Recibido: 2 agosto de 2021
Aceptado: 3 octubre de 2021

DETECCIÓN DE DIFICULTADES EN LA LECTURA UTILIZANDO RECONOCIMIENTO DE PATRONES EN EL HABLA PARA LA DETECCIÓN TEMPRANA DE LA DISLEXIA

DETECTION OF ERRORS IN THE READING-ALOUD UTILIZING PATTERN RECOGNITION IN SPEECH FOR THE EARLY DETECTION OF DYSLEXIA

B.Eng. Carlos Quintana, PhD. Orlando Maldonado, MSc. Gladys Quintana

Universidad de Pamplona
Ciudadela Universitaria. Pamplona, Norte de Santander, Colombia.
Tel.: 57-7-5685303, Fax: 57-7-5685303, Ext. 144
E-mail: carlos.quintana@unipamplona.edu.co, orlmaldonado@unipamplona.edu.co, gsquintanaf@unipamplona.edu.co

Resumen: En el presente artículo se presentan los resultados de una comparación de tres técnicas de reconocimiento de patrones aplicados en la clasificación de palabras aisladas que puedan simular errores comunes que ocurran en la lectura en voz alta de niños que padecen de dislexia en la etapa inicial de la lectura. Se hizo una base de datos con grabaciones extraídas de cinco locutores diferentes y muestreadas a una tasa de 16 kHz, la cual fue extendida utilizando aumento artificial de los datos. Las grabaciones fueron preprocesadas para eliminar el ruido y de estas fueron extraídos los MFCC. Se compara el rendimiento de los modelos de K-vecinos más cercanos (KNN), Perceptrón MultiCapa (MLP) y Redes Neuronales Convolucionales (CNN); obteniendo resultados del orden de 99 puntos para el reconocimiento de palabras distintivas, y resultados del orden de los 70 puntos para palabras fonéticamente similares.

Palabras clave: Dislexia, Reconocimiento del Habla, Reconocimiento de Patrones,Redes Neuronales

Abstract: This article presents the results of a comparison between three pattern recognition techniques applied to the classification of isolated words that simulates common errors that occur in the reading-aloud of children that manifest early stages of dyslexia. A database of recordings was made from five different speakers and sampled at 16 kHz, and later extended utilizing artificial data augmentation. The recordings were preprocessed to eliminate noise and the MFCC were later extracted. We compare the performance of the K-nearest neighbors (KNN), Multi-Layer Perceptron (MLP) and Convolutional Neural Networks (CNN), obtaining results in the order of 99 points for the recognition of phonetically distinctive words, and results in the order of 70 points for phonetically similar words.

Keywords: Dyslexia, Speech Recognition, Pattern Recognition, Neural Networks.

.

1. INTRODUCCIÓN

La dislexia en la etapa inicial de la lecturaes una dificultad de aprendizaje que no permite establecer un buen proceso lector, dificultando la construcción de lenguaje para establecer procesos de comunicación adecuados; se presenta en los niños de población promedio entre los seis y siete años, correspondientes a los grados de Primero y Segundo de Básica Primaria que no padecen compromisos cognitivos.

De acuerdo a la Asociación Internacional de la Dislexia, se estima que entre un 15 y 20 % de la población padece de una dificultad de aprendizaje del lenguaje.

La dislexia es la causa más común de dificultades en la lectura, la escritura y el deletreo, y se vuelve significativamente más difícil de tratar a más tiempo se deje sin atender. Esta condición se encuentra entre las principales causas de desmotivación y baja autoestima, así como fracaso y deserción escolar.

Una forma tradicional de realizar la detección de esta condición requiere de un profesional que haga extensas pruebas de lectura en voz alta a los niños, donde analice detenidamente los sonidos pronunciados porellos en búsqueda de algún error propio de esta condición, tras lo cual se le formulará al niño un procedimiento de apoyo basado en herramientas pedagógicas. Sin embargo, esta condición suele transcurrir desatendida, pues el maestro de aula regular pocas veces se detiene a intervenir pedagógicamente esta dificultad del aprendizaje al no contar con el conocimiento y/o los recursos que le permitan hacer una detección temprana de los problemas que el niño manifiesta en la lectura.

Este artículo resume los resultados del estudio realizado en Quintana (2021) acerca de la comparación del rendimiento de distintas técnicas de reconocimiento de patrones sobre unaserie de grabaciones de palabras aisladas que pudieran simular la lectura en voz alta de niños que presenten dislexia en la etapa inicial de la lectura y cometan alguno de los errores más comunes según Bastos (1983) y Fernández *et al*. (2006) (omisión o inserción de letras o sílabas, inversión de letras o sílabas) y basados en una prueba de lectura realizada por el programa de Licenciatura en Educación Infantil de la Universidad de Pamplona (Alvarado *et al*., 2019). Ejemplos de estos errores se presentan en la Tabla 1.

Tabla 1. Errores en la lectura según Bastos (1983) y ejemplos

Error	Ejemplos
Omisión	i**n**stituto → istituto
Inserción	maestro → ma**s**estro
Inversión estética	barco → **d**arco
Inversión silábica	arbol → **ra**bol
	estomago → esto**gamo**

Durante la realización de este proyecto fue decretada y mantenida la emergencia sanitaria debido a la pandemia de COVID-19. Buscando minimizar el contacto entre individuos, se realizaron grabaciones de los participantes del proyecto y voluntarios cercanos,de manera que se tuviesen conocimientos conclusivos para determinar la factibilidad de escalar el proyecto a pruebas con niños en el rango de edad estudiado.

2. RECONOCIMIENTO DE PATRONES

El reconocimiento de patrones es el proceso mediante el cual un sistema puede detectar y extraer patrones de un conjunto de datos utilizando algoritmos de aprendizaje de máquina (*machine learning*). Este puede definirse como la clasificación de la información en base a un conocimiento previamente obtenido, o en información estadística extraída de patrones y su representación.

El estudio de estos algoritmos ha venido siendo perfeccionado desde la década de los 1950s, y mantiene una clara distinción entre algoritmos *supervisados* y *no supervisados*.

El aprendizaje supervisado requiere presentar primero al sistema con muestras representativas y etiquetadas en una etapa denominada entrenamiento, para que luego este utilice el conocimiento adquirido y pueda catalogar nuevas muestras nunca antes vistas. Los primeros algoritmos de esta clase fueron desarrollados sobre modelos no paramétricos, como el K-Vecinos más cercanos (KNN), y para los años 1980s comenzaban a cobrar fuerza el paradigma tecnológico de las Redes Neuronales Artificiales.

Este algoritmo **KNN** basa su proceso de reconocimiento en criterios de distancia (vecindad) con otras entradas multidimensionales, asignando al resultado la clase a la que pertenezcan aquellos elementosmás cercanos. La variable k representa la cantidad de elementos cercanos que serán tenidos en cuenta para esta decisión (Berástegui y Galar, 2018).

Por su parte los modelos de Redes Neuronales Artificiales toman inspiración del sistema nervioso humano y su capacidad para memorizar y asociar hechos, de manera que pueda resolver problemas acudiendo a la experiencia adquirida (Camacho, 2016).

Usualmente estas se componen de múltiples unidades lógicas denominadas *perceptrones* o neuronas, organizadas en forma de capas (Fig. 1), de manera que cada una de las neuronas realiza un cálculo basada en las neuronas de la capa anterior.

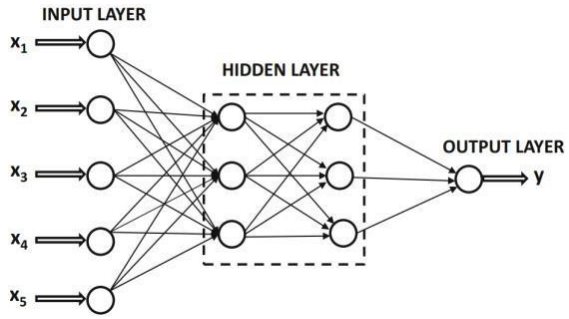

Fig. 1. Red neuronal multicapa con dos capas ocultas. Tomado de Aggarwal (2018).

El Perceptrón Multi-Capa (**MLP**) es la forma más básica que toma una red neuronal basada en capas. Los cálculos ocurren en las capas intermedias, que son llamadas capas ocultas.

El entrenamiento de una red neuronal se basa en el algoritmo de propagación hacia atrás o *backpropagation*, el cual se apoya en la regla de la cadena del cálculo diferencial, que calcula los gradientes de error en términos de sumas de productos de gradientes locales sobre los distintos caminos desde uno de los nodos o neuronas hacia la salida.

Las Redes Neuronales Convolucionales (**CNN**) se pueden considerar como una extensión de los MLP, las cuales fueron ideadas para el procesamiento de imágenes y otros datos con forma de retícula o matriz y fuertes dependencias locales en estas. Estos modelos toman su inspiración del funcionamiento biológico de la corteza visual y la detección de elementos visuales como bordes y orientación, y son quizás uno de los primeros grandes logros del campo del aprendizaje profundo.

La forma en que procesan los datos se basa en la especialización de sus componentes y el orden jerárquico de sus capas, así, y tomando como ejemplo la visión artificial, las primeras capas de una red pueden identificar líneas rectas que quizás impliquen bordes, las siguientes construyen sobre estas y pueden detectar esquinas y formas geométricas, y asísucesivamente. Esto se logra al reducir la dimensionalidad de los datos de entradamediante sus dos operaciones principales: convolución y agrupamiento (*convolution* y *pooling*). Estas capas se adicionan al modelo Multi-Capa haciendo un tipo depreprocesamiento de la información de entrada de manera que se optimice el proceso declasificación posterior.

Estas redes son utilizadas en gran medida para aplicaciones de reconocimiento de imágenes y visión artificial, pero otro tipo de información que comparte una estructura secuencial como texto, series temporales y otros datos espacio-temporales pueden representarse como casos especiales de este tipo de estructuras y ser procesados en una CNN, y trabajos previos como Abdel-Hamid *et al.* (2012), Huang *et al.* (2015) y Kubanek *et al.* (2019) han demostrado que este puede ser un método viable para clasificar señales de audio en forma espectrográfica.

3. CARACTERIZACIÓN DE LAS SEÑALES DE VOZ

La voz es una onda sonora que producen las cuerdas vocales, y es alterada por los demás órganos que conforman el aparato fonador humano (Hidalgo y Quilis, 2012).

Para adquirir las muestras de voz se utiliza un micrófono que transforme las ondas analógicas en señales digitales, y se han de tener en cuenta algunas características de la señal de entrada como el formato de la codificación de la muestra y la frecuencia de muestreo.

3.1 Preprocesamiento

El propósito de la fase de preprocesamiento es manipular la señal de forma que sea más fácil de analizar, eliminando la información que no es relevante, reduciendo el ruido de fondo y normalizando la intensidad de todas las muestras.

En este proyecto se realizó el siguiente procedimiento:

– Segmentación
– Filtro Promedio Móvil
– Filtro Paso Banda
– Filtro Preénfasis
– Normalización

3.2 Extracción de características

Un espectrograma es una forma de representación de una señal de audio. Esta nos permite observar en una misma gráfica las variables de tiempo, intensidad y frecuencia. Esta herramienta es indispensable para el reconocimiento del habla moderno.

Si queremos emplear características basadas en la representación espectrográfica (tiempo-frecuencia) de una muestra de audio, es menester simplificar los datos de entrada a su forma más sencilla, pero que aún contenga aquellas características que la hagan diferenciable. Algunos algoritmos utilizados paraeste propósito son los Coeficientes de Predicción Lineal (LPC), Predicción Lineal Perceptual (PLP), la Transformada Discreta Wavelet (DWT)y el caso de estudio del presente trabajo, los Coeficientes Cepstrales en la frecuencia de Mel (MFCC).

Esta última técnica ha gozado de una popularidad que la ha llevado a ser estándar en la

industria del reconocimiento del habla, de música y de sonidos ambiente por muchos años, gracias a que esta logra simular el proceso de decodificación del sonido que ocurre en el oído humano.

Esto lo logra a partir del principio de que este órgano no percibe las frecuencias de manera lineal, sino logarítmica. Esto es, que una persona percibe con más facilidad un cambio de tonos entre frecuencias bajas al mismo cambio en frecuencias más altas. Se teoriza que esta técnica es especialmente efectiva para la codificación de la información de los formantes de la señal de voz.

3.3 Aumento artificial de los datos

La técnica del aumento artificial de los datos es el proceso de crear muestras sintéticas al aplicar transformaciones sobre el conjunto de datos existente, de manera que podamos entrenar a un modelo de reconocimiento con un cuerpo de datos más diverso y hacerlo más resistente e invariante a las perturbaciones naturales de las señales. Esta técnica es especialmente útil cuando no contamos con las muestras suficientes que pueda requerir un modelo.

Basándonos en los hallazgos de Schlüter y Grill (2015), tenemos a nuestra disposición los efectos de:

– Alterar el tono de una grabación, haciéndolo perceptiblemente más agudo o más grave.

– Estirar la pista en el tiempo, mediante remuestreos y algoritmos de interpolación como el vocoder.

– Añadir ruido blanco aditivo gaussiano, que simule situaciones adversas ambientales.

4 METODOLOGÍA

Para el desarrollo de este proyecto se escogieron cuidadosamente una serie de palabras y variaciones de estas basadas en pruebas de lectura que han sido aplicadas en el pasado, de manera que se tuvieran ejemplos de cada uno de los errores descritos anteriormente. Estas son listadas en la Tabla 2.

Tabla 2. Lista de palabras utilizadas

Palabra	Variaciones
camisa	**t**amisa
faro	fa**rr**o
juego	**ju**go
arbol	**ra**bol, arb**lo**, ar**d**ol
chocolate	**co**colate, choc**h**olate, **cocho**late
trebol	**te**bol, **ter**bol, tre**d**ol
dado	da**o**, **b**a**b**o, **b**ado, da**b**o

Las grabaciones se realizaron en un ambiente controlado, cerrado y silencioso. Todas las palabras y variantes fueron pronunciadas en voz alta y de manera clara y vocalizada. Dado el contexto en el que se desarrolló este proyecto este se limitó a 5 locutores, cada uno aportando en promedio seis grabaciones distintas por cada variante, para un total de 30 grabaciones distintas por variante. Estas fueron muestreadas a una frecuencia de 16 kHz.

Se utilizó en gran medida la librería *librosa* 0.8.0 (McFee et al, 2015) para el lenguaje de programación Python 3.7.10. Esta poderosa librería nos permitió realizar la extracción de características de las muestras de audio basadas en los MFCC.

Para determinar el rendimiento de los clasificadores se hizo uso de la técnica de *Train-Test Split*, donde dividimos nuestro conjunto de muestras en dos subconjuntos que se llamarán *Entrenamiento* y *Prueba*. Como sus nombres lo indican, uno de los subconjuntos será utilizado para ajustar el clasificador a esos datos, y una vez entrenado, con el segundo conjunto analizaremos el comportamiento del modelo al predecir muestras que no ha visto.

Al inicio de la experimentación fueron separadas las grabaciones que hacen parte del conjunto de pruebas. Estas no serán presentadas al modelo al momento de entrenarlo ni serán aumentadas artificialmente, y serán utilizadas al final de todo el procedimiento. Se decidióseparar una proporción del 30% de estas, asegurándose de mantener una cantidad proporcionada en ambos conjuntos tanto de variantes como locutores

Al conjunto de entrenamiento fueron aplicadas las técnicas de aumento artificial de los datos, de manera que nuestro dataset se extendió en un factor de 24, es decir, obtuvimos 23 nuevas muestras artificiales por cada muestra original. (Garcia, A. P., et al, 2016).

Para el entrenamiento de las redes neuronales fue necesario dividir nuevamente el conjunto de entrenamiento en dos subconjuntos de entrenamiento y validación. Esto no fue necesario para el modelo de KNN.

En la Tabla 3 se listan los modelos que obtuvieron los mejores rendimientos experimentalmente en una cantidad de tiempo razonable, así como detalles sobre su arquitectura, y que serán los que serán comparados a continuación.

Tabla 3. Modelos a comparar

Modelo	Cantidad de vecinos
KNN	5 vecinos
Modelo	**Capas ocultas**
MLP	50 50 50
CNN	50C-P 50 50 50

5. RESULTADOS

A continuación se presentan los resultados de los experimentos realizados en los distintos modelos clasificadores.

En la Tabla 4 se muestra el rendimiento del modelo KNN con 5 vecinos al clasificar las muestras ingresadas. En la Tabla 5 se muestra el rendimiento del modelo MLP y en la Tabla 6 se muestra el rendimiento del modelo CNN.

Para estos dos últimos modelos de redes neuronales se ilustra también el rendimiento sobre los conjuntos con los que fueron entrenados así como la métrica adicional del error cuadrático medio entre estos. (MSE). Elrendimiento de estos modelos se resume en la Figura 2.

Tabla 4. Rendimiento del modelo KNN

Palabra	Número de clases	Precisión
camisa	2	87,5%
faro	2	70,4%
juego	2	100,0%
arbol	4	75,9%
chocolate	4	91,8%
trebol	4	67,4%
dado	5	41,0%
Promedio		72,9%

Tabla 5. Rendimiento del modelo MLP

Palabra	Número de clases	Precisión		
		val	test	mse
camisa	2	83,8%	68,8%	0,313
faro	2	84,2%	55,6%	0,444
juego	2	100,0%	100,0%	0,000
arbol	4	84,1%	72,4%	0,586
chocolate	4	98,5%	93,9%	0,122
trebol	4	90,1%	53,1%	0,816
dado	5	78,2%	57,4%	1,180
Promedio		86,0%	67,3%	0,582

Con estos resultados podemos ver que el modelo que tiene un rendimiento a lo largo de todas las palabras es el modelo CNN, seguido del modelo de KNN y por último el modelo de MLP.

Tabla 6. Rendimiento del modelo CNN

Palabra	Número de clases	Precisión		
		val	test	mse
camisa	2	100,0%	93,8%	0,063
faro	2	95,4%	70,4%	0,296
juego	2	100,0%	100,0%	0,000
arbol	4	94,8%	72,4%	0,483
chocolate	4	100,0%	93,9%	0,122
trebol	4	68,6%	55,1%	0,510
dado	5	52,4%	41,0%	1,492
Promedio		85,1%	71,0%	0,564

Si analizamos individualmente los resultados de cada palabra nos podemos dar cuenta de que hay palabras y variaciones que de manera consistente les cuesta más a los modelos diferenciar, como las palabras y variaciones de *dado* y *trebol*. Esto se fundamenta en que hay palabras donde los cambios fonéticos que ocurren son más difíciles de percibir a nivel frecuencial. Esto tiene que ver con la forma en que el aparato fonador genera distintos tipos de sonidos, así como el movimiento natural de este y sus limitaciones hacen que ciertos sonidos se vean influenciados por aquellos inmediatamente anteriores o siguientes. Es por esto que vemos cómo las palabras y variaciones de *juego* y *chocolate* consiguen resultados consistentemente mejores.

El tiempo de ejecución de los algoritmos presentados fue considerablemente distinto entre sí. Mientras que el algoritmo de KNN no tomó más de un par de segundos en completarse, la red MLP tomó tiempos de casi 2 minutos entre generación, entrenamiento y prueba; y la red CNN llegó a tardar más de 20 minutos realizandoeste mismo proceso. Esto nos demuestra la magnitud de complejidad que adiciona cada uno de los modelos, y nos hace conscientes de que estos requerirán de un gran poder computacional.

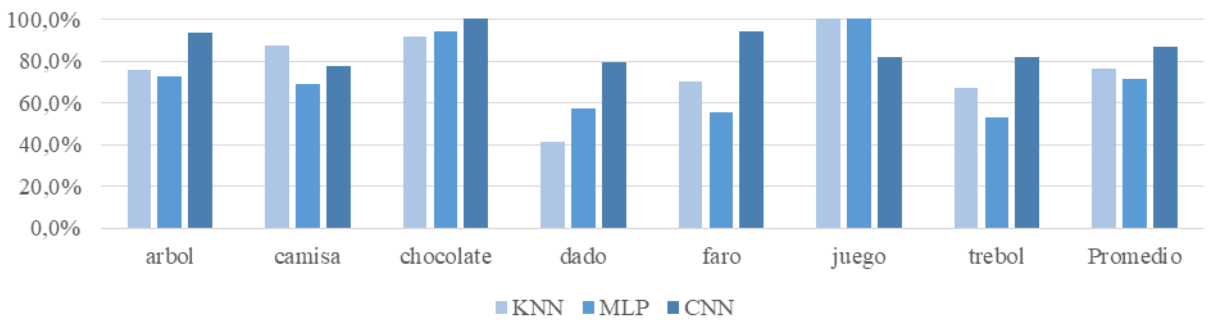

Fig. 2. Rendimiento de los modelos de clasificación sobre las muestras de prueba.

6. RECONOCIMIENTO

Al grupo de trabajo del Semillero *Huellas del Saber*, del Grupo de Investigación *Investigación Pedagógica* del programa de Licenciatura en Educación Infantil de la Universidad de Pamplona, por sus esfuerzos investigativos sobre el fenómeno de la dislexia en la etapa inicial de la lectura.

Al PhD. Valerio Velardo y la comunidad de The Sound of AI, por sus invaluables aportes prácticos y teóricos al público, en materia de blogs, artículos, video tutoriales y repositorios open-source, de los cuales este proyecto se benefició en gran medida.

7. CONCLUSIONES

El proceso presentado para la extracción de características de unas grabaciones de audio y los modelos aquí presentados son excepcionalmente eficaces para el problema de reconocimiento de palabras aisladas, las cuales sean lo suficientemente distintivas fonéticamentelas unas de las otras, llegando a los ordenes de 95a 99 puntos. Estos modelos, sin embargo, se ven debilitados ante la presencia de muestras ambiguas, donde los cambios se reduzcan a fonemas muy puntuales pero se mantengan el acento, la cadencia y demás propiedades de la palabra. Para el reconocimiento de palabras fonéticamente similares se obtuvieron valores experimentales del orden de los 70 puntos para los modelos de KNN y MLP, y del orden de los 78 puntos para los modelos de CNN.

Recomendamos ampliar el trabajo presentado realizando una recolección de grabaciones de audio más extensa, con una mayor cantidad de locutores y variaciones de las palabras, y con equipos de mejor calidad. Pese a los esfuerzos realizados para sacar el mayor provecho de la base de datos presentada, lo cierto es que esta no puede hacer frente a los requerimientos del aprendizaje profundo moderno.

Recomendamos también aprovechar los servicios de computación en la nube ofrecidos al público, como una herramienta para continuar el entrenamiento de redes neuronales con topologías más grandes y complejas.

REFERENCIAS

Abdel-Hamid, O., Mohamed, A. R., Jiang, H., y Penn, G. (2012). *Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition*. 2012 IEEE international conference on Acoustics, speech and signal processing (ICASSP), pp. 4277-4280

Aggarwal, C. C. (2018). *Neural networks and deep learning*. Springer.

Alvarado, R., Caicedo, M. y Gelvez, M. (2019). *Los DBA: "Herramienta Pedagógica para la intervención de las dificultades de aprendizaje en la lectura*". Universidad de Pamplona.

Bastos, V. (1983). *La Dislexia y su Tratamiento*. Universidad de Pamplona.

Camacho C, C. (2016). *Desarrollo de un Sistema de reconocimiento de habla natural basado en redes naturales profundas*. Universidad Autónoma de Madrid.

Dougherty, G. (2013). *Pattern Recognition and Classification*. Springer-Verlag New York.

Fernández, F., Llopis, A. y DeRiego, C. (2006). *La dislexia: origen, diagnóstico y recuperación*. (16ª edición). Madrid: Morata.

Garcia, A. P., Suarez, O., & Castellanos, W. (2016). ERAAE virtual library. Paper presented at the CHILECON 2015 - 2015 IEEE Chilean Conference on Electrical, Electronics Engineering, Information and Communication Technologies, Proceedings of IEEE Chilecon 2015, 911-916. doi:10.1109/Chilecon.2015.7404681

Gelvez, L. y Maldonado, J. (2012) *Aplicación de Redes Neuronales Morfológicas al reconocimiento de vocablos simples*.Revista Colombiana de Tecnologías de Avanzada, **Vol. 19**(1), pp. 13-20.

Hidalgo N, A. y Quilis M, M. (2012). *La voz del lenguaje: fonética y fonología del español*. Tirant Humanidades

Huang, J. T., Li, J., y Gong, Y. (2015). *An analysis of convolutional neural networksfor speech recognition*. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4989-4993.

Katagiri, S. (2003). *Speech Pattern Recognition using Neural Networks*. En Chow, W. y Juang, B. H. (Eds.), Pattern Recognition in Speech and Language Processing CRC Press LCC. pp. 115-147.

Kubanek M, Bobulski J, y Kulawik J. (2019) *A Method of Speech Coding for Speech Recognition Using a Convolutional Neural Network*. Symmetry. **Vol. 11**(9), pp. 1185.

McFee, B., Raffel, C., Liang, D., P.W. Ellis, D., McVicar, M., Battenberg, E. y Nieto, Oriol. (2015). *librosa: Audio and Music Signal Analysis in Python*. Proceedings of the 14th Python in Science Conference (SCIPY 2015).

Quintana, C. (2021). *Detección de dificultades*

en la lectura inicial en niños de primer y segundo grado de Básica Primaria utilizando reconocimiento de patrones en el habla para la detección temprana de la dislexia. Universidad de Pamplona

Rabiner, L. y Juang, B. H. (1996). Fundamentals of Speech Recognition. Prentice-Hall International Inc.

Sen, S., Dutta, A. y Dey, Nilanjam. (2019). Audio Processing and Speech Recognition. Springer.

Suarez, O. J., Díaz, N. H., & Garcia, A. P. (2020). A real-time pattern recognition module via matlab-arduino interface. Paper presented at the Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology, doi:10.18687/LACCEI2020.1.1.646

SITIOS WEB

The International Dyslexia Association (IDA). Dyslexia Basics. https://dyslexiaida.org/ dyslexia-basics/. (10 de noviembre de 2021).

Brownlee, J. (26 de agosto de 2020). Train-Test Split for Evaluating Machine Learning Algorithms. Machine Learning Mastery. https://machinelearningmastery.com/train- test-split-for-evaluating-machine-learning- algorithms/. (10 de noviembre de 2021).

Smith, J. (2011). Spectral Audio Signal Processing. WK3 Publishing. http://ccrma.stanford.edu/~jos/sasp/. (10 de

noviembre de 2021).

ANEXOS

Glosario de siglas y abreviaciones:

CNN (en inglés Convolutional Neural Network) Red Neuronal Convolucional.

KNN (en inglés K-Nearest Neighbor) K-Vecinos más cercanos.

MFCC (en inglés Mel Frequency Cepstral Coefficients) Coeficientes Cepstrales en la frecuencia de Mel.

MLP (en inglés Multi-Layer Perceptron) Perceptrón Multi-Capa.