PROJECT REPORT

ON

“WEBSITE FOR LOST AND FOUND”

Department of Computer Science Engineering
Netaji Subhas University of Technology

Submitted in partial fulfilment of the requirements of the degree

of

Bachelor of Engineering/Master of Technology
in Computer Science Engineering

Under the guidance of:
Dr. Shampa Chakraverty
Ms. Kanika Bhatia

By:
Dipanshu Kumar (2019PCS2004)
Aniket Tomar (2019PCS2017)
Vaibhav (2017UCO1620)
Hemant Vashist (2017UCO1657)

ACKNOWLEDGEMENT

The present work will remain incomplete unless we express our feelings of gratitude towards a number of persons who delightfully co-operated with us in the process of this work.

This project would not have taken shape, without the guidance provided by our Soft Computing Professor, Dr. Shampa Chakraverty , Department of Computer Science & Engineering who helped in our project and resolved all the technical as well as other problems related to the project and, for always providing us with a helping hand whenever we faced any bottlenecks, in spite of being quite busy with her hectic schedules. It was her mentorship that encouraged us to expedite our project process and could complete it in time. Her precious suggestions and constructive guidance has been indispensable in the completion of this project work.

Dipanshu Kumar (2019PCS2004)
Aniket Tomar (2019PCS2017)
Vaibhav (2017UCO1620)
Hemant Vashist (2017UCO1657)

TABLE OF CONTENTS

PROBLEM STATEMENT

Problem title: “Lost and Found”

Design a website which uses biometric datasets for faces and fingerprints which can be used for identification of individuals.

Aadhar number can be used as the basis for identification of individuals.
In case of victims such as minors, heavily – injured, traumatized, their biometric traits can be used to fetch the records, so as to locate their permanent residence and/or contact their guardians.

STUDENTS INVOLVED

Dipanshu Kumar (2019PCS2004)

Aniket Tomar (2019PCS2017)

Vaibhav (2017UCO1620)

Hemant Vashist (2017UCO1657)

Division of Work

Hemant (2017UCO1657)

- Data collection, generation and preprocessing
- Deciding a strategy to extract features from biometric data
- Using these features in an algorithm to map them to demographic details of the users and fetching these details.
- Building an API to be used in the website.
- Integration of the API with the client code.

Vaibhav (2017UCO1620)

- Website layout and server integration
- Layout and pages in the website from a user's perspective
- Server
- Integration with Firestore Storage
- Integration of flask

Dipanshu (2019PCS2004)

- Team management
- JS
- Data collection and preprocessing
- HTML
- Design layout
- Connect the website to firebase cloud

Aniket (2019PCS2017)

- Data collection
- Firebase for firestore storage
- Flask for backend
- Testing
- website layout

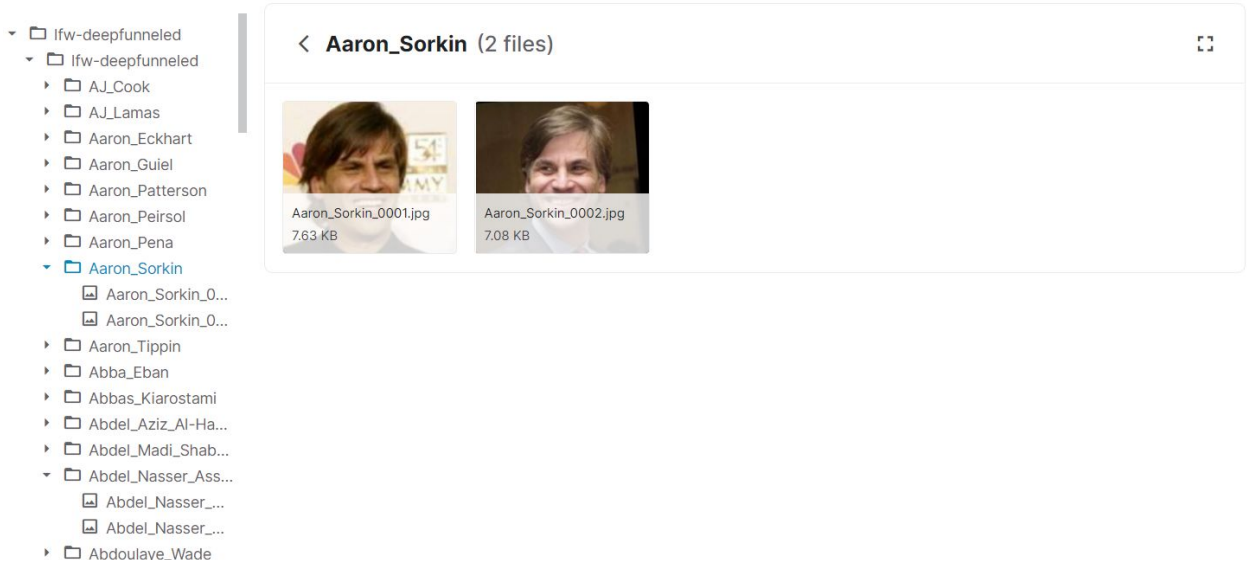
PROGRESS ACHIEVED

- Data collection, generation and preprocessing
- Working model on the data obtained to get the desired result
- Full website design with a user interface and integration with an API
- API which interacts with the data repository and the ML models and systems
- Improve model accuracy by tweaking model parameters and re-annotating the dataset.
- Learnt about Genetic Algorithms and when can a problem be modelled to it

DATASETS USED

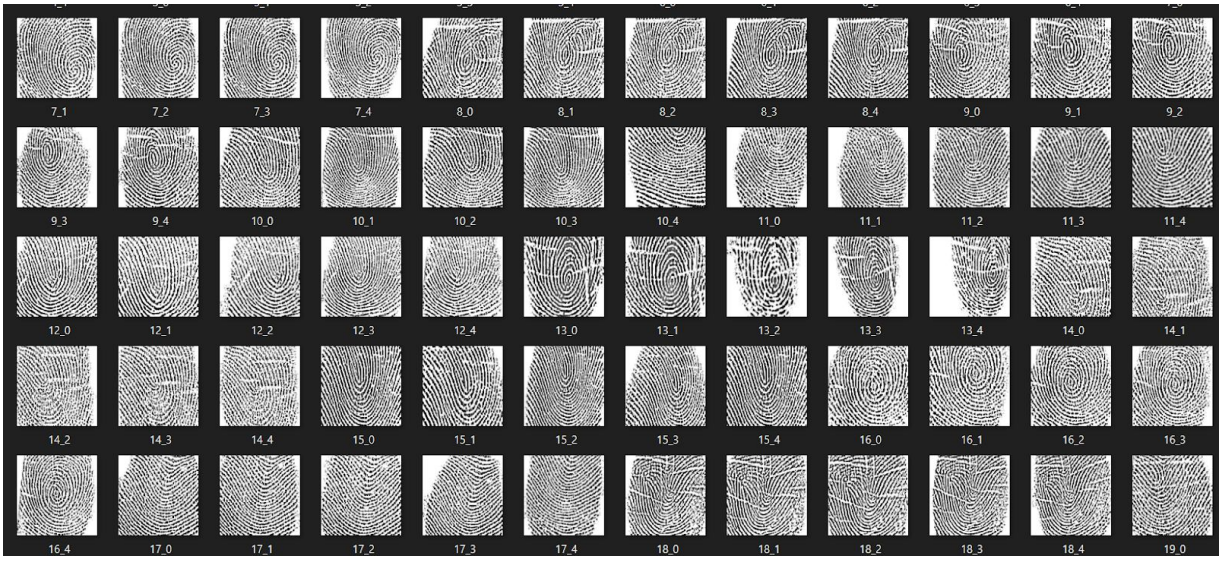
- **Faces dataset**

  We used Labelled Faces in the Wild(LFW) dataset available on Kaggle. It contains over 13,000 face images of about 5700 people. However, we used only 50 samples(people) due to computational constraints.



- **Fingerprint dataset**

  The dataset used was [Fingerprint Color Image Database] on Mathworks. It contains five fingerprint samples each (for each finger) of 50 subjects.



- **Demographic dataset**

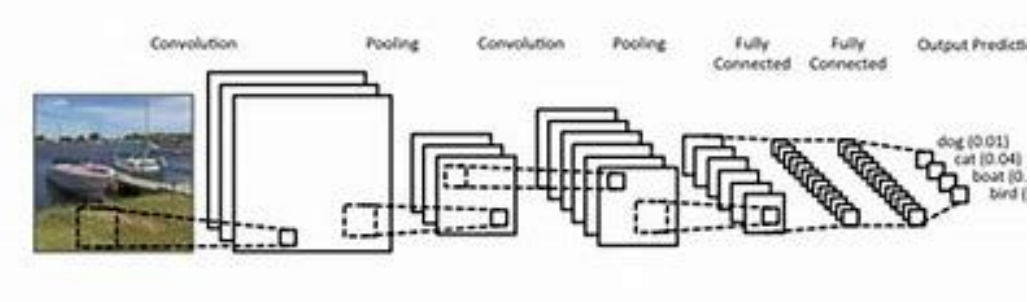
  The dataset was generated by combining different fields from different sources. The name was taken from the LFW dataset while other details were generated by scraping random addresses, 12-digit AADHAAR numbers and phone numbers. https://docs.google.com/spreadsheets/d/1OY--nKqRqpaivzPzNAcFU7vyha1_o2iQAgKo6-C3VtA/edit#gid=0

	A	B	C	D	E	F	G	H
1	Index	FirstName	LastName	DOB	Gender	PhoneNumber	AadhaarNumber	Address
2	0	Tommy	Franks	26-01-1989	M	9855544513	990818051631	State Bank of India-Tis Hazari Court, Lala Hardev Sahai Marg, Civil Lines, New Delhi, North, Delhi, 110054, IND
3	1	Halle	Berry	13-07-1997	F	9155568538	993813166289	District Court Tis Hazari Gate No 4, Khanna Market Road, Civil Lines, New Delhi, North, Delhi, 110054, IND
4	2	Saurav	Ganguly	13-08-2012	M	8555005776	989756849639	Tis Hazari Metro Station, Lala Hardev Sahai Marg, Civil Lines, New Delhi, North, Delhi, 110054, IND
5	3	Derek	Jeter	13-08-2012	M	7555860946	992104337296	Cafe Buddy's, Lala Hardev Sahai Marg, Civil Lines, New Delhi, North, Delhi, 110054, IND
6	4	Pedro	Malan	29-11-1991	M	7555436894	997056779137	Tis Hazari Metro Station Gate No 2, Lala Hardev Sahai Marg, Civil Lines, New Delhi, North, Delhi, 110054, IND
7	5	Hitomi	Soga	20-11-1990	F	9755536429	994008797686	Tis Hazari Metro Station, Lala Hardev Sahai Marg, Civil Lines, New Delhi, North, Delhi, 110054, IND
8	6	Geoff	Hoon	28-05-1988	M	7555356145	992507537373	Tis Hazari Metro Station Gate No 1, Lala Hardev Sahai Marg, Civil Lines, New Delhi, North, Delhi, 110054, IND
9	7	Sophia	Loren	26-01-1995	F	7555524829	993544070017	St Stephen's Hospital-ER, St Stephen's Hospital Road, Civil Lines, New Delhi, North, Delhi, 110054, IND
10	8	Li	Zhaoxing	24-06-1999	M	8555304167	996784834265	St Stephen's Hospital, St Stephen's Hospital Road, Civil Lines, New Delhi, North, Delhi, 110054, IND
11	9	Sheila	Copps	05-05-1985	F	7555225465	994004756782	St Mark's Cathedral, Shivajinagar, Bengaluru, Karnataka, 560001, IND
12	10	Amer	al-Saadi	24-08-1991	M	9655530151	995164764058	Gallery Blue Spade, Ashoka Nagar, Bengaluru, Karnataka, 560001, IND
13	11	Richard	Armitage	14-04-1997	M	8555087956	997298705384	Mahatma Gandhi Park-Gate 1, Shivajinagar, Bengaluru, Karnataka, 560001, IND
14	12	Zinedine	Zidane	20-01-1990	M	8555283106	993423541415	Resorts in Bangalore, Mahathma Gandhi Road, Shivajinagar, Bengaluru, Karnataka, 560001, IND
15	13	Martha	Stewart	21-09-1996	F	9855558285	992435586615	Hotel Jobs at Bangalore, Mahathma Gandhi Road, Shivajinagar, Bengaluru, Karnataka, 560001, IND
16	14	Katherine	Harris	15-11-1965	F	8555806961	996151513452	ST. Mark's Cathedral, Shivajinagar, Bengaluru, Karnataka, 560001, IND
17	15	Monica	Seles	29-03-1992	F	9455545659	992747090827	Mahatma Gandhi Statue, NH-44, Shivajinagar, Bengaluru, Karnataka, 560001, IND
18	16	Mohammad	Khatami	14-11-1999	M	7555706048	995208095638	Victoria, Cubbon Park Road, Sampangi Rama Nagar, Bengaluru, Karnataka, 560001, IND
19	17	Carmen	Electra	21-11-1995	F	9055520875	995272779841	St Mark's Cathedral, Mahathma Gandhi Road, Shivajinagar, Bengaluru, Karnataka, 560001, IND
20	18	Ana	Palacio	18-11-1998	F	7555998300	987781528800	Kotak Mahindra Bank-Sri M L Subbaraju Road, Lavelle Road, Shivajinagar, Bengaluru, Karnataka, 560001, IND
21	19	Keanu	Reeves	24-04-1986	M	8555171876	988569832178	Shree Govindam Optical and Departmental, Station Road, Chandpole, Jaipur, Rajasthan, 302001, IND
22	20	Alvaro	Uribe	09-11-1996	M	8555619183	989185200740	Khandelwal Computers, Station Road, Chandpole, Jaipur, Rajasthan, 302001, IND
23	21	Carlos	Moya	09-12-1996	M	7555031278	993063086259	Jai Govind General Store, Station Road, Chandpole, Jaipur, Rajasthan, 302001, IND
24	22	Gerhard	Schroeder	09-03-1996	M	7085619682	661119915959	No. 3, Ground Floor, Part 1 Market,Greater Kailash I, Greater Kailash, Delhi - 110048
25	23	Luis	Figo	09-01-1996	M	8065482991	459115941591	5/A Gulab Estate, Safed Pool, Sakinaka, T2 Link Road, Andheri Kurla Road, Mumbai, Maharashtra, 400072
26	24	Art	Howe	10-05-1996	M	9485239053	554412548648	No. 51/11, 7th Avenue, Soundariya Colony, Anna Nagar West (Extension), Chennai. 600101

Soft Computing Technique Employed

Convolutional Neural Networks

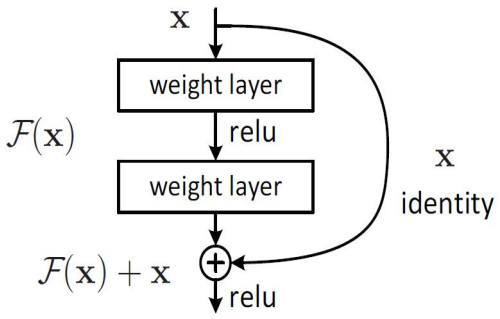
A Convolutional Neural Network (CNN) is composed of one or more convolutional layers (often with a subsampling step) and then followed by one or more fully connected layers as in a standard multilayer neural network. The architecture of a CNN is designed to take advantage of the 2D structure of an input image (or other 2D input such as a speech signal). This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features.



Approach Towards Problem

The objective of the project is to uniquely identify individuals and return their demographic details (Gender, Age, Name etc.) on the basis of their biometric input i.e. fingerprints and face images.

This is achieved by using Single Invariant Feature Transform to store the minutiae points of the fingerprint samples present in the repository as key points and get their descriptors for calculating difference scores and generating keypoint visualizations.



The face recognition system is composed of a CNN which captures the data points present in the face in the convolutional layers of the network through residual learning.

Proposed Methodology

Fingerprint Recognition

The scale-invariant feature transform (SIFT) is a feature detection algorithm in computer vision to detect and describe local features in images. It introduces invariance to image scale and rotation for while extracting features from it.

We used ORB (Oriented Fast and Rotated BRIEF) which is a library in OpenCV to implement SIFT in Python.

We also used Harris Corner Detection which is a corner detection operator that is used in computer vision algorithms to extract corners and infer features of an image.

Pseudocode for fingerprint recognition:

FOR EACH SAMPLE FINGERPRINT IN THE DATA REPOSITORY:

1. Skeletonize the image to a binary pixel matrix.
2. Remove stray dots(noise) from the image and perform Harris Corner Detection to get minutiae points in the image.
3. These minutiae points are the Keypoints to be used in SIFT and their descriptors are calculated using ORB.
4. These keypoints and descriptors are then fed into a cost function which calculates difference_score between the test image and the sample images. This uses Brute Force Feature Matcher which calculates cost based on the distance, illumination difference and orientation difference of the keypoints.
5. Return average score of the difference_scores in step 4.
6. If average_score < THRESHOLD, the image is classified as a match.

Face Recognition

We have used a pretrained model for this task which is a ResNet network with 29 conv layers. It's essentially a version of the ResNet-34 network from the paper Deep Residual Learning for Image Recognition by He, Zhang, Ren, and Sun with a few layers removed and the number of filters per layer reduced by half.

The network was trained from scratch on a dataset of about 3 million faces. This dataset is derived from a number of datasets. The face scrub dataset (http://vintage.winklerbros.net/facescrub.html), the VGG dataset (http://www.robots.ox.ac.uk/~vgg/data/vgg_face/), and then a large number of images that were scraped from the internet. The dataset was cleaned and preprocessed by removing labeling errors, which meant filtering out a lot of parts in the images from VGG. This was done by repeatedly training a face recognition CNN and then using graph clustering methods and a lot of manual review to clean up the dataset. In the end about half the images are from VGG and face scrub. Also, the total number of individual identities in the dataset is 7485.

The network training started with randomly initialized weights and used a structured metric loss that tries to project all the identities into non-overlapping balls of radius 0.6. The loss is basically a type of pair-wise hinge loss that runs over all pairs in a mini-batch and includes hard-negative mining at the mini-batch level

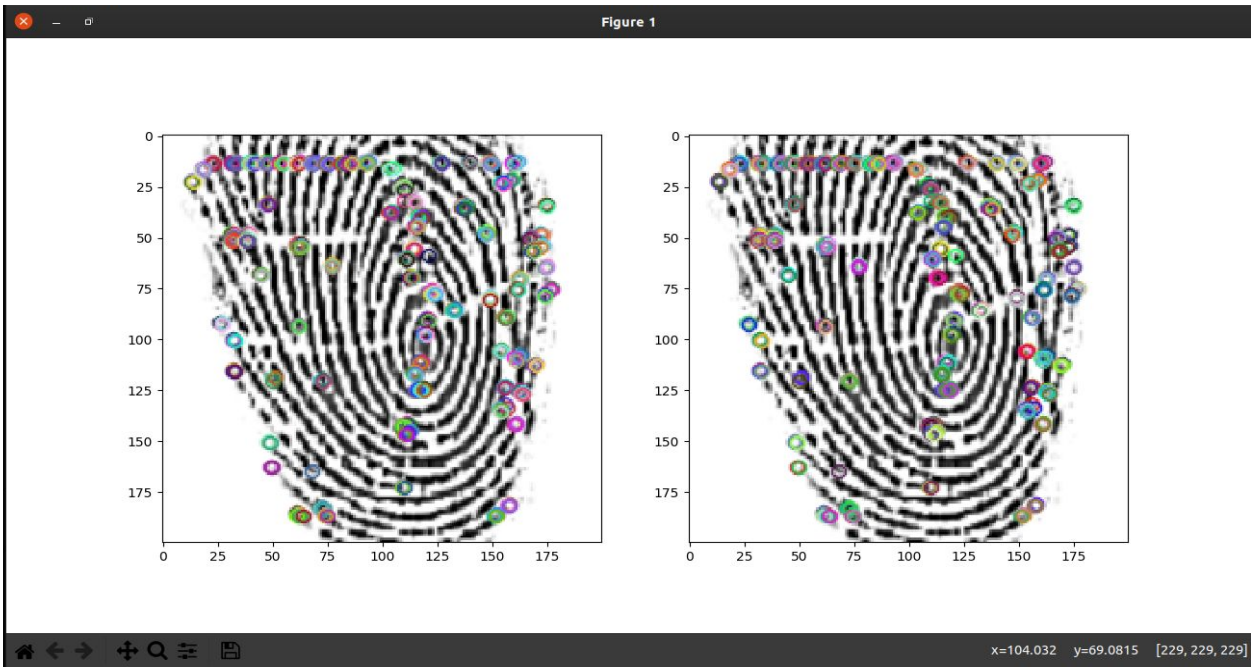
Working of the models

Fingerprint Recognition

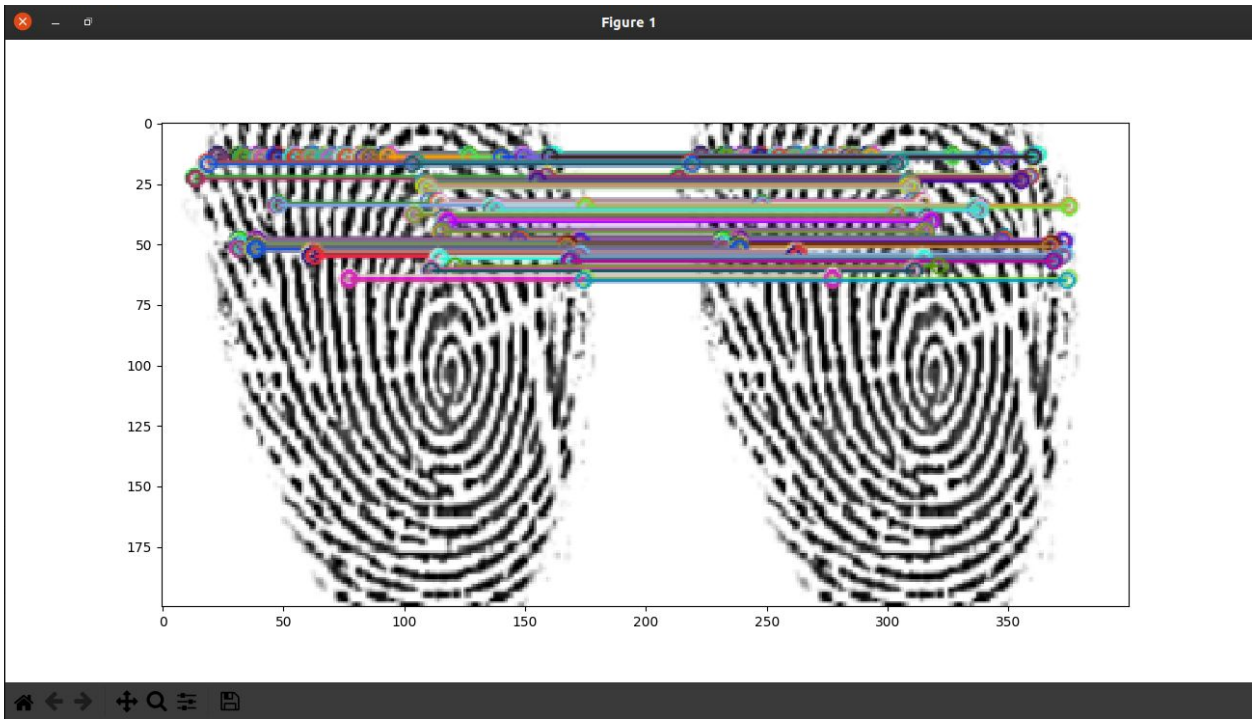
The fingerprint image of a ring finger selected for testing

```
127.0.0.1 - - [14/Aug/2020 15:15:03] "POST /fingerprint HTTP/1.1" 200 -
Image path: fingerprint_test/13_3.jpg
Subject number: 0                Score:  55.857142857142854
Subject number: 1                Score:  56.4
Subject number: 2                Score:  58.53658536585366
Subject number: 3                Score:  51.73170731707317
Subject number: 4                Score:  53.67857142857143
Subject number: 5                Score:  48.47826086956522
Subject number: 6                Score:  51.74074074074074
Subject number: 7                Score:  55.607142857142854
Subject number: 8                Score:  56.31707317073171
Subject number: 9                Score:  53.888888888888886
Subject number: 10               Score:  58.03448275862069
Subject number: 11               Score:  63.26315789473684
Subject number: 12               Score:  56.291666666666664
Subject number: 13               Score: 20.0
127.0.0.1 - - [14/Aug/2020 15:17:42] "POST /fingerprint HTTP/1.1" 200 -
```

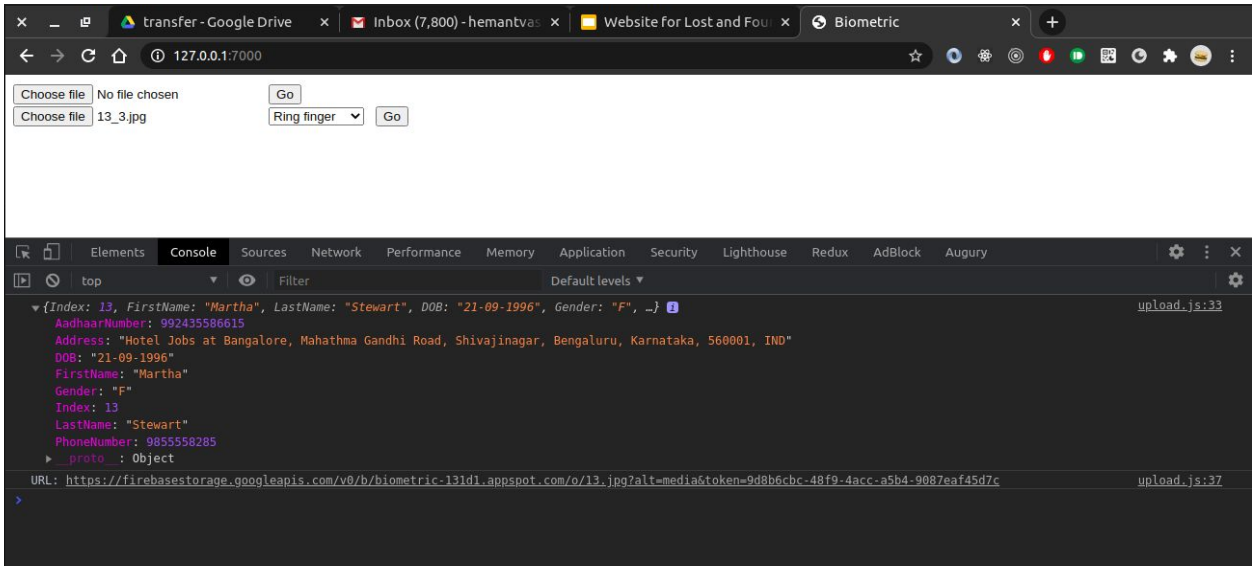
Each image in the repository is matched with the test finger and their corresponding scores are returned. The first sample with score < THRESHOLD is a match.



The minutiae points detected in test image and the matched image.



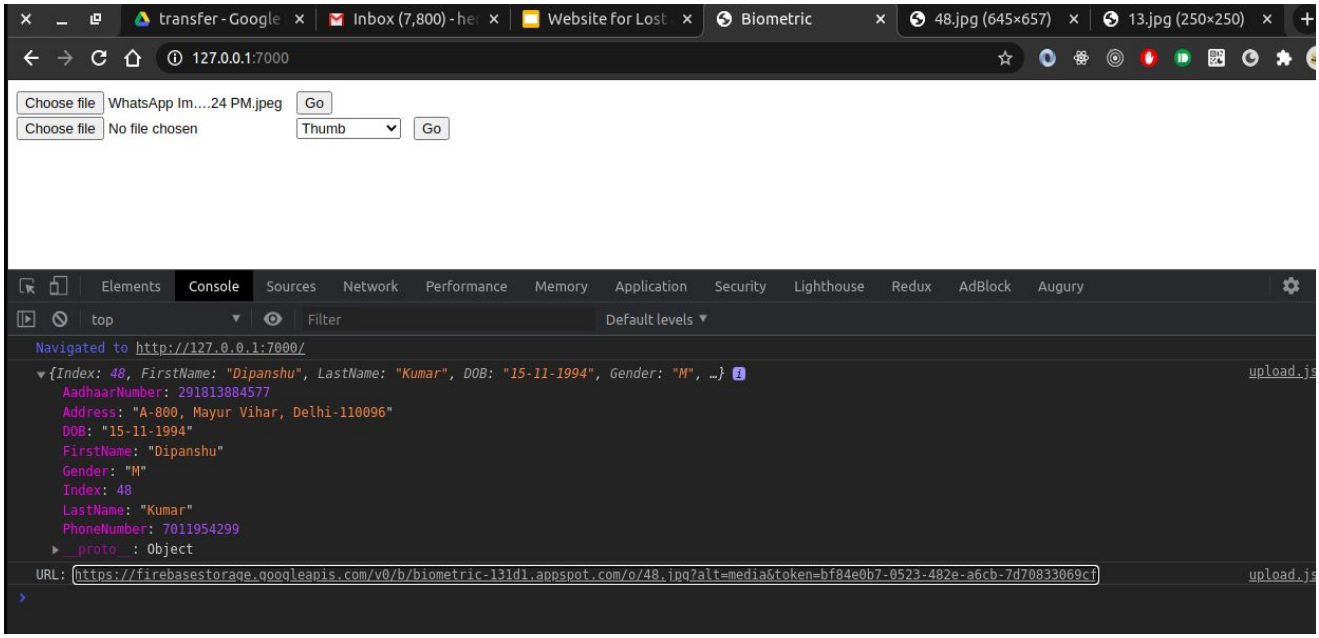
The mapping of the minutiae points which actually matched in the test image(left) and the matched image(right).



The details of the individual returned by the model with the url to their face image.

Face recognition

Test face image used for the identification of an individual



The details of the individual returned along with a url to their original face image.

PERFORMANCE

Metrics (out of 50 test samples)	Fingerprint	Face
Accuracy	93.33 %	95.56 %
Precision	83.33 %	83.33 %
Recall	71.43 %	83.33 %
F1 Score	76.92	83.33

FUTURE SCOPE

1. This project can be used as a base for realizing a scheme to be implemented in other projects of greater level such as helping people during any disaster management so that more and more people can come forward to volunteer for help.

2. It can be used for extensibility in various ways, that is we can add more features in the current project so that people can easily find required information in one place only.

3. The project currently uses a dataset of 250 samples (50 individuals with 5 fingerprint and 1 face picture each). The model can be adapted to more real world scenarios by introducing more varied fingerprint and face data.

4. The fingerprint model is computationally slow. It could be optimised by storing the keypoints in a data structure already before the user uploads a fingerprint image.

REFERENCES

1. https://datasetsearch.research.google.com/
2. https://ieeexplore.ieee.org/abstract/document/8322288/
3. https://ieeexplore.ieee.org/abstract/document/8308186
4. https://pypi.org/project/numpy/
5. https://pandas.pydata.org/
6. https://matplotlib.org/
7. https://www.kaggle.com/jessicali9530/lfw-dataset
8. https://miro.medium.com/max/600/0*wGPpgnPImtwLb8NX.png
9. Pre test and post test: https://docs.google.com/spreadsheets/d/1NyAHSQPTO7_cOK7FipZlXzJEfRBvB6X5MPyIeVUy2v8/edit?usp=sharing
10. Individuals' dataset generated: https://docs.google.com/spreadsheets/d/1OY--nKqRqpaivzPzNAcFU7vyha1_o2iQAgKo6-C3VtA/edit?usp=sharing

CHALLENGES FACED

1. A lot of unstructured data.
2. Too much internship load
3. Not able to communicate properly due to lockdown
4. Different availability hours of team members
5. Too much load of assignments from college

LEARNINGS FROM EACH OTHER

2017UCO1657

- Learnt web scrapping , work organisation, and selenium from 2017PCS2004
- Learnt algorithmic design and problem solving from 2017UCO1620
- Learnt about punctuality and data collection from 2019PCS2017

2019PCS2004

- Learnt python, flask and firebase from 2017UCO1657
- Learnt web development and JS from 2017UCO1620
- Learnt about punctuality and data collection from 2019PCS2017

2017UCO1620

- Learnt firebase, flask and vanillaJS from 2017UCO1657. Also got an idea about the techniques to recognize biometrics.
- Learnt web scraping , work organisation and selenium from 2019PCS2004
- Learnt about punctuality and data collection from 2019PCS2017

2019PCS2017

- Learnt biometric recognition, python,flask, and firebase from 2017UCO1657
- Learnt web development and algorithmic design from 2017UCO1620
- Learnt web scraping , work organisation and selenium from 2019PCS2004

SKILLS LEARNT WHILE PURSUING PROJECT

2017UCO1657 - flask,firebase,biometric recognition, matplotlib,genetic algorithm

2017UCO1620 - web development,flask,vanillaJS,biometric recognition

2019PCS2004 - biometric recognition,flask, web development,apache, nodeJS

2019PCS2017 - biometric recognition,flask,python,JS,selenium, nodeJS