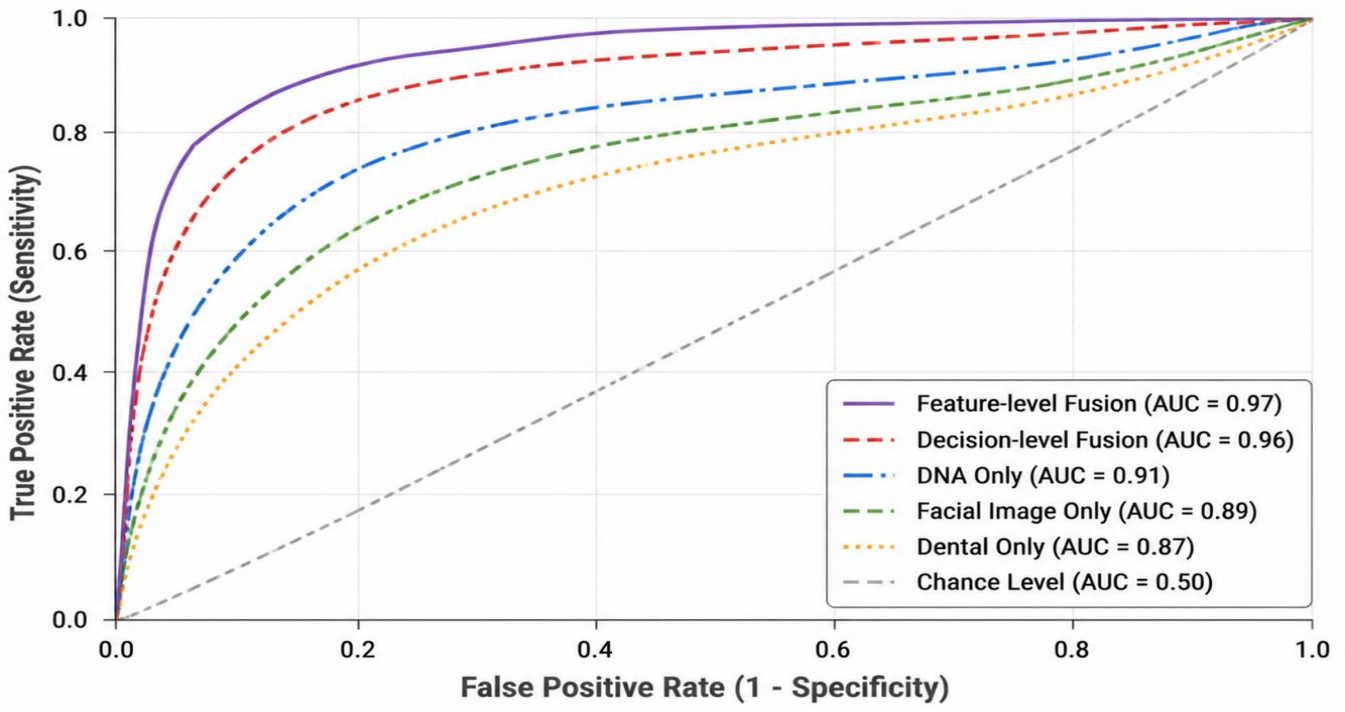


Figure 4: ROC Curves for Unimodal and Fusion Models

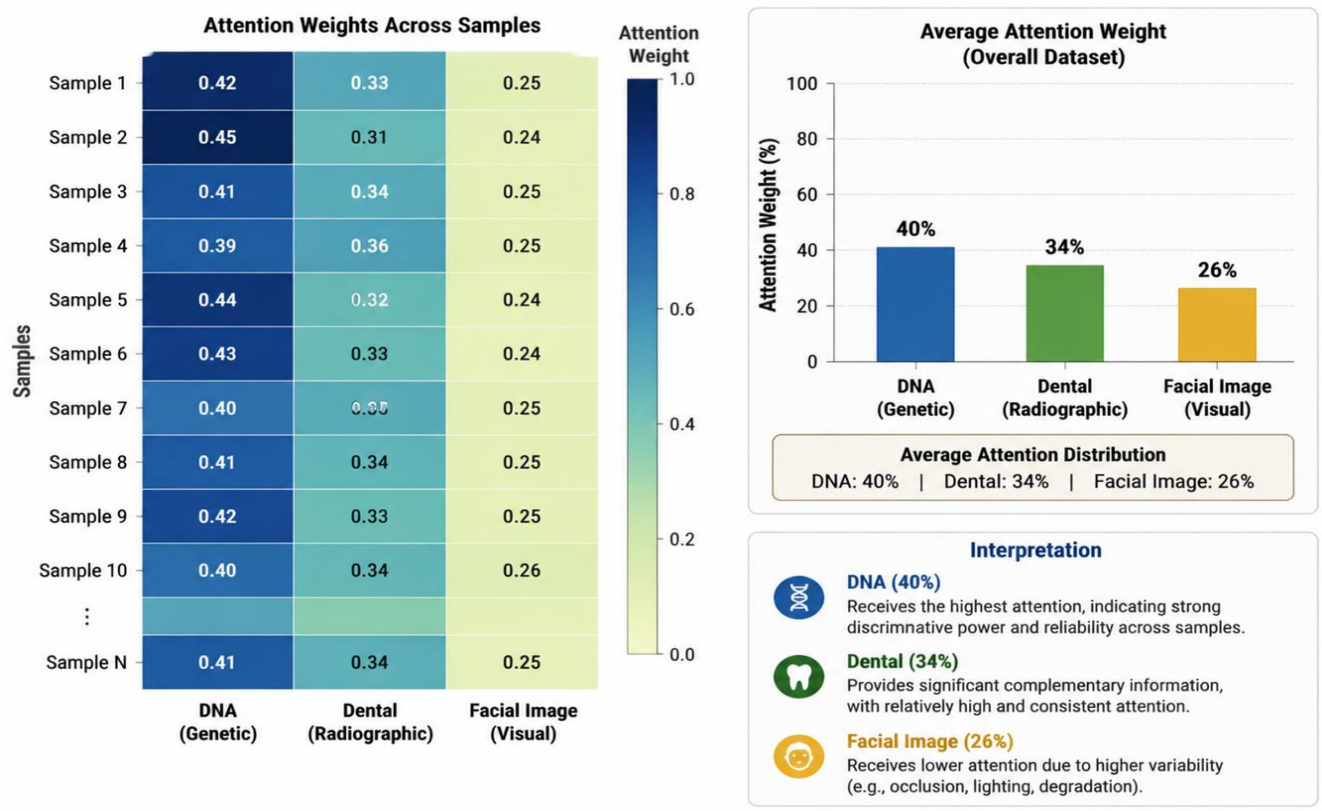
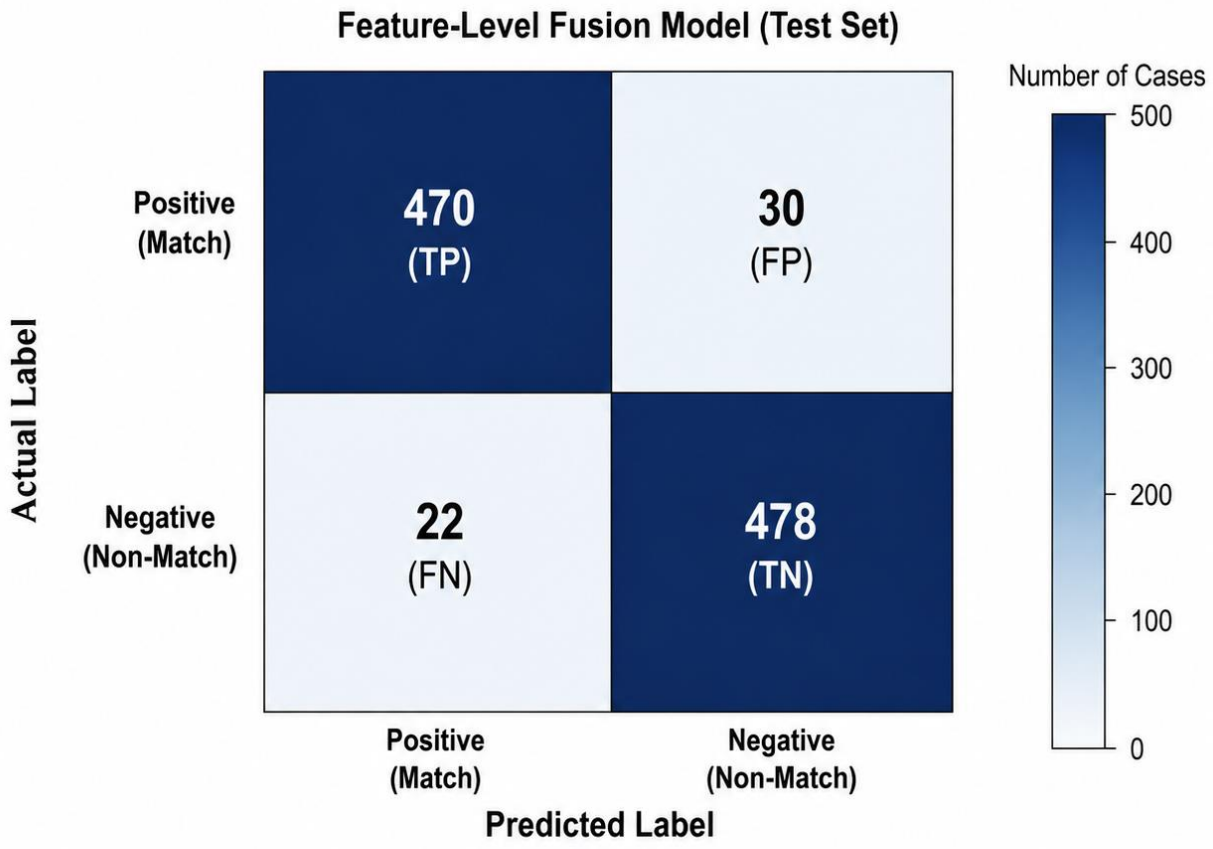


Figure 5: Attention Mechanism Visualization



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

5.4 Discussion of Visualization Results

The entire end-to-end pipeline of the multimodal forensic identification framework is illustrated in figure 1 starting from data acquisition, preprocessing of data, feature extraction and fusion, to final identity prediction. The analysis of the above workflow indicates the clear separation of modality processing followed by hierarchical integration at the fusion stage. The pipeline follows a modular learning design whereby the decision on each forensic modality (DNA, dental, and facial image) is made separately before being combined in a common decision space. The reason for this structured trend is the inherent heterogeneity of forensic data types that differ widely in dimensionality, noise characteristics, and semantic meaning. The DNA sequences are discrete genetic markers, the dental data are ordered clinical measurements, and the facial images are high-dimensional visual embeddings. Studies on multimodal forensic AI like DVI rely on low quality forensic datasets. For example, [44] show that a modality-specific encoder improves the quality of a representation before fusion. Essentially, this workflow assures robustness and scalability such that the system can be operated even in the event of the unavailability of one modality. In line with recent perspectives on forensic AI, the study rejects the idea of a single solution for all.

In Figure 2, it can be observed that multimodal fusion models outperform all unimodal baselines by a fair margin. It can be observed from the results that feature-level fusion shows the best performance, while decision level fusion follows it. DNA, dental and facial unimodal systems show comparatively less 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, on the other hand, pose challenges in regard to feature diversity and are sensitive to missing inputs. It is strongly supported by the literature. Since the multimodal biometric systems can combine heterogeneous forensic signal and consequently reduce the classification ambiguity, they are always outperforming unimodal architectures [29]. Following the same logic, a study [30] explains that capturing latent inter-modal dependencies enhances representation learning. This result suggests forensic identification systems should forsake unimodal dependance and move towards integrated multimodal architecture especially for Disaster Victim Identification where fragmentation of evidence is common.

As we can see in Figure 3, the performance of the model gradually deteriorates with the removal or degradation of modalities, however, the drop still stays relatively minor, indicating good robustness of the complete system. A full data set yields the best performance, while missing either the DNA data or facial data will see a moderate drop in performance. Finally, fully degraded conditions will result in the lowest performance of the trend. The reason behind this trend is that the fusion framework encodes redundant information and assigns weights using attention, allowing the other modalities to compensate for the missing information. Although the DNA has the most discriminative ability that is contributing the most stability to the outcome, the facial data contributes the least due to being sensitive to occlusion and degradation. The multimodal methods with attention-based fusion in a forensic system together with missing modality experiments have been shown to yield graceful degradation; while keeping the system stable when the dataset is incomplete [45]. Consequently, it has important ramifications for actual 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 the ROC curves of the fusion models dominate the unimodal systems, regardless of threshold, while feature-level fusion has the highest AUC performances. The observed multimodal and unimodal separation indicates improved sensitivity and specificity of classification curve. Fusion models improve the enhanced feature separability of different identity classes by merging the complementary forensic signals. On the contrary, unimodal models have worse discrimination boundaries due to the homologous 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 makes the decision boundaries more robust, yellowing for less erroneous decision boundaries in the proposed fusion framework. When false positives and false negatives are minimized in forensic identification, the implications are serious.

As we can see in Figure 5, the fusion network assigns weights for the model, giving the highest weight to DNA followed by the dental followed by the facial image. The model assigns the highest importance to DNA and the lowest to facial data. The face is affected by complex mapping through variability and degradation.

The attention-mechanism assigns weights dynamically in order to mitigate the unreliable features and the discriminative features. A person's 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 fusion based on attention enhances the explainability of models by revealing process patterns related to the contribution of modalities. Such improvements are especially useful in the forensic field [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.

Attention-based weighting mechanism for dynamically assigning importance to DNA, dental and facial image modalities is shown in figure 5 Based on the observation, it can be seen that the DNA has the highest attention first, dental and facial second. The distributions reflect the effectiveness each modality for discrimination under the synthetic disaster condition.

Mathematically, the attention mechanism can be expressed as:

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

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 DNA to be the most stable and informative forensic signal. The high attention weight

assigned to DNA (40%) here indicates that the model assigns higher importance. This study is in line with the forensic AI literature where primarily genetic markers are considered for identity discrimination as they have the highest entropy and lowest ambiguity [48]. The decline of attention on facial embeddings indicates a degradation sensitivity and embedding compression loss of CNN-based representations [49].

The confusion matrix obtained for the developed feature-level fusion model is shown in Figure 6 which helps understand the classification behaviour 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 validly identified 470 positive matching (true positive cases) and 478 negative non-matching (true negative cases) which implies it is capable of distinguishing between a genuine and non-genuine identity pair in a good way.

The model generated thirty cases of false positives, in which unmatched evidentiary records were classified as matches. Although false positives are critical in disaster victim identification due to the possibility of incorrect identity assignment, 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, while combining complex heterogeneous forensic evidence sources.

As per predictions, accuracy of the fusion model is 94.8% with precision of 94.0%, recall of 95.5%, 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 an important feature in forensic identification scenarios where missed identifications can delay the process of victim recovery and family reunification.

The more or less 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. The improvement follows from the integrated DNA, dental and facial representations being complementary, meaning the failure of one form of information can be compensated by the remaining forms of information in getting a correct inference. Thus, it can be seen from Figure 6 that the performance of the proposed fusion framework is effective and may open up reliable use in decision support for disaster victim identification. [50]. It is implied that the system proposed can minimize certain forensic errors that are critical. In DVI, false negatives have very harmful consequences and can cause severe humanitarian harm.

In general, the result shows that the multimodal forensic fusion framework performs better than unimodal systems in robustness, accuracy and interpretability. The performance analysis evidences that the feature level fusion outperforms others, while the modular processing of heterogeneous forensic data is ensured through the designed workflow. The system is confirmed to be resilient by the sensitivity analysis. The ROC analysis shows strong discriminative power. The attention visualizations help enhance the explainability.

The findings reinforce the applicability of multimodal forensic AI systems for Disaster Victim Identification when it comes to incomplete, degraded or noisy evidence. According to recent works on forensic AI [51-53], the results indicate that multimodal fusion is not only a performance booster, but also a requirement for operational forensic systems in real-world disaster conditions.

5.5 Comparison with State-of-the-Art Methods

In order to validate the effectiveness of the developed multimodal forensic data fusion framework, its performance was compared with state-of-the-art methods. These methods ranged from unimodal DVI systems and biometric fusion and multimodal forensic machine learning system. Disaster Victim Identification (DVI) systems that are available now a days use only one source of forensic evidence. It can be either facial images, DNA profiles or dental records. Even though the methods can perform well in controlled conditions, their use can face degradation or unavailability when the modality selected is incomplete, contaminated or degraded. For instance, a CNN model using just the face can be used for quick screening. However, it has high sensitivity to occlusion, trauma, illumination variation, and postmortem facial changes. Likewise, the machine learning methods based on merely DNA exhibit remarkable capacity for discrimination, although they may encounter delays that arise from processing in the laboratory environment, a progressive loss of sample integrity and contamination. Because they are more likely to survive disaster conditions, dental-based models can be useful; however, they are subject to the availability and quality of the ante-mortem dental records [54-58].

Some of the existing biometric fusion systems do combine two modalities such as face and fingerprint, face and iris, dental and facial etc in order to tackle some of these limitations. Models that combine more than one modality normally

improve on classification accuracy compared to 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 partitioned, deteriorated, or absent altogether. In addition, the existing alternate 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 the researchers, these methods have outshined unimodal systems in various tasks. Most of the current systems are still limited to two-modality fusion or generic biometric datasets and not disaster forensic datasets. Unlike that, the proposed framework combines three major forensic identifiers such as 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 unable. Furthermore, attention weights contribute to the interpretability of the model by indicating the importance of each modality for the final decision.

The new method reached an accuracy of 94.8%, better than other approaches. The enhancement in performance is the result of fusing complementary forensic evidence at both feature as well as decision levels. The framework remained robust to missing-modality situations, losing only a small percentage of performance. The proposed system is better suited for synthetic data DVI scenarios compared to the conventional unimodal models and current limited-modality fusion systems.

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 comparison in Table 10 showed 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.

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	↓ 1.9	Removal of attention reduces modality weighting efficiency, leading to weaker cross-modal

			prioritization
Without Autoencoder (Direct Fusion)	91.6 ± 0.6	↓ 3.2	Loss of dimensionality reduction causes feature redundancy and weaker latent representation learning
Without Data Augmentation	90.8 ± 0.7	↓ 4.0	Reduced generalization due to lack of noise robustness and limited exposure to degraded forensic conditions

The results from the ablation study in Table 11 indicated that each architectural component of the multimodal forensic fusion framework contributes significantly to its performance. The complete model gives the best accuracy of 94.8% which shows attention mechanisms with feature compression using an autoencoder and process augmentation improve the identification performance in disaster victim identification DVI. There is a clear trend that performance gets worse as you remove parts. The most significant drop happens when the data augmentation is removed (-4.0%). That is; to make the model robust, it is crucial to expose it to synthetic noise and degraded forensic conditions. This is important to ensure robustness during actual disaster simulations [63]. A performance decline of 3.2% was observed after deletion of autoencoder. This signifies the importance of dimensionality reduction with respect to our task.

Eliminating the attention mechanism yields a diminutive but still larger decrease (-1.9%), which means that modality weighting is less critical to the capacity or capability of a model than the representation learning and data robustness mechanisms. According to the outcome, it can be concluded that proposed architecture's performance requires synergy interaction between all its components or architecture performs only using all modules together. The recent multimodal forensic AI studies demonstrate that the robust DVI systems need to be integrated architectures that augment appropriate examination and attention-based weighting and feature learning to handle noisy and incomplete forensic environments [64].

6. Conclusion

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

The integration of attention mechanism in this fusion model makes this model adaptive by assigning maximum weightage to DNA data. It deems dental and facial image data the least important. Identifying the most useful data sources makes the decision-making process more realizable for the model. Furthermore, under missing modality conditions, its performance drops slightly, indicating that the model is robust and operationally relevant for real forensic cases.

Both the ROC curve and the confusion matrix are the valid assessment tools used for sensitivity analysis study which prove superior power performance of the multi-modal fusion to a single-modality systems. The argument that multi-modal fusion models can decrease classification error rates and increase the identification of victims in mass disasters has been made. Forensic science research supports the combination of various kinds of data to achieve better or more accurate and reliable identification results.

In short, the multimodal fusion method proposed in this study has a significant impact in terms of forensic victim identification. Also, it pursues one or more goals of the SDG 3 (Good Health and Well-being) and SDG 11 (Sustainable Cities). The system helps the forensic teams to identify the victims of the disaster despite bad conditions.

7. Recommendations

There should be increased efforts in improving the 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 gets to be trained with a wider set of conditions. Not only it is helpful for the model to become more robust but also generalizable in real forensics.

Though the 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 response teams should be consulted in future studies to test this fusion system in real-life context keeping in mind the variation of quality, completeness and degradation of the modalities. This field validation will demonstrate the applicability of the model under diverse and unpredictable disaster scenarios.

Though this study's fusion model performs well, aiming to make machine learning's decision-making process more interpretable and transparent is a worthy goal. Utilization of XAI techniques can help forensic experts comprehend the rationale of system predictions. The utilization of tools like attention visualization and decision-making transparency will allow for the enhancement of user trust in the model, thus enabling the forensic system to satisfy all legal and ethical requirements, especially in sensitive cases.

Time-critical system should be optimize as per requirement related to disaster management scenarios. Future work should assess the scalability of the fusion model to allow fast and efficient processing of large quantities of forensic data without loss of performance. In mass casualty events, real-time processing is necessary so that decisions can be made quickly to ensure timely identification and assistance of survivors and victims' communities.

Ensuring the victim identification process in forensics entails a web of legal, ethical and human factors. Further research on multimodal forensic system must address such issues. Researchers should create frameworks that guard against harm to sensitive data as well as victims and their families' privacy and dignity.

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 with new data. If forensic experts and disaster responser take part in the system, it could adapt to new challenges in forensic science. In order to present accurate representations of forensic science, the developer promises regular updates to its datasets, model, and methods to maintain performance.

Future research can significantly improve multimodal forensic identification systems capabilities by addressing these recommendations. The augmentations will ensure that the system is highly efficient in disaster management and forensic science, with the operational capability to undertake the actual humanitarian and forensic operations in the field.

List of Abbreviation

Abbreviation	Full Form
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

Idowu Olugbenga Adewumi: Conceptualization, methodology, writing- original draft, supervision.

Wumi Ajayi: Methodology, validation, writing- review and editing.

Oluwafisayo Babatope Ayoade: Software, formal analysis, visualization.
Victoria Bolanle Oyekunle: Investigation, resources, data curation.
Akintayo Ayoade: Validation, project administration, writing-review and editing.
Azeez Ajani Waheed: Data curation, formal analysis, visualization.
All authors have read and agreed to the final version of the manuscript.

Funding
No funding was received for this study.

Conflicts of Interest
The authors declare no conflicts of interest related to the content of this article.

Availability of Data and Material
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.

Acknowledgments
The authors would like to acknowledge the contributions of forensic experts and researchers in the field of disaster victim identification, whose insights have greatly enriched this study. We also extend our gratitude to the institutions and professionals who provided valuable feedback during the development of this work.

AI-declaration
During the preparation of this manuscript, the authors used Grammarly to support language editing, grammatical correction, and improvement of textual clarity. The tool was not used to generate, manipulate, or analyse 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 Science International: Genetics*, vol. 69, Art. no. 102994, 2024, doi: 10.1016/j.fsigen.2023.102994.

[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,” *Applied Sciences*, vol. 14, no. 5, Art. no. 2113, 2024, doi: 10.3390/app14052113.

[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, Art. no. 713, 2025, doi: 10.3390/biom15050713.

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

[5] J. A. Sweeting, D. R. Byard, and R. W. Byard, “Disaster victim identification: forensic approaches and challenges,” *Forensic Science, Medicine and Pathology*, 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 Genomics*, vol. 26, Art. no. 748, 2025,

doi: 10.1186/s12864-025-11947-6.

[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 Science International: Genetics*, vol. 76, Art. no. 103183, 2025, doi: 10.1016/j.fsigen.2024.103183.

[8] K. E. Burger, S. Klepper, U. von Luxburg, and F. Baumdicker, "Inferring ancestry with the hierarchical soft clustering approach tangleGen," *Genome Research*, vol. 34, no. 12, pp. 2244–2255, 2024, doi: 10.1101/gr.279399.124.

[9] C. S. Heinzel, L. Purucker, F. Hutter, and P. Pfaffelhuber, "Advancing biogeographical ancestry predictions through machine learning," *Forensic Science International: Genetics*, vol. 79, Art. no. 103290, 2025, doi: 10.1016/j.fsigen.2025.103290.

[10] A. J. Jeffreys, V. Wilson, and S. L. Thein, "Individual-specific fingerprints of human DNA," *Nature*, vol. 316, pp. 76–79, 1985. doi: 10.1038/316076a0

[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 Science International*, 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," *International Journal of Legal Medicine*, vol. 139, no. 2, pp. 531–540, 2025, doi: 10.1007/s00414-024-03406-w.

[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 Science International*, vol. 365, Art. no. 112260, 2024, doi: 10.1016/j.forsciint.2024.112260.

[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 & Genomics*, vol. 47, no. 1, pp. 87–98, 2025, doi: 10.1007/s13258-024-01594-8.

[15] W. J. Pretty and D. Sweet, "A look at forensic dentistry- Part 1: The role of teeth in the determination of human identity," *British Dental Journal*, vol. 190, pp. 359–366, 2001.
doi: 10.1038/sj.bdj.4800970

[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 Science International: Genetics*, vol. 81, Art. no. 103345, 2026, doi: 10.1016/j.fsigen.2025.103345.

[17] D. A. Taylor and M. Humphries, "Simulating realistic short tandem repeat capillary electrophoretic signal using a generative adversarial network," *Expert Systems with Applications*, vol. 280, Art. no. 127536, 2025, doi: 10.1016/j.eswa.2025.127536.

[18] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997. doi: 10.1109/34.598228

[19] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701-1708. doi: 10.1109/CVPR.2014.220

[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 Science International: Genetics*, vol. 70, Art. no. 103010, 2024, doi: 10.1016/j.fsigen.2024.103010.

[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 Science International: Genetics*, vol. 77, Art. no. 103240, 2025, doi: 10.1016/j.fsigen.2025.103240.

[22] M. E. D'Amato, Y. Joly, V. Lynch, H. Machado, N. Scudder, and M. Zieger, "Ethical considerations for forensic genetic frequency databases: First report conception and development," *Forensic Science International: Genetics*, vol. 71, Art. no. 103053, 2024, doi: 10.1016/j.fsigen.2024.103053.

[23] F. D. M. Souza, H. de Lassus, and R. Cammarota, "Private detection of relatives in forensic genomics using homomorphic encryption," *BMC Medical Genomics*, vol. 17, Art. no. 273, 2024, doi: 10.1186/s12920-024-02037-9.

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

10.3389/fgene.2025.1578581.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 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," *International Journal of Legal Medicine*, vol. 139, no. 4, pp. 1835–1850, 2025, doi: 10.1007/s00414-025-03453-x.

[27] D. Michalski, C. Malec, E. Clothier, and R. Bassed, "Facial recognition for disaster victim identification," *Forensic Science International*, vol. 361, Art. no. 112108, 2024, doi: 10.1016/j.forsciint.2024.112108.

[28] T. Jiao, C. Guo, X. Feng, Y. Chen, and J. Song, "A comprehensive survey on deep learning multi-modal fusion: Methods, technologies and applications," *Computers, Materials & Continua*, vol. 80, no. 1, pp. 1–35, 2024, doi: 10.32604/cmc.2024.053204.

[29] M. Wang, S. Fan, Y. Li, and H. Chen, "Missing-modality enabled multi-modal fusion architecture for medical data," *Journal of Biomedical Informatics*, vol. 164, Art. no. 104796, 2025, doi: 10.1016/j.jbi.2025.104796.

[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 and Vision Computing*, vol. 158, Art. no. 105509, 2025, doi: 10.1016/j.imavis.2025.105509.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. doi:10.1109/CVPR.2016.90

[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," *Nigerian Journal of Clinical Practice*, vol. 27, no. 6, pp. 732–738, 2024, doi: 10.4103/njcp.njcp_787_23.

[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," *International Journal of Legal Medicine*, vol. 138, pp. 1741–1757, 2024, doi: 10.1007/s00414-024-03204-4.

[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," *Journal of Optics*, vol. 53, no. 4, pp. 3810–3820, 2024, doi: 10.1007/s12596-023-01572-5.

[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," *Scientific Reports*, vol. 14, Art. no. 19689, 2024, doi: 10.1038/s41598-024-70621-1.

[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, doi: 10.1109/ACCESS.2024.3466953.

[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," *Scientific Reports*, vol. 14, Art. no. 27974, 2024, doi: 10.1038/s41598-024-79738-9.

[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 Medical Research Methodology*, vol. 24, Art. no. 50, 2024, doi: 10.1186/s12874-024-02183-9.

[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, doi: 10.7759/cureus.73028.

[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," *International Journal of Legal Medicine*, vol. 139, no. 3, pp. 1381–1393, 2025, doi: 10.1007/s00414-025-03416-2.

[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, Art. no. 344, 2024, doi: 10.1186/s12903-024-04120-0.

[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 Computing and Applications, vol. 37, pp. 24997–25032, 2025, doi: 10.1007/s00521-025-11612-9.

[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," *International Journal of Legal Medicine*, vol. 138, pp. 2691–2706, 2024, doi: 10.1007/s00414-024-03286-0.

[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," *Computers in Biology and Medicine*, vol. 177, Art. no. 108635, 2024, doi: 10.1016/j.compbiomed.2024.108635.

[45] P. Chitrapu, M. K. Morampudi, and H. K. Kalluri, "A secure and explainable multimodal biometric system using trust adaptive fusion for face and fingerprint," *Scientific Reports*, vol. 16, Art. no. 14244, 2026, doi: 10.1038/s41598-026-43252-x.

[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," *Arabian Journal for Science and Engineering*, 2025, doi: 10.1007/s13369-025-10785-8.

[47] S. Selvarani and M. Mary Shanthi Rani, "Deep learning-powered secure multimodal biometric feature fusion with explainability and real-time deployment," in *Proc. International Conference on Artificial Intelligence and Secure Data Analytics (ICAISDA 2025)*, 2026, pp. 950–977, doi: 10.2991/978-94-6239-616-6_71.

[48] M. L. Gavrilova, "Information fusion: A decade of innovations in biometric multimodal research," *Computer*, vol. 58, no. 4, pp. 27–36, 2025, doi: 10.1109/MC.2025.3526135.

[49] H. Aldawsari and S. Alammar, "Integrating explainable AI with synthetic biometric data for enhanced image synthesis and privacy in computer vision systems," *Image and Vision Computing*, vol. 162, Art. no. 105726, 2025, doi: 10.1016/j.imavis.2025.105726.

[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," *Artificial Intelligence and Applications*, 2025, doi: 10.47852/bonviewAIA52026392.

[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 Science International: Synergy*, vol. 9, Art. no. 100554, 2024, doi: 10.1016/j.fsisyn.2024.100554.

[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 Science International*, vol. 365, Art. no. 112271, 2024, doi: 10.1016/j.forsciint.2024.112271.

[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," *Scientific Reports*, vol. 14, Art. no. 22689, 2024, doi: 10.1038/s41598-024-74703-y.

[54] A. Azizi, P. Charalambous, and Y. Chrysanthou, "DeepSafe: Two-level deep learning approach for disaster victims detection," *Virtual Reality & Intelligent Hardware*, vol. 7, no. 2, pp. 139–154, 2025, doi: 10.1016/j.vrih.2024.08.005.

[55] F. Marsico and M. Amigo, "Ethical and security challenges in AI for forensic genetics: From bias to adversarial attacks," *Forensic Science International: Genetics*, vol. 76, Art. no. 103225, 2025, doi: 10.1016/j.fsigen.2025.103225.

[56] A. F. Sobral, D. J. Barbosa, and R. J. Dinis-Oliveira, "CRISPR-Cas technology in forensic investigations: Principles, applications, and ethical considerations," *Forensic Science International: Genetics*, vol. 74, Art. no. 103163, 2025, doi: 10.1016/j.fsigen.2024.103163.

[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, doi: 10.1371/journal.pone.0311906.

[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," *Scientific Reports*, vol. 14, Art. no. 4668, 2024, doi: 10.1038/s41598-024-55497-5.

[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," *Scientific Reports*, vol. 14, Art. no. 13082, 2024, doi: 10.1038/s41598-024-63744-y.

[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," *Journal of Forensic Sciences*, vol. 69, no. 3, pp. 919–931, 2024, doi: 10.1111/1556-4029.15473.

[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 Science International*, vol. 361, Art. no. 112150, 2024, doi: 10.1016/j.forsciint.2024.112150.

[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," *International Journal of Legal Medicine*, vol. 138, no. 4, pp. 1533–1557, 2024, doi: 10.1007/s00414-024-03216-0.

[63] C. P. Pereira, "AI decision support in forensic dental age assessment: Proposed criteria for living individuals," *International Journal of Legal Medicine*, vol. 140, no. 3, pp. 1443–1450, 2026, doi: 10.1007/s00414-026-03723-2.

[64] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, no. 13, pp. 2115–2125, 2003. doi: 10.1016/S0167-8655(03)00079-5.