DEVELOPMENT OF A GENETIC ALGORITHM BASED GEODESIC ACTIVE CONTOUR FOR IRIS BASED ETHNICITY PREDICTION SYSTEM

.

ABSTRACT

The developments in science and technology have made it possible to use biometrics in applications where it is required to establish or confirm the identity of individuals. Among all possible biometric characteristics, the use of iris texture for recognition of individuals has been proven to be highly reliable. However, existing iris prediction systems have suffered from inability to handle more constrained acquisition (processing non-ideal iris images), high processing time and inappropriate parameter settings which usually results in inaccurate segmentation and poor classification results. This research therefore developed an improved segmentation and classification algorithms for iris-based ethnicity prediction system featuring the three major tribes in Nigeria. Six hundred (600) iris images from three major tribes in Nigeria (Yoruba, Hausa and Ibo) were locally captured for the database. Genetic Algorithm based Geodesic Active Contour (GAGAC) and standard Geodesic Active Control (GAC) were used for iris segmentation while Standard Support Vector Machine (SVM) and Galactic Swarm Optimisation SVM (GSOSVM) was used for iris classification. GAGAC and GSOSVM were used in the designing of the iris-based ethnicity prediction system at segmentation and classification stage. The developed iris-based ethnicity prediction system gave an improved predictive performance over the conventional one. The developed system can be used in different areas where higher security authentication is required.

Keywords: Segmentation, classification, Support Vector Machine, ethnicity, security

1. INTRODUCTION

Ethnicity has been discovered not to have any standard scientific definition for its measurement hence a fuzzy variable to automate and measure. Without a standard scientific definition, the tendency of misclassifying individuals into ethnic separations is very high (Kaneshiro *et al.* 2011). While individual may adopt new language, change religion or alter their places of residence, changing individual ethnicity is much more difficult, even when it is based upon such characteristics (Chandra 2006).

It has been observed that in this part of the world, identification of individual's ethnicity does not go beyond physical factors which in most cases can be manipulated or spoofed to carry out nefarious activities.(Falohun et al. 2012) When identification does not go beyond what is seen and touched, then such means of identification can be faulted and the integrity of such is questionable; as a result, the interest in biometrics for identification has increased to enhance security and the process of verifying individual's identity (Kim *et al.* 2012).

Biometric techniques include identification based on physiological characteristic such as face, fingerprints, geometry, hand geometry, hand veins, palm, iris, retina, ear and voice and behavioral traits like gait, signature and keystroke dynamics. These traits are unique to every individual although the physiological ones are more reliable and stable than the behavioral because they are non-alterable (Makinde *et al.* 2019). In biometrics, there are three main types of modalities: hard, soft and hidden. The hard ones are considered classic such as fingerprint, iris, face or signature among others. The soft are traits like height, weight, age, gender, ethnicity, skin color, voice and eye color. The hidden ones, also called intrinsic, are based on medical data, as bio-signals, Magnetic Resonance Imaging (MRI) images or X-ray images (Yamakawa *et al.* 2007).

Among the various biometric technologies (fingerprint, iris, face, palm print, hand geometry, gait and many more), iris is highly accurate, reliable and fool-proof because irises are highly distinctive and of stable characteristics throughout lifetime. (Falohun et al., 2010) Just like fingerprints, irises are unique to each individual and have little similarities between ethnic groups, The iris is the colored portion of the eye visible around the pupil and it is covered by the cornea. And it is the only internal part of the body visible from the outside. Iris is an internally protected organ whose texture is stable from birth to death. It is one of the most secured mechanisms when security is concerned (Lagree and Bowyer 2011). Many researchers have been working from the last decade to extend the application of iris recognition system in several areas like tracing criminals, terrorist and missing children; ethnicity, age and gender prediction; accurate diagnosis of eye defect and ascertaining state of health (Abbasi *et al.* 2013).

A significant number of iris segmentation techniques have been proposed in the literature. Most popular techniques are based on the use of: Integro-differential operator, Hough transform and Active Contour (Samir and Arun 2009). The performance of an iris segmentation technique is greatly dependent on its ability to precisely isolate the iris from the other parts of the eye. Integro-differential operator and Hough transform rely on curve fitting approach on the edges in the image and perform better with good quality, sharply focused iris images. Also, Active Contour cannot naturally handle changes in the topology of the evolving contour. However, under challenging conditions (non-uniform illumination, motion blur, off-angle), the edge information may not be reliable (Luo *et al.* 2007). It was reported that most failures to match in iris recognition system result from inaccurate iris segmentation.

Most of the existing segmentation algorithms (such as Integro-differential operator, Hough transform and Active Contour) assumed that the iris is circular and elliptical in shape resulting in under-segmentation and over-segmentation (Lagree and Bowyer 2011; Zhang *et al.* 2011; Gugulethu *et al.* 2016; Latinwo *et al.* 2016; Singh *et al.* 2017; Latinwo *et al.* 2018). Recent segmentation algorithm like Active Shaped Model, Randomized elliptical Hough Transform and Active Contours and Geodesic Active Contour (GAC) assumed non circular and non elliptical shape of iris. However, GAC supports accurately estimating the radius of the iris and its centre thereby lessens the concerns related with the traditional models but the time required to segment the iris is high though it gives better accuracy (Minal *et al.* 2012).

This research optimized GAC as segmentation algorithm using an adaptive strategy and a global optimization technique, Genetic Algorithm (GA) to automatically determine the regularization parameters rather than the conventional manual method for each iris image in the dataset which reduces the segmentation time and increases the accuracy.

Furthermore, for an improved iris based ethnicity prediction system with higher accuracy, various methods were usually used for classification namely: Hamming distance, Euclidean distance, Normalized correlation, Support Vector Machine (SVM) and Artificial Neural Network among others but SVM is a machine learning technique based on structural risk minimization (minimizing classification error) and widely adopted in various fields of classification because of its robustness and ability to learn both simple and highly complex classification models even though it has some limitations which can be easily overcome through optimization of its parameters. The most common problem encountered in setting up the SVM model was how to select the kernel function and its parameter values. Inappropriate parameter settings lead to poor classification results (Keerthi and Lin 2003). Hence, global optimization meta-heuristic technique, Galactic Swarm Optimization was used to optimized the Support Vector Machine (SVM) for parameter determination

The performance of an iris segmentation technique is greatly dependent on its ability to precisely isolate the iris from the other parts of the eye. Most failures to match in iris prediction system result from inaccurate iris segmentation. Most of the existing segmentation algorithms assume that the iris is circular or elliptical in shape resulting in under-segmentation or over-segmentation (Lagree and Bowyer 2011; Zhang *et al.* 2011; Gugulethu *et al.* 2016; Latinwo *et al.* 2016; Singh *et al.* 2017; Latinwo *et al.* 2018). Recent segmentation algorithm such as Geodesic Active Contour (GAC) assumes non circular and non elliptical shape of iris but the time required to segment the iris is high though it gives better accuracy (Minal *et al.* 2012).

Furthermore, Support Vector Machine (SVM) which is one of the state-of-the-art classification algorithms because of its robustness and ability to learn both simple and highly complex classification models suffer inappropriate parameter settings leading to poor classification results thereby reducing and preventing its usage in many real-life applications where classification accuracy is ultimate (Keerthi and Lin 2003).

The aforementioned problems in GAC algorithm necessitated optimization using Genetic Algorithm (GA) to automatically determine the regularization parameters rather than traditional manual method for each iris image in the dataset. Also, the major problem of support vector machine which is inappropriate parameter setting necessitated the employment of good optimizing performance of galactic swarm optimization (GSO) to properly tune the parameters of SVM and the RBF kernel in order to boost the classification performance.

The use of iris for recognition of individual has been proven to be highly reliable among all possible biometric characteristics (Samir and Arun 2009). However, existing iris prediction systems have suffered from inability to handle more constrained iris acquisition (processing non-ideal iris images considering the fact that iris is non-circular and non-elliptical in shape), high processing time and inappropriate parameter settings which usually results in inaccurate segmentation and poor classification results. This necessitated the designing of an iris-based ethnicity prediction system using the locally captured irises of three major tribes in Nigeria (Yoruba, Hausa and Ibo). The formulated segmentation and classification algorithms were employed at both segmentation and classification stage of the system.

The aim of this research is to develop a Genetic Algorithm based Geodesic Active Contour (GAGAC) segmentation approach and Galactic Swarm Optimization Support Vector Machine (GSOSVM) classification algorithm for iris-based ethnicity prediction system.

2. LITERATURE REVIEW

2.1 Ethnicity

Ethnicity can be defined as the fact or state of belonging to a social group that has a common national or cultural tradition. The definition for race is sometimes equated with ethnicity; as human race can be defined as a group of people with certain inherited features that distinguish them from other groups of people. All men of whatever race are classified by the anthropologist or biologist as belonging to one specie, homo-sapiens. Ethnicity can also be defined as a vast group of people loosely bounded together by historical, socially significant elements of their morphology and/or ancestry. It can serve as the connections between physical features, races and personal characteristics (Ratcliffe 2010).

A variety of means are available for identifying a person, in order to associate data with them. Such as name-based, code-based, knowledge-based, token-based and biometric-based techniques (Romen and Yampoolskiy 2008)) (Sabhanayagam *et al.* 2018).

2.2 Related works on analyzing hard biometric to predict soft biometric

Lagree and Bowyer (2007) confirmed the possibility of predicting ethnicity based on iris texture. This is possible if there are similarities of the iris texture of a certain ethnicity, and these similarities differ from ethnicity to ethnicity and obtained 91% correct Asian / Caucasian ethnicity classification.

Michael *et al.* (2015) explored an approach to gender prediction from iris images using different types of features (including a small number of very simple geometric features, texture features and a combination of geometric and texture features) and a more versatile and intelligent classifier structure with an accuracy of up to 90% in the BioSecure Database.

Orike *et al.* (2016) proposed a gender and ethnicity identification system in Nigeria using fingerprint technology to capture the fingerprints of a group of people in other to identify and verify their identities through the use of trained classifiers. The result showed that over 98% test cases accurately identified persons ethnicity and gender.

Juan *et al.* (2016) predicted gender directly from the same binary iris code that could be used for recognition and found that the information for gender prediction is distributed across the iris, rather than localized in particular concentric bands and was able to achieve 89% correct gender prediction using the fusion of the best features of iris code from the left and right eyes.

Latinwo *et al.* (2016) analyzed iris texture by performing dimensionality reduction and extracting unique feature codes of images for efficient ethnicity classification using iris images from African and two Asian datasets.

Singh *et al.* (2017) performed ethnicity and gender classification on iris images by presenting a novel supervised autoencoder based approach. The model was evaluated on two datasets each for ethnicity and gender classification. The results obtained demonstrate its effectiveness in comparison to existing approaches and state-of-the-art methods

Sarfaraz *et al.* (2018) predicted human ethnicity from facial images using neural networks. The research is done for three major ethnicities: Mongolian, Caucasian and Negro. The accuracy of the model obtained through artificial neural network is 82.4% whereas the accuracy obtained by deploying convolution neural network is 98.6%.

Aworinde and Onifade (2019) evaluated the performance of feature extraction techniques that can determine ethnicity of an individual using fingerprint biometric technique and deep learning approach. Hence, fingerprint images of one thousand and fifty-four (1054) persons of three different ethnic groups (Yoruba, Igbo and Middle-Belt) in Nigeria were captured.

From the researches reviewed in this sub-section, it can be deduced that hard biometric traits have been analyzed by several researchers to predict soft biometric for the purpose of identification and classification.

This research will address the issue of iris image segmentation for an improved texture extraction, as it is reported that most match failures in iris recognition system result from inaccurate iris segmentation. However, iris segmentation is the most time-consuming step in the iris recognition system and so become the bottleneck in real time environments. Iris segmentation is difficult task and faces some challenges such as specular reflection, contrast enhancement, blurred images and occlusion. Two main challenges of iris segmentation of realistic eye images are addressed: segmentation accuracy and processing speed.

A significant number of iris segmentation techniques have been proposed in the literature. Most popular techniques assumed that irises are circular or elliptical in shape, hence focusing on determining model parameters that best fit these hypothesis in the segmentation process resulting in challenging processing of non-ideal iris images resulting in under-segmentation and over-segmentation (Arun and Samir 2006). They mostly are based on using an Integro-Differential Operator, Hough transform and Active Contour. The performance of an iris segmentation technique is greatly dependent on its ability to precisely isolate the iris from the other parts of the eye. Integro-differential operator and Hough transform techniques rely on curve fitting approach on the edges in the image. Such an approach researches well with good quality, sharply focused iris images. However, under challenging conditions (e.g., nonuniform illumination, motion blur, off-angle, etc.), the edge information may not be reliable (Samir and Arun 2009).

It is reported that most failures to match in iris recognition system result from inaccurate iris segmentation. For instance, even an effective feature extraction method would not be able to obtain useful information from an iris image that is not segmented accurately. For better performance of the iris recognition system correct segmentation method plays vital role.

2.3 Related Works on Ethnicity Prediction

Ethnicity classification is an old topic in social science. It is often assumed to be a fixed trait based on ancestry. In natural science, few attempts have been made to perform automatic ethnicity classification based on human images. (Qiu *et al.* 2006). A number of researches have been done on ethnicity identification and classification using different approaches and subsequently achieving varying results. The corresponding limitation of such previous researches are identified and presented in this section.

Qiu *et al.* (2006) used global texture information of iris images to develop a novel ethnic classification method. Their research investigated Asian and Non-Asian iris images motivated from the arguement that the characteristics of their iris images are different. Iris database from CASIA v2 (2400 Asian eye images), UPOL (384 European eye images) and UBIRIS (1198 European eye images) were used. Gabor filters and Iris Textons technique was used for classification with 89.95% and 91.02% recognition accuracy. It was observed that the selection of samples form all the databases already introduced a natural bias towards the result. The trained algorigthm separated the images based on differences in lighting rather than iris texture. Howerve, the research was limited to Caucasians subjects and the algorithm separated the irises based on lighting rather than texture.

Lagree and Bowyer (2007) examined the possiblity of predicting ethnicity based on iris texture. An iris image data set representing 120 persons and 10-fold person disjoint cross validation was used and 91% correct Assian/Caucassian ethnicity classification was obtained. It was observed that the research only looked at very broad ethnicity classification whereas more research could be done to examine more categories. Sequnetial Minimal Optimization was used for feature extraction while SVM was used for classification. It was as well observed that the performance of the classifier employed has not been tested on subjects of multiple ethnic background.

Tariq *et al.* (2009) carried out research on gender and ethnicity identification using silhouetted face profiles. In the research, 441 tested images show that silhouetted face profiles have a lot of information as it relates to ethnicity identification. The research employed shape context-based matching for classification and achieved an average accuracy of 71.66% for ethnicity classification. It was observed that there exists varying degree of accuracy on different ethnic groups; while some ethnic groups attained a relatively low average accuracy, some others attained much higher average accuracy in classification.

Lyle *et al.* (2010) carried out Soft Biometric Classification using Periocular Region features with the goal to investigate the effectiveness of local appearance features extracted from the periocular region images. Gender and ethnicity information were extracted from the periocular region images using grayscale pixel intensities and periocular texture composed by Local Binary Patterns as features while SVM was used as classifier. 91% accuracy on ethnicity was achieved using 5-fold cross validation. The research is limited to only two classes of ethnicity which is insufficient for classification.

Demirkus *et al.* (2010) presented a prototype video tracking and person categorization system using face and person soft biometric features to tag people while tracking them in multiple camera views. In the research, Support Vector Machine (SVM) classifiers were used for the purpose of classification. However, it is observed that only frontal faces could be captured for classification which seem not to be sufficient for a system like this.

Zarei and Duxing (2012) predicted ethnicity of individuals from iris textures using artificial neural netresearchs.A dataset acquired using LG 4000 sensor camera was used with 120 subjects each with 10images of 5 from left and 5 from right eyes. The result showed 93.3% with netresearch 1 (person disjoint) and 97.7% with netresearch 2 (non-person disjoint).Iit was observed that the weight and biases of the netresearch are only recorded and saved when the validation error was at the minimum thereby introducing a potential bias (when iterating using netresearch 2) to the overall perfomance of the model.

Maneet *et al.* (2017) performed ethnicity and gender classification on iris images by presenting a novel supervised autoencoder based approach (Deep Class-Encoder). It was evaluated on two datasets (ND-Iris-0405 and a combined multi ethnicity Iris dataset(CASIA, IMP and ND_Iris-0405 cocnsisting of images of Chinese,Indian and Caucasian) each for ethnicity and gender classification. The model achieved a classification accuracy of 97.38% (with Neural Network). It was observed that the present of artifacts like hair bangs renders samples challenging for ethnicity leading to high misclassification rate.

Latinwo *et al.* (2018) classified iris images from Nigeria, China and Hong Kong origin using Self-Organizing Feature Maps(SOFM) blended with Principal Component Analysis based feature extraction and prepocessed using Hough transform and Histogram Equalisation. Left and right irises of 240 subjects constituting 480 imges were acquired online from CUIRIS (Nigeria), CASIA (China) and CUHK (Hong Kong) datasets. 93.75% Correct Classification Rate was obtained with varying thresholds. It was observed that the research was not performed on a verys large scale of data. More of the strength and weaknesses of SOFM algorithms can be investigated using large datasets.

Sarfaraz *et al.* (2018) predicted ethinicity for three major ethnicities: Mongolian, Caucasian and Negro with Neural Network using FERET dataset. The accuracy of the model obtained through artificial neural network is 82.4% whereas the accuracy obtained by deploying convolution neural network is 98.6%. It was observed that the cost in terms of time required for feature extraction and training that network was much and the research could very well be extended for other known ethnicities.

Aworinde and Onifade (2019) proposed a soft computing model of soft biometric traits for gender and ethnicity classification using Gabor filter and KNN for feature extraction and classification. Histogram equalization was used for preprocessing of the captured fingerprint dataset. Result from the research proved to be 96% accuracy in predicting person’s ethnicity and gender with an average recognition time of less than 2seconds. However, the research employed appearance based technique of identification in that focus was majorly on the ridge arrangement and type for classification into various ethnic divides; equally, the dataset used for the research is quite small for validation of the model.

Wong *et al.* (2020) predicted the ethnicity of Canadians with varying performance by specific ethnic Categories employing the automated machine learning approach using only name and census location features. Using census 1901, the multiclass and binary class classification machine learning pipelines were developed. The 13 ethnic categories examined were Aboriginal (First Nations, Me´ tis, Inuit, and all-combined)), Chinese, English, French, Irish, Italian, Japanese, Russian, Scottish, and others. Machine learning algorithms included regularized logistic regression, support vector, and naïve Bayes classifiers. The census had 4,812,958 unique individuals. For multiclass classification, the highest performance achieved was 76% F1 and 91% accuracy. For binary classifications for Chinese, French, Italian, Japanese, Russian, and others, the F1 ranged 68–95% (median 87%). The lower performance for English, Irish, and Scottish (F1 ranged 63–67%) was likely due to their shared cultural and linguistic heritage. Adding census location features to the name- based models strongly improved the prediction in Aboriginal classification (F1 increased from 50% to 84%).

Bessudnov *et al.* (2021) develop a machine learning classifier that predicts perceived ethnicity from data on personal names for major ethnic groups populating Russia. We collect data from VK, the largest Russian social media website. Ethnicity has been determined from languages spoken by users and their geographical location, with the data manually cleaned by crowd workers. The classifier shows the accuracy of 0.82 for a scheme with 24 ethnic groups and 0.92 for 15 aggregated ethnic groups. It can be used for research on ethnicity and ethnic relations in Russia, in particular with VK and other social media data.

Fangzhou Xie (2022) provided a new R package, rethnicity for predicting ethnicity based on names. The Bidirectional Long Short-Term Memory (Bi-LSTM), a recurrent neural network architecture commonly used for natural language processing, was chosen as the model for our study. The Florida Voter Registration was used as the training and testing data. Special care was given for the accuracy of minority groups by adjusting the imbalance in the dataset. The models were trained and exported to C++ and then integrated with R using Rcpp. Additionally, the availability, accuracy, and performance of the package were compared with other solutions.

It was observed from review of related researches that most researches done on ethnicity identification were approached from facial/iris/fingerprints biometric points of view for the purpose of ethnic classification and which are not without major drawbacks as identified by Latinwo *et al.* (2016) ranging from susceptibility to poor image quality which is associated with high failure enrol rate, environmental dependency in terms of illumination, high cost of iris scanners amongst other factors. Also, a person’s finger changes size or form/pattern over time or placed in different directions and the fingerprint scanner does not take this into consideration. When these changes occur, an individual can have difficulty identifying themselves.

However, it is abundantly evident that in all the researches reviewed on using iris, it was observed that improper segmentation always lead to incorrect feature extraction thereby reducing the recognition performance and most adopted techniques in literature assumed that the papillary, the limbic and eyelid boundaries are circular and elliptical in nature. Hence, they focused on determining model parameters that best fit these hypothesis (Arun and Samir 2006). Only few in the literature do not assume circular or elliptical boundaries among which is GAC (Geodesic Active Contour) which also face the problem of over-segmentation.

In this research, high segmentation time which greatly reduces the performance of GAC was addressed and overcome by developing a Genetic-Algorithm based Geodesic Active Contour (GAGAC) where the regularization parameters were determined automatically instead of manually for each image. Galactic Swarm Optimization (GSO), a new global optimization metaheuristic technique was used to improve the classification performance of Support Vector Machine (SVM) in parameter determination (GSOSVM).

Proper parameters setting can improve the SVM classification accuracy. The parameters that should be optimized include penalty parameter C and the kernel function parameters such as the gamma (g) for the radial basis function (RBF) kernel. The penalty parameter C determines the trade-off between minimizing the training error and maximizing a classification margin. Moreover, the kernel parameters define the nonlinear mapping from the input feature space to a high-dimensional feature space. One of the research objectives is to optimize the parameters without degrading the SVM classification accuracy.

2.4 Performance Evaluation Metrics

Some well known biometric performance evaluation metrics that can be examined in evaluating a prediction system are:

i. Segmentation Accuracy (SA): This is the percentage of images that are correctly segmented.

ii. Segmentation Time (ST): This is the total time use to segment the entire image in the system in seconds.

iii. False Positive Rate (FPR): is the proportion of negative cases incorrectly identified as positive cases in the data (i.e the frequency at which a system incorrectly categorizes negative as positive). It measures when it's actually no, how often does it predict yes?.

iv. Sensitivity: is the probability that the test indicates the presence of an iris image belonging to an ethnicity in a created database.

v. Specificity: is the probability that the test indicated the presence of an iris image belonging to an ethnicity but tested negative for that ethnicity in the created database.

vi. Precision: is used to measure how often the system predict yes to a biometric trait when it was actually no.

vii. Accuracy: Overall, how often is the classifier correct? (Newberg, 2006)

The selection of SVM parameters is actually an optimization problem in which search algorithms are used to find the best configuration of parameters for given problem (De-Miranda *et al.* 2012 and Gaspar *et al.* 2012). The use of evolutionary algorithm for parameter optimization is very much faster and often gives better results (Friedrichs and Igel 2005 and Rossi and de Carvalho 2008). There are several methods to adjust the SVM parameter among which are genetic algorithm (GA) and particle swarm optimization (PSO) and Galactic Swarm Optimization (GSO).

3. METHODOLOGY

3.1 Research Approach

For a reliable iris-based ethnicity prediction system, constant improvement is required in segmentation and classification of the basic stages involved in iris recognition system, which has been confirmed to be the most accurate personal biometric identification scheme. There are five basic stages targeted towards achieving the aim of this research:

i. Construction of locally acquired iris database.

ii. Formulation of GAGAC and GSO-SVM Algorithms

iii. Designing an Iris-based Ethnicity Prediction System using the two formulated Algorithms

iv. Software implementation and testing using Matrix Laboratory (MATLAB R2020a) software.

v. Performance evaluation of the segmentation algorithms in (ii) using Segmentation Accuracy (SA), Segmentation Time (ST) and the performance of the developed iris-based ethnicity prediction system evaluated using, False Positive Rate (FPR), Sensitivity, Specificity, Precision, Accuracy, classification time and statistical t-test of significance.

3.2 Data Collection and Construction of Locally Acquired Iris Database

A unique and indigenous iris data set was collected for the selected tribes (Yoruba, Ibo and Hausa). The dataset was collected in an uncontrolled environments, different light illuminations and varied distance capturing both left and right irises of one hundred (100) subjects per tribe. The data set for Yoruba was collected in Osun State. Selected set of subjects was interviewed to ensure that they are pure Yorubas comprising of Oyo, Ijesa, Ijebu, Ekiti, Igbominas. The dataset for Ibos was collected at Abia, Enugu and Anambra states to ensure that the selected, interviewed subjects were pure Ibo people while the dataset for Hausa was collected at Sabo area at Osun State (due to the insecurity in the country) using the influence of the Seriki of Hausa in Osogbo who ensure total cooperation from the subjects. In the database, All Yoruba Ibo and Hausa dataset was named using YORL001 – YORL100, IBOL001 – IBOL100 and HAUL001 – HAUL100 for left irises while YORR001 – YORR100, IBOR001-IBOR100 and HAUR001 – HAUR100 was used to name the right irises respectively.

Altogether, six hundred (600) iris images comprising of both left and right irises of 100 subjects from three major tribes in Nigeria (Yoruba (200 sample images), Hausa (200 sample images) and Ibo (200 sample images)) were locally captured using CMITECH Imager (camera) with high resolution in an uncontrolled environment to construct a database used in the development of the iris based ethnicity prediction system. Some desirable properties were considered during image acquisition:

i. High resolution and good sharpness to enable accurate segmentation

ii. Good lighting condition i.e under controlled light intensity to prevent image distortion.

An application software (CMIRIS SDK Version 1.2.6 for Windows OS) was used to acquire and pre-process the iris image before actual image enhancement were employed and the following algorithm steps was applied:

i. Capture iris: to capture a good image on the screen

ii. Re-snap image: to re-snap an image that was wrongly or not well captured

iii. Crop iris: to manually cut off other parts of the eyes and the face mistakenly captured to reduce captured image to the required one

iv. Gray Scale: to change the image to gray scale form.

v. Save image: All the captured images were saved into work environment with their respective iris identity (Yoruba, Ibo and Hausa) to afford the iris image to be further processed. Figure 3.1 (a - c) presented selected original iris images of Yoruba, Ibo and

Hausa people with dimension 640x480 while Figure 3.2 (a - c) presented selected cropped iris images of Yoruba, Ibo and Hausa with dimension 314x353.

3.3 Formulation of GAGAC and GSO-SVM Algorithm

Two optimized algorithms were formulated in this research: GAGAC, which eventually overcome the problem of over-segmentation in Geodesic Active contour (GAC) using the adaptive strategy of Genetic Algorithm (GA) and GSOSVM which

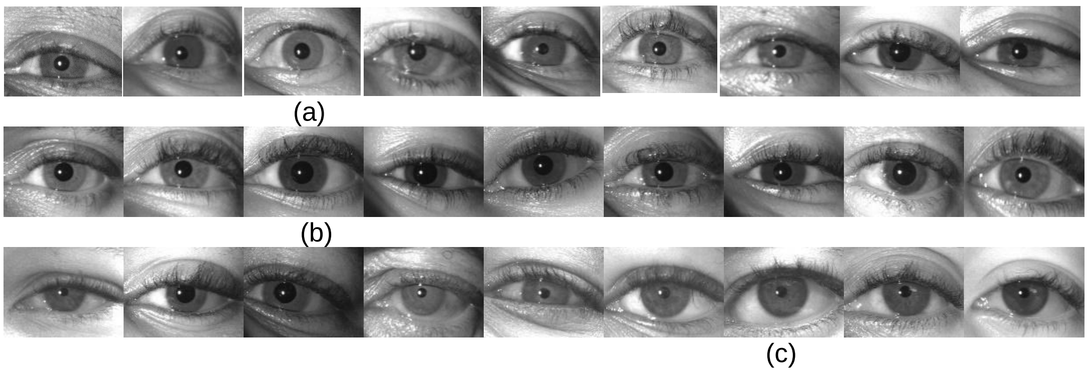


Figure 3.1: Selected original iris images (a) Yoruba (b) Ibo and (c) Hausa.

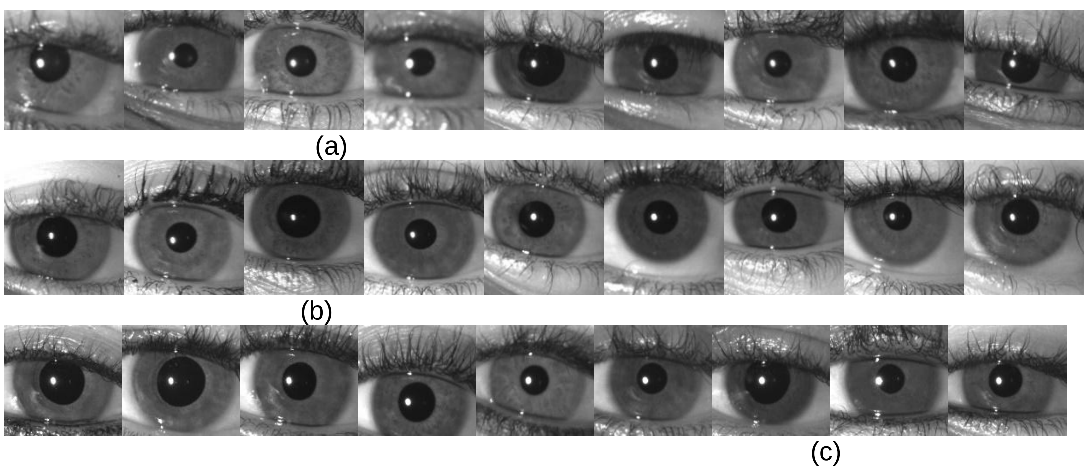


Figure 3.2: Selected cropped iris images (a) Yoruba (b) Ibo and (c) Hausa.

also displayed a good classification performance using the advantages of GSO to optimize SVM.

3.3.1 Algorithm of GAGAC (Genetic Algorithm based GAC)

This research came up with a new algorithm optimizing Geodesic Active Contour with Genetic Algorithm (GAGAC) at segmentation level which allowed GA to perform an automatic search for the optimal values of the regularization parameters (σ (expansion weight) for Gaussian, *k* (number of iterative time step) *and* α (contour weight)) which were normally provided by user for each image in the stopping function algorithm) that played an important role in GAC. Genetic Algorithm which is an adaptive strategy and global optimization algorithm was used to provide the best optimal value for the three parameters and these values were supplied as inputs in GAC algorithm for segmenti*ng all the acquired images. The formulated algorithm for GAGAC is as shown in Algorithm 3.1.

3.3.2 Flow diagram of GSOSVM

In this research, a new algorithm was developed by optimizing a well known classification algorithm which is Support Vector Machine that is able to handle high dimensional datasets.

Radial Basis Function (RBF) kernel was used as the appropriate kernel function because of the following reasons: It has fewer controllable parameters than the polynomial kernel, maps samples into a higher dimension and has less numerical difficulties. For the SVM that utilized RBF as the kernel function, there are two parameters, (C and γ) to be optimized. The goal is to identify the best (C, γ).

Algorithm 3.1:

Algorithm for GAGAC Segmentation:

Phase 1: Find the Stopping function: K

Inputs: Determine the optimal value of the three parameters of GAC (σ, *k* and α)

Step 1: Find the best σ for Gaussian, best *k and* best α using GA

t := 0;

Create initial population $B_0 = (b_{1,0}, \ldots, b_{m,0})$;

WHILE *stopping condition not fulfilled* DO

BEGIN

(∗ proportional selection ∗)

FOR i := 1 TO m DO

BEGIN

x := Random[0,1];

$k := 1$:

WHILE $k < m \,\&\, x < \sum_{j=1}^{k} f(b_{j,t}) / \sum_{j=1}^{m} f(b_{j,t})$ DO

k := k + 1;

$b_{i,t+1} := b_{k,t}$

END

(∗ one-point crossover ∗)

FOR i := 1 TO m − 1 STEP 2 DO

BEGIN

IF Random[0,1] ≤ P_C THEN

BEGIN

pos := Random{1,...,n − 1};

FOR k := pos + 1 TO n DO

BEGIN

aux := b i,t+1 [k];

$b_{i,t+1}[k] := b_{i+1,t+1}[k]$

$b_{i+1,t+1}[k]$:= aux

END

END

END

(∗ mutation ∗)

FOR i := 1 TO m DO

FOR k := 1 TO n DO

IF Random[0,1] < P_M THEN

invert $b_{i,t+1}[k]$;

t := t + 1

END

Step 2: Filter the image with Gaussian filter σ,(*G(x,y)*)

$$G(x,y)=\frac{1}{\sqrt{2\pi}x\sigma}\times e^{-\frac{x^2}{2\sigma^2}}$$

Step 4: Implement the equation for stopping function:

$$K(x,y)=\frac{1}{1+\left(\frac{\left\|\nabla(G(x,y)\times I(x,y))\right\|}{k}\right)^{\alpha}}$$

Phase 2: Generating ψ, zeroth level set:

Step 1: Input segmented pupil image.

Step 2: Create pupil mask having radius greater than pupil radius.

Step 3: Generate ψ according to

$$\psi(x,y)=\begin{cases}0,\ if\,(x,y)\quad is\quad on\quad the\,curve\\ <0, if\,(x,y)\quad is\quad inside\ the\,curve\\ >0. if\,(x,y)\quad is\quad outside\,the\,curve\end{cases}$$

Step 4: Display it on input eye image.

Phase 3: Perform segmentation:

Step 1: Maximum iterations = Input from user

Step 2: ε = Input from user

Step 3: Propagation =1(constant)

Step 4: Initialize ψ

Step 5: Evolve ψ according to discrete implementation equation,

$$\frac{\psi_{i,j}^{t+1}-\psi_{i,j}^{t}}{\Delta t}=-cK'_{i,j}\left\|\nabla\psi^{t}\right\|-K'_{i,j}(\varepsilon\kappa_{i,j}^{t}\left\|\nabla\psi^{t}\right\|)+\nabla\psi_{i,j}^{t}\cdot\nabla K_{i,j}^{t}$$

Step 6: Increment Δt according to Courant-Friedrichs-Lewy (CFL) condition.

Step 7: Check number of iterations and convergence.

Step 8: If number of iterations < maximum iterations or convergence is not reached Go back to step 6.

Step 9: Else Exit

Step10: Display final contour.

Phase 4: Estimation of radius

Step 1: Create mask by binarization of final contour.

Step 2: Calculate angle for all the values of final extracted contour.

Step 3: Check calculated angle if it is less than 182 and greater than 179.

Step 4: If yes then angle = 180°

Step 5: If no, Check calculated angle if it is less than or equal 212 and greater than 208.

Step 6: If yes then angle = 210°

Step 7: If no, Check calculated angle if it is less than or equal 152 and greater than 150.

Step 8: If yes then angle = 150°

Step 9: If no, Check calculated angle if it is less than or equal 32 and greater than30.

Step 10: If yes then angle = 30°

Step 11: If no, Check calculated angle if it is less than or equal 1 and greater than -1.

Step 12: If yes then angle = 0°

Step 13: If no, Check calculated angle is less than or equal -29 and greater than -31.

Step 14: If yes then angle = -30°

Step 15: Calculate Euclidean distance from pupil center using each angle.

Step 16: Take average of this 5 distances.
Step 17: Draw circle with this radius±20.

Where C is a regularization parameter that controls the "flexibility" of the hyperplane while **γ** is the kernel parameter that controls the correlation among support vectors. Selecting an improper value for **γ** may cause an overfitting.

The combined variable (C, γ) of the SVM penalty factor C and the RBF kernel parameter γ was used as the search target of the GSOSVM algorithm so as to find the combinatorial variable value which has the highest classification accuracy of SVM. It is the ultimate objective of the GSOSVM algorithm to optimize the SVM parameter model. In this research, accuracy rate of classification of SVM was taken as an evaluation criterion.

All the necessary steps employed to perform the operation of the GSOSVM algorithm were summarized in Figure 3.3, where the two phases of the GSO algorithm are presented using PSO flow diagram presented in Figure 2.27. The pseudocode of GSOSVM was presented in Algorithm 3.2 where w is the coefficient vector and b is the offset. ξ_i is a slack variable introduced when linear is not separable; C is the penalty factor used to represent the penalty index for misclassification. C determines the learning ability of SVM and the experience risk coordination degree. $\mathrm{sgn}(x)$ is a sign function, α_i^* is the Lagrange coefficient corresponding to the support vector, b^* is the classification threshold. γ is the RBF kernel function parameter which affects the distribution of complexity of sample data in the characteristic space.

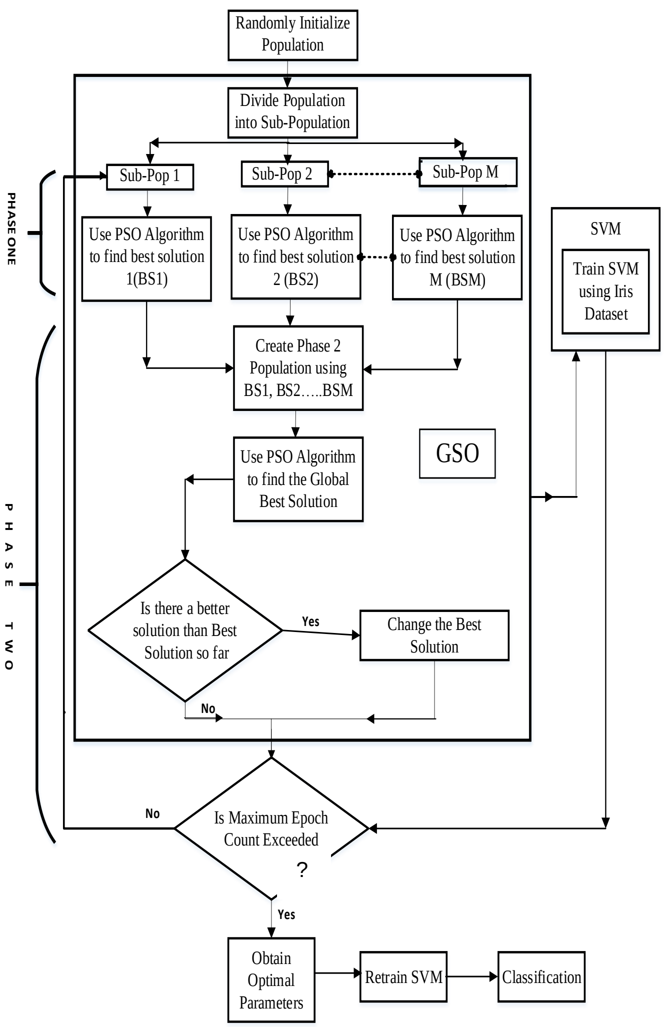


Figure 3.3: GSOSVM Flow Diagram

?

Algorithm 3.2. GSOSVM Algorithm

Inputs: Determine the vaious training and testing data

Output: Determine the calculated accuracy

Select the optimal values of Cost penalty C and gamma γ (Search target of GSO)

Apply GSO to find the optimal values for C and γ of SVM.

1. Level 1 Initialization: $x_j^{(i)}$, $v_j^{(i)}$, $P_j^{(i)}$, $g^{(i)}$ within [xmin, xmax]D randomly

2. Level 2 Initialization: $v^{(i)}$, $p^{(i)}$ within [xmin, xmax]D randomly
3. The population is divided into M subpopulations, $x_i \subset : i = 1, 2, ...M$

4. The population is initialized randomly, $x_j^{(i)} \in x_i : j = 1, 2, \ldots N$

5. Begin PSO: Level 1

For each of th e M subpopulations (subswarms), calculate the position ar

for k $\leftarrow$ 0 to L1 do

$$v_{i,j}^{t+1} = \omega v_{i,j}^{t} + c_1 r_1 (p_{i,j} - x_{i.j}^{t}) + c_2 r_2 (p_{g,j} - x_{i,j}^{t})$$

$x_{i.j}^{t+1} = x_{i.j}^{t} + v_{i,j}^{t+1}$

if $f(x_j^{(i)}) < f(p_j^{(i)})$

then $p_j^{(i)} \leftarrow x_j^{(i)}$

if $f(p_j^{(i)}) < f(g^{(i)})$

then $g^{(i)} \leftarrow p_j^{(i)}$

if $f(g^{(i)}) < f(g)$

then $g \leftarrow g^{(i)}$

End PSO

Begin PSO: Level 2

Initialise Swarm $y^{(i)} = g^{(i)} : i = 1, 2, \ldots, M;$

for k $\leftarrow$ 0 to L2 do

$v^{(i)} \leftarrow \omega_2 v^{(i)} + c_3 r_3 (p^{(i)} - y^{(i)}) + c_4 r_4 (p_{g,j} - y^{(i)})$

$y^{(i)} \leftarrow y^{(i)} + v^{(i)}$

if $f(y^{(i)}) < f(p^{(i)})$ then $p^{(i)} \leftarrow y^{(i)}$

if $f(p^{(i)}) < f(g)$ then $g \leftarrow p^{(i)}$

End PSO

Return g, f(g) for C and γ .

While (Stopping condition is not met) do else

Implement SVM model applying the optimal value of C and γ from GSO

Given a set of training data $T = \{x_i, y_i\}_{i=1}^{n}$. Let $\{x_i, y_i\}$, $1 \leq i \leq n$ where each datum must conform to the criteria $x_i \in R_d$, $y_i \in \{+1, -1\}$ where d denotes the number of dimensions of the input data and n represent the number of training data.

$$Min \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \xi_i$$

$$y_i (w^T \times x_i + b) + \xi_i - 1 \geq 0$$

where $\xi_i \geq 0$ i=1,2……..n

$$f(x)=\text{sgn}\{(w^T\times x_i)+b\}=\text{sgn}\{\sum_{i=1}^{n}\alpha_i^* y_i(x_i^T\times x)+b^*=\text{sgn}\{\sum_{i=1}^{n}\alpha_i^* y_i K(x_i,x)+b^* \quad K(x_i,x)=f(x)^T\times f(x)$$

and $K(x_i,x)=e^{-\|x_i-x\|^2/2\gamma^2}$ (RBF kernel)

End while

Return accuracy, where f(x)=accuracy, x=(C, γ)

3.4 Implementation Procedure for the Iris-based Ethnicity Prediction System using the Two Formulated Algorithms

The process flow diagram shown in Figure 3.4 was the representation of the basic stages involved in the development of the iris-based ethnicity system. The training and testing data were acquired locally considering three major tribes (Yoruba, Ibo and Hausa) in Nigeria and normalized into a uniform pixel. Image enhancement of the image using refinement technique was done to intensify and improve the images to make their features visible. This was followed by segmentation using Genetic Algorithm based Geodesic Active Contour (GAGAC)) process which isolated the iris from the eye region and located the inner and outer boundaries of the iris, upper and lower eyelid detection revealing the details of the iris features.

The normalization process was done using Daugman's Rubber Sheet Model to transform the segmented iris region so that it has a fixed dimension for the comparison template. The feature extraction/encoding was performed using Log Gabor Filters (since field suggests that natural images are better coded by filters that have Gaussian transfer function when viewed on the logarithmic frequency scale) for feature reduction, removal of noise/unwanted elements and generation of unique feature codes before classification. Classification between the training and testing data was done using GSO-SVM algorithm. The iris-based ethnicity prediction system was implemented using Matrix laboratory R2020a. The computational environment for this research is CPU (Intel i5-4210u 1.70 GHz) RAM (8 GB RAM) OS (not lower in version than Windows 7 Pro 64-bit). The performance of the iris-based ethnicity prediction system was evaluated using some standard biometric prediction metrics. The graphical user interface (GUI) was as shown in figure 3.4:

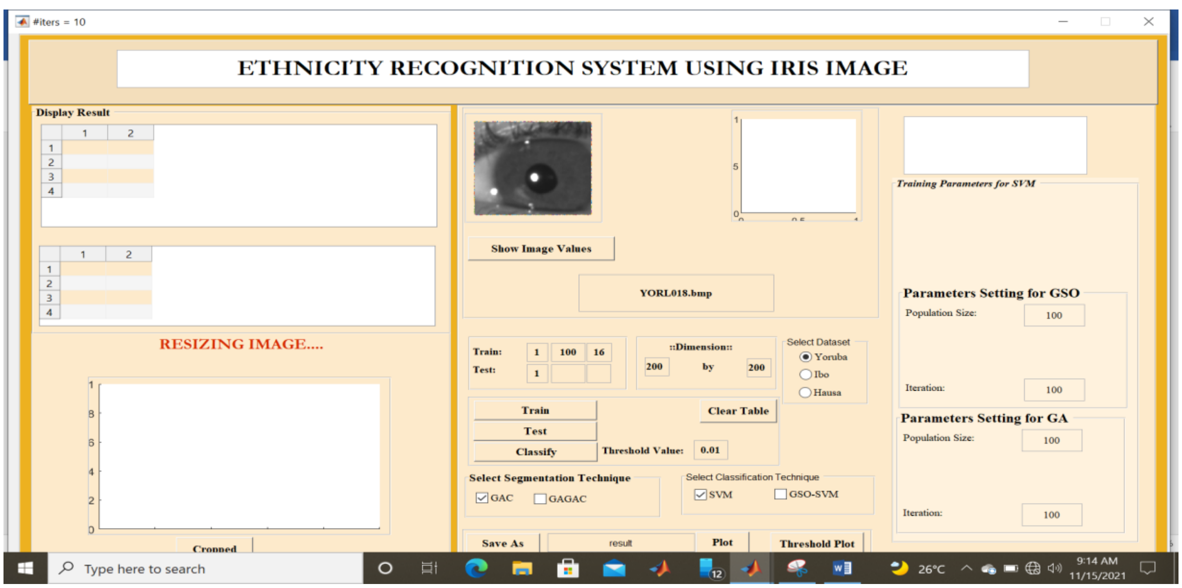


Figure 3.4: Graphical user interface of the design system

3.4.1 Image acquisition

The locally acquired iris database in section 3.2 was used to constitute the training and testing dataset. K-fold cross validation data splitting strategy was used in order to build a more generalized system which enable all images to be used as both trained and test images. In k-fold cross-validation, the original

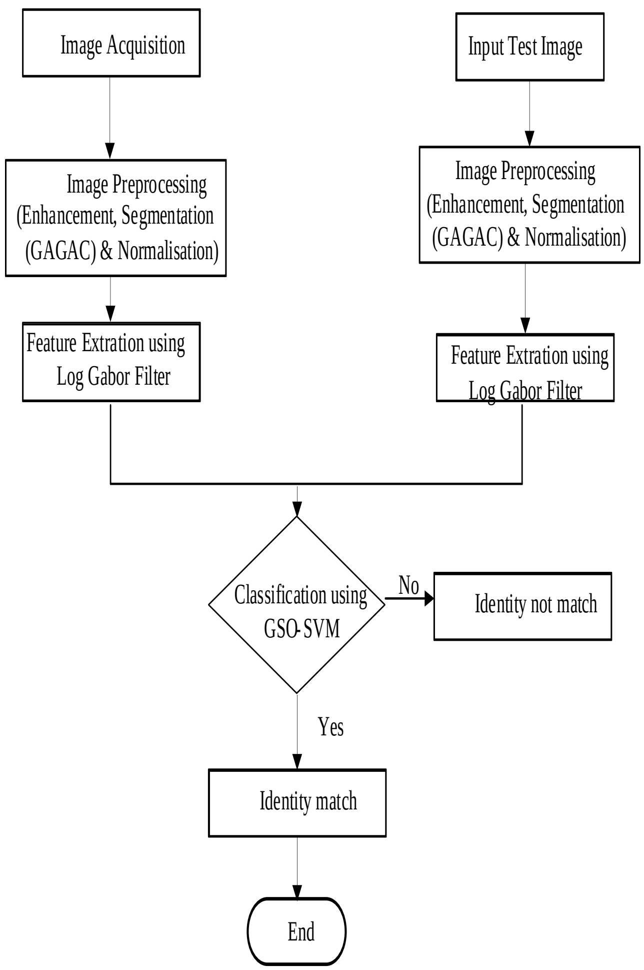


Figure 3.5: Flow Diagram for the Ethnicity Prediction System

samples of the dataset were randomly partitioned into k subsets of (approximately) equal size, and the experiment is run k times. For each time, one subset was used as the testing set, and the other k - 1 subsets were used as the training set. The average of the k results from the folds can then be calculated to produce a single estimation. In this research, the value of k was set to 10.

3.4.2 Image preprocessing

The Image preprocessing considered are: iris image enhancement, segmentation and normalization.

i. **Image enhancement:** In order to take advantage of the characteristics of the pupil described above, the image was enhanced using histogram equilization in order to obtain a reliable binary image of the pupil.

ii. Segmentation: This stage was taken care of using GAC and the formulated GAGAC as described in section 3.3.1. The segmentation methods in this research involves three major steps. First, the approximate location of the iris center is detected. Second, the iris region is extracted. Finally, the reflections (noises) are removed from the iris region. The significance of this steps is its robustness to realistic noises caused by non-ideal imaging settings such as reflections, blurred boundaries, gaze-deviation, and eyelids occlusion. Sample of segmented iris images was presented in Figure 3.6(a - c), sample of iris centre localization was presented in Figure 3.7(a-c) while sample of noise removal from images were presented in Figure 3.8(a – c).

iii. **Normalization:** Once the iris region is successfully segmented from an eye image, the next step is to transform the iris region so that it has fixed dimensions for the comparisons of templates. The normalization process produces iris regions, which have the same constant dimensions, so that two photographs of the same iris under different conditions will have characteristic features at the same spatial location. Daugman's rubber sheet model was employed here. Algorithm for daugman's rubber sheet model is as shown in Algorithm 2.2 while sample of normalized iris images were presented in Figure 3.9.

3.4.3 Feature encoding

For accurate recognition result, the most discriminating information present in an iris pattern must be extracted. Only the significant features of the iris will be encoded so that comparisons between irises can be made. The product of the normalization stage was encoded using Log Gabor Filters considering its maximum suitability for bandwidth. the feature code ranges for each ethnic group was presented in Table 3.1.

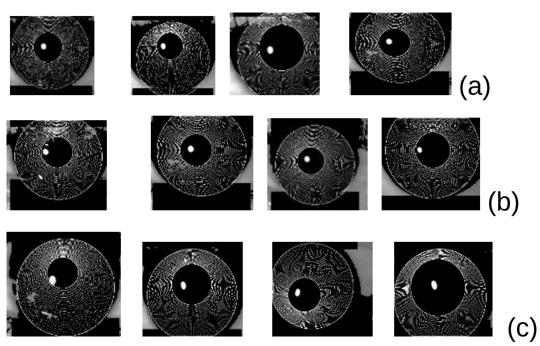


Figure 3.6: Segmented iris images (a) Yoruba (b) Ibo and (c) Hausa

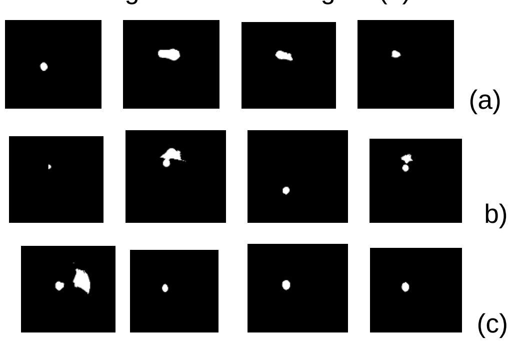
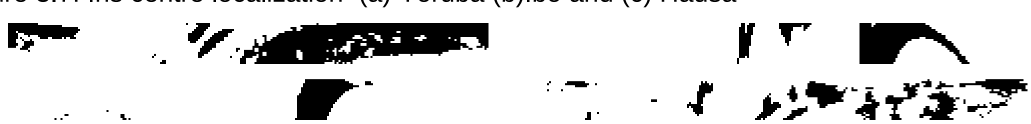
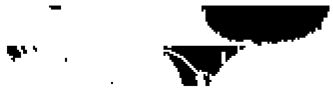


Figure 3.7: Iris centre localization (a) Yoruba (b)Ibo and (c) Hausa





(a)

(b)

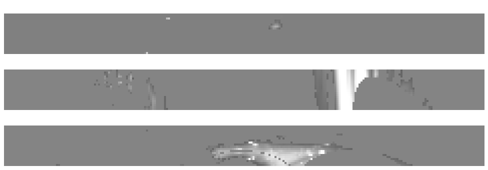
(a)

(b)

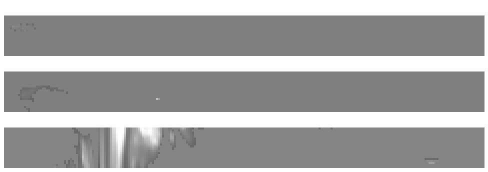
(c)

Figure 3.8: Noise removal of iris images (a) Yoruba (b) Ibo and (c) Hausa

(a)



(b)



(c)

Figure 3.9 : Normalized Iris Images (a) Yoruba (b)Ibo and (c) Hausa

Table 3.1: Feature Codes Range for each Ethnic Group

Feature code (F)	Hausa	Ibo	Yoruba
0.100000000<F<0.200000000	200	0	0
0.700000000<F<0.800000000	0	200	0
2.00000000<F<3.000000000	0	0	200

3.4.4 Classification algorithm

The feature vectors obtained using Log Gabor Filters was subjected to classification using SVM and the formulated algorithm (GSOSVM).

3.5 Performance Evaluation Metrics for iris-based Ethnicity Prediction System

The performance of the developed iris based ethnicity prediction system was measured using these metrics:

i. False Positive Rate: $$\frac{FP}{(FP+TN)}\times 100 \qquad 3.1$$

ii. Sensitivity: $$\frac{TP}{(TP+FN)}\times 100 \qquad 3.2$$

iii. Specificity: $$\frac{TN}{(TN+FP)}\times 100 \qquad 3.3$$

iv. Precision: $$\frac{TP}{TP+FP}\times 100 \qquad 3.4$$

v. Accuracy: $$\frac{TP+TN}{TP+FN+TN+FP}\times 100 \qquad 3.5$$

FP indicates the number of images that are impostors but incorrectly accepted by the system. TN indicates the number of images that are impostors and are correctly rejected by the system. FN indicates the number images that are valid but incorrectly rejected by the system. TP indicates the number of images that are valid and are accepted by the system

4. RESULTS AND DISCUSSION

4.1 Presentation of Results

This research developed iris-based ethnicity prediction system using the Genetic Algorithm Geodesic Active Contour (GAGAC) segmentation algorithm and Galactic Swarm Optimization Support Vector Machine (GSOSVM) as classifier. System performance was verified using sensitivity, specificity, precision, accuracy and classification time as metrics to measures the predictive capabilities of the designed prediction system. The results of each metrics were based on the concepts of the confusion matrixes (true and false positive values (TP and FP) and true and negative values (TN and FN) achieved by the system against the actual outcomes

The evaluation results of the techniques (GAGAC and GSOSVM) were based on the selected three major tribes in Nigeria (Yoruba, Hausa and Ibo). Statistical analysis was also carried out using t-test to analyze the result obtained for segmentation time and accuracy along with performance of FPR, SEN, SPEC, PRE and ACC for validation purposes.

4.2 Results of Iris-based Ethnicity Prediction System.

Parameter regularization which is a technique used to reduce errors by fitting the functions appropriately on the given algorithm to avoid overfitting were carried on the standard GAC and SVM using GA and GSO respectively. The results presented in Table 4.1 depicts the performance of GAGAC segmentation technique and GSOSVM as classifier based on the three selected tribes. The result from Table 4.1(a-c) showed that increase in threshold value resulted to increase in classification performance. The system achieved a better performance at threshold 0.75 with 188, 192 and 190 correctly classified (TP) and 12, 8 and 10 irises misclassified (FN) to other ethnic groups while 6, 2 and 4 irises were wrongly classified (FP) as Yoruba, Ibo and Hausa respectively.

Furthermore, it was discovered from Table 4.1a-c that GAGAC/GSOSVM at threshold value of 0.75 attained a better classification performance for Yoruba, Ibo and Hausa respectively. This result illustrated that GAGAC/GSOSVM outperformed GAC/SVM in terms of better false positive rate, sensitivity, specificity, precision and accuracy.

4.3 Discussion of results

This research focused on segmentation and classification problems in processing non-ideal iris images considering the non circular and non elliptical nature of the iris which are challenging tasks resulting in inaccurate segmentation and poor classification. The research was limited to solving the problem of high segmentation time and low accuracy in GAC using Geodesic Active Contour optimized with Genetic Algorithm (GAGAC) in determining the parameters automatically rather than the conventional method. Also, the major problem of Support Vector Machine which is parameter setting was addressed by employing the good optimization performance of Galactic Swarm Optimization to boost the classification efficiency of Support Vector Machine.

This research focused on segmentation and classification problems in processing non-ideal iris images considering the non circular and non elliptical nature of the iris which are challenging tasks resulting in inaccurate segmentation and poor classification. The research was limited to solving the problem of high segmentation time and low accuracy in GAC using Geodesic Active Contour optimized with Genetic Algorithm (GAGAC) in determining the parameters automatically rather than the conventional method. Also, the major problem of Support Vector Machine which is parameter setting was addressed by employing the good optimization performance of Galactic Swarm Optimization to boost the classification efficiency of Support Vector Machine.

Table4.1:Performance based on GAGAC/GSOSVM

Ethnic Group	**TV**	**TP**	**FN**	**FP (o)**	**TN (o)**	**FPR (%)**	**SEN (%)**	**SPEC (%)**	**PREC (%)**	**ACC (%)**	**TIME (Sec)**
Y	0.1	194	06	22	378	5.50	97.00	94.50	89.82	95.33	63.19
O	0.4	192	08	16	384	4.00	96.00	96.00	92.31	96.00	59.81
R	0.6	190	10	10	390	2.50	95.00	97.50	95.00	96.67	64.77
	0.75	188	12	06	390	1.50	94.00	98.50	96.91	97.00	60.56

(a)

Ethnic Group	**TV**	**TP**	**FN**	**FP (o)**	**TN (o)**	**FPR (%)**	**SEN (%)**	**SPEC (%)**	**PREC (%)**	**ACC (%)**	**TIME (Sec)**
I	0.1	198	02	18	382	4.50	99.00	95.50	91.67	96.67	63.19
B	0.4	196	04	12	388	3.00	98.00	97.00	94.23	97.33	59.81
O	0.6	194	06	06	394	1.50	97.00	98.50	97.00	98.00	64.77
	0.75	192	08	02	398	0.50	96.00	99.50	98.97	98.33	60.56

(b)

Ethnic Group	TV	TP	FN	FP (o)	TN (o)	FPR (%)	SEN (%)	SPEC (%)	PREC (%)	ACC (%)	TIME (Sec)
H	0.1	196	04	20	380	5.00	98.00	95.00	90.74	96.00	63.19
A	0.4	194	06	14	386	3.50	97.00	96.50	93.27	96.67	59.81
U	0.6	192	08	08	392	2.00	96.00	98.00	96.00	97.33	64.77
	0.75	190	10	04	396	1.00	95.00	99.00	97.94	97.67	60.56

(c)

Table 4.2: Performance based on GAC/SVM

Ethnic Group	TV	TP	FN	FP (o)	TN (o)	FPR (%)	SEN (%)	SPEC (%)	PREC (%)	ACC (%)	TIME (Sec)
Y	0.1	178	22	30	370	7.50	89.00	92.50	85.58	91.33	116.01
O	0.4	176	24	26	374	6.50	88.00	93.50	87.13	91.67	121.77
R	0.6	174	26	22	378	5.50	87.00	94.50	88.78	92.00	117.26
	0.75	172	28	16	384	4.00	86.00	96.00	91.49	92.67	118.54

(a)

Ethnic Group	TV	TP	FN	FP (o)	TN (o)	FPR (%)	SEN (%)	SPEC (%)	PREC (%)	ACC (%)	TIME (Sec)
I	0.1	182	18	26	374	6.50	91.00	93.50	87.50	92.67	116.01
B	0.4	180	20	22	378	5.50	90.00	94.50	89.11	93.00	121.77
O	0.6	178	22	18	382	4.50	89.00	95.50	90.82	93.33	117.26
	0.75	176	24	12	388	3.00	88.00	97.00	93.62	94.00	118.54

(b)

Ethnic Group	TV	TP	FN	FP (o)	TN (o)	FPR (%)	SEN (%)	SPEC (%)	PREC (%)	ACC (%)	TIME (Sec)
H	0.1	180	20	28	372	7.00	90.00	93.00	86.54	92.00	116.01
A	0.4	178	22	24	376	6.00	89.00	94.00	88.12	92.33	121.77
U	0.6	176	24	20	380	5.00	88.00	95.00	89.80	92.67	117.26
	0.75	174	26	14	386	3.50	87.00	96.50	92.55	93.33	118.54

(c)

Where Y= Yoruba, I =Ibo, H= Hausa,

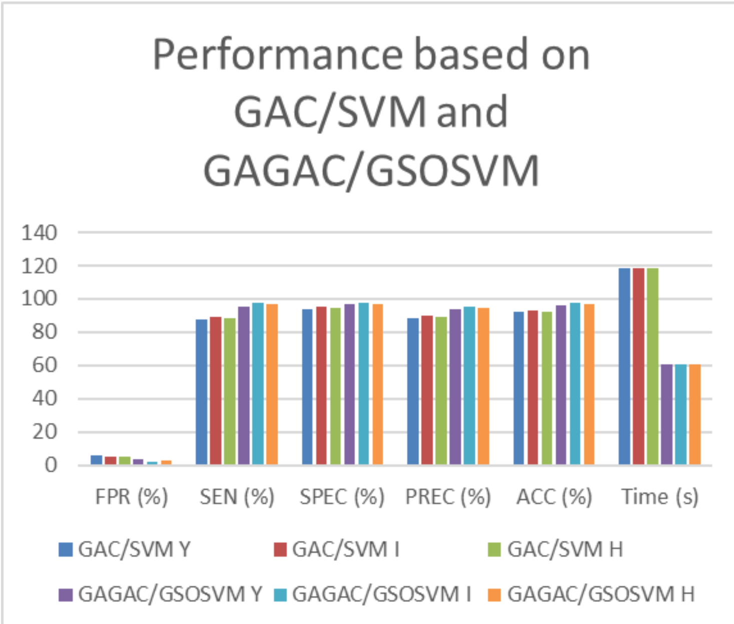


Figure 4.1: Graph showing the performance based on GAC/SVM and GAGAC/GSOSVM

The system performance was measured using the predictive capabilities of the developed iris-based ethnicity prediction system employing the standard GAC and the optimized GAC (GAGAC) segmentation algorithm in combination with the standard SVM and the optimized SVM (GSOSVM) as classifier. This was assessed through the elements of confusion matrix which are utilized to determine the metrics like sensitivity (SEN), specificity (SPEC), precision (PRE), accuracy (ACC) and classification time.

It can be inferred from the results achieved from Table 4.1 and 4.2 that GAGAC/GSOSVM outperformed GAC/SVM in terms of false positive rate, sensitivity, specificity, precision, accuracy and classification time. The improved performance displayed by GAGAC over conventional GAC during the segmentation stage can be traced to proper tuning of the parameters of GAC done by GA through optimization thereby improving the segmentation accuracy and time. Good segmentation performance displayed by GAGAC corroborated partly to the better achievement recorded with GAGAC/GSOSVM. It was reported in literatures that accurate iris segmentation always leads to correct feature extraction thereby resulting into most match success and better performance in iris-based prediction system and vice-versa (Masek and Kovesi, 2003 and Samir and Arun, 2009).

Also an improvement recorded in employing GSOSVM over standard SVM can be linked to the ultimate objective of GSO to maximize SVM parameters (penalty factor C and the RBF kernel parameter γ) accurately with the aim of overcoming the challenges and improving the accuracy rate of SVM which was greatly achieved as deduced in the results attained in GAGAC/GSOSVM combination. C parameter tells the SVM how much misclassifying should be avoided and a good value achieved through proper tuning of parameter C by GSO help the performance of the system.

It was discovered that accurate segmentation in GAGAC showed the reason for its better accuracy against GAC while the boost in classification experienced by SVM in GSOSVM reflects in the accuracy improvement over SVM.

This is evident in the performance displayed by GAGAC/GSOSVM over GAC/SVM. It also implies that the developed system utilized lesser time in segmenting the iris images and in classifying the images into Yoruba, Ibo or Hausa when GAGAC/GSOSVM was employed for segmentation and classification. This confirms the general statements that for better performance of an iris-based prediction system with good prediction values, correct segmentation is paramount along with good classifier (Aydin *et al.* 2011). Also, the use of evolutionary algorithm for parameter optimization is very much faster and often gives better results (Friedrichs and Igel, 2005 and Rossi and de Carvalho, 2008). Hence. Optimizing the parameters of GAC using GA and optimizing the parameters SVM with GSO reduces the overall classification time and increases the overall accuracy of the system.

5.0 CONCLUSIONS

This research developed an iris-based ethnicity prediction system using two formulated algorithms. The segmentation process was carried out using a geodesic active contour segmentation algorithm optimized with genetic algorithm (GAGAC) while classification of iris into its various ethnicity group was done using SVM optimized with galactic swarm optimization algorithm (GSOSVM). An extensive research has been carried out considering the problem of predicting ethnicity from non-ideal iris features. Most early research in iris segmentation assumed that the iris had a circular boundary. However, often the pupillary and limbic boundaries are not perfectly circular. Most datasets used by researchers in literature are preprocessed iris images capture under controlled environment and most of the algorithms may not perform well when they are subjected to real world system.

The techniques were developed to address the issue of segmenting accurately non ideal iris images, high processing time, inappropriate parameter settings which results in inaccurate segmentation and poor classification results. Optimization was carried out on the parameters of GAC using GA for segmentation and SVM using GSO to tuned the parameters to make the optimal separating hyperplane obtainable for classification to provide the best solution to the problem within the framework of available resources. To achieve the set objectives, iris images of the three major tribes in Nigeria were locally captured, preprocessed, segmented by GAGAC, normalized using Dagupan’s Rubber Sheet Model, code features generated by Log Gabor filter, and classified by GSOSVM.

In all the evaluations conducted, the formulated segmentation algorithm (GAGAC) achieved improved segmentation in term of accuracy and time compares to convention GAC, while the optimized SVM (GSOSVM) achieved a better classification results in terms of false positive rate, sensitivity, specificity, precision, accuracy and classification time. The experimental results showed by GAGAC/GSOSVM were examined and compared with standard GAC/SVM, using the performance metrics.

Conclusively, the statistical analysis results carried out validated that GAGAC significantly increased the accuracy and reduced the segmentation time while the GSOSVM classification technique also significantly illustrated that optimization of the parameters of SVM gave a significant improvement which commensurate to the gain of accuracy, precision and classification time achieved. This consequently implies that the optimization performed on GAC and SVM has positive impact on the efficiency of the developed iris-based ethnicity prediction system. It also indicated that the recorded higher accuracy for GAGAC/GSOSVM was not due to sampling error but that there is statistically significant evidence that the optimized algorithms (GAGAC and GSOSVM) performed better than the standard one (GAC and SVM).

6.0 **RECOMMENDATION**

This research has been able to developed an improved GAGAC segmentation and GSOSVM classification technique for an iris-based ethnicity prediction system. It is therefore recommended that the developed iris-based ethnicity prediction system can be employed to access government benefits and enhance airline security by verifying traveler identity, Researchers in the line of research can further look into the possibility of properly classifying offspring of mixed marriage, extending the prediction to other ethnic groups in Nigeria and people of different dialects within an ethnic group. Combining of multiple images or multiple biometrics (such as face and iris, iris and fingerprint, iris and voice) can be employed to improve performance of the ethnicity prediction system for better classification results. Future research in the same lines of thought of GSO optimizing SVM could be investigated using other global optimization algorithms such as GA, ACO, ABC, BA, or other heuristics algorithm in place of PSO variants for both exploration or exploitation stage.

REFERENCES

[1] **Abbasi A. A., Khan M. N. A., and Khan S. A.,** (2013): "A Critical Survey of Iris Based Recognition Systems", Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad, Pakistan. Middle-East Journal of Scientific Research, 15 (5) pp 663-668

[2] **Aworinde, H., and Onifade, O. F. W**. (2019). A Soft Computing Model of Soft Biometric Traits for Gender and Ethnicity Classification. *International Journal of Engineering and Manufacturing*, *9*(2), pp 54–63. https://doi.org/10.5815/ijem.2019.02.05

[3] Arun Ross and Samir. Shah (2006), "Segmenting Non-Ideal Irises Using Geodesic Active Contours," Proc. IEEE 2006 Biometric Symp., pp. 1-6, 2006.

[4] Aydin, I., Karakose, M. and Akin, E., 2011. A multi-objective artificial immune algorithm for parameter optimization in support vector machine. Applied Soft Computing 11, 120–129.

[5] **Bessudnov Alexey, Denis Tarasov, Viacheslav Panasovets, Veronica Kostenko, Ivan Smirnov, Vladimir Uspenskiy** (2021) "Predicting ethnicity with data on personal names in Russia" 1VK press release. 4 March 2021.https://vk.com/main.php?subdir=press&subsubdir= q4-2020-results

[6] **Caselles V., F. Catte, T. Coll, and F. Dibos**, (1993)"A geometric model for active contours," *Numer. Math.*, vol. 66, pp. 1–31, 1993.

[7] Chandra, K. (2006). What is Ethnic Identity and Does it Matter? *Annual Review of Political Science*, *9*, 397–424. **https://doi.org/** **https://doi.org/10.1146** /annurev.polisci.9.062404.170715

[8] **De-Miranda, P.B.C., Prudencio, R.B.C., Carvalho, A.C.P.L.F. and Soares, C.,** 2012. Combining a multi-objective optimization approach with meta-learning for SVM parameter selection, in: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Presented at the 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2909–2914.

[9] **Demirkus, M., Garg, K., and Guler, S.** (2010). Automated person categorization for video surveillance using soft biometrics. *Proceedings of the SPIE*, *7667*. https://doi.org/10.1117/12.851424

[10] **Falohun A.S.** (2012): "Development of a feature extraction method for Iris Recognition using enhanced inverse analytical fourier –mellin transform". PhD Thesis, Dept of Computer Science, LAUTECH, Ogbomoso.

[11] **Falohun A.S., Omidiora E.O, Olabiyisi O.S, Fenwa O.D and Adedeji O.T** (2010): "Web-Based Iris Recognition System", Proceedings of the First International Conference on Engineering and Technology, ISBN 978-2902-58-6

[12] **Fangzhou Xie** (2022) "rethnicity: An R package for predicting ethnicity from names SoftwareX, Volume 17, 2022,100965, ISSN 2352-7110,

[13] **Friedrichs, F. and Igel, C.,** (2005), "Evolutionary tuning of multiple svm parameters", Neurocomputing 64, pp 107–117.

[14] **Gaspar, P., Carbonell, J., Oliveira, J.L.**, 2012. On the parameter optimization of Support Vector Machines for binary classification. J Integr Bioinform 9, 201.

[15] **Gugulethu Mabuza-Hocquet, Fulufhelo Nelwamondo and Tshilidzi Marwala** (2016) “Ethnicity Prediction and Classification from Iris Texture Patterns: A Survey on Recent Advances”, 2016 International Conference on Computational Science and Computational Intelligence, Conference Paper · December 2016 DOI: 10.1109/CSCI.2016.0159

[16] **Juan Tapia Farias, Claudio Perez and Kevin W. Bowyer**, (2016), “Gender Classification From the Same Iris Code Used for Recognition”, IEEE Transactions on Information Forensics and Security 11(8):1-1 · August2016DOI: 10.1109/TIFS.2016.2550418

[17] **Kaneshiro, B., Geling, O., Gellert, K., and Millar, L.** (2011). The challenges of collecting data on race and ethnicity in a diverse, multiethnic state. *Hawaii Medical Journal*, *70*(8), pp168–171.

[18] **Keerthi, S. S., & Lin, C.-J.** (2003). Asymptotic behaviours of support vector machines with gaussian kernel. Neural Computation, 15, pp 1667–1689.

[19] **Kim, M., Moon, H., and Pan, S. B.** (2012). Framework of Human Identification using MultiModal Biometrics. *International Journal of Multimedia and Ubiquitous Engineering*, *7*(2), pp 509–514.

[20] **Lagree Stephen and Kevin W. Bowyer,** (2007), “Ethnicity Prediction Based on Iris Texture Features”

[21] **Lagree S. and Bowyer K.W.** (2011): “Ethnicity Prediction Based on Iris Texture feature”, Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556 USA.

[22] **Latinwo, B. M., Falohun, A. S., and Omidiora, E. O.** (2016). Principal Component Analysis - Based Ethnicity Prediction Using Iris Feature, *16*(4), pp 1–5. https://doi.org/10.9734/BJAST/2016/26131

[23] **Latinwo, B., Falohun, A., Omidiora, E., and Makinde, B.** (2018). Iris Texture Analysis for Ethnicity Classification Using Self-Organizing Feature Maps. *Journal of Advances in Mathematics and Computer Science*, *25*(6), pp 1–10

[24] **Luo, J., Ma, Y., Takikawa, E., Lao, S.H., Kawade, M., Lu, B.L.** (2007). *Person-specific SIFT features for face recognition, ICASSP, pp.563-566.*

[25] **Lyle, J. R., Miller, P. E., and Woodard, D. L**. (2010). Soft Biometric Classification Using Periocular Region Features. *IEEE*.

[26] **Makinde Bukola O., Olusayo D. Fenwa, Adeleye S. Falohun and Olufemi A. Odeniyi** (2019) “Performance Evaluation of Kernel-Based Feature Extraction Techniques for Face Recognition Systems”, FUOYE Journal of Engineering and Technology, Volume 4, Issue 1, March 2019 ISSN: 2579-0625 (Online), 2579-061

[27] **Maneet Singh, Shruti Nagpal, Mayank Vatsa, Richa Singh, Afzel Noore and Angshul Majumdar,** (2017), “Gender and Ethnicity Classification of Iris Images using Deep Class-Encoder”, 2017 IEEE International Joint Conference on Biometrics (IJCB), 978-1-5386-1124-1, pp 666-673

[28] **Michael Fairhurst, Meryem Erbilek and Marjory Da Costa-Abreu,** (2015), “Exploring Gender Prediction From Iris Biometrics”; 2015 International Conference of the Biometrics Special Interest Group (BIOSIG), DOI: 10.1109/BIOSIG.2015.7314602 Gesellschaft für Informatik e.V., Bonn, Germany

[29] **Minal K. Pawa, Sunita S. Lokhande and V. N. Bapat** (2012) “Iris Segmentation using Geodesic Active Contour for Improved Texture Extraction in Recognition” *International Journal of Computer Applications (0975 – 888) Volume 47– No.16, June 2012* 40

[30] **Orike, S., Anireh, V. I. E., and Ibironke, A. S.** (2016). A Gender and Ethnicity Identification System in Nigeria using the Fingerprint Technology. Proceedings of the World Congress of Engineering, I.

[31] **Qiu X. C., Sun Z. A., and Tan T. N.** (2006): "Global texture analysis of iris images for ethnic classification", Springer LNCS 3832: International Conference on Biometrics, pp 411-418, June 2006.

[32] **Qiu X. C., Sun Z. A., and Tan T. N.** (2006): "Global texture analysis of iris images for ethnic classification", Springer LNCS 3832: International Conference on Biometrics, pp 411-418, June 2006.

[33] **Romen V. Yampoolskiy and Venu Govindaraju** (2008), "Behavioural biometrics , a survey and classification", International Journal of Biometrics, 2008, Vol.1, no.1. DOI: 10.1504/IJBM 2008.018665 PG 81-113

[34] **Ratcliffe, P.** (2010): "Ethnic group": ISA Editorial Arrangement of sociopedia.isa

[35] **Rossi, A.L.D. and de-Carvalho, A.C.P**., (2008). "Bio-inspired Optimization Techniques for SVM Parameter Tuning". 10th Brazilian Symposium on Neural Netresearchs, 2008. SBRN '08. Presented at the 10th Brazilian Symposium on Neural Netresearchs, 2008. SBRN '08, pp. 57–62.

[36] **Sabhanayagam T., V. Prasanna Venkatesan and K. Senthamaraikannan,** (2018): "A Comprehensive Survey on Various Biometric Systems", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 5 (2018) pp. 2276-2297

[37] **Samir Shah and Arun Ross** (2009), "Iris Segmentation Using Geodesic Active Contours", IEEE Transactions on Information Forensics And Security,Vol.4, No.4, pp.824-836, Dec2009.

[38] **Sarfaraz Masood, Shubham Gupta, Abdul Wajid and Suhani Gupta,** (2018), "Prediction of human ethnicity from facial images using neural netresearchs", Advances in Intelligent Systems and Computing · June 2018 DOI: 10.1007/978-981-10-3223-3_20

[39] **Singh Maneet, Nagpal Shruti, Vatsa Mayank, Singh Richa, Noore, Afzel and Majumdar, Angshul.** (2017). "Gender and ethnicity classification of Iris images using deep class-encoder". 2017 IEEE International Joint Conference on Biometrics (IJCB), pp 666-673. 10.1109/BTAS.2017.8272755.

[40] **Tariq, U., Hu, Y., and Huang, T. S.** (2009). Gender and ethnicity identification from silhouetted face profiles. In *Proceedings - International Conference on Image Processing, ICIP*. https://doi.org/10.1109/ICIP.2009.5414117

[41] **Wong KO, Zaïane OR, Davis FG, Yasui Y.** A machine learning approach to predict ethnicity using personal name and census location in Canada. PLoS One. 2020 Nov 18;15(11):e0241239. doi: 10.1371/journal.pone.0241239. PMID: 33206667; PMCID: PMC7673495.

[42] **Yamakawa T.; Taniguchi, K.; Momen T.; Kobashi, S.; Kondo and K; Hata, Y**. (2007). *Biometrics Personal Identification Using Sole Information*, IEEE International Conference on Systems, Man and Cybernetics, pp. 3438-3442.

[43] **Zarei A. and Duxing M.** (2012): "ANN for prediction of ethnicity based on Iris Texture", Department of Electrical Engineering, University of the Pacific Stockton, CA, USA.

[44] **Zhang H., Z. Sun, T. Tan, and J. Wang** (2011). "Ethnic classification based on iris images". Chinese Conference on Biometric Recognition, pp. 82-90, Springer Berlin Heidelberg, 2011.