applied system innovation

MDPI

Article

Biometric-Based Key Generation and User Authentication Using Voice Password Images and Neural Fuzzy Extractor

Alexey Sulavko *, Irina Panfilova, Daniil Inivatov, Pavel Lozhnikov, Alexey Vulfin and Alexander Samotuga *

Department of Comprehensive Information Security, Omsk State Technical University, 644050 Omsk, Russia; panfilova_2015@bk.ru (I.P.); dinivatov@gmail.com (D.I.); vulfin.am@ugatu.su (A.V.)

* Correspondence: sulavich@mail.ru (A.S.); samotugasashok@mail.ru (A.S.); Tel.: +79-5339-490-54 (Alexey Sulavko)

Abstract: This work is devoted to the development of a biometric authentication system and the generation of a cryptographic key or a long password of 1024 bits based on a voice password, which ensures the protection of a biometric template from compromise. A new hybrid neural network model based on two types of trigonometric correlation neurons was proposed. The model is capable of recording correlation links between features and is resistant to data extraction attacks. The experiments were conducted on our own AIC-spkr-130 dataset and the publicly available RedDots, including recordings of user voices in different psycho-emotional states (sleepy state, alcohol intoxication). The results show that the proposed neural fuzzy extractor model provides an equal error probability level of EER = 2.1%.

Keywords: neural fuzzy extractor; correlation neurons; voice recognition; biometric template protection; autoencoders; automatic machine learning

check for updates

Academic Editor: Friedhelm Schwenker

Received: 8 December 2024
Revised: 3 January 2025
Accepted: 14 January 2025
Published: 17 January 2025

Citation: Sulavko, A.; Panfilova, I.; Inivatov, D.; Lozhnikov, P.; Vulfin, A.; Samotuga, A. Biometric-Based Key Generation and User Authentication Using Voice Password Images and Neural Fuzzy Extractor. *Appl. Syst. Innov.* **2025**, *8*, 13. https://doi.org/10.3390/asi8010013

Copyright: © 2025 by the authors. Published by MDPI on behalf of the International Institute of Knowledge Innovation and Invention. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Today, society is on the threshold of a "digital" revolution. Digitalization is happening everywhere and concerns almost all spheres of activity. Trends related to the globalization of remote access technologies (e.g., telemedicine and distance learning) are developing. More and more professions are becoming remote. In such an information environment, it is extremely important to prove the authenticity of the virtual image of a remote user (employee, partner, student, etc.).

Modern security systems increasingly use biometric parameters as a basis, but their effectiveness and convenience remain questionable: special equipment is required to recognize a fingerprint or iris, and the accuracy of facial identification can be affected by numerous parameters, from lighting, posture, and emotions, to changes in appearance by wearing accessories or a mask.

One option for solving the problem of reliable contactless authentication is based on the use of the speaker's voice parameters. However, speech parameters are variable depending on the psychophysiological state (PPS), which can negatively affect accuracy.

A voice image can be associated with a long cryptographic key or password directly used for authentication. Such a binding is conducted using special artificial intelligence (AI) models that allow us to not only perform authentication (recognition of a biometric image) but at the same time protect the cryptographic key or password and the biometric image during storage and transmission over communication channels.

Biometric authentication has become a critical component of ensuring secure access to systems and data. This approach requires resistance to attempts to extract knowledge from it. The field of biometric template protection (BTP) has evolved to encompass various approaches to hiding, encrypting, and transforming biometric images. The main ones are homomorphic encryption, which allows for encrypted data processing, revocable biometrics, and AI models for linking a key and a biometric image, such as a fuzzy extractor (vault, commitment, embedder) and a neural fuzzy extractor.

The aim of this study is to develop a neural network model for biometric-based key generation and user authentication based on voice images. In this paper, a new type of neuron called trigonometric correlation neurons is proposed. A correlation neuron is a neuron that can record deviations in the correlation dependence of input data [1]. Thus, a correlation neuron does not analyze the values of features, but instead, the correlation relationships between them. Moreover, the knowledge of such a neuron can be stored in open form since it is resistant to key extraction and biometric data extraction attacks.

The first correlation neuron model [1] showed high results only for the case when the input features had a strong mutual correlation dependence (positive or negative). In this paper, we propose a neuron model based on trigonometric functions, which allows for working with features that have both weak and strong dependence.

2. State of the Art

Modern developments aim to increase accuracy and resistance to external factors. Researchers test the proposed methods using voice archives with different recording lengths, noise levels, and the number and age of subjects. One of the key metrics in this area is considered to be the Equal Error Rate (EER). EER is an important indicator because it allows one to quantify the balance between the False Rejection Rate (FRR) and the False Acceptance Rate (FAR). Authentication systems are divided into text-dependent and text-independent ones, that does not allow for a direct comparison of their accuracy.

There are several known basic approaches to extracting features from voice images, in particular, i-vector and x-vector. The first is based on a Gaussian mixture (EER = 9.23% [2]). X-vector is based on the use of neural networks (NNs) and usually shows higher performance (EER = 5.21% [2]). Experiments comparing these methods were conducted on voice archives, such as NIST SRE 2016, VoxCeleb 2017 and SWBD.

Short-Time Fourier Transform (STFT) is also often used to analyze speech signal features. The paper [3] presents a model for speaker recognition based on short speech segments lasting less than one second. The experiments used their own dataset, including recordings of pronunciations of numbers in English collected from six students. A total of 3000 audio files were recorded in WAV format. The authors proposed an improved multi-frequency cepstral coefficient (MFCC) algorithm that uses both low-frequency and mid- and high-frequency components. These features were used in a two-dimensional convolutional neural network (2D-CNN), which built a three-dimensional tensor for processing acoustic data. The accuracy of the model on the test dataset was 99.6%, measured using the F1-score metric, which considered both the precision and recall of the model.

The authors of the ResNet architecture made a fundamental contribution to the development of deep learning and speaker recognition. This architecture uses "residual blocks" with skip connections between layers, which solves the gradient decay problem and simplifies the training of deep neural networks. This concept [4] was successfully tested on two models based on a 34-layer architecture with AM-Softmax and AAM-Softmax loss functions. The results showed a significant reduction in errors compared to most of the works presented at Interspeech 2020, reaching EER = 5.19% on the VoxSRC 2020 set.

The paper [5] presents an improved Res2Net architecture that can recognize more complex patterns by implementing a multi-scale approach in the "residual blocks" and dividing the feature map into smaller segments. The best model consisted of a 50-layer Res2Net network with the AM-SoftMax loss function and achieved an EER of 0.83% on the VoxCeleb dataset. One problem is the lack of phonetic information in speech segments when authenticating with a voice password or short speech fragments. In such cases, the authors of the study [6] recommended extracting vector representations using x-vector from deep NN layers, which resulted in a 14% decrease in EER to 13.35%. It was also determined that training the PLDA (Probabilistic Linear Discriminant Analysis) model, which is used at the speaker classification stage, on short speech segments contributes to an additional reduction in the equiprobable error by 5% (x-vector extraction occurring on the seventh layer with a dimension of 150).

The RecXi model [7] was proposed in the study [7]. It includes three Gaussian inference layers, each of which extracts individual speech components. The test resulted in a 9.56% reduction in the EER value compared to the base SOTA model. The experiment was conducted using VoxCeleb and SITW.

One of the most recently proposed approaches to improve the information content of speech data is the multimodal Vision-Guided Speaker Recognition (VGSR) training [8], in which the face recognition system acts as a teacher, passing on knowledge useful for distinguishing faces to speaker recognition models. During training, the voice identification model learns to more effectively extract unique features of the speaker based on the data that the face can provide about the personality. The authors suggested a distillation method that includes a hyperparameter to prevent an over-dependence of the voice model on the face model, addressing overfitting. In testing on the CN-Celeb and VoxCeleb1 datasets, the system showed a 10–15% reduction in EER compared with traditional methods using only voice data.

Authentication systems working with text-dependent data are actively studied using various databases. One of these is RedDots. In [9], attention was paid to speaker verification using short speech segments. The experiment was conducted on a subset of 100 speakers who pronounced numerical sequences. The authors focused on the first eight phrases common to all subjects. A four-level decomposition using a wavelet transform was applied to these segments, and the resulting features were then transferred to the neural network for classification. Using the Hilbert transform to improve the phase characteristics of the sound made it possible to achieve EER = 4.85%.

In [10], the authors focus on using Gaussian Mixture Models (GMMs) with Maximum A Posteriori (MAP) adaptation and Nuisance Attribute Projection (NAP) to compensate for channel effects. The experiments were conducted on 10 standard phrases from the RedDots archive. The focus was on the tests for the "Impostor Correct" case, in which the non-target speaker pronounces the correct phrase. This test allows simulating situations when the attacker knows the correct text, which complicates the authentication task. Using bootstrap aggregation allowed the EER to further reduce the EER from 3.3% to 2.6%.

The paper [11] proposes a method for text-dependent speaker verification based on hidden Markov models (HMMs) and i-vector. The experiments were conducted using the RedDots database. For the experiment, digit-specific HMM was used, which improved the quality of speech segmentation and reduced errors compared to other methods, such as x-vector. The authors also applied uncertainty normalization and LDA regularization. This approach allowed them to obtain the results EER = 1.52% for men and 1.77% for women (Table 1).

Table 1. Comparison of modern biometric authentication methods.

Methods	Modality, Dataset	Obtained Results
i-vector, x-vector [2]	Voice, NIST SRE 2016, VoxCeleb, SWBD	EER = 9.23%, EER = 5.21%
2D-CNN (MFCC) [3]	Voice, authors' own archive	F1-score = 99.6%
ResNet (AM-Softmax, AAM-Softmax) [4]	Voice, VoxSRC 2020	EER = 5.19%
Res2Net [5]	Voice, VoxCeleb	EER = 0.83%
PLDA + x-vector [6]	Voice	EER = 13.35%
RecXi [7]	Voice, VoxCeleb, SITW	EER reduction by 9.56%
VGSR [8]	Voice + Face, CN-Celeb, VoxCeleb1	Reduce EER by 10–15%
Wavelet-based classification [9]	Voice, RedDots	EER = 4.85%
GMM (MAP + NAP) [10]	Voice, RedDots	EER = 2.6%
HMM + i-vector [11]	Voice, RedDots	EER = 1.52% (men), 1.77% (women)
Homomorphic encryption [12]	Face	Matching time 2.38 ms, memory 32.8 KB
Lattice-based fuzzy extractors [13]	Face	Entropy equal to 45 bits, high security
Fuzzy vault + DTW [14]	Voice	Impossibility of template inversion
RBOMP with i-vector and prime factorization [15]	Voice	EER = 3.43%, ARM attack protection
Hybrid method: homomorphic encryption + cancelable biometrics [16]	Face	Authentication time is about 7 s
Hybrid scheme: cancelable biometrics + biocryptography [17]	Face, CVL, FEI, FERET.	Accuracy: CVL = 99.47%, FEI = 98.10%, FERET = 100%
Neural fuzzy extractor with correlation neurons [18]	Voice, RedDots	EER = 2.64%
Fusion of four systems and tandem features [19]	Voice, RedDots	EER = 1.96–2.28% for male, 2.7–3.48% for female
TD-SV based on GMM-UBM [20]	Voice, RedDots	EER = 3.06% with 5 db noise, 2.7% with 10 db noise

There are several areas of research aimed at ensuring the confidentiality of biometric data during processing by a biometric system and the knowledge of machine learning models: the use of homomorphic encryption methods, fuzzy extractors (vault, commitment, embedder, etc.), and neural fuzzy extractors, as well as cancelable biometrics.

Homomorphic encryption allows us to perform calculations on encrypted data without revealing their contents but requires significant computing resources [21]. Fuzzy extractors allow you to restore encrypted keys even with small changes in biometric data, providing flexibility, but have limitations in resistance to more significant changes in patterns [22]. A neural fuzzy extractor, based on neural networks, provides high accuracy in converting biometrics into code, but its reliability directly depends on the quality of training and the amount of data [23].

To improve the performance of homomorphic encryption, an approach called asynchronous encryption is proposed [12]. Its use allowed for performing operations on encrypted data in parallel with the processing of other data, which led to a reduction in processing time. Also, by optimizing the work with preprocessed data, it was possible to achieve an increase in the accuracy of personality recognition. This approach made it possible to achieve a template matching time of 2.38 ms with a memory size for storing one face image of 32.8 KB.

Lattice-based fuzzy extractors are also proposed to be used to protect templates [13]. The goal of the study was to create a reliable system capable of restoring a cryptographic key in the presence of certain errors in biometric data. In the work, dense lattices such as E8 and Leech were used to create stable templates that could maintain accuracy even when changed. Experiments have shown that their method reduces the likelihood of data leaks and provides a high level of security with an entropy of 45 bits, applicable to modern face recognition systems, such as ArcFace.

In [14], the authors propose a BTP approach based on generating cryptographic keys from a voice template using a fuzzy vault scheme. The main goal is to ensure that the template cannot be easily inverted and compromised. The system uses dynamic time warping (DTW) to align voice characteristics. The fuzzy vault stores the template by adding chaff points, which makes it computationally difficult for attackers to distinguish between genuine and fake points.

In [15], the authors propose a new approach to voice BTP using Random Binary Orthogonal Matrices Projection (RBOMP). The system works with i-vectors that are transformed using orthogonal matrices to create hashes that prevent template inversion. As an additional protection, the prime factorization method was introduced, which provides higher privacy and protection against multiple-record attack (ARM). The test results showed that the proposed system maintains a high level of protection (EER = 3.43%) without significant performance losses.

A hybrid method based on a combination of homomorphic encryption and cancelable biometrics can be used to protect biometric templates [16]. Cancelable biometric methods used included BioHashing, MLP hashing, and IoM hashing. Testing was conducted on MOBIO and LFW samples using modern ArcFace, ElasticFace, and FaceNet face recognition models. The results show that the proposed method provides comparable performance to homomorphic encryption, while reducing computational costs by reducing the template size. This made it possible to achieve a duration of the authentication procedure equal to seven seconds.

The authors of [17] developed a hybrid biometric template protection scheme by integrating cancelable biometrics and biocryptography. Face recognition was performed using image preprocessing methods using the Tree Structure Part Model, feature extraction using the Ensemble Patch Statistics technique, and classification using a multi-class linear SVM. The testing was conducted on three databases: CVL, FEI and FERET. The system achieved an accuracy of 99.47% for CVL, 98.10% for FEI and 100% for FERET, which confirms its effectiveness compared to existing methods.

In one of the latest works, whereby the authors used a neural fuzzy extractor, a new approach for biometric authentication was implemented based on the use of correlation neurons [18]. Voice images of speakers from the RedDots database were used as biometric data. The main advantage of the proposed model was the use of feature correlation to improve authentication accuracy. The system achieved an error rate of EER = 2.64% with 4096 neurons and six synapses. Table 1 provides information about the main methods described in the current section.

In publicly available scientific papers, there is no widespread use of neural fuzzy extractor for protecting voice images, but it has been sufficiently tested for such types of images as faces [24] and fingerprints [25]. However, its application to voice images is justified by the universality of working with feature vectors extracted from a “raw” biometric image, which makes the method independent of the method of presenting biometric data. Only the feature extraction algorithm is modality-dependent.

3. Materials and Methods

3.1. Voice Image Databases

Among the most common corpora are NIST SRE, TIMIT, VoxCeleb and VoxCeleb2.

NIST SRE is aimed at creating speaker identification systems based on telephone conversations. More than 6800 subjects participated in its formation. Recordings lasting from 10 to 60 s were made for each subject [26]. TIMIT includes 6300 audio recordings of phonemes pronounced by 630 speakers in English. Unfortunately, this number of pass phrase examples for each individual subject is insufficient for our study [27].

VoxCeleb and VoxCeleb2 are focused on text-independent speaker recognition, which makes them unsuitable for our experiment, since our task requires a text-dependent procedure. These datasets contain recordings of interviews and public speeches, where speakers pronounce arbitrary, rather than fixed, pass phrases.

RedDots includes recordings of 100 speakers who spoke 22 similar phrases/sentences during a weekly session, resulting in a data corpus of 124,800 audio files. This is the only suitable corpus, but the authors did not take into account the speaker's state when forming the database.

We created our own dataset, AIC-spkr-130, focused on voice password authentication with biometric template protection. Voice passwords were short phrases that consist of one, two, or three (but short) words (e.g., "access control" and "allow access"). Each subject chose a password from a pre-prepared dictionary. Subjects' passwords are not unique (i.e., one key phrase could be used by several subjects). The proposed dataset includes recordings of 130 speakers aged 18 to 50 years in WAV format with a sampling frequency of 8 kHz and a sample size of 16 bits (mono). Later, we expanded the dataset by including recordings of the same speakers and passwords in an altered psychophysiological state such as (sleepy or alcoholic intoxication states).

According to the guidelines of the Russian Ministry of Health, alcohol intoxication was modeled at three stages: the first is a blood alcohol content of 0.2–0.3‰, which does not have a significant effect on the body; the second is 0.3–0.5‰, where mild euphoria and decreased coordination begin; the third is 0.5–1‰, which corresponds to mild intoxication with impaired perception and reasoning ability. The volume of alcohol that the subject had to consume was calculated using the Widmark Formula (1):

$$c = A/m \times r, \quad (1)$$

where c is the concentration of alcohol in the blood in ‰, A is the mass of the drink consumed in grams, m is the body mass in kilograms, and r is the Widmark coefficient.

The sleepy state was achieved by taking sedatives of plant origin (motherwort, mint, or valerian), which reduced the heart rate by 3–5% compared to the norm. In each PPS, the subjects pronounced 8 passwords at least 60 times for each phrase, which provided a sufficient amount of data for analysis.

3.2. Extracting Features from a Voice Image

3.2.1. Preprocessing of Voice Recordings

Fourier spectrograms can be fed to the input of deep convolutional neural networks (CNNs) for biometric feature extraction [28]. However, large input data sizes require large sample sizes for training a multilayer NN, so instead of spectrograms, we used the amplitude spectrum averaged over all windows (Figure 1), which integrates information about local characteristics of the voice recording and smooths out random outliers. The spectrum obtained because of the transformation depends on both the speech characteristics of the speaker and the message itself, so if the user pronounces the same phrase, their spectra will be similar. In case of replacing the voice phrase or the speaker, significant changes will

be noticeable in the graph. In this study, different representations of speech images are used based on different STFT window functions: rectangular, Gaussian, Blackman, Bartlett, and Hamming.

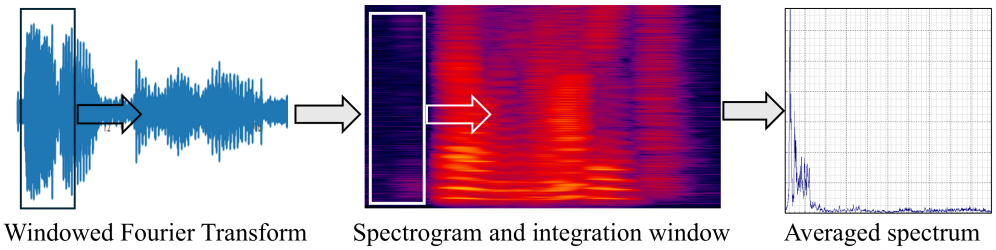

Figure 1. Construction of the averaged spectrum.

Different representations can be used both to train different classifiers that can be combined into a single ensemble and to form a single feature vector. We can obtain many highly correlated feature pairs and process them with a single neural fuzzy extractor capable of extracting additional information from correlations [1].

3.2.2. CNN Architecture for Voice Phrase Feature Extraction

This paper proposes to combine multilayer CNNs and neural fuzzy extractors using the stacking principle. CNNs are used to generate features, which are then fed to the input of the neural fuzzy extractor. This approach improves the quality of classification by combining the strengths of both architectures: deep neural networks for image analysis and feature extraction, and neural fuzzy extractors for rapid additional training of the system on new user data and image verification while ensuring the confidentiality of biometric templates.

To extract features, we used the autoencoder architecture, which is a deep neural network that allows for extracting informative features. It includes an encoder that compresses input data into a more compact form, which allows us to reduce the dimensionality of the feature space, and a decoder that is trained to restore them. During training, the average spectrum of the voice password is fed to the input and output of the autoencoder, which allows the network to learn to highlight key features of the voice data.

The autoencoder is based on the VGG19 architecture [29], which was modified to handle averaged spectra. It includes one-dimensional convolutions, layers with batch normalization, and a fully connected layer with a linear activation function (Figures 2 and 3).

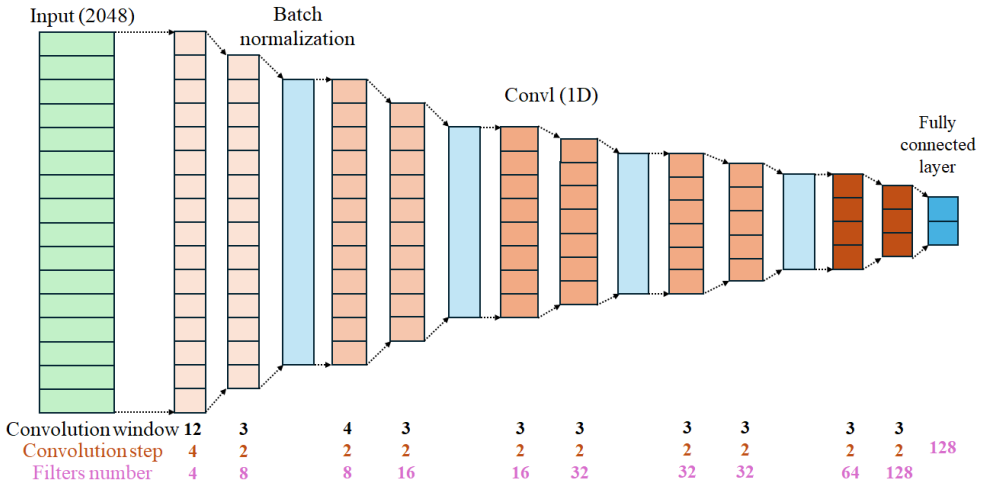

Figure 2. Encoder architecture for feature extraction.

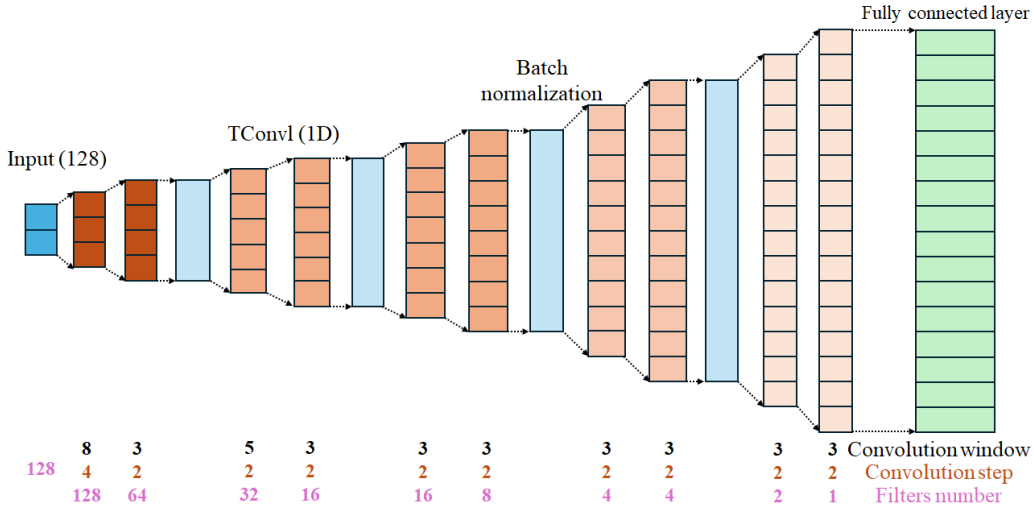

Figure 3. Decoder architecture.

The training sample was the VoxCeleb2 voice dataset (200,000 short utterances that have a similar duration to voice passwords—from 1.5 to 5 s), the optimization algorithm was Adam, and the error function was binary cross-entropy.

3.2.3. Ensembling Pre-Trained CNNs

To create the ensemble, five similar autoencoder architectures were formed and trained on 500 epochs. Each of them was trained under identical conditions: the same sample, optimizer, and error function. However, each CNN was fed different image representations as input. Representations were the averaged spectra calculated using different window functions: rectangular, Gaussian, Blackman, Bartlett, and Hamming.

Different window functions produce similar spectra, which, after processing by the encoder, produce correlated features. This solution ensures the extraction of both strongly and weakly correlated feature vectors, which are necessary for the operation of neurons with trigonometric measures of proximity. The cosine measure is effective for weakly correlated data, while the cotangent measure improves the processing of highly correlated features, which increases the accuracy of calculations.

During the user authentication process, this speech signal is converted into 5 averaged spectra, each of which is fed to the input of the corresponding encoder, at the output of which the feature vector is calculated. The vector data are combined and fed to the input of the neural fuzzy extractor, which at the output produces the personal key or speaker (Figure 4).

3.3. Fuzzy Neuro-Extractor

3.3.1. Neuron Model

The study [1] demonstrated that using meta-features generated by correlations between pairs of features allows the neural fuzzy extractor to reliably separate images of legitimate ("Genuine") and illegitimate ("Impostor") users. Formally, the space of meta-features can be obtained by mapping (2).

$$a'_l = f(a_i, a_j) \tag{2}$$

where i and j are the feature numbers of the original feature vector $(i \neq j)$, a'_l is a meta-feature, i.e., a feature obtained by synthesizing two (in this case) or more original features using a functional transformation f, and l is the meta-feature number.

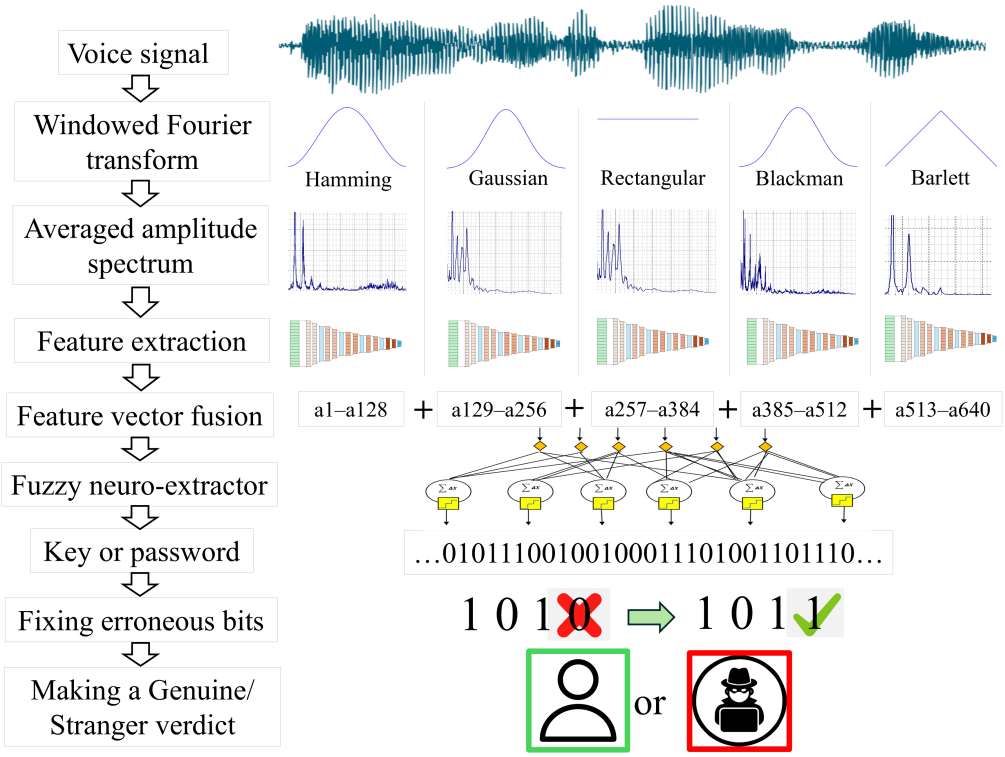

Figure 4. Authentication process based on a CNN committee and NNBCC.

In this paper, we proposed to use an analogue of the cosine distance (cosine similarity) [30] as a functional transformation f for pairs of features with weak correlation dependence. The mentioned above functional transformation operates at the level of individual pairs of features of the original vector (3):

$$a'_l = \cos(\widehat{\bar{d}, \bar{v}}). \tag{3}$$

The cosine is determined between two vectors (Figure 5): vector $\bar{d}$ connects the "center of mass" (the intersection of the average values of two features) with a point representing the desired image, and vector $\bar{v}$ has a length similar to vector $\bar{d}$ but is located at the beginning of the trigonometric circle, which has the same center. Using the "center of mass" as the reference point of the specified vectors allows us to avoid using the "Genuine" image for the same purpose. Here, the coordinates of the position of the legitimate image in the feature subspace remain unknown, and the image itself is confidential.

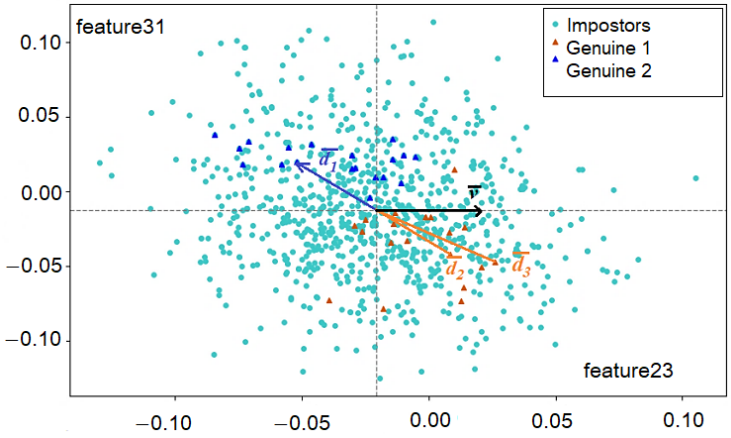

Figure 5. Calculation and visualization of the distance from the "center of mass" to the images

In general, metric (3) allows us to avoid using the Euclidean distance, which has a significant drawback: the Euclidean metric does not take into account the direction in which the "Genuine" image is located relative to the "center of mass". Figure 5 shows that close distances are often demonstrated by images localized in different parts of the subspace ($\bar{d}_1$ and $\bar{d}_2$), while images localized in one area, on the contrary, can be characterized by completely different distances (d_2 and d_3).

However, in subspaces where the features are highly correlated, the vectors $\bar{d}$ and $\bar{v}$ may be nearly collinear. The cosine measures only the angle between the vectors but does not take into account their length. In addition, if the vectors lie almost on the same line, the cosine may not vary much, even if the angle between them is large enough. In this regard, an additional metric is needed that takes into account the correlation links in the subspaces of pairs of features, but works within the same trigonometric circle. The following transformation can serve as such a metric (4):

$$a'_l = \cot(\widehat{\bar{d}, \bar{v}}), \tag{4}$$

The cotangent of the angle between the $\bar{d}$ and $\bar{v}$ vectors provides a more detailed representation of the angular relationship between the vectors, especially if they are located close to each other on the correlation line (arrangement of images in subspaces at different functionals presented on Figure 6). As can be seen, the proper region of the class of images in the space of correlated features is "stretched out", and the metric (4) allows for a better separation of images of correlated classes by sectors.

The combined use of cosine and cotangent of an angle allows for a comprehensive approach to the analysis of feature vectors, especially in highly correlated data.

Then, the simplest trigonometric correlation neuron will be built based on one of the two metrics, summing up the input meta-features obtained using the described transformations. The neuron makes a decision on the sum of the input values of the meta-features according to the two-level threshold activation function $\varphi(y)$ (5):

$$\varphi(y) = \begin{cases} 1, y \geqslant T_2 \\ 0, T_1 < y < T_2 \\ -1, y \leqslant T_1 \end{cases} \tag{5}$$

where T_1 and T_2 are the decision thresholds for one of the three function values. The activation function values must be converted into binary states of the type "01", "10", "11" in order to associate the external binary key with the biometric image. The conversion must be conducted using a set of tables for converting the activation function states into binary codes. There are 24 variants of such tables in total. To associate one state with the desired binary code, it is necessary to randomly select one of the 6 conversions. Then, each neuron will produce ≈ 2 bits of information.

The described transformations allow the proposed modification of the neural fuzzy extractor to produce long cryptographic keys, significantly exceeding those obtained using the neural fuzzy extractor trained in accordance with the Russian Federation standard GOST R 52633.5 [31]. The standardized neural fuzzy extractor is severely limited in terms of key length since it does not allow for the duplication of features during assembly and training because of its susceptibility to the Marshalko attack [32]. Also, the "standard" neural fuzzy extractor is susceptible to another attack associated with binary outputs of neurons in one bit of information [33]. Using the activation function (5) allows us to avoid this vulnerability, since when more than one bit of information appears at the neuron output, the attack described in [33] becomes nonlinear and extremely difficult to implement.

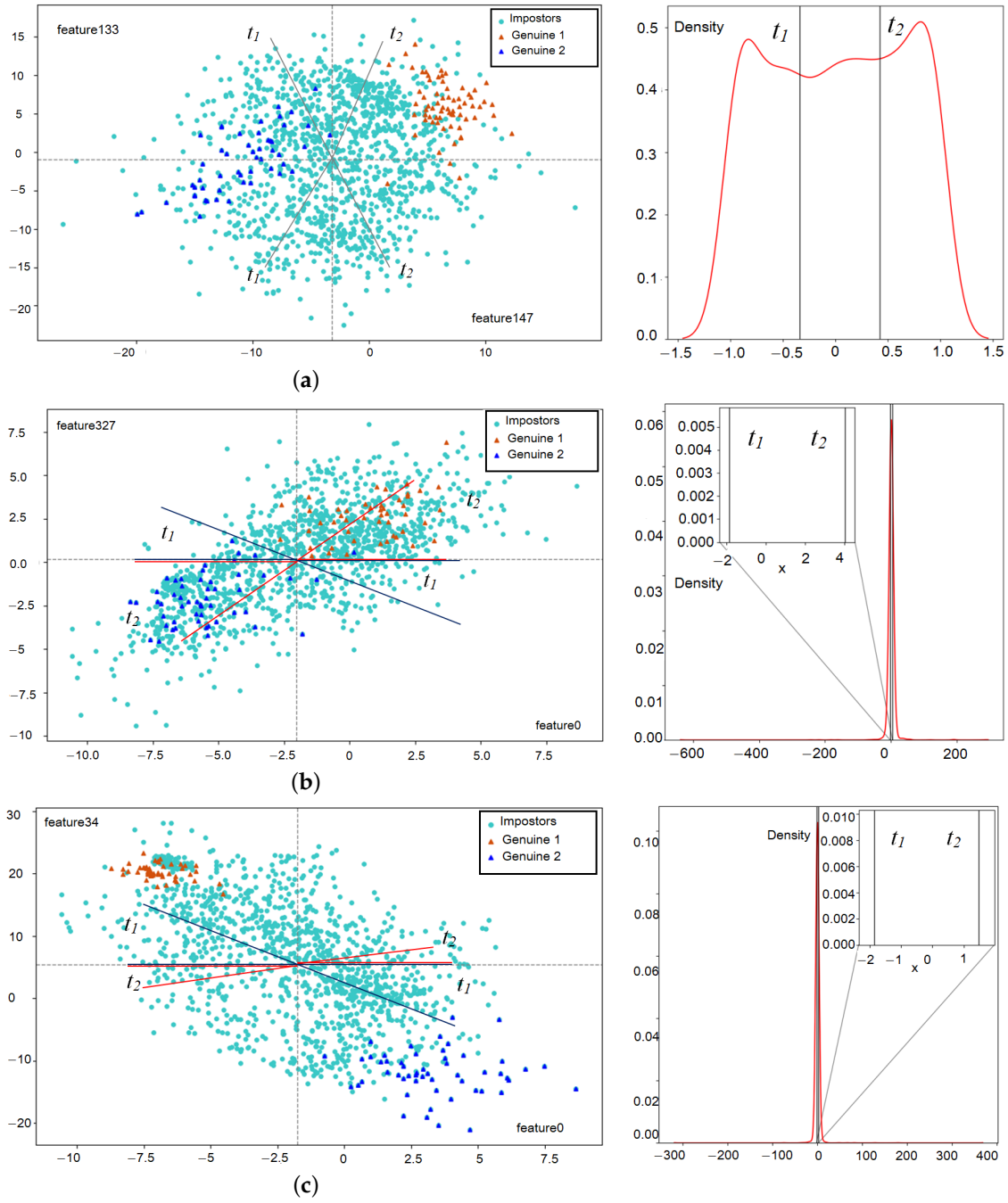

Figure 6. Arrangement of images in subspaces at different functionals: (**a**) Thresholds t_1 and t_2 in the subspace of a pair of features (symmetrical regarding the y-axis) and on the probability density plot of meta-features obtained using the metric (2). (**b**) Thresholds t_1 (blue lines) and t_2 (red lines) in the positively correlated subspace of a pair of features (symmetric about the y and x axes) and on the probability density plot of meta-features obtained using the metric (4). (**c**) Thresholds t_1 (blue lines) and t_2 (red lines) in the negatively correlated subspace of a pair of features (symmetric about the y and x axes) and on the probability density plot of meta-features obtained using the metric (4).

The trigonometric correlation neuron receives specially selected pairs of features, for each of which, it is necessary to determine its own thresholds t_1 and t_2. The selection of pairs within one neuron is conducted according to two criteria:

1. Each pair generates one type of meta-features: cosine or cotangent. The input of one neuron receives either pairs in the subspace of which the correlation dependence of features is expressed (cotangent type of meta-features) or such pairs in whose

subspaces the correlation dependence is absent or is weak according to the Chaddock scale (cosine type of meta-features).

2. Relative to its thresholds t_1 and t_2 in the subspace, each pair of all "Genuine" images entered one of the three sectors $([-\infty, t_1], (t_1, t_2), [t_2, \infty])$. The threshold values T_1 and T_2 for the activation function of neurons are saved as the average values of the ancient thresholds t_1 and t_2 for the subspaces of pairs of features obtained using the calibration algorithm that described in Section 3.3.2. The following formulas are used to calculate threshold values: $T_1 = \frac{1}{k}\sum_{z=1}^{k} t_{1z}$ and $T_2 = \frac{1}{k}\sum_{z=1}^{k} t_{2z}$, where k is the number of synapses (inputs) of a neuron; z is the number of a synapse (input) of a neuron; and t_{1z} and t_{2z} are the thresholds for a pair of features entering the input of a neuron, obtained sequentially because of the algorithm's operation.

The private thresholds of each pair of features are not secret; they are preliminarily considered in the process of the biometric system. Such a sample does not contain data of the registered subjects and must be representative in order to objectively reflect the potential conditions of the current system.

3.3.2. Calibration of the Biometric System

It is proposed to calculate the thresholds according to the following algorithm (Figure 7):

1. The coef that is the correlation coefficient of a pair of features is calculated;
2. The type of meta-feature is determined in accordance with the correlation: positive correlation $(coef \geqslant 0.5)$ − metric (4) (+), negative correlation $(coef \leqslant -0.5)$ − metric (4) (−), weak correlation $(-0.5 < coef < 0.5)$ − metric (3).
3. An empirical probability density function $(f(.))$ of meta-features is constructed, calculated according to the corresponding formula;
4. The $f(.)$ function is integrated in order to find the distribution function $F(.)$;
5. The interval $[0, 1]$ is divided into m equal sectors $([0, \frac{1}{m}, \frac{2}{m}, \ldots, 1])$;

For each value of $z \in [0, m-1]$ from the obtained set, the corresponding value of the distribution function t_z for which $F(t_z) = z$ is calculated. The obtained values t_z are the desired thresholds.

For each pair of features, the thresholds calculated using the specified algorithm are saved and then used to calculate the thresholds of the neuron activation function.

3.3.3. Training Neural Fuzzy Extractor

The procedure of synthesis and training of a neural fuzzy extractor is carried out automatically without using the backpropagation method. The structure of the neural fuzzy extractor should be built on the basis of $N = \frac{S}{q}$, where N is a number of neurons, S is the desired length of the cryptographic key, and q is the number of bits produced by one neuron and calculated by Formula (6):

$$q = [\log_2 z], \tag{6}$$

where z is a number of thresholds dividing the probability density function of meta-features into equal sectors $(z = m - 1)$, and [.] is a rounding operation to a larger value. The neural fuzzy extractor is built separately for each subject, i.e., for each class of the training set of biometric data.

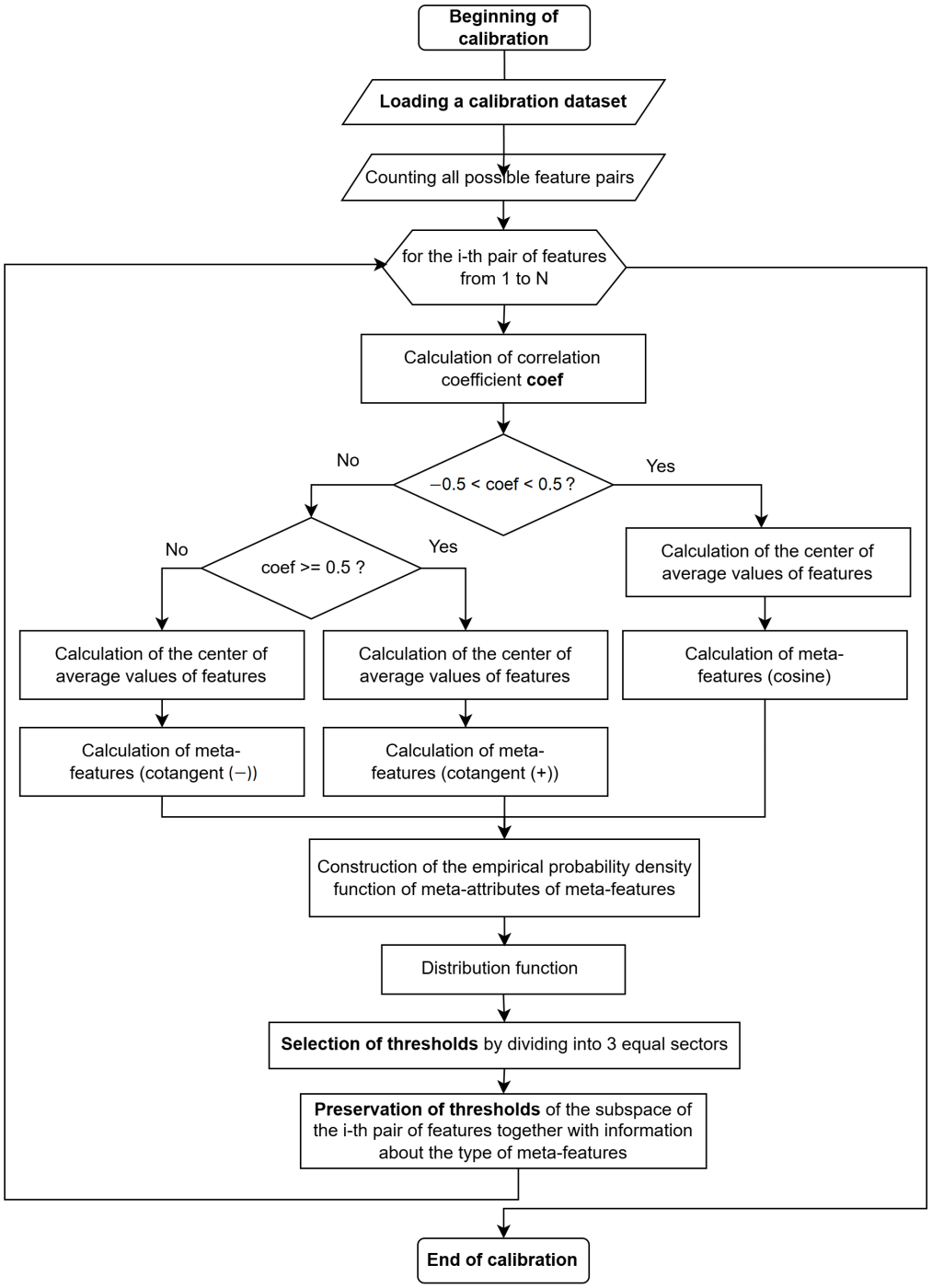

Figure 7. Algorithm for calibration of fuzzy neural extractor parameters.

The first stage of the neural fuzzy extractor training algorithm is to divide the subspace of feature pairs of the training set "Genuine" into three equal sectors in accordance with the threshold values obtained at the calibration stage. Here, it is necessary to identify such feature pairs for which all "Genuine" images are located in one of the three sectors. The result of the first stage will be 3 groups of feature pairs, from which three types of neurons are "assembled" (Figure 8).

It is important to consider that each of the three groups can contain neurons summing cosine meta-features, as well as cotangent ones (additionally differing in the "direction" of the correlation: positive—"+" and negative—"−"). It is advisable to have an equal ratio of "cosine" and "cotangent" neurons within one group.

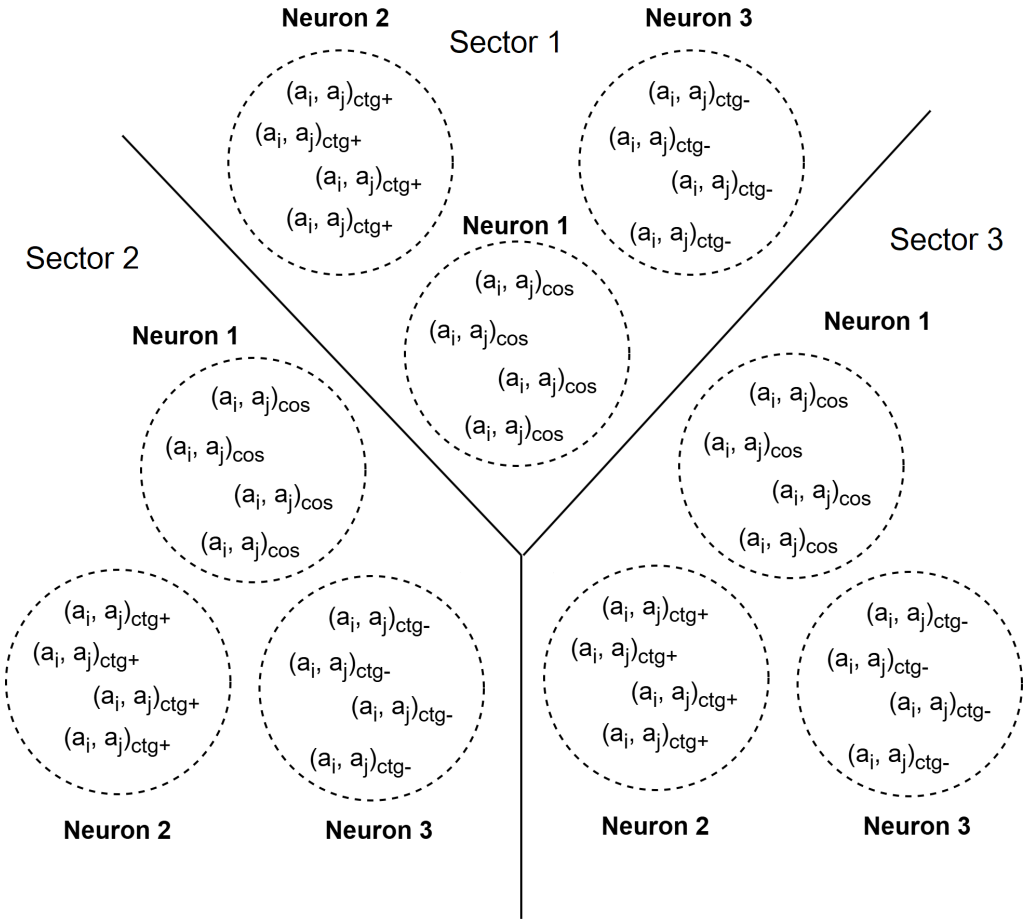

Figure 8. Schematic of neuron formation from three sectors and three types of neurons.

For each neuron, pairs of features from a certain (one) group are randomly selected. The number of pairs of features is equal to the number of synapses of the neuron. The inputs of each neuron must be unique, and no pair can be reused in another neuron. This requirement must be met to eliminate the possibility of implementing an attack of this type [32].

The result of assembling the neural fuzzy extractor is a structure in which three types of neurons, collected from three groups of pairs of features, are present in equal proportions (based on the length of the key).

4. Experimental Results

4.1. Metrics

This study uses the Equal Error Rate (EER) metric to evaluate the performance of the neural fuzzy extractor within the proposed biometric system. To calculate the EER, a graph of the dependence of FAR on FRR is plotted at different decision thresholds, and the point at which these two curves intersect is determined. The EER value reflects the threshold level at which FRR and FAR are equal. The EER was used in this study because of its ability to reflect the optimal state of the system between security and convenience.

4.2. Experimental Setup

This study proposes an approach to protect biometric data based on the use of a neural fuzzy extractor built using two types of trigonometric neurons: cosine and cotangent. The main goal of the model is to ensure reliable and secure transformation of voice data into cryptographic keys or passwords, minimizing the risk of compromising biometric templates.

The experimental part includes the preliminary extraction of 128-dimensional feature vectors from voice images using autoencoders based on convolutional neural networks (CNNs) described in Section 4. These features are fed to the input of neurons based on trigonometric proximity measures. As shown in Section 5, the cosine measure should give better results when processing data with low correlation, while the cotangent measure, on the contrary, should demonstrate advantages when working with highly correlated features. For optimal data processing, cotangent neurons were additionally adapted to separately process positively and negatively correlated inputs. To evaluate the efficiency of the neural fuzzy extractor with the two types of trigonometric neurons, cosine and cotangent, experiments were conducted on two datasets: AIC-spkr-130 and RedDots.

The AIC-spkr-130 dataset includes records that are divided into two groups: "Impostors" (class #60), used for calibration, and legitimate users, randomly selected in the amount of 30 subjects for training the neural fuzzy extractor. Here, the following voice images of the AIC-spkr-130 were used for training, which were obtained under the conditions of the "normal" state of the subjects. Images obtained under the altered psychophysiological state of the subjects were used as test images.

For each converter, the FRR and FAR indicators are estimated, where FRR is determined based on "Genuine" images that are not involved in training, and FAR is estimated on all other classes, except for the one for which the neural fuzzy extractor is trained. A similar procedure was performed for the RedDots dataset. Here, the first class of data is used for calibration ("Impostors"), and 30 random images are used to train the neural fuzzy extractor.

To better understand the effectiveness of the neural fuzzy extractor with different types of trigonometric neurons, comparative experiments are conducted using neurons of exclusively cosine and exclusively cotangent types. This will allow us to more accurately determine under what conditions each metric shows the best results and what advantages their combined use provides.

4.3. Results

To build the neural fuzzy extractor, a key length of 1024 bits was chosen, providing a sufficient level of cryptographic security and preventing the likelihood of successful cryptanalysis using modern computing power.

The only configurable parameter of the neural fuzzy extractor is the number of synapses of one neuron η. This parameter directly affects the level of the first type of error, the False Rejection Rate (FRR). Increasing the number of synapses allows the neuron to more accurately represent and distinguish the features of training images, which helps to reduce the FRR. Thus, the more the synapses, the better the system copes with identifying legitimate users. Which leads to a decrease in the number of false rejections. The number of neurons similarly affects the FAR. Although FAR is also important, in this context, its reduction is not a priority, since the focus is on minimizing FRR.

As part of the experimental evaluation, different values of η were tested in order to find the optimal balance between FAR and FRR and the best EER value (Table 2).

Table 1 shows that the minimum error for the cosine-based neuron network on both datasets is achieved at $\eta = 4$ and is 0.055 for AIC-spkr-130 and 0.15 for RedDots, respectively, which indicates limited effectiveness of measure (3) with increasing model complexity. Using cotangent as the basis of neurons leads to some decrease in authentication errors, especially with a small number of synapses. The lowest error values are recorded at $\eta = 4$: 0.044 for AIC-spkr-130 and 0.061 for RedDots. The combination of neurons based on two measures (cosine and cotangent) showed the best results, demonstrating the minimum error EER ≈ 0.021 for AIC-spkr-130 and EER ≈ 0.032 for RedDots. Obvi-

ously, the combination of these two functionals is optimal in terms of error minimization. This is because of the correct selection of metrics underlying neurons that depend on the correlation of feature pairs. Thus, the hypothesis that the trigonometric functions of cosine and cotangent are better suited for low and highly correlated feature pairs is confirmed.

Table 2. EER values at different numbers of synapses of the fuzzy extractor neuron.

Measure	Number of Synapses, η	Dataset	
		AIC-spkr-130	**RedDots**
Cosine	4	0.055	0.15
	6	0.061	0.147
	8	0.081	0.155
Cotangent	4	0.044	0.061
	6	0.049	0.117
	8	0.047	0.132
Cosine + cotangent	4	0.021	0.032
	6	0.061	0.069
	8	0.065	0.071

It is worth mentioning that increasing the number of synapses in the cases of both datasets does not lead to a significant decrease in EER, but causes its growth. This may indicate that increasing the complexity of the model does not always contribute to improving its characteristics. Thus, increasing the number of neurons and, accordingly, the key length is more advantageous than increasing the number of inputs to the trigonometric neuron.

The best experimental results presented in the table are shown in Figure 9, which characterize the dependence of FRR and FAR on the Hamming distance (the number of positions in which the i-th generated code differs from the code associated with the biometric image). The maximum value of the Hamming distance is equal to the key length. The intersection point of the obtained graphs will describe the EER value for a given ratio of neural fuzzy extractor configurations.

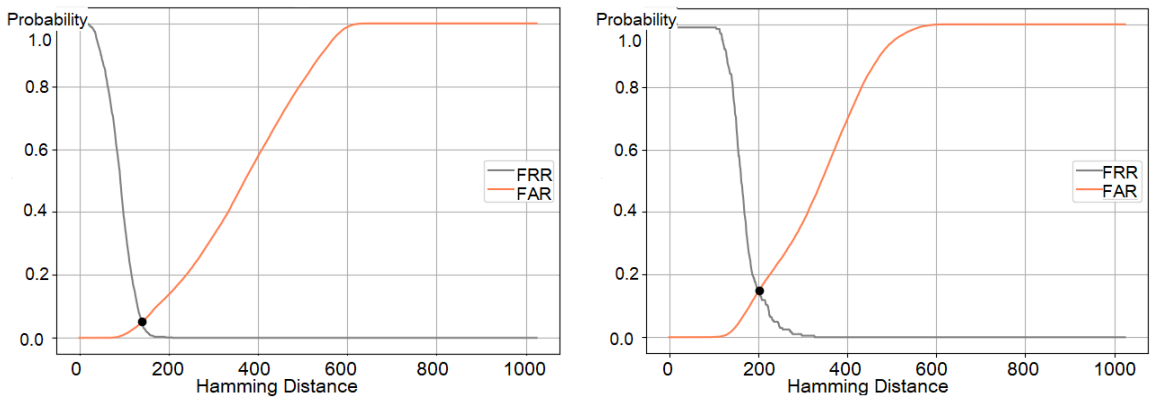

Figure 9. *Cont.*

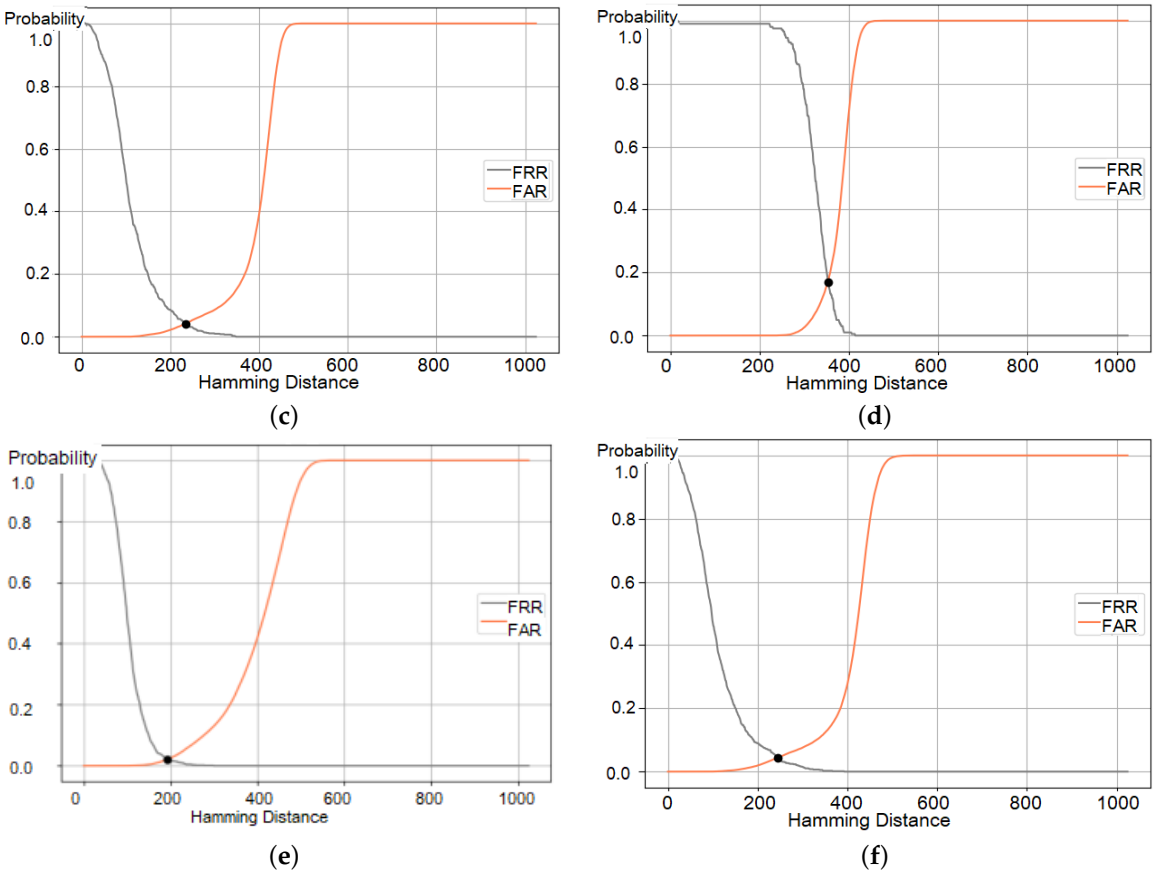

Figure 9. Best EER values according to the results of the experiment: (**a**) AIC-spkr-130 (EER ≈ 0.055), (**b**) RedDots (EER ≈ 0.15), (**c**) AIC-spkr-130 (EER ≈ 0.044), (**d**) RedDots (EER ≈ 0.061), (**e**) AIC-spkr-130 (EER ≈ 0.021), (**f**) RedDots (EER ≈ 0.032).

5. Discussion

As can be seen from Table 3, the obtained results in terms of solution accuracy correspond to the world level, but they surpass it in terms of security, since the proposed solution allows for protecting biometric templates from compromise.

Table 3. Results achieved in the verification task using RedDots as an example.

Methods	Used Part of RedDots	EER, %
Wavelet transform, NN [9]	The first 8 standard phrases	4.85
GMM, MAP, NAP, Bagging [10]	10 standard phrases (ImpostorCorrect)	2.6
HMM, i-vector, digit-specific models [11]	33rd and 34th phrases, (ImpostorCorrect)	1.52–1.77
Fusion of four systems and tandem features [19]	Part 1 (ImpostorCorrect)	2.28 for male, 3.48 for female
Fusion of four systems and tandem features [19]	Part 4 (ImpostorCorrect)	1.96 for male, 3.22 for female
TD-SV based on GMM-UBM [20]	Part 1 (TargetWrong, ImpostorCorrect, ImpostorWrong)	3.06 with 5 db noise, 2.7 with 10 db noise

Testing of the neural fuzzy extractor trained in accordance with the GOST R 52633.5 standard [34] was also conducted (Table 4).

Table 4. Comparison of the proposed solution with the neural fuzzy extractor trained according to GOST R 52633.5 [34] on AIC-spkr-130 and RedDots dataset (EER, %).

Methods	Key Length	EER, %	RedDots
Neural fuzzy extractor trained in accordance with GOST R 52633.5 [34]	160 bit	2.7	3.5
Proposed solution	1024 bit	2.1	3.2

As can be seen, the proposed neural fuzzy extractor model based on trigonometric correlation neurons outperforms the basic one in both accuracy and key length. However, to confirm its stability under real operating conditions, it was necessary to evaluate its operation under noisy conditions.

For this purpose, experiments were conducted to evaluate the effect of noise on the efficiency of the proposed solution. For this purpose, office noises (conversations, typing, etc.) were added to the AIC-spkr-130 corpus records with coefficients of 0.1 and 0.25 (Figure 10). The specified coefficients mean the ratio of the average volume level of the superimposed noises to the average volume level of the original file. The noise superimposition process consisted of adjusting the level of the added sounds in accordance with a specified coefficient, which determines the proportion of the noise level relative to the level of the original voice. Enrichment of the original voice images was achieved by scaling the amplitude of the noise signal and subsequent summation of audio signals in the time domain. The resulting corpora were used as a test sample, which led to an increase in the relative EER by 9% and 17% to 2.29% and 2.46%, respectively.

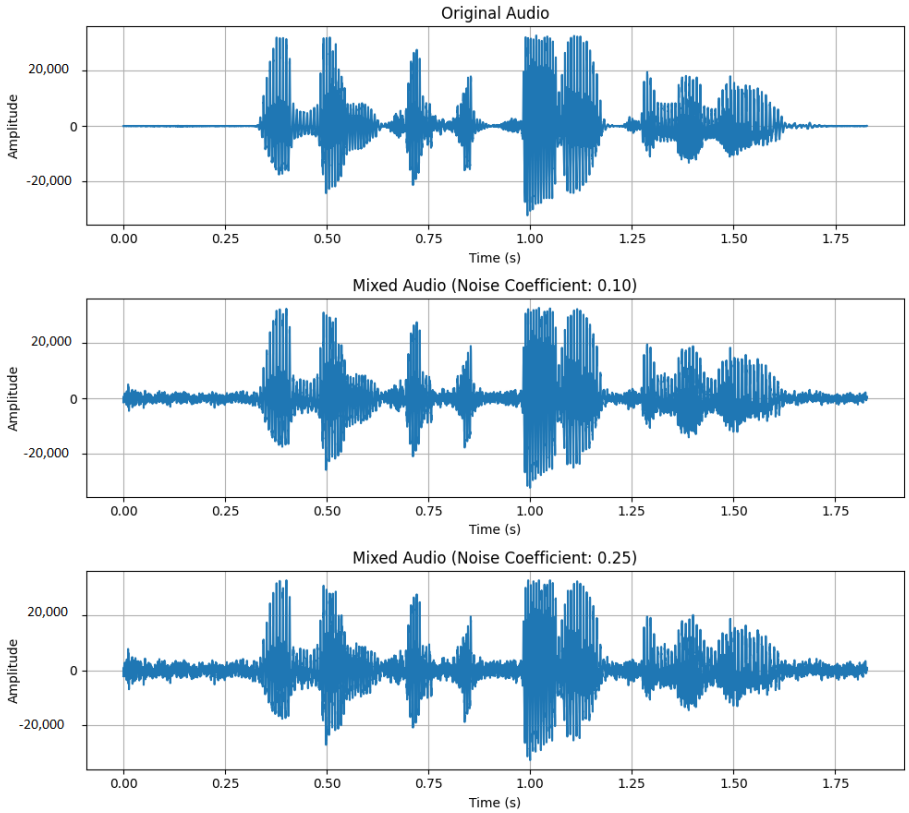

Figure 10. Comparison of oscillograms of the original voice image and its noisy versions.

6. Conclusions

A neural fuzzy extractor model that allows for generating cryptographic keys or passwords based on voice data for voice biometric systems was proposed. The proposed

model protects user biometric template data from being compromised. The new model is based on new types of neurons that are based on trigonometric measures of proximity and allow us to separate input data by correlation level.

This study showed that the considered trigonometric measures of proximity (3) and (4) have opposite properties to separate classes of images in a two-dimensional feature space. The measure based on the cosine distance (3) shows good results if the features are independent or have a weak correlation dependence. The measure based on the cotangent distance is more advantageous to use if the features are highly correlated. Based on these measures of proximity, two types of neurons are proposed, each of which should process features with a corresponding level of dependence. Moreover, the inputs of a neuron based on a cotangent measure of proximity should be set in such a way that they process positively or negatively correlated data separately. Each neuron is partially connected to avoid the possibility of an attack based on observing common connections between neurons [32].

The experiments used two datasets: publicly available RedDots, and the proposed AIC-spkr-130 dataset collected by us as part of the research, which takes into account the psychophysiological states of the subjects (the speakers pronounced password phrases not only in a normal state but also in a sleepy state, and after drinking alcohol). The neural fuzzy extractor model showed a fairly high result—EER = 0.021 (AIC-spkr-130) and EER = 0.032 (RedDots)—corresponding to the world level, while the model turned out to be resistant to changes in the speaker's state at the time of authentication. In addition, it ensures the confidentiality of biometric data and the user's key used in training the model without using cryptography. The length of the key generated by the proposed model is several times higher than the key length for the model trained according to the GOST R 52633.5 standard (1024 versus 160 bits). The model is also resistant to a selection attack [33], since each neuron can generate two binary states at the output, unlike the model from the GOST R 52633.5 standard.

Future research will be related to the transfer of the proposed neural fuzzy extractor model to other biometric modalities.

Author Contributions: A.S. (Alexey Sulavko), P.L., I.P. and A.V.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, writing—original draft preparation, writing—review and editing, supervision, project administration, funding acquisition. D.I. and A.S. (Alexander Samotuga): validation, formal analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the state assignment of the Ministry of Science and Higher Education of the Russian Federation, grant number (theme No.) FSGF-2023-0004.

Data Availability Statement: Data that are used in the present research, such as datasets from aiconstructor, are partially available to download from the following link: http://en.aiconstructor.ru/page40799810.html (accessed on 20 October 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Sulavko, A. Biometric-Based Key Generation and User Authentication Using Acoustic Characteristics of the Outer Ear and a Network of Correlation Neurons. *Sensors* **2022**, *22*, 9551. [CrossRef] [PubMed]
2. Snyder, D.; Garcia-Romero, D.; Sell, G.; Povey, D.; Khudanpur, S. X-Vectors: Robust DNN Embeddings for Speaker Recognition. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 5329–5333. [CrossRef]
3. Li, Y.; Chang, S.; Wu, Q. A short utterance speaker recognition method with improved cepstrum—CNN. *SN Appl. Sci.* **2022**, *4*, 330. [CrossRef]

4. Heo, H.S.; Lee, B.J.; Huh, J.; Chung, J.S. Clova Baseline System for the VoxCeleb Speaker Recognition Challenge 2020. *arXiv* **2020**, arXiv:2009.14153. [CrossRef]
5. Xiao, X.; Kanda, N.; Chen, Z.; Zhou, T.; Yoshioka, T.; Chen, S.; Zhao, Y.; Liu, G.; Wu, Y.; Wu, J.; et al. Microsoft Speaker Diarization System for the VoxCeleb Speaker Recognition Challenge 2020. In Proceedings of the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021. [CrossRef]
6. Kanagasundaram, A.; Sridharan, S.; Sriram, G.; Prachi, S.; Fookes, C. A Study of x-Vector Based Speaker Recognition on Short Utterances. In Proceedings of the Interspeech 2019, Graz, Austria, 15–19 September 2019; pp. 2943–2947. [CrossRef]
7. Liu, T.; Lee, K.A.; Wang, Q.; Li, H. Disentangling Voice and Content with Self-Supervision for Speaker Recognition. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 50221–50236. [CrossRef]
8. Jin, Y.; Hu, G.; Chen, H.; Miao, D.; Hu, L.; Zhao, C. Cross-Modal Distillation for Speaker Recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 12977–12985. [CrossRef]
9. Sarma, K.; Pyrtuh, F.; Chakraborty, D. Speaker Verification System using Wavelet Transform and Neural Network for short utterances. *Asian J. Converg. Technol.* **2020**, *6*, 30–35. [CrossRef]
10. Aronowitz, H. Speaker recognition using common passphrases in RedDots. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 5405–5409. [CrossRef]
11. Maghsoodi, N.; Sameti, H.; Zeinali, H.; Stafylakis, T. Speaker Recognition With Random Digit Strings Using Uncertainty Normalized HMM-Based i-Vectors. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *27*, 1815–1825. [CrossRef]
12. Jindal, A.K.; Shaik, I.; Vasudha, V.; Chalamala, S.R.; Ma, R.; Lodha, S. Secure and Privacy Preserving Method for Biometric Template Protection using Fully Homomorphic Encryption. In Proceedings of the 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Guangzhou, China, 29 December 2020–1 January 2021; pp. 1127–1134. [CrossRef]
13. Zhang, K.; Cui, H.; Yu, Y. Facial Template Protection via Lattice-Based Fuzzy Extractors. Cryptology ePrint Archive, Paper 2021/1559, 2021. Available online: https://eprint.iacr.org/2021/1559.pdf (accessed on 8 December 2024).
14. Monrose, F.; Reiter, M.; Li, Q.; Wetzel, S. Cryptographic key generation from voice. In Proceedings of the 2001 IEEE Symposium on Security and Privacy, S&P 2001, Oakland, CA, USA, 14–16 May 2001; pp. 202–213. [CrossRef]
15. Chee, K.Y.; Jin, Z.; Cai, D.; Li, M.; Yap, W.S.; Lai, Y.L.; Goi, B.M. Cancellable speech template via random binary orthogonal matrices projection hashing. *Pattern Recognit.* **2018**, *76*, 273–287. [CrossRef]
16. Ghouzali, S.; Bousnina, N.; Mikram, M.; Lafkih, M.; Nafea, O.; Al-Razgan, M.; Abdul, W. Hybrid Multimodal Biometric Template Protection. *Intell. Autom. Soft. Comput.* **2021**, *27*, 35–51. [CrossRef]
17. Sardar, A.; Umer, S.; Rout, R.K.; Sahoo, K.S.; Gandomi, A.H. Enhanced Biometric Template Protection Schemes for Securing Face Recognition in IoT Environment. *IEEE Internet Things J.* **2024**, *11*, 23196–23206. [CrossRef]
18. Sulavko, A.; Inivatov, D.; Vasilyev, V.; Lozhnikov, P. Authentication based on voice passwords with the biometric template protection using correlation neurons. *Inf. Control Syst.* **2024**, 21–38.
19. Alam, M.J.; Kenny, P.; Gupta, V. Tandem Features for Text-Dependent Speaker Verification on the RedDots Corpus. In Proceedings of the Interspeech 2016, Francisco, CA, USA, 8–12 September 2016; pp. 420–424. [CrossRef]
20. Sarkar, A.K.; Sarma, H.; Dwivedi, P.; Tan, Z.H. Data Augmentation Enhanced Speaker Enrollment for Text-dependent Speaker Verification. In Proceedings of the 2020 3rd International Conference on Energy, Power and Environment: Towards Clean Energy Technologies, Shillong, India, 5–7 March 2021; pp. 1–6. [CrossRef]
21. Alaya, B.; Laouamer, L.; Msilini, N. Homomorphic encryption systems statement: Trends and challenges. *Comput. Sci. Rev.* **2020**, *36*, 100235. [CrossRef]
22. Fuller, B.; Meng, X.; Reyzin, L. Computational fuzzy extractors. *Inf. Comput.* **2020**, *275*, 104602. [CrossRef]
23. Rathgeb, C.; Kolberg, J.; Uhl, A.; Busch, C. Deep Learning in the Field of Biometric Template Protection: An Overview. *arXiv* **2023**, arXiv:2303.02715. [CrossRef]
24. Jindal, A.K.; Chalamala, S.; Jami, S.K. Face Template Protection Using Deep Convolutional Neural Network. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 575–5758. [CrossRef]
25. Kuznetsov, O.; Zakharov, D.; Frontoni, E. Deep learning-based biometric cryptographic key generation with post-quantum security. *Multimed. Tools Appl.* **2024**, *83*, 56909–56938. [CrossRef]
26. Sadjadi, O. NIST SRE CTS Superset: A large-scale dataset for telephony speaker recognition. *arXiv* **2021**, arXiv:2108.07118.
27. Zue, V.; Seneff, S.; Glass, J. Speech database development at MIT: Timit and beyond. *Speech Commun.* **1990**, *9*, 351–356. [CrossRef]
28. El-Moneim, S.A.; Nassar, M.A.; Dessouky, M.I.; Ismail, N.A.; El-Fishawy, A.S.; El-Samie, F.E.A. Cancellable template generation for speaker recognition based on spectrogram patch selection and deep convolutional neural networks. *Int. J. Speech Technol.* **2022**, *25*, 689–696. [CrossRef]

29. Mahum, R.; Irtaza, A.; Javed, A. EDL-Det: A Robust TTS Synthesis Detector Using VGG19-Based YAMNet and Ensemble Learning Block. *IEEE Access* **2023**, *11*, 134701–134716. [CrossRef]
30. Xia, P.; Zhang, L.; Li, F. Learning similarity with cosine similarity ensemble. *Inf. Sci.* **2015**, *307*, 39–52. [CrossRef]
31. Akhmetov, B.; Ivanov, A.; Alimseitova, Z. Training of neural network biometry-code converters. *News Natl. Acad. Sci. Repub. Kazakhstan Ser. Geol. Tech. Sci.* **2018**, *1*, 61–68.
32. Marshalko, G.B. On the security of a neural network-based biometric authentication scheme. *Math. Probl. Cryptogr.* **2014**, *5*, 87–98. [CrossRef]
33. Bogdanov, D.S.; Mironkin, V.O. Data recovery for a neural network-based biometric authentication scheme. *Math. Probl. Cryptogr.* **2019**, *10*, 61–74. [CrossRef]
34. Malygin, A.; Seilova, N.; Boskebeev, K.; Alimseitova, Z. Application of artificial neural networks for handwritten biometric images recognition. *Comput. Model. New Technol.* **2017**, *5*, 31–38.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.