Journal of Techniques, ISSN: 2708-8383, Vol. 6, No. 1, 2024, Pages 85-92
DOI: https://doi.org/10.51173/jt.v6i1.1893

JOURNAL OF TECHNIQUES

Journal homepage: http://journal.mtu.edu.iq

RESEARCH ARTICLE - ENGINEERING

Classification of Dyslexia Among School Students Using Deep Learning

Alia Hussein[1*], Ahmed Talib Abdulameer[1], Ali Abdulkarim[1], Husniza Husni[2], Dalia Al-Ubaidi[3]

[1]Technical College of Management - Baghdad, Middle Technical University, Baghdad, Iraq

[2]School of Computing, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

[3]Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor 81310, Malaysia

[*] Corresponding author E-mail: dac2005@mtu.edu.iq

Article Info.

Article history:

Received
15 August 2023

Accepted
12 December 2023

Publishing
31 March 2024

Abstract

Dyslexia is a common learning disorder that affects children's reading and writing skills. Early identification of Dyslexia is essential for providing appropriate interventions and support to affected children. Traditional methods of diagnosing Dyslexia often rely on subjective assessments and the expertise of specialists, leading to delays and potential inaccuracies in diagnosis. This study proposes a novel approach for diagnosing dyslexic children using spectrogram analysis and convolutional neural networks (CNNs). Spectrograms are visual representations of audio signals that provide detailed frequency and intensity information. CNNs are powerful deep-learning models capable of extracting complex patterns from data. In this research, raw audio signals from dyslexic and non-dyslexic children are transformed into spectrogram images. These images are then used as input for a CNN model trained on a large dataset of dyslexic and non-dyslexic samples. The CNN learns to automatically extract discriminative features from the spectrogram images and classify them into dyslexic and non-dyslexic categories. This study's results demonstrate the proposed approach's effectiveness in diagnosing dyslexic children. The CNN accurately identified dyslexic individuals based on the spectrogram features, outperforming traditional diagnostic methods. Spectrograms and CNNs provide a more objective and efficient approach to dyslexia diagnosis, enabling earlier intervention and support for affected children. This research contributes to the field of dyslexia diagnosis by harnessing the power of machine learning and audio analysis techniques. Facilitating faster and more accurate identification of Dyslexia in children, ultimately improving their educational outcomes and quality of life.

Publisher: Middle Technical University

Keywords: Dyslexia; CNN; Deep Learning; Types of Dyslexia; Spectrograms.

1. Introduction

Dyslexia is a learning disorder related to language, particularly affecting reading skills. It is the most prevalent reading difficulty, with approximately one in ten children having Dyslexia. According to the International Dyslexia Association, around 15 to 20% of the world's population exhibits symptoms of Dyslexia. This condition leads to difficulties in processing word-sounds and identifying letter-sounds. Individuals with Dyslexia may struggle with recognizing letter reversals and often get confused between different letters[1].

Consequently, Dyslexia results in challenges with phonics, leading to reduced reading proficiency, limited vocabulary development, and hindered acquisition of other language-related knowledge. Dyslexia is a learning disorder that affects children's reading and writing abilities. Early and accurate diagnosis of Dyslexia is crucial for providing appropriate interventions and support to affected children. Traditionally, diagnosing Dyslexia has relied on the expertise of specialists, which can be time-consuming and costly. However, advancements in technology, particularly in the field of machine learning, offer promising solutions for automating the diagnosis process [2, 3]. One such technological approach is the utilization of spectrograms and convolutional neural networks (CNNs) for diagnosing dyslexic children. Spectrograms are visual representations of sound that provide detailed information about the frequency and intensity of different audio signal components [4]. CNNs, on the other hand, are deep learning models that excel at extracting meaningful patterns and features from complex data. The combination of spectrograms and CNNs presents a powerful tool for analyzing audio data[5], such as speech signals, and identifying patterns associated with Dyslexia. By training CNNs on a large dataset of spectrograms from both dyslexic and non-dyslexic children, the network can learn to differentiate between the two groups based on distinct patterns present in the spectrogram images. The advantages of using spectrograms and CNNs for dyslexia diagnosis are twofold. Firstly, analyzing audio signals in the frequency domain allows for a more comprehensive assessment of the underlying auditory processing deficits associated with Dyslexia. Secondly, CNNs have shown remarkable capabilities in learning complex patterns and can leverage this ability to accurately classify dyslexic and non-dyslexic children based on the features extracted from spectrograms[1]. This research explores the effectiveness of spectrograms and CNNs for diagnosing dyslexic children. By transforming raw audio signals into spectrogram images and feeding them into a CNN model, we can investigate the model's performance in accurately identifying dyslexic individuals. The results of this study will contribute to the growing body of research on leveraging machine learning

Nomenclature & Symbols			
CNN	Convolution Neural Network	Adam	Adaptive Moment Estimation
ASR	Automatic Speech Recognition	LPD	Letter position dyslexia
ReUL	Rectified Linear Unit	STFT	Short Time Fourier Transform

techniques for dyslexia diagnosis. They may potentially lead to more accessible and efficient diagnostic tools for identifying Dyslexia in children. The main contribution of this paper is a framework (CNN model for deep learning) to diagnose early-stage dyslexic students. The recorded voices of the students are transformed into spectrograms. Then, a spectral plot is created, containing discriminative features for classifying between healthy readers and children with Dyslexia. This study focuses on Iraqi children who have Dyslexia and still face significant difficulties in basic reading, particularly word recognition. The participating children with Dyslexia in this study are elementary school students aged between 7 and 14 years, whose reading level has been determined and suggested by their teachers. The structure of this research is as follows: Introduction: The first paragraph serves as the introduction. Related Work: The second paragraph discusses some related works on dyslexia detection methods. Convolutional Neural Network (CNN) Explanation: The third paragraph provides a simplified explanation of the Convolutional Neural Network (CNN) algorithm used in the proposed model, followed by a second part explaining the Spectral Plot algorithm. Iraqi Dataset and Dyslexia Types: The fourth paragraph presents the Iraqi dataset and practical details on dyslexia types, which are being explored scientifically for the first time in Iraq and the Arab region. Research Experiment and Results: The fifth paragraph illustrates the research experiment and its outcomes. Conclusion: The sixth paragraph outlines the conclusions drawn from the research. This paper aims to diagnose Dyslexia, which is the identification of individuals who have difficulty in reading, writing, and spelling despite having intelligence and appropriate education. The overall objective of this study is to identify Dyslexia in children through a reading model for children who have Dyslexia in the Arabic language. Speech recognition (ASR) is important in enhancing children's interest in learning to read using the computer. The opportunity to use ASR technology is available to assist children, especially those with learning difficulties.

2. Related Works

Despite the limited research on speech recognition issues for individuals with speech impediments in English and the absence of such research in Arabic to the best of our knowledge, efforts have been exerted to enhance the accuracy of recognizing children's speech, addressing various aspects of this challenging task. These efforts encompass improvements in training data, feature extraction techniques, and effective model architecture. These are some previous studies related to Dyslexia:

In 2017, the research presented by Husniza Husni, et al. [6], Dyslexic children read with a lot of highly phonetically similar errors which is a challenge for speech recognition (ASR). Listening to highly phonetically similar errors is indeed difficult even for a human. To enable a computer to 'listen' to dyslexic children's reading is even more challenging as we have to 'teach' the computers to recognize the readings and adapt to the highly phonetically similar errors they make when reading. This is even more difficult when segmenting and labelling the read speech for processing before training an ASR. Hence, this paper presents and discusses the effects of highly phonetically similar errors on automatic transcription and segmentation accuracy and how the spoken pronunciations somehow influence it. Several 585 files of dyslexic children's reading are used for manual transcription, force alignment, and training. The recognition of the ASR engine using automatic transcription and phonetic labelling obtained an optimum result, with 23.9% WER and 18.1% FAR. The results are almost similar to the ASR engine using manual transcription 23.7% WER and 17.9% FAR. In the research introduced by Manoj Kaushik, et al. [7], a deep learning model is employed for early-stage diagnosis of Specific Language Impairment (SLI). The method involves using utterances containing seven distinct types of vocabulary. A thorough examination is conducted across various age groups. Moreover, the approach is not influenced by gender, age, or speaker, ensuring its independence. Remarkably, the achieved accuracy reaches an impressive 99.09%. In the research, Mahmoud Gharaibeh [8], Dyslexia is a condition characterized by difficulties in accurately and rapidly decoding words while reading. In this research, three assessment tools were developed, piloted, and evaluated for their validity and reliability: The Rapid Automatized Scale (RANS), the Arabic Reading Ability Scale (ARAS), and the Phonological Awareness Scale (PAS). These instruments were tested on a sample of 700 students aged 8 to 9 years old. The participants' results on the three instruments led to the formation of four groups: Double Deficit (DD), Rapid Automatized Naming Deficit (RAND), Phonological Awareness Deficit (PAD), and No Deficit The instruments were found to have content validity, with support from published reports, and the RANS was further revised by educational experts. The study discovered a significant negative correlation between Phonological Awareness (PA) test scores and RAN scores (mistakes and time) ($r = -.44$; $p < .001$) and a significant positive correlation between RAN mistakes and RAN time ($r = .47$; $p < .001$). The RANS demonstrated acceptable internal reliability with a Cronbach's alpha coefficient of $\alpha = .85$ ($> .70$, which is considered acceptable). Additionally, high inter-rater reliability was observed for all three instruments ($r \geq .86$, $p < .001$). Based on the findings, it was concluded that these three instruments are capable of predicting reading difficulties and Dyslexia in Arabic-speaking populations. In [5], this study aims to determine and classify the back-to-eye movement (retrieving words/re-reading) and skipping lines while reading from electrooculography (EOG) signals. For this aim, EOG signals were recorded while reading a text from healthy and dyslexic children. In this study, a method to assist in the diagnosis and follow-up of Dyslexia is proposed by determining skipping lines and back-to-eye movement (retrieving words/re-reading) while reading. Using the proposed method, skipping lines while reading and back to eye movement (retrieving words/re-reading movements) were determined from EOG signals, and spectrogram images of these movement signals were obtained using the Short Time Fourier Transform (STFT) method. These spectrogram images were classified using the 2-dimensional Convolutional Neural Network (2D-CNN) classifier. The 2D-CNN model has classified the skipping lines signals while reading and back-to-eye movement (retrieving words/re-reading) signals with 99% success. The findings show that the method proposed in diagnosing and following Dyslexia can give positive results using these EOG signals. In 2023, Shankar Parmar and Chirag Paunwala [9] show that Dyslexia is a neurological disorder affecting reading and writing abilities. The age range of 7–12 years old is ideal for detecting Dyslexia in its primitive stages. A method utilizing Electroencephalogram has been suggested to assist in the early identification of Dyslexia. Evaluating the effectiveness of different methods for detecting Dyslexia. The Wavelet Scatter Transform approach achieved high accuracy rates of 96.96% and 97.12% for two datasets.

3. Proposed Model

Diagnosing Dyslexia among school students using speech recognition techniques, CNN (Convolutional Neural Networks), and spectrogram analysis is an intriguing approach. ASR (Automatic Speech Recognition) involves converting spoken language into written text [10]. These techniques can be used to transcribe speech samples from school students, providing a textual representation of their spoken words.

3.1. Convolutional Neural Networks (CNNs)

CNNs are deep learning models known for their strong performance in pattern recognition and visual tasks[11]. In the context of speech recognition, CNNs can be trained on spectrogram representations of speech data, including the spectrogram itself. CNNs learn to extract meaningful features and patterns from the spectrogram, enabling accurate speech recognition and transcription [12]. Convolutional Neural Networks (CNNs) are considered one of the best algorithms used for language identification using Spectrograms of audio (Language Identification for Audio Spectrums). Spectrograms are used to represent raw audio as input to the CNN [13], which is then used for identifying recorded sounds. CNNs were chosen in this research due to the following advantages:

- Minimal pre-processing requirement: This method requires very little pre-processing. Raw audio data is directly fed into the neural network, with spectrogram representations being formed as each batch is fed into the network.
- Ability to work with short audio segments: Another advantage is that the technique can work with short audio clips (around one or two seconds) for efficient classification. This is crucial for voice assistants who need to identify Dyslexia among school students.
- Convolutional neural networks (CNNs) compare the image part by part. So they finish the similarity better than whole-picture matching schemes[14].

3.1.1. Tensor flow

To obtain satisfactory outcomes from the model, it is necessary to pre-process the input data. Pre-processing involves cleaning and adapting the data to make it suitable for the model. Data augmentation is a frequently used technique in image pre-processing, which allows us to increase the dataset size by utilizing the existing data. To begin with, a crucial initial stage involves data normalization. Normalization refers to the process of adjusting data values to a standardized scale. In our specific case, we will pre-process our images by normalizing the pixel values from 0 to 1. Initially, the pixel values are within the range of 0 to 255.

3.1.2. Activation functions

There are several different activation functions. One of the commonly used functions is the ((Rectified Linear Unit) ReLU). It has been proven that the ReLU function generally performs better than other activation functions and is widely used today [15]. The definition of the ReLU activation function is shown below in the equation:

$$\boldsymbol{f(x)} = \begin{cases} 0 \text{ for } x < 0 \\ 1 \text{ for } x \geq 0 \end{cases} \quad (1)$$

Another activation function is SoftMax, which is commonly used for the output layers of neural networks. The SoftMax function takes the activations of all n neurons in the layer and creates a probability distribution consisting of n probabilities. Given an input vector x containing the activations of each neuron, the SoftMax function, denoted as softmax(x), produces a vector of the same length as x containing the computed probabilities[16, 17]. The activation in neural cells is defined using the SoftMax function.

$$\boldsymbol{\sigma(x)i} = \frac{e^{xi}}{\sum_{j=1}^{n} e^{xj}} \quad (2)$$

$$\boldsymbol{i} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{n \; and \; x} = (\boldsymbol{x1}, \boldsymbol{x2}, \dots, \boldsymbol{xn}) \in \boldsymbol{Rn}$$

3.2. Spectrogram analysis

Spectrogram analysis refers to examining the frequency content of an audio signal using spectrograms. By analyzing spectrogram representations, which capture frequency components over time, it becomes possible to identify specific patterns and characteristics related to Dyslexia. Variations or deviations in the spectrogram of students with Dyslexia compared to those without Dyslexia can provide valuable insights for diagnosis [18].

By combining speech recognition techniques, CNNs trained on spectrogram representations, and spectrogram analysis, it is possible to develop a system capable of transcribing spoken words and analyzing spectrogram patterns associated with Dyslexia. This approach can help in early detection and diagnosis of Dyslexia among school students, enabling appropriate intervention and support [19]. There are several methods to convert audio into a spectrogram. One common approach is to use a programming library like NumPy, which makes it easy to create spectrograms. The audio signal is converted into a spectrogram as shown in Fig. 1.

3.3. Long Short-Term Memory (LSTM)

networks can be applied in multiple ways to support individuals with Dyslexia, particularly in the context of language processing and reading difficulties. There are several ways in which LSTM can assist individuals with Dyslexia, and one of them is Predictive Text Tools. LSTM models can be used to create predictive text tools. These tools can predict and suggest the next word a user intends to type based on their previous typing patterns. For dyslexic individuals who struggle with spelling and word prediction, this can be a valuable aid [20].

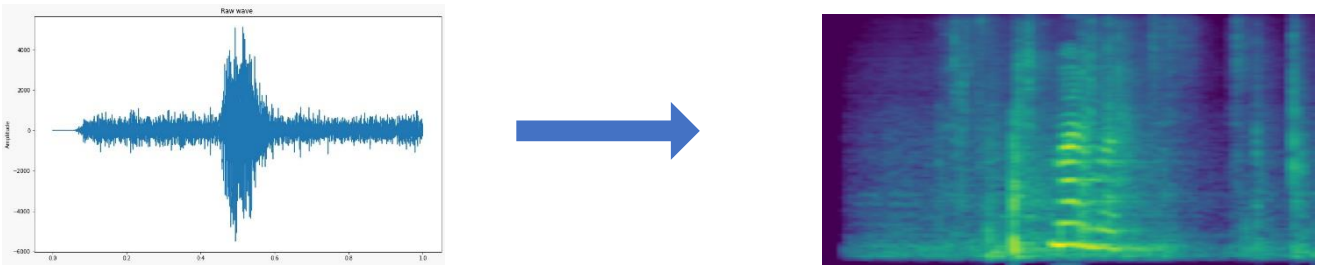

Fig. 1. Show the conversion of the audio signal into a spectrogram

4. Data Set

This study focuses on Iraqi children who have difficulties in reading and struggle with word recognition. The participants in the study are dyslexic children in elementary school between the ages of 7 and 14, who have a similar reading level as determined and suggested by their teachers. Most automatic speech recognition applications use readings from children ranging from 5 to 18 years old. Previous studies have included several children ranging from a few to several thousand. Therefore, the participation of 20 dyslexic children and 80 non-dyslexic children in this research is considered representative based on different studies. The participants were selected from three primary schools: Al-Kawthar School in the Ghazaliya area, Al-Asirah School, and Abu Hanifa Al-Numen School in the Amiriya area. These schools are among the elementary schools that provide special dyslexia classrooms, which is a pioneering program launched and fully supported by the Ministry of Education.

The study revolves around facilitating the learning process for dyslexic children in Arabic, specifically in word recognition. Word recognition is the foundational level of reading, where beginners typically recognize simple Arabic words during the learning process. Therefore, considering that this study is related to dyslexic children whose reading level corresponds to or resembles that of 6 to 8-year-olds, word recognition is most relevant in this context. This research identifies, for the first time, the types of developmental Dyslexia in the Arabic language based on recorded voices of children who have Dyslexia. These types include Dyslexia related to letter position, attentional Dyslexia, visual Dyslexia, neglect dyslexia, surface dyslexia, phonological Dyslexia, and deep Dyslexia [21]. Developmental Dyslexia has many forms, among which is shown in Table 1.

Table 1. Utterances in the database used

S. No	Description	Size of class	Language	Utterances
1	Letter position dyslexia LPD	10	Arabic Dyslexia of the Arabic language	"همام"،"اصطفوا", 'أمتحان', 'بحار', 'ضربه', 'ضرير', 'ضيع', 'فادي', 'فتحه', 'فلفل" "هامم"،"اصفطوا"،"امحتان"،"باحر"،"ضبره"،"ضيرر"،"ضعي"،"فايد"،"فحته"،"لفلف"
2	Deep dyslexia	10	Arabic Dyslexia of the Arabic language	"قطر"،"مرمر "،"ضوء"،"ضرير"،"ضار"،"سلاح"، "بابا"، "الشمس"، "السادس"، "اذا" "مقطور"،"تمر"،"وضوء"،"سرير"،"فلاح"،"ماما"،"المشمش"،"مسدس"،"اذون"
3	Neglect dyslexia	10	Arabic Dyslexia of the Arabic language	"جمعة"، "جمع" ، "مشمش" ، "جاء"، "حصار" ، "حرزتم ،"حراره" "اكلت"،"جسرين""جرح" "نجرح"،"اكلتم"،"جسر"،"جمع"،"شمش"،"جاءت"،"حصارت"،"حرز"،"حرار"
4	Surface dyslexia	10	Arabic Dyslexia of the Arabic language	"لقلق"،"كي"،كوب" "كتكوت"،"كتان"،"كسره"،"كرم"، "كره"، "كرار"، "كمال" "لقلقه"،"كي"،"كوبه"،"كتوة"،"كتانه"،"كسر"،"كرمة"،"كر"،"كراره"،"كمال"
5	Vowel letter dyslexia	10	Arabic Dyslexia of the Arabic language	"عمان"، جاهز"موجود"،"معطف"،"مقطع"، "مقر"،"ملاك" ,"مكروهان" ,"حبي" ,"امل" "عامان"،"جهاز"،"مجود"،"معطوف،"مقاطع"،"مقير"،"ملك"،"مكروهون"،"حبو"،"امال"

5. Experiments

The system is designed in stages to handle the sequential processing of sound and feature extraction using the spectrogram algorithm. The extracted features are stored as spectral image files (PNG) to represent the sound properties. The next stage involves building the proposed system as a classifier using the Convolutional Neural Network (CNN) algorithm. The training phase begins until satisfactory results are achieved. The system stores the parameters for use in the language prediction process during the testing phase. Python programming language was used in building the system, as it is easy to learn and open source. The proposed system includes the following basic stages:

- First stage: Initial processing of audio files
- Second stage: Preparing the database for the audio files of the speakers
- Third stage: Feature extraction using the spectrogram algorithm.
- Fourth Stage: Building the classifier and predicting the speaker's language using Long Short-Term Memory (LSTM) algorithms[22]. The structure of the proposed system is illustrated in the Fig. 2.

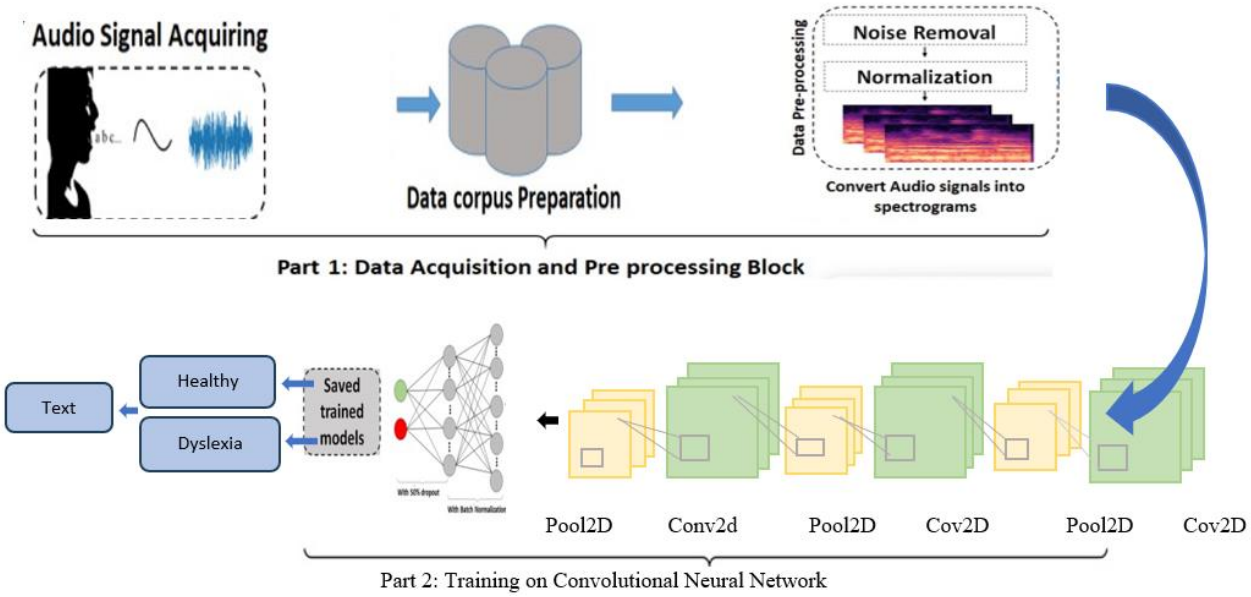

Fig. 2. Speech-to-text model architecture

5.1. Audio recording and pre-processing

Pre-processing the data is a fundamental step in identifying Dyslexia among school students because it ensures that the data is well-prepared for specific types of analysis. In this research, the pre-processing was done in two steps:

- Removing silence: Eliminating silent portions from the audio files, to create a robust and efficient system for identifying the speaker's language, the first step in the initial processing involves eliminating silent segments from the audio files. This is a critical task as it allows us to extract relevant features solely from the audio content, simplifying the language identification process and reducing computational complexity. The removal of silent periods is achieved by segmenting the audio file into smaller audio frames, each lasting 5 milliseconds. Frames with a sound intensity below -30 decibels are removed, while frames with a sound intensity exceeding -30 decibels are retained. Consequently, a final file is generated, devoid of any silent intervals.
- Standardizing the file length to one or two seconds, the database contains audio files with varying durations. Therefore, after removing silent periods from the audio, it was necessary to proceed with the second step of the initial processing, which involves making all the audio files have the same length for consistent sample sizes and achieving the desired outcomes.to achieve this, the audio files were standardized, and a specific length of two seconds was designated for each audio file. If an audio file's duration exceeded two seconds, only the initial two seconds were kept. Conversely, if an audio file was shorter than two seconds, it was duplicated until it reached the required length.

5.2. Experimental result

This is a large-scale study that examined the topic of Dyslexia in the Arabic language. The experiment was conducted using the Spectrogram features of each audio file, where only 10 short words were used for each model. The dataset consisted of 8000 audio files for the 'Speech-To-Text' model. Five CNN models were trained specifically for processing Arabic words, with the first model being trained on the following word:

['أصطفوا', 'أمتحان', 'بحار', 'ضربه', 'ضرير', 'ضيع', 'فادي', 'فتحه', 'فلفل', 'همام']

The dataset is categorized into two groups based on the labels: children with Dyslexia and healthy children. The database contains a total of 8000 audio utterance clips, with 6400 clips belonging to healthy children and 1600 clips belonging to children with Dyslexia. To prepare the data for the 2D-CNN architecture, the entire audio clip dataset is transformed into a spectrogram image dataset. The performance evaluation of the dyslexia-trained model involves analyzing various standard statistical metrics, including accuracy, sensitivity (recall), specificity, precision, and f-1 score. These metrics are computed based on the model's predicted outcomes, which can be categorized into four groups: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP corresponds to the model's accurate identification of Dyslexia in the audio utterances, while TN represents the model's correct identification of healthy audio utterances without dyslexia patterns. FP denotes cases where the model incorrectly predicts Dyslexia in healthy audio, and FN indicates instances where the model wrongly predicts a healthy audio as having Dyslexia. By examining these statistical measures, the performance and effectiveness of the dyslexia-trained model are assessed. A comprehensive analysis is conducted using the predicted scores from the proposed dyslexia identification model. The different scores, based on which the models' performance is analyzed, are shown in the equations below. The different scores on which models' performance is analyzed are shown in equations 3-7.

$$Accuracy\ (Acc) = \frac{Total\ correct\ predictions}{Total\ Number\ of\ subjects} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (4)$$

$$\text{Recall or Sensitivity} = \frac{(TP)}{(TP+FN)} \quad (5)$$

$$\text{Specificity} = \frac{(TN)}{(TN+FP)} \tag{6}$$

$$\text{f1 Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{7}$$

This study represents a pioneering research journey, where these carefully selected short and vital Arabic words were chosen to understand the challenges and intricacies of Dyslexia in the context of Arabic speech processing. The main goal of the 'Speech-To-Text' model was to accurately convert spoken Arabic words into textual representations. Such a capability finds numerous real-world applications, including voice recording services, language learning tools, and improving accessibility for individuals with Dyslexia. It's worth noting that the successful application of these CNN models with Spectrogram features on this small but carefully curated set of words lays the foundation for future efforts with larger and more diverse datasets, thus advancing the field of Arabic speech recognition and dyslexia research.

5.3. Data distribution

In the experiments of the algorithm classification, the 2D spectral data of ten short words for each model, by hundreds of different people. We divided the 2D spectral data into the training set, the verification set, and the testing set. The number of training sets is about three times that of the testing set.

Each spectrogram size: (775,308)

5.4. Train result

We trained our proposed CNN models, and the results for model one is presented in Figs. 3 & 4. From Figure A, we can see that the loss of the training and validation sets is always between 2.3 and 3.5, and there are no irregular up-and-down violent fluctuations. Both the training and validating accuracy are rising and eventually reaching a stable value; there is no longer a trend of large value changes. If we increase the epoch, the training loss and the validating loss gradually become smaller and eventually stabilize. To sum up, our proposed CNN can overcome the vanishing gradient in training and validation and can fully extract features of spectral data from end to end, which is conducive to the correct classification of spectra.

5.5. Cross-validation result analysis

The Adam optimization algorithm is utilized with a learning rate of 0.01. The network's architecture consists of a total of 56,127,946 trainable parameters. To improve the variety of healthy and Dyslexia spectrograms, the audio signals are first converted into dB-scaled spectrograms, and then shuffling is applied to the data.

During training, each fold undergoes 100 epochs, with a standard batch size of 32. The model achieves impressive results, with an average training accuracy of 99 % and an average testing accuracy of 99 %. The obtained average training loss is 0.031039, while the average testing loss is 0.051909.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 773, 306, 32)      896

 max_pooling2d (MaxPooling2D  (None, 386, 153, 32)     0
 )

 conv2d_1 (Conv2D)           (None, 384, 151, 64)      18496

 max_pooling2d_1 (MaxPooling  (None, 192, 75, 64)      0
 2D)

 conv2d_2 (Conv2D)           (None, 190, 73, 128)      73856

 max_pooling2d_2 (MaxPooling  (None, 95, 36, 128)      0
 2D)

 flatten (Flatten)           (None, 437760)            0

 dense (Dense)               (None, 128)               56033408

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 10)                1290

=================================================================
Total params: 56,127,946
Trainable params: 56,127,946
Non-trainable params: 0
_________________________________________________________________
```

Fig. 3. Speech-to-text model summary

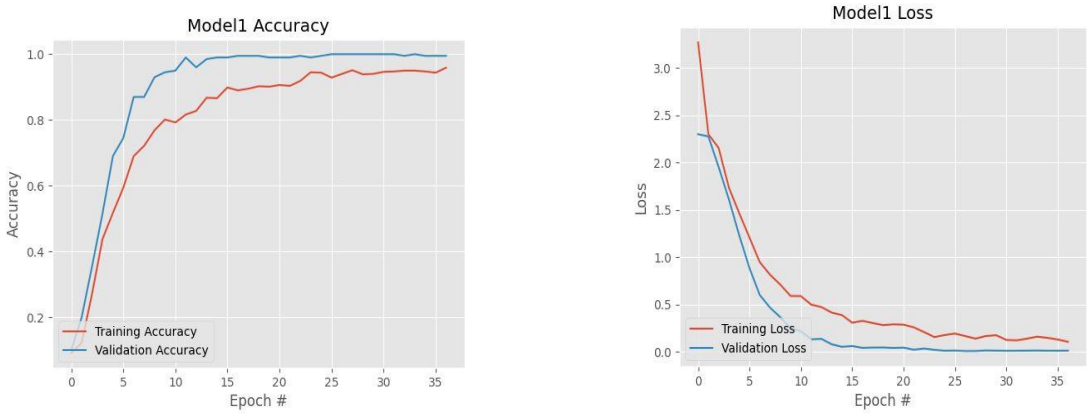

Fig. 4. Training and validation accuracy and loss

6. Results and Discussion

After we trained our models as a classifier in Python, the truth matrix (Confusion Matrix) can be created using the "Scikit-Learn" module.

By visualizing the confusion matrix, as we can see in Fig. 5, we can observe the model's performance and assess its accuracy by analyzing the diagonal values, which correspond to the number of correct classifications. This allows us to determine how well our model is performing and how accurate its predictions are which is shown in Table 2.

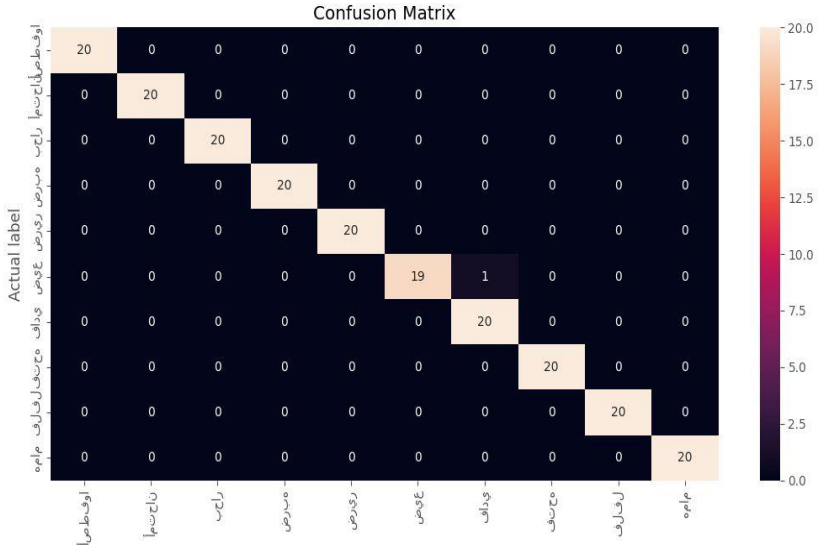

Fig. 5. Confusion matrix for the proposed model

Table 2. Evaluation Metrics for the Proposed Model

Model	Accuracy%	Val-Accuracy%	f1-Score %
1	98	99	99
2	99	100	100
3	95	100	100
4	98	100	100
5	97	98	98

7. Conclusion

In this study, we used Convolutional Neural Networks (CNNs) to automatically diagnose Dyslexia in children through speech signals that may have a role in this field. Recorded speech sounds were used and converted into images using a spectrogram to form the input signals for the CNN. It was found that the CNN could recognize speech signals for children with learning difficulties with higher accuracy using training data. These results suggest using CNNs to analyze speech signals to recognize children with learning difficulties. It has been found that the convolutional neural network can recognize speech signals for children with learning difficulties with a test accuracy of up to 99% using training data. In the end, we successfully constructed our speech-to-text models capable of recognizing basic voice commands. Given vast data and powerful computational systems, we can create even more advanced models, yielding superior outcomes.

Acknowledgement

This study was submitted as a component of the prerequisites for earning a Master's degree in Information Technology Management. I wish to seize this moment to extend my appreciation to all those who aided and provided assistance during the completion of this project, which underscores the academic excellence of my esteemed university.

References

[1] K. Singh, V. Goyal, and P. Rana, "Existing Assistive Techniques for Dyslexics: A Systematic Review," Artificial Intelligence for Accurate Analysis and Detection of Autism Spectrum Disorder, pp. 94-104, 2021, doi: 10.4018/978-1-7998-7460-7.ch007.

[2] R. Werth, "Dyslexia: Causes and Concomitant Impairments," Brain Sciences, vol. 13, no. 3, p. 472, 2023, doi: https://doi.org/10.3390/brainsci13030472.

[3] J. Stein, "Theories about developmental dyslexia," Brain Sciences, vol. 13, no. 2, p. 208, 2023, doi:org/10.3390/brainsci13030472.

[4] O. Pronina and O. Piatykop, "The recognition of speech defects using convolutional neural network," in CTE Workshop Proceedings, 2023, vol. 10, pp. 153-166, doi: DOI: https://doi.org/10.55056/cte.554

[5] F. Latifoğlu, R. İleri, and E. Demirci, "Assessment of dyslexic children with EOG signals: Determining retrieving words/re-reading and skipping lines using convolutional neural networks," Chaos, Solitons & Fractals, vol. 145, p. 110721, 2021, doi: https://doi.org/10.1016/j.chaos.2021.110721.

[6] Husni, Husniza, Nik Nurhidayat Nik Him, Mohamad M. Radi, Yuhanis Yusof, and Siti Sakira Kamaruddin. "Automatic transcription and segmentation accuracy of dyslexic children's speech." In AIP Conference Proceedings, vol. 1891, no. 1. AIP Publishing, 2017., doi:https://doi.org/10.1063/1.5005387.

[7] M. Kaushik, N. Baghel, R. Burget, C. M. Travieso, and M. K. Dutta, "SLINet: Dysphasia detection in children using deep neural network," Biomedical Signal Processing and Control, vol. 68, p. 102798, 2021, doi: https://doi.org/10.1016/j.bspc.2021.102798.

[8] M. Gharaibeh, "Predicting dyslexia in Arabic-speaking children: Developing instruments and estimating their psychometric indices," Dyslexia, vol. 27, no. 4, pp. 436-451, 2021, doi: https://doi.org/10.1002/dys.1682.

[9] S. Parmar and C. Paunwala, "A novel and efficient Wavelet Scattering Transform approach for primitive-stage dyslexia-detection using electroencephalogram signals," Healthcare Analytics, vol. 3, p. 100194, 2023, doi: https://doi.org/10.1016/j.health.2023.100194.

[10] M. Khudhair and A. Talib, "Improving Low Resources Arabic Speech Recognition using Data Augmentation," in 2022 Fifth College of Science International Conference of Recent Trends in Information Technology (CSCTIT), 2022: IEEE, pp. 60-65, doi: DOI: 10.1109/CSCTIT56299.2022.10145613.

[11] O. L. Usman, R. C. Muniyandi, K. Omar, and M. Mohamad, "Advance machine learning methods for dyslexia biomarker detection: A review of implementation details and challenges," IEEE Access, vol. 9, pp. 36879-36897, 2021, doi: DOI: 10.1109/ACCESS.2021.3062709.

[12] I. S. Isa, M. A. Zahir, S. A. Ramlan, L.-C. Wang, and S. N. Sulaiman, "CNN comparisons models on dyslexia handwriting classification," ESTEEM Academic Journal (EAJ), vol. 17, pp. 12-25, 2021.

[13] A. M. Badshah et al., "Deep features-based speech emotion recognition for smart affective services," Multimedia Tools and Applications, vol. 78, pp. 5571-5589, 2019, doi: https://doi.org/10.1007/s11042-016-4041-7.

[14] S. Revay and M. Teschke, "Multiclass language identification using deep learning on spectral images of audio signals," arXiv preprint arXiv:1905.04348, 2019, doi:https://doi.org/10.48550/arXiv.1905.04348.

[15] F. Ramo and M. N. Kannah, "Intelligence System for Multi-Language Recognition," Journal of Education and Science, vol. 31, no. 1, pp. 93-110, 2022, doi: DOI: 10.33899/edusj.2021.129868.1156.

[16] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in 2013 IEEE international conference on acoustics, speech and signal processing, 2013: IEEE, pp. 8609-8613, doi: 10.1109/ICASSP.2013.6639346.

[17] N. L. Hakim, T. K. Shih, S. P. Kasthuri Arachchi, W. Aditya, Y.-C. Chen, and C.-Y. Lin, "Dynamic hand gesture recognition using 3DCNN and LSTM with FSM context-aware model," Sensors, vol. 19, no. 24, p. 5429, 2019, doi: https://doi.org/10.1002/dys.1682.

[18] G. Atkar and P. Jayaraju, "Speech synthesis using generative adversarial network for improving readability of Hindi words to recuperate from dyslexia," Neural Computing and Applications, vol. 33, pp. 9353-9362, 2021, DOI: 10.33899/edusj.2021.129868.1156.

[19] Rochford, M. Visual Speech Recognition Using a 3D Convolutional Neural Network (Doctoral dissertation, California Polytechnic State University), 2019, doi: https://doi.org/10.15368/theses.2020.7.

[20] Jasira, K. T., & Laila, V. DyslexiScan: A Dyslexia Detection Method from Handwriting Using CNN LSTM Model. In 2023 International Conference on Innovations in Engineering and Technology (ICIET) (pp. 1-6). IEEE, , DOI: 10.1109/ICIET57285.2023.10220750.

[21] N. Friedmann and M. Haddad-Hanna, "Types of developmental dyslexia in Arabic," Handbook of Arabic literacy: Insights and perspectives, pp. 119-151, 2014, (2023, July), doi: DOI 10.1007/978-94-017-8545-7_6.

[22] M. Sameer, A. Talib, A. Hussein, and H. Husni, "Arabic Speech Recognition Based on Encoder-Decoder Architecture of Transformer," Journal of Techniques, vol. 5, no. 1, pp. 176-183, 2023, DOI: https://doi.org/10.51173/jt.v5i1.749 .