Pak. j. life soc. Sci. (2025), 23(2): 872-880 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences

www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2025-23.2.00218

RESEARCH ARTICLE

Development of biometric identification tools for solving information security problems based on multimodal neural networks

Anna Poluyan[1*], Sofya Petrenkova[1], Kseniia Korovina[1]

[1]Don State Technical University, Rostov-on-Don, Russia

ARTICLE INFO

Received: Aug 19, 2025

Accepted: Oct 12, 2025

Keywords

Authentication
Biometric identification
Biometrics
Vulnerabilities
Multimodality
Unimodality

***Corresponding Author:**

AnnaPoluyan1@mymail.academy

ABSTRACT

Biometric identification is currently one of the most pressing issues in information security. Biometric identification, unlike traditional methods (passwords, tokens), offers a fundamentally different approach to authentication based on the uniqueness of human physiological and behavioral characteristics. Its implementation is accelerating in the commercial sector (for example, fingerprint payments), at the government level (electronic passports with biometrics) and in consumer electronics (facial recognition in smartphones). According to Juniper Research, by 2026, more than 4 billion devices will use biometric identification. The biometric technology market, according to MarketsandMarkets, will reach $82.9 billion by 2027, reflecting the growing demand for reliable and convenient solutions to ensure the security of biometric data. However, traditional methods are vulnerable to, for example, spoofing, adversarial attacks (fakes and synthesized voices), noise or low data quality. The article proposes measures to solve the problem of biometric identification using artificial intelligence tools, namely multimodal language models, which are currently one of the best methods in the field of machine learning. A comparative analysis is conducted and the advantages of using a multimodal approach compared to unimodal systems are indicated. Data protection measures are proposed and the effectiveness of this approach is assessed.

INTRODUCTION

Biometric identification is a process of identity verification based on the analysis of unique biological characteristics, which can be divided into:

- Physiological: fingerprints, iris structure, facial geometry, palm vein pattern, DNA.

- Behavioral: typing dynamics, gait, and voice patterns.

The key advantage of biometrics is that biometric features are intrinsically linked to the user: they cannot be forgotten (like a password) or accidentally lost (like a token). However, digital representations of this data (fingerprint templates, 3D face models) can be compromised if leaked from databases. For example, in 2019, 28 million biometric records were leaked from the BioStar 2 access control system. This highlights the importance of securing biometric templates through cryptography and conversion to irreversible forms (hashing).

Unlike a password, a biometric template cannot be "changed" after a compromise, which requires special storage approaches (for example, the use of cancelable biometrics or the use of cryptographic techniques). FIDO2 standards provide for the storage of biometric data only on the user's local device (for example, in the Secure Enclave of the iPhone), which minimizes the risk of centralized leaks.

In Today's World, Technology Faces Challenges:

- Technical: false positives/failures, vulnerabilities to spoofing attacks (for example, 3D masks to deceive facial recognition systems).

- Ethical: risks of privacy violations (collection and storage of biometric data), discrimination due to errors in the recognition of certain groups of the population (Kaspersky, 2019; Habr, 2023).

Integrating biometrics into security systems requires compliance with standards (ISO/IEC 19794 for data templates) and regulatory norms (GDPR governing the processing of personal data).

MATERIALS AND METHODS

Biometric identification, based on the analysis of unique physiological and behavioral characteristics, faces a number of key threats, such as:

• Vulnerability to spoofing attacks (3D masks, synthesized voices, digital distortion).

• Attacks on voice biometrics (attacks based on speech synthesis, voice deep fake).

• Attacks on behavioral biometrics (imitation of typing dynamics, Gan synthesis of the mouse gesture).

• Irreversibility of biometric template compromise.

• Ethical risks associated with the use of artificial intelligence (discrimination, privacy issues).

Facial recognition is one of the most popular methods of biometric identification. With the development of computer vision and deep learning, the facial recognition system has moved from classical methods (for example, algorithms based on histograms or Eigenfaces) to modern neural networks. To date, the ANPR (Detection-Segmentation-Recognition) process has been significantly improved due to the introduction of neural network components, which demonstrate significantly higher accuracy compared to classical approaches.

The Main Components of A Modern Facial Recognition System Include the Following Technologies (Poluyan Et Al., 2023):

Face detection - determining the area of the image where the face is present. Modern algorithms such as SSD (Single Shot MultiBox Detector) based on Caffe provide high speed and accuracy even in real time.

Alignment and normalization - Bring the face image to a standard view using keypoint detection algorithms (e.g., dlib with shape_predictor_68_face_landmarks model). This eliminates variations in position, scale, and orientation.

Face Embedding is the transformation of an image of a face into a compact vector of fixed dimension (e.g., a 128-dimensional vector) that encapsulates unique facial features. This is done by using deep neural network models such as dlib_face_recognition_resnet_model_v1.

Classification - Using machine learning techniques such as Support Vector Machine (SVM), embeddings are compared to determine whether a given person is owned by a particular user or not.

At the same time, unimodal systems, the most common today, have significant limitations in the accuracy of the results obtained and ensuring the security of the processed data. Modern multimodal neural networks have surpassed their predecessors, they are pretrained on huge data arrays, data identification occurs on two or more biometric features, which makes it possible to increase the accuracy, stability of algorithms and improve reliability against threats of unauthorized access (Varga and Moore, 1990; Soleymani et al., 2018). The main components of biometric data recognition by the Ministry of Taxes and Taxes are being developed in two directions: sequentially and in parallel (Hammad et al., 2018; Stefanidi et al., 2020). Sequential architecture is a waterfall model in which data from different modalities is processed sequentially, then combined for analysis. At each successive iteration, each successive modality is already working with the reduced data sets. The advantage of this approach is a reduction in the computational load and the ability to dynamically change the order of modalities. The disadvantages include the fact that errors are not corrected at the early stages of the system's operation, the optimal sequence depends on the order in which the modalities are processed.

In parallel construction of multimodal neural networks, biometric information of different nature is analyzed simultaneously (Lin et al., 2015; Soleymani et al., 2018; Park et al., 2019). The algorithm of biometric identification parallel to the MHC can be described as follows:

1.At the initial stage, the initialization process takes place, i.e. the necessary processing modules for each biometric modality are loaded into the system, on which the starting weight coefficients are set, the parameters for the normalization of estimates and the processing time are determined.

2.Thereafter, all biometric data is independent in parallel streams under the following conditions:

- A failure in one modality does not affect the processing of others;
- All data is timestamped;
- Each thread evaluates the reliability of the result.

3.For each modality obtained, key features are detected, accuracy data are recorded, a data quality score is calculated, and the consistency of the extracted features is checked.

4.Based on the results of the implementation of paragraphs 1-3, the weight coefficients of the obtained modalities are checked and adjusted according to the following formula:

$$K_m = \frac{Modality\_reliab}{\sum Modality\_reliab}$$

where Km is the weight of the modality, Modality_reliab is the certainty of the modality, ∑Modality_reliab is the sum of the certainties of all modalities.

The following formula is often used in the calculation (Neubeck and Van Gool, 2006): Modality_reliab

$$Modality\_reliab = a * QalitySore + b * ClassifierConfidence + c * ConsistencyScore$$

where a+b+c=1, QalitySore is the technical quality of the data; – "certainty of the classifier" – consistency with other modalities. $ClassifierConfidenceConsistencyScore$

Their combination allows you to dynamically adapt the weights so that the most reliable modalities have a stronger impact on the result. At the same time, modalities with low quality receive a reduced weight, if the modality fails, its weight is redistributed, the minimum total weight of active modalities should be ≥ 0.5.

The merge mechanism is implemented with the help of. The method of merging with weight normalization, which ensures that the sum of the weights is equal to one, retains the influence of each modality.

1. After adaptive weighting of the results, they are combined according to the formula

Final_result = ∑(Km×Norm_score), where K is the weight of the modality, Norm_score is the normalized estimate.

2. The final stage is the decision-making block, which consists of the following steps:

- Comparison of the final result with the threshold value (in test mode, the indicator was 0.80, but can be adjusted);
- Identification of possible attacks;
- making a final decision (identified / not identified).

RESULTS AND DISCUSSION

Based on the above material, a program for biometric identification and face recognition has been developed (Fig. 1). The developed program uses two key technologies:

OpenCV - a framework for image and video processing, which allows you to perform face detection, image conversion, creation of blobs for neural networks and work with ML libraries (for example, SVM).

Dlib is a computer vision library that provides pre-trained models for face recognition, such as a ResNet model for extracting 128-dimensional embeddings and a model for determining key points of the face dlib_face_recognition_resnet_model_v1 res10_300x300_ssd_iter_140000. Converts the face image to an embedding vector. Embeddings are normalized (L2 normalization) and are used to train an SVM model that classifies embeddings as user-owned ("GOOD") or non-user-owned ("BAD").

```
cmake_minimum_required(VERSION 3.19)
project(Diplom LANGUAGES CXX)

find_package(Qt6 6.5 REQUIRED COMPONENTS Core Widgets)
find_package(OpenCV REQUIRED)
find_package(OpenSSL REQUIRED)
find_package(dlib REQUIRED)
find_package(Eigen3 REQUIRED)

qt_standard_project_setup()

qt_add_executable(Diplom
    WIN32 MACOSX_BUNDLE
    main.cpp
    mainwindow.cpp
    mainwindow.h
    mainwindow.ui
    .gitignore
)

target_link_libraries(Diplom
    PRIVATE
        Qt6::Core
        Qt6::Widgets
        OpenSSL::SSL
        OpenSSL::Crypto
        ${OpenCV_LIBS}
        dlib::dlib
        Eigen3::Eigen
)

target_include_directories(${PROJECT_NAME} PRIVATE
    ${OpenCV_INCLUDE_DIRS}
    ${OPENSSL_INCLUDE_DIR}
)

include(GNUInstallDirs)

install(TARGETS Diplom
    BUNDLE  DESTINATION .
    RUNTIME DESTINATION ${CMAKE_INSTALL_BINDIR}
    LIBRARY DESTINATION ${CMAKE_INSTALL_LIBDIR}
)

qt_generate_deploy_app_script(
    TARGET Diplom
    OUTPUT_SCRIPT deploy_script
    NO_UNSUPPORTED_PLATFORM_ERROR
)
install(SCRIPT ${deploy_script})
```

Fig.1. Example of the Implementation of the Biometric Identification and Face Recognition Program

Using an SSD detector allows you to quickly and accurately find areas with faces. The advantage of this model is its ability to process a video stream in real time, which is critical for access control systems. The detector converts the input image into blob format (using the cv::d nn::blobFromImage function), after which the neural network outputs the coordinates of the rectangles in which the faces are detected.

The dlib_face_recognition_resnet_model_v1 model is trained on a huge number of images and is capable of converting a face image into a 128-dimensional vector, where the distance between the vectors for the same face is minimal and significant for different faces. This allows such embeddings to be used for clustering and subsequent classification.

The support vector method (SVM) is one of the classical machine learning algorithms for binary classification (Goodfellow et al., 2014; Eykholt et al., 2018; Keanini, 2019). In this case, the SVM is trained to distinguish between two classes: "self-face" and "alien face" based on the following algorithm:

1. Pre-processing of embeddings. Embeddings are normalized to the L2 norm, which improves the stability of the classification.

2. SVM parameters. A linear kernel is used, since embeddings are usually linearly separated in space.

3. Training. The Adam (Adaptive Moment Estimation) algorithm was used for training, it calculates the adaptive learning rate based on the first and second moments of gradient descent (k1, k2). After the training sample is formed, the SVM is trained, and the model is saved for later use in real time. The initial learning rate is 0.0001, the coefficient of adaptation of moments k1=0.9 is gradient

averaging (smoothes out the "noise"), k2=0.999 is taking into account the squares of the gradient for adaptive step adjustment. Weights are updated with learning and L2 normalization. These parameters were experimentally selected for biometric identification tasks.

4. Image normalization is used to improve the quality of feature extraction and reduce the influence of external factors (lighting, noise). Converting images to blob format and then L2 normalizing embeddings results in more consistent features, which improves recognition accuracy.

5. To assess the quality of training, the Binary Cross-Entropy loss function was used, which allows the algorithm to understand the moments of deviation from the correct learning process and adjust the weights. It is calculated according to the formula.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N} \quad y_t \cdot log\ log\ (y_p) + (1 - y_t) \cdot log\ log\ (1 - y_p)$$

where N is the total number of samples in the dataset;

yt is the actual label for each sample (either 0 or 1);

yp is the predicted probability (model output);

log is a function of the natural logarithm.

The proposed MHC during testing contained two layers for processing one-dimensional data, creating a convolution core with a single dimension. The first filter contained from 32 to 128 neurons, the second from 16 to 64. The output fully connected neural network has 1 to 3 hidden layers, with 128, 64, 32 neurons, respectively. The best result was obtained in a model with filters of 64 and 32, respectively, and having 3 hidden layers, with 20 learning epochs. It took the longest time to train this model with equal numbers of epochs. A graph of loss and accuracy in the training process on the training and validation samples is presented in Figures 2 and 3, respectively.

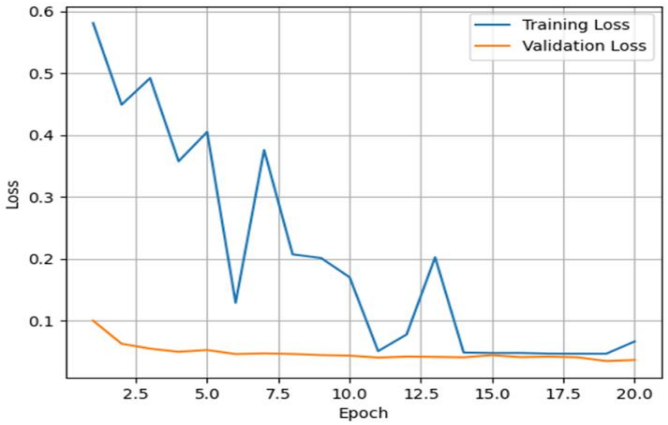

Fig. 2. Loss Graph in Training and Validation Sampling

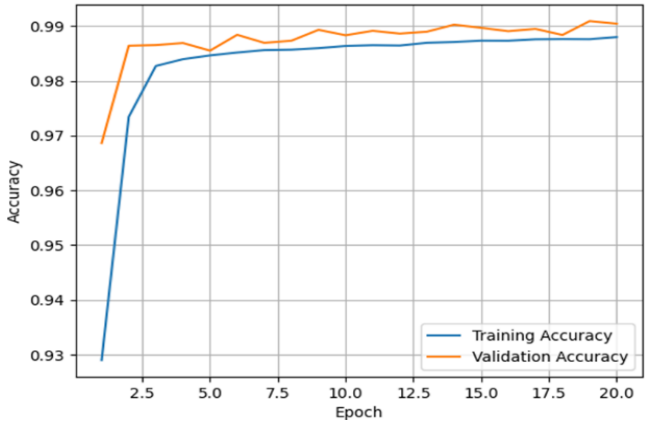

Fig. 3. Accuracy Graph on Training and Validation Samples

Recently, attacks on neural networks that process biometric data have become more and more widespread. The study (Bal, 2024) demonstrates the Fast Gradient Sign Method (FGSM) attack. This

method uses the gradient of the loss function to identify the least noticeable changes that are most disorienting to the model. Adding invisible harmful noise to the input data leads to an erroneous classification of the object. The FGSM method is used not only as a means to demonstrate the vulnerability of AI models, but also as part of the creation of more resilient models. For example, in research practice, the FGSM method is used for Adversarial Training, where the model is trained not only on ordinary data, but also on modified examples created using FGSM. This allows you to increase the resistance of the model to attack influences.

To ensure the security of biometric data, it is proposed to use a combination of static rules and ML analysis, which allows you to stop attacks in real time without noticeable impact on the system. The algorithm for implementing the attack protection unit in a multimodal biometric system is as follows:

1. At the start of the system, attack detectors for each modality, signature analysis of known attacks, and heuristic rules for detecting suspicious activities are loaded and initialized.

2. The implementation of the process of stream data processing at the physical layer is presented in Fig. 4.

```
def check_physical_artifacts(sensor_data):
    # Для видеопотока
    if sensor_data.modality == 'face':
        # Проверка параметров изображения
        if detect_photo_attack(frame):
            return AttackAlert.PHOTO_ATTACK
        if detect_video_replay(frame_sequence):
            return AttackAlert.REPLAY_ATTACK

    # Для аудиопотока
    elif sensor_data.modality == 'voice':
        if detect_tts_artifacts(audio_waveform):
            return AttackAlert.TTS_SYNTHESIS
        if detect_voice_conversion(audio_features):
            return AttackAlert.VOICE_CONVERSION
```

Fig.4. The Process of Stream Data Processing at the Physical Level

3. Comparison of timestamps of synchronization of modalities is based on the principle of a limited time window of cross-correlation (Fig.5), where:

- For each modality, the exact timestamp of the moment of data capture is recorded, the maximum discrepancy between the extreme labels in the group of modalities is calculated, which should not exceed 0.1 seconds.
- If the threshold is exceeded, an alarm is generated.
- The system switches to an increased verification mode (re-verification of biometric data, activation of additional verification methods, event logging).

```
# Сравнение временных меток:

def check_temporal_alignment(modality_results):
    max_time_diff = 0.1  # 100 мс
    timestamps = [m.timestamp for m in modality_results]

    if max(timestamps) - min(timestamps) > max_time_diff:
        return AttackAlert.TIMING_ANOMALY

# Анализ согласованности:

def check_crossmodal_consistency(results):
    # Пример: сравнение движений губ и аудио
    lip_movement = results['face'].lip_activity
    voice_activity = results['voice'].vad

    if correlation(lip_movement, voice_activity) < 0.7:
        return AttackAlert.LIP_SYNC_ANOMALY
```

Fig.5 Timestamp comparison

4. An assessment of risks aimed at prompt response to possible threats is proposed:

Table 1. Assessment of Risks Aimed at Prompt Response to Possible Threats

Threat level	Actions
Low (threshold < 0.3)	Logging, weight increase of other modalities
Medium (0.3< threshold <0.8)	Request an additional authentication factor
High (threshold ≥ 0.8)	Block attempts, transfer information to the administrator, collect evidence

5. Implementation of adaptive system protection (Adversarial Training):

- Daily update of signature databases;
- Training detectors on new threats and attacks (20% of the data from the training sample were specially added "hacked" samples);
- Anomaly detection – the input sample is checked to what extent the input sample is identical to normal data;
- Regular logging and auditing (report on attack attempts, automatic collection of evidence, immediate notification when the threshold increases> 0.7);
- Automatic calibration of thresholds;
- account lockouts in the event of a series of attempted attacks.

The main quality metrics were Precision = = $0.991\frac{\mathrm{TP}}{TP+FP}$

Recall (completeness) = = $0.987,\frac{TP}{TP+FN}$

where TP is the number of positive positives, FP is the number of false positives, and FN is the number of false negatives.

As we can see from the tests carried out in a laboratory environment, the proposed algorithm for ensuring the security of biometric data has a number of advantages over unimodal systems, such as the use of integration of signature analysis and heuristic rules for detecting attacks, adaptive adjustment of threshold values for new threats, and a balanced risk assessment system.

Table 2. Comparative Characteristics of the Effectiveness of the Protection System

Criteria	Developed algorithm	Signature method	Neural networks (based on unsupervised learning)
Detecting Known Attacks	96,7%	84,2%	89,4%
Detect unknown threats	81%	18,5%	73,6
Average Detection Time (ms)	65	48	82
Hardware Requirements	Average	Low	High

The detection accuracy of known attacks is 96.7% when tested on the Wide Multi-Channel Presentation Attack Database (WMCA) public datasets, which contain 1679 10-second videos, of which 347 contain real people, and 1332 are examples of attacks on biometric data; The LFW dataset contains 13,233 images of faces collected from the web. This dataset consists of 5,749 identifications with 1,680 people with two or more images. For unknown attacks, 81% is achieved through the integration of anomaly analysis and machine learning techniques.

When developing this program, the developers faced the following difficulties:

1. High Computational Load. Real-time video stream processing requires significant computing resources. To solve this problem, the system implements:

- Frame Drop: Not every frame is processed by expensive embeddings detection and extraction operations.
- Asynchronous processing: Some operations, such as saving images or training a model, can be performed on a separate thread or pause the video stream to avoid dropping FPS.

2. Data balance and quality. To successfully train an SVM model, a balanced dataset is required (Demetrio et al., 2018; Skylight Cyber, 2019; Purchina et al., 2023). Normalizing embeddings helps smooth out differences caused by variations in lighting and posture.

CONCLUSION

The development of biometric identification tools, in particular, facial recognition systems, is a complex but extremely promising area that can significantly improve the level of information security. Modern methods based on the MNC make it possible to create high-precision systems that are resistant to changes in environmental conditions and potential attacks. The results of the study showed that the combination of several modalities significantly increases the level of security and reduces the likelihood of unauthorized access. However, the successful implementation of such systems requires not only technical solutions, but also strict compliance with regulatory and ethical standards.

REFERENCES

Bal H., 2024. Deep dive into IDS-IPS architecture: Building a robust network defense. Hostomize. URL: https://hostomize.com/blog/ids-ips-architecture/

Demetrio L., Biggio B., Lagorio G., Roli F. and Armando A., 2018. Exploring adversarial examples in malware detection. URL: https://arxiv.org/abs/1810.08280

Eykholt K., Evtimov I., Fernandes E., Li B., Rahmati A., Xiao C., Prakash A., Kohno T. and Song D., 2018. Robust physical-world attacks on deep learning visual classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp: 1625–1634.

George A., Mostaani Z., Geissenbuhler D., Nikisins O., Anjos A. and Marcel S., 2019. Biometric face presentation attack detection with multi-channel convolutional neural network. IEEE Transactions on Information Forensics and Security, 15: 42–55.

Goodfellow I., Shlens J. and Szegedy C., 2014. Explaining and harnessing adversarial examples. IRL: https://arxiv.org/abs/1412.6572

Habr, 2023. Methods of attack on AI. Habr Articles. URL: https://habr.com/ru/articles/778234/

Hammad M., Liu Y. and Wang K., 2018. Multimodal biometric authentication systems using convolutional neural network based on different level fusion of ECG and fingerprint. IEEE Access, 7: 26527–26542.

Kaspersky, 2019. Attacks on artificial intelligence. Kaspersky Whitepaper. URL: https://media.kaspersky.com/ru/business-security/attacks-on-artificial-intelligence-whitepaper.pdf

Keanini T.K, 2019. The state of machine learning. Cisco Blogs. URL: https://blogs.cisco.com/security/the-state-of-machine-learning-in-2019

Lin T.Y., RoyChowdhury A. and Maji S., 2015. Bilinear CNN models for fine-grained visual recognition. Proceedings of the IEEE International Conference on Computer Vision, pp: 1449–1457.

Matveev Yu.N., 2012. Technology of biometric identification of a person by voice and other modalities [Tekhnologiya biometricheskoy identifikatsii lichnosti po golosu i drugim modal'nosti]. Instrument Engineering, 3: 46–61.

Neubeck A. and Van Gool L., 2006. Efficient non-maximum suppression. 18th International Conference on Pattern Recognition (ICPR'06), vol. 3, pp: 850–855.

Park D.S., Chan W., Zhang Y., Chiu C.C., Zoph B., Cubuk E.D. and Le Q.V., 2019. SpecAugment: A simple data augmentation method for automatic speech recognition. arXiv preprint arXiv:1904.08779.

Poluyan E.N., Tseligorova V.V., Galushka N.M. and Kodatsky N.M., 2023. [Modern approaches in biometric recognition]. Modern Science: Actual Problems of Theory and Practice. Series: Natural and Technical Sciences, 8-2: 123–127.

Powers D.M.W., 2020. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.

Purchina O.A., Poluyan A.Y. and Fugarov D.D., 2023. The use of artificial intelligence in ensuring information security. In: Physics and Mechanics of New Materials and Their Applications (PHENMA 2023). Southern Federal University, Rostov-on-Don, Taganrog, Russia, pp: 239–240.

Skylight Cyber, 2019. Cylance, I kill you! Skylight Cyber. URL: https://skylightcyber.com/2019/07/18/cylance-i-kill-you/

Soleymani S., Dabouei A., Kazemi H., Dawson J. and Nasrabadi N.M., 2018. Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification. 24th International Conference on Pattern Recognition (ICPR), pp: 3469–3476.

Soleymani S., Torfi A., Dawson J. and Nasrabadi N.M., 2018. Generalized bilinear deep convolutional neural networks for multimodal biometric identification. 25th IEEE International Conference on Image Processing, pp: 763–767.

Stefanidi A.F., Topnikov A.I. and Priorov A.L., 2020. Bimodal identification of personality based on facial and speech biometrics. Proceedings of the 17th International Conference, Penza, Russia, pp: 125–129.

Varga A.P. and Moore R.K., 1990. Hidden Markov model decomposition of speech and noise. International Conference on Acoustics, Speech, and Signal Processing. IEEE, pp: 845–848.