School of Computer Sciences

CDS590 – Consultancy Project & Practicum

Final Report

Enhancing Handwriting Images Dataset using GAN-based Augmentation of CNN Models for Classification Dyslexia Severity

Ahmed Adel Sanhan Al-Haidary
P-COM0113/22

Supervisor: Umi Kalsom Binti Yusof

Mentor: Dr. Iza Sazanita Isa

Practicum place: Persatuan Dyslexia Malaysia (PDM)

SEM 1 2023/2024

DECLARATION

"I declare that the following is my own work and does not contain any ***unacknowledged*** work from any other sources. This project was undertaken to fulfill the requirements of the Consultancy Project & Practicum for the Master of Science (Data Science and Analytics) program at Universiti Sains Malaysia".

Signature	:	Ahmed.
Name	:	Ahmed Adel Sanhan Al-Haidary
Date	:	31/1/2024

ACKNOWLEDGEMENTS

First and foremost, I extend my sincere gratitude to Allah the Almighty, the Most Gracious, and the Most Merciful, for bestowing His blessings upon me throughout my academic journey and for guiding me toward the successful completion of this project. May His blessings be upon His final Prophet Muhammad (peace be upon him), his noble family, and esteemed companions.

I am deeply indebted to my project supervisor, Associate Professor Dr. Umi Kalsom Binti Yusof, for her invaluable guidance, unwavering support, and exceptional mentorship. Her expertise, encouragement, and insights have played a pivotal role in shaping my understanding of project management, enhancing my analytical skills, and refining the quality of my work. I am profoundly grateful for her dedication and commitment to my academic development.

I extend my heartfelt appreciation to my mentor, Dr. Iza Sazanita Isa, for her continuous encouragement, insightful feedback, and belief in my capabilities. Her mentorship has been instrumental in my personal and professional growth, instilling in me the confidence to overcome challenges and pursue excellence in my endeavors. I am truly honored to have had the privilege of learning from her wisdom and expertise.

I wish to express my deepest gratitude to my beloved Mother, whose unwavering support and encouragement have been my source of strength and inspiration throughout my Master's study. Her love, sacrifices, and guidance have been instrumental in shaping my character and academic journey. Additionally, I extend my special thanks to my father for his enduring support and encouragement. Their selfless sacrifices and unwavering belief in my potential have been the cornerstone of my academic achievements.

Finally, I am grateful to my family, friends, and all those who have supported and encouraged me along this journey. Their encouragement, understanding, and unwavering support have been invaluable to me, and I am deeply appreciative of their presence in my life.

ABSTRACT

Dyslexia, a complex neurodevelopmental disorder, poses formidable challenges to individuals in acquiring proficient reading skills despite adequate instruction and cognitive abilities, with its accurate diagnosis remaining elusive due to its multifaceted nature and symptom overlap with other learning disorders. This study endeavors to develop an automated dyslexia diagnosis system leveraging dyslexic handwriting images and cutting-edge deep-learning methodologies, focusing on implementing advanced techniques, notably Convolutional Neural Networks (CNNs), to construct a robust dyslexia diagnosis model. To counteract data imbalance challenges inherent in dyslexic handwriting datasets, data augmentation utilizing Generative Adversarial Networks (GANs) will be integrated to enhance model accuracy and generalization capability, promising more precise and timely dyslexia identification and equipping individuals with essential resources to confront reading impediments effectively. The project's outcomes encompass a meticulously trained CNN-based model adept at accurately categorizing dyslexia severity levels. Experimental findings demonstrate the successful implementation of automated dyslexia severity classification using advanced deep learning techniques and GAN-based data augmentation, wherein the LeNet-5 model, trained with an augmented dataset enriched with additional words generated by GANs, outperformed both the original DenseNet201 model and the LeNet-5 model trained solely with original and generated words within the original dataset's existing classes without introducing new words, achieving an impressive F1 score, recall, precision, and accuracy of 91%, underscoring the efficacy of data augmentation techniques in enhancing the accuracy and robustness of CNN models for dyslexia diagnosis and classification.

Keywords: Dyslexia, , Data Augmentation, Deep Learning, CNN, Transfer Learning.

ABSTRAK

Disleksia, satu gangguan neuroperkembangan yang kompleks, memberikan cabaran yang besar kepada individu dalam memperoleh kemahiran membaca yang mahir walaupun dengan arahan yang mencukupi dan keupayaan kognitif, dengan diagnosis yang tepat sukar didapati kerana sifat yang kompleks dan simptom yang bertindak balas dengan gangguan pembelajaran lain. Kajian ini bertujuan untuk membangunkan sistem diagnosis disleksia automatik dengan menggunakan imej tulisan tangan disleksia dan metodologi pembelajaran mendalam yang terkini, dengan tumpuan kepada pelaksanaan teknik-teknik terkini, terutamanya Rangkaian Neural Konvolusi (CNN), untuk membina model diagnosis disleksia yang kukuh. Bagi menangani cabaran ketidakseimbangan data yang ada dalam dataset tulisan tangan disleksia, pembesaran data menggunakan Rangkaian Generatif Antagonis (GANs) akan disatukan untuk meningkatkan ketepatan model dan keupayaan pengamiran, menjanjikan pengenalpastian disleksia yang lebih tepat dan tepat pada masanya serta memperlengkapi individu dengan sumber-sumber penting untuk menghadapi halangan dalam membaca dengan berkesan. Hasil projek merangkumi model CNN yang terlatih dengan teliti dan mahir dalam mengkategorikan tahap keparahan disleksia secara tepat. Penemuan eksperimen menunjukkan pelaksanaan berjaya pengelasan keparahan disleksia secara automatik menggunakan teknik-teknik pembelajaran mendalam terkini dan pembesaran data berasaskan GAN, di mana model LeNet-5, yang dilatih dengan dataset yang ditambah dengan perkataan tambahan yang dijana oleh GANs, mengungguli kedua-dua model asal DenseNet201 dan model LeNet-5 yang dilatih hanya dengan perkataan asal dan yang dijana dalam kelas-kelas asal dataset tanpa memperkenalkan perkataan baru, mencapai skor F1, panggilan semula, ketepatan, dan ketepatan yang mengagumkan sebanyak 91%, menegaskan keberkesanan teknik pembesaran data dalam meningkatkan ketepatan dan kekerasan model CNN untuk diagnosis dan pengelasan disleksia.

Kata Kunci: Disleksia, Pembesaran Data, Pembelajaran Mendalam, CNN, Pembelajaran Pemindahan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

PDM	- Persatuan Dyslexia Malaysia
ISIS	- Rotary Club of Gombak and the Institute of Strategic and International Studies
GANs	- Generative Adversarial Networks
OCR	- Optical Character Recognition
CNN	- Convolutional Neural Network
SVM	- Support Vector Machine
MLP	- Multilayer Perceptron
HDF5	- Multilayer Perceptron
FID	- Fréchet Inception Distance
KID	- Kernel Inception Distance
HWD	-Handwriting Distance
HDF5	- Hierarchical Data Format, version 5
HiGAN+	- Handwriting Imitation Generative Adversarial Network

CHAPTER 1

INTRODUCTION

1.1. Background

Persatuan Dyslexia Malaysia (PDM) is a non-profit organization established in 1995 to advance the education and welfare of individuals with dyslexia and specific learning disabilities in Malaysia. It originated from a successful public seminar on dyslexia awareness organized in 1993 by the Rotary Club of Gombak and the Institute of Strategic and International Studies (ISIS). Following the seminar, PDM was officially registered with the Registrar of Societies Wilayah Persekutuan(Persatuan Dyslexia Malaysia, 2023).

PDM actively engages in various initiatives to support individuals with dyslexia. The organization conducts seminars for teachers and parents, offers counseling services, and provides special classes for dyslexic children. These efforts contribute to raising awareness, facilitating early diagnosis and specialized treatment, and empowering individuals with dyslexia to overcome their learning challenges. Additionally, PDM advocates for dyslexia-related training in teacher education programs and aims to establish remedial teacher training courses and services in areas where resources are lacking(Persatuan Dyslexia Malaysia, 2023). By fostering collaboration with kindred associations and individuals interested in specific learning disabilities, PDM aims to exchange knowledge, share best practices, and collectively address the challenges faced by individuals with dyslexia, ultimately working towards creating a more inclusive society(Persatuan Dyslexia Malaysia, 2023).

Figure 1.1: Persatuan Dyslexia Malaysia Head Quarters(Persatuan Dyslexia Malaysia, 2023)

1.2. Problem Statement

Dyslexia, a neuro-developmental disorder characterized by difficulties in learning to read despite adequate instruction, intelligence, and socio-cultural background, poses significant challenges to affected individuals. As the most common type of learning disorder, dyslexia impact 4a child's academic achievement, specifically in the area of reading. The term dyslexia is often used interchangeably with developmental dyslexia and specific reading disability, indicating that the developmental progress is typical except for reading difficulties. The term itself, derived from Greek roots, reflects the inherent struggle individuals with dyslexia face in dealing with words(Wajuihian & Naidoo, 2012).

However, the precise identification and diagnosis of dyslexia remain problematic. The complexity of dyslexia, its overlapping symptoms with other learning disorders, and the lack of standardized diagnostic criteria contribute to challenges in accurately identifying and addressing dyslexia in individuals (Snowling et al., 2020). Furthermore, the impact of dyslexia extends beyond academic performance, affecting individuals' self-esteem, confidence, and overall well-being(Wajuihian & Naidoo, 2012). Handwriting-based screening tests have shown their efficiency for dyslexia diagnosis (Van Waelvelde et al., 2012). However, current manual diagnostic methods present significant drawbacks: they are expensive, expert-dependent, and time-consuming. These limitations hinder the widespread adoption and accessibility of dyslexia diagnosis, delaying intervention and support for affected individuals.

Therefore, there is a pressing need to develop effective assessment methods and interventions to better support individuals with dyslexia and provide them with the necessary tools to overcome their difficulties with words and reading.

The current dataset poses several challenges due to its small size and severe imbalance across dyslexia severity levels. There is a limited number of students detected for severe and moderate dyslexia, primarily attributed to conventional evaluation processes that are either too general or unable to identify potential cases accurately. Additionally, the conventional augmentation methods used in dataset preparation contribute to low CNN performance by introducing biases and inadequately representing the diversity of dyslexic handwriting patterns.

To address these challenges, it is essential to have an automated model to diagnose dyslexia accurately and efficiently. The purpose of this project is to develop such a model using dyslexic handwriting images, extracting dyslexic handwriting features, and using advanced deep-learning techniques. However, to ensure the model's accuracy and generalizability, data augmentation using advanced techniques, such as Generative Adversarial Networks (GANs), will be implemented. This technique will help address data imbalance issues and generate additional augmented data that is representative of the dyslexia handwriting patterns. By incorporating data augmentation, the automated model can overcome the subjectivity and limitations of current diagnostic methods, enabling more precise and timely identification of dyslexia in individuals.

1.3. Research Questions

- How can data augmentation techniques using Generative Adversarial Networks (GANs) be effectively implemented to address data imbalance in dyslexia handwriting images and improve the performance of the model?
- What deep learning techniques can be employed to develop an automated dyslexia diagnosis model using dyslexia handwriting images and extracted features?
- What is the impact of integrating data augmentation techniques with GANs and deep learning models on the accuracy and reliability of dyslexia diagnosis?
- How does the proposed automated dyslexia diagnosis model compare to existing methods in terms of accuracy, efficiency, and reliability?

1.4. Objectives of Project

The primary aim of this research is to enhance GAN-based augmentation by integrating the generation of synthetic dyslexia handwriting images. This augmentation aims to improve the performance of Convolutional Neural Networks (CNNs) in the automated detection of potential dyslexia risk among school children. To achieve this goal, the objectives are:

1- Implement data augmentation techniques with Generative Adversarial Networks (GANs) to address data limitation and imbalance.
2- Develop a convolutional neural networks (CNN) model for dyslexia diagnosis using dyslexia handwriting dataset.

1.5. Benefit of Project

One of the key contributions of this project lies in the collection and curation of dyslexia handwriting images from Persatuan Dyslexia Malaysia (PDM). This dataset will serve as a valuable resource, capturing the diverse characteristics and patterns of dyslexic handwriting. The availability of such a dataset will not only support the development of the automated model but also contribute to future research and advancements in dyslexia assessment and intervention.

This thesis represents a significant advancement in the detection of dyslexia risk, with the introduction of a transformative approach that is powered by Generative Adversarial Network (GAN)-based augmentation models. By addressing the challenges posed by imbalanced handwriting datasets, this research fundamentally redefines the landscape of dyslexia identification.

Through the integration of advanced GAN-based augmentation models, this project sets new benchmarks in precision and objectivity for dyslexia risk detection. By overcoming the limitations and subjectivity inherent in existing diagnostic methods, it establishes a robust and dependable system for identifying dyslexia with unparalleled accuracy.

The methods presented in this thesis are designed to speed up the identification of dyslexia, allowing organizations like Persatuan Dyslexia Malaysia (PDM) to efficiently assess a larger number of patients, facilitating the early detection of numerous dyslexia cases. This early identification allows PDM to promptly initiate treatment, significantly improving treatment efficiency and outcomes.

A pivotal contribution of this project is the creation of a comprehensive dataset comprising dyslexia handwriting images from Persatuan Dyslexia Malaysia (PDM). This rich resource encapsulates diverse dyslexic handwriting characteristics, serving as a cornerstone for automated model development and fostering future research in dyslexia assessment and intervention strategies.

The utilization of GAN-based augmentation models isn't confined solely to dyslexia; its adaptability extends to numerous medical domains facing similar data limitations. This pioneering approach lays the groundwork for a scalable solution applicable across varied medical fields, addressing data scarcity and imbalance issues in diagnostic practices.

CHAPTER 2

RELATED WORKS

2.1 Introduction

This section provides an overview of the past literature in the research domain of dyslexia recognition, with a focus on the utilization of data science and analytics techniques. The use of data science and analytics has emerged as a promising approach to improving dyslexia recognition and understanding. In this section, we explore previous studies that have employed various data science techniques and tools in the context of dyslexia recognition. The discussion contains the application of data science and analytics techniques, to facilitate accurate and efficient dyslexia identification.

2.2 Dyslexia

Dyslexia is a neuro-developmental disorder characterized by difficulties in reading, spelling, writing, and speaking, despite adequate instruction, intelligence, and socio-cultural background. It is one of the most common learning disorders, affecting an estimated 15-20% of the global population (International Dyslexia Association, 2021). Dyslexia can have significant negative consequences on academic achievement, self-esteem, and social-emotional development (Al-Barhamtoshy & Motaweh, 2017)

Early detection and intervention are crucial in addressing the challenges posed by dyslexia and mitigating its negative impact on individuals. Traditional diagnostic methods rely on costly assessments conducted by experts, which may not be accessible to all individuals in need. Therefore, there is a growing interest in developing automated and efficient diagnostic tools using advanced technologies and computational approaches.

Eye-tracking technology has emerged as a promising avenue for dyslexia recognition. Eye movement patterns during reading differ between individuals with dyslexia and typical readers. Dyslexic readers exhibit longer fixations, shorter saccade duration, and more regressions, indicating difficulties in decoding and recognizing printed words (Raatikainen et al., 2021). Machine learning methods have been applied to eye movement data, showing success in identifying dyslexic readers (Raatikainen et al., 2021).

Brain imaging studies have also contributed to understanding dyslexia at a neural level. Various studies have revealed subtle and spatially distributed variations in brain anatomy among individuals with dyslexia (Tamboer et al., 2016). Machine learning techniques, such as support vector machines, have been utilized to differentiate between individuals with and without dyslexia based on these neuroanatomical variations (Tamboer et al., 2016).

Additionally, the analysis of handwriting patterns has gained attention as a potential diagnostic tool for dyslexia. Researchers have developed computational models and classifiers to identify dyslexic handwriting based on various metrics(Isa et al., 2019; Spoon et al., 2019). Machine learning algorithms, including convolutional neural networks, have shown promising results in accurately classifying dyslexic handwriting (Isa et al., 2019; Zahia et al., 2020).

2.3 Dyslexia Classification Techniques

This section provides a systematic and comprehensive review of the relevant works that have utilized data science and analytics techniques for dyslexia detection. The aim is to examine the effectiveness and relevance of these techniques in improving dyslexia screening and diagnosis.

Pattern recognition techniques have been widely employed for dyslexia detection, particularly using handwriting images. Isa, (Isa et al., 2019) developed an automated detection system based on pattern recognition and optical character recognition (OCR) methods. The system achieved an accuracy of 73.33% by extracting features from handwriting images using OCR and utilizing machine learning algorithms for classification. Similarly, Rosli (Rosli et al., 2021a), utilized a convolutional neural network (CNN) for dyslexia handwriting recognition, achieving a remarkable accuracy of 95.34%. These studies highlight the potential of pattern recognition techniques in dyslexia detection.

Deep learning techniques, particularly CNNs, have shown remarkable performance in dyslexia detection. Asvestopoulou (Asvestopoulou et al., 2019) , developed DysLexML, a screening tool based on various ML algorithms, including CNNs, for analyzing eye movement data recorded during silent reading. The accuracy achieved by DysLexML was 97% and 84% using linear SVM, showcasing the effectiveness of deep learning in dyslexia detection. Alqahtani (Alqahtani et al., 2023) , employed deep learning models for dyslexia prediction using diverse datasets, achieving high accuracy rates ranging from 86.14% to 94%. These findings highlight the potential of deep learning techniques in improving dyslexia screening.

Machine learning techniques have also been utilized for dyslexia detection. Yogarajah and Bhushan (Yogarajah & Bhushan, 2020) conducted an automated diagnosis of dyslexia based on handwriting samples using a CNN. Their model achieved an accuracy of 75% using a dataset of children's handwriting samples. Spoon (Spoon et al., 2019), developed a system that employed computer vision techniques and machine learning algorithms to classify handwriting samples as indicative of dyslexia or not, achieving an accuracy of 77.6%. These studies demonstrate the potential of machine learning techniques in dyslexia detection.

The reviewed literature demonstrates the effectiveness of data science and analytics techniques in dyslexia detection. Pattern recognition techniques based on handwriting analysis have provided reasonable accuracy rates, although there is room for improvement. Deep learning techniques, particularly CNNs, have shown exceptional performance in dyslexia detection, with high accuracy rates reported across different datasets. Machine learning approaches have also yielded promising results, although they exhibit slightly lower accuracy rates compared to deep learning techniques.

Table 2.1: Classification Techniques for Dyslexia Diagnosis

Study	Year	Reference	Technique	Data Type	Accuracy
Isa et al. (n.d.)	2019	(Isa et al., 2019)	Pattern Recognition, OCR	Handwriting Images	73.33%
Rosli et al. (n.d.)	2021	(Rosli et al., 2021a)	Transfer Learning, LeNet-5	Handwriting Images	95.34%
Yogarajah and Bhushan (n.d.)	2020	(Yogarajah & Bhushan, 2020)	Deep Learning, CNN	Handwriting Images	86.14%
Alqahtani et al. (2023)	2023	(Alqahtani et al., 2023)	Deep Learning, CNN	Brain Images	77%
Asvestopoulou et al. (n.d.)	2019	(Asvestopoulou et al., 2019)	GANs, and SVM	Eye Movements	97%
Spoon et al. (n.d.)	2019	(Spoon et al., 2019)	CNN	Handwriting Images	77.6%

2.4 Machine Learning Techniques for Handwriting Classification

In the realm of data science and analytics, three research papers provide valuable insights into the performance and comparison of various techniques. In the first paper by Wang et al.(Wang et al., 2021), the authors compared traditional machine learning (SVM) and deep learning (CNN) algorithms for image classification. SVM achieved an accuracy of 0.88 on the large MNIST dataset and

0.86 on the small COREL1000 dataset, while CNN achieved higher accuracies of 0.98 and 0.83, respectively. This suggests that traditional machine learning performs better on small sample datasets, while deep learning excels on large sample datasets.

Moving to character recognition, the study by Ben Driss et al.(Driss et al., 2017) compared multilayer perceptron (MLP) and convolutional neural network (CNN) models. The real-time CNN outperformed MLP, exhibiting twice the relevance in character classification.

In the context of text classification, the study by Kamath et al. (Kamath et al., 2018) compared traditional machine learning techniques (Naive Bayes, Logistic Regression, Support Vector Machine, Random Forest Classifier, and MLP) with deep learning approaches (Convolutional Neural Network). Logistic regression achieved the best accuracy among traditional machine learning algorithms. However, CNN consistently outperformed all other techniques, exhibiting significantly higher accuracies. Notably, traditional machine learning algorithms showed better accuracy with raw data, while CNN performed better with processed data.

Synthesizing the findings, it is evident that deep learning techniques, particularly CNN, excel in complex tasks like image classification and character recognition. They consistently achieve higher accuracies and demonstrate better recognition capabilities. Traditional machine learning techniques have their strengths in handling small sample datasets and can be effective in certain contexts. However, CNN consistently outperforms them in various scenarios.

2.5 Deep Learning for Handwriting Classification

The domain of handwriting image classification has undergone a transformative journey, employing deep learning to address various challenges, including distinguishing between native and non-native Arabic handwriting, identifying dyslexia symptoms, and characterizing demographic traits.

Rahmanian and colleagues explored demographic classification based on handwriting patterns using advanced CNN architectures, including DenseNet201, InceptionV3, and Xception. These models showcased significant improvements in gender and handedness classifications. Particularly, DenseNet201 displayed an impressive accuracy of 84% for gender and 99.14% for handedness classification. The study highlighted the capabilities of deep CNN architectures in effectively extracting features and enhancing demographic classification accuracy (Rahmanian & Shayegan, 2021).

Models like GoogleNet, LeNet, and DenseNet201 present a compelling case for their integration into dyslexia risk detection frameworks. Their proven track record in accurately identifying distinct handwriting traits aligns well with the nuances associated with dyslexic handwriting, making them pivotal in enhancing dyslexia risk detection accuracy and sensitivity.

The study conducted by Isa and colleagues aimed at detecting dyslexia symptoms through handwriting analysis. They compared various CNN models, including CNN-1, CNN-2, CNN-3, and LeNet-5, highlighting their efficacy in identifying potential dyslexia indicators among school children. Leveraging data augmentation and preprocessing techniques, these models achieved impressive performances, exceeding 87% accuracy. This research accentuated the significance of CNN models in identifying dyslexia-related patterns in handwriting (Isa et al., 2021).

Rosli et al. focused on developing a transfer learning CNN model using LeNet-5 for dyslexia handwriting recognition. Their meticulously designed model achieved a remarkable accuracy of 95.34% in classifying three classes of dyslexic handwriting. This achievement emphasized the potential of transfer learning in effectively recognizing distinct dyslexia-related handwriting patterns. The use of data augmentation and fine-tuning hyperparameters notably contributed to the model's robustness and accuracy(Rosli et al., 2021b).

Almisreb and team evaluated seven deep transfer learning models to discern between native and non-native Arabic handwriting. Among the models, GoogleNet emerged as a standout performer, showcasing exceptional accuracy rates of 93.2% and 95.5% for native handwriting classification using normal and augmented datasets, respectively. This validated the robustness and efficacy of transfer learning, especially GoogleNet, in differentiating Arabic handwriting styles(Almisreb et al., 2022).

2.6 Data Augmentation for Deep Learning

Supervised learning faces challenges when dealing with imbalanced datasets, particularly in medical diagnostics like dyslexia detection. Class imbalance often leads to misclassification biases, where the minority class, such as patients with dyslexia, faces greater misclassification due to overemphasized representation of the majority healthy class (Johnson & Khoshgoftaar, 2019).

Traditional augmentation techniques designed for handwriting recognition have been extensively studied. Heil and Breznik explored 22 augmentation methods in handwritten stenography recognition. Their study revealed that small rotations, shifting, shearing, and scaling were suitable augmentations, while larger rotations and certain alterations adversely affected recognition performance (Heil & Breznik, 2023).

However, applying traditional augmentation methods in dyslexia detection presents challenges. These techniques, while effective in general recognition tasks, may inadvertently alter critical features essential for dyslexia risk assessment. Modifications such as rotations and scaling, while useful, might disrupt handwriting elements crucial for accurate dyslexia risk identification.

To address these challenges, the utilization of Generative Adversarial Networks (GANs) has emerged. Wang and Perez highlighted the effectiveness of GAN-based augmentation in image classification, emphasizing the preservation of essential features while expanding datasets. GANs offer a unique approach to generating handwriting samples that retain critical features relevant to dyslexia risk detection.

Leveraging GANs for data generation involves a min-max strategy that successfully produces counterfeit samples resembling the original data distribution. This approach holds promise in augmenting limited datasets, enhancing the ability to identify dyslexia-related patterns without compromising critical features(Perez & Wang, 2017)
.

2.7 GANs techniques for Handwriting Augmentation

There are several Gans that has been created to increase the amount of data for handwritten image such as : HiGAN can generate variable-length handwritten words/texts conditioned on arbitrary textual contents, which are unconstrained to any predefined corpus or out-of-vocabulary words. Moreover, HiGAN can flexibly control the handwriting styles of synthetic images by disentangling calligraphic styles from the reference samples. Experiments on handwriting benchmarks validate our superiority in terms of visual quality and scalability when comparing to the state-of-the-art methods for handwritten word/text synthesis.

The use of Generative Adversarial Networks (GANs) in generating realistic handwritten text images has gained significant traction due to its potential to augment datasets and improve text recognition systems. Several notable studies demonstrate the diverse capabilities and applications of GANs in synthesizing handwriting styles, conditioning text, and addressing challenges in data augmentation for text recognition tasks.

ScrabbleGAN introduced a semi-supervised approach for synthesizing handwritten text images of varying lengths. Leveraging a novel generative model, ScrabbleGAN demonstrated the ability to produce versatile word images, allowing control over text styles like cursive writing and pen stroke thickness (Fogel et al., 2020). This approach showcased improved performance over supervised handwritten text recognition systems by generating images across various lexicons and styles.

Eltay et al. proposed an adaptive data augmentation technique using GANs to address class imbalance issues in text recognition. By generating balanced augmented data, this approach improved text recognition accuracy on imbalanced handwritten Arabic text datasets, showcasing the effectiveness of GANs in dealing with such problems.

GANwriting introduced a method conditioning the generative process on calligraphic style features and textual content to produce realistic handwritten word images(Kang et al., 2020). This model's versatility allowed rendering any input word, mimicking calligraphic features in a few-shot setup, and demonstrated superior quality in generating realistic handwritten word images.

HiGAN+ presented a novel approach for handwriting imitation using GANs, focusing on disentangling representations to generate diverse and realistic handwritten texts (Gan et al., 2022). It emphasized the generation of handwriting images conditioned on arbitrary textual content and calligraphic styles, showcasing superior visual quality, style transferability, and scalability compared to existing GANs.

SLOGAN enriched handwriting training samples by synthesizing diverse styles based on latent vectors. Using a style bank, this method parameterized handwriting styles and manipulated them to generate new styles (Luo et al., 2022). SLOGAN demonstrated its superiority in generating diverse handwriting styles, improving recognizer training, and potentially aiding domain adaptation tasks.

Among these approaches, HiGAN+ stands out for its disentangled representation, scalability, and style transferability in generating diverse handwriting styles. For the current project, HiGAN+ presents a compelling choice due to its emphasis on a diverse and realistic generation of handwriting images, ideal for augmenting dyslexia handwriting datasets.

Given its superior visual quality, scalability, and disentangled style representations, HiGAN+ is chosen as the model for augmenting the dyslexia handwriting dataset. This choice aligns with the project's objective of generating diverse, realistic, and disentangled handwriting images, conditioned on arbitrary textual content. HiGAN+ demonstrates a remarkable ability to synthesize handwriting images, providing a strong foundation for augmenting the dyslexia dataset with diverse handwriting styles and arbitrary textual contents.

Table 2.2: GANs Model Comparison

	WER (word error rate)	**FID (Fréchet Inception Distance)**
ScrabbleGAN	23.61	23.78
GANwriting	17.26	20.55
SLOGAN	14.95	12.06
HiGAN+	**12.37**	**9.6546**

2.8 Relevant Work on few analytical tools

The field of data science and analytics relies heavily on efficient tools and IDEs for effective coding, exploration, and analysis. In this review, we examine several popular programing languages and IDEs used in data science, drawing insights from the research papers provided.

2.8.1 Python :

Python is widely used in data science and analytics for several reasons. It has a simple and readable syntax, modularity, object-oriented design, and a large library of standard extensions(Nagpal & Gabrani, 2019). Python offers high-level data structures and supports dynamic typing, automatic memory management, and exceptions. It is free for commercial purposes and runs on most modern computers. Python's features like less coding and high readability, portability and flexibility, platform independence, and a balance of low-level and high-level programming make it popular for data analytics applications.

2.8.2 Jupyter Lab:

Jupyter Lab is a web application-based IDE for data science that provides a visually intuitive organization of code and allows for interactive output. It offers features like writing code, creating visualizations, and adding text explanations using the markdown editor (Saabith et al., 2021). JupyterLab is easy to use and provides functionalities such as auto code completion, debugging, and auto-save. It supports over 40 programming languages, making it versatile. Pros of JupyterLab include visually intuitive organization, easy replication of notebooks, code presentation, and server hosting. However, some cons include the lack of code style correction, inconsistent kernel behavior, limited third-party app integration, and difficulty in viewing changes in GitHub.

2.8.3 PyCharm:

PyCharm is a Python IDE developed by JetBrains. It offers a code editor with features like syntax highlighting, code analysis, and quick fixes. PyCharm supports various frameworks like Django and Flask, making it suitable for web development(Saabith et al., 2021). It provides a powerful debugger, integrated unit testing, a version control system, and the ability to customize the interface. Pros of PyCharm include data science scripting, frontend development, intuitive and simple IDE, and integration with databases and version control. Some cons include higher memory usage, the need for configuration setup, and the cost of the full version.

2.8.4 Google Colaboratory:

Google Colaboratory, commonly referred to as Colab, is a cloud-based service built upon Jupyter Notebooks, aimed at democratizing machine learning education and research. It offers a runtime environment tailored for deep learning tasks and initially provides users with free access to a robust GPU, making it an attractive platform for conducting analytical work(Carneiro et al., 2018). However, for extended runtimes or to access higher GPU capabilities and additional RAM, Colab may charge users, introducing a flexible pricing model to accommodate varying computational needs .

Colab notebooks facilitate seamless collaboration and sharing among users, fostering a conducive environment for teamwork and knowledge dissemination. These notebooks support both Python 2 and 3 runtimes, pre-configured with essential machine learning and artificial intelligence libraries, including TensorFlow, Matplotlib, and Keras(Carneiro et al., 2018).

2.8.5 Visual Studio:

Visual Studio is an IDE developed by Microsoft that supports Python, and R along with other programming languages. It offers features like IntelliSense code completion, code snippets, debugging, profiling, unit testing, and integration with Git. Visual Studio provides flexibility, portability, and the ability to work with different platforms. Pros of Visual Studio include smart code completion, Git integration, customization options, and support for various technologies. However, some cons include limited search features, lower performance, and the need to switch runtime environments.

2.9 Summary

The literature review provides insights into the challenges associated with dyslexia diagnosis, including the limitations of traditional diagnostic methods and the need for automated tools. It discusses various techniques, such as eye-tracking technology, brain imaging studies, and handwriting analysis, that have been explored for dyslexia recognition. Furthermore, it reviews different machine learning and deep learning approaches, highlighting their effectiveness in dyslexia detection and classification based on handwriting patterns.

One notable finding from the literature review is the potential of Generative Adversarial Networks (GANs) in data augmentation for dyslexia diagnosis. GANs offer a unique approach to generating handwriting samples, addressing challenges such as data imbalance and dataset size limitations. Among the reviewed GANs techniques, HiGAN+ stands out for its ability to generate diverse and realistic handwriting images, making it a compelling choice for augmenting dyslexia datasets.

Based on the insights gathered from the literature review, the project aims to utilize HiGAN+ for data augmentation tasks and implement both LeNet-5 and DenseNet201 models for dyslexia classification. The project will be implemented using Python programming language, with model training conducted on Google Colab and image preprocessing performed using Jupyter Lab. By leveraging advanced deep learning techniques and GAN-based data augmentation, the project aims to develop an accurate and efficient automated model for dyslexia diagnosis.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

I This chapter delves into the detailed methodology employed to achieve the objectives of the project. Each phase of the project's methods is elucidated, beginning with an overview of the proposed project framework in Section 3.2, which outlines the development of an automated dyslexia severity classification system utilizing data augmentation generated from Generative Adversarial Networks (GANs). Following this, Section 3.3 provides an in-depth exploration of data collection and description.

Subsequently, Section 3.4 elucidates the image processing techniques utilized, encompassing data vectorization, image scaling, image normalization, conversion to grayscale, and conversion to HDF5 format. The subsequent section, Section 3.5, delineates the data augmentation methods using GANs employed in the project, detailing their implementation and significance.

Moving forward, Section 3.6 outlines the evaluation metrics utilized to assess the quality of the generated images, providing insights into the effectiveness of the augmentation techniques employed. Section 3.7 elaborates on the development of Convolutional Neural Network (CNN) models and the experimental procedures undertaken in the project.

In Section 3.8, the performance metrics utilized to evaluate the efficacy of the CNN models are explicated, offering a comprehensive understanding of the model's performance and capabilities. Finally, the chapter concludes with Section 3.9, which details the experimental setup utilized to conduct all experiments and generate the images essential for the project's execution.

3.2 Project Framework

Figure 3.1 illustrates the comprehensive project framework employed throughout the entirety of the study. The framework initiates with the collection of the primary dataset, denoted as PDM. Subsequently, to prepare the dataset for data augmentation and input into Convolutional Neural Networks (CNNs), a series of preprocessing steps are executed. Initially, each word in the dataset is resized and labeled, followed by the conversion of images to grayscale. Furthermore, to ensure uniformity, the images are standardized to a height of 64 pixels. To facilitate efficient data handling, the preprocessed images are then converted into HDF5 format.

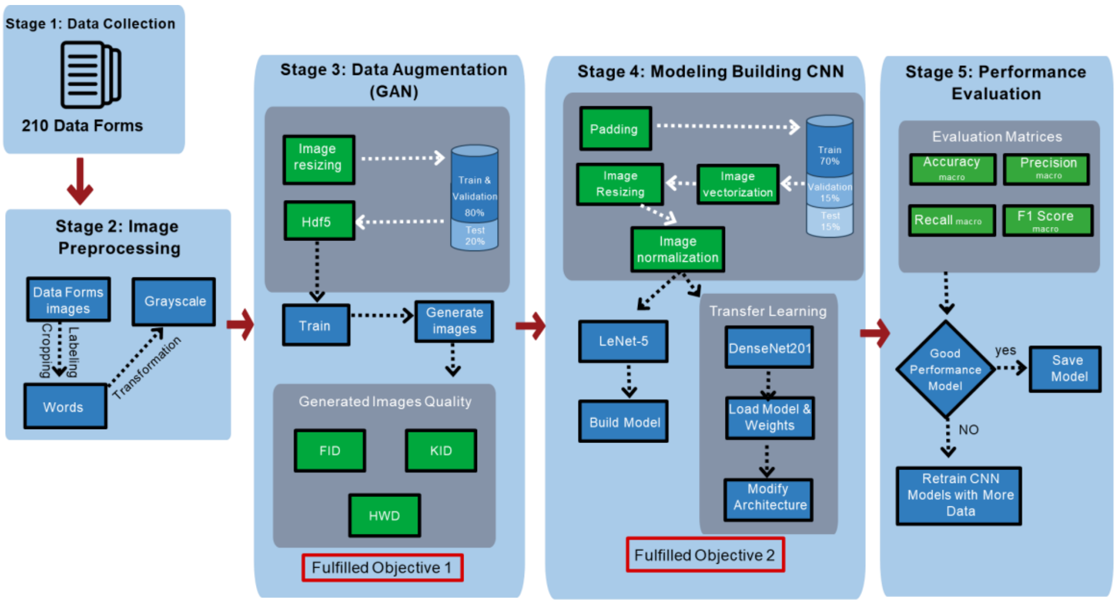


Figure 0.1: Project Framework

The next phase of the framework involves data augmentation utilizing Generative Adversarial Networks (GANs). The augmented dataset is generated by training the GAN model on the preprocessed images. Upon completion of the augmentation process, the quality of the generated images is evaluated using three distinct metrics: Fréchet Inception Distance (FID), Kernel Inception Distance (KID), and Handwriting Distance (HWD).

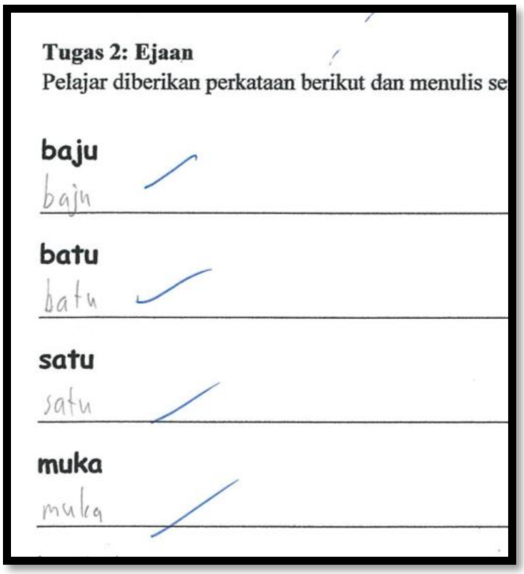
Following the assessment of image quality, CNN models are developed using both the original and augmented images. However, before model development, additional preprocessing steps are performed on the images. These include resizing and vectorizing the images, followed by normalization to enhance model performance.

Subsequently, the CNN models' performance is evaluated using a set of standard metrics, including Accuracy, Recall, Precision, and F1 Score. If the performance of the models falls below the desired threshold, additional data from the augmented dataset is incorporated to enhance model performance iteratively.

3.3 Data Collection

The dataset utilized in this research project was procured in collaboration with the Persatuan Dyslexia Malaysia (PDM), also known as the Dyslexia Association of Malaysia. The dataset comprises handwritten images from 210 samples, each reflecting distinct writing styles and characteristics. Within this dataset, 38 subjects were identified as having a high risk of dyslexia, 26 exhibited a moderate risk, and 42 showed a low risk. Additionally, 104 subjects were categorized as having no indications of dyslexia.

The data collection involved subjects completing paper-based forms containing 11 prescribed words for assessment, allowing an evaluation of their potential dyslexia. All handwritten images are stored in JPG format, accumulating to a total dataset size of 30MB.


Tugas 2: Ejaan
Pelajar diberikan perkataan berikut dan menulis se

baju
baju

batu
batu

satu
satu

muka
muka

Figure 3.2: Data Form

Professionals affiliated with PDM meticulously labeled each image within the dataset. The severity of dyslexia was systematically assigned to individual images, guided by several key handwriting indicators validated by experts in the field:

1- Incorrect Spelling and Capitalization: Observable errors in spelling and capitalization.
2- Inappropriate Sizing and Spacing of Letters: Difficulties in maintaining proper letter size and spacing while copying words.
3- Mix of Cursive and Print Letters: Presence of a blend of cursive and print letters in words.
4- Omitting Letters and Words: Instances of omitting letters or entire words from sentences.
5- Poor Letter Formation: Issues with forming letters clearly and accurately.
6- Difficulty with Spacing Text: Challenges in maintaining consistent spacing between words or letters.
7- Difficulty with shape discrimination and Letter Spacing: Struggles in differentiating shapes and maintaining appropriate letter spacing.
8- Trouble Organizing Words from Left to Right: Difficulties in organizing words correctly from left to right.
9- Inability to Stay Within Margins: Writing extending beyond the margins provided.
10- Incomplete Sentences: Writing that lacks completeness in sentence structure.

Each handwriting sample was rigorously assessed based on these established dyslexia indicators to ascertain the severity and presence of dyslexia-related characteristics.

3.4 Image Preprocessing

3.4.1 image resizing

The initial dataset contained handwriting images with substantial irrelevant information and noise, posing challenges for accurate dyslexia classification by deep learning models. To address these challenges and improve the model's emphasis on essential handwriting features, extensive preprocessing steps were taken.

3.4.2 Handwriting Image Segmentation

Each original handwriting image, which contains multiple words, was subjected to a segmentation process. This involved the separation of each word within the image to generate 11 distinct images, one for each word. This segmentation aimed to reduce the inclusion of irrelevant data and noise, thereby facilitating the model's classification process.

Adobe Illustrator was employed to perform resizing, labeling, and cleaning operations on the segmented images. The initial image size of 1654 x 2340 pixels was transformed into labeled words with a standardized size, featuring a height of 64 pixels while preserving the original width. This resizing process substantially reduced the dataset's size, thereby optimizing it for computational efficiency without compromising essential handwriting features.

Furthermore, the images underwent a meticulous cleaning process to eliminate extraneous elements. Notations from instructors, such as correct mark signs and student erasing marks, were removed. Additionally, the computer-generated text that students were required to copy was erased to prevent any confusion during model training.

As a result of these preprocessing steps, the dataset's size was substantially reduced from its original 30MB to 6.32 MB. This reduction in data size was instrumental in enhancing the dataset's relevance by focusing on crucial handwriting features while eliminating unnecessary noise and irrelevant information.

3.4.3 Transformation to Grayscale

To further refine the dataset for accurate dyslexia classification, grayscale conversion was performed on the segmented handwriting images. Grayscale conversion simplifies the data representation by reducing complexity while enhancing contrast and clarity. This preprocessing step standardizes the appearance of handwriting images, making them more uniform and compatible with subsequent analysis tools and algorithms.

By converting the handwriting images to grayscale, irrelevant color variations due to factors like lighting conditions and ink quality were eliminated. This normalization process ensures consistency across the dataset, facilitating feature extraction and pattern recognition tasks. Moreover, grayscale images are computationally efficient, requiring fewer resources for processing compared to colored images.

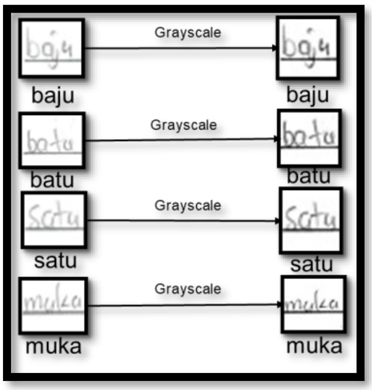


Figure 3.3: Original image & Grayscale

3.4.4 Transformation to Hierarchical Data Format (HDF5)

In preparation for the application of Generative Adversarial Networks (GANs) to augment the dataset, the dataset underwent conversion to the HDF5 format. This conversion process involved organizing the segmented handwriting images into a structured HDF5 architecture tailored for efficient data storage and retrieval. The architecture comprises datasets for images, labels, word IDs, and corresponding seek indices and lengths. The HDF5 format facilitates seamless integration with GANs, enabling the generation of synthetic handwriting samples to augment the dataset for improved model performance and robustness.

3.4.5 Padding

Padding plays a crucial role in maintaining the integrity and structure of handwritten words. Since words are inherently rectangular, with varying widths and heights, padding ensures that the images retain their original shape and aspect ratio, thereby preserving essential word information. By adding padding to the images, specifically by extending the height to match the width, the resulting images become square-shaped. This square shape prevents distortion when resizing the images to the required input size of the CNN model. As a result, the features extracted from the images remain consistent and unaffected by stretching, leading to a more accurate representation of the handwriting characteristics. The padding process, exemplified by the code snippet provided, ensures that each word maintains its spatial properties, facilitating the CNN model's ability to discern subtle patterns and features crucial for dyslexia severity classification.

3.4.6 Data Vectorization

In conjunction with the preprocessing techniques applied to the dyslexia handwriting images, data vectorization stands as a key step in preparing the dataset for model training and evaluation. Data vectorization involves converting the raw image data into a structured format compatible with machine learning algorithms, facilitating efficient processing and analysis. In this context, each handwriting image undergoes vectorization to transform it into a numerical representation suitable for classification tasks. Additionally, categorical encoding is applied to the target labels using one-hot encoding techniques, transforming them into binary vectors representing the presence or absence of each dyslexia severity level. By aligning with the preprocessing techniques, such as grayscale conversion and resizing, data vectorization ensures consistency and compatibility with the subsequent stages of model development and evaluation.

3.4.7 Data Normalization

Data normalization involves scaling the numerical features to a standardized range, typically between 0 and 1, to mitigate the effects of varying feature scales and gradients during model training. In this context, each pixel value of the handwriting images undergoes normalization to ensure consistency and comparability across all images. By dividing each pixel value by 255, the maximum pixel intensity value, grayscale pixel values are rescaled to the range [0, 1], effectively normalizing the pixel intensities. This normalization process aligns with previous preprocessing techniques, such as grayscale conversion and resizing, to maintain uniformity and compatibility throughout the dataset. By incorporating data normalization into the preprocessing pipeline, the dataset's features are appropriately scaled, thereby facilitating efficient model convergence and improving the overall performance and robustness of dyslexia severity classification models.

3.5 Data Augmentation

The original dyslexia dataset exhibits a notable imbalance in its distribution, with a significant disparity between the number of individuals categorized as normal, high risk, moderate risk, and low risk. Specifically, the population labeled as normal far surpasses those identified with high dyslexia risk by a factor of three and exceeds those categorized with low risk by a factor of ten. Such an imbalanced dataset poses inherent challenges in training deep-learning models. The substantial overrepresentation of the normal class within the dataset often leads to models prioritizing learning patterns specific to this majority class while potentially overlooking or underrepresenting the minority classes. This skewed focus may result in an overfit model that fails to generalize effectively to new or unseen data, significantly compromising its reliability and applicability in real-world scenarios. Addressing this imbalance becomes crucial in ensuring the model's capacity to discern and appropriately categorize individuals across the dyslexia severity range.

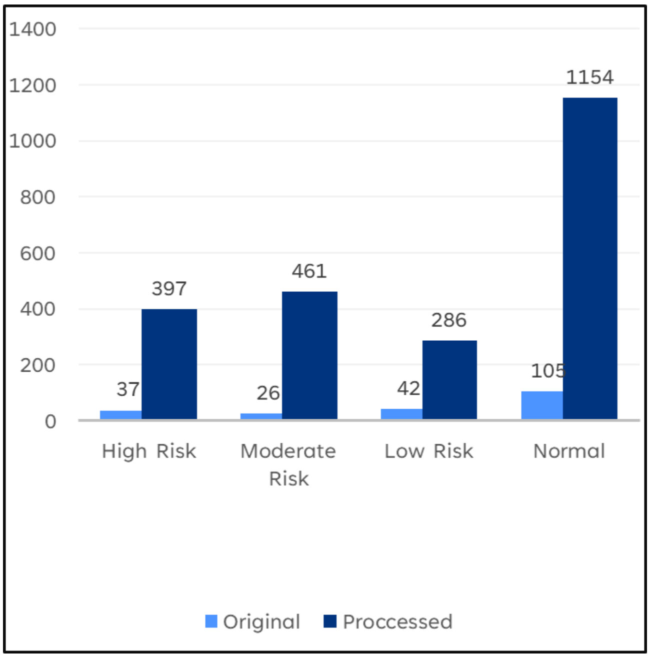


Figure 3.4: Data Distribution

Addressing the challenge of imbalanced data in dyslexia handwriting datasets necessitates effective augmentation techniques. HiGAN+ stands out as an innovative approach within generative adversarial networks (GANs), focusing on disentangled representations to synthesize diverse and lifelike handwritten texts. Its emphasis on generating handwriting images conditioned on arbitrary textual content and calligraphic styles showcases superior visual quality, style transferability, and scalability, outperforming existing GAN models.

In the context of dyslexia classification models, HiGAN+ presents an enticing choice due to its emphasis on generating diverse and realistic handwriting images. This unique approach offers significant potential for augmenting dyslexia handwriting datasets, enabling the integration of varied handwriting styles reflective of the dyslexic population. The model's disentangled representation, scalability, and style transferability make it a compelling solution for enhancing the diversity and richness of dyslexia handwriting datasets.

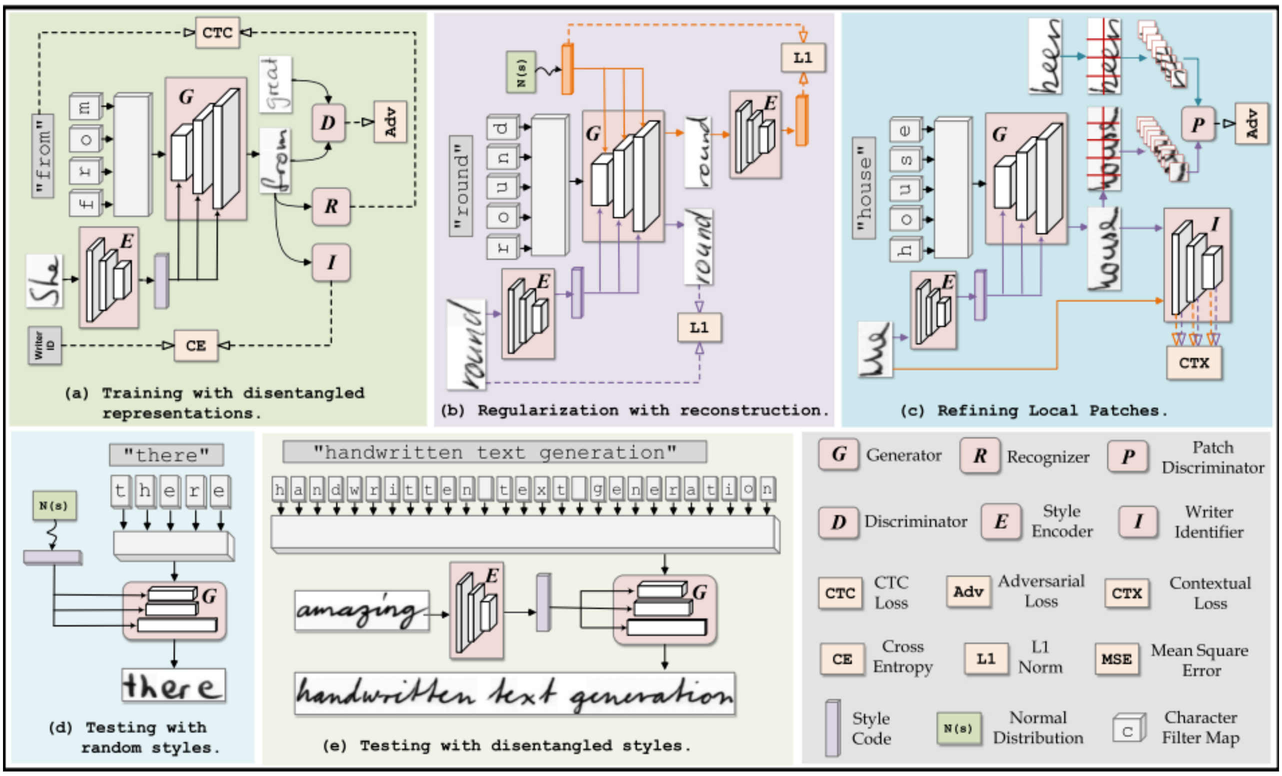


Figure 3.5: HIGANs+ Framework (Gan et al., 2022)

3.6 GAN Evaluation Matrices

3.6.1 Fréchet Inception Distance (FID)

The Fréchet Inception Distance (FID) is a metric used to assess the quality of generated samples produced by Generative Adversarial Networks (GANs). It operates by embedding a set of generated samples into a feature space defined by a specific layer of the Inception Net or any Convolutional Neural Network (CNN). Within this feature space, the embedding layer is treated as a continuous multivariate Gaussian distribution, allowing for the estimation of mean and covariance parameters for both the generated and real data(Borji, 2019).

The Fréchet distance, also known as the Wasserstein-2 distance, is then computed between these two Gaussian distributions. This distance serves as a quantifiable measure of the quality of generated samples, with lower FID values indicating smaller distances between synthetic and real data distributions. FID is lauded for its strong discriminability, robustness, and computational efficiency.

3.6.2 Kernel Inception Distance (KID)

The Kernel Inception Distance (KID) stands as a pivotal metric in the realm of generative adversarial networks (GANs), designed to assess the quality of generated images. It represents an extension of the Fréchet Inception Distance (FID) and introduces the concept of Maximum Mean Discrepancy (MMD) with the kernel trick(Bińkowski et al., 2018).

KID operates by measuring the dissimilarity between the distributions of real and generated images directly in feature space. Leveraging the kernel trick enhances its ability to capture both global and local characteristics within these distributions. This nuanced approach provides a more comprehensive evaluation of image generation quality compared to traditional metrics.

3.6.3 Handwriting Distance (HWD)

The Handwriting Distance (HWD) is a specialized metric designed for evaluating Handwriting Text Generation (HTG) models. Operating within the feature space of a network trained to extract handwriting style features from variable-length input images, HWD enables nuanced comparisons of subtle geometric handwriting features. Leveraging a perceptual distance measure, it accurately assesses handwriting realism by utilizing robust style features extracted from a pre-trained backbone trained on synthetic text image datasets. HWD is adept at handling variable-length text images and exhibits numerical stability, even when computed on a limited number of samples. Through extensive experimental evaluation, HWD has proven effective in capturing differences in handwriting styles and surpasses traditional metrics like Fréchet Inception Distance (FID) in expressing the performance of Styled HTG models. Overall, HWD offers a tailored and practical evaluation score for assessing the realism and faithfulness of generated text images in the HTG domain, contributing to advancements in this field (Pippi et al., 2023).

3.7 Modeling

3.7.1 LeNet-5

LeNet-5, a Convolutional Neural Network (CNN) architecture, assumes a central role in the classification of dyslexia from handwriting images, owing to its customized design and remarkable capabilities. CNNs, like LeNet-5, boast a hierarchical structure containing convolutional layers, pooling operations, activation functions, and fully connected layers. This detailed architecture empowers CNNs to autonomously learn discriminative features directly from raw data, a pivotal advantage when interpreting the fine patterns present in dyslexic handwriting.

The convolutional layers within LeNet-5 play a crucial role in detecting local patterns within the handwriting images. These layers act as filters, scanning the input images to identify distinctive features that are indicative of dyslexic handwriting characteristics. Additionally, pooling operations employed in LeNet-5 serve to down sample the feature maps, effectively reducing their dimensionality while retaining essential information. Meanwhile, activation functions introduce non-linearities, enhancing the network's capacity to capture complex relationships between features.

This intricate process of feature extraction facilitates the identification of distinguishing characteristics specific to dyslexic handwriting, thereby enabling accurate classification.

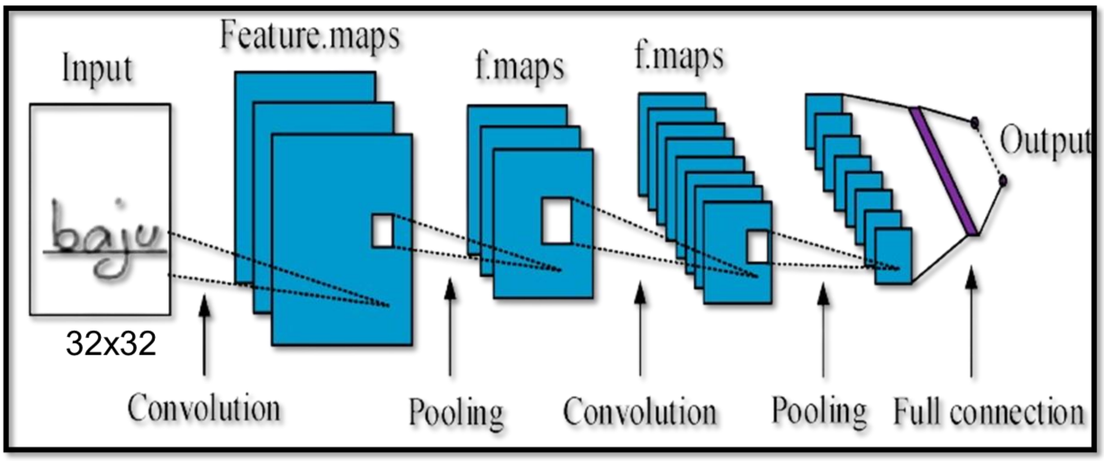


Figure 3.6: CNN Framework

3.7.2 Transfer Learning using DenseNet201.

In the realm of dyslexia classification, the utilization of transfer learning emerges as a strategic approach due to the scarcity of available data. Transfer learning capitalizes on insights gleaned from established models, amplifying performance, and circumventing the necessity for exhaustive data acquisition. This project implements DenseNet201, a powerful transfer learning model pre-trained on the ImageNet dataset. By leveraging DenseNet201 alongside the SGD optimizer and categorical cross-entropy loss functions, the classification process is streamlined, facilitating the development of accurate dyslexia classification systems in resource-limited scenarios.

DenseNet201 stands as a robust foundation for extracting intricate features from handwriting images. Pretrained on a vast array of images, it possesses a deep understanding of diverse visual features, enabling it to determine slight patterns inherent in dyslexic handwriting. Fine-tuning DenseNet201 for dyslexia classification involves customizing its top layers to align with the specific task requirements. A new SoftMax layer, tailored to the desired number of classes, is appended to the pre-trained architecture. This customized classification layer seamlessly integrates with DenseNet201, forming a cohesive network capable of discerning dyslexia-related features from handwriting images.

Furthermore, the adaptation of DenseNet201 involves selectively freezing the weights of pre-existing layers while keeping the classification layer trainable. This strategy enhances the model's capacity to adapt to dyslexia-specific patterns while minimizing overfitting risks. By optimizing the model's architecture and training parameters, DenseNet201 becomes a potent tool for developing accurate and efficient dyslexia classification systems.

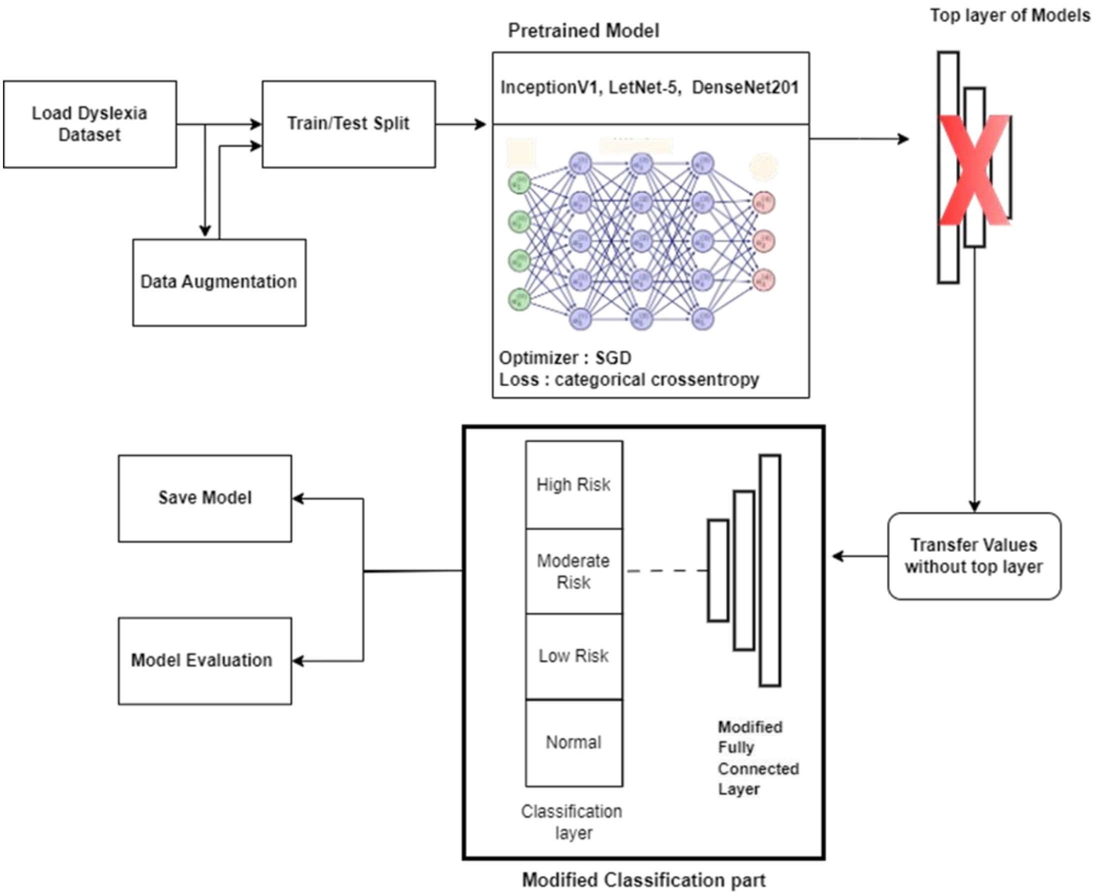


Figure 3.7: Transfer Learning Framework

3.8 CNN Evaluation

The evaluation of the transfer learning model for dyslexia severity involves several key metrics that measure its performance across multiple classes, ensuring a comprehensive understanding of its effectiveness in classifying dyslexia severity levels. These metrics serve as vital indicators, shedding light on various aspects of the model's performance and enabling informed decision-making in dyslexia research and intervention strategies.

Accuracy stands as a fundamental metric, offering an overall measure of the model's performance by indicating the ratio of correctly predicted samples to the total number of samples across the four dyslexia severity levels. It provides valuable insight into the model's ability to classify severity levels accurately and is instrumental in assessing its overall efficacy.

Precision, another critical metric, assesses the proportion of accurately predicted samples within each severity category, including high, moderate, low, and no indications, among the predicted positive samples for that category. By quantifying the model's ability to precisely classify samples within each severity level, precision offers valuable insights into its ability to avoid misclassifications and false positives.

Similarly, recall measures the true positive predictions among the actual positive samples for each severity level. It complements precision by providing a measure of the model's ability to capture all positive instances within each severity category, ensuring comprehensive coverage of relevant samples and minimizing false negatives.

The F1 Score, a harmonic mean of precision and recall, provides a balanced assessment for each severity category, offering a single metric that considers both precision and recall simultaneously. By combining these two metrics, the F1 Score offers a comprehensive evaluation of the model's performance, accounting for both false positives and false negatives and providing a nuanced understanding of its accuracy and precision across different severity categories.

3.9 Experimental Setup

3.9.1 Tools & language used

In this study, we used different tools and programming languages to handle our data and build our models. Each tool had its job to do, and together they helped us conduct our research effectively.

Data Preprocessing Tools:

To start, we used Adobe Illustrator to clean up and resize our dataset, making sure each word looked neat and was the right size. We also used Snip and Skitch apps to trim down unnecessary white spaces around the images, making our dataset more efficient.

Development Environment:

All our data preprocessing tasks, like converting data and making it ready for our models, were done in Jupyter Notebook. It's a handy tool that lets us write code, see results, and explain our work all in one place.

Model Training Platform:

When it came to training our models, we relied on Google Colab, a powerful platform with big computers called GPUs. These GPUs helped our models learn faster, making our research more efficient.

Programming Languages:

Python was our main language throughout the project. It's easy to read and understand, and it has lots of tools we need, like OpenCV for working with images and NumPy for handling data.

Libraries and Frameworks:

We used different libraries to help us build and train our models. Pillow helped us work with images, Keras made it easier to create our CNN models like LeNet-5 and DenseNet201, and PyTorch helped us build our GANs model called HIGAN+.

3.9.2 Experimental Data Configuration

In our project, we implemented two distinct data configurations for training purposes. Each configuration was designed to optimize the training process and facilitate robust model development.

For the evaluation of Generative Adversarial Networks (GANs) performance, we divided the dataset into two subsets: a training set comprising 80% of the data and a testing set consisting of the remaining 20% as shown in the below table. This split allowed us to assess the efficacy of the GANs model in generating synthetic handwriting samples.

Table 3. 1: GAN Dataset Class Distribution

Class	Normal	Low	Moderate	High
Training & Validation	923	219	351	300
Testing	231	67	110	97
Total Images	1154	286	461	397

On the other hand, for training Convolutional Neural Network (CNN) models like LeNet-5 and DenseNet201, we adopted a different approach. Here, we partitioned the dataset into three distinct subsets: 70% for training, 15% for validation, and another 15% for testing. This tripartite split ensured that our CNN models were trained on a sizable portion of the data, validated on a separate subset to tune hyperparameters, and finally tested on an independent subset to evaluate their performance.

Moreover, we conducted three experiments to explore the effectiveness of our models further. The first experiment utilized the original dataset exclusively, while the second experiment involved training with a balanced dataset generated from the GANs model. Lastly, the third experiment incorporated the balanced data along with 5 additional words, each consisting of 200 different variations, totaling 1000 new words for each class. These experiments enabled us to assess the impact of dataset variations on model performance and generalization capabilities.

Table 3. 2: CNN Training Data Distribution

Experiment	Experiment1	Experiment2	Experiment3
Image Type	Original	Generated Word	5 new words
Training	1608	6300	9100
Validation	345	1350	1950
Testing	345	1350	1950
Total Images	2298	9000	13000

3.9.3 Experimental Scenario

In our experimental setup, we started the development of two distinct Convolutional Neural Network (CNN) models: LeNet-5 and DenseNet201, each trained with three different datasets to explore various training scenarios and assess model performance thoroughly.

For the training of both LeNet-5 and DenseNet201 models, we employed a standardized training protocol consisting of 100 epochs, utilizing a learning rate of 1e-03 and the Adam optimizer. To evaluate model performance, we utilized the Categorical Cross-Entropy loss function, a widely adopted metric for multi-class classification tasks.

LeNet-5 Configuration:

The LeNet-5 architecture comprises two sets of convolutional layers followed by average pooling layers for down sampling. The network also includes fully connected layers with hyperbolic tangent (tanh) activation functions, culminating in a SoftMax layer for multi-class classification. This configuration was chosen for its simplicity and effectiveness in handling handwritten digit classification tasks.

DenseNet201 Configuration:

In contrast, the DenseNet201 architecture, pre-trained on the ImageNet dataset, was utilized as a feature extractor. We appended a SoftMax layer with the number of output neurons corresponding to the number of classes in our dyslexia severity classification task.

Both models were meticulously trained and evaluated across three distinct datasets: the original dataset, a balanced dataset generated from GANs, and a modified dataset with additional words. Each experimental configuration aimed to investigate the models' robustness, generalization capabilities, and adaptability to varying dataset compositions.

3.10 Summary

This chapter delves into the methodology through five distinct steps. Initially, the dataset was acquired from Persatuan Dyslexia Malaysia, providing the foundation for our research. Subsequently, image preprocessing techniques were employed to enhance and extract relevant features from the images, essential for both Generative Adversarial Networks (GANs) and Convolutional Neural Network (CNN) procedures. This preprocessing stage ensured that our images were clean and properly formatted, facilitating their conversion into a machine-readable format.

Following preprocessing, the GANs model was trained on the four classes within our dataset individually. Moreover, leveraging a pre-existing DenseNet201 model and a custom-built LeNet-5 CNN model, we conducted independent training sessions with identical configurations and utilizing varying dataset sizes across three separate experiments. Consequently, a total of six CNN models were trained and evaluated according to predefined evaluation criteria.

CHAPTER 4:

RESULTS AND DISCUSSION

4.1 Introduction

This study focuses on augmenting the potential dyslexia risk model by increasing the diversity of handwriting image datasets through synthetic data generation. Synthetic data serves as a crucial tool for data augmentation, particularly in scenarios where real-life data collection is constrained, offering opportunities to enhance the accuracy of classification and machine learning models within limited environments.

In this chapter, we delve into the experimentation and validation of the synthetic data generated for augmenting the dyslexia risk model. The analysis encompasses a comprehensive examination of the synthetic data, aiming to gain insights into its quality and utility.

The chapter unfolds with a detailed presentation of the experimental setup, outlining the training configuration, including the datasets utilized, as well as the software employed. Furthermore, we scrutinize the performance dynamics of the generator and discriminator throughout the Generative Adversarial Network (GAN) training process, shedding light on their efficacy and convergence.

Moving forward, we delve into the evaluation of synthetic data. Leveraging methodologies proposed in prior chapters, we assess the quality of synthetic data through visual inspection and similarity measures, providing a nuanced understanding of its fidelity to real data. Additionally, we explore the augmentation potential of synthetic data in improving classification performance, progressively integrating it into the original training dataset while meticulously documenting the behavior of the classifier.

Finally, this chapter culminates with a comprehensive summary, encapsulating key findings and insights derived from the experimentation process, thereby elucidating the efficacy and implications of synthetic data augmentation in enhancing the dyslexia risk model.

4.2 Result of Implementing HiGAN+ for Data Augmentation

The dataset we obtained from Persatuan Dyslexia Malaysia revealed a significant challenge: severe imbalances and limitations across different dyslexia severity classes. To mitigate this issue, we adopted a data augmentation strategy using HiGAN+, a specialized Generative Adversarial Network (GAN), to synthesize additional handwriting images. By replicating existing images, HiGAN+ effectively augmented the dataset, ensuring a more balanced representation of each dyslexia class.

The primary objective of data augmentation was to equalize the number of images across all classes, enabling our Convolutional Neural Network (CNN) models to learn features for each class label with equal priority. This adjustment was crucial for preventing bias in the model's learning process, ensuring that all classes received adequate attention during training.

Moreover, to further bolster the dataset's size and diversity, we introduced five new words(dua, gua, ubat, anda, abu) to each class. Each word comprised 200 images, resulting in a total of 1000 additional images per class. This deliberate augmentation strategy aimed to enrich the dataset with varied handwriting samples, enhancing the model's ability to generalize across different writing styles and patterns.

The distribution of the original dataset and the newly generated images is outlined in the table below, providing insight into the augmentation process and its impact on dataset composition.

Table 4. 1: Dataset Distribution

Data Type	Original	Balanced	Balanced+5 Words
Normal	1154	2250	3250
Low	286	2250	3250
Moderate	461	2250	3250
High	397	2250	3250
Total Images	2298	9000	13000

The outcomes of our GANs experiments were promising, demonstrating the model's proficiency in capturing essential features such as font intensity, letter spacing, and authorial style to generate realistic handwriting images. The model's ability to replicate handwriting styles accurately underscores its effectiveness in creating synthetic data to augment our dataset.

Visual inspection of the generated data revealed striking similarities to the original handwriting samples, with discernible variations in font styles and letter formations faithfully reproduced. The figure below provides a comparative analysis between the original and generated data, showcasing the GANs' capacity to mimic handwriting styles across different dyslexia severity classes.

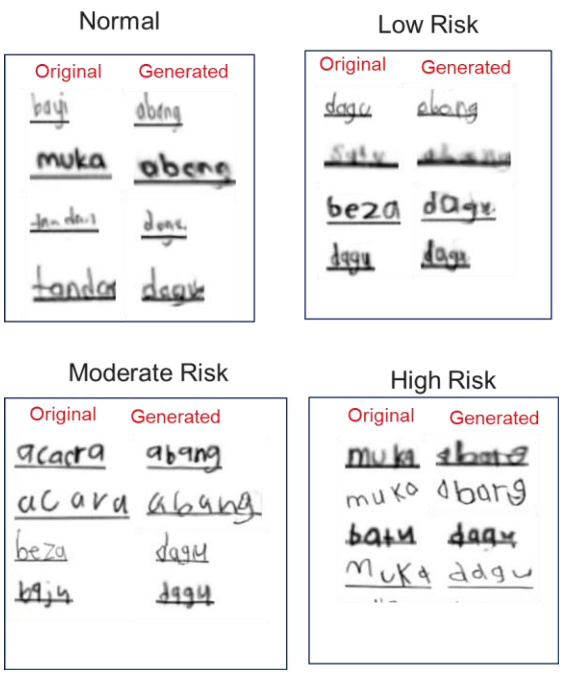


Figure 4. 1: Generated Image example

Furthermore, to illustrate the diversity and authenticity of the generated data, we present an example of the newly synthesized word "dua" across various dyslexia severity classes. This exemplifies the GANs' adaptability in generating handwriting images that encompass a wide spectrum of styles and characteristics, effectively enriching our dataset with diverse and representative samples.

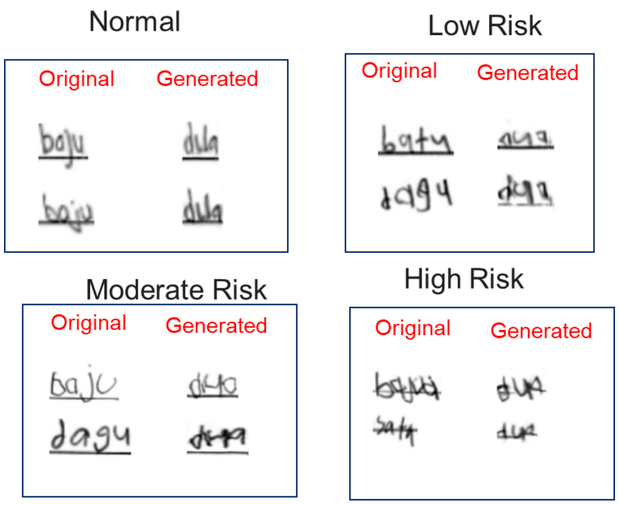


Figure 4. 2: Generated new word example.

In evaluating the quality of generated images produced by Generative Adversarial Networks (GANs), three distinct metrics were utilized: Fréchet Inception Distance (FID), Kernel Inception Distance (KID), and Handwriting Distance (HWD). Each metric offers unique insights into the fidelity and realism of the generated samples as shown from the below table.

Table 4. 2: Generate Image Quality Evaluation

Class	FID	KID	HWD
Normal	69.448	0.0756	0.647
Low	82.021	0.0654	0.634
Moderate	62.349	0.0461	0.799
High	57.778	0.0333	0.701
Total Images	57.119	0.0590	0.584

The evaluation of image generation quality metrics provides valuable insights into the performance of our Handwriting Text Generation (HTG) models. Each metric employed - Fréchet Inception Distance (FID), Kernel Inception Distance (KID), and Handwriting Distance (HWD) - offers distinct considerations and implications.

FID primarily assesses general image characteristics, disregarding specific handwriting shapes. While it offers insights into image distribution similarity, its reliance on a backbone network trained on ImageNet may yield misleading values for handwritten text images due to differing aspect ratios. Additionally, FID's sensitivity to dataset size bias, evident in our results where larger classes yield better FID scores, underscores its limitations in small dataset contexts.

KID addresses FID's bias towards large datasets by measuring dissimilarity in feature space directly, yielding improved results. However, similar to FID, KID was not explicitly designed for assessing image quality and inherits FID's limitations in this regard.

In contrast, HWD offers tailored evaluation for HTG models, operating within a feature space specifically trained to extract handwriting-style features. By utilizing perceptual distance for nuanced comparison of handwriting geometric features, HWD offers robust style feature extraction, variable-length text image handling, and numerical stability, as demonstrated by our results. Our HWD scores for each class - Normal, Low, Moderate, and High - have a small number which indicates the effectiveness of our HTG model in generating realistic handwriting images across various dyslexia severity levels.

4.3 Classification Performance of CNN Models

In this section, we delve into the detailed evaluation and analysis of the performance metrics for the pre-trained CNN models across multiple experiments. Each experiment focused on measuring crucial metrics such as accuracy, precision, recall, and F1 score to assess the efficacy of the models in dyslexia severity classification tasks. The results of these experiments are presented comprehensively in Tables 4.2, 4.3, and 4.4, with emphasis placed on identifying the top-performing models highlighted in bold.

The subsequent sections will provide a thorough examination of each experiment, elucidating the methodologies employed, the observed outcomes, and the implications derived from the findings.

4.3.1 Result of Experiment 1

Experiment 1 focused on training the LeNet-5 and DenseNet201 models using a dataset comprising 1608 images out of a total of 2298 images. The performance of both models was evaluated based on key metrics including accuracy, precision, recall, and F1-score, as outlined in Table 4.3. The results indicate that DenseNet201 outperformed LeNet-5 across all metrics, demonstrating superior accuracy, precision, recall, and F1-score.

Table 4. 3: Experiment 1 Result

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
LeNet-5	69	55	56	54
DenseNet201	**76**	**69**	**67**	**67**

However, despite the better performance of DenseNet201, observations from the training process reveal some noteworthy insights. Firstly, both models exhibited signs of overfitting, as evidenced by the disparity between training and validation accuracies. This overfitting phenomenon can be attributed to the limited size of the dataset and the inherent class imbalance within it as shown in figure .

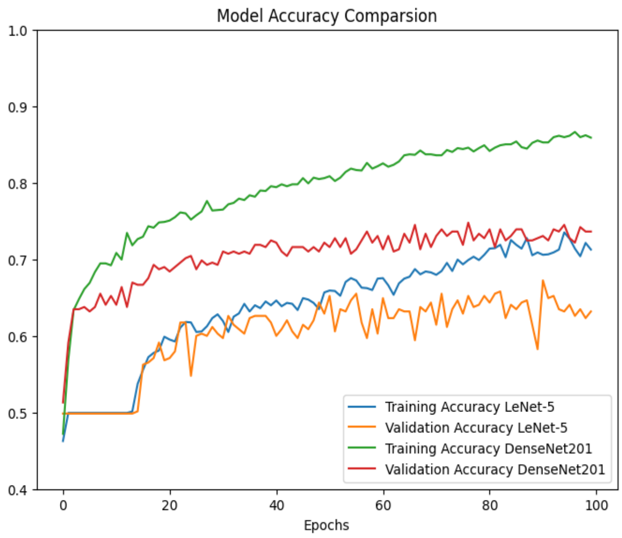


Figure 4. 3: Experiment 1 Model Comparison

Despite the overfitting, DenseNet201 showcased relatively lower levels compared to LeNet-5. This improvement in generalization can be attributed to the utilization of pre-trained weights in DenseNet201, which provided a more robust starting point for feature extraction and classification. Consequently, DenseNet201 demonstrated enhanced capability in correctly identifying instances across different classes, particularly evident in the lower class, where it doubled the correct predictions compared to LeNet-5.

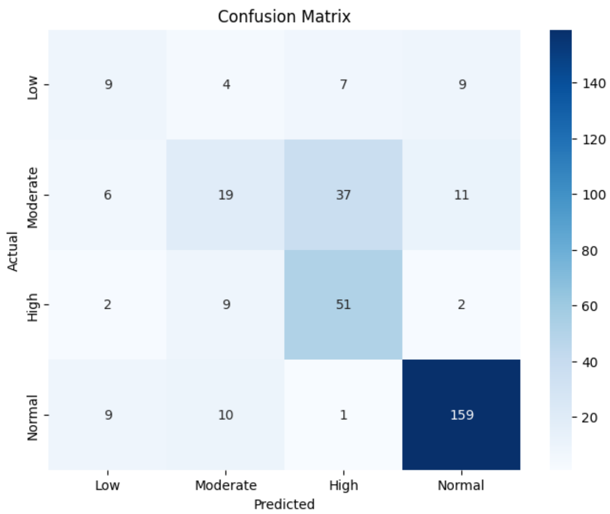


Figure 4. 4: Experiment 1 LeNet-5 Confusion Matrix

Further insights into the models' performance can be gleaned from the confusion matrices depicted in Figure 3. While LeNet-5 correctly classified 238 images, DenseNet201 achieved superior results by correctly classifying 263 images. These visual representations underscore the models' ability to correctly identify instances across various dyslexia severity levels, with DenseNet201 exhibiting more robust classification capabilities.

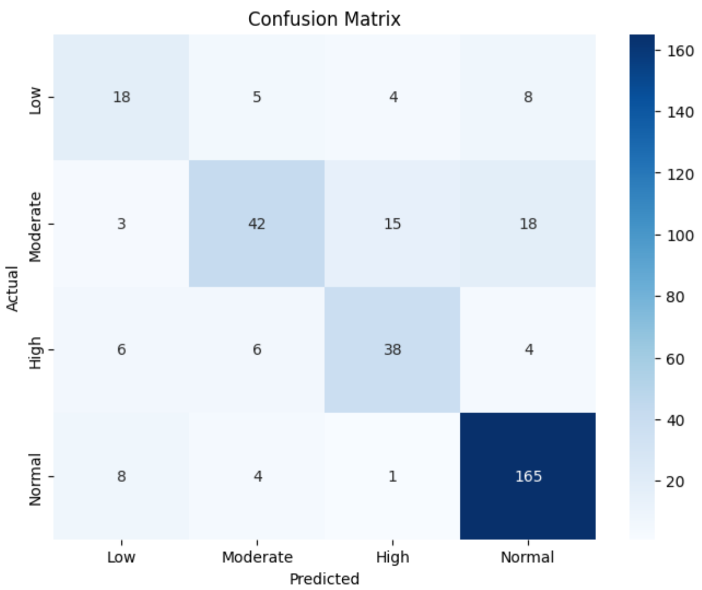


Figure 4. 5: Experiment 1 DenseNet201 Confusion Matrix

4.3.2 Result of Experiment 2

Experiment 2 delved into training the LeNet-5 and DenseNet201 models using an augmented dataset obtained from the GANs model. The dataset was significantly expanded, growing from 2298 to 9000 images, with each class containing 2250 balanced images. Of these, 6300 images were utilized for model training, while the remaining 2700 images were allocated for evaluation and testing purposes.

Upon evaluating the performance of both models using metrics such as accuracy, precision, recall, and F1-score, as delineated in Table 4.3, it was observed that LeNet-5 outperformed DenseNet201. The accuracy of the LeNet-5 model reached 88%, surpassing the accuracy of the DenseNet201 model, which reached 73%.

Table 4. 4: Experiment 2 Result

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
LeNet-5	**88**	**88**	**88**	**88**
DenseNet201	73	74	73	73

Remarkably, the significant reduction in overfitting was noted after augmenting the data and balancing the dataset, reflected in the consistent values of F1-score mirroring the accuracy. Both LeNet-5 and DenseNet201 models exhibited improved generalization performance, with the F1-score aligning closely with the accuracy metrics. Additionally, the training and validation accuracies depicted in the accompanying figure displayed a convergence trend, indicating reduced overfitting.

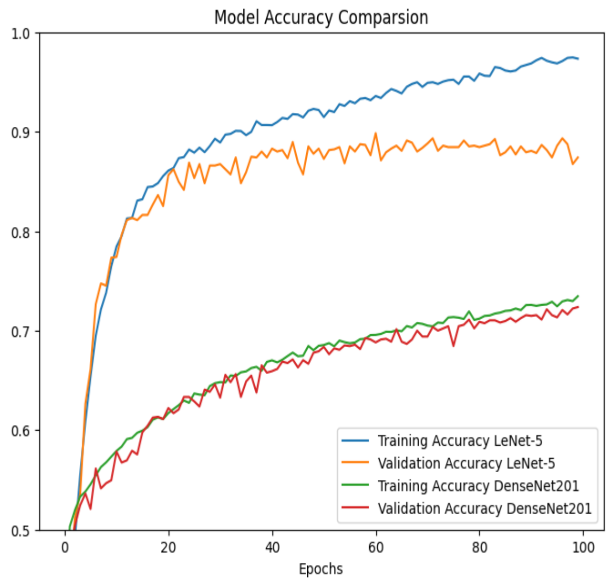


Figure 4. 6: Experiment 2 Model Comparison

Furthermore, LeNet-5 demonstrated superior classification capability, correctly identifying 1186 out of 1350 images, as illustrated in the figure.

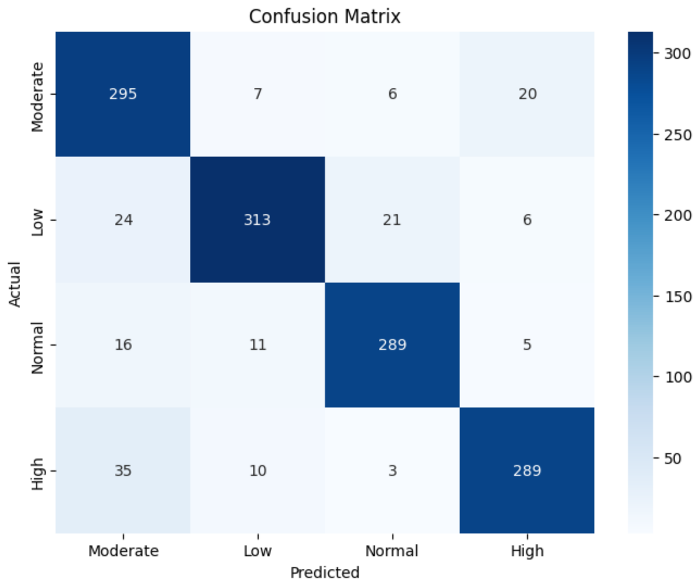


Figure 4. 7: Experiment 2 LeNet-5 Confusion Matrix

In contrast, DenseNet201 recognized 989 images, with a notable focus on normal and low severity classes over moderate and high severity classes.

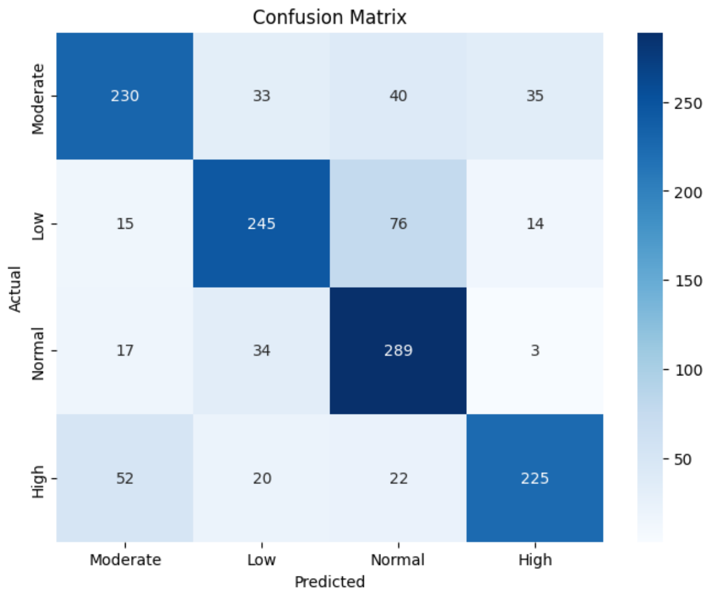


Figure 4. 8: Experiment 2 DenseNet201 Confusion Matrix

4.3.3 Result of Experiment 3

Experiment 3 explored the training of LeNet-5 and DenseNet201 models using an augmented dataset enriched with five additional words obtained from the GANs model. This augmentation significantly expanded the dataset, swelling from 2298 to 13000 images, with each class now containing 3250 balanced images. Within this enlarged dataset, 9100 images were used for model training, while the remaining 3900 images were reserved for evaluation and testing purposes. Notably, LeNet-5 continued to showcase superior performance compared to DenseNet201, achieving an accuracy of 91%, while DenseNet201 attained an accuracy of 86%.

Table 4. 5: Experiment 3 Result

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
LeNet-5	**91**	**91**	**91**	**91**
DenseNet201	86	86	86	86

In the context of Experiment 3, LeNet-5 demonstrated a remarkable 22% improvement compared to its performance with the original dataset.

Figure 4. 9: Experiment 3 Model Comparison

This significant enhancement underscores the model's adaptability and capability to leverage the augmented dataset effectively. The confusion matrix vividly illustrates LeNet-5's strength, with an impressive accuracy in predicting 1798 out of 1950 images. This outcome suggests that the simplicity of LeNet-5's architecture facilitated robust feature extraction and classification, enabling it to capitalize on the additional data and achieve notable performance gains. Furthermore, the model's ability to maintain high accuracy despite the augmented dataset indicates its resilience to overfitting and its capacity to generalize well to diverse handwritten text images.

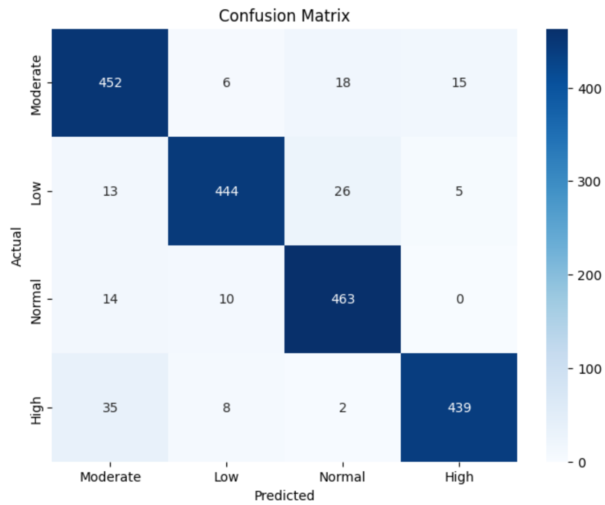


Figure 4. 10: Experiment 3 LeNet-5 Confusion Matrix

On the other hand, DenseNet201 also showcased improvement in performance, albeit to a lesser extent compared to LeNet-5, with a 12% enhancement observed when trained with the original dataset. Despite this progress, the confusion matrix highlights that DenseNet201 achieved correct predictions for 1669 out of 1950 images, indicating slightly lower accuracy compared to LeNet-5.

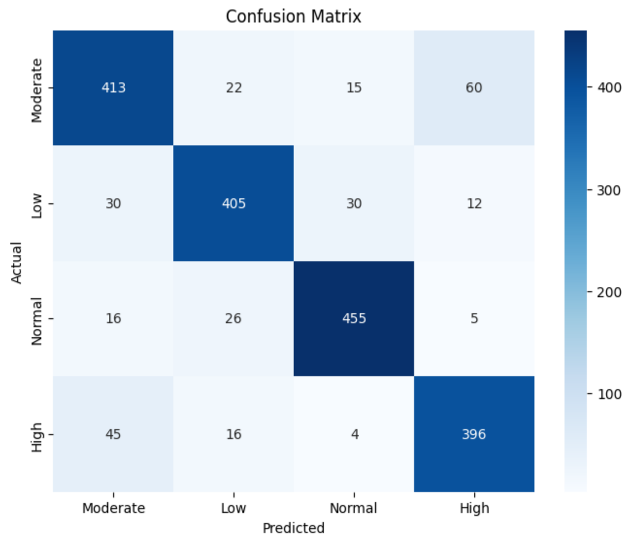


Figure 4. 11: Experiment 3 DenseNet201 Confusion Matrix

4.3.4 Selection of The Optimal CNN

Based on the provided results as shown in the figure below, it is evident that the LeNet-5 model, trained with the augmented dataset and further enhanced with the additional 5 words, outperforms both the original DenseNet201 model and the LeNet-

5 model trained solely with generated data. The F1 score, recall, precision, and accuracy metrics consistently demonstrate superior performance for the LeNet-5 model with the augmented dataset and additional words. With an F1 score of 91%, a recall of 91%, a precision of 91%, and an accuracy of 91%, this model achieves the highest performance across all metrics compared to the other configurations.

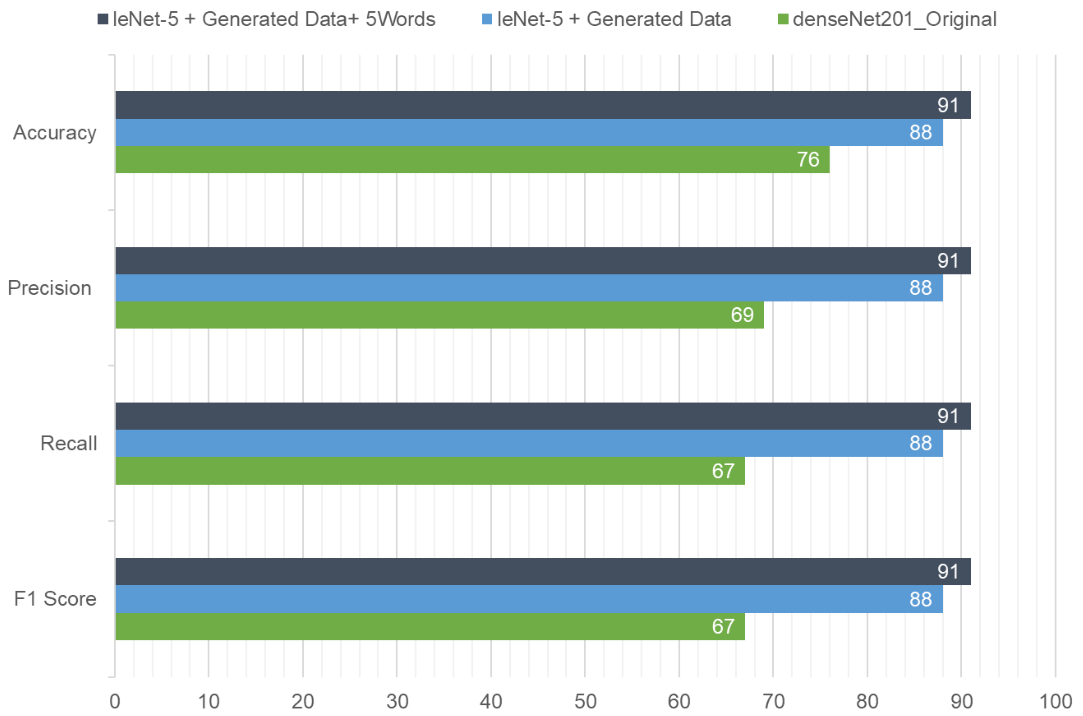


Figure 4. 12: CNN performance

The simplicity and customized architecture of LeNet-5 likely contributed significantly to its superior performance. Its fewer parameters and layers enabled better generalization to the augmented dataset, preventing overfitting and effectively capturing essential features. Additionally, LeNet-5's design, specifically optimized for handwritten digit recognition, may have inherently possessed features suitable for classifying handwritten text images. The stability of LeNet-5's training process further enhanced its performance compared to DenseNet201, which struggled to converge effectively to the augmented dataset due to its deeper and more complex architecture.

Moreover, the transfer learning approach used for DenseNet201 might not have been as effective for handwriting classification, as features learned from ImageNet may not have directly transferred to handwritten text images. As evidenced by the significant increase in performance with the augmented dataset, we can conclude that augmentation methods have proven to be highly effective in improving the performance of the CNN models for handwriting classification.

4.3.4 CNN Test Result

To evaluate the performance of our CNN models in identifying different levels of dyslexia severity, a comprehensive test was conducted. The test involved feeding images from each class of the original dataset individually to the models without labels to measure their accuracy. The overall results are summarized in the table below.

Table 4. 6: Test Accuracy Result

Model	High	Moderate	Low	Normal	Accuracy
DenseNet201	70.0	59.0	41.26	91.42	65.42
LeNet-5	85.6	77.22	75.2	95.5	83.38

These results illustrate the ability of each model to correctly classify images into their respective dyslexia severity levels. LeNet-5 outperforms DenseNet201 across all severity levels, achieving higher accuracy rates. For instance, in identifying high-severity cases, LeNet-5 attained an accuracy of 85.6%, while DenseNet201 achieved 70.0%. Similarly, for moderate and low-severity cases, LeNet-5 demonstrated higher accuracy compared to DenseNet201. Overall, LeNet-5 exhibited superior performance in accurately classifying dyslexia severity levels.

Additionally, image samples demonstrating the predictive class for different words within the high-severity class were generated using both LeNet-5 and DenseNet201 models. These samples provide visual insights into how each model predicts the severity level for handwritten text images. An example of such image samples for the high severity class prediction using LeNet-5 and DenseNet201 models is presented below.

Figure 4. 13: High Class Image Prediction Using DenseNet201

Figure 4. 14: High Class Image Prediction Using leNet-5

4.4 Summary

Based on the findings of this chapter, it is evident that our approach utilizing GANs for data augmentation has significantly improved the classification of dyslexia severity levels. This augmentation strategy enhanced the models' ability to learn and generalize patterns across different dyslexia severity levels, resulting in improved classification performance. The additional synthetic data generated by the GANs effectively supplemented the original dataset, providing the CNN models with a more diverse and comprehensive training dataset. Consequently, the models demonstrated enhanced capability in accurately discerning dyslexia severity levels from handwritten text images.

Furthermore, the remarkable accuracy achieved by the created models in predicting dyslexia severity levels indicates the effectiveness of our approach. By leveraging GANs for data augmentation, we have successfully addressed the challenge of limited and imbalanced datasets, allowing the models to achieve high accuracy in dyslexia classification.

CHAPTER 5

CONCLUSION & LESSON LEARNED

5.1 Conclusion

In summary, this thesis addresses the critical need for accurate and accessible dyslexia diagnosis tools by proposing and implementing an automated dyslexia severity classification system. Dyslexia, characterized by difficulties in reading despite adequate intelligence and instruction, presents a significant challenge to individuals affected by this neurodevelopmental disorder. Through a comprehensive exploration of Persatuan Dyslexia Malaysia (PDM) and the challenges within dyslexia diagnosis, this study underscores the limitations of current diagnostic methods and advocates for advanced assessment tools.

The framework developed in this thesis encompasses data collection, preprocessing, data augmentation using Generative Adversarial Networks (GANs), and the development of Convolutional Neural Network (CNN) models. Notably, the study emphasizes the pivotal role of data augmentation in mitigating dataset imbalances and biases, thereby enhancing model accuracy and generalizability.

Reviewing related literature sheds light on various dyslexia recognition approaches, including data science techniques, eye-tracking technology, brain imaging studies, and handwriting analysis. The efficacy of pattern recognition techniques, deep learning models, and machine learning algorithms in identifying dyslexia-related patterns is thoroughly discussed.

The methodology section delineates the project's intricate methodology, from the development of CNN models such as LeNet-5 and DenseNet201 to the implementation of data augmentation using GANs and experimental procedures. The results underscore the effectiveness of data augmentation, with the LeNet-5 model surpassing DenseNet201 in dyslexia severity classification.

In conclusion, the paper demonstrates the successful implementation of automated dyslexia severity classification using advanced deep learning techniques and GAN-based data augmentation. The experiments revealed that the LeNet-5 model, trained with an augmented dataset enriched with additional words generated by GANs, outperformed both the original DenseNet201 model and the LeNet-5 model trained solely with generated data. The LeNet-5 model achieved an impressive F1 score, recall, precision, and accuracy of 91%. In contrast, the DenseNet201 model, while showing improvement compared to its original configuration, achieved an accuracy of 86%. These results underscore the efficacy of data augmentation techniques in enhancing the accuracy and robustness of CNN models for dyslexia diagnosis and classification.

5.2 Challenges

Throughout this research journey, I encountered several hurdles related to understanding the subject matter and handling the dataset. Firstly, diving into dyslexia studies was entirely new to me, requiring an extensive learning process to grasp the disease and its fundamental concepts from scratch. Similarly, navigating generative models for the first time posed an extra challenge, demanding considerable effort to comprehend these models and choose the best fit for the project's needs. Additionally, working with the dataset proved demanding due to its complexities. Cleaning and preparing the data manually took a substantial amount of time, as the dataset required meticulous handling to ensure its accuracy and readiness for analysis. These obstacles, stemming from the need for both domain knowledge and dataset refinement, significantly shaped the learning curve of this project.

Moreover, implementing GANs for data augmentation presented its own set of challenges. GANs models are inherently complex and require a deep understanding of their architecture and training process. Furthermore, leveraging GANs for data augmentation necessitated high-performance hardware, such as V100 GPUs, to handle the computational demands of training these models effectively. Additionally, GANs training is notorious for taking a long time to converge, further extending the project timeline and requiring patience and perseverance throughout the training process.

5.3 Future Works

Future endeavors could focus on training the GAN model with comprehensive datasets encompassing a wide range of alphabets, characters, and dyslexia-relevant features in handwriting images. By incorporating diverse and extensive datasets, the GANs can be further optimized to enhance their accuracy and robustness in generating handwriting word images and classifying dyslexia likelihood. This approach would contribute to a more comprehensive understanding of dyslexia detection and provide insights into improving the performance of automated dyslexia severity classification systems.

REFERENCES

Al-Barhamtoshy, H. M., & Motaweh, D. M. (2017). Diagnosis of Dyslexia using computation analysis. *2017 International Conference on Informatics, Health & Technology (ICIHT)*, 1–7.

Almisreb, A. A., Md Tahir, N., Turaev, S., Saleh, M. A., & Al Junid, S. A. M. (2022). Arabic Handwriting Classification using Deep Transfer Learning Techniques. *Pertanika Journal of Science & Technology*, *30*(1).

Alqahtani, N. D., Alzahrani, B., & Ramzan, M. S. (2023). Deep Learning Applications for Dyslexia Prediction. *Applied Sciences*, *13*(5), 2804.

Asvestopoulou, T., Manousaki, V., Psistakis, A., Nikolli, E., Andreadakis, V., Aslanides, I. M., Pantazis, Y., Smyrnakis, I., & Papadopouli, M. (2019). Towards a robust and accurate screening tool for dyslexia with data augmentation using GANs. *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, 775–782.

Bińkowski, M., Sutherland, D. J., Arbel, M., & Gretton, A. (2018). Demystifying mmd gans. *ArXiv Preprint ArXiv:1801.01401*.

Borji, A. (2019). Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding*, *179*, 41–65.

Carneiro, T., Da Nóbrega, R. V. M., Nepomuceno, T., Bian, G.-B., De Albuquerque, V. H. C., & Reboucas Filho, P. P. (2018). Performance analysis of google colaboratory as a tool for accelerating deep learning applications. *IEEE Access*, *6*, 61677–61685.

Driss, S. Ben, Soua, M., Kachouri, R., & Akil, M. (2017). A comparison study between MLP and convolutional neural network models for character recognition. *Real-Time Image and Video Processing 2017*, *10223*, 32–42.

Fogel, S., Averbuch-Elor, H., Cohen, S., Mazor, S., & Litman, R. (2020). Scrabblegan: Semi-supervised varying length handwritten text generation.

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4324–4333.

Gan, J., Wang, W., Leng, J., & Gao, X. (2022). HiGAN+: Handwriting Imitation GAN with Disentangled Representations. *ACM Transactions on Graphics (TOG)*, *42*(1), 1–17.

Heil, R., & Breznik, E. (2023). A Study of Augmentation Methods for Handwritten Stenography Recognition. *Iberian Conference on Pattern Recognition and Image Analysis*, 134–145.

International Dyslexia Association. (2021, July 26). *How common are language-based learning disabilities*. https://dyslexiaida.org/frequently-asked-questions-2/

Isa, I. S., Rahimi, W. N. S., Ramlan, S. A., & Sulaiman, S. N. (2019). Automated detection of dyslexia symptom based on handwriting image for primary school children. *Procedia Computer Science*, *163*, 440–449.

Isa, I. S., Zahir, M. A., Ramlan, S. A., Wang, L.-C., & Sulaiman, S. N. (2021). CNN comparisons models on dyslexia handwriting classification. *ESTEEM Academic Journal (EAJ)*, *17*, 12–25.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1), 1–54.

Kamath, C. N., Bukhari, S. S., & Dengel, A. (2018). Comparative study between traditional machine learning and deep learning approaches for text classification. *Proceedings of the ACM Symposium on Document Engineering 2018*, 1–11.

Kang, L., Riba, P., Wang, Y., Rusinol, M., Fornés, A., & Villegas, M. (2020). GANwriting: content-conditioned generation of styled handwritten word images. *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, 273–289.

Luo, C., Zhu, Y., Jin, L., Li, Z., & Peng, D. (2022). SLOGAN: handwriting style synthesis for arbitrary-length and out-of-vocabulary text. *IEEE Transactions on Neural Networks and Learning Systems*.

Nagpal, A., & Gabrani, G. (2019). Python for data analytics, scientific and technical applications. *2019 Amity International Conference on Artificial Intelligence (AICAI)*, 140–145.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *ArXiv Preprint ArXiv:1712.04621*.

Persatuan Dyslexia Malaysia. (2023). *Persatuan Dyslexia Malaysia*. https://www.dyslexiamalaysia.org/en/

Pippi, V., Quattrini, F., Cascianelli, S., & Cucchiara, R. (2023). HWD: A Novel Evaluation Score for Styled Handwritten Text Generation. *ArXiv Preprint ArXiv:2310.20316*.

Raatikainen, P., Hautala, J., Loberg, O., Kärkkäinen, T., Leppänen, P., & Nieminen, P. (2021). Detection of developmental dyslexia with machine learning using eye movement data. *Array*, *12*, 100087.

Rahmanian, M., & Shayegan, M. A. (2021). Handwriting-based gender and handedness classification using convolutional neural networks. *Multimedia Tools and Applications*, *80*, 35341–35364.

Rosli, M. S. A. Bin, Isa, I. S., Ramlan, S. A., Sulaiman, S. N., & Maruzuki, M. I. F. (2021a). Development of CNN transfer learning for Dyslexia handwriting recognition. *2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 194–199.

Rosli, M. S. A. Bin, Isa, I. S., Ramlan, S. A., Sulaiman, S. N., & Maruzuki, M. I. F. (2021b). Development of CNN transfer learning for Dyslexia handwriting recognition. *2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 194–199.

Saabith, S., Vinothraj, T., & Fareez, M. (2021). A review on Python libraries and Ides for Data Science. *Int. J. Res. Eng. Sci*, *9*(11), 36–53.

Snowling, M. J., Hulme, C., & Nation, K. (2020). Defining and understanding dyslexia: past, present and future. *Oxford Review of Education*, *46*(4), 501–513.

Spoon, K., Siek, K., Crandall, D., & Fillmore, M. (2019). Can we (and should we) use AI to detect Dyslexia in children's handwriting? *Proceedings of the International Conference on Machine Learning AI for Social Good Workshop*, 1–6.

Tamboer, P., Vorst, H. C. M., Ghebreab, S., & Scholte, H. S. (2016). Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia. *NeuroImage: Clinical*, *11*, 508–514.

Van Waelvelde, H., Hellinckx, T., Peersman, W., & Smits-Engelsman, B. C. M. (2012). SOS: a screening instrument to identify children with handwriting impairments. *Physical & Occupational Therapy in Pediatrics*, *32*(3), 306–319.

Wajuihian, S. O., & Naidoo, K. S. (2012). Dyslexia: An overview. *Optometry & Vision Development*, *43*(1).

Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognition Letters*, *141*, 61–67.

Yogarajah, P., & Bhushan, B. (2020). Deep learning approach to automated detection of dyslexia-dysgraphia. *The 25th IEEE International Conference on Pattern Recognition*.

Zahia, S., Garcia-Zapirain, B., Saralegui, I., & Fernandez-Ruanova, B. (2020). Dyslexia detection using 3D convolutional neural networks and functional magnetic resonance imaging. *Computer Methods and Programs in Biomedicine*, *197*, 105726.

APPENDIX A

Include in the appendices reference materials that are too lengthy for the main thesis. If the appendix includes tables and figures, label them as Table A.1, Figure A.1, and so on.

3.2 Activities Plan and Gantt Chart

The project will follow a systematic timeline and activities plan to ensure the successful completion of the dyslexia diagnosis system. The Gantt chart below outlines the key activities and their corresponding start dates, end dates, and duration:

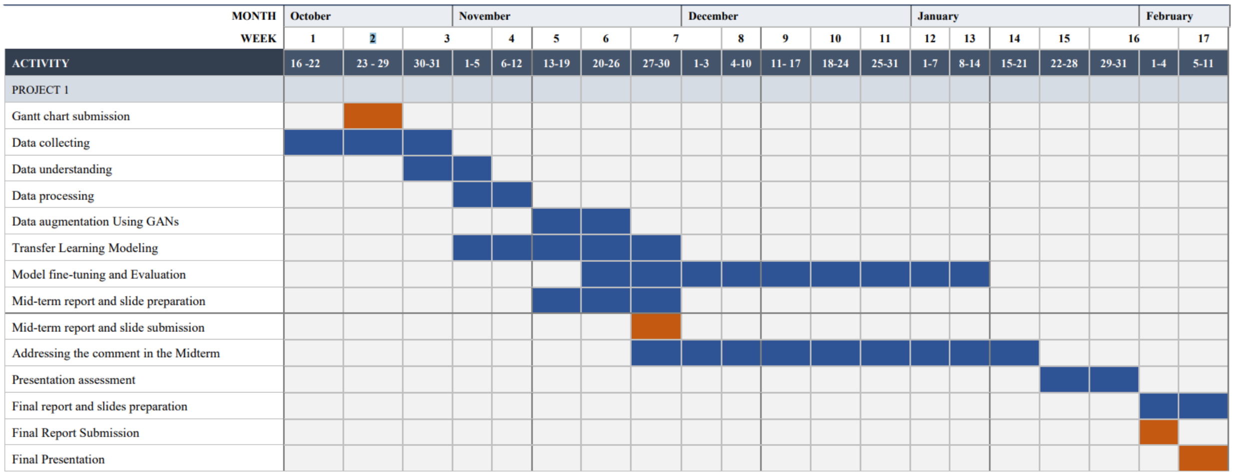


Figure A.8: Gantt Chart