KYC Using Deep Learning

Sui Mervin Maher, Department of Computer Science, REVA University, mervinjane@gmail.com and

Dr. Prasanth T, Associate Professor, REVA University, prasanth.t@reva.edu.in

Abstract— Know Your Customer (KYC) is a preventive measure used by many sectors such as bank, telecommunication, cryptocurrencies, etc. KYC allows a user to link their business information along with their Government issued ID's. This system focuses on creating a system which can automeatically verify all the details of the Government issued ID's and verify them. Tensor Flow, Keras and object detection model will help the system to detect all the objects correctly on the Government issued ID's and after that it can verify the integrity of the ID's. Face Detection system would be used to check if the image on the ID matches the person doing KYC. If any suspicious behavior is caught then a manual KYC check request will be sent. The coding for the system will be done in using Python Language and also use Flask framework. This system will allow companies to perform KYC automatically without the supervision of any human which will make the task of KYC easy and quick.

Keywords-component; KYC, Deep Learning, Face Detection

1. INTRODUCTION

KYC is a process of linking clients Government ID details with their respective accounts. The clients provides their Government ID's and those details are checked and accordingly the account is verified for KYC. Using Deep Learning, we can set up a face detection model along with object detection model to analyze the Government issued documents and to verify their integrity and compare it. Images of various dimensions, clarity, position are filtered using various models and then authentication is checked. The process of KYC is done automatically but manual authentication is provided as needed. KYC is an upcoming method of verification which can be used in various fields such as Banking, Cryptocurrencies, Government websites etc. which requires verification of Government IDs.

2. LITERATURE SURVEY

1. In [2018] Prakash Chandra Mondal etal., " Transaction Authorization from Know Your Customer (KYC) Information in Online Banking" A KYC application is proposed for authorization of information in online banking. KYC is done to authorize transactions before any financial transaction is done from the online banking application using Challenge Question i.e. CQ. Security Questions are examples of Challenge Question (CQ). Man in the middle type of attacksare not possible for this system. The use of additional hardwares made the cost of the application high.

2. In [2018] Prakash Chandra Mondal etal., " Know Your Customer (KYC) based authentication method for financial services through the internet" developed an effective system capable of authenticating services relating to finance which can also be accessed through internet. The use of Dynamic KYC is basedon MultiFactor Authentication (MFA) method which ensuresthat all the accesses over the internet for the financial services are secured. It can be used on private and public devices. This system also uses additional hardware which increases it's cost.

3. In [2019] Marc PIC etal., "Remote KYC: Attacks and Counter-Measures" They developed a system to do analysis of the user documents for securing them from different types attacks which are traditional and also the new ones. Examples of the attacks are Complete Photography Replacement, Face Swapping and Face Morphing. This system is very accurate in detecting any false matches between two images which are different.

4. In [2019] Piyush Yadav etal., "Transforming the Know Your Customer (KYC) Process using Blockchain" Distributed Ledger Technology is used as a new solution for the KYC process. This system reduces the cost of traditional KYC

process. It also saves a lot of time for the the completion of this KYC process thereby being very time efficient. Payment wallets are yet to use this type of systems.

5. In [2020] N. Sundareswaran etal., "Optimised KYC Blockchain System". Advanced Encryption Algorithm is used to build the KYC system which works on the concepts of Blockchain Technology. This system reduces the need to store the data by 20%. But the downside of this KYC blockchain system is that another encryption techniques might be more beneficial.

6. In [2020] Abdullah Al Mamun etal., "Secure and Transparent KYC for Banking System Using IPFS and Blockchain Technology". A hash value is generated when a user opens a bank account. The user is asked to complete the KYC process to generate this hash value. This is done using the InterPlanetary File System network and the sharing of this system is done using blockchain technology.As the hash value is used to open banks in multiple banks, this system makes the work of user easy and simple. Thereby this sytem is very time efficient. It also reduces the workload of the user and the cost that the user has to spend over the KYC process. However, as this system is capable of handling large amounts of data, It is not suggested to use this system for single bank accounts.

3. LITERATURE SURVEY RESULTS

After literature survey on various KYC and Deep Learning Projects, various advantages and disadvantages of all the projects are summarized. Most of the existing systems for KYC are not having automatic authentication. Because of manual verification in the existing systems for the process of KYC, the systems take a lot of time to process the data of the user and to verify it. Also they require more time for authentication as manual verification is done. This increases a lot of time and a lot of cost for the existing systems.

After literature survey we need to propose a KYC system such that it can handle automatic verification using object detection and face detection models. Also it will save time and cost as no manual authentication is required.

4. SYSTEM ARCHITECTURE

In this chapter we have listed the system architecture of KYC Using Deep Learning and the working of the KYC system.

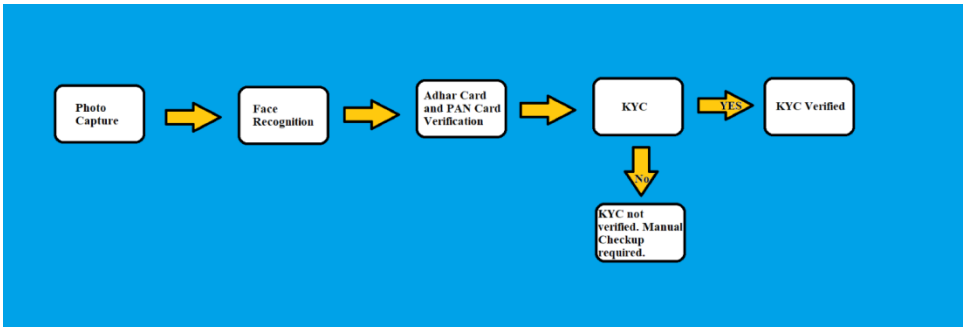

Fig1.System Architecture of KYC System

The KYC system starts with a photo capture. The user will be asked to capture their photo and then fill their Government id details i.e. Aadhar card and PAN card. Image on the Government ID will be fetched using Face Recognition model and it will be compared with the Photo captured at first. The object detection model will be used to extract the details from the Government IDs and verify it with the details manually entered by the user. Based on the verification of both the face recognition model and object detection model the system will give its results. If both the models return the value true then the KYC is verified else the KYC is not verified and manual checkup is required.

5. PROPOSED SYSTEM

In theproposed system. a KYC is used to authenticate the details of the user by comparing it with the details given on their Government IDs i.e. Aadhar card and PAN card. Different methods are used for Face Detection, Object Detection and Image Processing. All these methods are used to compare the details. If all of them return the value TRUE then KYC is verified else KYC is not verified and manual checkup is required.

6. METHODOLOGY

We are using DeepFace library for Face Recognition phase. DeepFace library provides us with many pre-trained models which can be used for image processing. For retaining the highest accuracy while also keeping in mind the same platform of programming, we use FaceNet and VGGFace models. FaceNet divides an image into 128 vectors numbers which represents most important features of a face. VGGFace has 3 stages i.e. Convolution, MaxPool and SoftMax. Convolution adds elements to neighbouring Kernels. MaxPool provides maximum values of the patches. SoftMax changes vectors of numbers to probability. In Face Recognition phase, there are 3 stages i.e.FaceAttribute, FaceRecognition and FaceVerification. FaceAttribute is used to retrieve the value of gender using the face attributes. FaceRecognition is used to identify the details of the face of the captured photo by comparing it with the photos in the image database. FaceVerification is used to verify two different images of the same person.

After finishing face detection, we are performing object detection for the Government IDs i.e. Aadhar Card and PAN Card. The data has been split between the training and the testing data and the model is trained using tensorflow. Pre-trained model SSD ResNet50 V1 FPN 640x640 (RetinaNet50) is used. This is an object detection model which allows us to create our custom object detection model. After training the data, we will use the custom model and test it on our system. First a random Aadhar card and PAN card to check if it is verified or not. After verification, the image will be skewed in order to rotate it to be properly aligned. After aligning, the image is converted into black and white so that the characters can be easily detected better. At last Object Character Recognition is performed to extract out the characters from Aadhar Card and PAN Card. If there is any error in matching the fields then the system will return FALSE else it will return the fields.

7. RESULT

In this chapter we have listed the result of the KYC System.

Result-1: Front-end

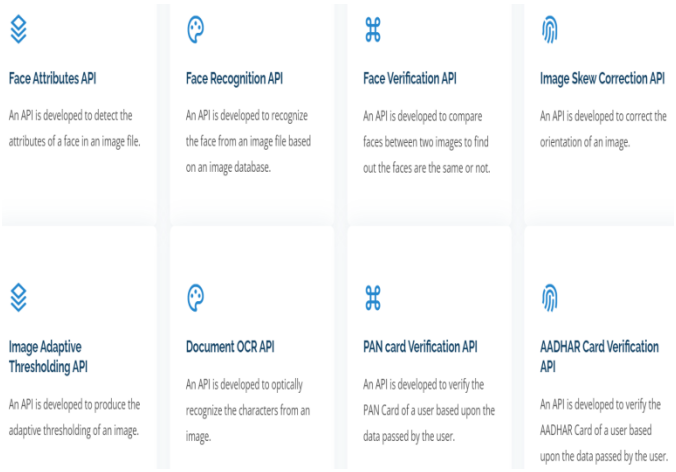

Fig 2. Front-end of KYC system

The front-end of the system is created using html, css and javascript. The KYC system includes 8 different API's which includes functionality for the entire KYC process.

Result-2: Face Attributes API

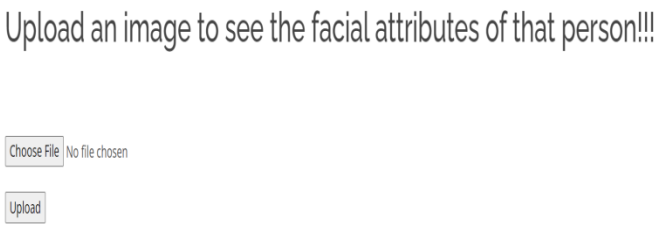

Fig 3. Face Attributes API

Face Attributes API will detect all the attributes such as age, race, gender, etc from the uploaded image which will help us get key features of the person.

Result-3: Face Recognition API

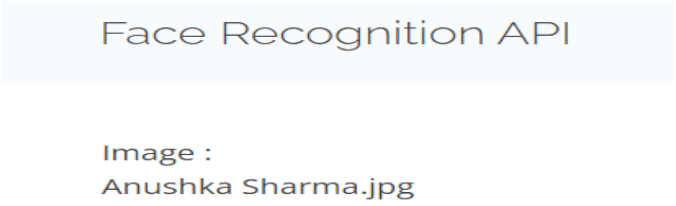

Fig 4. Face Recognition API

Face Recognition API will detect the name of the person who's image is uploaded comparing the image with other images of the same person in the database.

Result-4: Face Verification API

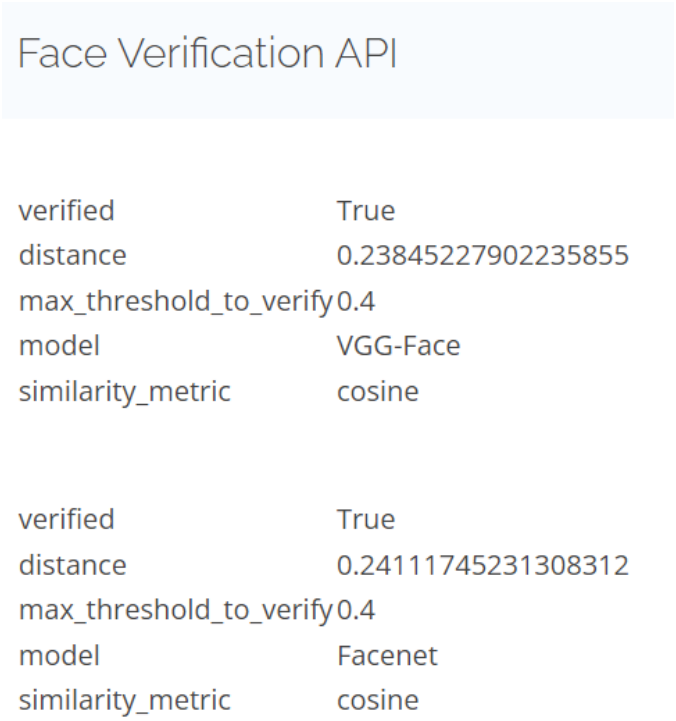

Fig 5. Face Verification API

Face Verification API compares two images and checks if both the images are of the same person or not. If both the images are of the same person then the system will return verified field as True else it will return False.

Result-5:Image Skew Correction API

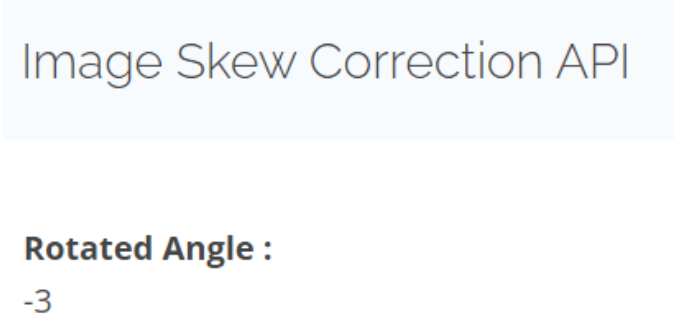

Fig 6. Image Skew Correction API

Image Skew Correction API will perform rotation of the image in such a way that the image fits perfectly on the screen which will make the work of KYC system easy. It will also display the angle by which the image is rotated.

Result-6: Image Adaptive Thresholding API

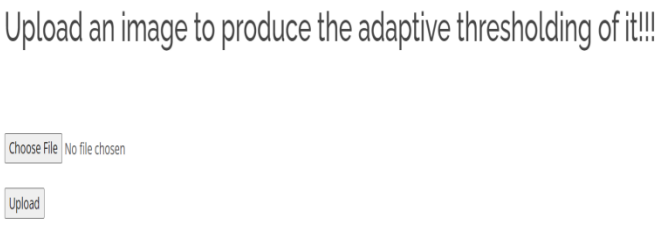

Fig 7. Image Adaptive Thresholding API

Image Adaptive Thresholding API takes the rotated image from the Image Skew Correction API and converts the image into a black and white image so that the system can retrieve the text from the image easily

Result-7: Document OCR API

Fig 8. Document OCR API

Document OCR API will extract all the text from the image which will be used to verify Government ID's for KYC.

Result-8: PAN Card Verification API

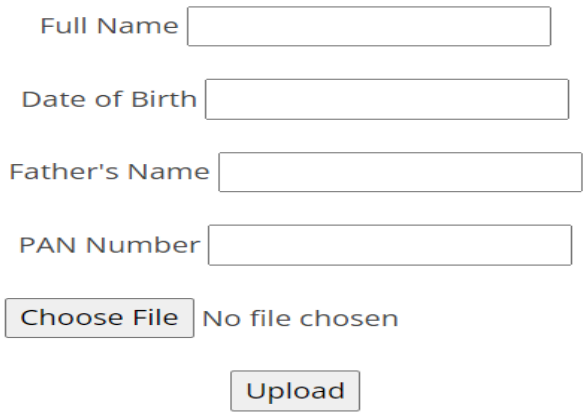

Fig 9. PAN Card Verification Form

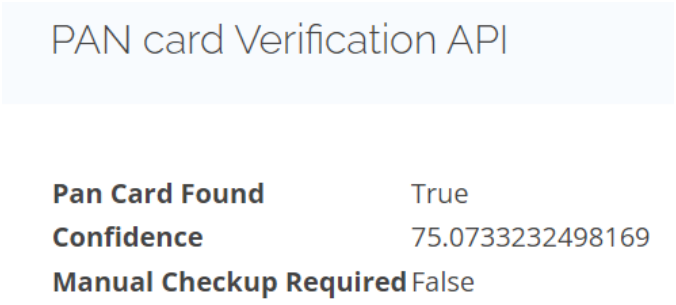

Fig 10. PAN Card Verification API

PAN Card Verification API is used to verify if the PAN Card is valid or not. The user will be asked to fill a form and upload the image of their PAN card. If the details entered by the user matches with that retrieved by the KYC system then the system would say manual checkup is not required meaning KYC verified, else manual checkup is required.

Result-9: Aadhar Card Verification API

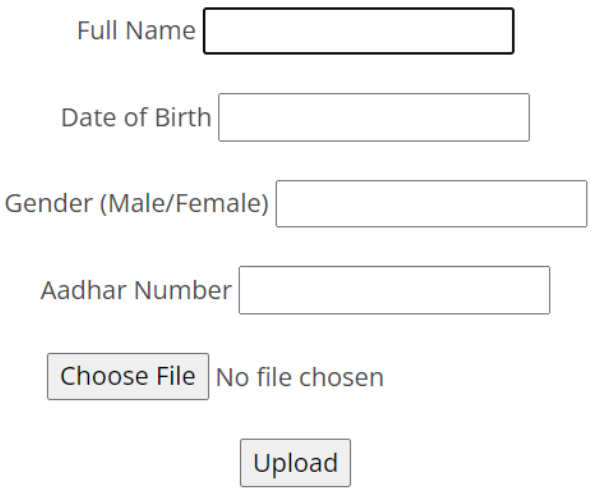

Fig 11. Aadhar Card Verification Form

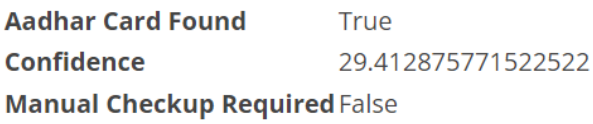

Fig 12. Aadhar Card Verification API

Aadhar Card Verification API is used to verify if the Aadhar Card is valid or not. The user will be asked to fill a form and upload the image of their Aadhar card. If the details entered by the user matches with that retrieved by the KYC system then the system would say manual checkup is not required meaning KYC verified, else manual checkup is required.

Conclusion

In the proposed system, aKYC system is created using various different Deep Learning models such as face detection, object detection, image processing, etc. This system would provide automatic authentication of the Government issued id's which could save time and cost of manual authentication though for some rare cases manual authentication would be provided if necessary.As it reduces the time for KYC, more people could benefit out of it by being able to get their documents verified faster than before which could also portray better image of the respective application or the website.

References

[1] José Parra Moyano and Omri Ross, "KYC Optimization using Distributed Ledger technology", Springer -Business & Information systems Engineering, , pp-411-423, vol.59,2018.

[2] José Parra Moyano and Omri Ross, "KYC Optimization using Distributed Ledger technology", Springer -Business & Information systems Engineering, , pp-411-423, vol.59,2018.

[3] G. Zyskind, O. Nathan and A. Pentland, "Decentralizing Privacy: Using Blockchain to Protect Personal Data", IEEE Security and Privacy Workshops, San Jose, CA, 2015.

[4] Mauro Isaja and John Soldatos, Distributed ledger technology for decentralization of manufacturing processes,IEEE Conference on Industrial Cyber-Physical Systems (ICPS) 2018 at, St. Petersburg, Russia, pp 696- 701,2018..

[5] Rui Yuan, Yu-Bin Xia, Hai-Bo Chen, Bin-Yu Zang, Jan Xie and ShadowEth: Private Smart Contract on Public Blockchain, Springer, Journal of Computer Science and Technology, Issue 3, pp 542–556. vol 33,2018.

[6] XiaoqiLi ,PengJiang, ,XiapuLuo,TingChen and QiaoyanWen, "A survey on the security of blockchain systems", Elsevier Future Generation Computer systems", ,Aug 2017

[7] Sein Myung, Jong and Hyouk Lee, Ethereum smart contract-based automated power trading algorithm in a microgrid environment, Springer Journal of Super computing, PP 1-11,2018

[8] PetarMaymounkov and David Mazieres, "Kademlia: A peer-topeer information system based on the xor metric", Business & Information Systems Engineering Journal, 2002

[9] Basha, S. M., Poluru, R. K., & Ahmed, S. T. (2022, April). A Comprehensive Study on Learning Strategies of Optimization Algorithms and its Applications. In *2022 8th International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-4). IEEE.

[10] Shbair, Wazen& Steichen, Mathis & François, Jérôme and State, Radu"Blockchain Orchestration and Experimentation Framework: A Case Study of KYC", IEEE/IFIP Network Operations and Management Symposium,

[11] Liehuang Zhu, Yulu Wu, Keke Gai, Kim-Kwang and Raymond Choo "Controllable and trustworthy blockchain-based cloud data management", Elsevier, Future Generation Computer Systems, pp. 527-535, vol 91, 2019

[12] PaulJ.Taylor,TooskaDargahi,AliDehghantanha,Reza,M.Parizi,Kim Kwang and RaymondChoo, ,A systematic literature review of blockchain cyber security.,DigitalCommunication and networks,", Elsevier Feb 2019

[13] WjatscheslavBaumung, and VladislavFomin, Framework for enabling order management process in a decentralized production network based on the blockchain-technology, Elsevier, Procedia CIRP, PP 456-460, vol 79, 2019