International Journal of Computer Science and Information Security (IJCSIS),
Vol. 14, No. 7, July 2016

Support Vector Machine, Multilayer Perceptron Neural Network, Bayes Net and k-Nearest Neighbor in Classifying Gender using Fingerprint Features

S.F. Abdullah
Optimisation, Modelling, Analysis, Simulation and Schedulling (OptiMASS) Research Group
Universiti Teknikal Malaysia Melaka
76100 Durian Tunggal, Melaka, Malaysia.

Z.A. Abas
Optimisation, Modelling, Analysis, Simulation and Schedulling (OptiMASS) Research Group
Universiti Teknikal Malaysia Melaka
76100 Durian Tunggal, Melaka, Malaysia.

A.F.N.A. Rahman
Optimisation, Modelling, Analysis, Simulation and Schedulling (OptiMASS) Research Group
Universiti Teknikal Malaysia Melaka
76100 Durian Tunggal, Melaka, Malaysia.

W.H.M Saad
Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka
76100 Durian Tunggal, Melaka,Malaysia.

***Abstract*—A scientific study of fingerprints, lines, mounts and shapes of hands are called dermatoglyphics. Dermatoglyphics features from fingerprint are statistically differ between the gender, ethnic groups, region and age categories From the previous study of gender classification in forensic area, the process of feature extraction is done manually and classify using a statistical approach. The features extracted were; ridge count (RC), ridge density (RD), ridge thickness to valley thickness ratio (RTVTR) and white lines count (WLC). The sample use consists of 300 respondents where each respondent gives 10 different fingerprints. Four classifiers which are Bayes Net, Multilayer Perceptron Neural Network (MLPNN), k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) are used in order to evaluate the performance of the proposed algorithm. The overall performance of the classifier is 95% of the classification rate. From all classifiers, SVM emerges as the best classifier for proposed algorithm.**

Keywords*—*fingerprint, gender classification, SVM, MLPNN, k-NN, Bayes Net

I. INTRODUCTION

Fingerprint and other skin ridges are called dermatologlyphics. This term is a combination of the ancient Greek words ‘derma’ meaning a skin and ‘glyph’ for carving which refers to the natural ridges on the surface of the skin epidermis that do not change their size or shape over time. According to some researches, special characteristic from fingerprint ridges has been proven statistically to differentiate gender, ethnic group and age categories [1][2] of their owners.

Fingerprint has become the most popular biometrics used in security area since a long time ago. This is due to their high acceptability, immutability and uniqueness [3]. The immutability of the fingerprint refers to the patterns that remain unchanged over time, whereas the uniqueness is related to the differences between the individual ridge details across the whole fingerprint image. Fingerprint is always associated with the criminology especially in forensics [4] and it has been used and accepted since 1975 as an important way to recognise a person’s gender [5].

There are two levels of features in the fingerprint structure which are the global feature and the local feature. Local feature is a tiny unique characteristic of fingerprint ridge, as shown in Figure 1. It refers to the ridge and valley details which carry the information of individuality of the fingerprint while global feature refers to the pattern that carry the information of the fingerprint class that can be seen with the naked eye.

Fig.1 The ridge and valley

Some research and publication works discovered and appraised the weakness of the gender classification accuracy [3][6][7][8], while some researcher analysed the correlation of fingerprint with gender of an individual by proposing a new classification method to overcome gender classification problem [4][6][7][8]. Furthermore, several publications enhance the accuracy by comparing the classification rate

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

using different classifier due to the great potential of fingerprint as an efficient classification method [9][10].

A new algorithm based on the Acree's theory, focusing on fingerprint global feature extraction was proposed by Abdullah *et. al.* and has been implemented for gender classification [8][11]. The algorithm automatically conducts the ridge calculation process from the 25mm2 square box in order to enhance the gender classification process previously done in forensic laboratory. The classification result of this study was validated with previous descriptive statistical method used by forensics expert to classify the gender. The algorithm was able to achieve 74.50% correctly classified gender when compared to previous manually method. However, this work cannot be compared with other gender classification algorithm by other researchers because of the different approach of classification method used. All the previous researchers used the data mining approach in order to evaluate their proposed algorithm.

Thus, this study aims to evaluate the performance of the same algorithm using the different approach of classification, specifically the data mining approach. Bayes Net, Multilayer Perceptron Neural Network (MLPNN), Support Vector Machine (SVM) and k- Nearest Neighbor (k-NN) were used as a classifier in order to determine the best classifier for a gender classification algorithm proposed by Abdullah *et al*. [8].

The rest of this paper is organized as follows. Section 2 presents the methodology that has been done in this study, while the result of this experiment is shown in Chapter 3. Chapter 4 discusses the result and the comparison of the performance and lastly, Chapter 5 presents the conclusion.

II. METHODOLOGY

The sample of this study consist of 3000 fingerprint images where 1430 fingerprint were taken from male samples and another 1570 are from female. The database of the extracted features, taken from Abdullah *et. al.* [8] was used in order to evaluate the performance of the proposed algorithm.

There are four features extracted which consist of Ridge Count, Ridge Density, Ridge Thickness to Valley Thickness Ratio and White Lines Count. The classifiers used in this experiment were Bayes Net, Multilayer Perceptron, Neural Network (MLPNN), Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN).These four classifiers are run using open source software Waikato Environment for Knowledge Analysis (WEKA).

The accuracy and the confusion matrix of each classifier was compared with four different tests as shown in Figure 3. The test included 1 fold cross validation, 80% train and 20% test, 70% train and 30% test, and lastly 60% train and 40% test. The results of each test were in form of accuracy and the confusion matrix.

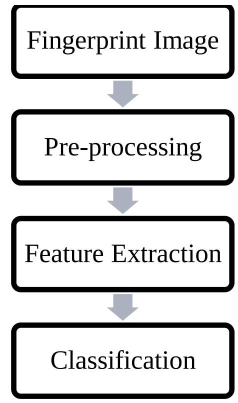

Fig.2 The block diagram of the methodology process

Figure 2 shows the methodology of the gender classification process used in this study. In the initial stage, each fingerprint was manually collected using data collection forms and was scanned using Fuji Xerox Docuscan C4250 before undergoing some of preprocessing technique and feature extraction process. This is done in order to extract the relevant and important data as previously conducted by Abdullah *et al.* [8][11][12]. In the meantime, the final step of this study compared the accuracy of selected classifier. Figure 3 shows the four different classifiers with the four different tests used in evaluating the performance of the proposed algorithm.

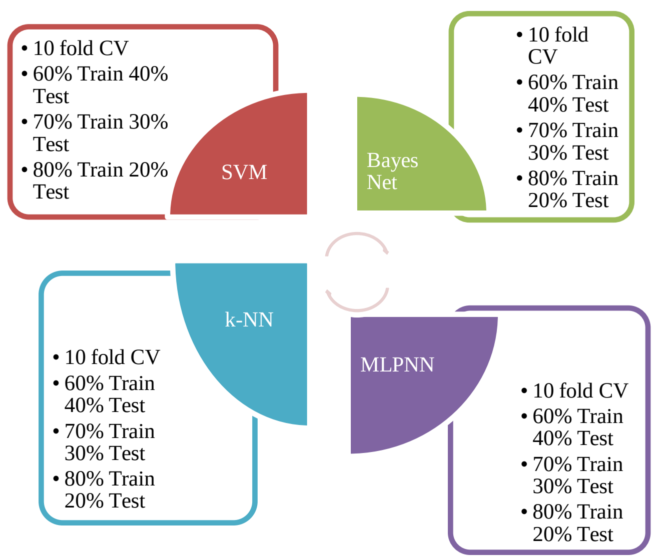

Fig. 3 Classifier with different test option

III. RESULT

The result of each classifier is given in Table I and the result is illustrated in bar chart as shown Figure 4. It can be seen that Multilayer Perceptron Neural Network (MLPNN) classifier has a higher classification rate for all different test compared to k-Nearest Neighbor (k-NN) which shown the lowest classification rate. The Multilayer Perceptron Neural Network (MLPNN) classifier provides 100% accuracy for 70%

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

train and 30% test option. It can be said that the higher percentage of training data provide the lowest accuracy when tested with Multilayer Perceptron Neural Network (MLPNN).

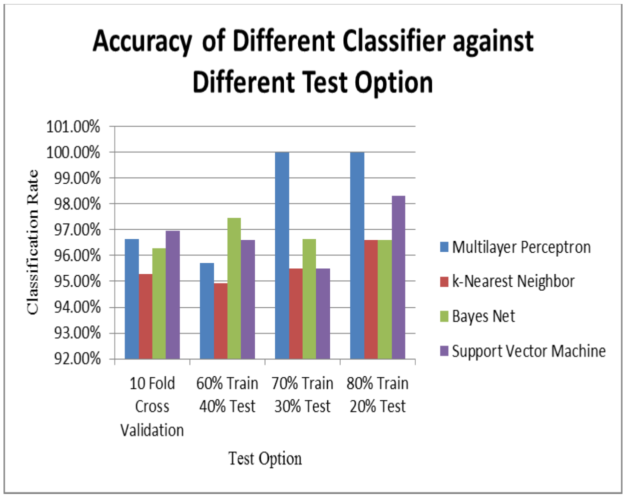

Fig 4. Accuracy of different classifier against different test

The performance of the classification rate of MLPNN changed drastically for each case and shown a fluctuation in accuracy. It is due to overfitting problem which always happens to neural network model. The error on the training set was driven to a very small value when a new large data presented to the network. The network memorised the training data model, but it has not learned to generalise the new situation of data. To overcome this over fitting problem, the number of features needs to be reduced, sustaining the most important features. Compared to the other classifier, there were some increase and decline of the accuracy when it was tested with the different test option.

The 10 fold cross validation case is a process of evaluating the predictive models by portioning the original data into 9 folds (training set) as the train model and 1fold (test set) to evaluate it. This portioning process will be repeated for another 9 times. Each fold will have the opportunity to be as the test set for each time. The accuracy of each test was recorded and the mean of these accuracies is calculated. For this case, it can be seen that for all classifiers provide almost the same accuracy which was 95.9% of the classification rate.

From the result, Support Vector Machine (SVM) provides the highest classification with 96.95% classification rate compared to other three classifiers tested with the 10 fold cross validation. All classifiers shown the slight difference in classification rate and all were able to achieve classification rate of above 95%.

IV. Discussion

The result of this study was compared with the previous result of gender classification in other countries. In year of 2006, Badawi *et al.* [10] have conducted his study of gender classification using fingerprint in India. This study used three different types of features which are RTVR, Ridge Count (RC) and White Lines Count (WLC) and the Neural Network (NN) as classifiers. The result from this study provides 87.46% classification rate. This is in contrast with Manish *et al.* [13] that only used two features which are RTVTR and Ridge Density (RD) and was able to achieve 88.00 % of classification rate using Support Vector Machine (SVM) as a classifier.

Meanwhile Gupta *et al.* [14] and Rajesh *et. al.* [7] used the same method of feature extraction, but different hybridization method. Gupta *et al*. [14] combined the two methods, Discrete Wavelet Transform (DWT) and Back Propagation Neural Network (BPNN) in order to classify the gender and from their study, they were able to achieve 91.45 % classification rate. Different to Rajesh *et al.* [7], the use of Discrete Wavelet Transform (DWT) was used in analysing fingerprints in frequency domain analysis and Gaussian Mixture Model (GMM) for classifying the dominant features by using rank. GMM was used because of the ability to approximate the distribution of the patterns of an image. They achieved 92.60% at the 3rd level of DWT decomposition.

Gender classification accuracies of the proposed method and the published result are shown in Table II. It is shown that the result of the proposed method shown a 96.95% classification rate, which is higher from the previous study done before.

V. Conclusion

In conclusion, the performance of the proposed algorithm was tested using four different types of classifiers with four different tests. There are four classifiers used in this study, they are Bayes Net, Multilayer Perceptron Neural Network (MLPNN), k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM). Furthermore, four different test cases were used to test the classifier. They are 10 fold cross validation case, 70% train 30% test case, 60% train 20% test case and lastly, the 60% train 40% test case. From the result, we can conclude that overall classification for all classifiers achieved more than 90% classification rate. However, SVM emerged as the best classifier for the proposed algorithms. Our future work will be continued with the study and comparison of different features involved in this gender classification by using the fingerprint.

Acknowledgment

This work was supported by research grants RAGS/2013/FTMK/ICT02/01/B00039 under the Ministry of Higher Education of Malaysia for financial support.

References

[1] Nigeria, Yaba Lagos. "Analysis, Design and Implementation of Human Fingerprint Patterns System "Towards Age & Gender Determination, Ridge Thickness To Valley Thickness Ratio (RTVTR) & Ridge Count On Gender Detection." Analysis 1, no. 2 (2012).

[2] Gnanasivam, P., and Dr S. Muttan. "Estimation of age through fingerprints using wavelet transform and singular value decomposition." International Journal of Biometrics and Bioinformatics (IJBB) 6, no. 2 (2012): 58-67.

[3] Wadhwa, Ravi, Maninder Kaur, and Dr KVP Singh. "Age and Gender Determination from Finger Prints using RVA and dct coefficients." IOSR Journal of Engineering (IOSRJEN) (2013).

[4] Nte N. D. 2012. An Evaluation of the Challenges of Forensic Investigation and Unsolved Murders in Nigeria. African Journal of Criminology and Justice Studies: AJCJS, Vol. 6, No. 1/2, p.143.

[5] Reddy G. G. 1975. Finger dermatoglyphics of the Bagathas of Araku Valley (AP), India. American journal of physical anthropology, Vol. 42, No. 2, pp.225-228.

[6] Dongre, Ganesh B., and Dr SM Jagade. "A Review and Study on Fingerprint Based Gender Classification using Classified Techniques." International Journal of Scientific & Engineering Research 6, no. 3 (2015): 596.

[7] Rajesh, D. Gnana, and M. Punithavalli. "Wavelets and Gaussian mixture model approach for gender classification using fingerprints." In Current Trends in Engineering and Technology (ICCTET), 2014 2nd International Conference on, pp. 522-525. IEEE, 2014.

[8] S. F. Abdullah, A.F.N.A. Rahman, Z.A.Abas and W.H.M Saad, "Development of a Fingerprint Gender Classification Algorithm Using Fingerprint Global Features" International Journal of Advanced Computer Science and Applications(IJACSA), 7(6), 2016.

[9] Ceyhan, Eyup Burak, and Seref Sagiroglu. "Gender inference within Turkish population by using only fingerprint feature vectors." In Computational Intelligence in Biometrics and Identity Management (CIBIM), 2014 IEEE Symposium on, pp. 146-150. IEEE, 2014.

[10] Badawi, Ahmed M., Mohamed Mahfouz, Rimon Tadross, and Richard Jantz. "Fingerprint-Based Gender Classification." In IPCV, pp. 41-46. 2006.

[11] Abdullah, S. F., A. F. N. A. Rahman, Z. A. Abas, and W. H. M. Saad. "Multilayer Perceptron Neural Network in Classifying Gender using Fingerprint Global Level Features." Indian Journal of Science and Technology 9, no. 9 (2016).

[12] Abdullah, S.F., A.F.N.A.Rahman, and Z. A. Abas. "Classification Of Gender By Using Fingerprint Ridge Density In Northern Part Of Malaysia." Journal of Engineering and Applied Sciences.10 (22) (2015).

[13] Verma M, Agarwal S. Fingerprint based male-female classification. Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08. Springer Berlin Heidelberg. 2009; 1:251–7.

[14] Gupta S, Rao AP. Fingerprint based gender classification using Discrete Wavelet Transform and Artificial Neural Network. International Journal of Computer Science and Mobile Computing. 2014 Apr: 3(4):1289–96.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE I. ACCURACY AND CONFUSION MATRIX OR DIFFERENT CLASSIFIER IN DIFFERENT TEST OPTION

<table>
<tr><th>Classifier</th><th>Test Option</th><th>Confusion Matrix</th><th>Accuracy</th></tr>
<tr><td>Multilayer Perceptron Neural Network</td><td>10-fold cross validation</td><td>139 4 | a = L 6 151 | b = P</td><td>96.62%</td></tr>
<tr><td></td><td>60% Train 40% Test</td><td>49 5 | a = L 0 66 | b = P</td><td>95.76%</td></tr>
<tr><td></td><td>70% Train 30% Test</td><td>38 0 | a = L 0 52 | b = P</td><td>100%</td></tr>
<tr><td></td><td>80% Train 20% Test</td><td>31 0 | a = L 1 28 | b = P</td><td>100%</td></tr>
<tr><td>k-Nearest Neighbors (1 Nearest Neighbors)</td><td>10-fold cross validation</td><td>136 7 | a = L 7 150 | b = P</td><td>95.27 %</td></tr>
<tr><td></td><td>60% Train 40% Test</td><td>51 3 | a = L 3 63 | b = P</td><td>94.91 %</td></tr>
<tr><td></td><td>70% Train 30% Test</td><td>36 3 | a = L 1 50 | b = P</td><td>95.50%</td></tr>
<tr><td></td><td>80% Train 20% Test</td><td>29 2 | a = L 0 29 | b = P</td><td>96.61 %</td></tr>
<tr><td>Bayes Net</td><td>10-fold cross validation</td><td>135 8 | a = L 3 154 | b = P</td><td>96.28 %</td></tr>
<tr><td></td><td>60% Train 40% Test</td><td>51 3 | a = L 0 66 | b = P</td><td>97.45 %</td></tr>
<tr><td></td><td>70% Train 30% Test</td><td>36 3 | a = L 0 51 | b = P</td><td>96.622 %</td></tr>
<tr><td></td><td>80% Train 20% Test</td><td>30 2 | a = L 0 28 | b = P</td><td>96.61 %</td></tr>
<tr><td>Support Vector Machine</td><td>10-fold cross validation</td><td>140 3 | a = L 6 151 | b = P</td><td>96.95 %</td></tr>
<tr><td></td><td>60% Train 40% Test</td><td>53 1 | a = L 3 63 | b = P</td><td>96.61 %</td></tr>
<tr><td></td><td>70% Train 30% Test</td><td>38 1 | a = L 3 48 | b = P</td><td>95.50 %</td></tr>
<tr><td></td><td>80% Train 20% Test</td><td>32 0 | a = L 1 27 | b = P</td><td>98.30 %</td></tr>
</table>

TABLE II. COMPARISON OF FINGERPRINT GENDER CLASSIFICATION ACCURACIES

<table>
<tr><th></th><th colspan="3">Ahmed Badawi et al.</th><th>Manish Verma et al.</th><th>Gnanasivam et al.</th><th>Gupta et al.</th><th>Rajesh et al.</th><th colspan="4">Proposed Method</th></tr>
<tr><td>Features used</td><td colspan="3">RTVTR, WLC, RC, RCAPT</td><td>RTVTR, RW, RD</td><td>DWT SVD</td><td>DWT</td><td>DWT</td><td colspan="4">RC,RD,WLC RTVTR</td></tr>
<tr><td>Classifier</td><td>FCM</td><td>LDA</td><td>NN</td><td>SVM</td><td>KNN</td><td>Back Propagation NN</td><td>GMM</td><td>Bayes Net</td><td>SVM</td><td>MLPNN</td><td>k-NN</td></tr>
<tr><td>Accuracy</td><td>56.47</td><td>84.52</td><td>87.64</td><td>88</td><td>88.28</td><td>91.45</td><td>92.67</td><td>96.28</td><td>96.95</td><td>96.62</td><td>95.27</td></tr>
</table>

https://sites.google.com/site/ijcsis/
ISSN 1947-5500