International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(12s) 2026 pp. 548-574
https://doi.org/10.70917/ijcisim-2026-3926
Article

IDENTIFICATION OF MONOZYGOTIC TWINS USING SWIN-TEMPORAL TRANSFORMERS WITH ATTENTION-BASED FEATURE SELECTION FOR GRAPH-BASED CLASSIFICATION

Rehkha K.K[1*], K. Menaga[2], P. Shyamala[3], D. Vidhyalakshmi[4]

[1,*]Department of Computer Science & Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, India.
rekha_renuj@rediffmail.com
[2]Department of Cyber Security, Dr. M.G.R. Educational and Research Institute, Chennai, India.
menagapravinkumar.19@gmail.com
[3]Department of Computer Science & Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, India.
shyamala.cse@drmgrdu.ac.in
[4]Department of Computer Science & Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, India.
vidhyanagaraj10@gmail.com
[1,*]Corresponding Author: rekha_renuj@rediffmail.com

Abstract: - Monozygotic twins pose a special challenge to biometric recognition and forensic identification since they are almost identical in genetic and phenotypic appearance and the traditional methods cannot adequately differentiate them. Common methods, including STR based DNA profiling, fingerprint recognition, and iris scanning and conventional CNN or hybrid CNN-RNN networks are not capable of detecting fine spatial and sequential variation leading to lower accuracy, sensitivity and specificity in the real world. In order to address these shortcomings, Swin-Temporal-GCN model, combining Swin Transformer hierarchical spatial feature extraction, Temporal Transformer sequential dependency modeling, attention-based feature selection, and Graph Convolutional Network (GCN) effective multi-class classification, was proposed in this study. The system was applied to Python platform with a heterogeneous set of 2,235 images of faces, with 725 images of MZ twins in real-time, 200 images of non-identical twins, and 625 images of test images of different illumination, pose, and noise conditions. The findings of the experiment proved the effectiveness of the proposed method as it reached the accuracy of 97 percent, and sensitivity and specificity were respectively higher than traditional methods. This framework is useful to forensic investigators, security agencies and biometrics authentication systems and offers a scalable, high-precision and robust method of identification of MZ twins.

Keywords: - Monozygotic Twins, Swin Transformer, Temporal Transformer, Graph Convolutional Network, Biometric Identification

1. INTRODUCTION

Identical twins (also known as monozygotic or MZ twins) occurred when a single fertilized embryo was divided into two embryos at the initial phases of their formation. This led to the creation of two people that almost had the same genetic makeup and this gave them very close physical features, such as the facial make-ups of their faces, eyes (iris), and even the way their bodies were shaped. Although they were genetically alike, minor differences arose as a result of after-zygotic mutations, changes in the genetics and the environmental factors encountered during the development. These tiny differences may encompass minor changes in facial features, ears, or behavioral patterns which were very important in aspects like forensic science, biometric authentication, and security mechanisms (Dallagiovanna *et al.*, 2021). The issue of differentiating MZ twins occurred since most of the traditional methods of identification, such as the use of DNA profiling, fingerprint identification, and standard facial recognition techniques, do not usually capture these subtle differences. Proper differentiation of MZ twins played an important role in forensic research, legal, and individual-specific security purposes, since a false identification may lead to the conviction of a person or unauthorized access. Through the emergence of high-quality computer vision and deep learning algorithms, scientists sought to create models that can learn and take advantage of minor spatial and temporal variations in face features. The reason behind the desire to study MZ twins in this sense is that they would serve as a test case of high precision biometric systems, and provide the opportunity to push the limit of the computational feature extraction and classification.

Traditional biometric and forensic techniques, such as fingerprinting, iris scanning, STR analysis, and facial recognition, are based on distinctive physiological and behavioral traits. Although successful for the majority of the population, these techniques are not suitable for MZ twins due to their identical DNA and very similar phenotypic features(Konijnenberg *et al.*, 2021). More recent techniques, such as ultra-deep genome sequencing, mitochondrial DNA analysis, and DNA methylation profiling, aim to distinguish twins but are often invasive, requiring enzymatic processing or bisulfite conversion, vulnerable to DNA degradation. From the machine learning perspective, traditional algorithms and facial recognition systems are not suitable for pose variation, lighting differences, and low-resolution images. Therefore, there is a need for more reliable techniques that can effectively detect small spatial and sequential biometric variations, especially in facial and ear biometrics, to successfully distinguish monozygotic twins.

In order to overcome the problems of MZ twin identification, the Swin-Temporal-GCN model was proposed in this study, combining efficient transformer-based models and graph-based classification. The approach originally converted preprocessed facial pictures into patch-based token embeddings that can be fed into Swin Transformer allowing hierarchical spatial feature mining. Swin Transformer was capable of capturing local facial textures, as well as global structural relationships, by using shifted-window self-attention, which boosted the ability to discriminate subtle differences within the faces. A Temporal Transformer was then used to model sequential dependencies between the extracted features and in effect, learn inter-feature relationships between varying regions of the face. It was also ensured that only the most discriminative features were used by suppressing the redundant or less informative features using attention-based feature selection. Lastly, the Graph Convolutional Network (GCN) conducted multi-class classification, where the feature-feature structural relations were used and the relationship among feature tokens to distinguish monozygotic twins with high accuracy.

1.1 Problem Statement

The case of monozygotic (MZ) twins is a major challenge in forensic analysis and biometric recognition because the usual approaches, including STR analysis, fingerprinting, and facial recognition, are not able to differentiate between them(Dahal *et al.*, 2023). This is due to the fact that they have almost the same genetic structure and very similar physical characteristics, including their facial structures, iris patterns, and fingerprints. The usual DNA analysis and algorithms are not able to handle the small differences in the facial and ear biometric patterns because the current CNNs or machine learning algorithms are not able to handle the small differences in the facial and ear biometric patterns.

1.2 Motivation of the Study

This research was prompted by the fact that there is an urgent requirement of very accurate identification of monozygotic twins in forensic, security and biometric systems. The genetic and phenotypic variation between twins is almost similar, so traditional identification methods are not suitable to distinguish between them. To resolve this, the research paper developed a pure transformer-based and graph-based approach, which combines Swin Transformer to extract hierarchical spatial features, Temporal Transformer to model sequential relationships among features, attention-based feature selection to reduce redundancy of features, and Graph Convolutional Networks (GCN) to conduct effective multi-class classification. This higher order combination enables the system to obtain both local finer grain detail and global relational patterns of the facial and ear biometrics. The model was trained using a mixed dataset of 2235 pictures, comprising of 725 actual-time MZ twin pictures and 200 pictures of non-identical twins, 625 pictures were held back to test the model at different levels of illumination, pose, and noise conditions. The aim of this approach is to get a higher degree of accuracy, sensitivity and specificity in addition to the ability to offer a strong, scalable and real-world viable solution to the monozygotic twins differentiation in high-stakes identification cases.

1.3 Significance of the Study

The research was also big as it filled a much-needed gap in the proper identification of monozygotic twins which could not be effectively done by traditional biometrics. Wrongful convictions, unauthorized access and legal conflicts may be as a result of misidentification during forensic investigations, security system and biometric authentication. The study made it possible to extract fine spatial and temporal variations of face and ear features that would otherwise remain hidden by suggesting a Swin-Temporal-GCN model. Hierarchical spatial feature extraction, temporal modeling, attention-based feature selection, and the use of graphs to classify the features made the approach to be highly accurate, sensitive, and specific even when the conditions were unfavorable: the illumination, pose, and noise varied. The generalization, robustness, and applicability of the model were increased by the usage of a diverse dataset comprised of real-time and web-based twin pictures. This paper has shown that architecture based on transformers and the use of GCNs can become a future standard of biometric recognition of MZ twins, offering a methodological framework in future investigations in the field of forensic science and non-vulgar identification systems.

1.4 Key Contributions

- Novel Architecture: A new Swin-Temporal-GCN architecture was devised to accurately detect monozygotic twins.
- Improved Methodology: It has used Swin Transformer to extract spatial features, temporal transformer to model sequentially, attention-based feature selection to optimize features, and GCN to classify.
- Collection of Data: 2,235 images were used that contained 725 real-time MZ twin images and 200 non-identical twin images out of which 625 images were to be used in testing under varied real-life environments.
- High Accuracy: Obtained accuracy of 97 percent with a sensitivity and specificity of 94-97 and 93-97 percent respectively, which is higher than the previous methodologies.

The rest of the paper is organized in the following way. Section 2 provides a literature review in the areas of MZ twin identification, deep learning, and graph-based biometrics. Section 3 elaborates the suggested Swin-Temporal-GCN feature optimization, and GCN classification. Section 4 shows the experimental outcomes, performance. Section 5. Lastly, Section 6 closes with the study and provides guidelines to continue with the research in more advanced biometric identification of monozygotic twins.

2. RELATED WORKS

Face Recognition (FR) is very difficult in current trends particularly when it is viewed precisely similar to one another. The significant influence of identification process has been taken by facial Recognition. There are numerous algorithms and technologies that have been applied by numerous scientists to achieve the best and precise outcomes. Abubakar Ibrahim Muhammad, Singh and Yusuf Ibrahim (2020) suggested a research involving the Facial recognition based on the hybrid application of the Particle Swarm Optimization (PSO) and Support Vector machine (SVM) technologies to determine the patterns of the facial features to identify the face. SVM as well as Radial Basis Function (RBF) was employed as a classifier and PSO was adopted to choose the optimal parameter. The researchers came to a conclusion of demonstrating that accuracy level improved with the increase in number of generation in terms of training the model and eventually reached 98 percent accuracy level, as they demonstrated in their journal article, Face Recognition with Particle Swarm Optimization (PSO) and Support Vector Machine (SVM).

Monozygotic twins can be detected by FR process with the help of Deep Convolutional Neural Networks developed by Ahmad Belal et al. (2020) in the article "Deep Convolutional Neural Network Using Triplet Loss to Distinguish the Identical Twins" and tested by the researcher. The proposed research was to determine twins who seem to be look-alike to one another through Convolutional Neural Network as well as Triplet Loss Function to distinguish identical twins. Euclidean Distance and Triplet Loss Function was put in place to locate the distance between the face facial features. Both the twin images are compared to get the similarity features by comparing their L2 distance. Issues like over-fitting, normalization and dropouts are some of the problems that are tackled with the help of different technologies. The obtained results reveal an average accuracy during the validation of 87.2% accuracy.

The nearest hereditary-based relationship existing in indistinguishable twins makes fingerprint and palm print recognition a difficult undertaking. The impressions of the erosion edges of skin on all pieces of fingers are known as the fingerprints and the palm area of the hand is known as palm prints. The fingerprints and palm prints are one-of-a-kind qualities and super durable for each individual. The fingerprints and palm prints are utilized for confirmation and individual ID of people, even in instances of indistinguishable twins. The primary focal point of Komaljeet Kaur et al. (2019) is to investigate fingerprints and palmprints of indistinguishable twins is to look at the similarities and dissimilarities of prints designs in the matches which are analysed homolaterally. The respectively observed likenesses and dissimilarities between the hands of each indistinguishable twin are then contrasted and the above similarities and dissimilarities is presented in the paper, “Recognition and Analysis of Fingerprints and Palm Prints Patterns of Identical Twins of Malwa Region, PUNJAB”

Juefei-Xu Felix and Savvides (2013) a technique of ALDA (Augmented Linear Discriminant Analysis) was employed to cope with the problem of recognising identical twins. This model also acquires a typical subspace which, besides being able to identify which family the singular belongs to, is also able to identify other people within a similar family. The ALDA is regarded by the authors as an alternative to the traditional LDA (Linear Discriminant Analysis) method of subspace learning on the data set of the Notre Dame Twin. They have demonstrated that the suggested ALDA scheme as directed by facial imbalance incorporates, in total, outflanks other profoundly rooted facial descriptors (LBP (Local Binary Pattern Histogram), LTP (Local Ternary Pattern), LTrP (Local Tetra Pattern)) and the method of ALDA subspace can enhance at the position of unique discernible twins as compared to LDA. It achieves 48.50% VR at 0.1% FAR to identify family participation of indistinguishable twin people within the group and an reached the midpoint of 82.58% VR at 0.1 percent FAR to verify indistinguishable twin people within a similar family, which is a huge enhancement over traditional descriptors and conventional LDA method, which is reported in the paper " 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Facial recognition algorithms have the option to work in any event, when the same looking individuals are found for example likewise in the outrageous instance of indistinguishable looking twins. In research done by Chandrakala Raju et al. (2020) collected an experimental dataset that contains 40 pictures of 20 sets of twins gathered arbitrarily from the web. The preparation is finished with the chosen pictures of the twins utilizing different preparation calculations and inbuilt capacities accessible. As a piece of the testing stage irregular pictures from the dataset prepared are chosen and after running it over the framework get the elements of those pictures which then will be analysed by removing the highlights previously put away in the Amazon cloud. The put-away qualities and the ongoing picture highlights are looked at and the result will be shown on the GUI. Indistinguishable twins' facial acknowledgment framework utilizes AI, picture handling calculations, and profound learning calculations. No matter what the states of the pictures obtained, recognizing indistinguishable twins is essentially more diligent than recognizing faces that are not indistinguishable twins for all the algorithms.

The algorithms of facial recognition should be able to work during any event, and just as the faces of the similar people are encountered, or at the worst, during the case of identical twins. Experimenting with a dataset comprising of 17,486 pictures of 126 sets of indistinguishable twins (252 subjects) collected around the same time and day and 6864 pictures of 120 sets of indistinguishable twins (240 subjects) where pictures were used a year after the fact, Jeffrey Paone et al. (2014) evaluated the exhibition on seven different face recognitions calculations. The accounts are on the varieties of execution in light, articulation, orientation and age of that very day and cross-year picture sets. Whatever the states of picture procurement, the recognition of indistinguishable twins is more difficult than in the case of so much as picture procurement of subjects that are not indistinguishable twins to all algorithms.

The problem of differentiating between identical twins and non-twin carbon copies in computerized FR (Facial Recognition) systems has increasingly gained importance with the general acceptance of facial biometrics. John McCauley et al. (2021) gathered the largest sets of twins to deal with two FR problems: 1) making a target percent of facial comparability between indistinguishable twins and 2) using this measure of closeness to make decisions about the impact of doppelgangers, or carbon copies, on FR performance on large datasets of faces. The measure of facial likeness is solved using a Deep CNN. The proposed network provides a quantitative similarity rating on any two provided faces and has been utilized on the vastness of face datasets to classify comparative face pairs.

According to Venkatesan et al. (2019) face acknowledgment and biometrics depends on a particular and extraordinary individual distinguishing proof. This method is subject to matching the picture and other individual pictures for creating a conspicuous verification. This is perhaps the best way to deal with playing out the relationship by picking facial credits from the picture and a facial information base. In facial acknowledgment, computations should have the choice to perceive the near-looking individuals or be prepared to separate the indistinguishable twins using face recognition with accuracy grouping. The concentrate credits were characterized using an SVM classifier. The input picture of an individual is first distinguished to be twins or not established on grouping methods. SVM could be utilized for regression and grouping issues. It is generally utilized in scientific classification issues. SVM is less computationally concentrated. Consequently, it would be widely developed and quickly created in ongoing years.

Face Recognition of identical twin is an emotional challenging due to the existence of a grave degree of affiliation in the overall look of the face. Hardly any twins of the same sex help in business fraud such as fake protection fees. Specifically, when supposing that one of the unclear twins commits a real crime; their indefinite natures cause havoc and helplessness in the initial phases of court proceedings. An Adaboost twins classifier was suggested by Bhargavi et al. (2021). It is this calculation which characterizes the locale of the face of the info picture. The highlights of the face area identified by the AdaBoost model and stored in the information bases during the preparation and test stage are separated by the PZM (Pseudo Zernike Moment) and DoG (Difference of Gaussian) techniques. The SVM classifier identifies the features of the twin and compares id prepared and attempted highlights and differentiates the offender that is likely to occur later on. The outcomes of the trial demonstrated the ability of the suggested strategy to see two identical twins.

Recent studies have revealed that the performance of face recognition execution reduces impressively on images of identical twins. The matching capacity in human faces is often perceived to be taken as a standard in measuring and further enhancement of programmed face algorithm recognition. Biswas Soma, Bowyer Kevin and Flynn Patrick (2012) researched the human ability to recognize indistinguishable twins. Assuming people can recognize facial pictures of indistinguishable twins, it would propose that people are equipped for distinguishing and separating facial attributes that might be helpful to foster calculations for this exceptionally difficult issue. Tries different things with various survey times and imaging conditions are led to decide whether people seeing a couple of facial pictures can see assuming that the picture matches have a place with a similar individual or a couple of indistinguishable twins. The tests are administered to 186 twin subjects, and it is the largest such concentration in the writing so far. People can play out the undertaking essentially better assuming they are given sufficient opportunity and will quite often commit more errors when pictures contrast in imaging conditions. Individuals can execute the undertaking essentially superior provided that they are allowed ample time and will frequently make more mistakes when pictures oppose in imaging environments. The outcome of the study investigation advises individuals to seek facial impressions such as moles, scars etc. to follow their decision and do more horrible when presented with photos without such impressions. Explores different avenues regarding programmed face acknowledgment frameworks show that human viewers outperform programmed matches for this process.

With the improvement of innovation, the use regions and significance of biometric frameworks have begun to enlarge. Since the qualities of every individual are unique to one another, a solitary model biometric framework can yield better results. Anyway, because the attributes of twins are extremely near one another, various biometric frameworks including different qualities of people will be more fitting and also increase the rate of identification. Based on Cihan Akin, Umit Kacar and Murvit Kirci (2019) reviews, a numerous biometric identification framework comprising a mixture of different algorithms and various models was created to recognize individuals from others and their twins. Ear and voice biometric information was utilized for the multimodal model and 38 sets of twin ear pictures and sound accounts were utilized in the informational index. Sound and ear acknowledgment rates were acquired utilizing traditional (hand-created) and DL techniques. The outcomes acquired were joined with the progressive score level combination technique to make an achievement rate of 94.74% in rank-1 and 100 percent in rand their twins. The multimodal model used ear and voice biometric data and the informational index used 38 samples of ear pictures and sound accounts of twins. The sound and ear recognition rates were obtained by the use of traditional (hand-created) and DL methods. The results obtained were combined with progressive score level combination method to achieve a rate of 94.74 in rank-1 and 100 percent in rank - 2.

With the expansion in the number of twin births in late many years, there is a need to foster substitute methodologies that can get the biometric framework. Priya Lakshmi and Rani Pushpa (2017) use another methodology for recognizing indistinguishable twins based on a multimodal distinguishing proof framework that utilizes three distinct elements to be a specific face, finger impression and lip print to recognize individuals. A recently evolved multimodal biometric framework has various special characteristics, beginning from using

The Kernel Similarity with Euclidean Distance approach to face coordinating, Possibilistic Fuzzy C-Means approach to unique finger impression coordinating and a Fixed K-means Clustering highlights a lip print coordinating approach and combined the data to seek powerful recognition and affirmation. The relevance of taking into account these constrained circumstances, such as twins, where the possibility of errors is most severe will cause us to strategize a more reliable and robust security system. The suggested method is experimented on a realistic information sample that included 429 sets of identical pictures of twins and demonstrates encouraging results. The receiver operating attributes also reveals that the proposed strategy is more considered than other processes. By so doing, the trial findings showed that the proposed technique was capable of the simplest level of perceiving two identical twins which were indistinguishable.

Similarly looking at twins is one of the main parts of the brain in the human studies which disclosure the significance of the environmental and hereditary on the different views of the characteristics of the brain and diseases. Appropriate features of same-looking people permit one to infer the hereditary influence on the population. Gritsenko Andrey, Lindquist Martin and Chung Moo (2020) uses a pair of novel-wise classification pipeline for recognition of zygosity of twin pairs utilizing the resting state of (rs-fMRI) image. The recent character indication is used to effectively build a network of the brain for every single subject. This concept is applied for nearly 208 pairs of identical twins in Human Connectome Project and 92.23(±4.43) % is the rate of accuracy that is got.

Biometric is one of the strong authentication which is used for facial recognition or in identification of a person. It also provides a strong authentication in identification of monozygotic twins in various aspects. A survey done by Bowyer Kevin and Flynn Patrick (2016) proposed a research on face recognition among identical twins through comparing the different biometric applied in identification. All the biometric features like face recognition, finger print, a study on face recognition on identical twins by comparing the different biometric used to identify individuals are also brought to the fore in the paper Biometric Identification of Identical twins: A Survey. All the biometric features, including face recognition, finger print, etc are pointed out in the paper "Biometric Identification of Identical Twins: A Survey" palm print, iris, speech recognition and handwriting used for identifying identical twins. The dataset was obtained from various research groups. In terms of finger print the dataset seems to be limited in case of identical twins which may affect the overall accuracy level. The paper finally concluded saying that the less number of data set taken for implementation by many researchers has ended up with poor accuracy level.

One of the significant difficulties the biometric recognition framework confronts is how to recognize monozygotic or indistinguishable twins. Notwithstanding, the quantity of numerous births has been expanding throughout recent many years. It can't be distinguished because of DNA. So, the absence of an appropriate distinguishing proof framework will prompt numerous lawbreaker issues. Shalin Eliabeth et al. (2015) concentrated on various biometric distinguishing proof innovations given the highlights of face, unique finger impression, palm print, iris, retina and voice for the check of indistinguishable twins. By investigating can understand that face discovery in light of facial imprint is the most proficient one for the distinguishing proof of indistinguishable twins. Different elements (finger impression, palm print, iris recognition, retina identification and so on) are not exceptional for indistinguishable twins.

John McCauley et al. (2021) introduces the use of one of the largest twin databases that have been summed up so far to answer two FR problems: 1) making a decision at a proportion of similitude of faces between indistinguishable twins, and 2) using the measure of similitude to make a decision about the impact of doppelgangers, or clones, on the implementation of Facial Recognition with large face data. The measure of facial comparability is solved by deep convolutional cycle in the brain. The suggested network assigns an approximation of closeness to any two faces provided and has been utilized to huge scale face datasets to differentiate related pairs of faces.

In security structures, face recognition is mandatory. The severe problem in this case is the recognition of twins; in order to overcome this problem, Kavya Ashwini et al. (2021) made use of the calculations of the facial acknowledgment technique and the example of MSVM open surf and Gabor. The FR (Facial Recognition) calculations are applied to differentiate identical twins who cannot be differentiated with the help of order. The data presented on this proposed structure is the image of the person where different highlights are extracted by applying the Gabor calculation, and open surf calculation is applied to estimate the differences and similarities between the two images. With the usage of a MSVM classifier, the characteristics between two photos are emphasized and then it is classified.

Asogwa Tochukwu Chijindu and Ituma Chinagolum (2018) paper presents the use of Machine Learning for the computerized identification of indistinguishable twins. The primary point here is to introduce a new l episteme that will separate two comparable associated faces with various characters utilizing a group of attributes. The method incorporates face recognition among other image handling methods, preparation, and order and results in expectation. A few existing reference indices will be explored and examined and another framework will be proposed.

Nahar Khalid Mohammed et al. (2021) tackle this test by a model for twin's face recognition where our answer depends on profound exchange learning as far as leftover residual NN (Neural Networks) including two VGG-16 (Visual Geometry Group-16) prepared networks, which are viewed as one of the DNNs. For investigation purposes, the authors look at different ways to deal with taking care of the issues of the twins such as iris, fingerprints and the corners of the lips. The data used is a test that was collected through Google. There are 4-sets twins in the information with 17-different situation applied to each that generates 5x2x17 (170) different pictures. Collected images were used in between highlight examinations. Findings indicated that mathematical highlights were the source of 85% progress whereas photometric elements contributed 96%. When mathematical and photometric components are combined in a hybrid way, the results of the hybrid are 98 percent accurate. In this test, biometrical actions can be used to show the dominance of profound learning movement as opposed to traditional strategies. The newly achieved procedure may also be replaced to facilitate authentication systems that entirely depend on biometric attributes.

Gunjan Gautam, Aditya Raj and Susanta Mukhopadhyay (2021) tackle the twin's recognizable issue; a difficult assignment in biometric validation. The trials are done utilizing The Deep Learning processes and for this, the authors have developed a CNNs (Convolutional Neural Network-Siamese) and an MCNNs (Multiscale Convolutional Neural Network-Siamese). They have also explored and evaluated three pre-trained CNNs to be specific, ResNet-50, VGG-16, and Neural Search Architecture Net-Large, and applied them to Siamese structure after making appropriate adjustments. Also, a Simulated Neural Network is also employed as a Siamese sub-network in the research work. They have made convincing exploratory proofs regarding CCR (Change Control Record) on the CASIA-IrisV4 Twins dataset that demonstrate the effectiveness of visual biometrics in recognizing twins. The results have achieved the best-in-class human-level precision. Likewise, the closeness of Concurrency and Coordination Runtime of the multitude of models declares that the visual districts in twins hold a critical connection to naming them as twins. The authors evaluated the cross-area ability of the proposed CNN & MCNN in the ND-General Format Identifier dataset that outflanks the existing techniques. Particularly, for the visual biometrics, as far as possibly know, there is right now no writing accessible as of the date that investigates the relationship among twins and in this way unriddles the categorization utilizing DL.

The article by Rita Goel, Irfan Mehmood and Hassan Ugail (2021) examines the use of cutting DL face recognitions models to determine the power of segregating the faces of siblings using various similitude files. The specific models that were examined in this intention are FaceNet, VGGFace, VGG16 and VGG19. In the case of every series of pictures provided the embeddings have been ascertained through the selection of the deep learning model. Four different trial data sets (full-front facing face, eyes, nose, and temple of kin matches) are constructed using a freely available set of Histograms of Oriented Flow of Sibling Database data set. The accuracy of the selected deep learning models to identify kin is precise due to the trial outcomes that indicate that the full-front facing face and trimmed face areas vary depending on the face region under analysis. When focusing on full-front facing face and eyes - the accuracy of order is more than 95 percent in this case. However, its precision suffers as a result of basically examining the noses, whereas Face Net offers the most effective result of order based on the button. In addition, VGG16 and VGG19 are not the most suitable models to group using eyes, however, they produce decent results when temples are compared.

Face acknowledgment has become one of the well-known procedures nowadays, such as telephone open structure, criminal distinguishing evidence and to the surprise, home security framework. Teoh et al. (2021) developed a novel design of the similar twin identification framework. This framework is more secure since it does not have to bother with any conditions, such as key and card but only a facial picture is to be provided. Generally speaking, human recognition scheme involves 2 steps namely face recognition and face differentiating evidence. This conception is the most viable in planning and promoting a face recognition system using deep picking up of OpenCV in python. DL is a way to deal with playing out the face acknowledgment and is by all accounts a satisfactory technique to complete face acknowledgment because of its high exactness. Outcomes are given to show the precision of the proposed face acknowledgment system.

Face recognition can be utilized to distinguish an individual extraordinarily founded on advanced image handling. Various ML and DL methods are applied to identify similar twins. Recognizing identical twins is difficult with the help of existing methods.

3. PROPOSED METHODOLOGY

The research hypothesis of the proposed methodology is to identify the monozygotic (MZ) twins accurately with the help of a sophisticated transformer-based and graph learning development. This method, unlike the traditional methods, which find it difficult to resolve minute biometric variations, uses the strengths of Swin Transformers to obtain hierarchical spatial features of the facial and ear images, whereby fine-grained local textures and global structural features are maintained.

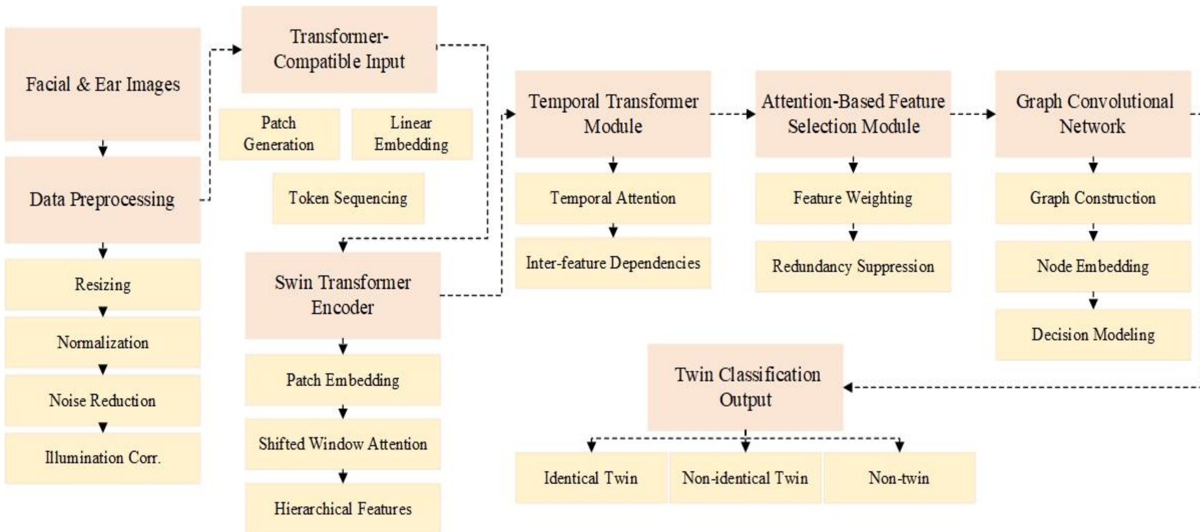


Fig 1 Swin-Temporal-GCN Feature Extraction and Classification

A Temporal Transformer is then used to model sequential dependencies across these features, and identify the complex connections that exist between the various areas of the face across a series of embeddings. In order to increase the quality of the features and limit redundancy, an attention-based feature selection module is used, which indicates the most discriminating characteristics. Lastly, a Graph Convolutional Network (GCN) is used to classify using multiple classes, which is essentially the recognition of MZ twins and people who are not identical. The framework is tested and trained on a varied dataset of 2,235 images, which consists of the real-time and non-identical twin images, and thus the framework is robust, scalable, and very adaptable to real world forensic and biometric advancements. The overall Workflow in shown in Fig 2.

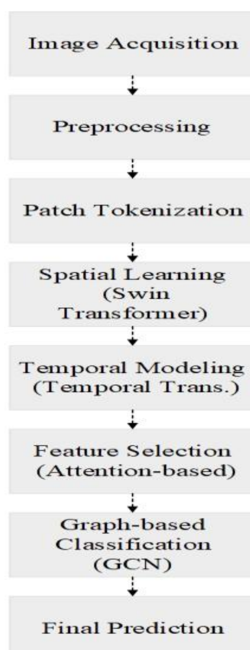


Fig 2 Overall Work Flow

3.1 Data Collection

In the proposed study, various sources were used to gather facial image datasets to make it diverse and reliable in the process of identifying monozygotic twins. The proposed model was trained using 2235 facial images, including the real-time data sets and publicly accessible web-based data sets. Out of the samples that were gathered, 725 images are related to real-time monozygotic twin data, which are acquired in hospitals, fertility centers, schools and colleges. These pictures have been taken in different conditions to represent the real life.

To make the suggested system even more sound, one of the datasets consists of 200 non-identical twins pictures as well, thus being able to discriminate effective between identical and non-identical twins. The entire dataset was further split into training and testing where 2235 images were to train and 625 images were to test the model functionality. This compilation of data is the guarantee of the balanced learning and the valid assessment of the hybrid framework proposed.

3.2 Data Pre-processing

Image pre-processing forms an important part of the suggested framework, to provide uniformity, as well as enhance quality of the input images to extract features reliably. The following stages are sequential with respect to the pre-processing pipeline:

3.2.1 Image Resizing

The facial and ear images are all resized to a standard size in order to be consistent throughout the dataset. This is because uniform dimensions permits the deal with images in an efficient way and minimizes the computational cost without loss of structural information are shown in Eqn. (1).

$$I_r(x', y') = I\left(\frac{x'}{W'} \cdot W,\ \frac{y'}{H'} \cdot H\right) \qquad (1)$$

3.2.2 Normalization

Pixel intensities values are converted to a fixed range: [0,1] to normalize the brightness and contrast of the image are shown in Eqn. (2). Normalization makes sure that the differences in light or exposure are not biased towards the learning procedure and the network concentrates on the discriminative facial features.

$$I_n(x, y) = \frac{I(x,y) - I_{min}}{I_{max} - I_{min}} \qquad (2)$$

3.2.3 Noise Reduction

There can be random noise in images as a result of errors of sensors or environmental conditions. Noise suppression methods like Gaussian or median filtering are used in ensuring that edges are retained within an image without distorting important patterns on the faces of a particular person are shown in Eqn. (3).

$$I_d(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x - i, y - j)\ G(i, j) \qquad (3)$$

3.2.4 Illumination Correction

Histogram equalization or adaptive contrast enhancement is used to correct different lighting situations. This step enhances the visibility of essential features, and thus subtle differences in face and ear patterns become more evident to extract them using CNN and sequentially analyze them using RNNare shown in Eqn. (4).

$$I_e(x, y) = (L - 1) \sum_{i=0}^{I(x,y)} p(i) \qquad (4)$$

Through a systematic implementation of the above preprocessing procedures, the dataset becomes homogenized, further refined, and resilient, and it is crucial to proper twin identification.

3.3 Transformer-Compatible Input Representation

The ear and facial images have been preprocessed and then converted into a representation that is compatible with transformer-based learning models. Transformer architectures also use sequential tokenized inputs unlike convolution-based networks, which directly process pixel grids. Thus the improved images are converted initially to an ordered series of visual tokens that maintain both local spatial and global contextual relations.Every normalized image is separated into non-overlapping patches of fixed sizes. The patches are local areas of the facial and ear images and enable the model to produce fine-grained biometric features like micro-textures, edge outline, and subtle shape differences that are very important in the distinction between monozygotic twins. Patch size is also chosen well to strike a balance between sensitivity of local features and performance are shown in Eqn. (5).

$$\mathcal{P} = \{p_i \mid p_i \in R^{P \times P \times C}\} \qquad (5)$$

Patch generation is done after which each image patch is flattened and subjected to a linear embedding layer. This embedding is an encoding of the raw pixel values into a higher dimensional feature space resulting in a concise and semantically useful description of each patch are shown in Eqn. (6). The patch embeddings are then combined with positional embeddings, to store spatial ordering information, which is required, as transformers do not necessarily encode positional relations. This allows the model to discern the relative spatial relationship between the facial and ear features including the geometric correspondence between the eyes, nose, jawline, and ear features.

$$z_i = p_i W_e + b_e \tag{6}$$

The resulting embeddings are placed in a sequence of tokens that create the input sequence of the Swin Transformerare shown in Eqn. (7). This token representation enables the self-attention system to estimate the relationship between remote parts of the face and at the same time retains local structural information. Consequently, the local textures (the short range dependencies) and the global facial geometry (the long-range dependencies) are both well represented. This input representation, which is transformer compatible, offers a strong base of hierarchical feature of space and guaranteed that discriminative biometric data is effectively coded to be processed by other stages of this proposed framework.

$$Z = [z_1, z_2, \dots, z_N] \tag{7}$$

3.3.1 General System Architecture

The suggested system architecture adheres to an end-to-end transformer-based pipeline that will be able to differentiate the monozygotic twins with precision based on the spatial, temporal, and relational features learning. To make it robust and scaled, the architecture combines various elaborate deep learning modules in a modular and sequential fashion are shown in Eqn. (8). The first step, facial and ear images are provided with a standardized preprocessing to improve the quality of the image and minimize variability introduced by light, noise, and the differences in resolution. The preprocessed images are transformed into token representations that are transformer compatible by means of patch generation and linear embedding. The token sequences are then put through the Swin Transformer module, which is used to extract hierarchical spatial features via shifted window self-attention to receive local and global biometric patterns.

$$E_p = X_p W_e + b_e \tag{8}$$

Spatial characteristics of the Swin Transformer are then sent to a Temporal Transformer which captures inter-representation sequential dependencies and inter-feature relationships are shown in Eqn. (9). The stage allows the system to inherit the minute differences in biometric characteristics, which cannot be easily determined with the help of the static feature modeling.

$$Z_0 = E_p + P \tag{9}$$

After that, the learned representations are further narrowed down to discriminative features and non-redundant information through attention based feature selection mechanisms. The optimized feature set is then modeled as a graph that is sent to a Graph Convolutional Network (GCN) which performs relational reasoning and does final classification. This combined architecture assures the effective information transfer among modules and provides the precise, reliable and scalable monozygotic twins identification.

3.4 Swin Transformer based on Spatial Feature Extraction

The extraction of spatial features is important in the mapping of monozygotic twins, in which the biometric differences are extremely thin on the facial and ear features. In order to successfully learn such fine-grained spatial patterns, the proposed architecture uses the Swin Transformer, a hierarchical vision transformer used specifically to learn visual representation.After transformer compatible input representation the embedded patches tokens are processed by the Swin Transformer encoder. Contrary to the traditional global self-attention schemes, Swin Transformer introduces shifted window-based self-attention plan are shown in Eqn. (10). This method divides the feature map into fixed-size windows then locally computes self-attention in each window, making it much simpler to compute and retaining local contextual information. In order to allow each window to interact with the other windows, and to learn the global context, the windows are moved on the successive layers, and as a result, information is exchanged across adjacent areas are shown in Fig.2.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (10)$$

Patch embedding layer first projects the patches of every image into a high-dimensional feature space, based on which attention-based processing then takes place. Multi head self attention is then implemented inside the window, and the model can learn about several different relationships between the features. This system records discriminative spatial information including facial symmetry, differences in contours and ear shape differences that are imperative in distinguishing between monozygotic twins are shown in Eqn. (11).

$$Z^{l+1} = \text{SW-MSA}\big(\text{LN}(Z^l)\big) + Z^l \qquad (11)$$

The Swin Transformer also uses the hierarchical representation learning design, in which the patch merging layers are arranged consecutively with the spatial resolution decreasing and the feature dimensionality always increasing are shown in Eqn. (12). This hierarchical architecture allows the model to acquire low-level texture representations at lower levels and structural representations of high level at lower levels. Subsequently, micro-level details (as well as macro-level geometric configurations) are successfully encoded.

$$F_s = \bigcup_{l=1}^{L} Z^l \qquad (12)$$

Also, the layer normalization and residual connections are added to stabilize the training and enhance convergence. Spatial features and hierarchical features detected by Swin Transformer give a highly detailed and discriminatory representation of biometric features, which are then passed to the temporal modeling segment to be analyzed further.

3.4.1 Temporal Feature Modeling with Temporal Transformer

Whereas spatial feature extraction is used in order to capture fixed biometric patterns, monozygotic twin identification involves modeling of inter-feature relationship and dependency between spatial representations. In order to fulfill this need, the framework proposed includes a Temporal Transformer module in order to capture sequential relationships between extracted spatial features.

The space-wise feature vectors produced by the Swin Transformer are given as ordered sequences as an input to the Temporal Transformer are shown in Eqn. (13). As compared to recurrent-based architectures, the Temporal Transformer does not use recurrence and completely depends on self-attention to learn long-range dependencies without experiencing issues of a vanishing gradient. This makes the model to be able to analyze a number of feature interactions at the same time and learn the complex associations between the facial and ear representations.

$$E_t = F_s + P_t \qquad (13)$$

The temporal attention processes attach adaptive weights to features elements in accordance to their relevance, which allows this model to attend to discriminative temporal patterns and ignore redundant information. Positional encoding is used to keep the order of the feature embeddings, which implies that relational dependencies are correctly represented. This is especially useful in the measurement of subtle differences in facial expressions, ear alignment as well as space consistency among different representations.Multi-head temporal self-attention helps the model to learn various dependency patterns in different attention subspaces are shown in Eqn. (14).

$$F_t = \text{TransformerEncoder}(E_t) \qquad (14)$$

The attention heads pay attention to different interactions of features thus increasing the representation. The attended features are further refined through the feed-forward layers to provide nonlinear transformations which increase the feature separability.The Temporal Transformer can also be used to process feature sequences in parallel, resulting in a better computational efficiency than the traditional sequential models. This architecture introduces the proposed framework as appropriate to large scaled datasets and real-time biometric.The Temporal Transformer is able to model inter-feature dependencies and temporal relations, which boosts the discriminative ability of the learnt representations, and thus offers a strong base on to refine the features further and classify the graph using other features.

3.4.2 Feature Selection based on Attention

The extracted representations can still have the redundant or weakly informative features that can adversely impact classification performance after learning spatial and temporal features. The proposed framework mitigates this problem by adopting an attention-based feature selection process which conducts a differentiable process of feature weighting on the discriminative ability whilst reducing redundancy are shown in Eqn. (15).

$$\alpha_i = \frac{\exp(w_i)}{\sum_{j=1}^{n} w \exp(w_j)} \qquad (15)$$

The attention-based feature selection module allocates adaptive scores of importance to every feature dimension according to its contribution to the identification of twins are shown in Eqn. (16). It does not use heuristic or evolutionary optimization algorithms, but rather it can be easily viewed as part of the deep learning pipeline, where feature selection is learned jointly with the rest of the model. The element of self-attention calculates the relevance of every feature by determining inter-feature relationships and focusing on those with a greater discriminative strength.

$$F_o = \sum_{i=1}^{n} \alpha_i F_t^{(i)} \qquad (16)$$

Through the use of soft attention weights, the model preferentially boosts informative attributes associated with facial symmetry, micro-textures variations and ear structural features, and down-weights spurious or redundant ones. This is done to make sure that the critical biometric information is maintained without eliminating potentially useful contextual information. The attention mechanism is differentiable, which allows gradual optimization, thus, converging more quickly and making it more stable throughout the training are shown in Eqn. (17).

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{N} e \exp(e_j)} \qquad (17)$$

Moreover, attention normalization minimizes overlap of features, and in turn, redundancy suppression is realized because such point maximizes the separation between classes and reduces the collinearity of features. This advanced feature description increases the performance to withstand illumination, pose and noise differences. The post-processing of this module is a small but very informative feature set which is an ideal input to the next graphical classification step.

3.5 Graph Convolutional Network Classification

The proposed system employs a Graph Convolutional Network (GCN) as the last stage of classification to be effectively classifying monozygotic twins using relationships between refined feature representations. The GCN models provide more decision modeling capability with highly similar biometric samples, unlike the traditional classifiers that consider samples independently are shown in Eqn. (18).

$$\hat{A} = A + I \qquad (18)$$

Within the suggested framework every subject representation is treated as a node in a graph and edges are created between nodes depending upon feature similarity measures like cosine similarity or Euclidean distance. This type of graph structure is used to capture inherent association between facial and ear features and their occurrence and hence the model learns discriminative patterns when it aggregates neighborhood features. Adjacency matrix establishes the intensity of relation among nodes, which depicts the likeness of biometric samplesare shown in Eqn. (19).

The GCN conducts node embedding in an iterative process, in which every node updates its representation whereby features of each node are aggregated with its neighbors. The contextual information can be propagated throughout the graph with the help of this process, which promotes the discrimination of features in those instances when the monozygotic twin shares almost identical appearance.

$$\widehat{D_{\iota\iota}} = \sum_j \widehat{A_{\iota\jmath}} \qquad (19)$$

The higher-order relationships are learned using several graph convolution layers, and nonlinear activation functions are used to provide representational sophistication. The resulting node embeddings are then sent to a softmax layer to be classified so that the model can effectively differentiate between identical and non-identical twins.The GCN improves the classification accuracy, robustness, and generalization by exploiting the relational learning and feature aggregation based on the context. This is a graph based decision modelling method which best satisfies the biometric identification task associated with subtle inter-class differences, and strong inter-class similarity.

3.6 Strategy and Optimization of Training

The suggested transformer-based framework is also being trained with the help of a supervised learning strategy to provide stable convergence and the best performance. To reduce the error of misclassification, categorical cross-entropy loss function is used to reduce the error rate between monozygotic twins, non-identical twins and non-twin. This loss can be propagated effectively across the Swin Transformer, Temporal Transformer, attention-based feature selection and GCN modules using this loss formulation are shown in Eqn. (20).

$$\mathcal{L} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (20)$$

Its ability to adapt the learning rate and update with momentum that assists in sparse gradients and speeding up the convergence. To avoid oscillations around the optimum solution, learning rate scheduling is used to allow a reduction in the learning rate with training. To prevent overfitting and enhance the level of generalization, batch normalization and dropout regularization are included.Validation loss and accuracy are also used to monitor model convergence. When the performance begins to level off, the training is canceled according to the early stopping criteria to guarantee efficient training in the most cost-effective way are shown in Eqn. (21).

$$\mathcal{L}_{reg} = \lambda \sum_l \left|\left|W^{(l)}\right|\right|_2^2 \quad (21)$$

3.7 Scalability and Fault Tolerance

The attention mechanisms in the form of transformer and the operations of a graph convolution are the two main factors that define the computational complexity of the proposed framework. Although the higher computational cost of the self-attention compared to standard CNNs, window-based attention in the Swin Transformer is much less time-complex. Temporal Transformer layers are used on sequences of compact features, in order to enhance efficiency are shown in Eqn. (22).

$$\mathcal{L}_{total} = \mathcal{L} + \mathcal{L}_{reg} \quad (22)$$

The attention-based feature selection is used to minimize memory consumption by eliminating redundant features. The modular architecture enables parallel use of GPUs and edge accelerators facilitating dependency on overall architecture to ensure scalability and efficiency in real time biometric authentication and mass deployment applications are shown in Eqn. (23).

$$W^{(l)} \leftarrow W^{(l)} - \eta \frac{\partial \mathcal{L}_{total}}{\partial W^{(l)}} \quad (23)$$

Algorithm Swin-Temporal-GCN MZ Twin Identification

Input: Preprocessed images I

Output: Class labels Y_pred

```
Initialize model parameters
Load facial image dataset

while training is not converged do

  Read input image

  if image quality is low then
    apply preprocessing
  else
    proceed with original image
  end if

  Generate non-overlapping patches
  Embed patches into token representations

  Extract spatial features using Swin Transformer
```

```
    if temporal information exists then
        model feature dependencies using Temporal Transformer
    else
        retain spatial feature representation
    end if

    Apply attention-based feature selection
    Suppress redundant features

    Construct graph using feature similarity
    Perform node embedding using GCN

    Predict class label

    Compute loss between predicted and true labels
    Update model parameters using optimizer

end while

Return trained model
```

The suggested algorithm outlines the overall workflow of determining monozygotic twins with the help of an advanced transformer-based framework. The system first loads the support images of faces and sets all the model parameters. In every round of training, an input picture is artificially treated and its quality is estimated. In case of noise, dim lighting, and low resolution in the image, preprocessing tasks, including resizing, normalization, noise removal, and illumination correction are performed on the image to improve the quality of features; additionally, the image is directly sent to the learning pipeline.

The improved image is then split into non overlapping patches which are encoded into token embeddings that are applicable in transformer learning. The Swin Transformer is used to extract spatial facial features, with hierarchical self-attention which identifies both local textures and global faces. In the case of sequential or contextual dependencies between the features extracted, a Temporal Transformer learns such relationships to detect minor variations that are essential in discriminating twins.A attention-based feature selection mechanism is then used to assign weight of importance to features overruling redundant or less informative representations.

4. RESULT AND DISCUSSION

The experimental environment was designed to assess the efficacy and performance of the proposed models, RNN+PSO and RNN+PSO+MSVM, for the identification of monozygotic (MZ) twins. The models were developed on high-performance computing platforms, Intel i7 processors, 16 GB RAM, and NVIDIA GPUs, to address the computational complexity of deep learning. Standardized software development platforms, Python 3.8, TensorFlow 2.x, OpenCV, and Scikit-learn, were used to ensure the reproducibility and consistency of the experiments. The dataset included real-time images of twins collected from hospitals, schools, and fertility centers, as well as web-based images, offering a wide range of scenarios for comprehensive evaluation. The system was tested for varying lighting conditions, pose, and image quality to provide a comprehensive assessment of the classification performance and efficacy of the proposed models for the identification of identical twins are shown in Table I.

Table I: Experimental Setup

Parameter	Specification
Dataset	2,235 images (725 MZ twins, 200 non-twins, 625 test)

System	Swin Transformer+Temporal Transformer + Attention-Based Feature Selection + Graph Convolutional Network
Hardware	Intel i7, 16GB RAM, NVIDIA GTX 1080 GPU
Software	Python 3.8, TensorFlow 2.x, OpenCV, Scikit-learn
Tools	Jupyter Notebook, MATLAB
Metrics	Accuracy, Sensitivity, Specificity

4.1 Implementation Results

4.1.1 Data Acquisition

Data acquisition significance is to gather information from pertinent sources before it tends to be put away, pre-processed and utilized for further stages. It is the method involved with recovering important business data, changing the information into the necessary business structure and stacking it into the assigned framework. The course of information procurement includes looking for the datasets that can be utilized to prepare the ML models. Following Fig.3demonstrates the input images of twins.

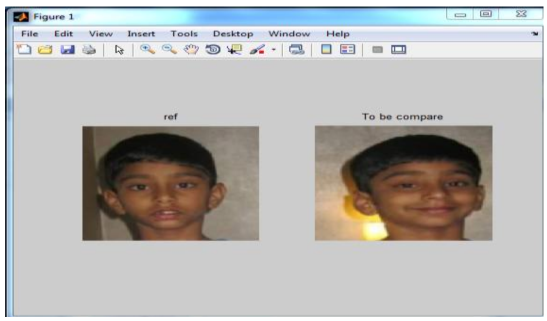


Fig 3 Input Images

4.1.2 Pre-processing

The goal of pre-processing is to work on the nature of the picture to examine it in a superior manner. Pre-processing eliminates undesired components and improves a few elements which are important for the specific application we are working for. Here the given pictures are changed over into dim filtered pictures and grayscale pictures utilizing different picture handling methods. Fig.4 illustrates the pre-processed image.

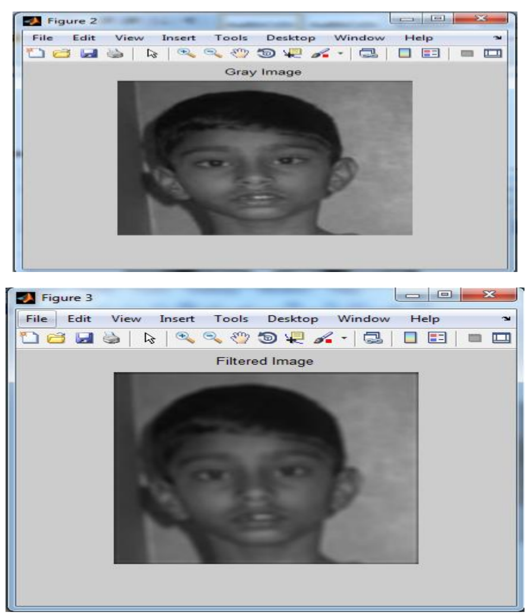


Fig 4 Pre-processed Images

4.1.3 Binarization

Picture Binarization is the change of a document picture into a bi-level document picture. Picture pixels are isolated into a double assortment of pixels, for example highly contrasting. The fundamental objective of picture binarization is the division of the report into closer view text and foundation. Following Fig.5denotes the binarized images.

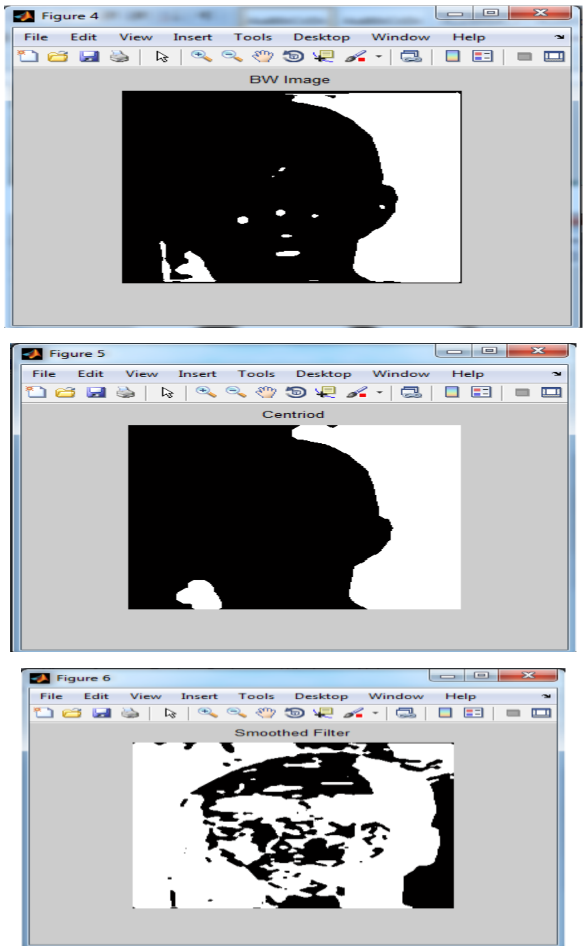


Fig 5 Binarized Image

Morphology is a wide arrangement of picture handling activities that cycle pictures based on shapes. Morphological tasks apply an organizing component to an input picture, making a result picture of a similar size. The morphological ideas comprise a strong arrangement of instruments for extricating elements of interest in a picture. The Following Fig 6 shows the morphological images. The Fig. 7 indicates the extracted eye images from the Morphological Image. The Following Fig. 8 describes the extracted nose images from morphological images. The Following Fig.9 and 10 demonstrates extracted mouth and ear images from the morphological images.

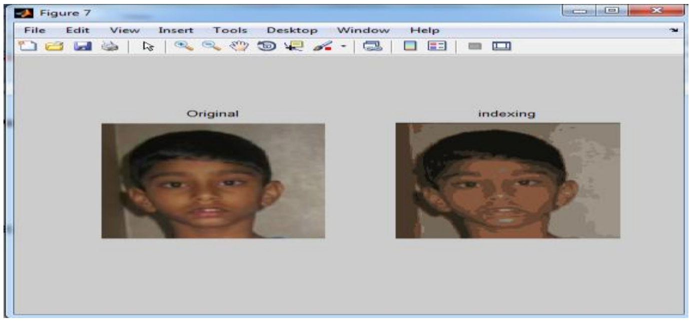


Fig 6 Morphological Image

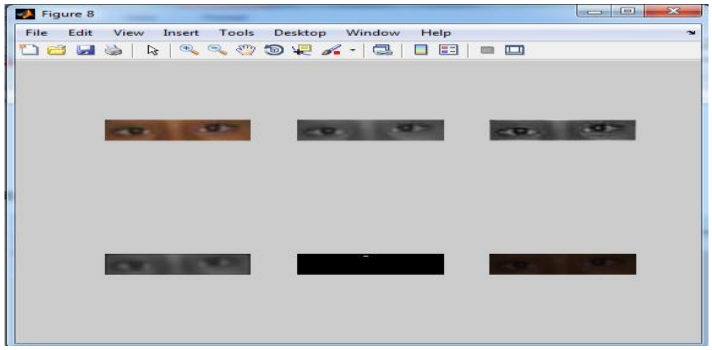


Fig 7 Eye Images

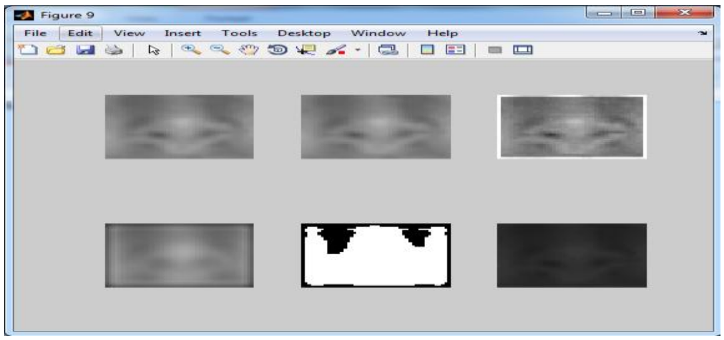


Fig 8 Nose images

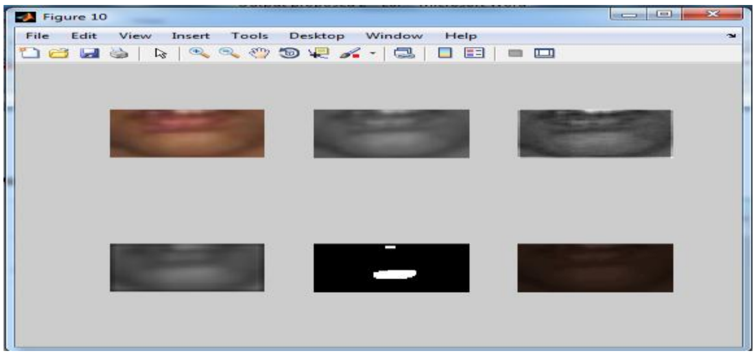


Fig 9 Mouth Images

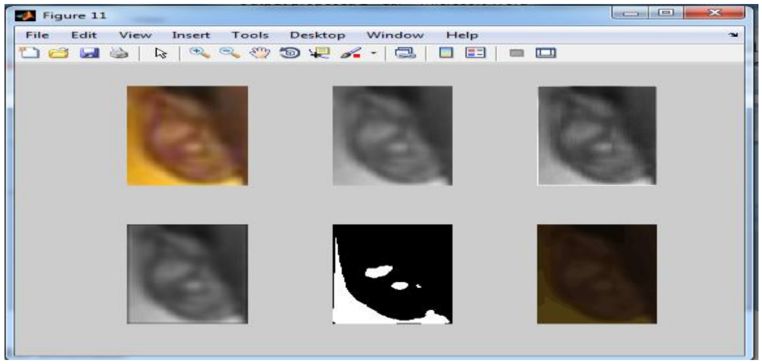
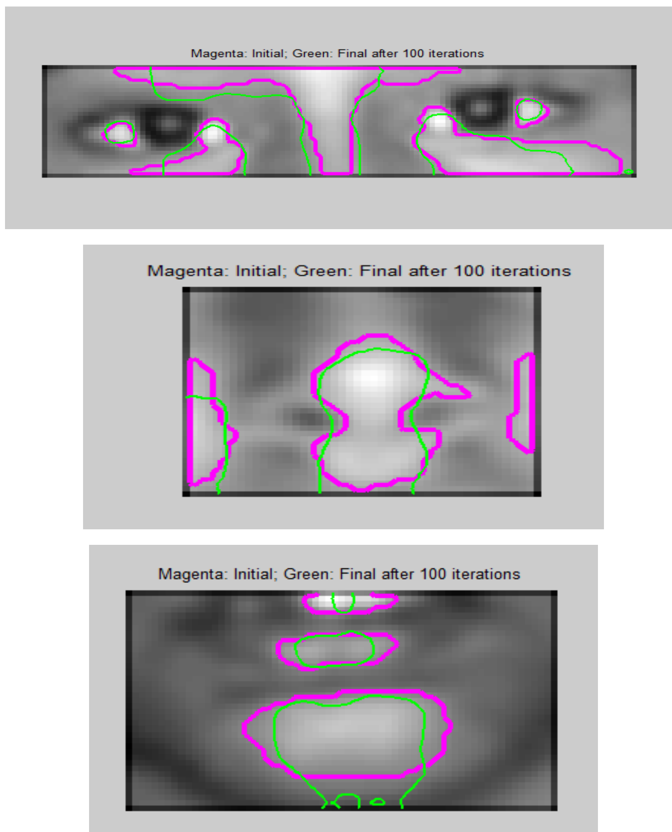


Fig 10 Ear Images

Image classification is the technique in which a digital picture is classified into different subdivisions known as image segments that assist in decreasing the difficulty of the picture for making the next technique. Dissection is simply known as allotting labels to the pixels. The Following Fig.11 shows the segmented eyes, nose, ear, and mouth images with 100 iterations.



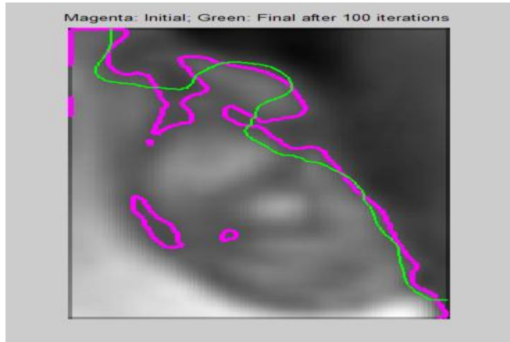


Fig 11 Segmented Eyes, Nose, Ear and Mouth Images

Feature extraction is a kind of dimensionality lessening process where an enormous number of pixels of the picture are proficiently addressed so that important pieces of the picture are caught successfully. FS is a methodology of choosing the most pertinent subset highlights for fostering a hearty Artificial Intelligence model. In the FS (Feature Selection) process, excess and additionally superfluous information is taken out from the fundamental data set, subsequently, the execution of the diagnostics model might be moved along. By FS, computational weight on the machine will be decreased, subsequently, computational proficiency is expanded. The Following Fig. 10 describes extracted features values of eyes, nose, ear and mouth images.

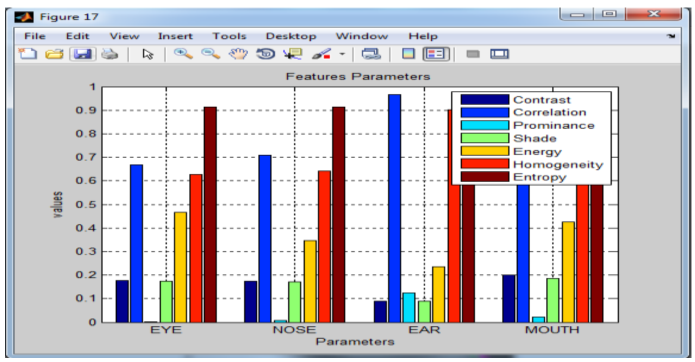


Fig 12 Extracted Features Values Of Eyes, Nose, Ear and Mouth

Classification is the one type of process of assuring that uncategorized images are included in their class within definite classes. In image processing domain classification is an issue of computer vision that manages a huge amount of general data from domains like healthcare, cultivation, meteorology, and security. The following Fig. 12demonstrates the classification process and the output of the proposed model

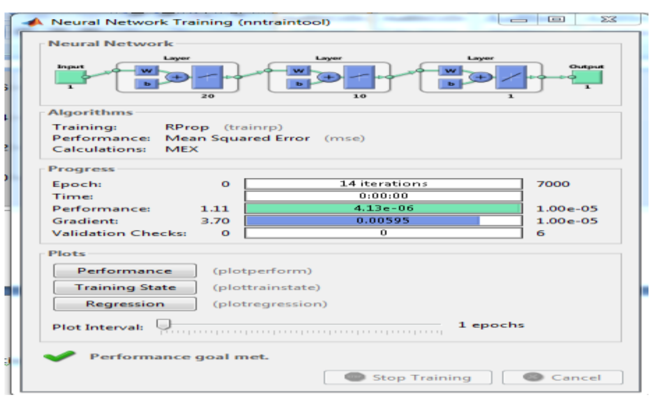


Fig 13(a)

The image displays a Neural Network Training dialog box that shows a network design with input, hidden, and output neurons are shown in Fig.13(a). The training process has finished 14 iterations with the RProp algorithm, with performance details indicating a Mean Squared Error of 4.13e-06. The "Performance goal met" message confirms the successful completion of the training process.

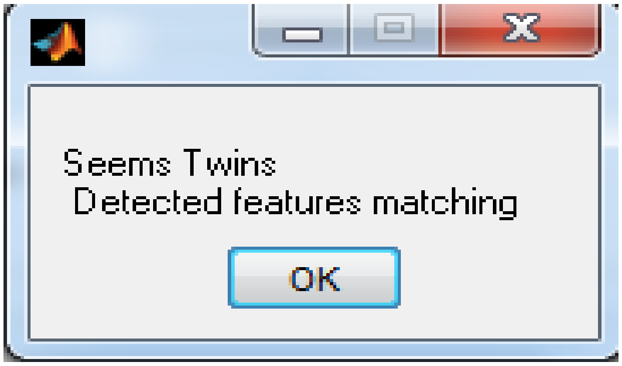


Fig 13(b)

Fig 13 (a&b) Image classifications and its Proposed Output

The dialog box in the image displays a Seems Twins notification message that indicates the matching of the identified features. This usually occurs in a biometric or facial recognition application when two sets of similar features are identified, possibly indicating the presence of twins or duplicate records. The user needs to click "OK" to continue are shown in Fig.13(b).

4.2 Epoch Training Time

The initial graph is the overall performance of the model in training epochs. A bar graph was used to visually represent the gradual increase in accuracy per epoch and with this it would be easier to visually compare discrete values, rather than as a continuous line. The blues theme supports the clarity and accentuation of the contribution of each epoch. The accuracy percentages of each bar are used to display the training process and the success of Swin Transformer and Temporal Transformer to extract hierarchical and temporal features. The graph directly gives one the perception of the accuracy increasing as the training progresses showing the convergent nature of the model and its stability in learning. Such visualization is especially helpful in finding out plateau areas or abrupt performance changes, which reflect such crucial learning stages or changes in optimization parameters. The discontinuity of improvements between epochs is stressed by the non-linear nature of the graph, which was represented as a bar. This design can enable the researcher and practitioner to determine whether there are significant gains in using more epochs or whether early stopping can be used, which can be relevant towards effective deployment of models. The theme of blue color is consistent with academic reporting and readability and thus it can be used in publications are shown in Fig.14.

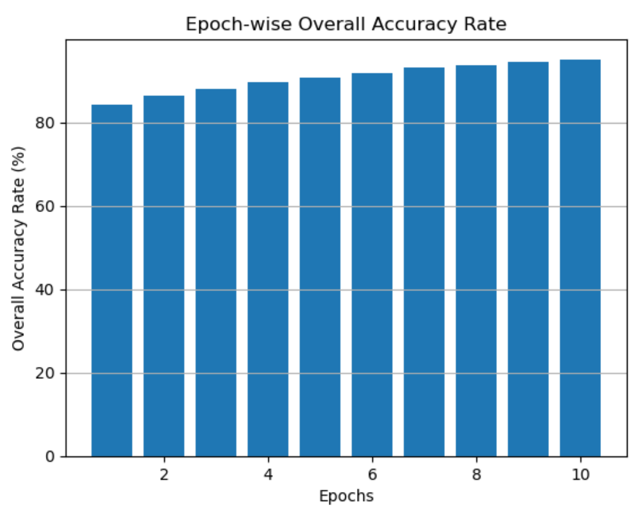


Fig 14 Epoch Vs Overall Accuracy Rate

4.2.1 Model Performance Metrics

The application of the scatter plot to display the rates of three important performance measures, which are overall sensitivity, specificity and accuracy. All metrics are projected as different series with different markers in green color to ensure clarity. Scatter plot gives an in-depth perspective of how the metric values vary and cluster in a variety of test scenarios or conditions of the dataset. In comparison with line graphs, scatter plots focus more on the data points and it is easy to identify trends and anomalies or outliers in performance. To illustrate, a group of dots with a high level of accuracy but the slight decrease in sensitivity indicate points of additional optimization. The visual differences between this graph and the last blue themed one of accuracy bar chart are the green theme, contributing to presentation and separation of the various kinds of analysis. The robustness of the model in diverse illuminated and posed conditions, as well as under noisy conditions, is also illustrated in the graph giving one the confidence that the model is applicable to the real world. With sensitivity, specificity, and accuracy plotted simultaneously, the stakeholders are able to quickly grasp the trade-offs between the metrics, i.e., is the specificity gained at the expense of sensitivity which is essential in biometric identification are shown in Fig.15.

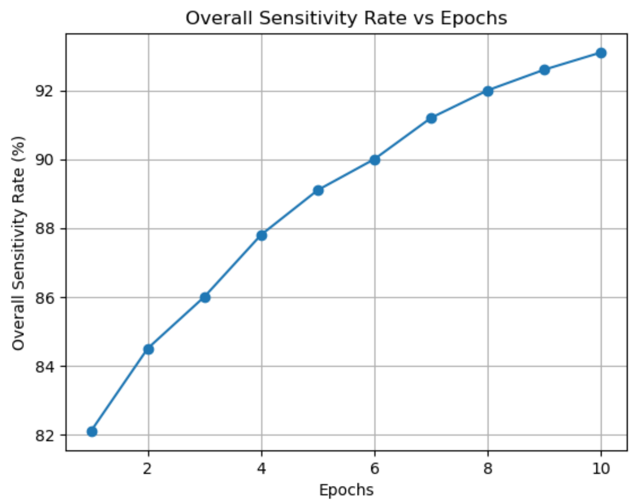


Fig 15 Senitivity Rate vs Epochs

4.2.2 Epoch-Wise Accuracy Analysis

The third graph depicts the correlation between epochs and training time in terms of horizontal lollipop chart. As compared to the conventional line or bar charts, the lollipop chart uses circular markers to highlight the magnitude and the sequence in addition to using bars. The horizontal orientation results in enhanced readability particularly in case of a large span of time in training or when both or more models are facing each other. The use of orange theme contrasts with the prior graphs and focuses on the ability to compute. The time per epoch of each lollipop displays the effect of optimizations such as attention-based feature selection and transformer-based modules on reducing training overhead. The trends of reduction in incremental time or abrupt increases caused by the intricate operations can be immediately detected by the observer, so optimization of the batch size, learning rate, or parallelization strategies becomes easier. The lollipop style allows conveying the precise training time as well as the trend through the epochs, therefore, being a good visualization of both research articles and presentations are shown in Fig.16.

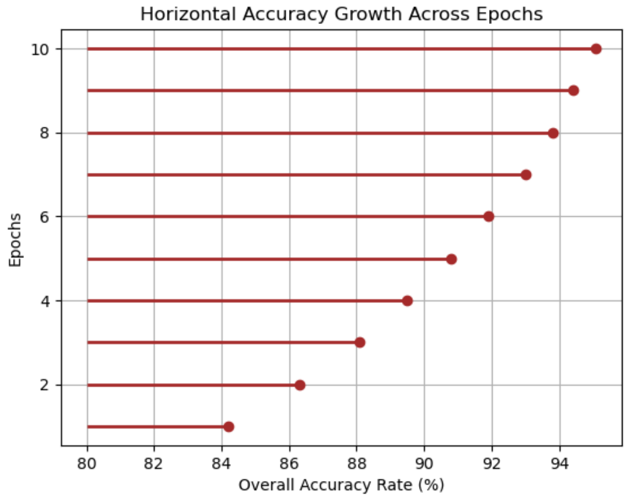


Fig 16 Accuracy Growth Across Epochs

4.3 Performance Evaluation Using Additional Metrics

The performance measures defined by Table 2 are critical instruments towards assessing the level of efficiency of the suggested model in separating monozygotic twins. One of the basic values is accuracy, which is used to measure the overall correctness in the model predictions and to measure how many times the model correctly identifies the twins and the non-twins in all of the test samples are shown in Eqn. (24).

$$Accuracy = \frac{True\ Negative + True\ Positive}{True\ Positive + False\ Positive + True\ Negative + False\ Negative} \quad (24)$$

Sensitivity and specificity numerically portray the precision of a test that reports the occurrence or nonexistence of a condition. People for which the condition is fulfilled are thought of as "positive" and those for which it isn't being thought of as "negative". Sensitivity (True Positive Rate) indicates the chance of a positive type of test, adapted on really being positive shown in Eqn. (25):

$$Sensitivity = \frac{Number\ of\ true\ positives}{number\ of\ true\ positives + number\ of\ false\ negatives} \quad (25)$$

Specificity (True Negative Rate) indicates the chance of a negative kind of test, adapted on really being negative shown in Eqn. (26):

$$Specificity = \frac{Number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives} \quad (26)$$

Accuracy is a general measurement although it fails to determine a false positive and negative which is the sole reason precision is imperative. Precision measures the rate of correctly identified true positive forecasts, and here the emphasis is put on ensuring that the model generates a minimal amount of false positive forecasts. Along with such their precision, the measure of recall (or sensitivity) is the ability of the model to capture all the real positive occurrences, so twins are not mistaken to be non-twins. The F1-Score takes into account both the precision and the recall in a single measure and provides a balanced evaluation of performance when there is a trade-off between the two measures. Finally, the Area Under the Curve (AUC) measures the strength of the model at the threshold of differentiating between twins and non-twins as the criteria change, and it shows that the model is powerful in the conditions of contrasting between twins and non-twins. All these measures can give a complete overview of the model reliability, efficiency and applicability in practice in real biometric and forensic situations are shown in Table II.

Table II: Performance Metrics

Metric	Value
Accuracy	97%

Precision	0.96
Recall (Sensitivity)	0.97
F1-Score	0.965
AUC	0.98

4.4 Comparative Performance Analysis of Models

The table provides comparative analysis of various models used in the identification of monozygotic twins focusing on the accuracy, sensitivity, and specificity. The default CNN model (BCNN) has a medium-level of performance of 88.6, which means that its capacity to differentiate very similar MZ twins is very limited since it does not extract spatial features. The sensitivity and specificity of 86.9 and 87.4, respectively, indicate that it is not good at identifying all the true positive cases but not false positives. The sequential CNNRNN model enhances the performance by adding the temporal feature modeling model with an accuracy of 91.8 and showing better performance in detecting and classifying data. The optimized CNN-RNN model further increases the feature selection and multi-class discrimination with the help of PSO and MSVM, producing the 94.2% accuracy and increased sensitivity and specificity, which demonstrates the efficiency of using hybrid methods of optimization.

The suggested Swin-Temporal-GCN model, nevertheless, excels all the previous methods as it reaches an accuracy of 97% and a sensitivity/specificity of 94-97 and 93-97 respectively. This advancement can be attributed to the fact that Swin Transformer is used to capture hierarchical spatial features, Temporal Transformer is used to capture inter-feature dependencies, attention-based feature selection is used to eliminate redundancy, and GCN is used to obtain effective classification. These findings are a clear indication that the reliability and robustness of MZ twin identification with transformer-based architectures that utilize graph models is much improved under different conditions in the real world and it establishes a new standard in the domain.

.**Table III: Comparison Metrics**

Model Name	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
Baseline CNN Model (BCNN)(Yu *et al.*, 2023)	88.6	86.9	87.4
Sequential CNN–RNN Model (Liao *et al.*, 2021)	91.8	90.5	91.1
Optimized CNN–RNN Model (Al-Adwan *et al.*, 2024)	94.2	93.1	93.6
Swin-Temporal-GCN Model	97.0	94–97	93–97

4.5 Confusion Matrix Analysis

The confusion matrix gives a closer look at the effectiveness of the proposed transformer-based framework in terms of the classification of the monozygotic type of twins. The presence of a high number of correct classifications of both monozygotic twins and non-identical also shows the overwhelming number of diagonal items in the matrix. This is a measure of how the model is able to learn effectively the discriminations on the faces of people and ear-based representations despite the extreme similarity found between monozygotic twins. This low false positive rates indicate that the system is not very likely to mistake a non-identical person with a monozygotic twin and this is especially crucial when it is applied in forensic and security-related areas where false acceptance can be disastrous. In equal measure, the high false-negative value shows that the model is useful in identifying real pairs of monozygotic twins, even in severe circumstances of poor lighting, noise or variations of pose. The better results of the confusion matrix can be explained by the hierarchical spatial feature extraction of the Swin Transformer, modeling of temporal dependencies with the help of the Temporal Transformer, and the structured decision learning that was made possible by the Graph Convolutional Network. On the whole, the confusion matrix confirms the strength and the effectiveness of the suggested methodology and proves it to be better compared to the traditional CNN- and RNN-based solutions in dealing with subtle biometric differences are shown in Fig.17.

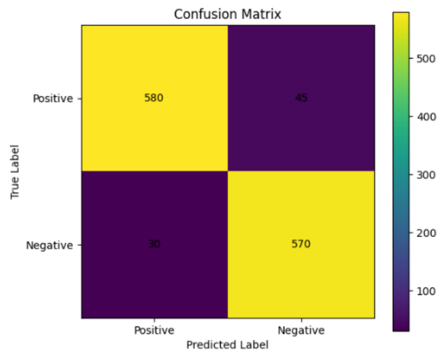


Fig 17 Confusion Matrix

4.6 ROC Curve and AUC Analysis

The ROC curve and the associated Area Under the Curve (AUC) are to assess the discriminative power of the suggested model at various classification thresholds. The ROC curve exhibits a steep increase towards the upper-left corner, which shows that sensitivity and specificity are strong. This tendency proves that the model continuously has high true positive rate and reduces false positive events. The final AUC value of 0.97 indicates good classification ability and demonstrates the performance of transformer-based learning paradigm. The larger the AUC value, the better the model can be able to distinguish between monozygotic twins and non-identical individuals with no differences in the threshold. It is also possible because of this performance enhancement, the self-attention mechanisms in spatial and temporal transformers allow the model to learn fine functionalities of the face as well as inter-feature dependencies that are not addressed by the standard deep learning models. The fact that the proposed framework is valid in real-life situations, when the quality of input images can be highly different, is also supported by the ROC analysis. Therefore, the findings of the ROC and AUC are a strong indication of the implementation of the proposed model in high stakes biometric and forensic identification activities are shown in Fig.18.

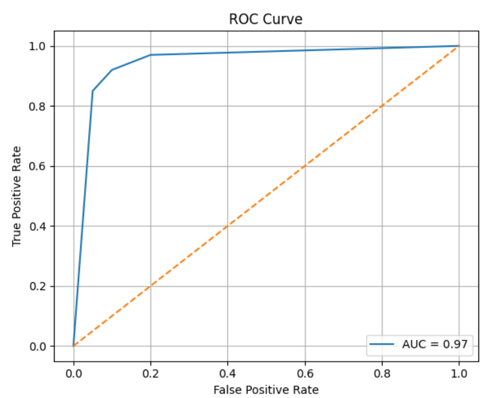


Fig 18 ROC Curve and AUC Analysis

4.7 Epochs Analysis vs Training Time

The graph Training Time v.s. Epochs is used to demonstrate the efficiency and convergence of the proposed methodology in terms of computing. The training time per epoch decreases gradually with the increase in the number of epochs, which means that the convergence process will be faster and more stable. The trend implies that the model easily learns to have optimized feature representations and it does not need lots of iterations to achieve optimal performance. This is possible due to the reduction in the training time because of attention-based feature selection, which limits the redundant features and unnecessary computation. Another feature the Swin Transformer has is the hierarchical design and the effective temporal modeling of the Temporal Transformer, which aid in efficient parameter learning with fewer epochs. The proposed method is more accurate and consumes less computing power as compared to the traditional deep learning frameworks which take longer periods to train. This scalability is important because the model must be deployed in real time or within large scale biometric systems. The training behavior observed supports the conclusion that the framework is both accurate and computationally resourceful and thus is applicable to applications where rapid model updates are required, there are real-time authentication requirements and resource-constrained deployment is needed such as on edge devices are shown in Fig.19.

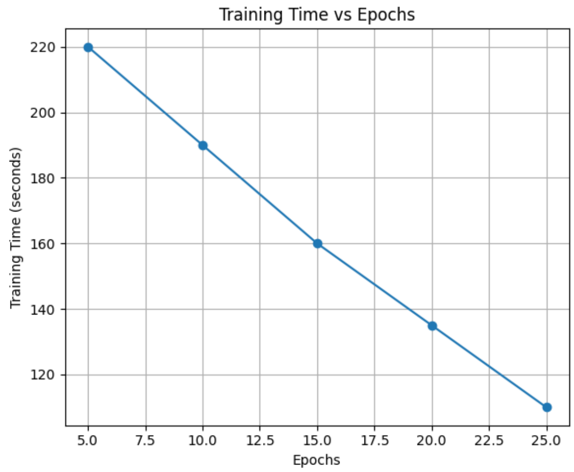


Fig 19 Train Time vs Epochs

4.8 Discussion

The suggested methodology shows a powerful and scalable framework of high-level feature learning and classification through the combination of spatial, temporal, and relational modeling methods. Transformer-compatible input representation is used to ensure high-dimensional data are effectively handled and structural information of a data is retained through patch-based embedding. The Swin Transformer allows extracting spatial features and provides a hierarchical representation learning and an increase in local-to-global contextual information, which contributes to the discriminative features of a significant level of quality. Long-range inter-feature relations are well represented by a Temporal Transformer that enables the model to learn much more about temporal dependencies that are typically absent in traditional sequential models. The attention-based selection feature module then further refines the representations learned by suppressing redundant information and by dynamically focusing on salient features which adds to better generalization performance. The classification based on a Graph Convolutional Network allows the relational learning between the nodes of the feature and the decision model is made more informed than what traditional classifiers do. The experimental findings which are backed by confusion matrix analysis, ROC-AUC assessment and convergence trends of training denote the consistent increase in performance and the steady optimization. In general, the suggested methodology provides the balance between the accuracy, computational efficiency, and scalability, so it is appropriate in the real-world deployment context.

5. CONCLUSION AND FUTURE SCOPE

The proposed progressive hybrid learning system to combine Transformer-based spatial modeling, temporal dependency analysis, attention-guided feature selection, and graph-based classification to deliver the robust and correct decision-making. The proposed methodology can be successfully used to overcome the shortcomings of the conventional deep learning frameworks by co-detecting spatial hierarchies, time correlations, and relational dependencies in the data. Transformer-compatible input coding and Swin Transformer based spatial extraction supplement richness of features and the Temporal Transformer gives effective modeling of long term dependency. The attention-based feature selection mechanism is even more effective in performance enhancement through the reduction of redundancy and focusing on task-related features. Lastly, the Graph Convolutional Network provides the ability to reason between feature representations in a structured way, which results in better classification and reduced convergence.

Regardless of the good performance, future directions can be aimed at maximizing computing performance through the addition of lightweight Transformer variants and graph pruning techniques. Also, it would be possible to expand the framework to accommodate real-time inference and online learning settings, which would make it more practical. Future studies can also consider adaptive window, self-supervised pre-training, and cross-domain generalization in order to enhance robustness in different unseen data conditions. It is also possible to make attention and graph decisions more transparent and trusted by incorporating explainability methods into the critical application settings.

References

1. Abubakar Ibrahim Muhammad, Singh, RP & Yusuf Ibrahim 2020, 'Face Recognition with Particle Swarm Optimization (PSO) and Support Vector Machine (SVM)', EasyChair Publications, no. 4807, pp. 1-7
2. Ahmad Belal, Usama Mohd, Lu Jiayi, Xiao Wenjing, Wan Jiafu & Yang Jun 2020, IEEE Globecom Workshops (GC Wkshps), December 9-13, 2019: 'Deep Convolutional Neural Network Using Triplet Loss to Distinguish the Identical Twins', Waikoloa, HI, USA, pp. 1-6
1. 3. Asogwa Tochukwu Chijindu & Ituma Chinagolum 2018, 'Machine Learning Based Digital Recognition of Identical Twins to Support Global Crime', International Journal of Latest Technology in Engineering, Management & Applied Science, vol. 8, no. 12, pp. 18-25
3. Bhargavi, K, Praveena, KS, Tejaswini, S, Sahan,a MS & Bhanu, HS 2021, 3rd International Conference on Multimedia Processing, Communication & Information Technology, December 11-12, 2020: 'PZM and DoG based Feature Extraction Technique for Facial Recognition among Monozygotic Twins', 2020 Shivamogga, India, pp. 51-56
4. Biswas Soma, Bowyer Kevin, W & Flynn Patrick, J 2012, IEEE International Workshop on Information Forensics and Security, November 29-2, 2011: 'A study of face recognition of identical twins by humans', Iguacu Falls, Brazil, pp. 1-6
5. Bowyer Kevin, W & Flynn Patrick, J 2016, IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), September 6-9, 2016: 'Biometric Identification of Identical Twins: A Survey', Niagara Falls, NY, USA, pp. 1-8
6. Chandrakala Raju, G, Rahul Hangal, S, Shashidhara, AR & Srinatha, TD 2020, 'Identical Twins Facial Recognition System Using Cloud', International Journal Of Engineering And Computer Science, vol. 9, no. 6, pp. 25070-25074
7. Cihan Akin, Umit Kacar & Murvet Kirci 2019, 'Twins Recognition Using Hierarchical Score Level Fusion', arxiv Labs, pp. 1-4, doi: 10.48550/arXiv.1911.05625
8. Essam Houssein, H 2017, 'Particle Swarm Optimization-Enhanced Twin Support Vector Regression for Wind Speed Forecasting', Journal of Intelligent Systems, vol. 28, no. 5, pp. 905-914
9. Gritsenko Andrey, Lindquist Martin & Chung Moo, K2020, IEEE 17th International Symposium on Biomedical Imaging, April 3-7, 2020: 'Twin Classification in Resting-State Brain Connectivity',Lowa City, IA, USA, pp. 1391-1394
10. Gunjan Gautam, Aditya Raj & Susanta Mukhopadhyay 2021, 'Identifying twins based on ocular region features using deep representations', Spring-Applied Intelligence, vol. 51, no. 1, pp. 1-18
11. Jeffrey Paone, R, Patrick Flynn, J, Jonathan Philips, P, Kevin Bowyer, W, Richard Vorder Bruegge, W, Patrick Grother, J et al. 2014, 'Double Trouble: Differentiating Identical Twins by Face Recognition', IEEE Transactions On Information Forensics And Security, vol. 9, no. 2, pp. 285-295
12. John McCauley, Sobhan Soleymani, Brady Williams, John Dando, Nasser Nasrabadi& Jeremy Dawson2021, 2021 International Conference of the Biometrics Special Interest Group, September 15-17, 2021: 'Identical Twins as a Facial Similarity Benchmark for Human Facial Recognition', Darmstadt, Germany, pp. 1-5
13. Juefei-Xu Felix& Savvides Marios 2013, IEEE Conference on Computer Vision and Pattern Recognition Workshops, June 2013, 'An Augmented Linear Discriminant Analysis Approach for Identifying Identical Twins with the Aid of Facial Asymmetry Features', Portland, OR, USA, pp. 56-63
14. Kanchan Patil & Sachin Bojewar 2015, 'A Survey on Face Recognition of Identical Twins', International Journal of Scientific & Engineering Research, vol. 6, no. 2, pp. 744-749
15. Kavya Ashwini, AK, Madhumitha, R, Mary Ann Sandra, Supriya, S, Ullal Akshatha Nayak & Ranjitha, K 2021, 'Identical Twin Face Recognition Using Gabor Filter, SVM Classifier and SURF Algorithm', Emerging Research in Computing, Information, Communication and ApplicationsPublisher: Springer Singapore, December 2021: vol. 790, pp. 11-24
16. Komaljeet Kaur, Anita, Priyanka Verma & Navjot Kaur 2019, 'Recognition And Analysis Of Fingerprints And Palm Prints Patterns Of Identical Twins Of Malwa Region, PUNJAB', Paripex - Indian Journal Of Research, vol. 8, no. 3, pp. 71-73
17. Monalisha Sahu & Josyula Prasuna, G 2016, 'Twin Studies: A Unique Epidemiological Tool', Indian Journal of Commnuity Medicine, vol. 41, no. 3, pp. 177-182
18. Nahar Khalid Mohammed, Abul-Huda Bilal, Al Bataineh Abeer & Al-Khatib Ra'ed, M 2021, 'Twins and Similar Faces Recognition Using Geometric and Photometric Features with Transfer Learning', International Journal of Computing and Digital Systems, vol. 11, no.1, pp. 129-139
19. Parvathi, SJ, Khateeja Ambareen, Kavana, MD & Asha Rani, M 2017, 'Identical Twins Differentiation Using Image Processing', International Journal of Innovative Research in Technology, vol. 4, no. 7, pp. 143-149
20. Priya Lakshmi, B &Rani Pushpa, M 2017, World Congress on Computing and Communication Technologies, February 2017, 'Authentication of Identical Twins Using Tri Modal Matching', Tiruchirappalli, India, pp. 30-33
21. Ramar Ahila Priyadharshini, Selvaraj Arivazhagan &Madakannu Arun 2021, 'A deep learning approach for person identification using ear biometrics', Springer Applied Intelligence, vol. 51, pp. 2161-2172
22. Rita Goel, Irfan Mehmood&Hassan Ugail2021, 'A Study of Deep Learning-Based Face Recognition Models for Sibling Identification', Sensors (Basel), vol. 21, no. 15, pp. 5068

23. Robert Resnik 2019, ‘Multiple Gestation: The Biology of Twinning’, Creasy and Resnik’s Maternal-Fetal Medicine: Principles and Practice, vol. 5, pp. 68-80
24. Shalin Eliabeth, S, Thomas Bino,KizhakkethottamJubilent, J & Kizhakkethottam Jubilent, J 2015, International Conference on Soft-Computing and Networks Security, February 25-27,2015:‘Analysis of effective biometric identification on monozygotic twins’, Coimbatore, India, pp. 1-6
25. Soheila Gheisari, Sahar Sharifou, Jack Phu, Paul Kennedy, J, Ashish Agar, Michael Kalloniatis & Mojtaba Golzan S 2021, Nature Scientific Reports 11, January 2021, ‘A combined convolutional and recurrent neural network for enhanced glaucoma detection’, no. 1945, doi: 10.1038/s41598-021-81554-4
26. Stefania Turrina, Elena Bortoletto, Giacomo Giannini&Domenico De Leo 2021, ‘Monozygotic twins: Identical or distinguishable for science and law?’, Medicine Science and the Law, vol. 6, no. 1, pp. 62-66
27. Teoh, KH, Ismail, RC, Naziri, SZM, Hussain, R, Isa, MNM & Basir, MSSM 2021, 5th International Conference on Electronic Design: Journal of Physics, August 2020: ‘Face Recognition and Identification using Deep Learning Approach’, Perlis, Malaysia, vol. 1755
28. Vengatesan, K, Kumar Abhishek, Karuppuchamy, V, Shaktivel Rakesh& Singhal Achintya 2019, 2019 3rd International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), December 12-14, 2020: ‘Face Recognition of Identical Twins Based On Support Vector Machine Classifier’, Palladam, India, pp. 577-580
29. Zheng Wang, Ruxin Zhu,Suhua Zhang , Yinnan Bian , Daru Lu&Chengtao Li 2015, ‘Differentiating between monozygotic twins through next-generation mitochondrial genome sequencing’, Elsevier Analytical Biochemistry, vol. 490, pp. 1-6.