Detección de Parkinson mediante Espectrogramas en Color y Redes Neuronales Convolucionales

Renata Guatelli[1], Veronica Aubin[1], Marco Mora[2], José Naranjo-Torres[2], and Alesio Sinopoli[1]

[1] Universidad Nacional de La Matanza, San Justo 1754, Argentina
{rguatelli,vaubin}@unlam.edu.ar
[2] Laboratorio de Investigaciones Tecnológicas en Reconocimiento de Patrones, Universidad Católica del Maule, Chile
{mmora,jnaranjo}@ucm.cl

Abstract. En este trabajo se propone una estrategia de aumentación de datos para los repositorios de espectrogramas utilizados en la detección de la Enfermedad de Parkinson. Esta estrategia consiste en crear espectrogramas a partir de una señal de voz considerando distintas paletas de colores. Se utilizan 13 paletas de colores provistas por la herramienta colormap de Matlab. Para la evaluación de los resultados se consideran los modelos de CNN AlexNet, VGG 16, ResNet 50, Inception v3 y Squeezenet. De los experimentos se observa que las CNN mejoran la performance de clasificación y disminuyen la variabilidad de los resultados cuando utiliza el conjunto de datos aumentado.

Keywords: Deteccion Parkinson · Espectrogramas en Color · Aumentación de Datos · Redes Neuronales Convolucionales.

1 Introducción

La Enfermedad de Parkinson (EP) es un trastorno neurodegenerativo del sistema nervioso central de progresión lenta e irreversible. Los síntomas principales de la enfermedad son el temblor, la rigidez muscular, la inestabilidad postural y la lentitud del movimiento. Los síntomas de la EP se ven reflejados en forma temprana en el habla y en la voz.

Diferentes trabajos han propuesto el procesamiento de señales de voz para obtener parámetros acústicos como método objetivo y no invasivo para la detección de la EP [1]. En particular, un enfoque reciente es estudiar la representación visual del espectro de frecuencias de señales de voz (espectrogramas) mediante Redes Neuronales Convolucionales (CNN).

En [2] se presenta una metodología para clasificar EP a partir de muestras de audio en tres idiomas diferentes: español, alemán y checo. Compara dos enfoques, uno donde se extraen características a partir de las señales de voz, que se clasifican utilizando un SVM, y otro que utiliza espectrogramas para entrenar

una CNN en un idioma, para hacer transferencia de aprendizaje a cada uno de los idiomas restantes. Los mejores resultados los obtuvieron al entrenar con el idioma español, y luego hacer transferencia al alemán obteniendo una presición del 77.3% y al checo 76.7%. En [3] se utilizan para la detección de la EP una arquitectura ResNet previamente entrenada usando las bases de datos ImageNet y SVD. Para la clasificación utiliza los espectrogramas de audios de las vocales con fonación sostenida de la base PC-GITA [4]. La precisión obtenida en el conjunto de validación es superior al 90%. En [5] se combinan metodologías de aprendizaje profundo con metodologías de aprendizaje automático (SVM, random forest, perceptrón multicapa) para clasificar a los pacientes de Parkinson. Presenta la comparación de tres métodos para clasificar EP utilizando la base de datos PC-GITA [4]. El primer método se basa en aprendizaje por transferencia aplicado a espectrogramas de grabaciones de voz. El segundo método evalúa características profundas extraídas de espectrogramas de voz usando clasificadores de aprendizaje automático. El tercer método evalúa características acústicas simples de grabaciones usando clasificadores de aprendizaje automático. El enfoque basado en el segundo método dio la precisión más alta alcanzando un 99,7% aplicando un perceptrón multicapa a grabaciones de monólogo. En [6] se propone la detección de patologías de voz de personas con enfermedades neurodegenerativas como Parkinson y Alzheimer. Las muestras se obtuvieron de las bases de datos SVD Saarbrucken [7] y PC-GITA [4]. Se utiliza una arquitectura CNN obteniendo una precisión en la clasificación de más 95%.

La construcción de bases de datos para detectar la EP es un proceso complejo debido a que, el paciente debe ser evaluado clínicamente por diversos especialistas (neurólogo, fonoaudiólogo, entre otros), y seguir protocolos de grabación del sonido estrictos para los sujetos sanos y enfermos de la muestra. Esto implica que obtener los datos de un sujeto requiere mucho tiempo, que existen pocas bases de datos públicas, y que las bases de datos tienen pocas muestras. Lo anterior pone dificultades para abordar el problema con CNN.

En este trabajo se propone una estrategia de aumentación de datos para los repositorios de espectrogramas utilizados en la detección de la EP. Esta estrategia consiste en generar imágenes del mismo espectrograma con distintas paletas de colores. Esta estrategia permite aumentar drásticamente el número de muestras para el conjunto de entrenamiento de las redes, mejorar la precision de clasificación y disminuir la variabilidad de los resultados.

La estructura de este trabajo es la siguiente: La sección 2 presenta el proceso de aumentación de datos de espectrogramas utilizando paletas de colores, la sección 3 muestra resultados de detección de la EP con diversos modelos de redes CNN y finalmente, la sección 4 presenta las conclusiones de esta investigación.

2 Estrategia de Aumentación de Datos

Un espectrograma permite representar a lo largo del tiempo las variaciones de frecuencia y amplitud de una señal de sonido. Es una representación en tres dimensiones: tiempo, frecuencia y amplitud. Comúnmente el espectrograma se

representa a través de un gráfico en dos dimensiones: tiempo (eje horizontal) y frecuencia (eje vertical), donde la tercera dimensión (amplitud) es representada mediante el uso de una escala de colores.

En este trabajo se consideraron los sonidos de la base de datos presentada en [8]. 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. La edad de los pacientes con Parkinson varía entre 38 y 79 años con una duración media de la enfermedad de 6 años. Para generar los espectrogramas, se aplicó a las señales de la base de datos la Short-Time Fourier Transform (STFT), utilizando la escala de grises para representar la amplitud. La base de datos original contiene diferentes vocalizaciones, en particular se utilizaron las grabaciones de la vocal A sostenida, generándose finalmente 135 espectrogramas en escala de grises, de los cuales 58 corresponden a enfermos y 77 a personas sanas.

Como estrategia de aumentación de datos, ademas de la paleta gris, se crearon espectrogramas con las paletas de color disponibles en la herramienta colormap de Matlab. Se consideraron 13 paletas de colores: “autumn”, “bone”, “cool”, “copper”, “gray”, “hot”,“hsv”, “jet”, “parula”, “pink”, “spring”, “summer” y “winter”. Es interesante mencionar que algunas paletas de colores son útiles para resaltar ciertos detalles del sonido, las paletas “copper” y “bone” resaltan las formas de crestas y valles, mientras que “jet” o “hsv” da una indicación de la inclinación de las pendientes. Se excluyeron las paletas “colorcube”, “flag”, “lines”, “prism” y “white”, pues generan imagenes pixeladas y ruidosas. Considerando la estrategia de aumentacion de datos, se generaron 1755 espectrogramas, 754 de personas enfermas y 1001 de personas sanas.

3 Resultados

Para mostrar los beneficios de la estrategia de aumentación de datos propuesta, se realizaron experimentos con diversos modelos de CNN que son representativos de las arquitecturas existentes: AlexNet [9], VGG 16 [10], ResNet 50 [11], Inception v3 [12] y Squeezenet [13]. En estos modelos se trabaja en modo de transferencia de aprendizaje. Para todas las redes anteriores, se realizaron experimentos con dos conjuntos de datos, el primero con espectrogramas en escala de grises y el segundo con el conjunto de datos aumentado considerando los espectrogramas en color. Para obtener una medida de performance objetiva se consideró un esquema de validación cruzada de tres conjuntos: 70% entrenamiento 10% validación y 20% test. Se realizaron 10 repeticiones del esquema de validación cruzada para informar el promedio de las repeticiones.

Para obtener los resultados se utilizó Matlab R2018b y Deep Learning Toolbox. Se realizaron experimentos con diversos hiperparámetros para las redes, de los cuales fueron seleccionados los siguientes:

- Conjunto Original: épocas 25, mini-bacht 32 y razón de aprendizaje 0.0001.
- Conjunto Aumentado: épocas 35, mini-bacht 32 y razón de aprendizaje 0.0001.

La tabla 1 presenta los resultados obtenidos para cada red. Se informa el máximo, el mínimo, la distancia entre el máximo y el mínimo (Rango), y el promedio de las repeticiones de la clasificación del conjunto de test. Las columnas Orig corresponden a los resultados con los espectrogramas en escala de grises originales, y las columnas Aum corresponden a los resultados del conjunto de datos aumentado. De la tabla se observa que para todos los modelos de redes, la estrategia

Table 1. % Acierto Clasificación Conjunto de Test con diversos modelos de CNN

	Máximo % acierto		**Mínimo % acierto**		**Rango % acierto**		**Promedio % acierto**	
	Orig	**Aum**	**Orig**	**Aum**	**Orig**	**Aum**	**Orig**	**Aum**
AlexNet	88.46%	91.74%	65.38%	81.20%	23.08%	10.54%	77.69%	87.64%
VGG16	96.15%	98.01%	61.54%	92.88%	34.62%	5.13%	76.92%	95.98%
Inception V3	65.38%	86.89%	38.46%	78.92%	26.92%	7.98%	54.62%	83.13%
ResNet50	88.46%	90.03%	61.54%	86.89%	26.92%	3.13%	75.38%	88.09%
SqueezeNet	61.54%	84.05%	46.15%	73.79%	15.38%	10.26%	58.08%	80.09%

de aumentación de datos permitió mejorar tanto la razón de acierto como la dispersión de los resultados. La red VGG16 obtuvo los mejores indicadores, con un promedio en la tasa de acierto de 95.98%, un máximo de 98.01% y un mínimo de 92.88%.

4 Conclusiones

Este trabajo ha propuesto una estrategia de aumentación de datos para los repositorios de espectrogramas utilizados en la detección de la Enfermedad de Parkinson. La estrategia consiste en crear espectrogramas a partir de una señal de voz considerando distintas paletas de colores. Se probaron diferentes arquitecturas de CNN, y para todas ellas se observa una mejora en los indicadores de performance para el conjunto de datos aumentado. Los elevados niveles de performance alcanzados muestran que la estrategia de aumentación de datos es pertinente y que las CNN resuelven el problema con elevados niveles de precisión.

5 Agradecimientos

Los autores agradecen al proyecto de investigación C224, DIIT- UNLaM, Hospital Posadas y Hospital Rivadavia por permitirnos utilizar la base de datos generada. Además, se agradece al Laboratorio de Investigaciones Tecnológicas en Reconocimiento de Patrones LITRP (www.litrp.cl) de la Universidad Católica del Maule, Chile, por proporcionar los servidores de cómputo de alto rendimiento donde se realizaron los experimentos.

References

1. Siddharth Arora, Ladan Baghai-Ravary, and Athanasios Tsanas. Developing a large scale population screening tool for the assessment of parkinson's disease using telephone-quality voice. *The Journal of the Acoustical Society of America*, 145(5):2871–2884, 2019.
2. Juan Camilo Vásquez-Correa, Tomas Arias-Vergara, Cristian D Rios-Urrego, Maria Schuster, Jan Rusz, Juan Rafael Orozco-Arroyave, and Elmar Nöth. Convolutional neural networks and a transfer learning strategy to classify parkinson's disease from speech in three different languages. In *Iberoamerican Congress on Pattern Recognition*, pages 697–706. Springer, 2019.
3. Marek Wodzinski, Andrzej Skalski, Daria Hemmerling, Juan Rafael Orozco-Arroyave, and Elmar Nöth. Deep learning approach to parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 717–720. IEEE, 2019.
4. Juan Rafael Orozco-Arroyave, Julián David Arias-Londoño, Jesús Francisco Vargas-Bonilla, Maria Claudia Gonzalez-Rativa, and Elmar Nöth. New spanish speech corpus database for the analysis of people suffering from parkinson's disease. In *LREC*, pages 342–347, 2014.
5. Laiba Zahid, Muazzam Maqsood, Mehr Yahya Durrani, Maheen Bakhtyar, Junaid Baber, Habibullah Jamal, Irfan Mehmood, and Oh-Young Song. A spectrogram-based deep feature assisted computer-aided diagnostic system for parkinson's disease. *IEEE Access*, 8:35482–35495, 2020.
6. Nam Trinh and OBrien Darragh. Pathological speech classification using a convolutional neural network. In *in Proc. IMVIP*, pages 72–75, 2019.
7. William John Barry and Manfred Pützer. Saarbruecken voice database. institute of phonetics, university of saarland. urlhttp://www.stimmdatenbank.coli.uni-saarland.de/ash, 2007.
8. Monica Giuliano, Silvia Noemí Perez, Maldonado Maldonado, Pablo Bondar, Daniela Linari, Dario Adamec Adamec, María Inés Debas, Carlos Morales Morales, Leticia de León, Aldo Yaco Yaco, Joice Birelli Birelli, Macarena Martínez Ribaya, María Lis Lacaze, and Jorge A. Gurlekian. Construction of a parkinson´s voice database. In *International Conference on Industrial Engineering and Operations Management*. IEOM Society International, 2021.
9. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
10. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
11. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
12. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
13. Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.