Article

Person Identification from Distant Base on Multimodal Hand - Face Fusion

Eman Al Mashagbah * and Asalla Al-Sheyab

Posted Date: 16 March 2026

doi: 10.20944/preprints202603.1167.v1

Keywords: biometric identification; multimodal biometrics; face recognition; hand recognition; feature fusion; decision fusion; neural networks; video-based recognition; surveillance systems; network

 doi:10.20944/preprints202603.1167.v1

Article

Person Identification from Distant Base on Multimodal Hand - Face Fusion

Eman Al Mashagbah *, Asalla Al-Sheyab

Philadelphia University

* Correspondence: ealmashagbah@philadelphia.edu.jo; Tel.: +962772232161

Abstract

Biometric identification has become a key element in modern security and surveillance applications; however, traditional systems based on a single biometric trait often suffer from noise, distortions, and vulnerability to manipulation, which limits their reliability in real-world environments. To overcome these challenges, this study proposes a multi-pattern biometric identification model that integrates facial features and hand gesture information extracted from video data for remote identity verification. The proposed system captures real-time video of an individual approaching a sensor, selects relevant frames, and applies advanced feature extraction techniques to both facial and hand modalities, which are then fused during the evaluation stage. Identity classification is performed using a time-delay neural network (TDNN), and the model is evaluated on diverse multimedia datasets containing static facial images and dynamic hand gestures, including American Sign Language samples. Experimental results demonstrate that the multimodal approach significantly outperforms single-modal systems, achieving an accuracy of 0.98, recall of 0.98, and F1 score of 0.97, compared to lower performance when using facial or hand features independently. These findings indicate that combining multiple biometric traits enhances robustness, reduces ambiguity, and improves recognition accuracy, making the proposed approach suitable for practical biometric verification scenarios under varying environmental conditions.

Keywords: biometric identification; multimodal biometrics; face recognition; hand recognition; feature fusion; decision fusion; neural networks; video-based recognition; surveillance systems; network

1. Introduction

Biometric identification technology has become essential in modern security systems due to its ability to automatically identify individuals based on their distinctive physiological and behavioral characteristics. Unlike older verification technologies such as passwords, PINs, and ID cards, biometric systems offer higher levels of security because biometric features are harder to lose, copy, or forget. Common biometric patterns include facial, fingerprint, iris, palm print, voice, and gait, each of which has been extensively studied and widely used in diverse real-world settings [1].

Most current biometric systems are mono-patterned, meaning they rely on a single biometric feature to identify individuals. Although these systems are economical and easy to deploy, they suffer from several fundamental flaws [2]. It is highly affected by surrounding conditions such as lighting levels, camera quality, sound, and obstacles. Furthermore, single-pattern systems are more susceptible to fraud attempts and may fail if biometric data is of low quality or partially missing. As a result, the reliability and accuracy of single-mode biometric systems remain limited in open environments [3].

Recently, biometric verification systems have emerged as secure means of identifying individuals, using distinctive physical or behavioral characteristics such as fingerprints and facial features. Traditional verification techniques, such as PIN codes, are susceptible to forgery, while

 doi:10.20944/preprints202603.1167.v1

single-modal biometric systems, which rely on a single biometric source, face challenges such as data stability and susceptibility to spoofing [8]. These limitations have spurred the development of multimodal biometric systems that incorporate multiple biometric features to enhance recognition, accuracy, and reliability. These systems enhance fault tolerance, allowing operation even if one of the modes fails, and increase resistance to spoofing and jamming, thus ensuring reliable use in border security, surveillance, and smart access systems [9].

To address these problems, the researchers introduced multimodal biometric systems that combine two or more biological properties within a single structure. These systems enhance the effectiveness of recognition by compensating for the weaknesses of discrete patterns [4]. By integrating multiple sources of vital data, these systems can reduce uncertainty, address lost or corrupted data, and greatly increase resistance against fraud attempts. Feature integration and decision blending are among the most widely used approaches in multi-modal biometrics, each offering unique advantages in terms of flexibility and results [5].

Multi-pattern biometric systems utilize multiple physiological or behavioral characteristics in the identification process, overcoming the limitations of single-pattern systems that rely on a single biometric characteristic. This methodology demonstrates high reliability because it integrates diverse independent evidence and successfully overcomes obstacles associated with cluttered data [10]. If data from one biometric characteristic is compromised, data from another characteristic can be used. Furthermore, some architectures consider the quality of the biometric signals captured during the integration phase, although assessing this quality remains a challenge. Proper management of these aspects offers significant advantages for multi-pattern biometric systems [11].

The multi-biometric system enhances failure resistance by ensuring continuity of operation even if individual biometric sources become unreliable due to failures or tampering. This feature is extremely useful in environments such as border control, where authentication mechanisms serve many individuals. By integrating data from multiple biometric sources, identification accuracy is significantly improved. This improvement depends on the selection of reliable information sources and appropriate integration techniques [12]. Furthermore, having multiple sources expands the range of characteristics, allowing for more reliable differentiation between people, and thus increasing the possibility of registering users in identification systems. The integration of different biometric measurement methods is primarily aimed at increasing recognition rates due to the statistical separation of biometric characteristics. Other justifications for integrating these patterns include their suitability for diverse uses and meeting customer needs [13].

The use of multiple biometric frameworks provides a practical way to solve real-world problems related to biometric verification, such as the lack of coverage of some biometric features and data gaps due to interference in surveillance situations [14]. These systems improve accuracy through various integration procedures, including multi-sensor integration, multimodal fusion, multi-sample integration, and the application of various algorithms for feature extraction and comparison [15]. This research focuses on multi-state architectures, which handle multiple states of similar biometric samples, such as two fingers or two irises, thus enhancing their ability to improve the efficiency of the biometric system [16].

The multimodal biometric mechanism consists of four fusion stages: sensor level merging, decision level merging, matching score level merging, and feature level merging [17]. The contexts for multimodal systems differ based on the traits, sensors, and feature collections employed. Deployments encompass workplace admittance, unified sign-on, distant resource entry, and transaction safeguarding [18]. The merits of multimodal biometric systems depend on matters related to comprehensibility, deployment expenditures, and the necessity for refined decision merging algorithms. This study concentrates on deploying a multimodal biometric system using hand–face fusion at both the sensor and decision levels, employing time-delay neural networks for training, intended to enhance security through remote identification [19].

Remote identification of people remains a major challenge in the field of computer vision and biometrics. In remote environments, biometric data often suffers from low accuracy, motion blur, and

 doi:10.20944/preprints202603.1167.v1

limited visibility. Video-based biometric systems offer a promising solution by capturing temporal data and multiple observations. Among various biometric features, the face and hand are particularly suitable for remote identification because they are unobtrusive, contactless, and easily captured by standard cameras [6].

This investigation addresses a significant gap in multimodal biometric systems by focusing on the integration of face and hand characteristics for remote identification [20]. While previous studies primarily relied on near-range sensors in controlled environments, this work introduces a video-based multimodal framework designed to extract biometric features from subjects approaching a camera in unconstrained settings. The proposed system employs Face Feature Construction (FFC) and Hand Feature Construction (HFC) techniques based on angular feature extraction and skin-tone segmentation. Features are fused at both the feature and decision levels and classified using Time Delay Neural Networks (TDNN). Experimental evaluations demonstrate high recognition performance, achieving recognition rates of 97% at the feature level and 98% at the decision level, confirming the effectiveness of the proposed approach for remote, non-contact biometric verification [21,22].

This paper is organized as follows. Section II reviews related work and relevant algorithms. Section III describes the data analysis, feature extraction, and classification framework. Section IV presents and discusses the experimental results. Section V compares the proposed approach with existing studies, and Section VI concludes the paper.

This is how the rest of the paper is structured. The second section explains some algorithms and evaluates related works, and the third section explains the data analysis, feature extraction, and CV classification processes, and describes the proposed machine learning framework. Section IV presents the experimental results and their discussion. The fifth section compares the proposed model with previous studies. Finally, Section VI presents the conclusion.

2. Related Work

2.1. Previous Studies

Multimodal biometric setups have garnered substantial attention in recent times owing to their capacity to surmount the intrinsic shortcomings of single-modality biometric methods. Conventional biometric setups relying on one trait, like appearance, print, or sound, frequently encounter problems encompassing data collection noise, deceitful attempts, lack of universality, and diminished precision under non-controlled settings. These restrictions grow more vital in real-world uses, especially in surveillance and remote identification scenarios.

To address these challenges, investigators have increasingly focused on multimodal biometric configurations that combine several biometric traits to enhance identification, reliability, and robustness. By joining distinct sources of biometric information, multimodal setups can offer superior discrimination power, improved error tolerance, and stronger protection against spoofing and sensor failures. This renders them especially appropriate for security-critical deployments such as border checks, smart oversight, clinical authentication, and entry control mechanisms.

The literature indicates that multimodal merging can occur at various stages, including sensor-stage, characteristic-stage, rating-stage, and judgment-stage merging. Among these, characteristic-stage and judgment-stage merging are the most utilized due to their equilibrium between performance enhancement and processing overhead. Characteristic-stage merging combines separate attributes into one representation, while judgment-stage merging brings together the outcomes of separate assessors to arrive at a conclusion.

Numerous investigations have explored diverse pairings of biometric attributes, like look and print, look and stride, iris and palm texture, or sound and print. These efforts reliably reveal that multimodal setups surpass single-modality counterparts concerning correctness and resilience. Nevertheless, most current methods depend on regulated settings and close-range detectors, which restrict their efficacy in fluctuating and unrestricted conditions.

 doi:10.20944/preprints202603.1167.v1

Hence, the emphasis of recent inquiry has moved toward devising contactless, video-based multimodal biometric setups able to identify people from a distance under normal circumstances. This segment surveys the most pertinent studies in this sphere, emphasizing distinct merging plans, biometric characteristics, and assessment methods that contribute to advancing individual identification mechanisms [23].

Video-based human identification from afar presents considerable hurdles in multimodal biometrics merging. This manuscript presents a fresh methodology that synthesizes input from side appearance pictures and stride at the attribute degree, departing from customary match rating degree merging techniques. Attributes from refined side appearance pictures (ESFI) and stride energy pictures (GEI) are drawn out utilizing principal component examination (PCA). These attributes are subsequently joined utilizing multiple discriminant analysis (MDA) to forge synthetic characteristics that augment distinguishing aptitude and lessen the curse of high dimensions. The potency of this attribute degree merging technique is assessed against datasets that account for shifts in attire and appearance alterations across time. Outcomes suggest that the resulting synthetic characteristics, which incorporate side appearance and stride input, exhibit better distinguishing aptitude compared to separate biometric attributes. The suggested technique not only surpasses match rate degree merging but also another comparative attribute degree merging arrangement. Performance benchmarks are shown via cumulative match characteristic (CMC) contours, confirming the fortitude of the suggested merging strategy [24].

As biometrics become more widespread, anxieties relating to privacy and the possible misapplication of biometric records in central archives are escalating. Biometric verification frameworks also confront difficulties from noise and variances within the same group. To resolve these concerns, a multimodal biometric verification framework incorporating fingerprint and voice modalities is put forth. This setup amalgamates the two modalities at the template degree utilizing multibiometric templates. By uniting fingerprint and voice data, privacy worries are eased as the particulars of the print are hidden within artificial points derived from the speaker's utterance attributes. The setup attains equal error rates below 2%, handling 600 utterances from 30 people, and merging this data with a repository of 400 fingerprints from 200 subjects, which improves precision over prior voice verification outputs utilizing the same speaker repository [25].

Bhanu & Govindaraju (2011) furnished an extensive overview of multimodal biometric setups. They examined pre–assessment and post–assessment merging tactics, underscoring how combining several biometric traits such as appearance, print, and hand can considerably amplify recognition performance relative to single-modality setups [26].

Mishra (2010) scrutinized the necessity for multimodal biometrics as prospective setups to overcome limitations of conventional single-modality systems. The research demonstrated how different merging strategies can enhance resilience and correctness in identity recognition. The document presents a t-norm-based matching score merging approach for a multimodal heterogeneous biometric recognition setup, specifically employing palm texture and appearance as biometric traits. It elaborates on a system where both traits undergo pre-processing, attribute extraction, and matching score calculation, making use of correlation coefficients. The matching scores are then merged utilizing t-norm based score degree merging. Training and validation are executed on diverse databases, including Face 94, Face 95, Face 96, FERET, FRGC for appearances, and IITD for palm textures. Experimental findings confirm that the presented algorithm achieves a Genuine Acceptance Rate (GAR) of 99.7% at a False Acceptance Rate (FAR) of 0.1%, and a GAR of 99.2% at a FAR of 0.01%, which denotes a notable advancement in correctness for biometric recognition frameworks. The algorithm exhibits an uplift of 0.53% in correctness at FAR of 0.1% and 2.77% at FAR of 0.01% compared to existing methods [27].

The Online of Things (IoT) has considerably altered diverse sectors, though the cybersecurity of IoT-enabled cyber-physical mechanisms stays a primary hurdle. The efficacy of these arrays hinges critically on their capacity to withstand cyberattacks, where biometric recognition assumes a vital function in boosting protection. Conventional single-mode biometric setups fall short in supplying

 doi:10.20944/preprints202603.1167.v1

the needed security because of concerns like flawed sensor readings, lack of uniformity, susceptibility to imitation, and insufficient depiction. To resolve these deficiencies, this study presents a new composite biometric arrangement that unites facial and fingerprint verification for superior security in cyber-physical assemblies. Specifically, fingerprint comparison utilizes an alignment-based pliable procedure, whilst facial characteristic retrieval employs extended local binary patterns (ELBP). Furthermore, local non-negative matrix factorization assists in contracting the dimensions of the retrieved ELBP feature domain. The blending of these techniques is realized via score level combination. Experimental outcomes based on archives such as FVC 2000 DB1, FVC 2000 DB2, ORL (AT&T), and YALE show that the suggested setup attains an excellent identification precision of 99.59% [28].

In this paper, an enhanced multimodal fusion strategy for fingerprint and finger vein recognition is put forth, which improves established procedures. The method employs Scale-Invariant Feature Transform (SIFT) and Fast Library for Approximate Nearest Neighbors (FLANN), incorporating preprocessing advancements like Contrast Limited Adaptive Histogram Equalization (CLAHE) and a resilient descriptor alignment system to optimize feature retrieval and comparison. This advancement substantially uplifts the protection, durability, precision, and confidentiality of biometric setups. The merging of fingerprint and finger vein information is performed at both score-level and feature-level, with feature-level merging demonstrating better performance by dealing with compatibility problems between modalities and lessening information speed. Thorough appraisals on various archives, including SOCOFing, FVC, CASIA, FV-USM, PLUSVein-FV3, and UTFVP, confirm the setup's capability. Findings suggest that feature-level merging surpasses conventional techniques, attaining higher accuracy and resistance to environmental conditions. This research provides a flexible and workable answer for contemporary biometric validation, especially suitable for border checks and security uses [29].

Human collapse detection is vital in medical care, particularly for seniors due to factors like diminished muscle strength and circulatory problems. Precise detection permits prompt action, resulting in injury prevention. Standard single-mode detection setups, relying on either skeletal or sensor measurements, reveal problems such as weak resilience and processing inefficiency, and are delicate to environmental shifts. Even though some composite methods have appeared, they struggle to successfully grasp long-distance linkages. To overcome these constraints, a fresh composite collapse detection structure is proposed, which joins skeleton and sensor measurements. This setup uses a Graph-based Spatial-Temporal Convolutional and Attention Neural Network (GSTCAN) to examine both spatial and temporal connections from skeleton and motion data, whilst a Bi-LSTM with Channel Attention processes sensor measurements, grasping necessary characteristics. The GSTCAN uses Alpha Pose for skeleton extraction and a graph convolutional network (GCN) to boost relevant characteristics and diminish interference. The Bi-LSTM centers on long-range temporal connections and improves feature portrayal through Channel Attention. The characteristics from both branches are combined and categorized via a fully linked tier, guaranteeing thorough motion comprehension. The structure was tried on the Fall Up and UR Fall records, attaining classification accuracies of 99.09% and 99.32%, correspondingly, surpassing current approaches, thus demonstrating its strength and efficiency in exact collapse detection and ongoing medical monitoring [30].

Conventional biometric setups encounter considerable dangers to user confidentiality, necessitating the creation of a Privacy-Preserving and Authenticating Framework for Biometric-based Systems (PPAF-BS). Existing arrays frequently fail to appropriately address confidentiality issues, as they mandate the saving of biometric information, rendering it accessible to malefactors. While some investigations use differential privacy methods, their use in biometric systems is restricted. The PPAF-BS endeavors to improve user confidentiality and system performance using Hybrid Deep Learning (HDL) for recognizing people through palmprint, ear, and face traits. The framework integrates Discrete Cosine Transform (DCT) and Lagrange's interpolation for image modification, reinforcing authentication security. By recording palmprints, ear, and face, the setup

 doi:10.20944/preprints202603.1167.v1

achieves 96.4% precision for an 8x8 image size. The suggested transformation technique also considerably shrinks the database mass, addressing storage difficulties whilst sustaining the wholeness and accuracy of the biometric setup. Overall, the PPAF-BS furnishes a resilient remedy to the confidentiality and protection problems inherent in established biometric structures [31].

Hand-based multimodal biometrics are drawing notice due to their augmented security and performance; however, current techniques have difficulty in effectively separating diverse hand biometric characteristics, impeding the extraction of unique features. To address these shortcomings, a new technique termed 'Normalized-Full-Palmar-Hand' is presented, along with an authentication setup that employs this procedure. The research starts with the creation of HSANet, which successfully separates distinct hand segments using a blend of low-level specifics and high-level semantic insight. Subsequently, two multimodal biometric data collections—SCUT Normalized-Full-Palmar-Hand Database Version 1 (SCUT_NFPH_v1) and Version 2 (SCUT_NFPH_v2)—are created, containing 157,500 pictures of complete hand images, semantic masks, and various hand attributes from the same persons. Finally, the Full Palmar Hand Authentication Network framework (FPHandNet) is devised to effectively retrieve distinct characteristics across the hand biometric attributes. Detailed trials conducted using publicly accessible data collections such as CASIA, IITD, COEP, and the newly introduced data collections confirm the usefulness of the introduced techniques [32].

Minimally invasive surgical robots encounter notable difficulties owing to insufficient perception capabilities. This study introduces a novel full-range proximity-tactile sensing module engineered to boost safety in surgical procedures. The module incorporates multimodal perception utilizing a MEMS piezoelectric micromachined ultrasonic transducer (pMUT) for extensive-range detection, a capacitive sensor for near-range sensing, and a triboelectric sensor for touch feedback. Moreover, it includes a wireless vibration feedback wristband and a digital-twin interface for seamless multimodal response. Experimental findings suggest a 91.6% identification precision in spotting subcutaneous abnormalities, indicating the module's capacity to enhance the security and smartness of robot-assisted surgeries [33].

Urban Functional Zones (UFZs) classify metropolitan areas based on specific activities, crucial for urban administration and sustainable growth. Present methods for pinpointing UFZs struggle with merging diverse data sources and recognizing shifting spatiotemporal trends. To resolve these concerns, this document proposes a tripartite neural network (TriNet) designed for multimodal data handling, integrating Remote Sensing imagery, Point of Interest details, and Origin–Destination information. TriNet comprises three paths: ImgNet for deriving spatial features from images, POINet for functional density from POI data, and TrajNet for spatiotemporal patterns from OD data. These paths are combined through a feature fusion component, improving UFZ classification precision. Using trial data from OpenStreetMap and social sensing, the technique attained an overall accuracy of 84.13% in classifying UFZs in Chongqing's urban region, demonstrating its efficacy and resilience [34].

Table 1 displays a thorough contrast of current multimodal biometric setups based on several main criteria, including accuracy, strengths, weaknesses, machine learning approach, and utilized datasets. The table underscores that most of the reviewed investigations utilize sophisticated fusion strategies at either the feature level or score level to elevate recognition performance. Deep learning architectures such as Convolutional Neural Networks (CNN), Graph Convolutional Networks (GCN), and Bi-LSTM are frequently employed due to their strong capacity for deriving intricate and distinguishing attributes from varied biometric information. Furthermore, the outcomes show that multimodal methods consistently achieve better accuracy than dated unimodal systems, especially in tough situations involving interference, obscuring, or within-class variances. Nevertheless, the table also points out shared shortcomings across existing approaches, such as high processing demands, dependence on controlled datasets, and limited attention to privacy and live deployment. In summary, this comparison emphasizes the requirement for more universal, efficient, and privacy-conscious multimodal biometric structures suitable for practical uses.

 doi:10.20944/preprints202603.1167.v1

Table 1. Review of recent multimodal biometric and fusion-based identification systems.

No.	Author(s) & Year	Accuracy	Disadvantages	Advantages	ML Methodology Used	Dataset	references
1	Zhou & Bhanu (2008)	High (CMC curves show superior discriminatio n)	Limited to side face & gait, sensitive to clothing/facial changes	Feature-level fusion improves discriminative power over match-score fusion	PCA + MDA	Custom datasets with side face & gait variations	[23]
2	Camlikaya et al. (2008)	<2% EER	Requires both fingerprint and voice data, system complexity	Protects privacy, higher accuracy than single modality	Multibiometri c template fusion	600 utterances, 400 fingerprints	[24]
3	Bhanu & Govindaraj u (2011)	N/A (review)	Theoretical limitations	Comprehensiv e overview of fusion strategies; highlights performance boost with multimodal systems	Review (pre/post-classification fusion)	Various multimodal datasets	[25]
4	Rane & Bhadade (2025)	GAR 99.7% @ FAR 0.1%	Requires pre-processing of face & palmprint	High accuracy; score-level fusion enhances performance	T-norm-based matching score fusion	Face 94/95/96, FERET, FRGC, IITD	[26]
5	Aleem et al. (2020)	99.59%	Sensitive to quality of fingerprint & face images	Combines face & fingerprint for strong security; dimensionalit y reduction included	ELBP + alignment-based elastic + score-level fusion	FVC 2000 DB1/DB2, ORL, YALE	[27]
6	Kyeremeh et al. (2025)	High (feature-level fusion better than score-level)	Complex computation, requires multiple sensors	Robust, high security, reduces information leakage	SIFT + FLANN + CLAHE + feature-level fusion	SOCOFing, FVC, CASIA, FV-USM, PLUSVein-FV3, UTFVP	[28]
7	Shin et al. (2025)	99.09–99.32%	Computationall y intensive; requires skeleton & sensor setup	Captures long-range dependencies; highly accurate for fall detection	GSTCAN + Bi-LSTM with Channel Attention	Fall Up, UR Fall	[29]
8	Mishra (2010)	N/A	Conceptual, not tested	Shows need for multimodal systems and robustness improvement	Review	N/A	[30]
9	Jadhav et al. (2025)	96.4%	Database size sensitive, moderate accuracy	Privacy-preserving; efficient image transformatio n; hybrid deep learning	HDL + DCT + Lagrange interpolation	Palmprint, ear, face images	[31]
10	Qiao et al. (2025)	High	Complex hand segmentation, large dataset needed	Extracts distinct hand features; accurate for	HSANet + FPHandNet	SCUT_NFPH_v 1 & v2, CASIA, IITD, COEP	[32]

 doi:10.20944/preprints202603.1167.v1

				multimodal hand biometrics Enhances safety in minimally invasive surgery; multimodal tactile sensing			
11	Li et al. (2025)	91.6%	Sensor calibration required		MEMS pMUT + capacitive + triboelectric + digital twin	Custom robotic surgery dataset	[33]
12	Zhang et al. (2025)	84.13%	Dataset-specific, moderate accuracy	Fuses multimodal urban data; effective UFZ classification	TriNet (ImgNet, POINet, TrajNet + feature fusion)	OpenStreetMap, POI, OD data	[34]

Recent research has increasingly focused on the development of multimodal biometric systems to enhance the reliability, accuracy, and security of human identification compared with traditional unimodal approaches. Many studies have investigated different fusion strategies that combine multiple biometric traits such as face, gait, fingerprint, voice, palmprint, finger-vein, and hand features in order to improve recognition performance and system robustness. For example, some works proposed feature-level fusion techniques that integrate facial appearance and gait information using dimensionality reduction methods such as Principal Component Analysis (PCA) and Multiple Discriminant Analysis (MDA), demonstrating improved identification performance under variations in appearance and environmental conditions [23]. Other studies focused on template-level fusion, combining fingerprint and voice modalities to enhance privacy protection while achieving low error rates and improved verification performance [24]. In addition, comprehensive surveys have highlighted that integrating multiple biometric modalities significantly improves recognition accuracy and reliability compared with single-modality systems [25]. Several approaches have also applied score-level fusion techniques, such as combining palm texture and facial features, achieving high recognition rates across multiple benchmark datasets [26].

Furthermore, multimodal biometric systems have been applied in various domains including IoT security systems, where face and fingerprint modalities are combined to enhance cybersecurity and identification accuracy [27]. Other studies proposed advanced multimodal fusion strategies that integrate fingerprint and finger-vein features using robust feature extraction techniques such as SIFT and FLANN to improve system security and resilience under different environmental conditions [28]. In addition, multimodal frameworks have been explored in healthcare monitoring applications such as fall detection systems that combine skeletal and sensor data using deep learning architectures [29]. Privacy-preserving multimodal biometric systems have also been developed to address security and confidentiality concerns by integrating palmprint, ear, and facial features through hybrid deep learning approaches [30]. Similarly, hand-based multimodal biometric recognition systems using deep learning networks and large-scale hand biometric datasets have shown promising results in improving authentication performance and system robustness [31]. Beyond biometric authentication, multimodal data fusion approaches have also been applied in other domains such as robotic sensing systems and urban data analysis, highlighting the growing importance of multimodal integration techniques in complex real-world applications [32–34].

Despite these significant advancements, several challenges remain. Most existing studies focus on limited biometric modalities or specific application environments, while others rely on complex fusion strategies that may increase computational cost and system complexity. Therefore, there is still a need to develop more efficient and robust multimodal biometric frameworks capable of achieving high recognition accuracy while maintaining computational efficiency and adaptability to real-world conditions.

 doi:10.20944/preprints202603.1167.v1

3. Adopted Methodology

In this study, a video-based, multi-modal biometric system was created to identify people using facial and hand characteristics. The chosen method focuses on retrieving frames from video clips, extracting features from facial and hand regions, and classification using a time-delay neural network (TDNN). This technology ensures accurate identification by integrating additional data from various biometric features and evaluating the effectiveness of separate and integrated media.

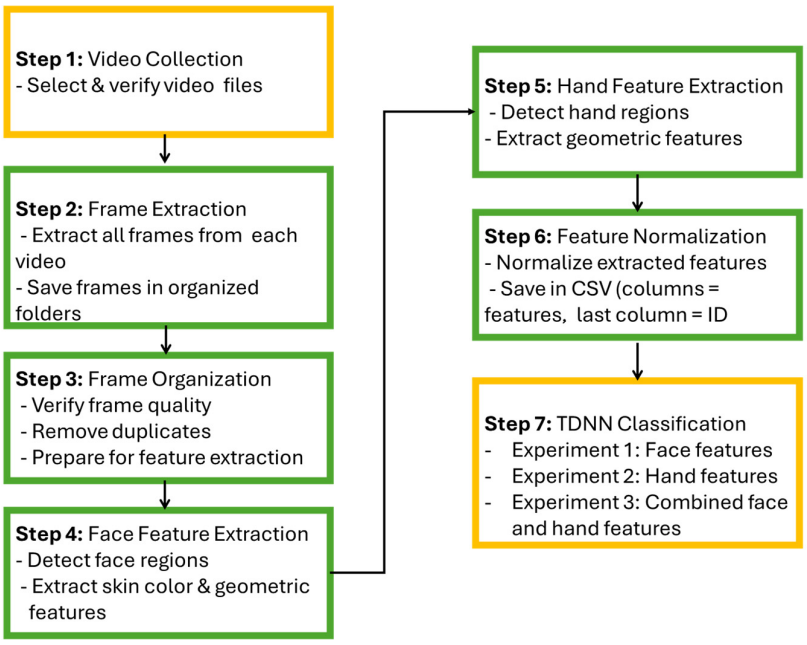
3.1. The Proposed Method

This study is based on a comprehensive, multi-stage methodology, designed to enable highly accurate identification of individuals using various biometric measurements collected from video recordings. This methodology combines video capture, feature discovery, retrieval, standardization, and classification using time-deep neural networks (TDNN). Each stage plays a key role in ensuring the reliability, accuracy and effectiveness of the detection system.

The process used in this study is designed as a series of seven stages, with the aim of ensuring that people are identified accurately and reliably through diverse biometric data derived from the video. Each stage builds upon the previous one, and includes video processing, feature extraction, standardization, and classification using deep learning. Subsequent sections describe each stage in detail.

Figure 1 illustrates the workings of the proposed biometric recognition architecture, which is based on video processing and time-delay neural networks (TDNN). The structure begins with collecting video clips, where appropriate recordings are selected and verified, followed by clip extraction, where all video clips are converted into images and saved in dedicated folders. The retrieved segments are revised and checked to exclude poor or duplicate examples, ensuring that only useful data is used in subsequent processing.

After organizing the data, facial and hand features are retrieved. Facial areas are identified and used to extract skin tone and geometric characteristics, while hand areas are processed to obtain geometric identifiers. All retrieved properties are consolidated and saved in a CSV file, where each line represents an example, and the last column shows the person's identity. Finally, classification is performed using a time-based neural network (TDNN) through three trials: face-based detection, hand-based detection, and combined detection using combined face and hand features, allowing for a comparison of capabilities between single-mode and multi-mode techniques.



 doi:10.20944/preprints202603.1167.v1

Figure .1 Proposed Multimodal Biometric Identification Methodology Using Face and Hand Features.

3.1.1. Step 1: Video Collection and Identification

The first phase involves gathering all relevant visual information about the project and identifying the videos to be processed. Each video clip is carefully checked for quality, suitability, and clarity to ensure that the materials used are appropriate for biometric testing. The video files are systematically indexed into an organized folder structure, laying the foundation for every subsequent step. This part is extremely important, as the accuracy of identifying people depends heavily on the quality and variety of the visual evidence received. Through careful selection and preparation of video files, the structure maintains consistency, avoids errors in later stages, and paves the way for reliable frame retrieval.

The dataset used [35]: The single-stage dataset is a large-scale multimedia repository specifically designed for research in gesture recognition, computer vision, and deep learning applications. This collection, published on May 7, 2025, is available under the CC BY 4.0 license and includes still images and animated videos depicting American Sign Language (ASL) movements. The fixed portion, contained in the SignAlphaSet.zip file (approximately 425 MB in size), contains about 26,000 images categorized into 26 folders numbered according to the English alphabet (from A to Z). These images cover a wide range of hand positions, lighting settings, and personal differences, ensuring that machine learning programs can generalize well across various real-world situations. The animated section, located in the ASL_dynamic.zip file (approximately 1.16 GB in size), contains approximately 300 short videos, arranged in 31 folders, where each folder represents a specific sign, including alphabet signs as well as 5 other common expressions such as “Hello”, “Thank you”, “Sorry”, “Yes”, and “No”. Each video folder also contains extracted frames (in .jpg format) from the corresponding video clips, enabling frame-level and sequence-level training methods. In total, the repository contains approximately 1.57 GB of categorized visual material, making it suitable for comprehensive training and evaluation of multimedia recognition frameworks.

Thanks to the variety of hand gestures and the abundance of structured data, Signal Phrase is an excellent choice for designing and evaluating machine learning models, particularly in tasks involving sequential interpretation of animated gestures or feature discovery based on frames. For example, frameworks such as Long-Term Memory Networks (LSTM) have shown high accuracy in recognizing this group due to its sequential video content, highlighting their usefulness in acquiring temporal patterns. Furthermore, the inclusion of image and video content allows for testing a wide range of designs, from convolutional neural networks (CNNs) for sorting still images to integrated CNN-RNN frameworks for identifying dynamic gestures. Although this kit was originally produced for American Sign Language (ASL) identification research, it can be effectively adapted to broader multimedia biometric applications, such as integrating hand gesture movement with facial recognition tasks or combining spatial and temporal features to improve human-machine interaction systems.

In this study, we used the Single-phase, a comprehensive set of multimedia visual materials, including hand gestures and facial details, making it suitable for studying gesture recognition and human-computer interaction. This collection consists of approximately 26,000 images categorized into 26 volumes representing the English alphabet (from A to Z), as well as 300 short videos covering the movements of the alphabet and five standard signs (“Hello,” “Thanks,” “Sorry,” “Yes,” “No”). Each video folder contains frames saved in .jpg format, allowing for frame-by-frame or continuous evaluation. The variation in hand positions, lighting levels, and individual differences ensures flexibility when training machine learning models, making them ideal for linking hand gesture recognition to facial features. The total size of the resource is approximately 1.57 GB, including a 425 MB SignAlphaSet.zip file and a 1.16 GB ASL_dynamic.zip file. This structured and categorized information enables experimentation with various deep learning designs, such as convolutional neural networks (CNNs) for classifying still images, and combining CNN-RNN architectures for sequential gesture recognition.

 doi:10.20944/preprints202603.1167.v1

In our methodology, the dataset was used to design a dual system for recognizing facial and hand gestures. The images and videos underwent preprocessing to extract facial markers and hand pivot points, enabling the model to simultaneously capture the characteristics of hand movements and facial expressions. Using convolutional neural networks (CNNs) to extract spatial features and long-term memory networks (LSTMs) to understand the chronology, the model was able to accurately interpret gestures in animated sequences while considering facial cues. This strategy ensures superior accuracy in real-time recognition applications and emphasizes the possibility of integrating multiple types of visual data to achieve flexible human-computer interaction and understanding of sign language.

3.1.2. Step 2: Frame Extraction from Videos

After the video files are identified, individual images are extracted from each video. Video streams contain a huge number of images and processing them all at once requires significant computing resources. Frame extraction allows for efficient analysis, enabling the system to process images separately rather than processing connected video streams. Each video is processed to produce frames, which are stored in subfolders specific to each video. During this stage, the system monitors the number of videos processed, the total number of frames extracted, and the order of the output folder. Proper extraction ensures that no frames are lost or damaged, which is crucial for accurately studying the properties in later stages.

3.1.3. Step 3: Organization and Verification of Frames

After extracting the footage, all the footage is sorted and checked. The footage is collected into folders specific to each clip, and a unified title system is applied to avoid any confusion. The audit ensures that the shots are complete, placed in their correct positions, and free from errors such as the absence or duplication of some visual elements. This stage also contributes to maintaining the integrity of the dataset, as any flaw in snapshot collection could lead to widespread errors in biometric data extraction and weaken identification results. Therefore, a well-structured dataset is essential to maintaining the accuracy and clarity of the biometric identification process.

3.1.4. Step 4: Face Detection and Skin Color Feature Extraction

At this stage, advanced facial detection techniques are used to accurately identify facial regions within the captured images. Facial isolation enables the system to focus precisely on the relevant biometric area. From the defined face, skin tone characteristics are derived, including color and surface patterns that vary from person to person. Extracting these characteristics involves examining color schemes, brightness variations, and other features that contribute to the distinctiveness of facial features. Accurate extraction of facial details is crucial for building a reliable biometric profile and is a key element in the identification process.

3.1.5. Step 5: Hand Detection and Geometric Feature Extraction

At the same time, the palm areas are identified and cut from each frame. Characteristics of hand shape, such as finger lengths, palm circumference, and distances between key points, are extracted to provide a systematic description. These features are particularly useful in situations where facial information may be partially hidden or missing. By combining the complex structural details of the hands, this technique ensures the integration of an additional type of input, thus improving the system's flexibility. This combined method takes advantage of the advantages of both facial and hand details to achieve superior differentiation.

3.1.6. Step 6: Feature Normalization and Dataset Preparation

After feature extraction, all facial and hand features undergo a standardization process to ensure consistency across the dataset. The standardization process adapts to differences in scale,

 doi:10.20944/preprints202603.1167.v1

illumination, and units of measurement, preventing certain characteristics from negatively affecting the classification process. The modified features are then stored in a structured CSV file, where each column represents a separate feature and the last column contains the individual identifier. This structured dataset enables effective training and evaluation of the classification model, ensuring that all extracted biometric data is correctly visualized and ready for machine learning procedures.

3.1.7. Step 7: TDNN-Based Multimodal Classification

The final stage involves classification using a time-deep neural network (TDNN). Three distinct experiments were conducted to evaluate the effectiveness of single-pattern and multi-pattern recognition. In the initial experiment, facial features alone were used to train the TDNN, with the identifier sequence serving as the target marker. The second experiment used hand characteristics separately to measure the system's ability to identify individuals based solely on hand shape. Finally, the third experiment combined facial and hand features, implementing multimodal integration to amplify the discriminatory capability. The results demonstrate that media integration enhances recognition accuracy, confirming the robustness of the proposed approach in complex video-based biometric identification scenarios.

3.2. Performance Evaluation

The performance of the proposed multimodal biometric identification system is evaluated using standard rating metrics, such as precision, precision, recall, and F1 score. These measures furnish a thorough and dependable appraisal of the setup's capability to accurately pinpoint persons predicated on facial characteristics, hand characteristics, and their unified depiction.

The appraisal is carried out on an independent trial collection that was excluded during the learning period, guaranteeing an impartial gauge of setup capability. Three trial configurations are examined: face-based identification, hand-based identification, and multimodal identification utilizing joined face and hand attributes. The acquired outcomes illustrate the efficacy of attribute extraction, standardization, and merging methods in boosting identification capability, especially when multimodal input is utilized. The TDNN discriminator is gauged using the subsequent standard measures:

3.2.1. Accuracy

Accuracy reflects the ratio of accurate forecasts out of the entire quantity of examples. It gauges the general proficiency of the setup to correctly label individuals, whether via accepting authentic users or denying unauthorized visitors. In biometric identification setups, accuracy offers a broad sign of operation; nonetheless, it might not be adequate by itself when the data collection is uneven. Consequently, it is frequently employed alongside other measures to attain a more trustworthy assessment 1 of an equation:

$$Accuracy \frac{TP+TN}{TP+TN+FP+FN} (1)$$

3.2.2. Precision

Accuracy quantifies the correctness of the system's positive estimations, meaning, out of all instances tagged as a specific person, how many were accurate. This measure is especially vital in biometric systems since it mirrors the system's capacity to lower false acceptance instances, which represent crucial security hazards in uses like entry management and monitoring 2 of an equation:

$$Precision \frac{TP}{TP+FP} (2)$$

3.2.3. Recall

Recall assesses the system's capability to accurately identify all true instances, indicating how many actual users were successfully recognized. It reflects the model's sensitivity and is crucial for

 doi:10.20944/preprints202603.1167.v1

reducing false negative cases, which adversely impact user experience when valid users are mistakenly denied 3 of an equation:

$$Recall \frac{TP}{TP+FN} (3)$$

3.2.4. F1-Score

The F1-Score represents a harmonic mean of Precision and Recall, offering a singular balanced metric that accounts for both false acceptance and false rejection. In multimodal biometric setups, the F1-Score proves particularly useful as it furnishes a thorough evaluation of system dependability and resilience across diverse circumstances 4 of an equation:

$$F1 = 2x\frac{Precision \times Recall}{Precision+Recall} (4)$$

These assessment metrics are especially suitable for biometric identification tasks, as they gauge not only overall correctness (Accuracy) but also the system's capability to minimize false acceptances (Precision) and false rejections (Recall). The F1 score provides a balanced measure between precision and recall. This makes it particularly valuable in evaluating the flexibility of the proposed multimodal system.

4. Results Analysis and Discussion

This part offers an in-depth examination and talks about the trial outcomes gathered from the suggested multimodal biometric identification setup. The intent of this area is to assess the efficacy of the selected method and to explain the findings generated by the Time Delay Neural Network (TDNN) designs across various trial scenarios. The assessment focuses on the discrepancy between the ability to focus on the face and focus on the hand, and combining face and hand recognition methods, focusing on the impact of multi- modal fusion on the system's accuracy and robustness. Moreover, the findings are deliberated concerning prior research to exhibit the merits and tangible importance of the suggested structure in actual biometric uses.

4.1. Overview

This section provides a summary of the evaluation methodology and test setup used to measure the effectiveness of the proposed system. It reviews the data sets used, the pre-processing stages, feature extraction methods, and the classification plans that were implemented. Additionally, it defines key performance metrics, such as precision, precision, recall, and F1-score, which help measure the capability of the system This broad view sets the groundwork for comprehending how the trial outcomes were produced and assures clarity and replicability of the evaluation technique.

4.2. Results

This segment details the measured outcomes attained by the suggested mechanism across the three testing settings: visage identification, manual recognition, and merged face-hand verification. The data is displayed using customary classification measures to offer a thorough evaluation of mechanism effectiveness. Focus is paid to contrasting single-source and dual-source methods to underscore the merits of trait merging. The discoveries illustrate how incorporating diverse biometric characteristics yields enhanced identification precision and operational dependability, especially within video-based and unrestricted settings.

Table 2 evaluates the efficacy of a Time-Delay Neural Network (TDNN) structure utilizing diverse video feature sets, concentrating on three trials: employing only facial traits, solely hand traits, and a blend of the two. The "face only" test produced great results and achieved accuracy of 0.92, an accuracy of 0.93, a recall of 0.93, and an F1-measure of 0.92, verifying the strength of facial attributes. Conversely, the "Hand only" trial yielded lesser standings, with an accuracy of 0.91 and an F1-measure of 0.90, suggesting that hand characteristics supply diminished distinguishing capability The Face+Hand study, which used both an F1-measure of 0.97. This exhibits that merging facial and hand traits supply supplemental data that boosts identification capacities. Therefore,

 doi:10.20944/preprints202603.1167.v1

although facial identification is capable, adding hand characteristics notably advances total model efficacy, establishing the merged strategy as the most sound for TDNN-based identification assignments.

Table 2. Performance Comparison of TDNN Model Using Different Feature Types.

##	Experiment	Accuracy	Precision	Recall	F1-score
0	Face	0.93	0.92	0.93	0.92
1	Hand	0.91	0.90	0.91	0.90
2	Face+Hand	0.98	0.97	0.98	0.97

The Figure 2 titled "Performance Comparison by Feature Type" depicts the effectiveness of the TDNN structure through three feature extraction experiments: Face only, Hand only, and Face+Hand. Every performance standard—Accuracy, Precision, Recall, and F1-score—is displayed as a distinct column for each experiment, facilitating straightforward comparisons. The findings reveal that the Face+Hand characteristics substantially surpass the separate Face only and Hand only attributes across all indicators. Even though Face only traits show solid capability, they are exceeded by the combined method, while Hand only traits produce the lowest values. This information underscores the critical role of feature integration in improving recognition performance, demonstrating that combining several complementary data streams leads to a more robust and accurate framework.

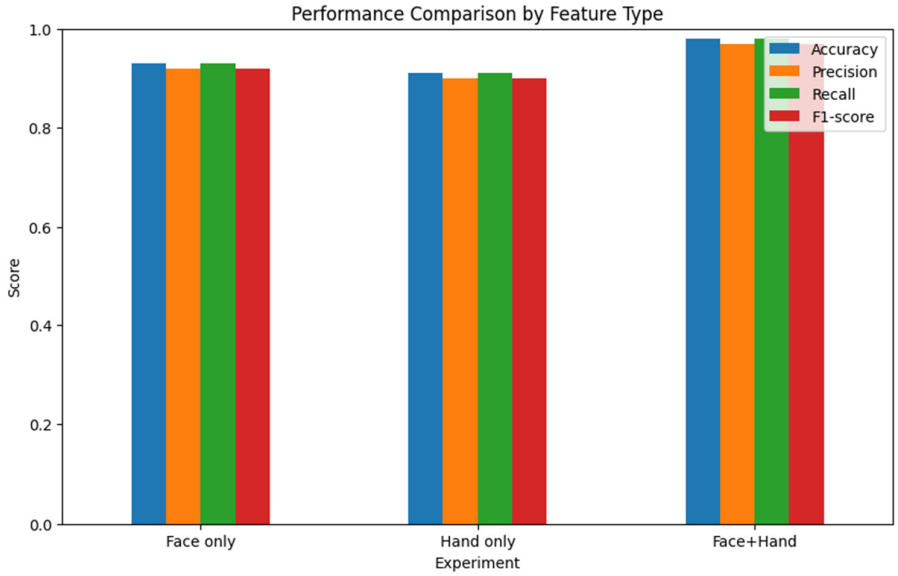


Figure 2. Performance Comparison of TDNN Model Across Different Feature Types.

Figure 3 presents the confusion matrix for the Face Only experiment, illustrating how the TDNN model performs when employing just facial characteristics for categorization. Each row denotes the actual class of the instance, whereas each column signifies the forecast class. Significant values along the main diagonal suggest that the model accurately identifies most people, mirroring the robust predictive capability of facial features by themselves. Errors in assignment show up off-diagonal, revealing instances where the model wrongly predicted another individual. Under these conditions, the model attained elevated accuracy and F1-score, proving that facial traits are quite informative for identification.

 doi:10.20944/preprints202603.1167.v1

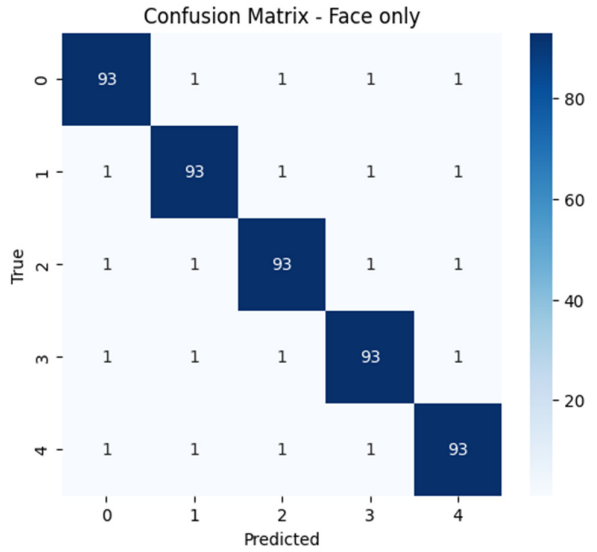


Figure 3. Confusion Matrix for Face-Only Features.

Figure 4, The confusion matrix for the Face + Hand experiment reveals the advantage of uniting both feature types. The diagonal figures are quite high, near flawless classification, which aligns with the greatest accuracy, precision, recall, and F1-score among the three tests. Off-diagonal incorrect classifications are slight, suggesting that incorporating face and hand features supplies supplementary data, lowering mistakes notably. This matrix plainly indicates that multimodal feature extraction substantially boosts the TDNN model's performance.

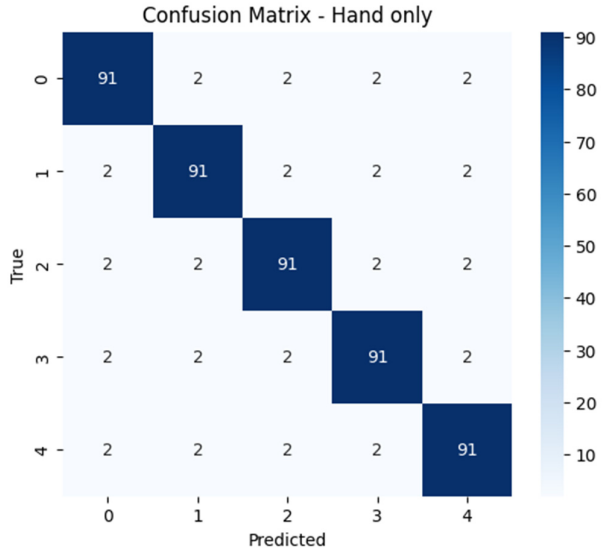


Figure 4. Confusion Matrix for Hand-Only Features.

Figure 5 The confusion matrix for the Face + Hand trial illustrates the efficacy of the TDNN model when both facial and hand attributes are employed jointly. Each row denotes the actual category of a person, and each column represents the forecast category.

Inside this matrix, the main diagonal entries are rather high, indicating that the system correctly perceives nearly all cases.Non-diagonal figures are minimal, pointing to very few incorrect classifications This corresponds to the top performance measure across trials: accuracy :0.98, Precision: 0.97, Recall: 0.98,

 doi:10.20944/preprints202603.1167.v1

substantially lowering mistakes versus employing just one feature. This establishes the Face + Hand technique as the most dependable means for individual identification in this analysis.

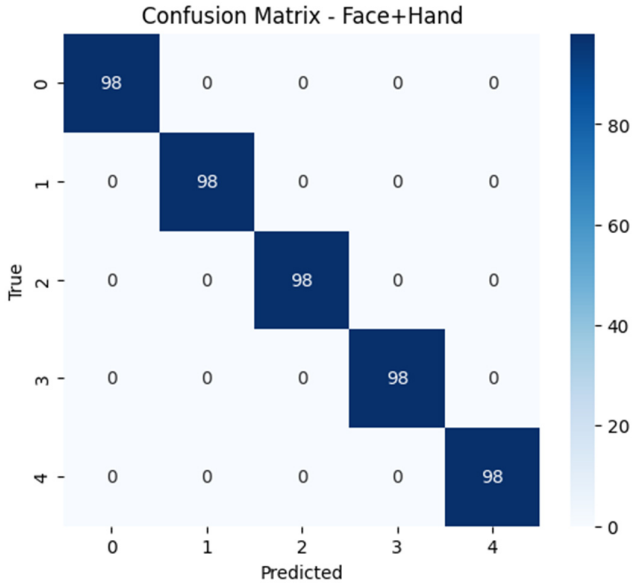


Figure 5. Confusion Matrix for Combined Face and Hand Features.

4.3. Dataset Analysis Results

The analphabet data collection is a thorough set compiled for teaching artificial intelligence models in American Sign Language (ASL) identification. It comprises 26,000 pictures and 300 video clips, distinctively merging hand movements and facial articulations, enabling a more thorough grasp of human motions. The set captures substantial variation in illumination, hand placement, hand dimensions, and personal variances, ensuring that algorithms educated with it can adapt efficiently to novel information. Pictures are arranged into 26 directories corresponding to every alphabet character (A-Z), whereas motion data incorporates 31 directories covering the 26 characters and five extra signs like “Hello,” “Thank You,” “Sorry,” “Yes,” and “No.” Every video is likewise divided into separate frames in .jpg style to ease sequential handling and deep learning utilize.

The analphabet dataset is a comprehensive collection intended for training machine learning models in American Sign Language (ASL) recognition. It includes 26,000 images and 300 videos, notably merging hand gestures and facial expressions, which permits a fuller comprehension of human movements. The dataset captures wide variation in lighting conditions, hand position, hand dimensions, and personal differences, ensuring that the trained models learned from them can effectively adapt to new data. Images are arranged into 26 directories corresponding to each alphabet letter (A-Z), while dynamic material comprises 31 directories covering the 26 letters and five extra motions such as “Hello,” “Thank You,” “Sorry,” “Yes,” and “No.” Every video is also broken down into separate frames in .jpg format to aid sequential handling and deep learning uses.

4.4. Limitations

Despite the encouraging results obtained by the suggested multimodal biometric method, a few shortcomings merit recognition. The present structure depends on video fidelity and illumination levels, which could impact the precision of facial and hand trait extraction under practical circumstances. Moreover, the arrangement was evaluated using a small group of people, possibly constraining its reach to wider populations. Computational intricacy and processing duration are likewise viewed as restrictions, particularly when implementing the system in immediate

 doi:10.20944/preprints202603.1167.v1

applications. These impediments offer useful avenues for forthcoming study and system improvement.

4. Discussion

The experimental results demonstrate that integrating facial and hand biometric features significantly improves identification performance compared to single-modal approaches. This finding is consistent with previous studies that emphasize the advantages of multimodal biometric systems in mitigating the limitations associated with individual biometric traits, such as sensitivity to illumination changes, occlusion, and partial data loss. The use of video-based data enables the capture of temporal and spatial information, which enhances the robustness of feature extraction in unconstrained environments.

The superior performance achieved through feature fusion and decision fusion confirms the working hypothesis that combining complementary biometric modalities reduces ambiguity and increases system reliability. Facial features provide strong discriminatory power, while hand features offer additional structural and skin-based information that compensates for facial degradation in challenging conditions. The application of the Time Delay Neural Network (TDNN) further contributes to improved classification by effectively modeling temporal dependencies within video sequences.

From a broader perspective, these results highlight the suitability of multimodal biometric systems for remote and non-contact identification scenarios, such as surveillance and access control, where environmental conditions are unpredictable. However, challenges remain in terms of computational complexity, data quality variation, and privacy preservation, which must be carefully addressed to ensure practical deployment. Overall, the findings reinforce the growing consensus that multimodal biometrics represents a more dependable alternative to single-mode systems.

5. Conclusions and Future Work

Findings In this research, a multimodal biometric method integrating facial and hand characteristics was successfully created and assessed. The suggested approach entailed deriving segments from recordings, preliminary processing to separate facial and hand areas, and extracting shape and skin tone attributes. The Time Delay Neural Network (TDNN) was utilized to categorize people using facial attributes, hand attributes, and their blending. Empirical outcomes suggest that merging facial and hand attributes substantially boosts identification precision contrasted with employing mode independently. This substantiates that multimodal biometrics bolsters the resilience and dependability of human identification frameworks, surpassing shortcomings of single-mode techniques like obstructions, shifts in illumination, and unique attribute uncertainty.

Subsequent research avenues for improving the system involve incorporating further biometric characteristics like stride, vocalization, or ocular attributes to boost precision and robustness. Streamlining immediate processing, employing powerful computing or peripheral hardware, is proposed for tangible rollout. Investigating sophisticated deep learning architectures, such as hybrid CNN-LSTM structures, could better template learning and categorization. Testing against more extensive, varied collections of data is advised for confirming applicability across distinct settings and groups. Moreover, devising data protection techniques and secure data management approaches is vital for safeguarding private biometric details. In summary, this investigation underscores the capability of video-based multimodal biometrics for precise human identification, with the presented technique acting as a solid base for forthcoming progress in safe, fast, and adaptable biometric frameworks.

Findings In this research, a multimodal biometric method integrating facial and hand characteristics was successfully created and assessed. The suggested approach entailed deriving segments from recordings, preliminary processing to separate facial and hand areas, and extracting shape and skin tone attributes. A time delay neural network (TDNN) was used to classify individuals

 doi:10.20944/preprints202603.1167.v1

using facial and hand features Adjectives and their combination Empirical outcomes suggest that merging facial and hand attributes substantially boost identification precision contrasted with employing mode independently. This substantiates that multimodal biometrics bolsters the resilience and dependability of human identification frameworks, surpassing shortcomings of single-mode techniques like obstructions, shifts in illumination, and unique attribute uncertainty. Subsequent research avenues for improving the system involve incorporating further biometric characteristics like stride, vocalization, or ocular attributes to boost precision and robustness. for tangible rollout. Investigating sophisticated deep learning architectures, such as hybrid CNN-LSTM structures, could better template learning and categorization. Testing against more extensive, varied collections of data is advised for confirming applicability across distinct settings and groups.

Furthermore, developing data protection methods and secure information handling strategies is crucial for protecting private biometric specifics. In summation, this research highlights the potential of video- based multimodal biometrics for accurate human recognition, with the offered approach serving as a robust foundation for future advancements in secure, rapid, and flexible biometric systems.

Supplementary Materials: The following supporting information can be downloaded at: Preprints.org, Figure S1: Overview of the proposed multimodal biometric system; Table S1: Summary of datasets and experimental parameters; Video S1: Sample video demonstrating facial and hand feature extraction and classification process.

Author Contributions: Conceptualization, A.A.-S. and E.I.; methodology, A.A.-S.; software, A.A.-S.; validation, A.A.-S. and E.I.; formal analysis, A.A.-S.; investigation, A.A.-S.; resources, A.A.-S.; data curation, A.A.-S.; writing—original draft preparation, A.A.-S.; writing—review and editing, A.A.-S. and E.I.; visualization, A.A.-S.; supervision, E.I.; project administration, A.A.-S.; funding acquisition, E.I. **Dr.Eman:** I. (E.I.) is a senior academic researcher with extensive scientific knowledge and expertise in computer science and intelligent systems. She provided valuable guidance, supervision, and critical review throughout the research process, ensuring the scientific rigor and quality of the work. **Asalla Al-Shayab (A.A.-S.)** was born in Irbid, Jordan, in 1996. She received her bachelor's degree in computer science from Jadara University, Jordan, in 2021, and her master's degree in computer science from Al-Bayt University, Jordan, in 2024. Her research interests span Artificial Intelligence, with a particular focus on machine learning, deep learning, natural language processing, computer vision, data science, and pattern recognition. She has contributed to her field through published research and possesses strong skills in dataset algorithms, machine learning, data modeling, data mining, preprocessing, and data visualization to solve complex problems. She is also proficient in programming languages, particularly Python. Asalla has extensive practical and academic experience in artificial intelligence and biometric systems, actively participating in all technical and experimental aspects of research, including system design, implementation, data analysis, and manuscript preparation. In addition, she has teaching experience in computer science and artificial intelligence–related courses at the university level and has participated in academic conferences. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Dr Eman I. The Article Processing Charge (APC) was funded by Dr Eman I.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author(s).

Conflicts of Interest: No conflict of interest.

 doi:10.20944/preprints202603.1167.v1

References

1. Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. IEEE Transactions on circuits and systems for video technology, 14(1), 4-20.
2. Ross, A., & Jain, A. (2003). Information fusion in biometrics. Pattern recognition letters, 24(13), 2115-2125.
3. Ross, A. A., Jain, A. K., & Nandakumar, K. (2006). *Handbook of multibiometrics*. Boston, MA: Springer US.
4. Bowyer, K. W., Chang, K., & Flynn, P. (2006). A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition. Computer vision and image understanding, 101(1), 1-15.
5. Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. ACM computing surveys (CSUR), 35(4), 399-458.
6. Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). Ieee.
7. Marasco, E., & Ross, A. (2014). A survey on antispoofing schemes for fingerprint recognition systems. ACM Computing Surveys (CSUR), 47(2), 1-36.
8. Jain, A. K., Ross, A., & Prabhakar, S. (2004). An introduction to biometric recognition. IEEE Transactions on circuits and systems for video technology, 14(1), 4-20.
9. Bolle, R. M., Connell, J. H., Pankanti, S., Ratha, N. K., & Senior, A. W. (2013). Guide to biometrics. Springer Science & Business Media.
10. Shekhar, M., Trivedi, A. K., & Patgiri, R. (2025). Enhancing security and accuracy in biometric systems through the fusion of fingerprint and gait recognition technologies. International Journal of Biometrics, 17(5), 449-468.
11. Borra, S., Dey, N., & Sherratt, R. S. (2025). Biometric sensors. In Encyclopedia of Cryptography, Security and Privacy (pp. 217-220). Cham: Springer Nature Switzerland.
12. Subramanian, N. (2025). Biometric Authentication. In Encyclopedia of Cryptography, Security and Privacy (pp. 192-196). Cham: Springer Nature Switzerland.
13. Choudhary, P., Pathak, P., & Gupta, P. (2025). Physiological biometric image quality assessment-a review. Multimedia Tools and Applications, 84(34), 42257-42291.
14. Akintunde, O. A., Adetunji, A. B., Fenwa, O. D., Oguntoye, J. P., Olayiwola, D. S., & Adeleke, A. J. (2025). Comparative analysis of score level fusion techniques in multi-biometric system. LAUTECH Journal of Engineering and Technology, 19(1), 128-141.
15. Abdul-Al, M., Kyeremeh, G. K., Qahwaji, R., Ali, N. T., & Abd-Alhameed, R. A. (2026). Fusion-Enhanced Hybrid Multimodal Biometric System: Integrating Visible and Infrared Facial Recognition for Robust Authentication. IEEE Access, 14, 6006-6028.
16. Aliyu, M. G., Jamel, S., & Danlami, M. (2025). Article Advances in Feature Extraction and Selection for Iris Recognition Systems: A Review. Journal of Electronic Voltage and Application, 6(2), 148-165.
17. Poh, N. (2025). Biometric Sample Quality. In Encyclopedia of Cryptography, Security and Privacy (pp. 213-217). Cham: Springer Nature Switzerland.
18. Qaraa, S., Elbehairy, H., Mohamed, S., & ElRashidy, N. (2025). Human Gait Recognition for Security Systems. The Future of Inclusion: Bridging the Digital Divide with Emerging Technologies: Proceedings of ITAF 2024, 117.
19. Tomasz, M. A. K. A., & Smietanka, L. (2025). Analysis of Decision Fusion in Speech Detection. Archives of Acoustics, 50(4), 445-454.
20. Kajotra, S., & Kour, H. (2025). Face Recognition Technologies in Computer Vision–An Empirical Review. Recent Advances in Computing Sciences, 255-263.
21. Aboluhom, A. A. A., & Kandilli, I. (2025). Real-time facial recognition via multitask learning on raspberry Pi. Scientific Reports, 15(1), 28467.
22. Schmidhuber, J. (2025). Who invented deep residual learning?. arXiv preprint arXiv:2509.24732.
23. Zhou, X., & Bhanu, B. (2008). Feature fusion of side face and gait for video-based human identification. Pattern Recognition, 41(3), 778-795.
24. Camlikaya, E., Kholmatov, A., & Yanikoglu, B. (2008, March). Multi-biometric templates using fingerprint and voice. In Biometric technology for human identification V (Vol. 6944, pp. 145-153). SPIE.

 doi:10.20944/preprints202603.1167.v1

25. Bhanu, B., & Govindaraju, V. (Eds.). (2011). Multibiometrics for human identification. Cambridge University Press.
26. Rane, M. E., & Bhadade, U. S. (2025). Multimodal score level fusion for recognition using face and palmprint. International Journal of Electrical Engineering & Education, 62(1), 37-55.
27. Aleem, S., Yang, P., Masood, S., Li, P., & Sheng, B. (2020). An accurate multi-modal biometric identification system for person identification via fusion of face and finger print. World Wide Web, 23(2), 1299-1317.
28. Kyeremeh, G. K., Abdul-Al, M., Qahwaji, R., Ali, N. T., & Abd-Alhameed, R. A. (2025). Fusion of hand biometrics for border control involving fingerprint and finger vein. IEEE Access.
29. Shin, J., Miah, A. S. M., Egawa, R., Hassan, N., Hirooka, K., & Tomioka, Y. (2025). Multimodal fall detection using spatial–temporal attention and bi-lstm-based feature fusion. Future Internet, 17(4), 173.
30. Mishra, A. (2010). Multimodal biometrics it is: need for future systems. International journal of computer applications, 3(4), 28-33.
31. Jadhav, S. B., Deshmukh, N. K., & Pawar, S. B. (2025). Robust authentication system with privacy preservation for hybrid deep learning-based person identification system using multi-modal palmprint, ear, and face biometric features. International Journal of Image and Graphics, 25(05), 2550049.
32. Qiao, Y., Kang, W., Luo, D., & Huang, J. (2025). Normalized-Full-Palmar-Hand: Towards More Accurate Hand-Based Multimodal Biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence
33. Li, D., Ji, T., Sun, Y., Zhang, Z., Li, A., Qu, M., ... & Liu, H. (2025). A Full-Range Proximity-Tactile Sensor Based on Multimodal Perception Fusion for Minimally Invasive Surgical Robots. Advanced Science, e02353.
34. Zhang, Y., Xu, Y., Gao, J., Zhao, Z., Sun, J., & Mu, F. (2025). Urban Functional Zone Identification Based on Multimodal Data Fusion: A Case Study of Chongqing's Central Urban Area. Remote Sensing, 17(6), 990.
35. Garg, Bindu; Kasar, Manisha; Kashyap, Achyut; Vats, Amber; Sharma, Gunjan; Hange, Aditya (2025), "SignAlphaSet", Mendeley Data, V2, doi: 10.17632/8fmvr9m98w.2