FINGER MOTION IN CLASSIFYING OFFLINE HANDWRITING PATTERNS

By:

YEOH SHEN HORNG

(Matric no: 120431)

Supervisor:

Dr. Loh Wei Ping

May 2017

This dissertation is submitted to
Universiti Sains Malaysia
As partial fulfillment of the requirement to graduate with honors degree in
BACHELOR OF ENGINEERING (MECHANICAL ENGINEERING)

School of Mechanical Engineering
Engineering Campus
Universiti Sains Malaysia

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed………………………………………………………… (**YEOH SHEN HORNG**)

Date: 7th June 2017

Statement 1

This journal is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by giving explicit references. Bibliography/ references are appended.

Signed………………………………………………………… (**YEOH SHEN HORNG**)

Date: 7th June 2017

Statement 2

I hereby give consent for my journal, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available outside organizations.

Signed………………………………………………………… (**YEOH SHEN HORNG**)

Date: 7th June 2017

Acknowledgement

I would like to express my deepest appreciation to all whose help, support and inspiration resulted in the accomplishment of this journal. A special gratitude goes to my Final Year Project (FYP) supervisor, Dr. Loh Wei Ping, whose contribution in stimulating suggestions and giving encouragement has helped me to complete this project successfully especially in journal writing. Besides her fruitful guidance, she provided the equipment required for the experimental data collection.

Further, I also like to extend acknowledgement with high appreciation the crucial role of the experiment participants for the data collection process. They provided me with the case study raw data for analysis. Without them, this project would not have been possibly completed.

Last but not least, I would like to give my appreciation for my fellow friends, who have help me in getting the participants for the data collection process.

Table of Contents

List of Tables

List of Figures

Nomenclature/ Symbols

GH = Gender and handedness	a = LN
MR = Male and right handed	b = LP
FR = Female and right handed	c = Norm
ML = Male and left handed	d = Slightly Inconsistent
FL = Female and left handed	e = Extremely Inconsistent

Abstrak

Pengecaman tulisan luar talian merujuk kepada keupayaan mesin untuk menerima dan mentafsir input tulisan tangan individu daripada imej ditangkap atau diimbas. Dalam kajian-kajian lepas, pengelasan tulisan tangan luar talian ditentukan berasaskan sepenuhnya pada corak tulisan itu. Setakat pengetahuan kami, tiada kajian yang cenderung meramalkan ciri tulisan berdasarkan pergerakan jari. Oleh itu, kajian ini mengaitkan pergerakan jari kepada corak tulisan tangan. Khususnya, objektif kajian ini termasuklah: (i) untuk menentukan sama ada sifat-sifat gerakan jari dapat membezakan corak tulisan tangan. (ii) mengklasifikasi corak tulisan tangan dengan kecondongan ayat berdasarkan pergerakan jari yang berbeza. (iii) untuk menyiasat penyata peraturan-hujah antara gerakan jari dan kecondongan tulisan tangan. Kajian ini melibatkan 30 subjek dan pengekstrakan ciri-ciri corak tulisan tangan dan video pergerakan jari ketika penulisan. Data ini akan melalui tiga peringkat analisis perlombongan data: prapemprosesan data, pengelasan dan interpretasi data. Data yang telah dipraproses akan dikelaskan dengan algoritma J48. Ketepatan ramalan pengelasan selepas dilatih boleh mencapai sehinga 98%. Hasil kajian memperlihatkan bahawa sudut ibu jari memainkan peranan utama dalam pengelasan kecondongan ayat Inggeris.

Abstract

Offline handwriting recognition refers to the ability of a machine to receive and interpret a previous individual-made handwritten input from a photographed or scanned image. In previous studies, the offline handwriting classification is determined solely based on the handwriting patterns. To the best of our knowledge, no studies were found to predict the English words inclination based on the finger motions. Therefore, this study aims to relate the finger movements to handwriting patterns. The specific objectives include: (i) to determine whether finger motion attributes can distinguish patterns of handwriting, (ii) classify handwriting patterns by sentence inclination based on different finger motion, (iii) to investigate the rule-reasoning statements between the finger motion and the handwriting inclinations. This study involves the features extractions from handwriting patterns of 30 subjects with recorded videos of finger movements during writings. Raw data undergo three stages of data mining analyses; data preprocessing, data classification and data interpretation. The preprocessed data is classified using the J48 tree algorithm. The correctly classified accuracy prediction after trained could achieve up to 98 %, Finding revealed that the angle of thumbs plays a significant role in classification of the inclination of the English sentence.

1.0 Introduction

Handwriting recognition is the ability of a computer to receive and interpret intelligible handwritten input from various sources on paper documents, photographs and even online touch screen devices. People use handwriting very often in daily life. Basically, the handwriting recognition can be categorized into offline recognition and online recognition. The former is based on the written text on paper material while the latter is based on the dynamics of writing from touch screen devices [1]. Existing studies in handwriting recognition mostly focused on text recognition based on the handwriting pattern.

Data mining is defined as the extraction of data from large-scale data. Thus, it is possible to show the relationship between the data for predictions. Data mining also helps in decision making processes in institutions with the development of new strategies after it is analyzed using statistical methods. In other words, data mining is the discovery of the data in database.

Offline handwritings were very commonly studied and used for soft-biometric identification and forgery detection. Offline handwritings are usually processed and analyzed from the image processing algorithm perspectives. Ideally, the fingers' motion while writing also give impacts to an individual's handwriting patterns. However, the prediction of handwriting patterns based on finger movements has not been reported. In other words, the offline handwriting data mining application is mostly done on the handwritten text alone. Abundance of handwriting experimental data are solely recorded with very little informative discovery in return. On the other hand, data mining concept for predicting collected handwriting samples based on finger movement is something new. To the best of our knowledge, no work has investigated the relationships between finger movements and the inclination of the handwriting.

The lacking from the previous studies motivates the goal of this project i.e. (i) to determine whether finger motion attributes can distinguish patterns of handwriting, (ii) classify handwriting patterns by sentence inclination based on different finger motion, (iii) to investigate the rule-reasoning statements between the finger motion and the handwriting inclinations.

In this project, the finger motions in writing English sentence is experimentally captured on two bases: finger motion video and handwriting images. The collected data will be employed for data mining analysis in three stages aided by Waikato Environment for Knowledge Analysis (WEKA) software. The finger motions video is initially captured with a sports camera (16MP camera) and the handwriting samples are recorded. Data preprocessing works are required to remove irrelevant information from the study data. Emphasis are on the angles of fingers during handwriting process and angle of the pen. Preprocessed data undergo data classification analyses, of which the extracted features are grouped by similar patterns into classes and later subjected to rule-reasoning analysis to predict the handwriting patterns.

2.0 Literature Review

Different methods were previously reported in recognizing handwriting patterns. Among the past studies, Assaleh, et al. [1] had proposed successful handwritten Arabic alphabet recognition by tracking the hand motion. In their study, a camera was used to capture the video of hand motion and projected into accumulated set of images. The resulting feature is later classified on K-Nearest Neighbor (KNN) algorithm. The technique was proven superior to the results obtained by the classical Hidden Markov models (HMW)-based scheme. Udhan et al. [2] proposed a system that utilized Artificial Neural Network (ANN) for alphabets recognition from the finger gesture path by attaching a colored sticker on a finger. As the finger is moved while writing an alphabet, the trajectory of the finger is recorded and is processed to obtain the handwritten pattern. The alphabet is often recognized by the feature point matrix. An online recognition system that used leap motion controller was suggested by Vikram, et al. [3]. The leap motion controller was used to capture the 3D finger movement which will be processed by using dynamic time warping (DTW) algorithm to recognize the handwriting patterns in real time. Putra, et al. [4] proposed a method to improve the recognition accuracy without relying on the normalization technique. The researcher created the handwritten characters into graphs with string representation based on structural approach.

The analysis of handwriting pattern had successfully gained attention in fields such as forgery detection and identification via handwriting recognition. Verma, et al. [5] proposed a method to detect offline signature forgery using the global and geometric feature. The author approached the problem in two steps. Initially, a set of signatures were obtained from the subject and fed to the system. These signatures are pre-processed and then the pre-processed images were used to extract relevant geometric features that can distinguish signatures of different persons. These are used to train the handwriting recognition system. The mean value of these features was computed to test the signature image by feeding them to the system. The tested image was pre-processed for extraction of geometric features. These values were compared with the mean features that were used to train the system. The authors calculated the Euclidean distance using the mean and standard deviation of all features. The maximum Euclidean and the minimum Euclidean values of the training sample were used to set the acceptance range.

Meanwhile, a robust prediction of writer's gender, age range and handedness had been considered using SVM classifier and features such as pixel density, pixel distribution and gradient local binary patterns. Besides, a combination method that uses Fuzzy MIN and MAX rules combined membership degree which boost the accuracy of the prediction system [7].

In offline handwriting analysis, Murat and Seher [8] detected the gender of the writer based on the handwriting analysis. The gender detection of a writer is performed by utilizing 133 study attributes. The data analysis used J48 decision tree and ID3 for gender detection.

3.0 Methodology

In this study, data mining is applied to the experimental case study data on finger motion and the offline handwriting. The process can be categorized into four main stages (Figure 1).

Each and every stage in the flow will be detailed in the following subsections.

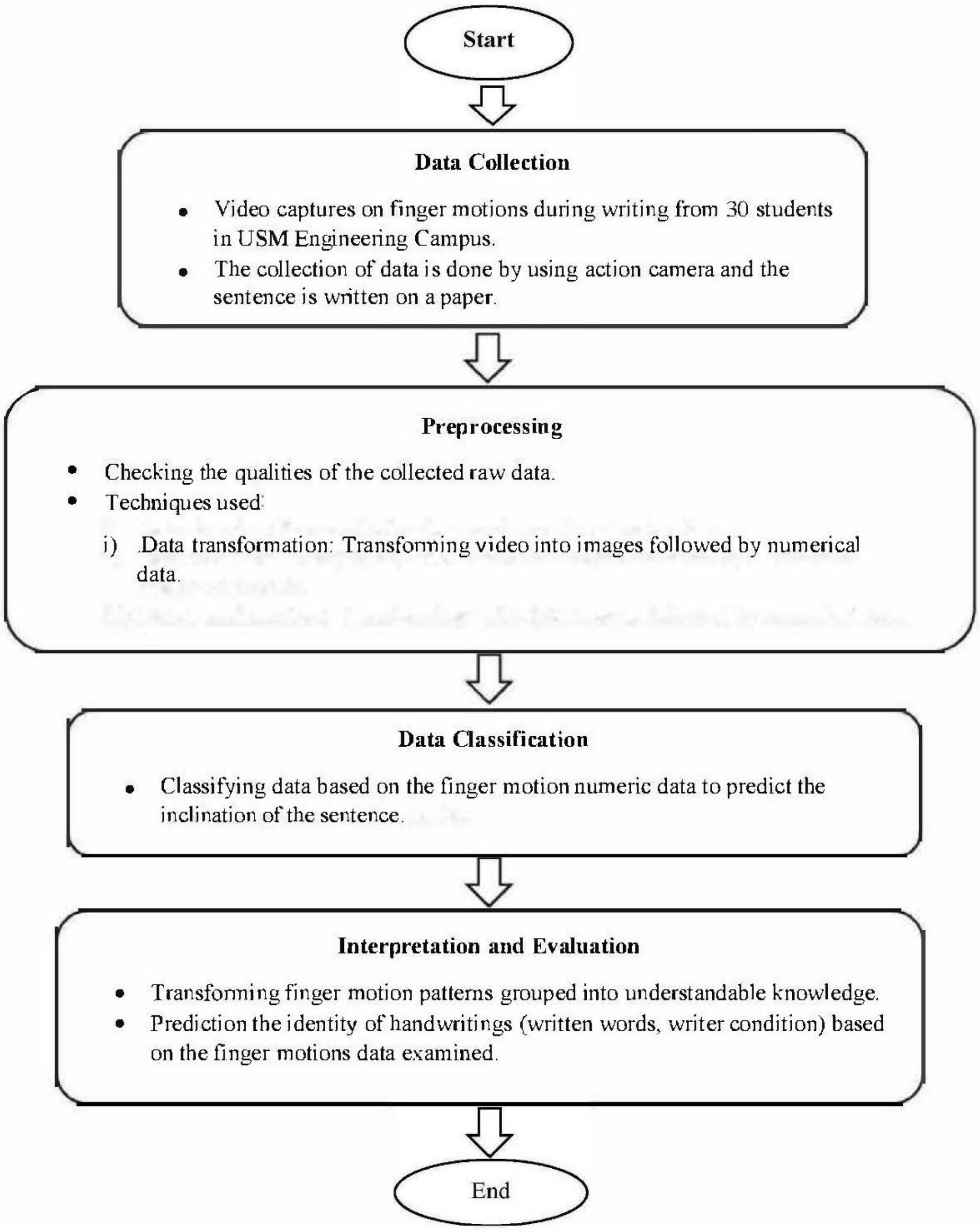


Figure 1: Flow of the project

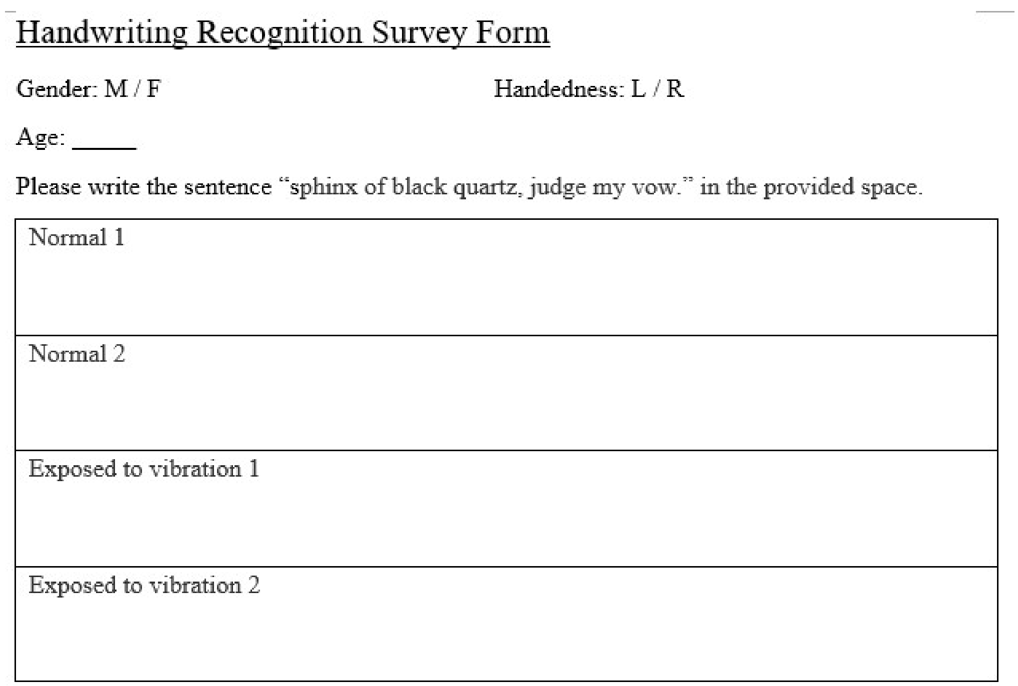
3.1 Data Collection

The first stage of the project involves experimental data collection. The data collection process was carried out in BioMotion Capture Laboratory at School of Mechanical Engineering involving

30 students from USM Engineering Campus as the study subjects. The materials and equipment used include a sports camera, tripod stand, table, chair, gel ink pen and survey form. The camera used throughout the experiment is a 16 MP sports camera that could record full HD videos at 60 fps. The entire experimental setup is shown in Figure 2. The data collection process consists of two sessions; the video recording and the offline handwriting patterns.

Figure 2: Different views of experimental setup, a) top view, b) front view, c) side view

The camera is positioned in front of the writing paper, inclined at 49° to the vertical. Several conditions were set: positions of all equipment throughout the data collection process being fixed and the same pen used throughout the data collection process. Each participant was required to write the phrase "sphinx of black quartz, judge my vow." under camera capture during the writing process on a provided survey form (Figure 3).



Handwriting Recognition Survey Form

Gender: M / F Handedness: L / R

Age: _____

Please write the sentence "sphinx of black quartz, judge my vow." in the provided space.

Normal 1
Normal 2
Exposed to vibration 1
Exposed to vibration 2

Figure 3: Survey form that is used in data collection process

3.2 Data Description

The raw data were recorded into 14 attributes: pen angle 1, pen angle 2, thumb1, thumb2, middle1, middle2, index1, index2, time1, time2, L1, L2, GH and inclination, 484 instances, 60 missing values. These attributes measurement scales include ordinal, numeric and nominal attributes. Details of data description are summarized in Table 1.

Table 1: Detail description of attributes in dataset

Attributes	Description	Scale type	Data range

time*i*	The time at which the image frame is extracted from the videos.	Numeric	0 - 25 (s)
thumb*i*	The angle of the thumb relative to the horizontal axis.	Numeric	-180 – 180 (°)
index*i*	The angle of index finger relative to the horizontal axis.	Numeric	-180 – 180 (°)
middle*i*	The angle of middle finger relative to the horizontal axis.	Numeric	-180 – 180 (°)
pen*i*	The angle of pen relative to the horizontal axis	Numeric	-180 – 180 (°)
L*i*	The total length of the written sentences.	Numeric	0 – 13 (cm)
GH	Gender and handedness of the participant.	Nominal	{MR, ML, FR, FL}
Inclination	The angle of inclination of the whole written phrase.	Ordinal	{LP ($0.5° \leq x \leq 2°$), LN ($-2° \leq x \leq -0.5°$), Norm ($-0.5° < x < 0.5°$), Slightly inconsistent, Extremely inconsistent}
i	Number of repeats	Numeric	i =1, 2

3.3 Data Preprocessing

The real-world database is highly susceptible to noise, missing and inconsistent data. Low quality data may lead to low quality data mining results. Thus, data preprocessing is required to improve the data quality so that qualitative mining results can be obtained. The preprocessing techniques can be applied by selecting a suitable filter in WEKA to filter out the undesirable data. However, preprocessing technique is not required in this study as the extracted data is clean. In addition, the preprocessing technique used for filling the missing values is not necessary as this would result in inaccurate data (considering each person write in different ways). Data integration merges data from multiple sources into a coherent data store such as data warehouse. Data reduction reduce data size by aggregating, eliminating redundant features or clustering. Basically, this technique is used to reduce numerous amount of data into meaningful parts. Data transformation converts a set of data values from one data format of a source data system into another data format. In this study, the data transformation involves the conversion of image into numeric values. The data preprocessing levels is shown in Figure 4.

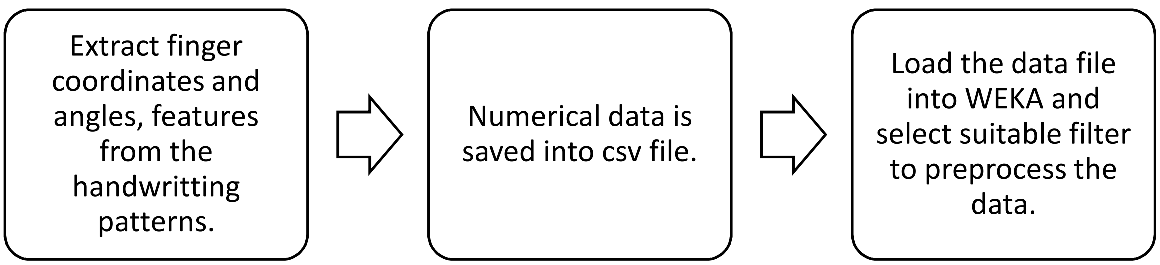


Figure 4: Preprocessing levels

The numerical data is extracted from videos and the handwriting patterns are sorted into attributes and instances. The study attributes are designed from the angles of fingers and pen extracted from video captured frames and the observed handwriting features (i.e. inclination and length) as detailed in the following subsections. Upon extraction of numerical data, the data is saved into .csv file readable by Waikato Environment for Knowledge Analysis (WEKA) machine learning software (preprocessing and classification analyses. In WEKA, the raw data is preprocessed with the available preprocessing tool known as filter.

3.3.1 Handwriting video-framed image feature extraction

The features extracted from the handwriting video-framed images include the angles of thumb, index finger, middle finger and pen using Adobe Photoshop CC 2015. In Adobe Photoshop, the ruler tool is used for measurement. For instance, the angle of the fingers relative to the horizontal axis is displayed on option bar as shown in Figure 5.

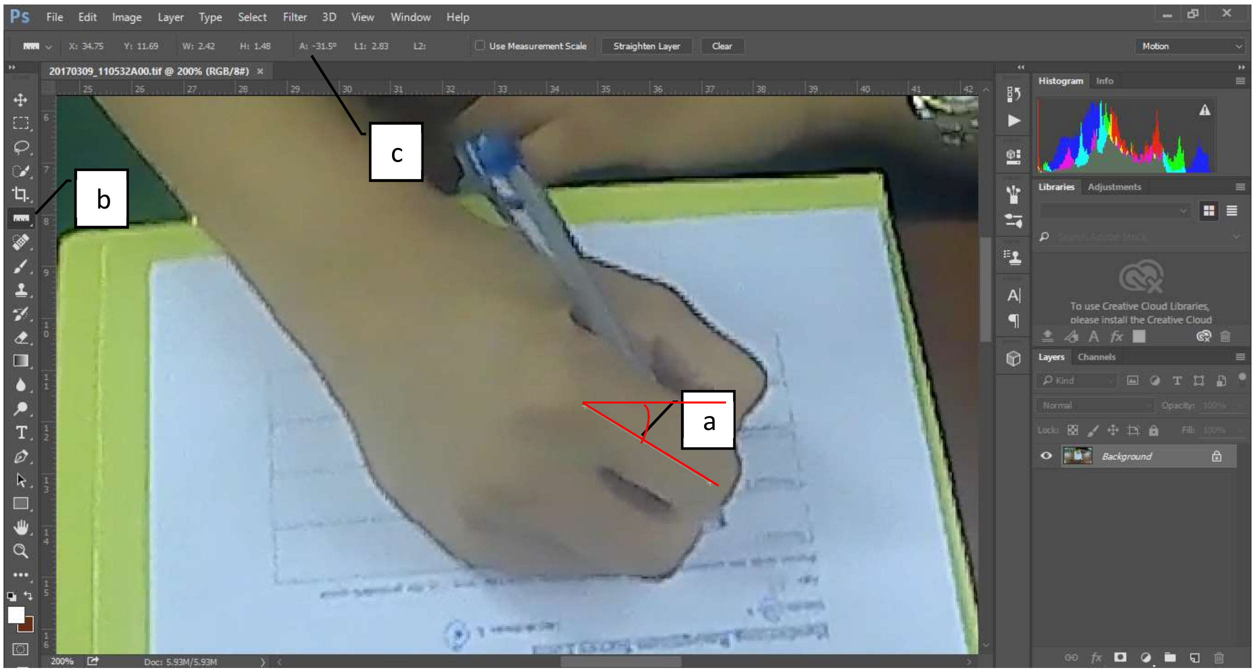


Figure 5: The tools used for measurements, a) Line drawn with ruler tool, b) ruler tool in Adobe Photoshop, c) angle of the drawn line with respect to horizontal axis

3.3.2 Handwriting features

The features measurements of handwriting are measured on simple metric ruler and protractor. These features include the total length and the inclination of the sentence. The sample of handwriting with the extracted features is shown in Figure 6.

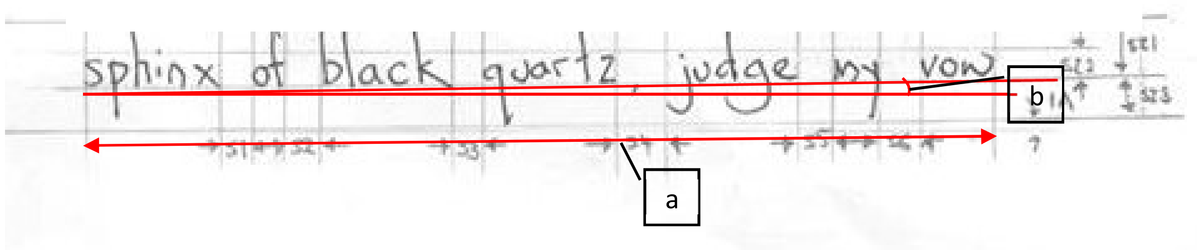


Figure 6: Handwriting patterns feature extraction, a) total length of sentence, b) sentence inclination

3.4 Data Classification

Data classification extracts models describing important data classes, to predict the class labels. Data classification involves two steps: learning step and classification step. In learning step, a classifier is built to describe a predetermined set of data classes or concepts, where a classification algorithm builds the classifier by learning from a training set as their corresponding class labels. The class label attribute is categorical as each value serves as a category or class. In this study, the learning step also known as supervised learning as each training tuple is provided with the corresponding class label. In the second step, the classifier that is built in previous step is used for classification and a test set is provided to test the accuracy of the built classifier. It is imperative to avoid the testing of the built classifier with training set. This is because the classifier will tend to overfit the data (i.e., during learning it may incorporate some particular anomalies of the training data that are not present in the general data set overall), resulting in optimistic predictive accuracy. WEKA is built-in with several groups of classifiers: Bayes, Clojure, Functions, Lazy, Meta, Mi, Misc, Pyscript, Rules, Scripting, Sklearn, Timeseries and Trees classifiers. Each of these classifiers has its own algorithms for data classification. For instance, J48, CDT and ID3 are among the algorithms under the decision trees classifier and these three algorithms work differently from each other.

This project involves a case study dataset deployed on 10 folds cross validation for classification as per the default setting in WEKA. Cross validation (a.k.a. rotation estimation) partitions dataset as training and testing. Thus, this process is repeated for ten times (i.e. 10 folds) by different partitioning the datasets for training and testing and the results are averaged.

J48 tree classifier from tree classifier is used for classification. Decision tree induction is the learning of decision trees from class-labeled training tuples. A decision tree is a flowchart-like tree structure (Figure 7) which consist of root node (starting node), interior nodes (non-leaf nodes) and leaf nodes (terminal node) that hold a class label. These nodes are connected by branches that represent the outcome of the test. Given a tuple, X, for which the associated class label is unknown, the attribute values of the tuple are tested against the decision tree. A path is traced from the root to a leaf node, which holds the class prediction for that tuple forming sets of rule-reasoning statements. This study choses J48 for several reasons. Firstly, it is simple and appropriate for exploratory knowledge discovery. Besides, decision trees can handle multidimensional data as in the case study (484 rows × 14 columns).

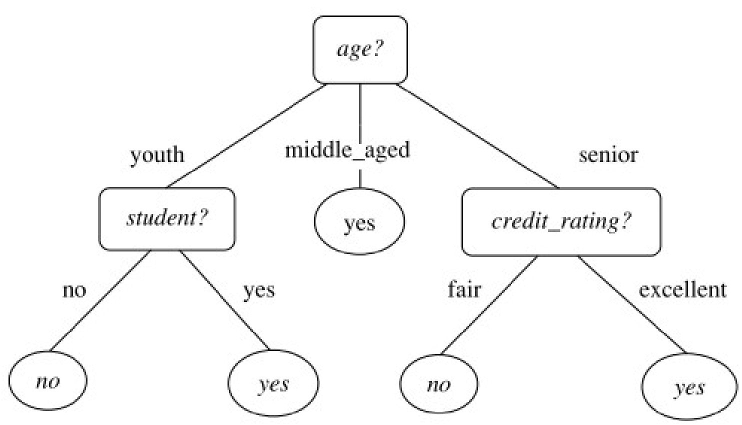


Figure 7: Example of decision tree classification [13]

3.5 Data Interpretation and Evaluation

The summary results in WEKA involves the classification prediction accuracy, kappa statistic, root mean square error and mean absolute error. The percentage of the correctly classified instances (or prediction accuracy) determines how accurate the classifier could predict the instances correctly into actual class after being trained.

Kappa statistic compares an observed accuracy with an expected accuracy. It takes into account random chance (i.e. agreement with a random classifier), which generally means it is less misleading than simply using the accuracy in predicting the instances correctly as metric. Observed accuracy is the number of instances that were classified correctly throughout the entire confusion matrix while expected accuracy is defined as the accuracy of any random classifier would be expected to achieve based on the confusion matrix. The kappa statistic, observed accuracy and expected accuracy are computed as shown from equations (1) to (3).

$$kappa = \frac{observed\ accuracy - expected\ accuracy}{1 - expected\ accuracy} \tag{1}$$

$$observed\ accuracy = \frac{TP + TN}{Total\ number\ of\ instances} \tag{2}$$

$$expected\ accuracy = \frac{(TN + FP) * (TN + FN) + (FN + TP) * (FP + TP)}{Total * Total} \tag{3}$$

where TP = true positive, TN = true negative, FP = false positive, FN = false negative

The classification analyses in this study were mainly discussed by the percentage of correctly classified instances. This is because percentage accuracy of the correctly classified instances is easily understood and the percentage count could conveniently convert to compute by number of instances to select the most suitable classifier.

4.0 Results

On data preprocessing stage, raw data were transformed into 14 attributes, 484 instances and 60 missing values. The missing values are maintained in its original form without any preprocessing work to filter or impute them. This is due to the imputation on the missing values will change the nature of original data and therefore may yield an inaccurate result. Classification analyses performed using J48 tree algorithm on the study data resulted in 98.1 % accuracy as shown in Figure 8. A further visual investigation on tree diagram shows the size of tree 69 while the size of leaves was 39 (Figure 8). From the results of J48 algorithm, the list of rules created with J48 is presented in Appendix B.

```
Number of Leaves  :     39

Size of the tree :      69


Time taken to build model: 0.12 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.05 seconds

=== Summary ===

Correctly Classified Instances         475               98.1405 %
Incorrectly Classified Instances         9                1.8595 %
Kappa statistic                          0.9723
Mean absolute error                      0.0176
Root mean squared error                  0.0838
Relative absolute error                  6.5106 %
Root relative squared error             22.8422 %
Total Number of Instances              484
```

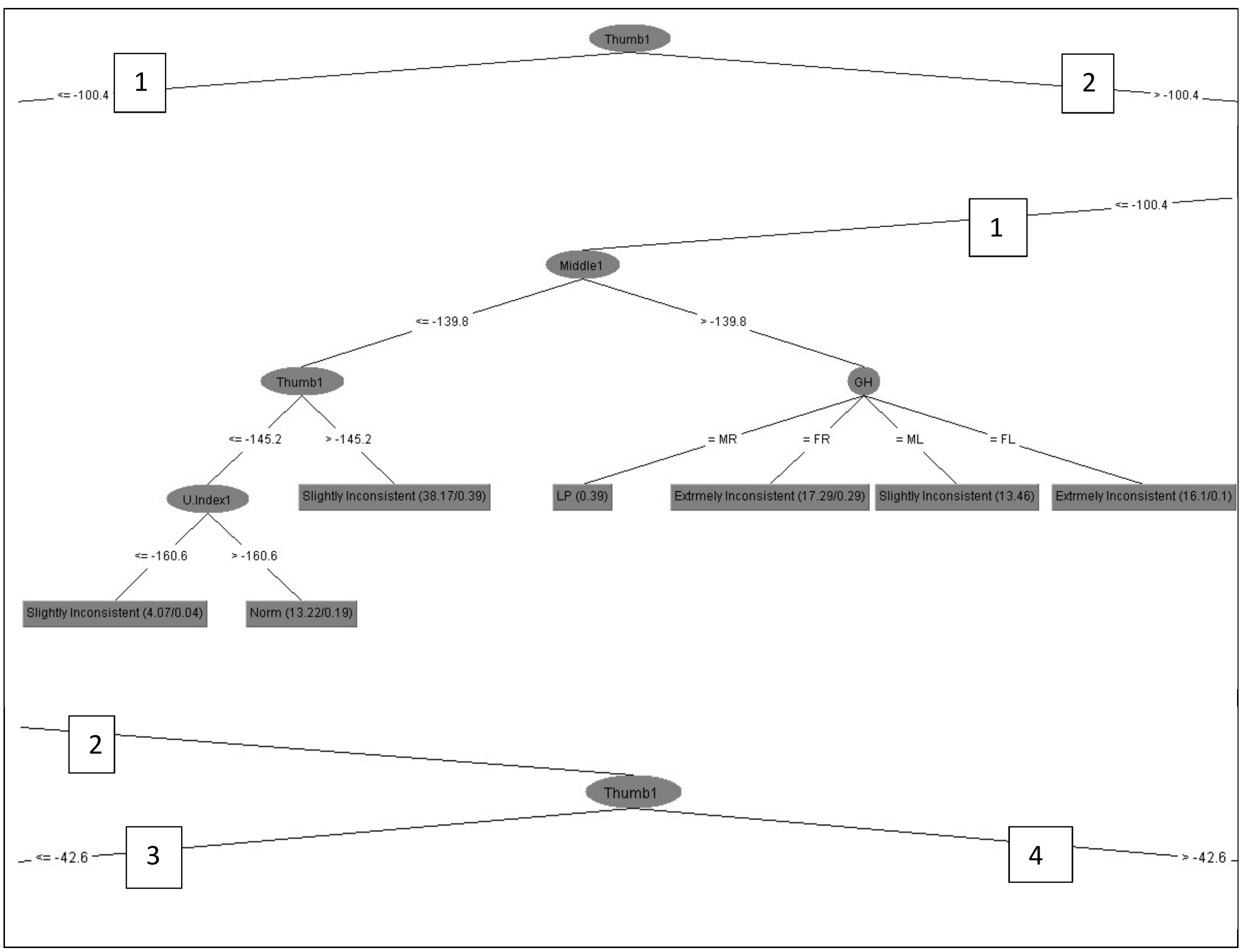
Figure 8: Summary of the results of J48 in WEKA

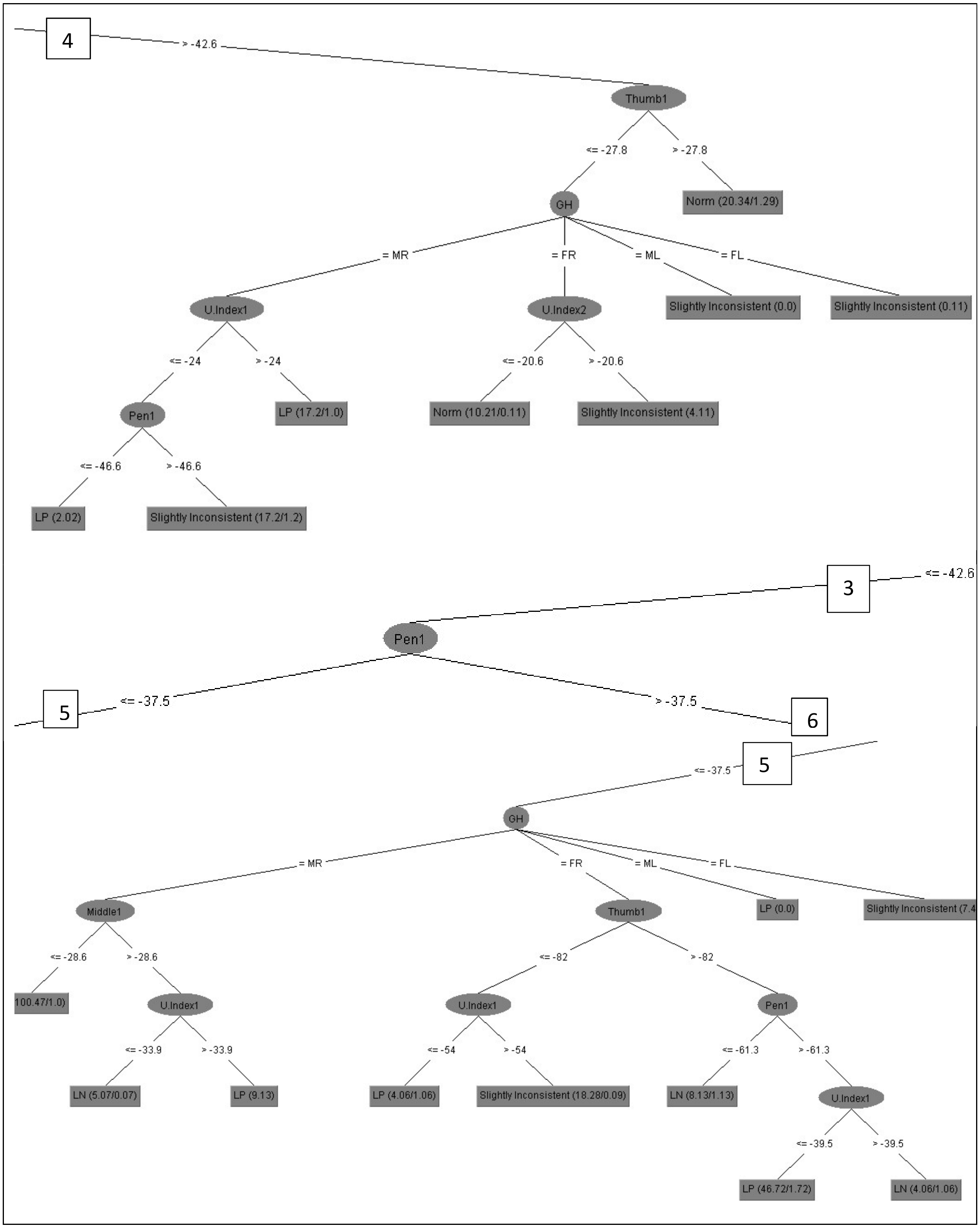
Based on the J48 algorithm results, the kappa statistic, mean absolute error and root mean square error were found to be 0.9723, 0.0176 and 0.0838 respectively. The low values of mean absolute and root mean square errors imply the high accuracy of the classifier in classifying the instances. The kappa statistic (range from 0 to 1) of 0.9723 shows a very high agreement between the observed accuracy and the expected accuracy. The entire decision tree constructed is shown in

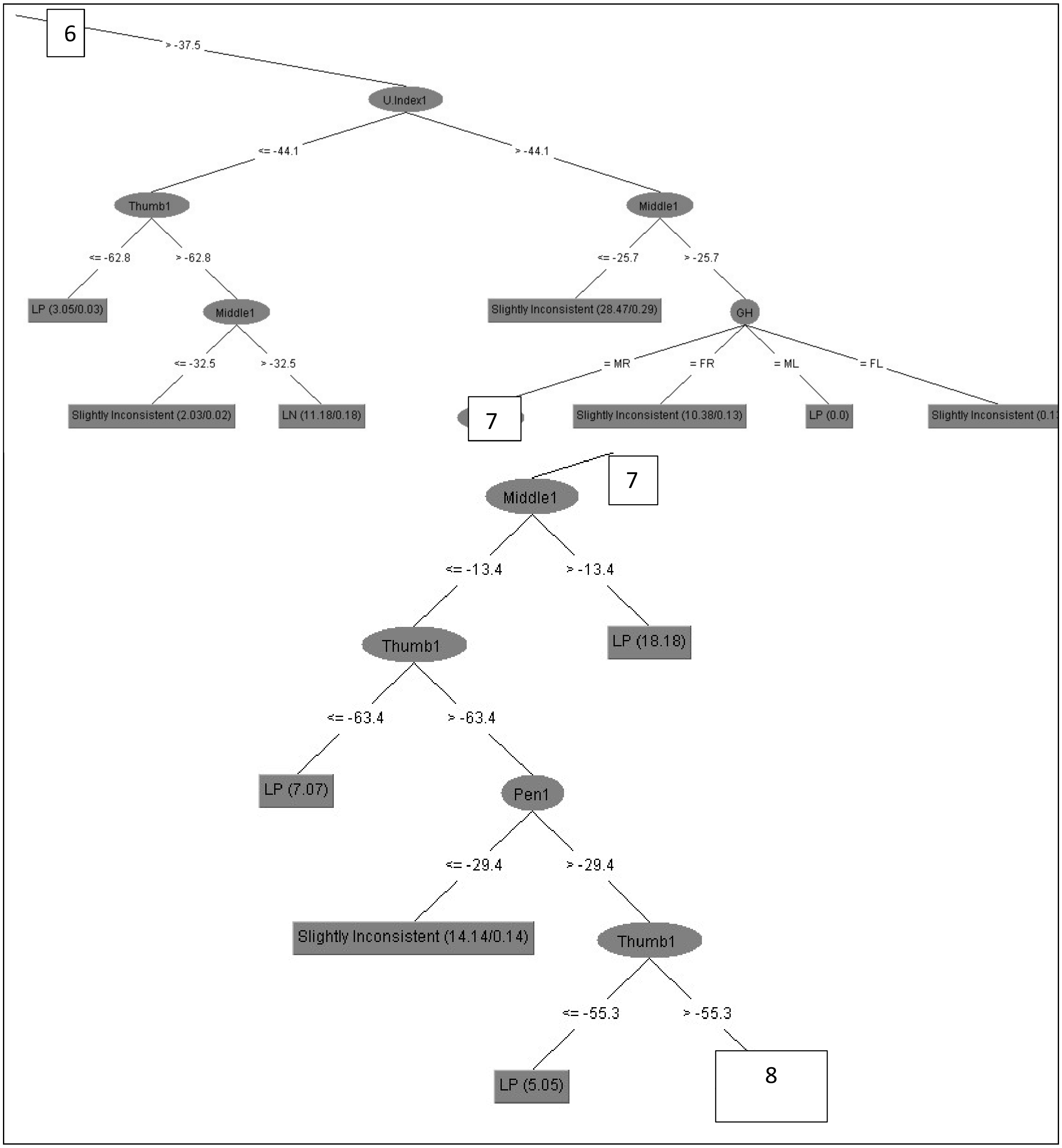
Figure 9. Based on the figure, it is found that angle of thumb plays a major role in the classification, followed by angle of middle finger and the angle of pen. In addition, it is noticed that attributes L1 and L2 do not appearing in the decision tree. It shows that the classification can be done solely based on the finger movement attributes.

The confusion matrix from the J48 algorithm results is shown in Legend | a = LN, b = LP, c = Norm, d = Slightly inconsistent, e = Extremely inconsistent

Figure 10. Based on the figure, the green color shows the number of correctly classified instances while the red one is the number of misclassified instances. There are total of 475 out of 484 instances are correctly classified while only while only 9 instances are misclassified.




4
> -42.6
Thumb1
<= -27.8
> -27.8
GH
Norm (20.34/1.29)
= MR
= FR
= ML
= FL
U.Index1
U.Index2
Slightly Inconsistent (0.0)
Slightly Inconsistent (0.11)
<= -24
> -24
<= -20.6
> -20.6
Pen1
LP (17.2/1.0)
Norm (10.21/0.11)
Slightly Inconsistent (4.11)
<= -46.6
> -46.6
LP (2.02)
Slightly Inconsistent (17.2/1.2)
3
<= -42.6
Pen1
5
<= -37.5
> -37.5
6
5
<= -37.5
GH
= MR
= FR
= ML
= FL
Middle1
Thumb1
LP (0.0)
Slightly Inconsistent (7.4
<= -28.6
> -28.6
<= -82
> -82
100.47/1.0)
U.Index1
U.Index1
Pen1
<= -33.9
> -33.9
<= -54
> -54
<= -61.3
> -61.3
LN (5.07/0.07)
LP (9.13)
LP (4.06/1.06)
Slightly Inconsistent (18.28/0.09)
LN (8.13/1.13)
U.Index1
<= -39.5
> -39.5
LP (46.72/1.72)
LN (4.06/1.06)


6
> -37.5
U.Index1
<= -44.1
> -44.1
Thumb1
Middle1
<= -62.8
> -62.8
<= -25.7
> -25.7
LP (3.05/0.03)
Middle1
Slightly Inconsistent (28.47/0.29)
GH
<= -32.5
> -32.5
= MR
= FR
= ML
= FL
Slightly Inconsistent (2.03/0.02)
LN (11.18/0.18)
7
Slightly Inconsistent (10.38/0.13)
LP (0.0)
7
Middle1
<= -13.4
> -13.4
Thumb1
LP (18.18)
<= -63.4
> -63.4
LP (7.07)
Pen1
<= -29.4
> -29.4
Slightly Inconsistent (14.14/0.14)
Thumb1
<= -55.3
> -55.3
LP (5.05)
8

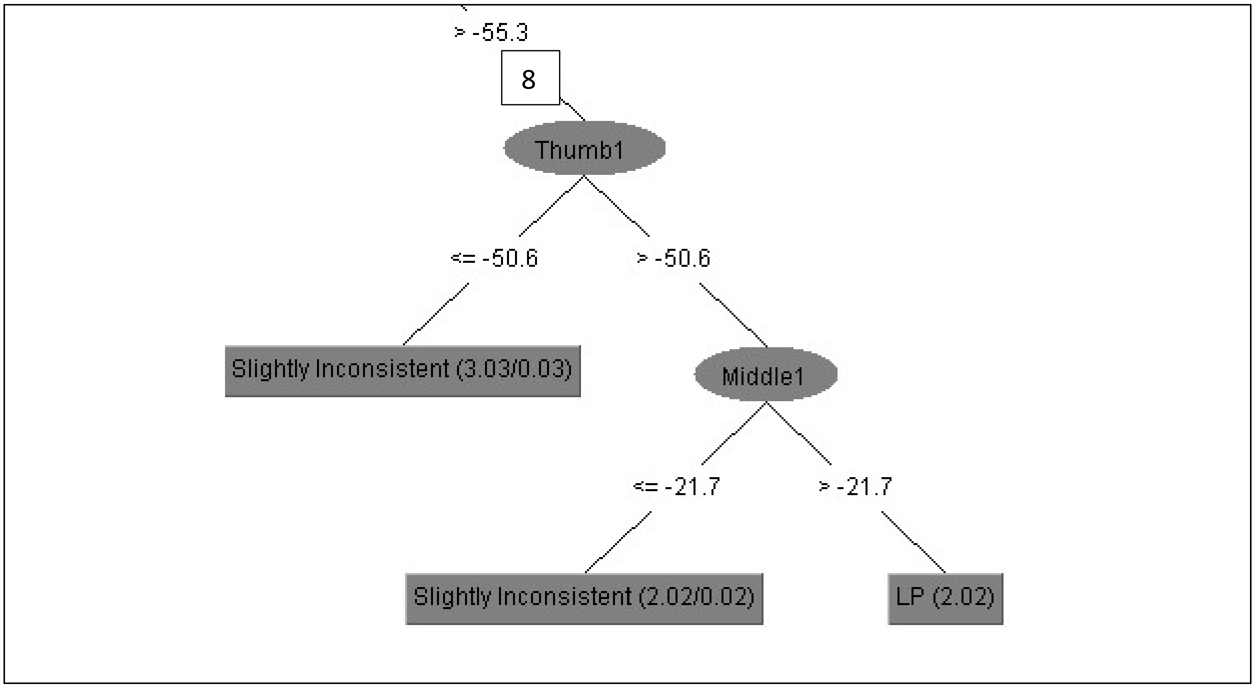


Figure 9: Tree visualization of J48 algorithm

a	26	3	0	0	0	89.66%
b	0	212	0	1	0	99.53%
c	1	0	42	1	0	95.45%
d	1	1	1	162	0	98.18%
e	0	0	0	0	33	100.00%
	92.86%	98.15%	97.67%	98.78%	100.00%	98.14%
	a	b	c	d	e	

Legend | a = LN, b = LP, c = Norm, d = Slightly inconsistent, e = Extremely inconsistent

Figure 10: Confusion matrix from J48 decision tree result

5.0 Discussion

The raw experimental data were transformed into 14 attributes (i.e. pen1, pen2, thumb1, thumb2, index1, index2, middle1, middle2, time1, time2, L1, L2, GH and inclination) of 484 instances as discussed in Section 3.2. The dataset being classified on different groups of classifiers (LP, LN, Norm, Slightly inconsistent and Extremely inconsistent) were targeted for the purpose of checking on most appropriate classifiers to suit the case study data. The accuracies from each classifier were reflected in Table 2. With reference to Table 2, it is shown that the classifiers in tree classifier group shows the most consistent results in all embedded algorithms ranging from 87.8 % to 99.8 % with merely 12% variation as shown in Table 2. Figure 11 indicates that not all gender and handedness reflect all the five classes groupings (i.e. LP, LN, Norm, Slightly inconsistent and Extremely inconsistent). Based on Figure 11, the classifier shows 100% accuracy in estimating the FL while the accuracy in classifying the data as MR, ML and FR are 98.86%, 97.83% and 96.85% respectively. If MR were to be extracted and classified using J48, the algorithm could classify data perfectly into classes a, c and d. Meanwhile, the accuracy in classifying the data into class b is 98.17%. On the other view, when ML is extracted, the accuracy in classifying the data into classes c and d are 96.97% and 100% respectively. From the analysis, the obvious finding is that the incorrect classification into MR, ML and FR is much likely affected by the inconsistencies in the overall data. This is because around 40% of the data were of inconsistent inclination. Such inconsistent inclination can be further segregated into slightly inconsistent and extremely inconsistent, with slightly inconsistent class contributes the most (70%) of the inconsistent class itself.

Table 2: Accuracy of each classifier from different group of classifiers

Classifier group	**Classifier**	**Accuracy**
Bayes	BayesNet	87.3967
	NaiveBayes	65.9091
	NaiveBayesMultinomial Text	44.0083
	NaiveBayesUpdateable	65.9091
Clojure	Clojure	44.0083
Lazy	IB1	73.3471
	IBk	73.3471
	IBkLG	73.3471
	KStar	100
Tree	BFTree	98.3471
	CDT	92.562
	FT	95.2479
	J48	98.1405
	J48Consolidated	83.0579
	J48graft	98.1405
	LADTree	87.8099
	LMT	100
	NBTree	96.281
	RandomForest	99.7934
	RandomTree	99.5868
	REPTree	94.0083
	SimpleCart	98.9669
Rules	ConjunctiveRule	60.124
	Decision Table	85.9504
	DTNB	88.843
	FURIA	97.1074
	JRip	92.562
	MODLEM	86.5702
	NNge	69.6281
	OneR	76.6529
	PART	97.1074
	Ridor	95.6612
	ZeroR	44.0083

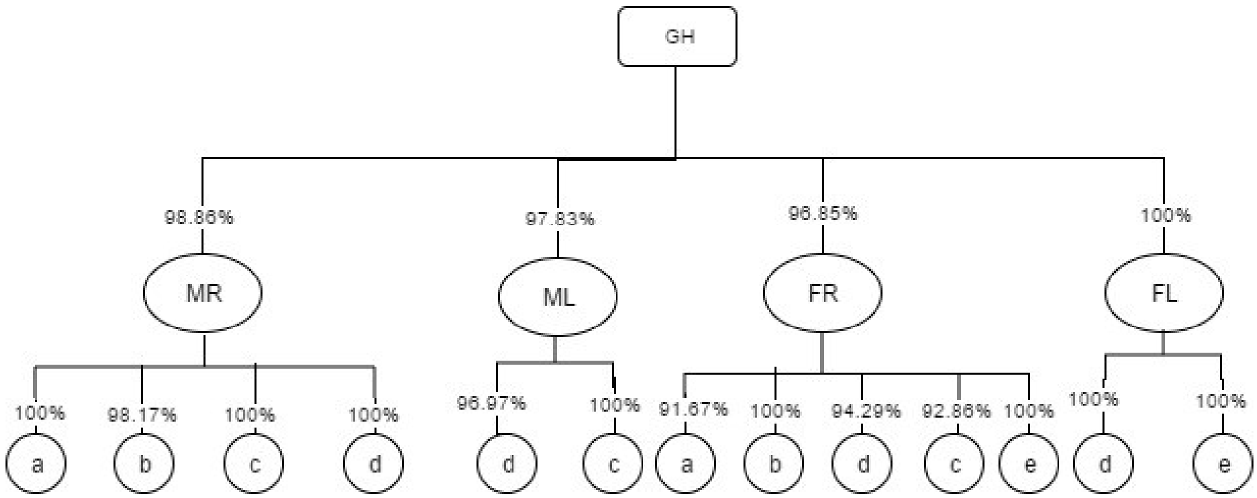


Figure 11: Tree diagram of classification from J48

6.0 Conclusion

The features extracted from finger motions include angles of thumb, middle, index fingers and pen were used to relate to inclination patterns found in offline handwritings. The study was performed to meet three objectives: (i) to determine whether finger motion attributes can distinguish patterns of handwriting, (ii) classify handwriting patterns by sentence inclination based on different finger motion, (iii) to investigate the rule-reasoning statements between the finger motion and the handwriting inclinations. Offline handwritings are commonly categorized from similar sample image patterns but has not been understood from fingers motion detections.

As such, dataset was collected from video finger motion during offline handwriting along with images of handwritings samples. Raw data were transformed into numeric information consisting of 14 attributes, 484 instances and 60 missing values. The dataset is being classified using several classifiers include Bayes, Clojure, Functions, Lazy, Meta, Mi, Misc, Pyscript, Rules, Scripting, Sklearn, Timeseries and Trees classifiers. Data undergo two levels of analyses using data mining approach with the aid of WEKA tool. The results from classification analyses showed that Tree classifier algorithms are most consistent in terms of correct classification accuracies. J48 algorithm is being selected from Tree classifier for further analysis. The reason was that this algorithm is found capable to represent data classes by decision tree structure on 98.1405% accurate prediction. The results from the tree diagram showed that the attributes L1 and L2 were missing from the tree. The main findings from this study is that the classification of the offline handwriting inclination can be solely determined from finger motion attributes without the need of the handwriting features such as length of the sentences.

Further studies should focus on single word prediction based on the finger motion attributes. The data mining on the word prediction based on finger motion attributes is not yet reported so far. Thus, there is much to be explored from this aspect.

References

[1] K. Assaleh, T. Shanableh, and H. Hajjaj, "Recognition of handwritten Arabic alphabet via hand motion tracking," J. Franklin Inst., vol. 346, no. 2, pp. 175–189, 2009.

[2] S. Udhan and P. P. Futane, "Hand Motion Tracking for Alphabet Recognition using ANN," vol. 3, no. 1, pp. 289–293, 2013.

[3] S. Vikram, L. Li, and S. Russell, "Handwriting and Gestures in the Air, Recognizing on the Fly," CHI 2013 Ext. Abstr., vol. ACM 978-1-, 2013.

[4] M. E. W. Putra and I. S. Suwardi, "Structural Off-line Handwriting Character Recognition Using Approximate Subgraph Matching and Levenshtein Distance," Procedia Comput. Sci., vol. 59, pp. 340–349, 2015.

[5] A. C. Verma, D. Saha, and H. Saikia, "Forgery Detection in Offline Handwritten Signature Using Global and Geometric," Int. J. Comput. Electron. Res., vol. 2, no. 2, pp. 182–188, 2013.

[6] A. Piyush Shanker and A. N. Rajagopalan, "Off-line signature verification using DTW," Pattern Recognit. Lett., vol. 28, no. 12, pp. 1407–1414, 2007.

[7] N. Bouadjenek, H. Nemmour, and Y. Chibani, "Robust soft-biometrics prediction from off-line handwriting analysis," Appl. Soft Comput. J., vol. 46, pp. 980–990, 2016.

[8] M. Topaloglu and S. Ekmekci, "Gender detection and identifying one's handwriting with handwriting analysis," Expert Syst. Appl., vol. 79, pp. 236–243, 2017.

[9] T. Yamasaki and T. Hattori, "A new data tablet system for handwriting characters and drawing based on the image processing," 1996 IEEE Int. Conf. Syst. Man Cybern. Inf. Intell. Syst. (Cat. No.96CH35929), vol. 1, pp. 428–431, 1996.

[10] P. Braido and X. Zhang, "Quantitative analysis of finger motion coordination in hand manipulative and gestic acts," Hum. Mov. Sci., vol. 22, no. 6, pp. 661–678, 2004.

[11] M. Liwicki, S. Ebert, and A. Dengel, "Bridging the gap between handwriting recognition and knowledge management," Pattern Recognit. Lett., vol. 35, no. 1, pp. 204–213, 2014.