

1
2 **DEVELOPMENT OF A GENETIC ALGORITHM BASED**
3 **GEODESIC ACTIVE CONTOUR FOR IRIS BASED ETHNICITY**
4 **PREDICTION SYSTEM**
5
6
7

8 **ABSTRACT**
9

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.

10
11 *Keywords:* Segmentation, classification, Support Vector Machine, ethnicity, security
12
13

14 **1. INTRODUCTION**
15

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

22 | “It has been observed that in this part of the world, identification of individual’s ethnicity
23 | does not go beyond physical factors which in most cases can be manipulated or spoofed to
24 | carry out nefarious activities.” (Falohun *et al.* 2012) When identification does not go beyond

25 what is seen and touched, then such means of identification can be faulted and the integrity
26 of such is questionable; as a result, the interest in biometrics for identification has increased
27 to enhance security and the process of verifying individual's identity (Kim *et al.* 2012).

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

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

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

61 | Most of the existing segmentation algorithms (such as Integro-differential operator,
62 | Hough transform and Active Contour) assumed that the iris is circular and elliptical in shape
63 | resulting in under-segmentation and over-segmentation (Lagree and Bowyer 2011; Zhang *et al.*
64 | 2011; Gugulethu *et al.* 2016; Latinwo *et al.* 2016; Singh *et al.* 2017; Latinwo *et al.* 2018).
65 | Recent segmentation algorithm like Active Shaped Model, Randomized elliptical Hough
66 | Transform and Active Contours and Geodesic Active Contour (GAC) assumed non circular
67 | and non elliptical shape of iris. However, GAC supports accurately estimating the radius of
68 | the iris and its centre thereby lessens the concerns related with the traditional models but
69 | the time required to segment the iris is high though it gives better accuracy (Minal *et al.*
70 | 2012).

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

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

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

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

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

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

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

121 2. LITERATURE REVIEW

122 2.1 Ethnicity

123 | "Ethnicity can be defined as the fact or state of belonging to a social group that has a
124 common national or cultural tradition. The definition for race is sometimes equated with

125 ethnicity; as human race can be defined as a group of people with certain inherited features
126 that distinguish them from other groups of people. All men of whatever race are classified by
127 the anthropologist or biologist as belonging to one specie, homo-sapiens. Ethnicity can also
128 be defined as a vast group of people loosely bounded together by historical, socially
129 significant elements of their morphology and/or ancestry. It can serve as the connections
130 between physical features, races and personal characteristics” (Ratcliffe 2010).

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

134 **2.2 Related works on analyzing hard biometric to predict soft biometric**

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

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

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

147 Juan *et al.* (2016) predicted gender directly from the same binary iris code that could be
148 used for recognition and found that the information for gender prediction is distributed
149 across the iris, rather than localized in particular concentric bands and was able to achieve
150 89% correct gender prediction using the fusion of the best features of iris code from the left
151 and right eyes.

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

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

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

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

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

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

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

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

196 **2.3 Related Works on Ethnicity Prediction**

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

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

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

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

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

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

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

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

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

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

268 Aworinde and Onifade (2019) proposed a soft computing model of soft biometric
269 traits for gender and ethnicity classification using Gabor filter and KNN for feature extraction
270 and classification. Histogram equalization was used for preprocessing of the captured

271 fingerprint dataset. Result from the research proved to be 96% accuracy in predicting
272 person's ethnicity and gender with an average recognition time of less than 2seconds.
273 However, the research employed appearance based technique of identification in that focus
274 was majorly on the ridge arrangement and type for classification into various ethnic divides;
275 equally, the dataset used for the research is quite small for validation of the model.

276 Wong *et al.* (2020) predicted the ethnicity of Canadians with varying performance by
277 specific ethnic Categories employing the automated machine learning approach using only
278 name and census location features. Using census 1901, the multiclass and binary class
279 classification machine learning pipelines were developed. The 13 ethnic categories
280 examined were Aboriginal (First Nations, Me'tis, Inuit, and all-combined), Chinese, English,
281 French, Irish, Italian, Japanese, Russian, Scottish, and others. Machine learning algorithms
282 included regularized logistic regression, support vector, and naïve Bayes classifiers. The
283 census had 4,812,958 unique individuals. For multiclass classification, the highest
284 performance achieved was 76% F1 and 91% accuracy. For binary classifications for
285 Chinese, French, Italian, Japanese, Russian, and others, the F1 ranged 68–95% (median
286 87%). The lower performance for English, Irish, and Scottish (F1 ranged 63–67%) was likely
287 due to their shared cultural and linguistic heritage. Adding census location features to the
288 name- based models strongly improved the prediction in Aboriginal classification (F1
289 increased from 50% to 84%).

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

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

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

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

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

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

335 **2.4 Performance Evaluation Metrics**

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

- 338 i. Segmentation Accuracy (SA): This is the percentage of images that are correctly
339 segmented.
- 340 ii. Segmentation Time (ST): This is the total time use to segment the entire image in
341 the system in seconds.
- 342 iii. False Positive Rate (FPR): is the proportion of negative cases incorrectly identified
343 as positive cases in the data (i.e the frequency at which a system incorrectly
344 categorizes negative as positive). It measures when it's actually no, how often does
345 it predict yes?.
- 346 iv. Sensitivity: is the probability that the test indicates the presence of an iris image
347 belonging to an ethnicity in a created database.
- 348 v. Specificity: is the probability that the test indicated the presence of an iris
349 image belonging to an ethnicity but tested negative for that ethnicity in the
350 created database.
- 351 vi. Precision: is used to measure how often the system predict yes to a biometric trait
352 when it was actually no.
- 353 vii. Accuracy: Overall, how often is the classifier correct? (Newberg, 2006)

354 The selection of SVM parameters is actually an optimization problem in which search
355 algorithms are used to find the best configuration of parameters for given problem (De-
356 Miranda *et al.* 2012 and Gaspar *et al.* 2012). The use of evolutionary algorithm for
357 parameter optimization is very much faster and often gives better results (Friedrichs and Igel
358 2005 and Rossi and de Carvalho 2008). There are several methods to adjust the SVM
359 parameter among which are genetic algorithm (GA) and particle swarm optimization (PSO)
360 and Galactic Swarm Optimization (GSO).

361

362 **3. METHODOLOGY**

363 **3.1 Research Approach**

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

- 368 i. Construction of locally acquired iris database.
- 369 ii. Formulation of GAGAC and GSO-SVM Algorithms
- 370 iii. Designing an Iris-based Ethnicity Prediction System using the two formulated
371 Algorithms
- 372 iv. Software implementation and testing using Matrix Laboratory (MATLAB R2020a)
373 software.
- 374 v. Performance evaluation of the segmentation algorithms in (ii) using Segmentation
375 Accuracy (SA), Segmentation Time (ST) and the performance of the developed iris-
376 based ethnicity prediction system evaluated using, False Positive Rate (FPR),
377 Sensitivity, Specificity, Precision, Accuracy, classification time and statistical t-test of
378 significance.

379 **3.2 Data Collection and Construction of Locally Acquired Iris Database**

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

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

- 399 i. High resolution and good sharpness to enable accurate segmentation
- 400 ii. Good lighting condition i.e under controlled light intensity to prevent image
401 distortion.

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

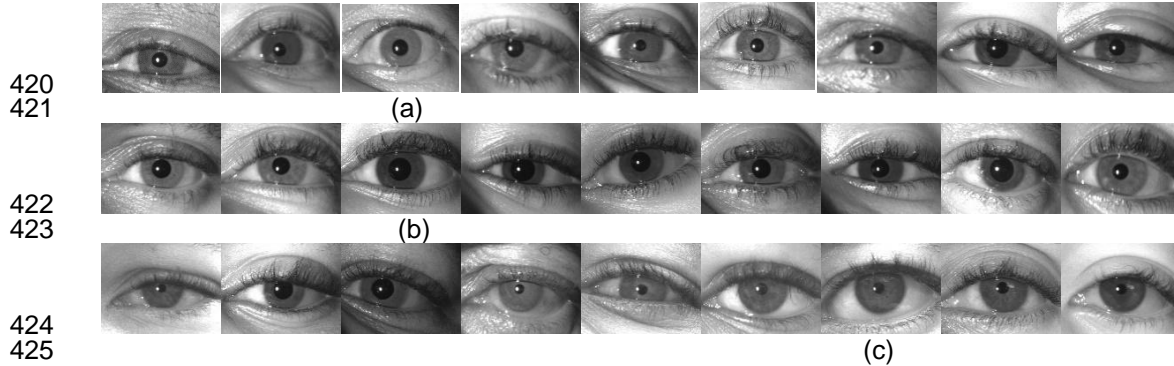
- 405 i. Capture iris: to capture a good image on the screen
- 406 ii. Re-snap image: to re-snap an image that was wrongly or not well captured
- 407 iii. Crop iris: to manually cut off other parts of the eyes and the face mistakenly
408 captured to reduce captured image to the required one
- 409 iv. Gray Scale: to change the image to gray scale form.
- 410 v. Save image: All the captured images were saved into work environment with their
411 respective iris identity (Yoruba, Ibo and Hausa) to afford the iris image to be further
412 processed. Figure 3.1 (a - c) presented selected original iris images of Yoruba, Ibo and

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

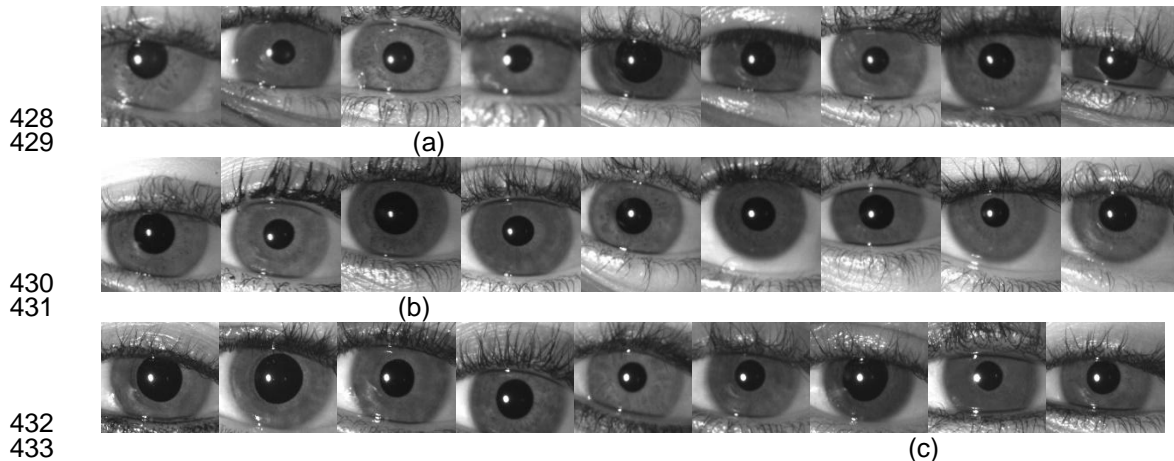
415 3.3 Formulation of GAGAC and GSO-SVM Algorithm

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

419



426 Figure 1: Selected original iris images (a) Yoruba (b) Ibo and (c) Hausa.
427



434 Figure 2: Selected cropped iris images (a) Yoruba (b) Ibo and (c) Hausa.

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

438 3.3.1 Algorithm of GAGAC (Genetic Algorithm based GAC)

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

448 3.3.2 Flow diagram of GSOSVM

449

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

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

457 Algorithm 3.1:

458 Algorithm for GAGAC Segmentation:

459 Phase 1: Find the Stopping function: K

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

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

462 t := 0;

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

464 WHILE stopping condition not fulfilled DO

465 BEGIN

466 (* proportional selection *)

467 FOR i := 1 TO m DO

468 BEGIN

469 x := Random[0,1];

470 k := 1;

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

472 k := k + 1;

473 $b_{i,t+1} := b_{k,t}$

474 END

475 (* one-point crossover *)

476 FOR i := 1 TO m - 1 STEP 2 DO

477 BEGIN

478 IF Random[0,1] $\leq P_C$ THEN

479 BEGIN

480 pos := Random{1, ..., n - 1};

481 FOR k := pos + 1 TO n DO

482 BEGIN

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

484 $b_{i,t+1}[k] := b_{i+1,t+1}[k]$

485 $b_{i+1,t+1}[k] := aux$

486 END

487 END

488 END

489 (* mutation *)

490 FOR i := 1 TO m DO

491 FOR k := 1 TO n DO

492 IF Random[0,1] $< P_M$ THEN

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

494 t := t + 1

495 END

496 Step 2: Filter the image with Gaussian filter $\sigma,(G(x,y))$

497
$$G(x, y) = \frac{1}{\sqrt{2\pi x\sigma}} \times e^{-\frac{x^2}{2\sigma^2}}$$

498 Step 4: Implement the equation for stopping function:

499
$$K(x, y) = \frac{1}{1 + \left(\frac{\|\nabla(G(x, y) \times I(x, y))\|}{k} \right)^\alpha}$$

500 Phase 2: Generating Ψ , zeroth level set:

501 Step 1: Input segmented pupil image.

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

503 Step 3: Generate ψ according to

504
$$\psi(x, y) = \begin{cases} 0, & \text{if } (x, y) \text{ is on the curve} \\ < 0, & \text{if } (x, y) \text{ is inside the curve} \\ > 0, & \text{if } (x, y) \text{ is outside the curve} \end{cases}$$

505 Step 4: Display it on input eye image.

506 Phase 3: Perform segmentation:

507 Step 1: Maximum iterations = Input from user

508 Step 2: ε = Input from user

509 Step 3: Propagation =1(constant)

510 Step 4: Initialize ψ

511 Step 5: Evolve ψ according to discrete implementation equation,

512
$$\frac{\psi_{i,j}^{t+1} - \psi_{i,j}^t}{\Delta t} = -cK'_{i,j} \|\nabla \psi^t\| - K'_{i,j} (\varepsilon K'_{i,j} \|\nabla \psi^t\|) + \nabla \psi^t_{i,j} \cdot \nabla K'_{i,j}$$

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

514 Step 7: Check number of iterations and convergence.

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

517 Step 9: Else Exit

518 Step10: Display final contour.

519

520 Phase 4: Estimation of radius

521 Step 1: Create mask by binarization of final contour.

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

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

524 Step 4: If yes then angle = 180°

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

526 Step 6: If yes then angle = 210°

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

528 Step 8: If yes then angle = 150°

529 Step 9: If no, Check calculated angle if it is less than or equal 32 and greater than 30.

530 Step 10: If yes then angle = 30°

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

532 Step 12: If yes then angle = 0°

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

534 Step 14: If yes then angle = -30°

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

536 Step 16: Take average of this 5 distances.

537 Step 17: Draw circle with this radius ± 20 .

538

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

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

547 All the necessary steps employed to perform the operation of the GSOSVM
548 algorithm were summarized in Figure 3, where the two phases of the GSO algorithm are
549 presented using PSO flow diagram presented in Figure 2. The pseudocode of GSOSVM
550 was presented in Algorithm 3.2 where w is the coefficient vector and b is the offset. ξ_i is a
551 slack variable introduced when linear is not separable; C is the penalty factor used to
552 represent the penalty index for misclassification. C determines the learning ability of SVM
553 and the experience risk coordination degree. $\text{sgn}(x)$ is a sign function, α_i^* is the Lagrange
554 coefficient corresponding to the support vector, b^* is the classification threshold. γ is the
555 RBF kernel function parameter which affects the distribution of complexity of sample data in
556 the characteristic space.
557

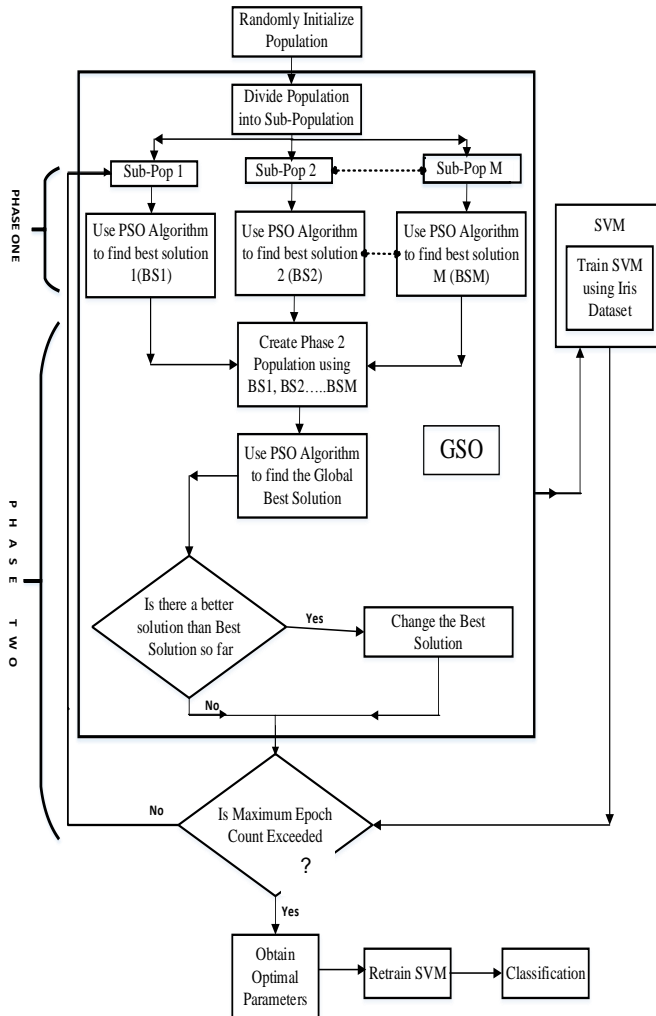


Figure 3: GSOSVM Flow Diagram

558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574

Algorithm 3.2. GSOSVM Algorithm

Inputs: Determine the various 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: $v_j^{(i)}, P_j^{(i)}$ within [xmin, xmax]D randomly
3. The population is divided into M subpopulations, $x_i \subset: i = 1, 2, \dots, M$

609 $f(x) = \text{sgn}\{(w^T \times x_i) + b\} = \text{sgn}\{\sum_{i=1}^n \alpha_i^* y_i (x_i^T \times x) + b^*\} = \text{sgn}\{\sum_{i=1}^n \alpha_i^* y_i K(x_i, x) + b^*\} \quad K(x_i, x) = f(x)^T \times f(x)$

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

611 End while

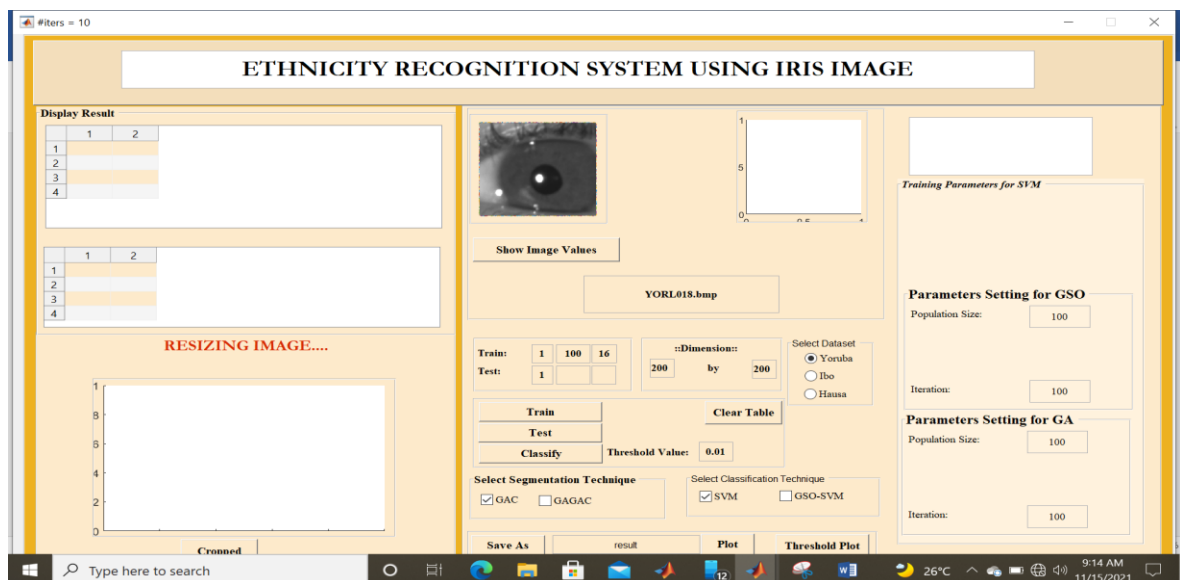
612 Return accuracy, where $f(x)$ =accuracy, $x=(C, \gamma)$

613

614 3.4 Implementation Procedure for the Iris-based Ethnicity Prediction 615 System using the Two Formulated Algorithms

616 The process flow diagram shown in Figure 4 was the representation of the basic
617 stages involved in the development of the iris-based ethnicity system. The training and
618 testing data were acquired locally considering three major tribes (Yoruba, Ibo and Hausa) in
619 Nigeria and normalized into a uniform pixel. Image enhancement of the image using
620 refinement technique was done to intensify and improve the images to make their features
621 visible. This was followed by segmentation using Genetic Algorithm based Geodesic Active
622 Contour (GAGAC)) process which isolated the iris from the eye region and located the inner
623 and outer boundaries of the iris, upper and lower eyelid detection revealing the details of the
624 iris features.

625 The normalization process was done using Daugman's Rubber Sheet Model to
626 transform the segmented iris region so that it has a fixed dimension for the comparison
627 template. The feature extraction/encoding was performed using Log Gabor Filters (since
628 field suggests that natural images are better coded by filters that have Gaussian transfer
629 function when viewed on the logarithmic frequency scale) for feature reduction, removal of
630 noise/unwanted elements and generation of unique feature codes before classification.
631 Classification between the training and testing data was done using GSO-SVM algorithm.
632 The iris-based ethnicity prediction system was implemented using Matrix laboratory R2020a.
633 The computational environment for this research is CPU (Intel i5-4210u 1.70 GHz) RAM (8
634 GB RAM) OS (not lower in version than Windows 7 Pro 64-bit). The performance of the iris-
635 based ethnicity prediction system was evaluated using some standard biometric prediction
636 metrics. The graphical user interface (GUI) was as shown in figure 4:



637

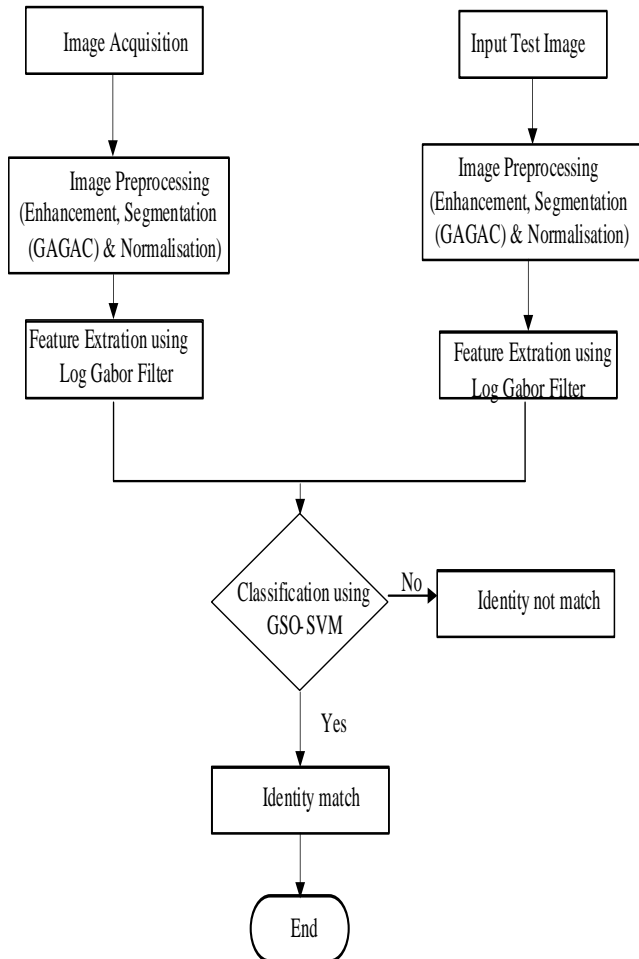
638 **Figure 4: Graphical user interface of the design system**

639

3.4.1 Image acquisition

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

644
 645



646

647 Figure 5: Flow Diagram for the Ethnicity Prediction System

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

653 3.4.2 Image preprocessing

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

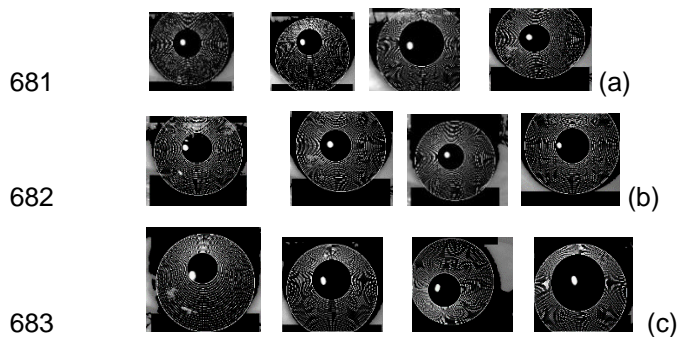
656 i. **Image enhancement:** In order to take advantage of the characteristics of the pupil
 657 described above, the image was enhanced using histogram equilization in order to obtain a
 658 reliable binary image of the pupil.

659 **ii. Segmentation:** This stage was taken care of using GAC and the formulated
 660 GAGAC as described in section 3.3.1. The segmentation methods in this research involves
 661 three major steps. First, the approximate location of the iris center is detected. Second, the
 662 iris region is extracted. Finally, the reflections (noises) are removed from the iris region. The
 663 significance of this steps is its robustness to realistic noises caused by non-ideal imaging
 664 settings such as reflections, blurred boundaries, gaze-deviation, and eyelids occlusion.
 665 Sample of segmented iris images was presented in Figure 6(a - c), sample of iris centre
 666 localization was presented in Figure 7(a-c) while sample of noise removal from images were
 667 presented in Figure 8(a – c).

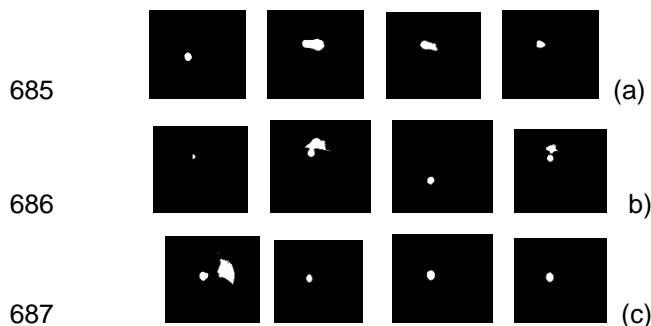
668 **iii. Normalization:** Once the iris region is successfully segmented from an eye image,
 669 the next step is to transform the iris region so that it has fixed dimensions for the
 670 comparisons of templates. The normalization process produces iris regions, which have the
 671 same constant dimensions, so that two photographs of the same iris under different
 672 conditions will have characteristic features at the same spatial location. Daugman’s rubber
 673 sheet model was employed here. Algorithm for daugman’s rubber sheet model is as shown
 674 in Algorithm 2.2 while sample of normalized iris images were presented in Figure 9.

675 **3.4.3 Feature encoding**

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



684 Figure 6: Segmented iris images (a) Yoruba (b) Ibo and (c) Hausa



688 Figure 7: Iris centre localization (a) Yoruba (b)Ibo and (c) Hausa



691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

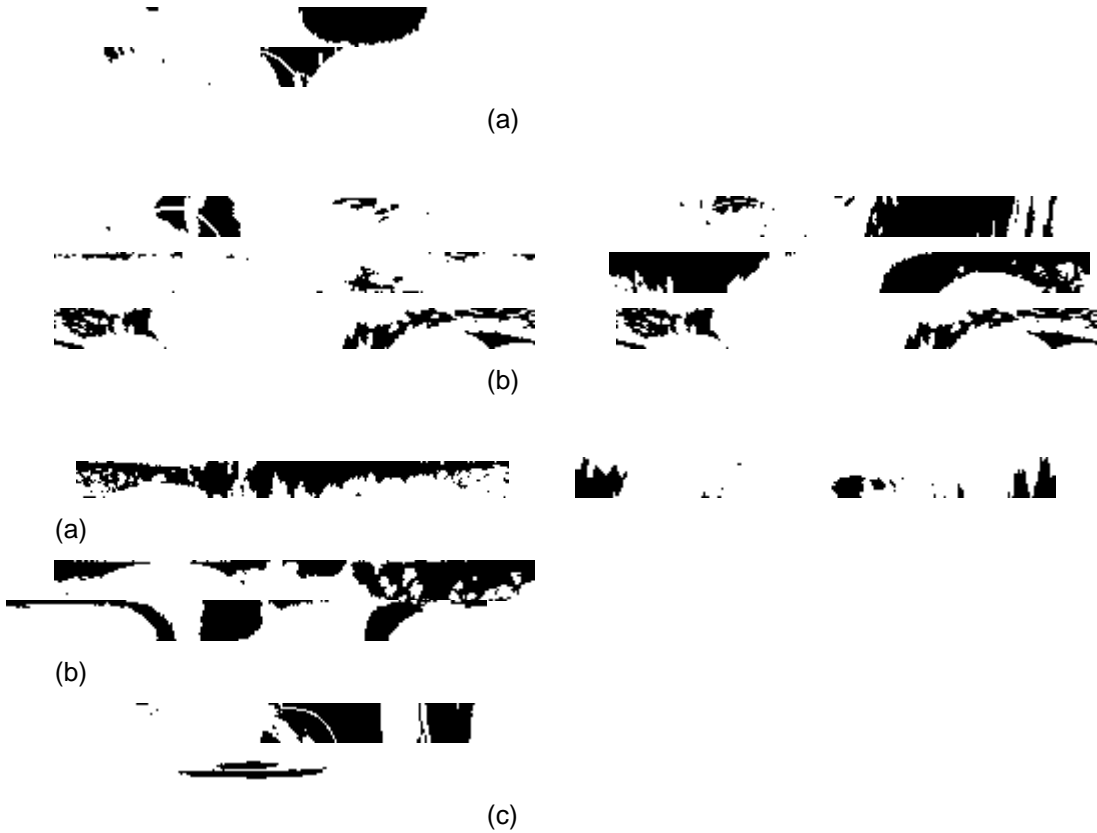


Figure 8: Noise removal of iris images (a) Yoruba (b) Ibo and (c) Hausa

708

709

710

711

712

713

714

715

716

717

718

719

720

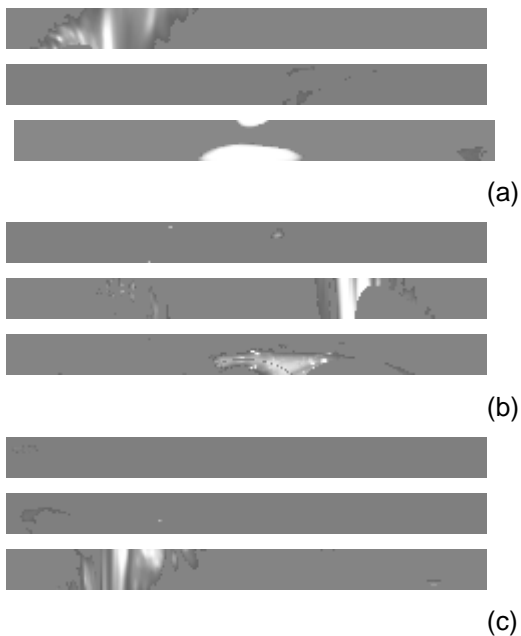


Figure 9 : Normalized Iris Images (a) Yoruba (b)Ibo and (c) Hausa

721

722

Table 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
2.000000000<F<3.000000000	0	0	200

723

724 3.4.4 Classification algorithm

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

727 3.5 Performance Evaluation Metrics for iris-based Ethnicity Prediction System

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

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

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

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

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

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

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

739

740 4. RESULTS AND DISCUSSION

741 4.1 Presentation of Results

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

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

754 **4.2 Results of Iris-based Ethnicity Prediction System.**

755 Parameter regularization which is a technique used to reduce errors by fitting the
 756 functions appropriately on the given algorithm to avoid overfitting were carried on the
 757 standard GAC and SVM using GA and GSO respectively. The results presented in Table 4.1
 758 depicts the performance of GAGAC segmentation technique and GSOSVM as classifier
 759 based on the three selected tribes. The result from Table 4.1(a-c) showed that increase in
 760 threshold value resulted to increase in classification performance. The system achieved a
 761 better performance at threshold 0.75 with 188, 192 and 190 correctly classified (TP) and 12,
 762 8 and 10 irises misclassified (FN) to other ethnic groups while 6, 2 and 4 irises were wrongly
 763 classified (FP) as Yoruba, Ibo and Hausa respectively.

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

768 **4.3 Discussion of results**

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

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

787 Table2: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

788 (a)

789

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

790 (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

791 (c)

792 **Table 3: 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

793 (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

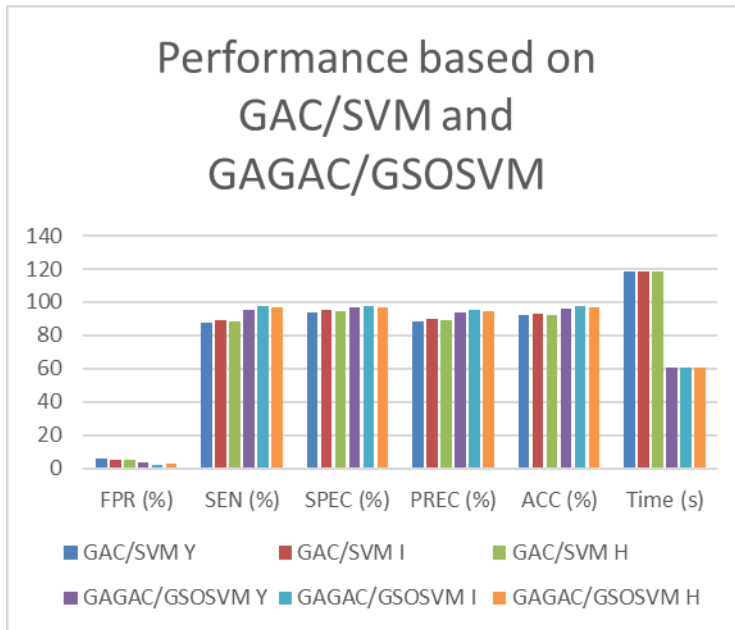
794 (b)

795

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

796 (c)

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



798

799

800

Figure 10: Graph showing the performance based on GAC/SVM and GAGAC/GSOSVM

801

802

803

804

805

806

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.

807

808

809

810

811

812

813

814

815

816

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).

817

818

819

820

821

822

823

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.

824

825

826

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.

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

837
838
839

5.0 CONCLUSIONS

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

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

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

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

875
876
877

6.0 RECOMMENDATION

878 This research has been able to developed an improved GAGAC segmentation and
879 GSOSVM classification technique for an iris-based ethnicity prediction system. It is therefore

880 recommended that the developed iris-based ethnicity prediction system can be employed to
881 access government benefits and enhance airline security by verifying traveler identity,
882 Researchers in the line of research can further look into the possibility of properly classifying
883 offspring of mixed marriage, extending the prediction to other ethnic groups in Nigeria and
884 people of different dialects within an ethnic group. Combining of multiple images or multiple
885 biometrics (such as face and iris, iris and fingerprint, iris and voice) can be employed to
886 improve performance of the ethnicity prediction system for better classification results.
887 Future research in the same lines of thought of GSO optimizing SVM could be investigated
888 using other global optimization algorithms such as GA, ACO, ABC, BA, or other heuristics
889 algorithm in place of PSO variants for both exploration or exploitation stage.
890

891 REFERENCES

- 892
- 893 [1] **Abbasi A. A., Khan M. N. A., and Khan S. A.,** (2013): “A Critical Survey of Iris Based
894 Recognition Systems”, Shaheed Zulfikar Ali Bhutto Institute of Science and
895 Technology (SZABIST), Islamabad, Pakistan. *Middle-East Journal of Scientific*
896 *Research*, 15 (5) pp 663-668
- 897 [2] **Aworinde, H., and Onifade, O. F. W.** (2019). A Soft Computing Model of Soft
898 Biometric Traits for Gender and Ethnicity Classification. *International Journal of*
899 *Engineering and Manufacturing*, 9(2), pp 54–63.
900 <https://doi.org/10.5815/ijem.2019.02.05>
- 901 [3] Arun Ross and Samir. Shah (2006), “Segmenting Non-Ideal Irises Using Geodesic
902 Active Contours,” *Proc. IEEE 2006 Biometric Symp.*, pp. 1-6, 2006.
- 903 [4] Aydin, I., Karakose, M. and Akin, E., 2011. A multi-objective artificial immune
904 algorithm for parameter optimization in support vector machine. *Applied Soft*
905 *Computing* 11, 120–129.
- 906 [5] **Bessudnov Alexey, Denis Tarasov, Viacheslav Panasovets, Veronica Kostenko,**
907 **Ivan Smirnov, Vladimir Uspenskiy** (2021) “Predicting ethnicity with data on personal
908 names in Russia” 1VK press release. 4 March
909 2021.<https://vk.com/main.php?subdir=press&subsubdir=q4-2020-results>
- 910 [6] **Caselles V., F. Catta, T. Coll, and F. Dibos,** (1993)“A geometric model for active
911 contours,” *Numer. Math.*, vol. 66, pp. 1–31, 1993.
- 912 [7] Chandra, K. (2006). What is Ethnic Identity and Does it Matter? *Annual Review of*
913 *Political Science*, 9, 397–424. <https://doi.org/> <https://doi.org/10.1146/annurev.polisci.9.062404.170715>
- 914
- 915 [8] **De-Miranda, P.B.C., Prudencio, R.B.C., Carvalho, A.C.P.L.F. and Soares, C.,**
916 2012. Combining a multi-objective optimization approach with meta-learning for SVM
917 parameter selection, in: 2012 IEEE International Conference on Systems, Man, and
918 Cybernetics (SMC). Presented at the 2012 IEEE International Conference on
919 Systems, Man, and Cybernetics (SMC), pp. 2909–2914.
- 920 [9] **Demirkus, M., Garg, K., and Guler, S.** (2010). Automated person categorization for
921 video surveillance using soft biometrics. *Proceedings of the SPIE*, 7667.
922 <https://doi.org/10.1117/12.851424>
- 923 [10] **Falohun A.S.** (2012): “Development of a feature extraction method for Iris
924 Recognition using enhanced inverse analytical fourier –mellin transform”. PhD Thesis,
925 Dept of Computer Science, LAUTECH, Ogbomoso.
- 926 [11] **Falohun A.S., Omidiora E.O, Olabiyisi O.S, Fenwa O.D and Adedeji O.T** (2010):
927 “Web-Based Iris Recognition System”, *Proceedings of the First International*
928 *Conference on Engineering and Technology*, ISBN 978-2902-58-6
- 929 [12] **Fangzhou Xie** (2022) “rethnicity: An R package for predicting ethnicity from names
930 SoftwareX, Volume 17, 2022,100965, ISSN 2352-7110,
- 931 [13] **Friedrichs, F. and Igel, C.,** (2005), “Evolutionary tuning of multiple svm parameters”,
932 *Neurocomputing* 64, pp 107–117.

- 933 [14] **Gaspar, P., Carbonell, J., Oliveira, J.L.**, 2012. On the parameter optimization of
934 Support Vector Machines for binary classification. *J Integr Bioinform* 9, 201.
- 935 [15] **Gugulethu Mabuza-Hocquet, Fulufhelo Nelwamondo and Tshilidzi Marwala**
936 (2016) "Ethnicity Prediction and Classification from Iris Texture Patterns: A Survey on
937 Recent Advances", 2016 International Conference on Computational Science and
938 Computational Intelligence, Conference Paper · December 2016 DOI:
939 10.1109/CSCI.2016.0159
- 940 [16] **Juan Tapia Farias, Claudio Perez and Kevin W. Bowyer**, (2016), "Gender
941 Classification From the Same Iris Code Used for Recognition", *IEEE Transactions on*
942 *Information Forensics and Security* 11(8):1-
943 1 · August 2016 DOI: 10.1109/TIFS.2016.2550418
- 944 [17] **Kaneshiro, B., Geling, O., Gellert, K., and Millar, L.** (2011). The challenges of
945 collecting data on race and ethnicity in a diverse, multiethnic state. *Hawaii*
946 *Medical Journal*, 70(8), pp168–171.
- 947 [18] **Keerthi, S. S., & Lin, C.-J.** (2003). Asymptotic behaviours of support vector machines
948 with gaussian kernel. *Neural Computation*, 15, pp 1667–1689.
- 949 [19] **Kim, M., Moon, H., and Pan, S. B.** (2012). Framework of Human Identification
950 using MultiModal Biometrics. *International Journal of Multimedia and*
951 *Ubiquitous Engineering*, 7(2), pp 509–514.
- 952 [20] **Lagree Stephen and Kevin W. Bowyer**, (2007), "Ethnicity Prediction Based on Iris
953 Texture Features"
- 954 [21] **Lagree S. and Bowyer K.W.** (2011): "Ethnicity Prediction Based on Iris Texture
955 feature", Department of Computer Science and Engineering, University of Notre
956 Dame, Notre Dame, Indiana 46556 USA.
- 957 [22] **Latinwo, B. M., Falohun, A. S., and Omidiora, E. O.** (2016). Principal Component
958 Analysis - Based Ethnicity Prediction Using Iris Feature, 16(4), pp 1–5.
959 <https://doi.org/10.9734/BJAST/2016/26131>
- 960 [23] **Latinwo, B., Falohun, A., Omidiora, E., and Makinde, B.** (2018). Iris Texture
961 Analysis for Ethnicity Classification Using Self-Organizing Feature Maps. *Journal of*
962 *Advances in Mathematics and Computer Science*, 25(6), pp 1–10
- 963 [24] **Luo, J., Ma, Y., Takikawa, E., Lao, S.H., Kawade, M., Lu, B.L.** (2007). *Person-*
964 *specific SIFT features for face recognition, ICASSP*, pp.563-566.
- 965 [25] **Lyle, J. R., Miller, P. E., and Woodard, D. L.** (2010). Soft Biometric Classification
966 Using Periocular Region Features. *IEEE*.
- 967 [26] **Makinde Bukola O., Olusayo D. Fenwa, Adeleye S. Falohun and Olufemi A.**
968 **Odeniyi** (2019) "Performance Evaluation of Kernel-Based Feature Extraction
969 Techniques for Face Recognition Systems", *FUOYE Journal of Engineering and*
970 *Technology*, Volume 4, Issue 1, March 2019 ISSN: 2579-0625 (Online), 2579-061
- 971 [27] **Maneet Singh, Shruti Nagpal, Mayank Vatsa, Richa Singh, Afzel Noore and**
972 **Angshul Majumdar**, (2017), "Gender and Ethnicity Classification of Iris Images using
973 Deep Class-Encoder", 2017 IEEE International Joint Conference on Biometrics
974 (IJCB), 978-1-5386-1124-1, pp 666-673
- 975 [28] **Michael Fairhurst, Meryem Erbilek and Marjory Da Costa-Abreu**, (2015),
976 "Exploring Gender Prediction From Iris Biometrics"; 2015 International Conference of
977 the Biometrics Special Interest Group (BIOSIG), DOI: 10.1109/BIOSIG.2015.7314602
978 Gesellschaft für Informatik e.V., Bonn, Germany
- 979 [29] **Minal K. Pawa, Sunita S. Lokhande and V. N. Bapat** (2012) "Iris Segmentation
980 using Geodesic Active Contour for Improved Texture Extraction in Recognition"
981 *International Journal of Computer Applications (0975 – 888) Volume 47– No.16, June*
982 *2012* 40
- 983 [30] **Orike, S., Anireh, V. I. E., and Ibironke, A. S.** (2016). A Gender and Ethnicity
984 Identification System in Nigeria using the Fingerprint Technology. *Proceedings of the*
985 *World Congress of Engineering*, I.

- 986 [31] **Qiu X. C., Sun Z. A., and Tan T. N.** (2006): "Global texture analysis of iris images for
987 ethnic classification", Springer LNCS 3832: International Conference on Biometrics,
988 pp 411-418, June 2006.
- 989 [32] **Qiu X. C., Sun Z. A., and Tan T. N.** (2006): "Global texture analysis of iris images for
990 ethnic classification", Springer LNCS 3832: International Conference on Biometrics,
991 pp 411-418, June 2006.
- 992 [33] **Romen V. Yampolskiy and Venu Govindaraju** (2008), "Behavioural biometrics , a
993 survey and classification", International Journal of Biometrics, 2008, Vol.1, no.1. DOI:
994 10.1504/IJBM 2008.018665 PG 81-113
- 995 [34] **Ratcliffe, P.** (2010): "Ethnic group": ISA Editorial Arrangement of sociopedia.isa
- 996 [35] **Rossi, A.L.D. and de-Carvalho, A.C.P.,** (2008). "Bio-inspired Optimization
997 Techniques for SVM Parameter Tuning". 10th Brazilian Symposium on Neural
998 Netresearchs, 2008. SBRN '08. Presented at the 10th Brazilian Symposium on Neural
999 Netresearchs, 2008. SBRN '08, pp. 57–62.
- 1000 [36] **Sabhanayagam T., V. Prasanna Venkatesan and K. SenthamaraiKannan,** (2018):
1001 "A Comprehensive Survey on Various Biometric Systems", International Journal of
1002 Applied Engineering Research ISSN 0973-4562 Volume 13, Number 5 (2018) pp.
1003 2276-2297
- 1004 [37] **Samir Shah and Arun Ross** (2009), "Iris Segmentation Using Geodesic Active
1005 Contours", IEEE Transactions on Information Forensics And Security, Vol.4, No.4,
1006 pp.824-836, Dec2009.
- 1007 [38] **Sarfaraz Masood, Shubham Gupta, Abdul Wajid and Suhani Gupta,** (2018),
1008 "Prediction of human ethnicity from facial images using neural netresearchs",
1009 Advances in Intelligent Systems and Computing · June 2018 DOI: 10.1007/978-981-
1010 10-3223-3_20
- 1011 [39] **Singh Maneet, Nagpal Shruti, Vatsa Mayank, Singh Richa, Noore, Afzel and**
1012 **Majumdar, Angshul.** (2017). "Gender and ethnicity classification of Iris images using
1013 deep class-encoder". 2017 IEEE International Joint Conference on Biometrics (IJCB),
1014 pp 666-673. 10.1109/BTAS.2017.8272755.
- 1015 [40] **Tariq, U., Hu, Y., and Huang, T. S.** (2009). Gender and ethnicity identification from
1016 silhouetted face profiles. In *Proceedings - International Conference on Image*
1017 *Processing, ICIP.* <https://doi.org/10.1109/ICIP.2009.5414117>
- 1018 [41] **Wong KO, Zaiane OR, Davis FG, Yasui Y.** A machine learning approach to predict
1019 ethnicity using personal name and census location in Canada. PLoS One. 2020
1020 Nov 18;15(11):e0241239. doi: 10.1371/journal.pone.0241239. PMID: 33206667;
1021 PMCID: PMC7673495.
- 1022 [42] **Yamakawa T.; Taniguchi, K.; Momen T.; Kobashi, S.; Kondo and K; Hata, Y.**
1023 (2007). *Biometrics Personal Identification Using Sole Information*, IEEE International
1024 Conference on Systems, Man and Cybernetics, pp. 3438-3442.
- 1025 [43] **Zarei A. and Duxing M.** (2012): "ANN for prediction of ethnicity based on Iris
1026 Texture", Department of Electrical Engineering, University of the Pacific Stockton, CA,
1027 USA.
- 1028 [44] **Zhang H., Z. Sun, T. Tan, and J. Wang** (2011). "Ethnic classification based on iris
1029 images". Chinese Conference on Biometric Recognition, pp. 82-90, Springer Berlin
1030 Heidelberg, 2011.