



UNIVERSIDAD NACIONAL ANDRÉS BELLO
FACULTAD DE CIENCIAS DE LA REHABILITACIÓN
ESCUELA DE KINESIOLOGÍA

Reconocimientos y clasificación de patrones de
marcha neurodegenerativa mediante variables
temporoespaciales y machine learning. Esclerosis
Múltiple, Parkinson, Huntington.

TESIS PARA OPTAR AL GRADO ACADÉMICO DE LICENCIADO EN
KINESIOLOGÍA

IGOR IRAZZOKY ROJAS

DOCENTE GUÍA

KLGO. DIEGO ROBLES CRUZ

SANTIAGO 2017

Agradecimientos

A todas las personas que estuvieron detrás de este hermoso trabajo y fueron parte del proceso de estudio, a la Universidad Nacional Andrés Bello, a la Facultad de Ciencias de la Rehabilitación, a la Escuela de Kinesiología por abrir sus puertas ante nuestras necesidades de querer conocer y aprender de este lindo camino como la Kinesiología y el estudio del movimiento humano; por este proceso conocimos personas que nos ayudaron a seguir y nos inspiraron en el desarrollo del conocimiento.

Gracias a los docentes que en algún minuto fueron parte de esta escuela y que por circunstancias individuales no están presentes, no solo aportando sus conocimientos sino también de manera emocional. Gracias en forma directa a nuestras familias, a nuestros docentes cercanos, quienes se volvieron parte importante en nuestro proceso formativo y se volvieron amigos durante este periodo, considerando que fueron ellos quienes siempre nos impulsaron a seguir adelante, a pesar de no utilizar el lenguaje verbal, solo pensando en estar presentes, el amor que sentimos hacia éstas personas nos inspira a seguir viviendo y crecer en el ámbito que nos apasiona.

Gracias de sobremanera a los amigos y familiares que estuvieron presentes.

Dedicatoria.

Este estudio ésta dedicado a mis profesores Diego Robles Cruz y Esteban Montero Álvarez, quienes me guiaron y me dieron la oportunidad de acercarme al mundo de la investigación. A mi madre Silvia Rojas Trincado, y a mis amigos incluyendo a mi entrenador en la enseñanza media, quienes me mostraron la carrera, y por sobre todo a la vida por guiarme a ella y a esta escuela en este período de tiempo en particular.

Dando por referencia lo que se denomina encuentro físico, un evento en un momento determinado.

Muchas gracias.

Resumen

La expresión de la marcha humana asociada a afecciones neurológicas es de compleja descripción y se representa por ser marchas de tipo fluctuante y con patrones de movimiento complejos. La misión de describir marchas neurodegenerativas además de lograr su correcta clasificación a través de métodos robustos es uno de los objetivos de este proyecto de investigación. El uso de información sencilla como parámetros temporo espaciales es el llamado del sitio web physionet.org el cual libera esta base de datos obtenidos en la Clínica Neurológica del Hospital General de Massachussetts y levanta el desafío científico de identificar y clasificar los patrones de marcha de tres patologías neurodegenerativas de manera objetiva bajando el rango de error de un análisis clínico específico. Para dar solución a esta problemática, se propone el uso de redes neuronales (ARNA) y máquinas de soporte vectorial (MSV).

Índice general

I	Preliminares	12
1.	Introducción	13
1.1.	Problema de la Investigación	16
1.1.1.	Planteamiento del problema	16
1.2.	Objetivos	17
1.2.1.	Objetivo General	17
1.2.2.	Objetivos específicos	17
II	Marco Teórico	19
2.	Enfermedades neurodegenerativas.	20
2.0.1.	<i>Enfermedad de Parkinson (EP)</i>	20
2.0.2.	<i>Enfermedad de Huntingtons (EH)</i>	21
2.0.3.	<i>Esclerosis Lateral Amiotrófica (ELA)</i>	22
2.1.	<i>Clustering</i>	23
2.1.1.	<i>Análisis de Redes Neuronales Artificiales (ARNA)</i>	23
2.1.2.	<i>Máquinas de Soporte Vectorial (MSV)</i>	24
2.2.	Características del Aprendizaje Automatizado	25
2.2.1.	<i>Matriz de Confusión</i>	25

<i>ÍNDICE GENERAL</i>	8
2.2.2. Receiver Operating Characteristic	25
III Marco Experimental	28
3. Materiales y Metodología	29
3.1. Diseño de Estudio	29
3.1.1. Descriptivo	30
3.2. Clasificación	31
3.2.1. Experimentos Evaluados	32
3.3. Materiales y Métodos	33
3.3.1. Sujetos.	33
3.3.2. Análisis Estadístico	36
3.3.3. Clustering	36
IV Discusión de Resultados	37
4. Resultados	38
5. Discusión	47
6. Conclusión	49

Índice de cuadros

2.1. Matriz de confusión	27
3.1. Porcentaje Clases	33
3.2. División por Género	34
3.3. Peso por Clase	34
3.4. Resultados de Normalidad	35
3.5. Tabla de Distribución Muestral de la Edad	35
4.1. Patologías vs Grupo Control	43
4.2. Patologías vs Control y Ganglios Basales vs ELA	44
4.3. Comparación entre Patologías	45
4.4. Severidad y Velocidad	46

Índice de figuras

3.1. Histograma Edad	34
4.1. Distribución de la Marcha	39
4.2. Balanceo y Doble Apoyo	40
4.3. Intervalos de Apoyo Derecho e Izquierdo	40
4.4. Fases de Balanceo Derecho e Izquierdo	41
4.5. Velocidad de la Marcha	42

Parte I

Preliminares

Capítulo 1

Introducción

Las enfermedades neurológicas son patologías de alta prevalencia, entre ellas las enfermedades neurodegenerativas, las cuales afectan la calidad de vida y pérdida de la funcionalidad de los pacientes. Se estima una tasa de muerte de 6,8 al año, y teniendo un nivel de gasto alto, esto sólo en la población europea un valor de 139.000.000.000 al año [1].

La literatura ha descrito las patologías neurodegenerativas como enfermedades o desórdenes que afectan al sistema nervioso central o periférico de manera progresiva. Estos desórdenes abarcan; Alzheimer, Temblor Esencial, Enfermedad de Parkinson, Demencia de Cuerpos de Lewy, Demencia fronto-temporal, Esclerosis Lateral Amiotrófica, Enfermedad de Huntington y desorden de Creutzfeld-Jakob [2]. Las enfermedades con mayor prevalencia son la demencia y sus derivados que están asociados en gran medida a la Enfermedad de Parkinson, la Esclerosis Lateral Amiotrófica, Alzheimer y la Enfermedad de Huntington [2]. Estas afecciones principalmente alteran el patrón de movimiento, exceptuando el Alzheimer, siendo de gran interés para la investigación en la ciencia del movimiento. La Enfermedad de Huntington (EH) es difícil determinar su incidencia y pre-

valencia debido a que es hereditaria y corresponde mayoritariamente a una población específica [3], siendo una enfermedad genética dominante, la literatura describe que ha aumentado considerablemente de manera inespecífica, hace dos décadas una prevalencia de uno por millón y resaltando la dificultad del seguimiento poblacional y variante en el mundo, buscando la prevalencia por mortalidad en hospitales e instituciones mentales [3], la principal causa de muerte en los pacientes con EH son neumonías y problemas cardíacos, describiendo que los criterios son estrictamente no comparables, debido a estas variables y limitaciones es difícil demostrar hoy en día la prevalencia de estos sujetos [3], aún así, la literatura ha descrito que en los sujetos con EH aumenta su prevalencia al doble por cada 150 años, 6 generaciones aproximadamente, alcanzando solo en algunas poblaciones de norteamérica a 4,09 por millón, existiendo controversia con otros estudios en distintas poblaciones. Si consideramos el aumento de las poblaciones y migraciones es una enfermedad de relevancia mundial, del punto de vista económico y gran valor teórico [3].

La incidencia y prevalencia de la Enfermedad de Parkinson (EP) descrita ha ido en aumento en la última década, correspondiendo actualmente a 100 por millón detallando que es mayoritariamente en la población de la tercera edad [2]. La EP es una enfermedad que de acuerdo con el estado actual de los conocimientos relatados está subdiagnosticada, correspondiendo a un desorden neurodegenerativo heterogéneo debido a un proceso ideopático de las vías dopaminérgicas de evolución progresiva, produciendo lesiones multifocales principalmente en los ganglios basales a nivel de la sustancia negra y del locus coeruleus, causando la dificultad en la ejecución de movimientos y la rigidez [4].

La Esclerosis Lateral Amiotrófica (ELA) es una afección neurodegenerativa que afecta las células nerviosas del cerebro y la zona anterior del cordón medular produciendo un daño a las motoneuronas [5]. La literatura describe esta condi-

ción como multifactorial. Su frecuencia es relativa en el mundo en Japón alcanza 0,31 personas/año y en Suecia 100000 personas por año [2], la literatura detalla en Buenos Aires Argentina en el Hospital Ramos Mejias muestra una incidencia anual de 6/1000000 al año [6]. Siendo que en la literatura publicada en Taiwan muestran una incidencia muy relativa entre 868 por 100000 al año [7]. No cabe duda que la ELA se hace cada vez más relevante para la población mundial debido a su aumento en la incidencia el gran gasto económico, y por sobretodo la alta mortalidad que han ido en aumento [8, 9].

Debido a la condición y deterioro neural estas patologías describen en su sintomatología pérdida del control motor, afectando la marcha de la población, hasta llegar a la dependencia de un ventilador mecánico y luego la muerte, siendo la pérdida de control motor el comienzo de su sintomatología y de gran importancia para su pesquisa temprana.

La Marcha Humana es un proceso de locomoción en el cual el cuerpo humano, en posición erguida, se mueve hacia adelante proyectando su centro de masas, alternativamente por ambas piernas [10, 11]. La marcha humana es un fenómeno complejo que involucra a estructuras y funciones del sistema neuro-musculo-esquelético (SNME) [12]. Es un sistema multifactorial, el que es influenciado por variables tan simples como la masa corporal, el género e incluso factores mentales como la depresión [12, 13, 14, 15]. Si bien el proceso de la marcha es de carácter individual, las semejanzas entre distintos sujetos son tales, que puede hablarse de un patrón característico de marcha humana normal, el cual experimenta modificaciones debido a la influencia de diversos factores y particularmente, bajo determinadas condiciones patológicas [12].

La Marcha Humana ha sido altamente investigada y estudiada en condiciones normales y patológicas describiendo modelos y patrones [16, 17, 18, 19]. Así, a su vez ha sido descrita y clasificada con distintas técnicas como análisis visual,

cronómetro, cinta métrica, interruptores plantares, plataformas instrumentadas, técnicas de análisis cinemático y mediante clustering, utilizado para analizar la muestra de la recopilación de datos de algunas de estas pruebas.

Se ha descrito que el fenómeno físico de la marcha es altamente impredecible, debido a lo anteriormente mencionado y su multi-factorialidad imposibilita el hecho de asegurar que un dato temporal coincidirá con otra medición, es por esto que las técnicas lineales convencionales se ven limitadas.

1.1. Problema de la Investigación

1.1.1. Planteamiento del problema

La literatura ha descrito que la marcha humana puede ser clasificada con diversas técnicas y modelos, desde un simple péndulo hasta clasificadores complejos de aprendizaje automatizado, estos últimos son conocidos como clasificadores o clustering [20]. Las técnicas de clasificación son métodos de análisis de multivarianza mostrando su mayor potencial en el reconocimiento de patrones. Las redes neuronales es un tipo de clasificador supervisado, caracterizadas por hacer su análisis y solución de problemas a través de los ajustes de los pesos conocidos, entregando como salida una matriz de confusión y una Receiver Operating Characteristic o curva ROC, que demuestran la calidad de la clasificación.

Se ha relatado ampliamente la marcha patológica en enfermedades neurodegenerativas a través de distintos algoritmos computacionales [11, 21], inclusivamente en la misma población [22, 23, 24], estableciendo un interés mundial, dando una relevancia al método de clasificación y descripción de la marcha humana. Sin embargo en la actualidad se desconocen los diversos motivos de análisis, poca comparación entre bases de datos en los distintos modelos y clasificaciones, adicionalmente de la pobre descripción de la marcha en base de datos, y la refuta

de los procesos metodológicos.

Es por esto, que es de suma importancia describir la marcha neurodegenerativa y la clasificación de patrones a través de métodos rigurosos en distintos tipos de bases de datos alternando los sujetos de estudio y sus variables, con el propósito de valorar el real aporte de estos parámetros temporo espaciales como atributos o características que permitan la correcta clasificación.

¿Es posible describir y clasificar la marcha neurodegenerativa en las patologías de ELA, EP, EH. mediante machine learning y las variables temporoespaciales?

1.2. Objetivos

1.2.1. Objetivo General

Discriminar distintos patrones de marcha asociadas a tres enfermedades neurodegenerativas (Parkinson, Huntington y Esclerosis Lateral) mediante el estudio de variables temporoespaciales y clasificadores supervisados como redes neuronales artificiales y máquinas de soporte vectorial.

1.2.2. Objetivos específicos

- Determinar el comportamiento de las variables temporoespaciales para cada enfermedad neurodegenerativa.
- Discriminar las variables de mayor relevancia para el uso de los clasificadores.
- Clasificar los distintos tipos de marcha según tipo de patología en base a las distintas técnicas de clustering.

Parte II

Marco Teórico

Capítulo 2

Enfermedades neurodegenerativas.

En la actualidad se conocen más de 100 enfermedades neurodegenerativas; Alzheimer, Parkinson, Huntington y Esclerosis Lateral Amiotrófica son las más importantes por su frecuencia y/o gravedad. Las enfermedades neuro-degenerativas son muy heterogéneas en cuanto a sus síntomas y hallazgos anatomopatológicos. La adecuada clasificación de las enfermedades es imprescindible en medicina para el correcto diagnóstico y pronóstico de los pacientes y para la investigación. La clasificación de las enfermedades neurodegenerativas es de alta complejidad debido a que se solapan entre sí [25].

2.0.1. *Enfermedad de Parkinson (EP)*

Corresponde a un desorden neurodegenerativo heterogéneo debido a un proceso ideopático de las vías dopaminérgicas de evolución progresiva, produciendo lesiones multifocales principalmente en los ganglios basales a nivel de la sustancia negra pars compacta y del locus coeruleus [4, 26], causando la dificultad en la ejecución de movimientos y la rigidez, produciendo alteraciones en la marcha caracterizada por un patrón de marcha hipocinetica-rígida [27].

Existe una pérdida de las sinergias entre la postura-dinámica que da lugar a la flexión del tronco y las articulaciones, disminución del braceo, de la altura y longitud del paso, reducción de la velocidad y la cadencia que puede permanecer inalterada o incluso aumentada en casos como resultado de un mecanismo compensatorio, caracterizando la marcha [4, 28].

El peak de incidencia está en los sujetos 80-84 años, siendo una afección propia de los sujetos pertenecientes a la tercera edad, la literatura relata el inicio de la enfermedad es a los 50-60 años. La incidencia y prevalencia descrita ha ido en aumento en la última década, correspondiendo actualmente a 100 por millón detallando que es mayoritariamente en la población de la tercera edad. La EP es una enfermedad de acuerdo con el estado actual de los conocimientos relatados está subdiagnosticada [2].

2.0.2. *Enfermedad de Huntingtons (EH)*

Siendo una enfermedad genotípica dominante, clínicamente se caracteriza por movimientos anormales deterioros cognitivos y disturbios psiquiátricos, los desórdenes motores son heterogéneos e hiperkinesias severas (chorea, tic y distonia) que puede coexistir con transtornos de disminución del movimiento (acinesia y bracinesia) [29]. Andrzejewski describe que los pacientes con EH disminuyen notablemente su cadencia y su velocidad de la marcha [30]. Es difícil determinar su incidencia y prevalencia debido a que es hereditaria y corresponde mayoritariamente a una población específica, mostrando sus primeros síntomas alrededor de los 30 años y demostrando el peak de esta enfermedad a los 40-59 y el promedio de supervivencia de los pacientes que padecen EH son 20 años. La literatura describe que ha aumentado considerablemente de manera inespecífica [2], hace dos décadas una prevalencia de uno por millón y resaltando la dificultad del seguimiento poblacional y variante en el mundo, buscando la prevalencia por

mortalidad en hospitales e instituciones mentales [3], la principal causa de muerte en los pacientes con EH son neumonías y problemas cardíacos, describiendo que los criterios son estrictamente no comparables, debido a estas variables y limitaciones es difícil demostrar hoy en día la prevalencia de estos sujetos, aún así la literatura ha descrito que los sujetos con EH aumenta su prevalencia al doble por cada 150 años, 6 generaciones aproximadamente, alcanzando sólo en algunas poblaciones del Norte de América a 4,09 por millón, aún existiendo controversia con otros estudios en distintas poblaciones [2].

2.0.3. *Esclerosis Lateral Amiotrófica (ELA)*

Es una afección neurodegenerativa que afecta las células nerviosas del cerebro y la zona anterior del cordón medular produciendo un daño a las motoneuronas. La literatura describe esta condición como multifactorial. Hadzipasic refiere disminución de la amplitud de la señal en la electromiografía además de la pérdida de sinergia muscular en la marcha entre el tibial anterior y gastrocnemio [31]. Hausdorff además describe en esta afección la existencia de una menor velocidad y cadencia de la marcha, agregando una disminución en el largo del paso [23]. Su frecuencia es relativa en el mundo, en Japón alcanza 0,31 personas/año y en Suecia 100000 personas por año, la literatura detalla en Buenos Aires Argentina en el Hospital Ramos Mejias muestra una incidencia anual de 6/1000000 al año [32]. Siendo que lo descrito publicada en Taiwán muestran una incidencia muy relativa entre 868 por 100000 al año [33].

La bibliografía describe que el padecimiento de esta enfermedad está en la edad adulta, dándose el peak a los 75-79 años y una sobrevida de 2,7 años promedio [2].

2.1. *Clustering*

Son algoritmos matemáticos diseñados en solucionar problemas de clasificación, los cuales pueden ser utilizados en distintos componentes o problemas de interés [34]. Son herramientas usadas en la minería de datos (también conocida como knowledge discovery in databases) para su análisis.

2.1.1. *Análisis de Redes Neuronales Artificiales (ARNA)*

Es un método inspirado en el funcionamiento de redes neuronales biológicas en las que un número de elementos simples de cómputo no lineal se interconectan, a través de ecuaciones matemáticas, aceptados patrones de inputs y entregando patrones de outputs calculando así una clasificación a través de patrones numéricos, tomando decisiones que con el tiempo se pueden transformar en automatizadas y no supervisadas. Las ARNA son de utilidad en reconocimiento de patrones y modelado y filtrado no lineal de señales. Estas redes presentan algún tipo de plasticidad por lo cual hoy en día son muy utilizadas en la ciencia donde el problema puede ser impreciso o variante en el tiempo.

En la literatura las ARNA han sido utilizadas ampliamente, Llajaruna clasificó el proceso de equimosis para el peritaje médico legal, disminuyendo de un 80 % a un 1,2 % el error en las imágenes de muestreo [35]. Nimmrich las utilizó como herramienta para tratar de clasificar las ondas cerebrales (delta, theta, alpha y beta) muestreadas a través de electroencefalograma de enfermedades neurodegenerativas, como diferentes tipos de demencia, Parkinson y Huntingtons [36].

En la marcha, la ciencia relatada describe que se ha usado las ARNA para clasificar los patrones de marcha normal, Pourhedayat clasificó a través de un modelo simulando los ganglios basales por un modelo matemático en los sujetos con

Huntington [37], adicionalmente Leporace analizó la fuerza de reacción del suelo a través de 17 sujetos sanos y una simulación con redes neuronales obteniendo resultados similares a otros estudios con errores altos [38]. La clasificación en los cambios de los patrones de marcha por la edad fueron realizados por Begg y Kamruzzaman en Australia, dando buenos resultados y un aprendizaje de un 83 % en la clasificación [39].

Los estudios hechos por Yoo demostraron que entregando los modelos de los patrones de la marcha humana ya descritos se logra un reconocimiento de los movimientos normales logrando una automatización a través de las ARNA [40].

2.1.2. Máquinas de Soporte Vectorial (MSV)

Son algoritmos de aprendizaje basados en la creación de un hiperplano, en primera instancia fueron utilizados para resolver problemas binarios, pero al poco tiempo de ser utilizados comenzaron a ser usados en tareas de mayor complejidad como multclasificación, agrupamiento y regresión. Las MSV deben su éxito no a disminuir el error como lo hacen o se espera en la mayoría de los clasificadores, sino que buscan modelos que tengan pocos errores en el futuro, es decir, aumenten su aprendizaje.

En la diecisieteava conferencia de machine learning Klinkenberg and Joachims presentaron un método conocido como «Support Vector Machines» [41], utilizado recientemente por Ahmed para crear una nueva clasificación en los sujetos recién nacidos que presentan Encefalopatía Hipóxica Isquémica, obteniendo una curva de aprendizaje del 87 %, concluyendo con éxito su investigación [42]. Dutta el año 2010 clasificó correctamente las Arritmias Cardíacas muestreadas con Electrocardiograma, diferenciándola en tres grupos y obteniendo un rango de alta clasificación entre el 95,51-96,12 % [43].

La bibliografía científica relatada por Zheng a través de varios estudios de machine learning, describen que el análisis de la marcha y su clasificación a través de máquinas de soporte vectorial y otros métodos utilizados en la minería de datos en las enfermedades neurodegenerativas [44], pueden ser clasificados invitando a la población mundial de científicos del campo a utilizar estas herramientas de cuantificación y clasificación.

2.2. Características del Aprendizaje Automatizado

2.2.1. *Matriz de Confusión*

La matriz de confusión es una tabla (ver tabla 2.1), que muestra la predicción y el error de los datos muestreados pudiendo predecir el error estadístico tras haber hecho un cálculo de probabilidades sobre el límite del error aceptado [45], es decir, sobre esta tabla están todos los casos y posibles resultados de las pruebas realizadas. El análisis sensitivo de la muestra de datos marca los resultados como 1 siendo positivo y 0 siendo negativo.

Es una tabla que muestra especificidad, sensibilidad y los errores conocidos, mediante las instancias que dieron casos positivos siendo verdaderos y los casos falsos siendo realmente negativos, los resultados no correspondientes estos datos reales son marcados como error, conocidos como error tipo 1 y tipo 2, marcados como casos positivos siendo verdaderamente negativos y los negativos siendo realmente positivos, existiendo entre estos la menor probabilidad posible. Buscando un resultado de un 80 % o superior.

2.2.2. Receiver Operating Characteristic

El análisis mediante Receiver Operating Characteristic o de Curvas ROC se inició alrededor de la segunda guerra mundial, gracias al problema que se planteaba

con las señales recogidas por el radar, ya que no discriminaban si eran señales de aviones o ruidos. Es una herramienta para estimar la validez de una prueba diagnóstica y permite evaluar la precisión y clasificar categorías, facilitando la toma de decisiones; la validación de estas pruebas es a través de una matriz de confusión.

Curva ROC que representa las curvas Especificidad y Sensibilidad.
graficadas en la función.

$p(x) > 0,5$ decimos que $y = 1$.

puede resultar insuficiente en este caso, pues bastantes individuos podrían ser clasificados como positivo siendo negativo. Se llama sensibilidad a la curva.

$(t) = P(p(x) > t | y=1), 0 \leq t \leq 1$.

Variando t ; la curva Se va dando la proporción de individuos a los que se detecta caso negativo.

Para $t = 0$ todos los individuos resultarían negativos, y para $t = 1$ todos resultarían positivos.

Se llama especificidad a la curva

$Es(t) = P(p(x) < t | y = 0), 0 \leq t \leq 1$.

Variando t ; la curva.

Para $t = 0$ todos los individuos resultarían positivos, y para $t = 1$ todos casos negativos.

La curva ROC (Receiving Operating Characteristic) resume las dos curvas de sensibilidad y especificidad.

- El área bajo la curva ROC es siempre mayor o igual que 0,5. Un valor a partir de 0,8 se considera como que la discriminación es buena un valor a partir de 0,9 se consideraría como muy bueno. La discriminación sería perfecta si el área vale 1 [46].

	Casos a predecir	
caso real	SI	NO
SI	Verdaderos positivos	Falsos negativos
NO	Falsos positivos	Verdaderos negativos

Cuadro 2.1: Matriz de confusión

En la tabla 2.1 muestra los casos explicados, los casos predichos y los casos reales. Explicando los casos que fueron predichos y coincidieron con casos reales, y los casos que no fueron predichos.

Parte III

Marco Experimental

Capítulo 3

Materiales y Metodología

3.1. Diseño de Estudio

El diseño de estudio corresponde a una investigación cuantitativa, no experimental, transversal descriptivo, no probabilístico por conveniencia, enmarcado en el contexto de una investigación desarrollada por investigadores de la Clínica Neurológica del Hospital General de Massachussetts y revelado a la comunidad científica por el banco de bases de datos web physionet.org con la autorización de los investigadores. Explicado en el algoritmo 3.1 definiendo el proceso a utilizar en este estudio.

Este proyecto se enmarca en el desafío impuesto por el sitio web physionet.org, página de importancia universal que revela problemáticas científicas de distinta índole a la comunidad de investigadores. En este caso se utilizó la base de datos referente a determinar la capacidad de identificación de patrones de marcha neurodegenerativa.

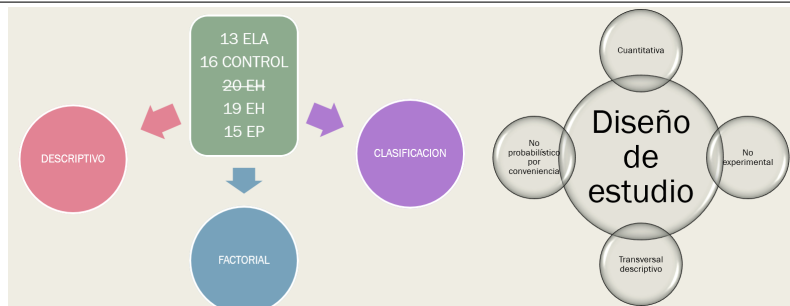
Algoritmo 3.1 Diagrama de procesos.

Figura la cual representa el diseño de estudio y sus diferentes visiones. La figura izquierda señala la eliminación de un sujeto con EH con valores cuestionables para el estudio.

3.1.1. Descriptivo

Este estudio cuenta con 16 casos control, 19 casos con Enfermedad de Huntington, 15 con Enfermedad de Parkinson y 13 pacientes con Esclerosis Lateral Amiotrófica. Se realizó un análisis descriptivo, de la cual se utilizaron 13 variables temporo espaciales.

A continuación se enumeran dichas variables:

- Intervalo de zancada izquierdo en segundos.
- Intervalo de zancada derecha en segundos.
- Intervalo de balanceo izquierdo en segundos y expresado en términos de porcentaje.
- Intervalo de balanceo derecho en segundos y expresado en términos de porcentaje.
- Intervalo apoyo izquierdo en segundos y expresado en términos de porcentaje.

- Intervalo de apoyo derecho en segundos y expresado en términos de porcentaje.
- Intervalo de doble apoyo en segundos y expresado en términos de porcentaje.
- Velocidad de la marcha en metros por segundo. En esta última variable se tomó el tiempo desde los 20 segundos iniciales de marcha.

Adicionalmente se efectuó análisis descriptivo de los distintos tipos de marcha con el propósito de identificar las variables que mejor expresaran la variabilidad de cada una de estas.

Para el análisis descriptivo se utilizó el programa Stata® y Excel® en el que se tomaron las variables de edad, peso, altura, severidad de la patología con escalas diferentes.

3.2. Clasificación

Previo a la clasificación se agruparon los grupos de estudio según el tipo de patologías con el propósito de evaluar la capacidad de los métodos de clasificación entre los distintos componentes asociados a cada tipo de marcha. Para esta tarea se utilizó un Perceptrón Multicapa Retroalimentado, además de Máquinas de Soporte Vectorial (MSV) para cada experimento, generándose 12 experimentos a evaluar.

Los algoritmos de clasificación fueron implementados en el software Weka 3.9®

3.2.1. Experimentos Evaluados

Se generaron distintos experimentos para nutrir el análisis de los datos y la capacidad de distinguir los distintos tipos de marcha a clasificar.

1. Discriminación entre las patologías neurodegenerativas (sin grupo control).
2. Discriminación entre grupo control v/s ELA.
3. Discriminación entre grupo control v/s EP.
4. Discriminación entre grupo control v/s EH.
5. Discriminación entre grupo EH v/s ELA.
6. Discriminación entre grupo ELA v/s EP.
7. Discriminación entre grupo EP v/s EH.
8. Se agrega la variable severidad dentro de todos los grupos.
9. Se agrega la variable velocidad sin considerar la variable severidad dentro de todos los grupos.
10. Comparación de las tres patologías v/s el grupo control.
11. Comparación de las tres patologías v/s el grupo control considerando todas las variables.
12. Comparación del comportamiento de enfermedades asociadas a los Ganglios Basales v/s ELA.

Clases	Porcentaje
Control	25 %
EH	31 %
ELA	20 %
EP	24 %

Cuadro 3.1: Porcentaje Clases

El porcentaje de afección por clase en la muestra detallando que un 25 % corresponde al grupo control, un 20 % a los sujetos con ELA, un 24 % a sujetos con Parkinson y un 31 % sujetos con EH.

3.3. Materiales y Métodos

3.3.1. Sujetos.

Los sujetos con ELA (25 %), EP (24 %) y EH (31 %), son usuarios ambulatorios de la Clínica Neurológica del Hospital General de Massachussetts, a los cuales se les solicitó caminar durante 5 min; se contabilizó y se registró el tiempo a partir de los 20 segundos en adelante, se colocó un sensor de resistencia en la planta del pie detectando la caída del pie y el contacto con el suelo determinando las variables relatadas.

En la prueba de normal de Shapiro-Wilk aplicada a las variables descriptivas se demostró que las variables normales son el peso y la altura respectivamente en esta muestra, la edad se comportó como variable no normal con un valor Z de 0,03849 demostrado en la tabla 3.1, teniendo en cuenta que en la muestra 36 sujetos son mujeres, lo que corresponde a un 56,25 % de la muestra y solo 28 sujetos son varones lo que corresponde a un 43,75 % de la muestra, en el gráfico de la figura 3.2 se logra observar la división por género en los cuatro tipos de sujetos. El promedio de la velocidad de la marcha fue de 1,5 m/s con un SD de 0,2807.

Clases/Género	Hombre	Mujer
Control	7 %	39 %
EH	21 %	39 %
ELA	7 %	8 %
EP	36 %	14 %

Cuadro 3.2: División por Género

El gráfico anterior relata una división por género entre clases, mostrando que de manera equivalente un 39 % en mujeres con Huntington y grupo control, y sólo un 14 %, en comparación con los varones que son el grupo mayoritario junto a los sujetos con ELA ambos de un 36 %, y el grupo control solo un 7 %.

Clase/IMC	Bajo Peso	NormoPeso	SobrePeso	Obesidad
Control	31 %	63 %	6 %	31 %
EH	40 %	40 %	15 %	40 %
ELA	8 %	38 %	31 %	8 %
EP	7 %	93 %	0 %	0 %

Cuadro 3.3: Peso por Clase

La división de sujetos por rango de peso establecido a través del índice de masa corporal y la guía de elige vivir sano, influyendo esto en la marcha humana y produciendo un error en la clasificación de los patrones característicos. En los sujetos con HD un 80 % corresponde a sujetos con bajo o normo peso a diferencia del 54 % de los sujetos con ELA representan a un grupo con obesidad o sobrepeso.

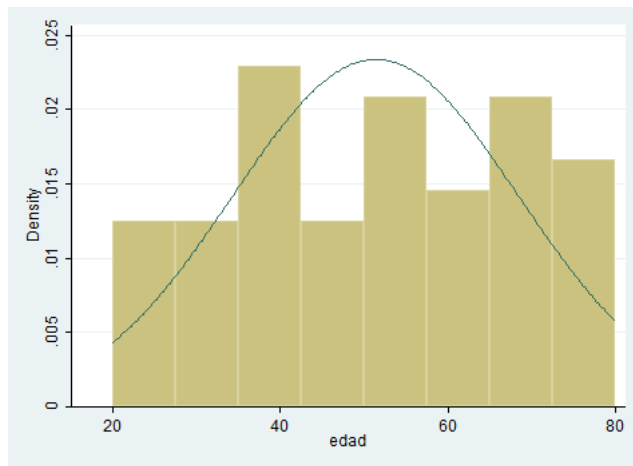


Figura 3.1: Histograma Edad

La figura 3.4 muestra un histograma de la Edad de la muestra la cual se visualiza como un comportamiento no normal, causado por las diferencias etarias entre patologías.

Normalidad	
Edad	0,03849
Altura	0,93788
Peso	0,39474
Velocidad	0,70535

Cuadro 3.4: Resultados de Normalidad

En la tabla anterior se ve la prueba de normalidad de Shapiro-Wilk, demostrando que la edad no es una variable normal. La variable de la edad se demuestra como no normal debido a que la afecciones de distintas patologías ocurre en distintos rangos etarios.

Variable	P50	P25	P75	RIQ
Edad	52.5	38.5	66	27.5

Cuadro 3.5: Tabla de Distribución Muestral de la Edad

La tabla demuestra que el rango intercuartílico es de 27.5 que va desde 52.5 hasta los 66 años de edad.

En la tabla 3.1 se demuestra que la manera correcta para relatar las variables de la edad que pueden ser influyentes en la marcha y los modelos de clasificación no se comporta de manera normal por lo que se establece una tabla de descripción junto a los percentiles y el rango intercuartílico.

Para la recopilación de los datos en el sitio web physionet.org, consideró 16 sujetos control, 19 HD, 15 PD, 13 ELA.

Las 13 variables temporales verificadas fueron:

- Intervalo de zancada izquierda en segundos.
- Intervalo de zancada derecha en segundos.
- Intervalo de balanceo izquierdo en segundos.
- Intervalo de balanceo izquierdo en porcentaje.
- Intervalo de balanceo derecho en segundos.
- Intervalo de balanceo derecho en porcentaje.

- Intervalo apoyo izquierdo en segundos.
- Intervalo apoyo izquierdo en porcentaje.
- Intervalo de apoyo derecho en segundos.
- Intervalo de apoyo derecho en porcentaje.
- Intervalo de doble apoyo en segundos.
- Intervalo de doble apoyo en porcentaje.
- Velocidad de la marcha.

3.3.2. Análisis Estadístico

El análisis estadístico de carácter descriptivo de la muestra fue utilizado incluyendo gráficos y análisis de centralización y dispersión separados por clase, no se consideró utilizar un análisis estadístico inferencial.

3.3.3. Clustering

La utilización de los clasificadores fue estandarizada utilizando el programa Weka 3.9[®] y con un grado de entrenamiento del setenta por ciento, de esta manera fue clasificado cada uno de los grupos ya relatados. los outcomes fueron definidos por la clase como variable de salida y los atributos seleccionados ya relatados como variables temporoespaciales junto a las variables de velocidad de la marcha y severidad del estado de los pacientes utilizados por escalas diferentes seleccionadas por patología.

La clasificación de los datos fue a través de un perceptrón multicapa (multi-layer perceptron) tipo a que cuenta con la primera capa que variaba según la cantidad de salidas y entradas utilizando los valores por defecto solo exigiendo el porcentaje de entrenamiento. Por contraparte las MSV denominadas en el programa SOM y un nivel de entrenamiento del 70 % de la misma forma.

Parte IV

Discusión de Resultados

Capítulo 4

Resultados

En los resultados descriptivos obtenidos se visualizó una diferencia en la carga de peso de los sujetos controles, respecto a los sujetos con patologías neurodegenerativas, mostrado en las imágenes 4.1; 4.2; 4.3. Además se puede visualizar una distribución más amplia en el comportamiento patologías neurodegenerativas.

Para las pruebas de normalidad de las trece variables definidas como variables continuas, sólo la velocidad se comportó como normal, con un valor $z=0,66982$ en la prueba de Shapiro-Wilk.

Adicionalmente se demostró que las variables temporales de la marcha por si sola son suficientes para diferenciar los sujetos del grupo control respecto de las patologías a través de los métodos de clasificación, generando como resultado una curva ROC de 0,971. Pero al contrastar o intentar diferenciar los distintos tipos de marcha neurodegenerativas las curvas de aprendizaje son bajas, llegando a un máximo en la curva ROC de 0,6. Sin embargo, cabe destacar que entre las patologías existe una excepción en el grupo con ELA, en el cual se logra alcanzar un valor 0,8 en la curva ROC.

Por otra parte, podemos señalar que las patologías neurodegenerativas asociadas

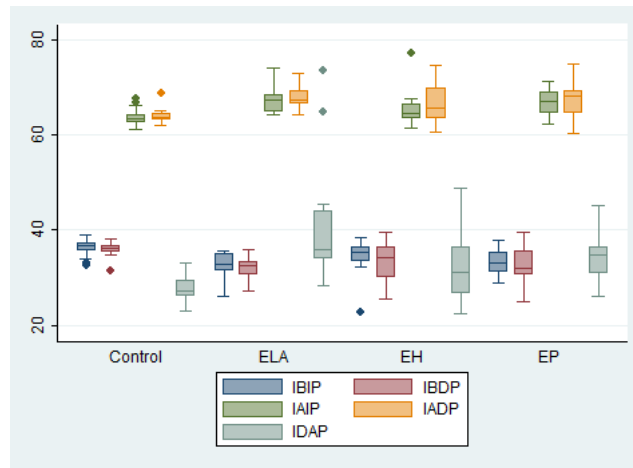


Figura 4.1: Distribución de la Marcha

En la figura 4.1 se observan 5 variables las cuales corresponden todas a porcentajes, IBIP (Intervalo Balanceo Izquierdo Porcentaje), IBDP (Intervalo Balanceo Derecho Porcentaje), IAIP (Intervalo Balanceo Izquierdo Porcentaje), IADP (Intervalo Apoyo Derecho Porcentaje), IDAP (Intervalo Doble Apoyo Porcentaje).

a los Ganglios de la base llegan a un máximo de 0,625, por lo que se buscaron distintas estrategias para generar mejorías en la discriminación con el propósito de mejorar el resultado final. Entre ellas, el experimento de clasificación entre patologías asociadas a los Ganglios de la base respecto a sujetos con ELA llegando a una curva de aprendizaje de 0,65 con redes neuronales y 0,5 con MSV.

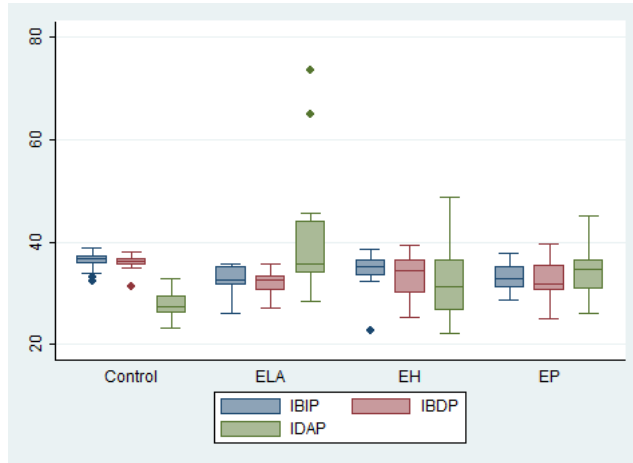


Figura 4.2: Balanceo y Doble Apoyo

En la figura 4.2 se observa la gráfica de cajas de la fase de balanceo y doble apoyo, logrando observar que la distribución de la carga de peso es mayor en los sujetos con HD además de la diferencia asimétrica temporal entre ambos hemisferios.

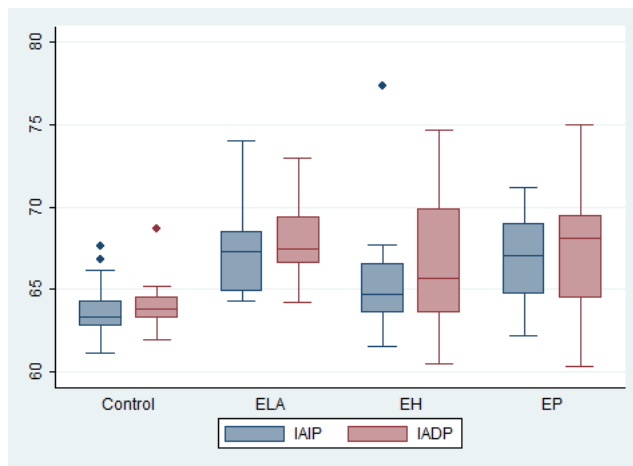


Figura 4.3: Intervalos de Apoyo Derecho e Izquierdo

En la distribución que se observa en la figura 4.3 del gráfico de cajas se observa un mayor predominio en el apoyo derecho en las enfermedades influyentes de los ganglios basales.

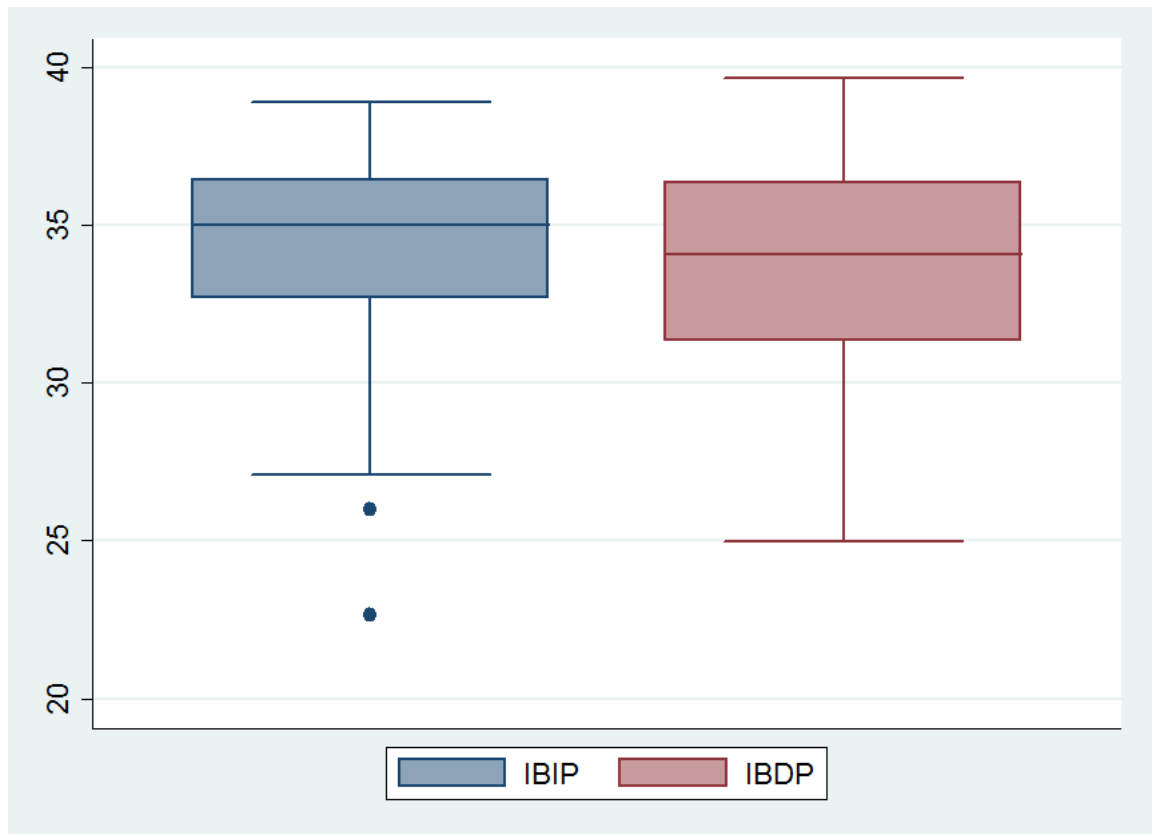


Figura 4.4: Fases de Balanceo Derecho e Izquierdo

En la figura 4.4 no se observan grandes diferencias en la distribución global de los sujetos en la fase de balanceo de ambos hemicuerpos.

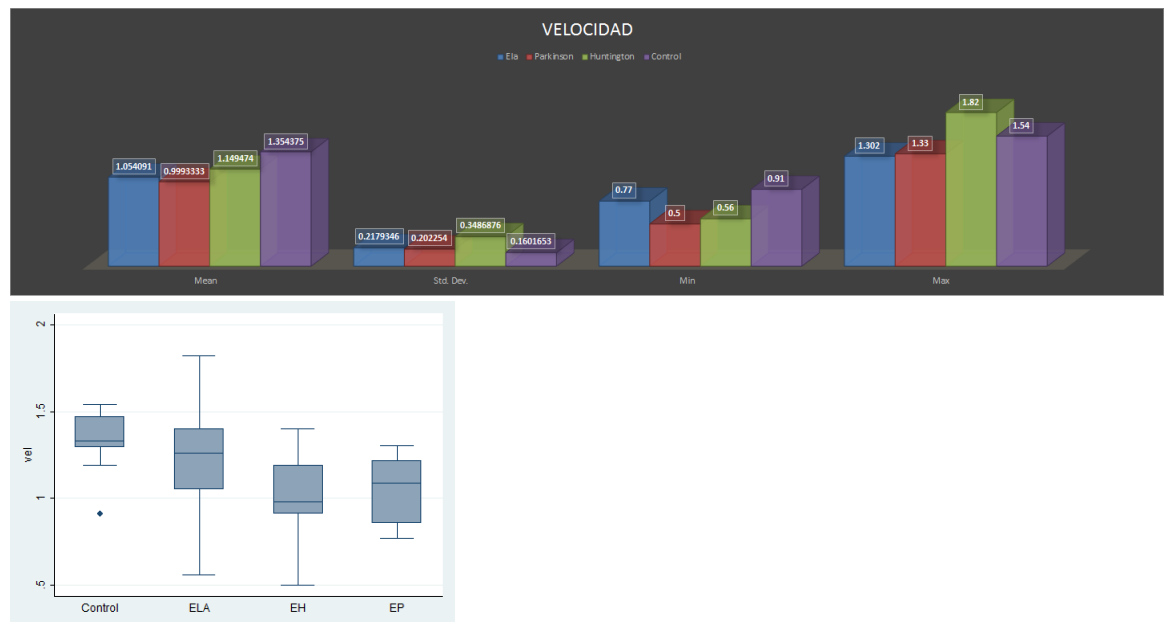


Figura 4.5: Velocidad de la Marcha

En la velocidad de la marcha clasificada se observa una distribución mucho menor en los sujetos control, la mayor distribución en la velocidad de la marcha está en los sujetos con ELA, mientras que en los sujetos con Parkinson tienen una mayor distribución comparada con la normal, pero más centralizada a diferencia de HD y ELA

ARNA PATOLOGIAS SIN GRUPO CONTROL				
Correctly Classified Instances				4 28,57%
Incorrectly Classified Instances				10 71,43%
ROC AREA	ELA	HD	PD	CLASE
0,600	1	3	0	0 ELA
0,455	1	2	0	0 HD
0,469	1	5	1	1 PD

MSV PATOLOGIAS SIN GRUPO CONTROL				
Correctly Classified Instances				3 21,43%
Incorrectly Classified Instances				11 78,57%
ROC AREA	PD	HD	ELA	CLASE
0,575	0	4	0	0 ELA
0,500	0	3	0	0 HD
0,500	0	7	0	0 PD

ARNA PATOLOGIAS JUNTO GRUPO CONTROL				
Correctly Classified Instances				38 59,38%
Incorrectly Classified Instances				26 40,63%
ROC AREA	CONTROL ELA	HD	PD	CLASE
0,971	15	0	0	1 CONTROL
0,887	0	9	3	1 ELA
0,625	3	2	9	6 HD
0,589	0	3	7	5 PD

MSV PATOLOGIAS JUNTO GRUPO CONTROL				
Correctly Classified Instances				27 42,19%
Incorrectly Classified Instances				37 57,81%
ROC AREA	CONTROL ELA	HD	PD	CLASE
0,766	5	0	11	0 CONTROL
0,740	0	6	7	0 ELA
0,495	4	2	13	1 HD
0,709	1	0	11	3 PD

Cuadro 4.1: Patologías vs Grupo Control

Muestra un experimento realizado en weka® el cual trata de clasificar dos grupos (patologías sin un grupo control y patologías con un grupo control) a través de dos métodos diferentes, los cuales dio como resultado que existe una gran influencia en el grupo control para los métodos algorítmicos de aprendizaje.

ARNA CONTROL HD		
Correctly Classified Instances	8	72,73%
Incorrectly Classified Instances	3	27,27%
ROC AREA	CONTROL HD	CLASE
0,708	1	2 CONTROL
0,708	1	7 HD

ARNA CONTROL PD		
Correctly Classified Instances	6	66,67%
Incorrectly Classified Instances	3	33,33%
ROC AREA	CONTROL PD	CLASE
0,950	4	0 CONTROL
0,950	3	2 PD

ARNA ELA CONTROL		
Correctly Classified Instances	7	77,78%
Incorrectly Classified Instances	2	22,22%
ROC AREA	CONTROL ELA	CLASE
0,900	4	0 CONTROL
0,900	2	3 ELA

ARNA GB ELA			
Correctly Classified Instances	10	71,43%	
Incorrectly Classified Instances	4	28,57%	
ROC AREA	ELA	GB	CLASE
0,650	1	3	3 ELA
0,650	1	9	9 GB

MSV CONTROL HD		
Correctly Classified Instances	9	81,82%
Incorrectly Classified Instances	2	18,18%
ROC AREA	CONTROL HD	CLASE
0,771	2	1 CONTROL
0,771	1	7 HD

MSV CONTROL PD		
Correctly Classified Instances	7	77,78%
Incorrectly Classified Instances	2	22,22%
ROC AREA	CONTROL PD	CLASE
0,800	4	0 CONTROL
0,800	2	3 PD

MSV ELA CONTROL		
Correctly Classified Instances	6	66,67%
Incorrectly Classified Instances	3	33,33%
ROC AREA	CONTROL ELA	CLASE
0,675	3	1 CONTROL
0,675	2	3 ELA

MSV GB ELA			
Correctly Classified Instances	10	71,43%	
Incorrectly Classified Instances	4	28,57%	
ROC AREA	ELA	GB	CLASE
0,500	0	4	4 ELA
0,500	0	10	10 GB

Cuadro 4.2: Patologías vs Control y Ganglios Basales vs ELA

Es un experimento de redes neuronales y soporte vectorial, el cual muestra los resultados de las patologías y grupos control, adicionalmente un grupo en el cual se separó en ganglios basales y ELA, dando así como resultado observar que las curvas ROC son suficientes para la clasificación entre una afección y un sujeto control, pero no logra determinar un patrón de marcha en la curva de aprendizaje entre enfermedades inclusive de diferente clase.

ARNA HD ELA				MSV HD ELA			
Correctly Classified Instances	7	70%		Correctly Classified Instances	7	70%	
Incorrectly Classified Instances	3	30%		Incorrectly Classified Instances	3	30%	
ROC AREA	ELA	HD	CLASE	ROC AREA	ELA	HD	CLASE
0,958		4	0 ELA	0,625		1	3 ELA
0,958		3	3 HD	0,625		0	6 HD

ANN PD HD				MSV PD HD			
Correctly Classified Instances	8	80%		Correctly Classified Instances	6	60%	
Incorrectly Classified Instances	2	20%		Incorrectly Classified Instances	4	40%	
ROC AREA	HD	PD	CLASE	ROC AREA	HD	PD	CLASE
0,708		6	0 HD	0,500		6	0 HD
0,708		2	2 PD	0,500		4	0 PD

ARNA PD ELA				MSV PD ELA			
Correctly Classified Instances	5	62,50%		Correctly Classified Instances	6	75%	
Incorrectly Classified Instances	3	37,50%		Incorrectly Classified Instances	2	25%	
ROC AREA	ELA	PD	CLASE	ROC AREA	ELA	PD	CLASE
0,867		2	1 ELA	0,733		2	1 ELA
0,867		2	3 PD	0,733		1	4 PD

Cuadro 4.3: Comparación entre Patologías

Muestra una comparación entre patologías las cuales demuestran específicamente entre ellas que algunas sí son suficientes para la clasificación. (el verdadero cuestionamiento de esto es que la cantidad muestral es demasiado baja, produciendo un gran sesgo en el entrenamiento)

ARNA SEVERIDAD					MSV SEVERIDAD						
Correctly Classified Instances	9 47,37%				Correctly Classified Instances	9 47,37%					
Incorrectly Classified Instances	10 52,63%				Incorrectly Classified Instances	10 52,63%					
ROC AREA	CONTROL	ELA	HD	PD	CLASE	ROC AREA	CONTROL	ELA	HD	PD	CLASE
0,943	4	0	1	0	CONTROL	0,871	3	0	2	0	CONTROL
0,617	0	1	1	2	ELA	0,458	0	1	3	0	ELA
0,590	0	1	3	2	HD	0,538	0	1	4	1	HD
0,633	0	1	2	1	PD	0,633	0	1	2	1	PD

ARNA MSV					MSV VELOCIDAD						
Correctly Classified Instances	6 31,58%				Correctly Classified Instances	6 31,58%					
Incorrectly Classified Instances	13 68,42%				Incorrectly Classified Instances	13 68,42%					
ROC AREA	CONTROL	ELA	HD	PD	CLASE	ROC AREA	CONTROL	ELA	HD	PD	CLASE
0,829	3	1	0	1	CONTROL	0,843	3	0	2	0	CONTROL
0,500	2	1	1	0	ELA	0,575	1	1	2	0	ELA
0,667	0	1	0	5	HD	0,603	0	1	4	1	HD
0,650	0	2	0	2	PD	0,633	1	1	1	1	PD

ARNA SEVERIDAD CON VELOCIDAD					SMO SEVERIDAD CON VELOCIDAD						
Correctly Classified Instances	37 57,81%				Correctly Classified Instances	35 54,69%					
Incorrectly Classified Instances	27 42,19%				Incorrectly Classified Instances	29 45,31%					
ROC AREA	COTROL	ELA	HD	PD	CLASE	ROC AREA	CONTROL	ELA	HD	PD	CLASE
0,978	14	0	1	1	CONTROL	0,898	13	0	2	1	CONTROL
0,875	0	8	3	2	ELA	0,740	0	6	6	1	ELA
0,672	2	1	9	8	HD	0,591	4	2	12	2	HD
0,574	0	3	6	6	PD	0,650	2	0	9	4	PD

Cuadro 4.4: Severidad y Velocidad

Las últimas tablas incluyen las variables de severidad y velocidad dando un buen pronóstico y gran clasificación con una curva ROC cercanas o superiores al 70 %, se observa que los cambios incluyendo ambas variables son muy influyentes en el patrón, mejorando la clasificación.

Capítulo 5

Discusión

Uno de los posibles cuestionamientos que se realizan en este estudio, es posiblemente la poca variabilidad de modelos que se utiliza para la clasificación, debido a que la literatura relata que existen múltiples tipos de clasificadores y subclases de éstos [34], además de agregar que no se correlaciona con las publicaciones referentes [44, 47], las cuales alcanzaron un bajo error y un alto poder de clasificación con redes neuronales y soporte vectorial, logrando establecer patrones de marcha en enfermedades neurodegenerativa y una fuerte correlación.

A diferencia de Lascano quien publicó su estudio y una gran predicción entre sujetos con Parkinson y normales [12], Dutta logró implementar las redes neuronales de manera efectiva en sujetos el 2009 a través de redes neuronales del mismo tipo que este estudio teniendo una excelente clasificación [47], adicionalmente los resultados publicados Zheng el mismo año con la misma base de datos relata que los métodos de machine learning son efectivos para la correcta clasificación entre sujetos con enfermedades neurodegenerativas [44].

Este estudio a pesar de haber logrado describir las variables temporo espaciales de la marcha y la variación en los sujetos pertenecientes a la base de datos del

Hospital General Clínico Neurológico de Massachussetts a través de los gráficos de distribución, las curvas roc y el aprendizaje realizado por el software weka 3.9®. en la totalidad de los experimentos (exceptuando en el caso de los sujetos con EH vs control) siempre tuvo un área roc mayor, las RNA por sobre MSV, vease las tablas 4.1; 4.2; 4.3;4.4.

En las tablas mostradas en este estudio se logra apreciar que el área roc mostrada en la figura 4.1 entre sujetos sin grupo control en las RNA alcanza un valor de 0,6 siendo insuficiente para la clasificación mientras que lo demostrado cuando se le agrega los sujetos control llega a un valor de 0,97, mientras que los valores alcanzados con MSV son 0,575 sin sujetos control y 0,766 mostrando un área de predicción baja además de ser muy inferior a las demostradas por los otros algoritmos de aprendizaje.

Los datos relatados en el experimento de los sujetos control vs las patologías cada uno por sí solo, demostró una predominancia en las RNA como sistema que demostró un aprendizaje mayor relatando un área roc 0,95 sobre 0,8 en los sujetos con Parkinson vs los control. En este mismo experimento no se obtuvo una adecuada clasificación entre los ganglios basales y ELA debido que se alcanza 0,65 en RNA y 0,5 en MSV, por contraparte los clasificados con EH vs el grupo control obtuvo un 0,7 en RNA y en MSV un 0,77 mejorando solo la clasificación en este punto, cabe destacar que en este aspecto al dividir la base de datos disminuyó considerablemente la cantidad de sujetos, lo que pudo haber afectado a al aprendizaje automatizado. Por lo tanto cuando se amplió la base de datos entregándose nuevas variables e incluyendo los sujetos, mejoró considerablemente su aprendizaje (vease en la tabla 4.4), relatando que las variables de severidad y velocidad pudiesen alertar a la automatización entregándoles valores que serian capaces de clasificarlas, pudiendo sobrevalorar los resultados.

Capítulo 6

Conclusión

No fue capaz de discriminar patrones de marcha neurodegenerativa a través de los procesos de machine learning, mientras que se logró determinar una clasificación entre sujetos sanos y con enfermedades neurodegenerativas con las técnicas de clustering utilizadas, refiriendo que es de mejor elección las RNA sobre MSV bajo las condiciones del software, también se demostró que las variables de severidad y velocidad entregan información de gran influencia para la clasificación, Mientras que el análisis descriptivo ratificó la disminución de la longitud del paso en los sujetos con Parkinson y Huntington que puede correlacionado a lo publicado anteriormente por Curtze, debido a factores como la disminución del balance y la percepción de las personas [48]. Se realizó la revisión de la metodología y los distintos tipos de análisis, además de encontrar visiones diferentes con otros autores[44], no se logró identificar patrones de marcha suficientes a través de métodos algorítmicos y está abierto el debate de los métodos computacionales, que como kinesiólogos no especialistas en el área tecnológica podemos hacer un aporte desde el punto de vista de la salud, se logró establecer una clasificación en la marcha normal, pero sigue siendo insuficiente en la marcha patológica.

Se verificó y analizó las variables y demostró su importancia en los distintos métodos de clasificación, objetando los métodos anteriores relatados en la literatura, además del estudio de las bases de datos, concluyendo que no representa una base de datos lo suficientemente grande, el método de ser clasificación de datos masivos parece ser insuficiente para esta población.

Los datos de los estudios previos pueden estar sobrevalorados, incluyendo variables a los clasificadores que se les pudo haber indicado un diagnóstico previo antes de clasificar, dando como resultado valores sumamente esperanzadores que pudiesen estar alterados, adicionalmente la gran cantidad de limitaciones que subyace en este estudio, poca cantidad de sujetos a utilizar, recordando que estas herramientas son utilizadas en base de datos denominadas bigdata con una alta cantidad de sujetos a estudiar, a diferencia de este estudio que los clasificó fue una baja cantidad para el entrenamiento de los algoritmos de aprendizaje automatizado. la falta de información en el momento de recopilación de datos como los criterios de inclusión y exclusión o la ignorancia en parametros de las variables definidas, la deficiencia en el soporte Weka 3.9® de no poder programar los parametros y crear redes neuronales propias.

Bibliografía

- [1] World health organization. neurological disorders public health challenges. switzerland, 2006
- [2] De Pedro-Cuesta J, Rabano A, Martínez-Martín P, Ruiz-Tobar M, Alcalde Cabero E, Almazán-Isla, et al. Comparative incidence of conformational, Neurodegenerative Disorders. *Journal Pone*. 2015; 15.
- [3] Michael Conneally P. Huntington disease: Genetics and Epidemiology. *Am J Hum Genet*, 1984; 36, 506-5026.
- [4] Cano de la Cuerda R, Macías Jiménez A, Crespo Sánchez V, Morales Cabezas M. Escalas de valoración y tratamiento fisioterápico en la enfermedad de Parkinson. *Fisioterapia*. 2004; 26 (4): 201-210.
- [5] Hausdorff J M, Letratanakul A, Cudkovicz M E, Peterson A L, Kaliton D and Goldberger A L. Dynamic marker of altered gait rhythm in amyotrophic lateral sclerosis. *J Appl Physiol*. 2000; 88: 2045-2053.
- [6] Berttini M, Gisella M, Monachelli G, Rodríguez G, Rey R C, Martínez L, Sica R E. Epidemiology of amyotrophic lateral sclerosis Patients in a centre in Buenos Aires. *Arq Neuropsiquiatr*. 2011; 69 (6): 867-870.
- [7] Tsai C P, Wang K C, Hwang C S, Lee I T, Lee C T. Incidence, prevalence and Medical expenditures of classical amyotrophic lateral sclerosis in Taiwan, 1999-2008. *Jfma*. 2013; 114, 612-619.

- [8] Jennum P, Ibsen R, Perdersen S W, Kjellberg J. Mortality, health, social and economic consequences of amyotrophic lateral sclerosis: a controlled national study. *J Neurol.* 2012; 260: 785-793.
- [9] Nakken O, Lindstrom J C, Tysnes O B, Holmoy T. Mortality trends of amyotrophic lateral sclerosis in Norway 1951-2014: an age-period-cohort study. *J Neurol.* 2016. 263 (12): 2378-2385
- [10] Prat Pastor J M, Sanchez Lacuesta J J, Hoyos Fuentes J V, Viosca Herrero Enrique, Soler Gracia C, Comín Clavijo M, et al. *Biomecánica de la Marcha Normal y Patológica.* Valencia, Instituto de Biomecanica de Valencia; 2005.
- [11] Trojaniello D, Ceratti A, Pelosin E, Avanzino L, Mirelman A, Hausdorff, et al. Estimation of step-by step spatio-temporal parameters of normal and impaired gait using shank-mounted magneto-inertial sensors: application to elderly, hemiparetic parkinsonian and choreic gait, *Jneuroengrehab.* 2014 11: 152
- [12] Lescano C N, Herrera C V, Rodrigo S E. *Clasificación de patrones de marcha humana mediante redes neuronales.* SABI. 2011
- [13] Deschamps T, Thomas-Ollivier V, Sauvaget A, Bulteau S, Fortes-Bourbousson M, Vachon H. Balance characteristics in patient with major depression after a two-month walking exercise program: a pilot study. *Gait & posture.* 2015; 42: 590-593
- [14] Aguiar L, Santos-Rocha R, Vieira F, Branco M, Andrade C, Veloso A. Comparison Between overweight due to pregnancy due to added weight to simulate body mass distribution in pregnancy. *Gait & posture.* 2015; 42: 511-517.

- [15] Julien D, Gauvin L, Richard L, Kestens Y, Payette H. Associations between walking and depressive symptoms among older adults: do purposes and amounts of walking matter? results from the voisinage study. *menpa*. 2015; 8: 37-43.
- [16] Perry J. Gait analysis: Normal and pathological Function. New Jersey: thorofare; 1992.
- [17] Cifuentes C, Martínez F, Romero E. Análisis teórico y computacional de la marcha normal y patológica: una revisión. *Med*. 2010; 18 (2): 182-196.
- [18] Kuo A D. The six determinants of gait and the inverted pendulum analogy: a dynamic walking perspective. *human movement science*. 2007; 26 :617-656.
- [19] Winters T F, Gage J R, Hicks R. Gait patterns in spastic hemiplegia in children and young adults. *the journal bone and joint surgery*. 1987; 69A (3) : 437-441.
- [20] Gioufios G, Grieve D W. The use of neural networks to recognize patterns of human movement: gait patterns. *Clin biomech*. 1995; 10 (4): 179-183.
- [21] Pourhedayat A, Sarbaz Y. A grey box neural network model of basal ganglia for gait signal of patients with huntington disease. *JBCN*. 2016; 7 (2): 107-114.
- [22] Hausdorff J M, Cudkovicz M E, Firtion R, Wei J Y, Goldberger A. Gait variability and basal ganglia disorders: stride-to stride variations of gait cycle timing in parkinson's disease and huntington's disease. *movement disorders*. *Mov Disord*. 1998; 13 (3): 428-437.

- [23] Hausdorff J M, Lertratanakul A Cudkowicz M, Peterson A, Kaliton D, Goldberger A L. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *JapplPhysiol*. 2000; 88: 2045-2053.
- [24] Hausdorff J M, Mitchell Susan, Firtion R, Peng C K, Cudkowicz M E, Wei J Y, Goldberger A L. Altered fractal dynamics of gait: stride-interval correlations with aging and huntington's disease. *J Appl Physiol*. 1997; 262-269.
- [25] Rodriguez-Antiguedad A Z. Enfermedades neurodegenerativas. *Minusval*. 2004: 17-19.
- [26] Haeri M, Sarbaz Y, Gharibazadeh S. Modeling the Parkinson's tremor and its treatments. *Yjtbi*. 2005; 236: 311-322.
- [27] Villadóniga M, San Millán A, Cabañes-Martínez L, Avilés-Olmos I, Del Álamo-De Pedro M, Regidor I. Análisis cuantitativo de la marcha en pacientes con enfermedad de Parkinson avanzada. *Rev Neurol* 2016; 63: 97-102.
- [28] Sarbaz, Banie M, Pooyan M, Gjaribzadeh S, Towhidkhah F, Jafari A. Modeling the gait of normal and Parkinsonian persons fot imporving the diagnosis. *neulet*. 2012; 509: 72-75
- [29] Martínez P A, García-Rui, Feliz C E, Garcia C, Del Val J, Herranz A. Reaction time and rhythm of movement in huntington's disease. *NS*. 2016; 362: 115-117.
- [30] Andrzejewski K L, Dowling A V, Stamler D, Felong T J, Harris D A, Wong C, et al. Wearable sensors in huntington disease: A pilot study. *Jiurnal of of Huntington's disease*. 2016; 5: 199-206.

- [31] Hadzipac M, Ni W, Nagy M, Steenrod N, McGinley M J, Kaushal A, et al. Reduced high-frequency motor neuron firing, emg fractionation, and gait variability in awake walking ALS. *PNAS*. 2016: E7600-E7609.
- [32] Bettini M, Gargiulo-Monachelli G M, Rodriguez G, Rey R C, Martinez P L, Sica R E. Epidemiology of amyotrophic laterak sclerosis patients in a centre in Buenos Aires. *Arq Neuropsiquiatr*. 2011; 69 (6): 867-870.
- [33] Tsai C, Wang K, Hwang C, Lee I, Lee C T. incidence, prevalence, and medical expenditures of classical amyotrophic lateral sclerosis in taiwan, 1999-2008. *Journal of the formosan medical association*. 2015; 114: 612-619.
- [34] Sierra A B. Aprendizaje automático: conceptos básicos y avanzados. Madrid: Pearson educación SA; 2016.
- [35] Llajaruna G M, Mauricio D S, Gutiérrez L. atación de equimosis en el peritaje médico legal Peruano mediante redes neuronales artificiales y procesamiento de imágenes. *Claio*. 2012; 24-28.
- [36] Nimmrich V, Draguhn A, Axmacher N. Neuronal network oscillations in neurodegenerative diseases. Springer. 2015.
- [37] Pourhedayat A, Sarbaz Y. A grey box neural network model of basal ganglia for gait signal of patients with huntingtons disease. *Basic and clinical neuroscience*. 2016; 7 (2): 107-114.
- [38] Leporace G, Batista L A, Metsavaht L, Nadal J. Residual analysis of ground reaction forces simulation during gait using neural networks with different configurations. *Conf Proc IEEE Eng Med Biol Soc*. 2015: 2812-2815

- [39] Begg R, Kamruzzaman J. Neural networks for detection and classification of walking pattern changes due to ageing. *Australasian physical & Engineering Sciences in Medicine*. 2006; 29 (2): 188-195
- [40] Yoo J, Hwang D, Moon K, Nixon M S. Automated human recognition by gait using neural network. IPTA. 2008.
- [41] Bifet A, Adaptive stream mining: pattern learning and mining from evolving data streams. Vol 207. Nueva Zelanda. IOS Press; 2010.
- [42] Ahmed R, Temko A, Marnane W, Lightbody G, Boylan G. Grading hypoxic-ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine. *Clin Neurophysiol*. 2015: <http://dx.doi.org/10.1016/j.clinph.2015.05.024>
- [43] Dutta S, Chatterjee A, Munshi S. Correlation technique and least square support vector machine combine for frequency domain based ECG beat classification. *IPEM*. 2010; 32: 1161- 1169.
- [44] Zheng H, Yang M, Wang H and McClean S. Machine learning and statistical approaches to support the discrimination of neuro-degenerative diseases based on gait analysis. *SCI*. 2009; 189: 57-70.
- [45] Ahlemeyer S A and Coleman S. A practical guide to data mining for business and industry. Alemania: Wiley; 2014.
- [46] Cuadras C M. Nuevos Métodos de análisis multivariante. Barcelona, España: CMC Editions; 2014
- [47] Dutta S Chatterjee A and Munshi S. An automated hierarchical gait pattern identification tool employing cross-correlation-based feature extraction and recurrent neural network based classification. *Expert Systems*. 2009; vol 26 (2): 202-217.

- [48] Curtze C, Nutt J G, Carlson-kuhta P, Mancini Martina, Horak F B. Objective gait and balance impairments relate to balance confidence and perceived mobility in people with parkinson's disease. *APTA*. 2016; 96 (11): 1734-1743.