

Procesamiento de Señales de Voz para la Clasificación de la Enfermedad de Parkinson Utilizando Redes Neuronales Convolucionales

Renata S. Guatelli(1), Verónica I. Aubin(1),
Gabriela M. Barkats Von Willei(2), María del C. Dalmasso(3)
Universidad Nacional de la Matanza. San Justo. Argentina.
(1) Departamento de Ingeniería e Investigaciones Tecnológicas.
(2) Departamento de Ciencias de la Salud.
(3) Escuela de artes y medios de comunicación
{ [vaubin](mailto:vaubin@unlam.edu.ar); [gbarkats](mailto:gbarkats@unlam.edu.ar); [mdalmasso](mailto:mdalmasso@unlam.edu.ar); [rguatelli](mailto:rguatelli@unlam.edu.ar) }@unlam.edu.ar

RESUMEN

La enfermedad de Parkinson es un desorden neurodegenerativo del sistema nervioso, de causa desconocida, y curso crónico, progresivo e irreversible. Si bien aún no se conoce una cura, existen tratamientos que pueden mejorar significativamente la calidad de vida de los pacientes, si su detección es temprana.

La enfermedad generalmente se diagnostica en etapas tardías, estando aproximadamente un 60% de las neuronas degeneradas, de modo que al iniciar la terapia la probabilidad de detener el progreso de la enfermedad es improbable.

Los síntomas más comunes son la aparición de temblores y rigidez muscular. Debido a la rigidez de los músculos se producen alteraciones de la voz las cuales tienen gran potencial para el diagnóstico no invasivo y precoz de la enfermedad. El bajo costo de este diagnóstico en comparación con los estudios clínicos lo haría accesible a un mayor número de personas.

En este trabajo se presentan los avances del proyecto “Procesamiento de Señales de Voz para la Clasificación de la Enfermedad de Parkinson Usando Herramientas Tecnológicas”. Se aplicaron distintos modelos de aprendizaje profundo para la clasificación de enfermos de Parkinson y no enfermos utilizando espectrogramas de las señales de voz. Se logró una precisión del 95.65% en la clasificación.

Palabras clave: *Detección de Parkinson, Análisis de la voz, Espectrograma, Máquina de aprendizaje extremo.*

CONTEXTO

Esta investigación se realiza como parte del Proyecto de investigación PROINCE-C252 “Procesamiento de Señales de Voz Para la Clasificación de la Enfermedad de Parkinson Usando Herramientas Tecnológicas”, se desarrolla conjuntamente entre el Departamento de Salud y el de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza (UNLaM). Propone el análisis de señales de voz como método objetivo y no invasivo para el diagnóstico precoz de la enfermedad de Parkinson (EP). El mismo es continuación de los proyectos C199, C224 y C239 de la UNLaM.

1.INTRODUCCIÓN

La EP es un trastorno degenerativo crónico del sistema nervioso central, asociada con una disminución progresiva de las habilidades motoras y la integración sensoriomotora, de progresión lenta y, por el momento, irreversible. Si bien puede aparecer en adultos jóvenes, su prevalencia es en mayores de 65 años. En [1] se hace referencia a que este trastorno provoca altas tasas de discapacidad, y que en estadios avanzados requiere asistencia permanente. El diagnóstico temprano es esencial para el aumento de la neuro-protección y mejoría en el pronóstico

necesarios para sostener una mejor calidad de vida [2].

En los últimos años se han llevado a cabo una serie de estudios con el objetivo de lograr la detección temprana de la EP considerando diferentes biomarcadores, tales como: imágenes de resonancia magnética, tomografía por emisión de positrones y la tomografía por emisión de fotón único, la voz, la escritura, la marcha, el movimiento al pulsar los botones del teclado de las computadoras, entre otros.

La literatura consultada sostiene un fuerte consenso acerca de que las alteraciones de la voz aparecen con frecuencia en etapas tempranas, antes que las complicaciones motoras propias de la enfermedad [2]. Se estima que entre un 60-80% de pacientes con EP presentan alteraciones de la voz. Medir las distorsiones o cambios en la voz proporciona información útil para diferenciar entre voces patológicas y voces sanas [3], así como para identificar mediante esquemas de clasificación el estadio de la enfermedad. Los sistemas de diagnóstico basados en el análisis de la voz presentan interesantes ventajas pues permiten el diagnóstico no invasivo y precoz de la enfermedad [4].

Con el avance de la tecnología y la disponibilidad de grandes cantidades de datos, la inteligencia artificial ha evolucionado hacia el aprendizaje profundo donde las relaciones entre los datos surgen de forma automática, a través de la cooperación de un conjunto de elementos simples que forman parte de una red. Las redes neuronales convolucionales (CNN) son un componente esencial del aprendizaje profundo.

Trabajos recientes abordan el diagnóstico de la EP mediante CNN [5]. No obstante, las CNN son una solución efectiva si se dispone de grandes conjuntos de datos y en el entrenamiento se requiere estimar millones de parámetros. Por lo cual, es necesario servidores de cómputo equipados con procesadores de múltiples cores, y disponibilidad de aceleradores como las GPU, para disminuir el excesivo tiempo de

entrenamiento requerido por las CNN. Es importante considerar, que en la detección del Parkinson a partir de señales de voz, se cuenta con un número reducido de muestras para analizar, pues las bases de datos disponibles tienen muy pocos pacientes (esto se explica porque el protocolo de adquisición de muestras es complejo, intervienen muchos especialistas y es de alto costo).

Las Máquinas de Aprendizaje Extremo (ELM) corresponden a un paradigma de entrenamiento para Redes Neuronales supervisadas, semisupervisadas y no supervisadas [6]. La literatura presenta diferentes casos de otras problemáticas, donde las ELM tienen un nivel de precisión similar a las CNN y las Máquinas de Soporte Vectorial, pero con un tiempo de entrenamiento considerablemente más bajo [7].

Un espectrograma es una representación visual de la señal de sonido la cual considera el tiempo, frecuencia y amplitud. La representación del habla a través de espectrogramas ha demostrado ser estable y robusta, incluso en presencia de elevados niveles de ruido.

En [8] se demostró que las vocales sostenidas tienen información discriminativa suficiente para la clasificación de la EP mediante el uso de modelos de aprendizaje automático

La línea de investigación de este trabajo es discernir entre pacientes con EP y personas sin EP a partir del uso de técnicas de aprendizaje profundo aplicadas a los espectrogramas de señales de voz obtenidas a partir de grabaciones de la vocal /a/ realizadas en un ambiente controlado. Asimismo, se comparan las ventajas del uso de ELM sobre las CNN para la clasificación. El avance en el análisis de la voz de personas con y sin EP apunta a brindar herramientas útiles, que servirán de soporte al médico para realizar el diagnóstico y seguimiento de la enfermedad.

2.LINEA DE INVESTIGACION Y DESARROLLO

El estudio de la voz aplicado a la clasificación entre enfermos y no enfermos de Parkinson, se aborda en el proyecto de investigación C252 antes mencionado, desde dos enfoques. Un enfoque busca estudiar distintos algoritmos para extraer características de la voz, y luego utiliza dichas características como entrada de clasificadores. El otro enfoque emplea el aprendizaje automático sin buscar descriptores, y trabajar en el dominio original de la información (señales o espectrogramas de la voz). Los avances desarrollados sobre esta última línea son los que se presentan en este artículo.

El esquema general del proceso de trabajo planteado en esta investigación se presenta en la Figura 1 donde se observa que, partiendo de los sonidos de la base de datos de audios de voz, transformados en espectrogramas, se generó una nueva base de datos, sobre la cual se aplicaron técnicas de Deep Learning para clasificar entre enfermos y no enfermos de Parkinson. Esta información servirá de soporte al médico para realizar el diagnóstico y seguimiento de la enfermedad.

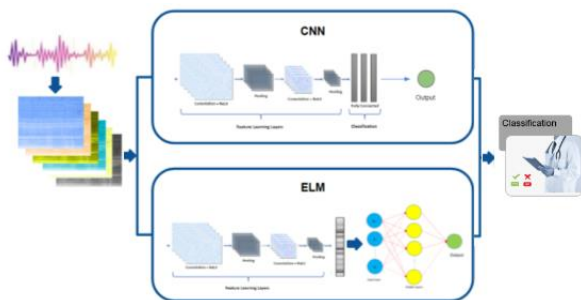


Figura 4 Proceso general de trabajo planteado en esta investigación

Para los experimentos realizados se utilizó la base de datos con voces de personas con y sin EP presentada en [9]. La cual está disponible en el repositorio de la UNLaM (<https://repositoriocyt.unlam.edu.ar/>) con acceso público. Este repositorio contiene 55 sujetos de enfermos de Parkinson (24 mujeres y 31 varones) evaluados neurológicamente con la Escala Unificada de Calificación de la Enfermedad de Parkinson (UPDRS) y 64 sujetos sin Parkinson. En el momento de

realizar la grabación los enfermos de Parkinson se encontraban medicados, es decir en "on", en estado que los síntomas son de baja intensidad. La edad de los pacientes con Parkinson varía entre 38 y 79 años, con una duración media de la enfermedad de 6 años.

Base de datos de espectrogramas

Para generar los espectrogramas, se aplicó a las señales de la base de datos de audios la Short-Time Fourier Transform (STFT), utilizando la escala de grises para representar la amplitud. La base de datos original contiene diferentes fonaciones, en particular en este paper se utilizaron las grabaciones de la vocal /a/ sostenida, generando finalmente 135 espectrogramas en escala de grises, de los cuales 58 corresponden a enfermos y 77 a personas sanas.

Aumentación por color

Para obtener un buen entrenamiento de las redes neuronales convolucionales es necesario contar con un gran volumen de datos. Como estrategia de aumentación de datos, además de la paleta gris, se crearon espectrogramas con 13 paletas de color. Cabe mencionar que algunas paletas de colores son útiles para resaltar ciertos detalles del sonido, las paletas "copper" y "bone" de MatLab resaltan las formas de crestas y valles, mientras que "jet" o "hsv" da una indicación de la inclinación de las pendientes. Considerando la estrategia de aumentación de datos, se generaron 1755 espectrogramas, 754 de personas enfermas y 1001 de personas sanas. Se realizaron experimentos que confirmaron la validez de esta estrategia de aumentación de datos.

Aumentación por segmentación del audio

Esta segunda metodología para aumentar la cantidad de muestras consistió en generar fragmentos de los sonidos originales. Los 135 sonidos originales tenían una duración variable entre 3 a 5 segundos. Se realizaron cortes de 1 segundo con 50% de solapamiento de los sonidos originales. Con esta estrategia

de aumentación de datos se obtuvieron 1168 muestras de sonido de 1 segundo cada una. Para cada uno de estos fragmentos se generaron los espectrogramas en escala de grises y en las 13 paletas de color seleccionadas. De esta forma, se obtuvieron 1168 espectrogramas en escala de grises y 15.184 espectrogramas en color.

Experimentos

Con el objetivo de validar la aumentación por segmentación de audio, para analizar la clasificación de los enfermos de Parkinson se utilizaron 5 arquitecturas de CNN pre-entrenadas: AlexNet, VGG-16, SqueezeNet, Inception V3 y ResNet-50. Para comparar las tasas de acierto y tiempo entre CNN y ELM se realizaron 4 tipos de experimentos:

Experimento 1 (EXP1): Con espectrogramas en escala de grises de sonidos originales.

Experimento 2 (EXP2): Con espectrogramas en color de sonidos originales.

Experimento 3 (EXP3): Con espectrogramas en color de fragmentos de sonido.

Experimento 4 (EXP4): Con espectrogramas en color de sonidos originales y fragmentos.

3.RESULTADOS

En esta investigación se estudiaron distintos métodos de aprendizaje profundo que utilizan espectrogramas de señales de voz para la clasificación de enfermos y no enfermos de Parkinson. Los sistemas de diagnóstico basados en el análisis de la voz presentan ventajas pues permiten el diagnóstico no invasivo y precoz de la enfermedad.

Se probaron diferentes arquitecturas de CNN, aplicadas al repositorio ampliado con la primera estrategia de aumentación de datos. Se aplicó la técnica de transferencia de aprendizaje debido a la poca cantidad de muestras disponibles. Para todas las arquitecturas, se observó una mejora en los indicadores de performance para el conjunto de datos aumentado pasando de 91.30% al 95.65%. Los niveles de performance alcanzados muestran que la estrategia de aumentación de datos a través de la utilización de diferentes paletas de colores, es

pertinente y que las CNN resuelven el problema con niveles de precisión aceptables. Se realizó una comparación entre la metodología de CNNs aplicadas a los espectrogramas obtenidos de las señales de voz, con el análisis acústico realizado por en la otra línea de investigación del mismo proyecto. De esta forma, se pudo comprobar que los resultados obtenidos fueron similares, con una leve mejora para la técnica de las CNNs, que además presenta la ventaja de no requerir un pre - procesamiento de las señales de audio para encontrar los mejores descriptores de las muestras analizadas.

Para comprobar la reducción de tiempo de las ELM sobre las CNN manteniendo niveles aceptables de accuracy, se realizaron experimentos midiendo los tiempos de entrenamiento, validación y test, así como también se registraron las medidas de rendimiento de cada modelo: accuracy, eficiencia y especificidad para los 4 experimentos, antes mencionados, en las 5 arquitecturas consideradas. El tiempo de entrenamiento de las CNN pasó de 38.4 seg (EXP1). a 5.857,4 seg.(EXP4) al aumentar la cantidad de muestras de 135 a 16.939. Mientras que para las ELM pasó de 8 seg. a 28.8 seg. respectivamente.

De los experimentos, se observa, como se mencionó, que el aumento de muestras, en general, permite mejorar la calidad del entrenamiento, obteniéndose mejores valores de performance. Además se pudo comprobar que el clasificador basado en ELM alcanzó un nivel de precisión similar al de las CNN pero con un tiempo de entrenamiento sustancialmente más reducido.

A continuación se detallan los trabajos publicados a partir de los resultados obtenidos en esta investigación:

Mayo 2020 - "Predicción de la enfermedad de Parkinson utilizando redes neuronales convolucionales". - Guatelli, R., Aubin, V. I., Pérez, S. N.. In XXII Workshop de Investigadores en Ciencias de la Computación.

Octubre 2021 - "Detección de Parkinson mediante espectrogramas en color y redes neuronales convolucionales". - Guatelli, R., Aubin, V. I., Mora, M., Naranjo-Torres, J., Sinopoli, A. In II Simposio Argentino de Imágenes y Visión (SAIV 2021)-JAIIO

Enero 2022 - "Classification of Parkinson's disease patients based on spectrogram using local binary pattern descriptors". - E Gelvez - Almeida, A Vásquez-Coronel, R Guatelli, V Aubin and M Mora. In Journal of Physics: Conference Series. IOP Publishing.

Noviembre 2022 - "Análisis comparativo entre CNN y Modelos Logísticos para detección de la Enfermedad de Parkinson utilizando la voz" - Renata S. Guatelli, Monica Giuliano, Verónica Aubin, Luis Fernández, María Laura Pepe, Silvia N. Perez - In 10o Congreso Nacional de Ingeniería Informática / Sistemas de Información

Octubre 2023 - "Detection of Parkinson's disease based on spectrograms of voice recordings and Extreme Learning Machine random weight neural networks". Autores: R Guatelli, V Aubin, M Mora, J Naranjo, A Mora-Olivar - In Journal Engineering Applications of Artificial Intelligence, 125, 106700.

En esta línea de investigación en progreso se continúa trabajando para aumentar el número de voces de la base de datos. Hacer un análisis perceptual de las muestras. Estudiar la clasificación de los estadios de la enfermedad. Realizar estudios preliminares para validar el uso de teléfonos celulares en la adquisición de los sonidos de pacientes con Parkinson. Continuar con el estudio y análisis de medidas acústicas derivadas de la señal de voz y algoritmos de deep learnig para diferenciar la voz patológica de la voz sana. Se propone además realizar vínculos con grupos de investigación de Argentina o el exterior, que se encuentren trabajando en temas afines a este proyecto. Difundir los resultados de la investigación.

4.FORMACIÓN DE RECURSOS HUMANOS

El equipo de investigación está formado por docentes investigadores de la Universidad Nacional de la Matanza pertenecientes al Departamento de Ingeniería e Investigaciones Tecnológicas y el Departamento de Salud. El equipo de trabajo está conformado por un director, un codirector, cuatro docentes investigadores, una alumna de grado, dos becarias y un investigador externo de la Universidad Católica del Maule, Chile.

En base a esta investigación se realizó la Tesis de Maestría en Informática "Detección de la Enfermedad de Parkinson basada en Máquinas de Aprendizaje Extremo a cargo de Lic. Renata Silvia Guatelli. Directora Dra. Mg. Verónica Inés Aubin, cuya defensa fue llevada a cabo en septiembre de 2023.

5.BIBLIOGRAFÍA

- [1] Jara, G., Fariás, P. (2023). Percepción de las dificultades de voz por parte de usuarios con enfermedad de Parkinson. Revista de Investigación en Logopedia. 13(1) doi: 10.5209/rlog.81566
- [2] Becker, G., Müller, A., Braune, S., Büttner, T., Benecke, R., Greulich, W., ... & Thümler, R. (2002). Early diagnosis of Parkinson's disease. Journal of neurology, 249, iii40-iii48.
- [3] Tsanas, A., Little, M. A., McSharry, P. E., Spielman, J., & Ramig, L. O. (2012). Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE transactions on biomedical engineering, 59(5), 1264-1271.
- [4] Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gurgun, F., Delil, S. & Kursun, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828-834.
- [5] Hireš, M., Gazda, M., Drotar, P., Pah, N. D., Motin, M. A., & Kumar, D. K. (2022). Convolutional neural network ensemble for Parkinson's disease detection from voice recordings. Computers in biology and medicine, 141, 105021.
- [6] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. Neurocomputing, 70(1-3), 489-501.
- [7] Chorowski, J., Wang, J., & Zurada, J. M. (2014). Review and performance comparison of SVM-and ELM-based classifiers. Neurocomputing, 128, 507-516.
- [8] Sakar B E, Isenkul M E, Sakar C O, Sertbas A, Gurgun F, Delil S, Apaydin H, Kursun O 2013 Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings IEEE Journal of Biomedical and Health Informatics 17(4) 828

[9] Giuliano, M., Perez, S. N., Maldonado, M., Bondar, P., Linari, D., Adamec, D. A., & Gurlekian, J. A. (2021). Construction of a parkinson's voice database. In International Conference on Industrial Engineering and Operations Management (p. 940). Sao Paulo: IEOM Society International.