

Utilización de grabaciones con teléfonos móviles para la predicción de la enfermedad de Parkinson

Silvia N. Pérez Mónica Giuliano
Universidad Nacional de La Matanza
sperez@unlam.edu.ar mgiuliano@unlam.edu.ar

Resumen

Este estudio investiga el procesamiento de señales de voz para detectar la enfermedad de Parkinson (EP). Se analizaron los datos demográficos y registros de fonación vocal /a/ de la base de datos mPower para identificar a los pacientes con EP. El procesamiento automático de las fonaciones se realizó utilizando el software libre OpenSmile obteniéndose 62 características o parámetros. Utilizando R, también software libre, se logró el ajuste de un modelo parsimonioso que redujo de 62 a 5 características significativas de fonación, las que se consideraron además del género y la edad. Se utilizaron redes neuronales de percepción multicapa (MLP) y modelos de regresión logística (LR) para obtener un modelo con alta capacidad de predicción. Para medir esta, se utilizó como métrica el área bajo la curva ROC, la que resultó superior a 0,82 en todos los modelos ajustados. Este trabajo contribuye a la monitorización de pacientes con EP a partir de la grabación de algunas características de fonación recogidas mediante un teléfono móvil.

Introducción

Se considera que la enfermedad de Parkinson (EP) pertenece a un grupo de trastornos que tienen su origen en la degeneración de neuronas, por lo que se la describe como una enfermedad neurodegenerativa [1]. La ausencia de una prueba diagnóstica de referencia y la falta de pericia clínica son características que pueden hacer subestimar los casos incipientes en una población determinada y provocar, muchas veces, un retraso considerable en su diagnóstico. Si bien el diagnóstico definitivo es histopatológico, se considera que un diagnóstico preliminar en etapas tempranas resulta ser un desafío y una enorme responsabilidad, considerando las implicancias de pronóstico vital y las limitaciones motrices futuras para cada paciente diagnosticado. En el caso de la EP y otras enfermedades neurodegenerativas, diversos estudios [1,2] han identificado anomalías en grabaciones de voz de los pacientes, poniendo de manifiesto la posibilidad de usar técnicas de análisis de señales acústicas para el diagnóstico precoz de la

enfermedad. Se considera que el diagnóstico a través del análisis de trastornos del habla utilizando sistemas de tele-diagnóstico y tele-monitoreo, no solo evitaría las complicaciones de los métodos invasivos, sino que, además, es de bajo costo y fácil administración. Además del hecho de que la evaluación clínica tradicional para monitorear la progresión es costosa, los pacientes con EP pueden tener problemas para asistir a las visitas médicas de control. Por lo tanto, se necesita una herramienta clínica económica y objetiva para seguir la progresión de síntomas de la EP. Los crecientes avances de la tecnología de la información y las telecomunicaciones ofrecen una buena oportunidad para la supervisión a distancia.

Los trastornos de la voz se pueden medir con herramientas acústicas y hay una gran cantidad de investigaciones que apuntan a la evaluación automática de patrones de voz patológicos [3, 4] y su relación con la EP [2, 5, 6, 7]. En este contexto, se encuentran diversos trabajos que recopilan grabaciones de voz para extraer diferentes características o parámetros. Para determinar su influencia en la EP, se midieron las siguientes señales del habla: Frecuencia fundamental o tono [2, 8, 5, 9]; Brillo, fluctuación, HNR [2, 5, 8, 9]; Coeficientes Cepstrales de Frecuencia de Mel (MFCC) [10, 7] y medidas no lineales de disfonía, tales como: Entropía del período de lanzamiento (PPE), Análisis de Fluctuación Detendida (DFA), Entropía de Densidad del Período de Recurrencia (RPDE) [2, 9, 11]. Algunas características están altamente correlacionadas entre sí, lo que sugiere el uso de métodos de selección de características [9, 11,12].

Como se expuso anteriormente, el análisis de datos de pacientes con EP conlleva el trabajo de aplicación de distintos tipos de software: en principio, uno que permita extraer características de la voz a partir de una grabación, y en segundo término se requiere software estadístico que permita aplicar métodos estadísticos avanzados. En este trabajo se utilizó OpenSmile para las primeras tareas, y R para la segunda parte. En las siguientes subsecciones se describe el contexto de análisis acústico en que se inscribe el trabajo, y el software que permite extraer características o parámetros a partir de la onda acústica.

En este trabajo se continúa el estudio realizado en [13], analizando características asociadas a dicho análisis. El objetivo es presentar y describir el modelo parsimonioso de selección de variables que permitan diferenciar fonaciones de la vocal /a/ realizadas por EP y no EP.

Análisis acústico

Las medidas físicas de la voz humana se basan en el empleo de diversos parámetros acústicos que reflejen las tres dimensiones perceptibles del sonido: amplitud, tono y estructura temporal. La amplitud (cuyo principal parámetro es la intensidad) es una medida de la presión sonora al transmitirse la voz en el medio aéreo, expresada en decibelios (dB), dependiente de la amplitud de la vibración de las cuerdas vocales y de la presión subglótica [14], el tono se expresa mediante la F0 de la señal vocal, medida en hertzios (Hz); por último, los parámetros derivados del tiempo tienen relación con la tasa y rapidez de la vocalización. La representación de las variables físicas es altamente compleja a causa de la variabilidad de la energía espectral de la voz.

La voz es evaluada secuencialmente mediante la selección de los parámetros y los más utilizados se detallan a continuación.

- La frecuencia fundamental, F0 o también llamada glótica es la frecuencia de cierre y apertura de la glotis debida a la presión del flujo de aire traqueal, caracteriza la emisión de sonidos sonoros (fonación).
- El espectro del pulso glótico generado (espectro de la fuente glotal) será afectado por los filtros del tracto vocal, para producir sonidos vocálicos. Es decir: el tracto vocal es un filtro o conjunto de filtros mecánicos que enfatiza algunas de las componentes del espectro glotal, especialmente las componentes que coinciden con sus propias frecuencias de resonancia. Esas componentes o zonas de componentes de mayor energía en el espectro resultante, son los formantes. Por lo tanto, la modulación del espectro de la fuente glotal y las resonancias del tracto vocal afectarán la ubicación de los formantes. Los espectros de los sonidos vocálicos están bien diferenciados entre sí y caracterizan cada vocal.
- Los anchos de banda están definidos [15] como una función de la pérdida de energía debida al calor de conducción, a la viscosidad, al movimiento de las paredes de las cavidades resonantes, a la radiación del sonido desde los labios y a la parte real de la impedancia glótica. Acústicamente, las vocales son los sonidos más simples: están representadas por patrones estacionarios y la configuración articulatoria es casi estable.
- El *Jitter* es una medida que representa la variabilidad de frecuencia fundamental ciclo a ciclo. La escala de medición es logarítmica con un intervalo de normalidad que va desde el 0,1 al 1% y un intervalo de alteración que va desde el 1 al 10%.
- El *Shimmer* indica la variabilidad en decibelios de la amplitud pico a pico, de 0 a 0,3 dB se

considera que pertenece al intervalo de normalidad mientras que si está entre 0,3 y 1,3 dB pertenece al de alteración.

- La relación armónica/ruido (HRN) es el cociente entre la energía espectral causada por las componentes armónicas de la señal y la correspondiente al ruido. El rango es de 4 a 9 para la zona de normalidad y de 4 a -1 en la zona de alteración.
- La amplitud del *Cepstrum* evalúa la correlación de ciclos, tiene una escala lineal con una zona de normalidad que va desde 0,2 a 0,6 y una zona de alteración desde 0,2 a -0,2.

La frecuencia fundamental es la principal unidad de análisis acústico. Sus valores varían a lo largo de la vida y según los sexos. Niños y niñas tienen una frecuencia media de 240 Hz hasta la pubertad, en la que los varones tienen un descenso hasta los 110 Hz (la voz se torna más grave), mientras que las mujeres se mantienen en 210 Hz. Hacia la tercera edad aumenta la frecuencia de los hombres a 140 Hz y disminuye en las mujeres a 190 Hz en promedio. Estos valores se modifican en la voz senil, reduciéndose la F0 en las mujeres a 175 Hz aproximadamente, mientras que se incrementa en el hombre a 130 Hz a los 70 años y a 160 Hz a los 90 años [16].

Los audios se analizan utilizando OpenSMILE. Cabe aclarar que varios estudios muestran diferencias en las mediciones de parámetros según el software utilizado [17], por lo que los valores utilizados en la parametrización son dependientes de este.

Software utilizado para la extracción de características

OpenSMILE es una herramienta de extracción de características y reconocimiento de patrones ampliamente utilizados que se aplica para una gran variedad de casos de uso. SMILE es un acrónimo para la interpretación de voz y música por extracción de grandes espacios. La herramienta OpenSMILE permite extraer grandes espacios de funciones de audio en tiempo real. Combina funciones de recuperación de información musical y procesamiento de voz. Está escrito en C++, los componentes del extractor de características pueden interconectarse libremente para crear características nuevas y personalizadas, todo a través de un simple archivo de configuración. Se pueden agregar nuevos componentes a OpenSMILE a través de una sencilla interfaz de plug-in binario y una API completa [18].

La técnica de extracción de parámetros que se emplea está vinculada principalmente a la obtención de los parámetros MFCC (Mel-Frequency Cepstral Coefficients). Se obtienen 62 parámetros que se pueden agrupar según la siguiente lista:

- 10 informan sobre F0

- 10 relacionadas al volumen
- 6 medidas y desviaciones típicas del *Jitter*, el *Shimmer* y la tasa de HRN
- 18 relacionadas con las frecuencias F1, F2, F3
- 18 consideradas otras características.

La obtención de parámetros en el sistema MFCC es la más utilizada para el análisis de la voz.

Metodología

El estudio actual utilizó un conjunto de datos de acceso abierto. Estos son contribución de usuarios de la aplicación móvil Parkinson mPower como parte del estudio mPower desarrollado por Sage Bionetworks y descrito en Synapse (<https://www.synapse.org>).

En las siguientes subsecciones se describe el conjunto de datos, así como las técnicas empleadas para el análisis.

Los datos

Los datos codificados del estudio, que consisten en respuestas a encuestas y mediciones de sensores móviles, se almacenaron en Synapse para su distribución controlada a los investigadores [19]. De la base de datos disponible en la plataforma mPower-Synapse¹, se extrajeron 62 parámetros de voz utilizando el software OpenSmile. Estas características se adjuntaron a otros datos: edad, sexo, diagnóstico de EP (sí/no), años desde el diagnóstico, años de medicación. De las grabaciones de fonación se recogieron los audios y el momento de la medicación en el que se realizó la grabación.

Inicialmente, había 51420 grabaciones de voz diferentes. Se identificó un total de 2253 personas mayores de 35 años. Para cada uno, se seleccionó una grabación de la fonación de la vocal /a/. El diagnóstico de EP se consideró como variable de clasificación. Posteriormente, los casos inconsistentes entre el diagnóstico, el año desde el diagnóstico (año de diagnóstico), el año de inicio de la medicación y el momento de toma de medicación en que se hizo el registro, fueron eliminados de la base resultando en 2222 casos.

Métodos estadísticos

La parametrización de cada grabación se realizó en 62 funciones lineales utilizando el software OpenSmile. Se realizaron análisis estadísticos para identificar las variables relevantes que predicen el diagnóstico de la EP a través de

¹ These data were contributed by users of the Parkinson mPower mobile application as part of the mPower study developed by Sage Bionetworks and described in Synapse [[doi:10.7303/syn4993293](https://doi.org/10.7303/syn4993293)]. (<https://www.synapse.org>)

estos parámetros y algunas variables demográficas. Para ello, se utilizaron las siguientes técnicas: Análisis de Componentes Principales (PCA), Análisis de Varianza (ANOVA), redes neuronales Perceptrón Multicapa (MLP) y Regresión Logística (LR)[20].

Para comprobar la capacidad predictiva de los modelos de clasificación, se utilizó el área bajo la curva Característica de Operación (AUC-ROC). Este método mide la capacidad del modelo para clasificar correctamente a las personas con y sin la enfermedad.

Adicionalmente se calculó el porcentaje de casos bien clasificados considerando un punto de corte que maximizara sensibilidad (porcentaje de enfermos bien clasificados) y especificidad (porcentaje de casos sanos bien clasificados).

Resultados

Los audios de la base de datos mPower fueron parametrizados con OpenSmile, derivados en 62 variables específicas de parámetros de voz que luego se unieron a algunas variables demográficas para su posterior análisis. Se utilizaron los siguientes datos demográficos: edad, sexo, diagnóstico profesional, años desde el diagnóstico y año de inicio de la medicación. A partir de las grabaciones de fonación, se analizaron los audios y el punto de tiempo de la medicación en el momento de la grabación.

Reducción de variables

Se realizó un análisis cíclico de la reducción variable a partir de los 62 parámetros iniciales del fonema grabado.

En un principio, el Análisis de Componentes Principales (PCA) se realizó con el 76.1% de la varianza y se pasó de 62 a 33 componentes, considerando 18 factores con valores propios superiores a 1, las variables no explicadas y las más utilizadas en la bibliografía (1 ó 2 variables por factor más las no explicadas por el modelo).

Se realizó entonces el análisis ACP con el 69,5% de la varianza con 9 factores considerando la mayor varianza explicada y la contribución significativa de los autovalores con respecto a los anteriores (sin considerar las bajas contribuciones de los dos últimos factores con autovalor superior a 1). Quedan nueve variables, una para cada factor, la que tiene la mayor carga factorial o, en caso de tener que decidir entre media y desviación de una variable, se elige la media.

Se implementó una agrupación de K-medias en 9 factores PCA y permanecieron 6 grupos, uno de ellos con sólo 4 casos que fueron eliminados, dejando un total de 2218 casos y 5 grupos entre los cuales se contaban 940 casos con EP y 1278 sin EP.

Se aplicó ANOVA a las 9 variables diagnósticas de los 5 grupos y se eliminaron las variables sin diferencias en los valores medios, quedando 5 variables, a saber:

d1: Media de F0 logarítmico en una escala de frecuencia de semitonos.

d2: Promedio de la relación entre la energía del pico armónico espectral en la frecuencia central del primer formante y la energía del pico espectral en F0 en regiones con voz.

d3: Coeficiente de variación de la relación entre la energía del pico armónico espectral en la frecuencia central del primer formante y la energía del pico espectral en F0 en regiones con voz

d4: Media de la pendiente de regresión lineal del espectro de potencia logarítmico dentro de la entropía de la banda 0-500 Hz.

d5: Desviación media de las desviaciones en períodos F0 consecutivos individuales.

En resumen, los 62 parámetros iniciales se redujeron a estos 5 parámetros significativos para la diferenciación del diagnóstico de EP y no EP para los cuales se ajustaron modelos de predicción.

Análisis descriptivo

Un estudio de las características que resultaron del procedimiento de reducción de variables muestra que los 5 parámetros presentan diferencias entre los pacientes que tienen diagnóstico de enfermedad de Parkinson versus los que no lo tienen, aunque son más notorios para las variables que se corresponden con la frecuencia fundamental (F0). Esto condice con resultados mostrados por otros autores, que dan consenso en considerar el incremento en los niveles de la F0 como uno de los principales marcadores de alteración en la voz en la EP [15]. Estos autores, entre otros, pusieron de manifiesto que el incremento en la F0 es paralelo a la gravedad de los síntomas y al avance de la enfermedad. En la Tabla 1 se muestran estadísticos resumen de las 5 variables seleccionadas en el análisis previo. Puede observarse, aún sin dar una prueba estadística, que la diferencia de medias es manifiesta en las dos primeras características.

Tabla 1. Medidas resumen de las variables de fonación

	Con EP		Sin EP	
	media	desvío	media	desvío
d1	29,093	5,099	26,487	4,692
d2	-16,813	30,572	-9,488	22,270
d3	0,241	0,133	0,221	0,134
d4	0,091	0,023	0,085	0,022
d5	0,019	0,021	0,017	0,019

Para dar una visión de las diferencias entre variables según diagnóstico se representó mediante Infostat, en la Fig. 1, un diagrama de cajas (boxplot) para la característica d1 discriminada según corresponda a diagnóstico de EP o no, donde se aprecia el aumento en media de los valores para los pacientes de EP.

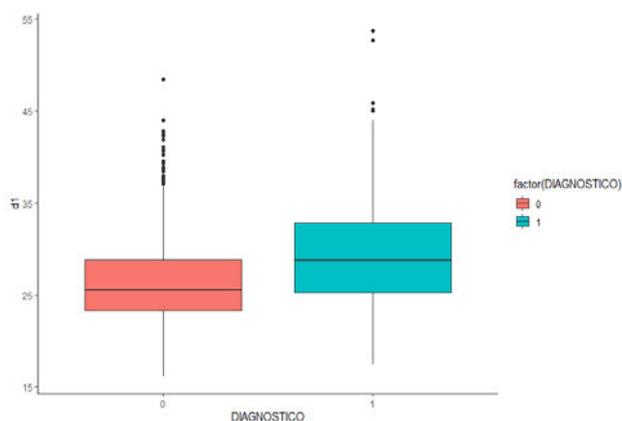


Fig. 1. Boxplot de la variable d1 según Diagnóstico

Se analizó también una posible relación entre género y diagnóstico de EP mediante tablas de contingencia, lo que dio un resultado significativo para la no independencia o relación. Esto se muestra en la Tabla 2 y resultó en un estadístico Chi.cuadrado-MV con p-valor menor a 0,0001. Cabe aclarar que hay 3 casos para los que no se registró género, y esto explica las variaciones en los totales.

Tabla 2. Diagnóstico versus género

	Sin EP	Con EP	Total
Mujer	242	332	574
Hombre	1034	607	1641
Total	1276	939	2215

Para ver la incidencia del género en las variables seleccionadas de la fonación, se muestra en la Fig. 2 que hay diferencia en la variable d1 también por género dentro de cada grupo de diagnóstico. Se aprecia también que los menores valores para la variable d1 se dan en el caso de las mujeres, en cuanto la media pero también en cuanto a los cuartiles.

Esto motiva la introducción de esta variable en los modelos propuestos.

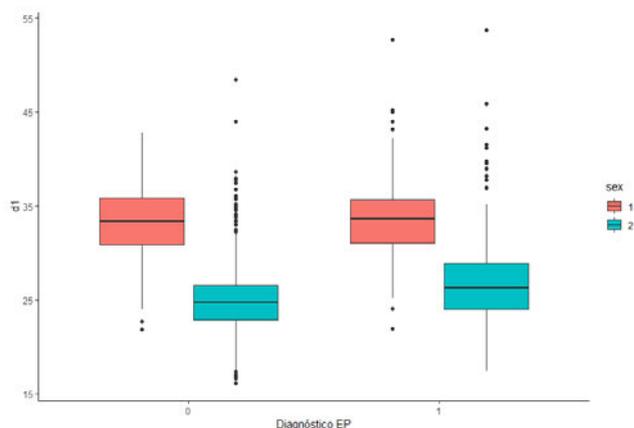


Fig. 2. Box.plots de la variable d1 según Diagnóstico y género (en rojo: mujer y en azul: hombre)

Modelos predictivos

Algunas de las técnicas de clasificación supervisada más utilizadas en el ambiente de ciencia de datos son los modelos de regresión logística y las redes neuronales [14]. Esto se debe a que tienen buen rendimiento para una amplia variedad de problemas. En este trabajo se utilizó el software R (www.r-project.org) para construir modelos predictivos.

Inicialmente, se separó el conjunto de datos en un 70% para entrenamiento de modelos (1552 casos) y un 30% para evaluar estos (666 casos).

Una red neuronal perceptrón de dos capas (MLP) es una red neuronal completamente conectada que consiste en una capa de entrada (que no se cuenta ya que sus neuronas son sólo para representación y, por lo tanto, no procesan), una capa oculta y una capa de salida (en este caso: EP o no-EP).

En un primer análisis se utilizaron redes neuronales multicapa sólo con las variables de fonación ajustando el modelo sobre los datos de entrenamiento. Posteriormente, sobre los datos de evaluación o testing se comparó la capacidad predictiva de los modelos MLP con 62, 33, 9 y 5 parámetros, utilizando como métrica el área bajo la curva ROC. Los valores obtenidos en todos los casos fueron similares, según se muestra en la Tabla 3. Este hecho valida la selección de 5 parámetros para construir nuestro modelo.

Tabla 3. AUC-ROC para MLP según número de variables

Número de variables	AUC – ROC para MLP
62	0.666
33	0.687
9	0.673
5	0.681

A partir de esto, sólo se consideraron las 5 variables de fonación además de género y edad y momento de medicación (med.t), con modelos de redes neuronales de 2 capas y de regresión logística.

Se ajustó con los datos de entrenamiento un modelo de regresión logística (RL) lo que dio un valor estimado de *odd* para el parámetro edad igual a 1.69. Dado que esta variable está medida en múltiplos de 5 años, este valor de *odd* puede interpretarse diciendo que la chance de tener diagnóstico de EP crece un 70% cada 5 años, si género y las otras variables se mantienen fijas.

Finalmente, se comparó la capacidad predictiva de modelos RL y MLP según alternativas de variables adicionales a las 5 elegidas de fonación, lo que se muestra en la Tabla 4. Puede observarse en todos los casos una alta performance de los modelos ajustados, con AUC por encima de 0.82 en todos los casos. El modelo MLP con 5 variables, edad, sexo y punto de medicación se observa con AUC = 0.972. Sin embargo, el resultado no converge con LR.

Tabla 4: Comparación de modelos predictivos

Número de variables de fonación	Otras variables	AUC MLP	AUC LR
5	Género/ edad	0.827	0.832
5	Género/edad/med.t	0.972	No converge

El modelo que incluye el momento de medicación considera en caso de estar medicado el paciente de EP si realizó la grabación “antes” de tomar la medicación, “después” o “en otro momento”. Los no pacientes de EP o los no medicados colocaron “en otro momento”.

En la Fig. 3 se observan las curvas ROC para los dos modelos implementados, RL y MLP, con 5 parámetros de voz, además de la edad y el sexo. El área bajo esa curva es AUC = 0.832 para la regresión logística y AUC= 0.827 para la red neuronal. Ambos modelos presentan una muy alta capacidad predictiva por lo que serán utilizados para la predicción.

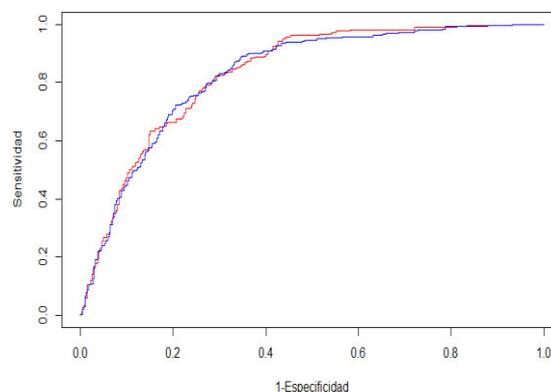


Fig. 3. Curvas ROC de modelo RL (rojo) y MLP (azul).

Dado que el objetivo de la implementación de estos modelos es la predicción, y que el modelo de regresión logística da como respuesta la probabilidad estimada de que el paciente tenga diagnóstico de EP, la clasificación con este modelo implica la elección de un punto de corte. Esto es, se debe seleccionar un valor según el cual, si la probabilidad estimada por el modelo resulta mayor que este, el paciente se clasifica como positivo para el diagnóstico de EP, en caso contrario se lo clasifica como negativo. Para elegir este punto de corte se evaluaron Sensitividad, tasa de bien clasificados positivos, y la Especificidad, tasa de bien clasificados negativos. En la Fig. 4 se muestran las curvas de ambos indicadores, graficados según sea el punto de corte elegido. Para maximizar ambos indicadores, el punto de intersección se muestra en el valor de abscisa 0.53.

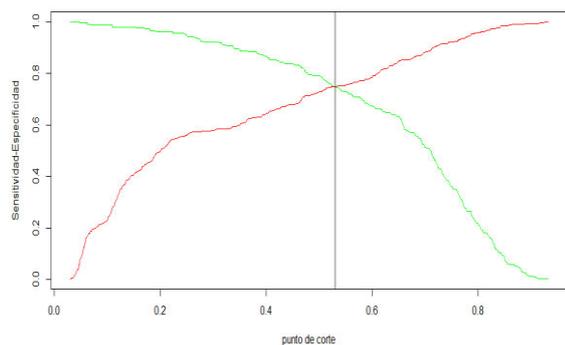


Fig. 4. Sensitividad y Especificidad vs punto de corte (RL)

Cabe aclarar que un estudio similar para el modelo MLP arrojó un valor óptimo de corte 0.54, similar al de RL. Considerando en cada caso el punto de corte elegido, el porcentaje de clasificación correcta (ACCU o exactitud) en el conjunto de testing con el modelo RL fue $ACCU = 0.7485$, mientras que con el modelo MLP este porcentaje de bien clasificados dio como resultado $ACCU = 0.756$.

Si bien puede creerse que la exactitud es la medida a preferir en estos casos, esta depende del punto de corte a partir del cual se considera que el paciente tiene predicción de EP, el cual puede ser subjetivo. Esto lleva a preferir el análisis de las curvas de sensibilidad y especificidad las que permiten seleccionar el punto de corte. El valor resultante del corte de las curvas permite controlar el error de clasificar erróneamente a una persona como EP, con la consecuencia de tratamientos innecesarios, y también el error de clasificar como sano a alguien que efectivamente tiene EP, lo que conlleva retrasar el tratamiento.

Finalmente, las distintas métricas utilizadas muestran una alta performance de ambos modelos, y muestra la factibilidad de usar esta metodología para monitoreo de pacientes.

Conclusiones

Este trabajo se encuadra en un proyecto de investigación que propone el análisis estadístico de señales acústicas de la voz como método objetivo y no invasivo para el diagnóstico precoz de la enfermedad de Parkinson (EP).

La contribución de este trabajo radica en la selección de cinco parámetros lineales sencillos de medir de la voz del software libre OpenSMILE. Además de la selección de 5 variables de los 62 originales, el estudio analizó la relevancia estadística de individual's edad y género. Se obtuvieron capacidades predictivas superiores a 0,82 para la identificación de diagnósticos de EP y no EP comparando redes neuronales y modelos de regresión logística.

Con la aplicación mPower App, las grabaciones de voz fueron recogidas por los participantes voluntarios con su

propio Smartphone, y estos datos se añadieron al modelo de predicción del diagnóstico de EP. Las grabaciones de los teléfonos móviles permiten el seguimiento de los pacientes con Parkinson a bajo coste y la detección de alteraciones que requieren asistencia a la consulta médica.

El análisis acústico de señales de voz demuestra aquí ser útil para la predicción de enfermos de EP diferenciando no enfermos de enfermos.

La profundización de este línea de estudio permitiría la predicción de la severidad de la enfermedad de Parkinson y se espera facilite el seguimiento remoto no invasivo de la progresión de la enfermedad.

Agradecimientos

Este trabajo se encuentra soportado por el Proyecto PROINCE-C224, Universidad Nacional de La Matanza, Argentina.

Referencias

- [1] Langston, J. "Parkinson's disease: Current and future challenges", *Neurotoxicology* 23 (4-5), 2002, pp. 443-450.
- [2] Tsanas, A., Little, M. A., McSharry, P. E., Ramig, L. O. "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech test", *IEEE Transactions of Biomedical Engineering* 57 (4). 2010. Pp. 884-893
- [3] Silva, D. G., Oliveira, L. C., Andrea, M. "Jitter estimation algorithms for detection of pathological voices", *EURASIP Journal of Advances in Signal Processing* 2009, pp.1-9.
- [4] Teixeira, J. P., Oliveira, C., Lopes, C. "Vocal acoustic analysis- Jitter, shimmer and HNR parameters", *Procedia Technolology* 9, 2013, pp. 1112-1122.
- [5] Dixit, V. Mittal, Y. Sharma. "Discrimination of People with Parkinson (PWP) disease on the basis of voice parameter analysis", *International Journal of Computer Applications* 94 (13), 2014, pp. 9-14.
- [6] Shirvan, R. A., Tahami, E. "Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method", in: 18th Conference on Biomedical Engineering, Theran, IEEE, 2011, pp. 278-283.
- [7] Benba, A., Jilbab, A., Hammouch, A. "Discriminating between patients with Parkinson and neurological diseases using cepstral analysis", *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24 (10), 2016, pp. 1100-1108.
- [8] Sakar, B. E., Isenkul, M. E., Sakar, C. O., Sertbas, A., Gungen, F., Delil, S., Apaydin, H., Kursun, O. "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings", *IEEE Journal of Biomedical and Health Informatics* 17 (4), 2013, pp. 828-834.
- [9] Peker, M., Sen, B., Delen, D. "Computer-aided diagnosis of Parkinson's disease using complex-valued neural networks and mRMR feature selection algorithm", *Journal of Healthcare Engineering* 6 (3), 2015, pp. 281-302.
- [10] Tsanas, A., Little, M. A., McSharry, P., Spielman, J., Ramig, L. O. "Novel speech signal processing algorithms for high accuracy classification of Parkinson's disease", *IEEE Transactions of Biomedical Engineering* 59 (5), 2011, pp. 1264-1271.
- [11] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., Ramig, L. O. "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease", *IEEE Transactions of Biomedical Engineering* 56 (4), 2009, pp. 1015-1022.

- [12] Schwab, P., Karlen, W. "PhoneMD: Learning to Diagnose Parkinson's Disease from Smartphone Data". arXiv preprint arXiv: 2018, pp. 1810.01485.
- [13] Giuliano, M., García-López, A., Pérez, S., Pérez, F. D., Sposito, O., & Bossero, J. "Selection of voice parameters for Parkinson's disease prediction from collected mobile data". In 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA). 2019 pp. 1-3).
- [14] Gamboa J, Jiménez-Jiménez FJ, Mate MA, Cobeta I. "Alteraciones de la voz causadas por enfermedades neurológicas". Revista Neurología, 2001; 33, pp. 153-68.
- [15] Klatt, DH. "Software for cascade/parallel formant synthesizer". Journal of the Acoustical Society of America, 1980; Vol 67, No3: 971-995.
- [16] Fernández S., Ruba D., Marqués M., Sarraqueta, L. "Voz del anciano". Rev Med Univ Navarra, 2006, V50, pp.44-48.
- [17] Elisei, N.G. "Análisis acústico de la voz normal y patológica utilizando dos sistemas diferentes: ANAGRAF y PRAAT". Interdisciplinaria, 2012, 29 (2), pp. 339-357.
- [18] Eyben F., Weninger F., Gross F., and Schuller B. "Recent developments in openSMILE, the munich open-source multimedia feature extractor". In Proc. of ACM MM 2013, Barcelona, Spain. New York, NY, USA: ACM, 2013, pp. 835-838.
- [19] Bot, B. M., Suver, C., Neto, E. C., Kellen, M., Klein, A., Bare, C., ... & Friend, S. H. "The mPower study, Parkinson disease mobile data collected using ResearchKit". Scientific data, 2016, 3, 160011.
- [20] Hastie, T.; Tibshirani, R. y Friedman, J. The Elements of Statistical Learning – Data Mining, Inference, and Prediction. Springer, Nueva York (2001).