UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

PSO-FuzzyNN Techniques in Gender Classification Based on Bovine Bone Morphology Properties

Nur Afiqah Sahadun, Nor Azizah Ali & Habibollah Haron
Faculty of Computing
Universiti Teknologi Malaysia
81310 UTM Johor Bahru, Johor, Malaysia
afiqah_sahadun@yahoo.com, nzah@utm.my, habib@utm.my

Submitted: 12/01/2018. Revised edition: 10/10/2018. Accepted: 21/10/2018. Published online: 30/05/2019
DOI: https://doi.org/10.11113/ijic.v9n1.215

***Abstract*—This simulation project aims to solve forensic anthropology issues by using the computational method. The positive identification on gender is such a potential field to be explored. Basically, gender identification in forensic anthropology by comparative skeletal anatomy by atlas and crucially affect the identification accuracy. The simulation identification method was studied in order to determine the best model, which reduce the total costs of the post-mortem as an objective. The computational method on simulation run improves the identification accuracy as proven by many studies. Fuzzy K-nearest neighbours classifier (FuzzyNN) is such a computational intelligence method and always shows the best performance in many fields including forensic anthropology. Thus, this intelligent identification method was implemented within the determining for best accuracy. The result of this proposed model was compared with raw data collection and standard collections datasets; Goldman Osteometric dataset and Ryan and Shaw Dataset (RSD) as a benchmark for the identification policy. To improve the accuracy of FuzzyNN classifier, Particle Swarm Optimization (PSO) feature selection was used as the basis for choosing the best features to be used by the selected FuzzyNN classification model. The model is called PSO-FuzzyNN and has been developed by MATLAB and WEKA tools platform. Comparisons of the performance measurement namely the percentage of the classification accuracy of the model were performed. The result show potential the proposed PSO-FuzzyNN method demonstrates the capability to the obtained highest accuracy of identification.**

***Keywords*—Gender classification, Forensic anthropology, Non-human**

I. INTRODUCTION

Gender Classification from cancellous bone has proven to be of great importance in the forensic anthropology in term of diagnosing decomposition or damage cancellous bone using fracture knowledge. Due to decreasing bone indicator resources which in turn lead to significant cost saving [1]. Thus, it was seen as a solution in order to optimize evidence material while contributing to positive identification. Traditional models for gender classification such as Fourier transform (FFT), discrete cosine transform (DCT) are based on qualitative data which is the observation of bone morphology from a frequency domain by the forensic anthropologist. These models contain considerable intra and inter-observer variability which can cause large errors of the gender classification due to human error. In addition, traditional models have raised issues with regard to accurate classification and challenge in management of data to identify optimum features and interpretation optimum features in a simple way. In this project, all these three issues were addressed by using a process model developed specifically for gender classification. This project used computational intelligence models, namely Fuzzy K-nearest neighbours classifier (FuzzyNN). To improve the accuracy of FuzzyNN classifier, Particle Swarm Optimization (PSO) feature selection was used as the basis for choosing the best features to be used FuzzyNN classification model.

This project is structured as follows: In section I, we present the introduction of the project. Then Section 2 provides the materials and method as the main body that used in this project. We present the final result for classification in Section 3. And last Section IV gives concluding remarks.

II. Materials and Methods

A. Subjects

There are three datasets, 1) Cancellous Bone of Bovine Fracture (CBBF) 2) Ryan and Shaw Dataset (RSD) and 3) Goldman Osteometric Online Dataset (GOOD) as shown in Table 1.

TABLE 1. Datasets

Data	**Female**	**Male**
Cancellous Bone of Bovine Fracture (CBBF)	7	8
Ryan and Shaw Dataset (RSD) Ryan and Shaw (2013)[2]	14	11
Goldman Osteometric Online Dataset (GOOD) Benjamin (2012)[1]	45	61

B. Pre-processing

There are five steps in data pre-processing which are data cleaning, declaration target output, normalization and division. Essentially, there is a need for data pre-processing in order to get the appropriate range for the model development when using predictor learning. Thus, high quality of data is the first and most important before running the experiment.

C. Fuzzy K-nearest Neighbours Classifier

Fuzzy K-nearest neighbours (FuzzyNN) classifier is an old-er approach based on the fuzzy-rough sets theory suitable for classification task especially [12].

1) Fuzzy K-nearest Neighbours for Gender Classification

Overall, the FuzzyNN for gender classification are divided into two; Part A and Part B.

Part A: get the k nearest neighbours of the test pattern x

Let **X={x,,** x2, ..., **x,}** be the set of already labeled data (training data), and C=(cI,cz, ... cc} is the result classification space. Let x be the unlabeled test data.

<u>Input x;</u>

Set K, 1 I K g n ;
Set the iteration counter count-1;
For all x_j **E X** (1 < j l n) Do
Compute 11 x-x, 11
If (i I K)
include x_j in the set of K-nearest neighbours and increase count by 1
else if (x_j is closer to x than any previous nearest neighbour)
Begin
Delete the farthest of the K-nearest neighbours
Include x_j in the set of K-nearest neighbours
End
End For

Part B: approximate x by the k-nearest neighbours

For all c_j **E** C (1 l i l c) Do
End For

where:

$$\sum X_n = \frac{\sum_{j=1}^{k} u_{ij} \ \left(\| x - x_j \|\right)^{\frac{-2}{m-1}}}{\sum_{j=1}^{k} \ \left(\| x - x_j \|\right)^{\frac{-2}{m-1}}} \tag{1}$$

Here, u_i, represents the membership of x, to the i” class, and m determines how heavily the distance is weighted when calculating each neighbour’s contribution to the membership value.

Part A is to choose some of the training data points that are similar to the test data point **as** its neighbours and part B is to use the membership functions of the selected neighbours to compute the approximated membership of the test data point. The fuzzy-rough nearest neighbour classification approach is a further generalization of the fuzzy nearest neighbour approach [ll]. The main idea of the algorithm is listed in the flowing?

Let X = (x_l, x_2, ..., **x,)** be the training data set, and x be the test data. Let C_d=(F_l,F_2) be a fuzzy partition on X,

where:
Here, u_i, represents the membership of x, to the i” class, and m determines how heavily the distance is weighted when calculating each neighbour’s contribution to the membership value.

Part A is to choose some of the training data points that are similar to the test data point **as** its neighbours and part B is to use the membership functions of the selected neighbours to compute the approximated membership of the test data point. Fuzzy-rough nearest neighbour classification approach is a further generalization of the fuzzy nearest neighbour approach [ll]. The main idea of the algorithm is listed in the flowing?

Let X = (x_l, x_2, ..., **x,)** be the training data set, and x be the test data. Let C_d=(F_l,F_2) be a fuzzy partition on X,

where:

$$F_1 = (f_{11}, f_{12}, \dots, f_{1n}) 0 \leq \sum_{j=1}^{n} f_{1j} \leq n \qquad (2)$$

$$F_2 = (1 - f_{11}, 1 - f_{12}, \dots, 1 - f_{1n}) \qquad (3)$$

Here f_{ij} means the membership degree that x, is similar to the test pattern **x** compared with all elements in the training set. Then we can approximate the output class **C,** over this fuzzy partition over X.

$$\mu_{\underline{\underline{Cc}}}(F_1) = inf_j max(1 - f_{i1}, u_{je}) \qquad 1 \leq j \leq n \qquad (4)$$

$$\mu_{Ce}(F_1) = sup_j min(f_{ij} - u_{je})^3 \qquad 1 \leq j \leq n \quad (5)$$

Finally **p ~ F(1)** and μ (F_1) can be used **as** the lower and upper approximation of the membership of F_1 to C, which is also the membership of **x** to C. Compared with the conventional fuzzy K-NN algorithm, this algorithm can also be divided into two parts. The first part is exactly the same **as** part A of conventional fuzzy K-NN algorithm. But its second part adopts a different approximation method to get the fuzzy-rough memberships of the test data point. This method includes not only the fuzzy uncertainties but also the rough uncertainties.

2) PSO Features Selection for Gender Determination

This research focused on applying Particle Swarm Optimization as a wrapper approach in feature selection. PSO is originally developed by Kennedy and Eberhart (1995) based on social life behavior. PSO is used to combine feature subsets and operators together to find the optimal subset of features. PSO originated in real social life by graphic simulate of bird of bird flocks searching food behaviors. All particles fly through the problem space guide by current optimum particles is a potential solution. Each particle has own position and velocity vector. All particles take the challenge by cooperating and competing in searching the optimum results and they will discover the best feature combinations as they fly in subset space.

Each random result and search for optimum results with updated the generations. PSO concept based on the social and the cognitive model while it updates the particles (Lin *et al*., 2008). In order to gain insight into how PSO works, Wang *et al.*, (2007) summarized it in a more understandable explanation. The performance of PSO is evaluated by the fitness function of each particle to meet termination criteria such as maximum generation. The best fitness in particles will be selected as the global best particle (gbest). Each particle will record its best previous position (pbest) and velocity (V). The value of V will be determined after particle position has been updated. V is responsible to control the PSO swarm [13].

III. EXPERIMENTAL RESULT

The following result is involve three datasets. The first dataset states the results for the long bone Malaysian bovine bone. The second long bone Old monkey world and third long bone of United State Adult population Dataset.

A. Trabecular Bones of Malaysian Bovine Population Data Set

The analysis of the result discussed the performance of accuracy of the classification techniques based on the experimental results is obtained in the classification process. The performance of classification accuracy is divided into two parts, namely classification accuracy for training and testing. Classification accuracy is visualized in the form of confusion matrix which is used as a supervised learning. Thus, the values of the determination accuracy obtain for Malaysian Bovine Bones dataset for FuzzyNN and PSO-FuzzyNN model, 70.60% and 91.67% respectively as shown in Table 2.

TABLE 2. Evaluation of FuzzyNN and PSO-FuzzyNN model

Model	Accuracy (%)
FuzzyNN	70.60%
PSO-FuzzyNN	91.67%

B. Trabecular Bones of Old World Monkey Population Data Set

Overall, the FuzzyNN and PSO-FuzzyNN determination models are developed. Then, the next step is to determine the best FuzzyNN and PSO-FuzzyNN model in order to obtain the best determination model for gender in forensic anthropology post-mortem process. The best FuzzyNN and PSO-FuzzyNN determination model was determined based on the evaluation parameters, which are the percentage of accuracy. The highest percentage of accuracy, sensitivity, specificity of the model will be considered as the best FuzzyNN and PSO-FuzzyNN determination model. Table 3 shows the FuzzyNN and PSO-FuzzyNN models with the the percentage of accuracy.

TABLE 3. Evaluation of FuzzyNN and PSO-FuzzyNN model

Model	Accuracy (%)
FuzzyNN	81.80%
PSO-FuzzyNN	83.33%

C. Trabecular Bones of Malaysian Bovine Population Data Set

The FuzzyNN and PSO-FuzzyNN determination models are developed. Then, the next step is to determine the best the FuzzyNN and PSO-FuzzyNN model in order to obtain the best determination model for gender in forensic anthropology post-mortem process. The best the FuzzyNN and PSO-FuzzyNN determination model was determined based on the evaluation parameters, which are percentage of accuracy.

TABLE 4. Evaluation of FuzzyNN and PSO-FuzzyNN model

Model	Accuracy (%)
FuzzyNN	89.83%
PSO-FuzzyNN	90.31%

IV. CONCLUSION

This gender determination in forensic anthropology studies focuses on positive identification. The method of simulation determination methods is analyzed to determine the best model in order to determine highest accuracy determination and reduce the post-mortem cost. In order to analyze the performance of Fuzzy K-nearest neighbours (FuzzyNN) and Particle Swarm Optimization- Fuzzy K-nearest neighbours (PSO-FuzzyNN), the other standard datasets was used as a benchmark. Comparisons of the performance measurement namely the percentage of the classification accuracy showed the performance of PSO-FuzzyNN by Trabecular Bones of Malaysian bovine population Data outperformed FuzzyNN by all dataset with 91.67%.

ACKNOWLEDGMENT

The author would like to thank the Ministry of Science, Technology and Innovation (MOSTI), Ministry of Higher Education (MOHE), and was conducted in collaboration with the Research Management Center (RMC) at the Universiti Teknologi Malaysia (UTM) under Vot Number R.J130000.7828.4F509.

REFERENCES

[1] Benjamin, M. Auerbach. Goldman Osteometric Dataset. http://web.utk.edu/~auerbach/GOLD.htm.

[2] Ryan, T. M., and Shaw, C. N. (2013). Trabecular Bone Microstructure Scales Allometrically in the Primate Humerus and Femur. *Proceedings of the Royal Society B: Biological Sciences*, 280(1758), 1-9.

[3] Guyomarc'h, P., & Bruzek, J. (2011). Accuracy and Reliability in Sex Determination from Skulls: A Comparison of Fordisc 3.0 and the Discriminant Function Analysis. *Forensic Science International*, 208(1), 180-e1.

[4] Giles, E., & Elliot, O. (1963). Sex Determination by Discriminant Function Analysis of Crania. *American Journal of Physical Anthropology*, 21(1), 53-68.

[5] Mays, S., & Cox, M. (2000). Sex Determination in Skeletal Remains. *Human Osteology in Archaeology and Forensic Scienc*e, 117-130.

[6] Bidmos, M. A., and Dayal, M. R. (2004). Further Evidence to Show Population Specificity of Discriminant Function Equations for Sex Determination Using the Talus of South African Blacks. *Journal of Forensic Sciences*, 49(6), 1165-1170.

[7] Kemkes-Grottenthaler, A. (2005). Sex determination by Discriminant Analysis: An Evaluation of the Reliability of Patella Measurements. *Forensic Science International*, 147(2), 129-133.

[8] Gualdi-Russo, E. (2007). Sex Determination from the Talus and Calcaneus Measurements. *Forensic Science International*, 171(2), 151-156.

[9] Wilkinson, G. S. (2009). Methods for Age Estimation and the Study of Senescence in Bats. *Ecological and behavioral Methods for the Study of Bats*, 315-326.

[10] Kemkes-Grottenthaler, A. (2005). Sex Determination by Discriminant Analysis: An Evaluation of the Reliability of Patella Measurements. *Forensic Science International*, 147(2), 129-133.

[11] Akhlaghi, M., Moradi, B., and Hajibeygi, M. (2012). Sex Determination Using Anthropometric Dimensions of the Clavicle in Iranian Population. *Journal of Forensic and Legal Medicine*, 19(7), 381-385.

[12] Bian, H., & Mazlack, L. (2003, July). Fuzzy-rough Nearest-neighbor Classification Approach. *Fuzzy Information Processing Society, 2003. NAFIPS 2003. 22nd International Conference of the North American IEE*E. 500-505.

[13] Rahman, S. A., Bakar, A. A and Hussein, Z. A. M. (2009). Filter-wrapper Approach to Feature Selection Using RST-DPSO for Mining Protein Function. *2nd Conference on Data Mining and Optimization*. 71-78.