IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Cancelable Biometrics using Deep Learning as a Cloud Service

TANUJA SUDHAKAR AND MARINA GAVRILOVA
Department of Computer Science, University of Calgary, Calgary, AB T2N 1N4, Canada.

Corresponding author: Tanuja Sudhakar (e-mail: tanuja.sudhakar1@ucalgary.ca).

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada by the Discovery Grant (DG) on Machine Intelligence for Biometric Security under Grant 10007544 and by the Strategic Partnership Grant on Biometric-enabled Identity Management and Risk Assessment for Smart Cities, Grant 10022972.

ABSTRACT Cloud computing is a technology that has gained rapid popularity in recent years. It has enabled use of immense computational power in a scalable and cost-efficient manner. Deployment of biometric technology in government and commercial organizations has become a standard security practice. However, independent biometric systems tend to be computationally and financially expensive, especially when user enrollment is high. A feasible solution is to create a biometric system on the cloud which can be used ubiquitously as an authentication service. In this paper, we propose a first cancelable biometric framework based on deep learning on the cloud. We establish that cloud is a good solution for biometric systems where intensive computation, quick response times, and high accuracy is required.

INDEX TERMS Biometric privacy, cloud computing, cancelable biometrics, deep learning.

I. INTRODUCTION

TODAY, many organizations are turning towards cloud service, due to its availability of resources on-demand and pay per usage benefits. This paper proposes a cancelable biometric system using deep learning on the Amazon Web Service (AWS) cloud platform, which can be used as a service by various clients. It also focuses on protection of biometric data on the cloud using cancelable biometrics.

Today, many organizations prefer biometric authentication instead of traditional passwords. According to a 2019 data breach investigations report by Verizon [1], it has been found that passwords account for 81% of the data breaches. This problem can be eliminated using biometrics, which provide a proof of identity. A biometric system utilizes one or more physical or behavioral traits such as irises, fingerveins, or fingerprints to recognize a person. Due to its distinctive features, biometrics are difficult to recreate or forge. Benefits include high accuracy, non-repudiation, and permanency. Due to such benefits, the global biometric market is expected to grow over 24.8 billion USD by 2021 [3]. However, as biometric systems mature, certain challenges emerge. When the enrollment rate is high, computational and matching complexity increases. With this comes increased costs and data storage. A good solution to tackle this issue is to offload the biometric module to the cloud, to cut costs and still maintain performance. Parallel computation using multiple servers and nodes distributes processing, thereby reducing response time. Thus, Biometrics-as-a-Service (BaaS) or cloud based biometric authentication emerged as a high-impact research area [2] [22].

Cloud based biometric authentication can streamline customer service and reduce fraudulent activity [2]. However, with cloud comes a different set of problems. The main challenges are ensuring privacy of biometric data on the cloud and security of biometric template during transmission to the cloud. In this work, a novel biometric system based on deep learning on the cloud has been proposed. To address the issue of privacy of biometric data on the cloud, we use cancelable biometrics. With regards to template security during transmission, a unique combination of biometric crossfolding and biometric-QR code embedding has been utilized, along with security protocols, to prevent spoofing and replay attacks.

Cancelable Biometrics (CB) is a privacy technique to intentionally distort user biometrics to create an indecipherable template [4]. The CB template is then stored and used for subsequent authentication of the user. Two major advantages of CB are: a) the original biometric is never used and b) the CB template can be cancelled or revoked in case of suspicious activity. In this work, CB templates are used in the matching module on the cloud to help preserve user privacy. Cancelability of crossfolded templates has been rendered through a cancelable module based on random projection on

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3003869, IEEE Access

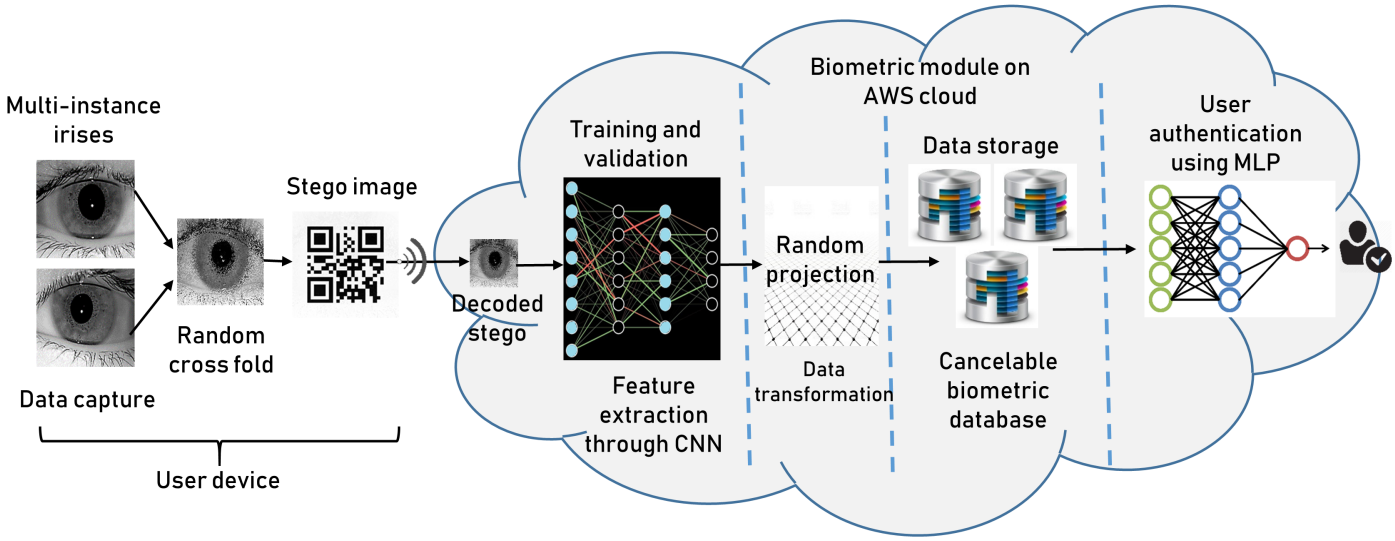


FIGURE 1: Proposed cancelable biometric system based on deep learning, offloaded to the cloud.

the cloud. Next, for safe transmission of biometrics to the cloud, a novel combination of steganography using QR codes and biometric crossfolding has been employed. Steganography is an ancient data hiding technique [23] where, *Stegano* means to conceal and *Graphe* means writing. Steganography has an advantage over cryptography because it conceals the fact that a secret message is being sent. The stego image looks like any other ordinary image which does not attract attention. In this work, a unique combination of steganography and cancelable biometrics is proposed for biometric security on the cloud. The main biometric engine based on Convolutional Neural Network (CNN) and the matching module based on Multi Layer Perceptron (MLP) have been designed on the cloud to create an accurate, time, and cost efficient cancelable system.

The goal of this work is to develop a novel multi-instance cancelable biometric system based on deep learning on the AWS cloud platform. Multi-instance biometrics have been utilized in this project due to certain advantages over unimodal biometrics. Multi-biometrics offers improved accuracy, reliability, and enhanced security [33] [27]. It helps to counter spoofing attacks by making it difficult for an adversary to simultaneously spoof multiple traits of a genuine user. Last, but most importantly, the use of deep learning models such as the CNN for feature extraction and MLP for user recognition has been explored in this work. CNN's properties of shared weights, translation invariance, convolutive layers, minimal preprocessing, and lower network complexity as compared to other deep learning architectures makes it a powerful tool in image recognition [24]. These reasons have motivated the use of CNN to extract features of the crossfolded biometrics. The biometric matching module further consists of a novel MLP model for user verification, which is another fully connected Artificial Neural Network (ANN). MLP has demonstrated high capability in classification, regression, and mapping due to its ability to distinguish data that is not linearly separable [26]. Moreover, the non-linear mapping property of mapping an N-dimensional input signal to an M-dimensional output signal, makes it a suitable matching module in this work.

Key contributions of this work are:

- A unique combination of biometric cross folding and steganography using QR codes has been proposed. This incorporates security of biometric templates over the transmission medium from spoofing attacks. The QR code which embeds the biometric also acts as a One Time Password (OTP), to prevent replay attacks.
- Novel use of CNN for feature extraction of cross folded biometrics on the cloud has been proposed in this work.
- Protection of biometric templates on the cloud has been rendered through the technique of random projection. This cancelable method has shown to be non-invertible, ensuring safety of original biometrics even if an adversary obtains the cancelable template and the user key.
- A novel MLP architecture has been deployed on the cloud for user recognition. Proposed MLP obtains very high accuracy and outperforms various classical machine learning methods.

Use of cancelable multi-biometric deep learning architecture on the cloud increases speed, quality of training, and is cost effective as can be seen from the experimental section.

II. LITERATURE REVIEW

Biometrics-as-a-Service (BaaS) is an emerging concept which is based on the idea of offering a biometric authentication solution as a cloud service. Applications, systems, and business security can be achieved by integrating biometric as a service authentication. Recent IEEE cloud computing special issue [2] explores how BaaS can mitigate fraudulent activity and provide quick biometric service in a cost-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3003869, IEEE Access

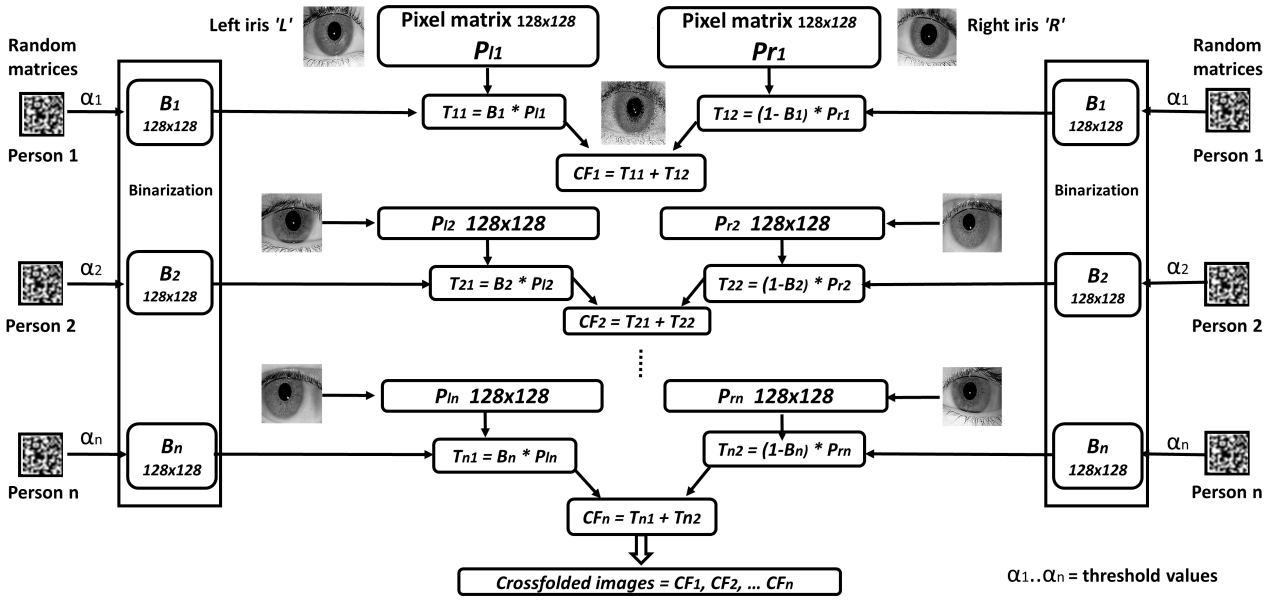


FIGURE 2: Generation of the crossfolded iris from a user's left and right iris images.

effective manner. In January 2020, it was proposed to use BAMHealthCloud for secure access to e-medical data [36]. There is also evidence that access control to documents requiring security clearance for defense, healthcare, and financial records can be conveniently provided using biometric authentication service [35]. BaaS can provide benefits for biometric user authentication such as a ubiquitous service, virtually unlimited hardware, and storage resources at an economical price [2] [22]. Based on the analysis and prediction of the market demands in the near future, BaaS can be used for biometric identification in law enforcement, forensics, and criminal identification by utilizing public biometric repositories on the cloud [34].

Cloud has already been used to offload biometric authentication in mobile phones [6]. In this work, a distributed Biometric Authentication in Mobile (BAM) was proposed on a cluster cloud. Authors employed a feed-forward neural network with backpropagation but did not address biometric template security on the cloud, nor security during transmission from the mobile. Work [7] also dealt with offloading the biometric processing in mobile to the cloud to tackle enrollment surge. It employed a classical texture based Local Binary Pattern (LBP) model for user recognition unlike our proposed approach which employs deep learning for biometric feature extraction and user verification.

While cloud computing offers many advantages, security and privacy of biometric data on the cloud is of high importance [8]. Two main challenges related to biometrics as a service are the protection of biometric data transmitted to the cloud and protection of biometric data stored on the cloud [34] [12]. In our work, we have addressed both the concerns of biometric data transmission and biometric data protection.

For biometric data protection on the cloud, we propose to use cancelable biometrics to preserve template privacy. Cancelable techniques like random projection have shown to be effective [9] [10], and have also been implemented on a cloud platform [11]. In this project, a combination of random cross folding and random projection has been employed to incorporate cancelability. For biometric data protection during transmission the cloud, a new technique of steganography by embedding the cross folded biometric image in a QR code has been proposed. Steganography has proved to be an effective technique to hide data during transmission, without making it look obvious like cryptography [13] [14]. In this work, the multi-instance biometrics concealed in a QR code is transmitted over a standard secure transmission protocol (SSH) to the cloud. At the cloud server, the QR code is decoded to obtain the crossfolded biometric image. The QR also acts as an OTP to prevent replay attacks.

Our proposed biometric module on the cloud operates on deep learning techniques such as CNN and MLP for feature extraction and user authentication, respectively. Therefore, a cancelable biometric system has been deployed on the cloud which can parallelize deep learning, provide high accuracy, cut costs, improve authentication speed and at the same time provide security to biometric data.

III. PROPOSED METHODOLOGY

The proposed method is a client-server model of a cloud-based cancelable biometric system. The biometric module and cancelable database are hosted on the cloud server, while the user or client site is where the biometric sensor is deployed. A web API connects the client to the cloud. Figure 1 illustrates overall architecture of the proposed cloud CB system. The framework is divided into five phases: *a) Data capture and cross folding, b) Steganography, c) Feature*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3003869, IEEE Access

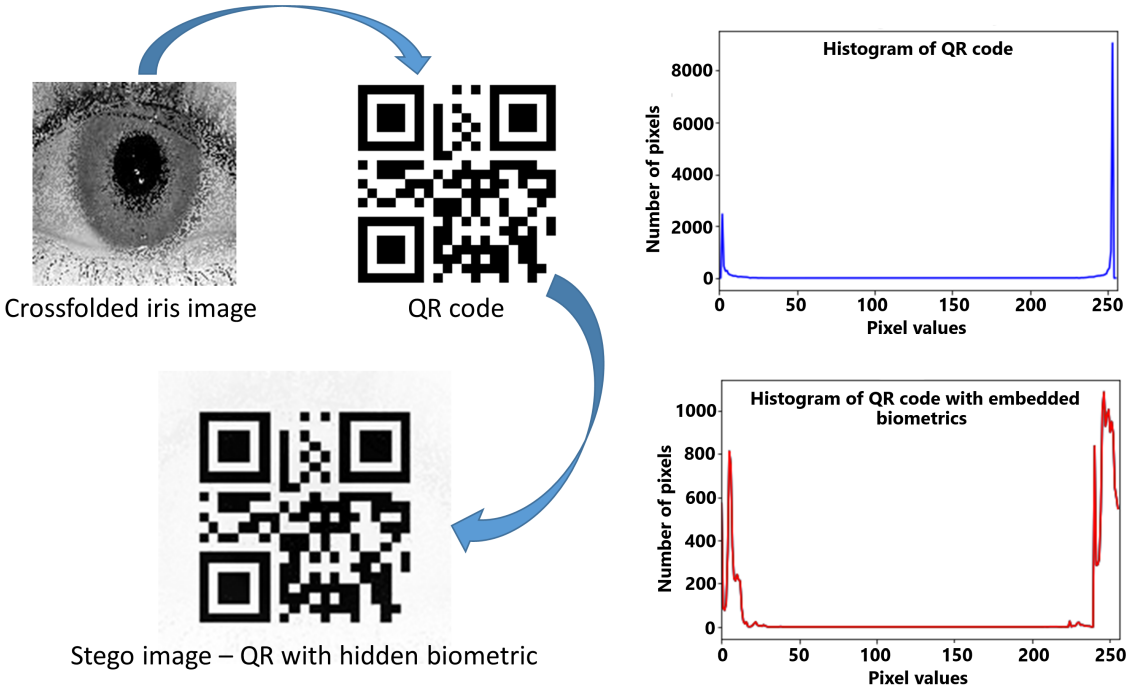


FIGURE 3: Proposed steganographic hiding of the crossfolded biometric image in a QR code cover image.

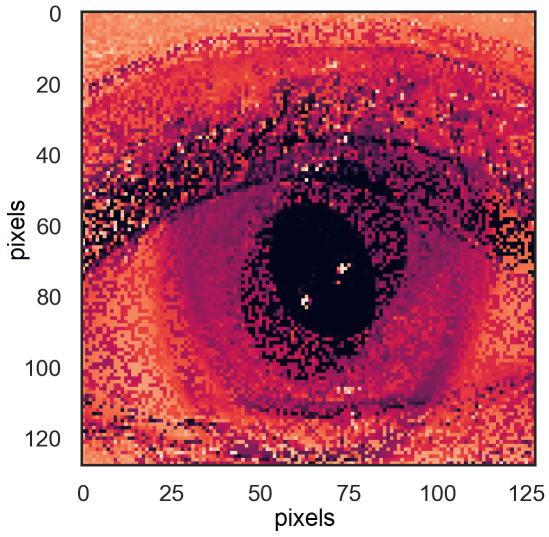


FIGURE 4: Sample of a crossfolded biometric image generated from multi-instance (left and right) iris images.

extraction through deep learning on the cloud, d) Cancelable template generation, and e) User verification through MLP.

A. DATA CAPTURE AND CROSS FOLDING (PHASE 1)

For this work, multi-instance biometrics of left and right irises, and middle and index fingerveins have been chosen. A new cross folding technique adapted from [10] using a randomly generated matrix G obtained from user key K has been utilized. The generated matrix G is converted into a binary matrix B of 128x128 size with cells being assigned a value of either 1 or 0, based on a threshold value α. The threshold α is calculated from the random matrix itself by averaging out matrix values to generate a mean value. If the value of the cell of the generated random matrix is greater than the threshold, it is assigned a value of 1, else 0. In this way, matrix G is converted into a binary matrix B. Subsequently, matrix B' or the one's complement of matrix B is generated. To obtain the cross fold, left iris pixel matrix Pl is multiplied to B to generate a binary left matrix $T11$. Similarly, right iris pixel matrix Pr is multiplied to B' to generate a binary right matrix $T12$. $T11$ and $T12$ are then fused to obtain the crossfolded iris/fingervein image CF as shown in the Figure 2. Figure 4 depicts a sample crossfolded iris image generated from multi-instance (left and right) iris images.

B. STEGO IMAGE AND TRANSMISSION TO THE CLOUD (PHASE 2)

In Phase 2, a novel technique of embedding the cross folded iris/fingervein image into a QR code via image in image steganography has been proposed. The QR code has been generated using an open source BSD license Python module - qrcode 6.1 [37]. The standard numeric encoding scheme is used to encode the random number sequence generating the QR code. Next, a Reed Solomon error correction parameter is set so that the QR code reader can discern if data has been read properly and corrects the noise. Third, the default parameters for version, box size, and border parameters are chosen to generate a QR code of 128x128px (the same as

that of crossfolded biometric images). Finally, default mask option and UTF-8 standard encoding are chosen. Python's QR code and OpenCV modules have been previously used in design of a security system for data hiding [38] [39].

In our work, we have proposed to embed the multi-instance crossfolded biometric template inside the QR code using steganography. Steganography provides safe and unidentifiable transmission of the biometrics to the cloud. The proposed crossfolding technique is resistant to template inversion and spoofing attacks owing to the complexity of the folded template. The proposed multi-instance biometric steganography inside a QR code increases biometric template protection by concealment. Also, the QR code performs another important function. It acts as a One Time Password (OTP) that is newly generated by the cloud server every time the crossfolded biometric needs to be transmitted. The QR code is then verified at the server end. This generation of a new QR for each transmission reduces the chance of replay attacks. Proposed procedure for crossfold biometric embedding in QR codes using steganography is as follows:

- After generation of the cross folded image in Phase 1, the client application requests access to the cloud server, and subsequently receives a QR code as a One Time Password (OTP).
- Image steganography is carried out by hiding the cross folded template in the QR code (cover image).
- The biometric image is hidden by substituting the least four significant bits of each pixel of the cover by the four most significant bits of the corresponding pixel in the biometric image. This creates a minimum deficiency in the cover that cannot be perceived by the human eye.
- The stego image is sent to the cloud via a Secure Shell (SSH) File Transfer Protocol (SFTP). At the cloud end, the cross folded biometric image is decoded from the stego image and OTP is verified.

Figure 3 shows the procedure of hiding a crossfolded iris image inside a QR code. The final QR image resembles the original QR code. However, it actually contains the biometric hidden inside it. This is possible by least significant bit substitutions as explained in step 2, making it unperceivable to the human eye. The different histograms of both images prove that they are actually different, even though they look similar as in Figure 3. This combination of steganography and QR embedding over SSH makes the proposed biometric transfer to the cloud more secure.

C. FEATURE EXTRACTION BY DEEP LEARNING ON THE CLOUD (PHASE 3)

Proposed cloud framework: To deploy the deep learning based biometric engine to the cloud, we utilize Amazon Web Service (AWS) cloud platform due to its popular and widely used deep learning environments. The system is deployed on Amazon Elastic Cloud compute EC2 G3 instance, which offers 4 NVIDIA Tesla M60 GPUs, each containing 8 GB of memory and 2048 parallel processing cores. The instance also comes with 64 virtual CPUs and a network bandwidth of 14 Gbps. AWS provides 488 GB of main memory for this instance. The 'elastic' nature of EC2 allows developers to easily scale up or down, depending on data traffic. The goal here is to parallelize deep learning using CNNs using the GPUs. To train the neural net, we propose to use a batch size of 128, or a batch of 32 samples for each GPU for data parallelism on the cloud. Data parallelism refers to each GPU training the whole model with its specific portion of the dataset. We first deploy the proposed model on a standalone system and then parallelize the model on the cloud to compare performance, cost, and time efficiency. The standalone system consists of Intel core i7 processor-8th Gen powered by a NVIDIA Cuda GPU.

CNN model: Feature extraction is a crucial part of a biometric system that affects its accuracy. Proposed feature extraction is performed using a Convolutional Neural Network on the cloud. The four major steps are: *a) Pixelating, b) Normalization, c) CNN training and testing, and d) Extraction of features from the dense layer of the CNN*. In the first stage, grey scale crossfolded iris images (128x128px) are pixelated. Gray-scale normalization is then carried out to reduce the effect of illumination differences by changing the range of pixel intensity values. The new range between 0 and 1 improves the speed of convergence of the neural network. A schematic representation of the CNN which is an adapted version of [16] is shown in Figure 5.

To train and test the CNN, data is split into 5 equal segments for fivefold cross validation. All class labels are encoded to one hot vectors, eg. class 2 is represented as [0,0,1,0,..]. Next, a random seed is initialized to pick up random data for training and testing. Conv1 and Conv2 consists of 2 sets of layers each with 32 filters and 64 filters. A kernel filter is parsed over the entire image transforming it. Maxpool is subsequently used to reduce dimensions/ downsample the output of Conv layer. Maxpool prevents overfitting and reduces computation cost. The Conv and Maxpool together help to decipher local features and combine them to form more global features. Figure 6 is a representation of crossfolded iris images after passing through filters of the CNN at layer1, layer3, and layer 6. The dropout layer is used to randomly ignore certain weights forcing the CNN to learn in a distributed way. For this model, we propose to use ReLU (Rectified Linear Unit) activation function as a rectifier to add non-linearity to the network. ReLU is chosen due to important advantages over activation functions like the Hyperbolic Tangent (TanH) and Sigmoid. ReLU is known to minimize the vanishing gradient problem present in TanH and Sigmoid functions [31], which prevents network learning due to the occurence of a vanishingly small gradient. ReLU is also known to converge to the global minima faster than TanH and Sigmoid, and is computationally more efficient to compute as it does not perform exponential operations [31]. Finally, fully connected layers like Flatten and Dense are used, after which SoftMax activation generates the probability of each class. After the model construction, a score

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3003869, IEEE Access

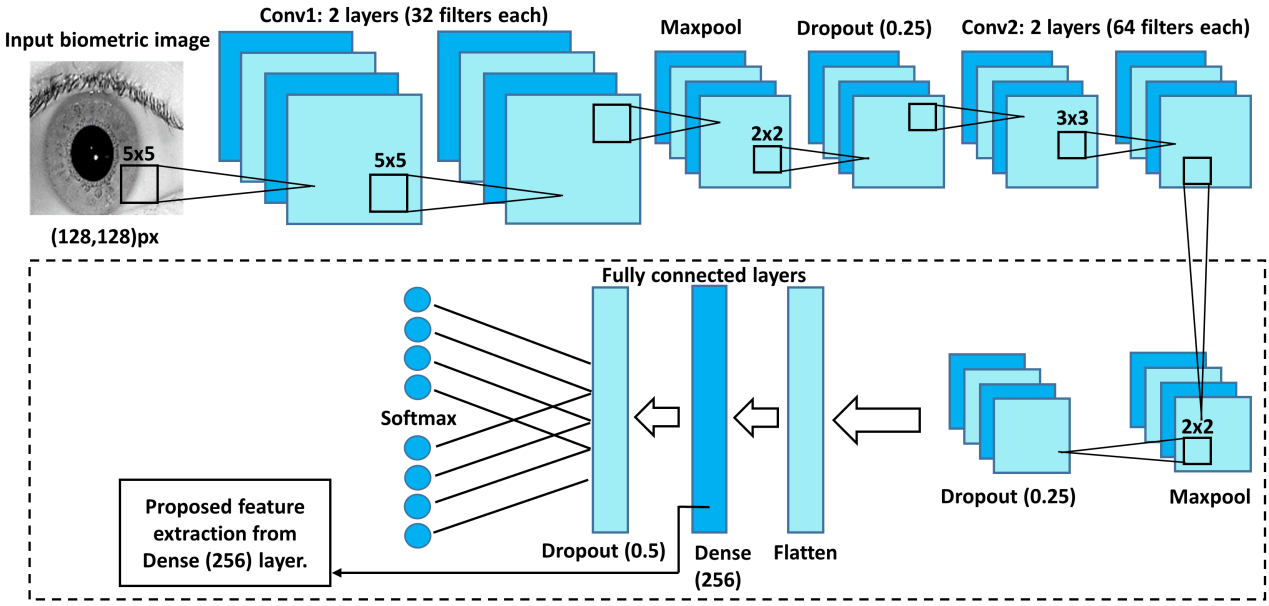
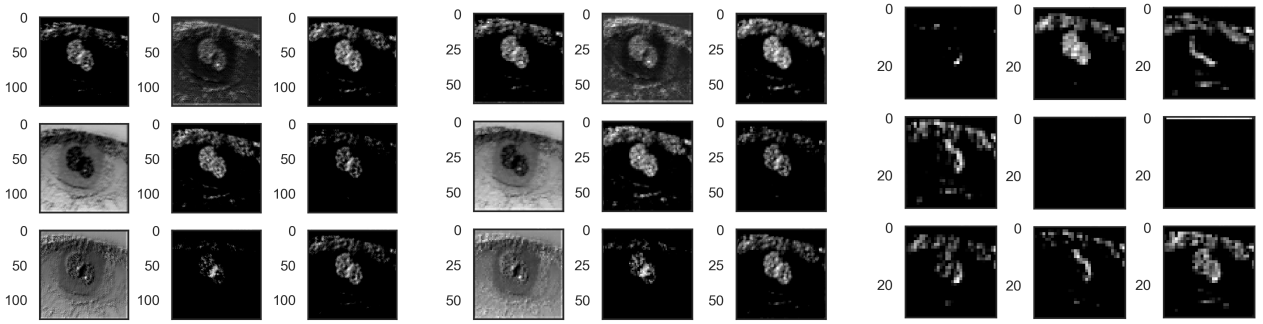


FIGURE 5: CNN for biometric feature extraction: a novel, compact, 6-layer CNN is proposed with adaptations inside the box.



(a) Layer 1 **(b)**Layer 3 **(c)**Layer 6

FIGURE 6: Activations of crossfolded iris images through various layers of the CNN.

function, loss function and optimization algorithm are set up. The loss function measures how well the model performs using known labels. The loss function used is categorical cross entropy [17] defined as (1):

$$-\sum_{c=1}^{M} y_{o,c} log(p_{o,c}) \tag{1}$$

where:

M - number of classes,

y - binary indicator (0 or 1) if class label c is the correct classification for observation o, and

p - predicted probability observation o is of class c.

The model's initial kernel values, weights and bias are set by an optimizer to minimize loss. In our research, RMSprop with default values has been chosen which monotonically decreases the CNN's learning rate. It is faster than stochastic gradient descent and less aggressive than the Adagrad method [32]. It also resolves Adagrad's radically diminishing learning rate problem [32]. An annealer is then used to help the optimizer converge faster to the global minimum of the loss function. When the learning rate *lr* is high, the optimizer may get stuck in a local minimum, which is undesirable. Thus, the *lr* has to be decreased in a slow and steady manner to reach the global minimum. We have chosen to reduce *lr* by half, if accuracy is not improved after 3 epochs, and set minimum *lr* to 0.00001. This type of adaptive learning is implemented by the ReduceLROnPlateau function on the RMSprop. The last step is fitting the model and calculating accuracy and loss after each epoch which is a forward and backward pass over an entire dataset through the neural network once. The CNN model is stored as a Keras model in a HDF5 file. This returns a model identical to the previous one (with same weights and biases) every time the CNN runs. The biometric features are then extracted from the final dense layer along with labels and stored as (1 x 256) vectors. Due to the reduced dimensional size of features from 16384 to 256 by the CNN, there is no need to use methods like Principle Component Analysis for dimension reduction,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3003869, IEEE Access

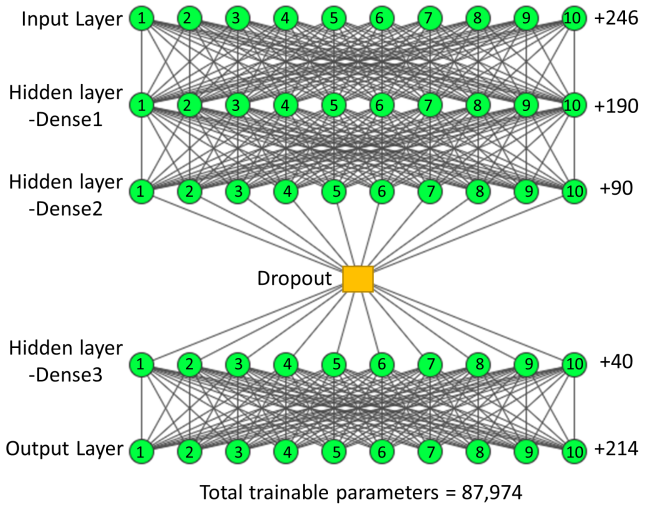


FIGURE 7: Proposed MLP model.

which is an added advantage. The deep learning model has been implemented on standalone and cloud architectures to compare accuracy and response times.

D. CANCELABLE TEMPLATE GENERATION (PHASE 4)

We propose to utilize cancelable biometrics to ensure privacy and security of biometric templates on the cloud. Biometric features extracted from the CNN are transformed into cancelable templates, through the use of random projection. According to the Johnson and Lindenstrauss lemma [15], if points in a vector space are of sufficiently high dimension, then they may be projected into a lower-dimensional space in a way which approximately preserves the distances between the points. The objective of RP is to maintain euclidean distance before and after projection, preserving statistical properties. It has proved to be effective in works [9] [10] [5] for cancelable template generation. The proposed technique comprises of two steps: *a) Generation of an orthogonal matrix (O) from userkey (K) and b) Multiplication of crossfolded feature matrix (CFm) and (O) to form the final cancelable database (CB = CB1, CB2, .. CBn) stored on the cloud.* The original biometrics remain safe since only the cancelable templates are stored and used for subsequent user authentication.

E. USER AUTHENTICATION (PHASE 5)

The user authentication module is deployed on the cloud platform, for which we propose a novel MLP model. The motive behind utilizing MLP, stems from its efficiency in classification, regression, and mapping functions [26]. This is due to its ability to distinguish data that is not linearly separable [26]. The non-linear mapping property of mapping an *N*-dimensional input signal to an *M*-dimensional output signal, makes it a suitable matching module in this work.

MLP model: Multi Layer Perceptron (MLP) is a type of feed-forward neural network. Similar to the CNN, it uses a supervised learning approach of backpropagation for training. The proposed MLP consists of three hidden layers consisting of 200, 100, and 50 fully connected nodes. Figure 7 is a visualization of the MLP model. Adam optimizer has been utilized along with the Relu activation function. Learning rate was initialized to 0.001 with momentum 0.9. A comparative study of MLP vs SVM (Support Vector Machine), k-NN (k-Nearest Neighbor), D-Tree (Decision Tree), and Naive Bayes classifiers has been performed to establish its superiority over the classical machine learning models. The performance of MLP on cloud vs stand alone system has also been evaluated.

MLP complexity: For a neural network with t training samples, f features, h hidden layers containing n neurons and o output neurons, the time complexity of backpropagation is calculated as $O(t * f * n^h * o * i)$, where i is the number of iterations [28]. Since proposed MLP has a small feature set of 256 and only 3 hidden layers (200, 100, and 50 neurons), 896 training samples, 224 output neurons for 200 epochs, the time complexity of backpropagation is found to be $O(896 * 256 * 350^3 * 224 * 200)$. This is much less than standard deep neural nets making it computationally very fast.

IV. EXPERIMENTS

A. EXPERIMENTAL SETUP

Iris data was acquired from IIT-D Database (v1) [18], collected by IIT-D Biometrics Research Lab using JIRIS JPC1000 in 2007. 176 males and 48 females in the 14-55 age bracket were studied. Each image (320x240 pixels, 225kb) was resized to 128x128 pixels of 17kb each. MMU was another iris dataset of 45 individuals obtained from Multimedia University [29]. Multi-instance middle and index fingervein images were acquired from FV-USM Dataset [30]. The infrared fingervein dataset consists of samples from 123 individuals (83 males and 40 females) in 22 to 50 age group. Images have an extracted Region Of Interest (ROI) and are rescaled to 128x128 pixels from 640x480 pixels, for better control of image processing through the layers of CNN. Figure 8 exhibits sample iris and fingervein images of the 3 datasets, with plots depicting label count for each sample. Images were pre-processed by grey scaling and pixelating.

B. EXPERIMENTAL RESULTS

We carry out experiments on both standalone and cloud platforms. Metrics of accuracy, loss, time, speed, and quality of neural network training have been compared for both scenarios. Proposed biometric feature extraction using CNN has been compared to other classical feature extraction methods. MLP performance for user verification is evaluated by Genuine Acceptance Rate (GAR), False Acceptance Rate (FAR), Equal Error Rate (EER), and Area Under the Curve (AUC). A comparison of MLP performance against other ML classifiers has also been carried out. The proposed system demonstrates high accuracy, speed, and cost-efficiency.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3003869, IEEE Access

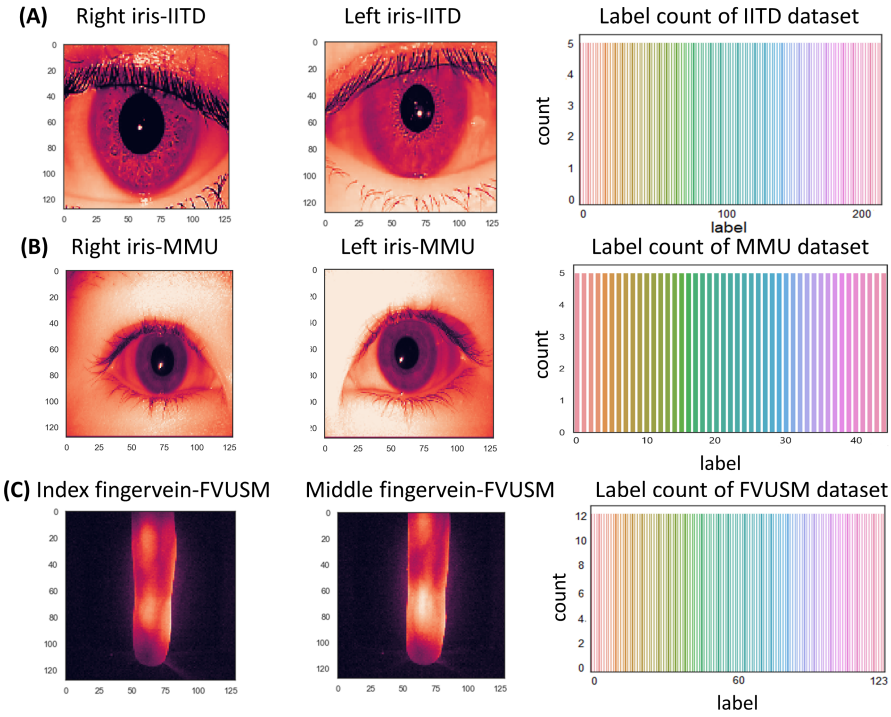
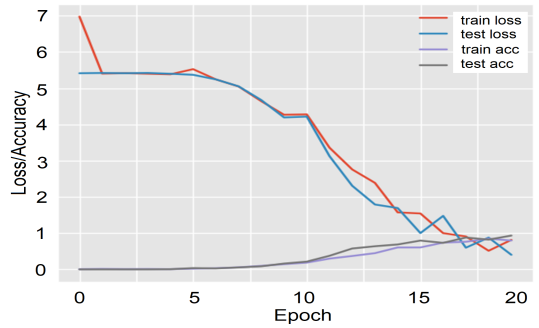
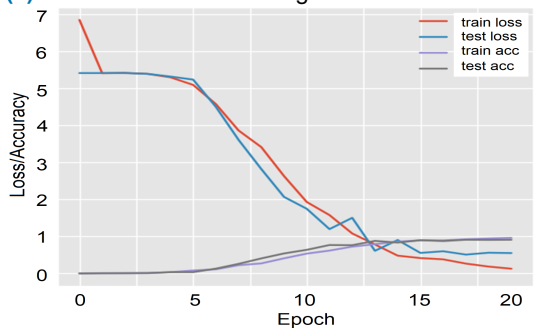


FIGURE 8: Data statistics: Figure 8(A) displays sample right and left iris images from IITD dataset. Figure 8(B) displays sample right and left iris images from MMU dataset. Figure 8(C) displays sample index and middle fingervein images from FVUSM dataset. (Right) Plots depicting labels vs number of samples per label for all the 3 datasets.



(a) Standalone - CNN training and validation curves.



(b) Cloud - CNN training and validation curves.

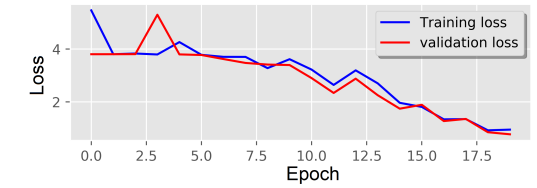
FIGURE 9: Standalone vs cloud architecture comparison of CNN performance on IITD data.

TABLE 1: Comparison of proposed biometric feature extraction using CNN vs other feature extraction methods.

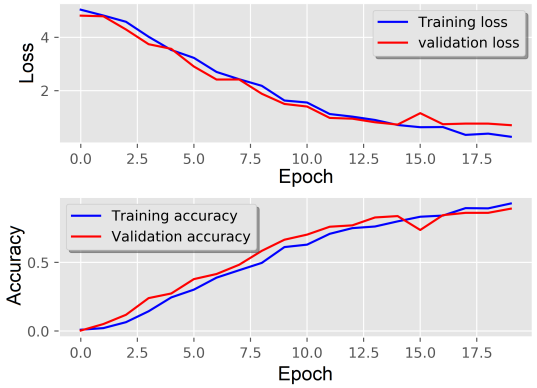
Dataset	Method	EER
IITD	5x5 Blocks of Avg Pixel Intensities	0.62
	Raw Pixel Intensities	0.50
	HOG	0.64
	Log Gabor [18]	0.38
	Proposed method using CNN	**0.12**
MMU	5x5 Blocks of Avg Pixel Intensities [25]	0.86
	Raw Pixel Intensities [25]	0.76
	Gabor + PCA	0.51
	Log Gabor	0.42
	Proposed method using CNN	**0.15**
FV-USM	5x5 Blocks of Avg Pixel Intensities	0.61
	Raw Pixel Intensities	0.59
	Gabor + PCA	0.39
	Log Gabor	0.06
	Proposed method using CNN	**0.05**

Summary of results: Figure 9 depicts the training, test accuracy, and loss curves for the proposed CNN on the IITD database. Figure 9(a) depicts CNN training on a standalone system. Figure 9(b) illustrates training of the CNN implemented on the cloud platform. It can be noted that the CNN achieves training accuracy of 97.3%, and validation accuracy of 90%. However, the CNN model on cloud learns the data faster and smoother starting from the fifth epoch. Cloud architecture also accelerates training time from 48.8s on the standalone system to 23.2s on the cloud (2x speed).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3003869, IEEE Access



(a)MMU iris dataset.



(b)FV-USM fingervein dataset.

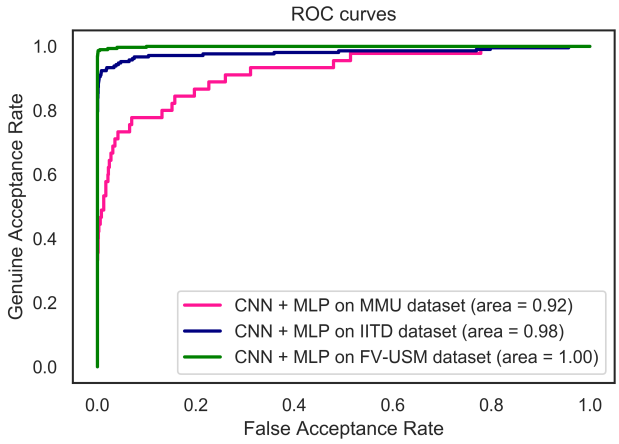
FIGURE 10: CNN training and validation curves on the AWS cloud.

TABLE 2: Comparison of classifiers for user verification.

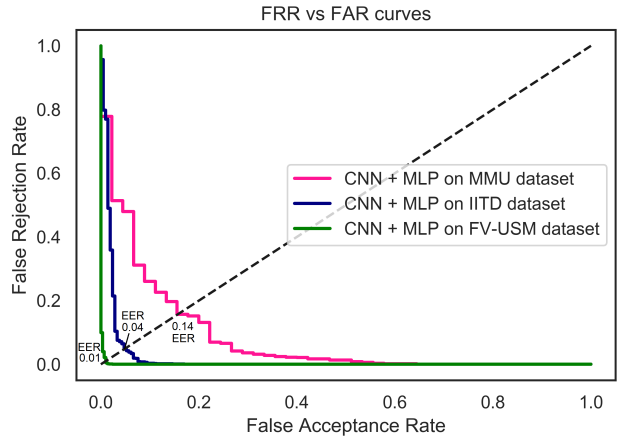
	EER on IITD dataset
Decision Tree	0.35
Gaussian Naive Bayes	0.33
k-Nearest Neighbor	0.15
SVM	0.12
MLP (Proposed)	**0.04**

Proposed CNN on cloud achieves 80% validation accuracy on MMU data, and 92% on FV-USM data (Figure 10). Table 1 demonstrates EER of the proposed CNN feature extraction is lower than classical feature extraction methods.

Biometric features are extracted from the final dense layer of the CNN and transformed to cancelable templates using random projection. CB templates are then stored in the cloud database, and are used for subsequent user authentication. We propose a MLP model for user authentication (matching module) on the cloud. In this work, we have followed a commonly used five-fold cross validation for training and testing. Average of five testing scenarios are reported in this section. Furthermore, additional tests have been performed on FV-USM fingervein dataset that include: ten-fold cross validation, stratified five-fold cross validation, and repeated random train test splits. The MLP obtains a high accuracy of 99.55% using 5-fold cross validation for FV-USM dataset



(a)GAR vs FAR plot on crossfolded cancelable data.



(b)FRR vs FAR and EER plot.

FIGURE 11: (a) Receiver Operating Characteristic (ROC) for MMU, IITD and FV-USM data (b) Detection Error Tradeoff (DET) curves with Equal Error Rate (EER) for three datasets.

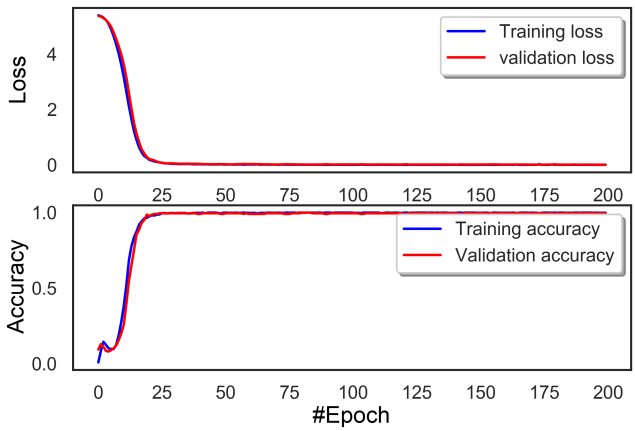


FIGURE 12: Proposed MLP attains 99% validation accuracy.

on both standalone and cloud architectures (Figure 12). Ten-fold cross validation resulted in 99.5% accuracy. Stratified cross validation (preserving percentage of samples for each class) obtained 99.49% accuracy. The repeated random train test split approach (hybrid of traditional train test split and k-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3003869, IEEE Access

TABLE 3: Speed and accuracy on cloud vs stand-alone.

Dataset	Platform	CNN time(s)	MLP(s)	Acc(%)
IITD	Cloud [11]	-	-	95.27
	Stand-alone	31.8	41.38	98
	Cloud (Proposed)	**23.2**	**10**	**98**
MMU	Stand-alone	8.5	6.5	92
	Cloud (Proposed)	**7**	**5**	**92**
FVUSM	Stand-alone	39.38	28.55	99.55
	Cloud (Proposed)	**30**	**13**	**99.55**

fold cross validation) results in 98.87% accuracy. The results demonstrate uniform consistently high accuracy in all testing scenarios.

Figure 11(a) establishes high AUC of 0.92 (MMU dataset), 0.98 (IITD dataset), and 1.00 (FV-USM dataset) by comparing GAR vs FAR. Figure 11(b) compares FRR and FAR values. Low EER of 0.14 (MMU), 0.04 (IITD), and 0.01 (FV-USM) is obtained for the proposed model on the cloud. We compare performance of proposed model to other ML algorithms like SVM, kNN, DTrees, and Naive Bayes (Table 2). MLP obtains lowest EER of 0.04, followed by SVM (EER 0.12) and DTrees (EER 0.35). Table 3 summarizes time and accuracy of standalone vs proposed cloud architecture. Proposed model achieves higher accuracy than recent work [11], where authors obtain 95.27% for IITD data on the cloud. It can be seen that time is significantly reduced while accuracy of biometric identification increased by using the cloud.

C. COST - BENEFIT ANALYSIS

Cloud solutions offer scalable and cost effective solutions, thanks to its pay-as-you-go feature. In this section, a cost-benefit analysis of the proposed system is performed. Two major costs are taken into consideration - a) Hardware costs and b) Processing costs. Two metrics are considered to evaluate the cost efficiency of the system. The first is a cost comparison between different biometric domains. The second metric deals with cost variations as the size of the data to be processed increases. Increase of data size leads to increase of processing requirement. AWS EC2 - Amazon Web Service Elastic Cloud Compute allows to scale VM instances as per processing requirements, and charges on hourly basis.

Total cost for the proposed system is given by C_{Tot} = C_H + C_P, where C_H denotes hardware and implementation costs, and C_P is the data processing cost on the Virtual Machines (VMs) on EC2. It should be noted that AWS charges only when the VMs start processing, no costs are levied during booting of the machines. C_P can be further calculated as C_P = n_I * c_I * t, where n_I is the number of VM instances, c_I is the hourly cost per instance and t is the running time in hours. Amazon provides different VM instances such as T2, M3, C3, G2, G3, P2, P3 with differing processing power and memory capacities. Proposed system makes use of the multi GPU and CPU g3.16xlarge instance. Assuming the proposed recognition system is running 24/7

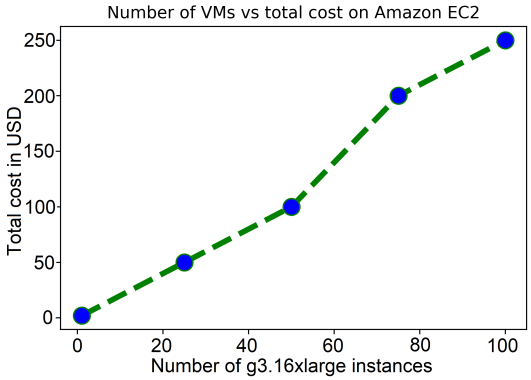


FIGURE 13: Number of VMs vs total cost in USD on AWS EC2.

at the price of 2.152 USD/hour for a reserved instance, the yearly cost would amount to 18,852 USD (C_P Processing cost). If high end iris/ fingervein scanner hardware devices were to be implemented, the total C_{Tot} would amount to approximately 20,000 USD. Table 4 compares the system to already existing biometric solutions in literature. Table 4 establishes that the proposed system incurs the lowest cost. Cloud solutions can reduce processing costs, and thereby provide a much more cost efficient solution. Figure 13 depicts how the cost increases with increase in data and processing. Overhead cost for additional memory usage for every 50 instances has also been taken into consideration. From the graph, it can be inferred that the cost increases linearly as the number of VM instances increase. This is particularly useful for small and medium enterprises, where the system can be scaled up without an exponential rise in cost. Therefore, it can be concluded that implementation of a cancelable multi-instance biometric recognition system on cloud is very cost effective. The following section analyzes the non-invertibility property of the genearted cancelable templates.

D. NON-INVERTIBILITY ANALYSIS

In the proposed system, CB templates are generated as a result of two transformations $\nabla 1$ and $\nabla 2$. $\nabla 1$ refers to the cross folding of left and right iris feature matrices of a user to obtain the biometric crossfold *CF*. This is through the generation of a random matrix *G* of *128x128* based on userkey *K*. This generated matrix is then converted to a binary matrix *B* of dimensions *128x128* based on a threshold value α generated from the matrix itself. $\nabla 2$ refers to the cancelable template generation through random projection. In this transform, an orthogonal matrix *O* is generated from *K*. Extracted features of the cross folded biometric *CF* are multiplied to the orthogonal matrix *O* to form the final cancelable template *CBn = On * CFn*.

An attacker cannot obtain original biometrics possessing both userkey and CB template because:

- Reconstruction of *G* of size *128x128* is nearly impossible due to the possible combinations in the range of $10^1 0^{128*128}$ (ten digit number with repetitions for each

TABLE 4: Cost comparison of different biometric solutions.

Biometric domain	Cost in USD/year	References	Technology
Iris recognition	1,100,000	[19]	RightPatient cloud based iris identification solution
Fingerprint	210,000	[20]	Fingerprint scanner
Facial recognition	126,000	[21]	Dense facial landmark
Iris/ fingervein recognition on cloud (Proposed)	**20,000**	**-**	**AWS cloud, CNN, MLP**

digit *(0-9)* and *16,384* such data cells for a *128x128* matrix). It is therefore difficult to calculate threshold value α which is generated from G itself. Therefore constructing the binary matrix B by applying threshold α to G has $2^{128*128}$ possible combinations.

- Further, reconstruction of the orthogonal matrix O from the CB template is very difficult. This is due to the decimal nature of the matrix. Combinations are in the order of $10^{10^{256}}$ (ten digit number with repetitions for each digit *(0-9)* and *256* such data cells for a *16x16* matrix) making it computationally unfeasible to obtain the same crossfolded feature matrix. Therefore, obtaining right and left iris features would be very hard from the crossfolded feature matrix due to the complexity of the generator matrix as described in the point above. This combination of the two transforms $\nabla 1$ and $\nabla 2$ make it a secure technique with very little risk of template inversion.

V. CONCLUSION

This paper proposes a novel cancelable biometric system based on deep learning, realized on the AWS cloud platform. In this work, the biometric engine, cancelable database, and deep learning module of the cancelable system are offloaded to the cloud. Parallel processing of convolutional neural network implemented on the cloud made it computationally faster than standalone systems, while maintaining accuracy. A novel cross folding and QR code steganography of multi-instance biometrics was utilized to protect biometric data from spoofing and replay attacks during transmission to the cloud. On the cloud, CNN was utilized to extract cross-folded biometric features that were converted to cancelable templates through the utilization of random projection. A novel MLP architecture was proposed for user verification. High accuracy of 99.55% was observed, outperforming other machine learning algorithms. Data parallelization on cloud was found to reduce processing time and cost when compared to other standalone biometric systems. We therefore establish that the proposed cancelable system on the cloud is accurate, secure, time, and cost efficient. Integrating BaaS into existing systems and networks is one of the future research directions.

ACKNOWLEDGMENT

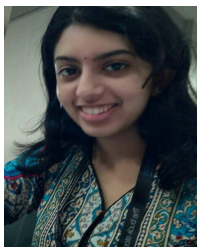
Authors would like to thank National Sciences and Engineering Research Council of Canada for partial support of this research in the form of Discovery Grant and Strategic Partnership Grant.

REFERENCES

[1] Verizon, *2019 Data Breach Investigations Report*, Accessed on: December, 2019 [Online]. Available: https://enterprise.verizon.com/resources/reports/dbir/.
[2] S. Barra, "Biometrics-as-a-service: cloud-based technology, systems, and applications," *IEEE Cloud Computing*, vol. 5, no. 4, pp. 33-37, 2018.
[3] TechSci, *Global Biometrics Market by Type for 2012-2022 Report*, Accessed on: December, 2017 [Online]. Available: https://www.techsciresearch.com/report/global-biometrics-market/1373.html.
[4] A. T. B. Jin and L. M. Hui, "Cancelable Biometrics," *Scholarpedia*, vol. 5, no. 1, pp. 9201, 2010.
[5] P. P. Paul and M. Gavrilova, "Rank level fusion of multimodal cancellable biometrics," *ICCICC*, pp. 80-87, 2014.
[6] F. J. Zareen, K. A. Shakil, M. Alam, S. Jabin, and S. Shakeel, "BAM-Cloud: a cloud based mobile biometric authentication framework," *arXiv: 1601.02781, v2*, 2017.
[7] A. S. Bommagani, M. C. Valenti, and A. Ross, "A framework for secure cloud-empowered mobile biometrics," *MILCOM*, pp. 255-261, 2014.
[8] B. Tapalina, S. Khalid, C. Nabendu, and C. Rituparna, "A survey of security and privacy issues for biometrics based remote authentication in cloud," *CISIM*, pp. 112-121, 2014.
[9] T. Sudhakar and M. Gavrilova, "Multi-instance cancelable biometric system using CNN," *Cyberworlds*, pp. 287-294, 2019.
[10] P. P. Paul and M. Gavrilova, "Multimodal cancelable biometrics", *IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing (ICCI*CC)*, Kyoto, pp. 43-49, 2012.
[11] P. Punithavathi, S. Geetha, and S. Shanmugam, "Cloud-based framework for cancelable biometric system," *CCEM*, pp. 35-38, 2017.
[12] A. Nigam and V. Singh, "A study on data transmission security threats in cloud," *IJIRCCE*, vol. 4, no. 5, pp. 8206-8209, 2016.
[13] K. Bailey and K. Curran, "An evaluation of image based steganography methods," *Multimedia Tools and Applications*, vol. 30, no. 1, pp. 55-88, 2006.
[14] A. Kumar and K. M. Pooja, "Steganography- a data hiding technique," *IJCA*, vol. 9, no. 7, pp. 19-23, 2010.
[15] M. Hassan, "Why is iris recognition biometrics becoming dominant?," Accessed: October, 2019 [Online]. Available: m2sys.com/blog/iris-recognition.
[16] Y. Ghouzam, "Introduction to CNN keras," Accessed: December, 2018 [Online]. Available: http://kaggle.com.
[17] P. Boer, D. Kroese, S. Mannor, and R. Rubinstein,"Tutorial on the cross-entropy method," *Annals of Ops Research*, vol. 134, pp. 19-67, 2005.
[18] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal identification," *CVPRW*, pp. 1-7, 2008.
[19] J. Thomas, "Novant makes 1.1M investment in patient-identification technology," *Charlotte Business Journal 2013*, Accessed: December, 2018 [Online]. Available: bizjournals.com/charlotte/novant-makes-11m-investment.
[20] S. McKenna and J. Sarage, "Biometric analytics cost estimating," *ICEAA*, 2015.
[21] Dense Facial Landmark, Faceplusplus, *Megvii*, 2019.
[22] V. Talreja, T. Ferrett, M. C Valenti, and A. Ross, "Biometrics as a service: A framework to promote innovative biometric recognition in the cloud," *ICCE*, pp. 1-6, 2018.
[23] C. Cachin, "An information-theoretic model for steganography," *2nd Workshop on Information Hiding*, pp. 41-56, 1998.
[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CVPR, arXiv:1409.1556*, 2014.
[25] A. Zeng, "Iris recognition", *princeton.edu*, Accessed: December, 2018 [Online]. Available: http://andyzeng.github.io/irisrecognition.
[26] H. Ramchoun, M. A. J. Idrissi, Y. Ghanou, and M. Ettaouil, "MLP: architecture optimization and training," *IJIMAI*, vol. 4, pp. 26-30, 2016.
[27] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*, New York: Springer-Verlag, 2006.

[28] Pedregosa et al, "Scikit-learn: machine learning in python," *JMLR*, vol. 12, pp. 2825-2830, 2011.
[29] Multimedia-University, *MMU Database*, Accessed: December, 2018 [Online]. Available: http://pesona.mmu.edu.my/ ccteo.
[30] M. Asaari, S. Suandi, and B. Rosdi, "Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics", *ESA journal*, vol. 41, no. 7, pp. 3367-3382, 2014.
[31] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep CNNs", *ACM*, vol. 60, no. 6, pp. 84-90, 2017.
[32] S. Ruder, "An overview of gradient descent optimization algorithms", *arXiv.1609.04747v2*, 2017.
[33] M. M. Monwar and M. Gavrilova, "Markov chain model for multimodal biometric rank fusion", *Signal, Image, and Video Processing*, vol. 7, no. 1, pp. 137-149, 2013.
[34] Markets and Markets, "Biometric-as-a-Service Market by Application Area (Government and Defense, Financial Services, Healthcare, Law Enforcement, and Human Resources), Modality (Unimodal and Multimodal), Offering (Solution and Services), and Region - Global Forecast to 2024", *Report by Markets and Markets*, Available: https://www.researchandmarkets.com/reports/4790748/, ID: 4790748, pages 69, June 2019.
[35] S. Barra, A. Castiglione, M. De Marsico, M. Nappi, and K. R. Choo, "Cloud-Based Biometrics (Biometrics as a Service) for Smart Cities, Nations, and Beyond", In *IEEE Cloud Computing*, vol. 5, no. 5, pp. 92-100, 2018.
[36] Kashish A. Shakil, Farhana J. Zareen, Mansaf Alam, and Suraiya Jabin, "BAMHealthCloud: A biometric authentication and data management system for healthcare data in cloud", *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 1, pp. 57-64, 2020.
[37] Pypi, Pure python QR Code generator QRCode 6.1, Version Jan 2019, Available: https://pypi.org/project/qrcode/, Last accessed: June 3rd, 2020.
[38] Li Fagen, Zhao Qinglan, Yang Shuntong, Zheng Dong, and Qin Baodong, "A QR Code Secret Hiding Scheme against Contrast Analysis Attack for the Internet of Things", *Security and Communication Networks, Hindawi*, vol. 2019, Article ID 8105787, 2019.
[39] N. S. T. Devi and Sateesh Kumar, "Design of Security System for Data Hiding using QR Codes", *International Journal of Scientific Engineering and Technology Research*, vol. 05, no. 01, pp. 130-134, January 2016.

MARINA GAVRILOVA is currently a Full Professor and an Associate Head of Department of Computer Science, University of Calgary. She is an International Expert in the area of biometric security, machine learning, pattern recognition, data analytics, and information fusion. She is a Co-Founder of the Biometric Technologies Laboratory and the SPARCS Laboratory for interdisciplinary computational sciences research. She has published three coauthored books, over 30 books of conference proceedings, and over 200 peer-reviewed articles in machine learning, biometric security, and multimodal cognitive system architectures. Her professional excellence and international stature was recognized by Senior ACM and Senior IEEE membership statuses, as well as the prestigious Canada Foundation for Innovation, the Killam Foundation, and the University of Calgary U Make a Difference Awards. She is the Founding Editor-in-Chief of the Transactions on Computational Sciences (Springer). She serves on the Editorial Board of the IEEE Transactions on Computational Social Sciences, IEEE Access, the Visual Computer, the International Journal of Biometrics, the International Journal of Cognitive Biometrics, and on the IEEE Transactions on Biometrics, Behavior, and Identity Science Steering Committee.



TANUJA SUDHAKAR received her B.Tech degree in Information Technology from the Anna University in 2016. She is currently pursuing the M.Sc. degree in Computer Science at the University of Calgary, Canada, under the supervision of Prof. Marina Gavrilova. She worked as a software developer in Mphasis, a Blackstone Company, from August 2016 to August 2018. Her research interests include cancelable biometrics, biometric security, computer vision, and deep learning.