# Machine Learning and Dyslexia: Diagnostic and Classification System (DCS) for Kids with Learning Disabilities

**Rehman Ullah Khan[1]\*, Julia Lee Ai Cheng[1], Oon Yin Bee[1]**

[1]*Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia*
*Corresponding author E-mail:krullah@unimas.my*

## Abstract

New generation is the future of every nation, but dyslexia which is a learning disability is spoiling the new generation. Most experts are using manual techniques to diagnose dyslexia. Machine learning algorithms are capable enough to learn the knowledge of experts and intelligently diagnose and classify dyslexics. This research proposes such a machine learning based diagnostic and classification system. The system is trained by human expert classified data of 857 school children scores in various tests. The data was collected in another fundamental research of designing special tests for dyslexics. Twenty-fifth percentile was used as threshold. The scores equal to the threshold and below were marked as indicators of children who were likely to have dyslexia while the scores above the threshold were considered to be indicators of children who were non-dyslexic. The system has three components: the diagnostic module is a pre-screening application that can be used by experts, trained users and parents for detecting the symptoms of dyslexia. The second module is classification, which classifies the kids into two groups, non-dyslexics and suspicious for dyslexia. A third module is an analysis tool for researchers. The results show that 20.7% of students seem to be dyslexic out of 257 in the testing data set which has confirmed by human expert.

*Keywords*: *Classification of dyslexics; Diagnosis of dyslexics; Dyslexia; Learning disabilities; Machine learning systems.*

## 1. Introduction

Dyslexia is a specific learning disability that is neurobiological in origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities [1]. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities and the provision of effective classroom instruction. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede the growth of vocabulary and background knowledge.

Malaysia is a relatively young nation, having a population of 31.66 million [2]. Five percent of the Malaysian population is primary school going children [2]. Education is one of the vehicles for achieving the country's vision. In Malaysia, 4-10 % of students have been found to have the signs and symptoms of dyslexia [3].

Despite the heterogeneity in reading disabilities, dyslexia has been the predominant focus in Malaysia[4]. Additionally, norm-referenced assessment for the Malaysian population is non-existent. Given that Malay is the national language and English is the lingua franca, bilingualism among Malaysians is expected [5]. Reading disabilities have detrimental consequences in children lives. All such children need help, but due to limited resources, logistics, and social stigma, they are unable to get the training that they require.

Therefore, there is a need for developing automated interventions to provide diagnosis to support children with reading disabilities including dyslexia. These automated interventions should be standardized, objective, repeatable, low-cost, and can be deployed outside of the clinic or at fingertips of the children. Therefore, we propose Diagnosis and Classification System to diagnose category and intensity of disability and then classify kids with learning disabilities based on the kids score from different tests. This system will provide the necessary literacy foundation in reading and writing starting at the earliest age possible (i.e., preschool). Given the uniqueness of the Malaysian education system that focuses on both English and Malay, the research and design of this Diagnostic and Classification System will contribute to the extant literature on technologies in learning difficulties, bilingualism, literacy diagnosis, and reading disabilities.

With this glaring need in research and practice, this research was conducted to provide an early diagnosis to schoolchildren in Malaysia, more specifically in a sample of children in Kuching, Sarawak. Given the lack of expertise in the area of diagnosis, the ultimate objectives of this research are

1. To develop and evaluate diagnostic and classification system for kids with learning disabilities.
2. To estimate dyslexia prevalence using predictions from the diagnostic and automated classification system.

## 2. Literature Review

Becoming literate is an important milestone in a child's scholarly life and a pathway to academic success into adulthood [6,7]. However, around the world and in Malaysia, there are children who continue to fall between the "cracks" and are left behind [8, 9]. Drawing from studies overseas, the researchers know that approximately 12% of the United States school population exhibit characteristics of reading disabilities that are heterogeneous [10]. Dyslexia affects about 10-15% of the school-age population [9]. Studies in Malaysia on dyslexia have reported similar findings [9].