Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

Study of Multimodal Identification Algorithms Using Modern Methods and Tools of Multivariate Analysis

Nataliya Boyko[1]

Abstract

This article aims to comprehensively investigate the theoretical and practical foundations, as well as the distinctive characteristics, underpinning the study of multimodal identification algorithms. This investigation will be conducted using state-of-the-art methods and tools of multidimensional analysis. The development of a multimodal algorithm using the method of modality fusion at the feature level encompasses the integration of various algorithms rooted in multivariate analysis. These include a combined voice activity detector, a face detector utilizing the MTCNN (multi-task cascade convolutional networks) architecture, fine-frequency cepstral coefficients, facial image features, and a decision-making module. To construct a multimodal identification algorithm, a framework for combining these algorithms based on multivariate analysis is proposed. Analysis of the acquired data indicates that "Test 1", utilizing facial image data, exhibits the highest performance indicators, approaching nearly 100%. Tests 2 and 3 involving voice signals exhibit a minor error in the pre-processing stage, attributed to the inherent delay experienced by participants during the video conference. The proposed multimodal algorithm, integrated within a biometric identification system, enables successful user verification research through the utilization of a combined multidimensional analysis algorithm. Furthermore, the algorithm showcases superior research outcomes in comparison to other analogous multimodal identification algorithms, as it yields precise results.

Keywords: *Algorithms, Modality, Multimodal Data, Multimodal Machine Learning, Multivariate Analysis.*

Introduction

The escalating impact of contemporary challenges and threats precipitates substantial destabilization of phenomena and processes across various domains of human activity, leading to deleterious transformations (Latysheva et al., 2020; Kovaleva et al., 2020). Amid this instability and uncertainty, the imperative to identify avenues for advancing information and communication tools, as well as the widespread utilization of computer systems, becomes paramount in implementing identification measures (Iatsyshyn et al., 2020; Shytyk and Akimova, 2020). These measures are based on a diverse range of methods and models for retrospective assessment of processes and phenomena and for making strategic forecasts. While the active phase of machine learning algorithm development and their successful integration into society and science has yielded some favorable outcomes, as evidenced by empirical investigations of machine learning algorithms for multimodal data, it is imperative to acknowledge that the challenges of classification in machine learning persist (Ostapenko et al., 2020; Popovych et al., 2021). These challenges persist despite significant advancements in the field and are further accentuated by the progress of modern science, thereby underscoring their enduring relevance and necessitating rigorous exploration.

The advancement of information technologies and computer hardware continues to intensify, leading to their increasing integration across diverse domains of society. Consequently, this integration exerts a significant influence on the processes and phenomena occurring within these domains, through the lenses of digitalization and qualitative transformations. In such a context, multimodal algorithms assume a pivotal role, as their utilization deepens and their functional scope expands. Innovative technologies and cutting-edge technical resources shift the focus from a human-centered paradigm to the software domain, necessitating the enhancement of methods for adapting human-machine interfaces and strengthening identification mechanisms. Undoubtedly, in the contemporary world, novel principles and opportunities for working with information that characterizes social phenomena and processes through computer systems

[1] PhD, Associated Professor at the Department of Artificial Intelligence, Lviv Polytechnic National University, Lviv, Ukraine. ORCID: https://orcid.org/0000-0002-6962-9363

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

are emerging. Consequently, the necessity for processing multimodal data, which entails working with diverse data types and integrating them into a unified model or algorithm, is substantiated.

Literature Review

The comprehensive examination of existing scientific advancements in this domain allows us to assert that, in the present circumstances, the study of multimodal identification algorithms using contemporary methods and tools of multidimensional analysis has become increasingly pertinent. This topic has garnered extensive deliberation within scientific circles and has found practical application. In particular, Nawal et al. (2024), Pandey et al. (2024), Park (2023), Pereira et al. (2023), Rajakumar and Ananth Kumar (2022) studied biometric elements and attributes, which allowed them to classify modern methods of object identification and visualize them in Figure 1. It is worth emphasizing that these methods pertain to biometric identification techniques and necessitate the utilization of multidimensional analysis methods and tools, which, in turn, form the basis for the development of multimodal identification algorithms.

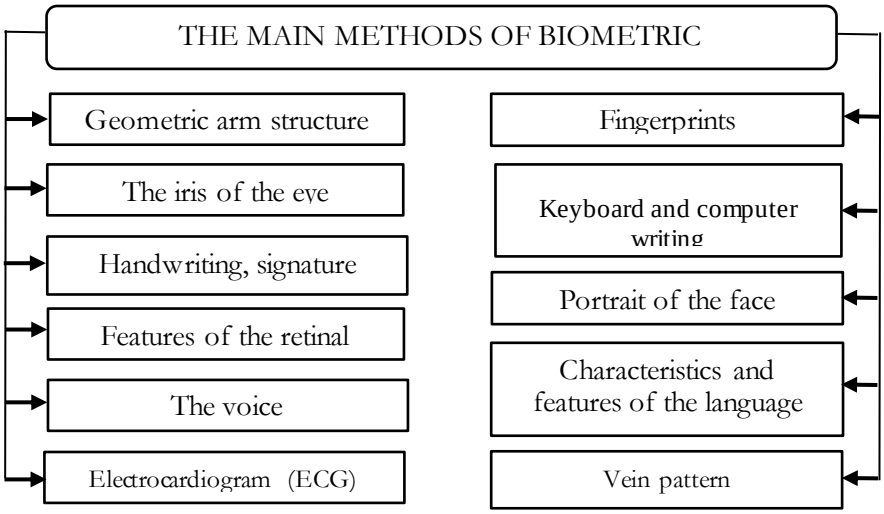

Figure 1. The Main Methods of Biometric Identification

Source: compiled based on (Nawal et al., 2024; Pandey et al., 2024; Park, 2023; Pereira et al., 2023; Rajakumar & Ananth Kumar, 2022)

The intricate facets surrounding the investigation of multimodal identification algorithms employing contemporary methods and tools of multivariate analysis have gained particular significance within the present context, further accentuated by the prevailing challenges and threats. Consequently, they have become a subject of fervent scientific discourse. Notably, Boyko (2021), acknowledging the accelerated integration of machine learning algorithms into diverse domains of existence, has directed scholarly inquiry in this direction. The utilization of data from cutting-edge technologies plays an active role in generating forecasts for multimodal data and facilitating the classification of phenomena and processes that permeate various strata of social life.

In their study, Abdullah et al. (2021) delved into the exploration of multimodal identification algorithms, employing contemporary methods and tools of multivariate analysis. Their investigation revealed the profound interplay between humans and modern computer technologies, which possess the capability to accomplish the task of emotion recognition from multimodal signals through meticulous examination and comparison. Furthermore, the efficacy of the multimodal processing carried out by the computer system is of paramount importance. Its accuracy is contingent upon factors such as the number of emotions observed, the extraction of pertinent features, and their alignment with established databases.

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

In their study, Wu et al. (2022) examined the viability of employing competitive discriminative domain adaptation algorithms for target detection in the fusion of infrared and visible images. This approach enables the detection of information from both modalities, infrared and visible light. Consequently, it significantly contributes to the accuracy of detection and identification algorithms.

In the context of advancing modern high-throughput omics measurement platforms for biometric research, Reel et al. (2021) highlighted the capacity of machine learning methods to integrate and analyze diverse omics datasets. This integration and analysis facilitate the identification of novel biomarkers that hold potential for identification purposes.

Based on the principles of multivariate analysis, Kumar and Jeyakumar (2022) extensively explore multimodal identification algorithms. They posit that the increasing demand for prompt utilization of digital information necessitates the quest for efficient systems capable of processing multimodal and multidimensional data. In this context, the problem of multivariate analysis assumes heightened relevance. This problem entails describing diverse registers within the research process and employing multivariate statistical methods to identify additional correlated sets of numerous variables (Sardinha & Pinto, 2019).

Simultaneously, Boyko (2023b) delve into the investigation of the merits of multimodality-based allocation algorithms and ascertain their potential integration with clustering strategies. Furthermore, they argue that these technologies prove suitable for working with multidimensional data. These assertions are substantiated by the findings of Gaonkar et al. (2021), who assert that multimodal data representation plays a pivotal role in data processing. The researchers contend that information fusion algorithms effectively mitigate heterogeneity gaps within datasets. Additionally, they identify various methods that leverage unimodal signals, including images, speech, text, iris, and fingerprints.

The perspective espoused by Behrad and Abadeh (2022) aligns with this notion, as they firmly believe that the utilization of multimodal data yields more precise outcomes. This stems from the additional information it provides, which holds strategic significance in the face of ambiguity and the imperative to construct the most accurate forecasts and identification processes.

In recent times, the exploration of innovative technologies in the realm of artificial intelligence has gained significant traction. This research focus has been prompted by the integration of machine learning methods, a development justified by the demands of the current era. As highlighted by Bayoudh et al. (2022), the abundance of vast datasets coupled with technological advancements creates an environment conducive to the rapid search for effective decision-making methods tailored to specific problems. Simultaneously, there is an escalating need to advance deep learning as a distinct approach within the field of machine learning, owing to its predictive capabilities and versatility. Additionally, as noted by Boyko (2023a), the utilization of multimodal data assumes paramount importance when there is a requirement to analyze dynamic changes and differentiated parameters of an object concurrently. Furthermore, the identification of digital surveillance objects requires the development of diverse models and algorithms to effectively represent multimodal data. Among the myriad contemporary methods and tools of identification, the author emphasizes the significance of multidimensional analysis, citing its inherent features and practical applicability as justifications. Building upon prior research in this domain, Ugryumov et al. (2020) also underscore the vital importance of system analysis methods, which prove highly relevant in optimizing and facilitating decision-making processes related to object identification.

Zhao et al. (2024) delve into the examination of multimodal data representation, emphasizing the identification of specific models that amalgamate multilevel, spatial, and temporal characteristics. Concurrently, Shoumy et al. (2020) assert the indispensability of employing state-of-the-art methods and tools, such as artificial intelligence and natural language processing, within the construction of multimodal identification algorithms. Additionally, they advocate for the integration of affective computing, which entails evaluating sentiments, emotions, and thought modeling, to enhance the algorithmic capabilities.

Consequently, the findings of comparative analysis within the domain of researching multimodal identification algorithms employing contemporary methods and tools of multidimensional analysis reveal

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

the diverse and multifaceted nature of scientific approaches to the identified issues. This diversity hinders the effective unification and selection of a singular consensus. Therefore, considering the aforementioned observations, it becomes evident that comprehensive investigations and extensive research efforts are imperative for addressing the intricacies associated with multimodal identification algorithms utilizing modern methods and tools of multidimensional analysis.

Research Objectives

The primary objective of this article is to thoroughly investigate the theoretical and practical underpinnings, as well as the distinct characteristics, of multimodal identification algorithms utilizing contemporary methods and tools of multidimensional analysis.

To accomplish this objective, the following tasks were undertaken: a comprehensive exploration and analysis of literature sources on the utilization of multimodal identification algorithms; an examination of sources discussing the integration of multimodal algorithms with artificial intelligence techniques, particularly those based on artificial neural networks; the development and implementation of multimodal algorithms utilizing specialized software and the creation of audiovisual datasets; the execution of a study involving test data to evaluate the performance of a multimodal identification algorithm; and an in-depth investigation of multimodal multivariate analysis algorithms for identification purposes.

The scientific novelty of this study is to proposes a scheme of a combined multimodal biometric identification algorithm based on multivariate analysis.

Input Data and Methods

Creating A Set of Video and Audio Data

The video dataset and audio were utilized to create an audiovisual set, employing specialized software such as Speech2Face. This software tools are specifically engineered for the generation, refinement, and evaluation of identification algorithms that rely on neural networks. The research materials consisted of image samples extracted from the video dataset obtained from the National Aviation University during a scientific and practical conference. The dataset encompassed a diverse range of participant variations, including changes in angles, positions within the frame, and other relevant factors. In Figure 2, we present a depiction of the image examples that were analyzed based on the compiled video database.

The authors conducted a comprehensive data collection and preliminary verification process to compile a dataset comprising 42 hours of real-time recordings from user conversations conducted through the Google Meet video-telephony service during video conferences. A total of 72 users participated in these sessions.

During the utilization of Speech2Face, it is important to acknowledge that the artificial neural network undergoes training to analyze the correlations between the audio and visual components of the voice. This enables the generation of images that reflect various physical attributes. The model receives the spectrogram of an audio voice recording as input, resulting in vector data representing facial characteristics. These vectors are subsequently decoded to generate the final face image. The algorithm leverages the AVSpeech dataset and the trained VGG-Face network, which facilitates the correlation of speech features with a range of biometric characteristics.

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

Figure 2. Image Analysis Based on The Collected Speech2Face Video Database

Source: compiled by the author

The utilization of the comprehensive Speech2Face dataset offers the potential to facilitate the development of a wide array of processes centered around the multidimensional analysis of unimodal and multimodal biometric identification algorithms. Consequently, the audiovisual databases of Speech2Face enable the integration of voice and face biometrics, combining them with other biometric parameters utilized as independent modalities (Rafatirad et al., 2022).

Creation Of the Speech2Face Audio Source Dataset for The Development of a Multimodal Identification Algorithm

To develop and evaluate the voice activity detector, it is essential to carefully plan and assemble a designated set of input data. For this purpose, the acquisition of speech signals using Speech2Face is undertaken, which comprises real speech recordings sourced from 22 distinct speakers. The speech data in this dataset is deliberately balanced in terms of gender, with 54% representing male speakers and 46% female speakers, respectively. Subsequently, each speaker is recorded for a duration ranging from 45 to 60 seconds, capturing the raw data through the utilization of Google Meet software. After the acquisition process, each audio fragment is meticulously analyzed to identify key characteristics such as noise, pauses, and interference.

During the preparatory and data collection phase, two principal datasets, namely Speech2Face, were identified and characterized. These datasets serve the purpose of developing and evaluating multimodal user identification algorithms. Furthermore, they can also be effectively utilized for the development and testing of a voice activity detector.

Testing Standard Speaker Identification Algorithms in The Speech2Face Dataset Based on A Neural Network

The utilization of convolutional neural networks (CNN) as a speaker identification algorithm involves the implementation of the VGG-M architecture (Rudregowda et al., 2024). This architecture is applied to train and test algorithms using the original speech signals derived from the pre-existing Speech2Face dataset. Before training and testing, the data from this dataset underwent a preprocessing stage, which comprised three key steps (Rafatirad et al., 2022). Firstly, a combined voice activity detector algorithm was employed to identify active speech regions. Subsequently, the speech signals were segmented into equal-duration fragments. Finally, frequency representations in the form of spectrograms and MTCNNs (multi-task cascade convolutional networks) were generated for these speech fragments (Masood et al., 2023). The preprocessing phase resulted in the creation of 33,128 speech fragments.

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

Classification Of Methods for Combining Biometric Parameters

Classical user identification systems are built based on using a single biometric modality, which is responsible for the type of biometric data used to verify and confirm the user. Such data may include, as a rule, fingerprints, face images, iris, audio, etc., but any unimodal user recognition system has inherent limitations. Such limitations include the lack of physical contact with the user, for example, when fingerprinting and iris analysis require physical contact with the user, which somewhat reduces the scope of practical application (Ammour et al., 2020).

Facial recognition systems are susceptible to variations in illumination levels and the presence of foreign objects during scanning, posing challenges to the accurate identification of facial features. The success of user identification is hindered by factors such as the quality of the photo registrar, sensitivity to age-related changes, facial expressions, and the angle of the user's face. Similarly, the effectiveness of a speaker identification system relies on the quality of the microphone and the data transmission channel. To address these issues, it becomes necessary to develop and implement solutions that enhance the reliability and robustness of the user recognition process (Wei et al., 2022).

The development of biometric identification systems encompasses the creation of multimodal algorithms, which involve the analysis of two, three, or more biometric parameters (Alay & Al-Baity, 2020). The fundamental concept behind multimodal algorithms is their ability to utilize a combined set of biometric parameters during user verification. This multimodal approach offers versatility, as it allows for the use of one or multiple biometric parameters depending on the specific verification scenario.

Multimodal biometric identification systems (MBIS) are comprised of four key modules: the data coverage module, feature generation module, comparison module, and decision-making module, as illustrated if Figure 3. In closed-set user identification tasks, the comparison module is not utilized, as it is specifically designed for verification tasks. Instead, it is substituted with the classification module (Lu et al., 2020).

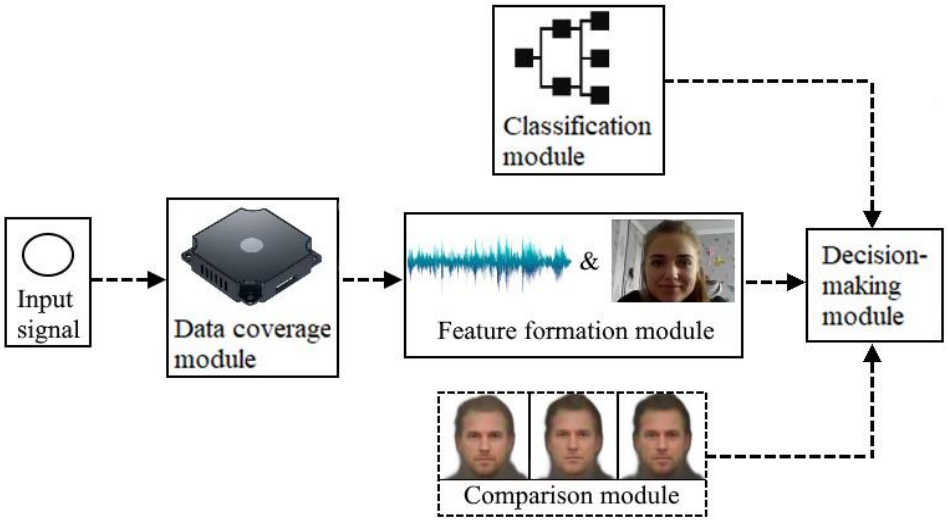

Figure 3. Component Modules for A Multimodal Biometric System

Source: compiled by the author

The Data Capture Module

It encompasses devices responsible for acquiring biometric data, such as cameras, microphones, fingerprints, and iris scanners, among others. This module also incorporates preprocessing techniques for the original data, including quality enhancement and the suppression of noise and interference.

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

The Feature Extraction Module

The feature extraction module is specifically designed to create a concise and informative representation of the provided data. In image analysis tasks, this module may employ features extracted using classical machine learning algorithms such as regression and clustering (e.g., binary patterns, support vector methods, an ensemble of decision trees, etc.). Similarly, in speech signal analysis tasks, feature extraction techniques can be utilized (Tanveer et al., 2023).

The Classification Module

The classification module is an essential component of a multimodal biometric identification system. It utilizes algorithms that search for patterns in the extracted feature data to accurately identify a user (Safavipour et al., 2022). In the case of algorithms based on convolutional neural networks (CNNs), the feature generation and classification functions are combined into a single module. Subsequently, the output of the multimodal system is further processed by a decision-making module, which analyzes the classifier's result and makes the final decision regarding the user recognition outcome (Tanveer et al., 2023).

Methods Of Designing Biometric System Schemes and Development of Multimodal Biometric Algorithms

Multimodal biometric systems are developed using two main approaches: sequential and parallel. In the sequential approach, the output results of one biometric modality analysis are utilized to narrow down the search area for potential users, and this information is then transferred to the input of the next unit for further analysis. On the other hand, in the parallel approach, biometric information of different natures is analyzed simultaneously. Furthermore, there are three levels at which the combination of modalities can occur: the feature level, the decision-making process, and the comparison level (Maiti et al., 2024).

In addition to the sequential and parallel approaches, there are other strategies for combining biometric data analysis, including multi-algorithmic systems and multi-sensor biometric systems. Multi-algorithmic systems involve the use of different algorithms to analyze the same modality. The selection of a specific algorithm is determined by the operating conditions of the system. For instance, in a face recognition system, one algorithm may be optimized for daylight conditions, while another algorithm may be designed for low-light or nighttime scenarios. These algorithms can operate concurrently, but the final decision regarding user identification is made by the decision-making module, which combines the results from multiple algorithms (Boykoa, 2023).

In addition to the analysis of multiple modalities and the utilization of multiple algorithms for a single biometric parameter, another approach commonly employed in the development of multisensor biometric systems is the use of multiple physical sensors with different functionalities to analyze a single biometric parameter. For example, a multisensor system may employ both a visible-range camera and an infrared thermal imager simultaneously to capture and analyze facial biometrics. This approach enhances the recognition process by leveraging complementary information provided by different sensors.

Furthermore, the multi-instance approach is another classification method used in combined biometric systems. This approach involves the analysis of several instances or variations of a single biometric modality. For instance, when analyzing facial biometrics, multiple instances of the face captured from different camera angles or positions can be considered (Boykoa, 2023).

Results

Research And Development of Multimodal Identification Algorithms

To develop a multimodal identification algorithm based on a convolutional neural network (CNN), the integration of modalities at the feature level is a crucial step. Two commonly used techniques for combining modalities in such algorithms are concatenation and bilinear combination. Concatenation involves merging features extracted from independent modalities into a single feature vector using a concatenation layer. This

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

approach allows the algorithm to leverage the information captured by each modality independently. Bilinear combination, on the other hand, combines modalities at the feature level by performing element-wise products between corresponding elements of the feature vectors. This approach captures the interactions between modalities and can enhance the discriminative power of the algorithm. Figure 4 illustrates these techniques and their integration within a multimodal identification algorithm based on a CNN.

The construction of the concatenation layer is based on mathematical operations as follows. Let us say that q_{m_1}- is a set formed during the modality analysis m_1 with *l* features, and a_{m_2} – is a set formed during the modality analysis m_2 with *k* features:

$$q_{m_1}=\{q_1,q_2,q_3 \ldots q_l\}$$

$$a_{m_2}=\{a_1,a_2,a_3 \ldots a_k\}$$

The next step is to form a combined set of features *p*, which is mathematically described by the following formulas:

$$p = \{q_{m_1}, a_{m_2}\},$$

$$p = \{q_1, q_2, q_3 \ldots q_l; a_1, a_2, a_3 \ldots a_k\}$$

$$p \in R^{k+l}$$

Following the concatenation stage, a subsequent fully connected layer is employed to reduce the dimensionality of the feature space. In this stage, the VGG-M architecture serves as the foundational algorithm, as depicted in Figure 5. The chosen architecture facilitates the examination of a multimodal algorithm centered on face and iris recognition. Through this algorithm, the spectral mapping of features is analyzed using the principles of image analysis, leveraging two-dimensional convolution kernels.

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

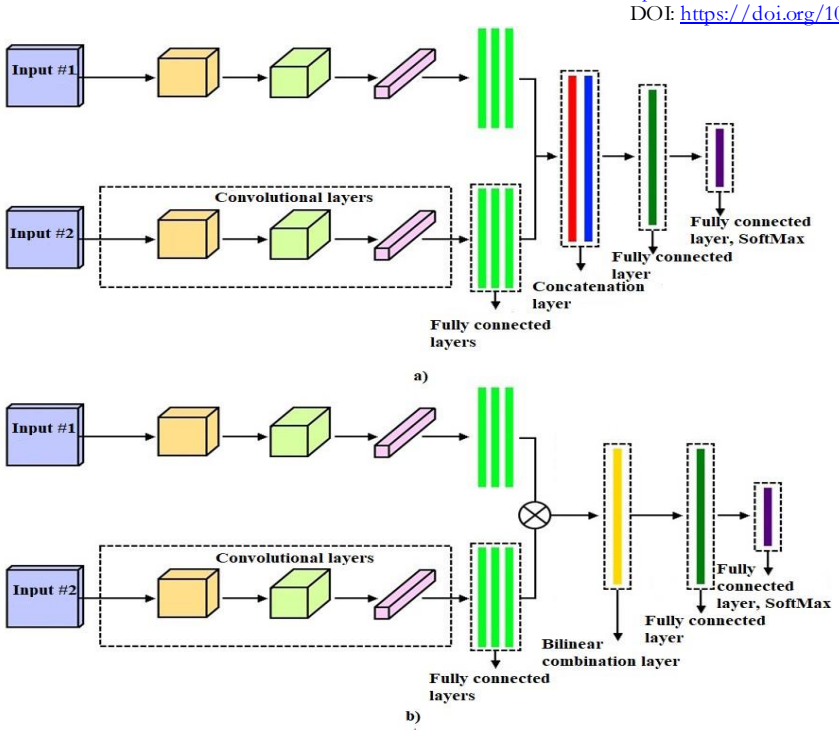

Figure 4. Combining Feature Level Modalities Using Concatenation (A) And Bilinear Combination (B) Operations

Source: compiled by the author

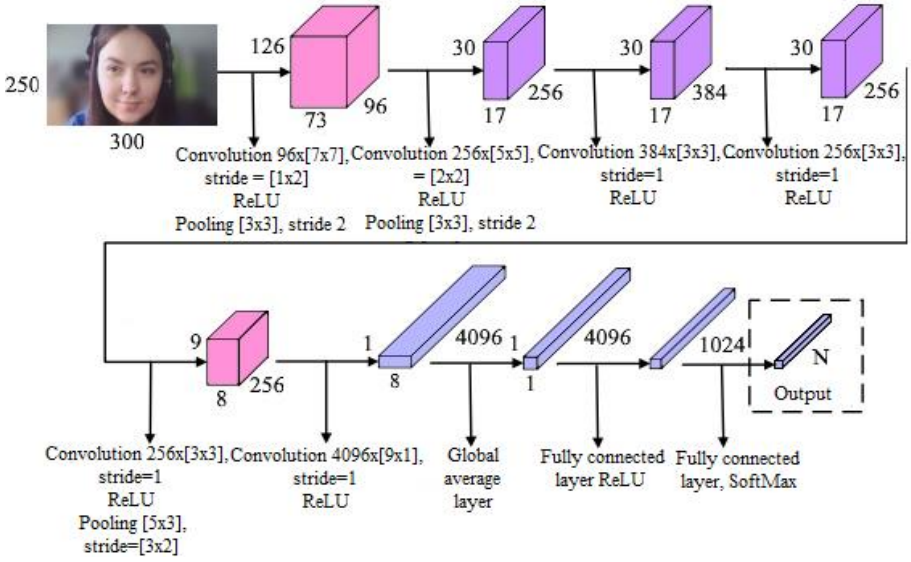

Figure 5. Architecture Of the Convolutional Neural Network VGG-M

Source: compiled by the author

The bilinear combination of feature sets q_{m_1} and a_{m_1} is described by the following formulas:

$$q_{m_1} = \{q_1, q_2, q_3 \dots q_l; a_1, a_2, a_3 \dots a_k\}, a_{m_2} = \{q_l, a_1, a_2, a_3 \dots a_k\},$$

$$p = q_{m_1} \times a_{m_2},$$

$$p = \{p_1, p_2, p_3 \dots p_n\}, p \in R^n$$

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

After that, the stage of normalization of the resultant vector p is performed:

$$f_n = \{1, p_n \geq 0 \ -1, \text{або}'$$

$$f = \{f_l, f_2, f_3 \ldots f_n\}, f \in R^n$$

$$x = f\sqrt{|p|},$$

$$z = \frac{x}{\|x\|_2},$$

With this approach, the number of features in the sets q_{m_1} and a_{m_1} should be equal.

An alternative approach involves the fusion of features extracted from various levels of convolutional layers within neural networks. This method entails combining multiple blocks of information with input data and is referred to as a multi-abstract combination. The structural diagram illustrating the configuration of the multi-abstract combination is presented in Figure 6.

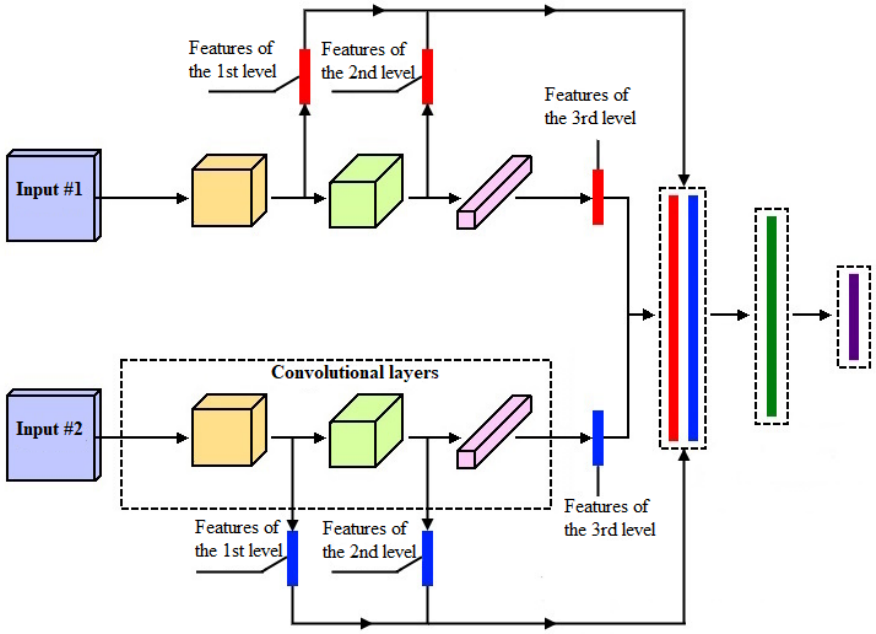

Figure 6. Building A Structural Scheme for Combining Features Based on A Multi-Abstract Combination

Source: compiled by the author

Development And Testing of Modalities at The Level of Feature Combination

The creation of a multimodal algorithm through the fusion of modalities at the feature level encompasses integrated algorithms employing multivariate analysis techniques. These algorithms encompass a combined voice activity detector, an MTCCN-based face detector (multi-task cascade convolutional networks), Mel-frequency cepstral coefficients, facial image features, a decision-making module, and more. The diagram illustrating the fusion of algorithms based on multivariate analysis for the development of a multimodal identification algorithm is depicted in Figure 7.

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

Combined Voice Activity Detector

The algorithm processes an audio fragment, also known as a phonogram, with a duration of 8 to 10 ms. Each detector analyzes the input sample, resulting in three independent predictions. These predictions are then fed into a generalized classifier, which makes the final decision on the assigned class for the input audio fragment, based on the "speech is present" or "speech is absent" principle.

The classifier in this model is an ensemble of "decision trees". To train this classifier, the Speech2Face dataset is prepared and the samples are divided into training and test sets. For the training set, a k-block cross-validation approach is used, where the parameter k determines the number of equal parts in which the training set is divided. The model is then trained on k-n parts, while the remaining part serves as the validation set. In this study, a value of 10 was chosen for the parameter k, ensuring that each of the k-n parts participates in the validation process.

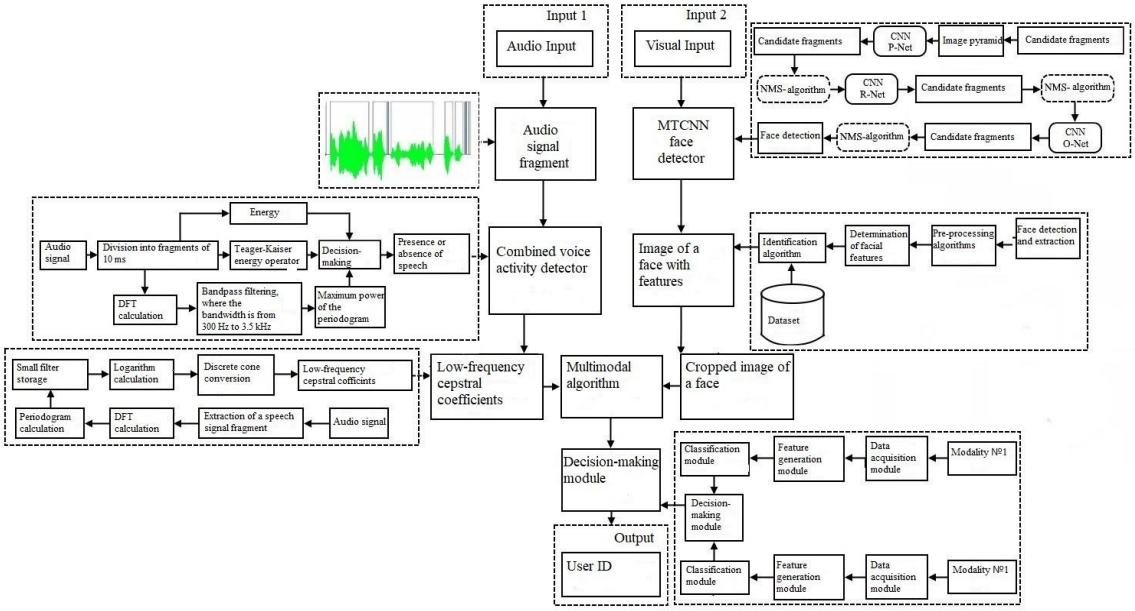

Figure 7. Scheme Of Combining Algorithms Based on Multivariate Analysis for The Development of a Multimodal Algorithm

Source: compiled by the author

Algorithm Of Mel-Frequency Cepstral Coefficients

Mel-frequency cepstral coefficients (MFCCs) are a highly effective method for processing and analyzing speech data. This algorithm concept finds applications in tasks such as speaker identification and face recognition systems. The Mel-frequencies, derived from the Mel-scale, establish a connection between human perception of sound pitch and its actual frequency measurement. The Mel-scale is utilized to construct a bank of triangular filters, which aggregate spectral energy within specific frequency ranges. These filters are applied to the periodogram, resulting in its collapse along the frequency axis.

The sequential steps of the algorithm are as follows. Firstly, a fragment of the speech signal is extracted and subjected to a discrete Fourier transform. Subsequently, the periodogram is computed. The speech signal is then passed through a low-pass filter, and the logarithm of the resulting signal is calculated. Finally, a discrete cosine transform is applied to the logarithm-transformed signal.

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

Mtccn Face Detection Algorithm

The face detection algorithm consists of a cascade of convolutional neural networks, which are described as follows. In the initial stage, the original image undergoes a resizing operation to construct an image pyramid. Subsequently, the pyramid is analyzed using a compact network that employs a cascade of a proposal network (P-Net). Each image within the pyramid is examined using a sliding window of size 12x12 pixels, which matches the input dimensions of the proposal network. The network generates output values that characterize candidate image fragments containing potential faces. To eliminate redundant overlapping fragments, the non-maximum suppression (NMS) algorithm is employed.

The remaining candidate fragments from the previous stage are resized to a unified format of 24x24 pixels and forwarded to a second, more sophisticated processing network known as R-Net. The R-Net scrutinizes these fragments and discards a larger proportion of false positive detections identified by the P-Net. The surviving candidates undergo further analysis using the NMS algorithm to eliminate redundant overlapping detections.

In the third stage, the candidate fragments are resized to a dimension of 48x48 pixels. These resized fragments undergo analysis by the output network (O-Net), which plays a crucial role in making the final decision and providing the coordinates of the detected faces. Furthermore, the O-Net also computes the coordinates of the five key facial landmarks, representing the positions of significant facial components.

Algorithm For Face Detection with Features

The video data from the Speech2Face database sets are subjected to processing using a Biologically Inspired Neural Classifier (BCM) algorithm. The algorithm follows a defined scheme for constructing a face image with distinctive features, as required by a biometric user identification system based on face analysis. The face detection process encompasses preprocessing stages and feature extraction, tailored to the specific classifier employed, to enhance the accuracy of the identification procedure. In parallel, the speech signals undergo preprocessing using a combined voice activity detector, which effectively removes pauses, noise, and interference, further refining the accuracy of the user identification algorithm.

The algorithm proceeds as follows: In the video stream, faces are detected and extracted, subsequently undergoing preprocessing to identify facial features critical for successful user identification. The processed images, previously identified by the algorithm, are stored in the dataset for reference and comparison purposes.

Algorithm Of the Decision-Making Module

The multimodal approach to combining modalities at the feature level is rooted in the analysis of biometric data from diverse sources, conducted in parallel to extract distinctive features that characterize each modality. Subsequently, these features are merged into a unified feature set, enabling effective classification and informed decision-making. The classification outcomes are then fed into the decision-making module, which orchestrates the final data processing based on principles of decisive operations.

The efficacy of the multimodal algorithm was evaluated through a series of tests that involved combining different datasets. Test 1 involved face image data, Test 2 utilized voice signals, and Test 3 incorporated voice signals with introduced distortions and interference. The outcomes of these test studies are presented in Table 1 for analysis and comparison.

Table 1. Results Of Multimodal Algorithm Testing

Test 1	Test 2	Test 3
99,83%	97,90%	88,73%

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

99,92%	98,02%	88,34%
99,90%	97,96%	88,47%

Source: compiled by the author

Based on the findings presented in Table 1, it can be inferred that Test 1, which utilized face image data, achieved the highest accuracy rates, approaching nearly 100%. This test outperformed the other tests in terms of accurately determining key parameters using the algorithm. However, Tests 2 and 3 involving voice signals exhibited a slight error in the pre-processing stage, which can be attributed to user delays during video conferences. Within a brief duration of 40-60 seconds, each participant in the video conference experienced intermittent pauses, and in some cases, external noise interference from concurrent processes occurred, significantly impacting the algorithm's signal processing accuracy. Nevertheless, the results of "Test 3" exhibited a marginal error ranging from 0.32% to 0.39%, while "Test 2" demonstrated an error range of 0.06% to 0.12%. From a speaker's perspective, the presence of pauses in the input signal is regarded as a negative factor since most neural networks rely on analyzing shorter speech fragments, typically lasting 2-3 to 4-5 seconds, to achieve high accuracy in user recognition during speech analysis.

Hence, the presented multimodal algorithm, integrated into a biometric identification system, proves to be effective in user verification research through its utilization of a combined multivariate analysis algorithm. Furthermore, comparative analysis reveals the algorithm's superior performance when compared to other similar multimodal identification algorithms.

The proposed multimodal biometric identification algorithm can function as a standalone system, leveraging logical operations such as "YES", "NO", "OR", and "AND" for combining modalities. In the case of employing the OR operator, a classification attempt is deemed successful if at least one modality correctly identifies the individual. Conversely, when using the AND operator, both modalities must accurately recognize the user for the attempt to be considered successful.

Discussion

In recent years, the significance of safeguarding personal data has escalated, underscoring the necessity for an identification system that integrates multiple biometric data sources. Such a system has been widely recommended for its potential to enhance accuracy and offer robust protection.

State-of-the-art techniques in the multivariate analysis have facilitated the exploration and advancement of multimodal algorithms, which encompass integrated algorithms and distinct stages of data acceptance, processing, and identification. By employing data pre-processing blocks and effectively combining modalities, these algorithms exhibit superior performance compared to conventional verification and identification approaches. An equally critical aspect of multimodal algorithms is the utilization of Artificial Neural Networks (ANNs) for feature fusion. Here, dedicated ANNs are employed for each modality to extract features that are specific to the respective modality. Soleymani et al. (2018) conducted a comprehensive analysis highlighting the enhanced efficiency of multimodal systems through the utilization of multiple Artificial Neural Networks (ANNs) with feature fusion at the feature level. Their research revealed that such systems outperform unimodal systems, attributing the improvement to the integration of bilinear and compact bilinear feature fusion techniques in multimodal biometric identification. To address this, the authors put forth a generalized compact bilinear fusion algorithm capable of weighted feature fusion and compact bilinear schemes. The results were conducted for the proposed algorithms on three challenging databases:

- The CMU Multi-PIE database consists of facial images under different illumination, viewpoints, and facial expressions recorded in four sessions where multiview facial images for 129 subjects were considered.

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

- The BioCop multimodal database is one of the few databases that allows for decoupled training and testing of multimodal fusion at the feature level and also allows for the use of the database in a separate period (2008, 2009, 2012, and 2013). The proposed algorithm is trained on 294 shared objects from the 2013 dataset and tested on the same objects from the 2012 dataset. tested on the same objects from the 2012 dataset. It is worth noting that although the databases are labeled as 2012 and 2013, the date of data collection for the common subjects in the datasets can vary from one to three years, which also has the advantage of investigating the effect of age progression.

- The BIOMDATA multimodal database is challenging because many samples are corrupted by blurring, occlusion, sensor noise, and shadows. The authors consider six biometric modalities: irises of the left and right eyes, as well as thumb and index fingerprints of both hands of both hands. The experiments were conducted on 219 subjects with samples for all six modalities. For each modality, four randomly selected samples are used for training, and the remaining samples are used for the test set.

Hence, the proposed generalized compact bilinear algorithm demonstrates its applicability in complex biometric identification algorithms, leveraging datasets such as CMU Multi-PIE, BioCop, and BIOMDATA multimodal databases.

Within the realm of artificial intelligence and multimodal algorithms, researchers and scientists actively seek alternative approaches to implement novel hybrid systems, aiming to enhance advanced identification and feature recognition technologies. In this context, Cherrat et al. (2020) put forward a hybrid system that combines the capabilities of tree-based efficient Artificial Neural Networks (ANNs), Softmax, and random forest classifier models within a multi-biometric fingerprint, vein, and face identification system. The authors present a comprehensive human recognition system employing ANN models alongside a multimodal biometric identification system that fuses fingerprint, finger skin, and face images. For the fingerprinting aspect, a traditional image preprocessing technique is employed to discern foreground and background using K-means and the DBSCAN algorithm. Subsequently, feature extraction is conducted through ANN and the screening method, with Softmax serving as the recognition function. Furthermore, the authors incorporate a radio frequency classifier for classification purposes.

The fusion of evaluations provided by these systems enhances the accuracy of user identification. To assess the effectiveness of the proposed algorithm, a thorough evaluation is conducted on the publicly available real-world multimodal biometric database SDUMLA-HMT, utilizing a GPU implementation. The experimental results obtained from these datasets demonstrate the substantial potential of the proposed identification biometric system. In comparison to other systems relying on unimodal, bimodal, and multimodal characteristics, the proposed work exhibits improved accuracy and efficiency in matching. Moreover, the experimental findings obtained from a real multimodal database reveal that the overall performance of the proposed multimodal system surpasses that of unimodal and bimodal biometric systems based on Convolutional Neural Networks (CNN) and various classifiers in terms of identification capabilities.

Biometric security is a critical concern in today's world, intensifying the challenges associated with data processing and utilization. Over the years, there has been a significant surge in research endeavors dedicated to biometrics. However, achieving enhanced recognition accuracy and speed in multimodal biometric technology continues to pose a notable challenge. Several studies have thus delved deeper into the exploration of multimodal biometric identification, especially in the context of smart cities. In their work, Rajasekar et al. (2022) propose an advanced multimodal biometric technique tailored for smart cities, utilizing a hierarchical combination of levels and modules. The authors' approach aims to address existing challenges by employing a multimodal fusion technique empowered by an optimized fuzzy genetic algorithm, thereby enhancing system performance. The analysis of the obtained results demonstrates that the proposed approach outperforms existing methods in various performance metrics, including false acceptance rate, false rejection rate, equal error rate, precision, prediction, and reliability. The authors introduce a scheme that achieves an impressive accuracy rate of 99.88% while significantly reducing the error rate to just 0.18%. The matching scores obtained for fingerprint and iris biometrics, considering a

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

substantial number of subjects, affirm the substantial potential of this integrated approach. The utilization of an optimized fuzzy genetic algorithm in multimodal biometrics enables an efficient fusion strategy, resulting in precise and reliable biometric recognition.

Conclusion

To devise a multimodal algorithm utilizing modality fusion, we curated a dataset comprising video and audio data using the Speech2Face software. These software tools were specifically designed for generating, developing, and evaluating identification algorithms based on neural networks. For our research materials, we gathered a collection of video dataset images.

The development of the multimodal algorithm through feature-level modality fusion entails the integration of various algorithms rooted in multivariate analysis. These include a combined voice activity detector, an MTCNN (multi-task cascade convolutional networks) face detector, Mel-frequency cepstral coefficients, a face image with associated features, and a decision-making module. We propose a scheme that combines these algorithms based on multivariate analysis to establish a robust multimodal identification algorithm.

Based on the obtained data, it is evident that Test 1, which involved face image data, achieved the highest processing rates, nearing nearly 100%. This test outperformed the others in accurately determining key parameters through the algorithm. In Tests 2 and 3, which utilized voice signals, a slight error was observed during the pre-processing stage. This error can be attributed to the delay introduced by participants during the video conference. Within a short time frame of 40-60 seconds, each participant occasionally paused, and in some instances, noise interference caused by external factors was present. These factors significantly impacted the accuracy of signal processing by the algorithm. However, the results of "Test 3" exhibited a negligible error in the tens range, ranging from 0.32% to 0.39%, while "Test 2" had an error rate of 0.06% to 0.12%.

The proposed multimodal algorithm, built upon a biometric identification system, enables successful user verification research through the utilization of a combined multivariate analysis algorithm. It showcases precise research outcomes when compared to other analogous multimodal identification algorithms.

References

Abdullah, S. S., Ameen, S. Y., Sadeeq, M. M. A., & Zeebaree, S. R. (2021). Multimodal emotion recognition using deep learning. Journal of Applied Science and Technology Trends, 2(2), 52–58. http://doi.org/10.38094/jastt20291

Alay, N., & Al-Baity, H. H. (2020). Deep learning approach for multimodal biometric recognition system based on the fusion of iris, face, and finger vein traits. Sensors, 20(19), 5523. http://doi.org/10.3390/s20195523

Ammour, B., Boubchir, L., Bouden, T., & Ramdani, M. (2020). Faceiris multimodal biometric identification system. Electronics, 9(1), 85. https://doi.org/10.3390/electronics9010085

Bayoudh, K., Knani, R., Hamdaoui, F., & Mtibaa, A. (2022). A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. The Visual Computer: International Journal of Computer Graphics, 38, 2939–2970. https://doi.org/10.1007/s00371-021-02166-7

Behrad, F., & Abadeh, M. S. (2022). An overview of deep learning methods for multimodal medical data mining. Expert Systems with Applications, 200, 117006. https://doi.org/10.1016/j.eswa.2022.117006

Boyko, N. (2021). Models and algorithms for multimodal data processing. WSEAS Transactions on Information Science and Applications, 20, 87–97. http://doi.org/10.37394/23209.2023.20.11

Boyko, N. (2023a). Research into machine learning algorithms for the construction of mathematical models of multimodal data classification problems. JCPEE, 11(2), 1–11. https://doi.org/10.23939/jcpee2021.02.001

Boyko, N. (2023b). Study of multimodal identification algorithms using modern methods and tools of multivariate analysis. Diqiu Kexue, 48(6), 31–44. https://doi.org/10.3799/dqkx.2023.8118

Cherrat, E., Alaoui, R., & Bouzahir, H. (2020). Convolutional neural networks approach for multimodal biometric identification system using the fusion of fingerprint, finger-vein, and face images. PeerJ Computer Science, 6, e248. http://doi.org/10.7717/peerj-cs.248

Gaonkar, A., Chukkapalli, Y., Raman, P., Srikanth, S., & Gurugopinath, S. (2021). A comprehensive survey on multimodal data representation and information fusion algorithms. In 2021 International Conference on Intelligent Technologies (CONIT), pp. 1–8. Hubli, India. http://doi.org/10.1109/CONIT51480.2021.9498415

Iatsyshyn, A., Iatsyshyn, A., Kovach, V., Zinovieva, I., Artemchuk, V., Popov, O., … Turevych, A. (2020). Application of open and specialized geoinformation systems for computer modelling studying by students and PhD students. Paper presented at the CEUR Workshop Proceedings, 2732, 893–908. https://doi.org/10.31812/123456789/4460

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

Kovaleva, V., Bukhteeva, I., Kit, O. Y., & Nesmelova, I. V. (2020). Plant defensins from a structural perspective. International Journal of Molecular Sciences, 21(15), 1–23. http://doi.org/10.3390/ijms21155307

Kumar, G., & Jeyakumar, V. (2022). Multimodal, Multianatomical, and Multidimensional Medical Image Retrieval System. Authorea. December 22, 2022. http://doi.org/10.22541/au.167169946.63655650/v1

Latysheva, O., Rovenska, V., Smyrnova, I., Nitsenko, V., Balezentis, T., & Streimikiene, D. (2020). Management of the sustainable development of machine-building enterprises: A sustainable development space approach. Journal of Enterprise Information Management, 34(1), 328–342. http://doi.org/10.1108/JEIM-12-2019-0419

Lu, L., Mao, J., Wang, W., Ding, G., & Zhang, Z. (2020). A study of personal recognition method based on EMG signal. In IEEE Transactions on Biomedical Circuits and Systems, 14(4), 681–691. http://doi.org/10.1109/TBCAS.2020.3005148

Masood, M., Nawaz, M., Malik, K. M., Javed, A., Irtaza, A., & Malik, H. (2023). Deepfakes generation and detection: State-of-the-art, open challenges, countermeasures, and way forward. Applied intelligence, 53(4), 3974–4026. https://doi.org/10.1007/s10489-022-03766-z

Maiti, D., Basak, M., & Das, D. (2024). Multimodal Biometric Integration: Trends and Insights from the Past quinquennial. International Journal of Computing and Digital Systems, 16(1), 189–198.

Nawal, S. H., Noria, T., & Abdelkader, I. K. M. (2024). Behavioral biometrics to detect fake expert profiles during negotiation. Multimedia Tools and Applications, pp. 1–20. Springer International Publishing. https://doi.org/10.1007/s11042-024-18644-8

Ostapenko, R., Herasymenko, Y., Nitsenko, V., Koliadenko, S., Balezentis, T., & Streimikiene, D. (2020). Analysis of production and sales of organic products in Ukrainian agricultural enterprises. Sustainability (Switzerland), 12(8). https://doi.org/10.3390/su12083416

Pandey, P., Raghav, Y. Y., Gulia, S., Aggarwal, S., & Kumar, N. (2024). Supervised and Unsupervised Learning Techniques for Biometric Systems. Supervised and Unsupervised Data Engineering for Multimedia Data, pp. 263–299. Wiley. https://doi.org/10.1002/9781119786443.ch12

Park, K. S. (2023). Authentication with Bioelectrical Signals. In Humans and Electricity: Understanding Body Electricity and Applications, pp. 249–273. Springer International Publishing. https://doi.org/10.1007/978-3-031-20784-6_11

Pereira, T. M., Conceição, R. C., Sencadas, V., & Sebastião, R. (2023). Biometric recognition: A systematic review on electrocardiogram data acquisition methods. Sensors, 23(3), 1507. https://doi.org/10.3390/s23031507

Popovych, I., Pavliuk, M., Hrys, A., Sydorenko, O., Fedorenko, A., & Khanetska, T. (2021). Pre-game expected mental states in men's mini-football teams: A comparative analysis. Journal of Physical Education and Sport, 21(2), 772–782. http://doi.org/10.7752/jpes.2021.02096

Rafatirad, S., Homayoun, H., Chen, Z., & Pudukotai Dinakarrao, S. M. (2022). Unsupervised Learning. In Machine Learning for Computer Scientists and Data Analysts: From an Applied Perspective, pp. 163–216. Cham: Springer International Publishing. http://doi.org/10.1007/978-3-030-96756-7

Rajakumar, G., & Ananth Kumar, T. (2022). Design of advanced security system using vein pattern recognition and image segmentation techniques. In Kumar, N., Shahnaz, C., Kumar, K., Abed Mohammed, M., Raw, R. S. (Eds.), Advance Concepts of Image Processing and Pattern Recognition. Transactions on Computer Systems and Networks, pp. 213–225. Springer Singapore. https://doi.org/10.1007/978-981-16-9324-3_12

Rajasekar, V., Predić, B., Saracevic, M., Elhoseny, M., Karabasevic, D., Stanujkic, D., & Jayapaul, P. (2022). Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm. Scientific Reports, 12(1), 1–11. http://doi.org/10.1038/s41598-021-04652-3

Reel, P. S., Reel, S., Pearson, E., Trucco, E., & Jefferson, E. (2021). Using Machine learning approaches for multi-omics data analysis: A review. Biotechnology Advances, 49. https://doi.org/10.1016/j.biotechadv.2021.107739

Rudregowda, S., Patilkulkarni, S., Ravi, V., Gururaj, H. L., & Krichen, M. (2024). Audiovisual speech recognition based on a deep convolutional neural network. Data Science and Management, 7(1), 25–34. https://doi.org/10.1016/j.dsm.2023.10.002

Safavipour, M. H., Doostari, M. A., & Sadjedi, H. (2022). A hybrid approach to multimodal biometric recognition based on feature-level fusion of face, two irises, and both thumbprints. Journal of Medical Signals and Sensors, 12(3), 177–191. http://doi.org/10.4103/jmss.jmss_103_21

Sardinha, T. B., & Pinto, M. V. (2019). Multi-Dimensional Analysis: Research Methods and Current Issues. Bloomsbury Academic. https://doi.org/10.5040/9781350023857

Shoumy, N. J., Ang, L. M., Seng, K. P., Rahaman, D. M., & Zia, T. (2020). Multimodal Big data effective analytics: A Comprehensive survey using text, audio, visual and physiological signals. Journal of Network and Computer Applications, 149, 102447. https://doi.org/10.1016/j.jnca.2019.102447

Shytyk, L., & Akimova, A. (2020). Ways of transferring the internal speech of characters: Psycholinguistic projection. Psycholinguistics, 27(2), 361–384. http://doi.org/10.31470/2309-1797-2020-27-2-361-384

Soleymani, S., Torfi, A., Dawson, J., & Nasrabadi, N. M. (2018, October). Generalized bilinear deep convolutional neural networks for multimodal biometric identification. In 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 763–767. IEEE. http://doi.org/10.1109/ICIP.2018.8451532

Zhao, F., Zhang, C., & Geng, B. (2024). Deep Multimodal Data Fusion. ACM Computing Surveys, 56(9), 1–36. https://doi.org/10.1145/3649447

Tanveer, H., Adam, M. A., Khan, M. A., Ali, M. A., & Shakoor, A. (2023). Analyzing the Performance and Efficiency of Machine Learning Algorithms, such as Deep Learning, Decision Trees, or Support Vector Machines, on Various Datasets and Applications. The Asian Bulletin of Big Data Management, 3(2), 126–136. https://doi.org/10.62019/abbdm.v3i2.83

Journal of Ecohumanism
2024
Volume: 3, No: 5, pp. 99 – 114
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v3i5.3877

Ugryumov, M., Chernysh, S., Strilets, V., Meniailov, I., & Ugryumova, K. (2020). Methods of machine learning in problems of system analysis and decision-making. Kharkiv National University named after V. N. Karazina. https://www.researchgate.net/publication/345100769_METODI_MASINNOGO_NAVCANNA_U_ZADACAH_SISTEMNOGO_ANALIZU_I_PRIJNATTA_RISEN

Wei, J., Huang, H., Wang, Y., He, R., & Sun, Z. (2022). Towards more discriminative and robust iris recognition by learning uncertain factors. IEEE Transactions on Information Forensics and Security, 17, 865–879. http://doi.org/10.1109/TIFS.2022.3154240

Wu, Z., Wu, X., Zhu, Yu., Zhai, J., Yang, H., Yang, Z., Wang, C., & Sun, J. (2022). Research on Multimodal Image Fusion Target Detection Algorithm Based on Generative Adversarial Network. Wireless Communications and Mobile Computing, 2022(1), art. ID 1740909. https://doi.org/10.1155/2022/1740909.