ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

Hand vein-based biometric authentication with convolutional neural networks and support vector machines

[1]Emile Beukes, [2]Johannes Coetzer

[1]Department of Mathematical Sciences, Stellenbosch University
Stellenbosch, South Africa

[2]Department of Mathematical Sciences, Stellenbosch University
Stellenbosch, South Africa

DOI: https://doi.org/10.5281/zenodo.6961864

Published Date: 04-August-2022

***Abstract*: The viability of utilising a convolutional neural network-based (CNN-based) feature extractor together with asupport vector machine (SVM) for the purpose of identity verification by means of near infra-red (NIR) images of individuals' dorsal hand vein patterns is investigated in this paper. More specifically, this study aims to determine whetherthe utilisation of an SVM, instead of a typical softmax classifier, may lead to an increase in system performance within the context of hand vein-based authentication using CNNs. The proficiency of a variety of novel hand vein-based authentication systems is *first* gauged by employing a softmax classifier, after which the most proficient system is selected, retrained and re-evaluated with a SVM instead of a softmax classifier. The most proficient system, in which the softmax classifier is replaced with a SVM, achieves an accuracy of** 98.90% **and** 99.23% **respectively within the context of the Bosphorus and the Wilches datasets.**

***Keywords:* biometric authentication, hand vein, deep learning, similarity measure network, siamese networks, two-channel networks, segmentation, individual specific, convolutional neural networks, support vector machine.**

I. INTRODUCTION

The need for trustworthy authentication systems has dramatically increased in recent years due to the rise in identity fraud. According to a survey by Shift Processing (2022), payment card fraud cost the worldwide economy anestimated $24.26 billion in 2018 alone and is still climbing. Figure 1 depicts the clear increase in credit card fraud in the USA.

The disadvantages associated with many popular authentication factors being employed today such as personal identification numbers (PINs), passwords and passphrases are that they may be easily stolen, rendering the client vulnerable to identity theft.

The rise in popularity of *biometric* authentication factors may therefore be attributed to the fact that they are generally more resistant to the vulnerabilities associated with other types of authentication factors. Popular biometric authentication factors include handwritten signatures, fingerprints and facial patterns, while less common biometric authentication factors include ear shell contours, retinal

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

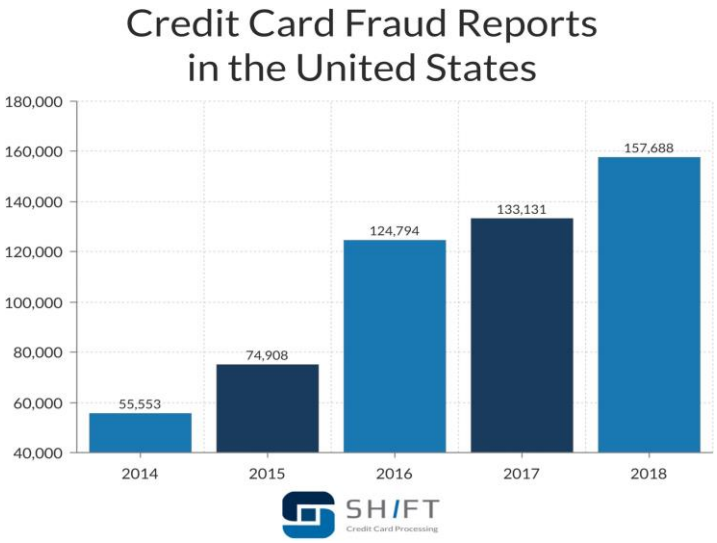

Fig. 1. Credit card fraud cases in the USA (courtesy of Shift Processing (2022)).

and iris patterns, as well as palm prints. One possible disadvantage associated with these factors is that they may *still* be falsified in some way or another, for example, a handwritten signature may be imitated. The acquisition of these authentication factors may also be intrusive and/or time-consuming.

The use of *hand veins* as an authentication factor has therefore gained significant traction due to its *extreme* resistance to forgery. In addition to the aforementioned, the acquisition of hand veins is no more intrusive than the acquisition of, for example, handwritten signatures or PINs. As a result, a number of hand vein-based authentication systems have been recently proposed, of which a relevant few are discussed in Section III.

The objective of the research presented in this paper isto determine whether or not the utilisation of an SVM, instead of a softmax classifier, may enhance system proficiency within the context of hand vein-based authenticationusing CNNs. A number of different combinations of neural network architectures, designs and image preprocessing techniques are investigated. The performance associatedwith each combination is first determined by utilizing a standard softmax classifier, after which a performance *ranking* is established over *all* the different combinations.

The softmax classifier associated with the most proficient system is subsequently replaced with an SVM, after which the amended system is retrained, re-validated and retested. A detailed comparison between the two aforementioned versions of the most proficient system is finally provided and discussed.

This paper is organised as follows. The contribution of this study to the current state of the art (SOTA) is presented in Section II, while related work is discussed in Section III. In Section IV, an overview of relevant network architectures is presented. Section V comprises of detailed descriptions and conceptualisations of the proposed systems. In Section VI, the employed image preprocessing strategies are detailed, and in Section VII, the employed data augmentation strategies are discussed. The experiments conducted in order to gauge the proficiency of the proposed systems are discussed in Section VIII. The results are presented and discussed in Section IX, while avenues for further research are envisioned in Section X.

II. CONTRIBUTION

The contributions of the research presented in this paper to the current SOTA are stated below.

(1) The hand vein-based authentication systems developed in this study are shown to achieve a *greater* proficiency for *both* of the employed databases when utilising an SVM instead of a softmax classifier, which constitutes the first and foremost contribution of the research presented in this paper.

(2) The protocol developed in order to determine the optimal hand vein-based authentication system from a variety of different systems *prior* to the application of an SVM constitutes another contribution to the current SOTA, which can be utilised in *any* field of deep learning.

(3) To the best of the authors' knowledge, two-channel networks have not yet been employed for the purpose of hand vein-based authentication. This constitutes another novel contribution to the fields of biometric authentication and deep learning.

(4) In addition to the aforementioned contribution, neither of the two *similarity measure network* architectures employed in this paper, that is two-channel networks and Siamese networks have, to the best of the authors' knowledge, been investigated within the context of hand-vein patterns acquired from the *dorsal* surface of the hand. Existing publications focus exclusively on hand-vein patterns acquired from the *palmer* surface of the hand.

(5) Furthermore, *binarised* hand vein patterns have not yet been employed within the context of hand vein based authentication via deep learning methods.

(6) An individual-specific approach, that is the scenario where a *separate* network is trained *for each* client enrolled into the system, has to the best of the authors' knowledge not yet been investigated within the context of hand vein-based authentication.

(7) It is demonstrated that the proficiency of the best system proposed in this paper is *comparable* to that of existing SOTA systems. This, in conjunction with the fact that the design of the aforementioned *best* system is fundamentally *different* to that of existing systems, constitute the final contribution of this research. It is therefore entirely plausible that, when the best system developed in this paper is *combined* with any of the existing SOTA systems, the combined system will most likely outperform the SOTA system.

III. LITERATURE STUDY

A number of existing systems that are relevant to the research presented in this paper are briefly discussed in this section.

Sathish et al. (2012) proposed a hand vein authenticationsystem that utilises a number of filtering techniques in conjunction with the 2D wavelet transform for the purpose of emphasising vein information in the input images. The preprocessed images are then subjected to adaptive thresholding in order to acquire a binarised hand vein image. The binarised hand vein samples of the claimed and questioned individuals are subsequently authenticated by utilising aselected optimal Hausdorff distance.

Krishnaveni et al. (2014) developed a hand vein-based authentication system based on minutia extraction and triplet triangulation. The hand vein images from the Bosphorus dataset are preprocessed by means of a unique region of interest (ROI) extraction protocol, followed by histogram equalisation, edge detection and morphological thinning.

An improved Gaussian matched filter is developed by Trabelsi et al. (2013) for the purpose of extracting features associated with hand veins. An artificial neural network (ANN) is subsequently employed for the purpose of feature matching and classification.

Beukes (2018) recently proposed a hand vein-based authentication system in which features are extracted through the calculation of the Radon transform. These features are subsequently matched using an Euclidean distance-based measure, after which authentication is achieved by employing either a score-based or a rank-based threshold.

A system for palm vein authentication called PVSNet is developed by Thapar et al. (2019), and involves two main steps. The first step constitutes an encoder-decoder network that is employed for the purpose of learning generative domain-specific features from a grey-scale ROI that is extracted from the input palm vein image. The second step constitutes an auto-encoded Siamese networkin which the convolutional layers are pre-trained in an unsupervised fashion by employing a triplet loss function. The margin of this loss function is adjusted for learning feature embeddings in such a way that it minimises the distance between embedding-pairs associated with the same individual and maximizes the distance for those associated with different individuals. PVSNet is shown to be effective on typical hand vein datasets.

Two different approaches to dorsal hand vein recognition, both of which utilise CNNs, are proposed by Al-johania and Elrefaei (2019). The first approach involves the utilisation of three popular pre-trained CNNs, namely AlexNet, VGG16 and VGG19, for feature extraction, in conjunction with error-correcting output codes, a SVM and a K-nearest neighbour algorithm for classification. The second approach involves transfer learning with the aforementioned pre-trained CNNs for feature extraction *and* classification. The proficiency of both approaches are gauged within the context of two public hand vein datasets, namely the Dr. Badawi dataset and the Bosphorus dataset.

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

Obayya et al. (2021) proposed a system where an optimal CNN within the context of the employed dataset is empirically determined through Bayesian optimisation by evaluating the proficiency of a number of different CNNs during training. The optimal CNN is employed for feature extraction on a suitable ROI, after which softmax classification is applied. Experimental results indicate that the proposed system outperforms SOTA palm vein authentication approaches at the time of publication.

IV. ARCHITECTURE

In this section, the network architectures employed in this study are discussed. While similarity measure networks (SMNs) are the main focus of this paper, a number of traditional CNNs are also employed in order to establish (1) the feasibility of utilising CNNs for identity *verification* rather than *classification* and (2) whether any proficiency may be gained by employing SMNs, rather than CNNs, for the purpose of hand vein-based authentication.

It is important to keep in mind that the networks proposed in this paper are designed within the context of individual-*specific* verification, where a tailor-made network is trained for *each* individual (client) in question. The motivation behind this approach is that (1) these types of networks may be fully trained in a very short period of time and (2) the enrolment of new clients into the system is seamless, since a new network is trained for each new client. This is in contrast with so-called individual-*independent* networks,in which case a single network is trained for *all* the clients in the system. This introduces the problem of the availability of a hand vein dataset containing a *sufficient* number of individuals before *any* training can occur. The aforementioned statement is verified through preliminary experiments conducted during the course of this study within the context of an individual-*independent* approach, where results indicate a substantial lack in the ability ofthe proposed systems to generalise across different sets of individuals. These experiments were conducted on the same two datasets employed in this paper, in which a total of *only* 100 individuals are available respectively.

A. Convolutional neural networks

Prior to the invention of CNNs, manually engineered features were extracted from images before being presented to the employed verifier. While manually engineered features and traditional image classification models can be superior to CNNs in certain scenarios (Hasan et al. (2019)), CNNs have the unique ability to *automatically* extract optimal features from images which are particularly well-suited for the CNN being employed. This is the principal reason for the rise in popularity of CNNs within the context of image classification.

The architecture of a typical CNN is illustrated in Figure 2, in which the CNN component comprises of a series of *convolutional* and *pooling* layers.

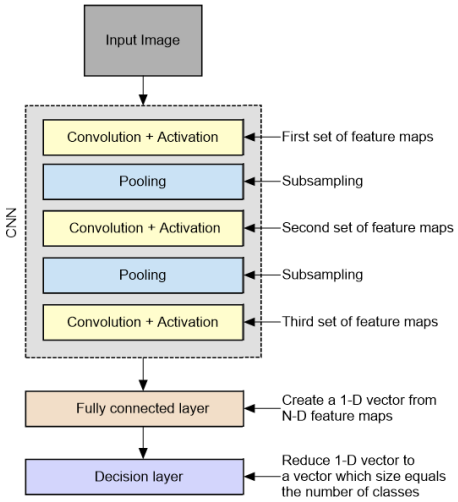

Fig. 2. The architecture of a typical CNN.

The purpose of the convolutional layers is to identify suitable features in the input image. This is achieved by convolving the input image with a kernel (or filter) that con-sists of a set of weights. Significant information is amplified by the filter, while background or insignificant information is gradually reduced over each successive feature map. An activation layer is usually applied, such as *ReLU*-activation,in order to avoid negative values in the feature maps.

The purpose of the pooling layers is to reduce the dimensionality of the feature maps by, for example, taking the largest value in each feature map (*max pooling*) or by taking the average of the values in each feature map (*average pooling*). This facilitates the retention of significant features while discarding irrelevant (background) information.

The purpose of the fully connected layer is to convert the multidimensional feature maps associated with the final convolutional layer into a one-dimensional vector, while the decision layer compresses the information in the aforementioned vector to a vector with a dimension that corresponds to the number of classes in the network. A *softmax* operation is typically applied to the output of the decision layer, so that each entry in the resulting vector may be considered to be the probability of the input image belonging to the class in question.

The training of a CNN therefore comprises of a number of iterations (*epochs*), where after each iteration, the weights of the layers are updated based on the difference between the true labels of the input images and the predicted probabilities in the output of the decision layers. In order to calculate this difference, a *loss function* is employed.

The goal of a CNN is therefore to gradually adjust the weights in the network in such a way that the predicted probabilities are as close as possible to the true labels.

B. Similarity measure networks

A similarity measure network invariably contains at least one CNN within its architecture. Two fundamental differences between a traditional CNN and a SMN is the way in which (1) input images are presented and (2) the CNNs are structured within the particular SMN.

The purpose of a SMN is to establish the *similarity* between two inputs, and is therefore typically associated with a *verification* problem, rather than *classification* problem. There are two classes within the context of a verification problem, typically referred to as the *positive* and *negative* class, where the positive class is associated with samples containing the object in question, while the negative classis associated with samples containing objects that are in some way different from the object in question.

The two SMN architectures employed in this study constitute a two-channel (2CH) network and a Siamese network. An overview of the aforementioned architectures are presented in Section IV-B1 and Section IV-B2 respectively.

1) Two-channel networks: The input for a 2CH network constitutes two greyscale (one-channel) images that are *stacked* in a depth-wise fashion, which produces a two- channel image. It is important to note that these two images (reference image and questioned image) are not *explicitly* compared within the network. Instead, the network produces a feature embedding, in which the difference between these images is *implicitly* contained. Two-channel networks are therefore often referred to as embedding networks. Figure 3 depicts the architecture of a typical 2CHnetwork.

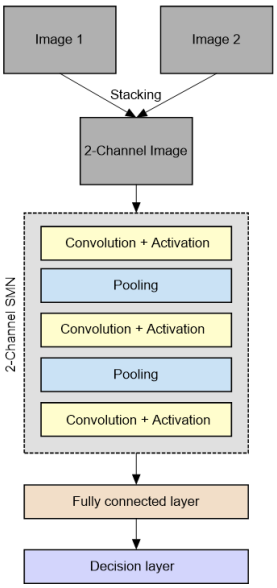

Fig. 3. The architecture of a typical 2CH network (adopted from Beukes and Coetzer (2020)).

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

2) Siamese networks: Siamese networks have two separate *branches* of CNNs into which each of the two images being compared is fed. Each branch subsequently yields a linear feature vector for the respective images, after which a difference vector is calculated between these feature vectors before being fed into the final decision layer. The architecture of a typical Siamese network is depicted in Figure 4.

It may therefore be said that Siamese networks learn the similarity between two images in an *explicit* fashion, which is in contrast with 2CH networks.

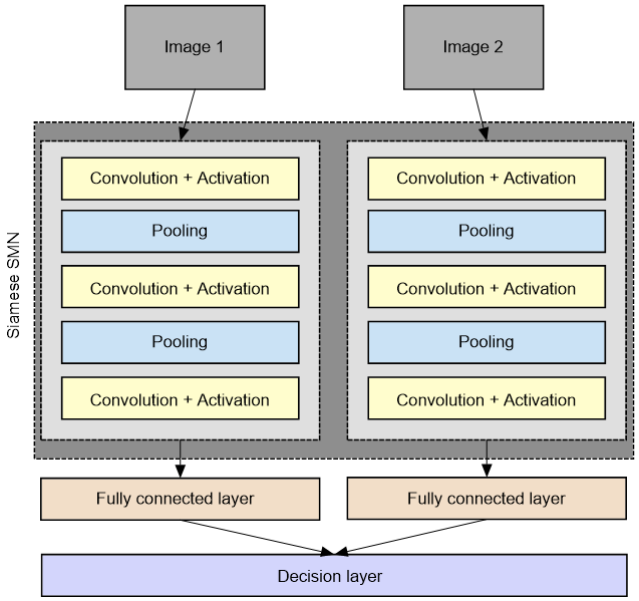

Fig. 4. The architecture of a typical Siamese network (adapted from Beukes and Coetzer (2020)).

C. Support vector machines

It is well known in the current literature that SVMs may perform exceptionally well within the context of many classification tasks. It is therefore conceivable that SVMs may be able to achieve a greater proficiency than that of a softmax classifier which is typically used within the context of CNN-based classification tasks. The reason for this is that an SVM estimates an *optimal* hyperplane which *maximally* separates the data samples in every class (see Figure 5). Additionally, the utilisation of an exhaustive grid search for the purpose of identifying an optimal set of hyper-parameters for the employed SVM within the context of the training data may *furthermore* enhance the performance of the SVM.

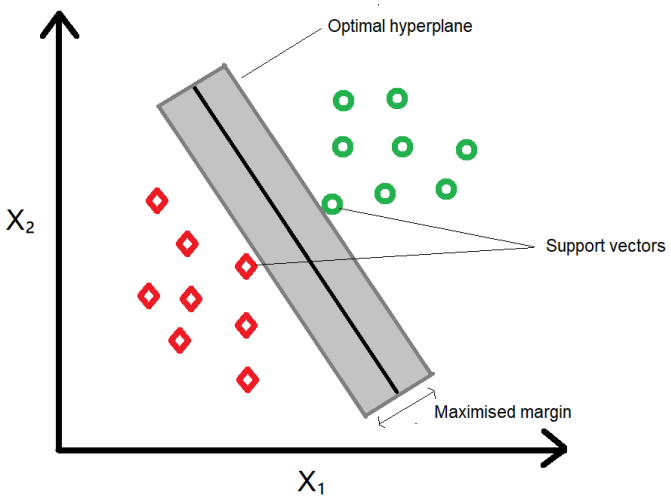

Fig. 5. The architecture of a typical two-class SVM.

This is in contrast to the softmax classification approach,where the probability of a test sample belonging to every available class is predicted. Said sample is subsequently classified to the class associated with the largest predicted probability. A softmax classifier therefore has two disadvantages when compared to a SVM, since (1) it doesn't attempt to maximally separate the training data and (2) the identification of an optimal set of hyper-parameters is not applicable within this context.

V. SYSTEM DESIGN

In this section, detailed descriptions of the networks proposed in this study are provided. These networks constitute a variety of CNN-based feature extractors which are first evaluated together with a given image preprocessing strategy (see Section VI) and a softmax classifier, after which the most proficient feature extractor and preprocessing strategy is retrained and re-evaluated with a SVM instead of a softmax classifier.

A. CNN-based feature extractors

This paper represents a significantly improved and expanded version of the proof of concept originally presented in Beukes and Coetzer (2020), where *only one* 2CH network with a softmax classifier is considered. In this version, the proposed networks comprise of three different neural network types, namely (1) four traditional CNNs, (2) four 2CH networks and (3) four Siamese networks.The differences between the four networks in each of the three aforementioned categories involve the composition and arrangement of the neural network *layers* that make up the CNN component of each of the three neural network types. For brevity, these compositions will henceforth be referred to as (1) *Standard* (ST), (2) *Batch Normalisation* (BN), (3) *Dropout* (DO) and (4) *Batch Normalisation and Dropout* (BD). *Standard* denotes a neural network with a similar composition of layers as the one depicted in Figure 2, *Batch Normalisation* denotes a *Standard* neural network composition with added *batch normalisation* layers, *Dropout* denotes a *Standard* neural network composition with added *dropout* layers and *Batch Normalisation and Dropout* denotes a *Standard* neural network composition with added *batch normalisation and dropout* layers. The utilisation of batch normalisation and dropout layers in a single network is motivated by the findings by Garbin et al.(2020), where it is reported that batch normalisation layers may be used in conjunction with dropout layers in order to improve system proficiency.

The four different CNN compositions within the context of each of the three different neural network types employed in this study are conceptualised in the aforementioned order in Figures 6, 7, 8 and 9, in which the kernel sizes are provided within each layer, while the change in image dimensions, as it progresses through the network, is provided to the right of each layer.

Within the context of the 2CH networks, the kernel sizes of the first convolutional layers in the CNNs conceptualized in Figures 6, 7, 8 and 9 is changed by design from (7 x 8) to (7 x 4), since a two-channel image is provided as input, instead of a one-channel image, as is the case for traditional CNN and Siamese SMNs.

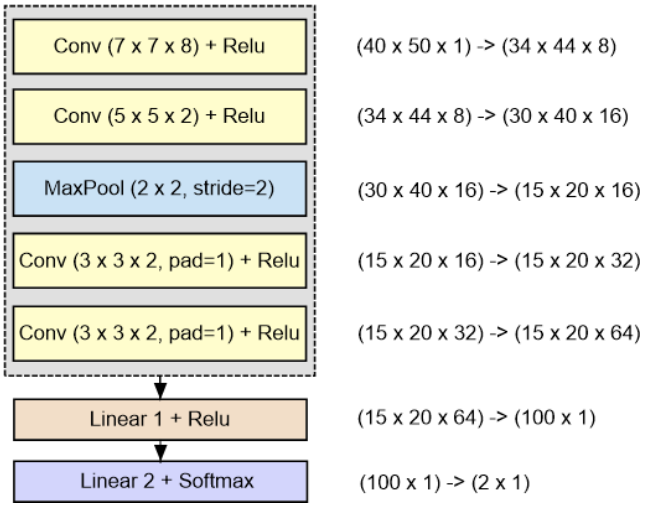

Fig. 6. Conceptualisation of the *Standard* neural network composition(adapted from Beukes and Coetzer (2020)).

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

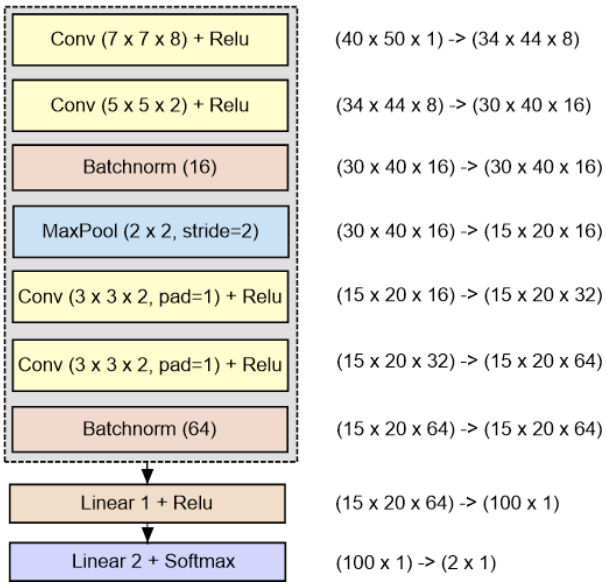

Fig. 7. Conceptualisation of the *Batch Normalisation* neural networkcomposition (adapted from Beukes and Coetzer (2020)).

Recall that the problem addressed in this study constitutes identity *verification*, rather than *classification*, and that a verification system, or verifier, may also be referred to as a *two-class* classifier. The first and second classes of a verifier are typically associated with authentic and imposter samples respectively. Each network architecture in this study is therefore designed to produce a 1D probability vector with *two* elements, in which the first and second entries are associated with the predicted probability of a questioned sample belonging to the client or an imposter respectively.

The optimal probability decision *threshold* for each client is selected during the validation stage, which is detailed in Section VIII-C. A questioned sample is subsequently accepted as authentic when the predicted probability in the *first* entry of the output vector is greater than or equal tothe selected threshold.

The employed loss function for all networks constitutes the mean squared error (MSE) loss, while averaged stochastic gradient descent (ASGD) with a constant learning rate of $1e^{-4}$ is employed as the optimisation algorithm.

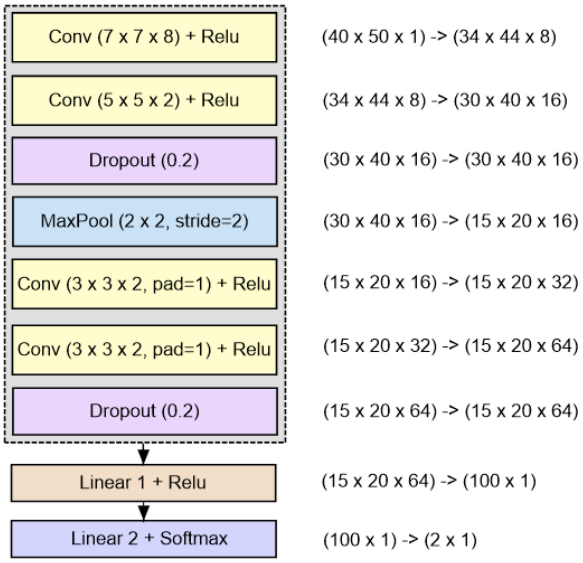

Fig. 8. Conceptualisation of the *Dropout* neural network composition (adapted from Beukes and Coetzer (2020)).

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

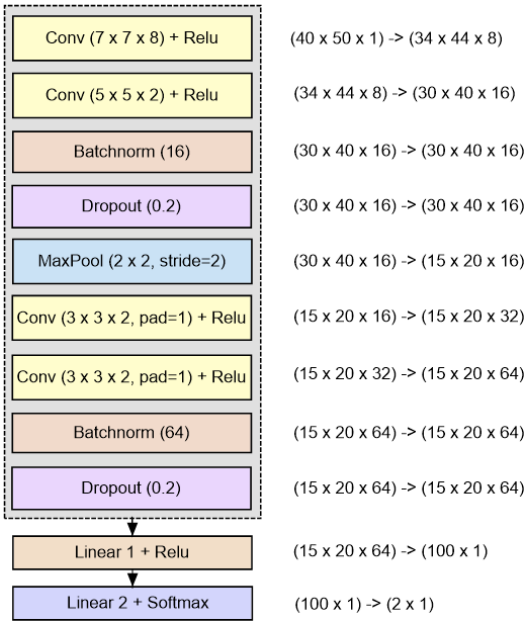

Fig. 9. Conceptualisation of the *Batch Normalisation and Dropout* neural network composition (adapted from Beukes and Coetzer (2020)).

B. SVM layer

In the event that the most proficient system has already been determined by employing a softmax classifier, said classifier is replaced with an SVM as illustrated in Figure 10.

It is important to note that the 2 x 1 output vector of the ”Linear2” layer depicted above does *not* constitute a probability vector, since the softmax activation component is replaced with a ReLU activation function. The output vector from the ReLU layer is therefore *directly* passedinto the appended SVM.

The preprocessing strategies employed in this study for the purpose of extracting a region of interest and enhancing the vein pixels are presented in the next section.

VI. IMAGE PREPROCESSING

The primary database employed in this study is referred to as the Bosphorus Hand Vein Database (Sankur (2011)). The proposed image preprocessing, data augmentation, andexperimental protocols (Sections VI, VII, and VIII) are therefore presented with the focus on the aforementioned

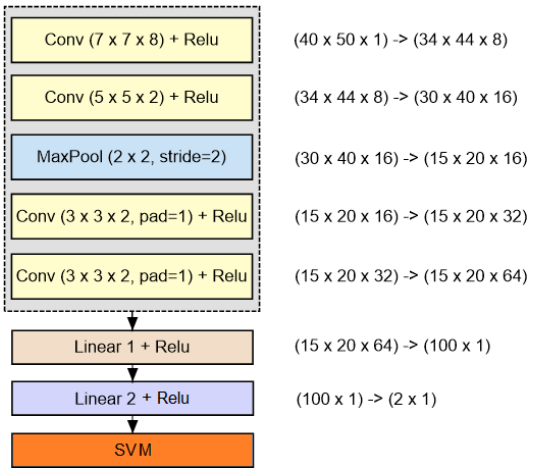

Fig. 10. Conceptualisation of a CNN-based feature extractor together witha SVM.

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

database. The first "DB1" database of the two so-called Wilches databases (Wilches-Bernal et al. (2020)) is also employed in this study as a secondary database for thepurpose of illustrating the stability of the proposed systems. The aforementioned protocols are largely similar for both databases, while the necessary adjustments for the Wilches database will be discussed in each corresponding section.

1) Bosphorus: Three NIR hand vein images were acquired from the dorsal surface of the left hand of each individual under four different environmental conditions. This results in a total of 12 left hand NIR images per individual. This process was repeated for a total of 100 different individuals, which results in a grand total of 1200 available left hand NIR images. In Figure 11, two of these images that belong to different individuals, are depicted.

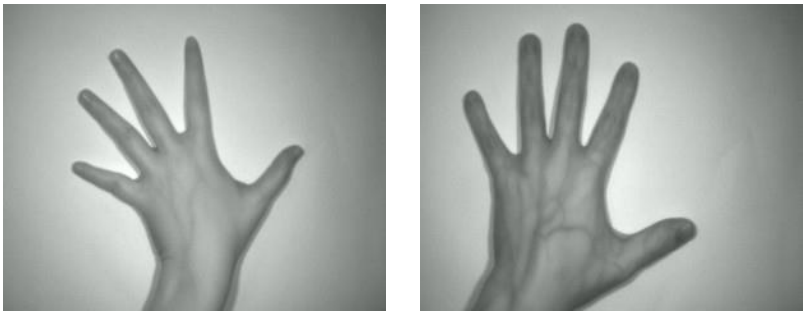

Fig. 11. Two images from the Bosphorus Hand Vein Database.

2) Wilches: This database contains 8 authentic hand vein images for each of 138 individuals, where 4 of these images are of the individual's left hand, while the other 4 are of the individual's right hand. In order to be consistent with the data partitioning and experimental protocol developed for the Bosphorus database, 100 out of the available 138 individuals are randomly selected, while also selecting *only*the 4 left hand images of each individual. In Figure 12, two images belonging to different individuals from the Wilches dataset are depicted.

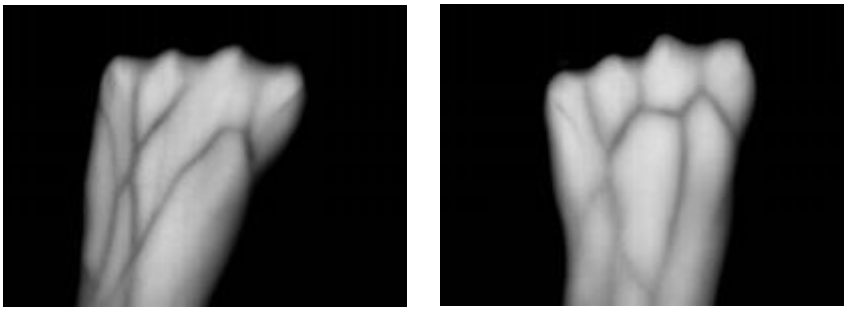

Fig. 12. Two images from the Wilches Hand Vein Database.

A. Extracting the ROI

A suitable region of interest (ROI) is first extracted from the input image, since the majority of visual information in each image is irrelevant within the context of hand vein-based authentication. The extracted ROI is subsequently subjected to three different preprocessing strategies, each of which vary in the extent of contrast enhancement. Each strategy is respectively evaluated within the context of all of the proposed networks. This constitutes a significant expansion when compared to the original proof of concept proposed in Beukes and Coetzer (2020), where *only one* preprocessing strategy is employed in conjunction with *only one* network architecture. The approach for extracting a suitable ROI within the context of the Bosphorus database is extensively detailed in Beukes (2018) and is therefore only summarised in Section VI-A1 for convenience. The ROI extraction protocol for the Wilches database is subsequently detailed in Section VI-A2.

1) Bosphorus: First, the image is binarised through Otsu thresholding in order to identify the pixels associated with the hand. The thumb and four fingers are subsequentlypruned by utilising suitable morphological image processing techniques. The dark area surrounding the hand is then removed, followed by the vertical alignment of the resulting binary object.

Finally, the wrist and the remainder of the thumb is pruned through suitable morphological image processing techniques. The original greyscale hand vein image associated with the resulting object is then also vertically aligned in order to obtain the original greyscale information associated with the resulting binary object. Theresults when the aforementioned process is applied to the images depicted in Figure 11 is shown in Figure 13.

Fig. 13. The initial ROIs of the images depicted in Figure 11.

Since the objects in Figure 13 are not rectangular, the center of mass is calculated for these objects, after which a bounding box of size 40 x 50 around the center of mass is selected, resulting in images of size 40 x 50 with an aspect ratio of 0.8. The negative of the resulting ROI is employed in order to ensure that the networks focus on the hand vein structure, instead of on the background information. The resulting images are depicted in Figure 14.

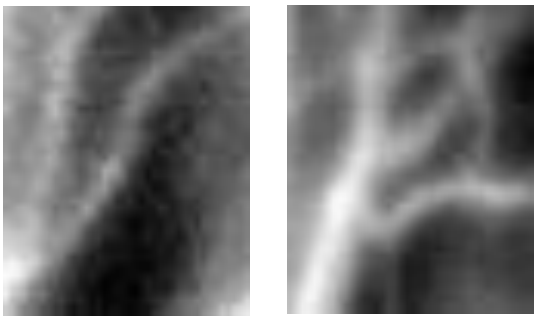

Fig. 14. The final ROIs of the images depicted in Figure 11.

2) Wilches: The image is first binarised through Otsu thresholding in order to identify the pixels associated with the hand. Next, the convex hull of the binary hand is determined, together with its center of mass. The final ROI is subsequently defined as the largest rectangle with the same aspect ratio as the final ROIs of the Bosphorus dataset(see Section VI-A1) that fits entirely inside the convex hull,centered on the center of mass of the convex hull. It is important to note that the actual dimensions of the ROI associated with each image may differ due to a difference in hand size. The resulting ROI is therefore finally resized to 40 x 50 pixels so as to correspond with the shape of the final ROIs within the context of the Bosphorus dataset. The negative of the resulting images when the aforementioned protocol is applied to the images depicted in Figure 12 are depicted in Figure 15.

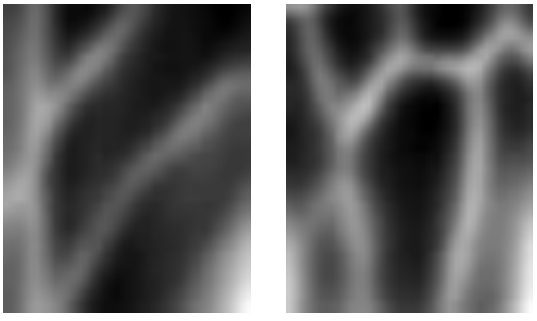

Fig. 15. The final ROIs of the images depicted in Figure 12.

The three preprocessing strategies that are respectively applied to the extracted ROIs are detailed in the next section.

B. Preprocessing the ROI

In this section the three preprocessing strategies respectively applied to the final ROI are discussed. These strategies are *only* discussed within the context of the Bosphorus database, since they are *identical* for both the Bosphorus and Wilches databases.

The objective of considering three preprocessing strategies with different degrees of contrast enhancement is to experimentally establish the optimal preprocessing strategy within the context of each of the proposed networks.

The first strategy is to apply no preprocessing and to present the extracted ROIs such as those depicted in Figure 14 *as is* to the proposed networks. The second strategy constitutes the application of a standard contrast enhancement technique to the extracted ROIs, and is discussed in Section VI-B1. The third strategy is to apply image segmentation to the extracted ROIs in order to obtain binary images, wherein pixels associated with hand veins are rendered white and pixels associated with the areas around the hand veins are rendered black. This segmentation strategy is discussed in Section VI-B2.

1) Contrast enhancement: The proposed contrast enhancement strategy entails the application of the contrast-limited adaptive histogram equalisation (CLAHE) method to the extracted ROIs. This strategy has been shown by Gupta and Kaur (2014) to be effective in enhancing local detail in a digital image. The results of the application of CLAHE with a clip limit of 0.6 and a window size of 10 x 10 on the extracted ROIs depicted in Figure 14 are scaled up for illustration purposes and shown in Figure 16.

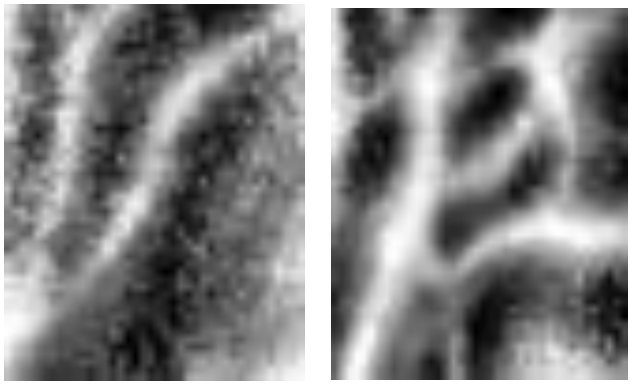

Fig. 16. The results of applying CLAHE to the images depicted inFigure 14.

2) Hand vein segmentation: The segmentation strategy, which is based on morphological reconstruction, is extensively detailed in Beukes (2018) and is therefore only summarised below for the purpose of reference.

In order to obtain the seed image for morphological reconstruction, the extracted ROI is first binarised by two different algorithms, each of which emphasises slightly different hand vein characteristics. These two algorithms constitute the morphological bottom-hat transform and the Laplacian transform. The morphological intersection of the two binarised ROIs is subsequently calculated and constitutes the final seed image that will be used for reconstruction.

In order to obtain the corresponding *mask* image for morphological reconstruction, the extracted ROI is binarised by *moving average thresholding* with a 9 x 9 window, where the local threshold value is equal to the mean of the pixel values in the window.

Once the seed and mask images are obtained, morphological reconstruction is conducted in order to obtain the final binarised hand vein image. Binarised versions of the images depicted in Figure 14 are scaled up for illustration purposes and shown in Figure 17.

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

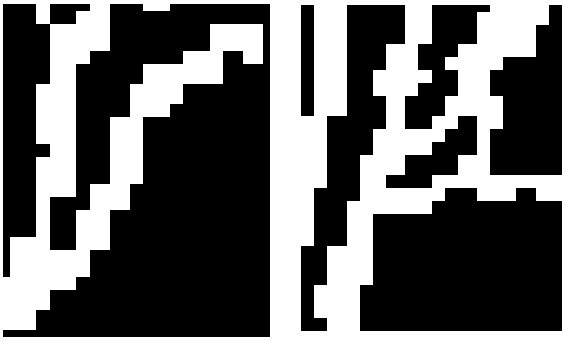

Fig. 17. Binarised versions of the ROIs associated with the images depicted in Figure 11.

VII. DATA AUGMENTATION

Data augmentation refers to the process of applying a set of transformations to each sample in a dataset in order to enlarge the size of the dataset in scenarios where the acquisition of additional *actual* samples is time consuming. This process is commonly applied within the context of deep learning, since the majority of deep learning architectures benefit from larger, as opposed to smaller training datasets. Shorten and Khoshgoftaar (2019) have also demonstrated that data augmentation may increase the generalisation ability of a neural network.

It is important to note that the employed augmentation protocol must simulate the expected types of variation that may occur during the acquisition of actual samples.The data augmentation strategy proposed in this study therefore constitutes image rotation and translation, as well as Gaussian smoothing, which introduces some randomnoise.

It is furthermore important to note that the proposed augmentation strategy depends on the number of samples acquired per client during enrolment.

A. Bosphorus

It is assumed that a total of 8 authentic hand-vein images are acquired per client during enrolment, which is randomly partitioned into 4 authentic training samples and 4 authentic validation samples. The remaining 4 images (from the available 12), are utilised as authentic test images for the purpose of simulation. The augmentation strategy is therefore identical for each set of authentic training, validation and test samples.

1) SMNs: Within the context of the SMNs proposed in this paper, each input image from the Bosphorus dataset, that is either (1) the original ROI (see Figure 14), or (2) the CLAHE contrast-enhanced ROI (see Figure 16), or (3) the binarised ROI (see Figure 17), undergoes the same data augmentation protocol proposed in Beukes and Coetzer (2020):

- 20 unique rotations through an angle θ, where $\theta \in \{-5 + \frac{k}{2} \mid k \in \{0, 1, 2, ..., 20\}\}$, followed by
- 20 random translations in the x and y directions, where $x, y \in \{-2, -1, 0, 1, 2\}$, and
- Gaussian smoothing with $\sigma = 0.4$.

A total of 20 augmented samples are subsequently produced for each image in the dataset within the context of the SMNs proposed in this paper.

2) CNNs: Within the context of the traditional CNNs proposed in this paper, each input image undergoes a more extensive data augmentation protocol, since, as opposed to SMNs, the input images are not paired up within this context. This protocol is outlined as follows:

- 1000 unique rotations through an angle θ, where $\theta \in \{-5 + \frac{k}{100} \mid k \in \{0, 1, 2, ..., 1000\}\}$, followed by
- 1000 random translations in the x and y directions,where $x, y \in \{-2, -1, 0, 1, 2\}$, and
- Gaussian smoothing with $\sigma = 0.4$.

A total of 1000 augmented samples are subsequently produced for each image in the dataset within the context of the traditional CNNs proposed in this paper.

B. Wilches

Due to the fact that only 4 authentic images are available for each individual within the context of the Wilches database, as opposed to 12 within the context of the Bosphorus database, minor adjustments are made to the data augmentation protocol outlined in Section VII-A. These changes are required in order generate the same number of *positive* augmented training, validation and test samples for each individual within the context of *both* databases, which is further discussed in Section VIII-A. The data augmentation protocol remains *unchanged* within the context of the training, validation and test *imposters*.

It is assumed that a total of 3 authentic hand-vein images are acquired per client during enrolment, which is randomly partitioned into 2 authentic training samples and 1 authentic validation sample. The remaining image (from the available 4), is utilised as an authentic test image for the purpose of simulation. Since the number of authentic samples are different within the context of training, validation and testing, the proposed augmentation protocol is adjusted so as to produce the same number of augmented samples as produced within the context of the Bosphorus database.

1) SMNs: Within the context of the SMNs proposed in this paper, each positive authentic *training* image from the Wilches dataset undergoes the following data augmentation protocol, resulting in a total of 80 (2 x 40) positive authentic training images.

- 40 unique rotations through an angle θ, where $\theta \in \{-5 + \frac{k}{4} \mid k \in \{0, 1, 2, ..., 40\}\}$, followed by
- 40 random translations in the x and y directions, where $x, y \in \{-2, -1, 0, 1, 2\}$, and
- Gaussian smoothing with $\sigma = 0.4$.

Each positive authentic *validation* and *test image* undergoes the following data augmentation protocol, resulting in 80 (1 x 80) positive authentic validation and test images:

- 80 unique rotations through an angle θ, where $\theta \in \{-5 + \frac{k}{8} \mid k \in \{0, 1, 2, ..., 80\}\}$, followed by
- 80 random translations in the x and y directions, where $x, y \in \{-2, -1, 0, 1, 2\}$, and
- Gaussian smoothing with $\sigma = 0.4$.

2) CNNs: Finally, within the context of the CNNs proposed in this paper, each positive authentic *training* image from the Wilches dataset undergoes the following data augmentation protocol, resulting in a total of 4000 (2 x 2000) positive authentic training images.

- 2000 unique rotations through an angle θ, where $\theta \in \{-5 + \frac{k}{200} \mid k \in \{0, 1, 2, ..., 200\}\}$, followed by
- 2000 random translations in the x and y directions,where $x, y \in \{-2, -1, 0, 1, 2\}$, and
- Gaussian smoothing with $\sigma = 0.4$.

Within the context of the CNNs proposed in this paper, each positive authentic *validation* and *test image* undergoes the following data augmentation protocol, resulting in 4000 (1 x 4000) positive authentic validation and test images:

- 4000 unique rotations through an angle θ, where $\theta \in \{-5 + \frac{k}{400} \mid k \in \{0, 1, 2, ..., 400\}\}$, followed by
- 4000 random translations in the x and y directions,where $x, y \in \{-2, -1, 0, 1, 2\}$, and
- Gaussian smoothing with $\sigma = 0.4$.

The experiments conducted in order to gauge the proficiency of the proposed networks within the context of a real-world scenario are detailed in the following section.

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

VIII. EXPERIMENTS

A. Data partitioning and cross-validation

The data partitioning and cross-validation protocol employed within the context of the SMNs investigated in this study and the Bosphorus database is identical to the one employed in Beukes and Coetzer (2020), and is briefly outlined below for the purpose of convenience. This protocol will only be detailed within the context of the Bosphorus database, since it is identical for the subset of the Wilches database employed in this study.

The 100 individuals in the employed database are randomly partitioned into 70 clients and 30 *test* imposters.

These test imposters represent criminals for the purpose of simulation, and are therefore typically *not* expected to be existing clients. The aforementioned protocol is repeated three times for the purpose of cross-validation.

Recall that a network is trained and validated for *each* client within the context of this study. A total of 35 training imposters may therefore be randomly selected fromthe other 69 clients, while the remaining 34 clients maybe utilised as validation imposters. This protocol is also repeated three times for each client in question for the purpose of cross-validation.

B. Training

1) Data: First, the data augmentation protocol outlined in Section VII is utilised in order to generate a sufficient number of authentic and imposter training samples. This results in 80 (see Section VII) authentic training and 8 400 (12 x 20 x 35) imposter training samples within the context of the SMNs proposed in this study, and 4 000 (see Section VII) authentic training and 420 000 (12 x 1 000 x 35) imposter training samples within the context of the traditional CNNs proposed in this study.

a) SMNs: The 80 augmented training samples are subsequently paired with (1) each other and (2) the 8 400 augmented imposter training samples by means of the Cartesian product, which respectively produces 6 400 (80 x 80) positive training and 672 000 (80 x 8 400) negative training pairs. It has been experimentally determined that 4 000 training samples (2 000 positive and 2 000 negative samples, ordered in an alternating fashion) is sufficient within the context of the SMNs proposed in this study.

b) Traditional CNNs: In order to produce a balanced training dataset within the context of the traditional CNNs proposed in this study, 2 000 out of the 4 000 augmented authentic training samples are randomly selected, while 2 000 out of the 420 000 augmented imposter training samples are also randomly selected. The 4 000 randomly selected test samples are finally joined and ordered in an alternating fashion.

2) Protocol:

a) CNN-based feature extractors: Recall from Section V that ASGD with a constant learning rate of $1e^{-4}$ constitutes the optimisation algorithm employed in this study. Each network is trained over 100 epochs. All batches contain 32 training *pairs* within the context of the 2CH and Siamese networks proposed in this study, and 32 training *images* within the context of the traditional CNNs proposed in this study. The samples in each batch are ordered in an alternating fashion. According to Section 8.4 in Goodfellow et al. (2016), neural networks should be initialized with caution in order to avoid non-convergence during training. The following alternative solution to this problem employed in Beukes and Coetzer (2020) is also employed in this study: Each network is initialized with random weights, after which training starts. If the training loss decreases insignificantly over the first 10 epochs, the network is reinitialized and training is restarted. It is experimentally determined that a maximum of 10 reinitializations is sufficient within the context of the networks developed in this study.

b) SVMs: In the event that the most proficient system has already been determined through the utilisation of a softmax classifier, said system is amended by replacing the softmax classifier with an SVM (see Figure 10) and retrained in the following fashion: the CNN-based feature extractor associated with the most proficient system is retrained (with the softmax activation layer removed), after which the raw 2 x 1 output vectors produced during the final (100th) epoch is collected together with their class labels for all training samples. These two sets subsequently constitute the training data for the SVM. The protocol for training the SVM is as follows: an exhaustive grid search for the best SVM hyper-parameters given the training data is conducted over a predefined set of feasible parameters. Once the best set of hyper-parameters is identified, the SVM is refitted on the training data using the optimal hyper-parameter set. This system, that is the most proficient system in which the softmax classifier is replaced with a trained, optimally-parameterised SVM, will henceforth be referred to as the "amended most proficient system".

ISSN 2349-7815

International Journal of Recent Research in Electrical and Electronics Engineering (IJRREEE)
Vol. 9, Issue 3, pp: (12-31), Month: July - September 2022, Available at: www.paperpublications.org

C. Validation

1) Data: First, the data augmentation protocol outlined in Section VII is utilised in order to generate a sufficient number of authentic and imposter validation samples. This results in 80 (see Section VII) authentic validation and 8 160 (12 x 20 x 34) imposter validation samples within the context of the SMNs proposed in this study, and 4 000 (see Section VII) authentic validation and 408 000 (12 x 1 000 x 34) imposter validation samples within the context of the traditional CNNs proposed in this study.

a) SMNs: The 80 augmented validation samples are subsequently paired with (1) each other and (2) the 8 160 augmented imposter validation samples by means of the Cartesian product, which respectively produces 6 400 (80 x 80) positive validation and 652 800 (80 x 8 160) negative validation pairs. In order to produce a balanced validation set within the context of the SMNs proposed in this study, a total of 4 000 validation samples (2 000 positive and 2 000 negative samples, ordered in an alternating fashion) are randomly selected from the available positive and negative samples.

b) Traditional CNNs: In order to produce a balanced validation dataset within the context of the traditional CNNs proposed in this study, 2 000 out of the 4 000 augmented authentic validation samples are randomly selected, while 2 000 out of the 408 000 augmented imposter validation samples are also randomly selected. The 4 000 randomly selected validation samples are finally joined and ordered in an alternating fashion.

2) Protocol: Recall from Section V that an optimal probabilistic threshold for accepting a questioned sample is determined during validation. It is therefore important to note that the validation stage within the context of this study is *not* employed in order to prevent overfitting by means of selecting an *early stopping criteria* (Prechelt (1998)), since experimental results show no evidence of overfitting.

A total of four different criteria are employed in this study for the purpose of identifying an optimal probabilistic threshold, and is determined as follows: The proficiency of each system is gauged at each of 50 discrete possible thresholds in the interval [0, 1], which involves the calculation of a false acceptance rate (FAR) and false rejection rate (FRR) at each considered threshold by authenticating samples in the validation set.

The FAR and FRR are defined as $100\,x\,\frac{F^+}{N}$ and $100\,x\frac{F^-}{N}$ respectively, where F^+ and F^- denotes the number of false positives and false negatives, while N and P denotes the total number of negative and positive samples in the employed dataset. The definitions of the four aforementioned criteria, which are based on the FAR, the FRR and the average error rate (AER), are given below, where $AER = \frac{(FAR+FRR)}{2}$

- The equal error rate (EER), that is the threshold where FAR = FRR,
- the *zero* FAR (FAR_{zero}), that is the smallest threshold where FAR = 0,
- the *zero* FRR (FRR_{zero}), that is the largest threshold where FRR = 0 and
- the *minimum* AER (AER_{min}), that is the threshold corresponding to the smallest AER.

In the event that the most proficient system has been identified, the amended most proficient system (see Section VIII-B2) is used to predict the labels of the validation samples.

D. Testing

1) Data: First, the data augmentation protocol outlined in Section VII is utilised in order to generate a sufficient number of authentic and imposter test samples. This results in 80 (see Section VII) authentic test and 7 200 (12 x 20 x 30) imposter test samples within the context of the SMNs proposed in this study, and 4 000 (see Section VII) authentic test and 360 000 (12 x 1 000 x 30) imposter test samples within the context of the traditional CNNs proposed in this study.

a) SMNs: The 80 augmented test samples are subsequently paired with (1) each other and (2) the 7 200 augmented imposter test samples by means of the Cartesian product, which respectively produces 6 400 (80 x 80) positive test and 576 000 (80 x 7 200) negative test pairs. In order to produce a balanced test set within the context of the SMNs proposed in this study, a total of 4 000 test samples (2 000 positive and 2 000 negative samples, ordered in an alternating fashion) are randomly selected from the available positive and negative samples.

b) Traditional CNNs: In order to produce a balanced test set within the context of the traditional CNNs proposed in this study, 2 000 out of the 4 000 augmented authentic test samples are randomly selected, while 2 000 out of the 360 000 augmented imposter test samples are also randomly selected. The 4 000 randomly selected validation samples are finally joined and ordered in an alternating fashion.

2) Protocol: During the test phase, the proficiency of the proposed networks are gauged by calculating the FAR, FRR and AER for the optimal thresholds associated with each of the four criteria discussed in Section VIII-C2. In the event that the most proficient system has been identified, the amended most proficient system (see Section VIII-B2) is used together with the optimal probabilistic threshold (see Section VIII-C2) to predict the labels of the test data.

The experimental results for the most proficient and *amended* most proficient systems are presented and discussed in Section IX.

IX. RESULTS

The experimental results are presented and discussed in this section, where the main focus constitutes an extensive comparison between the performance of the softmax classifier and the SVM within the context of the most proficient system. Note that the most proficient system is determined within the context of the Bosphorus database and amended for *both* databases.

The proficiency of three neural network architectures within the context of hand vein-based biometric authentication is investigated. Experiments are conducted on four different designs, together with three different preprocessing strategies and four different threshold selection criteria for each of the three architectures, resulting in a total of 144 unique system designs. Each experiment is cross-validated a total of 9 times by (1) shuffling the training and validation imposters for each individual three times and (2) employing three different holdout sets of test imposters.The five most proficient systems within the context of the softmax classifier are presented in the following section and briefly discussed.

A. Most proficient systems

In Table I, the results for the five most proficient systems are presented and ranked in ascending order based on the AER. The p-values of Welch's t-test (Welch (1947)) (also known as the unequal variances t-test) between the raw predictions of each consecutive system are rounded to 3 decimals and presented in the final column. For example, the p-value in the first row is associated with Welch's t-testbetween the results of the highest ranked and second highest ranked systems proposed in this paper. For convenience, the following table legend is defined:

- **Neural network architectures**
  - Siam: Siamese networks
- **Network design**
  - ST: Standard
  - BD: Batch Normalisation and Dropout
- **Preprocessing strategies**
  - None: No preprocessing
  - CLAHE: CLAHE-based contrast enhancement
  - Bin: Full binarisation

It is clear from Table I that the Siamese networks generally outperform the 2CH networks and traditional CNNs within the context of the hand vein-based biometric authentication systems investigated in this study. This may be attributed to the design of a Siamese network architecture, which involves the calculation of the element-wise difference between the linear feature vectors associated with the two input images being compared. These networks therefore minimise the loss based on an *explicit* comparison between the two images in question, which is not the case for 2CH networks or traditional CNNs.

The Siamese networks with added batch normalisation and dropout layers are also generally more proficient than the other Siamese networks (see Table I), which indicates that these layers aid in the discriminative ability of the networks.

TABLE I: THE RESULTS FOR THE FIVE MOST PROFICIENT SYSTEMS DEVELOPED IN THIS STUDY, RANKED ACCORDING TO THE AER, AND EVALUATED ON THE BOSPHORUS DATABASE. THE P-VALUES OF WELCH'S T-TEST BETWEEN THE RAW PREDICTIONS OF EACH CONSECUTIVE SYSTEM ARE ROUNDED TO 3 DECIMALS AND PRESENTED IN THE FINAL COLUMN.

System	Criterion	FAR (%)	FRR (%)	AER (%)	p-value
Siam (BD) + CLAHE	AER_{min}	1.01	2.41	1.71	0.000
Siam (BD) + Bin	AER_{min}	1.04	2.50	1.77	0.000
Siam (ST) + Bin	EER	1.56	2.16	1.86	0.751
Siam (BD) + CLAHE	EER	1.52	2.25	1.89	0.007
Siam (BD) + None	AER_{min}	1.19	2.58	2.20	0.067

The majority of systems in which either the CLAHE or the full binarisation preprocessing protocol is employed generally outperform those systems where no preprocessing is applied, which is a clear indication that these preprocessing strategies are useful in emphasising significant image features (see Table I).

If it is assumed that the null hypothesis may be rejected for p-values smaller than 0.05 within the context of Welch's t-test, then the null hypothesis may be rejected for the majority of the systems in Table I. In other words, the majority of differences in AERs between consecutive systems in Table I has not occurred by chance and has in fact occurred due to one system outperforming the next.

Finally, the most proficient system within the context of the softmax classifier constitutes a Siamese network with batch normalisation and dropout layers, together with the CLAHE preprocessing strategy and the AER_{min} criterion for determining an optimal probabilistic threshold. In the next section, the results for the most proficientand *amended* most proficient systems are presented and discussed for both databases employed in this study.

B. Softmax versus SVM

In Table II, the results of the most proficient and *amended* most proficient systems are presented for *both* databases employed in this study. The p-values of Welch's t-test between the raw predictions of these systems are rounded to 3 decimals and presented in the final row for each database respectively.

TABLE II: A COMPARISON BETWEEN THE AERS OF THE MOST PROFICIENT AND AMENDED MOST PROFICIENT SYSTEMS WITHIN THE CONTEXT OF *BOTH* DATABASES.

	Bosphorus	Wilches
Softmax (AER %)	1.71	1.24
SVM (AER %)	1.10	0.77
p-value	0.000	0.001

It is clear from the AERs presented in Table II that the *amended* most proficient system, that is the most proficient system in which the softmax classifier is replaced with an SVM, achieves a *greater* proficiency. The aforementioned statement is substantiated by the results of Welch's t-test in Table II, from which it is clear that the difference in AERs between the two systems has *not* occurred by chance,and that the SVM significantly outperforms the softmax classifier.

C. Comparison with prior work

In Table III, the proficiency of the *best* system developed in this study is placed into perspective by comparing itto existing SOTA systems that were also tested on the Bosphorus hand vein database. It is important to note that this contextualisation should not be interpreted as a *direct* comparison, since the experimental protocol vary significantly from one system to the next.

TABLE III: A COMPARISON BETWEEN THE MOST PROFICIENT SYSTEM DEVELOPED IN THIS STUDY AND EXISTING SOTA SYSTEMS THAT WERE ALSO EVALUATED ON THE BOSPHORUS DATABASE.

Publication	Preprocessing	Network	Accuracy (%)
This paper	ROI extraction and CLAHE	Siam (BD)	98.90
Arora et al. (2019)	ROI extraction, adaptive histogram equalisation and median filtering	Shannon feature transform and improved Hanman classifier	98.50
Krishnaveni et al. (2014)	ROI extraction, histogram equalisation, edge detection and morphological thinning	Minutia extraction, followed by triplet triangulation and matching	99.22
Trabelsi et al. (2013)	Improved Gaussian matched filtering	Artificial neural network	98.00
Al-johania and Elrefaei (2019)	None	Transfer learning with VGG19	99.25

It is however clear that the proficiency of the *best* system proposed in this study is *comparable* to that of existing SOTA systems. This, in conjunction with the *novelty* of the proposed system constitutes a significant contribution to the current state of the art. It is very conceivable that, when the *best* system proposed in this study is *combined* with any existing SOTA system, the resulting hybrid system will outperform the existing SOTA system.

X. FUTURE WORK

In this section, the avenues for possible future work within the context of the research conducted in this paper are envisioned.

First and foremost, the feasibility of the proposed systems should be investigated within the context of an individual-*independent* scenario, that is the case where a *single* network is trained for *all* the clients enrolled into the system, as opposed to the individual-*specific* scenario proposed in this paper. Although the proposed individual-specific approach has many advantages, as outlined in Sections II and III, and constitutes a significant contribution to the current state of the art, such an investigation will provide valuable perspective.

Another avenue for future work that warrants investigation is the use of *unbalanced* validation and/or test sets, in which case the number of positive samples significantly outnumber the number of negative samples. By doing so, the proficiency of the proposed systems may be gauged within the context of a *real-world* scenario, as opposed to the *idealised* scenario investigated in this study where *balanced* validation and test sets are used.

The feasibility of other SMN architectures such as pseudo-Siamese networks, central-surround two-stream networks and spatial pyramid pooling networks proposed by Zagoruyko and Komodakis (2015) may be investigated within the context of hand vein-based biometric authentication.

An investigation into alternative optimisation algorithms such as Adam, Adagrad, Adadelta and RMSprop investigated by Sun et al. (2020) and the use of adaptive learning rates may result in an increase in system proficiency.

Finally, an investigation into alternative classifiers, other than the employed softmax classifier or SVM, may also provide valuable insights, and may lead to an increase in system proficiency.

REFERENCES

[1] N. A. Al-johania and L. A. Elrefaei, "Dorsal hand vein recognition by convolutional neural networks: Feature learning and transfer learning approaches," *International Journal of Intelligent Engineering and Systems*, vol. 12, pp. 178–191, 2019.

[2] P. Arora, S. Srivastava, M. Hanmandlu, and S. Bhargava, "Robust authentication using dorsal hand vein images," *IEEE Intelligent Systems,*, vol. 34, pp. 25–35, 2019.

[3] Beukes, "Hand vein-based biometric authentication with limited training samples," Master's thesis, *Stellenbosch University*,, 2018.

[4] E. Beukes and J. Coetzer, "Hand vein-based biometric authentication using two-channel similarity measure networks," in *2020 International SAUPEC/RobMech/PRASA Conference*, University of Cape Town, Cape Town, South Africa, jan 2020, pp.1–6.

[5] Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning," *Multimedia Tools and Applications,*, vol. 79, p.12777–12815, 2020.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[7] S. Gupta and Y. Kaur, "Review of different local and globalcontrast enhancement techniques for a digital image," *International Journal of Computer Applications,*, vol.100, pp. 18–23, 2014.

[8] H. Hasan, H. Z. M. Shafri, and M. Habshi, "A comparison between support vector machine (SVM) and convolutional neural network (CNN) models for hyperspectral image classification," in *IOP Conference Series: Earth and Environmental Science, Volume 357, SustainableCivil and Construction Engineering Conference,*, Univer-sity Putra Malaysia, Kuala Lumpur, Malaysia, aug 2019.

[9] N. V. Krishnaveni, K. Sivasankari, and V. Vijayan, "Personal authentication using hand vein," *International Journal of Engineering Research and Technology (IJERT),*, vol. 3, pp. 2333–2339, 2014. [Online]. Available: https://www.ijert.org/personal-authentication-using-hand-vein

[10] M. I. Obayya, M. El-Ghandour, and F. Alrowais, "Contact-less palm vein authentication using deep learning with bayesian optimization," *IEEE Access,*, vol. 9, pp. 1940– 1957, 2021.

[11] L. Prechelt, "Automatic early stopping using cross valida- tion: quantifying the criteria," *Neural Networks*, vol. 11, pp. 761–767, 1998.

[12] S. Processing. (2022) Credit card fraud statistics. [Online]. Available: https://shiftprocessing.com/credit-card-fraud-statistics/

[13] B. Sankur. (2011) Bosphorus hand vein database. Bogazici University. [Online]. Available: http://bosphorus.ee.boun.edu.tr/hand/Home.aspx

[14] G. Sathish, S. Narmadha, S. Saravanan, and S. U. Mah- eswari, "Personal authentication system using hand vein biometric," *International journal of computer technologyand applications,*, vol. 3, pp. 383–391, 2012.

[15] Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of BigData,*, vol. 6, pp. 1–48, 2019.

[16] S. Sun, Z. Cao, H. Zhu, and J. Zhao, "A survey of opti- mization methods from a machine learning perspective," *IEEE Transactions on Cybernetics,*, vol. 50, pp. 3668– 3681, 2020.

[17] Thapar, G. Jaswal, A. Nigam, and V. Kanhangad, "Pvs- net: Palm vein authentication siamese network trained using triplet loss and adaptive hard mining by learning enforced domain specific features," in *2019 IEEE 5th International Conference on Identity, Security, and Be- havior Analysis (ISBA),*, Hyderabad, India, jan 2019.

[18] R. Trabelsi, A. Masmoudi, and D. Masmoudi, "A new multimodal biometric system based on finger vein and hand vein recognition," *International Journalof Engineering Research and Technology (IJET),*, vol. 5, pp. 3175–3183, 2013. [Online]. Avail-able: http://www.enggjournals.com/ijet/docs/IJET13-05- 04-297.pdf

[19] L. Welch, "The generalization of 'students's' prob- lem when several different population variances are involved," *Biometrika,*, vol. 34, pp. 28–35, 1947.

[20] F. Wilches-Bernal, B. Núñez-Álvares, and P. Vizcaya. (2020) A database of dorsal hand vein images. [Online]. Available: https://arxiv.org/abs/2012.05383

[21] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),*, Boston, MA, USA, jun 2015.