computers

MDPI

Article

Deep Feature Fusion of Fingerprint and Online Signature for Multimodal Biometrics

Mehwish Leghari [1,2,*], Shahzad Memon [1], Lachhman Das Dhomeja [1], Akhtar Hussain Jalbani [2] and Asghar Ali Chandio [2]

[1] Institute of Information and Communication Technology, University of Sindh, Jamshoro 76080, Pakistan; shahzad.memon@usindh.edu.pk (S.M.); lachhman@usindh.edu.pk (L.D.D.)
[2] Information Technology Department, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah 67480, Pakistan; jalbaniakhtar@quest.edu.pk (A.H.J.); asghar.ali@quest.edu.pk (A.A.C.)
* Correspondence: legharimehwish@quest.edu.pk

Abstract: The extensive research in the field of multimodal biometrics by the research community and the advent of modern technology has compelled the use of multimodal biometrics in real life applications. Biometric systems that are based on a single modality have many constraints like noise, less universality, intra class variations and spoof attacks. On the other hand, multimodal biometric systems are gaining greater attention because of their high accuracy, increased reliability and enhanced security. This research paper proposes and develops a Convolutional Neural Network (CNN) based model for the feature level fusion of fingerprint and online signature. Two types of feature level fusion schemes for the fingerprint and online signature have been implemented in this paper. The first scheme named early fusion combines the features of fingerprints and online signatures before the fully connected layers, while the second fusion scheme named late fusion combines the features after fully connected layers. To train and test the proposed model, a new multimodal dataset consisting of 1400 samples of fingerprints and 1400 samples of online signatures from 280 subjects was collected. To train the proposed model more effectively, the size of the training data was further increased using augmentation techniques. The experimental results show an accuracy of 99.10% achieved with early feature fusion scheme, while 98.35% was achieved with late feature fusion scheme.

Keywords: biometric fusion; feature-level fusion; multimodal biometrics; fingerprint and online signature recognition; convolutional neural network

Citation: Leghari, M.; Memon, S.; Dhomeja, L.D.; Jalbani, A.H.; Chandio, A.A. Deep Feature Fusion of Fingerprint and Online Signature for Multimodal Biometrics. *Computers* **2021**, *10*, 21. https://doi.org/10.3390/computers10020021

Academic Editor: M. Ali Akber Dewan
Received: 9 December 2020
Accepted: 3 February 2021
Published: 7 February 2021

1. Introduction

Biometric recognition has become an important part of daily life applications such as forensics, surveillance systems, attendance systems, unlocking the smart phones, Automated Tailored Machines (ATMs) and border and control systems in many countries due to the extensive research in this field [1]. The extensive use of the biometric data makes the end users concerned about the privacy of their biometric data. For example, in a password protected system if the password is compromised or hacked then the owner of the password can reset the password. In contrast to passwords and Personal Identification Numbers (PINs), the human biometrics is so naturally associated with the owner that it cannot be changed. For example, if facial data of someone is compromised or hacked then the actual owner cannot change his/her face. The biometric data like the facial data, fingerprint, hand geometry, voice, ear, iris and gait will remain unchanged for a user. Therefore, it is crucial to take all possible measures to protect the users' biometric data. Several measures have already been in practice which, ensure the privacy of the end users [2,3].

One of the solutions to ensure the end user's privacy is to combine the features from more than one biometrics [4]. The use of multimodal biometrics not only provides the reliability and accuracy but also ensures the security of the end user's data [5]. As the

data are stored in fused form, even if compromised, it will be more difficult to re-use the same data as the data fused from more than one biometrics trait does not pertain to any single biometric modality. Although the fusion of various biometric traits produces a more effective biometric system, fusing different biometric traits at feature level is complex due to incompatibilities in different features from the biometric modalities and a larger number of dimensions in the fused feature-set. Once the feature level fusion is achieved it yields more effective information as compared to other levels of fusion like sensor level and score level fusions [1].

The biometric may be explained as the measurement of the physiological or behavioral characteristics used to identify a person. Mainly, the biometrics is classified into two types: physiological and behavioral biometrics. Physiological biometric may be described as the calculations for the measurements of physical characteristics of a human body such as fingerprint, face, iris and hand geometry, whereas behavioral biometric may be defined as the measurements of the behavior of a person while performing some task, for example, voice, signature, gait and keystroke dynamics [6]. The physical biometric is comparably more accurate and reliable. However, there are some challenges to the security of the physical biometrics as those are more public and open in nature [6]. The face and fingerprints of the human beings are open to the public and can be easily captured by anyone. Once a high-resolution photograph is obtained, anyone can re-create and reuse human face or fingerprint [6]. While the behavioral biometrics is not as accurate and reliable as the physical ones, they are more personal to the owner and cannot be copied easily.

One of the solutions to enhance the security of biometric data is to use the multimodal biometrics. In a multimodal biometrics system, more than one biometric traits or modalities may be used to store the biometrics data in a fused form and thus provides more security as regeneration or replication of one biometric trait is very difficult from a fused biometrics template. Fusion of face with fingerprint, face with iris, fingerprint with iris and fingerprint with finger vein are among the common combinations for a multimodal biometrics system [7–9]. Therefore, it is essential to combine one of the physical traits with behavioral biometric traits to ensure the accuracy and privacy [10]. Furthermore, the combination of physical and behavioral biometric traits will increase the accuracy of a biometrics system and reduce the spoof attacks.

The biometrics data, if stored in an unencrypted form, is always prone to security attacks. One of the examples of such breach in security is "BioStar2" database [11]. The database stored biometrics data including fingerprints, facial recognition records and passwords of the users in an unencrypted form. The breach in the security in the database was discovered in 2019 by a team of security researchers. The services of the database were being used by thousands of the institutions of the developed and developing countries of the world including United Kingdom, United States of America, United Arab Emirates, Finland, Japan, Germany and many more. The data were obtained by security researchers just through the web browser by applying some operations. This shows that the biometric data of users can be compromised, thus it should be stored in such a fused and encrypted form, so that it cannot be reused, even if compromised.

Keeping in view the privacy of the users' biometric data, this research proposes and develops a feature level fusion scheme based on Convolutional Neural Network (CNN) for fingerprints and online signatures. This paper implemented two types of feature level fusion schemes using CNN—one is early fusion and the other is late fusion. Following are the main contributions of this study.

- A multimodal biometric dataset has been collected with 1400 fingerprints and 1400 online signatures from 280 subjects. As there are only a few databases that contain samples from both fingerprint and online signature biometric traits, there was a strong need to collect the data on a large-scale from the real users. For this reason, the data are collected from 280 subjects and utilized in this research. Furthermore, an Android application was developed to collect the online signatures.

- A CNN model is developed that fuses the features from fingerprints and online signatures into a single biometric template. Two types of fusion schemes: early and late fusion using CNNs are developed.
- Several experiments are performed by fine-tuning the hyper parameters of CNN using early and late fusion techniques on the new custom multimodal dataset, where the best results are obtained with early fusion and are compared with the related multimodal biometric systems.

Rest of the paper is organized as follows: Section 2 presents an overview of the research contribution in the field of biometric especially for the feature level fusion. Section 3 describes the database of fingerprints and online signatures collected for this research paper. Different steps of the proposed system based on both early feature fusion scheme and late feature fusion scheme, along with the experimental setup, are explained in detail in Section 4. Section 5 discusses the experimental results and compares the results with the related multimodal biometric systems while Section 6 concludes the paper and suggests some future directions for this research.

2. Related Work

The effectiveness of biometric fusion at the feature level is evident from the work of many researchers [4,7–9]. Biometric systems based on the fusion of features from multiple biometric traits offer more accuracy as compared with the accuracy achieved by using a single biometric trait. For example, Ahmad et al. [12] fused the features from the palmprint and face to achieve the higher accuracy rates than the palmprint or face alone. In their research the features were extracted using Gabor Filter. As the resultant feature vector offers a high dimensional data, hence the dimensions were reduced with the help of Principal Component Analysis (PCA) as well as Linear Discriminant Analysis (LDA). Then the features were combined serially and the Euclidian Distance algorithm was used to classify the fused template. The authors experimented on 400 images from the ORL face database and 400 images from Poly-U palmprint database. The Euclidean Distance classifier produced 99.5% accurate results for fused data. Similarly, Chanukya et al. [13] fused the features from fingerprint and ear biometrics. They extracted the features by applying modified region growing algorithm and Gabor filters to achieve an accuracy of 98% using a conventional neural network classifier.

It has always been important to select a good fusion technique to fuse the biometrics at the feature level. Various fusion schemes to fuse the biometric features have been adopted by the research society. Thepade et al. [14] fused the features from the human iris and palmprint. First, the images of palmprint and iris were fused together, then the features were reduced by using different transforms of energy distributions and comparison is made for various methods of the energy transform. The highest Genuine Acceptance Rate (GAR) achieved was 58.40% by Hartley transform. Another method for the fusion of iris and fingerprint was proposed by Guesmi et al. [15]. They extracted the features of iris and fingerprint using curvelet transform and selected the relevant features. The selected features were matched with the database using possibility matching and a GAR of 98.3% was achieved. Similarly, Xing et al. [16] proposed a new method for the feature level fusion of gait and face to be especially used by the images obtained from a Closed-circuit television (CCTV) camera. They extracted the eigenvalues from input images and minimized the difference between the values from same person. They fused the features by projecting the features of the same person into coupled subspace. To demonstrate their approach, they constructed a chimeric database comprising of 40 subjects using two publicly available databases of CASIA gait and ORL face. They used a nearest neighbor classifier and obtained an accuracy of 98.7%.

Different methods for the feature extraction of biometric modalities have been reported by the research community. For example, Sheikh et al. [17] used Dual Tree-Discrete Wavelet Transform (DT-DWT) to extract the features from the palmprint and finger knuckleprint. They used AdaBoost classifier and achieved the accuracy of 99.5%. In the same way,

Haghighat et al. [18] used Discriminant Correlation Analysis (DCA) for feature extraction from different sets of biometric modalities. It is proposed and experimentally proved by their research work that DCA assists in maximizing the difference between different classes and minimizing the difference between the same classes. They performed experiments on the feature fusion of face with ear and fingerprint with iris. They achieved the accuracies of over 99% by using a nearest neighbor classifier called k-Nearest Neighbors (kNN). In 2016 Jagadiswary et al. [19] presented their research on fusing features from the fingerprint, finger vein and the retina. In this work the feature extraction for the fingerprint was performed by extracting minutiae points from the fingerprints. The feature extraction for the retina was performed by extracting the blood vessel segmentation using threshold techniques based on Kirch's template. The blood vessel pattern of a finger vein was obtained using maximum curvature method and repeated line tracking method. The features were fused by concatenation method. The fused template was also encrypted to ensure further security. They achieved maximum GAR of up to 95%, the False Acceptance Rate (FAR) was reduced to as low as 0.01%. Some of the researchers even fused different feature extraction techniques to achieve the better results. To fuse the features from fingerprint, finger vein and the face, [8] calculated the Fisher vectors of the extracted features and then trained three classifiers including Support Vector Machine (SVM), Bayes classifier and kNN for recognition. Authors collected a dataset from 50 students with 5 samples from each student. The authors also tested the system using fake data and the maximum accuracy achieved by the system was 93%. It is evident from their experimental results that the kNN classifier outperformed the other two classifiers.

Some of the research has also demonstrated the power of biometric fusion by using different levels of fusions simultaneously. Azome et al. [20] combined the feature level and score level fusions. They proposed a multimodal biometric system based on iris and face. They applied the combination of various feature extraction methods to the face images from the ORL face database and iris images from the CASIA iris database. After feature combination they applied weighted score fusion to get the resultant accuracy. They used the nearest neighbor classifier and achieved up to 98.7% accuracy. Similarly, Toygar et al. [21] combined the features from the ear and the face profile for identical twins. They used both feature level and score level fusions to achieve the best recognition results. They used five separate feature extraction methods such as PCA, Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), Local Phase Quantization (LPQ) and Binarized Statistical Image Features (BISF). Their best achieved recognition results were 76% using kNN classifier. In the same way another noticeable research work was published by Sharifi et al. [22]. They fused the face and iris modalities at different levels to achieve a novel scheme for multimodal fusion. The modalities were fused at feature, score and decision levels. Features were extracted using Log-Gabor Transform technique. The LDA was used to reduce the size of the feature set. The best GAR achieved was 98.9%. In another research, the features from face, fingerprints and iris were fused together by Meena et al. [23]. Authors used a dataset of 280 subjects. Each subject contributed 4 samples for each of the 3 biometric traits. Hence a dataset of 2400 samples was used. Features were extracted by Local Derivative Ternary Pattern (LDTP). Two classifiers k-NN and SVM were trained to assess the accuracy and a maximum accuracy of 99.5% was acquired by the proposed system.

Recently, deep learning techniques have demonstrated the significant improvements over the traditional methods in the field of biometric recognition [24–26]. Deep feature level fusion techniques have been applied in several studies to fuse the features from different biometric traits for the multimodal biometric systems. In recent years, a multimodal biometric recognition using feature level fusion proposed by Xin et al. [8] combined the features from fingerprint, finger vein and the facial images for the recognition of persons. Zhang et al. [27] fused the periocular and iris features with the help of a deep CNN model for the mobile biometric recognition. For generating the specific and concise representation of each modality, they applied a maxout in the CNN and then merged the distinct features

of both modalities with the help of a weighted concatenation process. In the same manner, Umer et al. [28] combined the periocular and iris features for the biometric recognition of a person. They deployed different deep learning based CNN frameworks such as ResNet-50, VGG-16 and Inception-v3 for feature extraction and classification. They demonstrated that the performance of the system was improved by combining the features from various traits.

Similarly, Surendra et al. [29] developed a biometric based attendance system using a deep learning-based feature fusion technique on the iris features. The features from fingerprints and electrocardiogram (ECG) were fused by Jomaa et al. [30] to detect the presentation attacks. They deployed three CNN architectures including fully connected layers, 1-D CNN and 2-D CNN for the feature extraction from ECG and an EfficientNet for the feature extraction from the fingerprint. Recently, the periocular and the facial features have been combined by Tiong et al. [31] in a multi deep learning network for the facial recognition system. They further improved the recognition accuracy by combining the textural and multimodal features. The recent research discussed in this section has been summarized in Table 1.

Table 1. Summary of the recent research in the field of feature level fusion of biometrics.

Reference	Biometric Traits	Classifier	Accuracy or EER
Kaur et al. [9]	face and fingerprint	Neural Networks	94.4%.
Xin et al. [8]	fingerprint, face and finger vein	SVM, Bayes classifiers and kNN	93%
Ahmad et al. [12]	palmprint and face	Euclidean Distance	99.5%
Chanukya et al. [13]	fingerprint and ear	CNN	98%
Thepade et al. [14]	iris and palmprint	Hartley transform	GAR = 58.40%
Guesmi et al. [15]	iris and fingerprint	Possibility theory	EER = 0.3%
Xing et al. [16]	gait and face	Nearest Neighbor	98.7%
Sheikh et al. [17]	palmprint and finger knuckleprint	AdaBoost classifier	99.5%
Haghighat et al. [18]	face with ear and fingerprint with iris	kNN classifier	99%
Azome et al. [20]	iris and face	Nearest Neighbor	98.7%
Toygar et al. [21]	ear and profile face	kNN classifier	76%
Sharifi et al. [22]	face and iris		GAR = 98.9%
Meena et al. [23]	face, fingerprints and iris	kNN and SVM	99.5%
Zhang et al. [27]	periocular and iris	CNN	EER = 0.6%
Umer et al. [28]	periocular and iris	ResNet-50, VGG-16 and Inception-v3	CRR = 99.91%
Jomaa et al. [30]	fingerprints and electrocardiogram (ECG)	CNN	95.32%

As seen in Table 1 the literature on multimodal biometric fusion presents the diversity in fusion of biometric traits. However, most of the multimodal systems are based on the fusion of physiological traits only. As a person's most of the physiological traits are open to the public, it is beneficial to fuse the physiological biometric trait(s) with the behavioral biometric trait(s). To the best of our knowledge there has been no previous research on feature level fusion of fingerprints and online signatures using deep learning techniques. The feature level fusion of fingerprint and online signature was proposed in [32] however, they only proposed the fusion scheme and no experimental results were presented. Furthermore, the fingerprints and online signatures were fused at feature level using machine learning approach by Imran et al. [33]. However, they did not apply deep learning technique and used simple machine learning algorithms such as kNN and SVM. Their proposed system achieved an accuracy of 98.5% on a chimeric dataset synthesized

from Fingerprint Verification Competition (FVC) 2006 fingerprint dataset and MCYT-100 signature corpus. This research presents the feature level fusion of the fingerprint, the most common physiological biometric with the signature, the most common behavioral biometric for biometric identification using deep learning approach.

3. Dataset

A novel multimodal biometric dataset based on fingerprints and online signatures was collected for this paper. Most of the data were collected from the volunteer undergraduate students from four different universities and campuses including University of Sindh Dadu Campus, University of Sindh Mirpurkhas Campus, Quaid-e-Awam University of Engineering, Science & Technology Nawabshah and Sindh Agriculture University, Tandojam, Pakistan. Ages of the subjects were between 19 and 23 years. The dataset consist 1400 fingerprints and 1400 online signatures collected from 280 subjects. Each subject contributed five samples of the fingerprint from the same finger (i.e., ring finger of left hand) and five samples of their online signature. Fingerprint data were collected using SecuGen Hamster Plus fingerprint reader. The SecuGen Hamster Plus fingerprint reader uses Light Emitting Diode (LED) light source to capture the fingerprint images and saves the image in .bmp file format. The resolution of the captured image was 500 Dots Per Inch (DPI). Size of each fingerprint image was 260 × 300 pixels and the images were stored in a gray-scale format. A few samples of the fingerprints from the collected dataset are illustrated in Figure 1.

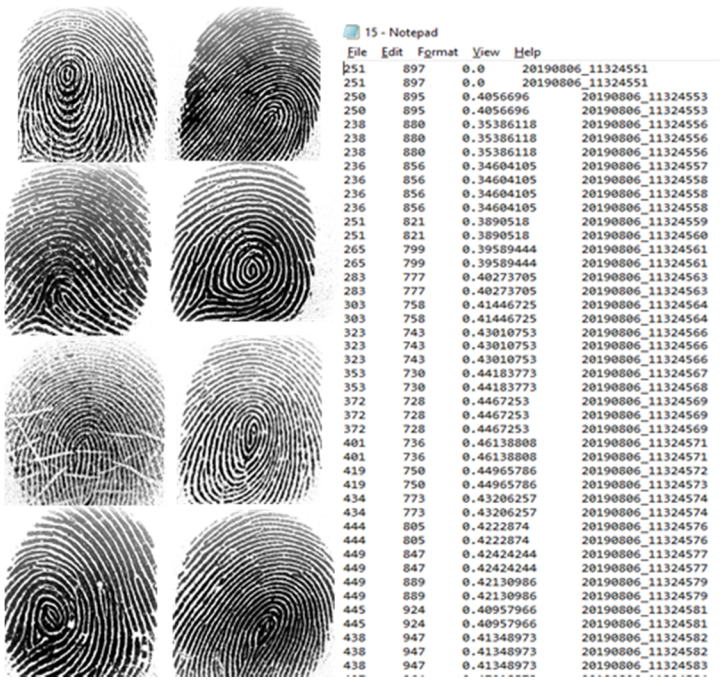

Figure 1. A view of the Multimodal Biometric Dataset collected for this paper: Fingerprint Samples (left), Online Signature Sample (right).

The other part of the dataset consists of online signatures. To collect online signature samples, an Android application was developed. The application was installed and used on a Samsung Galaxy Note 3 mobile phone with the S-Pen. The signatures were drawn on the screen of the smart phone with a pen and the application captured and stored the data in a text file in four columns namely x coordinates, y coordinates, pressure applied and the time data for each of the pixels of the signature. Each of the subject provided 5 samples of their signature. Figure 1 (right), presents a part of online signature template.

For the training of classifier, the data were separated into three sets: the training set, the validation set and the testing set. The training set consists of 60% of total data, the

validation set consists of 20% of the total data and the testing set also consists 20% of the total data. When five samples were divided into 60% for training, 20% for validation and 20% for testing the number of samples for each category became 3, 1 and 1, respectively. This small number of samples was not sufficient to train a deep learning model.

One of the problems encountered in this paper was the lack of samples for each class. The total of 5 samples have been collected for each subject or class. The justification for collecting the smaller quantity of samples was to represent the real-life situation. In real-life, it is unlikely for a person to provide 20 to 30 fingerprints or signatures for registering into a biometric system. Therefore, the data augmentation methods were applied to expand the data up to the three folds. For fingerprint images, data augmentation techniques such as rotation (not more than 10 degree), flipping, zooming and changing width, height and brightness or contrast of images were applied for expanding the size of training samples. Similarly, the augmented files for online signatures were created by slightly changing the values of x-coordinates and y-coordinates. This change made the effect of rotation of the signature, the increase in the height and width of the signature and the position of the signature. Similarly, the time data were changed slightly. However, the pressure data for each signature was not changed. After data augmentation, the dataset expanded up to 4200 fingerprints and 4200 online signatures for 280 classes with 15 samples of each modality per class.

4. Proposed Multimodal Biometric System Based on CNN

In the proposed multimodal biometric system, initially the fingerprint images and the online signature files were given as the input to the network and preprocessed to be used for the process of feature extraction. Preprocessing steps for the fingerprint include various methods from general image preprocessing techniques like image enhancement to more sophisticated techniques like thinning and binarization. First, the fingerprint image was enhanced in order to improve the quality of the image so that the ridges and the valleys on a fingerprint become clearly visible. After that, as all the images in the dataset were in gray-scale (0-255) so each image was binarized, means each pixel of the image was made equal to either 0 or 1. Some of the fingerprints during the preprocessing stage are illustrated in Figure 2.

In the next step, each ridge was thinned or skeletonized so that it becomes exactly one pixel in width. If the image was not enhanced, then the ridges were supposed to be distorted during the process of thinning and the one continuing ridge would have been shown as a broken ridge. Similarly, the preprocessing was accomplished for the online signature file by removing the duplicate values from the file. The values that were exactly same for all the four columns of the file were removed. After removal of the duplicate values, the zeros were appended to the end of the files to make the number of rows same in all the files.

After pre-processing, the task of feature extraction and classification is performed by the proposed CNN model. After the features are extracted from online signature and fingerprint by convolutional layers, the extracted features are fused together. Two types of feature fusion schemes; early and late fusion are proposed and developed in this paper. The early fusion technique achieved better accuracy than the late fusion technique. Both feature fusion techniques use same number of convolutional and fully connected layers. However, in the early fusion technique, the features are fused before the fully connected layers, while in the late fusion technique; the features are fused after the fully connected layers as illustrated in Figures 3 and 4. Each network is divided into two parts: fingerprint and online signature. The fingerprint part extracts features from the fingerprint images and online signature part extracts features from online signature file. The fingerprint part consists of five convolutional, five max pooling and two fully connected layers. The convolutional layers are used to extract the features from the fingerprint images. The low-level features were extracted by the initial convolutional layers, the later convolutional layers extract the middle and higher-level features, while the fully connected layers extract

more specific features. Similarly, online signature part consists of four convolutional layers, from which first and third layers were followed by a zero-padding layer and a pooling layer. Pooling layer in online signature part of the layer was also implemented using the max-pooling. In the feature fusion layer, the concatenation of features obtained by fingerprint part of the network and online signature part of the network has been performed using simple equation as below:

$$FF = concat(fp, os), \tag{1}$$

where the FF represents the feature vector that contains fused features from fingerprint and online signature. The fp was obtained by:

$$fp = W^T * K + b. \tag{2}$$

In Equation (2) the K represents the filter matrix of fingerprint, b denotes the network bias and the W^T stands for the weight metrics for the fingerprint. In the same way the os was obtained by the same equation.

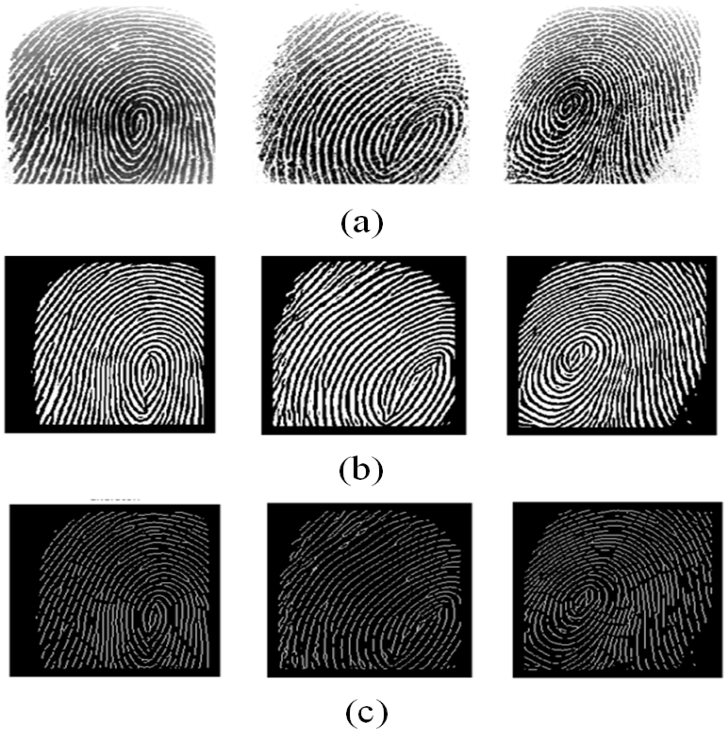

Figure 2. Some of the fingerprint images from proposed dataset after binarization, enhancement and thinning. (**a**) Presents original fingerprint images (**b**) presents the images after binarization and enhancement and (**c**) represents their corresponding images after skeletonization.

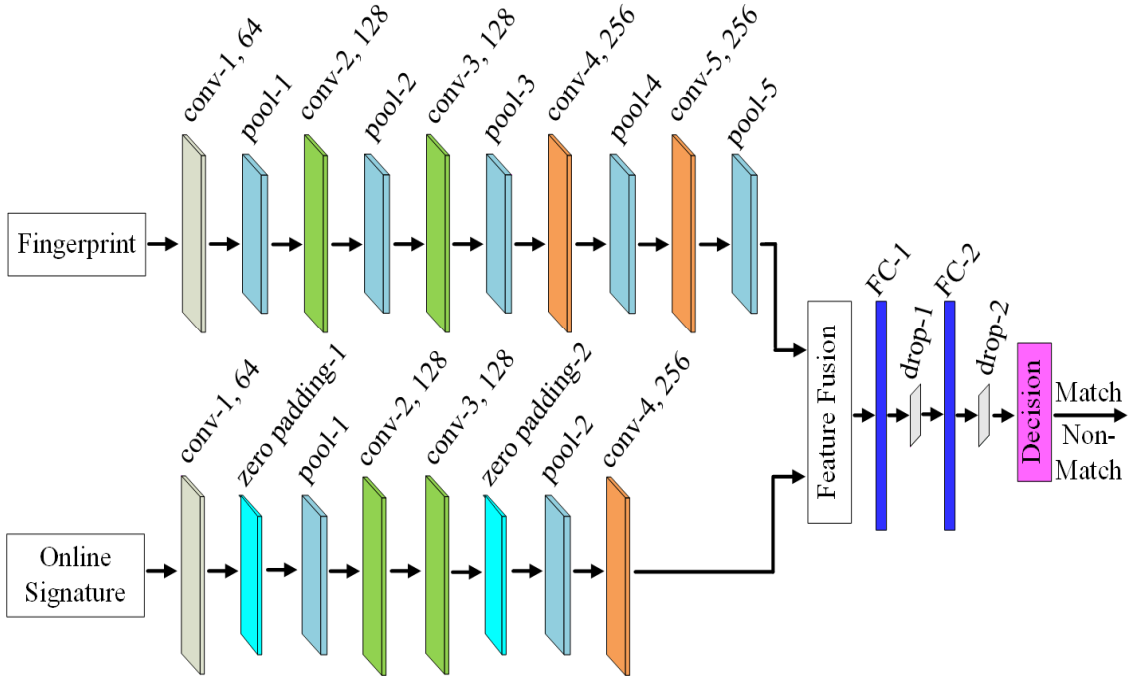

Figure 3. Architecture of the proposed deep feature fusion network for the early fusion of fingerprint and online signature recognition. "conv" represents the convolutional layers, "pool" represents the max-pooling layers, "FC" represents the fully connected layers, "drop" represents the dropout layers.

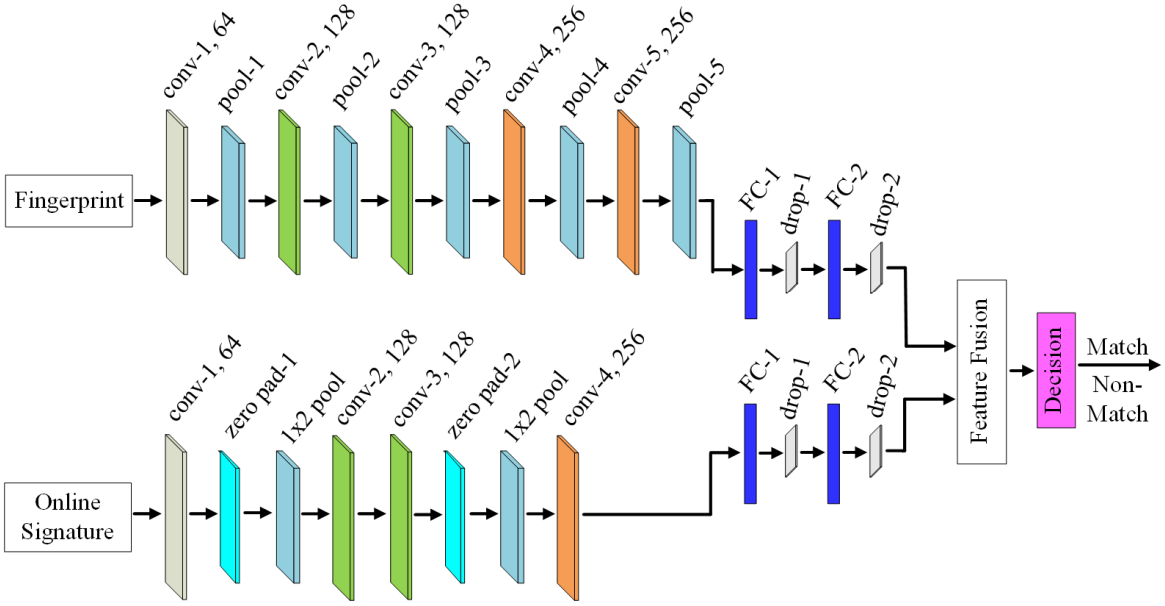

Figure 4. Architecture of the proposed deep feature fusion network for the late fusion of fingerprint and online signature recognition.

The dimensions of the online signature were fixed to 1×17, where 1 and 17 are the width and height of the signature, while for fingerprint image, the dimensions were fixed to $150 \times 150 \times 1$, where 150, 150 and 1 are the width, height and the depth of the fingerprint image. As the proposed feature fusion techniques combine the features of online signatures and fingerprint images and the CNN takes fixed size three-dimensional input, the dimensions of the online signature were reshaped to $1 \times 17 \times 1$ and passed to the input layer of the online signature network. Here the last 1 added shows the depth of the online signature. Though, reshaping of online signature from 1-D to 3-D increased

the network parameters but the fusion of two feature vectors, that is, fingerprint images and online signatures is possible when both have same dimensions. Therefore, the feature vectors of both fingerprint images and online signatures were fused together. The max pooling layer applied for online signature reduced the spatial dimensions of the feature vector. For online signature, the kernel size of the max pooling was set to 1×2. This reduced only the height of the signature without changing its width. Furthermore, the down-sampling of the signature to 1×8 after applying max pooling, reduced the accuracy of the overall system. The reason of the reduction in the accuracy was the smaller size of the signature data. The smaller number of features was extracted from the signature by the convolutional layers and the pooling layer further reduced the features. For that reason, two zero padding layers were applied to the signature before down-sampling by the pooling layer. The zero padding layers added an extra row and column of zeros to the signature data. The configuration of the online signature network is shown in Table 2. After adding zero padding layer to the signature data the accuracy of the system improved significantly.

Table 2. Configuration of the FingerNet and SigNet feature fusion model. 'F', 'K','P', 'S', 'T', 'B', 'L' and 'R' represent convolutional filters, kernel size, padding, stride, top, bottom, left and right values respectively.

Layer Type	FingerNet Parameters	Layer Type	SigNet Parameters
Input	$150 \times 150 \times 1$	Input	$1 \times 17 \times 1$
Convolutional	F: 64, K: 3×3, P: 'same'	Convolutional	F: 64, K: 3×3, P: 'same'
Max-Pooling	K: 2×2, S: 2×2	Zero Padding	T: 2, B: 2, L: 1, R: 1
Convolutional	F: 128, K: 3×3, P: 'same'	Max-Pooling	K: 1×2, S: 1×2
Max-Pooling	K: 2×2, S: 2×2	Convolutional	F: 128, K: 3×3, P: 'same'
Convolutional	F: 128, K: 3×3, P: 'same'	Convolutional	F: 128, K: 3×3, P: 'same'
Max-Pooling	K: 2×2, S: 2×2	Zero Padding	T: 2, B: 2, L: 0, R: 0
Convolutional	F: 256, K: 3×3, P: 'same'	Max-Pooling	K: 1×2, S: 1×2
Max-Pooling	K: 2×2, S: 2×2	Convolutional	F: 256, K: 3×3, P: 'same'
Convolutional	F: 256, K: 3×3, P: 'same'		
Max-Pooling	K: 2×2, S: 2×2		
Concatenate		[FingerNet | SigNet]	
Fully Connected		Neurons: 512	
Dropout		Ratio: 0.3	
Fully Connected		Neurons: 512	
Dropout		Ratio: 0.3	
Softmax		Classes: 280	

4.1. Early Feature Fusion Scheme

As discussed, two types of feature fusion schemes are proposed in this paper. One is based on early fusion scheme in which the features extracted by convolutional layers were fused or concatenated together before the fully connected layers. The concatenated feature vector was passed to two fully connected layers each followed by a ReLU layer. As the fully connected layers increase the training data by extracting more features from the data, so the dropout layers were added after both fully connected layers to randomly drop the learning parameters with a value of 0.3 to boost the training time of the network. Finally, a Softmax layer was used to distribute the results into the classes using probability distribution obtained for each class.

4.2. Late Feature Fusion Scheme

The second type of feature fusion scheme proposed for the fusion of fingerprint and online signature is the late feature fusion scheme. In this type of feature fusion scheme, the features extracted by convolutional layers were further passed to the two fully connected layers each followed by a dropout layer. Then the extracted features were concatenated and passed to the Softmax layer for the classification. Here again the dropout layers randomly drop the learning parameters with a value of 0.3 in order to improve the training time

of the network. Finally, a Softmax layer has been used for the classification based on the probability distribution obtained for each class.

5. Experiments and Results

This section discusses different parameters used during the performance of experiments for this research. Then the results obtained are discussed and compared with the related multimodal biometric systems.

5.1. Implementation Details

The proposed network was deployed with the help of the Keras library. Several experiments were performed by fine-tuning the hyper parameters of the CNN and the best results were achieved with the network configurations as shown in Table 2. To extract more useful features, a 3×3 filter was used by the convolutional layer. In the convolutional layer the number of filters and their sizes have the values as follows: {number_filters: filter_size}{64: 3}, {128: 3}, {128: 3}, {256: 3} and {256: 3}. To keep the same size of input and output of convolutional layer, each convolutional layer in fingerprint and online signature networks used the 'same' padding value with a stride of 1.

A window size of 2×2 with a stride value of 2×2 was used by the max-pooling layer. The number of filters in last convolutional layer was set to 256 for both fingerprint feature extraction sub-layer and online signature feature extraction sub-layer. Setting the output filter size equal to the 256 reduced the number of features and increased the accuracy. The feature level fusion was performed by feature concatenation operation on the output of both feature extraction sub-layers. Some experiments were also performed in which the number of outputs from the last convolutional layer was set to 512. However, increase in the output filter increased the dimensions of the fused vector and decreased the performance of the network. For that reason, the number of output filter from the last layer was set to 256 as it yielded the best performance.

Furthermore, an addition operation was also performed for fusion of the features from fingerprint feature extraction sub-layer and online signature feature extraction sub-layer to analyze the effects of addition on the performance. However, the addition operation on the extracted features did not increase the performance of the network in terms of accuracy.

5.2. Network Training

The training of the network was performed using Stochastic Gradient Descent (SGD) optimizer with a momentum rate of 0.9, learning rate of 0.005 and a weight decay of 0.00025. Different batch sizes were trialed during the network training but the best accuracy was achieved with a batch size of 64. The network was trained for 200 epochs. The network was trained on the real and augmented dataset of fingerprint and online signatures collected from 280 users.

5.3. Evaluation Metrics

The experimental results of this paper are presented in terms of the common evaluation metrics such as precision, recall, f1-scores and the accuracy scores. The precision was calculated by using following equation:

$$Precision = \frac{TP}{TP + FP}. \tag{3}$$

The *Precision* represents the percentage of the relevant results, where *TP* stands for the true positives and represents the number of actually correct matches that are classified as the correct matches by the classifier and *FP* stands for the false positives and represents the number of actually false matches that are misclassified as the correct matches by the classifier. Similarly, the recall is calculated by the following equation:

$$Recall = \frac{TP}{TP + FN}. \tag{4}$$

In Equation (4) the *Recall* represents the percentage of the total relevant results that are correctly classified by the classification algorithm. Here the *FN* stands for false negatives and represents the positive values that are incorrectly classified as the negative values by the classifier. In the same way *f*1-*score* represents the harmonic means of the recall and precision and is presented in Equation (5). Furthermore, the accuracy of the proposed model is obtained using the Equation (6).

$$f1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

5.4. Results

The early and late fusion networks were evaluated on the new collected dataset. Table 3 presents the results obtained from both types of feature fusion on proposed dataset in terms of recall, precision, f1-score and accuracy. From the early feature fusion scheme, the precision achieved is 98%. The recall obtained by the proposed model using early fusion is 99% while the f1-score is also 99%. Similarly, the precision achieved from the late fusion is 98%. And the same values of 98% have been achieved for both recall and f1-score from the late fusion scheme. The accuracy obtained with early and late feature fusion schemes is 99.10% and 98.35%, respectively. To further validate the performance of the proposed early and late fusion techniques on the newly developed dataset, five-fold and ten-fold cross validations have also been performed for each fusion technique where 97%, 96%, 95% and 95% validation scores were obtained with early and late feature fusion methods, respectively. Table 4 shows the results of 5-fold and 10-fold cross validation performed for early and late fusion techniques.

Table 3. Precision, recall and f1-score of the proposed early feature fusion and late feature fusion schemes.

Fusion Scheme	Precision	Recall	F1-Score	Accuracy
Early Feature Fusion	98%	99%	99%	99.10%
Late Feature Fusion	98%	98%	98%	98.35%

Table 4. 5-fold and 10-fold cross validation scores obtained with early and late fusion schemes.

k-Fold Value	Cross Validation Score: Early Fusion Scheme	Cross Validation Score: Late Fusion Scheme
5-fold	97%	95%
10-fold	96%	95%

5.5. Comparison with the Related Feature Level Fusion Methods

Table 5 presents related multimodal biometric fusions and compares their results with the results obtained from the proposed multimodal biometric fusion system. It can be learnt from Table 5, that most of the multimodal biometric fusion experiments have been performed using a limited amount of data [8]. Most of the researchers like [33,34] have used chimeric data by merging one biometric modality from one dataset and the other biometric modality from the other dataset to compose a larger multimodal dataset. Many research works have been using simple machine learning classifiers. Recently, it has become a common practice to use deep learning and convolutional neural networks for biometric recognition [7] instead of simple machine learning classifiers [8,9,33–37]. Research shows that the use of deep neural networks can increase the accuracy rates [38]. It is evident from Table 5, that the proposed multimodal biometric fusion is comparable with most of the related fusion schemes.

Table 5. Comparison of proposed deep learning based multimodal biometric fusion approach with some of the related multimodal fusion methods.

Reference	Biometric Traits	Database	Classifier	Accuracy
Xin et al. [8]	Fingerprint, face and finger vein	5 samples per trait from 50 subjects	SVM, Bayes classifiers and kNN	93%
Huang et al. [39]	Face and Gait	100 videos of 25 persons	Baysian and Nearest Neighbour classifier	65% to 100%
El-Alfy et al. [34]	Face and Gait	CMU PIE, FERET, AR Face database and USF Human ID gait database	SVM	88.6%
Gawande et al. [35]	Face and Gait	CASIA database	SVM	96.3%
Kondapi et al. [7]	Face and Iris	158136 Facial images of 550 persons	Convolutional Neural Network	95%
Milind et al. [36]	Face and Palmprint	Face 94, Face 95, Face 96 and IIT Delhi database	SVM	98%
Xinman et al. [37]	Face and Voice	100 subjects with 10 face images and 5 voice samples for each subject	SVM	93.6%
Proposed deep feature fusion of fingerprint and online signature	Fingerprint and Online Signature	1400 fingerprints and 1400 online signatures from 280 subjects	Convolutional Neural Network	99.1%

6. Conclusions and Future Work

In this paper, deep learning models based on the CNN architecture have been proposed for the feature level fusion of online signatures and fingerprints. Two feature fusion techniques, that is, early and late have been developed where the features extracted from both biometric modalities are fused together at convolutional and fully connected layers. The size of the input image for the fingerprint is fixed to $150 \times 150 \times 1$ and the size of the online signature file is 1×17. To fuse the features of fingerprint and online signature, the size of the signature was reshaped to $1 \times 17 \times 1$ before passing to the online signature network. To fuse the features of fingerprint image and online signature, different approaches were tried. However, the accuracy and other values for other evaluation metrics for the proposed system did not improve because of the width of the online signature's feature vector which was equal to 1. The problem was addressed and the accuracy and the values for other evaluation metrics for the system was increased by adding two zero-padding layers in the signature network. By this zero-padding technique, the extra zeros were added at all four sides of the feature vector, that is, top, bottom, left and right. In this way, the dimensions of the final feature vector became 4×4 in size. Similarly, the size of final feature vector of fingerprint was 4×4. These features have been fused by concatenation and passed to the fully connected layers for more abstract feature extraction and classification. The model was trained and tested on the new collected dataset and finally, the overall system achieved an accuracy of 99.10% with early fusion scheme and 98.35% with the late fusion scheme.

In future, low level characteristics or level 3 features of the fingerprint like ridge contours and active sweat pores may also be used for the fusion to ensure more accuracy and liveness of a user. In future one of the different-state-of-the-art cryptography techniques for biometrics may also be applied to the proposed system to further ensure the security of the fused biometric template.

Author Contributions: M.L. and A.A.C. collected the data, developed the model and performed the experiments. S.M., L.D.D. and A.H.J. supervised and verified all the experiments performed. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted as a part of PhD degree requirements of the corresponding author. The guidelines and approval of this study was given by Shahzad Memon.

Informed Consent Statement: Information consent was obtained from all subjects involved in this study. All the participants were informed about the purpose of the data collection.

Data Availability Statement: The multimodal biometrics dataset used in this research will be provided by sending an email to the corresponding author.

Acknowledgments: Authors would like to thank the administration and students of the University of Sindh Mirpurkhas campus, the University of Sindh Dadu Campus, the Quaid e Awam University of Engineering, Science and Technology Nawabshah and the Sindh Agriculture University Tandojam for providing full support and cooperation in the process of data collection.

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

References

1. Jain, A.K.; Nandakumar, K.; Ross, A.A. 50 years of biometric research: Accomplishments, challenges and opportunities. *Pattern Recognit. Lett.* **2016**, *79*, 80–105. [CrossRef]
2. Nandhinipreetha, A.; Radha, N. Multimodal biometric template authentication of finger vein and signature using visual cryptography. In Proceedings of the 2016 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 7–9 January 2016; pp. 1–4.
3. Datta, P.; Bhardwaj, S.; Panda, S.N.; Tanwar, S.; Badotra, S. Survey of Security and Privacy Issues on Biometric System. In *Handbook of Computer Networks and Cyber Security*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 763–776.
4. Nagar, A.; Nandakumar, K.; Jain, A.K. Multibiometric cryptosystems based on feature-level fusion. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 255–268. [CrossRef]
5. Ross, A.A.; Nandakumar, K.; Jain, A.K. *Handbook of Multibiometrics*; Springer: Berlin/Heidelberg, Germany, 2006.
6. Howell, K. IPSwitch. Available online: https://blog.ipswitch.com/3-reasons-biometrics-are-not-secure (accessed on 13 May 2020).
7. Kondapi, L.; Rattani, A.; Derakhshani, R. Cross-illumination Evaluation of Hand Crafted and Deep Features for Fusion of Selfie Face and Ocular Biometrics. In Proceedings of the 2019 IEEE International Symposium on Technologies for Homeland Security (HST), Boston, MA, USA, 5–6 November 2019; pp. 1–4.
8. Xin, Y.; Kong, L.; Liu, Z.; Wang, C.; Zhu, H.; Gao, M.; Zhao, C.; Xu, X. Multimodal Feature-Level Fusion for Biometrics Identification System on IoMT Platform. *IEEE Access* **2018**, *6*, 21418–21426. [CrossRef]
9. Sandhu, N.K.; Patterh, M.S. A Biometric Fusion Based on Face and Fingerprint Recognition using ANN. *Int. J. Recent Innov. Trends Comput. Commun.* **2017**, *5*, 88–92.
10. Dinca, L.M.; Hancke, G.P. Fall of One, the Rise of Many: A Survey on Multi-Biometric Fusion Methods. *IEEE Access* **2017**, *5*, 6247–6289. [CrossRef]
11. Scott, I. CPO Magazine, Breach of Biometrics Database Exposes 28 Million Records Containing Fingerprint and Facial Recognition Data. Available online: https://www.cpomagazine.com/cyber-security/breach-of-biometrics-database-exposes-28-million-records-containing-fingerprint-and-facial-recognition-data/ (accessed on 26 December 2019).
12. Ahmad, M.I.; Wai, L.W.; Satnam, S.D. Multimodal biometric fusion at feature level: Face and palmprint. In Proceedings of the CSNDSP, Newcastle Upon Tyne, UK, 21–23 July 2010; pp. 801–805.
13. Chanukya, P.S.; Thivakaran, T.K. Multimodal biometric cryptosystem for human authentication using fingerprint and ear. *Multimed. Tools Appl.* **2020**, *79*, 659–673. [CrossRef]
14. Thepade, S.D.; Bhondave, R.K.; Mishra, A. Comparing Score Level and Feature Level Fusion in Multimodal Biometric Identification Using Iris and Palmprint Traits with Fractional Transformed Energy Content. In Proceedings of the 2015 International Conference on Computational Intelligence and Communication Networks (CICN), Jabalpur, India, 12–14 December 2015; pp. 306–311.
15. Guesmi, H.; Trichili, H.; Alimi, A.M.; Solaiman, B. Novel biometric features fusion method based on possibility theory. In Proceedings of the 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC), Beijing, China, 6–8 July 2015; pp. 418–425.
16. Xing, X.; Wang, K.; Lv, Z. Fusion of gait and facial features using coupled projections for people identification at a distance. *IEEE Signal Process. Lett.* **2015**, *22*, 2349–2353. [CrossRef]
17. Oveisi, I.S.; Modarresi, M. A feature level multimodal approach for palmprint and knuckleprint recognition using AdaBoost classifier. In Proceedings of the 2015 International Conference and Workshop on Computing and Communication (IEMCON), Vancouver, BC, Canada, 15–17 October 2015; pp. 1–7.

18. Haghighat, M.; Abdel-Mottaleb, M.; Alhalabi, W. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1984–1996. [CrossRef]
19. Jagadiswary, D.; Saraswady, D. Biometric authentication using fused multimodal biometric. *Procedia Comput. Sci.* **2016**, *85*, 109–116. [CrossRef]
20. Azom, V.; Adewumi, A.; Tapamo, J.R. Face and Iris biometrics person identification using hybrid fusion at feature and score-level. In Proceedings of the 2015 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), Port Elizabeth, South Africa, 26–27 November 2015; pp. 207–212.
21. Toygar, Ö.; Alqaralleh, E.; Afaneh, A. Symmetric ear and profile face fusion for identical twins and non-twins recognition. *Signal Image Video Process.* **2018**, *12*, 1157–1164. [CrossRef]
22. Sharifi, O.; Eskandari, M. Optimal face-iris multimodal fusion scheme. *Symmetry* **2016**, *8*, 48. [CrossRef]
23. Meena, K.; Malarvizhi, N. An Efficient Human Identification through MultiModal Biometric System. *Braz. Arch. Biol. Technol.* **2016**, *59*. [CrossRef]
24. Minaee, S.; Abdolrashidi, A.; Su, H.; Bennamoun, M.; Zhang, D. Biometric recognition using deep learning: A survey. *arXiv* **2019**, arXiv:1912.00271.
25. Talreja, V.; Valenti, M.C.; Nasrabadi, N.M. Multibiometric secure system based on deep learning. In Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing (globalSIP), Montreal, QC, Canada, 14–16 November 2017; pp. 298–302.
26. Al-Waisy, A.S.; Qahwaji, R.; Ipson, S.; Al-Fahdawi, S.; Nagem, T.A. A multi-biometric iris recognition system based on a deep learning approach. *Pattern Anal. Appl.* **2018**, *21*, 783–802. [CrossRef]
27. Zhang, Q.; Li, H.; Sun, Z.; Tan, T. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2897–2912. [CrossRef]
28. Umer, S.; Sardar, A.; Dhara, B.C.; Rout, R.K.; Pandey, H.M. Person identification using fusion of iris and periocular deep features. *Neural Netw.* **2020**, *122*, 407–419. [CrossRef]
29. Surendra, I.; Sashank, T.S.; Praveena, M.A.; Manoj, R.J. Deep feature fusion for IRIS based on industrial biometric engineering. In Proceedings of the AIP Conference, Bangalore, India, 17–18 January 2020; p. 040003.
30. Jomaa, M.R.; Mathkour, H.; Bazi, Y.; Islam, M.S. End-to-End Deep Learning Fusion of Fingerprint and Electrocardiogram Signals for Presentation Attack Detection. *Sensors* **2020**, *20*, 2085. [CrossRef]
31. Tiong, L.C.O.; Kim, S.T.; Ro, Y.M. Implementation of multimodal biometric recognition via multi-feature deep learning networks and feature fusion. *Multimed. Tools Appl.* **2019**, *78*, 22743–22772. [CrossRef]
32. Leghari, M.; Memon, S.; Chandio, A.A. Feature-level fusion of fingerprint and online signature for multimodal biometrics. In Proceedings of the 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 3–4 March 2018; pp. 1–4.
33. Imran, M.; Kumar, H.; Jabeen, N.S.; Alaei, F. Accurate person recognition on combining signature and Fingerprint. *Int. J. Mach. Intell.* **2011**, *3*, 277–281.
34. El-Alfy, E.S.M.; BinMakhashen, G.M. Improved personal identification using face and hand geometry fusion and support vector machines. In Proceedings of the International Conference on Networked Digital Technologies, Dubai, United Arab Emirates, 24–26 April 2012; Volume 294, pp. 253–261.
35. Gawande, U.; Zaveri, M.; Kapur, A. A novel algorithm for feature level fusion using SVM classifier for multibiometrics-based person identification. *Appl. Comput. Intell. Soft Comput.* **2013**, *2013*, 515918. [CrossRef]
36. Rane, M.E.; Deshpande, P.P. Multimodal Biometric Recognition System Using Feature Level Fusion. In Proceedings of the 2018 IEEE Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018; pp. 1–5.
37. Zhang, X.; Dai, Y.; Xu, X. Android-Based multimodal biometric identification system using feature level fusion. In Proceedings of the 2017 IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Xiamen, China, 6–9 November 2017; pp. 120–124.
38. Zhong, D.; Shao, H.; Du, X. A Hand-Based Multi-Biometric via Deep Hashing Network and Biometric Graph Matching. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 3140–3150. [CrossRef]
39. Huang, Y.; Xu, D.; Nie, F. Patch distribution compatible semisupervised dimension reduction for face and human gait recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 479488. [CrossRef]