SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

An Efficient Iris Recognition Technique using CNN and Vision Transformer

Samihah Abdul Latif[1], Khairul Azami Sidek[1,*], Aisha Hassan Abdalla Hashim[1]

[1] Department of Electrical and Computer Engineering, Kuliyyah of Engineering, International Islamic University Malaysia, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 26 June 2023
Received in revised form 14 November 2023
Accepted 22 November 2023
Available online 7 December 2023

Keywords:
Convolutional Neural Network (CNN); Vision Transformer (ViT); hybrid model; iris recognition

ABSTRACT

The usage of biometric identification has increased in recent years, with numerous public and commercial organizations incorporating biometric technologies into their infrastructures. One of the technologies is iris recognition which has been used as a biometric recognition compared to other modalities to combat identity abuse due to its ability to eliminate risk of collisions or false matches even when comparing large populations. The use of CNN is proven to provide high accuracy; however, this technology involves the need for a large dataset and higher computational cost. Therefore, this study uses a combined model of Convolutional Neural Network (CNN) and Vision Transformer (ViT) in identifying and verifying an iris image. By using the proposed learning rate, it proves that the novel hybrid model is capable to achieve up to 93.66% accuracy in recognizing iris images. The cross-entropy loss function was implemented to reduce the loss and it was able to predict the class label more correctly. In addition, the model was thoroughly tested on three publicly available iris databases, achieving satisfactory iris recognition results. Furthermore, this model has the potential to be used in other biometrics such as face and retina recognitions.

1. Introduction

Performing iris recognition and identification as one of the biometric verifications has gained more attention during the past few years. The most trustworthy and efficient biometric method for identification and authentication is iris recognition [1]. Due to their stability, accuracy, and most reliable biometric identification technologies, hence it is widely applied in various applications, such as intelligent unlocking, border control, forensics, and healthcare [2].

In biometrics, iris became a highly informative object and high reliability as shown in Table 1, because the eye's highly distinctive qualities and easily accessible that do not alter throughout time [3,4]. Besides that, iris is not affected by age and does not require human contact with its scanning device [5]. Thus, this contactless biometric recognition gets attention from the public because they are relatively well known for their user-friendly and seamlessness [6].

* *Corresponding author.*
E-mail address: azami@iium.edu.my

https://doi.org/10.37934/araset.34.2.235245

Table 1
Comparison of biometric accuracy and reliability [7,8]

Biometric	Accuracy	Reliability	Uniqueness
Fingerprint	Very High	High	High
Facial	High	Medium	Low
Hand Geometry	High	Medium	Medium
Iris	Very High	High	High
Retina	Very High	High	High
DNA	Very High	High	None
Voice	Medium	Low	Low
Signature	Medium	Low	Low

Based on Table 1, iris scan has a very high accuracy that refers to how well a specific biometric can distinguish between individuals. The reliability of iris is high compared to other biometric which refers to how dependable it is for recognition purpose [9]. While the uniqueness of the iris is also high which means the iris is distinct from each other [10] make it extensively investigated for accurate and personal identification [11].

The iris recognition system has been developed using a variety of technologies. The most common is Convolutional Neural Network (CNN) that become state of the art in image recognition [12]. However, Self-attentive mechanisms-based Vision Transformers (ViT) have successfully classified and recognized images accuracy comparable to neural networks. Besides, the price of increasing accuracy is an exponential rise in the number of parameters without neural networks' built-in image-specific biases. Self-attention models, particularly ViTs, are an alternative to CNNs in computer vision. Transformers were first presented in 2018 for machine translation, but they have subsequently evolved into the primary architecture for Natural Language Processing (NLP) tasks including text summarization and speech recognition. They are responsible for the most current advancements in NLP, such as Google's BERT and Open Ai's GPT-3.

Many mobile vision tasks have been powered by lightweight CNNs. However, in [13] research, they do not rely on CNN at all and attempts have previously been made to use transformers for image recognition tasks. It produces results for image classification that are comparable to many cutting-edge CNNs [14]. Convolutional networks were either transformed using transformers or transformed by replacing specific modules in these attempts. It only used the common transformer design, which is the most common architecture used in NLP. The sole innovation they employed was to split an input image into multiple image patches, which were then fed into the transformer as usual input. The rest of the transformer's architecture, however, was unchanged. In the context of NLP, these image patches are treated similarly to words (tokens). As a result, the 16 by 16 input images are interpreted as 16 by 16 words in Ref. [13].

The outcomes demonstrate the uniqueness of this architecture that outperforms cutting-edge CNNs, which are referred to as the premier network design for image recognition. There is a lack of specific localization information caused by repetitive nature of down-sampling. ViT's emphasis on low-resolution features makes it unfit to capture images [15]. Although ViT has made comparable improvement, many researchers suggest hybridizing the computer vision model with CNN and ViT [15].

Inspired by [16] study that proposed a Progressive Multi-scale Vision Transformer (PMVT) model which obtained comparable performance based on the convolutional feature maps and transformer encoder where tokens received as input at different resolutions. Therefore, this work focuses on the development and implementation of combined CNN and ViT technologies in recognizing iris images. Considering the strengths of these two technologies, CNN were extracted in the features at the early stage and the next process will be completed using a transformer model.

2. Related Work

This section offers a comprehensive overview of commonly employed Deep Learning (DL) techniques in biometric recognition, along with an examination of studies focusing on the utilization of ViT in image recognition. DL offers a distinct advantage over traditional Machine Learning (ML) methods by enabling autonomous feature learning, resulting in significant time savings [17]. Furthermore, deep learning models automatically uncover hidden patterns within the data [18].

As a way to enhance the recognition of the iris system's Convolutional Neural Network model that has been pre-trained, with which the ImageNet Large Scale Visual Recognition Competition (ILSVRC) debuted, an iris segmentation phase-free transfer learning technique is presented [19]. The network is additionally optimized to avoid overfitting using Bayesian optimization and data augmentation. As a result, IRS using Deep Learning technique increased the ability of a trained model to be applied to new target tasks.

In [20] study, in order to recognize iris images, researchers are applying a hybrid network model that is based on EfficientNet-b0. In this model, iris segmentation, normalization, feature extraction, and matching are all integrated into a single network. The combination of iris databases comprised of MMU2 and CASIA Thousand demonstrated that the composite network framework's accuracy and efficiency outperformed the previous network framework. The model also has high parameter speed and efficiency.

An initial deep-learning facial recognition model was built to deliver quantitative data on low-capacity devices and its accuracy, size, and timings evaluated in [21] study. A number of databases, including both public and private datasets made especially for modelling the complexity of mobile scenarios were used, as the objective of this study is to cover as many scenarios as possible. Publicly accessible models and conventional methods were also evaluated to make a fair comparison. In addition, the assessment is strengthened using a variety of classifiers and detectors due to the significance of the template matching and face identification stages. To get performance information for the complete system when it was integrated into a mobile phone, A system implementation in the JAVA and Android programming languages was created and tested.

In [22] study presents some cutting-edge deep learning models are used in this review to combine these multimodal large data. With the expanding investigation of multimodal big data, there are still certain issues to be resolved. In order to educate readers on multimodal deep learning fusion approaches and inspire the development of new deep learning multimodal data fusion techniques, regardless of their originating community, this review gives a survey on the topic.

Liu *et al.,* [23] first introduced a unique condensed 2-channel CNN with little training data for quick and accurate iris identification and verification. This model is initially put forth as a high-performance basic iris classifier with three online augmentation schemes and radial attention layers. The weight distribution of the model is examined to perform branch and channel pruning. The CASIA-V4-Thousand database's promising results show that the suggested method can be applied in difficult iris recognition scenarios. In addition, the most crucial iris localization function, such as estimating parameterized pupillary and limbic iris boundaries, has been long neglected, resulting in superior CNN-based iris segmentation that is not fully utilized in iris recognition [24].

The comparison of related work shown in Table 2. Representative architectures that are widely used are summarized as fundamental to multimodal deep learning understanding. Following that, a summary of the current cutting-edge multimodal data fusion deep learning models is provided. Finally, some multimodal data fusion deep learning models challenges, and future topics are discussed.

Table 2
Comparison of related work

Author	Method(s)	Performance	Limitation
[19]	Transfer learning (AlexNet & DenseNet201)	Accuracy: AlexNet:97.22 DenseNet201: 98.81	The complexity of the model and the length of time needed to process it in algorithms and for computations are trade-offs
[20]	EfficientNet-b0	Accuracy: 98%	Iris location was deviated using pre-trained model
[21]	Deep learning for face recognition.	Time: 5s, FRR: 16.68% FAR: 0% FTA: 6.56% FNMR: 11.42%	The application is a bit slow but highly secure
[22]	Convolutional Neural Network for multimodal.	0.60% Error Equal Rate	Multi-modal recognition
[23]	2-ch Deep Convolutional Neural Network	(Verification EER) CASIA V1: 0.33% CASIA V3: 0.76% CASIA V4: 1.19%	The model has lower Floating-point operations and parameter

3. Methodology

This section presents the proposed of hybrid model. The architecture of CNN and ViT model is shown in Figure 1 that contains four major parts which is CNN layer, Encoder of Transformer, Linear Projection of Flattened Patches, and Multi-layer Perceptron.

3.1 The Proposed Model

The code is primarily based on CNN and ViT. The idea of a hybrid model is a pre-processing an input image with convolutional on the network. The CNN progressively reduces the spatial dimension of the input image and produces an output tensor of dimension w, h and c that denote the width, height, and channel. That is the number of features of the last layer. Algorithm 1 show steps in extracting the feature learning of iris images.

3.1.1 Algorithm 1: pre-processing data

```
i.    Input: read 3 channel images (RGB)
ii.   Output: TrainSetA, ValidSetA, TestSetA, TrainLabelA, ValidLabelA, TestLabelA
iii.  Calculate iris size
iv.   For (i in tqdm) Do
v.            if (i!=file.txt and i==image name) Do
vi.                   show() image
vii.                  crop and resize to 128x128
viii.                 show() cropped image
ix.                   Sort classified image
x.        Else show() not matching image
xi.   End For
```

Based on Algorithm 1, the input iris images will be crop and resized to 128 x 128 resolution. Most deep learning model architectures expect all input images to be of the same dimensions. The size of 128 x 128 pixels is sufficient to get the image iris feature and reduce the rate of image quality loss. In addition, accuracy will be lost if the image is cropped with a large size. The process of cropping and resizing or rescaling are part of data augmentation which are commonly used in previous studies [23,24].

The trainable linear projection layer is then applied to these output layer as in Figure 1. This layer serves as an embedding layer and produces vectors with fixed sizes. The sequence of output layer is then linearly concatenated using position embeddings to preserve the positioning information of the images.

The image patches' relative or absolute position in the sequence is crucial information that is injected. The 0th class is a crucial aspect of the position embedding module to be aware of. The 0th class concept is derived from the class token in Bidirectional Encoder Representations for Transformers (BERT). The class token will be served as the image representation that will be used for image recognition. This class learned as well as other classes, but its learning was unrelated to how it appeared. Instead, the model design has it hardcoded.

The transformer encoder is made up of several stacks of identical blocks. Each block begins with a Multi-Head Attention (MSA) layer and ends with a Feed-Forward layer. Each of the two sub-layers has a residual connection, which is followed by layer normalization. The model generates an output of embedded dimension D from all sublayers and embedding layers. The final vector of embedded dimension D is obtained by concatenating these attention heads and then passing them through a dense layer.

MLP, which stands for Multi-Layer Perceptron, is essentially a collection of layers that perform linear transformations. A two-layer classification network called the MLP concludes with a GELU (Gaussian Error Linear Unit). The output of the transformer is provided by the last MLP block, also referred to as the MLP head.

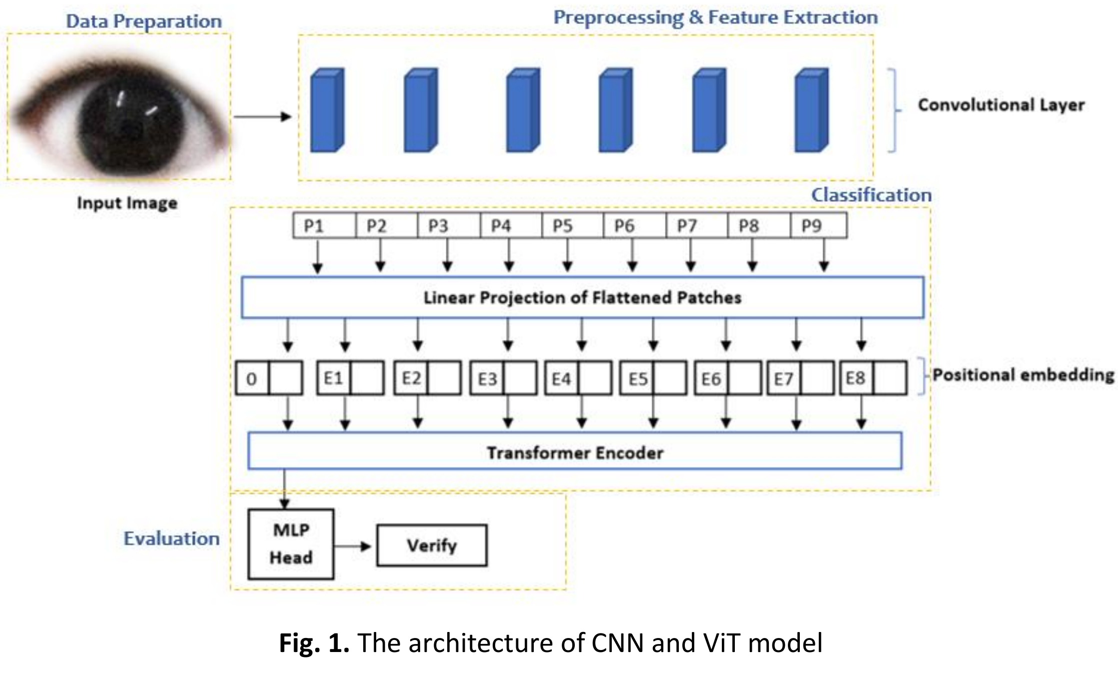


Fig. 1. The architecture of CNN and ViT model

In this experiment, the Cross Entropy loss function as in Eq. (1) was implemented to make this model able to predict the class label more correctly and to minimize the error.

$$H(p,q) = -\sum_{i=0}^{n} p(x_i)\log\big(q(x_i)\big) \tag{1}$$

where $p(x_i)$ represent the truth label and $q(x_i)$ is the SoftMax probability of the i^{th} class.

3.2 Databases

The Chinese Academy of Sciences (CASIA) iris databases are the most widely used; databases were being used 453 times in 305 articles, or in more than 44% of the research [27]. Despite being out of date, the CASIA Iris Database v.1 is still the most used and cited iris picture database (used mostly as a benchmark in 144 (21%) papers). Therefore, in this study, we are using CASIA Iris v.1 [28] which consist of 756 iris images for 108 persons.

MICHE was collected in a relatively unconstrained environment. The database contains 3732 images from 92 subject and 184 classes. The MICHE-I dataset contains images taken in unrestricted conditions [29]. As a result, their average quality is low, and the primary goal of the participating approaches was to try to address such data degradation.

BioSecure Multimodal Database (BMDB) iris database is a public database consisting of four eye images (two left and two right) with a resolution of 640*480 pixels, obtained from 200 subjects in two separate sessions. Iris patterns for each eye are unique for every individual and this is helpful in identifying an individual. Hence, the database will include 400 subjects, each with four eye images. The data in the file called “parameters.txt” are used to crop the eye images such that they only contain the iris area.

4. Results and Discussion

This section shows the investigation of the performance of ViT and CNN models in iris recognition during the experimentation.

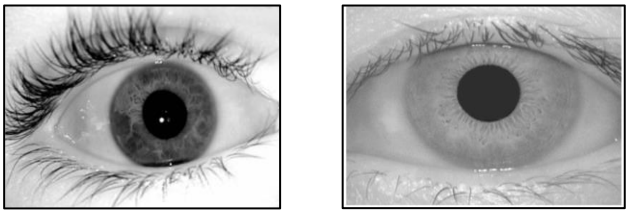
4.1 Experiment

In this section, the time taken will be highlighted when using three different databases and comparison between iris image quality. Table 3 show time taken for each image in testing phase. Among them, BMDB iris dataset only takes 0.12 seconds to verify the image while MICHE DB takes 0.16 seconds and CASIA, 0.19 seconds.

Table 3
Comparing time taken using same model and different dataset

Name of Database	Number of images (No. of test images)	Time (sec)
CASIA Iris V1	756 (750)	0.19
MICHE DB	3732 (750)	0.16
BMDB	1600 (750)	0.12

These comparison between a good quality image and a noise-corrupted iris image. Image quality also affects the processing time for image verification. In this experiment, BMDB Iris dataset get the shortest testing time per image due to the image quality as shown in Figure 2 (a). The CASIA dataset contains images that have disturbances where the iris is not very clear and is disturbed by eyelashes and ambient light as depict in Figure 2 (b). This difference is very significant because if the model is implemented in the crucial environment, the image processing time is very important and decisive for the success of the lightweight model.



(a) (b)

Fig. 2. (a) Iris image from BMDB [30], and (b) CASIA [28]. The sample iris images randomly picked from two databases

A deep learning model consumes data, makes predictions, and then determines whether the model's predictions were good or bad. The model's predictions were then compared to the labels, and the metric, the cross-entropy loss function, was used to calculate how much the predictions differed from the labels.

This model will be more adept at accurately predicting the labels the more data the model use to train it. This is because more data will allow a model more opportunities to make mistakes and learn from them. Assuming that the selection dataset is reasonable and does not have many entries that are highly like one another or many data points that are not representative. So, in this experiment, a complete dataset was used to train the model because the goal is to make it as accurate as possible at predicting the labels.

In theory, this model would perform better if the entire dataset was used for training rather than just 70% or 80% of the data. The problem is that if all the data is used for training, the user is unable to objectively assess the true performance of the model. Sure, the model could be evaluated using the data that it was trained on, but that would be problematic. As a result, in this experiment, 50% of the data is allocated to the training set, 25% to the validation set, and 25% to the test set.

4.2 Performance Analysis

The implementation of hybrid ViT and CNN model into iris recognition system to increase the accuracy of verification is in line with initial expectations. This model is able to provide data accuracy up to 93.66% as shown in Table 3. The BMDB database was chosen to be used in this experiment because the images obtained from it demonstrate that the processing time is the shortest. By using this dataset, two different learning rates are used to see how this learning rate affects the performance of this model.

Table 4 shows the accuracy rate and loss obtained. 800 iris images are assigned for training once the learning rate is set to 0.0001, and 400 iris images are designated for testing. A total of 50 epochs were used to obtain this high accuracy result. When the model is set to 25 epochs, the accuracy drops to 88.88%.

Table 4
Comparing the accuracy and loss using same model

Learning Rate	Epochs	Accuracy	Loss
0.0001	50	93.66%	7.74%
0.0001	25	88.88%	12.79%
0.001	50	85.12%	18.80%
0.001	25	76.29%	25.06%

The training phase is set to 25 epochs and used 0.0001 as the learning rate. Total samples data are 160 images, epoch is set to 25, batch size 5, number of iterations is 32 (160/5). Figure 3 shows the network performance during training and testing stage.

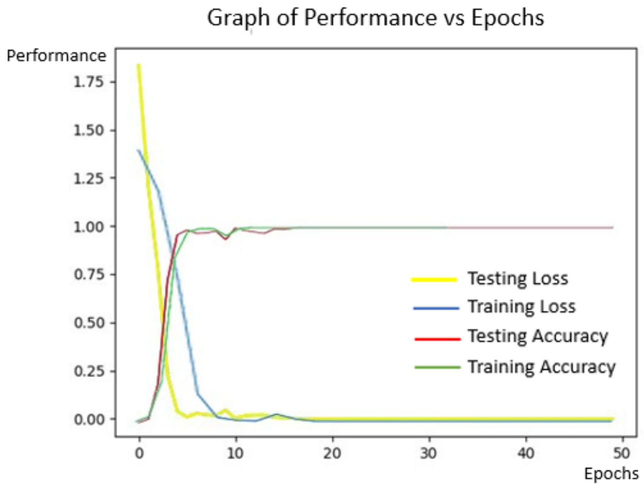


Fig. 3. Performance of the model using 0.0001 Learning Rate

Figure 3 above, depicts training and validation, the loss of validation is significantly lower compared to training loss, showing that it is simpler to predict values from the validation dataset than from the training dataset. One justification is that the data used for validation is limited which is only 400 iris images, but the training dataset (800 iris images) is widely represented, so the model therefore works exceptionally well considering these few instances. This is indicated that, when the loss function is minimized, the neural network model automatically makes better predictions, regardless of the specific characteristics of the task at hand.

Validation loss start with high value which is 1.75 and training loss start with 1.4. Both are decreased at 10 epochs and consistent until 50 epochs. While for training and validation accuracy, both are increased and consistent near to 1.00 until 50 epochs. This can be summarized that the model is learn the features well and able to verify the image up to 100 percent by using 0.0001 learning rate.

Figure 4 shows the performance of accuracy and loss by using 0.001 learning rate. As can be seen, training loss was started with high value which is 3.0 and validation loss is starting with 1.4.

This demonstrated that the learning rate was high, and it became difficult to learn small changes in the parameters required to fine-tune the model around the end of the evaluation process. As a result, the error flattens out very early. The gradient descent changes dramatically and hops around as soon as the learning rate is set to a high number. Therefore, this might cause divergence, which would raise the error.

When adjusting model weights during training, cross-entropy loss is used. Since reducing loss is the main goal, a model is better when the loss is low. In an ideal model, there is no cross-entropy loss. In this experiment, the function was successfully embedded this function to reduce the loss to 7.74%.

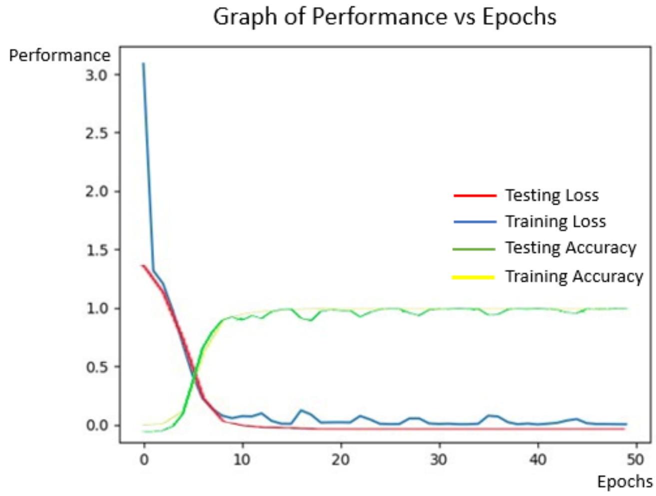


Fig. 4. Performance of the model using 0.001 Learning Rate

To best map inputs to outputs, a neural network learns or approaches a function using samples in the training dataset. The learning rate hyperparameter control how slow or fast the model learns. It precisely controls the amount of assigned error with which the model's weights are updated each time they are updated. The model will learn to predict the function as accurately as possible given the number of layers and the number of nodes per layer over an infinite number of training epochs and a precisely configured learning rate.

Typically, a model can learn more quickly if its learning rate is high, but at the cost of obtaining a less-than-ideal final set of weights. It might be possible for the model to learn a more ideal, or even a set of weights that is globally optimal in a moderate learning, but it will take a lot more time to train. In extreme cases, an excessively fast learning rate will provide weight updates that are excessively massive, which will cause the model's performance to fluctuate over training epochs. Diverging weights are thought to be the root of fluctuating performance. If the learning rate is too low, the problem might never be solved, or it might become trapped on a poor choice.

Even though, the iris recognition system with the CNN model which applies AlexNet and DenseNet201 achieved higher accuracy rates of 97.22% and 98.81% respectively, however, there is a trade-off with computational time and complexity of the algorithm [19]. Furthermore, to the best of our knowledge, there is no such model used to test iris recognition using ViT technology while the closest research is facial action unit detection that using ViT technology. The accuracy obtained is much lower which is 64.2% by using BP4D dataset [16]. Due to the limitation of the existing methods, therefore, this study suggested a novel hybrid model consisting of CNN and ViT technologies to overcome the aforementioned issues and achieved better accuracy outcome.

5. Conclusions

In this study, a novel hybrid model of ViT and CNN for computer vision was proposed in the context of iris recognition. This model was successfully trained and test by using the iris dataset to verify the legitimate user. Among three publicly available datasets, BMDB dataset obtained the shortest testing time per image due to the image quality. A well generalize network is developed and has been experiment by using high quality iris image to get high accuracy and low error. This experimental study also demonstrates that by utilizing 50 epochs and a learning rate of 0.0001 are sufficient to achieve a better outcome with an accuracy of 93.88%. In the future, numerous potential directions can be tested in different settings for robustness, scalability and security. This hybrid model also has the potential to be applied in various biometrics, such as face and retina recognitions.

Acknowledgement
This research was not funded by any grant.

References

[1] Ranjan, Shubhika, S. Prabu, P. Swarnalatha, G. Magesh, and Ravee Sundararajan. "Iris recognition system." *International Research Journal of Engineering and Technology* 4, no. 12 (2017): 864-868.
[2] Wang, Caiyong, Jawad Muhammad, Yunlong Wang, Zhaofeng He, and Zhenan Sun. "Towards complete and accurate iris segmentation using deep multi-task attention network for non-cooperative iris recognition." *IEEE Transactions on information forensics and security* 15 (2020): 2944-2959. https://doi.org/10.1109/TIFS.2020.2980791
[3] Odinokikh, Gleb, Mikhail Korobkin, Ivan Solomatin, Iurii Efimov, and Alexey Fartukov. "Iris feature extraction and matching method for mobile biometric applications." In *2019 International Conference on Biometrics (ICB)*, pp. 1-6. IEEE, 2019. https://doi.org/10.1109/ICB45273.2019.8987379
[4] Rana, Humayan Kabir, Md Shafiul Azam, Mst Rashida Akhtar, Julian MW Quinn, and Mohammad Ali Moni. "A fast iris recognition system through optimum feature extraction." *PeerJ Computer Science* 5 (2019): e184. https://doi.org/10.7717/peerj-cs.184
[5] Danlami, Muktar, Sapiee Jamel, Sofia Najwa Ramli, and Mustafa Mat Deris. "A framework for iris partial recognition based on Legendre wavelet filter." *International Journal of Advanced Computer Science and Applications* 10, no. 5 (2019). https://doi.org/10.14569/IJACSA.2019.0100506
[6] Ahmadi, Neda, Mehrbakhsh Nilashi, Sarminah Samad, Tarik A. Rashid, and Hossein Ahmadi. "An intelligent method for iris recognition using supervised machine learning techniques." *Optics & Laser Technology* 120 (2019): 105701. https://doi.org/10.1016/j.optlastec.2019.105701
[7] Al Rousan, Mohammad, and Benedetto Intrigila. "A Comparative Analysis of Biometrics Types: Literature Review." *Journal of Computer Science* 16, no. 12 (2020): 1778-1788. https://doi.org/10.3844/jcssp.2020.1778.1788
[8] Jain, Anil, Lin Hong, and Sharath Pankanti. "Biometric identification." *Communications of the ACM* 43, no. 2 (2000): 90-98. https://doi.org/10.1145/328236.328110
[9] Farouk, Rahmatallah Hossam, Heba Mohsen, and Yasser M. Abd El-Latif. "A Proposed Biometric Technique for Improving Iris Recognition." *International Journal of Computational Intelligence Systems* 15, no. 1 (2022): 79. https://doi.org/10.1007/s44196-022-00135-z
[10] Szymkowski, Maciej, Piotr Jasiński, and Khalid Saeed. "Iris-based human identity recognition with machine learning methods and discrete fast Fourier transform." *Innovations in Systems and Software Engineering* 17 (2021): 309-317. https://doi.org/10.1007/s11334-021-00392-9
[11] Jenadeleh, Mohsen, Marius Pedersen, and Dietmar Saupe. "Blind quality assessment of iris images acquired in visible light for biometric recognition." *Sensors* 20, no. 5 (2020): 1308. https://doi.org/10.3390/s20051308
[12] Wei, Yinyin, Xiangyang Zhang, Aijun Zeng, and Huijie Huang. "Iris Recognition Method Based on Parallel Iris Localization Algorithm and Deep Learning Iris Verification." *Sensors* 22, no. 20 (2022): 7723. https://doi.org/10.3390/s22207723
[13] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani *et al.,* "An image is worth 16x16 words: Transformers for image recognition at scale." *arXiv preprint arXiv:2010.11929* (2020).
[14] Alom, Md Zahangir, Mahmudul Hasan, Chris Yakopcic, Tarek M. Taha, and Vijayan K. Asari. "Inception recurrent convolutional neural network for object recognition." *Machine Vision and Applications* 32 (2021): 1-14. https://doi.org/10.1007/s00138-020-01157-3
[15] Adjei-Mensah, Isaac, Xiaoling Zhang, Adu Asare Baffour, Isaac Osei Agyemang, Sophyani Banaamwini Yussif, Bless Lord Y. Agbley, and Collins Sey. "Investigating vision transformer models for low-resolution medical image recognition." In *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp. 179-183. IEEE, 2021. https://doi.org/10.1109/ICCWAMTIP53232.2021.9674065
[16] Wang, Chongwen, and Zicheng Wang. "Progressive multi-scale vision transformer for facial action unit detection." *Frontiers in Neurorobotics* 15 (2022): 824592. https://doi.org/10.3389/fnbot.2021.824592
[17] Jamil, Amirah Hanani, Fitri Yakub, Azizul Azizan, Shairatul Akma Roslan, Sheikh Ahmad Zaki, and Syafiq Asyraff Ahmad. "A Review on Deep Learning Application for Detection of Archaeological Structures." *Journal of Advanced*

Research in Applied Sciences and Engineering Technology 26, no. 1 (2022): 7-14. https://doi.org/10.37934/araset.26.1.714

[18] Khan, Umair, William Pao, Nabihah Sallih, and Farruk Hassan. "Flow Regime Identification in Gas-Liquid Two-Phase Flow in Horizontal Pipe by Deep Learning." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 27, no. 1 (2022): 86-91. https://doi.org/10.37934/araset.27.1.8691

[19] Yow, Sue Chin, and Ahmad Nazri Ali. "Iris recognition system (IRS) using deep learning technique." *Journal of Engineering Science* 15, no. 2 (2019): 125-144. https://doi.org/10.21315/jes2019.15.2.9

[20] Hu, Qingqiao, Siyang Yin, Huiyang Ni, and Yisiyuan Huang. "An end to end deep neural network for iris recognition." *Procedia Computer Science* 174 (2020): 505-517. https://doi.org/10.1016/j.procs.2020.06.118

[21] Ríos-Sánchez, Belén, David Costa-da Silva, Natalia Martín-Yuste, and Carmen Sánchez-Ávila. "Deep learning for face recognition on mobile devices." *IET Biometrics* 9, no. 3 (2020): 109-117. https://doi.org/10.1049/iet-bmt.2019.0093

[22] Li, Hui, and Lei Zhang. "Multi-exposure fusion with CNN features." In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 1723-1727. IEEE, 2018. https://doi.org/10.1109/ICIP.2018.8451689

[23] Liu, Guoyang, Weidong Zhou, Lan Tian, Wei Liu, Yingjian Liu, and Hanwen Xu. "An efficient and accurate iris recognition algorithm based on a novel condensed 2-ch deep convolutional neural network." *Sensors* 21, no. 11 (2021): 3721. https://doi.org/10.3390/s21113721

[24] Wang, Caiyong, Yunlong Wang, Boqiang Xu, Yong He, Zhiwei Dong, and Zhenan Sun. "A lightweight multi-label segmentation network for mobile iris biometrics." In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1006-1010. IEEE, 2020. https://doi.org/10.1109/ICASSP40776.2020.9054353

[25] Wang, Wenshan, Yaoyu Hu, and Sebastian Scherer. "Tartanvo: A generalizable learning-based vo." In *Conference on Robot Learning*, pp. 1761-1772. PMLR, 2021.

[26] Takahashi, Ryo, Takashi Matsubara, and Kuniaki Uehara. "Data augmentation using random image cropping and patching for deep CNNs." *IEEE Transactions on Circuits and Systems for Video Technology* 30, no. 9 (2019): 2917-2931. https://doi.org/10.1109/TCSVT.2019.2935128

[27] Omelina, Lubos, Jozef Goga, Jarmila Pavlovicova, Milos Oravec, and Bart Jansen. "A survey of iris datasets." *Image and Vision Computing* 108 (2021): 104109. https://doi.org/10.1016/j.imavis.2021.104109

[28] CASIA, "CASIA-Iris-Mobile-V1.0 - Casia mobile database (datasets S1, S2 and S3)." http://biometrics.idealtest.org/dbDetailForUser.do?id=13

[29] De Marsico, Maria, Michele Nappi, Fabio Narducci, and Hugo Proença. "Insights into the results of miche i-mobile iris challenge evaluation." *Pattern Recognition* 74 (2018): 286-304. https://doi.org/10.1016/j.patcog.2017.08.028

[30] Ortega-Garcia, Javier, Julian Fierrez, Fernando Alonso-Fernandez, Javier Galbally, Manuel R. Freire, Joaquin Gonzalez-Rodriguez, Carmen Garcia-Mateo *et al.,* "The multiscenario multienvironment biosecure multimodal database (bmdb)." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, no. 6 (2009): 1097-1111. https://doi.org/10.1109/TPAMI.2009.76