UNIVERSITI TEKNOLOGI MARA

MULTICLASS DYSGRAPHIA SEVERITY CLASSIFICATION IN MALAY HANDWRITING USING ENHANCED IMAGE PROCESSING TECHNIQUES AND FEATURE-FUSION BASED MACHINE LEARNING

SITI AZURA BINTI RAMLAN

PhD

August 2026

UNIVERSITI TEKNOLOGI MARA

MULTICLASS DYSGRAPHIA SEVERITY CLASSIFICATION IN MALAY HANDWRITING USING ENHANCED IMAGE PROCESSING TECHNIQUES AND FEATURE-FUSION BASED MACHINE LEARNING

SITI AZURA BINTI RAMLAN

Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)

Faculty of Electrical Engineering

August 2026

CONFIRMATION BY PANEL OF EXAMINERS

I certify that a Panel of Examiners has met on 12th March 2026 to conduct the final examination of Siti Azura binti Ramlan on her Doctor of Philosophy thesis entitled "Multiclass Dysgraphia Severity Classification in Malay Handwriting Using Enhanced Image Processing Techniques and Feature Fusion–Based Machine Learning" in accordance with Universiti Teknologi MARA Act 1976 (Akta 173). The Panel of Examiners recommends that the student be awarded the relevant degree. The Panel of Examiners was as follows:

Irni Hamiza Hamzah, PhD
Associate Professor
Faculty of Electrical Engineering
Universiti Teknologi MARA
(Chairman)

Adi Izhar Che Ani, PhD
Senior Lecturer
Faculty of Electrical Engineering
Universiti Teknologi MARA
(Internal Examiner)

Syafeeza Ahmad Radzi, PhD
Senior Lecturer
Faculty of Electronics and Computer Technology and Engineering
Universiti Teknikal Malaysia Melaka
(External Examiner)

PROFESSOR DR HJH ZURAEDA IBRAHIM
Dean
Institute of Postgraduate Studies
Universiti Teknologi MARA
Date : 7 August 2026

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the results of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Postgraduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student	:	Siti Azura binti Ramlan
Student I.D. No.	:	2021356357
Programme	:	Doctor of Philosophy (Electrical Engineering) – EE950
Faculty	:	Electrical Engineering
Thesis Title	:	Multiclass Dysgraphia Severity Classification in Malay Handwriting Using Enhanced Image Processing Techniques and Feature Fusion–Based Machine Learning
Signature of Student	:	
Date	:	August 2026

ABSTRACT

Dysgraphia is a learning disability (LD) that causes difficulties in writing and commonly occurs among school-age children. One of its key symptoms is poor, inconsistent handwriting, typically identified through expert observation with prescribed scoring systems. However, such conventional assessments are time-consuming, prone to interruption, and may produce subjective or biased results. With advances in technology, automated dysgraphia detection methods have emerged, leading to various handwriting-based approaches. However, most existing studies emphasize dynamic handwriting features, such as tilt, angle, pressure level, and pen position, and are limited to binary classification (dysgraphia vs. non-dysgraphia).Consequently, multiclass severity classification using offline handwritten images remains underexplore. To address this gap, this study proposes a novel multiclass classification approach for assessing dysgraphia severity using offline Malay handwriting samples collected from children aged six to twelve years. A new dataset, the Malay Language Dysgraphia Handwriting Dataset (MLHSD) was developed to capture diverse handwriting characteristics. A robust preprocessing scheme based on a Red-Green-Blue-Black (RGBB) ruling-line removal algorithm was introduced to enhance image quality while preserving essential handwriting features. Subsequently, the Dual-Input Feature Fusion Model (H2FCD) was developed, integrating feature extraction from two complementary inputs: resized handwritten images with removed ruling lines (HIRRL) and histogram-based representations (Histo-HIRRL). These fused feature vectors were then classified across multiple machine-learning classifiers within a multiclass classification. Experimental results showed that the Support Vector Machine (SVM) achieved the best performance, with an accuracy of 99.26%, precision of 99.49%, recall of 99.52%, and F1-score of 99.51%. The robustness of the proposed model was further validated using unseen test data, achieving an accuracy of 88.89% and Matthews Correlation Coefficient (MCC) of 0.839, indicating strong agreement between predicted and actual classes. Additionally, expert validation by eight handwriting specialists yielded a moderately strong correlation coefficient of 0.640 between the model outputs and expert evaluations. These findings demonstrate that the proposed H2FCD model, combined with RGBB-based preprocessing and dual-input feature fusion, significantly improves multiclass classification of dysgraphia severity using offline Malay handwritten images, contributing to more reliable and objective screening tools for early intervention.

ACKNOWLEDGEMENTS

Alhamdulillah, all praise and gratitude be to Allah SWT, the Most Merciful and Compassionate, for granting me the strength, patience, and perseverance to complete this PhD journey. Without His blessings and guidance, this work would not have been possible.

First and foremost, I would like to express my deepest gratitude to my main supervisor, Associate Prof. Ir. Dr. Iza Sazanita Isa, for her invaluable guidance, continuous encouragement, and unwavering support throughout this research. My sincere appreciation is also extended to my co-supervisor, Dr. Ahmad Puad Ismail, as well as Associate Prof. Ir. Dr. Muhammad Khusairi Osman and Associate Prof. Ir. Dr. Zainal Hisham Che Soh, for their constructive feedback, insightful advice, and dedication, which have greatly enriched this work.

I am profoundly grateful to my beloved husband, Abdul Aziz Abdul Muttalib, for his endless love, patience, and encouragement, which have been a constant source of strength throughout this journey. To my dear "AZ" children, Anas Zufar, Alya Nur Zahira, late Asif Zahin and Aariz Zuhair, thank you for being my greatest source of joy and inspiration. My heartfelt appreciation also goes to my beloved parents, Ramlan Supeno and and my parents-in-law, and Abdul Muttalib Ahmad, whose prayers, understanding, and unconditional support have been the foundation of my life.

I would also like to acknowledge Majlis Amanah Rakyat (MARA) for the financial support provided, and Universiti Teknologi MARA (UiTM) for granting me study leave, which made this doctoral study possible.

Last but not least, my sincere appreciation extends to all family members, friends, colleagues, and every individual who has supported me, directly or indirectly, in completing this thesis. Your kindness and encouragement have truly made a difference throughout this journey. Alhamdulillah.

TABLE OF CONTENTS

		Page
CONFIRMATION BY PANEL OF EXAMINERS		**ii**
AUTHOR'S DECLARATION		**iii**
ABSTRACT		**iv**
ACKNOWLEDGEMENTS		**v**
TABLE OF CONTENTS		**vi**
LIST OF TABLES		**xi**
LIST OF FIGURES		**xiii**
LIST OF SYMBOLS		**xvi**
LIST OF ABBREVIATIONS		**xvii**
CHAPTER 1: INTRODUCTION		**1**
1.1	Background of the Study	1
1.2	Problem Statement	4
1.3	Objectives	7
1.4	Scope of Study	7
1.5	Significance of the Study	9
1.6	Thesis Outline	11
CHAPTER 2: LITERATURE REVIEW		**13**
2.1	Introduction	13
2.2	Overview of Dysgraphia	14
2.3	Screening Test for Dysgraphia	18
	2.3.1 Conventional Method	18
	2.3.2 Issues and Challenges of Dysgraphia Assessment	21
	2.3.3 Computer-Aided Method	22
2.4	Intelligence-based Method for Dysgraphia Classification Model	24
	2.4.1 ML-based Application in Dysgraphia Handwritten Classification and Recognition	24

2.4.2 DL-based Application in Handwritten Classification and Recognition 29
2.5 The Importance of Characterized Handwriting Datasets for Intelligence-based Methods 34
2.5.1 Diversity of Handwriting Characteristics 34
2.5.2 Limitations of Existing Dysgraphia Dataset 36
2.6 The Importance of Pre-Processing Techniques in Handwritten Images 38
2.7 Advancement In Feature Extraction Method for Dysgraphia Classification Models 41
2.7.1 Histogram-based Features Extraction 41
2.7.2 DL-based Features Extraction 44
2.7.3 Limitations of Single-Based Feature Extraction Method in Image Classification 46
2.7.4 Related Works on Feature Fusion-based Extraction Method 48
2.8 Classification Performance in Dysgraphia Detection based on Handwriting 50
2.9 Research Gap of Automated Features Extraction Method for Dysgraphia Classification Model 51
2.10 Summary 53

CHAPTER 3: RESEARCH METHODOLOGY 55
3.1 Introduction 55
3.2 Research Design 56
3.3 Phase 1: Development of New Dysgraphia Handwriting Dataset (MLHSD) 58
3.3.1 Data Acquisition 59
3.3.2 Labeling Dataset: Assigning Dysgraphia Severity Levels by PDM's Assessors 63
3.3.3 Dataset Conversion 66
3.3.4 Analysing Dataset using Cohen's Weighted Kappa and Spearman Correlation Coefficient 68
3.4 Phase 2: Proposed New Image Pre-processing Scheme for Removing Ruling Line 71

3.4.1 New Scheme of Ruling Line Removal Using RGBB-based Algorithm 73
3.4.1.1 Process 1: Image Rotation for Alignment 74
3.4.1.2 Process 2: Detection of Ruling Line and Line Replacement 78
3.4.1.3 Process 3: Handwriting Reconstruction 81
3.4.2 Evaluation Performance of the Proposed New Scheme of Ruling Line Removal Using RGBB-based Method 86
3.4.2.1 Evaluation Using Structural Similarity Index Measurement (SSIM) 86
3.4.2.2 Evaluation Using CNN Classification Model 88
3.5 Phase 3: The H2FCD model of Dysgraphia Severity Classification Based on Dual Input Images with CNN-Based Fusion Features 90
3.5.1 Dual-Input Image Preparation 91
3.5.2 Developing the Multiclass Dysgraphia Severity Classification Model (H2FCD) 93
3.5.2.1 Dual-Input Fusion Feature-based for H2FCD Classification Model 95
3.5.2.2 Machine Learning Classifier for H2FCD Classification Model 103
3.5.2.3 Model Training and Validation using k-fold Cross-Validation 105
3.6 Phase 4: Evaluation Performance of Proposed Multiclass Classification Model H2FCD 107
3.6.1 Classification Testing of the Proposed H2FCD Model using Metric Evaluation 107
3.6.1.1 Classification Performance using Metric Evaluation 108
3.6.1.2 Visualization Performance Using ROC-AUC Graph 112
3.6.2 Classification Based on Expert Evaluation 114
3.6.3 Output Analysis Using Spearman's Correlation Coefficient 115
3.7 Summary 116

CHAPTER 4: RESULTS AND DISCUSSION 117
4.1 Introduction 117

4.2 Phase 1: Result and Analysis of New Malay Dysgraphia Handwriting Dataset (MLHSD) 117
4.2.1 Analysis of MLHSD Dataset using Cohen's Weighted Kappa 118
4.2.2 Analysis of MLHSD Dataset Using Spearman's Correlation 118
4.2.3 Data Analysis Results using Pearson's Correlation Analysis 120
4.3 Phase 2: Results and Analysis of New Scheme of Ruling Line Removal Technique using RGBB-based Algorithm 121
4.3.1 Performance Evaluation Using Structural Similarity Index (SSIM) 122
4.3.2 Classification Performance of the Proposed New Scheme of Ruling Line Removal Technique 125
4.4 Phase 3: Training and Validation Performance Analysis of Multiclass Dysgraphia Severity Classification Model (H2FCD) 129
4.4.1 Performance of Multiclass Classification Dysgraphia Severity based on Dual-Input Fusion Feature-based using ML classifier 129
4.4.1.1 Precision, Recall and F1-score 129
4.4.1.2 Accuracy 131
4.4.1.3 Macro-Average AUC Value 132
4.4.1.4 ROC-AUC Graph for SVM 134
4.4.1.5 Result and Analysis of H2FCD Model with Other Model 136
4.5 Phase 4: Performance of Proposed Multiclass Classification Model H2FCD 139
4.5.1 Result and Analysis of Testing Classification Performance using H2FCD Model based on Metrics Evaluation 139
4.5.2 Result and Analysis of H2FCD Model Based on Expert Validation 143
4.6 Summary 147

CHAPTER 5: CONCLUSION **148**
5.1 Conclusion 148
5.2 Research Contribution 149
5.3 Implications for Research 151
5.4 Limitations and Future Work Recommendations 154

REFERENCES **157**
APPENDICES **180**
AUTHOR'S PROFILE **204**

LIST OF TABLES

Tables	Title	Page
Table 1.1	Type of Learning Disability	1
Table 1.2	Summary of Research Scope and Approach	9
Table 2.1	Symptoms of Dysgraphia Across Age Groups Provided by United State National Centre for Learning Disabilities	17
Table 2.2	Established Dysgraphia Screening Tools Using Conventional Method	19
Table 2.3	Dysgraphia Screening Tools Using Computer-Based Approach	22
Table 2.4	List of Previous Studies on Dysgraphia Recognition and Classification Using ML Model	26
Table 2.5	List of Previous Studies on Dysgraphia Recognition and Classification Using DL Model	32
Table 2.6	The Diversity of Dysgraphia Handwriting Characteristics	35
Table 2.7	Summarize of Previous Research on Dysgraphia Dataset For Handwriting Analysis	37
Table 2.8	Similar Studies on Removing Unwanted Entities From Handwriting Images	40
Table 2.9	List of Research Has Been on Dysgraphia Classification Using Feature Fusion-Based Extraction Method	49
Table 2.10	State-of-the-Art Methodology and Performance Comparison with Previous Research	52
Table 3.1	The Number of Participants For Each Age Group	62
Table 3.2	Distribution of Participants by Source, Gender, and Writing Dominance	63
Table 3.3	Dysgraphia Symptoms and Criteria	64
Table 3.4	Dysgraphia severity Levels According to the Error Weight Suggested by PDM	66
Table 3.5	The Number of Images Digitized Form and Cropped Image	67
Table 3.6	Level of Agreement For Weighted Cohen's Kappa Analysis	69

Table 3.7	Level of Spearman's Correlation Coefficient Analysis	70
Table 3.8	Level of Pearson correlation Coefficient Analysis	71
Table 3.9	Structural Similarity Index Analysis	87
Table 3.10	The Dataset Distribution of HIRRL and Histo-HIRRL Employed in the H2FCD Classification Model	95
Table 3.11	Parameter Setting For Residual of Feature Extraction	99
Table 3.12	Parameter Setting For Residual of Feature Extraction	100
Table 3.13	Activations and Number of Learnable Parameters for a New Composite Feature Extractor	102
Table 3.14	Parameter Setting for ResNet-18 of Feature Extraction	103
Table 3.15	Configuration Settings of Machine Learning Classifiers	105
Table 3.16	Data Partition for Each Fold Values	106
Table 3.17	Multiple Classes of Confusion Matrix	109
Table 3.18	List of Dysgraphia Severity Level for Each Category	109
Table 3.19	Interpretation of the Term Used in Confusion Matrix	110
Table 4.1	Agreement Analysis of Cohen's Weighted Kappa for Two Assessor Evaluation	118
Table 4.2	Spearman's Correlation Coefficient for Two Assessor Evaluations	119
Table 4.3	Correlations Between Age and Dysgraphia Category	120
Table 4.4	Correlations Between Gender and Dysgraphia Category	121
Table 4.5	Analysis of SSIM Results Based on Different Measurements	123
Table 4.6	Performance of Precision, Recall, and F1-Score for Dysgraphia Severity Classification for HIRL and HIRRL	128
Table 4.7	F1-Score Value for All Types of Image Feature Vectors Across Machine Learning Models	130
Table 4.8	Comparison of Existing Dysgraphia Detection Methods and the Proposed H2FCD Model Based on Feature Extraction, Classification, and Performance	137
Table 4.9	Precision, Recall and F1-Score for Each Class and Macro-Average Measurement	141
Table 4.10	Example of Misclassified Handwriting Images	142
Table 4.11	Results Image Categories Based on Severity Levels	144

LIST OF FIGURES

Figures	Title	Page
Figure 1.1	Number of MBPK by Program From 2020 to 2024	2
Figure 2.1	Example of Dysgraphia Handwriting in (a) Kanji and Kana (Left: Normal, Right: Dysgraphia)(b) Devagari Handwriting (c) Malay Language Handwriting	15
Figure 2.2	An Example of a Histogram Projection To Represent the Frequency of Pixel Intensity	42
Figure 3.1	Research Design For An Automated Dysgraphia Severity Classification Model	57
Figure 3.2	Overall Process of the Dysgraphia Dataset Development	59
Figure 3.3	A Sentence Produced By A Participant in the Writing Task Form	63
Figure 3.4	Evaluation Process For Assigning Severity Levels of Dysgraphia Symptoms Present in the Handwritten	64
Figure 3.5	Digitization Process	66
Figure 3.6	The Cropping Process of the Handwriting Sample	67
Figure 3.7	Handwritten Image With Ruling Line (HIRL)	72
Figure 3.8	Image Pre-Processing for Ruling Line Removal	73
Figure 3.9	Original image with Axis Misalignment	73
Figure 3.10	Overall Process For Image Rotation For Alignment in Proposed Scheme of Ruling Line Removal Using RGBB-Based Algorithm	75
Figure 3.11	Image Rotation Direction of +N ° or -N°	77
Figure 3.12	Overall Process For the Detection of Ruling Line and Line Replacement in the Proposed Scheme of Ruling Line Removal Using RGBB-Based Algorithm	79
Figure 3.13	RGB Vertical Histogram Based on the Rotated Image	80
Figure 3.14	Image After the Ruling Line Color Is Replaced with the Background Color	81

Figure 3.15	Flowchart of Pixel-Level Reconstruction From Neighbouring Pixel Reference	82
Figure 3.16	A Gap Was Detected in the Handwriting	83
Figure 3.17	After the First Straight Pixel Was Reconstructed	84
Figure 3.18	Process of Morphological Dilation And Erosion For Pixel Restoration	85
Figure 3.19	Classification of Dysgraphia Severity using HIRL and HIRRL as Input Images	89
Figure 3.20	Dysgraphia Severity Classification Process Based on HIRL and HIRRL	90
Figure 3.21	Phase 3 Process of Developing the Proposed H2FCD Classification Model	91
Figure 3.22	Handwriting Image with Removed Ruling Line (HIRRL)	92
Figure 3.23	Histogram Image (Histo-HIRRL) Preparation Process	93
Figure 3.24	The Framework of the Proposed H2FCD Classification Model	94
Figure 3.25	Feature Extraction Process for Both Input Based on Raw Image (HIRRL) and Historam Image (Histo-HIRRL)	96
Figure 3.26	The Feature Extractor Layer Structure of the Residual Architecture For HIRRL Uses in the Feature Extractor	98
Figure 3.27	A New Feature Extractor Layer Structure of 1 Block Residual Network for Histo-HIRRL Image Feature Extraction	101
Figure 3.28	Classification Performance Evaluation for the Dysgraphia Classification Model	107
Figure 3.29	Testing of the Proposed H2FCD Model With New Data Images	108
Figure 3.30	Examples of ROC Curves Based on the Model's Performance	113
Figure 3.31	An Example of a Handwriting Image for Expert Observation and Severity Classes	114
Figure 3.32	Output Analysis of the proposed H2FCD Classification Model Using Spearman's Correlation Coefficient	115
Figure 4.1	The Average Results of SSIM for Each Category of Dysgraphia Class	124

Figure 4.2	Analysis of the SSIM Range Value with the Number of Images Presented in Histogram	125
Figure 4.3	Accuracy Performance of Dysgraphia Severity Classification Based on HIRL and HIRRL	126
Figure 4.4	Accuracy Performance of Each Fold and Classifier Based on Image Features	132
Figure 4.5	Macro-Average AUC Performance of Classifiers Using Image, Histogram, and Fusion Feature Vectors	133
Figure 4.6	ROC-AUC Graph for SVM Model Using Image Features	134
Figure 4.7	ROC-AUC Graph for SVM Model Using Histogram Features	135
Figure 4.8	ROC-AUC Graph for SVM Model Using Fusion Features	136
Figure 4.9	Confusion Matrix of SVM Classifier with Fusion Feature Vector	140
Figure 4.10	Testing ROC-AUC Curve for SVM Classifier with Fusion Feature	142
Figure 4.11	Correlation Coefficient of Classification Output Class Between H2FCD Model and Expert Evaluation	146

LIST OF SYMBOLS

Symbols

α	Cronbach's Alpha (Reliability)
p	Statistically Significant Value
ρ	Spearman Correlation Coefficient
κ	Cohen's Weighted Kappa Coefficient
r	Pearson's Correlation Coefficient

LIST OF ABBREVIATIONS

Abbreviations

BOLD	Association for Children with Special Needs
BT	Bagged Tree
CCA	Connected Component Analysis
CNN	Convolutional Neural Network
CO_e	Classification Output from Expert Evaluation
CO_p	Classification Output from Proposed H2FCD Model
DL	Deep Learning
FPR	False Positive Rate
H2FCD	Handwriting and Histogram Features for Classifying Dysgraphia Severity
HIRL	Handwriting Image with Ruling Line
HIRRL	Handwriting Image with Removed Ruling Line
Histo-HIRRL	Histogram of Handwriting Image with Removed Ruling Line
HR	High-Risk
ISD	Senarai Semak Disleksia
KNN	K-Nearest Neighbor
LBP	Local Binary Patterns
LD	Learning Disabilities
LINUS	Literacy and Numeracy Screening
LSTM	Long Short-Term Memory

MCC	Matthews Correlation Coefficient
ML	Machine Learning
MOE	Ministry of Education
MR	Moderate-Risk
NGO	Non-Governmental Organizations
NLP	Natural Language Processing
OCR	Optical Character Recognition
PDM	Persatuan Disleksia Malaysia
REC	Research Ethics Committee
ResNet	Residual Network
RF	Random Forest
RNN	Recurrent Neural Network
RGBB	Red-Green-Blue-Black
ROC-AUC	Receiver Operating Characteristic – Area Under Curve
SDG	Sustainable Development Goal
SGDM	Stochastic Gradient Descent with Momentum
SLD	Specific Learning Disabilities
SMOTE	Synthetic Minority Oversampling Technique
SSIM	Structural Similarity Index Measure
SVM	Support Vector Machine
TPR	True Positive Rate
ViT	Vision Transformer

CHAPTER 1
INTRODUCTION

1.1 Background of the Study

The term learning disabilities (LD) includes a range of specific conditions, including dyslexia, dysgraphia, and dyscalculia, each of which presents distinct characteristics and learning challenges, as summarized in Table 1.1. These conditions significantly impact children, who frequently necessitate specialized educational interventions to facilitate their learning processes.

Table 1.1
Type of Learning Disability (Muktamath et al., 2021)

Type of Learning Disability	**Description**	**Affected Skill**	**Common Symptoms**
Dyslexia	A learning disorder characterized by difficulty in reading, particularly in word recognition and decoding	Reading	Difficulty recognizing words, slow reading, poor spelling, confusion with letters, trouble understanding text
Dysgraphia	A disorder that affects writing abilities, including handwriting, spelling, and organization of ideas	Writing	Poor handwriting, inconsistent spacing, difficulty forming letters, frequent spelling errors, trouble organizing thoughts in writing
Dyscalculia	A disorder that affects the ability to understand numbers and mathematical concepts	Mathematics	Difficulty understanding numbers, trouble with basic arithmetic, inability to recognize patterns, poor time and money management

The range of learning disabilities summarized in Table 1.1 contributes to the growing number of identified cases, as evidenced by the increasing trend of students with learning disabilities presented in Figure 1.1. According to Special Education Data (2024) from the Ministry of Education (MOE), the number of students identified with learning disabilities has increased markedly from 77,205 in 2020 to 113,059 in 2024 (Kementerian Pendidikan Malaysia, 2024). This trend highlights the crucial need for an early screening for intervention purposes (Vlachos & Avramidis, 2020). Children with

LD commonly struggle to focus on basic reading, writing, and counting (3M) skills. More importantly, writing as a multifaceted skill has presented significant complexity and has a profound influence on academic achievement (Šafárová et al., 2020). Consequently, early screening for dysgraphia is crucial to enable timely identification and the planning of appropriate interventions based on the severity of the handwriting difficulties (Alhusaini et al., 2021).

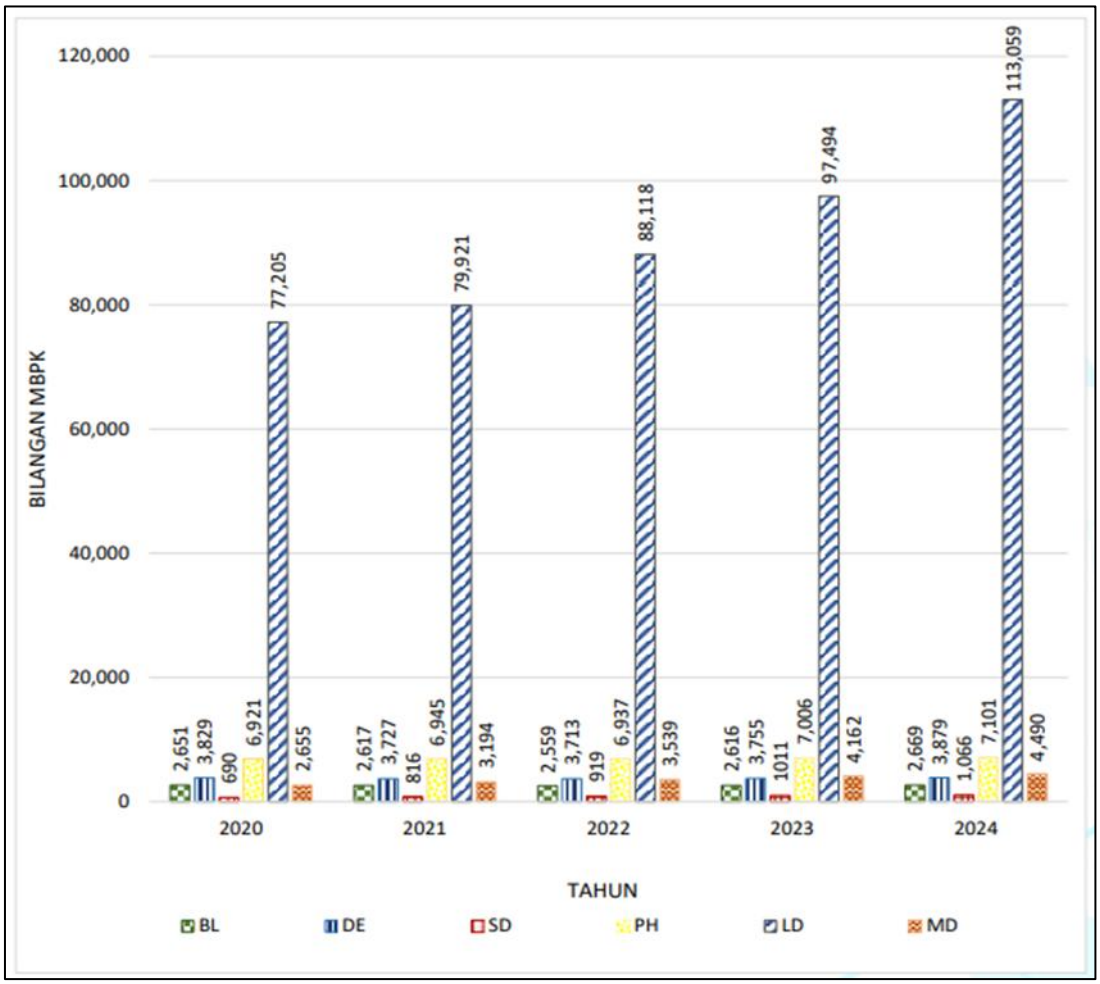


Figure 1.1 Number of MBPK by Program From 2020 to 2024 (Kementerian Pendidikan Malaysia, 2024)

Dysgraphia is conventionally identified during a screening session by following the prescribed protocol, i.e, with the presence of an expert or assessor to observe the candidate using scoring systems (Alhusaini et al., 2021; Šafárová et al., 2020; Vlachos & Avramidis, 2020). The screening sessions take place during a teaching and learning session and contain a number of examinations to ascertain a child's cognitive strengths and limitations. Numerous studies have aimed to examine the assessment process

through the use of screening test tools (Franken & Harris, 2021) questionnaires (Katarína Šafárová, Jiří Mekyska, 2021), and scoring systems (Rosenblum, 2018). There are numerous screening techniques for dysgraphia symptoms that have been researched by Van Waelvelde et al. (2012) and Hamstra-Bletz & Blöte, (1993). However, the techniques are exceedingly time-intensive, as teachers must constantly observe and analyze potential dysgraphia (P. Chung & Patel, 2015). Furthermore, this conventional assessment strategy necessitates the need for an expert to administer the assessment to children, and there is a significant probability that the evaluation process will be disrupted due to distractions or inability to pay sufficient attention. In addition, many researchers (Asselborn et al., 2018; Dui et al., 2020, 2021; Katarína Šafárová, Jiří Mekyska, 2021) have developed computer-based systems to enhance dysgraphia assessment approaches.

With the advancement of technology, research has increasingly focused on developing computerized tools and screening tests for dysgraphia. However, further investigations into writing tasks are necessary to enhance detection methods, particularly those utilizing artificial intelligence (Galli et al., 2019;Dimauro et al., 2020). In this context, machine learning (ML) has demonstrated considerable potential to contribute to dysgraphia detection, particularly within the specialized field of impairment assessment and assistive technology. Researchers have explored this ML system using a variety of inputs and methodologies for implementation across both offline and online handwriting domains (Jose & Kumar, 2021; Seki, 2020; Wong & Loh, 2019). Offline handwriting involves interpreting and converting handwritten text from physical paper into digital form through a manual scanning process (Jafarzadeh et al., 2024), while online handwriting is a digitized handwriting that captures digital pen movement, including sequential information, such as position, velocity, and acceleration (Wong & Loh, 2019). To date, only limited studies have explored the detection of dysgraphia using offline handwriting datasets, particularly in word or sentence form, based on machine learning models. Therefore, there is a high need to explore the detection of dysgraphia using offline handwriting datasets by implementing an ML model for better preparation of interventions for children.

Over the past few decades, there has been a significant increase in research on handwritten image classification using machine learning algorithms (W. Liu et al., 2020; Ahlawat & Choudhary, 2020; Darmatasia and Mohamad Ivan Fanany, 2017; Deore & Pravin, 2019). However, only a few studies have investigated the classification

of handwritten dysgraphia images across multiple classes and identified the most effective machine learning model. To address this gap, this study focuses on classifying dysgraphia severity levels using a CNN-based fusion feature vector, derived from dual-input images, in conjunction with a machine learning classifier. Hence, advancements in artificial intelligence technology continue to stimulate future studies in dysgraphia classification through improved feature extraction with precise pre-processing techniques and dataset preparation. Prior research suggests that there is a strong potential for classifying dysgraphia severity based on handwritten images using CNN-based feature extraction from multiple input images and image processing techniques. This research aims to enhance handwritten analysis technologies for special education to support early childhood development. This comprehensive approach is crucial for developing reliable tools that can facilitate the early identification of children with learning difficulties, with the goals of improving educational outcomes.

1.2 Problem Statement

The dysgraphia screening process is crucial for identifying the characteristics of dysgraphia symptoms and enabling prompt intervention that can significantly enhance the academic achievement of these children (Biotteau et al., 2019). However, the conventional screening approach requires experts to conduct assessments using observational and scoring method (Alhusaini et al., 2021; Šafárová et al., 2020; Vlachos & Avramidis, 2020), which are time-consuming, and may yield variable results depending on the experience of the assessors or screening experts (Raghuram et al., 2023). Therefore, with the recent increase in dysgraphia cases, conventional screening method has become increasingly difficult to address all screening candidates. This limitation has led to the development of computer-assisted screening systems to detect the severity levels of dysgraphia symptoms (Seki, 2020; Wong & Loh, 2019).

Datasets are essential for developing machine learning models, as the quality and characteristics of the dataset directly influence model performance. This is particularly important in dysgraphia handwriting analysis where the dataset may consist of either offline samples (paper-based, converted into digital form) (Jose & Kumar, 2021; Y. Wang et al., 2021) or online samples (captured in real-time through digital devices) (Seki, 2020; Wong & Loh, 2019). Although online handwriting features are commonly used for classifying dysgraphia severity (Ammour et al., 2021; Chakraborty

et al., 2020; Corbille et al., 2020; Karyna Korovai, 2020), compared to offline handwriting, currently, the dataset used for classifying dysgraphia handwriting severity is specific, as it only covering limited target groups of children with limited handwritten diversity, especially in the Malay language. To date, only non-image representations of handwritten datasets are publicly available in online repositories for research purposes (Drotár & Dobeš, 2020; Gemelli et al., 2023). Consequently, there is a need to establish a dataset of dysgraphia handwritten images in Malay to be employed in severity classification models.

Furthermore, the processing of handwritten images for classification models remains highly challenging. A critical step involves removing irrelevant elements such as ruling lines (Alipour et al., 2013), tables (Rani et al., 2021), and overlapping strokes (Mo & Zhang, 2022; Puri & Singh, 2019) which significantly affect model performance. However, existing removal techniques often introduce new issues, including the unintended erasure of handwriting content or the presence of residual artifacts that reduce recognition accuracy. Moreover, many current approaches are designed for structured layouts (Rani et al., 2021), and are not robust to handle varied or skewed handwritten documents. Progressive line removal methods may also eliminate essential handwriting features, further degrading performance (Mo & Zhang, 2022; Puri & Singh, 2019). These limitations are particularly critical for dysgraphia data, where handwriting is inherently inconsistent and fragile. Therefore, there is a critical need for a more robust ruling line removal approach that can effectively eliminate irrelevant elements while preserving important handwriting features and improving recognition reliability across diverse datasets (Rakesh et al., 2024). The process is aimed at improving the image quality and attaining the specific target conditions without losing any important information (Rakesh et al., 2024). Such preprocessing is essential to enhance image quality, retain meaningful information, and enable more accurate feature extraction for classification tasks.

The development of automated handwriting severity classification using CNN is highly crucial, as this could serve as a screening tool and a benchmark assessment for educational purposes. However, dysgraphia handwriting remains challenging due to its high variability and inconsistencies (P. Chung & Patel, 2015) which complicate feature extraction in CNN-based architectures (Devi & Kavya, 2023; Ghouse et al., 2022; Kunhoth, Al Maadeed, et al., 2023; Masood et al., 2023). Although deep feature extraction methods have been widely applied (Ahuja, 2020; Drotár & Dobeš, 2020;

Richard & Serrurier, 2020; Sihwi et al., 2019), the diversity of approaches indicates a need for more effective techniques to generate robust feature vectors, particularly for multiclass classification. Distinguishing multiple levels of dysgraphia severity requires capturing subtle and complex handwriting variations, which are often inadequately represented by single feature extraction methods. While several studies have explored alternative approaches (Biotteau et al., 2019; Danna et al., 2023; Sukiman et al., 2024; Vydeki et al., 2024), most rely on singular feature representations. In contrast, integrating multiple features from diverse pre-processed handwriting images has shown potential to provide richer information and improve classification accuracy (Alqahtani et al., 2023; Kunhoth, Al Maadeed, et al., 2023; Kunhoth et al., 2024; Masood et al., 2023) However, fusion features derived from dual-input images remain minimally explored in multiclass handwriting classification. For instance, Kunhoth et al., (2024) utilized various handwritten forms but was limited to binary classification and synthetic offline data, restricting real-world applicability. Therefore, there is a clear need to develop effective fusion-based feature extraction techniques for multiclass dysgraphia severity classification using offline handwritten data to improve classification performance.

Despite significant advancements in applying machine learning techniques for dysgraphia detection through handwriting analysis, evaluating classification performance using appropriate methodologies remains a challenge. Existing studies predominantly rely on model-level metrics such as accuracy and confusion matrices (Devi & Kavya, 2023; Ghouse et al., 2022; Kunhoth, Al Maadeed, et al., 2023; Kunhoth et al., 2024; Kunhoth, Maadeed, et al., 2023; Masood et al., 2023), with limited emphasis on assessing the statistical validity and reliability of the results. This narrow evaluation approach introduces important limitations. In particular, the use of the Matthews Correlation Coefficient (MCC) for performance validation is still restricted, as current applications are mainly confined to binary classification scenarios (Devi & Kavya, 2023). Consequently, MCC has not been adequately explored for multiclass dysgraphia severity classification. This gap highlights the need for more comprehensive evaluation strategies to ensure robust and reliable performance assessment. Furthermore, post-evaluation expert validation is essential to verify the significance and practical relevance of the classification outcomes.

Offline handwriting represents the primary form of written work produced during daily classroom activities. Therefore, the source of data is a significantly

practical and accessible source for assessing dysgraphia. Given the need for robust image preprocessing, removal ruling lines is crucial to produce significant features that can be extracted from high-quality handwriting images. Consequently, the effective process of feature extraction has become critical consideration in generating significant feature vectors. Thus, this study focuses on generating feature vectors by fusing features derived from dual-input images to classify the multiclass severity of dysgraphia.

1.3 Objectives

This study aims to classify the severity levels of dysgraphia in handwriting using a CNN-based feature extraction method with dual-input images and a machine learning classifier. The aim of this study can be accomplished through the following objectives:

i) To develop a new Malay Language Dysgraphia Handwriting Dataset (MLHSD) produced by school-age children.
ii) To formulate a new scheme for image pre-processing aimed at removing ruling lines from MLHSD images using the Red-Green-Blue-Black (RGBB)-based method.
iii) To model an enhanced multiclass classifier by leveraging the fusion of a CNN-based feature vector derived from dual-input of Handwriting Image with Removed Ruling Line (HIRRL) and Histogram Image of HIRRL (Histo-HIRRL).
iv) To evaluate the performance of the proposed multiclass classification model for dysgraphia severity levels metric-based evaluation, including confusion matrix, correlation analysis, and expert validation.

1.4 Scope of Study

Research scopes are essential to ensure that the research work is rational within the expected period. This study aims to classify the severity levels of dysgraphia using handwriting images by proposing a CNN-based model with fusion feature technique derived from dual-input images. The handwriting image dataset is highly important for developing a classification model for this study. To support the intended purpose of

the model a control group was included in the data acquisition process, which involved collecting data among school-aged children with dysgraphia symptoms aged between six to twelve years old, comprising a random selection of male and female students. A total of 178 handwriting samples were collected, consisting of 119 male and 59 female students. The handwriting samples selected for this study were subject to compulsory diagnostic assessment through observation by assesors from the Persatuan Dyslexia Malaysia (PDM). Data collection was conducted at PDM and primary schools under the Ministry of Education (MOE), Malaysia. Furthermore, this study was conducted using dysgraphia handwriting samples, which were limited tp the data accessible through PDM and MOE; however, the available data is still sufficient to achieve the research objectives. Moreover, the methodological approach used in this study was inspired by the manual method used by PDM's assessors in evaluating the severity levels of dysgraphia.

In this study, two main tools were used to accomplish the research objectives, related to designing a classification model of dysgraphia severity levels based on MLHSD images. One of the tools is Deep Learning Toolbox in MATLAB version R2023a which was used to develop classification model of dysgraphia severity levels. This tool provides a framework for designing and implementing deep neural networks with algorithms, machine learning algorithms and image processing techniques. Next, the data are analyzed using SPSS tool, which includes correlation and reliability measurements.

This research also focuses on classifying dysgraphia severity levels using handwriting image samples that consist of several sentences in the Malay Language. The study involves detecting and extracting feature vectors from selected sentences using an adopted Residual Network (ResNet) architecture as a deep feature extraction network. The extracted features are then classified using multiple machine learning classifiers, namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Bagged Trees (BT), and Random Forest (RF). Then, the model was trained, validated, and tested to determine its performance in terms of accuracy, precision, recall and F1-score. The testing classification performance is then validated using Matthew Correlation Coefficient (MCC). In addition, the model's performance is evaluated through expert validation involving eight experts who visually assess the processed images and assign severity ratings based on their professional expertise. The correlation analyses conducted in this study included Weighted Cohen's Kappa (WCK), Pearson's

Correlation Coefficient (PCC), Spearman's Correlation Coefficient (SCC) and MCC. The mean and standard deviation were reported to present the overall data analysis. Regardless of the study's depth, the accomplishment of this research in developing a classification model is highly significant and draws attention to this area of study. Table 1.2 summarizes the research scope with all components considered in this study. The significance of this study is presented in the following subsection.

Table 1.2
Summary of Research Scope and Approach

Component	**Description**
Dataset Size	178 Malay handwriting samples (119 male and 59 female)
Age Range	6–12 years old
Data Source	PDM and MOE primary schools
Preprocessing	Dual-input images (HIRRL and Histo-HIRRL)
Feature Extraction	Adopted ResNet architecture
Classification Methods	SVM, KNN, Bagged Trees (BT), Random Forest (RF)
Output Classes	High Risk, Moderate Risk, Low Risk
Evaluation Metrics	Accuracy, Precision, Recall, F1-score
Validation Method	MCC, Cohen's Kappa, Pearson, Spearman
Expert Validation	8 experts (visual assessment)
Tools Used	MATLAB (R2023a), SPSS

1.5 Significance of the Study

The classification of dysgraphia severity models enables users to aid in the screening process for detecting dysgraphia using children's handwriting images. This research aims to achieve these objectives through a comprehensive study with the goal of enhancing the performance of severity classification. Data collection, pre-processing, and model formulation processes are all critical steps in categorizing dysgraphia-related handwriting images. As a result, to obtain high performance and validity, an effective flow of the classification model is necessary. This study makes several notable contributions, as listed below:

a) Research on dysgraphia handwriting characteristics; specifically in Malay language; improving previous studies on how conventional methods, computer-assisted tools, and computing technology, such as machine learning, will supplement the understanding of dysgraphia detection, particularly in the using children's handwriting images as data input.
b) The data collection process consisting of several sentences contribute to a new database of dysgraphia handwriting images by school-aged children, which has the potential to be used for further research.
c) Feature extraction techniques and feature extraction algorithms for the offline handwriting may help the subsequent image classification process in addressing the imbalanced dataset in each class of severity level.
d) The employment of image processing techniques contributes to the feature extraction stage and supports the classification process during the training stage of model formulation.
e) A new automated machine learning-based model that can be used to assist in diagnosing dysgraphia severity in children based on their handwriting.
f) The finding of this research is based on the performance of the model in terms of accuracy, precision, and F1 score for classifying potential dysgraphia severity based on handwriting images. The classification performance was validated using MCC and the validity of the proposed model classification output through expert evaluation using statistical correlation coefficients.
g) The research design, sentence detection scheme, feature extraction algorithm, machine learning model formulation, research processes, and dataset development procedures can be utilized as references for developing any related deep learning model.

Consequently, this research may contribute to the development of improved dysgraphia screening approach. Furthermore, the findings of this research can be used in the Child Development department in hospitals as well as by the MOE or Non-Government Organizations (NGO) such as PDM and the Association for Children with Special Needs (BOLD) to support children with special needs. Since a computer-based system is recommended to be used with a screening test approach, the proposed study ensures fast, low error rate, and accurate diagnosis. Hence, the diagnosis outcome may facilitate the implementation of an intervention program.

1.6 Thesis Outline

This thesis is divided into five chapters, each outlining a key stage of the research. Chapter 1 introduces the background, problem statements, research aim, objectives, scope, and the significance of the study.

Chapter 2 reviews the previous and current studies related to this research study. This chapter also provides an overview of dysgraphia, including its definition, characteristics of dysgraphia handwriting, and examples of dysgraphia handwriting across various languages. The review continues with a conventional screening test for dysgraphia detection, including the issues and challenges associated with computer-aided approach for dysgraphia detection. The chapter then examines intelligence-based methods, including ML and deep learning (DL) implementation. This section explains about dysgraphia dataset used in ML and the limitations of existing dysgraphia dataset available on online platforms. The review also presents the pre-processing techniques involved in handwriting images, focusing on removing unwanted elements from the images. The feature extraction techniques are then explored by presenting the automated method of extraction, such as DL and fusion feature-based techniques used in previous studies. The limitation of current feature extraction strategies are also compared, particularly between single and fusion based techniques. The final section addresses the research gap in the extraction methods for dysgraphia handwriting classification model. The chapter concluded with a summary of the overall review conducted throughout the chapter.

Chapter 3 outlines the research methodology employed to develop a classification model for dysgraphia severity, utilizing combined features from dual-input handwriting images. This chapter includes the performance analysis approach and underlying assumptions. All four phases of research activities that align with the research objectives, including the performance measurements and data analysis are also presented in this chapter. The first phase covers Objective 1 after obtaining ethical approval, followed by the pre-processing phase, which involved implementation of image processing techniques, as presented in Phase 2. Meanwhile, the development process for classification model is explained in Phase 3. The final phase presents the confusion metric and statistical measure of correlation for model evaluation performance.

Chapter 4 presents and discusses the results and findings, organized according to the three main phases of the methodology: dataset development, pre-processing techniques, and classification and validation processes. In Phase 1, the statistical analysis of the correlation between two PDM assessors has been analyzed, and the correlation between severity and age and gender is also presented in this chapter. Meanwhile, Phase 2 presents the results of ruling line removal using the RGBB-based method, as measured by the Structural Similarity Index Measurement (SSIM). At the end of this, the classification performance between removed ruling line images and ruling line images is evaluated to observe the impact of the ruling line on classification performance. In Phase 3, the analysis of classification performance is presented to identify the best model in classifying the dysgraphia handwriting severity using confusion matrix measurements. In final phase, the testing classification performance is analyzed using confusion matrix and Matthew correlation coefficient (MCC). The validation of model performance is then analyze using Spearman's correlation coefficients to measure the correlation between classification output by proposed model and expert evaluation.

Finally, Chapter 5 summarizes the overall study by outlining the contributions of this research and suggesting possible directions for future work. In addition, this thesis outlines the comprehensive process of developing an automated dysgraphia severity classification system that utilizes handwriting images and combines feature vectors. The following chapters provide detailed discussions on the literature review, methodology, results, and conclusions. As a summary, this thesis concludes the processes involved in developing an automated dysgraphia severity classification system based on handwriting images including the utilization of dual-input images and the extraction and analysis of fused features. The following chapter 2 provides a more detailed discussion of the literature review, methodology, results, discussion, and conclusion.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Handwriting skills plays a crucial role in children educational development. However, children who exhibit the presence of dysgraphia symptoms may experience difficulties with handwriting. Dysgraphia significantly impacts children's writing abilities and is characterized by various writing challenges, including inconsistent letter formation, inadequate spacing, and disorganized strokes. The aforementioned issues have the potential not only to reduce the legibility of handwriting but also impact children's self-esteem, academic performance, and social engagement. In Malaysia, the number of children with learning disabilities including writing difficulties has been increasing annually (Kementerian Pendidikan Malaysia, 2024). Furthermore, conventional methods for assessing dysgraphia symptoms mainly rely on observations and evaluations conducted by educators, therapists, or specialists, who analyze handwriting samples (Danna et al., 2023). Although these methods are useful, they are time-consuming and often subjective, i.e different evaluators may render varying judgments. This creates a need for automated systems that can analyze handwriting more efficiently and objectively without compromising the accuracy. In the last decade, researchers have explored various approaches to dysgraphia detection, including the use of online and offline handwriting data (Wong & Loh, 2019), image processing (Sebastian et al., 2024), and machine learning (Dinusha et al., 2024; Sukiman et al., 2024; Yogarajah & Braj, 2020). Although deep learning and feature fusion have resulted in improved accuracy (Kunhoth, Al Maadeed, et al., 2023; Kunhoth et al., 2024; Masood et al., 2023), challenges remain including limited datasets, language dependence, and the absence of multiclass severity classification.

This chapter reviews related works in the field of dysgraphia detection. The review begins with an overview of dysgraphia, including its definition, symptoms, and impact on learning, and further discusses the conventional approaches for handwriting evaluation. Next, a review of online and offline handwriting datasets is presented, along with image processing and feature extraction techniques used in handwriting analysis. The chapter also highlights machine learning and deep learning approaches applied in

previous studies. Finally, the research gaps and limitations identified in existing works are summarized to provide the foundation for this study.

2.2 Overview of Dysgraphia

Specific Learning Disabilities (SLD) or Learning Disabilities (LD) is defined as a condition that affects the brain's ability to process, analyze, and understand information. A person with SLD faces challenge in acquiring certain academic skills, such as reading, writing, and mathematics (Dzalani & Shamsuddin, 2014). One form of the LD is dyslexia that causes problems in reading and comprehension. Meanwhile, other LDs such as dysgraphia impairs writing skills while dyscalculia affects understanding of mathematical concepts (Grigorenko et al., 2020). Although these LDs are often identified during childhood, these conditions frequently continue into older age. Therefore, research emphasizing early recognition of learning disabilities is highly important for informing customized educational strategies and interventions to assist students who are affected in their schooling (Cakiroglu, 2015). Across all LDs, dysgraphia is one of the learning disabilities that has received high attention in scholarly research. Many have used the term handwriting difficulties (Biotteau et al., 2019) or slow writing performance (Prunty & Barnett, 2017) among children to interpret dysgraphia. More specifically, dysgraphia has been characterized by lack of writing skills below expectation for their age and cognitive levels (P. J. Chung et al., 2020). In addition, students with dysgraphia who experience impairments spelling and writing skills (Dimauro et al., 2020; Katarína Šafárová, Jiří Mekyska, 2021; Vlachos & Avramidis, 2020) are often struggle to write words quickly and consistently. Therefore, dysgraphia impacts both the handwriting product (legibility of the written trace) and the writing process (movement that generates the trace) (Biotteau et al., 2019; Prunty & Barnett, 2017).

An individual with dysgraphia may have produced a poor handwritten product irrespective of language or handwriting characters. For example, the illegibility of handwritten products could be seen through the handwriting samples produced by children with dysgraphia, as presented in Figure 2.1. Another example illustrated Japanese character exhibiting dysgraphia symptoms as presented in Figure 2.1 (a). The left side shows normal kanji and Kana writing samples, whereas dysgraphia handwriting is shown on the right. Devanagari handwriting demonstrating dysgraphia

is presented in Figure 2.1 (b) while dysgraphia handwriting in the Malay language is shown in Figure 2.1 (c).d

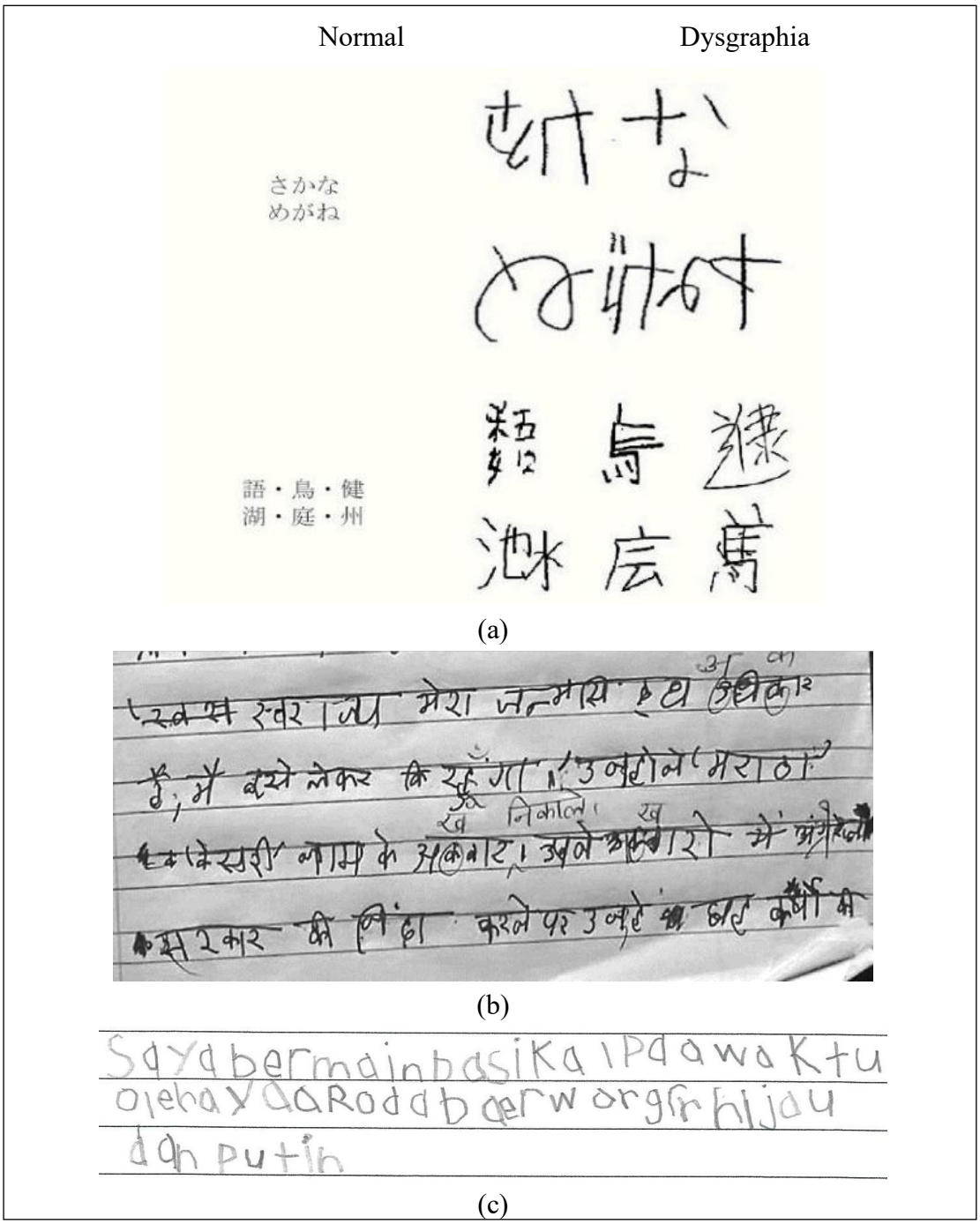


Figure 2.1 Example of Dysgraphia Handwriting in (a) Kanji and Kana (Left: Normal, Right: Dysgraphia)(Dyszy-Chudzińska, 2009)(b) Devagari Handwriting (Mulakaluri & Girisha, 2021) (c) Malay Language Handwriting

Handwriting characteristics vary across languages due to differences in script structure, stroke order, and visual complexity. For instance, Chinese characters are inherently logographic and compose of multiple radicals arranged in precise spatial

configurations, following strict rules for stroke sequence and layout (Z. Xu et al., 2025). In contrast, alphabetic scripts such as Latin rely on simpler graphemes where variations in slant, spacing, and letter formation are more common. Meanwhile, Arabic handwriting introduces additional complexity through its cursive nature, directionality (right-to-left), and multiple stylistic traditions, such as Naskh, Ruq'ah, and Diwani. Each style has unique stroke patterns and flow characteristics (Babushkin et al., 2024). Similarly, scripts such as Devanagari present distinctive challenges due to their horizontal headline (Shiro Rekha) and complex conjunct characters, which influence how handwriting difficulties are expressed and detected in children (Mulakaluri & Girisha, 2021; Patnaik et al., 2020; Puri & Singh, 2019). The same issue is further highlighted by Japanese writing systems, as the combination of kanji, hiragana, and katakana requires different cognitive and motor processing strategies, which have been shown to result in reading and writing disorders, such as dyslexia-dysgraphia (Dyszy-Chudzińska, 2009). These linguistic and stylistic differences directly impact the visual and kinematic features of handwriting, since effective analysis methods for one language may not generalize across others. Therefore, language-specific features and considerations are crucial for designing robust handwriting analysis systems, particularly for clinical applications such as the detection of dysgraphia. The features that represent the handwriting characteristic have a significant influence on identifying the symptoms present in handwritten images. These symptoms could be detected by analyzing the abnormal characteristics that may occur in dysgraphia handwriting products.

Typically, dysgraphia handwriting results in reversal letters (Biotteau et al., 2019; Fischer & Luxembourger, 2018; Katarína Šafárová, Jiří Mekyska, 2021), difficulty remembering how to shape letters (Biotteau et al., 2019; Prunty & Barnett, 2017), difficulty forming written sentences with proper grammar and punctuation (Biotteau et al., 2019), frequent omission of words, incorrect word sequence, incorrect verb and pronoun usage, and word ending mistakes (McCloskey & Rapp, 2017). Furthermore, as illustrated in Table 2.1, numerous additional indicators of dysgraphia have been identified across all age groups as documented in handwriting assignment guidelines provided by the United States National Center for Learning Disabilities (P. J. Chung et al., 2020).

Table 2.1
Symptoms of Dysgraphia Across Age Groups Provided by United State National Centre for Learning Disabilities (P. J. Chung et al., 2020)

Age Group	**Symptoms**
Pre-school children	1. An awkward grip or body position when writing. 2. Tire easily with writing. 3. Avoidance of writing and drawing tasks. 4. Written letters are poorly formed, inversed, reversed, or inconsistently spaced. 5. Difficulty staying within margins.
The school-age child	1. Illegible handwriting. 2. Switching between cursive and print. 3. Difficulty with word-finding, sentence completion, and written comprehension.
The teenager and young adult	1. Difficulty with written organization of thought. 2. Difficulty with written syntax and written grammar that is not duplicated with oral tasks.

According to prior research (Biotteau et al., 2019; Katarína Šafárová, Jiří Mekyska, 2021; Rosenblum et al., 2010), dysgraphia handwriting can be classified into two categories of characteristics :

a) Paper-based/offline handwriting – letter formation that affects readability such as missing or additional strokes, incorrect orientation, inversed, reversed, improper spacing, not consistent size, not able to follow the lines, letter erased or overwritten.
b) Real-time captured writing/online handwriting – writing process that is caused by poor control of handwriting movements such as inappropriate pencil grip or body position, velocity, pressure, easily pain during writing, hand posture, and pen tilt.

Based on this classification of handwriting types, studies on handwriting analysis can be divided into two main categories: online and offline handwriting. Although many studies have employed online handwriting samples (Ammour et al., 2021; Carbune et al., 2020; Corbille et al., 2020; Karyna Korovai, 2020; Seki, 2020; Wong & Loh, 2019), the results are contradictory to those studies that used offline

handwriting samples, as students naturally write on paper in most educational settings. In addition, online handwriting requires the use of additional devices, such as tablets, digital pens, and a camera, to capture features like pressure, speed, and stroke order in real-time (Corbille et al., 2020; Karyna Korovai, 2020). Although this approach could provide detailed information, such method contradictorily makes data collection more time-consuming, costly, and dependent on external hardware (Ghosh et al., 2022). These constraints can cause challenges, especially when working with larger groups of children in schools. On the other hand, offline handwriting simply involves collecting written samples on paper, which are later scanned. This approach is more practical because it reflects children's everyday writing tasks in classrooms, making the samples more natural and representative of real learning activities. Therefore, this study focuses on paper-based handwriting or offline handwriting, which handles inappropriate writing formation through sentence copying activities among primary school-aged children.

2.3 Screening Test for Dysgraphia

The screening tests for detection of dysgraphia symptoms have been widely proposed among educators. There are two types of different approaches for assessing dysgraphia using conventional and computer-based methods. The implementation of both methods may pose pros and cons and will be further discussed in the next sub-section.

2.3.1 Conventional Method

Typically, dysgraphia is screened through a specified procedure, such as through observation session (Franken & Harris, 2021; Hamstra-Bletz & Blöte, 1993; Katarína Šafárová, Jiří Mekyska, 2021; Van Waelvelde et al., 2012). The screening sessions are conducted under the observation of a teacher during a teaching and learning (T&L) session. A variety of assessments are planned to determine student's cognitive strength and limitations in various skills. The assessments included handwriting legibility, writing speed (Franken & Harris, 2021), emotional and physical well-being (Katarína Šafárová, Jiří Mekyska, 2021). As an example, several researchers have attempted to explore the effectiveness of screening test tools using questionnaire and scoring methods. Meanwhile, variety of screening tools have been proposed to identify

dysgraphia symptoms according to respective language used in different countries as presented in Table 2.2.

Table 2.2
Established Dysgraphia Screening Tools Using Conventional Method

Screening Tool	**Country**	**Criteria of Handwriting Evaluation In Tools**	**Handwriting Domain**
The Systematic Screening for Motoric-Handwriting Difficulties (SEMS) (Franken & Harris, 2021)	Bavaria, Germany	Letter formation, fluency of writing, letter size, regularity of letter size, spacing between words, sentence alignment, correct joins	Offline
Handwriting Proficiency Screening Questionnaire (HPSQ) and Minnesota Handwriting Assessment (MHA) (Alhusaini et al., 2021)	Riyadh City, Saudi Arabia	Rate, eligibility, form, alignment, size, spacing	Offline
Handwriting Proficiency Screening Questionnaire (HPSQ) for teachers and its modification for children (HPSQ-C) (Katarína Šafárová, Jiří Mekyska, 2021)	Czechia	legibility, performance time (amount of time copying, erasing during writing, frequently look whiteboard during writing), physical and emotional well-being (tired, pain, do not want to write)	Offline and online
Adapted from Minnesota Handwriting Assessment (MHA) tools (Sakamat et al., 2019)	Putra Height, Selangor, Malaysia.	Readability, consistency in size, word spacing, letter spacing, letter formation, time completed	Offline and online
Systematic Screening of Handwriting Difficulties (SOS) (Van Waelvelde et al., 2012)	Flanders (Belgium)	(1) the writing is too large, (2) widening of left-hand margin, (3) poor letter or word alignment, (4) insufficient word spacing, (5) acute turns in connecting joins to letters, (6) irregularities in joins and/or absence of joins, (7) collisions of letters, (8) inconsistent letter size, (9) incorrect relative height of the various kinds of letters, (10) letter distortion, (11) ambiguous letter forms, (12) correction of letter forms and, (13) unsteady writing trace	Offline
Detailed Assessment of Speed of Handwriting (DASH) (Barnett et al., 2009)	United Kingdom	(1) Copy Best (2) Alphabet Writing (3) Copy Fast (4) Graphic Speed, and (5) Free Writing.	Offline and online
The Concise Evaluation Scale for Children's Handwriting - BHK	Amsterdam	Size of letter, margin, word alignment, word spacing, joins of letter, collisions of letter, inconsistent of size, letter	Offline

Screening Tool	Country	Criteria of Handwriting Evaluation In Tools	Handwriting Domain
(Hamstra-Bletz & Blöte, 1993)		distortion, ambiguous letter form, corrected letter, unsteady writing trace	

As reported in the literature, different conventional methods have been frequently employed to screen for dysgraphia. One of the approaches was screening using questionnaires (Katarína Šafárová, Jiří Mekyska, 2021) to evaluate the handwriting proficiency, including written rate, alignment, size, spacing, and consistency. In addition, questionnaire-based approaches were combined with other methods, such as the Minnesota Handwriting Assessment (MHA), as proposed by Alhusaini et al. (2021) and Sakamat et al. (2019).

Other than that, screening tools employing motoric-based methods, such as the SOS score (Van Waelvelde et al., 2012) but lacking in the consideration for longer writing tasks. In contrast, the DASH measurement (Barnett et al., 2009) enables screening at both group and individual levels. Although research by Sakamat et al. (2019) shows that the development of writing abilities might be enhanced as the children mature in age and experience, however, dysgraphia handwriting often demonstrates poor structural performance from the outset, together with characteristics of unsteadiness that reflect poor motor skills (Hamstra-Bletz & Blöte, 1993). This finding is consistent with the results reported by Rosenblum (2018), who found that writing abilities are significantly associated with performance time and overall legibility. The study also noted a higher occurrence of erased/overwritten words and letters, and difficulty recognizing letters, and reduced efficiency in writing and less smooth handwriting. However, the observation method is ambiguous and requires screening tools to help the detection of dysgraphia symptoms among students. Teachers who are unsure about their own assessment of their pupils' handwriting abilities may utilize SEMS as a screening tool to substantiate their observations with training (Franken & Harris, 2021).

In Malaysia, dysgraphia screening is incorporated within the dyslexia screening test known as Instrumen Senarai Semak Disleksia (ISD) (Syahida et al., 2023). The ISD screening has been practically utilized at all schools across the country to identify children who may have learning difficulties in the mainstream education program. In the case of students under six years old, the Literacy and Numeracy Screening (LINUS)

Program was introduced to evaluate their reading, writing, and arithmetic abilities before they enter standard one (Dzalani & Shamsuddin, 2014). The ISD includes general assessment items that aim to identify signs of dyslexia, dysgraphia, and dyscalculia. However, it is not specifically designed to detect dysgraphia. This indicates that the unique handwriting issues related to dysgraphia, such as poor letter formation, slow writing speed, and problems with fine motor skills, may be overlooked. The ISD also relies heavily on teachers' subjective observations and primarily focuses on reading difficulties, rather than thoroughly evaluating writing performance. Similarly, the LINUS program emphasizes assessing literacy and numeracy skills without looking at the handwriting processes involved, making it less effective at detecting writing problems related to motor skills. Additionally, LINUS is given to young students when their fine motor skills are still developing. Students are only referred for additional assessments after they have experienced repeated failures, which can delay the identification and support of those with dysgraphia. As a result, both ISD and LINUS lack the detail and accuracy necessary to identify dysgraphia correctly, risking late recognition and inadequate support for affected children (Dzalani & Shamsuddin, 2014; Syahida et al., 2023; Mohd Ariffin et al., 2018).

Manual screening for handwriting difficulties is usually slow and relies heavily on expert observation and judgment. This can cause subjectivity and inconsistency. As a result, large-scale or early screening becomes difficult, especially in schools where time and resources are limited. Although conventional methods have been widely used for diagnosing dysgraphia symptoms, a migration from manual screening to a computer-aided approach is more promising. Such methods could result in the transition of conventional assessment scoring methods to efficient computer-based calculations.

2.3.2 Issues and Challenges in Dysgraphia Assessment

In assessing dysgraphia among children, a conventional method has been used as a screening tool to plan for intervention and improve their writing skills performance. However, conventional method is time-consuming since teachers have to monitor continuously and analyze possible dysgraphia symptoms methodically (P. Chung & Patel, 2015). Additionally, the conventional method requires an expert to administer the assessment to children. So, there is a high probability that the evaluation process will be aborted due to distraction or inability to pay attention. Besides, the assessment

procedure, which is typically based on manual evaluation through observation, may produce a biased diagnosis. Therefore, the evaluation needed to be conducted by multiple evaluators to avoid bias or error in diagnosing children with suspected dysgraphia symptoms (P. J. Chung et al., 2020). Meanwhile, numerous community-based studies have been done to support dysgraphia's impact on writing impairment (P. Chung & Patel, 2015; J. Mekyska et al., 2017) as well as to investigate its symptoms (Ashraf & Najam, 2020; Biotteau et al., 2019), diagnosis (Dimauro et al., 2020; Katarína Šafárová, Jiří Mekyska, 2021), and treatment (Biotteau et al., 2019). As a result, dysgraphia has recently been the topic of interest among researchers who are focusing on diagnosing, detecting, and classifying dysgraphia in order to enhance the quality of education for school students, particularly those with learning disabilities.

2.3.3 Computer-Aided Method

With recent technological advancements, many researchers have proposed significant screening tests using a computer-based model and framework. Research on a computerised system for analyzing the handwriting process and the written products for detecting handwriting disorders or dysgraphia is necessary and crucial as it offers many benefits (J. Mekyska et al., 2017). Thus, most research works has focused on developing computer-based dysgraphia detection tools and screening systems as reviewed in Table 2.3.

Table 2.3
Dysgraphia Screening Tools Using Computer-Based Approach

Approach	Detection Tools	Description
Mobile app/ tablet app	MyScript (Raza et al., 2017)	The user writes the term on MyScript first, and servers will then return the characters to be compared to achieve an accurate spelling of the term, and the score is increased or decreased accordingly.
	Play-Draw-Write (Dui et al., 2020)	Empty space is shown on the app's interface, with a symbol or sequence of symbols to copy, like a square, circle, line, and U.
Software system	TestGraphia (Dimauro et al., 2020)	Text analysis using document analysis algorithms

	Smart Analyzer software (Galli et al., 2019)	Smart Analyzer software – based on drawing gesture. Children were asked to copy the illustrated figure given (a circle, an equilateral cross and a square).
Digitizing tablet as assistive tools	HandAQUS (Katarína Šafárová, Jiří Mekyska, 2021)	HandAQUS, a tool for handwriting acquisition, is utilised in conjunction with digitising tablets (Wacom Intuos Pro L) and a special inking pen (Wacom Ink Pen) to detect the position of the pen (x and y axes), pressure on the surface, pen tilt, and azimuth measurements. Used HPSQ-C to check the child's self-evaluation.
	Wacom Intuos 3 Tablet(D'Antrassi et al., 2018)	Using Wacom Intuos 3 Tablet with the Intuos Ink Pen, and 5th grade ruled paper Using Terzi's method – classify alphabet letters in subsets of Euclidean geometrical elements (straight lines, curves and rotations)
Additional hardware	Electromyography (Govindhan et al., 2020)	Using Electromyography whereas the technique for determining the electrical potential created by muscle and analysing these signals to determine any irregularities in the human biomechanics. For normal adult (16 to 30 years and above 50 years)

According to the literature, computational methods can be classified into a few approaches, including mobile or tablet applications (Dui et al., 2020; Raza et al., 2017), software systems (Dimauro et al., 2020; Galli et al., 2019), and digitizing tablets as assistive tools (D'Antrassi et al., 2018; Katarína Šafárová, Jiří Mekyska, 2021). According to prior studies, most of the computational-based methods identified handwriting disorders by analysing drawing gestures and writing mechanism. As an example, D'Antrassi et al. (2018) investigated kinematic and qualitative parameters for detecting the symptoms of dysgraphia based on geometric elements, such as straight lines, curves, and rotations. Similarly, Galli et al. (2019) introduced an assessment approach for dyslexia and dysgraphia children through drawing tests using optoelectronic system with quantitative protocol. Furthermore, Dimauro et al. (2020) developed the TestGraphia software system to monitor children with dysgraphia and assisted doctors in making diagnoses, and found that the system yielded satisfactory performance. More advance work has been presented by Katarína Šafárová, Jiří Mekyska (2021) on the experimental research of graphomotor disabilities (GD) and developmental dysgraphia (DD). The research experimented pen position, tilt, pressure, and azimuth measurement through interdisciplinary collaboration among psychologists, educationists, and engineers in Brno, Czech Republic. The study by Govindhan et al. (2020) proposed significant detection tool incorporating EMG sensor to identify writing

abnormalities among normal adults aged between 16 to 33 years old and 50 years and above. Nevertheless, dysgraphia detection dysgraphia detection should ideally occur during early childhood (Dui et al., 2020) to avoid the risk of delay that will influence their performance in writing which becomes significant as it makes up a great proportion of daily activities in schools (Katarína Šafárová, Jiří Mekyska, 2021). Thus, extensive study on writing tests (Galli et al., 2019) for identifying dysgraphia was found necessary to improve the performance by utilizing a variety of technologies, including machine learning approaches (Dimauro et al., 2020).

2.4 Intelligence-based Method for Dysgraphia Classification Model

Conventional screening approaches tend to overlook varied signs, demonstrating the necessity for intelligent alternatives. Therefore, numerous researchers have begun to engage in this field with the goal of handling the limitations of conventional screening tests through the application of various intelligence-based methods. Artificial intelligence has been advancing in addressing numerous challenges related to handwriting recognition and classification. In the specialized field of dysgraphia, the classification model has received significant attention from researchers aiming to support the screening process for identifying the severity levels and informing intervention sessions. Innovations in machine learning and computer vision hold the potential to transform dysgraphia detection through the analysis of handwriting samples, and thus enabling the identification of patterns and irregularities. A more detailed discussion of this issue is presented in the following subsection.

2.4.1 ML-based Application in Dysgraphia Handwritten Classification and Recognition

Machine learning (ML) has grown in popularity and has begun to infuse people daily lives, including the specialized field of impairment identification and assistive technology. Based on the literature, ML underwent significant developments with the potential to aid in dysgraphia detection. The methodologies have a wide spectrum ranging from simple machine learning techniques to more advanced approaches by implementing ensemble models.

Numerous methodologies have been utilized in previous studies to identify and examine dysgraphia, incorporating both conventional machine learning and advanced feature extraction techniques. For example, Gaussian Naive Bayes (Bouhamoum et al., 2024), Random Forest (Asselborn et al., 2020; Dankovicova et al., 2019; Dinusha et al., 2024; Drotár & Dobeš, 2020; Galaz et al., 2020; Kunhoth, Maadeed, et al., 2023; Zvoncak et al., 2019), and Logistic Regression (Bouhamoum et al., 2024; Ling & Khalid, 2017) have been employed for the identification of dysgraphia handwriting. Further research by Bouhamoum et al. (2024) has explored synthetic dataset as additional input for Naive Bayes and the results demonstrated enhanced recall and accuracy. However, this approach demonstrates limitation in its ability to generalize to real-world situations due to restricted datasets.

In a similar direction, SVM classifiers have been utilized in most studies, which have shown encouraging accuracy rates (Dankovicova et al., 2019; Drotár & Dobeš, 2020; Kunhoth, Maadeed, et al., 2023; Ling & Khalid, 2017; Lomurno et al., 2023; Rosenblum & Dror, 2017; Sihwi et al., 2019; Zvoncak et al., 2019). For example, research by Rosenblum & Dror (2017) found that handwriting classification can achieve over 89.9% accuracy using features extracted from writing dynamics. However, classifications that frequently depend on the process of extracting the features may face challenges in multi-dimensional data contexts. To address such limitations, expert systems utilizing rule based reasoning, including forward chaining (Kurniawan et al., 2018) have been developed and attained diagnostic accuracy of 94.71%. While this study provides valuable insight, they failed to address the dependence on predefined rules, which constrains scalability and adaptability to various handwriting samples. Additionally, the acquisition of sufficient knowledge from a single expert could lead to inconsistencies in decision-making and detection results. This inconsistency arises from varying statements made by different experts and thus affects the reliability of the system's diagnosis.

Furthermore, ensemble models such as AdaBoost, Random Forest, and Support Vector Machine (SVM) (Dankovicova et al., 2019; Dinusha et al., 2024; Drotár & Dobeš, 2020; Kunhoth, Maadeed, et al., 2023) required large datasets and extensive hyperparameter tuning, which increases computational demands and often result in reduced performance when only limited data are available. Similarly, optical character recognition (OCR)-based methods incorporating tools such as OpenCV and the Discrete Wavelet Transform (DWT) are highly sensitive to variations in handwriting styles and

image quality (Villegas-Ch et al., 2023). Approaches that rely on external application programming interface (API)-based tools, such as the Google Cloud Vision API (Dinusha et al., 2024), also exhibit limited adaptability to diverse handwriting styles and introduce dependency on third-party services. Furthermore, optimization techniques such as XGBoost (J. Mekyska et al., 2019) and Bayesian optimization (Gavenciak et al., 2025) have shown tendencies to overfit and face challenges when dealing with imbalanced datasets. Even though countermeasures like the Synthetic Minority Oversampling Technique (SMOTE) and cross-validation can be applied, their effectiveness is highly dependent on the availability of large, well-labeled datasets and considerable domain expertise (Fernández et al., 2018). Collectively, these studies highlight the absence of an integrated framework capable of simultaneously achieving high accuracy, detection performance, and scalability of features used for dysgraphia detection.

Table 2.4 illustrates various performance levels in terms of accuracy, sensitivity, specificity, and additional metrics, highlighting differences in machine learning techniques and dataset characteristics.

Table 2.4
List of Previous Studies on Dysgraphia Recognition and Classification Using ML Model

Author (Year)	**ML model**	**Domain Type**	**Performance**	**Limitations**
Bouhamoum et al. (2024)	Gaussian Naive Bayes algorithm	Online and offline handwriting	accuracy of 0.94, recall of 0.85, precision of 0.85, and F1 score of 0.85.	Only tracks how the condition changes over time and uses simulated data, which may not accurately reflect real-world situations.
Rosenblum & Dror (2017)	SVM	Online handwriting	an accuracy of 89.9%.	limited to online data and relies only on handwriting dynamics and visual features from a digital pen, with limited use of special equipment.
Kurniawan et al. (2018)	Forward Chaining method	Offline handwriting	accuracy of 94.71%	Rule-based approach limits scalability and adaptability that are hard to adapt to new data

Author (Year)	ML model	Domain Type	Performance	Limitations
Karyna Korovai (2020)	K-means algorithm	Online handwriting	The study identified 12 dynamic features and 22 spatial features that effectively contributed to handwriting style clustering.	Unsupervised clustering only; lacks classification performance evaluation; generalizability uncertain
Asselborn et al. (2020)	Random Forest classifier	Online handwriting dataset	sensitivity of 78%	Only relies on a large set of 63 low-level features that require specialized digital equipment and only reports the sensitivity measurement, which gives an incomplete view of model performance.
Kunhoth, Maadeed, et al. (2023)	KNN, SVM, Random Forest, and AdaBoost.	online handwriting	accuracy of 80.8% using the AdaBoost classifier	Performance relatively low; features may be shallow; dataset size unclear
Drotár & Dobeš (2020)	AdaBoost	online handwriting	accuracy of 79.5%	Limited accuracy; no feature extraction optimization; dataset diversity unclear
Dinusha et al. (2024)	Random Forest	offline handwriting	accuracy of 87.3%	Only offline data used; no comparison with deep models; limited metrics reported
Lomurno et al. 2023)	Quasi-SVM	online handwriting	accuracy of 84.62%	Experimental algorithm with unclear scalability; dataset size not specified
Dankovicova et al. (2019)	Random Forest, SVM, and AdaBoost	online handwriting	of 67.1% for Random Forest, 66.1% for SVC, and 64.1% for AdaBoost.	Low performance; features may be insufficient; no optimization or feature selection used
Jiri Mekyska (2023)	gradient-boosting algorithm	Online handwriting	accuracy of 83.6%	Moderate accuracy; limited to online data only; dataset details unclear
Galaz et al. (2020)	Random Forests	Online handwriting	accuracy of 84%	Only one algorithm tested; no comparison or ablation study; limited metrics

Author (Year)	ML model	Domain Type	Performance	Limitations
Zvoncak et al. (2019)	support vector machine (SVM	Online handwriting	accuracy of 84.66%, with sensitivity at 88.70% and specificity at 82.53%.	Lacks feature extraction discussion; generalizability not validated on other datasets
Sihwi et al. (2019)	Support Vector Machine (SVM)	Online handwriting	accuracy of 82.51%	Limited accuracy; only one model; no feature optimization
Mekyska et al. (2019)	XGBoost.	Online handwriting	50% sensitivity and 90% specificity	Very low sensitivity; poor balance between classes; dataset imbalance issue

Table 2.4 summarizes that supervised models, such as SVM, Random Forest, and AdaBoost, dominate in the online handwriting domain, while unsupervised and hybrid approaches (Lomurno et al., 2023) demonstrated potential in feature extraction and clustering. Meanwhile, research on offline handwriting has demonstrated the successful outcomes of probabilistic and ensemble models (Dankovicova et al., 2019; Danna et al., 2023; Dinusha et al., 2024; Kaisar & Chowdhury, 2022; Rashid et al., 2023; Robaa et al., 2024; Sukiman et al., 2024). The differences in performance among various methodologies have highlighted numerous ML model proposed in the study field for dysgraphia handwriting recognition and classification. The implications suggest that the machine learning model holds significant promise for further exploration within the offline handwriting domain, particularly in the context of the performance obtained. However, conventional ML methods in handwriting analysis face several challenges. They rely heavily on a manual feature extraction process, such as stroke width, slant, and spacing, which is time-consuming. This approach often fails to capture subtle handwriting patterns. As a result, it limits accuracy and generalization across datasets. For example, Dankovicova et al. (2019) reported relatively low performance with Random Forest at 67.1%, SVM at 66.1%, and AdaBoost at 64.1% for online handwriting. This reflects poor adaptability, where ML methods also struggle with high handwriting variability due to intra-class differences and inter-class similarities, which could reduce classification accuracy. A study by Zvoncak et al. (2019) has proven this issue by achieving only 50% sensitivity using XGBoost. Similar to Asselborn et al. (2020), the study obtained 78% sensitivity with Random Forest,

highlighting the risk of false negatives in screening tasks. Furthermore, ML models often have scalability limitations. The model performs reasonably well in binary tasks but less effectively in multiclass severity classification. For instance, Kunhoth, Maadeed, et al. (2023) achieved 80.8% with AdaBoost. Even in offline handwriting, results remain moderate. As example, Dinusha et al. 2024) reported 87.3% accuracy using Random Forest, which is still below deep learning benchmarks.

Although ML models have achieved strong performance in handwriting classifications, their reliance on meaningful feature vectors from quality datasets poses limitations. This progression naturally leads to considering the implementation of deep learning approaches to enhance dysgraphia handwriting recognition and classification. Further investigation of deep learning applications in dysgraphia handwriting analysis is discussed in the next subsection.

2.4.2 DL-based Application in Handwritten Classification and Recognition

Deep learning (DL) works as a progressive framework in the fields of handwritten recognition and classification which is capable of handling a range of objectives and requirements. The deep learning-based methodology has been widely employed across a wide range of approaches, from lightweight deep layer algorithms to explainable deep learning frameworks. Nonetheless, limitations remain, and enhancements are necessary to refine the model by concentrating on the detection of learning disabilities symptoms through handwriting activities.

The most recent works proposed by Aldehim et al. (2024), Sebastian et al. (2024) and Yogarajah (2020) have demonstrated the extensive application of Convolutional Neural Networks (CNNs) in handwriting recognition and feature extraction. The research works are attributed to their outstanding ability for spatial pattern detection. Based on previous research, CNNs have shown an impressive ability to independently learn hierarchical features from handwriting images, thereby eliminating the need for a manual feature extraction process in preparing significant features. Therefore, the ability of CNNs to perform tasks effectively with large datasets and their flexibility in adapting varied handwriting styles highlights their significance in tackling handwriting-related challenges. Furthermore, hybrid CNN models that combine various algorithms such as CNNs with Long Short-Term Memory (LSTM) layers (Carbune et al., 2020; Masood et al., 2023) or Random Forest classifiers

(Alqahtani et al., 2023; Masood et al., 2023; Sesa et al., 2022) have been proven to enhance the classification performance. However, these methods come with a several drawbacks namely increasing complexity and computational requirements which could limit scalability for practical applications.

Several studies have investigated deep learning methods, including transfer learning (Robaa et al., 2024; Yogarajah & Braj, 2020), unsupervised learning (Saha et al., 2020), and hybrid models (Alqahtani et al., 2023; Masood et al., 2023), for classifying handwriting in dysgraphia. Transfer learning with transformer-based models has demonstrated promising results and facilitated the creation of an explainable AI framework (Robaa et al., 2024). However, explainable AI for dysgraphia is still at early stages, and current models are mostly trained on Latin-letter datasets, which limits their generalizability. Meanwhile, unsupervised techniques like variational autoencoders have been used to extract handwriting features (Saha et al., 2020). However, the technique does not combine well with supervised classifiers, which later reduces the effectiveness for dysgraphia diagnosis. Dynamic handwriting features are essential and rarely explored, with only limited work using RNNs and LSTMs to model temporal sequences (Zolna et al., 2019). Moreover, studies that combine traditional ML methods, such as SVM and Random Forest, with DL have shown strong potential (Kariyawasam, & Ratnayake, et al., 2019; Kariyawasam, & Samarasinghe, et al., 2019). But they often lack clear reasoning behind algorithm choices and demonstrate weak integration, which diminish its effectiveness. Despite advances in current research, it faces challenges related to dataset diversity, dynamic feature modelling, explainability, and scalability, which impede the development of robust models for dysgraphia detection.

CNNs are extensively applied in identifying dysgraphia due to the strong capability to extract spatial information from handwriting data. However, certain limitations remain in the use of CNN-based approaches. Sebastian et al. (2024) used a CNN-based Tesseract OCR on an offline handwritten dataset of capital and lowercase letters but hardly achieved 67.79% accuracy. The result of this study suggests that the algorithm may struggle with different handwriting styles. However, newer CNN (Vydeki et al., 2024;Aldehim et al., 2024) designs have demonstrated improved performance by achieving 91.8% and 99.5% accuracy, respectively. However, dataset variety and the hazards of overfitting may limit these outcomes. CNN models require large labeled datasets, therefore they may be less successful when the data is limited or when handwriting styles vary.

On the other hand, hybrid models that combined CNNs with other approaches have shown an improved performance, though complexity and resource limitations can be issues. As proven by Alqahtani et al. (2023) CNNs and SVMs are capable of achieving an accuracy of 99.33%. Similarly, the study by Masood et al. (2023) has shown that the CNN-LSTM model applied to online handwriting data reached 97.6% accuracy. However, both studies employed methods with limitations that may restrict their generalizability. These models employ CNNs for spatial feature extraction and LSTMs for capturing temporal patterns. Such advanced architectures require substantial computational resources and may be prone to overfitting, particularly when trained on small sample sizes.

Transformer-based and transfer learning methods are widely recognized for utilizing pre-trained information. As an example, Robaa et al. (2024) used a transformer-based model with transfer learning to develop an explainable AI framework which produced 99.65% accuracy on an offline letter-form dataset. However, the dataset's biases and lower generalizability may limit the study's applicability. These methods perform well and offer clear explanations, although they depend heavily on pre-trained models built with universal handwriting datasets. Nevertheless, this approach might not fully capture dysgraphia handwriting's complexities, limiting its applicability.

Experimental deep learning models, including MLPs and LSTMs, have been explored by Rashid et al. (2023), who achieved 80.3% accuracy with these models on an offline dataset of words and letters. Due to limitations of offline datasets, this strategy may be restricted. Due to poor feature extraction, these models performed worse than CNN-based approaches, despite being simpler and less computationally intensive.

Table 2.5 illustrates a variety of deep learning models utilized for dysgraphia handwriting recognition and classification, emphasizing both offline and online handwriting datasets.

Table 2.5
List of Previous Studies on Dysgraphia Handwriting Recognition and Classification Using DL Model

Author (Year)	**ML Model**	**Domain Type**	**Performance**
Sebastian et al. (2024)	CNN-based Tesseract OCR	offline handwriting dataset: upper and lower case letter	accuracy of 67.79%
Vydeki et al. (2024)	custom Convolutional Neural Network (CNN), VGG16, and ResNet50	offline handwriting	accuracy of 91.8%
Robaa et al. (2024)	Explainable AI framework. Using transfer learning and transformer-based	Offline handwriting - letter form	Precision of 99.65%
Rashid et al. (2023)	Multi-Layer Perceptron (MLP), and Long Short-Term Memory networks (LSTMs).	offline handwriting dataset (word and letter)	80.3% accuracy.
Ghouse et al. (2022)	Non-Discrimination Regularization in Rotational Region Convolutional Neural Network (NDR-R2CNN)	offline handwriting dataset	accuracy of 98.2%
Alqahtani et al. (2023)	hybrid CNN with machine learning classifiers, specifically Support Vector Machines (SVM) and Random Forest (RF)	offline handwriting dataset (letter)	The CNN-SVM model outperformed with an accuracy of 99.33%
Vilasini V et al., (2022)	Convolutional Neural Networks (CNN) and Vision Transformers (ViT)	offline handwriting dataset (letter)	CNN: accuracy of 79.47%, ViT : accuracy of 86.22%.
Aldehim et al. (2024)	convolutional neural network (CNN)	offline handwriting dataset	accuracy of 99.5%
Yogarajah (2020)	Convolutional Neural Network (CNN)	offline Hindi handwriting	average accuracy of (86.14 ± 1.02)%
Kunhoth, Al Maadeed, et al. (2023)	Convolutional neural network (CNN) models and fuses CNN features from different handwriting tasks. ML: support vector machine (SVM), AdaBoost, and Random Forest	offline handwriting Slovak orthography.	accuracy of 97.3% using a SVM
Masood et al. (2023)	hybrid model combining CNN and LSTM	online handwriting	accuracy of 97.6%

Author (Year)	ML Model	Domain Type	Performance
Saha et al. (2020)	Variational Autoencoder (VAE) Clustering	Offline handwriting dataset	successfully extracted and interpreted a significant number of distinguishable handwriting characteristics that are clinically recognized as symptoms of dysgraphia.
Zolna et al., (2019)	Recurrent Neural Network (RNN) with a 2-layer Long Short-Term Memory (LSTM) architecture to analyze handwriting dynamics.	Online handwriting dataset	RNN model correctly identified over 90% of children diagnosed as dysgraphic
Kariyawasam, & Samarasinghe, et al. (2019)	CNN	Offline handwriting	The screening accuracies for letter dysgraphia, dyslexia, and numeric dysgraphia were 88%, 58%, and 99%,
Kariyawasam, & Ratnayake (2019)	CNN and SVM	offline handwriting	90% accuracy
Kunhoth et al. (2024)	CNN with SVM and XGBoost algorithms	online and offline handwriting dataset	an accuracy of 88.8%

The literature discloses comprehensive progress in dysgraphia handwritten recognition and classification, especially using CNNs, hybrid models, and transformer-based methods. The literature summarized in Table 2.5 indicates that CNN-based models have achieved considerable popularity and have received significant recognition for their outstanding performance on offline handwriting datasets. These models have established a standard and provide encouragement for the advancement of enhanced models focusing on attaining greater accuracy and superior overall performance. However, their effectiveness is still limited when it comes to sentence-level handwriting data. Most previous studies have focused on letter or word recognition (Alqahtani et al., 2023; Rashid et al., 2023; Robaa et al., 2024; Sebastian et al., 2024; Vilasini V et al., 2022) and usually employ binary classification (Aldehim et al., 2024; Kariyawasam, & Ratnayake, 2019; Kariyawasam, & Samarasinghe, et al., 2019; Vydeki et al., 2024; Yogarajah & Braj, 2020), such as classifying dysgraphia writing apart from non-dysgraphia writing. This method does not fully meet the need for multiclass severity classification, which is important for understanding various levels of handwriting difficulties. Although feature fusion techniques (Alqahtani et al., 2023; Kunhoth, Al

Maadeed, et al., 2023; Kunhoth et al., 2024) have been introduced to improve the classification, CNN models still struggle with the complexity and variability of sentence-level handwriting. Therefore, the use of CNNs for multiclass severity classification using sentence-level handwriting remains underexplored, despite its importance in capturing the full spectrum of handwriting difficulties. Therefore, automated fusion feature techniques have a huge potential to help extract important features that distinguish between various severity classes.

2.5 The Importance of Characterized Handwriting Datasets for Intelligence-based Methods

In AI-based machine learning, particularly computer vision research that primarily works with image classification, it is highly crucial to have a characterized and specified dataset. This is because the research in writing analysis has various forms of datasets and different writing patterns. This section investigates the diversity of handwritten characteristics and presents the limitations of the handwriting dataset. The development and utilization of handwriting datasets for dysgraphia is crucial for advancing diagnostic and intervention tools for learning disabilities. Dysgraphia, in particular, affects writing abilities and hence resulting challenging in letter formation, spacing, and alignment. Several studies have focused on creating and analyzing handwriting datasets to improve the detection and understanding of these disorders.

2.5.1 Diversity of Handwriting Characteristics

Handwriting is regarded as unique to each individual exhibiting particular characteristics that form consistently similar pattern, thereby facilitating individual identification (Lewis, 2014). Moreover, research indicates that the uniqueness of handwriting can be quantitatively assessed through machine learning techniques, which evaluate characteristics like line spacing, slant, and the shapes of characters (Srihari et al., 2002). With various methodologies, handwriting exhibits variability influenced by factors such as age, gender, and cultural background, which play a significant role in shaping writing styles (Saunders et al., 2011). Therefore, automated comparison procedures can provide a statistical estimation of the likelihood that two individuals

possess indistinguishable writing profiles, underscoring the inherent variability present in handwriting (Saunders et al., 2011).

Table 2.6 presents the diversity of dysgraphia handwriting characteristics and criteria of each characteristic (Rosenblum, 2018; Biotteau et al., 2019; Lopez & Vaivre-Douret, 2023; Jusoh, 2024; Lopez & Vaivre-Douret, 2023)

Table 2.6
The Diversity of Dysgraphia Handwriting Characteristics

Handwriting Characteristics	**Criteria**
Variability in letter formation	• Inconsistent letter shapes and sizes. • Mixing of uppercase and lowercase letters within words • Use a combination of cursive and print styles in a disorganized manner
Variations in spacing and alignment	• Uneven spacing between words • Insufficient gaps between letters • Improper writing alignment of consistency
Range of writing speed and fluency	• Slow speed in writing due to difficulty in coordinating fine motor movements. • May write quickly but produce poorly formed, illegible letters. • Frequent pauses, erasures, and restarts disrupt the flow of writing.
Differences in pen pressure and stroke quality	• Apply excessive pressure, leading to deep imprints on the paper. • May write too lightly, making their handwriting faint or incomplete. • Variability in stroke thickness can result in inconsistent handwriting styles.
Variability in grip and writing posture	• Hold their writing tool with a tight, uncomfortable grip, leading to hand fatigue. • May use an unusual wrist or arm position, affecting writing control. • Frequent adjustments in grip may be observed as the individual struggles to maintain control over the pen or pencil.

In understanding the variety of handwriting characteristics associated with dysgraphia, it is crucial to develop the screening tools with an automated computer-aided system. The variety of handwriting exhibited by children with dysgraphia presents

challenges in assessing the level of severity. Consequently, numerous researchers have explored and investigated methodologies for recognizing and classifying dysgraphia through handwritten products. The following sub-sections will further discuss the existing dysgraphia dataset utilized for machine learning research, along with its limitations.

2.5.2 Limitations of Existing Dysgraphia Dataset

In recent machine learning (ML) research, dysgraphia recognition have been primarily implemented through handwriting analysis based on handwriting characteristics. Numerous studies have shown that handwriting analysis has potentially contributed to the high performance in identifying dysgraphia severity. Therefore, the dataset is important to realize the model of dysgraphia recognition or classification of severity.

Previous studies have shown that variety of datasets have been used in dysgraphia handwriting analysis involving different age groups and languages. For instances, the studies conducted by Asselborn et al. (2020) and Gargot et al. (2020) were conducted using French handwriting data, and other works were conducted using various handwriting datasets from different languages including Sri Lanka (Kariyawasam, Nadeeshani, Hamid, Subasinghe, Samarasinghe, et al., 2019), Hindi (Yogarajah & Braj, 2020) and Slovak (Drotár & Dobeš, 2020). These studies proved that data type plays a significant role in relation to the diversity of languages and participant backgrounds. Another highlight is the distinction between online and offline data sources.

Online data encompasses information gathered through digital instruments like tablets, allowing participants to document their responses, alongside dynamic and kinematic metrics such as velocity, pressure, speed, writing patterns, and angles. On the other hand, offline data is related to the act of writing on paper, which is subsequently transformed into a digital format via scanning and stored as JPG or PNG files. This review indicates that no specific benchmark model should be utilized in conjunction with the characteristics of a particular dataset. This results in significant challenges for comparison due to the extensive variety present in the execution of these experiments. Consequently, utilizing paper-based handwriting samples is inherently more straightforward and avoids the need for additional devices like tablets and styluses. The

act of writing on paper is a common practice engaged in by school children as they work towards mastering writing skills in alignment with their developmental stage.

Table 2.7 summarizes previous research on related to dataset used in handwriting analysis. In this field of study, most research highlights several important contextual factors, including the variety of languages, countries, data types, age groups, and sample sizes. Meanwhile, the factor of age does not hold the same significance as language in the research process. . In addition, the optimal sample size required for dysgraphia identification remains insufficiently established. Consequently, based on the previous studies, an analysis of the current literature highlights the challenge of limited benchmark datasets and models for analysis. Nonetheless, considering that a majority of the work is recent, it indicates that this field demonstrates significant potential and remains a promising area for further investigation. From all reviews, the only available studies that concentrated exclusively on the Malay language was (Rosli et al., 2021). However, the dataset was provided in the form of single characters of handwriting images. From this research, it is justified that the Malay language resources for dysgraphia research is notably limited and demands further development in this research area.

Table 2.7
Summarize of Previous Research on Dysgraphia Dataset For Handwriting Analysis

Authors	**Year**	**Data Type**	**Language/ Country**	**Age**	**Sample Size**
Jafarzadeh et al. (2024)	2024	Handwriting - image (words, sentences, character and digits)	Persian. Iran	not specified	400 native Persian writers. 44,000 words, 60,000 letters, and 6,000 digits.
Rashid et al. (2023)	2023	handwriting - digital tablet	Pakistan.	5 to 15 years old.	25 children.
Devillaine et al. (2021)	2021	handwriting - digital tablets.	Grenoble, France	school-aged children.	305 children.
Rosli et al. (2021)	2021	handwriting images – (character)	Malaysia	school-aged children	138,500 handwriting images.
Seman et al. (2021)	2021	handwriting images (character)	Chinese	not specified	not specified

Authors	Year	Data Type	Language/ Country	Age	Sample Size
Gargot et al. (2020)	2020	handwriting - digital tablets.	French	5 - 10 years old	280 children
Asselborn et al. (2020)	2020	handwriting - digital tablet.	French	5 and 15 years old	not specified
Yogarajah & Braj (2020)	2020	Handwriting - images	Hindi	7-9 years old	54 children, 267 letters
Richard & Serrurier (2020)	2020	Handwriting - images	Roman alphabet	not specified	1481 paragraphs
Drotár & Dobeš (2020)	2020	Handwriting - digital tablet	Slovak	8-15 years old	120 children
Kariyawasam, & Samarasinghe, et al. (2019)	2019	Handwriting - image	Sri Lanka	6-7 years old	5000 characters

Table 2.7 clearly demonstrates that the development of new Malay sentences and phrases is crucial for achieving the research objective. Based on the reviewed studies, the only study from Malaysia (Rosli et al., 2021) has examined dysgraphia classification and limited to the character level. To date, no research has extended this investigation to the word or sentence level in the Malay language. This gap highlights the need to expand further research beyond individual characters by using phrases and sentences since both represent how people write naturally. Meanwhile, the incorporation of handwriting in image format is significant for this research to imitate the conventional approach of writing tasks in educational environments that predominantly use paper. Therefore, identifying dysgraphia during the school years is significant by ensuring that the dataset includes individuals from this age group to enable effective and relevant data collection.

2.6 The Importance of Pre-Processing Techniques in Handwritten Images

Most related research on handwriting recognition has faced challenges in the preprocessing phase to prepare an affordable dataset for the recognition system (Ibrahim et al., 2023; Karthick et al., 2022b). Many techniques in image processing have been implemented to provide the finest data input before the recognition process, such as edge detection (Ley et al., 2021), morphology (Thi Thi Zin, 2021), binarization

(Alyahya et al., 2023; Rahman & Halim, 2022; Thi Thi Zin, 2021), and filtering (Najadat et al., 2019). Currently, the pre-processing technique is highly explored in handwriting recognition niche areas to enhance the quality of handwriting images.

As example, several studies have implemented several preprocessing steps such as binarization, edge detection, filtering, dilation, and morphological operations prior to executing their character handwriting recognition in pre-processing phase (2021a, 2021b; Kunhoth, Al Maadeed, et al., 2023; 2021; Thi Thi Zin, 2021). In another study, Najadat et al. (2019) utilized the binarization, filtering, and smoothing processes in preparing the Arabic handwritten character images for the recognition network using the Convolutional Neural Network (CNN). Another study on pre-processing phase was presented in a study by Fatimah et al. (2019) by resizing, binarizing, and thresholding the handwriting image in performing personality identification. Besides, the study by Longjam et al. (2023) mentioned that the implementation of the preprocessing step by applying greyscale conversion and resizing the image of multi-scripted signatures has significant effect on the classification model. In sum, all of the previous studies have presented the importance of pre-processing phase with aims to enhance the quality and achieve the specific targeted condition of the image without losing the information in handwriting recognition (Karthick et al., 2022a).

Pre-processing handwriting images is typically concentrated on enhancing the image quality by converting to binary images and segmenting the image based on the recognition system's target image. In addition to improving image quality, one important factor that must be considered in achieving strong handwriting recognition performance is the removal of irrelevant information from the images. This technique comprises the elimination of certain extraneous elements from the picture, including ruling lines (Alipour et al., 2013), table (Rani et al., 2021), and character strokes like Shiro Rekha in Devanagari words (Seki, 2020; Wong & Loh, 2019).

As summarized in Table 2.8, various techniques have been used to eliminate undesired entities including the ruling lines. The ruling lines were commonly seen in most handwriting images including Arabic, Mandarin, Tamil, Urdu and other languages. A variety of pre-processing techniques have emerged, aimed at removing ruling lines and irrelevant structures from both handwritten and printed documents. These advancements have markedly improved the accuracy of handwriting recognition and classification tasks.

Ruling line removal is a crucial preprocessing step that significantly improves the accuracy of handwriting recognition and classification. Various image processing techniques, such as line segmentation (Rani et al., 2021), edge detection (2021), mathematical morphology (Imtiaz et al., 2014), Hough Transform (Refaey, 2014, 2015), binarization (Alipour et al., 2013) , line thickness estimation (J. Kumar & Doermann, 2011), region-based extraction (Alipour et al., 2013), and connected component analysis (Khan et al., 2009) have been shown to effectively eliminate tables, gridlines, and ruling lines from handwritten and printed documents. By removing these structural artifacts, such methods help preserve the integrity of handwritten strokes, reduce background noise, and enhance the quality of extracted features. The implementation has enhances character classification and reduces recognition errors, ultimately supporting more accurate and robust machine learning and deep learning models for handwriting analysis across diverse languages and domains.

Table 2.8
Similar Studies on Removing Unwanted Entities From Handwriting Images

Author (Year)	**Method**	**Domain/Unwanted Entities**
Rani et al. (2021)	Line segmentation, edge detection techniques and morphological operations	Tables in both handwritten and pre-printed documents
Shobha Rani and Vasudev (2018)	1. Image enhancement: thresholding, median filtering, and order static filtering 2. Detecting lines: convolution with a rectangular structuring element 3. Post-processing: connected component analysis	Horizontal and vertical lines from pre-printed documents
Refaey (2015)	Hough Transform	Handwritten documents.
(Imtiaz et al., 2014)	Mathematical morphology techniques	Handwritten documents in various language
Alipour et al. (2013)	Binarization process, estimating the distance between the lines, finding the best left and right columns of the main lines, deleting pixels unrelated to the lines, and estimating the thickness of the main lines.	Handwritten documents
Kumar & Doermann (2011)	Integral-image feature and large-scale Support Vector Learning	Handwritten document images of Arabic language
Khan et al. (2009)	Connected component analysis to convert line components to the background	Handwritten and printed text

Author (Year)	Method	Domain/Unwanted Entities
(Puri & Singh, 2019)	Connected Component Analysis (CCA)	printed and handwritten mono-lingual Hindi, Sanskrit and Marathi documents
(Mekhaznia et al., 2020)	Employ Morphological Operations which uses dilation, erosion, opening, and closing to remove or refine line artefacts.	Staf-line on music score document image

2.7 Advancement in Feature Extraction Method for Dysgraphia Classification Models

The exploration of handwriting recognition and classification has evolved significantly through machine learning and deep learning techniques. Feature extraction is a fundamental step in a handwriting recognition system. The step transforms raw image data into a set of representative descriptors that can be effectively utilized by a classification model. Extending this perspective, recent studies have highlighted the implementation of advanced feature extraction methods. Recent works have specifically emphasized the classification models including CNNs, Optical Character Recognition (OCR) and hybrid models. The next section provides a thorough discussion on the respective feature extraction and evaluates their effectiveness as well as identifies any potential limitations.

2.7.1 Histogram-based Features Extraction

A histogram projection represents the distribution of pixel intensity values across an image in digital image processing (Wei et al., 2024). The depiction of the histogram projection is illustrated in Figure 2.2, wherein it is represented as a graph. The x-axis is designated to indicate intensity levels, specifically ranging from 0 to 255 in grayscale, while the y-axis delineates the frequency of pixels corresponding to each intensity level (Furht et al., 2018). This statistical representation summarizes the characterized composition and structural patterns of an image, which makes it useful for various computer vision tasks (Wei et al., 2024). The extraction of histogram information as numerical descriptors results in the formation of histogram features. These features are classified as handcrafted, capturing fundamental visual indications

such as contrast, brightness, edges, and texture that are crucial for representing and distinguishing different digital images' pixel distribution.

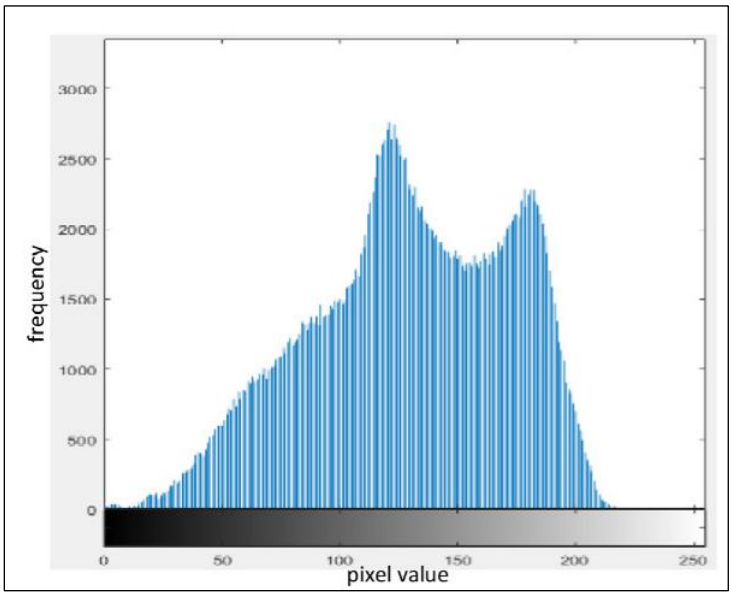


Figure 2.2 An Example of a Histogram Projection to Represent the Frequency of Pixel Intensity (Setiadi et al., 2023)

The importance of histogram features in image analysis stems from their ability to provide compact, interpretable, and robust representations of image content. By analyzing the distribution of pixel intensities, histogram-based methods can highlight structural and statistical patterns that support tasks like edge detection, image enhancement, segmentation, object tracking, and classification. For instance, an edge detection method using the standard deviation of histogram features was introduced by Najjar et al. (2022) to enhance the Canny algorithm, achieving superior visual quality and quantitative results compared to classical approaches. Meanwhile, a study by Mohan et al., (2025) demonstrated that histogram-based contrast enhancement significantly enhances the extraction of features from MRI images for accurate cancer classification. Similarly, earlier work by Guo (2009) also demonstrated the use of grayscale histogram features to distinguish between normal and abnormal tissues in CT images to support the image recognition task.

The integration of histogram-based features into CNNs and deep feature extraction has emerged as a promising approach in computer vision. The combination of handcrafted histogram features and deep models may have complementary strengths that could enhance the discriminative power of deep models. Traditionally used for

feature representation, histograms are now being adapted to improve the robustness and efficiency of CNNs by providing additional structural and statistical information. For instance, Nousi et al. (2019) has developed a visual object tracker that integrates CNN features with a Bag-of-Features histogram mechanism to improve the performance. Similarly, S. Wu et al. (2016) proposed a Feature Fusion Net (FFN) that fuses handcrafted color and texture histograms with CNN-derived features, yielding more compact and impactful. More recent work has shown that embedded learnable histogram layers directly within CNNs as demonstrated by Peeples et al. (2022). The joint optimization of histogram binning and deep feature extraction has shown improved image classification accuracy. With the same direction, studies by Gong et al. (2019), Z. Wang et al. (2018), and Yu et al. (2024) showed that learnable histogram layers enhanced semantic segmentation by adaptively determining optimal bin centers and widths during training. Recognizing these strengths, researchers have increasingly explored the benefits of incorporating histogram-based features into deep learning applications. A CNN demonstrates a remarkable ability to autonomously acquire hierarchical features directly from raw pixel data. However, it is important to note that they may neglect specific low-level statistical indications that handcrafted descriptors, such as histograms, are adept at capturing. The integration of histogram features into CNNs thus offers a hybrid approach that combines the interpretability and robustness of handcrafted features with the adaptability and discriminative power of deep learning.

Hybrid methods highlight the potential of integrating histogram features with deep learning. A study by Tsourounis et al. (2022) combined SIFT descriptors with CNNs to preserve spatial relations and achieve rotation invariance, enhancing classification and lip-reading tasks. In a similar objective, W. Chen et al. (2021) has evaluated feature purity using activation histograms and entropy, while S. Wu et al. (2016) has developed orientation-invariant histograms to improve rotation-invariant feature capture in CNNs. Although adding histogram features may not be directly interpretable, it can greatly improve the feature representations for accurately classifying complex and varied visual data, such as handwriting.

Despite these advances, the application of histogram-based features in the domain of handwriting analysis, particularly for detecting and classifying dysgraphia, remains largely underexplored. Most existing handwriting classification studies rely solely on deep learning models that learn features directly from raw images, often overlooking the potential benefits of incorporating handcrafted histogram features that

can capture subtle variations in stroke intensity, texture, and structural patterns (Gargot et al., 2020; Vydeki et al., 2024). Such characteristics are especially important in dysgraphia, where handwriting irregularities are often nuanced and localized. Integrating histogram-based features with deep CNN architectures could provide complementary structural and statistical information, enhance the discriminative power of the model and improve the accuracy of dysgraphia severity classification (Vydeki et al., 2024). This motivates the present study to investigate a hybrid approach that fuses features with deep CNN features to recognize and classify dysgraphia handwriting.

2.7.2 DL-based Features Extraction

In Deep Learning (DL) based classification model, automatic features extraction method has been introduced to overcome limitations of traditional hand-crafted techniques such as in OCR system. The findings from previous works on OCR-based methods have highlighted main challenges with variability of handwriting text (Sebastian et al., 2024), and difficulties of recognizing characters (Vydeki et al., 2024). Furthermore, both studies suggested that the overall performance of classification system has been impacted due to the complexity and inconsistencies introduced in a variety of feature extraction techniques.

On the other hand, DL-based feature extraction has experienced significant growth in the classification of handwriting image recognition. An advancement in feature extraction of CNNs has demonstrated significant improvement in handwriting classification. As an example, the research by Robaa et al. (2024) implemented MobileNetV3 integrated with Squeeze-and-Excitation (SE) blocks achieved an impressive test precision of 99.65%. This approach demonstrates the applicability of deep learning in recognizing complex handwriting features. Similar work by Yogarajah (2020) has highlighted the effective application of deep learning methods in handwriting analysis for identifying dyslexia and dysgraphia. The research has reported that by using CNNs feature extraction in dysgraphia can achieve an accuracy of 86.14% ± 1.02% for Hindi handwriting images. Meanwhile, the incorporation of dropout and batch normalization in detecting dyslexia using deep learning techniques has achieved an impressive classification performance of 96.4%, as proven by Aldehim et al. (2024). Although the method has successfully differentiated between normal and dyslexic handwriting, it does not specifically address dyslexia through the extraction and

analysis of reversal alphabet images. Consequently, this may affect the accuracy of the overall dyslexia-dysgraphia diagnosis.

Further improvement by Alqahtani et al. (2023) has proposed the identification of dyslexia through handwriting images by employing CNNs with Support Vector Machines (SVMs). The results show an impressive accuracy of 99.33% and this proved that combination of techniques enhances the classification accuracy. However, the absence of discourse regarding computational overhead and real-world applicability constrains its practical implementation. In addition, the classification work is limited by concentrating only on three distinct categories namely normal, reversal, and corrected all of which represent the types of character images.

Additionally, Zolna et al. (2019) have investigated dynamic handwriting analysis through the application of Recurrent Neural Networks (RNNs). The research analyzed RNNs in relation to CNNs and highlighted the significance of dynamic feature extraction as opposed to static image-based methodologies. As a results, the model has successfully identified more than 90% of children with the diagnosis, which presented the effectiveness of dynamic analysis in enhancing handwriting-based diagnosis.

Other than that, detection of dysgraphia by implementing statistical analysis has also shown promising results when CNN based feature extraction is implemented in conjunction with Mel Frequency Cepstral Coefficients (MFCCs) (Kariyawasam, & Samarasinghe, et al., 2019). The attained results show remarkable accuracy of up to 88% in the screening of alphabet dysgraphia letters. Thus, this study highlights the potential of integrating deep learning with interactive interventions using gamified environments.

The identification of dysgraphia has also explored feature extraction methods that analyze both spatial and kinematic aspects of handwriting. For example, Ghouse et al. (2022) introduced the NDR-R2CNN model, which integrates kinematic feature extraction and achieved a high accuracy of 98.2%. While this model improves analysis by filtering out irrelevant features, it has not been extensively compared with standard CNNs on real-world noisy datasets. Furthermore, kinematic measures alone were found to be insufficient for analyzing complex fractional-order deviation features, thus limiting their effectiveness. Similarly, the use of unsupervised deep learning remains limited with only one study by Saha et al. (2020) employing a Variational Autoencoder (VAE) to extract key handwriting features associated with dysgraphia. Although this approach is low-cost, non-invasive, and useful for identifying early signs of dysgraphia,

it lacks the ability to classify the severity levels of the condition. Thus, current feature extraction methods either fail to handle real-world noisy data, overlook complex handwriting deviations, or are unable to categorize dysgraphia severity, highlighting key gaps in existing approaches.

While the use of single-based models, such as CNNs and RNNs, for feature extraction demonstrates superiority over traditional OCR techniques, the persistent limitations in extracted features are due to the variability of handwriting features, which may pose challenges with diverse handwriting styles. This suggests that relying solely on single-model architectures may not be sufficient for capturing the full spectrum of handwriting characteristics associated with dysgraphia.

Moreover, most previous studies have mainly focused on binary classification and datasets limited to letter or word-level handwriting (Alqahtani et al., 2023; Sebastian et al., 2024; Vydeki et al., 2024). While these methods have shown good results they still exhibit limitations. To date, no studies have examined multiclass classification across three severity levels of dysgraphia for handwritten sentence data. The proposed classification must encompass a broader range of handwriting characteristics to accurately distinguish between complex variations in symptoms. This identified gap highlights the need for exploring fusion feature models. The integration of multiple complementary features has the potential to yield more nuanced representations, thereby enhancing the effectiveness of multiclass severity classification in the context of sentence-level handwriting analysis. Therefore, the fusion feature model is investigated to obtain greater variability in features (Kunhoth et al., 2024;Masood et al., 2023) which can yield improved differentiation in dysgraphia severity classes utilizing handwriting images.–

2.7.3 Limitations of Single-Based Feature Extraction Method in Image Classification

In image classification, the feature extraction stage is a crucial step that determine the accuracy of classification (G. Kumar & Bhatia, 2014; Liao et al., 2025). Broadly, feature extraction methods can be categorized into single-based extraction models and fusion-based extraction models.

Single-based extraction relies on a single feature extraction technique or model to represent image information. Although this approach can capture essential features

in handwriting images, it is inherently limited since it depends only on one type of representation. This feature may not be sufficient for complex handwriting images variations. These techniques may encompass conventional handcrafted features, such as Histogram of Oriented Gradients (HOG) which was employed by Wahid et al. (2021), as a handcrafted feature extraction for Bangla handwritten digits before the classification process using SVM, or contemporary deep learning models like Convolutional Neural Networks (CNNs), including architectures such as VGGNet or ResNet (Vydeki et al., 2024). These models are commonly used because of their ease of implementation and straightforwardness (Rawat & Wang, 2017). Nonetheless, they may have challenges in effectively generalizing in increasingly complex classification tasks, mainly when the image data includes varied textures or structural patterns (Bhatt et al., 2021). Thus, in the context of handwriting images, particularly regarding dysgraphia, substantial stylistic variation further complicates the identification of elements essential for the classification. Due to the constraints, the fusion-based extraction approach has been explored by integrating multiple feature extraction techniques at the feature level.

By extracting and combining features from varied sources such as texture, deep representation of image, and resolution information, it aims to exploit the complementary strengths of diverse feature modalities to improve the performance of the classification process. (Kunhoth et al., 2024; Waisi, 2024). For instance, the study by Anggraeny et al. (2023) has implemented Local Binary Patterns (LBP) to extract features of Javanese handwriting, combined with a Projection Profile Histogram. The study conducted by Kunhoth et al. (2024) has fused online and offline handwriting features into a feature vector by employing deep feature extraction pretrained by a CNN architecture and subsequently employing it within a classifier to differentiate between dysgraphia and non-dysgraphia handwriting. Therefore, classification performance can be significantly enhanced by extending feature extraction to encompass an extensive range of significant features and optimizing the utilization of all features in the classification process (Alevizos et al., 2024). Consequently, the fusion approach has enhanced classification performance by extracting more comprehensive and distinctive features, proving particularly effective in complex or varied image data scenarios.

2.7.4 Related Works on Feature Fusion-based Extraction Method

The feature extraction process is an essential task to determining the level of dysgraphia severity based on handwriting images. It is also important, especially when employing fusion techniques to integrate various data sources or feature types prior to classification. A variety of research has investigated diverse feature extraction methods, utilizing both deep learning and conventional machine learning models to enhance the performance of dysgraphia classification. Alongside the application of single deep learning feature extraction, numerous studies have explored the implementation of fused features for classifying and detecting dysgraphia through written inputs.

The study combining both online and offline handwriting features for dysgraphia diagnosis has been explored by Kunhoth et al. (2024). The EfficientNet-B7 was used to extract features of offline handwriting images, and online handwriting features were acquired by collecting dynamic pen stroke timing during writing. An accuracy of 88.8% was attained for these fusion features, facilitating the classification of dysgraphia, supported by a recall value of 90%. The implementation of transfer learning through feature extraction promises significant feature representation, which can enhance classification performance. In a similar context of utilizing a pretrained CNN model for feature extraction, a study conducted by Kunhoth, Al Maadeed, et al. (2023) investigated the classification of dysgraphia through various input modalities, achieving an impressive accuracy of 97.3% when employing SVM on the fused features of words, pseudo-words, difficult words, and sentences. This approach effectively combines features generated by CNNs with traditional machine learning classifiers, resulting in improved reliability.

In addition to integrating various modalities through fused features, the hybrid model has been examined by Masood et al. (2023), which employs CNN and LSTM. The fused features are subsequently classified utilizing Random Forest, and it has been asserted that performance has improved based on the extraction of multiple features, including stroke-based, shape-based, texture-based, and structural features. The integration of this methodology effectively encompasses both spatial and sequential dependencies, thereby enhancing its efficacy in the context of handwriting identifying abnormalities. While the model presents several advantages, it is essential to acknowledge that challenges, including selecting optimal hyperparameters and limitations incorporated into the dataset, significantly constrain its applicability. This

method offers a more interpretable framework for feature extraction than prior studies. Nonetheless, it is essential to acknowledge that this approach may face difficulties in generalizing to data that has not been encountered before.

The exploration of enhancing the categorization of handwritten features has led to the investigation of a hybrid approach that integrates CNN with Vision Transformers (ViTs) for the purpose of feature extraction, as documented in (Vilasini V et al., 2022). This investigation not only employed CNN and their hybrids or combinations with alternative CNN architectures but also integrated CNN with ViT for the purpose of feature extraction and the analysis of images at the pixel level. This study distinguishes itself from prior research that predominantly utilized pre-trained CNNs and handcrafted feature fusion techniques. Instead, it adopts the ViT framework, effectively capturing global dependencies within handwriting features by segmenting the image into fixed-size patches. This methodological shift results in enhanced sensitivity and specificity in the analysis. In contrast to the fusion of CNN and LSTM architectures (Masood et al., 2023), the interpretability associated with features derived from ViT (Vilasini V et al., 2022) presents a notable challenge. Furthermore, the effective application of these features may necessitate the utilization of extensive datasets.

Table 2.9
List of Research Conducted on Dysgraphia Classification Using Feature Fusion-Based Extraction Method

Authors	**Classification Result**	**Classification Method**	**Feature Extraction Method**	**Data Domain**
Kunhoth et al. (2024)	88.8% accuracy and 90% recall	SVM and XGBoost	EfficientNet-B7, to extract features from offline handwriting data. For online handwriting data, feature extraction focused on capturing dynamic aspects such as pen stroke timing and spatial features.	Both online and offline handwriting
Kunhoth, Al Maadeed, et al. (2023)	97.3% accuracy	SVM, AdaBoost, and Random Forest	DenseNet201 network pretrained on the ImageNet dataset as a feature extractor to generate CNN features from handwriting images	offline handwriting with different input of word, pseudo word, difficult word and sentences

Authors	Classification Result	Classification Method	Feature Extraction Method	Data Domain
Masood et al. (2023)	The CNN-LSTM model's accuracy is improved by the RF classifier	Random Forest	Employs a Convolutional Neural Network (CNN) as part of a CNN-LSTM model for feature extraction from handwritten characters	offline handwriting
Vilasini V et al. (2022)	96% sensitivity and specificity	Random Forest classifier detects dysgraphia using handwriting attributes	Employs Convolutional Neural Networks (CNN) and Vision Transformers (ViT) as automated deep learning methods for feature extraction from handwriting.	offline handwriting - character images

Table 2.9 presents previous studies that have categorized dysgraphia handwriting using feature fusion methodologies and the implementation of various classifiers. The use of feature fusion approaches in dysgraphia diagnosis has resulted in substantial improvement in accuracy and sensitivity. This technique provides an important advantage in that extraction to select or collect the best characteristics that will be used in the classification process, hence increasing classification performance. Furthermore, additional investigation into feature fusion algorithms may result in greater performance in classifying dysgraphia handwriting severity. Fusion-based extraction approaches performed better than single-based extraction, yielding more significant features that were employed in the classification process. Therefore, the fusion feature model is investigated to obtain greater variability in features, which can yield improved differentiation in dysgraphia severity classes utilizing handwriting images.

2.8 Classification Performance in Dysgraphia Detection based on Handwriting

Evaluating the classification performance of handwriting detection models is crucial for proving the model's strength and achieving meaningful results. While accuracy is the most frequently reported metric, researchers often employ more robust measures such as precision, recall, and the F1-score (Devi & Kavya, 2023; Ghouse et al., 2022; Kunhoth, Al Maadeed, et al., 2023; Kunhoth et al., 2024; Kunhoth, Maadeed, et al., 2023; Masood et al., 2023). Besides, most machine learning studies used Receiver Operating Characteristic (ROC) area under the curve (AUC) (Kunhoth, Al Maadeed, et

al., 2023; Kunhoth, Maadeed, et al., 2023) as a visual representation of results to provide a comprehensive evaluation of the model developed, particularly in conditions of class imbalance. Moreover, Matthews Correlation Coefficient (MCC) is a balanced performance metric used to evaluate the overall quality of a classification model (Devi & Kavya, 2023). MCC considers all four elements of the confusion matrix. This makes it a more reliable indicator of the model's accuracy for each class (Chicco et al., 2021).

In the context of existing research concerning the classification of dysgraphia through machine learning models utilizing handwriting images, it is notable that only Devi & Kavya (2023) has employed the Matthews Correlation Coefficient (MCC) as a balanced metric for binary classification. Nevertheless, in the context of multiclass severity classification, MCC implementation remains absent. Therefore, in addition to the ROC-AUC graph and F1-score as a balanced metric, this study used the MCC indicator to evaluate the multiclass classification model.

2.9 Research Gap of Automated Feature Extraction Method for Dysgraphia Classification Model

Feature extraction is a critical component in handwriting analysis, influencing the efficiency and accuracy of classification systems. Traditional hand-crafted techniques require manual identification of relevant features, which can be time-intensive and prone to inconsistencies. Automated feature extraction, particularly through deep learning, has gained prominence for its ability to derive meaningful representations from complex data. Residual networks, which incorporate skip connections, offer significant advantages by propagating input data properties across layers, thereby reducing errors and enhancing performance (Ramlan et al., 2023; Ramlan, Isa, Osman, et al., 2024). These methods have been shown to minimize background noise, preserve essential image features, and improve classification outcomes.

Table 2.10 provides a comprehensive summary of prior studies on dysgraphia detection and classification based on handwriting features. Most prior studies have focused on online handwriting analysis, emphasizing kinematic and dynamic characteristics to support the classification process. In contrast, research dedicated to offline handwriting, which involves converting handwritten physical content into digitized images, remains limited.

Table 2.10
State-of-the-Art Methodology and Performance Comparison with Previous Research

Author (Year)	Handwriting Type	Domain	Output (Binary/Multiclass)	Feature Extraction/ Feature Selection	Classifier	Performance
Drotár & Dobeš (2020)	Online	Slovak orthography	Binary	Based on online features of handwriting	AdaBoost	79.5% accuracy
Kunhoth, Al Maadeed, et al. (2023)	Offline	Slovak orthography	Binary	CNN features and fused CNN features	SVM, AdaBoost, and Random Forest	Highest performance by SVM, 97.3% accuracy
Masood et al. (2023)	Online	Parkinson's disease (PD) Dysgraphia English	Binary	CNN-LSTM feature extraction	Random Forest (RF)	The maximum accuracy of the proposed model is 96.1% and 97.6% for the uniform and normal distributions
Kunhoth, Maadeed, et al. (2023)	Online	Slovak orthography	Binary	In-air features extraction	KNN, SVM, Random Forest, and AdaBoost	AdaBoost classifier achieved a classification accuracy of 80.8%.
Devi & Kavya (2023)	Offline	English	Binary	Kekre-Discrete Cosine Transform with Deep Transfer Learning (K-DCT-DTL)	Deep Transfer Learning	The k-DCT-DTL approach attains 99.75% of the highest accuracy
Ghouse et al. (2022)	Offline	English	Binary	Non-Discrimination Regularization in Rotational Region CNN	CNN	Accuracy of 98.2 %
Devillaine et al. (2021)	Online	Drawing	Binary	Manual extraction	Random Forest, MLP, Extra Trees	An accuracy of 73%

Finally, the research gaps in automated feature extraction methods were synthesized. Key gaps identified include the lack of specialized handwriting datasets for dysgraphia, insufficient exploration of fusion-based feature extraction models in this domain, and limited attention to pre-processing techniques like line removal that could enhance classification performance. This review underscores the need for a more robust and tailored approach to dysgraphia detection using handwriting images that integrates

advanced pre-processing, diverse datasets, and hybrid feature extraction models to improve detection accuracy and overall dysgraphia assessment.

2.10 Summary

This chapter comprehensively reviews the literature and demonstrates substantial progress in dysgraphia detection and classification using conventional, machine-learning, and deep-learning approaches. Conventional assessment methods, which rely on expert observation of handwriting samples, are inherently subjective, time-consuming, and difficult to scale for large populations. Although intelligence-based methods have improved classification performance, several critical limitations remain. Most existing studies focus on binary classification and on character- or word-level handwriting data, with limited attention to multiclass severity classification using sentence-level handwriting, particularly in the Malay-language context. Additionally, existing datasets are often limited in diversity and size, restricting the generalizability of developed models.

Furthermore, conventional machine learning approaches rely heavily on handcrafted features, which are insufficient to capture the complex variability in dysgraphia handwriting. While deep learning models, particularly CNNs, have demonstrated strong capabilities in automatic feature extraction, their effectiveness remains constrained when dealing with highly variable and inconsistent handwriting patterns. Furthermore, single-feature extraction approaches are limited in representing the full spectrum of handwriting characteristics, whereas fusion-based approaches, although promising, remain underexplored, particularly those involving dual-input representations and sentence-level data.

Preprocessing challenges, particularly in removing ruling lines and irrelevant artifacts, further impact feature quality and classification performance. In addition, current evaluation practices are largely limited to conventional metrics such as accuracy and F1-score, with insufficient emphasis on robust statistical validation methods, particularly for multiclass classification. These gaps highlight the need for an integrated framework that combines effective preprocessing, robust feature extraction, fusion techniques, and comprehensive evaluation strategies.

To address these limitations, this study proposes a multiclass dysgraphia severity classification framework based on dual-input handwriting images, namely

Handwriting Images with Removed Ruling Lines (HIRRL) and Histogram-based representations (Histo-HIRRL). This approach leverages fusion-based feature techniques to capture meaningful characteristics of handwriting, enabling a more comprehensive feature representation for distinguishing subtle severity levels.

For feature extraction, an adopted Residual Network architecture, specifically ResNet-18, is selected. Residual networks are well-suited for this task due to their ability to mitigate vanishing gradients through skip connections, enabling effective learning of deep hierarchical features. Compared to deeper variants, they provide a more efficient architecture with fewer parameters, making it particularly suitable for moderate-sized datasets and reducing the risk of overfitting. This balance between model complexity and computational efficiency is crucial to ensuring stable performance in dysgraphia classification tasks with limited, highly variable handwriting data.

Furthermore, integrating multiple machine learning classifiers and comprehensive evaluation methods, including Matthews Correlation Coefficient (MCC) and expert validation, enhances the reliability and robustness of the classification results. Overall, this study aims to bridge the identified research gaps by developing a scalable, efficient, and interpretable framework for multiclass dysgraphia severity classification using offline handwriting data. This review highlights the extensive use of ML and DL techniques in handwriting analysis, emphasizing various preprocessing methods to enhance accuracy and support effective identification of dysgraphia through handwriting images.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

A computer-aided dysgraphia detection and classification system requires optimal image pre-processing algorithms and a reliable machine learning model to accurately detect and classify dysgraphia symptoms. This study aims to classify the severity levels of dysgraphia handwriting using a Convolutional Neural Network (CNN)-based fusion feature extraction from dual-input images and machine learning classifier. This chapter introduces a novel fusion-based approach that combines feature vectors from handwriting and histogram images using a deep CNN to perform multiclass classification of dysgraphia severity at the sentence level using machine learning classifier.

The chapter begins by discussing the setup for handwriting acquisition and data collection and introduces a new scheme for image pre-processing to remove ruling lines from handwriting images. This chapter then presents the performance of pre-processing by measuring the structural similarity index and classification performance between images with and without a ruling line, in order to select the best input for the classification model. Next, develop a new feature-based model, namely H2FCD (Handwriting and Histogram Features for Classifying Dysgraphia severity), by extracting features using a residual network layer applied to the newly generated input. The generated inputs include handwriting images without ruling lines in two forms; resized handwriting images and histogram images.

After that, the chapter investigates the performance of the fusion features-based model through the classification performance of dysgraphia severity by concatenating the feature vectors and substituting with several classifiers such as support vector machine (SVM), k-nearest neighbor (KNN), random forest (RF) and Bagged Tree (BT). Then, a new prototype of an automated tool was introduced to assist in classifying the severity of dysgraphia using offline handwriting input. Finally, the proposed model was validated by eight experts, and a correlation analysis was conducted between the experts' evaluations and the results of the proposed H2FCD model. The visualization of the overall research design is presented in detail in Section 3.2.

3.2 Research Design

This section presents a comprehensive research design by outlining the methodological procedure, which comprises four phases: developing a dysgraphia handwriting dataset, image pre-processing to remove ruling lines, multiclass classification using fused features and performance evaluation. The proposed research design is illustrated in Figure 3.1, depicting the overall flow and relationships among each phase and objective.

The first phase is aimed to develop a dysgraphia handwriting dataset named Malay Language Dysgraphia Handwriting Dataset (MLHSD) that started with ethical approval. Ethical approval is necessary for this research as it involves the collection and analysis of handwritten data produced by human participants, which may include sensitive personal information. To ensure the study followed the ethical standards and protected the rights and privacy of all participants, approval was obtained from the Research Ethics Committee (REC) of Universiti Teknologi MARA (UiTM) with reference number 600-TNCPI(5/1/6) REC/09/2021 (MR/830 (Appendix 1). The data collection approval was obtained from the MOE through the eRAS 2.0 system (Appendix 2). This approval is required to ensure compliance with governmental policies and ethical standards related to research conducted within educational settings and involving student participants. At this phase, the tasks involved data acquisition along with PDM (Appendix 3), which was then evaluated by PDM's assessors to classify dysgraphia severity in handwriting. Furthermore, all collected data from the paper writing were converted into images in digital format. This process ultimately yields a dataset of children's writing categorized into three levels of dysgraphia severity. The detailed process is explained in Section 3.3.

In Phase 2, the new scheme for the ruling line removal method, called the Red-Green-Blue-Black (RGBB)-based method, is presented and further explained in Section 3.4. This scheme was proposed for image rotation to facilitate alignment, detecting the ruling line, replacing pixel colors, and reconstructing handwriting due to displaced pixels during the removal of the ruling line process.

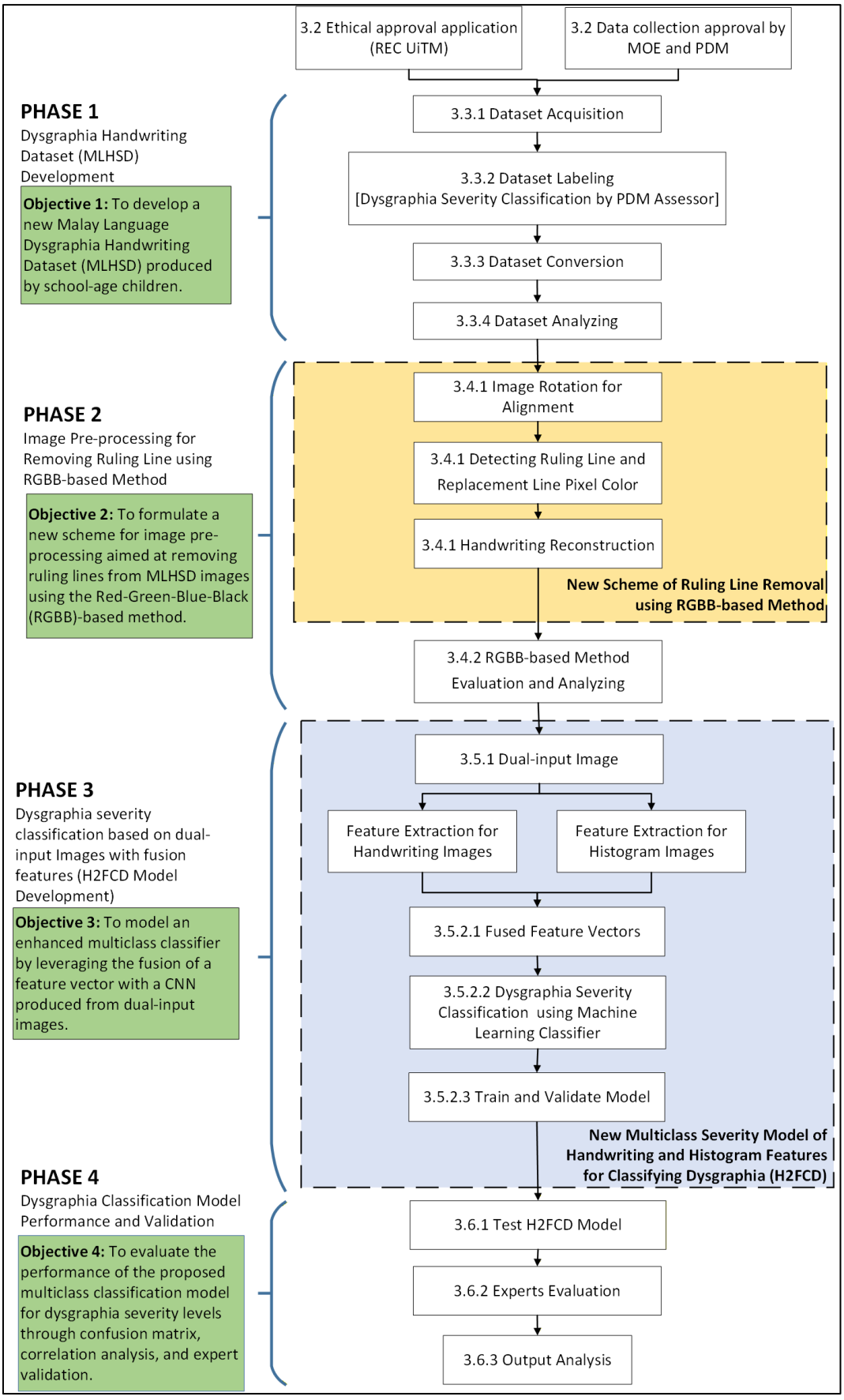


Figure 3.1 Research Design For An Automated Dysgraphia Severity Classification Model

In Phase 3, the dysgraphia severity classification was performed based on fused features of dual input images. Besides, the design of the classification model was developed based on the severity classes of handwritten images with removed ruling lines (HIRRL) obtained in Phase two. The dual input consists of HIRRL and the histogram image (Histo-HIRRL) generated from the handwritten image (in phase 2). Furthermore, both input images were extracted using deep feature extraction of residual network, and the features are sequentially fused to serve as input in the severity classification process. In the classification process, several machine learning classifier models were used, including Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Subspace Discriminant (SD), Random Forest (RF), and Bagged Tree (BT). Then, each classification model classifies the severity level of dysgraphia as High Risk, Moderate Risk, and Low Risk. The details of Phase 3 are explained in Section 3.5.

Finally, Phase 4 involves the performance and validation of the classification model. This final encompasses two parts which are testing the H2FCD model and expert evaluation. The model's performance is validated by eight experts, who assess handwriting images and compare the classification results with those obtained from the proposed H2FCD model, which uses fused features from dual input images. Finally, these two classification outputs are then measured to obtain the correlation analysis. Further explanation is presented in Section 3.6.

The research design serves as the main picture for the overall process described in this chapter. The details are presented in the appropriate section. The next section will explain Phase 1.

3.3 Phase 1: Development of New Dysgraphia Handwriting Dataset (MLHSD)

This section explains the process of dataset development and handwriting image preparation, including data acquisition, the labeling process by two experts' assessment, and the conversion process. In this study, this phase is required as an initial process to prepare the dataset used for supervised learning in automated dysgraphia classification model. Essentially, dataset development is crucial to ensure that a high-quality image of handwriting is prepared for model training in automated classification. Figure 3.2 illustrates the overall process of dataset development, beginning with data acquisition and continuing until the conversion of images into digital form and the completion of data analysis.

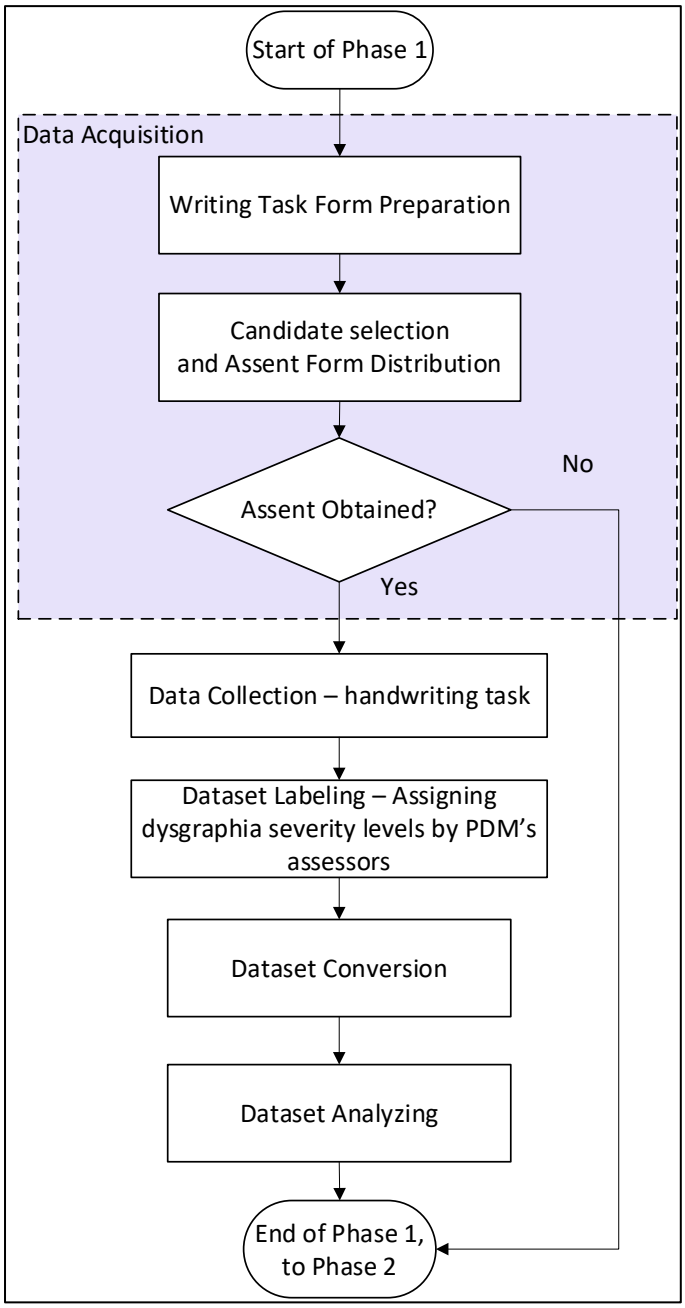


Figure 3.2 Overall Process of the Dysgraphia Dataset Development

3.3.1 Data Acquisition

The data acquisition process involved multiple sub-processes aimed at collecting handwriting data from children. The multiple processes encompass preparing forms, selecting and distributing the assent form, and collecting data after obtaining assent.

Firstly, the form preparation process includes designing child-friendly writing task sheets on A4 size paper specifically intended for children to use in writing tasks.

The form consists of demographic information and statements that participants were required to replicate and write in a designated space under each statement, as depicted in Appendix 4. The writing process was explained and guided by the instructors, i.e., dyslexia experts and schoolteachers. The writing task form must be thoroughly completed to uphold ethical standards. Concurrently, written assent was obtained from participants through a formal agreement form prior to participation (see Appendix 5). The assent form is designed to ensure that participants understand the purpose, procedures, risks, and confidentiality of the study before taking part in data collection. Through the assent form agreement, instructors explain the project in simple language, including why the research is being conducted, what the participant is required to do, and the voluntary nature of participation. The form also includes questions to confirm the child's understanding and agreement, along with spaces for their name, signature, or thumbprint. The objective of completing this process is to obtain the participant's willingness to join the study in an ethical and informed manner, ensuring respect for their rights and autonomy before any handwriting data is collected.

Data collection is a crucial process of collecting the handwriting produced by participants. This process was implemented at PDM and MOE primary schools. Among the students participating in this study were children who did not have any other learning disability, such as autism, or Attention Deficit Hyperactivity Disorder (ADHD). Meanwhile, the participants who were children with writing disorder issues have fulfilled the inclusion and exclusion criteria as listed below.

The inclusion criteria are:

i) Students aged between 6 and 12 years old.
ii) Left or right-handedness
iii) Gender

The exclusion criteria are:

i) Participants who have any combination of symptoms of other learning disabilities, such as autism and ADHD
ii) Participant's name
iii) Area of the intervention center

Based on the defined criteria, a quota sampling method was employed as a non-probability sampling technique due to the specific characteristics of the target population, namely, participants in special-needs subgroups of the learning-disability spectrum. This approach ensures appropriate representation aligned with the study objectives. Prior studies by Ileri et al. (2020), Dankovicova et al. (2019), and Kurniawan et al. (2018) report sample sizes ranging from 30 to 120, supporting the selected range. The estimated number of participants was further determined using the sample size formulas presented in Equations (3.1) and (3.2).

where,

α = confidence level (95%),

e = margin of error (5%),

p = population proportion (10%),

N = population size (400).

The sample size (n) is calculated according to the formula:

$$q=1+\left(\frac{z^2p(1+p)}{e^2N}\right) \tag{3.1}$$

$$n=\frac{z^2p(1-p)}{e^2}\times\frac{1}{q} \tag{3.2}$$

Based on the sample size calculation using a 95% confidence level, 5% margin of error, population proportion of 0.10, and a population size of 400, the estimated sample size obtained after applying finite population correction is approximately 103 participants. Therefore, the total number of participants (n = 178) exceeds the minimum required sample size of 103, indicating that the dataset is sufficiently large to support reliable statistical analysis and model development in this study.

Data collection refers to the collection of handwriting produced by participants among school-age children. This process was executed at the Malaysian Dyslexia Association (PDM) in Sungai Petani, Kedah and Ampang, Kuala Lumpur Branch, and Sekolah Kebangsaan Seberang Jaya 1, Seberang Perai, Pulau Pinang, and the participants were accommodated with letters of consent. A total of 178 participants, aged between 6 and 12 years old, participated in this data collection session (Rajput et al., 2023; Riley et al., 2025), as summarized in Table 3.1. A total of 178 children participated, ranging in age from 6 to 12 years. The largest group was at age 8 with 43 participants, followed by age 7 with 38 participants. Participation then gradually declined with increasing age, with the smallest group being those aged 10, comprising 14 participants. This distribution suggests a stronger representation of younger children (Coradinho et al., 2023), aged 7 to 9, providing a broad sample across primary school years for handwriting data collection.

Table 3.1
The Number of Participants For Each Age Group

Age	**Number of Participants**
6	18
7	38
8	43
9	27
10	14
11	20
12	18
Total	178

Meanwhile, the dataset distribution based on source, gender, and handwriting dominance is presented in Table 3.2. All 104 of the 178 participants from PDM have attended this session, while the remaining participants were from the MOE primary school. Moreover, of all 178 participants, there were 119 male students, and the remaining were female. Among all participants, only 3.93%, or seven participants who took part in writing, were left-handed.

Table 3.2
Distribution of Participants by Source, Gender, and Writing Dominance

	Criteria	Number of Participants
Location	PDM	104
	KPM school	74
Gender Distribution	Male	119
	Female	59
Handedness	Right	171
	Left	7

The participants were instructed to write on the provided paper by copying the given sentences (Biotteau et al., 2019; Rosenblum et al., 2004). Figure 3.3 presents the example of a sentence in the writing task form. Upon completion, the instructor gathered all submitted forms and proceeded to conduct a labelling process by assessing the severity level according to the handwritten products of the participants.

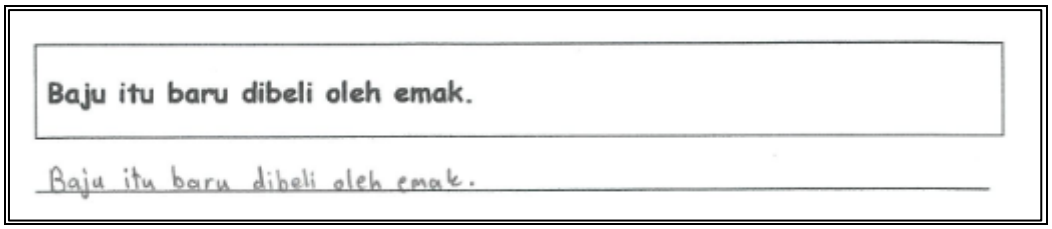


Figure 3.3 A Sentence Produced By A Participant in the Writing Task Form

The following section explains the overall process of labelling each data samples of dysgraphia severity levels by assessors who were experts in the evaluation process.

3.3.2 Labeling Dataset: Assigning Dysgraphia Severity Levels by PDM's Assessors

Once the dataset of handwritten images had been collected, the next process was to label each data sample by evaluating dysgraphia symptoms and severity levels. The aim of this task is to assign categories of the severity levels of dysgraphia symptoms presented in the handwriting for machine learning input training and testing in classification model. This task was performed manually by certified PDM assessors.

Figure 3.4 illustrates the process of assigning severity levels, which comprises observation results, reliability analysis, data selection, and dysgraphia categories. The classification results of dysgraphia severity were then used as a reference label for the training model in the next forward phase.

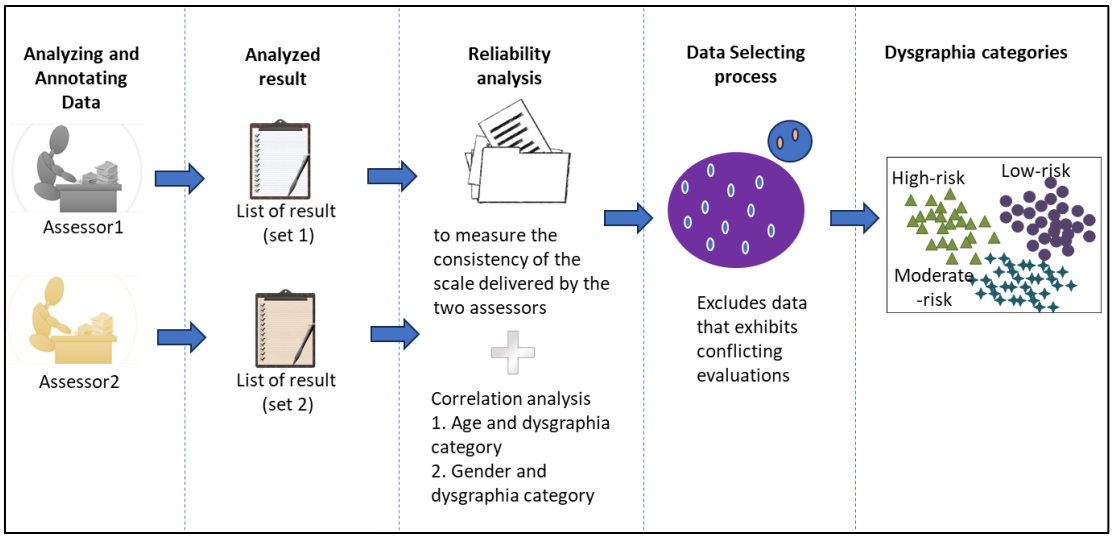


Figure 3.4 Evaluation Process For Assigning Severity Levels of Dysgraphia Symptoms Present in the Handwritten

In this stage, the labelling process was conducted through expert assessment by two professionals of PDM's assessor (interrater) to rate the severity levels of dysgraphia. The severity levels are evaluated accordingly into three classes namely as high-risk, moderate-risk, and low-risk (Sihwi et al., 2019) . The assessors evaluated the handwriting by analyzing each data sample and providing accurate annotations based on dysgraphia symptoms and criteria as explained in Table 3.3.

Table 3.3
Dysgraphia Symptoms and Criteria (Biotteau et al., 2019; P. J. Chung et al., 2020)

Symptoms	**Criteria**
Incorrect spelling and capitalization.	Capital letter error in the initial sentence. Spelling error for short words.
Inappropriate sizing and spacing of letters.	Inconsistent letter sizing. Inconsistent letter spacing.

Symptoms	Criteria
Mix of cursive and print letters words.	A change in writing style within a single word, such as shifting between uppercase and lowercase letters, alterations in letter slant or formation, and irregular pressure applied, resulting in varying thickness
Difficulty visualizing words before writing.	Inability to visualize a word before writing, such as spelling errors and misplaced letters.
Omitting letters and words from sentences	Letters omitted in a sentence Words omitted in a sentence
Poor letter formation	Unclear letter formation. Examples include letters with incomplete loops, irregular sizes, or ambiguous shapes that make it hard to distinguish between similar letters (e.g., "a" and "o" or "n" and "h").
Difficulty with spacing text.	Uneven spacing between words. This issue refers to irregular or inconsistent gaps between words in writing, which can make text harder to read. Difficulty distinguishing letter shapes
Has difficulty with shape discrimination and letter spacing	Difficulty distinguishing letter shapes. struggles to recognize or differentiate between visually similar letters, such as b/d, p/q, or m/n.
Trouble organizing words from left to right.	Disorganized word arrangement
Cannot stay within the margins.	Handwriting that does not stay within the designated boundaries of the page.
Do not write complete sentences.	Subject or predicate missing in a sentence.

Since more than one symptom might occurs in the handwriting, all the severity levels identified were calculated by sum of errors found in the handwriting task form (American Psychiatric Association, 2013; Jusoh, 2024). Consequently, the severity levels are categorized according to the weight of error in writing, as shown in Table 3.4 The errors that exceed 80% are classified as high risk. If the error is between 50% and 80%, it is classified as a moderate risk of dysgraphia, whereas less than 50% is classified as low risk. All dataset was manually organized and stored in separate folders according to the respective categories. The selection of three severity classes (high risk, moderate risk, and low risk) is based on the standardized assessment criteria established by the PDM, which certified assessors use to evaluate dysgraphia symptoms in children (Jusoh, 2024). These categories reflect the actual clinical classification applied during the assessment process.

Table 3.4
Dysgraphia severity Levels According to the Error Weight Suggested by PDM Jusoh, 2024)

Weight of Error in Writing	Dysgraphia Severity Class
> 80% of criteria severity marked	High Risk (HR)
50% - 80% of criteria severity marked	Moderate Risk (MR)
< 50%	Low Risk (LR)

3.3.3 Dataset Conversion

The process of image conversion refers to the process of converting the hardcopy format of all handwriting samples into a digital format. As depicted in Figure 3.5, the digitization process commences with scanning the hardcopy set using a Fuji Xerox ApeosPort 3560 machine scanner at 200 x 200 dpi and saving it in JPG format according to their respective category (severity level).

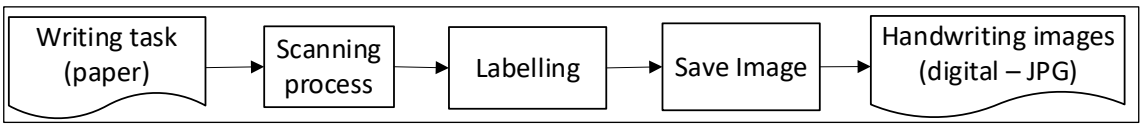


Figure 3.5 Digitization Process

Appendix 6 is an example of digitized samples and is saved in JPG format. Upon completion of the data digitization process, the cropping procedure is implemented to separate the written sentences into a single statement, as shown in Figure 3.6.

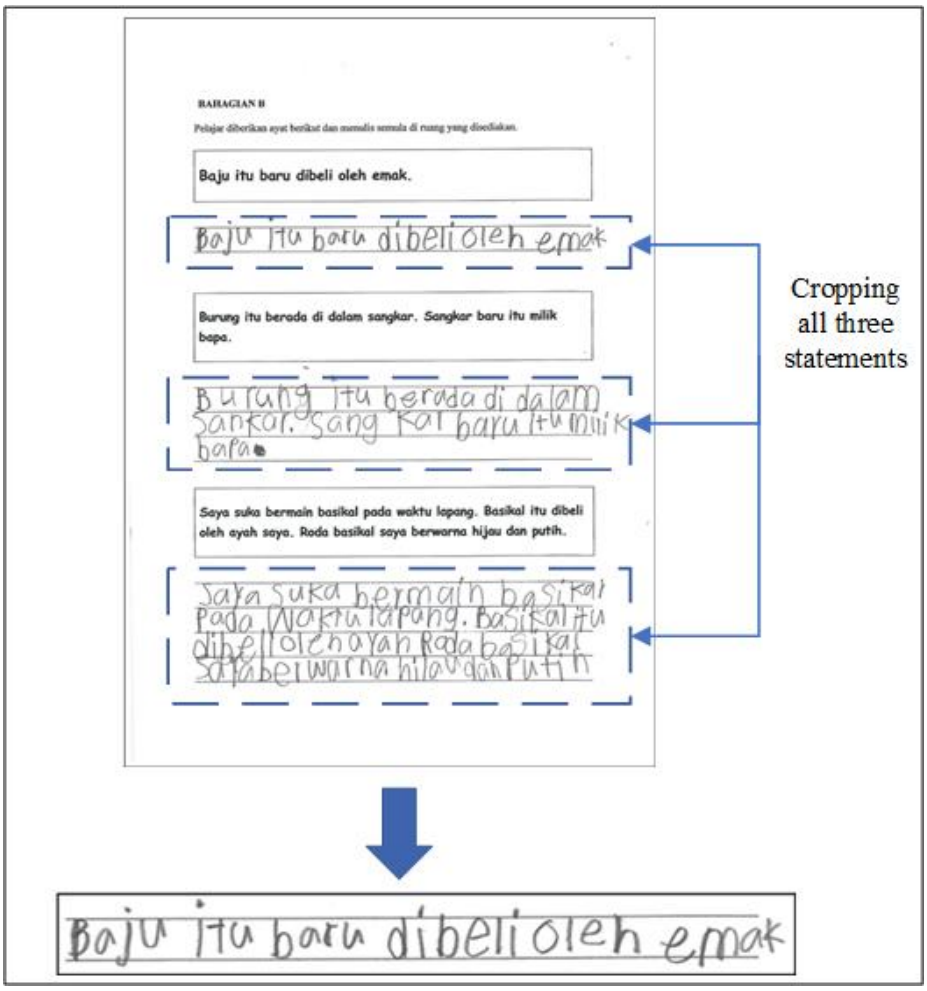


Figure 3.6 The Cropping Process of the Handwriting Sample

Based on Table 3.5, 172 handwriting samples of digitized images were gathered, 48 of which were categorized as high-risk, 59 as moderate-risk, and the remaining as low-risk. The cropped image contains 516 handwriting images, with 144 high-risk, 177 moderate-risk, and 195 low-risk.

Table 3.5
The Number of Images Digitized Form and Cropped Image

	High-risk	**Moderate-risk**	**Low-risk**	**Total**
Digitized form	48	59	65	172
Cropped image	144	177	195	516

3.3.4 Analysing Dataset using Cohen's Weighted Kappa and Spearman Correlation Coefficient

In this study, there are two assessors involved in the evaluation process to mitigate the potential bias that may arise from a single assessor's rating. In order to acquire the reliability of the data, correlation analysis is calculated to ensure the consistency of the ratings interpretation result.

This phase refers to a process of analyzing the collected dataset to measure the consistency of agreement and correlation between assessors in classifying the severity levels of dysgraphia. Therefore, the reliability of Cronbach's alpha, α, is measured to ensure the trustworthiness of the data before proceeding with further activities to determine interrater reliability between two assessors who completed the manual evaluation by ranking the severity of dysgraphia in handwritten samples. After that, both agreement and correlation analyses are measured using Weighted Cohen's Kappa (Sun, 2011) and Spearman's Correlation Coefficient (Schober & Schwarte, 2018) for the ordinal ranking level of dysgraphia severity assigned by two assessors. Eq. (3.3) is weighted Cohen's Kappa used to investigate the agreement between two assessors.

$$\kappa_w = 1 - \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} O_{ij}}{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} E_{ij}} \tag{3.3}$$

where,

k = number of severity categories (label 0, 1, 2 for low, moderate and high risk)

O_{ij} =The proportion of classes where Assessor 1 assigns category i and Assessor 2 assigns category j.

E_{ij} =The proportion of handwriting samples that would fall into Assessor 1 = i and Assessor 2 =j if both Assessors' decisions were random but followed their overall rating tendencies.

w_{ij} =The weight given to the level of agreement between Assessor 1 and Assessor 2

Table 3.6 shows the level of agreement for the weighted Cohen's kappa interpretation for each κ value. If the κ value closest to 1 indicates that both assessors

have almost perfect agreement in ranking the severity of dysgraphia for the handwriting image samples.

Table 3.6
Level of Agreement For Weighted Cohen's Kappa Analysis (Mary L. McHugh, 2012)

κ value	**Level of Agreement**
<0.00	Poor agreement
0.00< k < 0.20	Slight agreement
0.21 < k < 0.40	Fair agreement
0.41 < k< 0.60	Moderate agreement
0.61 < k< 0.80	Substantial agreement
0.81 <k< 1.00	Almost perfect to perfect agreement

Furthermore, this activity also implemented statistical analysis to measure the simple correlation coefficient as the reliability scores. The correlation coefficient,ρ, is presented as formulated in Eq. (3.4) and presented within the range of -1 to 1.

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{3.4}$$

where:

ρ = correlation coefficient,
n= number of rated items,
$d_i = R_{1i} - R_{2i}$ = difference between the ranks assigned by Assessor 1 and Assessor 2 to the item i.

The coefficient of the ρ value closest to 1 shows that the correlation between the raters is highly correlated and thus indicates a positive correlation. In contrast, the negatively correlated is presented when the ρ value is close to -1, which indicates a low agreement between the raters. Nevertheless, if the correlation coefficient is presented as 0, this indicates that there is definitely no correlation between the raters. In

conclusion, the overall interpretation of the correlation coefficient,ρ value, is tabulated in Table 3.7.

Table 3.7
Level of Spearman's Correlation Coefficient Analysis (Alsaqr, 2021; Mukaka, 2012)

$\boldsymbol{\rho}$ **value**	**Level of Correlation**
$0.00 < \boldsymbol{\rho} < 0.10$ or $-0.10 < \boldsymbol{\rho} < 0.00$	No or negligible correlation
$0.10 < \boldsymbol{\rho} < 0.39$ or $-0.39 < \boldsymbol{\rho} < -0.10$	Weak positive/negative correlation
$0.40 < \boldsymbol{\rho} < 0.69$ or $-0.69 < \boldsymbol{\rho} < -0.40$	Moderate positive/negative correlation
$0.70 < \boldsymbol{\rho} < 0.89$ or $-0.89 < \boldsymbol{\rho} < -0.70$	Strong positive/negative correlation
$0.90 < \boldsymbol{\rho} < 1.00$ or $-1.00 < \boldsymbol{\rho} < -0.90$	Very Strong positive/negative correlation

Next, the collected dataset was analyzed in terms of the relationship between age and the severity levels of dysgraphia symptoms among participants. Thus, the statistical measure of Pearson correlation,r was used to determine whether the age and gender of the participants had an inverse effect on the severity level. The Pearson correlation coefficient, is presented as formulated in Eq. (3.5) and presented within the range of -1 to 1.

$$r = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum x_i - \bar{x}^2 \sum y_i - \bar{y}^2}} \tag{3.5}$$

where:

r = correlation coefficient,
x_i= values of the x-variable in a sample,
$\bar{x}$= mean of the values of the x-variable,
y_i= values of the y-variable in a sample,
$\bar{y}$= mean of the values of the y-variable

The coefficient of the *r* value closest to 1 indicates that the correlation between the raters is highly significant and thus suggests a positive correlation. In contrast, a negatively correlated relationship is presented when the r value is close to -1, indicating a low level of association between the raters. Nevertheless, if the correlation

coefficient is presented as 0, this indicates that there is definitely no correlation between the raters. In conclusion, the overall interpretation of the correlation coefficient, *r* value, is tabulated in Table 3.8.

Table 3.8
Level of Pearson correlation Coefficient Analysis (Mukaka, 2012)

r value	**Level of Correlation**
0.00< r < 0.10 or -0.10 < r < 0.00	No or negligible correlation
0.10 < r < 0.39 or -0.39 < r < -0.10	Weak positive/negative correlation
0.40 < r < 0.69 or -0.69 < r < -0.40	Moderate positive/negative correlation
0.70 < r < 0.89 or -0.89 < r < -0.70	Strong positive/negative correlation
0.90 < r < 1.00 or -1.00 < r < -0.90	Very Strong positive/negative correlation

Data selection that is required to exclude any data that exhibits conflicting evaluations of two assessors. Once all the datasets have been analyzed accordingly, the next phase is performed image preprocessing techniques on the datasets, as further explained in Section 3.4. The MLHSD dataset was developed once all the activities in phase one are completed. The subsequent phase utilized the MLHSD dataset to achieve the second objective and beyond.

3.4 Phase 2: Proposed New Image Pre-Processing Scheme for Removing Ruling Line

This section describes in detail of the proposed image pre-processing scheme for removing the ruling line that appears on the MLHSD dataset, as depicted in Figure 3.7. Although the ruling lines are an important part of the writing task form, they need to be removed using several image-processing techniques, as their presence could affect the classification model's performance. This is because the ruling lines are part of the background and do not relate to the content of the extracted handwriting. These lines can be mistakenly interpreted as part of the handwriting strokes, which could cause confusion during feature extraction (Gold & Zesch, 2022). As a result, the extracted features may be distorted, thereby reducing the model's ability to learn and classify handwriting patterns effectively.

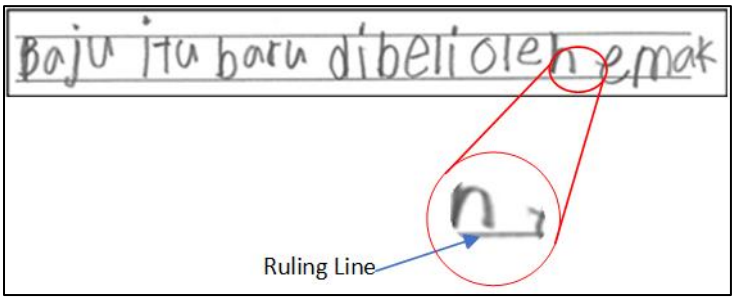


Figure 3.7 Handwritten Image With Ruling Line (HIRL)

This phase begins with the removal of ruling line using a new proposed scheme that employed the histogram projection of all pixel colors, including Red, Green, Blue, and Black, referred to as an RGBB-based method. This new scheme consists of three main steps: image rotation for alignment, detection of the ruling line, and line replacement. The next step is followed by handwriting reconstruction and finalized by analyzing performance, as presented in Figure 3.8. This approach is systematically executed to develop a unique information of handwriting features from data images through a comparative analysis to the ideal data image, thereby facilitating improvements in classification performance during phase 3.

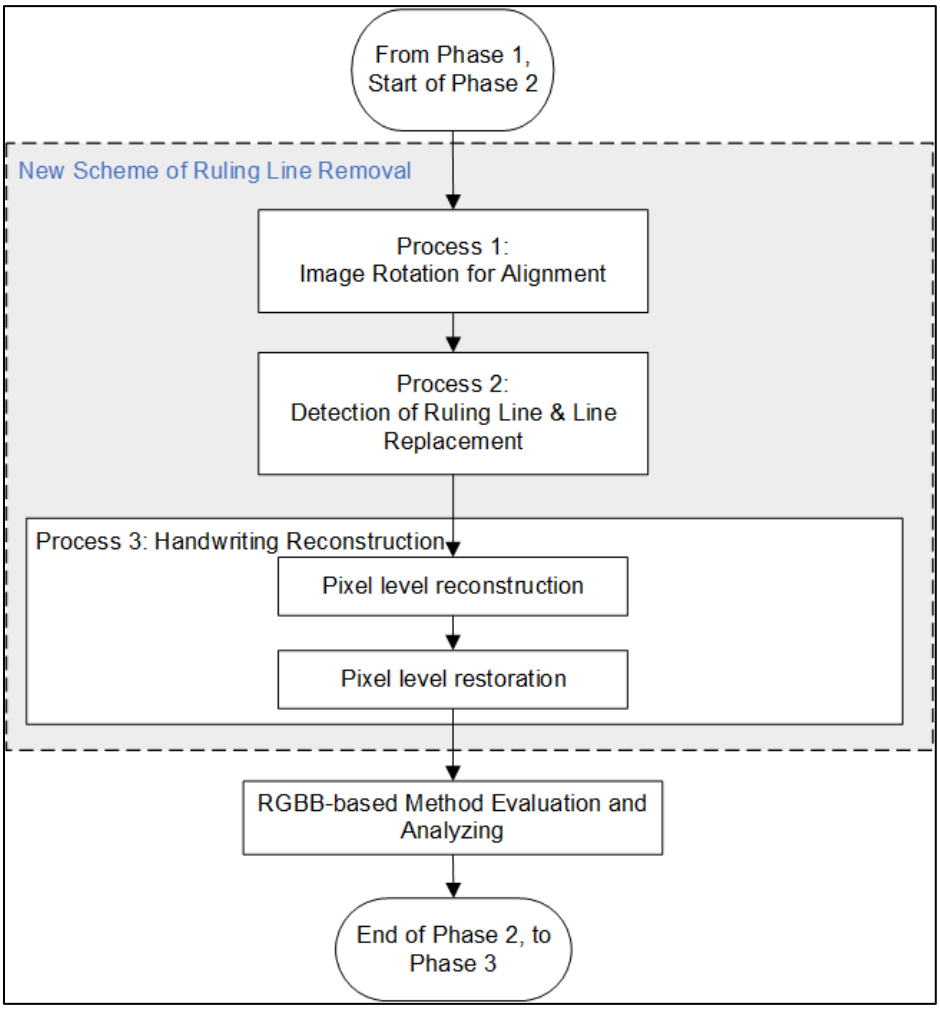


Figure 3.8 Image Pre-Processing for Ruling Line Removal

3.4.1 New Scheme of Ruling Line Removal Using RGBB-based Algorithm

The process of ruling line removal is essential to guarantee that the integrity of handwriting information remains unaffected by the visual characteristics of the lines. Therefore, a new scheme of ruling line removal was proposed using an RGBB-based algorithm and combining several image processing techniques. As presented in Figure 3.8, the proposed scheme for removing the ruling line has three subprocesses. It begins with image rotation and alignment, detecting the ruling line and replacing the ruling line pixel with a background color, and reconstructing the truncated letter. The following subsection provides an in-depth explanation of the proposed scheme for removing ruling lines.

3.4.1.1 Process 1: Image Rotation for Alignment

Figure 3.9 shows an example of an original image with a black ruling line that is misaligned with respect to the red reference axis. The small gap on the right side of the ruling line between the black and red lines indicates that a rotation is necessary to align them properly.

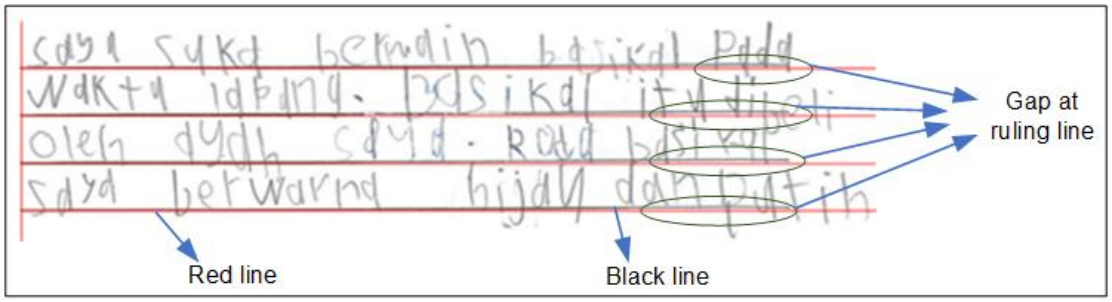


Figure 3.9 Original image With Axis Misalignment

Figure 3.10 shows the overall process for image rotation for alignment in the proposed scheme of ruling line removal using the RGBB-based algorithm. An image with a ruling line is rotated using histogram analysis. Thus, it chooses the optimal alignment to stop iterating and outputs the actual alignment image.

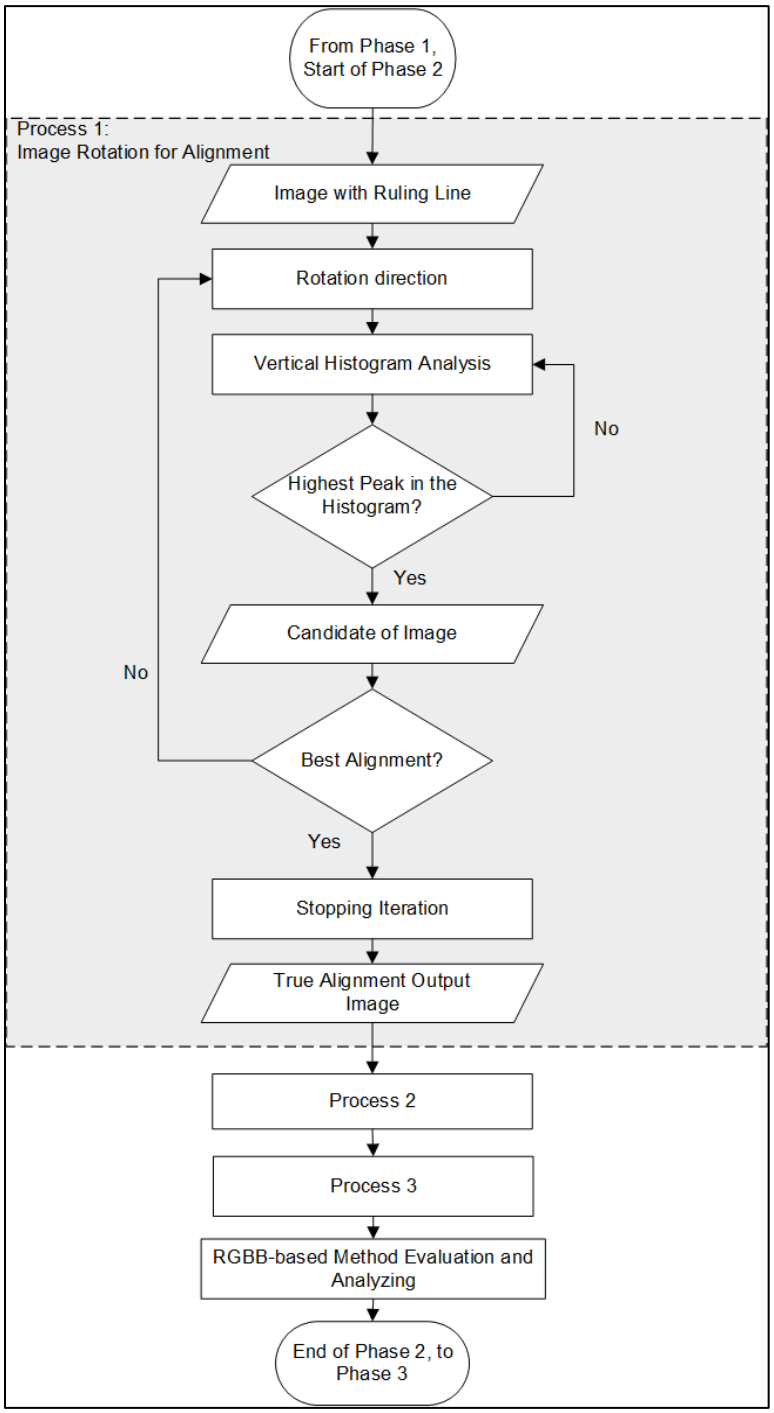


Figure 3.10 Overall Process For Image Rotation For Alignment in Proposed Scheme of Ruling Line Removal Using RGBB-Based Algorithm

The overall process of image rotation for alignment in the proposed scheme of ruling line removal is illustrated in the flowchart. Based on this flow, the detailed procedure is described as follows:

i) Present the original image to the alignment module.
ii) Generate three candidate images by applying clockwise rotation, no rotation, and counterclockwise rotation with a fixed step angle.
iii) Compute the vertical histogram for each candidate image. The histogram represents the column-wise distribution of pixel intensities.
iv) Identify the candidate with the highest histogram peak, as this indicates stronger alignment of text lines. Assign this candidate as the updated image.
v) Halve the step angle to refine the rotation process and repeat the procedure.
vi) Stop the iteration when the improvement in histogram peak is negligible or when the maximum iteration count is reached.
vii) The final output is the aligned image, which is then used for subsequent processing tasks.

Therefore, the first step in removing the ruling line is to ensure the images are as horizontally aligned as possible (Rehman & Saba, 2011). Algorithm 1 presents the pseudo-code for the image rotation scheme used to align the handwriting image. The illustration of the rotation process is shown in Figure 3.11 with +N° and -N° value of angle .

Algorithm 1: Image Rotation for Alignment

```
Input: image, search_range, step, window_size, line_length
Output: aligned_image, best_angle
Function score(image):
    hist_R ← per-row count of dark pixels in R channel
    hist_G ← per-row count of dark pixels in G channel
    hist_B ← per-row count of dark pixels in B channel
    hist_R ← moving_average(hist_R, window_size)
    hist_G ← moving_average(hist_G, window_size)
    hist_B ← moving_average(hist_B, window_size)
    peakiness ← avg(max(hist)/mean(hist) for hist in [hist_R, hist_G, hist_B])
    lines ← count_long_horizontal_lines(image, length=line_length)
    Return peakiness + lines
End
best_angle ← 0
best_score ← -∞
For angle From -search_range To +search_range STEP step:
```

```
17          rotated ← rotate(image, angle, fill=white)
18          s ← score(rotated)
19          If s > best_score:
20              best_score ← s
21              best_angle ← angle
            End
        End
22      aligned_image ← rotate(image, best_angle, fill=white)
```

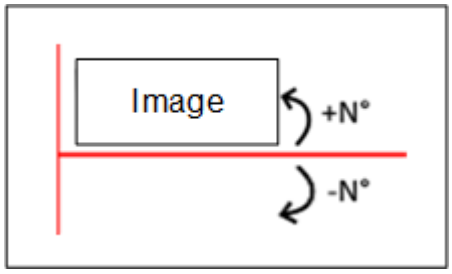


Figure 3.11 Image Rotation Direction of +N ° or -N°

The begins by defining a scoring function that evaluates the quality of alignment for a given rotated image. First, the per-row count of dark pixels is computed for each of the three RGB color channels (Steps 4–6). These histograms represent the horizontal distribution of text or ruling line structures within the image. To reduce local noise and emphasize consistent structures, each histogram is smoothed using a moving average filter with a defined window size (Steps 7–9).

Next, the optimal orientation was identified to improve text alignment for each image by analyzing the image using histogram analysis (Bao et al., 2022; Dos Santos et al., 2009). The maximum histogram is generated by summing pixel intensities column by column across the image and then selecting the tallest peak (Bao et al., 2022; Das & Panda, 2023). This peak reflects the strongest alignment of text lines and serves as the criterion for determining the optimal rotation angle. In order to achieve better text alignment, the algorithm computes a peakiness measure (Step 10) that captures the regularity of text lines by comparing the maximum histogram value to its mean. This value is averaged across the three channels, ensuring robustness against variations in color or illumination. Additionally, the algorithm detects and counts long horizontal lines within the image (Step 11), as properly aligned text should exhibit stronger horizontal continuity.

The overall score for a candidate rotation is then obtained as the sum of the peakiness measure and the line count. The main loop (Steps 16–21) systematically rotates the input image within a predefined search range using a fixed step size. For each candidate rotation, the image is rotated with white padding to preserve content integrity (Step 17), and its alignment score is computed. If the resulting score is higher than the previously recorded best score, the corresponding angle is stored as the current best alignment. The output of the image with true alignment was then used to detect the presence of a ruling line, with further explanation in the next sub-section.

3.4.1.2 Process 2: Detection of Ruling Line and Line Replacement

The detection of the ruling line was continued after it had been identified, as explained in the previous section. Figure 3.12 presents the step-by-step process for detecting the ruling line by generating a vertical projection histogram for each RGB channel. Therefore, the identified ruling lines were then replaced with white pixels, resulting in the MLHSD dataset without the ruling lines.

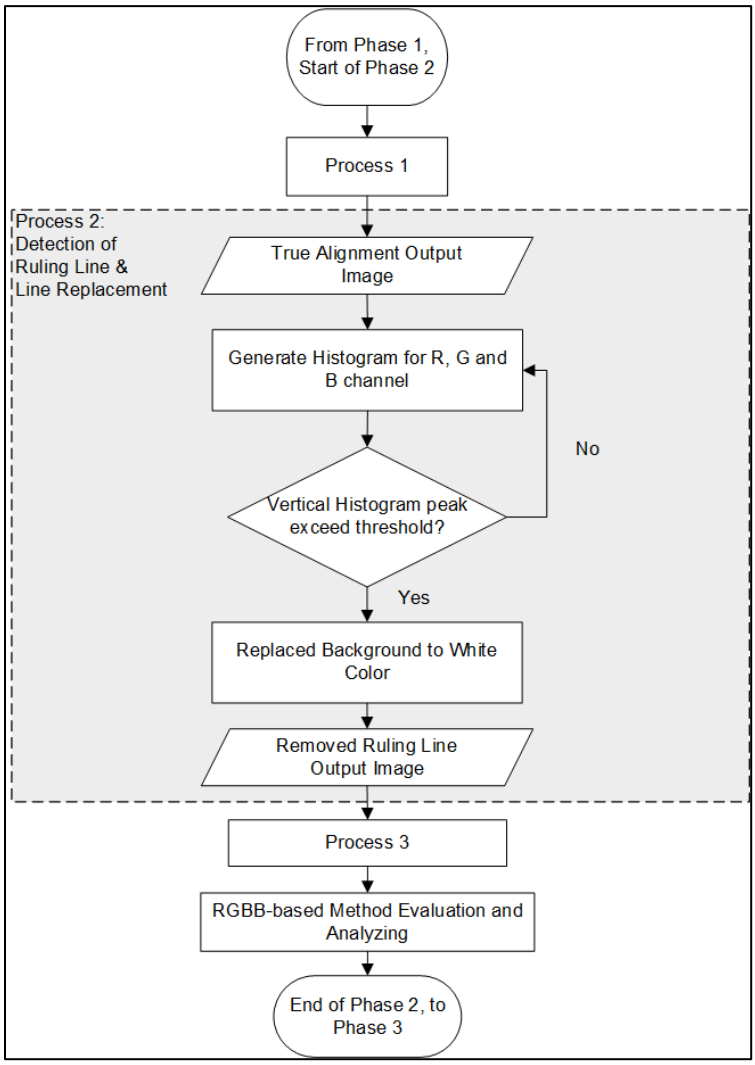


Figure 3.12 Overall Process For the Detection of Ruling Line and Line Replacement in the Proposed Scheme of Ruling Line Removal Using RGBB-Based Algorithm

The process starts with the true alignment output image. The detailed phases of the method are shown in Algorithm 2, which finds and removes horizontal ruling lines using pixel-intensity analysis and morphological operations.

Algorithm 2: Detecting Ruling Line and Color Replacement

Input: image (RGB), thresholds = {R, G, B}, window_size, min_line_length, line_width
Output: cleaned_image, num_lines
//Row-wise zero-pixel detection (per channel)
For each channel c in {R, G, B}:
    H_c[row] ← count of pixels == 0 in image[row, :, c]
    M_c ← moving_average(H_c, window_size)
    For each idx where M_c[idx] > thresholds[c]:
        Mark rows idx..idx+window_size-1 as to-be-replaced
    End
End
//Color replacement
cleaned_image ← copy(image)

```
For all marked rows:
    cleaned_image[row, :, :] ← 255 (white)
End
//Horizontal ruling line detection and removal
gray ←grayscale(cleaned_image)
bw ←adaptive_threshold(gray, invert=True, block_size=15, C=-2)
kernel←rectangular_structuring_element(width=min_line_length,
height=line_width)
horizontal_lines ← morphological_open(bw, kernel)
contours ← find_contours(horizontal_lines)
For each contour in contours:
    draw_white_over(cleaned_image, contour, thickness=2)
End
num_lines ← count(contours)
Return cleaned_image, num_lines
```

First, the image is analyzed by creating vertical histograms for the red, green, and blue (RGB) channels. These histograms show the distribution of pixel intensities column by column. This allows the algorithm to find consistent horizontal patterns that suggest the presence of ruling lines (Refaey, 2014). As depicted in Figure 3.13, three ruling lines are identified for each RGB color channel. Next, the peak value of each vertical histogram is compared to a predefined threshold.

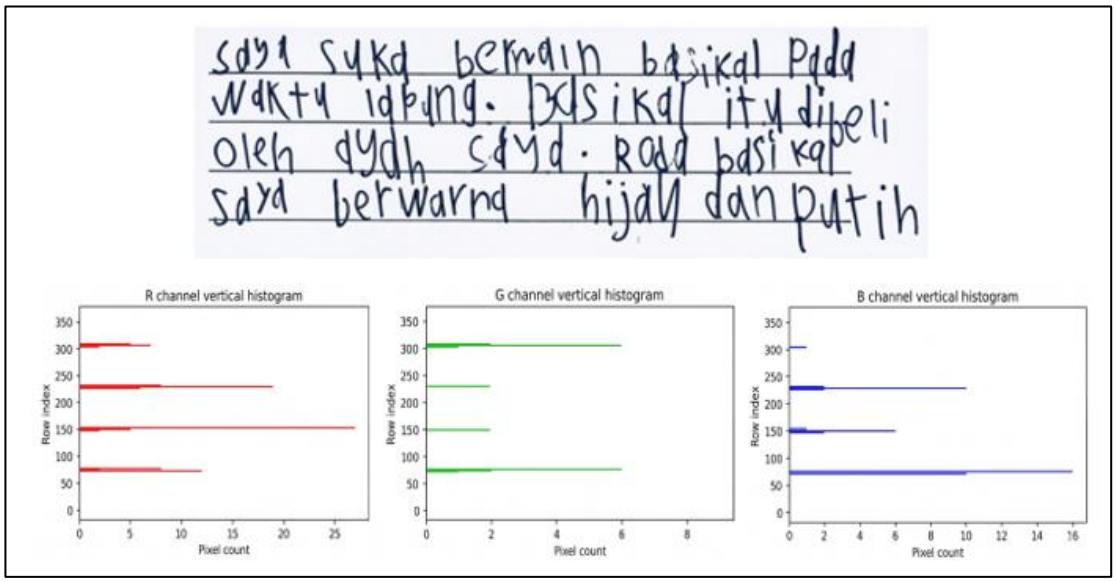


Figure 3.13 RGB Vertical Histogram Based on the Rotated Image

When a vertical histogram peak exceeds a predefined threshold, regardless of the color channel, the corresponding horizontal indices are replaced with the background color white, as illustrated in Figure 3.14.

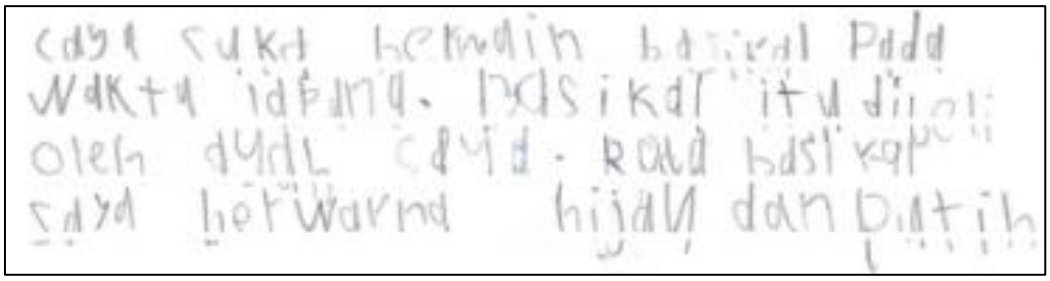


Figure 3.14 Image After the Ruling Line Color Is Replaced with the Background Color

This ensures that the handwriting stays visible while unwanted lines are removed. If the histogram peak does not exceed the threshold, the process continues unchanged, as no significant ruling line is found. In this case, the threshold is the minimum peak intensity needed to confirm the presence of ruling lines. The threshold was fixed by analyzing normalized intensity histograms and selecting a percentage of image width that consistently separates handwriting rows from ruling lines. The percentage thresholds are 30% for low-risk and moderate-risk handwriting and 40% for high-risk handwriting. The result of this process is a clean output image with the ruling lines removed, making it ready for the next steps in feature extraction and classification.

3.4.1.3 Process 3: Handwriting Reconstruction

After performing the image processing for removing horizontal ruled lines in Process 2, there will be situations where some characters or letters become disconnected as depicted in Figure 3.14. The procedure begins with the output image in three channels of RGB images that was obtained from Process 2, where ruling lines have already been removed.

The restoration phase employed a dual-strategy approach combining pixel-level reconstruction from neighbouring pixel reference and then morphological pixel refinement to ensure optimal handwriting continuity.

i) Pixel-level reconstruction from neighbouring pixel reference

The first reconstruction as shown in Figure 3.15 employs pixel-level reconstruction. This process starts with the identified pixel removed from the line removal step.

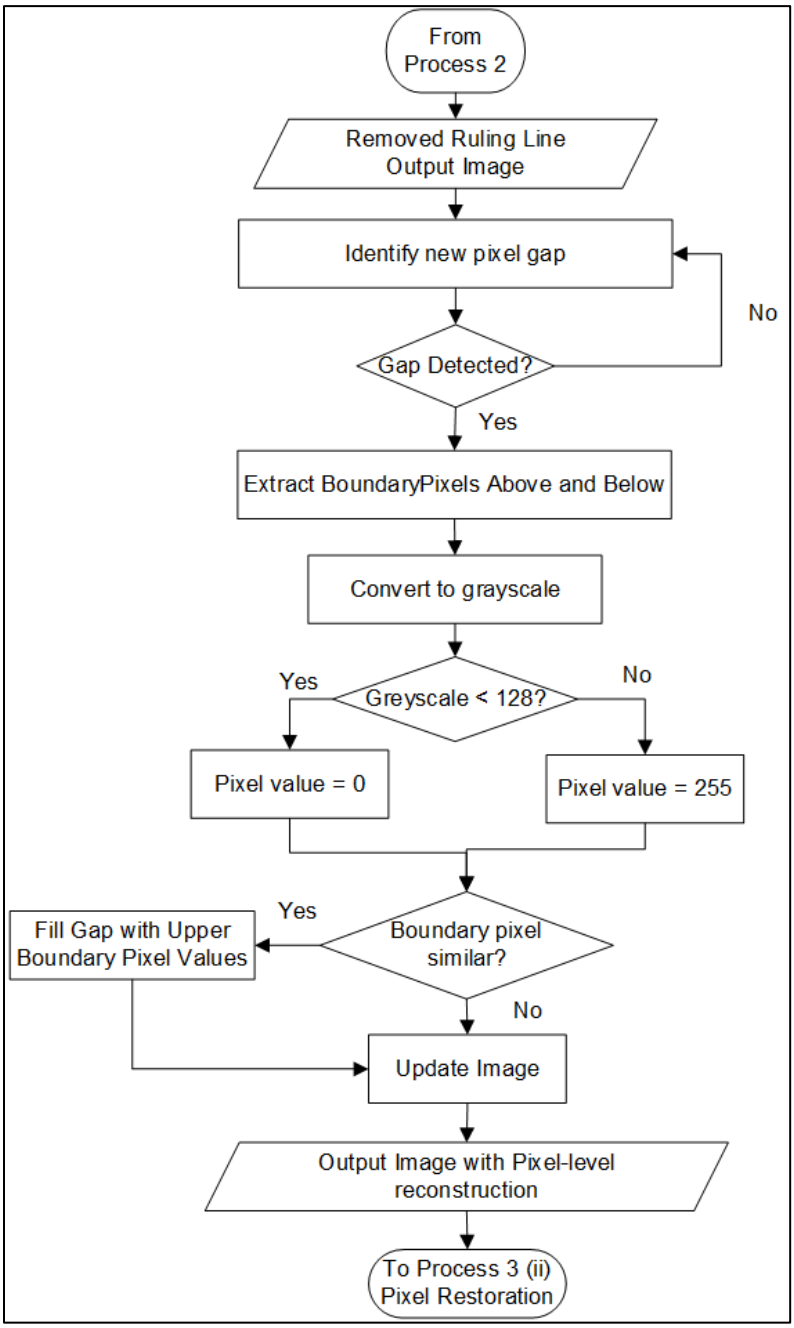


Figure 3.15 Flowchart of Pixel-Level Reconstruction From Neighbouring Pixel Reference

Based on the known removed pixel index, each of the pixels above and below was considered to be reconnected. The example is shown in Figure 3.16 where the letter “S” has a truncated part on the bottom curve.

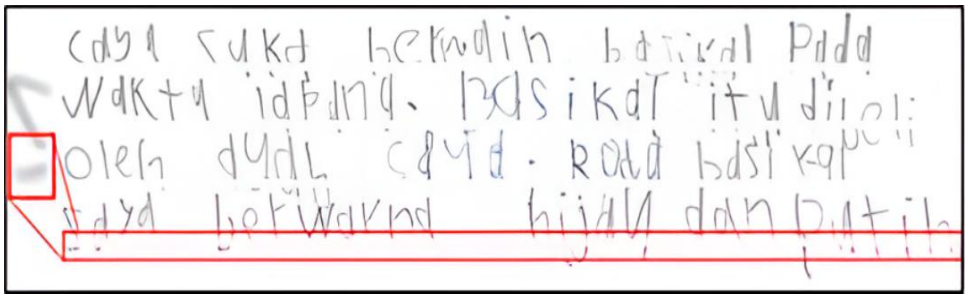


Figure 3.16 A Gap Was Detected in the Handwriting

The initial restoration stage employs a column-wise scanning algorithm to detect and bridge gaps created during the ruling line removal process (Alipour et al., 2013), as shown in Figure 3.15. The algorithm scans each column of the cleaned handwriting image to identify gaps in pixels where ruling lines were removed. Once a gap had been detected, the boundary pixels located above and below the gap were extracted and converted to grayscale by averaging their RGB values. The grayscale values were then binarized using a threshold of 128, which corresponds to the midpoint of the 8-bit grayscale range (0–255) and is a standard cutoff for separating foreground handwriting from background regions (Arvind et al., 2007). Moreover, this value was chosen because it represents the midpoint of the 8-bit grayscale range (0–255), making it a standard cutoff for separating dark (foreground, typically handwriting) and light (background) regions. Pixels with intensity values below 128 were assigned as black (value = 0), while pixels with values equal to or greater than 128 were classified as white (value = 255). The binary values of the upper- and lower-boundary pixels were compared to determine whether they belong to the same class. If both pixels were classified as foreground (handwriting) or both as background, the gap was considered part of a continuous handwriting stroke. In this case, the removed pixels within the gap were filled using the value of the upper boundary pixel to restore the stroke. If the boundary pixels belong to different classes, the gap was left unchanged to avoid incorrectly connecting unrelated characters. An example of the reconstructed handwriting is shown in Figure 3.17.

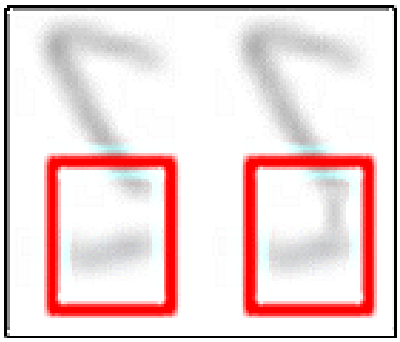


Figure 3.17 After the First Straight Pixel Was Reconstructed

After Pixel-level reconstruction from neighbouring pixel reference was implemented, the image still had missing details, such as curves and diagonal strokes, that the previous technique was unable to restore based on the assumption of the neighbouring pixel. The restoration's second phase used morphological dilation and erosion (Adamski et al., 2021).

ii) Morphological dilation and erosion for pixel restoration

Figure 3.18 depicts the restoration's second phase, which uses morphological dilation and erosion. The process began with the output from pixel-level reconstruction, which serves as the input image for subsequent refinement. The first step involved loading the handwriting image with ruling lines, followed by the conversion of the image to its inverse bitwise form to highlight the structural details of the writing.

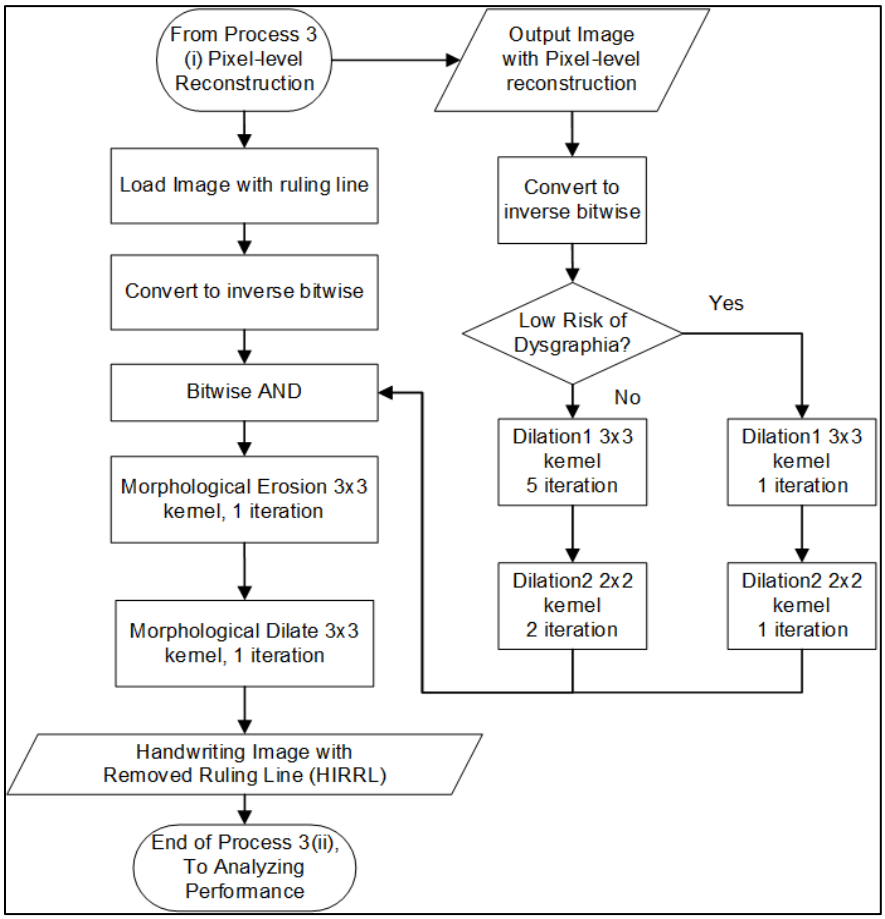


Figure 3.18 Process of Morphological Dilation And Erosion For Pixel Restoration

As depicted in Figure 3.18, the pixel restoration process involved a morphological dilation and erosion operations. The image derived from pixel-level reconstruction (PLR) underwent transformation to inverse binary, analogous to the process applied to the original image. In accordance with the risk level associated with the original image pertaining to dysgraphia, a varying degree of iteration was implemented. This subsequently involved executing an AND bitwise operation with the original transformed image. The resultant output image for line removal was generated after applying morphological erosion and dilation techniques to the new handwriting image with removed ruling line (HIRRL).

After pixel-level reconstruction, the image was converted back to its inverse bitwise form and then evaluated to determine if the subject has a low risk of dysgraphia. For handwriting with low dysgraphia risk, minimal enhancement was necessary to preserve the natural strokes. Consequently, dilation with a 3×3 kernel was applied once, followed by a secondary dilation with a 2×2 kernel for one iteration. This gentle process improves stroke visibility without causing distortion. Conversely, handwriting from

individuals with a higher risk of dysgraphia often appears weak, fragmented, or inconsistent. To strengthen these strokes, dilation with a 3×3 kernel was applied for five iterations and then dilation with a 2×2 kernel was applied for two iterations. These multiple steps were intentionally designed to reinforce weak strokes and improve legibility before the image was reprocessed through the main ruling line removal pipeline.

Lower iteration counts (one iteration each) were chosen for low-risk cases to prevent over-processing, ensuring that natural handwriting features stay complete. Higher iteration counts (five and two iterations) were assigned to high-risk dysgraphia cases after experimental evaluation showing that more enhancement was needed to recover the legibility of weaker strokes without causing excessive distortion. The final erosion and dilation steps, each with a single iteration, were applied consistently across all cases to refine the reconstructed handwriting and restore balanced stroke thickness after removing ruling lines. This adaptive approach ensures that preprocessing responds to the variability in handwriting quality, effectively balancing the goals of ruling line removal, stroke preservation, and improving legibility.

3.4.2 Evaluation Performance of the Proposed New Scheme of Ruling Line Removal Using RGBB-based Method

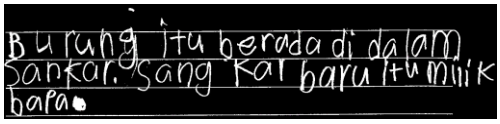
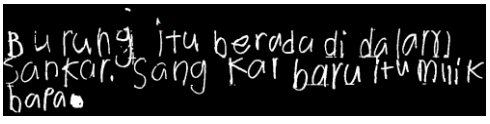
To assess the successful implementation of the proposed ruling line removal method, two evaluation techniques were employed: the Structural Similarity Index Measurement (SSIM) and a performance classification model, as described in the following sub-sections.

3.4.2.1 Evaluation Using Structural Similarity Index Measurement (SSIM)

This section evaluates the effectiveness of the proposed new ruling line removal method and illustrates its impact on preserving structural features by applying Structural Similarity Index (SSIM) as an assessment metric (Mudeng et al., 2022). Hence, the SSIM is a suitable instrument for evaluating the effectiveness of the ruling line removal technique in handwriting images. After eliminating the ruling lines, the algorithm assesses the structural similarity between the original handwriting image and the removed ruling line image (Sara et al., 2019). Therefore, the SSIM is essentially a

metric used to evaluate the effectiveness of a line removal operation in maintaining the core structural elements of the image (Peng et al., 2020). It constitutes a measurable evaluation of the extent to which the processed image maintains its initial structure, making it a valuable tool for evaluating the effectiveness of line removal schemes. Table 3.9 demonstrates the examples of handwriting images with the ruling line (HIRL) and removed ruling line (HIRRL).

Table 3.9
Structural Similarity Index Analysis

Handwriting Image With Ruling Line (Original)	**Handwriting Image With Removed Ruling Line**
	
	
	

The performance of newly proposed image preprocessing method for removing ruling line were evaluated using SSIM (Sara et al., 2019). The SSIM index was calculated by comparing the structural information between the original and final input images. This approach is often used to assess the quality of an image by comparing outputs and reference images (Nilsson & Akenine-Möller, 2020; Sara et al., 2019). The mathematical expression for SSIM involves several components, luminance, contrast, and structure components to provide a comprehensive measure of image similarity. Given that X is the original image and Y is the resultant image after applying the newly proposed image preprocessing method, the SSIM can be determined using Eq. (3.6) as follows:

$$SSIM(X,Y)=l(X,Y)\cdot c(X,Y)\cdot s(X,Y) \tag{3.6}$$

where,

$$l(X,Y)=\frac{2\mu_X\mu_Y+C_1}{\mu^2{}_X+\mu^2{}_Y+C_1}$$

$$c(X,Y)=\frac{2\sigma_X\sigma_Y+C_2}{\sigma^2{}_X+\sigma^2{}_Y+C_2}$$

$$s(X,Y)=\frac{\sigma_{XY}+C_3}{\sigma_X\sigma_Y+C_3}$$

Where

μ_X and μ_Y are the average intensities of images X and Y,

σ_X and σ_Y are the standard deviations of intensities in images X and Y,

σ_{XY} is the covariance of intensities between X and Y,

C_1 , C_2, and C_3 are constants to avoid instability when the denominators are close to zero.

In this study, the values of C1, C2, and C3 were set to 95, respectively. The constant were empirically chosen as 95 to stabilize SSIM computation and ensure consistent similarity measurement for the given dataset. The SSIM produces values ranging from -1 to 1, where a value of 1 indicates a perfect match, 0 indicates no similarity and -1 shows a complete mismatch. SSIM values closest to 1 indicates greater similarity evaluation of the proposed image preprocessing for ruling line removal. When SSIM returns yields a high value, it suggests that the processed image closely resembles the original image in terms of its structural features, such as edges, textures, and patterns. Therefore, SSIM is essential in image performance measurement since it provides a quantifiable measure of how well an image retains structural information during various image processing procedures.

By generating the HIRRL dataset after removing the ruling line, the study can then focus on Phase 3. In this phase, dysgraphia severity was classified using CNN-based fusion features with dual-input images.

3.4.2.2 Evaluation Using CNN Classification Model

The comparison of classification performance between input images of handwriting images with the ruling line (HIRL) and removed ruling line (HIRRL) is required to clarify the impact of the proposed new ruling line removal method. The

classification process was performed by assessing the performance of the two image categories, specified as HIRL and HIRRL. This investigation aims to verify the HIRRL image for use in the next feature extraction process. The input images must be produced prior to the feature extraction process, wherein both HIRL and HIRRL handwriting images were utilized to identify the severity determined by the written information. Figure 3.19 illustrates the comparison of the classification process, in which the input image is derived from the features and subsequently categorized into three classes.

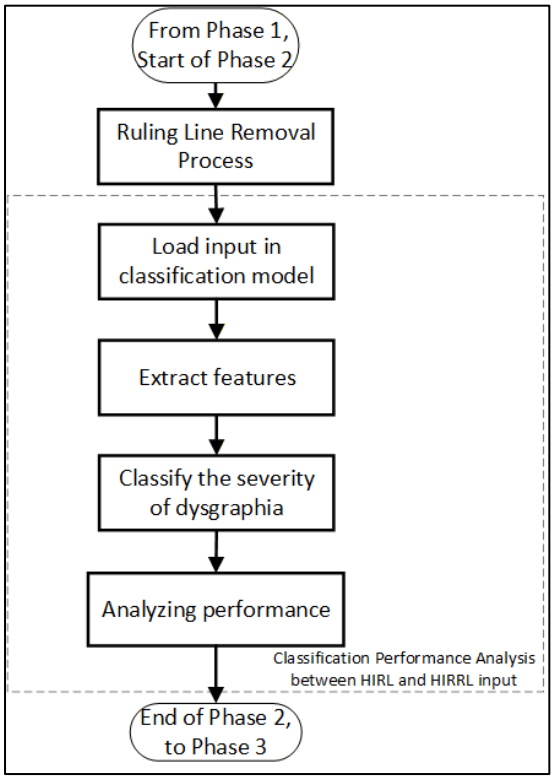


Figure 3.19 Classification of Dysgraphia Severity using HIRL and HIRRL as Input Images

The initiation of the process involved inserting the HIRL or HIRRL images into the feature extraction network, which employed a residual architecture. The features that were subsequently extracted were used to classify the severity of dysgraphia, with performance evaluated upon completion of the classification process.

Figure 3.20 shows the diagram of the classification process used for both HIRL and HIRRL input images. These images were used for feature extraction using a residual network architecture, as described in 3.5.2.1 (i), and for classification using an SVM model. The classification results fall into three classes: high risk, moderate risk, and low risk.

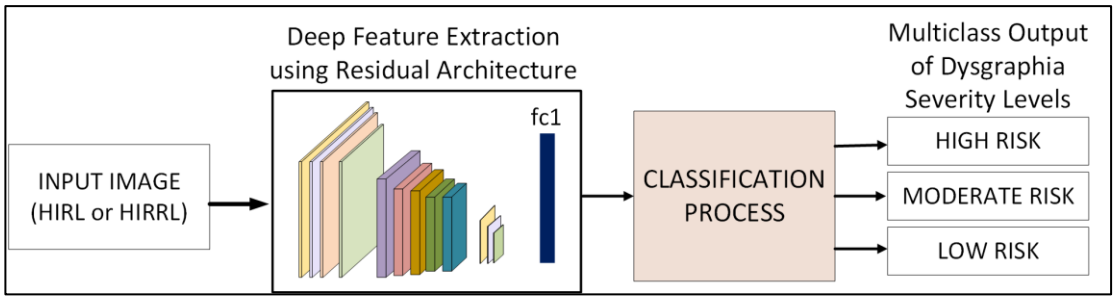


Figure 3.20 Dysgraphia Severity Classification Process Based on HIRL and HIRRL

Both classifications were then evaluated to observe the performance differences in implementing HIRL and HIRRL as input images. The evaluation was conducted by measuring the accuracy, precision, recall, and F1-score as mentioned in 3.6. The comparison results were used to assess the impact of ruling line removal on classification performance.

3.5 Phase 3: The H2FCD model of Dysgraphia Severity Classification Based on Dual Input Images with CNN-Based Fusion Features

This phase presents a proposed multiclass classification model namely H2FCD to classify dysgraphia severity levels using CNN-based model. The H2FCD was designed based on fused features extraction techniques which are capable of handling dual-input image in the form of histogram and handwriting images.

Figure 3.21 illustrates the overall process of Phase 3 which utilized the output from Phase 2 and aimed to achieve the most optimal performance in classification model by training and validating the model. The process begins with preparing the dualinput images taken from the Phase 2 images. Then, both handwriting and histogram images were extracted using a residual-based CNN architecture to obtain meaningful features. Consequently, the features obtained by residual networks were then fused by concatenating the feature vectors. All features are then used to classify dysgraphia severity. Four ML models were implemented using SVM, KNN, BT, and RF. The classifier model was trained and validated to determine the best-performing ML model. Further explanation is continued in the following sections.

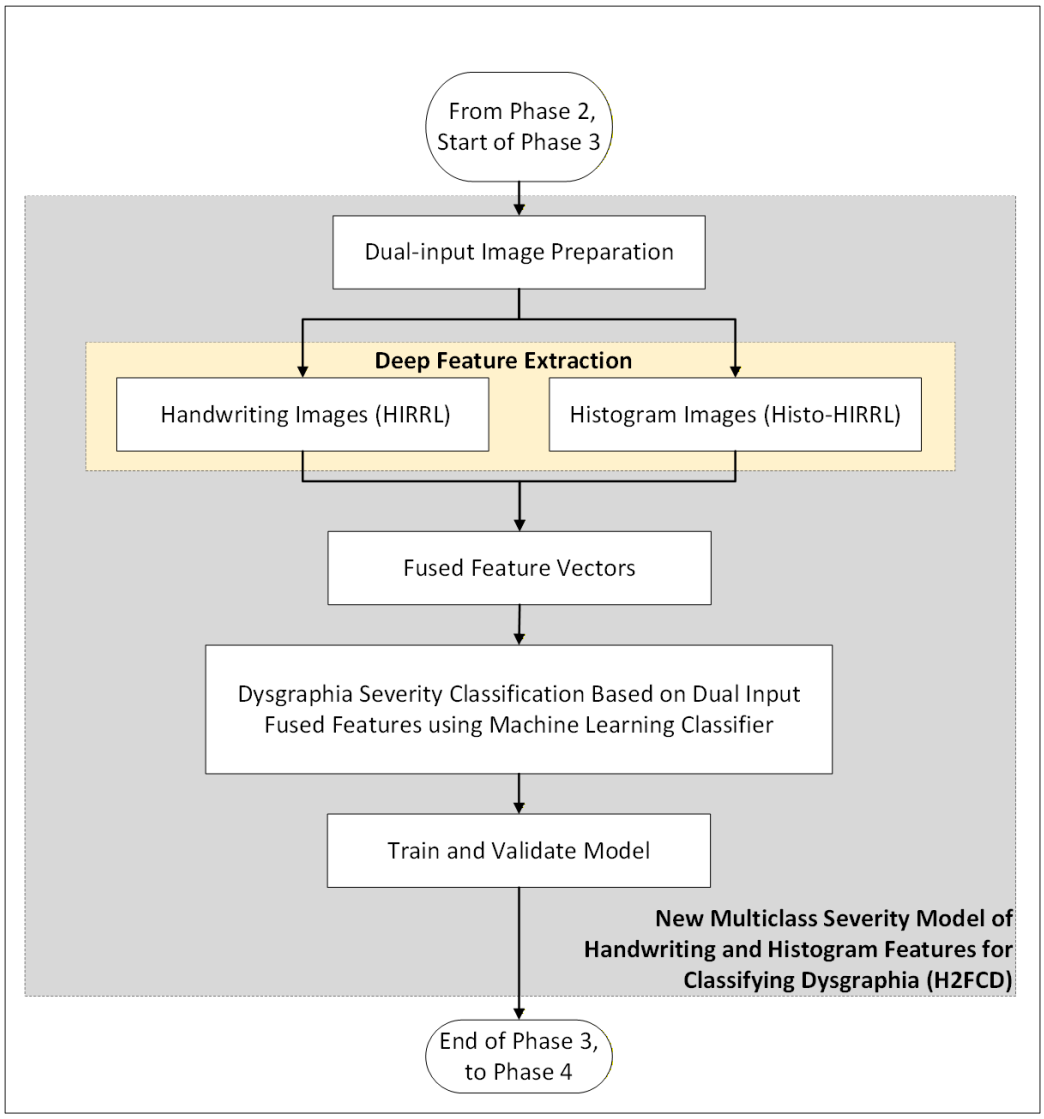


Figure 3.21 Phase 3 of Developing the Proposed H2FCD Classification Model

3.5.1 Dual-Input Image Preparation

Phase 3 began by preparing the input images which had been pre-processesed using newly proposed scheme in Phase 2. During this preparation task, all input images were resized to a standardize image size to generate two forms of datasets namely Handwriting Image with Removed Ruling Line (HIRRL) and Histogram Image from HIRRL (Histo-HIRRL). Further explanation of both datasets is explained as follows:

i) Handwriting Image with Removed Ruling Line (HIRRL)

This section outlines the procedure for creating dataset image for usage in 3.5.2. The input image comprises two handwritten images with removed ruling lines

(HIRRL). Figure 3.22 depicts the data image preparation process to generate a resized image ready for use in the subsequent classification process. Each procedure involves merging the three statements of handwriting into a single image to represent all the handwriting samples. However, the merged images were not resized to avoid loss of details and fine structure of the handwriting image. The original image underwent resizing, resulting in a 40% decrease in width and a 30% decrease in height.

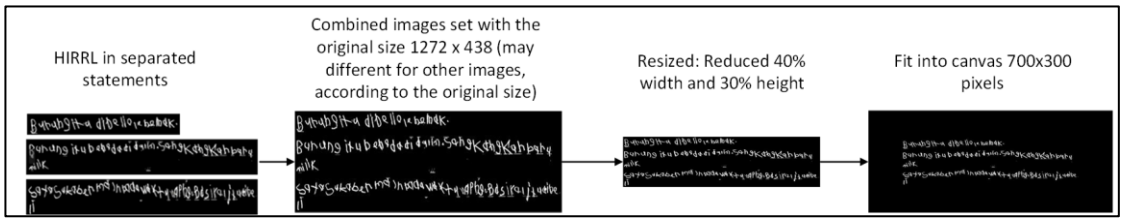


Figure 3.22 Handwriting Image with Removed Ruling Line (HIRRL)

Each input image was resized using fixed-percentage non-isotropic resizing (L. and M. X. Chen, 2009), where the width and height were independently scaled to 40% and 30% of their original sizes, respectively (C.-Y. Liu et al., 2019). This resizing method effectively reduces computational load and storage usage, although it incurs the expense of preserving the original aspect ratio. It resulted in geometric distortion being minimized, with a focus on significant regions (L. and M. X. Chen, 2009). Therefore, the significant features of handwriting, such as stroke and the skewness of handwriting, were retained.

Next, after the resizing process, the maximum width and height were calculated for all processed images to obtain resolution statistics. The *imresize()* function in MATLAB was then used to preserve image properties, such as type, color channels, and bit depth. Subsequently, the black background was increased to the standard image size of 700 * 300 pixels (Semma et al., 2021). This technique involves placing each image at the center of a fixed-size blank canvas while padding the surrounding area with a constant value of 255. This approach ensured that the original text dimensions and characteristics were not distorted, thereby preventing interference with the feature extraction process and enabling the generation of accurate and complete feature representations.

ii) Histogram Image (Histo-HIRRL)

This process purposely prepares histogram images produced from HIRRL images obtained in Phase 2 representing the optimal image form for improving classification performance. Figure 3.23 depicts the image dataset preparation process used to generate histogram images (Histo-HIRRL).

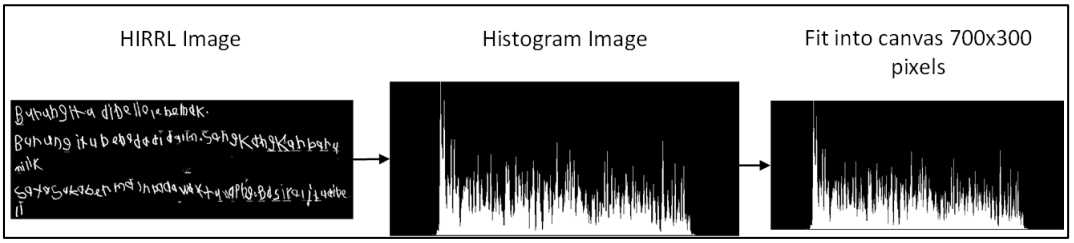


Figure 3.23 Histogram Image (Histo-HIRRL) Preparation Process

The Histo-HIRRL procedure involves combining images, which were then converted into histogram images to represent each handwriting sample dedicated to the individual respondent. The generated histogram images were then resized to dimensions of 700 by 300 pixels. The histogram image was employed as a second input modality, because it captured global intensity distributions and structural patterns that are not fully represented in raw handwriting images, which primarily focused on stroke shape and spatial features. In particular, histogram representations highlighted distribution-related characteristics such as stroke density and contrast variation. This alternative representation emphasizes the distribution of pixel intensities and structural variations (Furht et al., 2018), which may be overlooked in raw images. In the context of handwriting feature extraction for dysgraphia severity, such statistical variations are particularly valuable because irregularities in stroke pressure, spacing, and alignment often manifest as differences in pixel distributions. By incorporating histogram images alongside raw handwriting images, the model gained complementary feature sets, thereby improving overall classification accuracy (Kim & Youn, 2025).

3.5.2 Developing the Multiclass Dysgraphia Severity Classification Model (H2FCD)

In this section, a new multiclass dysgraphia severity classification model is proposed and explained in detail. An input for the fused features process, together with

the Histo_HIRRL image, for the next classification investigation is further explained in Section 1.6.4.1. Subsequently, a fusion dual-input feature-based classification model is executed to explore the performance of the ML classifier in classifying the severity of dysgraphia handwriting. Although ResNet-18 was used in this study as a feature extractor (Tang et al., 2025; Zhang et al., 2026) to enable flexible integration with multiple machine learning classifiers. This approach allowed the evaluation of different classifiers (SVM, KNN, RF, BT) on the same deep feature representations, which were more suitable for multiclass dysgraphia severity classification. Additionally, using ResNet-18 as a feature extractor reduced the risk of overfitting on a moderate-sized dataset and improved generalization compared to a fully end-to-end deep learning model (Francis & Prakash Verma, 2024; Zhang et al., 2026).

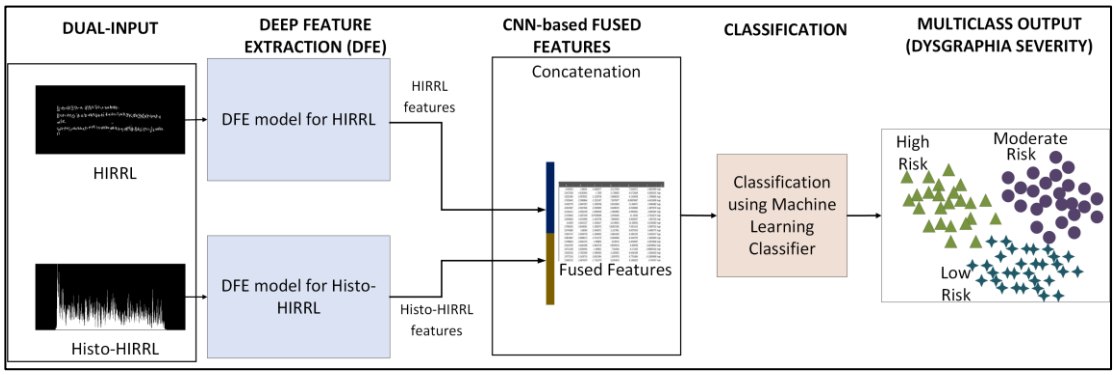


Figure 3.24 The Framework of the Proposed H2FCD Classification Model

The diagram of multiclass dysgraphia severity classification is shown in Figure 3.24. The proposed H2FCD classification model utilized deep feature extraction through CNN architecture combined with an ML classifier to generate multiclass dysgraphia severity outputs. Prior to classification, the dual-input datasets of HIRRL and Histo-HIRRL were used in parallel as attributes for supervised model training as explained in Section 3.5.1. Histo-HIRRL was created to fuse the feature vector with HIRRL features, followed by classification into three dysgraphia severity levels: high-risk, moderate-risk, and low-risk. Table 3.10 presents the dataset distribution of HIRRL and Histo-HIRRL employed in the H2FCD classification model. The training ratio was 80% from the total images (Yogarajah, 2020) out of 172, and the remaining images were reserved as a testing dataset employed in Section 3.6.1.

Table 3.10
The Dataset Distribution of HIRRL and Histo-HIRRL Employed in the H2FCD Classification Model

Dataset	The number of images	Ratio (%)
Training	136	80
Testing	36	20
Total	172	100

Both HIRRL and Histo-HIRRL images were extracted for the features via deep feature extraction using the residual network. The concatenated input data of future vectors were fed into each machine learning classifier model to accurately predict the corresponding label for each input. The performance of each ML model was evaluated to determine the best model for classifying dysgraphia severity. The measurement involved the accuracy, precision, recall, f1-score, ROC-AUC graph across different types of feature vectors; used, namely fusion feature, image feature, and histogram feature. The classification process is discussed in the following subsection.

3.5.2.1 Dual-Input Fusion Feature-based for H2FCD Classification Model

This section discusses the entire feature extraction process for two different inputs. Two CNN architectures were implemented in this study to resize raw handwriting images and histogram images. First, deep feature extraction was implemented using the ResNet-18 network to extract the HIRRL. The second part describes a new composite of Residual Networks for extracting features of Histo-HIRRL from handwriting images. Meanwhile, Figure 3.25 illustrates the process of feature extraction for both HIRRL and Histo-HIRRL handwritten inputs using neural network followed by concatenation of the CNN-based features.

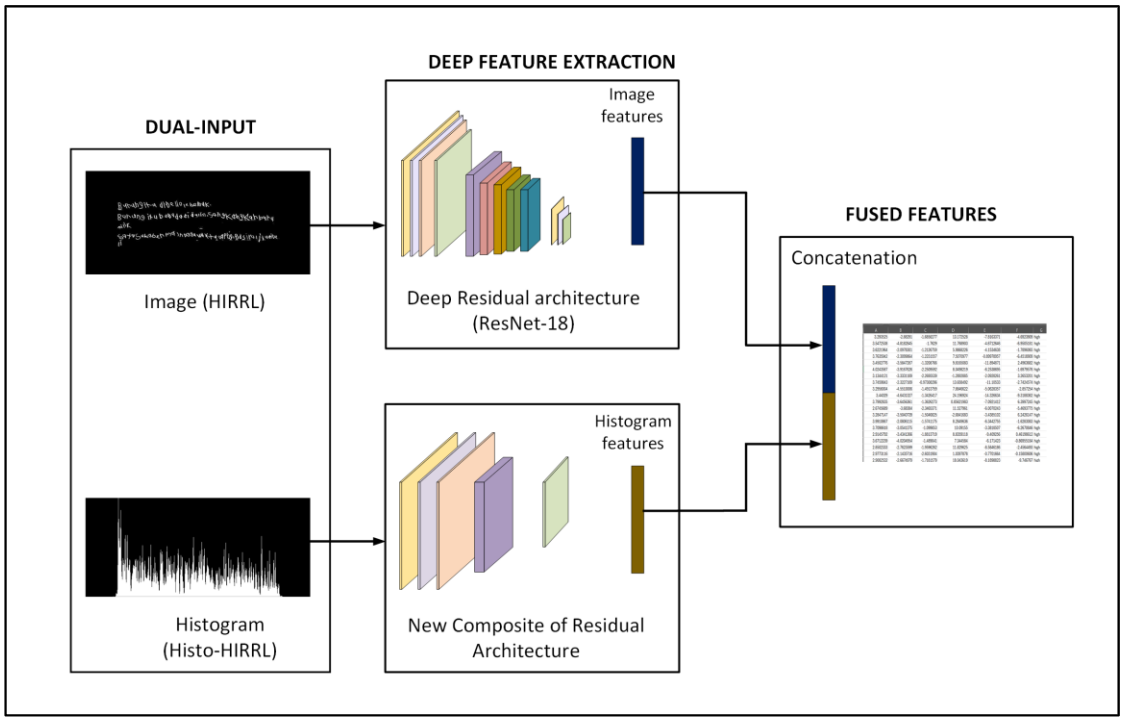


Figure 3.25 Feature Extraction Process for Both Input Based on Raw Image (HIRRL) and Historam Image (Histo-HIRRL)

The feature vectors generated from HIRRL features and Histo-HIRRL features were acquired from the completely linked layers, and subsequently assembled and concatenated to form a single set of combined features (Cengil & Çınar, 2022; Deniz et al., 2018; El Gannour et al., 2022; Ilhan et al., 2022; Pham & Won, 2019), as shown in the final stage of Figure 3.25. The inclusion of fuse features was crucial in enhancing the classification performance for the multiclass severity of dysgraphia. In the subsequent phase, the classification process was carried out, employing machine learning approaches to determine the optimal performance in categorizing the degree of dysgraphia. This was achieved by utilizing the fusion handwriting image-histogram features method, which considered two distinct types of input images.

The fusion feature is a combination feature vector of the HIRRL feature and the Histo-HIRRL feature; meanwhile, the image feature refers to a feature that is obtained only from the HIRRL feature extraction, and the histogram feature refers to a feature derived exclusively from Histo-HIRRL. The results were then compared to observe the performance for each of the feature vector implementations.

The first part of H2FCD classification model is feature extraction. The process of extracting a feature vector is crucial to ensure that the features are effectively utilized in the classification phase, which ultimately produces the classification result. This

section explains all the details of feature extraction techniques used in the H2FCD classification model. Two techniques were employed in the H2FCD model: the adopted Res-Net-18 and newly proposed composite residual network.

i) Feature Extraction Model for HIRRL using Adopted ResNet-18 Layer Network

For feature extraction, an adapted Residual Network architecture, specifically ResNet-18, was adopted due to its strong balance between representational capability and computational efficiency (Fateah et al., 2026; Zhang et al., 2026). Residual network was able to mitigate vanishing gradients through residual (skip) connections, enabling effective learning of hierarchical handwriting features such as edges, textures, and stroke patterns (Chi et al., 2021; Fateah et al., 2026). Compared to deeper variants, it has fewer parameters, making it more computationally efficient and less prone to overfitting when applied to a moderate-sized and highly variable dataset like MLHSD. This balance between model complexity and generalization makes ResNet-18 well-suited for multiclass dysgraphia severity classification. In the H2FCD design, the feature extraction utilizes residual architecture (Alyahya et al., 2023; Chi et al., 2021) to extract the input image and acquire the feature vector called HIRRL_features. Figure 3.26 illustrates the hierarchical arrangement that functions as a powerful feature extractor for the first input of rescaled images. The input layer consists of images that have been scaled to dimensions of 300 x 700 pixels. The input image undergoes an initial processing step, sending it through a convolutional layer. This layer applied 64 filters of size 7 × 7 to the input image with a stride of 2. Afterward, batch normalization and a Rectified Linear Unit (ReLU) activation function were applied. Before a 3x3 max-pooling layer was implemented with a stride of 2, consequently decreasing the spatial dimensions of the feature maps.This layer applied 64 filters of size 7 × 7 to the input image with a stride of 2

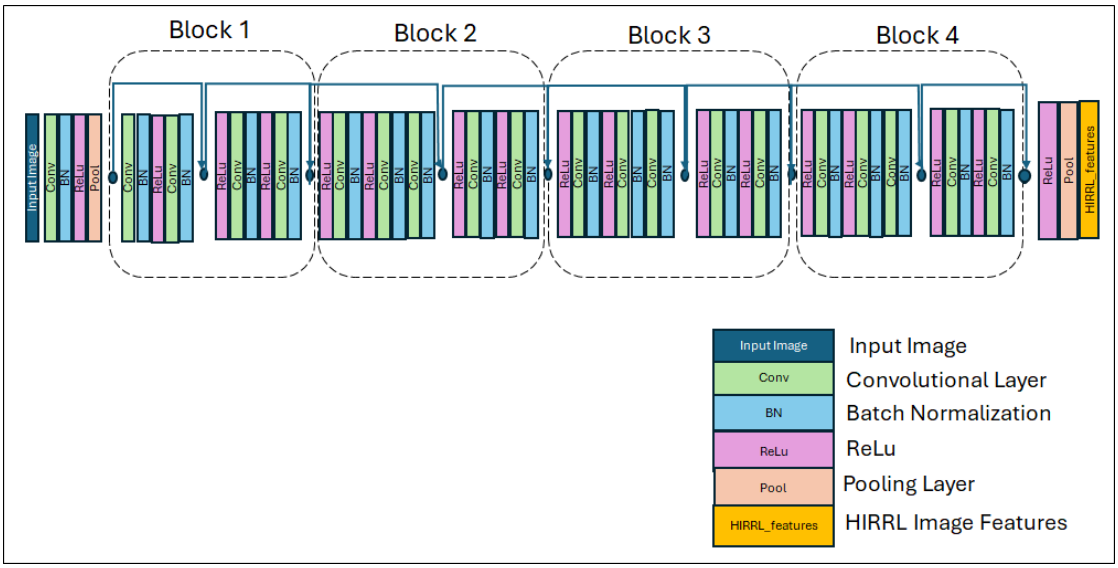


Figure 3.26 The Feature Extractor Layer Structure of the Residual Architecture For HIRRL Uses in the Feature Extractor

Subsequently, the network progresses through four different sets of residual blocks. The first residual block, Residual Block 1, consists of four convolutional layers, each with 64 filters and 3x3 kernels. The output of the second convolutional layer was merged with the original input of the block (skip connection) prior to applying the ReLU activation function. The second residual block, Residual Block 2, consists of three convolutional layers, each with 128 filters. It used a stride of 2 to reduce the size of the input. In addition, a skip connection was incorporated, comprising of a 1x1 convolution with 128 filters and a stride of 2 to ensure that the dimensions align prior to the addition operation. The subsequent sub-block employed two convolutional layers with a stride of 1 to reduce the size of the input, each comprising 128 filters. Following that, Residual Block 3 was built using the same structure as Residual Block 2, which included three convolutional layers with 256 filters. The first layer utilized a 3x3 convolution operation with a stride of 2 to achieve downsampling. The skip connection in this block additionally included a 1x1 convolution operation with 256 filters and a stride of 2. Furthermore, there was an additional sub-block comprising of two convolutional layers. Every layer had a stride of 1 and employs 256 filters. Residual Block 4 consisted of three convolutional layers, each containing 512 filters in the first sub-block, while maintaining the same structure. It utilized a 3x3 convolution operation with a stride of 2 in order to achieve downsampling. The skip connection in this block included a 1x1 convolutional operation with 512 filters and a stride of 2. The second sub-block comprises two convolutional layers with a stride of 1 x 1 and 512 filters.

After the residual blocks, the network utilized global average pooling, which entailed calculating the average values of each feature map across the spatial dimensions. Afterward, the tensor was sent into an HIRRL_features layer that consists of the feature vectors.

Table 3.11 provides an overview of the feature extraction model, showing how each layer transforms the input HIRRL image and contributes to the total learnable parameters. The model begins with conv1, followed by four residual blocks (res2–res5) that progressively reduce spatial size while increasing feature depth, with res5 containing the highest number of parameters. An average pooling layer then compresses the final features without adding parameters. Overall, the model learns a total of 11,302,272 parameters.

Table 3.11
Parameter Setting For Residual of Feature Extraction

Layer	**Activations**	**Number of Learnable**
Input Image (HIRRL)	300 x 700 x 1	-
Block (activations)	Components (short)	Learnable Parameters
conv1 (150×350×64)	conv1 + bn_conv1	9,600
res2 (75×175×64)	all res2 convs + batchnorms (4 conv + 4 BN)	148,224
res3 (38×88×128)	all res3 convs + batchnorms (3 blocks → 6 conv + 6 BN)	813,312
res4 (19×44×256)	all res4 convs + batchnorms (4 conv + 4 BN)	2,067,456
res5 (10×22×512)	all res5 convs + batchnorms (4 conv + 4 BN)	8,263,680
Average Pooling layer	1 x 1 x 512	-
Total	-	11,302,272

The parameters of the neural network model for dysgraphia handwriting images were individually configured for the network, as tabulated in Table 3.12. In this work, hyperparameters were selected for training ResNet-18 on input images sized 700×300 to ensure stable convergence, efficiency, and reliable generalization. Stochastic Gradient Descent with Momentum (SGDM) was used as an optimizer (Kondo & Iiduka,

2025). Momentum helps speed up convergence and reduce oscillations. Recent studies showed that it performed better in generalization compared to plain SGD (Mo & Zhang, 2022; Puri & Singh, 2019). Meanwhile, the specified parameters (learning rate, epochs, and mini-batch size) were retained, while all other training parameters were set to their default settings defined by the framework. The learning rate was fixed at 0.001. This conservative step size helps avoid divergence and overshooting, especially when training residual architectures on large image inputs (Puri & Singh, 2019). This network employed a total of 16 epochs. This number is usually enough for ResNet-18 to converge without overfitting because residual networks tend to converge faster (Rakesh et al., 2024). The mini-batch size was set to 16 to balance computational efficiency and generalization performance while considering the large image dimensions that limit GPU memory (Masood et al., 2023). Lastly, the network monitored performance every 30 iterations to ensure consistent evaluation during training without creating too much computational burden (Kunhoth, Al Maadeed, et al., 2023).

Table 3.12
Parameter Setting For Residual of Feature Extraction

Parameter	**Value setting**
Optimizer	SGDM
Learning rate	0.001
Epochs	16
Mini batch size	16
Frequency	30

ii) New Composite of Residual Network for Histo-HIRRL Feature Extraction

The Histo-HIRRL inputs were processed using a novel deep feature extractor layer structure based on a 1-block Residual Network (Bajić et al., 2022) and included a single skip connection as illustrated in Figure 3.27. The Histo-HIRRL inputs were derived from histogram-based representations and thus exhibited relatively low spatial complexity compared to natural images. Therefore, the input had low structural complexity and limited dataset size, deeper architectures with multiple residual blocks frequently causing unnecessary computational cost and elevated risk of overfitting

(Bejani & Ghatee, 2021). Secondly, residual learning (i.e., skip connections) maintained gradient stability and facilitated effective feature transmission, even in shallow networks. Skip connections enhance information flow and alleviated gradient vanishing, therefore promoting effective training despite constraints on network depth (G. Xu et al., 2025). By employing a 1-block structure, the model retains these advantages without incurring the diminishing returns often associated with deeper architectures (Ashwini et al., 2023).

The input layer consisted of histogram images with 300 × 700 pixels dimensions. The input image underwent convolutional processing using a layer composed of 16 filters and a stride of 2. Subsequently, batch normalization and the ReLU activation function were applied. The second convolutional layer comprised 32 filters, each with a filter size 3x3. The preceding input was downsampled using a stride of two. The third convolutional layer had a 3x3 filter size, consisting of 32 filters. The skip connection in this block consisted of a 1x1 convolutional operation with 32 filters and a stride of 2. The network employed global average pooling, which involved taking the average values in each feature map across spatial dimensions. The last layer is the Histo-HIRRL_features layer responsible for generating the ultimate feature vectors.

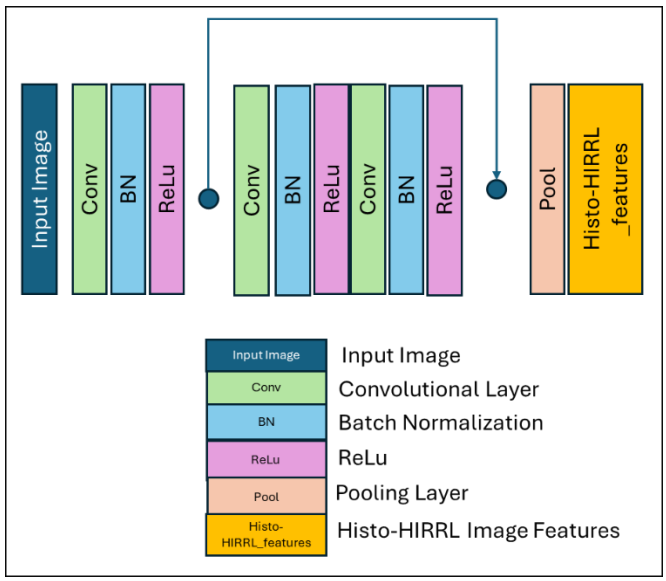


Figure 3.27 A New Feature Extractor Layer Structure of 1 Block Residual Network for Histo-HIRRL Image Feature Extraction

Table 3.13 presents the activation and number of learnables for this deep feature extractor for Histo-HIRRL images. The composite deep extractor was composed of multiple layers that modified the input image and contributed to the model's learning by

adapting parameters. The input image was a histogram with dimensions of 300 x 700 x 1 and no trainable parameters. The initial layer employed a 2D convolution technique, generating 16 feature maps with the same dimensions. The number of trainable parameters in this procedure was 416, and governed by the kernel size and input channels. Batch normalization maintains the dimensions of the feature maps while incorporating 32 trainable parameters to adjust the scale and shift. The second 2D convolution reduced the spatial dimensions by half to 150 x 350 and raises the number of feature maps to 32. This operation utilized a total of 4640 parameters that can be learned. Afterwards, an extra batch normalization layer was added, resulting in an additional 64 parameters. The final 2D convolution maintained the dimensions of the feature map while integrating 9248 parameters. After applying batch normalization with 64 parameters, a skip connection was created by conducting a 2D convolution with a 1x1 kernel, resulting in the inclusion of 544 more parameters. An average pooling layer was used to decrease the dimensions of the feature map to 75 x 175 x 32. This reduction was achieved without introducing any extra parameters.

Table 3.13
Activations and Number of Learnable Parameters for a New Composite Feature Extractor

Layer	**Activations**	**Number of Learnable**
Input Image (Histo-HIRRL)	300 x 700 x 1	-
2-D Convolution 1	300 x 700 x 16	416
Batch Normalization 1	300 x 700 x 16	32
2-D Convolution 2	150 x 350 x 32	4,640
Batch Normalization 2	150 x 350 x 32	64
2-D Convolution 3	150 x 350 x 32	9,248
Batch Normalization 2	150 x 350 x 32	64
2-D Convolution (Skip Connection)	150 x 350 x 32	544
Average Pooling layer	75 x 175 x 32	-
Total	-	1,275,011

This network was customized with individual parameters from the neural network model for dysgraphia histogram images, as presented in Table 3.14. This network used the SGDM optimizer. To estimate the step size needed to minimize the loss, an initial learning rate of 0.001 was chosen. This implies that the model iterated and modified the weights with each iteration. The model runs through 26 epochs, suggesting that the network processes the dataset 26 times in total during the phase (Goodfellow et al., 2017), using 16 minibatch sizes. A single epoch comprised 1251 iterations, with weights changing every 30 iterations. Several parameters influence extraction performance, such as the optimizer, learning rate, epochs, and iterations per epoch.

Table 3.14
Parameter Setting for ResNet-18 of Feature Extraction

Parameter	**Value Setting**
Optimizer	Stochastic Gradient Descent with Momentum (SGDM)
Learning rate	0.001
Epochs	26
Mini batch size	16
Frequency	30

Overall, this newly proposed composite residual network represents a novel and efficient feature extraction strategy tailored for histogram-based handwriting representations, improving classification performance while minimizing model complexity (W. Chen et al., 2021; S. Wu et al., 2016).

3.5.2.2 Machine Learning Classifier for H2FCD Classification Model

In this study, the ML classifier refers to a supervised machine learning model, where the model's objective is to accurately predict the correct label for a given input with fused features. The study utilized four machine learning models namely Support Vector Machine (SVM), K-nearest neighbour (KNN), Random Forest (RF), and Bagged Tree. All the models were trained and tested to accurately identify the severity

class in dysgraphia handwriting images. Among all classifiers the best performing machine learning classifier was selected for the H2FCD classification model.

i) Support Vector Machine (SVM)

This study utilized the multiclass SVM (Darmatasia and Mohamad Ivan Fanany, 2017) classifier with *k-fold* cross-validation to classify the severity of dysgraphia handwriting. The classifier was implemented using MATLAB's *fitcecoc* function, which employs the One-vs-One (OvO) method. This method constructs multiple binary classifiers for each pair of classes to determine the final classification. The *fitcecoc* function uses a linear kernel to map the input data to a higher-dimensional space, enabling the classification of linear decision boundaries.

ii) K-Nearest Neighbour (KNN)

This study assesses the efficacy of a multi-class KNN (W. Liu et al., 2020) classifier by employing *k-fold* cross-validation. The classifier utilizes MATLAB's *fitcknn* function to construct the cosine distance metric for identifying the closest neighbors. During each iteration of the cross-validation process, the dataset is divided into separate training and testing sets.

iii) Random Forest (RF)

This study assesses the efficacy of a multi-class classification model by employing a RF classifier (Masood et al., 2023) with *k-fold* cross-validation. The classifier is developed in MATLAB using the *fitensemble* function using the 'Bag' technique, which generates an ensemble of 50 decision trees for classification.

iv) Bagged Tree (BT)

This study utilizes a BT classifier (Sesa et al., 2022) to assess the effectiveness of multi-class classification through the implementation of *k-fold* cross-validation. The classifier is built in MATLAB using the *fitensemble* function, which use the 'Bag' method to construct an ensemble of 50 decision trees for classification.

Table 3.15
Configuration Settings of Machine Learning Classifiers

Classifier	MATLAB Function	Learning Type	Kernel / Distance Metric	Ensemble Method	No. of Trees / Neighbors	Key Parameters
Support Vector Machine (SVM)	fitcecoc	Supervised (Multiclass, OvO)	Linear Kernel	Not applicable	Not applicable	One-vs-One (OvO) strategy
K-Nearest Neighbour (KNN)	fitcknn	Supervised (Multiclass)	Cosine Distance	Not applicable	1 (default)	Distance metric: Cosine
Random Forest (RF)	fitensemble	Supervised (Multiclass)	Not applicable	Bagging ('Bag')	50 trees	Ensemble of decision trees
Bagged Tree (BT)	fitensemble	Supervised (Multiclass)	Not applicable	Bagging ('Bag')	50 trees	Bootstrap aggregation

Table 3.15 summarizes the configuration settings and implementation details of the machine learning classifiers employed in this study. These configurations ensure a fair and consistent evaluation across all models under the same validation framework.

3.5.2.3 Model Training and Validation using k-fold Cross-Validation

The model is then trained and validated to obtain the optimized classifier. In this study, the training process was validated using cross-validation. Cross-validation is valuable in machine learning classification applications due to its ability to ensure the reliability of the model and prevent bias caused by any individual class in the dataset (Y. Xu & Goodacre, 2018). In this study, the values of *k* were set to 2, 3, 4, and 5. Subsequently, each fold divided the dataset into separate training and testing sets. The training set was utilized to train the machine learning model, which then made predictions for both the training and test sets. The training and test accuracies were derived by comparing the predicted labels with the actual labels, which allowed the model's performance for each fold to be assessed.

Table 3.16 depicts the partition of data for each fold. *K-fold* cross-validation involved an iterative process that was repeated *k* times, with each iteration corresponding to one of the folds. During each iteration, the model was trained using a combination of *k-1* folds, where the model was exposed to slightly different subsets of the data in each iteration. This process helped the model discover patterns that were

more generalized. The unused fold was utilized as the test set in each iteration, enabling the evaluation of the model's performance on unseen data. Following the first training and testing, the procedure proceeded as follows: in each subsequent iteration, a different subset was chosen as the test set, while the model was trained on the remaining *k-1* subsets. This iterative process of rotation persisted until each fold had been utilized once as the test set, providing a comprehensive evaluation of the model over all subsets of the data.

The results from different *k-fold* configurations were compared to determine the most consistent and optimal validation strategy. The selected *k* value was then used to finalize the model before testing on unseen data, ensuring that the model demonstrated good performance.

Table 3.16
Data Partition for Each Fold Values

Number of Fold	Number of Data Partitions
2	68,68
3	45, 45, 46
4	34, 34, 34, 34
5	27, 27, 27, 27, 28

The model's precision and recall were assessed for each fold. The confusion matrix was computed to determine the True Positives (TP), False Positives (FP), and False Negatives (FN) for each class, which were subsequently utilized to calculate precision and recall. The analysis comprehensively assessed the classifier's ability to identify and remember instances from each class reliably. The mean training and test accuracies were computed across all folds, yielding a comprehensive performance metric for the classifier. In addition, the system generated average precision and recall numbers for each class, offering additional specific information about the performance of each individual class. The ROC curve displayed the relationship between the true positive rate and false positive rate for each class, using projected scores as a visual representation. This assessment approach utilized accuracy, precision, recall, and ROC-AUC metrics to evaluate the multi-class model classifier's performance objectively.

3.6 Phase 4: Evaluation Performance of Proposed Multiclass Classification Model H2FCD

This section presents the performance analysis for multiclass classification of dysgraphia severity using handwritten images as depicted in Figure 3.28. The proposed H2FCD model was tested on new data images. The measurements included a confusion matrix and an ROC-AUC graph showing accuracy, precision, recall, and F1-score. This phase also included an output analysis of the classification output from the H2FCD model and expert evaluation. The details of the evaluation are further explained in the next section.

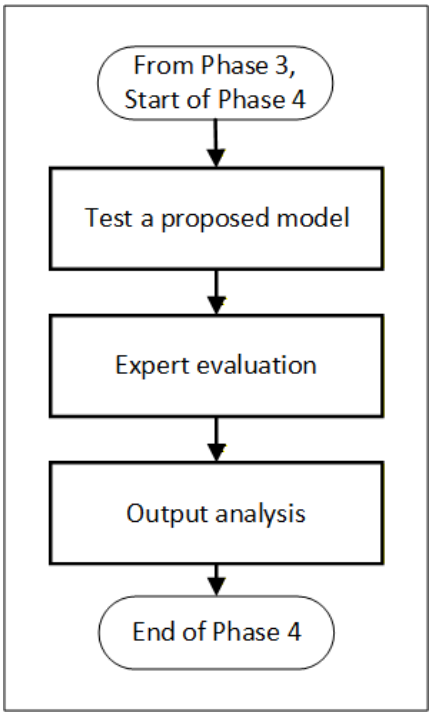


Figure 3.28 Classification Performance Evaluation for the Dysgraphia Classification Model

3.6.1 Classification Testing of the Proposed H2FCD Model using Metric Evaluation

After completing the training and validating the H2FCD model in Phase 3, the model was then tested with 36 new data images as listed in Table 3.10. Figure 3.29 shows the testing process of the proposed H2FCD model. The model testing process began by preparing the system to evaluate the H2FCD model using new handwriting

images. First, the new data images were loaded to ensure that the evaluation reflects the model's ability to generalize beyond the training dataset. Next, the previously trained H2FCD model was loaded into the system. Once both the model and data were ready, the testing phase proceeded by feeding the new images into the model to generate predictions based on the features it had learned. After obtaining these outputs, a performance analysis was conducted using appropriate evaluation metrics and a ROC-AUC graph.

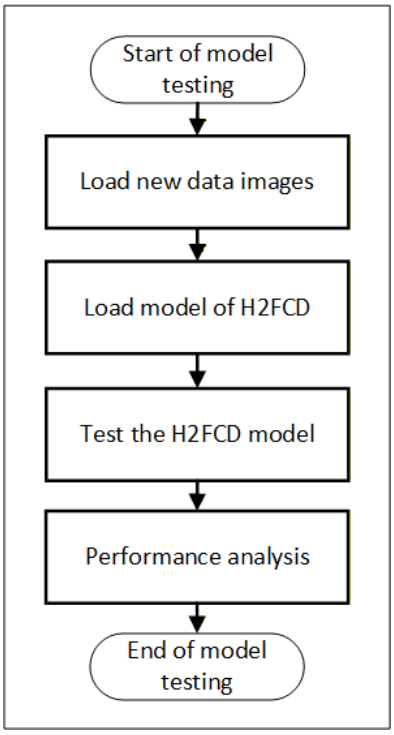


Figure 3.29 Testing of the Proposed H2FCD Model With New Data Images

The classification output obtained during testing was used as a reference in Section 3.6.3, which employed Spearman's Correlation Coefficient to measure the correlation with classification output from expert evaluation.

3.6.1.1 Classification Performance using Metric Evaluation

This section presents the classification performance metric used in this study, specifically to evaluate the performance of the investigation described in Section 3.5.2 and to assess the testing stage of the H2FCD model in Section 3.6.1.

This analysis encompasses the utilization of a confusion matrix within the context of multiclass classification (Grandini et al., 2020). The performance of

classification was calculated based on a confusion matrix, as shown in Table 3.17. In Phase 4, the primary objective is to evaluate the model by performing multiclass classification of dysgraphia severity into three classes: high risk, moderate risk, and low risk. The performance was assessed using a confusion matrix, which provided a comprehensive evaluation by considering all possible prediction outcomes.

Table 3.17
Multiple Classes of Confusion Matrix

		Predicted		
		High Risk (H)	**Moderate Risk (M)**	**Low Risk (L)**
Actual	High Risk (H)	P_{HH}	P_{HM}	P_{HL}
	Moderate Risk (M)	P_{MH}	P_{MM}	P_{ML}
	Low Risk (L)	P_{LH}	P_{LM}	P_{LL}

Table 3.18 presents the definitions associated with each prediction result that was obtained. P_{HH}, P_{MM}, and P_{LL} are the accurate prediction image numbers corresponding to their respective classes. All the remaining prediction images represent misclassifications into particular classes. The definitions of confusion matrix assessment components, including true positive, false positive, false negative, and true negative, are provided in Table 3.19. Classifier performance was assessed using these confusion matrix metrics.

Table 3.18
List of Dysgraphia Severity Level for Each Category

Label	**Interpretation**
P_{HH}	The number of predictions where images of a High Risk were correctly classified as a High Risk. This is also the TRUE POSITIVE for the High Risk class.
P_{HM}	The number of predictions where images of a High Risk were incorrectly classified as a Moderate Risk
P_{HL}	The number of predictions where images of a High Risk were incorrectly classified as a Low Risk
P_{MH}	The number of predictions where images of a Moderate Risk were incorrectly classified as a High Risk

Label	Interpretation
P_{MM}	The number of predictions where images of a Moderate Risk were correctly classified as a Moderate Risk. This is also the TRUE POSITIVE for the Moderate Risk class.
P_{ML}	The number of predictions where images of a Moderate Risk were incorrectly classified as a Low Risk
P_{LH}	The number of predictions where images of a Low Risk were incorrectly classified as a High Risk
P_{LM}	The number of predictions where images of a Low Risk were incorrectly classified as a Moderate Risk
P_{LL}	The number of predictions where images of a Low Risk were correctly classified as a Moderate Risk. This is also the TRUE POSITIVE for the Low Risk class.

Table 3.19
Interpretation of the Term Used in Confusion Matrix

Matrix Evaluation	Interpretation
True Positive (TP)	The TP is the number of predictions where data labeled to belong to a particular class was correctly classified as the said class. Example: P_{HH} is a number of predictions where images of a High Risk were correctly classified as a High Risk. $TP = P_{HH} + P_{MM} + P_{LL}$
False Negative (FN)	The total number of FN's for a class is the sum of values in the corresponding row except the TRUE POSITIVE values Example: $FN_H = P_{HM} + P_{HL}$
False Positive (FP)	The total number of FP's for a class is the sum of values in the corresponding row except the TRUE POSITIVE values Example: $FP_H = P_{MH} + P_{LH}$
True Negative (TN)	The total number of TNs for a certain class will be the sum of all columns and rows, excluding that class's column and row. For example, the calculation for True Negative for the High-Risk class. Example: $TN_H = P_{MM} + P_{ML} + P_{LM} + P_{LL}$

Various metrics were employed in this study to evaluate the performance of the proposed classifier model. Accuracy is a prevalent metric used to evaluate the effectiveness of a classifier model in machine learning. It is defined as a classification matrix that measures the proportion of correct predictions out of all predictions produced, as mentioned in (3.7). The accuracy of all case numbers was computed using

the equation (Zeng, 2025) based on fold cross-validation numbers in this study as follows:

$$\text{Accuracy}= \frac{\text{TP+TN}}{\text{TP+TN+FN+FP}} \tag{3.7}$$

Precision is a quantitative measure used to evaluate the correctness of forecasting a specific category inside a multi-class confusion matrix. The computation entails computing the ratio of the number of accurately recognized positive cases of the particular class to the total of accurately identified positive instances and inaccurately identified positive instances. Precision, essentially, refers to the quantity of correct positive forecasts. This study utilized this metric to ascertain the accuracy of classifying handwriting samples as high-risk and to quantify the total count of such samples. The precision (Zeng, 2025) can be determined by applying the specified calculation in (3.8):

$$\text{Precision}= \frac{\text{TP}}{\text{TP+FP}} \tag{3.8}$$

Recall, often refers to as sensitivity, is the evaluation of data's capacity to identify positive outcomes with respect to the true facts accurately. The precision of the positive predictions is assessed by the ratio of accurately anticipated positive events. This investigation calculated the number of occurrences of standard handwriting that the model correctly labeled. The recall equation (Zeng, 2025) is shown in the equation (3.9).

$$\text{Recall}= \frac{\text{TP}}{\text{TP+FN}} \tag{3.9}$$

The F1 Score (Heydarian et al., 2022) represents the model's harmonic mean, indicating that it has critical Precision and Recall levels. The F1-score is the harmonic mean of precision and recall, as seen in (3.10).

$$\text{F1-score}=2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{3.10}$$

The Matthews Correlation Coefficient (MCC) (Tamura et al., 2025) is a balanced performance metric used to evaluate the overall quality of a classification H2FCD model. MCC considers all elements of the confusion matrix (true positives, true negatives, false positives, and false negatives). This makes it a more reliable indicator of how well the model predicts each class. MCC produces a score between –1 and +1, where +1 indicates perfect classification, 0 represents performance no better than random guessing, and –1 reflects total disagreement between predictions and actual labels. In multi-class problems, MCC extends naturally and presents a single correlation-like value that summarizes the consistency of the model to assign samples to their correct classes. Because MCC incorporates all correct and incorrect predictions across all classes, it provides a balanced measure of classification performance. The MCC is calculated using (3.11)

$$\mathrm{MCC} = \frac{c\,s - \mathrm{t}\cdot\mathrm{p}}{\sqrt{(s^2 - \mathrm{p}\cdot\mathrm{p})\,(s^2 - \mathrm{t}\cdot\mathrm{t})}} \tag{3.11}$$

where,

C is the K×Kconfusion matrix with C_{ij}= (# samples whose true class = iand predicted class = j).

c=$\sum_k C_{kk}$ is the total number of correct predictions (trace of C).

$s = \sum_{ij} C_{ij}$ is the total number of samples.

t is the vector of true-class totals, t_k=$\sum_j C_{kj}$ (row sums).

p is the vector of predicted-class totals, p_k=$\sum_i C_{ik}$ (column sums).

3.6.1.2 Visualization Performance Using ROC-AUC Graph

Evaluating the performance of a classifier is crucial in machine learning to understand its effectiveness and making informed judgments about enhancing the model. The Receiver Operating Characteristic Curve (ROC-AUC) is an extremely valuable metric for assessing the performance of classifiers (Majnik & Bosnić, 2013). While the concept of ROC-AUC is straightforward in binary classification, its implementation in multi-class classification requires more intricate procedures. The

ROC curve visually depicts the relationship between the true positive rate (TPR), also known as recall, and the false positive rate (FPR) across varying threshold values. It vividly illustrates the equilibrium between sensitivity and specificity. The AUC measures the degree of distinguishability achieved by the model: an AUC of 1 indicates perfect classification, whereas an AUC of 0.5 suggests a lack of ability to discriminate.

The performance of different classifier models was evaluated to classify three categories of dysgraphia handwriting (low risk, moderate risk, and high risk). For multiclass classification, ROC curves were generated for the validation data of each class, and the maximum average AUC value of all classes was used as the selection criterion. The model with the highest average AUC value was selected as the superior model.

The AUC - ROC curve is a performance metric for classification issues with different threshold values. ROC is a probability curve, whereas AUC is a measure or degree of separability. It indicates how well the model can differ across classes. The larger the AUC, the more accurately the model predicts 0 classes as 0 and 1 classes as 1. By analogy, the greater the AUC, the greater the model's ability to classify dysgraphia severity. Figure 3.30 shows the example of the AUC – RPC curve, which indicates that ROC1 is a better classifier than ROC2.

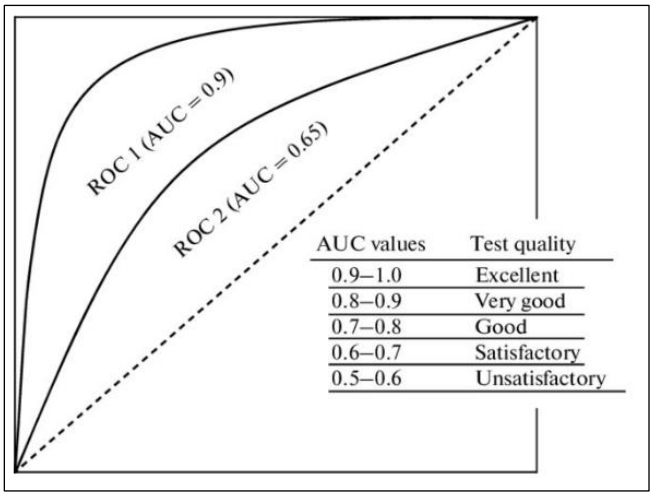


Figure 3.30 Examples of ROC Curves Based on the Model's Performance (Srinivasan & Mishra, 2024)

3.6.2 Classification Based on Expert Evaluation

This research involved implementing an expert evaluation procedure to validate the performance of the developed classification system. The evaluation process involved the participation of eight experts, as detailed in the following procedure:

i) A form was distributed to all experts as outlined in Appendix 7.
ii) Each handwriting image was subjected to expert observation.
iii) Experts assess the image by classifying it into one of the severity levels: high-risk, moderate-risk, or low-risk, using the provided evaluation form.
iv) All findings from expert observations were aggregated for analysis

An evaluation form was then designed to provide a standardized, method for completing the assessment. This form (Appendix 7) presents the processed image, and the evaluators were required to classify each image into one of the three categories: high-risk, moderate-risk, or low-risk, as shown in Figure 3.31.

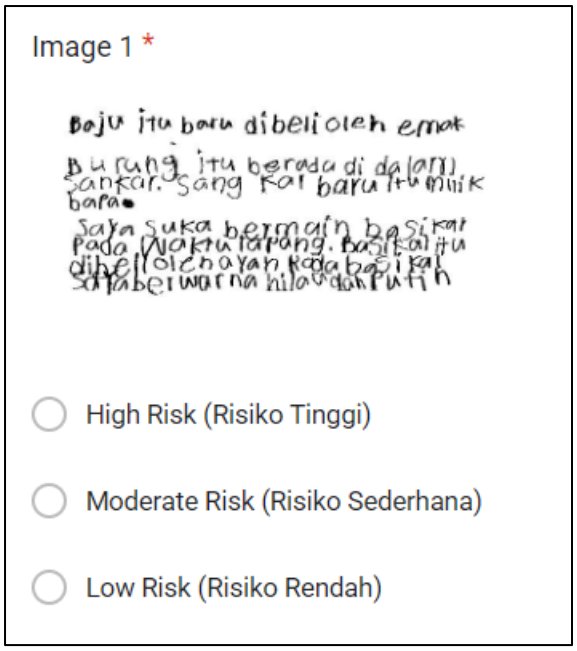


Figure 3.31 An Example of a Handwriting Image for Expert Observation and Severity Classes

The total of 30 images was provided for the expert evaluation process, presented as a list of handwriting images, as shown in the example in Figure 3.31. Expert judgments were based on the expert's experience and observation of the image, using

established criteria for dysgraphia symptoms. The resulting classification outputs were then analyzed using Spearman's Correlation Coefficient, as further described in the next section.

3.6.3 Output Analysis Using Spearman's Correlation Coefficient

This section presents the output analysis of the proposed H2FCD classification model using Spearman's Correlation Coefficient as depicted in Figure 3.32. This analysis aims to validate the proposed H2FCD model by measuring its correlation with expert classification outputs.

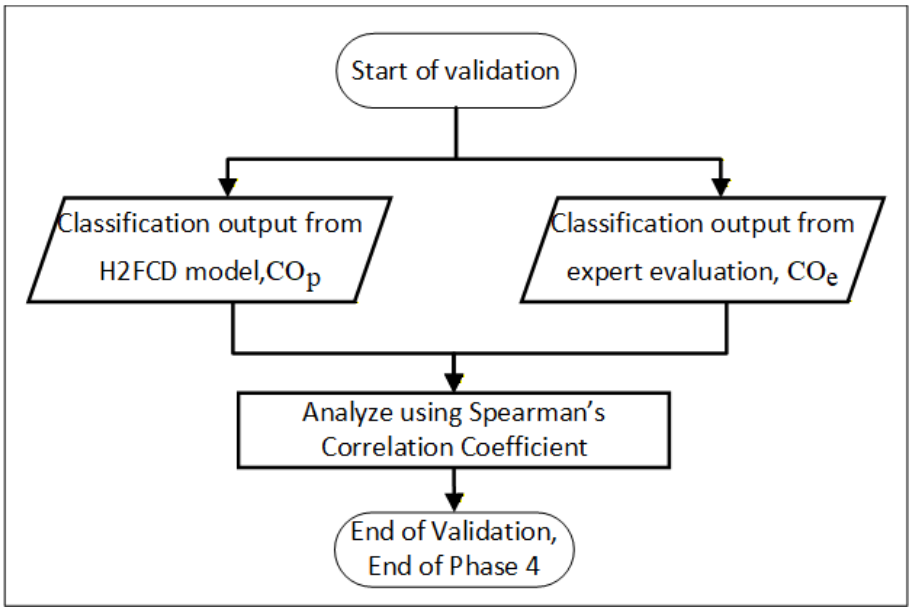


Figure 3.32 Output Analysis of the proposed H2FCD Classification Model Using Spearman's Correlation Coefficient

This analysis used the classification output from the H2FCD model, CO_p, and expert evaluation,CO_e, based on a total of 30 images. Spearman's correlation coefficient was employed to analyze the correlation between experts' judgment and the classification outputs generated by the proposed H2FCD model. This outputs served as a benchmark for comparing the classification output of the proposed system with that of experts and identifying potential areas for improvement.

3.7 Summary

This chapter has comprehensively presented the research methodology, including the proposed model and the evaluation methods employed in this study. In Phase 1, the database development that contain all handwriting images written by school-age children had been developed for supervised learning of the proposed model. The handwriting image was processed to be ready for use during phase 2. In Phase 2, a new scheme for ruling line was proposed to ensure that it did not influence the handwriting and to preserve the information depicted in the handwriting features. The removed ruling line removal images were then used as a feeder to extract and classify the handwriting image in Phase 3. A dual-input CNN-based fusion feature model namely H2FCD had been designed to classify dysgraphia severity using an ML classifier model. Finally, in Phase 4, the proposed classification model had been evaluated based on metrices evaluation and expert validation. All classification performance was measured using the confusion matrix concept and the ROC-AUC graph as visualization metrics. Expert evaluation process was conducted to validate the proposed model by analyzing the correlation coefficient between both CO_e and CO_p.

CHAPTER 4
RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents a comprehensive overview of the findings acquired at each phase of the research study, along with a thorough discussion that addresses all the objectives outlined in Section 1.3. The first phase involves the development of the dataset, which encompasses data collection, data description, and data analysis. The findings indicate a significant correlation between the evaluations conducted by the two assessors for each data handwriting sample following the data collection process. The second phase involved the ruling line removal procedure, which includes comparison between the original handwriting images and the removed line handwriting images to ensure minimal feature loss. A comparative analysis had also been conducted between HIRL and HIRRL inputs to examine the impact of removing the ruling line on classification performance. The third phase presents the classification performance of the proposed dual-input fusion feature-based model using an ML classifier. Finally, Phase 4 presents the validation of model performance through correlation analysis between the proposed H2FCD model and expert evaluations.

4.2 Phase 1: Result and Analysis of New Malay Dysgraphia Handwriting Dataset (MLHSD)

This section explains the outcome of objective 1, as outlined in 1.3. This section also presents the findings of dataset development, including a reliability analysis of the collected data and digitalization outcomes. Detailed data analyses are objectively discussed in 4.2.1, 4.2.2 and 4.2.3. The data analysis is presented to assess the agreement and correlation between two assessors who have evaluated and categorized the severity level of dysgraphia. The data analysis employed Cohen's Weighted Kappa to assess the level of agreement and Spearman's rank correlation coefficient to examine the association between the two assessors with the severity ratings they assigned. Additionally, a correlation analysis was performed between dysgraphia class and age, as well as between dysgraphia class and gender to observe the association between these

variables and the severity level rankings. As an initial analysis, the reliability of the data was measured, yielding a Cronbach's Alpha, α, of 0.987, which exceeds 0.9, indicating a high level of data reliability and trustworthiness. This result supports the suitability of the dataset for further statistical analysis.

4.2.1 Analysis of MLHSD Dataset using Cohen's Weighted Kappa

Table 4.1 presents the agreement between the two assessors which was evaluated using Weighted Cohen's Weighted Kappa, a reliability measure suitable for the severity level rating data assigned by both assessors. The analysis yielded a Weighted Kappa value of 0.961, indicating an extremely high level of agreement. According to the interpretation scale by Landis & Koch (1977), values between 0.81 and 1.00 represent almost perfect agreement, suggesting that the assessors' ratings were nearly identical. The standard error of 0.016 indicated a precise estimate, while the z-value of 16.332 with $p < 0.001$ confirmed that the agreement was statistically significant. The 95% confidence interval indicates that the lower bound is 0.931 and the upper bound is 0.992, further supporting the robustness of the result and reinforcing that the assessors demonstrated almost perfect consistency in their ratings.

Table 4.1
Agreement Analysis of Cohen's Weighted Kappa for Two Assessor Evaluation

Ratings	**Weighted Kappa**	**Asymptotic**			**95% Confidence Interval**	
		Std. Error	**z**	**Sig.**	**Lower Bound**	**Upper Bound**
Assessor 1 - Assessor 2	0.961	0.016	16.332	<0.001	0.931	0.992

4.2.2 Analysis of MLHSD Dataset Using Spearman's Correlation

Spearman's rank correlation coefficient was used to examine the correlation between the severity ratings assigned by the two assessors. Table 4.2 shows the analysis yielded a correlation coefficient ρ of 0.974, which is close to 1. This demonstrates a strong correlation between the ratings made by the two assessors. The analysis revealed a very strong positive correlation between the assessors, indicating that higher ratings by one assessor were closely associated with higher ratings by the other assessor. With

a p-value of < 0.001 for 178 data, this statistically significant result demonstrates that the two assessors ranked the severity levels in a highly consistent manner.

Table 4.2
Spearman's Correlation Coefficient for Two Assessor Evaluations

		Assessor1	**Assessor2**
Assessor1	Correlation Coefficient	1.000	0.974
	Sig. (1-tailed)		<0.001
	N	178	178
Assessor2	Correlation Coefficient	0.974**	1.000
	Sig. (1-tailed)	<0.001	
	N	178	178

In the context of research or assessments, these findings demonstrate that the measures are reliable and consistent across all observed assessment scenarios. The analysis results present a consistently high level of concordance between the two assessors in evaluating severity levels. The Cohen's Weighted Kappa value of $\kappa = 0.961$ demonstrated almost perfect agreement, reflecting a high degree of consistency in the ratings assigned. Complementing this, Spearman's rank correlation coefficient, ρ which achieved a value of 0.974 revealed a very strong positive association indicating that the ranking patterns of severity levels by both assessors were closely aligned. Collectively, these results provide strong statistical evidence that the assessors were highly consistent in both their ratings and the relative ordering of severity levels.

According to the correlation analysis, data with inconsistent ratings have been excluded. Out of 178 participants, 172 were chosen based on their handwriting and intraclass correlation interval; both raters had the same class identification. The assessors divided the data into three classes namely high-risk, moderate-risk, and low-risk that represent the severity levels of dysgraphia based on handwriting characteristics.

4.2.3 Data Analysis Results using Pearson's Correlation Analysis

Based on the 172 data analyzed in Section 4.2.2, further correlation analyses are conducted to examine the relationships between dysgraphia severity class, age, and gender. Table 4.3 tabulates the correlations between "age" and "dysgraphia_class" using Pearson correlation coefficient. This correlation measures the strength and direction of a linear relationship between two variables. The Pearson correlation coefficient between "age" and "dysgraphia_class" was -0.618. The variable "dysgraphia_class" had a -0.618 Pearson association with "age". The correlation coefficient's negative sign implies a negative relationship between age and dysgraphia class. The dysgraphia class decreases with age (or vice versa). The low sig-value (<0.001) indicates that this association was unlikely to occur by chance. Consequently, statistical evidence supports the conclusion that in the sample, there is a significant negative association between age and dysgraphia class. In this research section, a statistically significant negative correlation was discovered between age and dysgraphia class among the 172 participants.

Table 4.3
Correlations Between Age and Dysgraphia Category

		Age	**Dysgraphia_Class**
age	Pearson Correlation	1	-0.618
	Sig. (2-tailed)		<0.001
	N	172	172
Dysgraphia_class	Pearson Correlation	-0.618	1
	Sig. (2-tailed)	<0.001	
	N	172	172

Table 4.4 displays a statistical analysis of the association between dysgraphia class and gender, with a sample size of 172 participants. The Pearson correlation coefficient between dysgraphia_class and gender was found to be -0.188. The correlation coefficient's negative sign indicates a weak negative correlation between the two variables. This indicates that as the dysgraphia class rises (with a higher level of dysgraphia), gender corresponds to a specific pattern (male or female), and vice versa. Both correlations have a sig-value of 0.014, which is less than 0.05. This suggests that

the correlation is statistically significant, implying that the modest negative link between dysgraphia_class and gender could be attributable to chance rather than a meaningful or dependable discovery.

Table 4.4
Correlations Between Gender and Dysgraphia Category

		Dysgraphia_Class	**Gender**
Dysgraphia_class	Pearson Correlation	1	-0.188
	Sig. (2-tailed)		0.014
	N	172	172
gender	Pearson Correlation	-0.188	1
	Sig. (2-tailed)	0.014	
	N	172	172

The reliability, gender, and age correlation analyses are presented to support the examination of factors related to dysgraphia severity classification among schoolchildren. The intraclass correlation coefficient analysis indicates that both assessors demonstrated high agreement and correlation in identifying the severity level of the handwriting samples. Age showed a weak negative association with dysgraphia class, and a similarly weak relationship is observed between gender and dysgraphia class. Due to these findings, neither gender nor age will be incorporated as variables in the development of the dysgraphia severity classification model.

4.3 Phase 2: Results and Analysis of New Scheme of Ruling Line Removal Technique using RGBB-based Algorithm

Phase 2 focuses on presenting the results of new scheme for ruling line removal, which was implemented in various handwriting patterns by children with potential dysgraphia symptoms, as explained in Phase 1 in Section 3.3.3. This section provides analysis of the results related to objective 2, as mentioned in Section 1.3 which covers the following:

i) Performance evaluation of proposed ruling line removal method using SSIM
ii) Comparison of multiclass classification using the proposed new scheme of ruling line removal technique (RGBB-based) based on HIRL and HIRRL input images

4.3.1 Performance Evaluation Using Structural Similarity Index (SSIM)

SSIM is used to evaluate the structural similarity of handwriting images after removing ruling lines. It is calculated by comparing structural information in the original and removed ruling line images. This value is regarded as high and close to 1. A high Structural Similarity Index (SSIM) score indicates that the treated image closely resembles the original, indicating successful removal of ruling lines without significant distortions. The SSIM analysis (Nigam et al., 2024; Sara et al., 2019) has been successfully applied to all dataset images, resulting in several measurements including average, standard deviation, median, maximum, and minimum values that represent the overall SSIM for each image.

Table 4.5 shows the findings of the SSIM based on a quantitative metric measurement to evaluate the degree of similarity between two distinct types of handwriting images. The SSIM is scaled within a numerical range of -1 to 1, where a higher value signifies a greater degree of similarity, i.e; a value of 1 denoting the presence of identical images. This study presents a mean SSIM score of 0.859, indicating an acceptable similarity between the removed ruling line and the original image. The standard deviation was 0.038, indicating minimal variability in the data, with most values distributed very close to the mean. The median SSIM value was recorded at 0.865, indicating that a minimum of 50% of the image pairs demonstrate an SSIM score close to the perfect value. The highest recorded SSIM score was 0.961, which signifies an almost perfect degree of similarity in certain instances. In contrast, although the lowest value observed was 0.730, it still indicates a satisfactory level of similarity. As a summary, the findings presented herein indicate that all the images demonstrated significant structural similarity index, with a predominant number reflecting high quality or exhibiting minimal distortion.

Table 4.5
Analysis of SSIM Results Based on Different Measurements

Calculation	**SSIM**
Average	0.859
Standard Deviation	0.038
Median	0.865
Maximum	0.961
Minimum	0.730

In relation to SSIM, Figure 4.1 presents an average result of SSIM values, systematically classified according to all three levels of dysgraphia risks. The SSIM results serve as an average value measure of a similarity structure between handwriting samples and established reference standards, to represent an evaluation of handwriting quality after removing ruling lines.

The High-Risk category has an average SSIM value of 0.849, which, while still showing some similarity, suggests a greater deviation from the reference standard. The Moderate Risk category exhibits the average SSIM value at 0.869, suggesting that handwriting in this group is more structurally similar to the reference, indicating less severe but still notable dysgraphia traits. The Low-Risk category has a slightly lower SSIM score compared to High-Risk group with an average value of 0.858. This indicates that this category also demonstrates some degree of handwriting variability. In general, the findings of SSIM analyses show that all three categories of dysgraphia risk datasets possess the similarity structure even after the ruling lines had been removed.

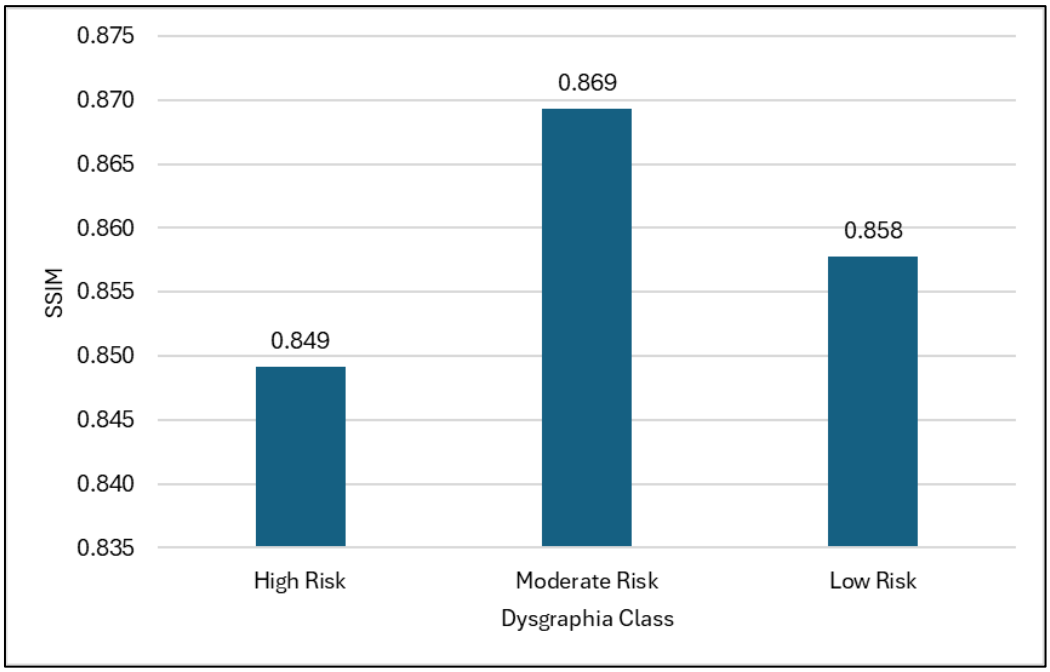


Figure 4.1 The Average Results of SSIM for Each Category of Dysgraphia Class

Figure 4.2 illustrates the distribution of images according to their SSIM values, systematically categorized into defined range intervals. The x-axis establishes the range of SSIM values, systematically partitioned into ten distinct intervals spanning from [0.730, 0.747] to [0.951, 0.968]. The vertical axis represents the quantity of images associated with each specified range of SSIM. The analysis revealed a significant aggregation of images situated within the mid-range spectrum, specifically in the interval [0.866, 0.883], which was characterized by the presence of the maximum quantity of images with a total of 118. The subsequent range was delineated as [0.849, 0.866], encompassing 97 images. The distribution demonstrated a bell-shaped configuration, characterized by a reduced number of images at both ends of the SSIM spectrum; specifically, four images were located within the lowest range [0.730, 0.747], while one image was situated in the highest range [0.951, 0.968]. Nonetheless, the bell-shaped shows slight positive skewness, with a greater proportion of images falling within higher SSIM ranges than lower ones. This indicates that most processed images achieve high structural similarity to the original images, reflecting the effectiveness of the proposed method in minimizing distortion during preprocessing. Only a small number of images were at the lower end of the SSIM range, suggesting that structural degradation occured in only a limited number of cases. Overall, the distribution confirmed that the proposed scheme for RGBB-based ruling line removal produced consistently high-quality outputs across diverse handwriting samples.

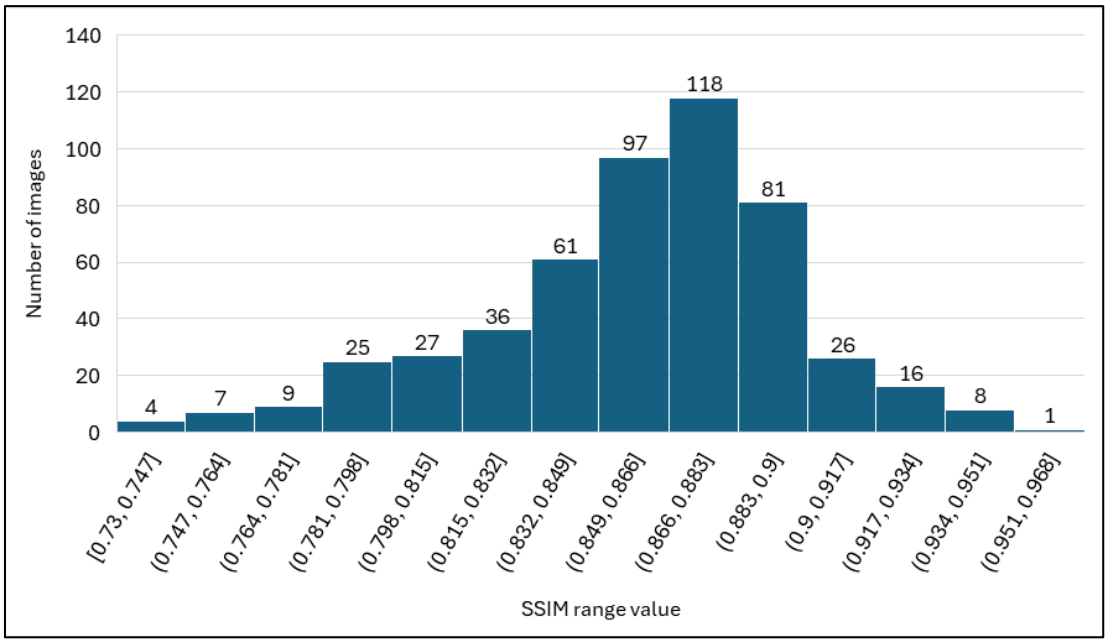


Figure 4.2 Analysis of the SSIM Range Value with the Number of Images Presented in Histogram

The procedure of removing ruling lines has a direct effect on the extraction of succeeding characteristics, ensuring that the structural features are preserved and aiding in the emphasis of important features. The significance of this lies in its ability to enable the use of precise and dependable feature extraction methods to process handwriting images. The next section discusses the findings of comparative achievements in evaluating the classification performance between two types of images: HIRL and HIRRL.

4.3.2 Classification Performance of the Proposed New Scheme of Ruling Line Removal Technique

In this study, feature extraction was performed using the ResNet-18 model, which is known for its ability to capture hierarchical image features through residual learning. The extracted features were subsequently classified using a Support Vector Machine (SVM) classifier to perform multiclass classification of dysgraphia severity levels on two input data sets: Handwriting image with ruling line (HIRL) and Handwriting image with removed ruling line (HIRRL). Figure 4.3 illustrates the accuracy percentages associated with two distinct inputs in the context of handwriting image classification performance, specifically HIRL and HIRRL. The classification performance was evaluated across four *k*-fold values: 2, 3, 4, and 5. In 2-fold cross-

validation, the classification performance using HIRL input achieved an accuracy of 91.91%. In comparison, the accuracy of classification based on the HIRRL input image is marginally superior, achieving 93.38%. In the context of a three-fold evaluation, the accuracy performance of HIRRL was markedly superior to that of HIRL, achieving accuracies of 94.85% and 91.18%, respectively. In a comparative analysis involving four-fold, implementing HIRRL input images demonstrated a superior accuracy of 94.12%, compared with 91.18% accuracy achieved by using HIRL input images. In the final analysis, within a fold of 5, the HIRRL input data achieves its peak accuracy of 94.89%, thereby exceeding the accuracy of the HIRL input image, which stood at 91.14%. The consistent application of the HIRRL indicated markedly improved performance across all cross-validation folds. This observation implies that eliminating the ruling line may significantly affect classification accuracy, particularly in handwriting-based classification methodologies.

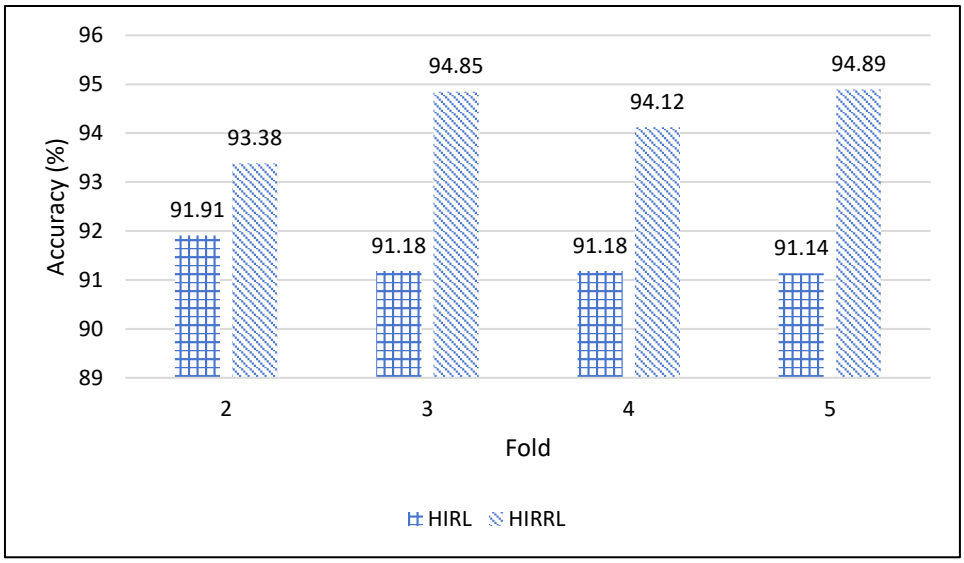


Figure 4.3 Accuracy Performance of Dysgraphia Severity Classification Based on HIRL and HIRRL

The findings contribute to a comparative analysis of two handwriting image classification approaches for dysgraphia severity using HIRL and HIRRL. Moreover, the numbers demonstrate that HIRRL consistently achieves higher accuracy across all tested fold cross-validation values of 2, 3, 4, and 5 compared to HIRL. This suggests that HIRRL provides more effective feature extraction and captures handwriting characteristics more accurately, thereby offering stronger discriminatory features for

differentiating dysgraphia severity levels and better representing handwriting style characteristics.

The superior performance of HIRRL provides a strong foundation for Phase 3 to enhance classification performance. This study provides a comprehensive analysis of automated classification methodologies used to assess dysgraphia severity by analyzing handwriting images. The implementation of feature extraction combined with a support vector machine classifier demonstrated the vital significance of ruling line removal in enhancing classification accuracy, as evidenced by the analysis results across diverse handwriting patterns.

These findings facilitate a comprehensive evaluation of the efficiency of each classification performance in addressing various handwriting styles, whereas HIRL and HIRRL yielded different findings. The comparison systematically compares the handwriting images input influenced, thereby exposing the strengths and weaknesses inherent to each input image used. As a conclusion, the RGBB-based ruling line removal is highly significant to enhance the classification performance with deep features of extraction using ResNet-18 and SVM classifiers. The accuracy performance is supported by several other measurements, namely precision, recall, and F1-score. This is all explained in the next section by presenting the number for each fold cross-validation.

Table 4.6 presents a comparative analysis of the performance metrics of two different classification models by implementing two different input images, namely HIRL and HIRRL. The evaluation is conducted through the evaluation of precision, recall, and F1-score, with varying *k*-fold cross-validation parameters (k = 2, 3, 4, 5) applied to different input images. The classification model when using HIRRL demonstrated superior performance relative to HIRL across all metrics for each k value. When k equals 2, the classification model using HIRRL exhibits superior performance metrics, particularly in precision, recall, and F1-score. This result signifies a greater accuracy in predictions when compared with the classification model when using the HIRL model, particularly under conditions of ruling line existence. With the number of folds increased to k = 3 and k = 4, the classification model using HIRRL consistently demonstrates exceptional performance, particularly in recall and F1-score. The performance results thereby illustrate its enhanced capability in classifying the multiclass severity of dysgraphia. The highest performance is achieved at k = 5, where the HIRRL input attains its optimal performance metrics, with precision at 96.23, recall

at 95.05, and F1-score at 95.63. In contrast, although HIRL input continues to show improvement; however, it remains inferior in both recall and F1-score metrics.

Table 4.6
Performance of Precision, Recall, and F1-Score for Dysgraphia Severity Classification for HIRL and HIRRL

	Precision		**Recall**		**F1-score**	
k-fold	**HIRL**	**HIRRL**	**HIRL**	**HIRRL**	**HIRL**	**HIRRL**
2	91.90	94.57	91.76	92.98	91.83	93.77
3	91.35	95.61	90.67	94.81	91.01	95.21
4	91.03	94.19	90.49	94.39	90.76	94.29
5	93.08	**96.23**	89.92	**95.05**	91.48	**95.63**

This finding of all results suggested that increasing *k* improves the performance of classification models; with the most accurate predictions when employed with 5 folds. HIRRL's consistent performance across the folds demonstrated its effectiveness in identifying relevant instances without missing positive cases. This improvement is due to the significant features of handwriting being better in classifying the dysgraphia severity by removing the ruling line, which presents the true characteristics of dysgraphia handwriting. As a result, HIRRL has been proven to be the preferred input for the classification model, owing to its superior ability to balance precision and recall across all k-folds, particularly as the number of folds increases.

The study demonstrates that HIRRL input is more effective than HIRL in handwriting recognition tasks, offering better precision, recall, and overall accuracy across various scenarios. This suggests that the HIRRL may incorporate more advanced techniques or optimizations to enhance performance in classifying dysgraphic handwriting. The findings substantiate the assertion that HIRRL serves as a more significant input image for the dysgraphia classification model. This research clarifies the implications of ruling-line removal in HIRRL, highlighting its advantages over HIRL.

In summary, the proposed scheme for ruling-line removal using an RGBB-based algorithm makes a significant contribution to enhancing the classification performance of the H2FCD model.

4.4 Phase 3: Training and Validation Performance Analysis of Multiclass Dysgraphia Severity Classification Model (H2FCD)

This section presents and discusses the findings of Phase 3, which covers Objective 3 as mentioned in Section 1.3. This section presents a performance analysis of multiclass classification based on a dual-input, fusion-feature model (H2FCD) using an ML classifier.

4.4.1 Performance of Multiclass Classification Dysgraphia Severity based on Dual-Input Fusion Feature-based using ML classifier

In this section, the results analysis presents comparison results of classification performance for three feature vector types, as follows:

i) Image feature: feature vector from HIRRL image only
ii) Histogram feature: feature vector from Histo-HIRRL image only
iii) Fusion feature: Fusion feature vectors from Image feature of HIRRL and Histogram feature of Histo-HIRRL

The performance of all three feature vector types were analyzed across five different ML models for SVM, KNN, SD, RF, and BT. The evaluation was conducted using 5-fold cross-validation, and the results for all models were presented in a metrics evaluation of accuracy, precision, recall, macro-average AUC, and an ROC-AUC graph. The findings demonstrate the performance of multiclass dysgraphia handwriting severity classification using dual input fusion features.

4.4.1.1 Precision, Recall and F1-score

Table 4.7 presents the performance of different machine learning models (SVM, KNN, RF, and BT) using three types of feature vectors: image, histogram, and fusion features, evaluated based on precision, recall, and F1-score across three dysgraphia severity classes.

Table 4.7
F1-Score Value for All Types of Image Feature Vectors Across Machine Learning Models

			SVM	**KNN**	**RF**	**BT**
Image Feature	Precision	high	97.78	100	98	100
		moderate	93.57	85.56	91.04	91.68
		low	97.33	88.85	88.67	86.77
	Recall	high	96.67	90	86.64	87.59
		moderate	94.92	88.59	92.18	86.79
		low	93.56	94.24	94.67	93.21
	F1-score	high	97.22	94.74	91.97	93.38
		moderate	94.24	87.05	91.61	89.17
		low	95.41	91.47	91.57	89.87
Histogram Feature	Precision	high	93.33	78.89	88.22	86.48
		moderate	88.21	84.97	87.78	86.34
		low	96.92	94.36	91.02	98.33
	Recall	high	91.43	95	90	93.06
		moderate	92	81.92	84.23	87.24
		low	91.89	92.75	90.63	91.75
	F1-score	high	92.37	86.2	89.1	89.65
		moderate	90.07	83.42	85.97	86.79
		low	94.34	93.55	90.82	94.93
Fusion Feature	Precision	high	**100**	96.67	98	97.5
		moderate	98.46	98.18	97.14	94.55
		low	**100**	100	98.18	100
	Recall	high	**100**	100	98	94
		moderate	98.57	97.5	95.14	97.78
		low	**100**	98.18	100	100
	F1-score	high	**100**	98.31	98	95.72
		moderate	98.51	97.84	96.13	96.14
		low	**100**	99.08	99.08	100

The results clearly show that the fusion feature consistently provides the best performance across all evaluation metrics compared to image and histogram features. This indicates that combining both image and histogram information enables the model to capture more comprehensive and discriminative handwriting characteristics, leading to improved classification accuracy.

Among all machine learning models, the Support Vector Machine (SVM) demonstrated the most stable and superior performance, particularly when using fusion features. The SVM achieved near-perfect scores, with several values reaching 100% in precision, recall, and F1-score across all severity classes.

In comparison, Random Forest (RF) also showed strong performance, especially with fusion features, but remained slightly less consistent than SVM. KNN and BT models achieve relatively good results; however, their performance was less stable and exhibited greater variation across feature types and classes.

Based on these findings, it can be concluded that combining fusion features with an SVM classifier provided the most effective and reliable approach for multiclass dysgraphia severity classification. The superior performance can be attributed to the ability of fusion features to enhance feature representation, while SVM effectively learns the optimal decision boundaries for classification.

4.4.1.2 Accuracy

Figure 4.4 presents the accuracy percentages for all four distinct machine learning classification models: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), and Bagged Trees (BT), using three feature types: image, histogram, and fusion. Generally, the results show that fusion features consistently achieved the highest accuracy across all classifiers, followed by image features, while histogram features produced the lowest performance. This indicates that combining image and histogram information provided more comprehensive feature representation, leading to improved classification accuracy.

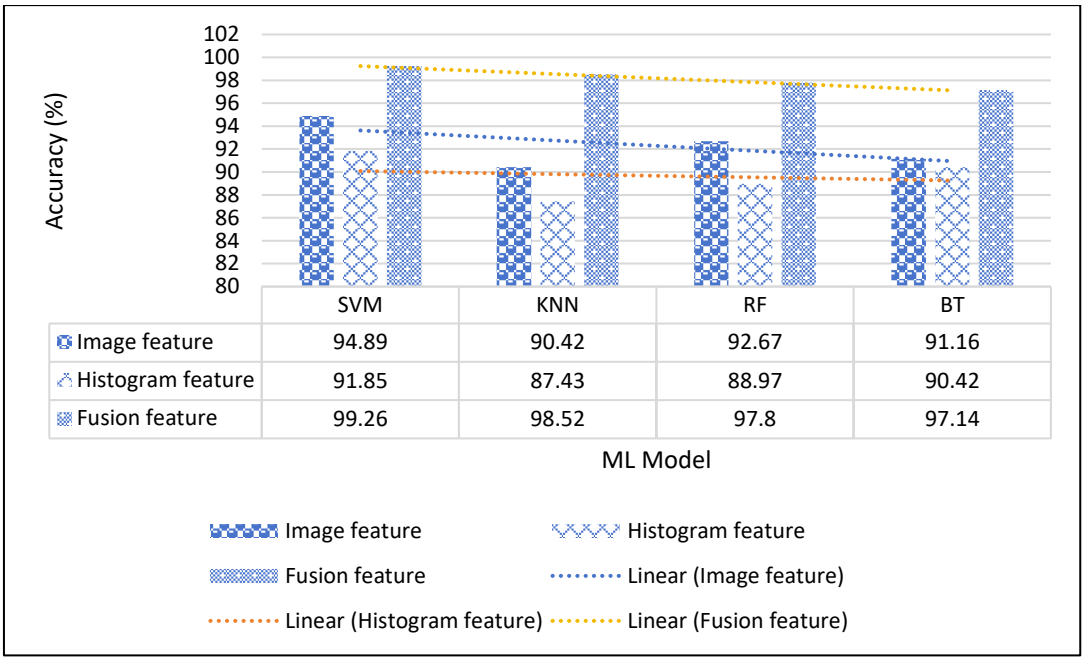


Figure 4.4 Accuracy Performance of Each Fold and Classifier Based on Image Features

Among the classifiers, SVM achieved the best performance, with a fusion feature accuracy of 99.26%. KNN also showed strong performance with fusion features (98.52%), followed by RF (97.80%) and BT (97.14%). In contrast, using histogram features alone yielded noticeably lower accuracy, particularly for KNN, highlighting their limitations when less informative features were used.

These findings are consistent with the earlier results for precision, recall, and F1-score, in which fusion features combined with an SVM also achieved the highest and most stable performance across all evaluation metrics. This confirms that integrating fusion features enhances the model's discriminative capability, while SVM effectively optimizes the classification boundaries, making it the most suitable approach for multiclass dysgraphia severity classification.

4.4.1.3 Macro-Average AUC Value

Figure 4.5 presents the results of the classification performance, namely the macro-average AUC value, for each type of feature. The figure illustrates the macro-average AUC performance of four machine learning classifiers (SVM, KNN, RF, and BT) using three different feature types: image, histogram, and fusion features. Overall, the results show that fusion features consistently achieved the highest AUC values

across all classifiers, reaching a near-perfect score of 1.0. This finding indicates that fusion features provided the most effective representation for distinguishing between dysgraphia severity classes.

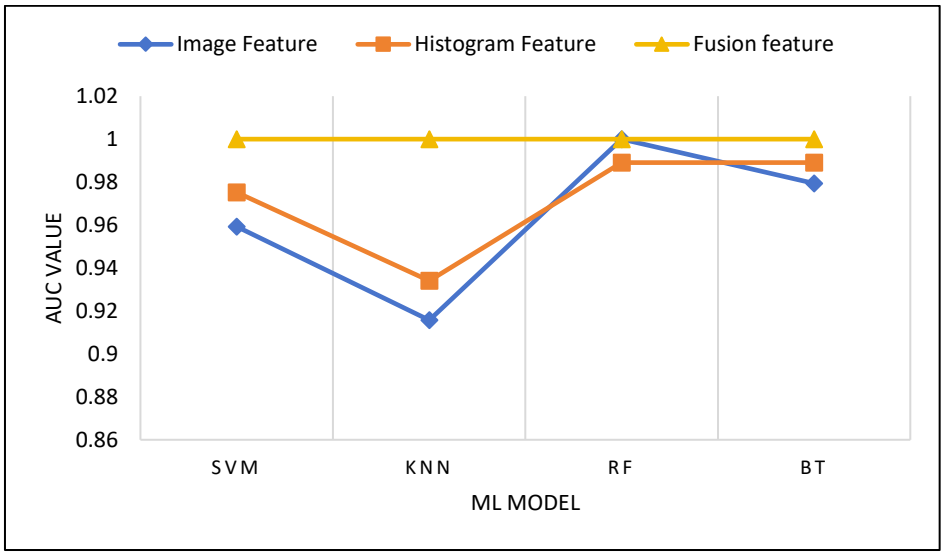


Figure 4.5 Macro-Average AUC Performance of Classifiers Using Image, Histogram, and Fusion Feature Vectors

Among the classifiers, SVM and RF demonstrated strong performance with image features, both achieving AUC values close to 1.0, whereas KNN showed comparatively lower performance, particularly with image and histogram features. Histogram features generally yielded moderate performance across all models, whereas fusion features significantly improved AUC for every classifier.

This trend is consistent with the earlier results on accuracy, precision, recall, and F1-score, in which fusion features combined with SVM also achieved the best and most stable performance. The consistently high AUC values confirm that the proposed fusion feature approach enhances the model's ability to correctly distinguish between classes, while SVM remains the most reliable classifier. Overall, fusion features demonstrated superior performance compared to other histogram and image features, as indicated by the overall macro-average AUC findings. Further evaluations of the proposed multiclass classification model H2FCD are presented in the next section.

4.4.1.4 ROC-AUC Graph for SVM

Figures 4.6, 4.7, and 4.8 present the ROC-AUC curves of the SVM classifier using three different feature types: image, histogram, and fusion features. ROC-AUC was used as a visual evaluation tool to assess the classification performance by illustrating the trade-off between the true positive rate and false positive rate. A curve closer to the top-left corner and an AUC value closer to 1 indicate better classification performance.

In Figure 4.6, the ROC curves obtained using image features demonstrate strong classification performance across all three classes (high, moderate, and low risk), with AUC values approaching 1.0. This indicates a high level of discriminative capability of the model. However, the slight deviation of the curves from the top-left corner suggests minor misclassifications, indicating that the model did not achieve perfect separation between classes when relying solely on image features.

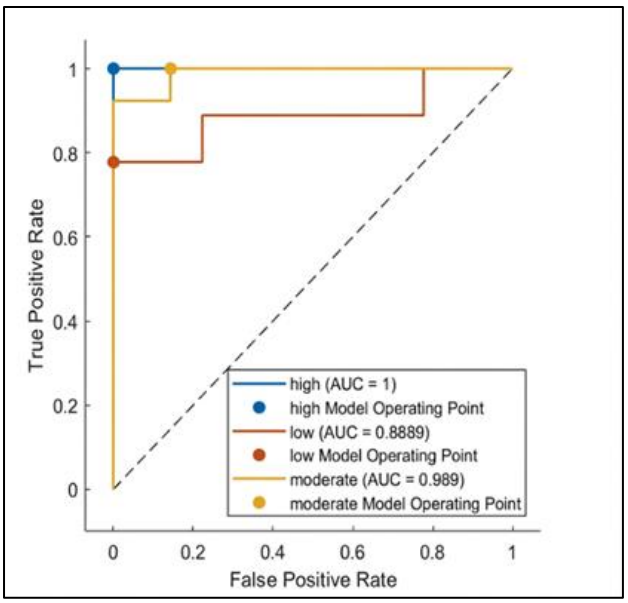


Figure 4.6 ROC-AUC Graph for SVM Model Using Image Features

In Figure 4.7, the ROC curves obtained using histogram features demonstrate reasonably strong classification performance; however, the AUC values were slightly lower compared to those achieved with image features, particularly for certain classes. This reduction in performance suggests that histogram features alone provide limited discriminative capability, as they may not sufficiently capture the structural and spatial characteristics required for effective classification.

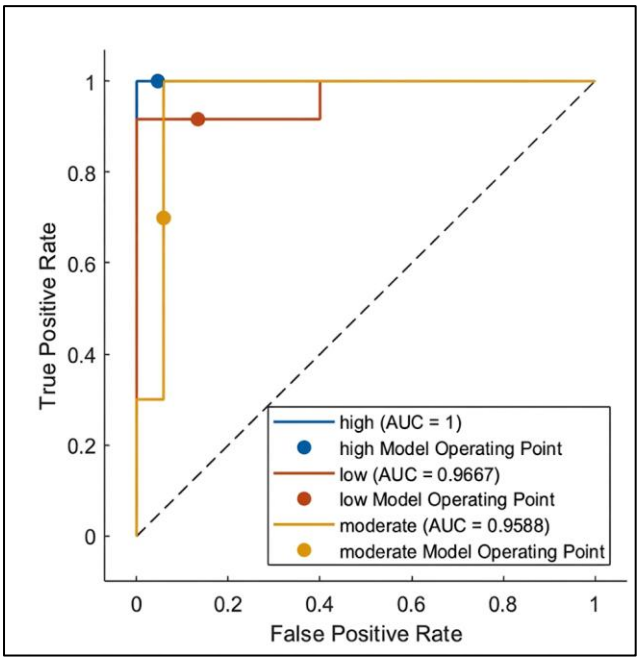


Figure 4.7 ROC-AUC Graph for SVM Model Using Histogram Features

In contrast, Figure 4.8 illustrates that using fusion features yields optimal classification performance, with all ROC curves converging to the top-left corner and achieving an AUC of 1.0 across all classes. This result indicates that the model attained maximal discriminative capability, reflecting an ideal balance between sensitivity and specificity. The absence of deviation from the optimal point further suggests that the classifier achieves complete separation among classes with no observable misclassification.

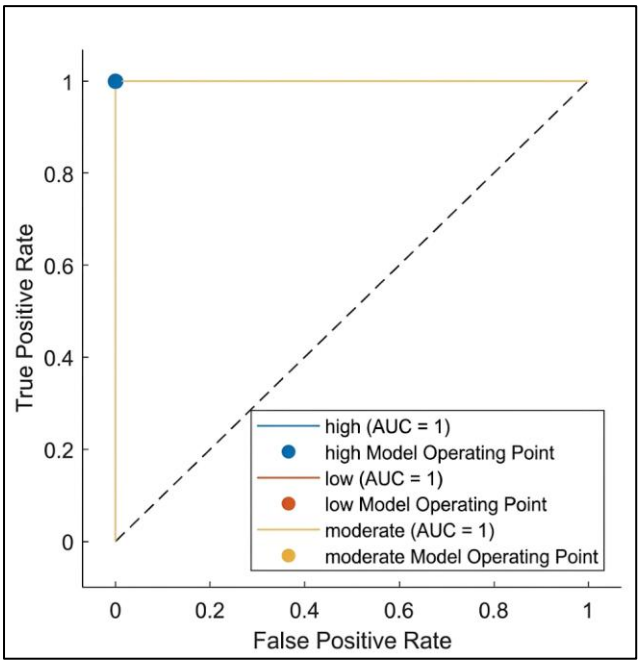


Figure 4.8 ROC-AUC Graph for SVM Model Using Fusion Features

These findings are consistent with previously reported results on accuracy, precision, recall, and F1-score, in which integrating fusion features with the SVM classifier yielded the highest and most stable performance across all evaluation metrics. This consistency reinforces the effectiveness of the proposed approach, demonstrating that integrating complementary feature representations significantly enhances the model's discriminative capability for accurately distinguishing among levels of dysgraphia severity.

4.4.1.5 Result and Analysis of H2FCD Model with Other Model

This section presents a comparative analysis of the proposed H2FCD model against existing fusion-based approaches. Among all the findings, fusion features have achieved exceptional accuracy and precision, as evidenced by ROC curves and macro-average AUC values. By utilizing data obtained from handwriting images and histogram images, accuracy can be enhanced slightly from 4.37% to 7.28%. Extracting the features of the writing from a scaled image without altering the actual scale size can yield more precise information for extraction. Histogram images enhance the transmission of information to machine learning models by accurately assessing the severity level of dysgraphia. Deep feature extraction can be utilized to supply the essential features in the classifier for the subsequent work of a classification model. The

precision and recall metrics may effectively showcase the model's accurate classification of dysgraphia writings into three risk classes: high-risk, moderate-risk, and low-risk.

By narrowing the scope of fusion features, the SVM machine learning model makes major contributions in areas where it can achieve high, consistent performance. This is evident in the SVM's accuracy, which was consistent compared to another model. The ROC-AUC graph demonstrates the SVM model's stability and efficiency, with consistent AUC values and a direct macro-average AUC, which was also high, indicating that the SVM was successful in diagnosing the severity of dysgraphia using handwriting images.

Table 4.8
Comparison of Existing Dysgraphia Detection Methods and the Proposed H2FCD Model Based on Feature Extraction, Classification, and Performance

Author (year)	**Handwriting Type**	**Domain**	**Output**	**Feature Extraction**	**Classifier**	**Performance**
Kunhoth, Al Maadeed, et al. (2023)	Offline (transformed from online)	Word, pseudoword, difficult word, sentence - Slovak	Binary	Deep features (DenseNet201) + feature fusion	SVM, AdaBoost, Random Forest	Accuracy: 97.3%
Brescia et al. (2026)	Online	Letter, word, pseudoword, difficult word, sentence - Slovak	Binary	Kinematic, temporal, robotic features	TabNet	Accuracy: 91.7%, Precision: 95.2%
Kunhoth et al., 2024)	Multimodal (online + offline)	Word, pseudoword, difficult word - Slovak	Binary	Deep features (EfficientNet-B7) + spatial & temporal features	SVM, XGBoost	Accuracy: 88.8%, Recall: 90.0%
Sharma et al. (2025)	Offline	English	Binary	Deep features (VGG16) + handcrafted features	Random Forest, SVM, Logistic Regression	Accuracy: 88.9%
Al Abadleh & Al-Shqeerat (2025)	Offline	English and Slovak	Binary	EfficientNet-B7 + SWIN Transformer (feature fusion)	Random Forest	Dysgraphia: 98.6%
Drotár & Dobeš (2020)	Online	Slovak	Binary	Online handwriting features	AdaBoost	Accuracy: 79.5%

Author (year)	Handwriting Type	Domain	Output	Feature Extraction	Classifier	Performance
Kunhoth, Maadeed, et al. (2023)	Online	Slovak	Binary	In-air feature extraction	KNN, SVM, RF, AdaBoost	Accuracy: 80.8%
Devi & Kavya (2023)	Offline	English	Binary	DCT + Deep Transfer Learning	Deep Learning	Accuracy: 99.75%
Ghouse et al. (2022)	Offline	English	Binary	CNN + regularization	CNN	Accuracy: 98.2%
Devillaine et al. (2021)	Online	Drawing	Binary	Manual features	RF, MLP, Extra Trees	Accuracy: 73%
Proposed Model	Offline	Malay language (Sentence)	Multiclass (High, Moderate, Low risk)	Deep features (ResNet-18) + histogram features	SVM	Accuracy: 99.26%

Table 4.8 presents a comparative performance analysis of existing studies and the proposed H2FCD model across the handwriting data domain, feature extraction, classifiers, and performance. Most existing studies rely on single-modality data, either offline or online. Offline approaches, such as Ghouse et al. 2022, and Kunhoth et al. (2024), utilize deep learning models (DenseNet201 and CNN), achieving high accuracies of 97.3% and 98.2%, respectively. While effective in capturing structural handwriting patterns, these methods depend on single-feature representations, limiting their ability to capture the full variability of dysgraphia.

In contrast, online approaches, such as Brescia et al. (2026), Drotár & Dobeš (2020), incorporate kinematic and temporal features but achieve lower performance (79.5%–91.7%) and require specialized hardware, thereby reducing their practical applicability. Multimodal approaches (Kunhoth et al., 2024) combine offline and online features, yet achieve only moderate performance (88.8%), indicating that feature fusion alone is insufficient without proper feature complementarity. Similarly, hybrid methods that combine deep and handcrafted features (Sharma et al., 2025) achieve 88.9%, suggesting that simple feature concatenation does not significantly improve performance.

The highest accuracy, reported by Devi & Kavya (2023) at 99.75%, was achieved using DCT with deep transfer learning. However, it is limited to binary classification, restricting its use for severity analysis. In contrast, the proposed H2FCD model employs an effective hybrid feature fusion strategy, integrating a residual network with histogram-based features to capture both global structural patterns and local pixel intensity distributions. This results in improved multiclass classification performance.

Therefore, the proposed H2FCD model has been successfully developed. Based on the findings in Section 4.4.1, HIRRL was used to enhance performance by developing a fusion-feature-based model that combined HIRRL feature vectors with a generated histogram image (Histo-HIRRL). The SVM model has demonstrated high performance, supported by strong ROC-AUC results, and has been selected for integration into the proposed model. The proposed model achieved 99.26% accuracy, comparable to the best-performing methods, while offering a key advantage in multiclass classification of dysgraphia severity (high, moderate, low).

Therefore, the proposed H2FCD model comprises a residual network for feature extraction and an SVM classifier. The experimental process had been completed, and a detailed analysis is provided in Section 4.4. This analysis examines all potential feature vectors and uses a machine learning model to categorize dysgraphia severity.

4.5 Phase 4: Performance of Proposed Multiclass Classification Model H2FCD

This analysis presents the validation findings as outlined in Objective 4 of Section 1.5. This section presents the performance of the proposed multiclass classification model H2FCD. The analysis is divided into two parts: a performance comparison with other studies, and a comparison of the classification outputs automatically generated by the H2FCD model and those of experts. The analysis is detailed in the following section.

4.5.1 Result and Analysis of Testing Classification Performance using H2FCD Model based on Metrics Evaluation

In this section, the correlation analysis is presented by demonstrating the classification results obtained using the proposed H2FCD model and experts'

evaluations. This section presents the findings of this study, outlining the proposed H2FCD performance. Model verification involves evaluating the model using 36 new input images to ensure that the classification task was accomplished successfully. Figure 4.9 presents the confusion matrix of the SVM classifier's performance using a fused feature vector. It shows the number of true predictions and misclassifications for each class. The diagonal position of the number represents the number of correctly classified data for each class. At the same time, the off-diagonal position represents the number of misclassifications, where the data of the class is incorrectly predicted as another class. According to the numbers presented in the confusion matrix, an accuracy of 88.89% was achieved, and 32 images were correctly predicted out of a total of 36 images.

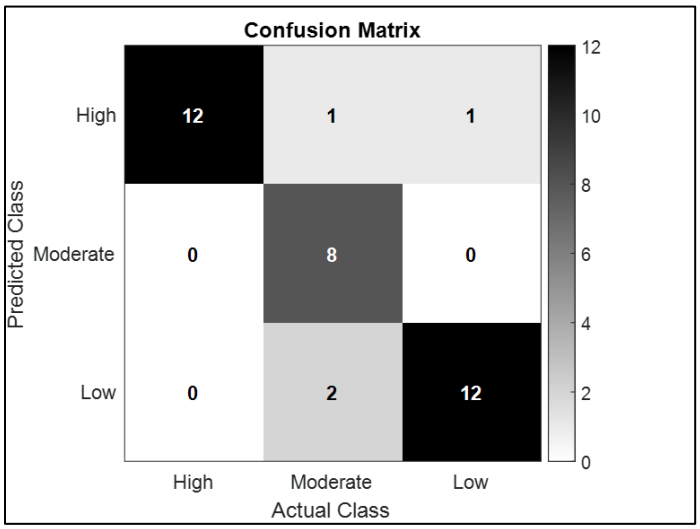


Figure 4.9 Confusion Matrix of SVM Classifier with Fusion Feature Vector

As tabulated in Table 4.9, macro-average precision was achieved at 90.47%, and macro-average recall was obtained at 88.35%. Therefore, the macro-average F1-score successfully achieved the value of 88.47%.

Table 4.9
Precision, Recall and F1-Score for Each Class and Macro-Average Measurement

Class	**High-Risk**	**Moderate-Risk**	**Low-Risk**	**Macro-average**
Precision	85.71	100	85.71	90.47
Recall	100	72.73	92.31	88.35
F1-score	92.31	84.21	88.89	88.47

Based on the multiclass MCC, the classifier achieved 0.839, indicating a strong overall correlation between the predicted and actual classes. Compared to the results reported in (Devi & Kavya, 2023), which are based on simpler binary classification tasks, this study addresses a more challenging multiclass problem. Therefore, the obtained MCC demonstrates robust generalization despite increased complexity and reflects that the classifier performs consistently well across all three severity classes (high, moderate, low), with minimal misclassification. Because MCC incorporates all elements of the confusion matrix and remains reliable even when class distributions differ, this high value demonstrates robust and well-balanced classification performance for the handwriting severity classification using the H2FCD model.

Figure 4.10 shows the ROC curve graph of the SVM classifier with the fusion feature vector. Based on the figure, the graphical curve represents the performance of the dysgraphia severity classification by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at several threshold settings. The findings provided three classes of curves, and all curves were close to the top-left corner of the plot, which shows that the performance of the classifier model is superior in classifying dysgraphia handwriting using fused feature vectors.

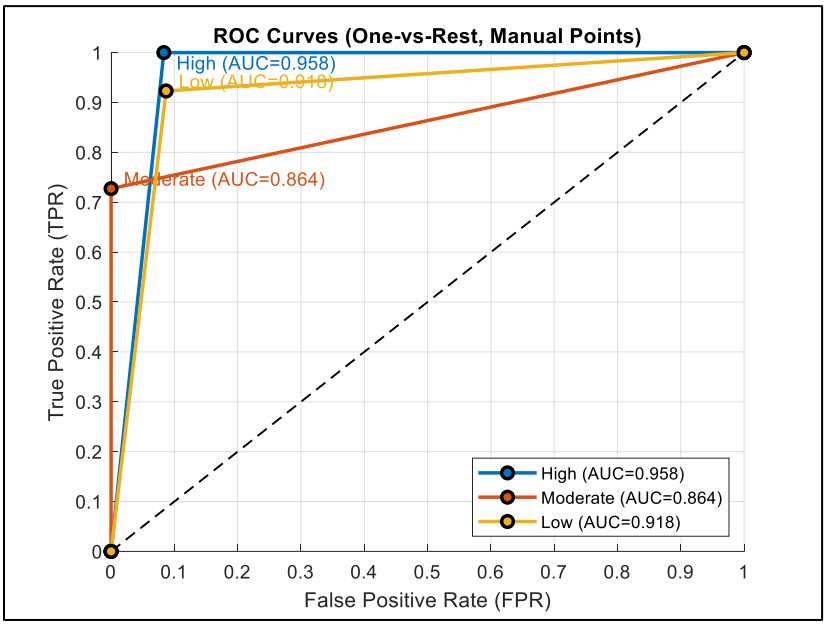
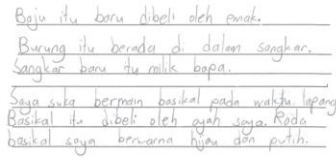
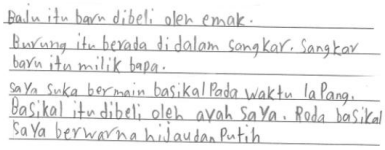
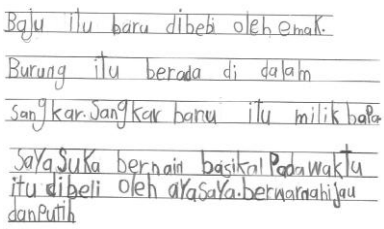
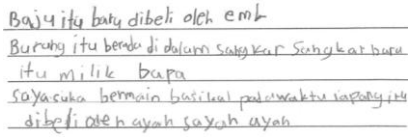


Figure 4.10 Testing ROC-AUC Curve for SVM Classifier with Fusion Feature

Table 4.10 presents several examples of misclassified handwriting samples, highlighting cases where the predicted class differs from the actual class. Most misclassifications occur between adjacent classes, particularly between moderate and low or moderate and high risk. This is likely due to the similarity in handwriting characteristics across these categories, such as stroke inconsistency and spacing irregularities.

Table 4.10
Example of Misclassified Handwriting Images

Image	Actual Class	Predicted Class
	Low risk	High risk
	Moderate risk	Low risk

	Moderate risk	High Risk
	Moderate risk	Low risk

These results indicate that while the proposed model achieves high overall performance, certain borderline cases remain challenging to distinguish, especially when handwriting features overlap between severity levels. Accordingly, the following section presents an expert evaluation of the model's classification outputs. The predicted results from the proposed H2FCD model are compared with expert assessments, and the level of agreement is quantified using correlation coefficients. This analysis serves to validate the consistency and reliability of the model in relation to human expert interpretation.

4.5.2 Result and Analysis of H2FCD Model Based on Expert Validation

Based on the H2FCD model's classification outputs, the experts evaluated the results, and the model's validity was supported by measuring the correlation coefficient, as demonstrated in Section 3.6.3 for each expert participating in this evaluation session. This analysis demonstrated the association between the classification outputs of the proposed model and those of the experts.

Table 4.10 shows the image category results by severity level, based on the proposed H2FCD model and the eight experts' evaluations. The column "CO_p" represents the predicted class from the model, while "CO_{e1}–CO_{e8}" indicates whether each expert categorizes the same classes (✓) or different classes (×) with the model's classification. Overall, a high number of agreement marks was observed across most samples, particularly for high- and low-risk classes, indicating strong consistency between the model predictions and expert judgments. However, some images were classified differently, especially within the moderate-risk category, reflecting the inherent difficulty in distinguishing borderline handwriting characteristics. These

results demonstrate that the proposed model generally achieves reliable agreement with expert evaluations while highlighting areas where classification remains challenging. The validity of the model is further supported by measuring the correlation coefficient, as demonstrated in Section 3.6.3 for each expert participating in this evaluation session. This validation demonstrated the association between the classification outputs of the proposed model and those of the experts.

Table 4.11
Results Image Categories Based on Severity Levels

		Testing Model	**Expert evaluation**							
Data Input	**True Label**	**CO_p**	**CO_{e1}**	**CO_{e2}**	**CO_{e3}**	**CO_{e4}**	**CO_{e5}**	**CO_{e6}**	**CO_{e7}**	**CO_{e8}**
Image 1	High-Risk	High-Risk	×	×	×	×	✓	×	×	✓
Image 2	High-Risk	High-Risk	×	✓	✓	×	×	×	×	✓
Image 3	High-Risk	High-Risk	✓	✓	✓	✓	×	✓	✓	✓
Image 4	High-Risk	High-Risk	×	✓	×	✓	×	✓	×	✓
Image 5	High-Risk	High-Risk	×	✓	×	×	×	×	×	✓
Image 6	High-Risk	High-Risk	✓	✓	✓	✓	✓	✓	×	✓
Image 7	High-Risk	High-Risk	✓	✓	✓	✓	✓	✓	×	✓
Image 8	High-Risk	High-Risk	✓	✓	✓	✓	✓	✓	×	✓
Image 9	High-Risk	High-Risk	✓	✓	✓	✓	✓	✓	×	✓
Image 10	High-Risk	High-Risk	✓	×	✓	✓	×	✓	×	✓
Image 11	Low-Risk	Low-Risk	✓	✓	✓	✓	✓	✓	✓	✓
Image 12	Low-Risk	High-Risk	×	×	×	×	×	×	×	×
Image 13	Low-Risk	Low-Risk	✓	✓	✓	✓	✓	✓	✓	✓
Image 14	Low-Risk	Low-Risk	✓	✓	✓	✓	✓	✓	✓	✓
Image 15	Low-Risk	Low-Risk	✓	✓	✓	✓	✓	✓	✓	✓
Image 16	Low-Risk	Low-Risk	✓	✓	×	✓	✓	✓	✓	✓
Image 17	Low-Risk	Low-Risk	✓	✓	×	✓	✓	✓	✓	×

		Testing Model	Expert evaluation							
Data Input	**True Label**	**CO_p**	**CO_{e1}**	**CO_{e2}**	**CO_{e3}**	**CO_{e4}**	**CO_{e5}**	**CO_{e6}**	**CO_{e7}**	**CO_{e8}**
Image 18	Low-Risk	Low-Risk	✓	✓	✓	×	✓	✓	✓	✓
Image 19	Low-Risk	Low-Risk	✓	×	×	✓	✓	✓	✓	×
Image 20	Low-Risk	Low-Risk	✓	✓	✓	✓	✓	✓	✓	×
Image 21	Moderate-Risk	Moderate-Risk	✓	×	×	×	×	×	✓	×
Image 22	Moderate-Risk	Low-Risk	×	×	×	×	×	×	×	×
Image 23	Moderate-Risk	Moderate-Risk	✓	✓	×	✓	×	×	×	×
Image 24	Moderate-Risk	Moderate-Risk	✓	×	✓	×	✓	✓	✓	✓
Image 25	Moderate-Risk	Low-Risk	✓	✓	✓	×	×	✓	✓	×
Image 26	Moderate-Risk	Moderate-Risk	×	✓	✓	✓	✓	✓	×	✓
Image 27	Moderate-Risk	High-Risk	×	×	×	×	×	✓	×	×
Image 28	Moderate-Risk	Moderate-Risk	×	×	✓	✓	✓	✓	✓	✓
Image 29	Moderate-Risk	Moderate-Risk	×	×	✓	✓	×	✓	×	×
Image 30	Moderate-Risk	Moderate-Risk	✓	✓	✓	✓	×	×	✓	×
		Number of images with equal classification output	20	20	19	20	17	22	15	19

CO_p: Classification Output from Model Testing, COe1:Classification Output from expert evaluation

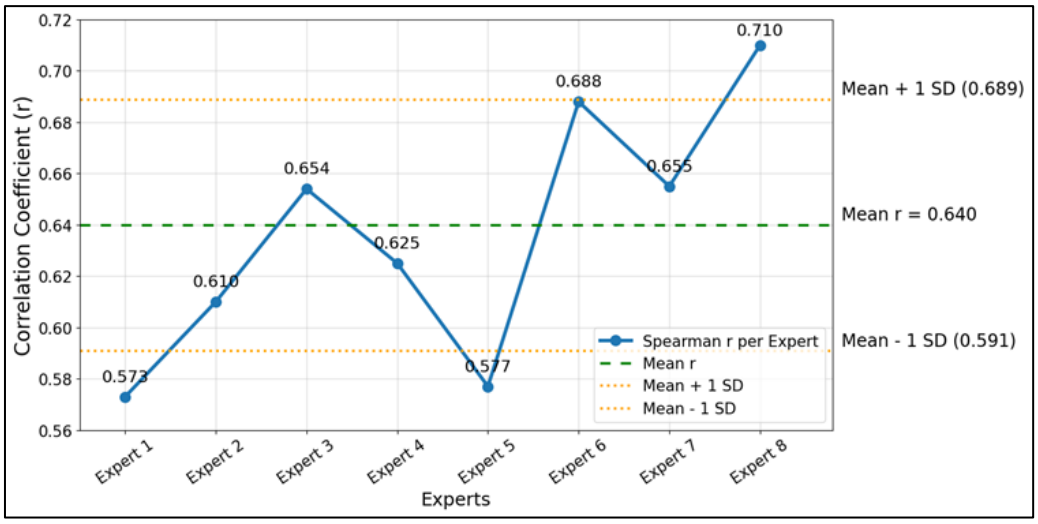


Figure 4.11 Correlation Coefficient of Classification Output Class Between H2FCD Model and Expert Evaluation

Figure 4.11 illustrates the Spearman correlation coefficients (r) between the proposed H2FCD model's classification outputs and the evaluations from eight experts. The correlation values ranged from 0.573 to 0.710, indicating a moderate-to-strong positive agreement between the model predictions and expert judgments. The mean correlation coefficient is $r = 0.640$, indicating an overall consistent alignment between automated classification and human evaluation.

The variability across experts was further analyzed using the mean ± standard deviation (SD), where the upper bound (mean + 1 SD = 0.689) and lower bound (mean − 1 SD = 0.591) provide insight into the dispersion of agreement levels. Experts 6 and 8 demonstrate higher agreement, exceeding the upper threshold, while Experts 1 and 5 judgement fell below the lower bound, indicating relatively lower agreement. This variation is expected due to subjective differences in human interpretation of handwriting characteristics. Nevertheless, all correlation values remained within an acceptable range, supporting the reliability of the proposed model.

The expert validation was conducted using a subset of 30 images selected from the unseen test dataset (Gómez-Carmona et al., 2024; X. Wu et al., 2022). This process evaluation involves a detailed manual assessment, which was time-consuming and cognitively demanding. A sufficient number of samples allowed experts to maintain focus and consistency, and the number was adequate for reliable statistical analysis, which can otherwise introduce bias and reduce annotation quality. The selected subset was designed to be representative of all dysgraphia severity classes, ensuring a balanced and meaningful validation process.

This approach is consistent with established practices in human-in-the-loop evaluation, where a representative subset of data is commonly employed to ensure high-quality expert annotations and reliable agreement analysis. Therefore, 30 samples is considered sufficient to provide robust and credible validation of the proposed classification model.

4.6 Summary

This chapter presents and discusses the study's results in accordance with the intended phases. The study yielded promising results in assessing the severity of dysgraphia by utilizing offline handwriting images, which were processed using image processing techniques and a deep network structure to extract handwriting features.

In phase 1, the dataset was developed by collecting handwritten samples. The study's findings included a reliability analysis for all data sets obtained from Cohen's weighted kappa and Spearman correlation coefficient analysis. The analysis showed high agreement between two assessors in evaluating the data samples for categorizing the severity levels of dysgraphia based on handwritten images. Phase 1 concluded with a total of 516 offline handwritten cropped images. The main step in phase 2 was image pre-processing, which includes the process of ruling line removal which was an essential step to extract the actual feature information from the offline handwritten images. It demonstrated promising SSIM results for a new RGBB-based ruling line removal scheme. Finally, Phase 3 involved three major analyses: multiclass classification performance for HIRL and HIRRL input images; multiclass classification performance for fusion feature-based HIRRL and Histo-HIRRL; and validating the proposed model by measuring testing performance and conducting correlation analysis between expert classification output and the proposed H2FCD model. Overall, the results contribute to the field, as this proposed model has the potential to be developed into a real-world application, thereby supporting SDG 4 of quality education through faster detection assistance.

CHAPTER 5
CONCLUSION

5.1 Conclusion

This research presents intelligence-based classification model for classifying dysgraphia severity based on handwriting images. The study extracted handwriting images and histogram images and utilized a Convolutional Neural Network (CNN)-based feature extraction approach within a classification model. The proposed approach achieved high accuracy performance in multiclass classification, which has three classes of severity: low-risk, moderate-risk, and high-risk. The validation of the proposed Handwriting and Histogram Features for Classifying Dysgraphia severity (H2FCD) model was conducted by comparing the classification output results with those of the experts.

Dysgraphia handwritten datasets involved collecting handwriting data from Persatuan Dyslexia Malaysia (PDM) and Ministry of Education (MOE) schools, then digitizing the handwritten images. The data were assessed manually by two evaluators, who identified and classified the data into severity levels. The dataset of MLHSD was successfully developed following reliability analysis using weighted Cohen's Kappa and Spearman correlation coefficients, which showed strong agreement and correlation between the two assessors. The Cronbach's alpha also indicates high reliability, with a value close to 1. The dataset achieved high reliability, with an agreement of 0.961 (Cohen's Kappa) and a correlation of 0.974 (Spearman's rank correlation), supported by a high Cronbach's alpha. The analysis confirms that the dataset is highly reliable and can be used with confidence in subsequent research (Section 4.2).

The research presented a new scheme for a Red-Green-Blue-Black (RGBB)-based method, which is a pre-processing method for removing ruling lines in dysgraphia handwritten images. The scheme achieved a good Structural Similarity Index Measurement (SSIM) result with an average SSIM of 0.80 (Section 4.3). It also improved classification performance for Handwriting Image with Removed Ruling Line (HIRRL) compared to Handwriting Image with Ruling Line (HIRL) images (Section 4.3.2), increasing classification accuracy from approximately 91% to above

94%, demonstrating its effectiveness in preserving structural information while removing noise.

In Phase 3, the proposed H2FCD model was successfully developed for multiclass classification of dysgraphia severity based on handwritten samples, utilizing fused feature vectors. This study demonstrates that the model achieves good performance, as evidenced by its accuracy, precision, recall, and F1-score, as well as an ROC-AUC graph. This study reveals that the Support Vector Machine (SVM) classifier model achieved the highest accuracy and has a very good AUC graph. This means that a dysgraphia severity classification model is more suitable when using an SVM as a classifier, particularly with fused features from dual input images that consist of handwritten images and histogram images, as explained in Section 4.4.

In the final phase, this model was successfully tested with unseen input images and validated by experts. The multiclass MCC value of 0.839 indicates a strong overall correlation between predicted and actual classes, indicating that the classifier performs consistently well across all three severity levels with minimal misclassification. The correlation coefficient analysis has presented a moderately strong correlation with a ρ value above 0.64, across the eight experts in this study (Section 4.5). The study further demonstrates that dual-input CNN feature fusion, combining handwritten images and histogram images, significantly enhances severity discrimination, improving classification accuracy by approximately 4.37% to 7.28%. This highlights the importance of integrating complementary feature representations to improve model performance.

Overall, this study contributes to the advancement of machine learning and image processing technologies in the field of handwriting analysis. In addition, this study is also compatible with various disability learning system models. The evaluation results and study findings demonstrate that the research objectives were successfully met. Based on these findings, further research in the same subject can be conducted to gather additional insights and enhance the performance classification model.

5.2 Research Contribution

This study aims to classify the severity levels of dysgraphia handwritten using a CNN-based feature extraction from dual-input images and a machine learning classifier to achieve optimal performance in classifying the severity levels. Generally,

the classification of dysgraphia severity based on handwriting involves several phases, depending on the purpose and objectives of the research. This study aims to classify dysgraphia severity into low-, moderate-, or high-risk categories. It was demonstrated that automatic classification can be accomplished using machine learning based on significant features fused from multiple input images. This work used CNN-based feature extraction for dual-input images, concatenating all features for the classification task.

The MLHSD dataset was created by initially collecting data through several processes, including preparing a writing form, receiving agreement from participants, and collecting handwriting data on paper as explained in Section 3.3. PDM's assessor then evaluated the data to categorize the handwriting into three categories: low-risk, moderate-risk, and high-risk. All handwriting forms on paper were then converted into digital image format by scanning the writing form and then organizing then into specific folders, saving the handwriting images according to the severity levels. The first objective was accomplished by measuring the correlation between two PDM assessors while evaluating the severity of dysgraphia in handwriting.

Image pre-processing begins after the dataset development is accomplished. This procedure involves the removal of the ruling line utilizing the new RGBB-based algorithm, which implements several image processing techniques as mentioned in Section 3.4. The second objective is achieved by measuring the SSIM for the input of the original image and the input image with removed ruling line. In addition, this technique for ruling line removal generated HIRRL image data, which was then used in the third phase, the classification phase. Consequently, two types of images were yielded from this phase: HIRL and HIRRL. These two images were then utilized to evaluate the impact of adapted classification performance, and the findings demonstrated that HIRRL can improve accuracy, precision, and recall values. Based on these results, HIRRL is shown to provide superior classification performance.

The third phase in Section 3.5 encompasses all the procedures required to fulfill the third objective, which is to classify the severity of dysgraphia using handwritten images. This approach consists of two major steps: feature extraction and classification. Prior to feature extraction, input images were provided, which include enhancing and producing histogram images. Histogram images was created for each piece of written data, and the images were then normalized to a standard size before being used in the classification process. Feature extraction is then performed using an automated CNN

based dual input fusion approach, which combines two input modalities: handwriting images and handwriting histograms. In this study, a composite residual network is introduced to extract features from the histogram images, enabling deeper representation learning and improved capture of pixel intensity distribution characteristics. This network complements the feature extraction from the original handwritten images, allowing the model to learn both global structural patterns and local intensity variations. Fusing features from both inputs enhances the model's discriminative capability. The inclusion of the composite residual network for histogram feature extraction plays a significant role in improving classification performance, as it provides richer, more complementary information than a single feature extraction source. Accordingly, this study successfully developed the H2FCD model using an SVM classifier based on a fused feature vector derived from dual-input images of handwriting and histograms.

The fourth objective is fulfilled by completing the validation procedure, described in Section 3.6. Performance indicators include accuracy, precision, recall, and F1-score, Matthew correlation (MCC) and ROC-AUC values. By demonstrating performance using MCC in multiclass classification, the model achieved meaningful classification of dysgraphia severity. The result classification demonstrates that feature fusion can produce a significant and persistent characteristic adequate to categorize the works in question as low-risk, moderate-risk, or high-risk.

This study successfully developed a new Malay Language Dysgraphia Handwriting Dataset (MLHSD) and employed the dataset into classification model of H2FCD. Prior to classifying process, MLHSD dataset undergoes pre-processing by removing ruling line using a new scheme of RGBB-based method. Then, a classification model was developed by leveraging the fusion features from dual-input by implementing CNN-based deep feature extraction. The performance evaluation is completed in the final phase using metric-based measurement, including model testing and validation of the classification output against expert evaluations.

5.3 Implications for Research

Research in this field is expanding rapidly, and the progress is encouraging and accelerating. This study's findings provide a wealth of empirical evidence and

theoretical support that may be applied to other studies. Accordingly, the implications of this study are presented in this section as follows:

i) The process of dataset development of dysgraphia children's handwriting was successfully executed through writing acquisition activities among school children and children from PDM. The dataset development process included analyses of inter-rater agreement between the two assessors, as well as correlation analyses examining the relationships between age and dysgraphia category and between gender and dysgraphia category.
ii) Digitized data images were collected and categorized based on dysgraphia severity level. The dataset has been published and used by other researchers (Ramlan, Isa, Ismail, et al., 2024)
iii) Handwriting images are used to classify the severity of dysgraphia since they are the most easily accessible data produced by schoolchildren when completing their schoolwork.
iv) A ruling line removal algorithm was explicitly created for dysgraphia handwriting images, by applying several image processing techniques.
v) The training model was obtained using a CNN-based fused feature fusion
vi) Research results, data analysis, procedure/process, algorithm and research design are used as a reference to develop such classification system in this field of research.

This study has impacted the development of science and technology knowledge, especially for the education system. The data collection, which consists of written samples of dysgraphia children, has contributed to the development of artificial intelligent model that can be implemented to help in detecting dysgraphia symptoms. Dataset development procedure also makes an exceptional contribution in developing data that involves physical data and which requires several appropriate processes and methods to build datasets as inputs for deep learning models or machine learning applications.

Furthermore, this study has a significant impact on the evolution in image processing techniques such as an innovative ruling line removal algorithm that has been successfully created specifically for dysgraphia handwriting images.

Besides, all of the experiments conducted in this study have shown positive impact on the development of the classification model. Several aspects of this study has contributed to the understanding of machine learning evolution. Furthermore, the results of the findings and performance values for the classification of dysgraphia severity were obtained using handwriting images with and without ruling lines. Moreover, the use of a residual network for feature extraction has also been investigated, and the performance of the residual and sequential network layers is compared (Ramlan et al., 2023). These results contribute to improved classification model performance specifically in handwriting analysis and recognition.

Aside from the classification model, features extraction is essential in determining dysgraphia severity level. In this study, new fused features were concatenated and designed to be employed in a classification model. The proposed CNN-based fused feature extraction was performed, and significant features were effectively obtained, allowing a classifier model to deliver accurate and precise classification results. The classification performance is presented in the confusion matrix evaluation and the AUC graph.

The results of the classification performance were categorized into three classes representing the severity of dysgraphia based on handwriting images features information. This multiclass classification has an impact on the knowledge domain of machine learning. This is because most successful models are classified into only two categories: yes or no. The categorization findings were divided into three classes based on the severity of dysgraphia as determined by handwriting image feature information. This multiclass classification affects the knowledge area of machine learning. This is because most successful models are classified into only two categories (binary classification)

The methodology utilized has an impact on system developers, particularly for automation systems that require detection and classification. Each stage/phase of the technique demonstrates the steps required in the development of a system similar to this. As a result, system developers can use this approach as a reference, as the methodology of this study is based on the findings of previous research literature evaluations.

This approach can serve as a resource for other academics who are interested in data collection, image preprocessing, feature extraction, multiclassification, and broader computer vision application. Moreover, the applied research design in this study

provides becomes a reference framework which can be utilized to construct new systems that require image classification domain.

Besides system developers, this study also has implications also for practitioners, including psychiatrists, schoolteachers, medical specialists, PDM officers, and other relevant NGOs. This is because the dyslexia-dysgraphia assessment is assisted by an automated system as proposed in this study. This assessment process shorten time required by the current observation and scoring system. This revolution further expanded the dysgraphia assessment method, which has become the primary aim for children educational well-being.

This study also impacts the education system in Malaysia, especially in facilitating earlier identification and intervention for children with dysgraphia based on detected severity levels. This is in line with SDG 4, which stated that the quality of education needs to be emphasized to ensure the success of these children in reaching their full potential to becoming human capital and national assets.

As this study proposes a handwriting image analysis approach for detecting dysgraphia symptoms using machine learning, the developed model is capable of classifying severity levels into three categories. Therefore, it can be used in any other fields that require handwriting characteristics analysis and information to identify severity levels. For instance, in terms of social health, this model could serve as a resource and guidance for various handwriting difficulties associated with other specific learning disorders and handwriting impairment issues.

5.4 Limitations and Future Work Recommendations

This study has been successfully completed while acknowledging several limitations which can be considered to provide opportunities for further investigations related to this research domain. At the beginning of the research design, the data collection process was carried out involving handwriting samples of school children using the Malay script commonly used among Malaysians. Therefore, this study is limited to Malay language handwriting for detecting the characteristics of dysgraphia or classifying dysgraphia severity levels. In the future, this work recommended to be extended to other languages commonly used in Malaysia.

In addition, this study relied on a small amount of data due to constraints in collecting data from children with special needs. This may introduce class imbalance

issues, which can affect model generalization. Although the model achieved strong performance, further work should incorporate data balancing techniques and data augmentation methods to improve robustness. In addition, approaches such as Generative Adversarial Networks (GANs) can be explored to synthetically generate dysgraphia handwriting samples, thereby increasing the diversity and quantity of training data. Therefore, it is suggested that dataset size should be increased in future studies to develop more accurate models with improved capability to classify a broader range of characteristics and diverse writing styles. Alternatively, further research should consider using a Generative Adversarial Network (GAN) to synthetically generate dysgraphia handwriting samples to increase the number of inputs required for training a supervised learning model.

Furthermore, the line removal method is designed explicitly for dysgraphia writings, which vary in severity levels. As described in Section 3.4.1, the morphological distribution of dilation and erosion sizes is based on severity level. This method can be enhanced further by developing a more generalized approach capable of identifying and removing ruling lines without requiring differentiation among dysgraphia severity classes.

Furthermore, the current model was developed using default hyperparameter settings, which may not fully optimize model performance. Future studies should consider systematic hyperparameter tuning and optimization techniques to further enhance classification accuracy and model efficiency. In addition to accuracy, the evaluation of performance using Matthews Correlation Coefficient (MCC) could be further improved, where a value exceeding 0.84 would indicate stronger and more reliable classification, particularly in imbalanced datasets.

Moreover, dysgraphia severity is classified into three categories: high risk, moderate risk, and low risk using CNN and several machine learning classifiers including SVM, RF, KNN, and BT. For further exploration, alternative machine learning models and more advanced CNN architectures can be explored to further enhance classification performance.

Although the proposed H2FCD model demonstrates high classification performance, it is important to emphasize that the system is designed as an assistive tool for experts, rather than a replacement for professional diagnosis. The model provides guidance in identifying dysgraphia severity level; however, final decisions should remain under the supervision of educators, clinicians, or specialists. Future work may

focus on integrating the system into decision-support frameworks that complement existing diagnostic procedures.

Finally, this study provided a strategy for classifying the severity levels of dysgraphia handwriting based on the students' handwriting input. This research could be integrated with other models that focus on other learning disabilities, such as reading and mathematical difficulties. Such integration could be widely used to identify an individual's learning difficulties and to provide a comprehensive system for diagnosing learning disabilities.

In conclusion, while the system exhibits certain limitations, the developed system has successfully fulfilled the objectives of the study and established a strong foundation to support user requirements in classifying dysgraphia severity levels.

REFERENCES

Adamski, M., Sarnacki, K., & Saeed, K. (2021). Binary handwriting image enhancement by directional field-guided morphology. *Information Sciences*, *551*, 168–183. https://doi.org/10.1016/j.ins.2020.11.019

Ahlawat, S., & Choudhary, A. (2020). Hybrid CNN-SVM Classifier for Handwritten Digit Recognition. *Procedia Computer Science*, *167*(2019), 2554–2560. https://doi.org/10.1016/j.procs.2020.03.309

Ahuja, N. J. (2020). A novel approach of Handwriting Analysis for Dysgraphia Type Diagnosis. *International Journal of Advanced Science and Technology Vol. 29, No. 3, (2020), Pp. 11812 – 11828 A*, *29*(3), 11812–11828.

Al Abadleh, A. H., & Al-Shqeerat, K. H. A. (2025). Hybrid Feature Extraction-based Learning Disabilities Identification Model. *Journal of Disability Research*, *4*(6). https://doi.org/10.57197/jdr-2025-0612

Aldehim, G., Rashid, M., Alluhaidan, A. S., Sakri, S., & Basheer, S. (2024). Deep Learning for Dyslexia Detection: A Comprehensive CNN Approach with Handwriting Analysis and Benchmark Comparisons. *Journal of Disability Research*, *3*(2). https://doi.org/10.57197/jdr-2024-0010

Alevizos, V., Edralin, S., Simasiku, A., Malliarou, D., Messinis, A., Papakostas, G., Xu, C., & Yue, Z. (2024). *Handwriting Anomalies and Learning Disabilities through Recurrent Neural Networks and Geometric Pattern Analysis*. http://arxiv.org/abs/2405.07238

Alhusaini, A., Melam, G., Takrouni, E. B., Kashoo, F. Z., & Buragadda, S. (2021). Relationship Between Teacher's and Therapist's Evaluations on Handwriting Performance in First-grade Children. *Iranian Rehabilitation Journal*, *19*(2), 147–155. https://doi.org/10.32598/IRJ.19.2.1236.1

Alipour, F., Faez, K., & Seifzadeh, S. (2013). Ruling lines removal in handwritten documents. *Iranian Conference on Machine Vision and Image Processing, MVIP*, 80–83. https://doi.org/10.1109/IranianMVIP.2013.6779954

Alqahtani, N. D., Alzahrani, B., & Ramzan, M. S. (2023). Detection of Dyslexia Through Images of Handwriting using Hybrid AI Approach. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 14, Number 10). https://doi.org/10.14569/IJACSA.2023.0141099

Alsaqr, A. M. (2021). Remarks on the Use of Pearson's and Spearman's Correlation Coefficients in Assessing Relationships in Ophthalmic Data. *African Vision and Eye Health*, *80*(1), 1–10. https://doi.org/10.4102/AVEH.V80I1.612

Alyahya, H. M., Ben Ismail, M. M., & Al-Salman, A. M. (2023). Intelligent ResNet-18 based Approach for Recognizing and Assessing Arabic Children's Handwriting. *International Conference on Smart Computing and Application, ICSCA 2023*. https://doi.org/10.1109/ICSCA57840.2023.10087682

American Psychiatric Association. (2013). *Diagnostic And Statistical Manual of DSM-5 TM*.

Ammour, A., Aouraghe, I., Khaissidi, G., Mrabti, M., Aboulem, G., & Belahsen, F. (2021). Online Arabic and French handwriting of Parkinson's disease: The impact of segmentation techniques on the classification results. *Biomedical Signal Processing and Control*, *66* (November 2020), 102429. https://doi.org/10.1016/j.bspc.2021.102429

Anggraeny, F. T., Via, Y. V., Mumpuni, R., Rahmania Hatta, H., Putri, N. E., & Bastian, J. (2023). Neural Network Classification in Javanese Handwriting Recognition using Projection Profile Histogram and Local Binary Pattern Histogram. *Technium Romanian Journal of Applied Sciences and Technology*, *16*(Special issue), 123–128. https://doi.org/https://doi.org/10.47577/technium.v16i.9970

Arvind, K. R., Kumar, J., & Ramakrishnan, A. G. (2007). Line Removal and Restoration of Handwritten Strokes. *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, *3*, 208–214. https://doi.org/10.1109/ICCIMA.2007.306

Ashraf, F., & Najam, N. (2020). An epidemiological study of prevalence and comorbidity of non-clinical dyslexia, dysgraphia and dyscalculia symptoms in public and private schools of pakistan. *Pakistan Journal of Medical Sciences*, *36*(7), 1659–1663. https://doi.org/10.12669/pjms.36.7.2486

Ashwini, B., Kaur, M., Singh, D., Roy, S., & Amoon, M. (2023). Efficient Skip Connections-Based Residual Network (ESRNet) for Brain Tumor Classification. *Diagnostics*, *13*(20). https://doi.org/10.3390/diagnostics13203234

Asselborn, T., Chapatte, M., & Dillenbourg, P. (2020). Extending the Spectrum of Dysgraphia: A Data Driven Strategy to Estimate Handwriting Quality. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-60011-8

Asselborn, T., Gargot, T., Kidziński, Ł., Johal, W., Cohen, D., Jolly, C., & Dillenbourg, P. (2018). Automated human-level diagnosis of dysgraphia using a consumer tablet. *Npj Digital Medicine*, *1*(1). https://doi.org/10.1038/s41746-018-0049-x

Babushkin, V., Alsuradi, H., Al-Khalil, M. O., & Eid, M. (2024). Analyzing Arabic Handwriting Style through Hand Kinematics. *Sensors*, *24*(19). https://doi.org/10.3390/s24196357

Bajić, F., Orel, O., & Habijan, M. (2022). A Multi-Purpose Shallow Convolutional Neural Network for Chart Images. *Sensors (Basel, Switzerland)*, *22*(20). https://doi.org/10.3390/s22207695

Bao, W., Yang, C., Wen, S., Zeng, M., Guo, J., Zhong, J., & Xu, X. (2022). A Novel Adaptive Deskewing Algorithm for Document Images. *Sensors (Basel, Switzerland)*, *22*(20). https://doi.org/10.3390/s22207944

Barnett, A. L., Henderson, S. E., Scheib, B., & Schulz, J. (2009). Development and standardization of a new handwriting speed test: The Detailed Assessment of Speed of Handwriting. *British Journal of Educational Psychology*, *2*(6), 137–157. https://doi.org/10.1348/000709909x421937

Bejani, M. M., & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificial Intelligence Review*, *54*(8), 6391–6438. https://doi.org/10.1007/s10462-021-09975-1

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K., & Ghayvat, H. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. In *Electronics (Switzerland)* (Vol. 10, Number 20). MDPI. https://doi.org/10.3390/electronics10202470

Biotteau, M., Danna, J., Baudou, É., Puyjarinet, F., Velay, J. L., Albaret, J. M., & Chaix, Y. (2019). Developmental coordination disorder and dysgraphia: Signs and symptoms, diagnosis, and rehabilitation. *Neuropsychiatric Disease and Treatment*, *15*, 1873–1885. https://doi.org/10.2147/NDT.S120514

Bouhamoum, R., Masmoud, M., Lyousfi, Y., Baazaoui, H., & Mehrotra, D. (2024). Towards an Intelligent Model for Dysgraphia Evolution Tracking. *Procedia Computer Science*, *246*(C), 3713–3722. https://doi.org/10.1016/j.procs.2024.09.185

Brescia, F., Santana, B. E. A., Diaz, M., Vessio, G., Ferrer, M. A., & Castellano, G. (2026). Integrating robotic kinematics and dynamics with online handwriting

features for dysgraphia classification. *Biomedical Signal Processing and Control*, *112*. https://doi.org/10.1016/j.bspc.2025.108560

Cakiroglu, O. (2015). Response to intervention: Early identification of students with learning disabilities. *International Journal of Early Childhood Special Education*, *7*(1), 170–182. https://doi.org/10.20489/intjecse.10399

Carbune, V., Gonnet, P., Deselaers, T., Rowley, H. A., Daryin, A., Calvo, M., Wang, L. L., Keysers, D., Feuz, S., & Gervais, P. (2020). Fast multi-language LSTM-based online handwriting recognition. *International Journal on Document Analysis and Recognition*, *23*(2), 89–102. https://doi.org/10.1007/s10032-020-00350-4

Cengil, E., & Çınar, A. (2022). The effect of deep feature concatenation in the classification problem: An approach on COVID-19 disease detection. *International Journal of Imaging Systems and Technology*, *32*(1), 26–40. https://doi.org/10.1002/ima.22659

Chakraborty, R., Mukherjee, D., Bhattacharyya, A., Mukherjee, H., Sur, M. K., Sen, S., & Roy, K. (2020). Online Handwritten Bangla and Devanagari Character Recognition by using CNN: A Deep Learning Concept. *2020 IEEE International Conference for Convergence in Engineering, ICCE 2020 - Proceedings*, 68–71. https://doi.org/10.1109/ICCE50343.2020.9290566

Chen, L. and M. X. (2009). Anisotropic Resizing of Model with Geometric Textures. *2009 SIAM/ACM Joint Conference on Geometric and Physical Modeling*, 289–294. https://doi.org/10.1145/1629255.1629292

Chen, W., Cong, C., & Huang, L. (2021). Convolutional Neural Network Image Feature Measurement Based on Information Entropy. *Laser & Optoelectronics Progress*. https://doi.org/10.3788/lop202158.2211004

Chi, X., Huang, S., & Li, J. (2021). Handwriting Recognition Based on Resnet-18. *Proceedings - 2021 2nd International Conference on Big Data and Artificial Intelligence and Software Engineering, ICBASE 2021*, 456–459. https://doi.org/10.1109/ICBASE53849.2021.00091

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews Correlation Coefficient (MCC) is More Reliable Than Balanced Accuracy, Bookmaker Informedness, and Markedness in Two-Class Confusion Matrix Evaluation. *BioData Mining*, *14*, 1–22. https://doi.org/10.1186/s13040-021-00244-z

Chung, P. J., Patel, D. R., & Nizami, I. (2020). Disorder of written expression and dysgraphia: definition, diagnosis, and management. *Translational Pediatrics*, *9*, S46–S54. https://doi.org/10.21037/TP.2019.11.01

Chung, P., & Patel, D. R. (2015). *Dysgraphia*. *8*(1), 27–36.

Coradinho, H., Melo, F., Almeida, G., Veiga, G., Marmeleira, J., Teulings, H. L., & Matias, A. R. (2023). Relationship between Product and Process Characteristics of Handwriting Skills of Children in the Second Grade of Elementary School. *Children*, *10*(3). https://doi.org/10.3390/children10030445

Corbille, S., Fromont, E., Anquetil, E., & Nerdeux, P. (2020). Integrating Writing Dynamics in CNN for Online Children Handwriting Recognition. *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, *2020-Septe*, 270–275. https://doi.org/10.1109/ICFHR2020.2020.00057

Dankovicova, Z., Hurtuk, J., & Fecilak, P. (2019). Evaluation of digitalized handwriting for dysgraphia detection using random forest classification method. *SISY 2019 - IEEE 17th International Symposium on Intelligent Systems and Informatics, Proceedings*, 149–153. https://doi.org/10.1109/SISY47553.2019.9111567

Danna, J., Puyjarinet, F., & Jolly, C. (2023). Tools and Methods for Diagnosing Developmental Dysgraphia in the Digital Age: A State of the Art. *Children*, *10*(12). https://doi.org/10.3390/children10121925

D'Antrassi, P., Perrone, I., Cuzzocrea, A., & Accardo, A. (2018). A composite methodology for supporting early-detection of handwriting dysgraphia via big data analysis techniques. In *Smart Innovation, Systems and Technologies* (Vol. 76, pp. 241–253). https://doi.org/10.1007/978-3-319-59480-4_25

Darmatasia and Mohamad Ivan Fanany. (2017). Handwriting Recognition on Form Document Using Convolutional Neural Network and Support Vector Machines (CNN-SVM). *2017 5th International Conference on Information and Communication Technology (ICoIC7)*, *0*(c), 3–5. https://doi.org/10.1109/ICoICT.2017.8074699

Das, M., & Panda, M. (2023). Seam carving, horizontal projection profile and contour tracing for line and word segmentation of language independent handwritten documents. *Results in Engineering*, *18*. https://doi.org/10.1016/j.rineng.2023.101110

Deniz, E., Şengür, A., Kadiroğlu, Z., Guo, Y., Bajaj, V., & Budak, Ü. (2018). Transfer learning based histopathologic image classification for breast cancer detection.

Health Information Science and Systems, *6*(1). https://doi.org/10.1007/s13755-018-0057-x

Deore, S. P., & Pravin, A. (2019). Histogram of oriented gradients based off-line handwritten devanagari characters recognition using SVM, K-NN and NN classifiers. *Revue d'Intelligence Artificielle*, *33*(6), 441–446. https://doi.org/10.18280/ria.330606

Devi, A., & Kavya, G. (2023). Dysgraphia disorder forecasting and classification technique using intelligent deep learning approaches. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *120*(September 2022), 110647. https://doi.org/10.1016/j.pnpbp.2022.110647

Devillaine, L., Lambert, R., Boutet, J., Aloui, S., Brault, V., Jolly, C., & Labyt, E. (2021). Analysis of graphomotor tests with machine learning algorithms for an early and universal pre-diagnosis of dysgraphia. *Sensors*, *21*(21). https://doi.org/10.3390/s21217026

Dimauro, G., Bevilacqua, V., Colizzi, L., & Di Pierro, D. (2020). TestGraphia, a software system for the early diagnosis of dysgraphia. *IEEE Access*, *8*, 19564–19575. https://doi.org/10.1109/ACCESS.2020.2968367

Dinusha, P., Sreekumar, S., & Lijiya, A. (2024). Detection of Potential Specific Learning Disabilities in Children through Handwriting Analysis Using Machine Learning. *2024 IEEE Region 10 Symposium, TENSYMP 2024*. https://doi.org/10.1109/TENSYMP61132.2024.10752261

Dos Santos, R. P., Clemente, G. S., Ren, T. I., & Calvalcanti, G. D. C. (2009). Text line segmentation based on morphology and histogram projection. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 651–655. https://doi.org/10.1109/ICDAR.2009.183

Drotár, P., & Dobeš, M. (2020). Dysgraphia detection through machine learning. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-78611-9

Dui, L. G., Calogero, E., Malavolti, M., Termine, C., Matteucci, M., & Ferrante, S. (2021). Digital Tools for Handwriting Proficiency Evaluation in Children. *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, (2), 1–4. https://doi.org/10.1109/bhi50953.2021.9508539

Dui, L. G., Lunardini, F., Termine, C., Matteucci, M., & Ferrante, S. (2020). A Tablet-Based App to Discriminate Children at Potential Risk of Handwriting Alterations in a Preliteracy Stage. *Proceedings of the Annual International Conference of the*

IEEE Engineering in Medicine and Biology Society, EMBS, *2020-July*, 5856–5859. https://doi.org/10.1109/EMBC44109.2020.9176041

Dyszy-Chudzińska, P. (2009). Developmental Dyslexia vs. Japanese Writing Systems' Neuronal Processing. *Investigationes Linguisticae*, *18*, 27. https://doi.org/10.14746/il.2009.18.2

Dzalani, H., & Shamsuddin, K. (2014). A review of definitions and identifications of specific learning disabilities in Malaysia and challenges in provision of services. *Pertanika Journal of Social Science and Humanities*, *22*(1), 1–18.

El Gannour, O., Hamida, S., Cherradi, B., Al-sarem, M., Raihani, A., Saeed, F., & Hadwan, M. (2022). Concatenation of pre-trained convolutional neural networks for enhanced covid-19 screening using transfer learning technique. *Electronics (Switzerland)*, *11*(1). https://doi.org/10.3390/electronics11010103

Fateah, N., Subhan, & Ifriza, Y. N. (2026). Optimizing Javanese script recognition using fine-tuned ResNet-18 and transfer learning. *IAES International Journal of Artificial Intelligence*, *15*(1), 443–453. https://doi.org/10.11591/ijai.v15.i1.pp443-453

Fatimah, S. H., Djamal, E. C., Ilyas, R., & Renaldi, F. (2019). Personality features identification from handwriting using convolutional neural networks. *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2019*, *6*, 119–124. https://doi.org/10.1109/ICITISEE48480.2019.9003855

Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, *61*, 863–905. https://doi.org/10.1613/JAIR.1.11192

Fischer, J. P., & Luxembourger, C. (2018). A synoptic and theoretical account of character (Digits and capital letters) reversal in writings by typically developing children. In *Education Sciences* (Vol. 8, Number 3). MDPI AG. https://doi.org/10.3390/educsci8030137

Francis, S. B., & Prakash Verma, J. (2024). Deep CNN ResNet-18 based model with attention and transfer learning for Alzheimer's disease detection. *Frontiers in Neuroinformatics* , *18*. https://doi.org/10.3389/fninf.2024.1507217

Franken, A. M., & Harris, S. R. (2021). Teachers' Perceptions of Handwriting Legibility Versus the German Systematic Screening for Motoric-Handwriting

Difficulties (SEMS). *OTJR Occupation, Participation and Health*, *41*(4), 251–258. https://doi.org/10.1177/15394492211033828

Furht, B., Akar, E., & Andrews, W. A. (2018). Creating Image Histograms. In B. Furht, E. Akar, & W. A. Andrews (Eds.), *Digital Image Processing: Practical Approach* (pp. 21–25). Springer International Publishing. https://doi.org/10.1007/978-3-319-96634-2_4

Galaz, Z., Mucha, J., Zvoncak, V., Mekyska, J., Smekal, Z., Safarova, K., Ondrackova, A., Urbanek, T., Havigerova, J. M., Bednarova, J., & Faundez-Zanuy, M. (2020). Advanced Parametrization of Graphomotor Difficulties in School-Aged Children. *IEEE Access*, *8*, 112883–112897. https://doi.org/10.1109/ACCESS.2020.3003214

Galli, M., Cimolin, V., Stella, G., De Pandis, M. F., Ancillao, A., & Condoluci, C. (2019). Quantitative assessment of drawing tests in children with dyslexia and dysgraphia. *Human Movement Science*, *65*(May 2018), 51–59. https://doi.org/10.1016/j.humov.2018.05.001

Gargot, T., Asselborn, T., Pellerin, H., Zammouri, I., Anzalone, S. M., Casteran, L., Johal, W., Dillenbourg, P., Cohen, D., & Jolly, C. (2020). Acquisition of handwriting in children with and without dysgraphia: A computational approach. *PLoS ONE*, *15*(9 September). https://doi.org/10.1371/journal.pone.0237575

Gavenciak, M., Mucha, J., Mekyska, J., Galaz, Z., Zvoncakova, K., & Faundez-Zanuy, M. (2025). Computer-Aided Diagnosis of Graphomotor Difficulties Utilizing Direction-Based Fractional Order Derivatives. *Cognitive Computation*, *17*(1). https://doi.org/10.1007/s12559-024-10360-7

Gemelli, A., Marinai, S., Vivoli, E., & Zappaterra, T. (2023). Deep-learning for dysgraphia detection in children handwritings. *DocEng 2023 - Proceedings of the 2023 ACM Symposium on Document Engineering*. https://doi.org/10.1145/3573128.3609351

Ghosh, T., Sen, S., Obaidullah, S. M., Santosh, K. C., Roy, K., & Pal, U. (2022). Advances in online handwritten recognition in the last decades. In *Computer Science Review* (Vol. 46). Elsevier Ireland Ltd. https://doi.org/10.1016/j.cosrev.2022.100515

Ghouse, F., Paranjothi, K., & Vaithiyanathan, R. (2022). Dysgraphia Classification based on the Non-Discrimination Regularization in Rotational Region Convolutional Neural Network. *International Journal of Intelligent Engineering and Systems*, *15*(1), 55–63. https://doi.org/10.22266/IJIES2022.0228.06

Gold, C., & Zesch, T. (2022). CNN-Based Ruled Line Removal in Handwritten Documents. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): 13639 LNCS* (pp. 530–544). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-21648-0_36

Gómez-Carmona, O., Casado-Mansilla, D., López-de-Ipiña, D., & García-Zubia, J. (2024). Human-in-the-loop machine learning: Reconceptualizing the role of the user in interactive approaches. *Internet of Things (Netherlands)*, *25*. https://doi.org/10.1016/j.iot.2023.101048

Gong, W., Chen, H., Zhang, Z., Zhang, M., Wang, R., Guan, C., & Wang, Q. (2019). A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. *Sensors (Switzerland)*, *19*(7). https://doi.org/10.3390/s19071693

Goodfellow, I., Bengio, Y., & Courville, A. (2017). Deep Learning. In *Deep Learning Book*. https://doi.org/10.13209/j.0479-8023.2014.011

Govindhan, A., Gnanavel, P., Venkat, S., & Vigneshvaran, B. (2020). Analysis of handwriting task using electromyography. *Journal of Physics: Conference Series*, *1706*(1). https://doi.org/10.1088/1742-6596/1706/1/012147

Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: an Overview*. 1–17. http://arxiv.org/abs/2008.05756

Grigorenko, E. L., Fuchs, L. S., Willcutt, E. G., Compton, D. L., Wagner, R. K., & Fletcher, J. M. (2020). Understanding, Educating, and Supporting Children With Specific Learning Disabilities: 50 Years of Science and Practice. *American Psychologist*, *75*(1), 37–51. https://doi.org/10.1037/amp0000452

Guo, Y. (2009). Gray-scale Histograms Feature Extraction Using Matlab. *Computer Knowledge and Technology*.

Hamstra-Bletz, L., & Blöte, A. W. (1993). A longitudinal study on dysgraphic handwriting in primary school. *Journal of Learning Disabilities*, *26*(10), 689–699. https://doi.org/10.1177/002221949302601007

Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. *IEEE Access*, *10*, 19083–19095. https://doi.org/10.1109/ACCESS.2022.3151048

Ibrahim, R. M., Nariman, G. S., & Majeed, H. D. (2023). Offline Kurdish Character Handwritten Recognition (OKCHR) using CNN with Various Preprocessing Techniques. *Journal of Engineering Science and Technology*, *18*(6), 3113 – 3127.

Ileri, R., Latifoglu, F., & Demirci, E. (2020). New Method to Diagnosis of Dyslexia Using 1D-CNN. *TIPTEKNO 2020 - Tip Teknolojileri Kongresi - 2020 Medical Technologies Congress, TIPTEKNO 2020*, 1–4. https://doi.org/10.1109/TIPTEKNO50054.2020.9299241

Ilhan, H. O., Serbes, G., & Aydin, N. (2022). Decision and feature level fusion of deep features extracted from public COVID-19 data-sets. *Applied Intelligence*, *52*(8), 8551–8571. https://doi.org/10.1007/s10489-021-02945-8

Imtiaz, S., Nagabhushan, P., & Gowda, S. D. (2014). Rule Line Detection and Removal in Handwritten Text Images. *2014 Fifth International Conference on Signal and Image Processing*, 310–315. https://doi.org/10.1109/ICSIP.2014.55

Jafarzadeh, P., Choobdar, P., & Safarzadeh, V. M. (2024). Khayyam Offline Persian Handwriting Dataset. *Computer Vision and Pattern Recognition*. https://doi.org/10.48550/arXiv.2406.01025

Jose, M., & Kumar, P. (2021). *Offline Cursive Handwriting Recognition using Convolutional Neural Network*. (August). https://doi.org/10.37896/jxu15.8/029

Jusoh, N. B. M. (2024). *Rubric Ciri-ciri Dysgraphia*.

Kaisar, S., & Chowdhury, A. (2022). Integrating oversampling and ensemble-based machine learning techniques for an imbalanced dataset in dyslexia screening tests. *ICT Express*, https://doi.org/10.1016/j.icte.2022.02.011

Kariyawasam, R., Nadeeshani, M., Hamid, T., Subasinghe, I., & Ratnayake, P. (2019). A Gamified Approach for Screening and Intervention of Dyslexia, Dysgraphia and Dyscalculia. *2019 International Conference on Advancements in Computing, ICAC 2019*, 156–161. https://doi.org/10.1109/ICAC49085.2019.9103336

Kariyawasam, R., Nadeeshani, M., Hamid, T., Subasinghe, I., Samarasinghe, P., & Ratnayake, P. (2019). Pubudu: Deep Learning Based Screening and Intervention of Dyslexia, Dysgraphia and Dyscalculia. *2019 IEEE 14th International Conference on Industrial and Information Systems: Engineering for Innovations for Industry 4.0, ICIIS 2019 - Proceedings*, 476–481. https://doi.org/10.1109/ICIIS47346.2019.9063301

Karthick, B., Singh, H., & Malarvel, M. (2022a). Pre-processing Techniques for Offline Tamil Handwritten Character Recognition. *7th International Conference on*

Communication and Electronics Systems, ICCES 2022 - Proceedings, 976–981. https://doi.org/10.1109/ICCES54183.2022.9835777

Karthick, B., Singh, H., & Malarvel, M. (2022b). Pre-processing techniques for Tamil online handwritten character recognition. *2nd IEEE International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering, ICATIECE 2022*. https://doi.org/10.1109/ICATIECE56365.2022.10047443

Karyna Korovai, O. M. (2020). Handwriting Styles Clustering: Feature Selection and Feature Space Analysis based on Online Input. *IEEE Third International Conference on Data Stream Mining & Processing*, 195–200.

Katarína Šafárová, Jiří Mekyska, V. Z. (2021). Developmental Dysgraphia: A New Approach to Diagnosis. *Educational Psychology: Concepts, Research and Challenges*, *28*(1), 143–160. https://doi.org/https://doi.org/10.18848/2327-7920/CGP/v28i01/143-160

Kementerian Pendidikan Malaysia. (2024). *Buku Data Pendidikan Khas Tahun 2024*.

Khan, A. R., Kurniawan, F., & Mohamad, D. (2009). An automated approach to remove line from text bypassing restoration stage. *2009 2nd International Conference on Computer, Control and Communication, IC4 2009*, *0*, 1–4. https://doi.org/10.1109/IC4.2009.4909214

Kim, D., & Youn, H. (2025). A Hybrid Deep Learning Approach Combining CNNs and Color Histograms for Improved Rock Classification. *Rock Mechanics and Rock Engineering*. https://doi.org/10.1007/s00603-025-04931-9

Kondo, Y., & Iiduka, H. (2025). *Accelerating SGDM via Learning Rate and Batch Size Schedules: A Lyapunov-Based Analysis*. https://doi.org/https://doi.org/10.48550/arXiv.2508.03105

Kumar, G., & Bhatia, P. K. (2014). A detailed review of feature extraction in image processing systems. *International Conference on Advanced Computing and Communication Technologies, ACCT*, 5–12. https://doi.org/10.1109/ACCT.2014.74

Kumar, J., & Doermann, D. (2011). Fast rule-line removal using integral images and support vector machines. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 584–588. https://doi.org/10.1109/ICDAR.2011.123

Kunhoth, J., Al Maadeed, S., Saleh, M., & Akbari, Y. (2023). CNN feature and classifier fusion on novel transformed image dataset for dysgraphia diagnosis in children. *Expert Systems with Applications*, *231*(May), 120740. https://doi.org/10.1016/j.eswa.2023.120740

Kunhoth, J., Al-Maadeed, S., Saleh, M., & Akbari, Y. (2024). Multimodal Ensemble with Conditional Feature Fusion for Dysgraphia Diagnosis in Children from Handwriting Samples. *Computer Vision and Pattern Recognition*. https://doi.org/https://doi.org/10.48550/arXiv.2408.13754

Kunhoth, J., Maadeed, S. Al, Saleh, M., & Akbari, Y. (2023). Exploration and analysis of On-Surface and In-Air handwriting attributes to improve dysgraphia disorder diagnosis in children based on machine learning methods. *Biomedical Signal Processing and Control*, *83*(January), 104715. https://doi.org/10.1016/j.bspc.2023.104715

Kurniawan, D. A., Sihwi, S. W., & Gunarhadi. (2018). An expert system for diagnosing dysgraphia. *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, *2018-Janua*, 468–472. https://doi.org/10.1109/ICITISEE.2017.8285552

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lewis, J. A. (2014). Handwriting Individuality. *Forensic Document Examination*, 45–53. https://doi.org/10.1016/B978-0-12-416693-6.00004-7

Ley, C. S., Syammi, A., & Ab Ghafar, B. (2021). Handwritten Character Recognition Using Convolutional Neural Network. *Progress in Engineering Application and Technology*, *2*(1), 593–611. https://doi.org/10.30880/peat.2021.02.01.058

Liao, J., Guo, L., Jiang, L., Yu, C., Liang, W., Li, K., & Pop, F. (2025). A machine learning-based feature extraction method for image classification using ResNet architecture. *Digital Signal Processing: A Review Journal*, *160*. https://doi.org/10.1016/j.dsp.2025.105036

Ling, Y. M., & Khalid, P. I. (2017). Handwriting difficulty screening tool based on dynamic data from drawing process. *Journal of Telecommunication, Electronic and Computer Engineering*, *9*(3–9), 27–31.

Liu, C.-Y., Chang, C.-C., Way, D.-L., & Tai, W.-K. (2019). Image resizing using fuzzy inferences. *IET Image Processing*, *13*(12), 2058–2066. https://doi.org/https://doi.org/10.1049/iet-ipr.2018.5298

Liu, W., Wei, J., & Meng, Q. (2020). Comparisions on KNN, SVM, BP and the CNN for Handwritten Digit Recognition. *Proceedings of 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications, AEECA 2020*, 587–590. https://doi.org/10.1109/AEECA49918.2020.9213482

Lomurno, E., Dui, L. G., Gatto, M., Bollettino, M., Matteucci, M., & Ferrante, S. (2023). Deep Learning and Procrustes Analysis for Early Dysgraphia Risk Detection with a Tablet Application. *Life*, *13*(3), 598. https://doi.org/10.3390/life13030598

Longjam, T., Kisku, D. R., & Gupta, P. (2023). Writer independent handwritten signature verification on multi-scripted signatures using hybrid CNN-BiLSTM: A novel approach. *Expert Systems with Applications*, *214*. https://doi.org/10.1016/j.eswa.2022.119111

Lopez, C., & Vaivre-Douret, L. (2023). Exploratory Investigation of Handwriting Disorders in School-Aged Children from First to Fifth Grade. *Children*, *10*(9). https://doi.org/10.3390/children10091512

Majnik, M., & Bosnić, Z. (2013). ROC analysis of classifiers in machine learning: A survey. *Intelligent Data Analysis*, *17*(3), 531–558. https://doi.org/10.3233/IDA-130592

Mary L. McHugh. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 276–282. https://doi.org/10.11613/BM.2012.031

Masood, F., Khan, W. U., Ullah, K., Khan, A., Alghamedy, F. H., & Aljuaid, H. (2023). A Hybrid CNN-LSTM Random Forest Model for Dysgraphia Classification from Hand-Written Characters with Uniform/Normal Distribution. *Applied Sciences (Switzerland)*, *13*(7), 1–14. https://doi.org/10.3390/app13074275

McCloskey, M., & Rapp, B. (2017). Developmental dysgraphia: An overview and framework for research. *Cognitive Neuropsychology*, *34*(3–4), 65–82. https://doi.org/10.1080/02643294.2017.1369016

Mekhaznia, T., Djeddi, C., Bensefia, A., & Jamil, A. (2020). A new approach of staff lines removal for manuscript documents. *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020*. https://doi.org/10.1109/ICDABI51230.2020.9325592

Mekyska, J., Bednarova, J., Faundez-Zanuy, M., Galaz, Z., Safarova, K., Zvoncak, V., Mucha, J., Smekal, Z., Ondrackova, A., Urbanek, T., & Havigerova, J. M. (2019). Computerised Assessment of Graphomotor Difficulties in a Cohort of School-aged

Children. *International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*, *2019-Octob*. https://doi.org/10.1109/ICUMT48472.2019.8970767

Mekyska, J., Faundez-Zanuy, M., Mzourek, Z., Galaz, Z., Smekal, Z., & Rosenblum, S. (2017). Identification and Rating of Developmental Dysgraphia by Handwriting Analysis. *IEEE Transactions on Human-Machine Systems*, *47*(2), 235–248. https://doi.org/10.1109/THMS.2016.2586605

Mekyska, J., Galaz, Z., Safarova, K., Zvoncak, V., Cunek, L., Urbanek, T., Havigerova, J. M., Bednarova, J., Mucha, J., Gavenciak, M., Smekal, Z., & Faundez-Zanuy, M. (2023). *Assessment of Developmental Dysgraphia Utilising a Display Tablet*. https://doi.org/https://doi.org/10.1007/978-3-031-45461-5_2

Mo, K., & Zhang, P. (2022). Interference Line Removal Using A Progressive Model. *2022 IEEE 5th International Conference on Artificial Intelligence and Big Data, ICAIBD 2022*, 402–405. https://doi.org/10.1109/ICAIBD55127.2022.9820420

Mohan, J., Rakesh, K., Raviteja, B., & Kumar, N. S. R. P. (2025). Medical Image Enhancement using Histogram Processing and Feature Extraction of Cancer Classification. *Indian Scientific Journal Of Research In Engineering And Management*, *09*(01), 1–9. https://doi.org/10.55041/ijsrem40445

Mohd Ariffin, M., Zarith Nabilah Tengku Othman, T., Shakirah Aziz, N., Mehat, M., & Izza Arshad, N. (2018). Dysgraphi Coach: Mobile Application for Dysgraphia Children in Malaysia. *International Journal of Engineering & Technology*, *7*(4.36), 440. https://doi.org/10.14419/ijet.v7i4.36.23912

Mudeng, V., Kim, M., & Choe, S. W. (2022). Prospects of Structural Similarity Index for Medical Image Analysis. *Applied Sciences (Switzerland)*, *12*(8). https://doi.org/10.3390/app12083754

Mukaka, M. M. (2012). Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal*, *24*(3), 69–71. www.mmj.medcol.mw

Muktamath, V. U., Hegde, P. R., & Chand, S. (2021). *Types of Specific Learning Disability*. https://doi.org/http://dx.doi.org/10.5772/intechopen.100809

Mulakaluri, S., & Girisha, G. S. (2021). Identifying Handwriting Difficulties in Children in Devanagari Script Using Machine Learning. In V. Bhateja, S. C. Satapathy, C. M. Travieso-González, & V. N. M. Aradhya (Eds.), *Data*

Engineering and Intelligent Computing (pp. 191–204). Springer Singapore. https://doi.org/https://doi.org/10.1007/978-981-16-0171-2_19

Najadat, H. M., Alshboul, A. A., & Alabed, A. F. (2019). Arabic Handwritten Characters Recognition using Convolutional Neural Network. *2019 10th International Conference on Information and Communication Systems, ICICS 2019*, (January), 147–151. https://doi.org/10.1109/IACS.2019.8809122

Najjar, F. H., Khudhair, K. T., Khaleq, A. H. A., Kadhim, O. N., Abedi, F., & Al-Kharsan, I. H. (2022). *Histogram Features Extraction for Edge Detection Approach*. 373–378. https://doi.org/10.1109/IICETA54559.2022.9888697

Nigam, S., Behera, A. P., Verma, S., & Nagabhushan, P. (2024). Deformity removal from handwritten text documents using variable cycle GAN. *International Journal on Document Analysis and Recognition*, *27*(4), 615–627. https://doi.org/10.1007/s10032-024-00466-x

Nilsson, J., & Akenine-Möller NVIDIA, T. (2020). *Understanding SSIM*. Retrieved https://research.nvidia.com/publication/2020-07_Understanding-SSIM

Nousi, P., Tefas, A., & Pitas, I. (2019). Deep Convolutional Feature Histograms for Visual Object Tracking. *International Conference on Acoustics, Speech, and Signal Processing*, 8375–8379. https://doi.org/10.1109/ICASSP.2019.8683004

Patnaik, S., Kumari, S., & Das Mahapatra, S. (2020). Comparison of deep CNN and ResNet for Handwritten Devanagari Character Recognition. *2020 IEEE International Conference for Convergence in Engineering, ICCE 2020 - Proceedings*, 235–238. https://doi.org/10.1109/ICCE50343.2020.9290637

Peeples, J., Xu, W., & Zare, A. (2022). Histogram Layers for Texture Analysis. *IEEE Transactions on Artificial Intelligence*, *3*, 541–552. https://doi.org/10.1109/TAI.2021.3135804

Peng, J., Shi, C., Laugeman, E., Hu, W., Zhang, Z., Mutic, S., & Cai, B. (2020). Implementation of the structural Similarity (SSIM) index as a quantitative evaluation tool for dose distribution error detection. *Medical Physics*, *47*(4), 1907–1919. https://doi.org/10.1002/mp.14010

Pham, T. T. D., & Won, C. S. (2019). Facial action units for training convolutional neural networks. *IEEE Access*, *7*, 77816–77824. https://doi.org/10.1109/ACCESS.2019.2921241

Prunty, M., & Barnett, A. L. (2017). Understanding handwriting difficulties: A comparison of children with and without motor impairment. *Cognitive*

Neuropsychology, *34*(3–4), 205–218. https://doi.org/10.1080/02643294.2017.1376630

Puri, S., & Singh, S. P. (2019). An efficient Devanagari character classification in printed and handwritten documents using SVM. *Procedia Computer Science*, *152*, 111–121. https://doi.org/10.1016/j.procs.2019.05.033

Raghuram, P., Ramachandran, P., Sethuraman, L., Sushmita, S., & Udayakumar, N. (2023). Prevalence of Handwriting Difficulty among Second Grade School Students in North Chennai: A Cross-sectional Study. *The Indian Journal of Occupational Therapy*, *55*(3), 91–93. https://doi.org/10.4103/ijoth.ijoth_4_24

Rahman, A. U., & Halim, Z. (2022). A graph-based solution for writer identification from handwritten text. *Knowledge and Information Systems*, *64*(6), 1501–1523. https://doi.org/10.1007/s10115-022-01676-7

Rajput, D., Wang, W. J., & Chen, C. C. (2023). Evaluation of a decided sample size in machine learning applications. *BMC Bioinformatics*, *24*(1). https://doi.org/10.1186/s12859-023-05156-9

Rakesh, S., Kushal Reddy, P., Prashanth, V., & Srinath Reddy, K. (2024). Handwritten text recognition using deep learning techniques: A survey. *MATEC Web of Conferences*, *392*, 01126. https://doi.org/10.1051/matecconf/202439201126

Ramlan, S. A., Isa, I. S., Ismail, A. P., Osman, M. K., & Che Soh, Z. H. (2024). Development of potential dysgraphia handwriting dataset. *Data in Brief*, *54*. https://doi.org/10.17632/39hr8dx76p.1

Ramlan, S. A., Isa, I. S., Osman, M. K., Ismail, A. P., & Che Soh, Z. H. (2024). Comparing CNN-based Architectures for Dysgraphia Handwriting Classification Performance. *Pertanika Journal of Science and Technology*, *32*(5), 2013–2032. https://doi.org/10.47836/pjst.32.5.05

Ramlan, S. A., Isa, I. S., Osman, M. K., Ismail, A. P., & Soh, Z. H. C. (2023). Classification Performance of Directed Acyclic Graph Network on Potential Dyslexia Handwriting Images. *Proceedings - 13th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2023*, 149–154. https://doi.org/10.1109/ICCSCE58721.2023.10237096

Rani, N. S., Pruthvi, T. R., Rao, A. G., & Bipin, N. B. J. (2021). Automated text line segmentation and table detection for pre-printed document image analysis systems. *2021 3rd International Conference on Signal Processing and Communication, ICPSC 2021*, 723–730. https://doi.org/10.1109/ICSPC51351.2021.9451779

Rashid, H. A., Malik, T., Siddiqui, I., Bhatti, N., & Samad, A. (2023). DYSIGN: Towards Computational Screening of Dyslexia and Dysgraphia Based on Handwriting Quality. *Proceedings of IDC 2023 - 22nd Annual ACM Interaction Design and Children Conference: Rediscovering Childhood*, 532–536. https://doi.org/10.1145/3585088.3593890

Rawat, W., & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, *29*(9), 2352–2449. https://doi.org/10.1162/neco_a_00990

Raza, T. F., Arif, H., Darvagheh, S. H., & Hajjdiab, H. (2017). Interactive mobile application for testing children with Dysgraphia. *ACM International Conference Proceeding Series*, *Part F1283*, 432–436. https://doi.org/10.1145/3055635.3056599

Refaey, M. A. A. (2014). Fast Detection and Removal Algorithms for Ruled Lines in Full-Color Scanned Handwritten Documents. *Proceedings of the International Conference on Computer Science, Computer Engineering, and Social Media*. https://doi.org/https://doi.org/10.13140/2.1.3788.9606

Refaey, M. A. A. (2015). Ruled lines detection and removal in grey level handwritten image documents. *2015 6th International Conference on Information and Communication Systems, ICICS 2015*, 218–221. https://doi.org/10.1109/IACS.2015.7103230

Rehman, A., & Saba, T. (2011). Document skew estimation and correction: Analysis of techniques, common problems and possible solutions. *Applied Artificial Intelligence*, *25*(9), 769–787. https://doi.org/10.1080/08839514.2011.607009

Richard, G., & Serrurier, M. (2020). *Dyslexia and Dysgraphia prediction: A new machine learning approach*. https://doi.org/10.48550/arXiv.2005.06401

Riley, R. D., Ensor, J., Snell, K. I. E., Archer, L., Whittle, R., Dhiman, P., Alderman, J., Liu, X., Kirton, L., Manson-Whitton, J., van Smeden, M., Moons, K. G., Nirantharakumar, K., Cazier, J. B., Denniston, A. K., Van Calster, B., & Collins, G. S. (2025). Importance of sample size on the quality and utility of AI-based prediction models for healthcare. In *The Lancet Digital Health* (Vol. 7, Number 6). Elsevier Ltd. https://doi.org/10.1016/j.landig.2025.01.013

Robaa, M., Balat, M., Awaad, R., Omar, E., & Aly, S. A. (2024). *Explainable AI in Handwriting Detection for Dyslexia Using Transfer Learning*. https://doi.org/10.48550/arXiv.2410.19821

Rosenblum, S. (2018). Inter-relationships between objective handwriting features and executive control among children with developmental dysgraphia. *PLoS ONE*, *13*(4), 1–14. https://doi.org/10.1371/journal.pone.0196098

Rosenblum, S., Aloni, T., & Josman, N. (2010). Relationships between handwriting performance and organizational abilities among children with and without dysgraphia: A preliminary study. *Research in Developmental Disabilities*, *31*(2), 502–509. https://doi.org/10.1016/j.ridd.2009.10.016

Rosenblum, S., & Dror, G. (2017). Identifying Developmental Dysgraphia Characteristics Utilizing Handwriting Classification Methods. *IEEE Transactions on Human-Machine Systems*, *47*(2), 293–298. https://doi.org/10.1109/THMS.2016.2628799

Rosenblum, S., Weiss, P. L., & Parush, S. (2004). Handwriting evaluation for developmental dysgraphia: Process versus product. *Reading and Writing*, *17*(5), 433–458. https://doi.org/10.1023/B:READ.0000044596.91833.55

Rosli, M. S. A. Bin, Isa, I. S., Ramlan, S. A., Sulaiman, S. N., & Maruzuki, M. I. F. (2021). *Development of CNN Transfer Learning for Dyslexia Handwriting Recognition*. (August), 194–199. https://doi.org/10.1109/iccsce52189.2021.9530971

Šafárová, K., Mekyska, J., Zvončák, V., Galáž, Z., Francová, P., Čechová, B., Losenická, B., Smékal, Z., Urbánek, T., Havigerová, J. M., & Rosenblum, S. (2020). Psychometric Properties of Screening Questionnaires for Children With Handwriting Issues. *Frontiers in Psychology*, *10*(January), 1–12. https://doi.org/10.3389/fpsyg.2019.02937

Saha, R., Mukherjee, A., Sarkar, A., & Dey, S. (2020). Extraction of common feature of dysgraphia patients by handwriting analysis using variational autoencoder. *Deep Learning: Research and Applications*, 85–104. https://doi.org/10.1515/9783110670905-005

Sakamat, N., Elaiza, N., Khalid, A., Asasi, P., Dengkil, K., Alam, S., & Ehsan, S. D. (2019). Handwriting Criteria Analysis of Lower Primary School. *E – Academia Special Issue GraCe, 2018*, 35–41.

Sara, U., Akter, M., & Uddin, M. S. (2019). Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. *Journal of Computer and Communications*, *07*(03), 8–18. https://doi.org/10.4236/jcc.2019.73002

Saunders, C. P., Davis, L. J., & Buscaglia, J. (2011). Using automated comparisons to quantify handwriting individuality. *Journal of Forensic Sciences*, *56*(3), 683–689. https://doi.org/10.1111/j.1556-4029.2011.01713.x

Schober, P., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Sebastian, C. J. M., Rogala, M. W. J., & Paglinawan, C. C. (2024). Neurological Disorder Rooted-Impaired Handwriting to Text Using Neural Network Image Processing for Special Education Applications. *2024 16th International Conference on Computer and Automation Engineering, ICCAE 2024*, 264–269. https://doi.org/10.1109/ICCAE59995.2024.10569849

Seki, Y. (2020). Collection of Online and Offline Handwritten Japanese Characters and Handwriting Classification Using the Data. *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, *2020-Septe*, 264–269. https://doi.org/10.1109/ICFHR2020.2020.00056

Seman, N. S. L., Isa, I. S., Ramlan, S. A., Li-Chih, W., & Maruzuki, M. I. F. (2021). *Classification of Handwriting Impairment Using CNN for Potential Dyslexia Symptom*. (August), 188–193. https://doi.org/10.1109/iccsce52189.2021.9530989

Semma, A., Lazrak, S., Hannad, Y., Boukhani, M., & El Kettani, Y. (2021). Writer Identification: The Effect of Image Resizing on CNN Performance. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *46*(4/W5-2021), 501–507. https://doi.org/10.5194/isprs-Archives-XLVI-4-W5-2021-501-2021

Sesa, O. A., Haikal, A. Y., Elhosseini, M. A., & Gad, H. H. (2022). *Smart Bagged Tree-based Classifier optimized by Random Forests (SBT-RF) to Classify Brain-Machine Interface Data*. 895–908.

Setiadi, D. R. I. M., Rustad, S., Andono, P. N., & Shidik, G. F. (2023). Digital Image Steganography Survey and Investigation (Goal, Assessment, Method, Development, and Dataset). *Signal Processing*, *206*. https://doi.org/10.1016/j.sigpro.2022.108908

Sharma, P., Agarwal, B., & Yadav, G. S. (2025). Hybrid Dysgraphia Detection Using Transfer Learning and Handcrafted Features. *Proceedings of 2025 8th International Conference on Emerging Technologies in Computer Engineering:*

Advances in Computing, Healthcare and Smart Systems, ICETCE 2025, 13–17. https://doi.org/10.1109/ICETCE66848.2025.11388145

Shobha Rani, N., & Vasudev, T. (2018). An efficient technique for detection and removal of lines with text stroke crossings in document images. In *Lecture Notes in Networks and Systems* (Vol. 14, Number January, pp. 83–97). https://doi.org/10.1007/978-981-10-5146-3_9

Sihwi, S. W., Fikri, K., & Aziz, A. (2019). Dysgraphia Identification from Handwriting with Support Vector Machine Method. *Journal of Physics: Conference Series*, *1201*(1). https://doi.org/10.1088/1742-6596/1201/1/012050

Srihari, S. N., Cha, S. H., Arora, H., & Lee, S. (2002). Individuality of handwriting. *Journal of Forensic Sciences*, *47*(4), 856–872.

Srinivasan, A., & Mishra, A. (2024). Receiver Operating Characteristic (ROC) Curve Analysis for Diagnostic Studies. In A. Srinivasan, A. Mishra, & P. Kumar-M (Eds.), *R for Basic Biostatistics in Medical Research* (pp. 253–258). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-6980-3_12

Sukiman, S. A., Azura Husin, N., Hamdan, H., & Azmi Murad, M. A. (2024). Handwriting Image Classification for Automated Diagnosis of Learning Disabilities: A Review on Deep Learning Models and Future Directions. *2024 19th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP)*, 1–6. https://doi.org/10.1109/iSAI-NLP64410.2024.10799245

Sun, S. (2011). Meta-analysis of Cohen's kappa. *Health Services and Outcomes Research Methodology*, *11*(3–4), 145–163. https://doi.org/10.1007/s10742-011-0077-3

Syahida, M. T. A., Suhaila, A. H. S., Filzah, M. R. N., Ismail; Zahian, Kamaludin; Adzhar, Mohamad Arif;, & Mahmud, M. (2023). Development of the Game-Based Pre-Screening Test for Students with Dyslexia. *Journal of Computer Science & Computational Mathematics*, *13*(4). https://doi.org/10.20967/jcscm.2023.04.006

Tamura, J., Itaya, Y., Hayashi, K., & Yamamoto, K. (2025). *Statistical Inference of the Matthews Correlation Coefficient for Multiclass Classification*. http://arxiv.org/abs/2503.06450

Tang, X., Wang, C., & Yin, S. (2025). Research on Occluded Facial Expression Recognition Algorithm Based on Hybrid Attention Mechanism and Multi-scale

Feature Extraction. *Proceedings of 2025 6th International Conference on Computer Information and Big Data Applications, CIBDA 2025*, 87–92. https://doi.org/10.1145/3746709.3746726

Thi Thi Zin, S. T. M. Z. P. T. O. (2021). Handwritten character recognition on android for basic education using convolutional neural network. *Electronics (Switzerland)*, *10*(8). https://doi.org/10.3390/electronics10080904

Tsourounis, D., Kastaniotis, D., Theoharatos, C., Kazantzidis, A., & Economou, G. (2022). SIFT-CNN: When Convolutional Neural Networks Meet Dense SIFT Descriptors for Image and Sequence Classification. *Journal of Imaging*, *8*(10). https://doi.org/10.3390/jimaging8100256

Van Waelvelde, H., Hellinckx, T., Peersman, W., & Smits-Engelsman, B. C. M. (2012). SOS: A screening instrument to identify children with handwriting impairments. In *Physical and Occupational Therapy in Pediatrics* (Vol. 32, Number 3, pp. 306–319). https://doi.org/10.3109/01942638.2012.678971

Vilasini V, Banu Rekha B, Sandeep V, & Vishnu Charan Venkatesh. (2022). Deep Learning Techniques to Detect Learning Disabilities Among children using Handwriting. *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, 1710–1717. https://doi.org/10.1109/ICICICT54557.2022.9917890

Villegas-Ch, W., Urbina-Camacho, I., & García-Ortiz, J. (2023). Detection of Abnormal Patterns in Children's Handwriting by Using an Artificial-Intelligence-Based Method. *Informatics*, *10*(2). https://doi.org/10.3390/informatics10020052

Vlachos, F., & Avramidis, E. (2020). The Difference between Developmental Dyslexia and Dysgraphia: Recent Neurobiological Evidence. *International Journal of Neuroscience and Behavioral Science*, *8*(1), 1–5. https://doi.org/10.13189/ijnbs.2020.080101

Vydeki, D., Bhandari, D., Patil, P. P., & Kulkarni, A. A. (2024). *Towards Accessible Learning: Deep Learning-Based Potential Dysgraphia Detection and OCR for Potentially Dysgraphic Handwriting*. https://doi.org/10.48550/arXiv.2411.13595

Wahid, M. F., Shahriar, M. F., & Sobuj, M. S. I. (2021). A Classical Approach to Handcrafted Feature Extraction Techniques for Bangla Handwritten Digit Recognition. *Proceedings of International Conference on Electronics, Communications and Information Technology, ICECIT 2021*. https://doi.org/10.1109/ICECIT54077.2021.9641406

Waisi, N. (2024). Deep Feature Fusion Method for Images Classification. *International Journal of Computational and Electronic Aspects in Engineering*, *5*(4), 148–153. https://doi.org/10.26706/ijceae.5.4.20241103

Wang, Y., Xiao, W., & Li, S. (2021). Offline Handwritten Text Recognition Using Deep Learning: A Review. *Journal of Physics: Conference Series*, *1848*(1). https://doi.org/10.1088/1742-6596/1848/1/012015

Wang, Z., Li, H., Ouyang, W., & Wang, X. (2018). Learnable Histogram: Statistical Context Features for Deep Neural Networks. *ArXiv*. https://doi.org/10.1007/978-3-319-46448-0_15

Wei, W., Xu, X., Hu, G., Shao, Y., & Wang, Q. (2024). Deep Learning and Histogram-Based Grain Size Analysis of Images. *Sensors*, *24*(15). https://doi.org/10.3390/s24154923

Wong, L. C., & Loh, W. P. (2019). Segregating Offline and Online Handwriting for Conditional Classification Analysis. *IOP Conference Series: Materials Science and Engineering*, *530*(1). https://doi.org/10.1088/1757-899X/530/1/012058

Wu, S., Chen, Y.-C., Li, X., Wu, A.-C., You, J.-J., & Zheng, W.-S. (2016, April 28). An Enhanced Deep Feature Representation for Person Re-identification. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. https://doi.org/10.1109/WACV.2016.7477681

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A Survey of Human-in-the-loop for Machine Learning. *Future Generation Computer Systems*, *135*, 364–381. https://doi.org/10.1016/j.future.2022.05.014

Xu, G., Wang, X., Wu, X., Leng, X., & Xu, Y. (2025). Development of residual learning in deep neural networks for computer vision: A survey. In *Engineering Applications of Artificial Intelligence* (Vol. 142). Elsevier Ltd. https://doi.org/10.1016/j.engappai.2024.109890

Xu, Y., & Goodacre, R. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, *2*(3), 249–262. https://doi.org/10.1007/s41664-018-0068-2

Xu, Z., Mittal, P. S., Ahmed, M. M., Adak, C., & Cai, Z. G. (2025). Assessing penmanship of Chinese handwriting: a deep learning-based approach. *Reading and Writing*, *38*(3), 723–743. https://doi.org/10.1007/s11145-024-10531-w

Yogarajah, P. (2020). Deep Learning Approach to Automated Detection of Dyslexia-Dysgraphia. *The 25th IEEE International Conference on Pattern Recognition*, 1–12. https://www.micc.unifi.it/icpr2020/

Yogarajah, P., & Braj, P. (2020). *Deep Learning Approach to Automated Detection of Dyslexia-Dysgraphia*.

Yu, W., Wang, H., Liu, J., & Wang, G. (2024). Multispectral image fusion technique based on decoupling of information. *IET Image Processing*. https://doi.org/10.1049/ipr2.13296

Zeng, G. (2025). Invariance Properties and Evaluation Metrics Derived from the Confusion Matrix in Multiclass Classification. *Mathematics*, *13*(16). https://doi.org/10.3390/math13162609

Zhang, C., Tu, C., Wang, Z., Cao, W., & Cao, W. (2026). ResNet18-ThunderSVM: Hybrid intelligence for handwritten digit recognition by fusing deep spatial features and high-performance classification. *Scientific Reports*, *16*(1). https://doi.org/10.1038/s41598-026-38258-4

Zolna, K., Asselborn, T., Jolly, C., Casteran, L., Marie-Ange~Nguyen-Morel, Johal, W., & Dillenbourg, P. (2019). *The Dynamics of Handwriting Improves the Automated Diagnosis of Dysgraphia*. http://arxiv.org/abs/1906.07576

Zvoncak, V., Mekyska, J., Safarova, K., Smekal, Z., & Brezany, P. (2019). New approach of dysgraphic handwriting analysis based on the tunable Q-factor wavelet transform. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings*, 289–294. https://doi.org/10.23919/MIPRO.2019.8756872

APPENDICES

APPENDIX 1

Ethical Approval Letter by REC, UiTM

www.uitm.edu.my

UNIVERSITI TEKNOLOGI MARA

Pejabat
Timbalan Naib Canselor
(Penyelidikan dan Inovasi)

Reference : 600-TNCPI(5/1/6)
Our reference : REC/09/2021 (MR/830)
Date : 30th September 2021

Ir Dr Iza Sazanita binti Isa
(Siti Azura binti Ramlan)
UiTM Cawangan Pulau Pinang
Kampus Permatang Pauh
13500 Permatang Pauh, Seberang Perai
Pulau Pinang

Dear Ir Dr Iza Sazanita,

APPROVAL LETTER - UiTM RESEARCH ETHICS COMMITTEE

Thank you for submitting your research proposal to Research Ethics Committee (REC). After considering your application, the Committee approved your proposal entitled "Classification of Dysgraphia Severity Using CNN Model and Image Processing Techniques" at Persatuan Disleksia Malaysia and Program Inklusif Disleksia Kementerian Pelajaran Malaysia in Kuala Lumpur, Selangor, Kedah, Pulau Pinang and Johor.

Details of the approval are as follows:

Ref. number:	REC/09/2021 (MR/830)
Approval Period:	30th September 2021 until 28th March 2024
Authorised personnel:	1. Ir Dr Iza Sazanita binti Isa 2. Siti Azura binti Ramlan

The UiTM Research Ethics Committee operates in accordance to the ICH Good Clinical Practice Guidelines, Malaysian Good Clinical Practice Guidelines and the Declaration of Helsinki. The approval of this project is conditional upon your continuing compliance with these guidelines and declaration.

We draw to your attention the requirement that a report on this research, must be submitted every 12 months from the date of the approval or on the completion of the project, whichever occurs first. Failure to submit reports will result in withdrawal of consent for the project to proceed. Amendments, if any, to the study documents are to be submitted to REC for approval.

If you require further information, please contact REC Secretariat at 03-55448069/03-55442794 or email at recsecretariat@uitm.edu.my.

Yours sincerely,

PROFESSOR DATO' DR ABU BAKAR ABDUL MAJEED
Chairman
UiTM Research Ethics Committee

c.c.: Rector, UiTM Pulau Pinang Branch

Universiti Teknologi MARA
Aras 3, Bangunan Wawasan
40450 Shah Alam, Selangor, MALAYSIA
Tel: (+603) 5544 2004/2255
Faks: (+603) 5544 2070

ISO 9001:2015 No. Sijil: 10120156

UiTM di hatiku

APPENDIX 2

Research Ethics Approval Letter by MOE

KEMENTERIAN PENDIDIKAN MALAYSIA
BAHAGIAN PERANCANGAN DAN PENYELIDIKAN DASAR PENDIDIKAN
ARAS 1-4, BLOK E8
KOMPLEKS KERAJAAN PARCEL E
PUSAT PENTADBIRAN KERAJAAN PERSEKUTUAN
62604 PUTRAJAYA

TEL : 0388846591
FAKS : 0388846579

Ruj. Kami : KPM.600-3/2/3-eras(15243)
Tarikh : 25 Januari 2023

SITI AZURA BINTI RAMLAN
NO. KP :

13700 PERAI
PULAU PINANG

Tuan,

KELULUSAN BERSYARAT UNTUK MENJALANKAN KAJIAN :
CLASSIFICATION OF DYSGRAPHIA SEVERITY USING CNN MODEL AND IMAGE PROCESSING TECHNIQUES

Perkara di atas adalah dirujuk.

2. Sukacita dimaklumkan bahawa permohonan tuan untuk menjalankan kajian seperti di bawah telah diluluskan dengan syarat :

" KELULUSAN INI BERGANTUNG KEPADA KEBENARAN PENGARAH JPN DAN PERTIMBANGAN PENTADBIR SEKOLAH. MAKLUMAT LENGKAP KAJIAN DAN PENGKAJI HENDAKLAH DINYATAKAN DALAM INSTRUMEN KAJIAN. PENGLIBATAN, PEMERHATIAN SERTA RAKAMAN VIDEO TERHADAP AKTIVITI PENGAJARAN DAN PEMBELAJARAN MURID DI DALAM BILIK DARJAH TIDAK DIBENARKAN. "

3. Kelulusan adalah berdasarkan kepada kertas cadangan penyelidikan dan instrumen kajian yang dikemukakan oleh tuan kepada bahagian ini. Walau bagaimanapun kelulusan ini bergantung kepada kebenaran Jabatan Pendidikan Negeri dan Pengetua / Guru Besar yang berkenaan.

4. Surat kelulusan ini sah digunakan bermula dari **30 Januari 2023** hingga **28 Julai 2023**

5. Tuan dikehendaki menyerahkan senaskhah laporan akhir kajian dalam bentuk *hardcopy* bersama salinan *softcopy* berformat pdf dalam CD kepada Bahagian ini.Tuan juga diingatkan supaya mendapat kebenaran terlebih dahulu daripada Bahagian ini sekiranya sebahagian atau sepenuhnya dapatan kajian tersebut hendak diterbitkan di mana-mana forum, seminar atau diumumkan kepada media massa.

Sekian untuk makluman dan tindakan tuan selanjutnya. Terima kasih.

"BERKHIDMAT UNTUK NEGARA"

Saya yang menjalankan amanah,

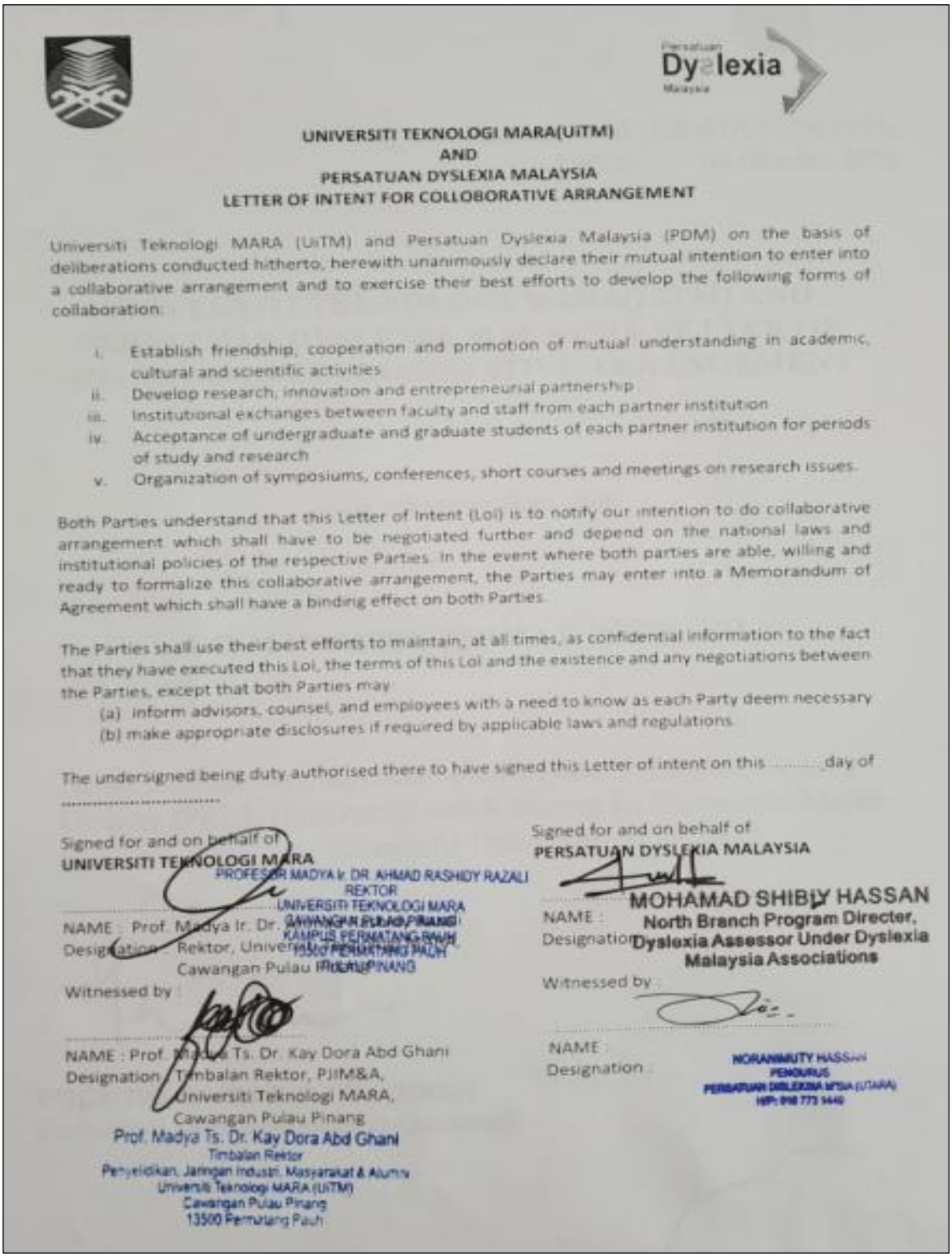
Ketua Penolong Pengarah Kanan
Sektor Penyelidikan dan Penilaian Dasar
b.p. Pengarah
Bahagian Perancangan dan Penyelidikan Dasar Pendidikan
Kementerian Pendidikan Malaysia

salinan kepada:-

JABATAN PENDIDIKAN JOHOR
JABATAN PENDIDIKAN KEDAH
JABATAN PENDIDIKAN PULAU PINANG
JABATAN PENDIDIKAN SELANGOR
JABATAN PENDIDIKAN WILAYAH PERSEKUTUAN KUALA LUMPUR

APPENDIX 3

Letter of Intent for Collaborative Arrangement (LOI)



UNIVERSITI TEKNOLOGI MARA(UiTM)
AND
PERSATUAN DYSLEXIA MALAYSIA
LETTER OF INTENT FOR COLLOBORATIVE ARRANGEMENT

Universiti Teknologi MARA (UiTM) and Persatuan Dyslexia Malaysia (PDM) on the basis of deliberations conducted hitherto, herewith unanimously declare their mutual intention to enter into a collaborative arrangement and to exercise their best efforts to develop the following forms of collaboration:

i. Establish friendship, cooperation and promotion of mutual understanding in academic, cultural and scientific activities
ii. Develop research, innovation and entrepreneurial partnership
iii. Institutional exchanges between faculty and staff from each partner institution
iv. Acceptance of undergraduate and graduate students of each partner institution for periods of study and research
v. Organization of symposiums, conferences, short courses and meetings on research issues.

Both Parties understand that this Letter of Intent (LoI) is to notify our intention to do collaborative arrangement which shall have to be negotiated further and depend on the national laws and institutional policies of the respective Parties. In the event where both parties are able, willing and ready to formalize this collaborative arrangement, the Parties may enter into a Memorandum of Agreement which shall have a binding effect on both Parties.

The Parties shall use their best efforts to maintain, at all times, as confidential information to the fact that they have executed this LoI, the terms of this LoI and the existence and any negotiations between the Parties, except that both Parties may

(a) inform advisors, counsel, and employees with a need to know as each Party deem necessary
(b) make appropriate disclosures if required by applicable laws and regulations.

The undersigned being duty authorised there to have signed this Letter of intent on this ..... day of .....

Signed for and on behalf of
UNIVERSITI TEKNOLOGI MARA

PROFESOR MADYA Ir. DR. AHMAD RASHIDY RAZALI
REKTOR
UNIVERSITI TEKNOLOGI MARA
CAWANGAN PULAU PINANG
KAMPUS PERMATANG PAUH
13500 PERMATANG PAUH
PULAU PINANG

NAME : Prof. Madya Ir. Dr. Ahmad Rashidy Razali
Designation : Rektor, Universiti Teknologi MARA,
Cawangan Pulau Pinang

Witnessed by :

NAME : Prof. Madya Ts. Dr. Kay Dora Abd Ghani
Designation : Timbalan Rektor, PJIM&A,
Universiti Teknologi MARA,
Cawangan Pulau Pinang

Prof. Madya Ts. Dr. Kay Dora Abd Ghani
Timbalan Rektor
Penyelidikan, Jaringan Industri, Masyarakat & Alumni
Universiti Teknologi MARA (UiTM)
Cawangan Pulau Pinang
13500 Permatang Pauh

Signed for and on behalf of
PERSATUAN DYSLEXIA MALAYSIA

MOHAMAD SHIBLY HASSAN
North Branch Program Directer,
Dyslexia Assessor Under Dyslexia
Malaysia Associations

NAME :
Designation :

Witnessed by :

NAME :
Designation :

NORANIMUTY HASSAN
PENGURUS
PERSATUAN DISLEKSIA MSIA (UTARA)
H/P: 016 773 1446

Pejabat Penasihat Undang-Undang
Office of the Legal Adviser

Pejabat Penasihat Undang-Undang

UNIVERSITI TEKNOLOGI MARA

Rujukan kami : 100-PUU(32/6/3194)
Tarikh : 04 Oktober 2022

UNIVERSITI TEKNOLOGI MARA(UiTM) AND PERSATUAN DYSLEXIA MALAYSIA LETTER OF INTENT FOR COLLOBORATIVE ARRANGEMENT

yang dikemukakan oleh

UITM KAMPUS PERMATANG PAUH

telah disemak dan diperakui oleh Pejabat Penasihat Undang-Undang (PPUU) Universiti untuk dibawa ke Mesyuarat Majlis Eksekutif Universiti.

PROF DR HAIDAR DZIYAUDDIN
Penasihat Undang-Undang Universiti

APPENDIX 4

Data Acquisition Form

SESI PENGUMPULAN DATA
Versi Bahasa Melayu
Data ini diperlukan mengumpul imej tulisan kanak-kanak normal, disgrafia atau kanak-kanak disleksia yang mempunyai masalah disgrafia.
BAHAGIAN A: Maklumat Pengambilan
Tanda √ untuk kotak yang berkaitan.

1. Jantina:
   ☐ Lelaki

   ☐ Perempuan
2. Umur: _____
3. *Menulis dengan tangan: kiri/kanan
4. Adakah pelajar mempunyai masalah pembelajaran

   ☐ Tidak

   ☐ Ya
5. Sekiranya Ya (soalan 4), sila nyatakan _____

   dan

   Keputusan diagnosis: ☐ Risiko Rendah

   ☐ Risiko Sederhana

   ☐ Risiko Tinggi

BAHAGIAN B

Tugas 1: Komposisi ayat
Pelajar diberikan ayat berikut dan menulis semula di ruang yang disediakan.

Baju itu baru dibeli oleh emak.

_____

_____

_____

Burung itu berada di dalam sangkar. Sangkar baru itu milik bapa.

_____

_____

_____

_____

Saya suka bermain basikal pada waktu lapang. Basikal itu dibeli oleh ayah saya. Roda basikal saya berwarna hijau dan putih.

_____

_____

_____

_____

APPENDIX 5

Agreement Form

REC 12 / 2020 Rev. 1

Jawatankuasa Etika Penyelidikan
Pusat Pengurusan Penyelidikan
Universiti Teknologi MARA
40450 SHAH ALAM
Tel: 03 – 5544-8069, Faks: 03 – 5544-2096/2767

BORANG PERSETUJUAN MENYERTAI PROJEK

Ibu bapa / wakil sah anda (LAR) telah memberi izin untuk anda menyertai projek yang bertajuk **Klasifikasi Tahap Disgrafia Menggunakan Model CNN dan Teknik Pemprosesan Imej**. Kami ingin memberi penerangan mengenai projek tersebut, supaya anda boleh membuat keputusan sendiri sama ada ingin menyertai projek tersebut ataupun tidak. Ada boleh tanya soalan sekiranya tidak faham. Anda boleh memilih sama ada untuk menyertai projek tersebut, atau tidak. Anda juga boleh mengambil lebih masa sebelum membuat keputusan.

Projek ini mengenai apa? (*Terangkan secara ringkas mengenai projek)*

Projek ini bertujuan untuk membangunkan satu model berasaskan komputer dengan mengaplikasikan konsep *deep learning* dalam pengesanan masalah penulisan dalam kalangan kanak-kanak disleksia (disgrafia). Seterusnya, model ini akan diukur dari segi tahap prestasinya dengan menggunakan data-data (input) sebenar. Oleh itu, proses pembangunan model pengesanan disgrafia ini memerlukan data-data (gambar tulisan tangan kanak-kanak disleksia) untuk dijadikan input dan mempunyai ketepatan yang tinggi bagi proses klasifikasi tahap masalah disgrafia.

Mengapa saya patut menyertai projek ini? (*Terangkan secara ringkas tujuan projek)*

Projek ini merupakan satu projek yang dibangunkan untuk membantu dalam pengesanan masalah penulisan tangan kanak-kanak. Penyertaan anda dalam sesi pengumpulan data ini dapat membantu pihak penyelidik dalam membangunkan model yang berprestasi tinggi dalam pengesanan tahap kesukaran penulisan tangan dalam kalangan kanak-kanak.

Apa yang perlu saya lakukan dalam projek ini? (*Terangkan secara ringkas kepada kanak-kanak ini peranannya dalam projek)*

Anda perlu mengikut tugasan yang diberikan dengan menulis semua perkataan yang diberikan secara tulisan dan juga lisan. Tulisan tangan anda akan dikumpulkan oleh penyelidik dan digunakan sebagai input dalam proses pembangunan model sistem ini.

Apa yang akan berlaku kepada saya dalam projek ini? (*Terangkan secara ringkas dan mengenai risiko)*

Anda hanya perlu duduk dan mengikut arahan yang diberikan dengan menulis perkataan atau ayat yang diberikan dalam masa yang telah ditetapkan.

Adakah saya perlu menyertai projek ini?
Anda tidak perlu menyertai projek ini jika anda tidak mahu. Sekiranya anda menyertai projek ini, anda boleh berhenti pada bila-bila masa tanpa membuat sesiapa marah. Jika anda ingin menyertai projek ini, sila tulis nama anda di bawah. Sila pastikan bahawa anda faham apa yang telah dijelaskan kepada anda.

Dengan siapa boleh saya bercakap mengenai projek ini?
Jika anda mempunyai apa-apa soalan, anda boleh menghubungi saya:

Nama Penyelidik: Siti Azura Ramlan

Nombor Telefon:

Adakah orang lain akan tahu tentang apa yang saya katakan atau lakukan dalam projek ini? *Terangkan secara ringkas mengenai kerahsiaan penyertaan penyelidikan*

Tiada orang lain yang akan tahu tentang apa yang anda katakan atau lakukan dalam projek ini kecuali penyelidik. Data yang diperolehi adalah bersifat rahsia.

Soalan Persetujuan:
Arahan kepada kanak-kanak: Sila bulatkan jawapan anda di bawah.

1.	Sudahkah projek ini dijelaskan kepada anda?	(Ya)/Tidak
2.	Adakah anda memahami projek ini?	(Ya)/Tidak
3.	Adakah anda mempunyai apa-apa soalan mengenai projek ini?	Ya/(Tidak)
4.	Jika anda telah bertanya, adakah anda memahami jawapannya?	(Ya)/Tidak
5.	Adakah anda faham bahawa anda boleh berhenti daripada projek ini pada bila-bila masa?	(Ya)/Tidak
6.	Adakah anda selesa untuk mengambil bahagian dalam projek ini?	(Ya)/Tidak
7.	Adakah anda selesa sekiranya suara anda direkodkan?	(Ya)/Tidak
8.	Adakah anda selesa sekiranya berada dalam video?	(Ya)/Tidak
9.	Adakah anda selesa sekiranya gambar anda diambil?	(Ya)/Tidak

(Sekiranya kanak-kanak tersebut tidak dapat membaca, cap ibu jari hendaklah diambil, sebagai ganti tandatangan)

Jika anda ingin mengambil bahagian, sila tulis nama dan turunkan tandatangan, atau letakkan cap ibu jari anda di dalam kotak.

Nama peserta [redacted]

Tandatangan

Tarikh 15/11/2021



Nama pengambil persetujuan

Tandatangan

NOR BIBI MASLINA JUSOH
Managing Officer, Training Officer,
Speaker & Dyslexia Assessor For North
Branch Under Dyslexia Malaysia Associations

Tarikh 15/11/2021

(Sekiranya peserta projek tidak boleh membaca, atau sekiranya penyelidikan membabitkan isu-isu sensitif, seorang saksi hendaklah menjadi saksi ketika persetujuan diambil. Saksi tersebut perlu melengkapkan bahagian di bawah)

Nama saksi ANUSHA A/P S RAVINDRAN

Tandatangan

Tarikh 15/11/2021

APPENDIX 6

Example of Data Acquisition Form

PH 9

SESI PENGUMPULAN DATA

Versi Bahasa Melayu

Tugasan ini disediakan untuk mengumpul imej tulisan kanak-kanak disgrafia atau kanak-kanak disleksia yang mempunyai masalah disgrafia.

BAHAGIAN A: Maklumat Pengambilan

Tanda √ untuk kotak yang berkaitan.

1. Jantina:
   - [x] Lelaki
   - [ ] Perempuan
2. Umur: 8
3. *Menulis dengan tangan: (kiri)/kanan
4. Tempoh Intervensi: 1 tahun ~~bulan~~
5. Keputusan diagnosis:
   - [ ] Risiko Rendah
   - [ ] Risiko Sederhana
   - [x] Risiko Tinggi

Evaluated by :
Name :
Signature :
NOR BIBI MASLINA JUSOH
Managing Officer, Training Officer,
Speaker & Dyslexia Assessor For North
Branch Under Dyslexia Malaysia Associati

Evaluated by :
Name :
Signature :

Noranimuty Hassan
PPM - 017 - 14 - 23031995
PERSATUAN DYSLEXIA MALAYSIA
NO. 73, JALAN PRESINT 9/3,
DARUL AMAN PERDANA,
06000 SUNGAI PETANI, KEDAH
013-204 8142 / 013-239 2113

*potong yang tidak berkaitan

BAHAGIAN B

Tugas 1: Komposisi ayat

Pelajar diberikan ayat berikut dan menulis semula di ruang yang disediakan.

Baju itu baru dibeli oleh emak.

Baju it baru dibeli oleh emak.

Burung itu berada di dalam sangkar. Sangkar baru itu milik bapa.

Burung itu ada berada di sangkar
sangkar baru itu milik
bapa.

10

Saya suka bermain basikal pada waktu lapang. Basikal itu dibeli oleh ayah saya. Roda basikal saya berwarna hijau dan putih.

saya suka bermain basikal pada waktu
saya berwarna hijau dan putih
berwarna hijau dan itu dibeli
oleh ayah saya roda basikal berwarna

1 2 3 4 5 6 7 8 9 10 11

*potong yang tidak berkaitan

APPENDIX 7

Expert Evaluation Form for Validating Model

Expert Evaluation for Dysgraphia Classification System using Handwriting Images (Penilaian Pakar bagi Sistem Pengelasan Disgrafia menggunakan Imej Tulisan Tangan)

Assalamualaikum and a very good day,

Thank you for your willingness to serve as an evaluator to classify the severity of dysgraphia based on the provided images for validation purposes. You are invited to participate as a subject matter expert in evaluating an automated image classification system. The purpose of this study is to compare the classification results obtained from the system with those determined by domain experts, such as yourself.

Below are the details regarding the validation process.

1. You will be asked to evaluate an images and classify each into one of the predefined categories (high, moderate and low risk of dysgraphia). Your responses will be used solely for research validation purposes.
2. All data collected will remain confidential and will be used for academic purposes only. No personally identifiable information will be published.
3. Your participation is voluntary. You may withdraw at any time without consequences.
4. By completing and submitting the evaluation form, you agree to participate in this study voluntarily.

****

Assalamualaikum dan salam sejahtera,

Terima kasih atas kesudian anda untuk menjadi penilai dalam mengklasifikasikan tahap keterukan disgrafia berdasarkan imej yang diberikan bagi tujuan pengesahan. Anda dijemput untuk mengambil bahagian sebagai pakar bidang dalam menilai sistem pengelasan imej secara automatik. Tujuan kajian ini adalah untuk membandingkan keputusan pengelasan yang diperoleh daripada sistem dengan yang ditentukan oleh pakar bidang seperti anda.

Berikut adalah butiran berkaitan proses pengesahan:

1. *Anda akan diminta untuk menilai imej dan mengklasifikasikan setiap satu*

kepada salah satu kategori yang telah ditetapkan iaitu risiko disgrafia tinggi, sederhana, dan rendah. Maklum balas anda akan digunakan semata-mata untuk tujuan pengesahan kajian.

2. *Semua data yang dikumpulkan akan dirahsiakan dan hanya digunakan untuk tujuan akademik. Tiada maklumat peribadi yang boleh dikenalpasti akan diterbitkan.*
3. *Penyertaan anda adalah secara sukarela. Anda boleh menarik diri pada bila-bila masa tanpa sebarang implikasi.*
4. *Dengan melengkapkan dan menghantar borang penilaian, anda bersetuju untuk menyertai kajian ini secara sukarela.*

Expert Details (Maklumat Pakar Penilai)

Please fill in all fields, and your data will be kept confidential.

(Sila lengkapkan semua ruangan, dan data anda akan dirahsiakan)

Name (*Nama*)*

Affiliation (*Organisasi/Institusi*)

Years of Experience in Dyslexia or Dysgraphia Assessment (*Tahun Pengalaman dalam Penilaian Disleksia atau Disgrafia*)

Area of Expertise (*Bidang Kepakaran*)

Contact No (for honorarium purposes)

No. Telefon (untuk tujuan honorarium)

- You will be given 30 images. The image should be contained the statement as below:

1. **Baju itu baru dibeli oleh emak.**
2. **Burung itu berada di dalam sangkar. Sangkar baru itu milik bapa.**
3. **Saya suka bermain basikal pada waktu lapang. Basikal itu dibeli oleh ayah saya. Roda basikal saya berwarna hijau dan putih.**

- Then, select the appropriate class based on your expert judgment and assist by followed criteria for each class.
- The classes are:
    - **High Risk - Errors occur in more than 80% of the writing**
    - **Moderate Risk - Errors occur in more than 50% but less than 80% of the writing**

- **Low Risk - Errors occur in less than 50% of the writing**

The errors are based on the dysgraphia handwriting characteristics as follows:

1. **Incorrect Spelling and Capitalization**
2. **Inappropriate Letter Size and Spacing**
3. **Mixing Cursive and Print Letters**
4. **Omission of Letters or Words**
5. **Poor Letter Formation**
6. **Difficulty Using Text Spacing**
7. **Difficulty Discriminating Letter Shapes and Spacing**
8. **Problems Arranging Words from Left to Right**
9. **Failure to Stay Within the Margin**
10. **Incomplete Sentence Construction**

****

Anda akan diberikan 30 imej. Setiap imej sepatutnya mengandungi ayat seperti di bawah:

1. ***Baju itu baru dibeli oleh emak.***
2. ***Burung itu berada di dalam sangkar. Sangkar baru itu milik bapa.***
3. ***Saya suka bermain basikal pada waktu lapang. Basikal itu dibeli oleh ayah saya. Roda basikal saya berwarna hijau dan putih.***

Sila pilih kelas yang sesuai berdasarkan penilaian profesional anda dan rujuk kriteria bagi setiap kelas seperti berikut:

Kategori Penilaian:

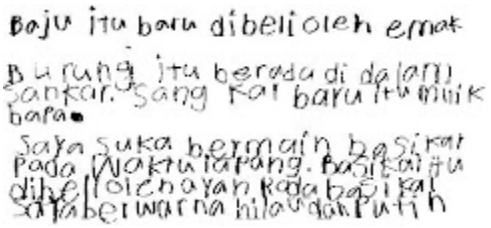
- ***Risiko Tinggi*** *– Ralat berlaku pada lebih 80% daripada keseluruhan tulisan.*
- ***Risiko Sederhana*** *– Ralat berlaku pada lebih 50% tetapi kurang daripada 80% daripada keseluruhan tulisan.*
- ***Risiko Rendah*** *– Ralat berlaku pada kurang daripada 50% daripada keseluruhan tulisan.*

Jenis ralat adalah berdasarkan ciri-ciri tulisan tangan yang berkaitan dengan disgrafia seperti berikut:

1. *Ejaan dan penggunaan huruf besar yang tidak betul*
2. *Saiz dan jarak huruf yang tidak sesuai*
3. *Campuran huruf sambung (cursive) dan huruf cetak (print)*
4. *Pengguguran huruf atau perkataan*
5. *Pembentukan huruf yang lemah*

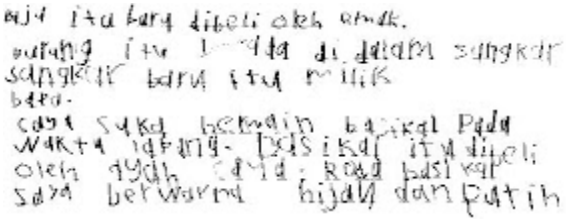
6. *Kesukaran menggunakan ruang teks dengan betul*
7. *Kesukaran membezakan bentuk dan jarak huruf*
8. *Masalah menyusun perkataan dari kiri ke kanan*
9. *Gagal menulis dalam ruang margin yang ditetapkan*
10. *Pembinaan ayat yang tidak lengkap*

Image 1*



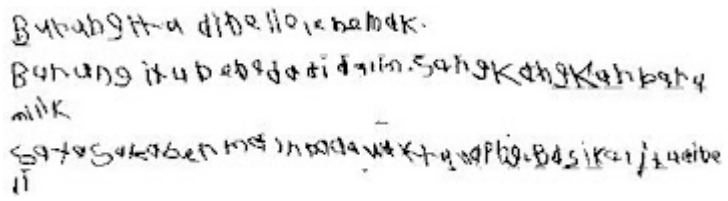
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 2*



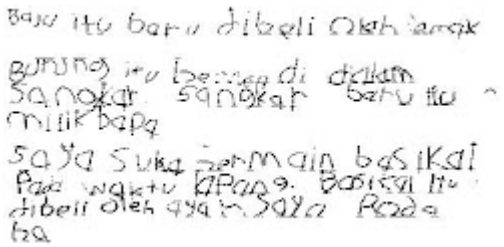
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 3*



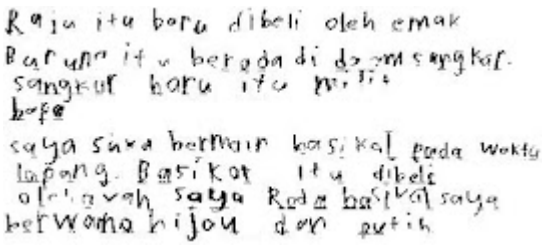
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 4*



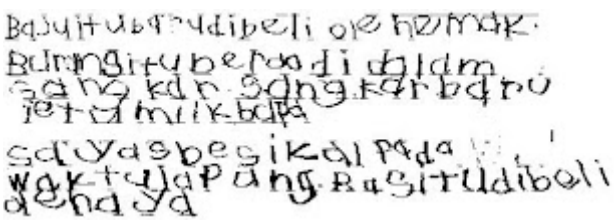
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 5*



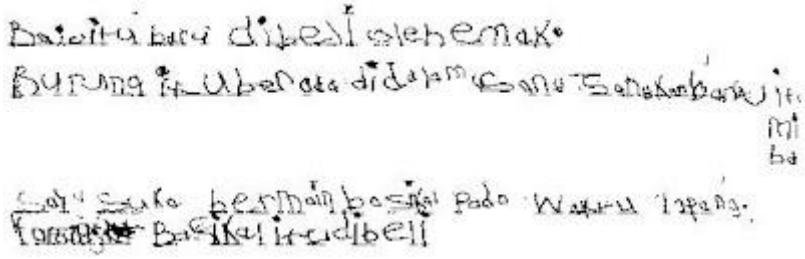
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 6*



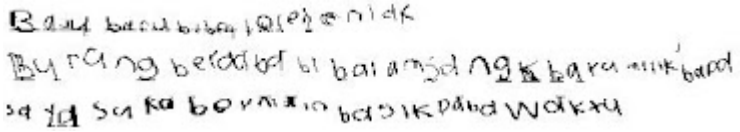
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 7*



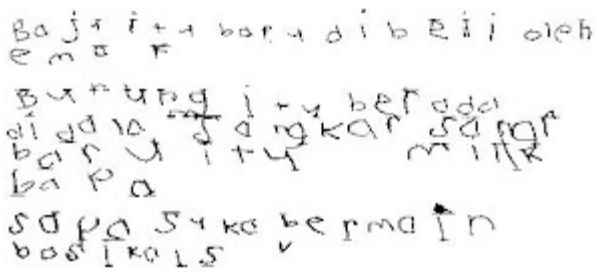
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 8*



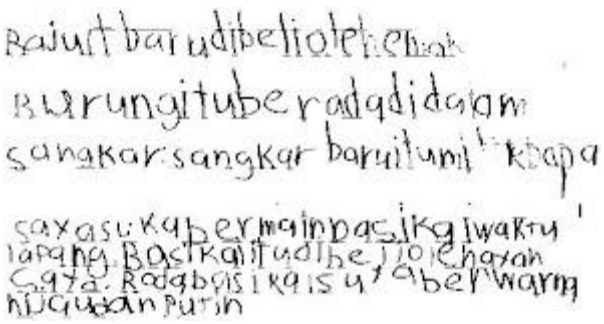
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 9*



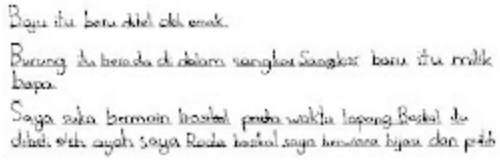
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 10*



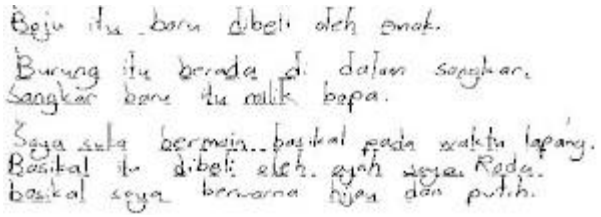
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 11*



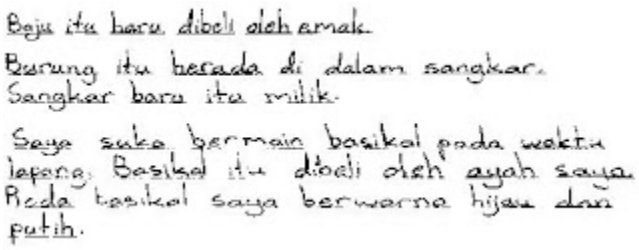
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 12*



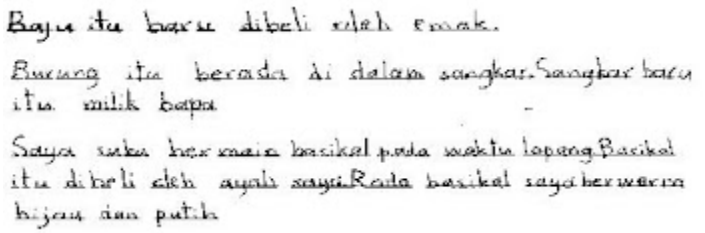
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 13*



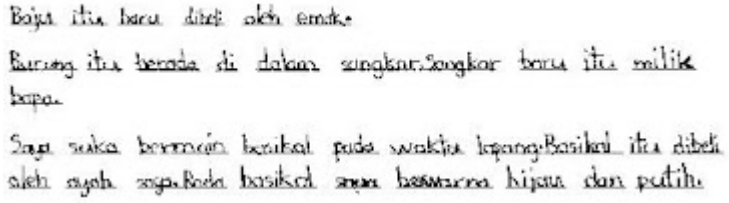
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 14*



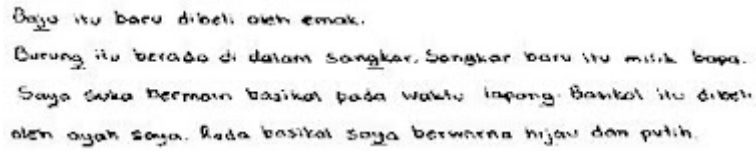
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 15*



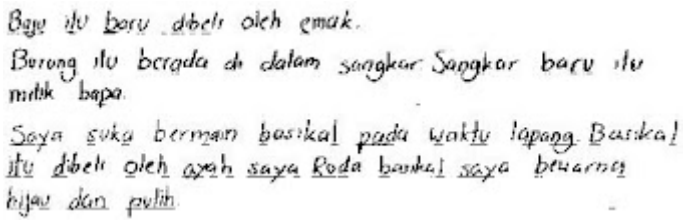
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 16*


Baju itu baru dibeli oleh emak.
Burung itu berada di dalam sangkar. Sangkar baru itu milik bapa.
Saya suka bermain basikal pada waktu lapang. Basikal itu dibeli oleh ayah saya. Roda basikal saya berwarna hijau dan putih.

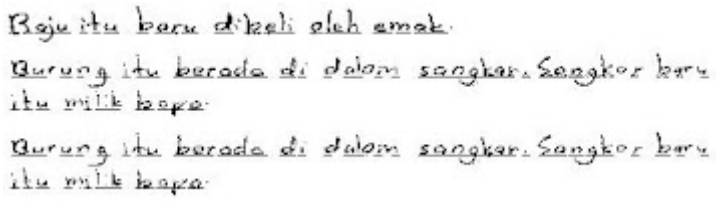
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 17*


Baju itu baru dibeli oleh emak.
Burung itu berada di dalam sangkar. Sangkar baru itu milik bapa.
Saya suka bermain basikal pada waktu lapang. Basikal itu dibeli oleh ayah saya. Roda basikal saya berwarna hijau dan putih.

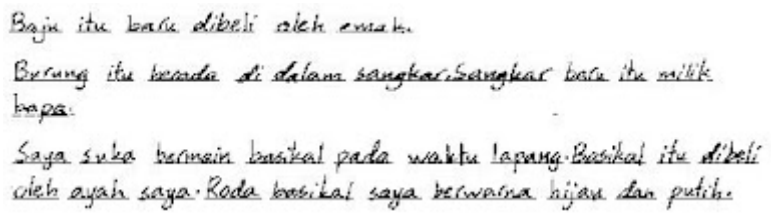
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 18*


Baju itu baru dibeli oleh emak.
Burung itu berada di dalam sangkar. Sangkar baru itu milik bapa.
Burung itu berada di dalam sangkar. Sangkar baru itu milik bapa.

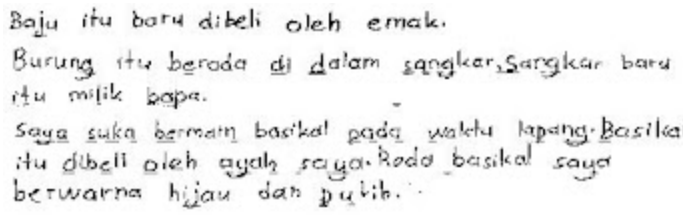
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 19*


Baju itu baru dibeli oleh emak.
Burung itu berada di dalam sangkar. Sangkar baru itu milik bapa.
Saya suka bermain basikal pada waktu lapang. Basikal itu dibeli oleh ayah saya. Roda basikal saya berwarna hijau dan putih.

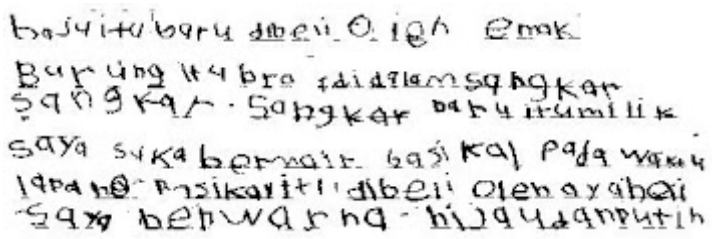
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 20*



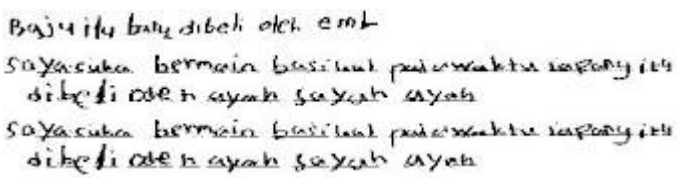
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 21*



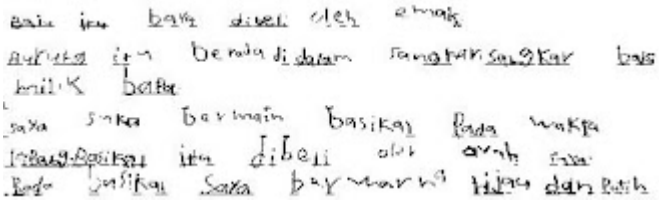
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 22*



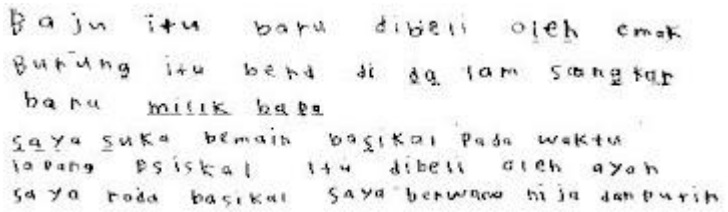
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 23*



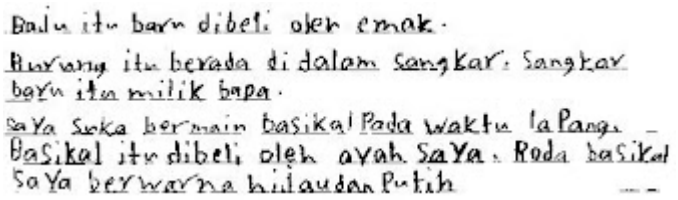
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 24*



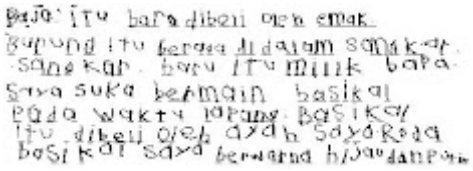
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 25*



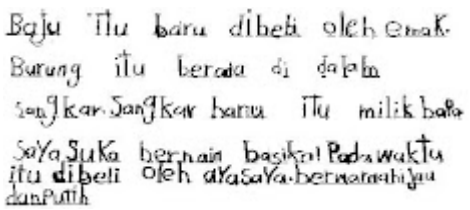
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 26*



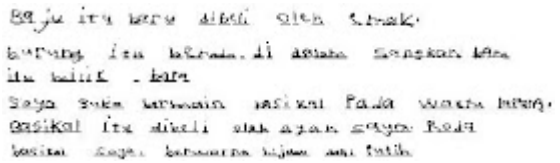
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 27*



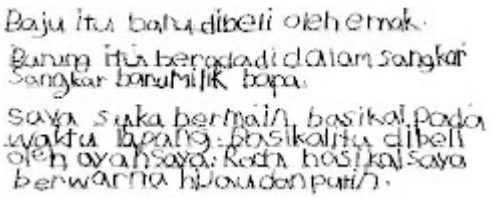
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 28*



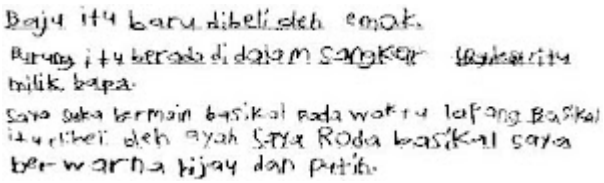
- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 29*



- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

Image 30*



- High Risk (Risiko Tinggi)
- Moderate Risk (Risiko Sederhana)
- Low Risk (Risiko Rendah)

APPENDIX 8

List of Experts

Name (Nama)	Affiliation (Organisasi/Institusi)	Years of Experience in Dyslexia or Dysgraphia Assessment (Tahun Pengalaman dalam Penilaian Disleksia atau Disgrafia)	Area of Expertise (Bidang Kepakaran)
Khairul Anwar Ishak	Persatuan Dyslexia Malaysia	8 tahun	Bahasa Melayu dan Matematik (Special needs)
Muhamad Zulkhairi Irwan bin Abd Rahman	Persatuan Dyslexia Malaysia	1 tahun	Special Needs
Hetla Pereira	Persatuan Dyslexia Malaysia	15 tahun	Special Need Teacher (Dyslexic & ADHD & mild Autism
Mia Ghazali	Inkubator Disleksia UUM	5 tahun	Pembantu Guru
Mera	Persatuan Dyslexia Malaysia	1 tahun	Mengajar kanak-2 berkeperluan khas terutama sekali dyslexia
Nur Syazana Abdul Salam	PDM	5 tahun	Mengajar
MOhamad Fadhil bin Zainudin	Persatuan Dyslexia Malaysia	1 tahun	Pengajar
Nurain Asyikin	Persatuan Disleksia Malaysia	3 tahun	Disleksia

AUTHOR'S PROFILE



Siti Azura Ramlan holds a Bachelor's degree in Computer Engineering and a Master of Science in Computer Science, with a specialization in research focused on Human-Computer Interaction, both obtained from Universiti Teknologi Malaysia (UTM), Skudai, Johor. She is currently pursuing a Doctor of Philosophy (PhD) at Universiti Teknologi MARA (UiTM), Malaysia. Her academic background reflects a strong foundation in both engineering and computer science, with research interests encompassing human-computer interaction, computer vision, image processing, and artificial intelligence applications. At the doctoral level, her work focuses on developing intelligent systems to detect handwriting difficulties (dysgraphia) and to support early childhood education and learning disability screening.

LIST OF PUBLICATION:

Ramlan, S. A., Isa, I. S., Ismail, A. P., Osman, M. K., & Che Soh, Z. H. (2024). Development of potential dysgraphia handwriting dataset. *Data in Brief*, *54*. https://doi.org/10.17632/39hr8dx76p.1

Ramlan, S. A., Isa, I. S., Jiad, N. N. A., Ismail, A. P., Osman, M. K., & Soh, Z. H. C. (2023). Comparison of deep learning model performance for handwritten character recognition of schoolchildren. In *Intelligent Multimedia Signal Processing for Smart Ecosystems* (pp. 327–350). Springer International Publishing. https://doi.org/10.1007/978-3-031-34873-0_14

Ramlan, S. A., Isa, I. S., Muttalib, A. A. A., Ismail, A. P., Osman, M. K., & Soh, Z. H. C. (2026). *A New Scheme of Ruling Line Removal Using RGBB Based Method for Potential Dysgraphia Handwriting Images*. 328–333. https://doi.org/10.1109/ic3ina68387.2025.11325687

Ramlan, S. A., Isa, I. S., Osman, M. K., Ismail, A. P., & Che Soh, Z. H. (2022). Investigating The Impact of CNN Layers on Dysgraphia Handwriting Image Classification Performance. *JOURNAL OF ELECTRICAL AND ELECTRONIC SYSTEMS RESEARCH*, *21*, 73–83. https://doi.org/https://doi.org/10.24191/jeesr.v21i1.010

Ramlan, S. A., Isa, I. S., Osman, M. K., Ismail, A. P., & Che Soh, Z. H. (2024). Comparing CNN-based Architectures for Dysgraphia Handwriting Classification Performance. *Pertanika Journal of Science and Technology*, *32*(5), 2013–2032. https://doi.org/10.47836/pjst.32.5.05

Ramlan, S. A., Isa, I. S., Osman, M. K., Ismail, A. P., Hisham, Z., & Soh, C. (2023). Investigation of Learning Rate for Directed Acyclic Graph Network Performance on Dysgraphia Handwriting Classification. *Investigation of Learning Rate for Directed Acyclic Graph Network Performance on Dysgraphia Handwriting Classification*, (March), 3–4. https://doi.org/10.1109/ICCSCE58721.2023.10237096

Ramlan, S. A., Isa, I. S., Osman, M. K., Ismail, A. P., & Soh, Z. H. C. (2023). Classification Performance of Directed Acyclic Graph Network on Potential Dyslexia Handwriting Images. *Proceedings - 13th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2023*, 149–154. https://doi.org/10.1109/ICCSCE58721.2023.10237096

Ramlan, S. A., Isa, I. S., Yogarajah, P., Osman, M. K., Ismail, A. P., & Soh, Z. H. C. (2026). H2FCD: Fusion of Offline Handwritten Image Features and Histogram Features for Multiclass Classification of Dysgraphia Severity. *IET Image Processing*, *20*(1). https://doi.org/10.1049/ipr2.70285

DATASET PUBLICATION:

1. Siti Azura Ramlan. (2023). Potential Dysgraphia Handwriting Dataset of School-Age Children. In Mendeley Data (Vol. V1).

RESEARCH GRANT:

1. Fundamental Research Grant Scheme (FRGS) - A Formulation on Optimum Resnet-CNN Layer Architecture Based on Dilated Method for Dysgraphia Severity Classification (RM71,300) -, Ministry of Higher Education, Malaysia (17 Sept 2021 - 16 Sept 2023)
2. International Research Grant (INT) – Study and Analysis on Dysgraphia Severity Symptom Based on Chinese Students Handwritten in Primary Schools (RM10,000) – Universitas Telkom Indonesia (01 Jun 2025 – 31 Mei 2027)

INNOVATION, INVENTION AND DESIGN EXHIBITION ACHIEVEMENT:

1. Potential dyslexia detection for primary school children using deep learning approach (2023) [Silver award]. Penang International Invention, Innovation, and Design (PIID 2023).
2. CNN-based model for potential dyslexia detection based on automated handwriting features extraction (2023) [Gold award]. International Exhibition & Symposium on Productivity, Innovation, Knowledge, Education & Design (I-Spike 2023).
3. Dysgraphia severity classification using support vector machine (SVM) based on automated handwriting image feature extraction (2024) [Gold award]. Melaka International Intellectual Exposition (MIIEX 2024).
4. Dysgraphia severity detection among school-age children using machine learning methods (2024) [Gold award]. International Innovation Competition on Languages, Education, and Societies (IINNCOME 2024).
5. Dysdetectv1: Dyslexia-dysgraphia severity classification using handwriting image and machine learning model (2025) [Gold award]. The Fourth International Competition on Sustainable Education (SUSED 2025).

INTELLECTUAL PROPERTY - COPYRIGHT:

1. H2FCD: Fusion Handwriting and Histogram Feature Vector for Multiclass Classification of Dysgraphia Severity – Notification Number: CRLY2025P05943, Intellectual Property Corporation of Malaysia.
2. Line Removal Techniques Using RGBB-Based Method for Dysgraphia Handwriting Images – Notification Number: CRLY2024P03329, Intellectual Property Corporation of Malaysia.
3. Dysgraphia Severity Classification Using CNN Model and Image Processing Techniques – Notification Number: CRLY2022P05504, Intellectual Property Corporation of Malaysia.

MEMORANDUM OF AGREEMENT

1. Memorandum of Agreement Between Universiti Teknologi Mara and Universitas Telkom. Project Title: Study and Analysis on Dysgraphia Severity Symptom based on Chinese Students Handwritten in Primary Schools. 2024-2026