International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 9, Issue 8, pp 22-30, August-2025
https://doi.org/10.47001/IRJIET/2025.908004

Advancements in Human Action Recognition: Leveraging Public Datasets and Biometric-Based Authentication Systems

[1]Ameera S. Mahmood, [2]Yaseen Hikmat Ismaiel

[1,2]Department Computer Science, College of Computer and Mathematic Science, Mosul University, Iraq
Authors E-mail: [1]ameera.22csp26@student.uomosul.edu.iq, [2]yaseen-hikmat@uomosul.edu.iq

***Abstract* - Human action recognition HAR is one of the critical areas of research in computer vision with several applications in security, healthcare, and human-computer interaction. The datasets made available to the public are used in the development of the field because they offer crucial tools for model calibration, credibility checks, and determination of efficiency. This paper presents a brief and important overview of HAR datasets, based on the categories of atomic actions, behaviors, interactions, as well as group activities. From the datasets, KTH, NTU RGB+D, and UCF101, the impact on action recognition is analyzed. Further, we consider the use of a biometric-based password system to better understand a more secure and convenient way of operating password systems. In this paper, a strategy of combining biometric authentication methods with multi-factor security mechanisms is presented. The paper reveals the need for enhancing the performance of the biometric system to generate good, secure passwords and emphasizes the privacy issues and weaknesses of the system. Last of all, the paper presents the future directions and proposals for behavioral biometric and multi-modal approaches to enhance security.**

Keywords: Human Action Recognition, Biometric Authentication, Public Datasets, Computer Vision, Action Recognition Datasets, Biometric Password Systems, Multi-Factor Authentication, Digital Security, Privacy, Behavioral Biometrics.

I. INTRODUCTION

Considering the trend in digital threats to security, biometric authentication systems have emerged as a significant alternative to passwords. Biometric authentication relies on human traits that are unique and cannot be reproduced as easily or done with imposter capabilities (e.g., fingerprints, facial traits, or iris patterns) [1]. Even though we have biometrics available to us to help mitigate those threats through authentication, unimodal biometric sources come with their limits, such as non-universality, sometimes inaccurate sensors in difficult environments, and susceptibility to spoofing using artificial tools.

Multimodal biometric systems have emerged to combine biometric features to get around these issues and provide better accuracy, robustness, and flexibility to address common problems. Biometric features are usually categorized by physiological, behavioral traits, and soft biometrics. Physiological characteristics include measurable (fingerprints, IK, visual facial, and facial characteristics). Behavioral characteristics display identifiable ways of behaving for each user (typing dynamics, identifier gait, and ways of interacting). Soft biometrics include non-unique characteristics that do not ability identification (e.g., height, gender, or age) that do improve performance significantly when paired with other modalities... Figure 1 shows three unique groupings of biometric features. Multimodal biometric recognition systems use several biometric features from these areas.

Fusion of biometric modalities can occur at four levels: sensor level, in which raw data from different sensors are fused; feature level, in which feature vectors are extracted and fused; score level, where match scores are computed and fused; and decision level, when a final decision is made based on the outputs of individual classifiers [2]. Typically, feature-level fusion is the best option, albeit it creates challenges regarding heterogeneous data fusion processing.

Simultaneously, Human Action Recognition (HAR) has emerged as one of the most favorable behavioral biometrics techniques that analyzes human movement and gesture information (body movements/gestures) containing RGB images, depth maps, or skeleton sequences. With the help of deep learning, the ability to extract complex spatial-temporal features from different HAR data has progressed significantly, enabling action-based authentication in dynamic scenarios to be more reliable [3].

This research aims to improve the recognition performance of biometric systems by using several biometric features. Furthermore, the study emphasizes the use of many layers of fusion to integrate various biometric variables.

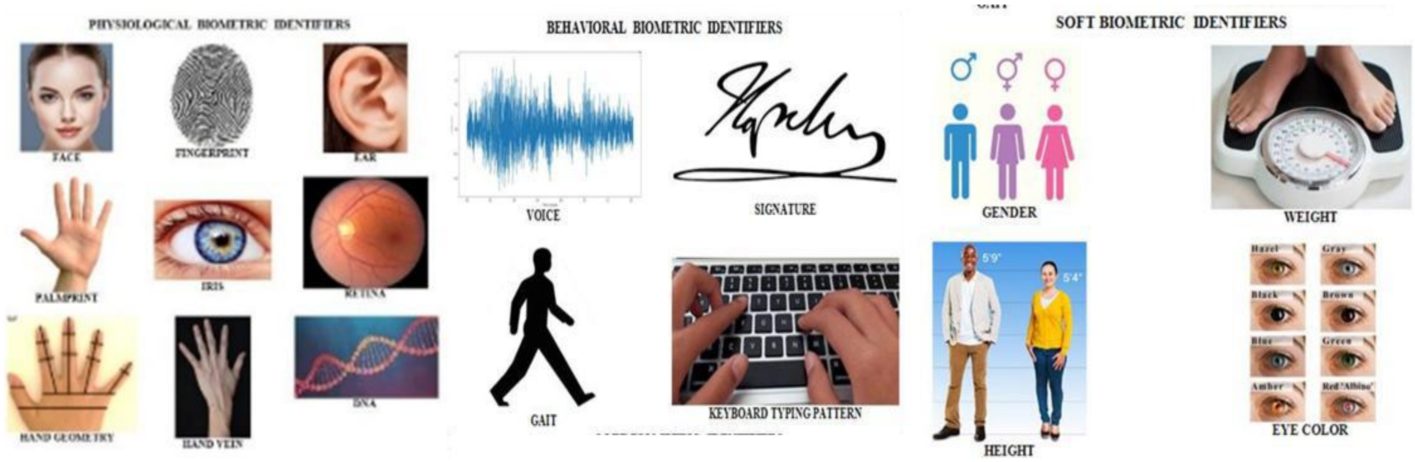

Figure 1: Physiological, behavioral, and soft biometric traits [4]

II. MULTI-BIOMETRIC SYSTEMS

By carefully merging data from several sources, a multi-biometric system overcomes the drawbacks of a single-biometric system. Since pooled information is more distinctive for individual identification than data obtained from a single source, leveraging many sources often enhances recognition performance and system dependability. Following the evaluation of fusion at different stages of the biometric identification process, the decision is taken regarding the types of information to combine [7].

2.1 Biometric Fusion

The multimodal framework performs much better thanks to the integration technique. The foundation of an effective multi-modal biometric framework is the use of a suitable fusion strategy to merge information obtained from various signals. Five distinct biometric fusion techniques are described in the literature: Four levels of fusion: feature-level, sensor-level, score-level, decision-level, and rank-level. Combining multiple cues is beneficial in a range of situations. Because feature-level and score-level fusion offer more gratifying and useful data integration, researchers looked into a variety of biometric modalities' fusion levels. They highlighted the possible contributions of these fusion levels to increased authentication accuracy [8].

2.2 Levels of Fusion

Figure 2 portrays biometric Fusion (BF) as an intrinsic part of a few biometric framework parts. It is fundamental in both distinguishing proof and check frameworks, giving shifting combination levels adjusted to individual necessities. This page gives a careful description of every combination level, including sensor-level, highlight-level, score-level, rank-level, and choice-level combinations. Every combination level is investigated on all the more likely to better understand its pertinence and applications in biometric distinguishing proof systems[3].

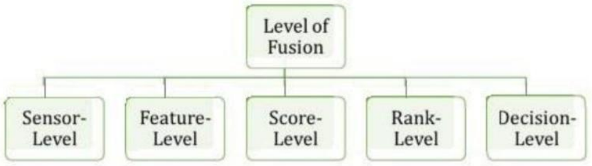

Figure 2: Fusion levels inside a multi-biometric system

Table 1 beneath sums up the various degrees of combination utilized in biometric recognition frameworks. Biometric Combination (BF) alludes to a group of approaches utilized at various phases of the recognition cycle. Understanding these combination levels is basic for further developing framework execution and guaranteeing biometric application reliability. The table gives a speedy portrayal of every combination level, including its motivation and regular methodologies utilized.

Table 1: Fusion levels inside a multi-biometric system [3]

Level of Fusion	Description
Sensor-level Fusion	Combination happens following information assortment, and is in many cases achieved utilizing multi-test or multi-sensor approaches. For instance, in face acknowledgment, this might incorporate

	coordinating many face pictures got according to different points of view into a single depiction through mosaicking or direct combination.
Feature-level Fusion	Joins components from the equivalent or particular information sources to improve biometric portrayals. This could involve consolidating highlight sets from a few biometric sorts or modalities to frame a uniform portrayal, which is often as possible utilized in multi-biometric frameworks to further develop security.
Score-level Fusion	Strategies for consolidating match scores from many matches using calculations, for example, max score combination or mean score combination. Since match scores from business frameworks are promptly accessible, this combination stage is irrefutable in writing.
Rank-level Fusion	This occurs after matching input probes with gallery templates, resulting in ranked lists of matched identities. Fusion approaches include regression models and Borda counts, which are appropriate for situations when features or matching scores are restricted.
Decision-level Fusion	Fusion at the decision level occurs when algorithms integrate results from numerous matches or classifiers. Examples include majority voting, which is useful in situations when only final judgments are accessible, such as commercial applications.

2.3 Data Protection in Biometric Systems

In biometric systems—especially those based on behavioral traits like Human Action Recognition (HAR)—ensuring data protection and user privacy is a critical concern. Unlike traditional passwords, biometric data cannot be easily changed once compromised. For instance, if a person's movement patterns or gestures are exposed, they can be misused without the person's knowledge. This highlights the importance of secure methods for storing, transmitting, and verifying such sensitive information.

One of the promising approaches for safeguarding biometric data is steganography, which hides sensitive data within seemingly harmless media, such as images or videos. Ismail et al. [33] introduced a method that uses two levels of steganography to embed protected information, making it significantly more difficult for attackers to detect or extract. This multi-layered strategy aligns well with the concept of fusion levels in biometric systems, where multiple sources or stages are combined to enhance security and reliability.

In another study, Ismail [34] proposed integrating data compression and encryption before embedding it into cover images. By compressing the data using Huffman coding and encrypting it, the approach not only reduces data size but also adds an extra layer of security against unauthorized access. Furthermore, a recent work [35] explored text-based steganography by encoding secret messages within specific positions of textual content, offering an alternative method for secure communication.

Although these techniques were initially applied in traditional digital security domains, their potential application in HAR-based biometric systems—especially in cloud environments or real-time data transmission—offers a valuable direction for enhancing data protection. Integrating such steganographic methods can ensure more secure and private biometric authentication processes.

III. HUMAN ACTION RECOGNITION FRAMEWORK

The data examined in previous Human Activity Recognition (HAR) research were divided into two main categories: vision-based and sensor-based. Whereas sensor-based HAR examines unprocessed data from wearable and monitoring devices, vision-based HAR concentrates on images or videos taken by optical sensors. Unlike wearable sensors, optical sensors are able to gather data as two-, three-, or four-dimensional images or videos. This makes them distinctive. Wearable technology, which is frequently employed in sensor-based HAR, tracks a variety of behaviors, including sitting, jogging, running, and resting. These sensors are limited, though, if the target is too far away or engaging in activities that the sensor is unable to pick up on [6].

The application of vision-based HAR in CCTV systems is well-established, and it has been thoroughly researched for activity and gesture identification. This method works especially well for interactive applications, security, and surveillance. Since vision-based HAR is easier to gather and more affordable than sensor-based data, it has become more popular in recent years. Consequently, computer vision for HAR is a common area of research interest [9].

ISSN (online): 2581-3048
Volume 9, Issue 8, pp 22-30, August-2025
https://doi.org/10.47001/IRJIET/2025.908004

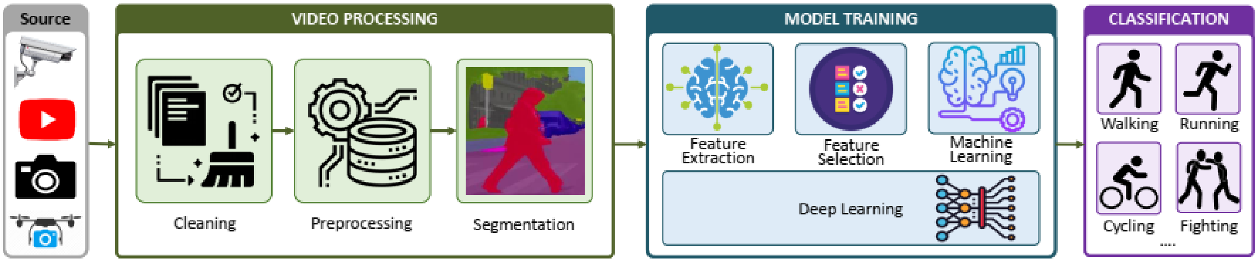

Figure 3: Human Action Recognition Framework

Typically, the framework for recognizing human activity consists of four key parts. First, utilizing optical sensing equipment, data is collected during the data collection phase. Second, crucial pre-processing actions for the gathered data are included in the pre-processing phase. Third, features are extracted from the dataset using methods like machine learning and deep learning during the learning or training phase. Lastly, the actions are identified and categorized during the activity recognition, also known as categorization, phase [9].

IV. HUMAN ACTION FEATURE EXTRACTION METHODS

4.1 Handcrafted Representation Method

Feature extraction is essential for Human action recognition (HAR). One of the oldest approaches includes: handcrafted features, which employ human-defined features such as Histogram of Oriented Gradients (HOG) and Depth Motion Maps (DMM). The descriptors mentioned above are calculated on colored images (RGB) or depth images in order to describe spatial and temporal behavior and are normally classified with classical machine learning models such as SVM [10].

4.1.1 Depth-Based Approaches

The prominence of depth-based methods was established by RGB-D sensors, e.g., Kinect. A major benefit of depth-based methods is that they can assist with accurate segmentation of the human body and the extraction of 3D motion characteristics from top, side, or front views. This includes multi-temporal DMMs, 3D depth projections, and normal vectors to improve detection capability in a convoluted environment [11][12].

Despite being efficient, handcrafted processes have limitations in terms of scalability and flexibility. The evolution of deep learning methods continues to complement or replace methods based on handcrafted sensor techniques, while simultaneously extracting hierarchical features and providing better accuracy than handcrafted alternatives. Numerous depth-based methods have been explored in the literature. For example, depth and color images can be captured using RGB-D cameras, allowing for high-quality descriptors like Depth Motion Maps (DMM) and Histograms of Oriented Gradients (HOG) for accurate action recognition [11].

Several proposed methods utilize 3D depth projection techniques and analyze actions by using depth data to create multiple views (top, side, and front) for improved spatial information [24]. One such method is the DMM (Depth Motion Maps) method, which can convert and utilize depth frames as motion maps and subsequently apply HOG to motion over time, creating efficient and succinct feature representations for different actions [12].To address variations in action speed, multiple temporal conditions were introduced within DMMs, allowing depth sequences to be segmented into multiple temporal sections, which improved the robustness of the recognition [13]. Also, normal vector-style methods compute surface normal in 3D space, which allows a better representation of the shape and motion of human limbs [14]. Lastly, local feature extraction techniques are a way to identify key points in depth images, such as interest points and the region of motion, using semi-local descriptors and space-time interest points (STIPs) [15].

4.1.2 Skeleton-Based Approaches

Skeleton-based methods for Human Action Recognition (HAR) are a class of techniques that represent human pose as a sequence of 2D or 3D joint locations over time. Because the representation is compact, it allows HAR processes to be efficiently computed, often with minimal storage use (to support computation of real-time performance). There are two types of skeleton-based methods: The trajectory-based method uses analysis of the spatial-temporal paths of body joints to extract features based on joint movement. They may use descriptors such as constrained time windows, and free-form curves [16]. We can call Volume-based methods those which ultimately compile spatio-temporal sequences as spatio-temporal volumes, that are made from features based on the pose, optical flow, gradients, and focus point detectors as a rich representation of the structure of motion [15].

4.1.3 Hybrid Feature-Based Approaches

Hybrid feature extraction approaches in human activity recognition (HAR) attempt to combine RGB, depth, and skeleton data to maximize performance. This process of fusing multiple modes enables models to harness both spatial and structural data.An example is a multi-sensory fusion approach where depth maps, skeletal joints, and color images are used, and basic action features are represented with descriptors that use Local Occupancy Patterns (LOP) [17].Another approach is the Hierarchical Hidden Markov Model (HHMM) that fused the skeletal motion with the depth cues, which produced a probabilistic model to combine multiple modalities of inputs [18].The random forest fusion that used joint motion energy and spatial interest points to improve classification, could also sometimes use a model like Naive Bayes Nearest Neighbor in their decision process [19]

4.2 Deep Learning Based Algorithms

To understand high-level data abstractions, deep learning, a type of machine learning, uses hierarchical algorithms. It has been extensively applied in a variety of AI fields, such as transfer learning, natural language processing, computer vision, and semantic parsing. Deep learning has gained popularity due to two main factors: the availability of large, high-quality labelled datasets and the switch from CPU-based to GPU-based training, which significantly speeds up the training of deep models [20].

The deep learning framework includes several supervised and unsupervised feature learning techniques, neural networks, and hierarchical probabilistic frameworks. Deep learning techniques have garnered significant attention recently due to their ability to use a wide range of data sources and outperform state-of-the-art methods on a variety of tasks. Deep learning-based features have replaced manually constructed features, leading to notable advances in computer vision [20].

Deep learning-based action recognition is gaining popularity due to its remarkable performance and ability to extract features from multi-dimensional datasets. Deep learning models input each feature into a deep network, which learns precise details across multiple layers, as opposed to traditional machine learning techniques, which rely on manually constructed features to recognize human actions. Even though these models are computationally costly during training and require a considerable quantity of data, their goal is to automatically extract important features for action detection. Several deep learning algorithms for action recognition exist, including Recurrent Neural Networks (RNNs), Autoencoders, Convolutional Neural Networks (CNNs), and Hybrid Models [21].

4.3 Attention-Based Methods

Attention-based methods are now critical to human action recognition by allowing models to focus only on the most informative spatial and temporal portions of the video data. Attention-based methods improve model performance by mimicking human perception, allowing models to differentiate which frames or sections of footage should be prioritized when recognizing that video, improving recognition classification and interpretation [22].

In addition, new model architectures like Transformers introduced conceptually new types of attention methods (e.g., multi-head and self-attention) that were originally developed for NLP but have been widely adapted and employed for vision tasks. Models like VideoBERT and VLM use both textual and visual modalities to enhance multimodal action understanding (e.g., visual recognition with text). Other frameworks like the GateHUB or PST^2 (Point Spatio-Temporal Transformer) implement attention over both time and space, thinking about incorporating these variations towards effective 3D action recognition [23].

With the growing abundance of attention-based architectures under different names and focuses, attention has been used to help enhance model performance in action recognition systems. Temporal attention models explicitly focus on identifying the most informative frames in an action sequence by focusing on the segments of an action that are most important for improving accuracy. Liu et al. showed that action recognition could be improved by simply emphasizing key temporal parts [23].

Similarly, purely spatial-temporal attention models can be developed in the same way to examine spatial and temporal processes in real-time, providing a larger array of features (event data) but highlighting action-relevant motion in both spatial and temporal dimensions. Originally conceived for natural language processing, transformer architectures have recently made significant advancements in the adaptation of these architectures for visual tasks. These models leverage key features such as multi-head attention, position encoding, and self-attention. A few notable adaptations include:

- GateHUB utilizes position-guided gating for cross-frame attention and gating of videos [4].
- Multiple vision-language models leveraging transformers have emerged. VideoBERT was inspired by word2vec's approach to jointly model visual and multiple textual modalities for a pretraining solution. VLM introduced two unigram sampling tasks, Masked Modality Modeling (MMM) and Masked Token Modeling (MTM), to improve multimodal comprehension. Similarly, ActBERT uses both local and global representations

from visual features and focuses on jointly modeling video and text representations.

- Sophisticated attempts like VATT and contrastive-based learning models like CBT employ contrastive learning paradigms ~e.g., Noise Contrastive Estimation (NCE), which help models retain fine-grained temporal allocations in video-language tasks. Indeed, there are also emergent and supporting models like UNIVL, which take multi-stream transformer encoders and provide action or predictions in comprehension and generation tasks for video and text data. Some recent undertakings also employ spatial-temporal factorization, which biases transformer models towards confined visual regions of parallel action, which can improve robustness or recognition in an action-centric dynamic scene [11].

V. PUBLICLY AVAILABLE DATASETS FOR HUMAN ACTION RECOGNITION

Researchers rely on publicly available datasets to train and validate Human Action Recognition (HAR) models. The datasets differ by their level of complexity, modality, and the type of actions they include (e.g., atomic actions such as waving or walking, object interactions such as using tools, and group activities). The well-used ***NTU RGB+D***, ****KTH***, and **UCF101** datasets have promoted future advancement of HAR systems by providing diverse video samples that are annotated and collected in several conditions.

Datasets are very useful benchmarks while creating and validating machine learning models when developing deep learning architectures that require a considerable amount of training data and labelled training data. Datasets typically are represented with only RGB, depth, and skeletal modalities, but have recently added multimodal dimensions as a recent trend in multi-sensor HAR is implemented [24][25].

VI. PASSWORD CREATION BASED ON BIOMETRIC FEATURES

Another improvement included in the given approach is the password creation using the biometric features, which is much more secure than traditional passwords, as well as convenient. Biometric systems use physical or behavioral characteristics like fingerprints, face, iris, or voice, and hence can hardly be forged, or someone else's biometric data obtained. While alphanumeric passwords can be forgotten, guessed, or cracked by brute force, biometric authentication processes use intrinsic features of an individual, making sure that access can only be granted to that particular person. They are gradually being incorporated into smartphones, computers, and even financial systems, where they allow for very easy use without the user having to remember a password or deal with password reset problems[26].

Biometric authentication systems offer stronger security than traditional passwords because they rely on unique personal traits that cannot be easily shared or replicated. Unlike passwords or passphrases, which are vulnerable to phishing, keylogging, or social engineering attacks, biometric data such as fingerprints or iris patterns is much harder to intercept or duplicate. As a result, biometric systems provide more effective protection against common security threats [27].

However, due to these and other challenges, there is steady progress in the advancement of biometric authentication technologies, including the multi-factor biometric systems and the behavioral biometric systems that can go ahead and improve the already existing security. Therefore, implementing biometric authentication together with other factors, for instance, Personal Identification Numbers or one-time passwords increases the strength of the solution since it is easier for a person to develop new threats as compared to a computer than to combine numerous security levels that are also trusted by users. In the future, given that the biometric technology is improving gradually, it is going to be the focal point of constructing safe passwords where convenience, effectiveness, and security are all factors to consider [28].

6.1 Types of Passwords and Their Security Implications

Passwords are an important component of creating strong security for an individual's information and other owned accounts and profiles. In this connection, different types of passwords have been invented in different eras with different levels of protection and with their special advantages. Passwords have evolved significantly from just simple numbers and letters to complicated alphanumeric ones, and even using biometric security features today, offer a wide range of passwords that suit individual' needs, security measures, and personal interests. In this table, we classify the various types of passwords depending on their descriptions, security level, and usability in order to determine which method fits best in securing certain areas of a digital environment [29].

Table 8: Different Types of Passwords

Password Type	Description	Security Level	Usability
Alphanumeric Passwords[30]	A combination of uppercase and lowercase letters and numbers.	Medium	Moderate
Special Character Passwords[30]	Includes special characters (e.g., @, #, $, etc.) along with letters and numbers.	High	Moderate
Passphrases[30]	A sequence of words or a sentence, often including spaces and special characters.	High	High (if long and complex)
Biometric-Based Passwords[30]	Uses physical traits such as fingerprints, face recognition, or voice patterns for authentication.	Very High	Very High (if accessible)
Two-Factor Authentication (2FA)[31]	Combines a password with a second authentication factor (e.g., a code sent via SMS, email, or a fingerprint scan).	Very High	Moderate to High
Graphical Passwords[30]	Passwords based on selecting images or drawing patterns on a grid or touchscreen.	High (if well-implemented)	High (user-friendly)
PINs (Personal Identification Numbers)[30]	Short numeric passwords, typically 4 to 6 digits, are often used for devices or ATM access.	Medium	Very High (easy to remember)
Behavioral Passwords[30]	Relies on unique patterns of user behavior (e.g., typing speed, swipe pattern, or mouse movements).	High (if biometric analysis is robust)	Moderate (needs monitoring)
One-Time Passwords (OTP)[30]	Temporary, single-use passwords are generated for each login or transaction, often delivered via SMS or email.	Very High	High (for transactions)
Cognitive Passwords[30]	Passwords based on personal knowledge, such as answers to security questions (e.g., "What was your first pet's name?").	Medium to Low	Moderate (if questions are unique)

VII. CONCLUSIONS

This work helps to stress the relevance of open-source data sources in the development of human action recognition. The reviewed datasets from the atomic actions to the group activities have made tremendous advancements possible for the action recognition systems in terms of accuracy and efficiency. The evaluation of biometric-based password systems also supports the fact that biometric technologies are capable of replacing the conventional password system. However, some issues like privacy and vulnerability to spoofing continue to be a major factor in curtailing its usage. In response to these difficulties, the future development and application of biometric technologies, especially multi-factor and behavioral biometric systems, are crucial to designing better authentication technologies. The integration of these systems with other new, increasing machine learning approaches is expected to transform the way digital security is addressed. More studies should be conducted to establish these improvements and identify the best solutions and applications that can provide both high security and privacy, while being easily used by the average user.

REFERENCES

[1] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, Present, and Future of Face Recognition: A Review," *Electronics*, vol. 9, no. 8, 2020, doi: 10.3390/electronics9081188.

[2] U. Sumalatha, K. K. Prakasha, S. Prabhu, and V. C. Nayak, "A Comprehensive Review of Unimodal and Multimodal Fingerprint Biometric Authentication Systems: Fusion, Attacks, and Template Protection," *IEEE Access*, vol. 12, pp. 64300–64334, 2024, doi: 10.1109/ACCESS.2024.3395417.

[3] H. A. Hussain and H. H. Abbas, "A Survey on Multi-biometric Fusion Approaches," *Kerbala J. Eng. Sci.*, vol. 03, no. 02, 2023.

[4] B. C. Arjun and H. N. Prakash, "Multimodal Biometric Recognition System Using Face and Finger Vein Biometric Traits with Feature and Decision Level Fusion Techniques," *Int. J. Comput. Theory Eng.*, vol. 13, no. 4, pp. 123–128, 2021, doi: 10.7763/IJCTE.2021.V13.1300.

[5] D. Cao, R. Liu, H. Li, S. Wang, W. Jiang, and C. X. Lu, "Cross Vision-RF Gait Re-identification with Low-cost RGB-D Cameras and mmWave Radars," *Proc.*

ACM Interactive, Mobile, Wearable Ubiquitous Technol., vol. 6, no. 3, 2022, doi: 10.1145/3550325.

[6] M. G. Morshed, T. Sultana, A. Alam, and Y. K. Lee, "Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities," *Sensors*, vol. 23, no. 4, pp. 1–40, 2023, doi: 10.3390/s23042182.

[7] E. Balraj and T. Abirami, "Performance Improvement of Multibiometric Authentication System Using Score Level Fusion with Ant Colony Optimization," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/4145785.

[8] N. Bala, D. R. Gupta, and A. Kumar, "Multimodal biometric system based on fusion techniques: a review," *Inf. Secur. J. A Glob. Perspect.*, pp. 1–49, Dec. 2021, doi: 10.1080/19393555.2021.1974130.

[9] R. R. Kumar *et al.*, "Report on the Follow-Up To the Regional Implementation Strategy of the Madrid International Plan of Action on Ageing in Lithuania," *Front. Neurosci.*, vol. 14, no. 1, pp. 1–13, 2021.

[10] R. Raj and A. Kos, "An improved human activity recognition technique based on a convolutional neural network," *Sci. Rep.*., vol. 13, no. 1, pp. 1–19, 2023, doi: 10.1038/s41598-023-49739-1.

[11] M. B. Shaikh and D. Chai, "RGB-D data-based action recognition: A review," *Sensors*, vol. 21, no. 12, pp. 1–25, 2021, doi: 10.3390/s21124246.

[12] M. F. Bulbul, S. Tabussum, H. Ali, W. Zheng, M. Y. Lee, and A. Ullah, "Exploring 3d human action recognition using STACOG on multi-view depth motion maps sequences," *Sensors*, vol. 21, no. 11, pp. 1–18, 2021, doi: 10.3390/s21113642.

[13] F. Shafizadegan, A. R. Naghsh-Nilchi, and E. Shabaninia, *Multimodal vision-based human action recognition using deep learning: a review*, vol. 57, no. 7. 2024. doi: 10.1007/s10462-024-10730-5.

[14] Y. Zhang *et al.*, "Extract latent features of single-particle trajectories with historical experience learning," *Biophys. J.*, vol. 122, no. 22, pp. 4451–4466, 2023, doi: 10.1016/j.bpj.2023.10.023.

[15] A. Sergiyenko, P. Serhiienko, and M. Orlova, "Local Feature Extraction in Images," *Information, Comput. Intell. Syst.*, no. 2, 2021, doi: 10.20535/2708-4930.2.2021.244191.

[16] Y. Wei *et al.*, "Elevating Skeleton-Based Action Recognition With Efficient Multi-Modality Self-Supervision," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, no. April, pp. 6040–6044, 2024, doi: 10.1109/ICASSP48485.2024.10447178.

[17] A. Moly, L. Struber, and G. Charvet, "Hierarchical Hidden Markov Model for Online Decoding in Brain-Computer Interface," pp. 1466–1470, 2024.

[18] M. Ameur, C. Daoui, and N. Idrissi, "Hierarchical hidden Markov models in image segmentation," *Sci. Vis.*, vol. 12, no. 1, pp. 22–47, 2020, doi: 10.26583/SV.12.1.03.

[19] J. S. Wibowo, E. N. Wahyudi, and H. Listiyono, "Performance Comparison of SVM, Naive Bayes, and Random Forest Models in Fake News Classification," *Eng. Technol. J.*, vol. 09, no. 08, pp. 4799–4804, 2024, doi: 10.47191/etj/v9i08.27.

[20] S. F. Ahmed *et al.*, *Deep learning modelling techniques: current progress, applications, advantages, and challenges*, vol. 56, no. 11. Springer Netherlands, 2023. doi: 10.1007/s10462-023-10466-8.

[21] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5. 2023. doi: 10.3390/computers12050091.

[22] R. Luiz and P. Bueno, "ATTENTION-BASED VIEW : PAST, PRESENT AND FUTURE," vol. 23, no. July 2023, pp. 1–41, 2024.

[23] R. Cui, A. Zhu, J. Wu, and G. Hua, "Skeleton-based Attention-aware Spatial-temporal Model for Action Detection and Recognition," *IET Comput. Vis.*, vol. 14, Feb. 2020, doi: 10.1049/iet-cvi.2019.0751.

[24] M. G. Morshed, T. Sultana, A. Alam, and Y. K. Lee, "Human Action Recognition: A Taxonomy-Based Survey, Updates, and Opportunities," Sensors, vol. 23, no. 4, pp. 1–40, 2023, doi: 10.3390/s23042182.

[25] G. Diraco, G. Rescio, P. Siciliano, and A. Leone, "Review on Human Action Recognition in Smart Living: Sensing Technology, Multimodality, Real-Time Processing, Interoperability, and Resource-Constrained Processing," Sensors, vol. 23, no. 11, pp. 1–26, 2023, doi: 10.3390/s23115281.

[26] E. Misini and U. Lajçi, "Biometric authentication," *Computer (Long. Beach, Calif)*, vol. 39, no. 2, pp. 96–97, 2022, doi: 10.1109/MC.2006.47.

[27] A. K. Jain, A. Ross, and K. Nandakumar, Introduction to Biometrics, Advances in Information Security, vol. 60, Springer, 2021. doi: 10.1007/978-1-4471-7307-2_1.

[28] M. Ganganna, "Review on Technology Advancements in Biometric Authentication Review on Technology Advancements in Biometric Authentication," no. July 2024.

[29] P. Fernando, C. Liyanage, and C. Karunatilake, "Challenges and Opportunities in Password Management : A Review of Current Challenges and Opportunities in Password Management : A Review of Current Solutions," no. August 2023, doi: 10.4038/sljssh.v3i2.96.

[30] N. A. Lal, S. Prasad, and M. Farik, "A Systematic Literature Review of the Types of Authentication," International Journal of Advanced Computer Science

and Applications (IJACSA), vol. 12, no. 7, pp. 832–849, 2021, doi: 10.14569/IJACSA.2021.0120784.

[31] W. Go, K. Internet, S. Agency, and J. Kwak, "Construction of a secure two-factor user authentication system using fingerprint information and password," no. April, 2014, doi: 10.1007/s10845-012-0669-y.

[32] L. E. Almeida, B. A. Fernández, D. Zambrano, and A. I. Almachi, "One-Time Passwords : A Literary Review of Different Protocols and Their Applications One-Time Passwords : A Literary Review of Different Protocols and Their Applications," no. January 2024, doi: 10.1007/978-3-031-48855-9.

[33] Y. H. Ismail, "Improved security using two levels of steganography," AIP Conf. Proc., vol. 3264, no. 1, pp. 030010-1–030010-6, 2023, doi: 10.1063/5.0191873.

[34] Y. H. Ismail,Y.S.Yousif, "Using compression and encryption to provide secure image steganography," AIP Conf. Proc., vol. 2398, no. 1, pp. 050033-1–050033-6, 2022, doi: 10.1063/5.0066466.

[35] Y. H. Ismail, A. A.Idres., "Proposed Method for Text Steganography," Journal of Modern Computing and Engineering Research, vol. 3, no. 1, pp. 45–51, 2024.

Citation of this Article:

Ameera S. Mahmood, & Yaseen Hikmat Ismaiel. (2025). Advancements in Human Action Recognition: Leveraging Public Datasets and Biometric-Based Authentication Systems. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(8), 22-30. Article DOI https://doi.org/10.47001/IRJIET/2025.908004

****