International Journal of Advanced Science and Technology
Vol. 29, No. 7s, (2020), pp. 638-649

Statistical Evaluation on the Performance of Dyslexia Risk Screening System Based Fuzzy Logic and WEKA

Ng Li Mun[1] and Nur Anida Jumadi[2]

[1]Department of Electronic Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn, Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia

[2]Department of Electronic Engineering, Faculty of Electrical and Electronic Engineering, Microelectronics & Nanotechnology-Shamsuddin Research Centre (MiNT-SRC), Institute of Integrated Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
[1]limum.ng@gmail.com, [2]anida@uthm.edu.my

Abstract

This paper presents a comparative study based on statistical analysis of dyslexia risk screening classification implemented using Fuzzy Inference System (FIS) and Waikato Environment for Knowledge Analysis (WEKA) data mining tool. A rapid dyslexia screening system based FIS has been previously developed using four dyslexia screening tests with initial objective to differentiate between normal, dyslexia and slow learner. However, the slow learner subjects were removed from the collected data because of its similar tests scores with dyslexia subjects. Using the collected data (n=104), the dyslexia risk screening via FIS is able to classify normal and dyslexia subjects. A comparative study based statistical evaluation is performed using Naïve Bayes, Decision Table and Random Forest (WEKA) to compare their accuracy, sensitivity, specificity and precision respectively. FIS has achieved 100 % for all the statistical performance in classifying dyslexia subject showing great improvement compared to the previous developed system. Meanwhile, all the three classifiers (WEKA) produce 100 % in terms of specificity and precision using training datasets. It is observed that the Random Forest (with training dataset n=200) shows the best classification performance among three classifiers in WEKA as it managed to reach 100 % in accuracy and sensitivity performance. This paper concludes that FIS and Random Forest are suitable to be used for dyslexic risk confirmation as both yield the best statistical performance. Future studies cover the development of standalone application and evaluation on the user acceptance and satisfaction along with the addition of IQ test in the second stage of FIS.

Keywords: *Dyslexia, Fuzzy Inference System, Screening, WEKA*

1. Introduction

Reading is an ability to look at and comprehend the meaning of written or printed matter by interpreting the characters or symbols of which it is composed [1], according to the definition of Oxford English Dictionary. It is as simple as it might be seemed, but the task of getting the meaning correctly can be a challenge to some of the people who are gifted with dyslexia. Dyslexia, can be known as reading disorder, whereby a dyslexic person struggles with learning in reading, writing, and spelling as well arithmetic [2]. Nonetheless, their intelligence is normal and they can be very creative and talented. Based

ISSN: 2005-4238 IJAST

on the statistics provided by Dyslexia International, it is revealed that around 10 % of the population possess dyslexia but there is a large number of the children who are yet to be identified due to the lacking of the awareness of this common learning difference among the public [3].

At the present time, the dyslexia screening process in Malaysia is administrated manually. Ministry of Education Malaysia has introduced instruments such as Literacy and Numeracy Screening (LINUS) and LINUS 2.0 as well as Instrumen Senarai Semakan Disleksia (ISD) in primary schools to evaluate dyslexic condition of school pupils [4]–[6]. LINUS and LINUS 2.0 consist of a list of constructs that is used for identifying the literacy and numeracy level of pupils. It is aimed to ensure that school pupils are able to acquire skills in reading, writing, spelling and mathematics after three years of primary school. The only difference between LINUS and LINUS 2.0 is that the former is included the mastery of Bahasa Malaysia language while the latter is added with the mastery of English language. However, it was terminated by the Ministry of Education Malaysia starting from 2019 onwards as the government has granted the freedom to every primary school to implement their own system in dyslexia screening [7]. Meanwhile, ISD is a checklist that examines the perception of mastery in reading, spelling and writing as well as to check symptoms of dyslexia among students.

Presently, there is seemed lacking of rapid screening tool for dyslexia as the existing screening approaches are manually conducted and taken more time to obtain the screening result. Most of the applications are used for intervention purpose but rarely focus on the screening. Dyslexia Association of Malaysia has established a screening manual named as Ujian Pengesanan Awal Disleksia Bahasa Melayu for the identification of dyslexia [8]. As an initiative, a rapid dyslexia risk screening system based fuzzy logic was developed based on four tests from Ujian Pengesanan Awal Disleksia Bahasa Melayu to rapidly screen the dyslexia risk level [9]. Nevertheless, the developed system demonstrates a moderate accuracy rate at 56.67 % in recognising between dyslexia and slow learner subjects. Therefore, a rapid dyslexia screening system with improved accuracy of fuzzy logic is proposed in this research work to enhance the reliability of the screening system towards dyslexia and non-dyslexia subjects. Based on the limitations, this paper is aimed to attain the following objectives which are: firstly, to design appropriate rules conditions of fuzzy logic that can differentiate between dyslexia and normal subjects via MATLAB Fuzzy Logic Toolbox; secondly, to conduct subject testing using the developed system; and lastly, to evaluate the performance of the developed Fuzzy Inference System with WEKA data mining tool .

The scopes of study in this research work comprises of four major components which are: (i) four tests (Rapid Naming, One-minute Reading, Two-minute Spelling and Pseudowords) from Ujian Pengesanan Awal Disleksia Bahasa Melayu will be utilised to rule out the dyslexic conditions in the development of dyslexia screening based Fuzzy Inference System; (ii) MATLAB Fuzzy Logic Toolbox will be used to design the entire dyslexia screening system; (iii) data collection will be involved four targeted schools (Preschool, School A, School B and School C); and (iv) statistical comparison based on system's performance in the aspects of accuracy, sensitivity, specificity and precision will be carried out between Fuzzy Inference System and WEKA data mining tool.

2. Research Methodology

A mathematical system that is embedded with fuzzy logic is proposed in this study to design the decision making of identification in dyslexia. The flowchart of the overall methodology is shown in Figure 1. First, the overall process begins with the software

ISSN: 2005-4238 IJAST

implementation using MATLAB Fuzzy Logic Toolbox to design the risk screening system. Then, subject testing is involved to collect the data for evaluating the performance of the developed system. The detailed of the data collection protocol is discussed in the following subsection. It is followed by data filtering to clean the dataset obtained to improve the quality of the results generated [10]. Finally, the data collected will be classified and statistically analysed comparatively using Waikato Environment for Knowledge Analysis (WEKA) data mining tool with the developed Fuzzy Inference System.

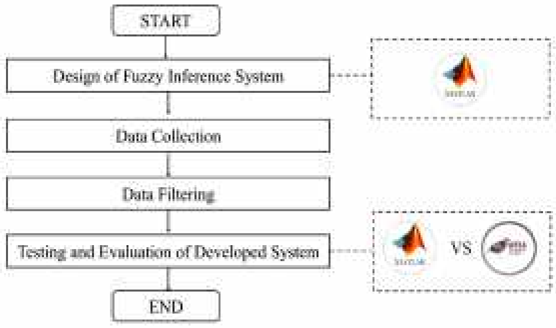

Figure 1. Overall Research Methodology Process

2.1. Participant and Data

This research work received the approval of the research ethical committee from RMC UTHM to conduct subject testing. A sample composed of 117 participants aged between 6 to 10 years old from kindergarten to Standard 4 in one preschool and three primary schools (these schools have special education and remedial classes) were collected. These participants consist of normal, dyslexia and slow learner category where the number of participants was 50, 58 and 9 for each category respectively. The participants' distribution is depicted in Figure 2.

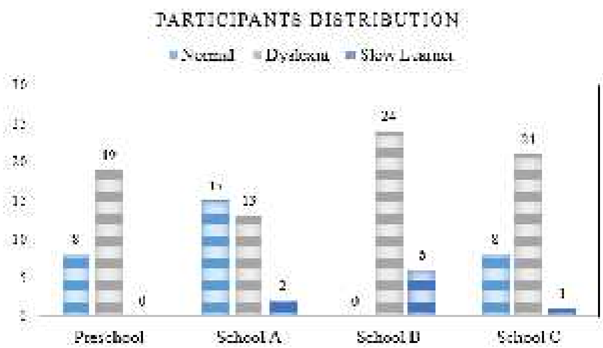

Figure 2. Participants Distribution of the Data Collection

The protocol of the data collection contained four adopted tests from the manual "Ujian Pengesanan Awal Disleksia Bahasa Melayu" which included Rapid Naming, One-minute Reading, Two-minute Spelling and Pseudowords. Each of the participants was required to undergo all the four tests in order to identify whether they were dyslexic or not. Then, the data collected was processed and analysed using developed Fuzzy Inference System and WEKA data mining tool with three types of classifiers comprising of Naïve Bayes, Decision Table and Random Forest to determine their statistical performance. Both of the systems were compared and their performances were evaluated accordingly. The protocol of data collection is shown in Figure 3.

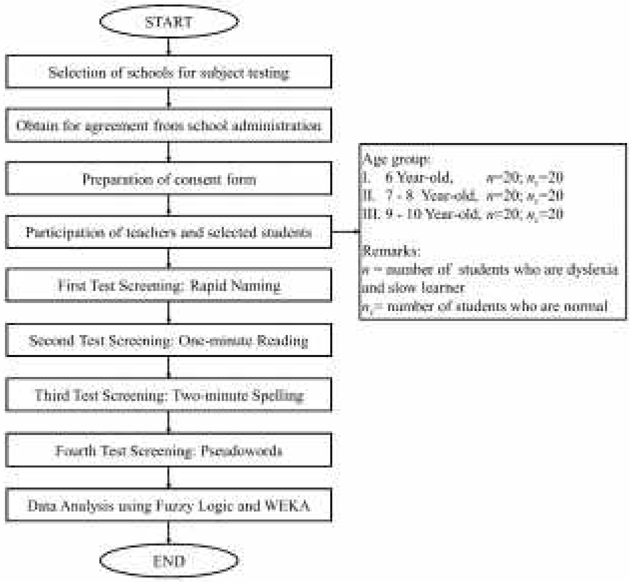

Figure 3. The Protocol of Data Collection

a. Data Filtering

Data filtering is a crucial step as to ensure the data collected is clean and suitable to be used in the evaluation of the developed system. Initially, there were a total of 117 subjects comprising of 77 dyslexia, 31 normal and 9 slow learners. These data were then fed into the developed fuzzy system. However, with the presence of 9 slow learner subjects, the accuracy obtained was not as high as expected though the accuracy attained was significantly improved compared to previously developed system. Therefore, the data filtering process was carried out to remove these 9 slow learner subjects. The reason is because the slow learner possess similar characteristics as dyslexics which are poor in reading and spelling [11] making the system difficult to differentiate between these two subject categories. Then, a second data filtering was done by removing four dyslexic subjects who had shown tremendous improvement after attending the remedial classes, Thus, the final data used in this study was n=104. The details on the data filtering distribution can be referred to Table 1.

Table 1. The Process of Data Filtering to Clean the Dataset

	Dyslexia	Normal	Slow Learner	Total
Before filtering	77	31	9	117
After filtering	77	31	0	108
Current data used	73	31	0	104

2.2. Fuzzy Inference System

Fuzzy logic is chosen to aid in decision making of dyslexia screening system in order to distinguish between dyslexia and non-dyslexia subjects. Fuzzy logic is a logical system that applies linguistic variables as they are defined by fuzzy sets to compute human reasoning [12]. It was introduced by Lotfi Zadeh in 1965 to facilitate the process of decision making in the form of degree of truth instead of the traditional Boolean logical system that works only true and false (1 and 0) [13].

In this study, MATLAB Fuzzy Logic Toolbox is selected to design the dyslexia risk screening system and it consists of four main components, which are FIS Editor, Membership Function Editor, Rule Editor and Rule Viewer. The input and output variables are specified in FIS Editor where the input variables are the scores obtained from Rapid Naming, One-minute Reading, Two-minute Spelling and Pseudowords whereas the output variable is the dyslexia risk.

The threshold value of dyslexia risk level is specified according to the range in Membership Function Editor to allow the mapping between input and output parameters. Then, the establishment of rule conditions for the identification between normal and dyslexic subject is created using Rule Editor. Finally, Rule Viewer allows the insertion of dataset to interpret the result desired based on the rules that have been defined previously. Figure 4 illustrates the overall process that takes place in MATLAB FIS.

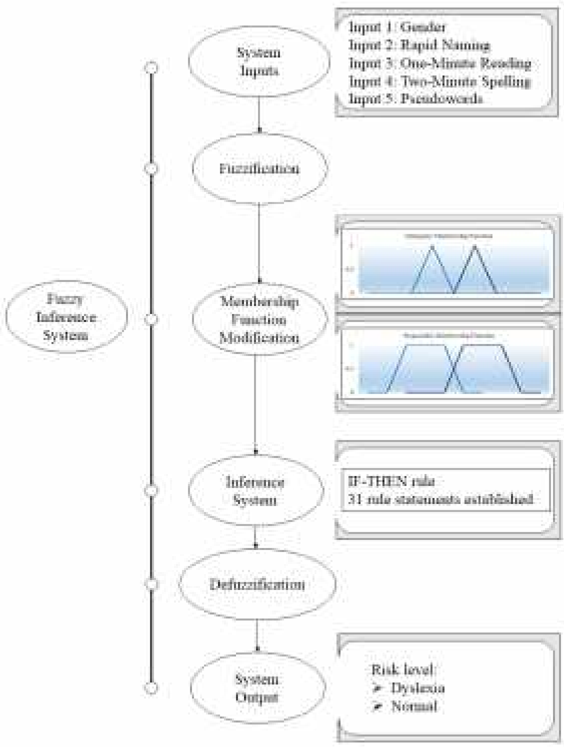

Figure 4. The Operation of Fuzzy Inference System

2.3. Waikato Environment for Knowledge Analysis (WEKA) Data Mining Tool

WEKA is a Java-based data mining tool developed by University of Waikato which consists of four main applications namely Explorer, Experimenter, Knowledge Flow and Simple Command Line Interface (CLI) to perform the desired data mining tasks [14]. In this research, only Explorer is utilised to perform analytics tasks because it has the functionality of data pre-processing and data classification using the pre-implemented machine learning algorithms. Dataset is divided into two subsets for training and testing purpose and stored in CSV format. The training dataset is comprised of Set I (n=100) and Set II (n=200) whereas the testing set is the collected data (n=104). The purpose of creating two different sample sizes for training data is to compare the classifiers' performance towards the testing data in dyslexia identification. Then, the file is converted to ARFF format to declare all the header and data information by assigning its relation and attributes. To make prediction in WEKA; Naïve Bayes, Decision Table and Random Forest are selected to measure each of their performance in classification as these three algorithms are widely used in classification cases [15]–[17]. In Explorer panel, the type of classifiers is pre-selected and then the training data is loaded to build up the model for prediction. Next, the testing dataset is loaded by selecting “Supplied test set” to create required prediction using the built model earlier. The result was saved by clicking the “Visualize Classifier Error”. The same process is repeated for all three selected classifiers using two sets of training data. The overall process of using WEKA is displayed in Figure 5.

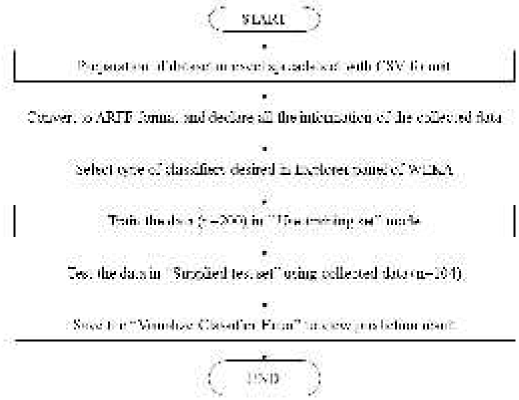

Figure 5. The Operation Flowchart of WEKA Data Mining Tool

2.3. Statistical Analysis

The statistical parameters used to evaluate the performance of the developed system are accuracy, sensitivity, specificity and precision. The confusion matrix and the formula of each of the parameters are depicted as below in Table 2.

Table 2. Confusion Matrix for Statistical Analysis

		Predicted	
		Negative	Positive
Actual	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

a. Accuracy

Accuracy is the ability to determine the dyslexic and normal cases correctly. The formula can be expressed mathematically as in (1).

$$Accuracy = [(TP+TN)/(TP+TN+FP+FN)] \times 100\ \% \qquad (1)$$

b. Sensitivity

Sensitivity is the ability of the system to identify the dyslexic cases correctly. It is expressed in (2) by calculating the proportion of true positive in dyslexic cases.

$$Sensitivity = [(TP)/(TP+FN)] \times 100\ \% \qquad (2)$$

c. Specificity

Specificity is the ability of the system to predict the normal cases correctly. It is expressed in (3) by calculating the proportion of true negative in normal cases.

ISSN: 2005-4238 IJAST

$$Specificity = [(TN)/(TN+FP)] \times 100\ \% \quad (3)$$

d. Precision

Precision is the ratio of correctly predicted dyslexic cases to the total predicted positive values and it is expressed in (4).

$$Precision = [(TP)/(TP+FP)] \times 100\ \% \quad (4)$$

3. Results and Discussion

This section presents the statistical analysis on the collected data based on accuracy, sensitivity, specificity and precision using developed Fuzzy Inference System and WEKA data mining tool.

3.1. Fuzzy Inference System

The outcome generated from the Fuzzy Inference System before data filtering and after data filtering is illustrated in Table 3. Table 3 presents the statistical analysis for three-stage evaluation on the performance of Fuzzy Inference System during data filtering process towards subjects of dyslexia, slow learner, normal and overall which is the combination of all the three subject categories.

As can be observed from the results in Table 3, during before data filtering (n=117) stage, the accuracy of developed system for identifying only dyslexia subjects was 94.81 %. The same result was obtained when the developed system was used towards dyslexia and slow learner subjects showing better performance was achieved compared to the previous developed system which was 56.67 % only. It can be discovered that the accuracy rate was increased 38.14 %. Besides, the system was able to recognise normal subject as the specificity obtained is 100 % with no error. However, the system was unable to classify when only slow learner subjects involved as the performance in all statistical analysis achieved 0 %. This result is not acceptable as the developed system classified the slow learner subjects as dyslexia subjects indicating misclassification. As a result, high false positive value was obtained resulting in low specificity rate. Based on the misclassification results, it shows the relevancy of implementing IQ test for the second stage of fuzzy logic as it aids in the distinction between the IQ level of slow learner and dyslexia. This is because the slow learner tends to have lower IQ level compared to the dyslexia subjects. Overall, the system has achieved an accuracy of 88.89 % for overall category (dyslexia, slow learner and normal).

A second evaluation on the system was conducted for n=108 after data filtering process as mentioned in section 2.1.1 and the accuracy result for overall category (dyslexia and normal subjects) was improved with an increment of 7.41 %. This indicates the significance of data filtering by eliminating the slow learner subjects in the evaluation of data collection. However, the result for dyslexia subjects remains unchanged at the rate of 94.81 % as the dyslexia subjects who had shown improvement were still counted in the testing data. For that, a final evaluation (n=104) was carried out by removing four ameliorated dyslexic subjects in the developed Fuzzy Inference System. The developed Fuzzy Inference System was successfully attained 100 % in all statistical analysis towards combination of dyslexic and normal subjects in the final evaluation. This implies that the system is able to correctly classify between normal and dyslexia subjects.

Table 3. Comparison of Performance of Fuzzy Inference System: (a) Before Data Filtering Process (n=117), (b) First Data Filtering Process (n=108) and (c) Current Data Used (n=104)

(a) Before data filtering process (n=117)				
Performance / Category	Statistical Analysis			
	Accuracy	Sensitivity	Specificity	Precision
Dyslexia (D) only	94.81 %	94.81 %	0 %	100.00%
Slow Learner (SL) only	0 %	0 %	0%	0 %
Normal (N) only	0 %	0 %	100.00 %	0 %
D+SL only	94.81 %	94.81 %	0 %	89.02 %
Overall (D+SL+N)	88.89 %	94.81 %	77.50 %	89.02 %
(b) First data filtering process (n=108)				
Performance / Category	Statistical Analysis			
	Accuracy	Sensitivity	Specificity	Precision
Dyslexia (D) only	94.81 %	94.81 %	0 %	100.00%
Normal (N) only	0 %	0 %	100.00 %	0 %
Overall (D+N)	96.30 %	94.81 %	100.00 %	100.00%
(c) Current data used (n=104)				
Performance / Category	Statistical Analysis			
	Accuracy	Sensitivity	Specificity	Precision
Dyslexia (D) only	100.00 %	100.00 %	0 %	100.00%
Normal (N) only	0 %	0 %	100.00 %	0 %
Overall (D+N)	100.00 %	100.00 %	100.00 %	100.00%

3.2. WEKA Data Mining Tool

For WEKA analysis, Naïve Bayes, Decision Table and Random Forest classification algorithms were applied to the collected data. The two training datasets namely as Set 1 (n=100) and Set 2 (n=200) were created based on the possible conditions that might occur by following the standard in Ujian Pengesanan Awal Disleksia Bahasa Melayu. Meanwhile, the testing data was comprised of 104 subjects with the distribution of 73 dyslexia subjects and 31 normal subjects. The classification performance was evaluated in terms of its accuracy, sensitivity, specificity and precision using Set 1 and Set 2 respectively. Table 4 and Table 5 provide the comparison overview on the classifier performance using Set 1 and Set 2 training data.

Table 4 and Table 5 reveal the effect of applying two samples of training datasets towards subjects of dyslexia only, slow learner only, normal only and combination of three subjects categories in terms of accuracy, sensitivity, specificity and precision. Using Set 1, the best classifier is Decision Table as the algorithm was able to output 100 % in accuracy, sensitivity, specificity and precision. Random Forest is the second best classifier as it has obtained 99.04 % in overall system's performance whereas Naïve Bayes showed the least in accuracy performance among three classifiers.

Table 4. The Summary of the Results Obtained from Testing Data (n=104) using Prediction Model Created by Set 1 (n=100) Training Data based on Naïve Bayes, Decision Table and Random Forest

Performance / Classifiers	Statistical Analysis			
	Accuracy	Sensitivity	Specificity	Precision
A. Naïve Bayes				

(i) Dyslexia (D) only	94.52 %	94.52 %	0 %	100.00%
(ii) Normal (N) only	0 %	0 %	100.00 %	0 %
(iii) Overall (D+N)	96.15 %	94.52 %	100.00 %	100.00 %
B. Decision Table				
(i) Dyslexia (D) only	100.00 %	100.00 %	0 %	100.00 %
(ii) Normal (N) only	0 %	0 %	100.00 %	0 %
(iii) Overall (D+N)	100.00 %	100.00 %	100.00 %	100.00 %
C. Random Forest				
(i) Dyslexia (D) only	98.63 %	98.63 %	0 %	100.00%
(ii) Normal (N) only	0 %	0 %	100.00 %	0 %
(iii) Overall (D+N)	99.04 %	98.63 %	100.00 %	100.00%

Table 5. The Summary of the Results Obtained from Testing Data (n=104) using Prediction Model Created by Set 2 (n=200) Training Data based on Naïve Bayes, Decision Table and Random Forest

Performance / Classifiers	Statistical Analysis			
	Accuracy	Sensitivity	Specificity	Precision
A. Naïve Bayes				
(i) Dyslexia (D) only	97.26 %	97.26 %	0 %	100.00 %
(ii) Normal (N) only	0 %	0 %	100.00 %	0 %
(iii) Overall (D+N)	98.08 %	97.26 %	100.00 %	100.00 %
B. Decision Table				
(i) Dyslexia (D) only	89.04 %	89.04 %	0 %	100.00 %
(ii) Normal (N) only	0 %	0 %	100.00 %	0 %
(iii) Overall (D+N)	92.31 %	89.04 %	100.00 %	100.00 %
C. Random Forest				
(i) Dyslexia (D) only	100.00 %	100.00 %	0 %	100.00 %
(ii) Normal (N) only	0 %	0 %	100.00 %	0 %
(iii) Overall (D+N)	100.00 %	100.00 %	100.00 %	100.00 %

In the second evaluation of WEKA performance, the sample size of training dataset has expanded to n=200 and yielded a better accuracy rate for Naïve Bayes and Random Forest. The Random Forest is said to be the best classifier in this case as it has successfully reached 100 % in accuracy, sensitivity, specificity and precision for dyslexia subjects and also overall category (dyslexia and normal). Meanwhile the accuracy for identifying dyslexia subjects using Naive Bayes has raised to 97.26 % compared its performance when using Set 1 training data. However, Decision Table shows opposite result with even though bigger quantity of training data was applied to the testing data, the accuracy towards dyslexia subject dropped from 100 % to 89.04 %. Also, using the same classifier, a decline of 7.69 % can be noticed for the accuracy calculated towards overall category (dyslexia and normal). This result could be possibly due to the nature of the algorithms itself and more research will be required to look into this issue [18]. The positive results generated in Naïve Bayes and Random Forest highlighted the success of undergoing data filtering process to ensure that the data are fit and suitable to be used. The outcome from two training datasets, it can be recognised that Random Forest and

Naïve Bayes produce better classification result when bigger sample size (n=200) is used.

3.3. Comparison between Fuzzy Inference System and WEKA Data Mining Tool Towards Combination of Normal and Dyslexia Subjects Classification

A comparative study was performed to evaluate the performance of Fuzzy Inference System and WEKA towards dyslexia and normal subjects' classification. Table 6 presents comparison table on statistical analysis performance between WEKA analysis and Fuzzy Inference System.

Table 6. Comparison between Fuzzy Inference System and WEKA Data Mining Tools Towards Combination of Normal and Dyslexia Subjects

Performance / Classifiers	Statistical Analysis			
	Accuracy	Sensitivity	Specificity	Precision
A. Fuzzy Inference System	100.00 %	100.00 %	100.00 %	100.00 %
B. WEKA				
I. Naïve Bayes				
(i) Set I (n=100)	96.15 %	94.52 %	100.00 %	100.00 %
(ii) Set II (n=200)	98.08 %	97.26 %	100.00 %	100.00 %
II. Decision Table				
(i) Set I (n=100)	100.00 %	100.00 %	100.00 %	100.00 %
(ii) Set II (n=200)	92.31 %	89.04 %	100.00 %	100.00 %
III. Random Forest				
(i) Set I (n=100)	99.04 %	98.63 %	100.00 %	100.00 %
(ii) Set II (n=200)	100.00 %	100.00 %	100.00 %	100.00 %

Based on Table 6, it can be seen that the best accuracy in identifying dyslexia and using training data of normal and dyslexic subjects is 100 % when Fuzzy Inference System and Random Forest (Set 2) were applied. Naïve Bayes has obtained the second best classifier as it was able to produced 98.08 % accuracy rate via training dataset n=200. For sensitivity measure, the rate of 100 % was achieved in Fuzzy Inference System, Decision Table (Set 1) and Random Forest (Set 2) meaning that these classifiers possess good ability in predicting the dyslexic condition. All the four classifiers were achieved 100 % in specificity as they are able to identify the patients who are not dyslexic which implying that these classifiers are useful in ruling the subjects who have dyslexic condition.

3. Conclusion and Recommendation

In conclusion, the implementation of Fuzzy Inference System and WEKA analysis for identifying between normal and dyslexia has been accomplished. It can be concluded that the best classifier of overall system (towards dyslexia and normal subjects) is Fuzzy Inference System and Random Forest (WEKA) due to its 100.00 % in accuracy, sensitivity, specificity and precision signifying that both classification algorithms exhibit the ability to distinguish between normal and dyslexia subject. Moreover, Fuzzy Inference System and Random Forest (n=200) are the best classifier for identifying the dyslexia subjects only as they exhibit ability to be used as the classification model for dyslexia screening with 100 % in all statistical analysis. Notably, classification may be varied depending on the types of applications to be performed [19]. Despite the fact that Naïve

ISSN: 2005-4238 IJAST

Bayes and Decision Table were unable to achieve 100 % accuracy, the results with accuracy rate greater than 90 % is considered acceptable in the measures of diagnostic accuracy [20].

As for the future work, a standalone application for rapid dyslexia screening system based fuzzy logic will be developed and tested for user acceptance and satisfaction on the functionality of the application. Besides that, implementation of Artificial Neural Network can be carried out to compare its performance with the current developed system. Finally, the second stage of fuzzy logic can be conducted to differentiate between dyslexia and slow learner subjects by adding IQ test in the proposed system.

Acknowledgments

Authors would like to express deeply gratitude to Research Management Centre (RMC) UTHM for providing funding via GPPS Grant No. H315 and TIER 1 Grant No. H226.

References

[1] M. W. Angus Stevenson, *Concise Oxford English Dictionary (12th Edition)*. 2011. (Stevenson, A. and Waite, M. (2011).
[2] G. R. Lyon, S. E. Shaywitz, and B. A. Shaywitz, "A Definition of Dyslexia," Ann. Dyslexia, vol. 53, no. 1, pp. 1–14, 2003.
[3] Dyslexia International, "Dyslexia International: Better Training, Better Teaching," 2014.
[4] N. Sani and A. R. Idris, "Implementation Of Linus Programme Based On The Model Of Van Meter And Van
[5] E. Luyee, F. Roselan, and N. Anwardeen, "Suitability of the Literacy and Numeracy Screening (Linus) 2.0 Programme in Assessing Children's Early Literacy," *The Malaysian Online journal of Educational Science*, vol. 3, no. 2.30, pp. 36-44, 2015.
[6] D. of S. Education, Instrumen Senarai Semak Disleksia. 2011.
[7] F. Bungga, "School education's LINUS programme to end next year," New Straits Times, 2018.
[8] S. Amirin, H. Azahari, S. Mohtaram, A. N. Che Pee and A. S. Shibghatullah, "Ujian Pengesanan Awal Disleksia Bahasa Melayu,", pp. 1-50, 2017.
[9] N. A. Jumadi, N. L. Mun, M. Morsin, and S. Mohtaram, "Dyslexia Risk Screening System based Fuzzy Logic," *Int. J. Eng. Technol.*, vol. 7, no. 4.30, pp. 165–169, 2018.
[10] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang, "Data cleaning: Overview and emerging challenges," Proc. ACM SIGMOD Int. Conf. Manag. Data, vol. 26-June-20, pp. 2201–2206, 2016.
[11] P. L. Brooks and S. A. J. Weeks, "A Comparison of the Responses of Dyslexic, Slow Learning and Control Children to Different Strategies for Teaching Spellings," Dyslexia, vol. 4, no. 4, pp. 212–222, 1998.
[12] L. A. Zadeh, "Fuzzy Sets," *Inf. Control*, vol. 8, pp. 338–353, 1965.
[13] L. A. Zadeh, "Fuzzy Sets As A Basis For A Theory Of Possibility," *Fuzzy Sets and Systems*, pp. 3-28, 1978.
[14] I. H. W. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer Peter Reutemann, "The WEKA Data Mining Software: An Update," SIGKDD Explor., vol. 11, no. 1, pp. 10–18, 2017.
[15] J. Kazmierska and J. Malicki, "Application of the Naïve Bayesian Classifier to optimize treatment decisions," *Radiother. Oncol.*, vol. 86, no. 2, pp. 211–216, 2008.
[16] R. Kohavi and D. Sommerfield, "Targeting Business Users with Decision Table Classifiers," pp. 1–5, 2011.

[17] A. Uzelac, N. Gligorić, and S. Krčo, "System for recognizing lecture quality based on analysis of physical parameters," Telemat. Informatics, vol. 35, no. 3, pp. 579–594, 2018.
[18] D. Bzdok, M. Krzywinski, and N. Altman, "Machine Leraning: A Primer," *Nat. Methods*, vol. 14, p. 1119, 2017.
[19] Information Resources Management Association, "Biomedical Engineering: Concepts, Methodologies, Tools, and Applications," pp. 137, 2017.
[20] A. Simundic, "Measures of Diagnostic Accuracy: Basic Definitions," *Electronic Journal of IFCC*, vol. 19, no. 4, pp. 203-211, 2009.