

# Dyslexia Detection System Based On Hand Writing

G. Roja

Department of Computer Science  
& Engineering (AI & ML),  
G Narayanamma Institute of  
Technology & Science (for  
women), (JNTUH), Hyderabad,  
India

P. Tejashwi

Department of Computer Science &  
Engineering (AI & ML),  
G Narayanamma Institute of  
Technology  
& Science (for women), (JNTUH),  
Hyderabad, India

P. Rakshitha

Department of Computer Science &  
Engineering (AI & ML),  
G Narayanamma Institute of  
Technology & Science (for  
women), (JNTUH),  
Hyderabad, India

S. Niharika

Department of Computer Science  
& Engineering (AI & ML),  
G Narayanamma Institute of  
Technology & Science (for  
women), (JNTUH),  
Hyderabad, India

O. Ruchitha

Department of Computer Science &  
Engineering (AI & ML),  
G Narayanamma Institute of  
Technology & Science (for women),  
(JNTUH),  
Hyderabad, India

**Abstract**—Dyslexia is a learning disability which hampers a child's capability to read, interpret sounds, connect letters, and understand written language even when the child has good eyesight and intelligence. It involves difficulty in phonological processing and word decoding rather than any visual problem. The difficulty is manifested through poor handwriting pattern and abnormal eye movements while reading. The proposed project entitled "Detection of Dyslexia in Children Using Handwriting and Eye Movement" employs a machine learning technique in order to detect dyslexia symptoms in children early. The system will perform the OCR of handwriting samples and then classify them through supervised learning algorithm like Artificial Neural Network (ANN) and Support Vector Machines (SVM). In order to improve the performance of the system, eye movements of individuals will be analyzed using deep learning approaches like CNN and Recursive Feature Elimination (RFE) for feature selection. Handwriting and eye movement analysis results will help to differentiate between dyslexic and non-dyslexic children. The project offers an inexpensive, impartial, and automatic tool for detecting dyslexia symptoms in children.

**Index Terms**—Dyslexia symptom; dyslexia handwriting; pattern recognition method; artificial neural network.

## I. INTRODUCTION

Dyslexia From the Greek language where Dys stands for difficulty and Lexia stands for reading, dyslexia can be defined as a learning disorder in which one finds it hard to read [1]. In the end, dyslexia does not mean that there is lack of intelligence but with the right teaching method, a dyslexia individual will be a successful person [2].

At the moment, the symptoms of dyslexia are being detected through the use of conventional techniques, including different screening tests. The test involves various assessments conducted through peer meeting and is performed to determine the cognitive strengths and weaknesses of the child in terms of phonemic awareness, visual-spatial skills, sound letter identification and so forth [3]. For example, the Ministry of Education Malaysia provides a checklist to detect the presence of dyslexic children, as well as their behavior and symptoms [4]. In brief, there are three key factors to be used in order to evaluate the potential dyslexic children: 1) the skill level in spelling, reading, and writing, 2) the strengths of students and 3) their weaknesses. There are some disadvantages of using the conventional techniques since they take too much time due to the fact that the assessors have to patiently observe and evaluate the potential dyslexic children. The conventional method also requires the participation of dyslexia experts because during the testing process, a dyslexic child becomes quickly bored and is not capable of focusing properly [5]. Besides, the process takes more time to be completed. Hence, a research by Mekyska et. al. (2017) suggests an automated diagnosis of the disorder to estimate the degree of difficulty and it is named the Handwriting Proficiency Screening Questionnaire (HPSQ). According to the findings, the proposed technique is more sufficient than the conventional one because the system can dynamically evaluate the handwriting.

As suggested by the study of Al-barhamtoshy (2017), a technique has been introduced for identifying the dyslexia symptom through brain activity analysis in two hemispheres (lobes), wherein the left lobe is verbal and arithmetic and the right one is spatial. The Electroencephalogram (EEG) signals have been used in order to characterize the electrical activity acquired through the scalp of the EEG through the use of metal electrode [7]. On the other hand, the study conducted by Mahmoodin (2015) involved brain activity analysis of dyslexic children under resting condition and writing condition through the EEG signal Power Spectrum Density (PSD). Writing common words activity has also been included in the study under the guidance of cues from the computers. However, there is a limitation of appropriate mathematical processing technique due to availability of huge amount of information in the results [8].

Another way of testing which was done through the use of auditory processing has been done to diagnose early signs of dyslexia, and the research is mainly targeted towards people with phonological dyslexia or those having trouble with sounding. The researchers analyze the data from the responses of the children who underwent phonological processing task using the MATLAB' classifier. This phonological processing task has been done through the use of prototypes that have been designed for the parents to diagnose the presence of dyslexic symptoms among children. Through the responses obtained from the children's data, parents are able to download the results of potential dyslexic symptoms among the children [9].

Eye movement testing, as suggested by Application et al., in 1985 is another test for detecting dyslexia's early symptoms. This test is done by stimulating a preset pattern of eye movement, which is measured in relation to dyslexic patients. An automated test to detect the degree of dyslexia is also suggested in the research. Eye Movement Detector (EMD) tests, Electrooculography (EOG), photoelectrocorneography, and video camera technique were applied in the research and produced output signals response in terms of eye movements. There are some constraints identified like being difficult to construct, expensive and requiring more time to construct [10].

A number of researchers have now suggested the use of handwriting methods as a screening test for detecting possible dyslexia symptoms since these are easy to analyse and allow for simple system development [11]. MATLAB software is the most recommended option for developing such a system because it is simple, inexpensive, and requires only a short development time while at the same time producing an output that is almost identical to that of other methods. In MATLAB there are a variety of processing tools that can be used, for example the Genetic Fuzzy System [12], the Penmanship Objective Evaluation Tool (POET) [13] [6], character recognition, neural networks and so on. Nevertheless, most of the screening tests proposed for the identification of dyslexia symptoms are still based on conventional approaches which involve human evaluation. These human evaluation methods have the limitations already mentioned. For this reason, the present study proposes the application of a pattern recognition technique using MATLAB software for the detection of dyslexiya.

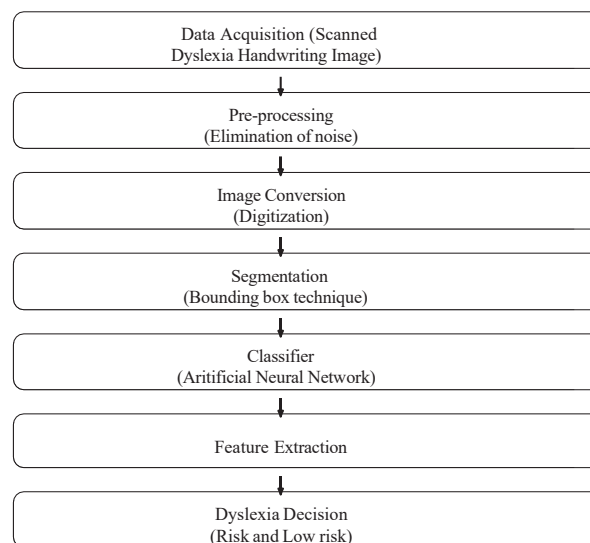


Fig. 1. Proposed pattern recognition techniques for dyslexic handwriting detection .

symptom with scanned handwritten images. The aim of this study is to detect early dyslexic symptom on primary school children age between 7 to 12 years old based on writing pattern approach.

## II. METHODOLOGY

### A. Image processing

The overall proposed work for pattern recognition method based on the handwriting image of dyslexic children is shown in Fig. 1. The scanned handwriting image is given as an input to the proposed system initially. The scanned images of dyslexic handwriting are collected from association of Dyslexia Malaysia (ADM), Sungai Petani, Kedah, Malaysia. The number of collected data sets is 30 scanned samples of dyslexic handwriting. In this project, several techniques are used to detect the dysgraphia and dyscalculia symptoms by image pre-processing and pattern recognition techniques:

- i. *Data acquisition:* The be selected dataset contains 30 samples of handwritten text collected from 30 dyslexic children that have been scanned by a scanner to get good quality scan for template dataset. The dataset comprises 8 selected lower-case of letters and numbers which are b, c, f, p, 2, 5, 6 and 7. The letters and numbers are chosen because this is the most common mistake made by dyslexic in writing.
- ii. *ge Conversion:* The data samples of handwritten text images scanned used at 300dpi which converts the data on the paper which being scanned into a bitmap image. These images are then store for pre-processing step.
- iii. *-processing:* The pre-processing of the image is the process to change and modification the image to make it suitable for recognition. The techniques used to enhance the image, firstly is converted from RGB to grayscale conversion. The coloured image means the pixel of value of the image contains a three-colour component which are red, green and

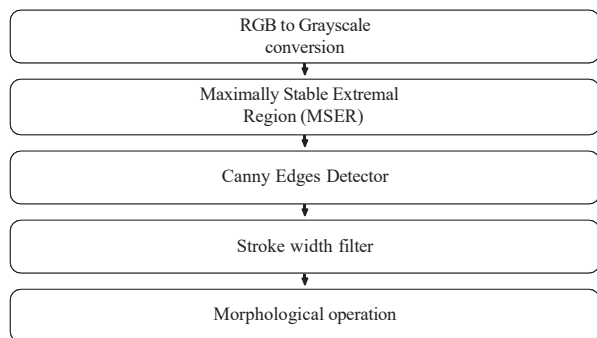


Fig. 2. Overall pre-processing methods

blue. This coloured image converted into grayscale because it is represented in a single matrix as the detection of the characters on a coloured image is more difficult to be proposed compared with grayscale image. Next, Maximally Stable Extremal Regions (MSER) detector is used since text characters usually have consistent colour. MSER region detector is also used for finding regions of similar intensities in the image. Thirdly, the Canny Edges detection is a multi-step algorithm that can detect edges with noise at the same time. It is used since the handwriting text is in clear background; it tends to produce high response to edge detection. Next step is filtering the character. Some of the remaining connected components being removed by using their region properties. The regionFilteredTextMask is used to eliminate regions that do not follow common text measurement. Furthermore, the filter character by using stroke width is used to remove regions where the stroke width exhibits too much variation. Final step is morphology which is used dilation and erode algorithms so that the character become wide to connect unconnected lines at the handwritten image make the character become thinner after dilation process to gain its original size.

1) *iv. Segmentation*: After the pre-processing completed, it is passed to segmentation phase in which the characters is separated from one another by using bounding box techniques [14]. Bounding box techniques consists of area, top, left, width and length. Then, the bounding objects are extracted in the feature extraction process.

2) *v. Features Extraction*: In features extraction process, an Optical Character Recognition (OCR) is used to recognize the characters from the input image. It is interconnection of several components which together perform the intended function [15]. After extracting the image and recognize the characters, the corrected characters are shown in the command window. The total of the correct detection will be calculated manually using Eq. (2) before classify using Artificial Neural Network as in Eq. (1).

### III. ANN CLASSIFIER

Artificial Neural Network (ANN) is a network of artificial neurons. Finally, the entire processed dyslexic handwriting image is further processed in ANN classifying network to classify the potential of dyslexia symptom. A Multilayer

TABLE I  
 THE DISTRIBUTION DATA FOR TRAIN, VALIDATION AND TEST.

| Remarks    | Distribution (%) | No Data |
|------------|------------------|---------|
| Train      | 80               | 192     |
| Test       | 10               | 24      |
| Validation | 10               | 24      |

Dyslexia Handwriting Image

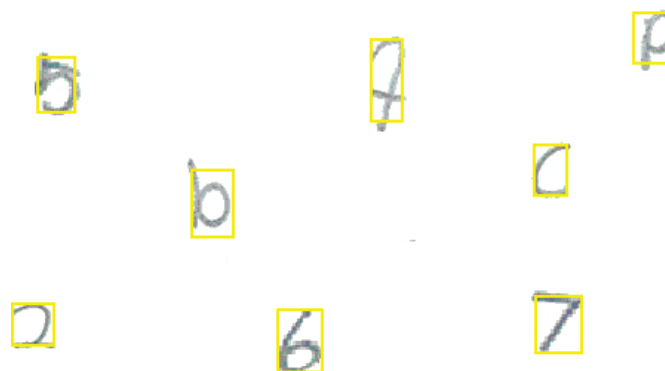


Fig. 3. Bounding box of detected characters

Perceptron (MLP) is used as the classifier to train and test the dataset. The dataset of the MLP is 240 which are after the extracted features process. Table I below shows the ratio use for the ANN is 0.8 for train, 0.1 for validation and 0.1 for testing. While for hidden nodes from 1 to 20 which to see the best performance and to get the average.

### IV. RESULTS AND DISCUSSION

First, the image processing part is quiet challenging because of handwritten can be any shape and types. So, there are several steps to process based on the methodology. After the preprocessing, each character is cropped on each letters and numbers that had selected which are for letters, 'b', 'c', 'f', 'p' while for numbers are '2', '5', '6', '7' at the segmentation stage. These numbers are selected because of they are common letters that the dyslexic always confused of the shape of the characters (based on the data that get from ADM). The cropped numbers and letters are shown in Fig. 3.

From OCR method, each of the characters detected from the samples produces the output of the correct numbers and letters. Fig. 4 shows example of the detected characters with the output from one of the subjects.

In order to determine the risk, there are 30 samples use from the data. Table II shows the data that obtained from the pre-processing and optical recognition method. From Table II shows the correct detection, C and manually calculation, M provides by the proposed method. The characters that correctly detected by the proposed method, C is fractionally divide to overall characters, O to find the accuracy of correct detection and manual calculation which is without using OCR method. The overall characters, O of the selected character of dyslexia handwriting are 8.

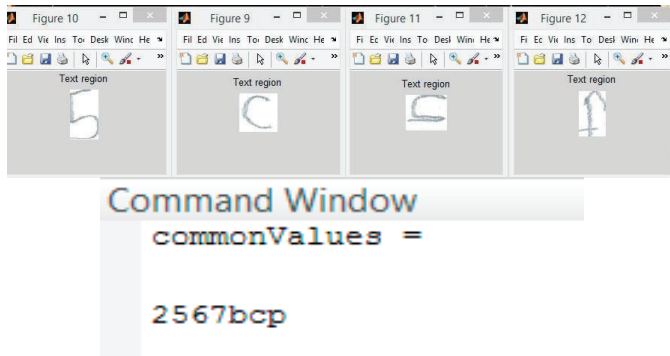


Fig. 4. (i) Detected each character as the output and (ii) display in command windows

$$\text{Accuracy}_{\text{Automated}}(\%) = \frac{\text{Corrected Character } C}{\text{Overall Characters } O} \times 100\% \quad (1)$$

$$\text{Accuracy}_{\text{Manual}}(\%) = \frac{\text{Manual Calculation } M}{\text{Overall Characters } O} \times 100\% \quad (2)$$

By compare the manually calculation, M and correct detection, C shows that the differences of corrected characters are slightly differences which mean that each sample, there are a few undetected correct characters. So, OCR method can be used for character recognition. The Table 3 below shows the accuracy of the difference similarity between C and M.

Next, the level of dyslexia will be train by using MLP neural network. Based on the level of dyslexia symptom are shown in Table IV.

The ANN inputs which extracted from the segmentation process of the image in step v), is set as attributes for the classifier. The followings attribute for ANN input and output layers are shown in Table V.

The ANN classifier is to train and test each process characters of the image which the characters has been extracted. These features were classified by using Multilayer Perceptron (MLP). From the result ANN, in the hidden nodes of morphological features are recorded in Table 5 shows that the performance was determined by its accuracy. For the hidden layers, sigmoid function was utilizing as the dynamic function. The best numbers was selected by maximum value of the classification accuracy on the test value is 0.7083 which the value of hidden nodes of ANN is 4.

Fig. 5 shows the train, validity and test graph which the x-axis is the hidden layer while the y-axis is the accuracy of the classification low risk (0) and risk (1) based on the features extraction. In ANN, good accuracy is in range from 80% to 100%. But, from the Table 6 show that the value in range from 50% to 75%. It can be concluded that the accuracy for classification using ANN are not in good range due to the samples are not much and need to add more features at image processing steps.

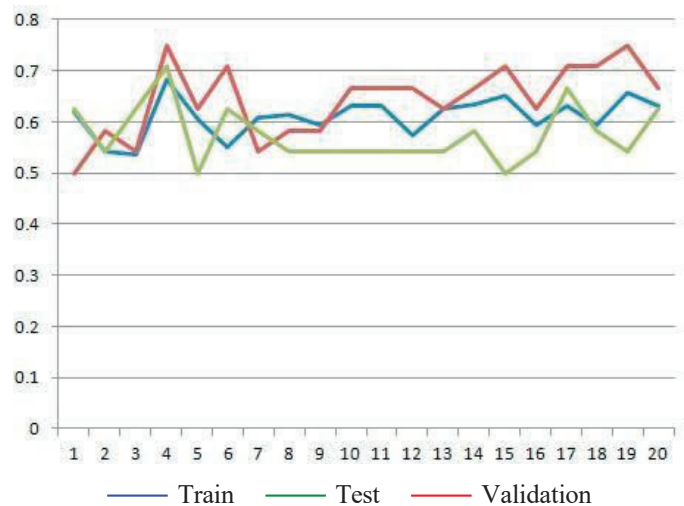


Fig. 5. Performance of ANN

## V. CONCLUSION

As a conclusion, the first objective which to develop an automated handwriting recognition system by using image processing technique and pattern recognition has been achieved. The

pattern recognition by using optical recognition method is one way to recognize the characters. An automated handwriting recognition system by using image processing technique and pattern recognition had shown significant results. The OCR method is good method to use for recognizing the solid characters for instances the words at the signboard etc. For further works, handwritten recognition by using OCR can be used but need additional method such as cursive method, skew method etc. for the program to run more efficient. This is because of it is not stable to detect and recognize the handwritten recognition due to the shape of the handwriting. For the second objective is to classify the levels of dyslexic symptoms based on its accuracy is also achieved. Based on the result, the performance of the classification accuracy is immediate. This is because of the ANN need a lot of samples to get the high accuracy and need to do more features of the characters.

From the findings of the data from Association of Dyslexia Malaysia (ADM), this project is able to obtain the dysgraphia and dyscalculia symptom, however for further improvement on detection of dyslexia symptom needed to consider all aspects from other evaluations, for example phonic, visual symptom etc. Compared to the other studies, this project has significant development, but more additional works require for improvements such as the physical of the dyslexia children experimental on hearing the sound, dyslexic visual surroundings and etc which need much time to develop and costly as well as more complicated.

## ACKNOWLEDGMENT

The author would like to thank Association of Dyslexia Malaysia (ADM), Sungai Petani, Malaysia and Universiti

TABLE II  
 PERCENTAGE CORRECTS CHARACTERS BY DYSLIXIC CHILDREN USING OPTICAL CHARACTER RECOGNITION.

| Subject | Correct detection, C | Accuracy (Automated), % | Manually Calculation, M | Accuracy (Manual), % |
|---------|----------------------|-------------------------|-------------------------|----------------------|
| 1       | 5                    | 62.5                    | 6                       | 75                   |
| 2       | 3                    | 37.5                    | 4                       | 50                   |
| 3       | 4                    | 50                      | 4                       | 50                   |
| 4       | 3                    | 37.5                    | 3                       | 37.5                 |
| 5       | 4                    | 50                      | 5                       | 62.5                 |
| 6       | 3                    | 37.5                    | 4                       | 50                   |
| 7       | 4                    | 50                      | 4                       | 50                   |
| 8       | 2                    | 25                      | 3                       | 37.5                 |
| 9       | 4                    | 50                      | 4                       | 50                   |
| 10      | 5                    | 62.5                    | 6                       | 75                   |
| 11      | 2                    | 25                      | 2                       | 25                   |
| 12      | 1                    | 12.5                    | 2                       | 25                   |
| 13      | 2                    | 25                      | 2                       | 25                   |
| 14      | 3                    | 37.5                    | 4                       | 50                   |
| 15      | 5                    | 62.5                    | 5                       | 62.5                 |
| 16      | 6                    | 75                      | 6                       | 75                   |
| 17      | 3                    | 37.5                    | 3                       | 37.5                 |
| 18      | 3                    | 37.5                    | 3                       | 37.5                 |
| 19      | 4                    | 50                      | 4                       | 50                   |
| 20      | 3                    | 37.5                    | 3                       | 37.5                 |
| 21      | 1                    | 12.5                    | 1                       | 12.5                 |
| 22      | 3                    | 37.5                    | 3                       | 37.5                 |
| 23      | 7                    | 87.5                    | 7                       | 87.5                 |
| 24      | 2                    | 25                      | 2                       | 25                   |
| 25      | 4                    | 50                      | 4                       | 50                   |
| 26      | 4                    | 50                      | 4                       | 50                   |
| 27      | 4                    | 50                      | 4                       | 50                   |
| 28      | 5                    | 62.5                    | 5                       | 62.5                 |
| 29      | 7                    | 87.5                    | 7                       | 87.5                 |
| 30      | 6                    | 75                      | 6                       | 75                   |

TABLE V  
 ATTRIBUTES FOR THE CLASSIFIER.

| Attributes              | Input layer |
|-------------------------|-------------|
| Area                    |             |
| Top                     |             |
| Left                    |             |
| Width                   |             |
| Height                  |             |
| Output layer (decision) |             |
| 0 (Low Risk)            |             |
| 1 (Risk)                |             |

TABLE III  
 THE ACCURACY OF THE DIFFERENCE SIMILARITY BETWEEN C AND M.

|                                                 | Samples | Accuracy, % |
|-------------------------------------------------|---------|-------------|
| Differences of similar accuracy between C and M | 22      | 22          |

$$22 \div 30 \times 100\% = \frac{22}{30} \times 100\% = 73.33\%$$

TABLE IV  
 LEVEL OF DYSLIXIA SYMPTOM

| Level of dyslexia | Percentage correct characters (%) | Correct characters |
|-------------------|-----------------------------------|--------------------|
| Risk (R)          | < 50                              | 0-4                |
| Low Risk (LR)     | ≥ 50                              | 5-10               |

Teknologi MARA for cooperation given upon completing this project. Also, thanks to all staff of Faculty of Electrical Engineering, Universiti Teknologi MARA Penang Campus for the supports.

## REFERENCES

- [1] J. Stein, "The Magnocellular Theory of Developmental Dyslexia," *Dyslexia*, vol. 7, no. 1, pp. 12-36, 2001.
- [2] International Dyslexia Association, *Dyslexia in the Classroom -What every Teacher needs to Know.*, vol. 1, no. 1. 2017.
- [3] A. Guide, E. Literacy, and R. Struggles, "The New Jersey Dyslexia Handbook," no. September, 2017.
- [4] Ministry of Education, "Instrumen Senarai Semak Disleksia," pp. 1-10, 2011.
- [5] H. Husni and Z. Jamaludin, *Masalah Pembacaan Kanak-Kanak Disleksia : Bagaimana Teknologi Membantu?* Institut Terjemahan & Buku Malaysia Berhad, 2015.
- [6] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and Rating of Developmental Dysgraphia by Handwriting Analysis," *IEEE Trans. Human-Machine Syst.*, vol. 47, no. 2, pp. 235-248, 2017.
- [7] H. M. Al-barhamtoshy, "Diagnosis of Dyslexia using computation analysis - IEEE Xplore Document," no. 1, 2017.
- [8] Z. Mahmoodin, "An Analysis of EEG Signal Power Spectrum Density Generated During Writing in Children with Dyslexia," pp. 6-8, 2015.
- [9] A. Poole and C. Aylward, "Lexa : A Tool for Detecting Dyslexia through Auditory Processing," pp. 6-10, 2017.
- [10] F. Application, P. Data, C. H. Smith, M. Rosen, H. Carl, P. A. Conf, and P. E. J. Eisenzopf, "Method and Means for Detecting Dyslexia," vol. 35, no. 7, pp. 886-890, 1985.
- [11] H. Van Waelvelde, T. Hellinckx, W. Peersman, and B. C. M. Smits-Engelsman, "SOS: A screening instrument to identify children with handwriting impairments," *Phys. Occup. Ther. Pediatr.*, vol. 32, no. 3, pp. 306-319, 2012.
- [12] A. M. Palacios, L. Sa'ñchez, and I. Couso, "Diagnosis of dyslexia with low quality data with genetic fuzzy systems," *Int. J. Approx. Reason.*, vol. 51, no. 8, pp. 993-1009, 2010.

TABLE VI  
 HIDDEN NODES FOR MORPHOLOGICAL FEATURES.

| Hidden layer | Train  | Validity | Test   | Epoch |
|--------------|--------|----------|--------|-------|
| 1            | 0.6198 | 0.5000   | 0.6250 | 8     |
| 2            | 0.5417 | 0.5833   | 0.5417 | 5     |
| 3            | 0.5365 | 0.5417   | 0.6250 | 2     |
| 4            | 0.6823 | 0.7500   | 0.7083 | 12    |
| 5            | 0.6042 | 0.6250   | 0.5000 | 6     |
| 6            | 0.5521 | 0.7083   | 0.6250 | 3     |
| 7            | 0.6094 | 0.5417   | 0.5833 | 3     |
| 8            | 0.6146 | 0.5833   | 0.5417 | 3     |
| 9            | 0.5938 | 0.5833   | 0.5417 | 2     |
| 10           | 0.6302 | 0.6667   | 0.5417 | 2     |
| 11           | 0.6302 | 0.6667   | 0.5417 | 2     |
| 12           | 0.5729 | 0.6667   | 0.5417 | 3     |
| 13           | 0.6250 | 0.6250   | 0.5417 | 2     |
| 14           | 0.6354 | 0.6667   | 0.5833 | 2     |
| 15           | 0.6510 | 0.7083   | 0.5000 | 2     |
| 16           | 0.5938 | 0.6250   | 0.5417 | 1     |
| 17           | 0.6302 | 0.7083   | 0.6667 | 3     |
| 18           | 0.5938 | 0.7083   | 0.5833 | 3     |
| 19           | 0.6563 | 0.7500   | 0.5417 | 3     |
| 20           | 0.6302 | 0.6667   | 0.6250 | 3     |

TABLE VII  
 AVERAGE CLASSIFICATION PERFORMANCE OF ANN.

| Dataset  | Average Accuracy | Train |
|----------|------------------|-------|
|          | 0.6102           |       |
| Validity | 0.6438           |       |
| Test     | 0.5750           |       |

- [13] S. Rosenblum, A. Y. Dvorkin, and P. L. Weiss, "Automatic segmentation as a tool for examining the handwriting process of children with dysgraphic and proficient handwriting," *Hum. Mov. Sci.*, vol. 25, no. 4–5, pp. 608–621, 2006.
- [14] Jameel, "a Review on Recognition of Handwritten Urdu Characters Using Neural Networks," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 9, pp. 727–730, 2017.
- [15] E. N. Bhatia, "Optical Character Recognition Techniques: A Review," vol. 4, no. 5, pp. 142–146, 2014.