Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

"TrustPass" -Blockchain based Trusted Digital Identity Platform towards Digital Transformation

Kalpa Dissanayake[1], Pavan Somarathne[2] and Ushan Fernando[3]

[1] Sri Lanka Institute of Information Technology

Received: 6 October 2021 Accepted: 5 November 2021 Published: 16 November 2021

Abstract

According to the United States Census Bureau, by June 2019 world population on earth was 7.5 billion, which exceeds the world population of 7.2 billion as of 2015. Each of these citizens needs to prove their identity in order to fulfill their day-to-day routine. In this current digital revolution whole world is transforming to digitalization. Therefore, proving someone?s identity in the digital space is a must, because being able to track a person digitally can result in elimination of the identity theft and most incidents related to online harassments, while focusing on data privacy and security of citizens, we have proposed ??Trust Pass??: Cyber Security Intelligence based trusted digital identity platform capable of registering and verifying service providers based on document validation neural network model (95.4

Index terms— cyber security intelligence, blockchain, cyber threat, three-factor biometric, data security and privacy, digital identity, neural networks.

1 I. INTRODUCTION

n day-to-day life verifying our identity or proving who we are is an inevitable factor. [1] Every Sri Lankan citizen should have a valid document with them that can prove their identity at all times. In Sri Lanka every citizen who is above 16 years of age must obtain a National Identity Card which can prove their identity anywhere in the country. [1] By the age of 18 citizens can apply for the driving license card which also can be used as a proof of identity. Furthermore, citizens can obtain and use their passport as a proof of identity as well. According to our survey [2] most people in Sri Lanka use their National Identity Card as their proof of identity. However, when entire world is moving towards a digital transformation being able to prove our identity remotely will give us a huge advantage. In traditional document-based identity proving system we can't do remote verifications to prove our identity we must physically present with our id document to prove our identity. This can be very troublesome on many occasions. according to our survey [2], clear majority of Sri Lankan citizens prefer to get services that require identity verifications via online remote delivery method rather than spending hours of their valuable time in government or private offices. Currently there is no any system that provide remote identity verifications for citizens of Sri Lanka. Many other countries have developed similar systems.

? Sing Pass in Singapore [3] ? Digital Id in Australia [4] ? Accenture KTDI in USA [5] ? Yoti in United Kingdom [6] ? SmartID in United Kingdom [7] Developing a comparable but more locally compatible digital identity system in a developing country like Sri Lanka can be a challenging task. Currently in Sri Lanka only two out of five people has digital literacy [8]. and moreover, Sri Lanka don't have a very sophisticated digital infrastructure. To develop a system that can overcome these challenges and facilitate maximum user convenience, we must study deep into other existing systems and popular research work.

2 II. BACKGROUND AND LITERATURE REVIEW a) Service Provider Management with Document

Validater Service Providers or the relying parties (RPs) like examination centers, Banks, Police etc. are one of the most required user groups which we are facilitating the services through our system. According to our survey [2] local and government service providers got very average level of significance because they do not have the facility to authenticate user details without manual inspection and it takes more time to accommodate the request. Since we are creating a secure, authenticated platform between users and RPs, RPs could be able to develop their performances without wasting users' time.

Because of that RPs need to register to our system to authenticate users through our system to their system. Since we are enabling RPs to register online, they must provide valid documents to Trust Pass system. We are introducing a Image Based Document Validator which can identify invalid document and invalidate the registration without wasting further personal and infrastructure resources in the registration process.

For the document validator we have created a Convolutional Neural Network (CNN). Performance of the CNN have gained tremendous success within last decades [9]. However, to gain a high accuracy and the performance the architecture of the CNN has a definite impact. On the other hand, we must have a large number of pre classified dataset to get a good output and it prevents the usage of many off the shelf state-ofart CNNS like Alex Net, VGG, ResNet being applied in classification problems and it may affect for the overfitting of the model too [10].

Although there are state-of-art CNNs in the market according to these researches [11] [12] that has been carried our regarding image classification Alex Net is the widely used CNN which has five convolutional layers and 3 fully connected layers.

3 b) Three Factor Biometric Authentication

On service providers requests citizens should be able to authenticate their details securely because of this implementing a fast, user-friendly, and secure authentication method that can be integrated with any smartphone is a crucial requirement of our system. according to statistics [13] at the end of 2018, more than 60% of smartphones are developed with an integrated fingerprint sensor and the present data suggests that by the 2023, more than 80% of smartphones will have some form of biometric hardware installed. [14] In this research [15] researchers were able to develop a face detection and recognition system that have an accuracy of 90%. According to their research to achieve a better accuracy and solid reliability they propose to integrate an iris scanner to the system. They have stated that without matching iris data this system is not suitable for use as an authentication method for ATM Machines or other high security systems.

Face Net is a face recognition model developed by google according to their research [16] Face Net is 99.63% accurate in distinguishing different faces and identifying them . Face Net is developed using a deep convolutional network with two different architectures The Zeiler & Fergus type [16] which can have many parameters and large number of Flops [16] and the Inception type [16] which can have few parameters. Zeiler & Fergus is more suitable for run in datacenter while Inception is proposed to run on mobile devices because of less memory usage.

Although accuracy of the biometric is a concern, we are primarily focused on developing a solution that will facilitate maximum security and verify the live presence of the user. According to this study [17]spoofing a 2D (two dimensional) face recognition which only incorporates a selfie camera and no special hardware similar to 3D (three dimensional) face recognition is a fairly easy task, getting access to user's photograph or recorded video clip is enough to launch a presentation attack [18].

4 c) Blockchain based Cyber Security Inteligence

The identity of living beings on earth depends on the characteristics of the body. The identity of the human body depends on the biological and social nature of the body. People have to deal with different people in their day-to-day activities and personal identity is very important. Humans' physical bodies reflect inherent traits and identities of humans. When a person thinks of himself as Who Am I, personal identity is reflected and this is unique ownership for each person. The characteristic that can be seen here is that the human's identity is indefinite and temporary. This can change over time. What is the answer we can give to the question of whether we were in a certain place, one day? It's really hard. This is due to the lack of definite identity. Therefore, physical identity is studied. Although efforts have been made to identify a person by his appearance, they have not been successful. Why is identity necessary? The main reason for the problem is the population. The Earth is now estimated to have a population of over 7 billion. Over time, the identification of identities can lead to many problems. Person identities can be mainly categorized into age, class, gender, national, regional, spiritual groups. The solution was to introduce document base methods such as National identity card, Passport, driving license, etc. There were various data privacy and security problems with these methods. Therefore, although it was possible to provide digitized solutions to personal identities, it was not possible to provide a reliable, effective, usability solution due to technical issues.

5 d) Analysis of user behavior and usage patterns

As the research block chain based trusted digital ID platform concerns regarding the usage of the digital identity by users. Usage pattern of the digital identity explains the user's usage of the digital identity platform. It's a fact that not every user's usage is equivalent, as the user's usage differ and vary from each other. In order to clearly analyze the data regarding the user's usage, analysis of usage patterns of the user can be introduced.

Analyzing of data includes data manipulation, data transformation, and data visualization in order to make a meaningful result from specific data set. These meaningful insight of data helps to make decisions. Therefore, it enables commercial fields, individuals and the governments make decisions from the insights, acquired from the data analysis. **??**19] The data analysis has the ability to come to apprehensive conclusions by the use of graphs.

Analysis of the usage pattern consists of the concept of collecting real usage data from the registration process. As an example, over one month of period. That real usage of data will have each user's

6 III. METHODOLOGY

The research which is discussed in this paper is a combination of improvements based on different key areas like Biometric Authentication, Neural Networks, Cyber Threat Intelligence. With the ambition of developing a Trusted Digital Identity Platform for Sri Lanka. But the idea and the essence of this research can be applied to any other domain or mobile application. Following given "Fig 1 **??** shows a high-level diagram of our system and it is followed by comprehensive description of each research component with the flow of the system.

7 a) Service Provider Management with document validator

One of the unique features of our system when comparing to existing solutions in the market is we are providing access to the service providers to register to our system online while providing the necessary documentation. Since a large number of service providers are using this feature, we have to eliminate illegitimate registration attempts to minimize wastage in resources and time of administrators.

In this research part, we are introducing a Document Validator that can identify business documents that are uploaded by customers to the system as valid business registration documents or not. In the Sri Lankan context, there are two main business registration documents namely Business Registration (BR) and Company Registration (CR). This document validator has the capability to identify the uploaded document as CR, BR, or An Image with low details or not a correct document.

8 1) Dataset and Preprocessing

We have created a dataset of 318 original images with 98 CR images 106 of BR images and 114 images which can be classified as either of these two categories. After the data augmentation using rotation, scale, shearing we collected around 1000 images. Data augmentation has been done in small amounts because the edges of the document get exempted otherwise.

9 2) Model creation and Training

The data set was divided into two parts, 25% as validation and 75% as training where 25% of the dataset is used to evaluate the model. Here we are creating a Convolutional Neural Network with a Sequential Model.

CNN architecture that has provided the best outcome contains following (Fig II **??**.

? Input layer: Loading of the input and producing an output that going to be an input for convolutional layers is carried out in this layer. We are using 375* 250 resized three channel (RGB) images as our input because The documents which we are classifying are mostly in A4 paper size. ? Convolutional layers: A set of learnable filters will be formed from the input image is the function of this layer. We have used two convolutional layers with the kernel size of 3x3 and the same padding. These two convolutional layers learn 32 filters in each one.

As the activation function we used Rectified Linear Unit (ReLU) for both of layers because given an value of z and the neuron's output is ?(z), if z>0 ?(z) = z, ifz<0 ?(z)= 0[20].

? Pooling layers: Pooling layers are responsible for downscaling the volume of the neural network by reducing small features. We have used one pooling layer after eachconvolutional layer. Both of them are max pooling layers which are set to 2 x 2 pooling windows with no strides. ? Flatten layer: Is used to flatten pooled feature map to a single column which can be fed to a fully connected layer or hidden layers. ? Hidden Layers: There were used to get the output as a single vector by inputting a single vector. In here we have used three hidden layers which first two have the ReLU as the activation function with 32 layers and 16 layers respectively while the output layer has Softmax as the activation function with three filters in it. We used the Softmax function for the multi-class classification because it scales the numbers and returns probabilities related to each class. Superviced is the trainig protocol we used for this classification. Adam is the optimizer that we have used for finding optimal model parameters which extends the functionality of Stochastic Gradient Decent.

10 b) Three Factor Biometric Authentication

Ensuring the trust of both citizens and service providers is the key priority of our system. Authentication is the process that allows both parties to verify their trust consequently, developing an authentication system in a way

that protect the trust of both parties is crucially important. To achieve this, we introduce 3 factor ? Confirm the live presence of the user using face recognition. ? Verify only a real person can authenticate to the system through liveness detection.

? Verify the authenticity of the user by using the biometrics available on the device or using pin number.

In face recognition, we have used Multi-Task Cascaded Convolutional Neural Network (MTCNN) [21] to extract faces from the video and Facenet model [16] to extract features of the face and create the face embeddings for the face recognition. Facenet is a deep convolutional neural network trained via a triplet loss function, according to this benchmark Facenet have an accuracy of 99.63% [22] compared to other similar face recognition models such as deepface model [23] by facebook with 98.37% accuracy [24] and openface model [25] with 92.92% benchmark accuracy [24], Facenet provides the highest accuracy. Another major concern in selecting Facenet method is that Facenet supports extra training data compared to other highperformance models like VarGFaceNet [26] we can train Facenet model with our own datasets to improve accuracy.

For the liveness detection we are using a combination of heuristic face movement detection and face texture analysis with a convolutional neural network CNN to differentiate real and spoofed faces. With the help of real time face contour detection in android ML kit [27] we can capture the face landmarks of the users face and predict the movement of the face. If any movement is detected, then the recorded video of the face will send the face recognition and texture-based liveness detection system hosted in the cloud. then convolutional neural network (CNN) will examine the texture of the detected face and differentiate if the detected face is real or spoof. CNN is trained using a dataset containing over10000 images containing both real and spoofed face images captured in different lightning conditions and reflections. Dataset contains images belongs to different skin colors while majority of images comprising brown skin color because we are specifically training this model to validate Sri Lankan citizens. CNN model architecture is similar to VGGNet model with less complex layers set because we need real time performance. Trust Pass is a digital identity system used in digital transformation. This system is used in the transformation of data in digital services between the citizen and the service provider. trust pass always protects users from real time threats. There are various difficulties in using traditional methods day to day life. DNA and fingerprints are used for the most important human identities. There are two main types of users in our system as citizen and service provider. There is the ability to authenticate the accuracy of documents such as the service provider's business registration. Therefore, the citizen will not meet fake service providers. In this process the accuracy of the documents is checked using image processing.

After authentication of user data, the security intelligence process minimizes threats. In this process, users are given a unique private key. The user's hash function is activated and the hash value is returned after the data is sent to the concealment mechanism. The hash value can be identified as the identity of each user. Introducing Deep locking malware using security intelligence. The user data is designed to not be compromised by an unauthorized person. Using deep neural networks (DNN) has deepened the locking malware mechanism. Here, cyber security intelligence is used to analyse the threat environment and minimize threat attacks. Attempts have been made to increase the accuracy of the user's documentation by using digital signatures. QR code is used to authenticate documents. The interplanetary file system is used to store system data. The hash function allows the user to retrieve stored data. **??**28] Ethereum virtual machine (EVM) has been used to power the IPFS process when using blockchain. The user can use deep unlocking to recover deeply locked data. The user's public key or recovery key is used. The security and privacy value of the user's data is always taken into consideration. User usage pattern analysis is a feature that is embedded in the system. Analyses the behaviour of users using the system. The purpose is to analyse the number of users using the system, the services received, and customer feedback. Therefore, system vulnerabilities can be identified. The overall system seeks to minimize the impact on the security and privacy of the user's real-time data. Our aim is to provide maximum security services to the user during digital transformation. It focuses on the threat environment and uses the cyber threat intelligence mechanism to minimize the threat impact on the system. Therefore, the behaviour of threats is monitored and analysed to prepare the system for future threats. Because the behaviour of threats is vulnerable, deep locking malware has been used to prevent this process from becoming a threat to threats throughout the system. Security intelligence has the potential to enhance the security and privacy of system data using the aforementioned user usage pattern analysis process conclusions. [29] The Blockchain based trusted digital identity platform is designed to ensure the privacy of the digital identity users. In this methodology the focused function is "Analyzing individual user's usage and usage pattern of digital ID by Collecting, Storing, and accessing data through digital identity management". In order to initiate the above-mentioned function, the first step is started from inputting the authenticated user's data, all authenticated user data will be gathered, and the collected authenticated data will be stored in a dataset. (The data for the dataset will be taken from the database). The next phase is to categorize the collected authenticated data relevant to the user. After gathering the relevant data, the gathered datasets will be categorized as mentioned in the high-level diagram. The analysis will mainly fall in to four types which are i). frequency of the daily usage of the user, Fig V **??**. frequented purpose of use (ex -bank, health), iii) analyzing customer feedback, and iv). identifying irregular usage of the user. Identifying the irregular users help to analyze the misuse of the identity as the user doesn't use the digital identity on regular basis.

11 d) Analysis of user behavior and usage patterns

Then, according to the categorized data in the dataset, the analysis process will initiate while analyzing, the user's usage will be identified according to the categorized data.

The categorized data will be analyzed using Arima Model [30] (Autograssive Integrated moving average). The Arima Model is also a form of Machine Learning. The analysis of user's usage will depend on the amount of the users who use the digital ID platform. The analysis of the usage patterns will be generated in to Arima Model (Statistical analysis model). Arima Model is a statistical analysis model used to predict the future values based on past values. Arima Model consists of different Models as an example Sarima and Sarimax can be defined. In here Sarimax Model of Arima Model will help to understand the data patterns and predict the analysis based on time series forecasting as the mentioned categories i). frequency of the daily usage of the user, ii). frequented purpose of use (ex -bank, health), iii) analyzing customer feedback, and iv). identifying irregular usage of the user, are predicted for duration of three months by using the dummy data in the dataset. In this Fig V here the data is stationary as the P -value is less than 0.05.

12 Figure VI: Stationary of the Data

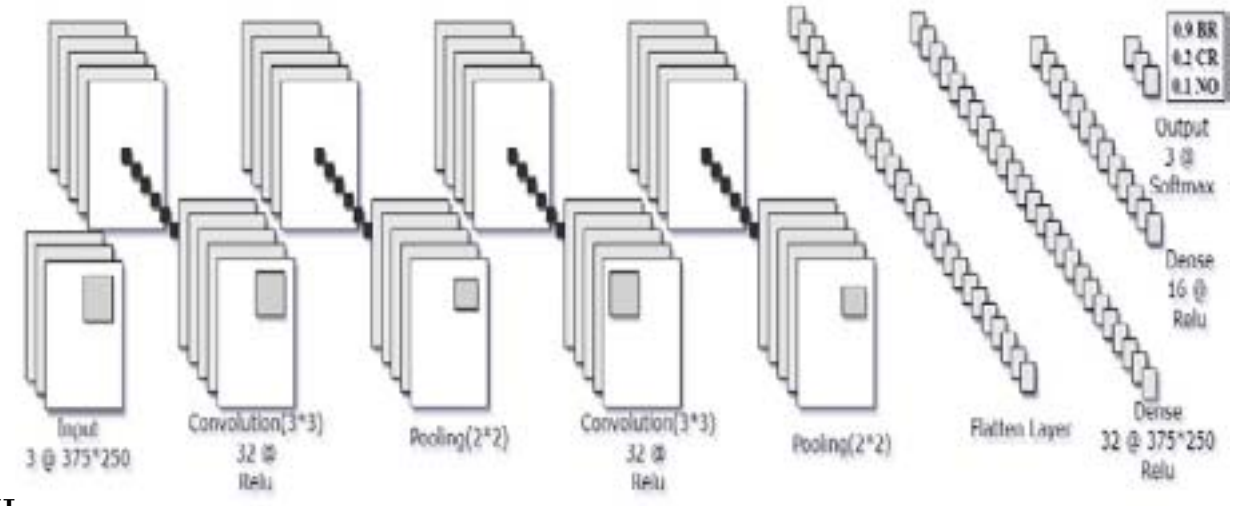
The graphs will be generated through the analysis which is done by using Arima Model; the generated graphs will help to define the analysis of the usage patterns. In this process, Security intelligence is considered to be the key to protecting a user's data and minimizing the impact of threats on privacy. Deep Locking Malware Algorithm protects user data from threats. Data security and privacy can also be enhanced by using the user's hash function and digital signature when storing data. Using Blockchain and Decentralized storage minimizes threats by storing user data. The primary purpose of these Trust Pass identity systems is

13 d) Analysis of user behaviour and usage patterns

The Arima model is used to perform the analyzation. [30] Arima model has the ability to convey the details of analyzation by adopting to time series of data as in the analyzation. It can be defined as a statistical analysis model which utilizes time series of data whether to comprehend the dataset properly or in order to predict the values of future accurately.

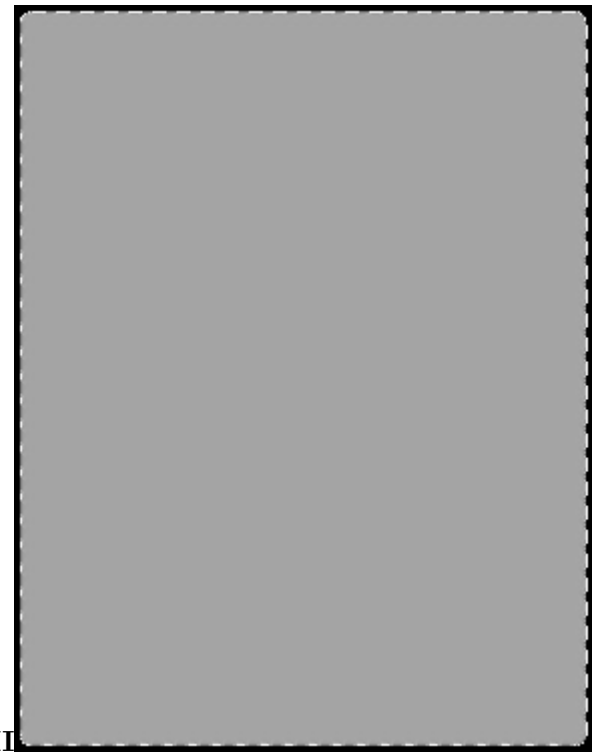
14 V. CONCLUSION

By paving thoughts toward the digital transformation of Sri Lanka, our intention is to create a decentralized digital Identity Platform that can be accessed from anywhere from any citizen that has enhanced security through cyber threat intelligence and threat intrusion analysis system. Our three-factor biometric feature extends the User side authorization to a next level on the other hand time and integrity of the Service providers would be saved with the automated document validator. All of the users will be secured using blockchain technology to minimize the threat from attackers while implementing user privacy.



II

Figure 1: Figure II :



III

Figure 2: Figure III :

V

Figure 3: Figure V :

Figure 4: ”

VI 

Figure 5: Figure VI :

IX

Figure 6: Figure IX :

VIII

Figure 7: Figure VIII :

Figure 8:

:

Figure 9: Figure X :

[Juniper Research: Mobile Biometrics] , *Juniper Research: Mobile Biometrics*

[Choudhari et al. ()] , S Choudhari , S K Das , S Parasher . 10.1109/TIFS.2014.2349158.28. *Interoperable Blockchain Solution For Digital Identity Management* 2014. p. .

[Smart and Deloitte (2021)] , " Smart , I D Deloitte , UK . https://www.deloitte.co.uk/smartid/ accessed Feb. 26, 2021.

[? Smartphone fingerprint sensor penetration worldwide 2014-2018 | Statista (2021)] *? Smartphone fingerprint sensor penetration worldwide 2014-2018 | Statista*, https://www.statista.com/statistics/804269/ accessed Aug. 19, 2021. (global-smartphone-finger print-sensor-penetration-rate/)

[Han et al. ()] 'A new image classification method using CNN transfer learning and web data augmentation'. D Han , Q Liu , W Fan . 10.1016/j.eswa.2017.11.028. *Expert Syst. Appl* 2018. 95 p. .

[Guo and Lan ()] 'A New Signature Based on Blockchain'. L Guo , C Lan . 10.1109/ICICAS51530.2020.00079. *2020 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*, 2020. p. .

[Chingovska et al. (2018)] 'Activation Functions in Neural Networks | by SAGAR SHARMA | Towards Data Science'. I Chingovska , A R Dos Anjos , S Marcel , J Stokkink , Pouwelse . 10.1109/Cybermatics_2018.2018.00230.20. https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6 *IEEE Conf. Internet Things*, 2018. 2018. 2018. Sep. 05, 2021. 9 p. . (Deployment of a Blockchain-Based Self-Sovereign Identity)

[Amofah ()] *ARIMA Model in Time Series Analysis*, L R Amofah . 2020.

[Sun et al. ()][Sun et al. ()] Authenticate $2 Trillion of Sales by 2023, Driven by Over 2,500$ 'Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification'. Y Sun , B Xue , M Zhang , G G Yen , J Lv . 10.1109/TCYB.2020.2983860. *IEEE Trans. Cybern* 2020. 50 (9) p. .

[Blockchain for Digital Identity | Accenture (2021)] *Blockchain for Digital Identity | Accenture*, https://www.accenture.com/us-en/services/blockchain/digital-identity accessed Feb. 24, 2021.

[Computer literacy statistics 2019 (annual) ()] *Computer literacy statistics 2019 (annual)*, http://www.statistics.gov.lk/ComputerLiterarcy/BuletinComputerLiteracy.pdf 2012-2015, 2019. 2019. Department of Census and and Statistics

[Taigman et al. ()] 'DeepFace: Closing the gap to human-level performance in face verification'. Y Taigman , M Yang , M Ranzato , L Wolf . 10.1109/CVPR.2014.220.24. *Labeled Faces in the Wild Benchmark (Face Verification) | Papers With Code*, 2014. p. .

[Kido et al. ()] 'Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN)'. S Kido , Y Hirano , N Hashimoto . 10.1109/IWAIT.2018.8369798. *Work. Adv. Image Technol* 2018. 2018. 2018. p. . (Int)

[Digital iD TM -ID on Your Phone -Australia Post (2021)] *Digital iD TM -ID on Your Phone -Australia Post*, https://www.digitalid.com/ accessed Feb. 24,2021.

[Digital identity as a force for good ? Yoti (2021)] *Digital identity as a force for good ? Yoti*, https://www.yoti.com/ accessed Feb. 24, 2021.

[Digital Identity Survay form (Responses) -Google Drive] *Digital Identity Survay form (Responses) -Google Drive*, https://docs.google.com/spreadsheets/u/2/d/e/2PACX-1vTXlGdWY4VCbIExzDiRWMpYhqG0-9p

[Singh and Goel] *Face Detection and Recognition S*, G Singh , A K Goel .

[Schroff et al. (2015)] 'FaceNet: A unified embedding for face recognition and clustering'. F Schroff , D Kalenichenko , J Philbin . 10.1109/CVPR.2015.7298682. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit* June. 2015. p. .

[Lee and Kwon ()] 'Going Deeper with Contextual CNN for Hyperspectral Image Classification'. H Lee , H Kwon . 10.1109/TIP.2017.2725580. *IEEE Trans. Image Process* 2017. 26 (10) p. .

[Zhang et al. ()] 'Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks'. K Zhang , Z Zhang , Z Li , Y Qiao . doi: 10.1109/ LSP.2016.2603342. 22. https://paperswithcode.com/sota/face-verification-on-labeled-faces-in-the *IEEE Signal Process. Lett* 2016. accessed Sep. 14, 2021. 23 (10) p. . (Labeled Faces in the Wild Benchmark (Face Verification) | Papers With Code)

[ML Kit | Google Developers (2021)] *ML Kit | Google Developers*, https://developers.google.com/ml-kit accessed Sep. 11, 2021.

[Baltrusaitis et al. (2016)] 'OpenFace: An open source facial behavior analysis toolkit'. T Baltrusaitis , P Robinson , L P Morency . 10.1109/WACV.2016.7477553. *IEEE Winter Conf. Appl. Comput. Vision, WACV 2016*, 2016. January 2018. 2016.

[Biggio et al. ()] 'Security evaluation of biometric authentication systems under real spoofing attacks'. B Biggio , Z Akhtar , G Fumera , G L Marcialis , F Roli . 10.1049/iet-bmt.2011.0012. *IET Biometrics* 2012. 1 (1) p. .

[SingPass Mobile (2021)] *SingPass Mobile*, https://app.singpass.gov.sg/ accessed Feb. 24, 2021.

[The Government Information Center (2021)] *The Government Information Center*, http://www.gic.gov.lk/gic/index.php/en/component/info/?id=416&task=info accessed Feb. 26, 2021.

[Tiao ()] 'Time Series: ARIMA Methods'. G C Tiao . *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, 2015. p. .

[Yan et al. ()] 'VarGFaceNet: An efficient variable group convolutional neural network for lightweight face recognition'. M Yan , M Zhao , Z Xu , Q Zhang , G Wang , Z Su . doi: 10.11 09/ICCVW.2019.00323. *Proc. -2019 Int. Conf. Comput. Vis. Work. ICCVW 2019*, (-2019 Int. Conf. Comput. Vis. Work. ICCVW 2019) 2019. p. .