

Predicción de Parkinson a partir de Datos de Tecleo

Rodrigo Domínguez Araya
Estudiante Universidad de Talca
Curicó, Chile
rdominguez@alumnos.utalca.cl

I. INTRODUCCIÓN

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo progresivo que afecta el sistema nervioso central, caracterizándose por la degeneración de neuronas dopaminérgicas en la sustancia negra del cerebro. Esta condición se

manifiesta en síntomas motores. A medida que la población envejece, la prevalencia de la enfermedad de Parkinson está aumentando, lo que

representa un desafío significativo para la salud pública. La detección temprana y el monitoreo de la progresión de la enfermedad son cruciales para el manejo adecuado de la condición. El desafío PhysioNet se centra en el análisis de datos relacionados con el comportamiento

motor de los pacientes mediante la recopilación de datos de escritura a

teclado. Se detallará el proceso de diseño y la evaluación de modelos de aprendizaje automático, específicamente un modelo de Gated Recurrent Unit (GRU), utilizado para clasificar a los pacientes según su condición.

II. DATOS

Contamos con muestras de tecleo de 85 pacientes los cuales 43 están sanos y 42 con la enfermedad de Parkinson.

Entre los datos que encontramos en los datasets se encuentran:

- pid – Id del paciente
- gt – True para Parkinson y Falso para no tiene Parkinson.
- updrs108 - Unified Parkinson's Disease Rating Scale part III (UPDRS-III)
- afTap - Alternating finger tapping resultado.
- sTap - Single key tapping resultado

- nqScore - neuroQWERTY index (nQi)
- Typing speed
- file_n – El CSV de los datos de tecleo personal del paciente.

Y el archivo perteneciente a cada paciente que incluye sus datos de tecleo:

- La tecla presionada.
- Cuanto mantuvo la tecla.
- Tiempo de Liberación de la Tecla desde el segundo 0.
- Tiempo en que presiono la Tecla desde el segundo 0.

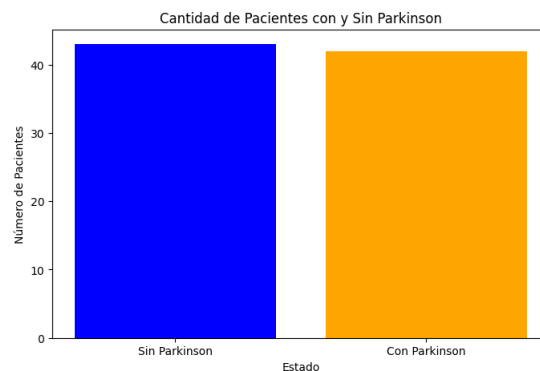


Ilustración 1: Pacientes con y sin EF

III. LIMPIEZA DE DATOS

Algunos pacientes cuentan con dos muestras de tecleo, por lo que en esos casos solo se usará una para que tengamos la misma cantidad de información para todos los pacientes.

Algunas muestras duran más que otras por lo que se usará la información recopilada hasta el segundo 563 para cada archivo ya que todos los pacientes teclearon mínimo 563 segundos. Los valores

anómalos se limpiaron usando desviación estándar e IQR.

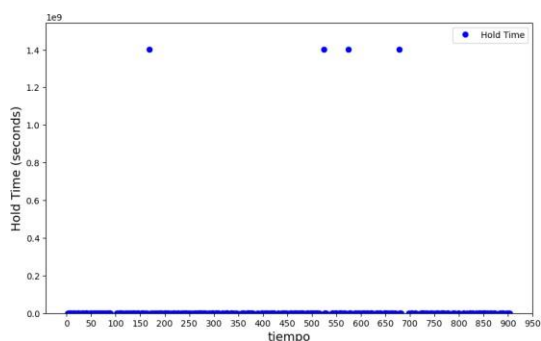


Ilustración 2: Gráfico Hold time vs tiempo sin limpiar; paciente 60

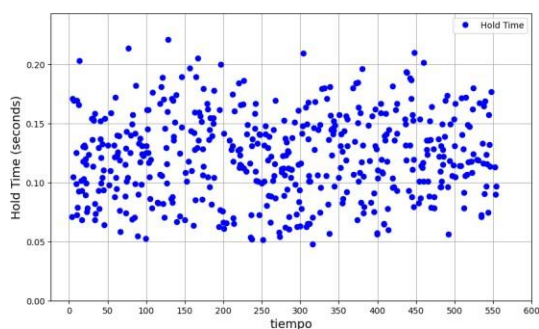


Ilustración 3: Gráfico Hold time vs tiempo limpio; paciente 60

IV. MANEJO DE LOS DATOS

Manejo de los datos: En base al archivo general se usó solo las columnas de pID, gt y typing speed ya que son los datos que se usaran para la resolución del problema. En cuanto a los CSV pertenecientes a la información de cada paciente, la información se usó para calcular varios nqScore en intervalos de 20 segundos para poder tener datos secuenciales por cada paciente y poder aplicar GRU.

V. NQSCORE

nqScore: El nqScore (neuroQWERTY index) es una métrica utilizada para evaluar la actividad motora de los

pacientes, especialmente en el contexto de trastornos neurológicos como la enfermedad de Parkinson. Este índice se calcula a partir de datos recopilados durante tareas de escritura a teclado,

analizando parámetros como el tiempo de retención de las pulsaciones de teclas, la duración de las presiones y los tiempos de liberación

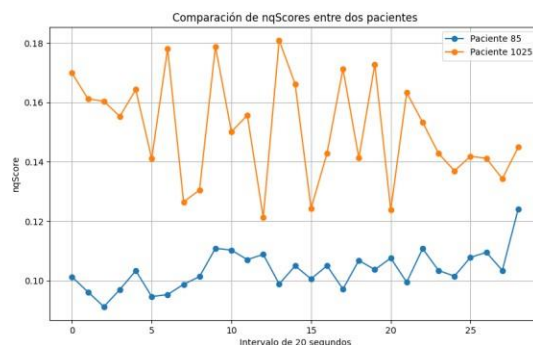


Ilustración 4: Comparación de nqScore de paciente 85 y 1025

En la ilustración 4 podemos ver graficados los 28 nqScore calculados del paciente 85 y 1025 cada puntaje se

calcula cada 20 segundos de un total aproximado de 560 segundos de tecleo que tiene cada paciente en su respectivo archivo. La ilustración nos muestra una clara separación entre ambos pacientes siendo el paciente 1025 una persona con Parkinson y el paciente 85 una persona sin Parkinson

VI. SELECCIÓN DE MODELOS

Selección de modelos: Para el análisis de los datos de los sujetos con Parkinson, se eligió un modelo de Red Neuronal Recurrente (RNN), específicamente la variante llamada Gated Recurrent Unit (GRU). La razón detrás de esta elección incluye:

- **Naturaleza Secuencial de los Datos:** Dado que los datos recopilados están relacionados con el tiempo (tecleos a intervalos de 20 segundos), los modelos RNN, y en particular GRU, son efectivos para aprender patrones temporales.
- **Conjunto de Datos limitados:** GRU es bueno cuando se trabaja con un tamaño de conjunto de datos limitado como lo es en este caso que se cuenta con 85 sujetos.
- **Flexibilidad en La Captura de Patrones Complejos:** Los datos de tecleo pueden estar influenciados por diversos

factores externos que afectan la manera en que los pacientes escriben y como las GRU son capaces de aprender patrones complejos y no lineales esto puede ser beneficioso para este caso.

VII. ENTRENAMIENTO Y EVALUACION DE MODELOS

Como se mencionó anteriormente se implementó el modelo de red neuronal GRU para clasificar la presencia de Parkinson en función de las puntuaciones de nqScore derivadas de los datos de tecleo. Los pasos realizados fueron:

- **Preparación de Datos:** Los datos fueron cargados desde archivos CSV (datos.csv y archivos aparte que contenía: Key Pressed, Hold Duration (s), Key Release Time (s), Key Press Time (s)). Las puntuaciones de nqScore fueron calculadas utilizando una función que analiza los tiempos de retención de las pulsaciones de teclas en el archivo de cada paciente, dividiendo los datos recopilados en 560 segundos aproximadamente en intervalos de 20 segundos y así generando varios nqScore por paciente para así tener datos secuenciales y poder usar GRU.
- **División de datos:** Se separaron los datos en conjuntos de entrenamiento y prueba. El 80% de los datos se utilizó para el entrenamiento del modelo, mientras que el 20% restante se reservó para evaluar la capacidad de predecir.
- **Evaluación del modelo:** Después del entrenamiento, el modelo fue evaluado utilizando el conjunto de prueba. Se calculó la precisión, así como otras métricas de rendimiento, como el recall y el F1-score. Estas métricas aparte de solo el accuracy son importantes para entender la capacidad predictiva del modelo que se entrenó.

VIII. RESULTADOS

Los resultados mostraron que el modelo alcanzó una precisión del 82,35%, con un reporte de clasificación que nos muestra de manera más detallada la capacidad del modelo de identificar correctamente a los pacientes con y sin Parkinson.

Precisión: El 80% de los pacientes predichos como sanos, realmente no tenían la enfermedad. El 83% de los pacientes predichos como enfermos, realmente tenían Parkinson.

Recall: El 67% de los verdaderos negativos fueron identificados correctamente y el 91% de los verdaderos positivos fueron correctamente identificados.

F1-Score: 73%, refleja un equilibrio entre precisión y recall, pero sugiere que hay margen para mejorar la detección de pacientes sanos. 87%, indica un buen equilibrio y una alta efectividad en la detección de pacientes con Parkinson.

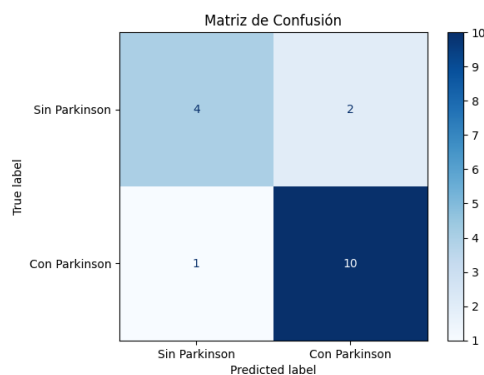


Ilustración 5: Matriz de confusión del modelo

El modelo mostro una precisión del 82% en general, lo que significa que en la mayoría de los casos podrá predecir bien, pero puede fallar, por lo que no debería tomarse como un diagnóstico médico, pero si pudiese usarse para hacerse una idea y en base a eso hacerse exámenes médicos para la detección de Parkinson.

También se probó el modelo con diferentes intervalos de tiempo y los resultados son:

Intervalo(s eg)	Accura cy	F1- scor e	Precisi on	Reca ll
0,2	0,35	0,26	0,18	0,5
0,5	0,41	0,39	0,51	0,51
5	0,76	0,72	0,76	0,70
15	0,76	0,74	0,74	0,74
20	0,82	0,8	0,82	0,79
50	0,71	0,69	0,69	0,70

Los resultados obtenidos muestran que el intervalo de tiempo utilizado para calcular el nqScore influye significativamente en el rendimiento del modelo GRU. Intervalos muy cortos, como 0.2 y 0.5 segundos, generan datos fragmentados y ruidosos, lo que afecta negativamente el rendimiento (Accuracy: 0.35 y 0.41, respectivamente). Esto podría deberse a que no se generan suficientes

datos en esos intervalos para calcular un nqScore representativo y útil para el modelo.

A medida que los intervalos aumentan, el modelo mejora su capacidad de capturar patrones relevantes, alcanzando su máximo rendimiento con un intervalo de 20 segundos (Accuracy: 0.82, F1-Score: 0.80). Esto sugiere que este intervalo proporciona el equilibrio óptimo entre la cantidad de datos disponibles y la capacidad de capturar patrones temporales significativos relacionados con la enfermedad de Parkinson.

Sin embargo, cuando el intervalo se extiende a 50 segundos, el rendimiento disminuye (Accuracy: 0.71, F1-Score: 0.69). Esto puede explicarse por la dilución de patrones específicos debido a la acumulación de información menos relevante o más heterogénea en intervalos muy largos. Por lo tanto, un intervalo de 20 segundos parece ser el más adecuado para este análisis, ya que permite capturar patrones robustos sin perder precisión ni recall.

IX. COMPARACION CON OTROS MODELOS

Modelo	Accuracy	F1-score	Precision	Recall
GRU	0,82	0,8	0,82	0,79
Random forest	0,53	0,51	0,71	0,64
Xtream Gradient Boosteam	0,65	0.65	0,68	0.69
SVM	0,53	0,51	0,71	0,64

- GRU (Gated Recurrent Unit)

Resultados: Accuracy 82%, F1-Score 0.80

Las GRU sobresalen en este problema debido a su capacidad para manejar datos secuenciales, como los nqScores calculados en intervalos de tiempo. Capturan dependencias temporales y patrones complejos, lo que las hace ideales para analizar cómo evolucionan los patrones de tecleo en pacientes con Parkinson. Además, su diseño más simple comparado con LSTM las hace eficientes para conjuntos de datos pequeños, como el de 85 pacientes.

- Random Forest

Resultados: Accuracy 53%, F1-Score 0.51

Random Forest no maneja datos secuenciales y se basa en características estáticas, como valores

promedio de nqScore. Aunque es efectivo para capturar relaciones no lineales, pierde información clave sobre la evolución temporal de los datos de tecleo, lo que resulta en un bajo rendimiento.

- XGBoost (Extreme Gradient Boosting)

Resultados: Accuracy 65%, F1-Score 0.65

XGBoost optimiza mejor que Random Forest al manejar relaciones no lineales y patrones complejos entre las características derivadas. Sin embargo, carece de mecanismos para procesar la estructura temporal de los datos, lo que limita su efectividad en comparación con las GRU.

- SVM (Support Vector Machine)

Resultados: Accuracy 53%, F1-Score 0.51

SVM no está diseñado para datos secuenciales y depende de relaciones claras entre características estáticas y etiquetas. En este caso, los datos son ruidosos y tienen una naturaleza temporal compleja que SVM no puede aprovechar. Esto, junto con el tamaño reducido del conjunto de datos, limita su rendimiento.

X. DISCUSIÓN

Los resultados obtenidos muestran que el modelo GRU, diseñado para manejar datos secuenciales, es adecuado para analizar los datos de tecleo y predecir la enfermedad de Parkinson. Sin embargo, la precisión del 82% indica que existen limitaciones importantes. La capacidad del modelo para identificar pacientes sanos es menor lo que sugiere un riesgo de falsos positivos (pacientes sin Parkinson diagnosticados con Parkinson). Esto puede ser atribuido a un tamaño reducido del conjunto de datos, limitando su capacidad para generalizar. Además, factores externos no controlados, como el nivel de habilidad en tecleo de los participantes, pueden haber influido en los resultados.

Las implicaciones de estos hallazgos son significativas, ya que este tipo de análisis tiene el potencial de proporcionar una herramienta no invasiva y económica para detectar EP en etapas iniciales. Sin embargo, el modelo no debe considerarse un sustituto de métodos diagnósticos médicos, sino como un complemento que puede guiar a los pacientes hacia pruebas adicionales.

XI. CONCLUSIONES

Este proyecto aporta una metodología innovadora para el análisis de datos de tecleo con fines médicos, mostrando que los modelos GRU pueden identificar patrones relacionados con la enfermedad de Parkinson. Los hallazgos destacan la importancia del manejo de datos secuenciales en aplicaciones médicas.

Se podría ampliar el tamaño del conjunto de datos para mejorar la generalización del modelo y explorar la integración de otros parámetros motores o cognitivos. Además, futuras investigaciones podrían investigar cómo incorporar esta herramienta en sistemas médicos existentes, promoviendo su uso en contextos clínicos como un soporte al diagnóstico temprano de la enfermedad de Parkinson.

XII. REFERENCIAS

- Arroyo-Gallego, T., Ledesma-Carbayo, M. J., Butterworth, I., Matarazzo, M., Montero-Escribano, P., Puertas-Martín, V., Gray, M. L., Giancardo, L., & Sánchez-Ferro, Á. (2018). Detecting Motor Impairment in Early Parkinson's Disease via Natural Typing Interaction With Keyboards: Validation of the neuroQWERTY Approach in an Uncontrolled At-Home Setting. *Journal of medical Internet research*, 20(3), e89. <https://doi.org/10.2196/jmir.9462>
- Roy, S., Roy, U., Sinha, D., & Pal, R. K. (n.d.). Imbalanced Ensemble Learning in Determining Parkinson's Disease Using Keystroke Dynamics. SSRN. Retrieved from <https://ssrn.com/abstract=4137564>
- Giancardo, L., Sánchez-Ferro, A., Arroyo-Gallego, T., Butterworth, I., Mendoza, C. S., Montero, P., Matarazzo, M., Obeso, J. A., Gray, M. L., & San José Estépar, R. (2016). Computer keyboard interaction as an indicator of early Parkinson's disease. *Scientific Reports*, 6, 34468. <https://doi.org/10.1038/srep34468>