e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

FACIAL FEATURE EXTRACTION FOR IDENTICAL TWINS FRAUD DETECTION USING MACHINE LEARNING ALGORITHMS

Dr. K. Suresh[*1], K. Chandusha[*2], Duvvala Suchideeksha[*3], Banavath Bharathi[*4], Avanigadda Karthik[*5]

[*1]Associate Professor Department Of Information Technology Malla Reddy College Of Engineering & Technology Hyderabad, India.

[*2]Assistant Professor, Department Of Computer Science Malla Reddy College Of Engineering & Technology Hyderabad, India.

[*3,4,5]Department Of Information Technology Malla Reddy College Of Engineering & Technology Hyderabad, India.

ABSTRACT

Identical twins, sharing nearly identical genetic makeup, present a unique challenge for traditional identification methods. This study explores the potential of machine learning (ML) algorithms to predict the presence of identical twins based on various biometric and genetic data. The research leverages a diverse dataset comprising genetic information, facial features, voice patterns, and other physiological attributes collected from a wide range of individuals, including identical twins and nontwin controls. We employ a multifaceted ML approach, including deep learning techniques, to analyse and extract distinctive patterns from the dataset. Feature selection and extraction methods are utilized to identify the most discriminative attributes for twin prediction.

In real words twins faces are exists and this twin s can utilize advantages to dupe people in examination or any other organizations. To detect such twins we are applying machine learning algorithm such as Naïve Bayes and Random forest which may get trained on possible Real and Twins faces. Additionally, this research investigates the transferability of models across different populations and explores the ethical considerations surrounding privacy and consent when dealing with sensitive genetic and biometric data.

I. INTRODUCTION

The main concept of our work is to develop an identical twin fraud detection approach that increases the accuracy of classification and identification using a class of naïve Bayes and Random Forest. Our method is based on the observation that twins have similar facial features. Trained with random forest to learn the facial features. A general face recognition model is developed based on a system comprising several features or attributes. Face detection, feature extraction, and face recognition are the key steps in any twin detection process of twins. Then trained with naïve Bayes and Random Forest to extract the facial features of an input image. It is applicable to all image data sets.

In our project, we have used algorithms such as Random Forest and Naïve Bayes in terms of accuracy The scope of a project focused on facial feature extraction for identical twins fraud detection using machine learning algorithms is broad and can have significant implications in various domains. Here are some aspects of the project's scope. In Identity Verification and Fraud Prevention the primary scope is to develop a system that enhances identity verification processes, particularly in scenarios where fraud detection is critical. Applications include online identity verification, secure access control systems, and preventing fraudulent activities such as account takeover. In Security systems the integration into security systems for facilities, airports, and other high-security areas where accurate identification is essential. We employ a multifaceted ML approach, including deep learning techniques, to analyse and extract distinctive patterns from the dataset. Feature selection and extraction methods are utilized to identify the most discriminative attributes for twin prediction.

Enhancing the reliability of facial recognition systems used for surveillance and security. In forensic science identifying individuals in forensic investigations, especially in cases involving twins. Differentiating between identical twins for forensic purposes. In Education and Awareness, scope for educating users, organizations, and the public about the capabilities, limitations, and ethical considerations of facial recognition technology. Raising awareness about the importance of responsible data hand ling and privacy safeguards.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

II. LITERATURE REVIEW

Literature Review: Biometric Data and Twin Recognition Abstract: This literature review explores the intersection of biometric data and twin recognition in the context of machine learning. It delves into the various biometric modalities, such as facial recognition, fingerprinting, and iris scans, and their potential utility in identifying identical twins. The review also investigates existing challenges and limitations in utilizing biometric data for twin recognition and provides insights into how machine learning can enhance the accuracy and reliability of twin identification.

Literature Review: Twin Studies and Genetics in Identical Twin Identification. Abstract: This literature review focuses on the genetic underpinnings of identical twin identification. It examines twin studies, genetic markers, and the heritability of physical traits in monozygotic twins. By analyzing the existing research, this review offers a comprehensive overview of the unique genetic factors that can be leveraged for accurate twin recognition, particularly through machine learning approaches.

Literature Review: Ethics and Privacy in Identical Twin Identification Using ML

Abstract: This literature review explores the ethical and privacy considerations associated with employing machine learning for identical twin identification. It investigates the potential risks of biometric data collection and the implications for individuals' privacy and consent. The review examines current ethical guidelines, legal frameworks, and best practices in twin identification to ensure that machine learning models are developed and applied responsibly and with utmost regard for ethical standards.

III. METHODOLOGY

The identical twin's facial recognition system has two main module designs: 1. Training Module Design

1. Testing Module

Design Firstly, the training module uses a local dataset as an input to the system for training the model. Upon selecting the dataset, the image processing phase is entered where the images of twins in the dataset are extracted and images are resized to 60*60 to just limit our system to get more clear view of just the face. There by remove all the unnecessary noise from the images. The next step is to convert the data to compatible format in order to allow lower layers of the algorithm to use this data as an input.Following which the normalizing of the data is done. The features are extracted using the random forest module as the model consists of multiple layers which will try to capture the important features excluding the noise and the negative values. In real words twins faces are exists and this twins can utilize advantages to dupe peoples in examination or any other organizations. To detect such twins we are applying machine learning algorithms such as Naïve Bayes and Random Forest which may get trained on possible Real and Twins faces. Once after training we can input face to this trained model to identify weather face is Real or Twin. Before training we are applying various image processing techniques such as applying Bilateral Filters to enhance image quality and then convert image to Black & White format and then apply Object detection technique to detect face from image. This processed image will be input to Machine learning algorithm to train a model.

For training we are using below images dataset

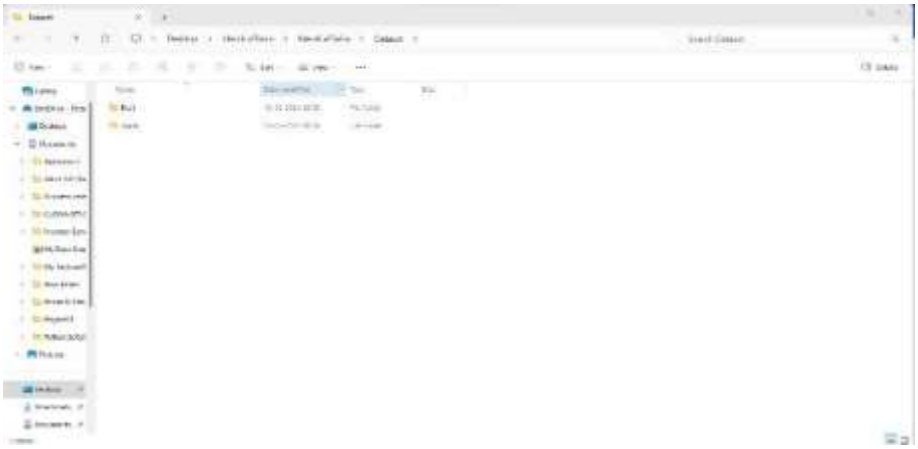

Figure 1:

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

In above screen we have two folders called Twins and Real and just go inside any folder to view images like below screen.

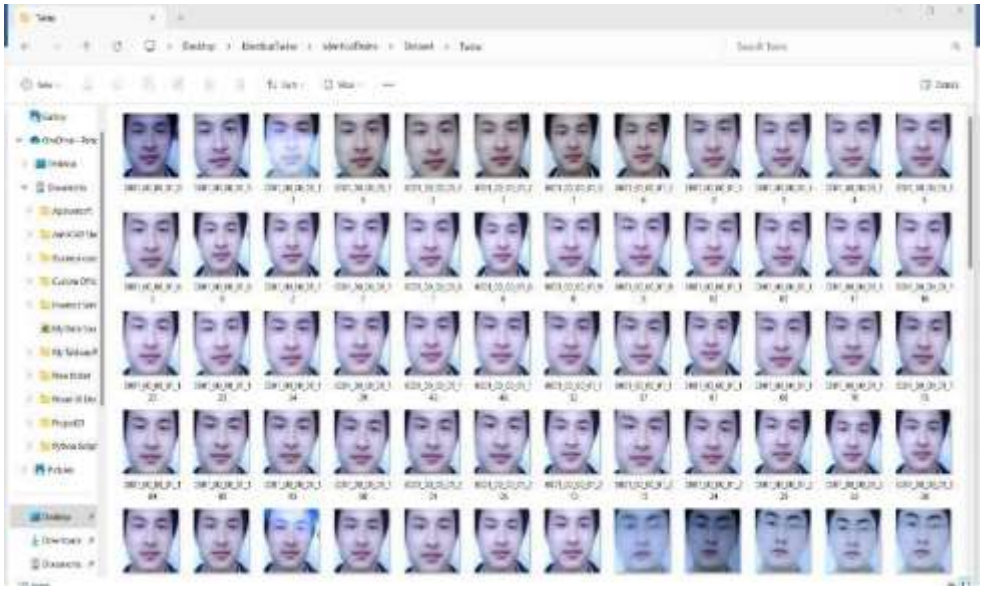

Figure 2:

So by using above images we will evaluate performance of both Random Forest and Naïve Bayes Algorithm.

To implement this project we have designed following.

1. **Input Dataset:** Using this module we will upload dataset to application and then apply filtration and object detection techniques.
2. **Facial feature extraction**: utilize facial landmarks detection or deep learning techniques to extract the features.
3. **Dataset Splitting**: using this module we will normalized and then shuffle and split dataset into train and test where application using 80% dataset for training and 20% for testing.
4. **Model training**: train the model using the training dataset, incorporating the extracted facial features. 80% processed train images will be input to Naive Bayes Algorithm to train a model and this model will be applied on 20% test images to calculate prediction accuracy
5. **Model Evaluation**: Assess the models performance on the testing dataset. 80% processed train images will be input to Random Forest Algorithm to train a model and this model will be applied on 20% test images to calculate prediction accuracy.
6. **Identical twins differentiation**: implementing the mechanism to identify suitable differences in facial features between identical twins. By using Comparison Graph module we will plot comparison graph between both algorithms .
7. **Twins or Real Face Detection**: using this module we will upload test images and then algorithm will predict weather image is real or belongs to twins.

The main concept of our work is to develop an identical twin fraud detection approach that increases the accuracy of classification and identification. Our proposed model used RGB Histogram on those different parts to take the initial decision whether this might be a potential twin candidate or not. The whole system of identical twin fraud detection divided into following steps- Input images, dataset pre- processing, feature extraction, training the model, classification.

In real words twins faces are exists and this twins can utilize advantages to dupe peoples in examination or any other organizations. To detect such twins we are applying machine learning algorithms such as Naïve Bayes and Random Forest which may get trained on possible Real and Twins faces. Once after training we can input face to this trained model to identify weather face is Real or Twin. This work has successfully provides a new pathway to support digital forensic.

In our project, we have used algorithms such as Random Forest and Naïve Bayes in terms of accuracy The scope

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

of a project focused on facial feature extraction for identical twins fraud detection using machine learning algorithms is broad and can have significant implications in various domains. Here are some aspects of the project's scope. In Identity Verification and Fraud Prevention the primary scope is to develop a system that enhances identity verification processes, particularly in scenarios where fraud detection is critical. Applications include online identity verification, secure access control systems, and preventing fraudulent activities such as account takeover. In Security systems the integration into security systems for facilities, airports, and other high- security areas where accurate identification is essential.

In forensic science identifying individuals in forensic investigations, especially in cases involving twins. Differentiating between identical twins for forensic purposes. In Education and Awareness, scope for educating users, organizations, and the public about the capabilities, limitations, and ethical considerations of facial recognition technology. Raising awareness about the importance of responsible data hand ling and privacy safeguards.

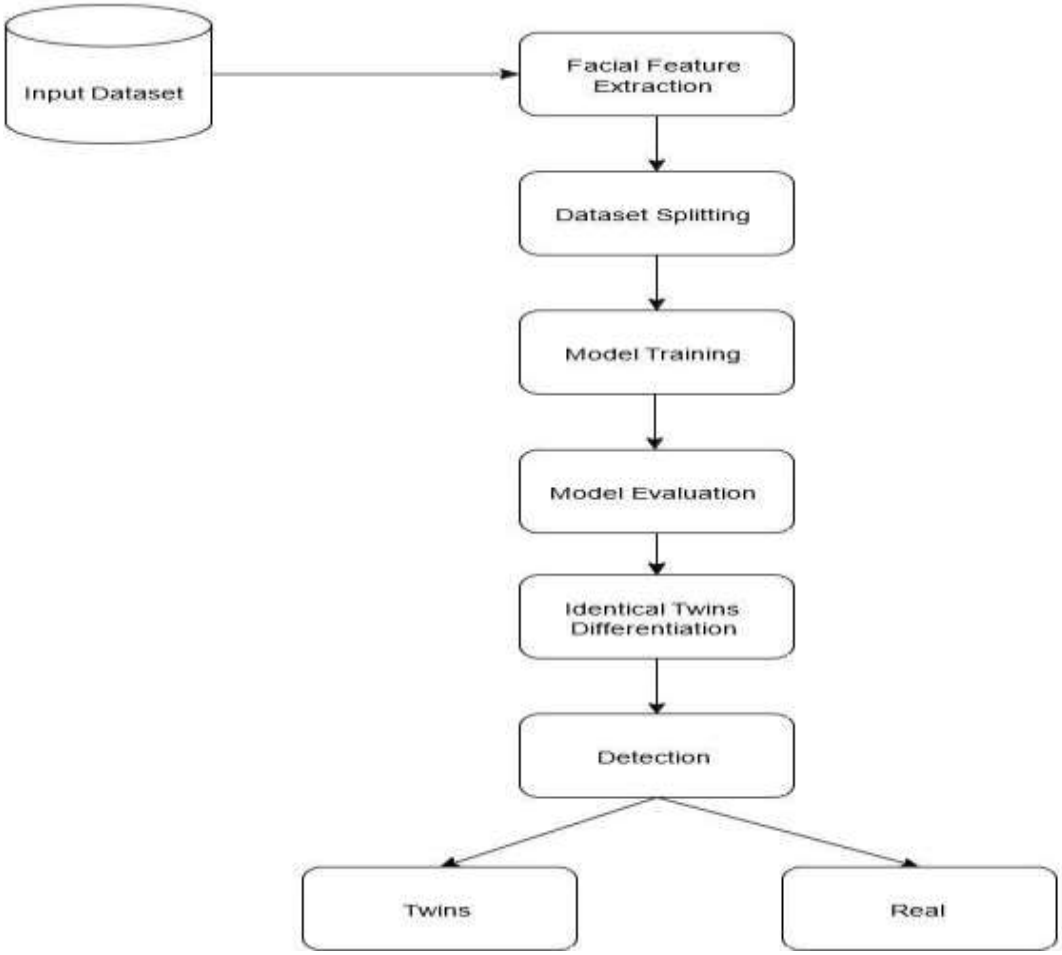

Figure 3:

IV. EARLIER WORK

The existing system for facial feature extraction in fraud detection typically involves the use of computer vision techniques and machine learning algorithms such as Support vector Machines(SVM) and Principal Component Analysis. But they have limitations that are overcome by Random Forest and Naïve Bayes Algorithms.

Identifying identical twins accurately is a challenging problem with several implications:

1. Forensics: In criminal investigations, the inability to distinguish between identical twins can lead to wrongful accusations or the release of a guilty party. Current forensic methods often rely on non-genetic traits that may not be unique to each twin.
2. Healthcare: In medical contexts, misidentification of twins can lead to incorrect treatments or prescriptions. This is particularly relevant in cases of organ transplantation, where genetic compatibility is crucial.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

3. Education and Research: Identical twins can provide valuable insights into the role of genetics in various traits and diseases. Accurate twin identification is vital for studies in behavioral genetics, epidemiology, and related fields.

Drawbacks:

Low Efficiency : The system exhibits low efficiency, suggesting that the computational processes and algorithms involved in facial feature extraction may not be optimized.

Low Accuracy : The system's low accuracy indicates shortcomings in the facial recognition models, implying that it struggles to reliably distinguish between identical twins based on genetic variations.

V. PROPOSED METHOD

The main concept of our work is to develop an identical twin fraud detection approach that increases the accuracy of classification and identification. Our proposed model used RGB Histogram on those different parts to take the initial decision whether this might be a potential twin candidate or not. The whole system of identical twin fraud detection divided into following steps- Input images, dataset pre-processing, feature extraction, training the model, classification.

In real words twins faces are exists and this twins can utilize advantages to dupe peoples in examination or any other organizations. To detect such twins we are applying machine learning algorithms such as Naïve Bayes and Random Forest which may get trained on possible Real and Twins faces. Once after training we can input face to this trained model to identify weather face is Real or Twin. This work has successfully provides a new pathway to support digital forensic investigation, employing artificial intelligence (machine learning) to improve existing face recognition systems Our model used Gray-Level Co- Occurrence Matrix Based on Contrast, Energy, Homogeneity, Dissimilarity, ASM, Correlation to take the ultimate result.

The "Facial Feature Extraction for Identical Twins Fraud Detection using Machine Learning Algorithms" project aims to address the challenges of twin identification through the following proposed solution:

Data Collection: A comprehensive dataset of identical twins, including their physical characteristics, biometric data, and genetic information, will be collected. This dataset will serve as the foundation for training the machine learning model. Machine Learning Model Development: State-of-the-art machine learning techniques, such as deep learning and feature extraction, will be employed to develop a model capable of identifying identical twins based on a combination of physical traits and biometric data. The model will be trained on the collected dataset to learn unique patterns that distinguish identical twins from other individuals.

Deployment and Integration: The trained model will be integrated into a user-friendly application that can be used in various domains, including forensics, healthcare, and research. This application will provide reliable and objective twin identification, helping to solve the problem of misidentification.

The purpose of the above project, focused on facial feature extraction for identical twins fraud detection using machine learning algorithms, can be summarized as follows:

Enhanced Identity Verification: Improve the accuracy and reliability of identity verification processes, especially in scenarios where traditional methods may struggle to distinguish between identical twins.

Fraud Prevention and Security Enhancement: Mitigate the risk of fraudulent activities, such as identity theft and unauthorized access, by developing a system that can effectively detect and prevent fraud attempts involving identical twins.

Biometric Security Advancement: Contribute to the advancement of biometric security by addressing the unique challenges associated with identical twins in facial recognition systems.

Improved Access Control Systems: Strengthen access control systems used in secure facilities, government institutions, and private organizations by providing a more reliable means of identifying individuals.

Legal and Ethical Compliance: Address legal and ethical considerations associated with the use of facial recognition technology, ensuring compliance with privacy regulations and protecting individuals' rights.

Public Awareness and Education: Raise awareness about the capabilities and limitations of facial recognition technology, promoting understanding and acceptance among users, organizations, and the general public.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

VI. RESULT AND ANALYSIS

Upload Twins Dataset

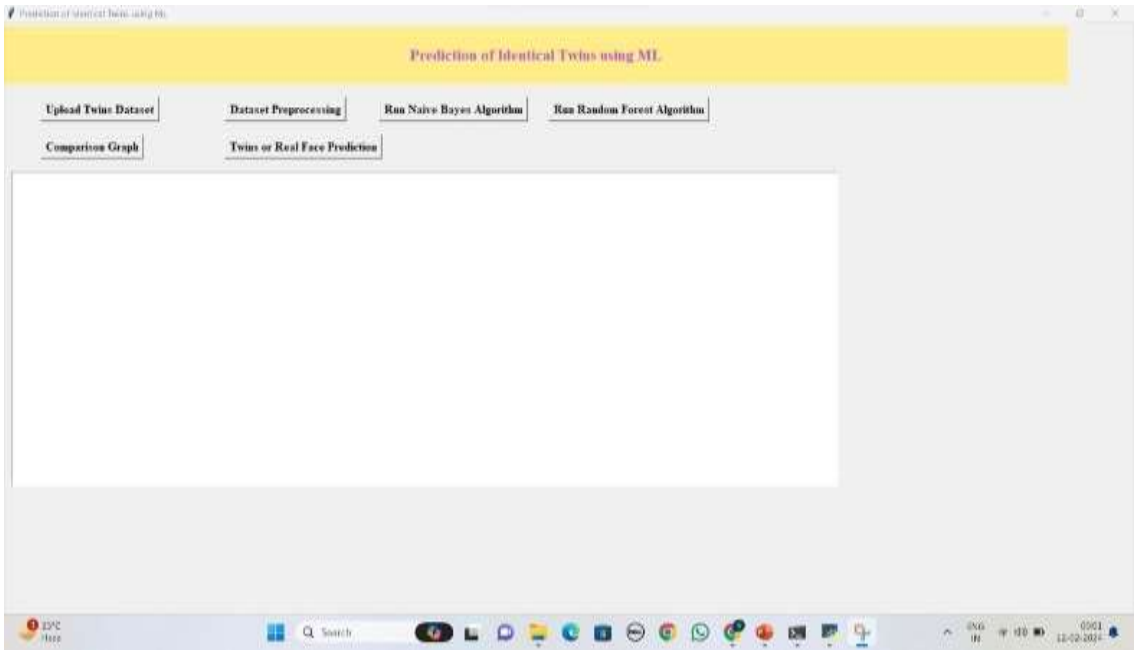

Figure 4:

In above screen click on 'Upload Twins Dataset' button to upload dataset and get below output

Selecting the Dataset

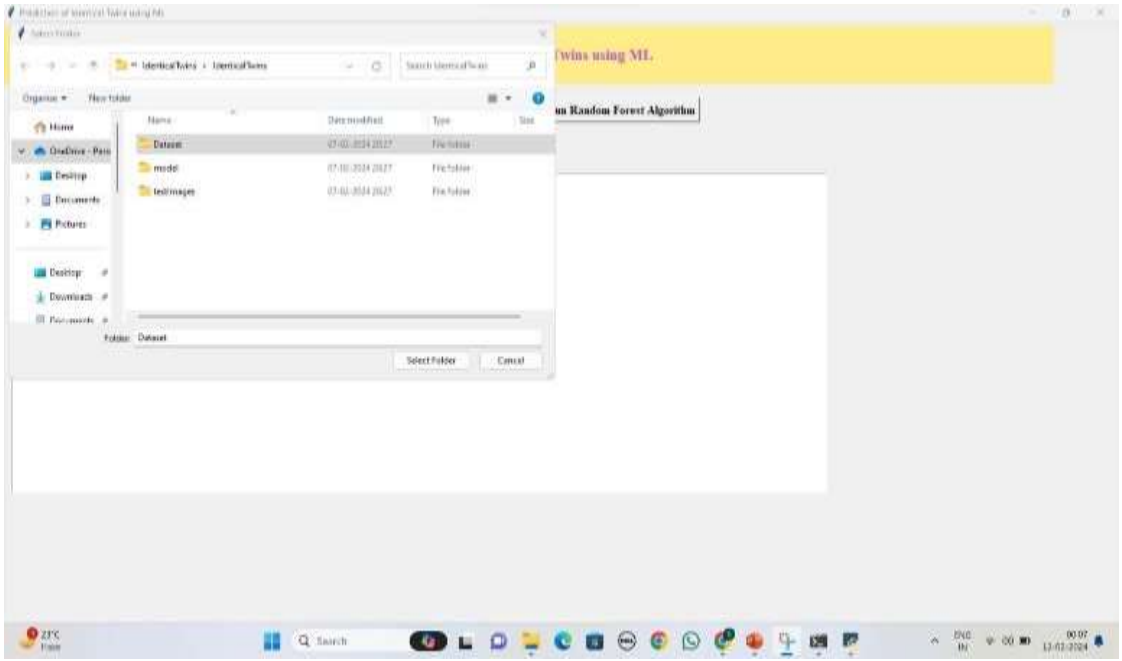

Figure 5:

In above screen selecting and uploading 'Dataset' folder and then click on 'Select Folder' button to load dataset and get below output .

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

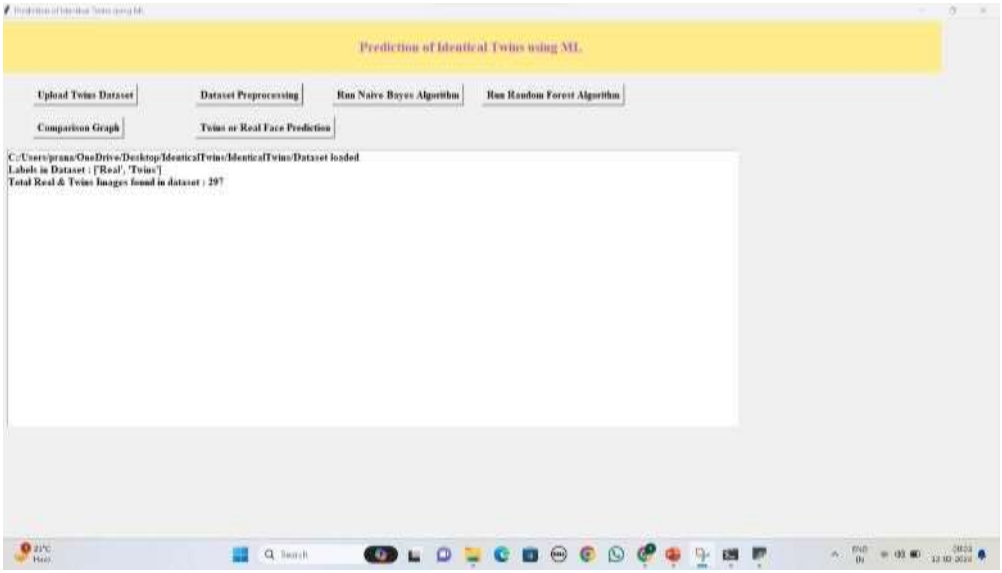

Figure 6:

In above screen we can see dataset loaded and we can see available labels and images in the dataset and now click on 'Dataset Preprocessing' button to normalize, shuffle and split dataset into train and test and will get below output.

Dataset Preprocessing

Figure 7:

In above screen we can see dataset processed and we can see total images used for train and test and now click on 'Run Naïve Bayes Algorithm' button to train Naïve Bayes and get below output.

Run Naïve Bayes Algorithm

Figure 8:

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

In above screen with Naïve Bayes we got accuracy as 99% and we can see other metrics also and in confusion matrix graph x-axis represents Predicted Labels and y- axis represents True Labels and green and yellow boxes contains Correct Prediction count and blue boxes represents incorrect prediction count which is 1 only and now close above window and then click on 'Run Random Forest' button to train Random Forest and get below output.

Run Random Forest Algorithm

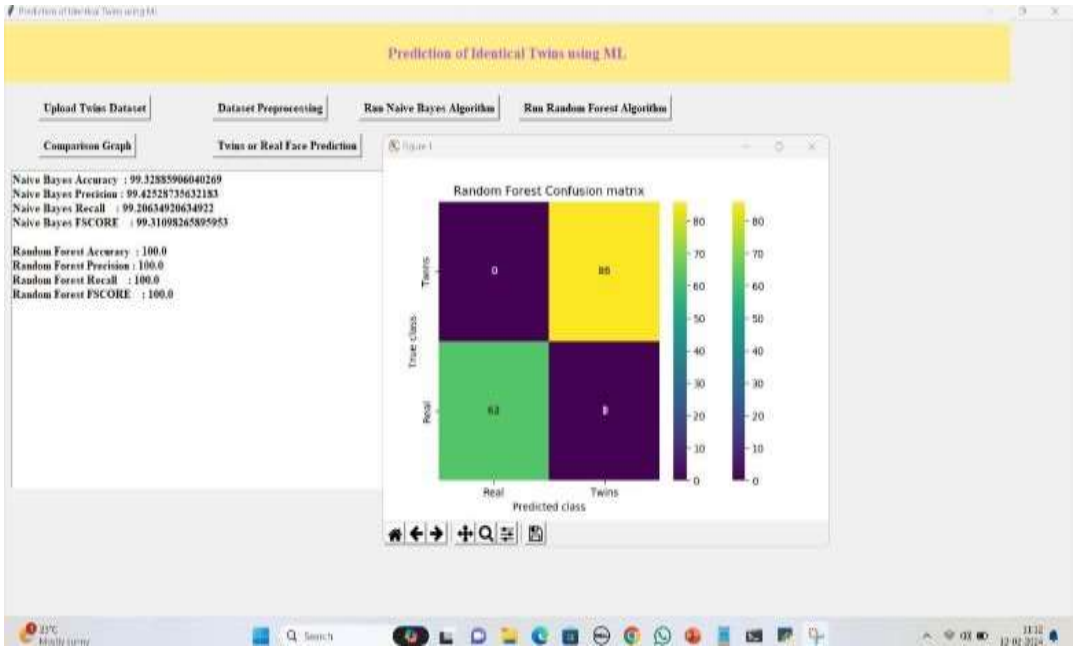

Figure 9:

In above screen with Random Forest we got 100% accuracy and we can see confusion graph also and now click on 'Comparison Graph' button to get below graph

Comparison Graph

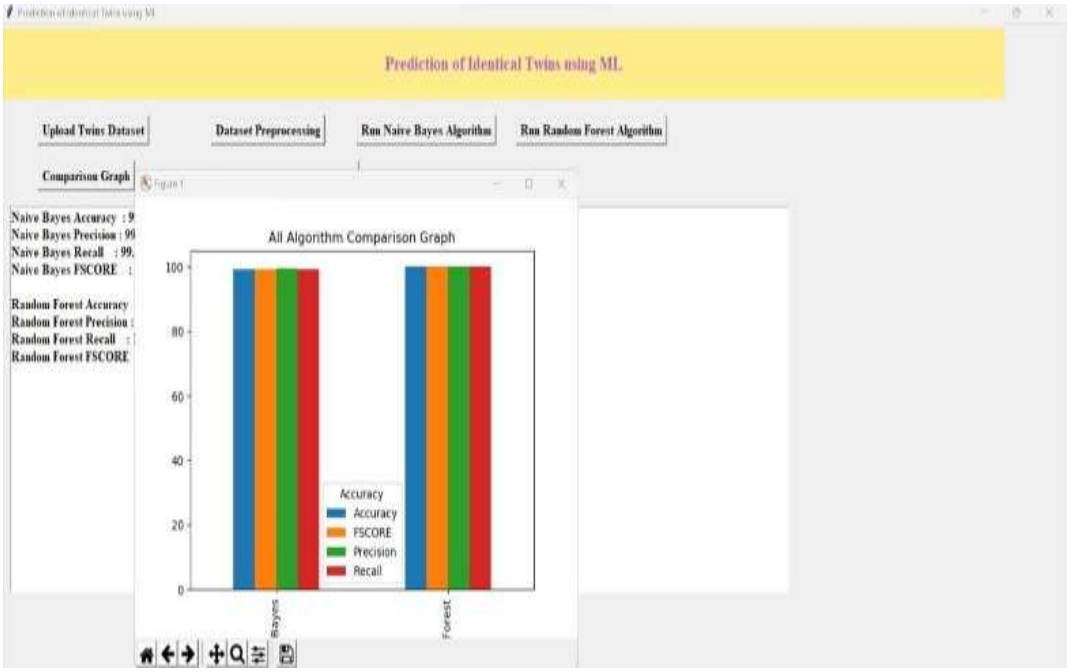

Figure 10:

In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different color bars and in both algorithms Random Forest got high performance and now click on 'Twins or Real Face Prediction' button to upload test image and get below output.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Twins or real face prediction

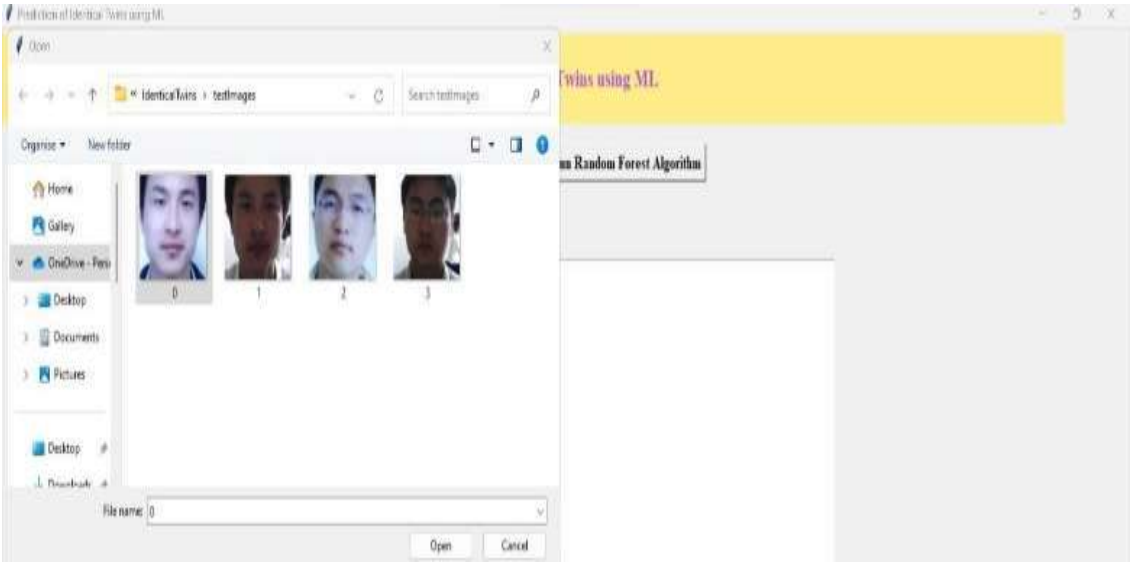

Figure 11:

In above screen selecting and uploading '0.jpg' image and then click on 'Open' button to load image and get below output.

Predicted as Twins

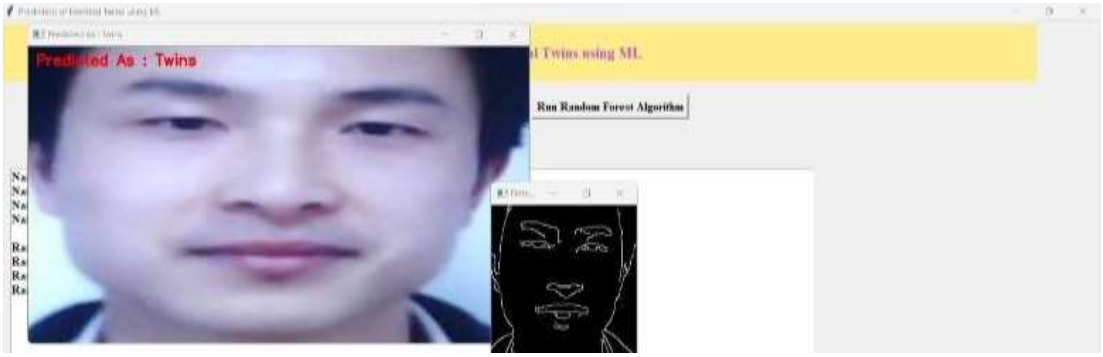

Figure 12:

In above screen in red color text we can see image predicted as Twins and we can see detected object in face in black and white color and similarly you can upload and test other images.

Predicted as Real

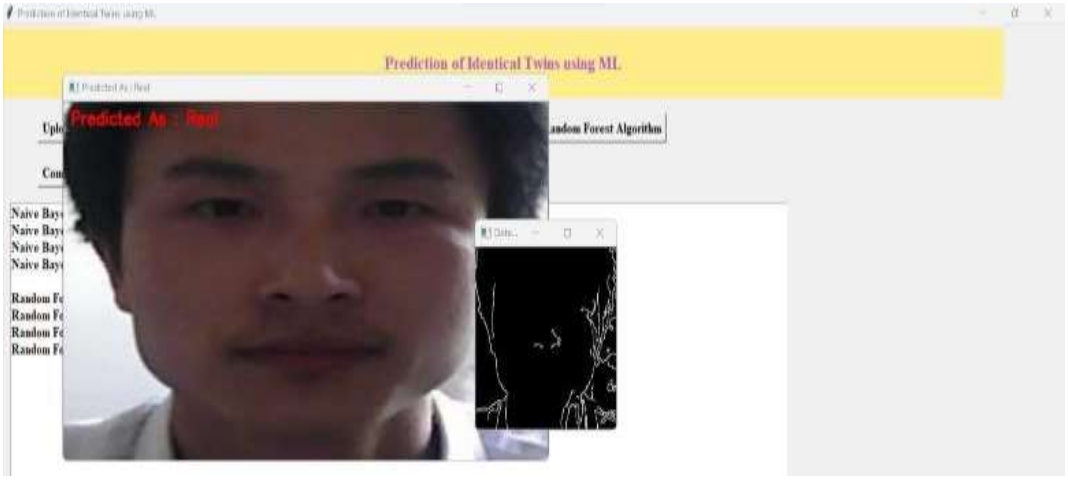

Figure 13:

In above screen image predicted as real.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

VII. CONCLUSION

This work has successfully provided a new pathway to support digital forensic investigation, employing artificial intelligence (machine learning) to improve existing face recognition systems. The current research work mainly concentrates on identifying the identical looking twins on basis of the features extracted.

The proposed system uses Python as the programming language for development of the project. The proposed system uses various number of frameworks like Open Computer Vision Library (OpenCV), TesnorFlow, NumPy, Keras module and Naïve Bayes and Random Forest Algorithm.

Flask frame work is used to build a Graphical User Interface for the system. It is used to connect to cloud and store in the database using the queries. Amazon Public Cloud is used to store the features in the cloud. Service used is Database as a service.

The practical performance of this project is analyzed and it shows that the system works perfectly for the limited dataset. This proposed work recognizes the identical twins and displays the result on the bases of features extracted, which may be any one among the three possible outputs of the system.

VIII. REFERENCES

[1] Recognition techniques for differentiate similar faces and twin faces R. Prema ; P. Shanmuga priya IEEE 2018.

[2] Face recognition using wavelet transforms-based feature extraction & spatial differentiation-based pre- processing Shridhar S Shanbhag; Sourabh B; K. Manikantan; S.Ramachandran IEEE 2018.

[3] Distinguishing Identical Twins by Face Recognition P. Jonathon Phillips, Patrick J. Flynn, Kevin W. Bowyer, Richard W. Vorder Bruegge, Patrick J. Grother, George W. Quinn, Matthew Pruitt IEEE 2018.

[4] Differentiating Identical Twins by Using Conditional Face Recognition Algorithms Swati Y. Dhote, A. D. Gotmare , M. S.Nimbarte IJSR 2015.

[5] Face Recognition: Issues, Methods and Alternative Applications Waldemar Wójcik, Konrad Gromaszek and Muhtar Junisbekov INTEC 2016.

[6] Face Recognition: From Traditional to Deep Learning Methods Daniel Saez Trigueros, Li Meng, Margaret Hartnett IJCT 2018.

[7] Performance Analysis of Human Face Recognition Techniques Sharmila; Raman Sharma; Dhanajay Kumar; Vaishali Puranik; KritikaGautham2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU) Year: 2019 | Conference Paper | Publisher: IEEE.

[8] Blur and Motion Blur Influence on Face Recognition Performance Katarina Knežević; Emilija Mandić RankoPetrović; BrankaStojanovi2018 14th Symposium on Neural Networks and Applications (NEUREL) Year: 2018.

[9] Alister Cordiner, Philip Ogunbona, and WanqingLi,"Face Detection Using Generalised Integral Image Features,"IEEE 2009.

[10] Janarthany Nagendrarajah, M. U. S. Perera, "Recognition of Expression Variant Faces – A Principle Component Analysis Based Approach for Access Control," IEEE 2010.