

Aplicación del aprendizaje automático para predecir la recuperación de pacientes con desórdenes prolongados de conciencia: reporte preliminar

Márquez Fernando¹, Grinberg Matías², Rivas María Elisa³, Jarmoluk Verónica³, Bonamico Lucas¹, Ron Melania⁴, Russo María Julieta¹, Olmos Lisandro¹

¹ Departamento de Medicina de Rehabilitación, Instituto de Investigaciones Neurológicas Raúl Carrea (FLENI), Buenos Aires, Argentina. E-mail: fmarquez@fleni.org.ar

² Instituto de Inteligencia Artificial Humana (HUMAI), Buenos Aires, Argentina.

³ Departamento de Terapia Física, Instituto de Investigaciones Neurológicas Raúl Carrea (FLENI), Buenos Aires, Argentina.

⁴ Departamento de Terapia Ocupacional, Instituto de Investigaciones Neurológicas Raúl Carrea (FLENI), Buenos Aires, Argentina.

Resumen. Comprender el pronóstico a largo plazo y predecir la recuperación de pacientes con Desórdenes Prolongados de Conciencia (DPC) continúa siendo un desafío en la actualidad. Nos proponemos explorar la viabilidad del uso de la tecnología de Aprendizaje Automático (AA) para predecir la recuperación de la conciencia en pacientes con DPC que participaron de un programa de neurorehabilitación. Se incluyeron 90 pacientes de los cuales el 26% (23/90) recuperaron la conciencia. Con 63 pacientes se entrenaron 7 tipos de modelos de AA observándose mejores desempeños al incluir variables recolectadas en el seguimiento. En la validación con el conjunto de test (n=27) la red neuronal entrenada con 37 variables logró la mejor exactitud balanceada (EB): 0.95 (Área Bajo la Curva: 0.96, Sensibilidad: 100%, Especificidad: 90%). Con la tecnología del aprendizaje automático fue posible predecir con alta exactitud, sensibilidad y especificidad la recuperación de la conciencia de pacientes con DPC incluyendo variables recolectadas durante el primer mes de internación. A pesar de las limitaciones dadas por la pequeña cantidad de pacientes, el enfoque demostró resultados iniciales prometedores que ameritan mayor investigación.

Palabras clave: Aprendizaje automático · Desórdenes prolongados de conciencia · Neurorehabilitación · Redes neuronales · Ciencia de datos

1 Introducción

Las lesiones cerebrales adquiridas son una de las principales causas de muerte y discapacidad entre la población joven en los países industrializados [1]. Entre las

secuelas más graves resultantes se encuentran los desórdenes prolongados de conciencia (DPC) que abarcan el Síndrome de Vigilia sin Respuesta (SVSR, previamente denominado Estado Vegetativo) [2] y el Estado de Conciencia Mínima (ECM) [3]. Ambos síndromes pueden convertirse en condiciones permanentes en las que los pacientes pueden sobrevivir durante muchos años sin ningún paso aparente hacia la recuperación de la conciencia pero también podrían ser estadios de transición hacia la Emergencia del Estado de Conciencia Mínima (EECM). En la búsqueda de marcadores pronósticos, estudios previos [4], [5], [6] han demostrado que una menor edad, una lesión de causa traumática y una transición rápida al ECM son predictores de un resultado favorable (es decir, la EECM). Por otro lado, una edad de más de 45 o 50 años, SVSR prolongado, etiología anóxica, fiebre de origen central, hiperactividad simpática paroxística y un alto grado de discapacidad funcional, evaluado por la Escala de Puntuación de Discapacidad (DRS) [7], se asocian a un mal pronóstico en pacientes en estos pacientes[8], [9], [10].

Más recientemente, algunos estudios concluyeron que el nivel de conciencia expresado por el puntaje total de la Escala de Recuperación de Coma Revisada (ERC) [11], [12] proporciona información pronóstica útil para la recuperación a los 12 meses en pacientes con DPC [13], [14] y a los 24 meses en pacientes en SVSR de etiología anóxica[15]. La Guía de DPC [16] recomienda el uso estandarizado de esta escala que consiste en 6 subescalas que evalúan de forma jerárquica la función auditiva, visual, motora, verbal/oromotor, comunicación y vigilia. La valoración del estado de conciencia utilizando el ERC-R se logra mejor mediante el análisis del perfil de rendimiento completo o total, que incluye los seis puntajes de la subescala. Para mejorar la calidad de registro de la ERC-R y la interpretación clínica de estos pacientes, Chatelle, Giacino y cols en 2016 reportaron un método empírico que permite identificar combinaciones de las subescalas que resultan imposibles o improbables [17]. En 2018, Portaccio y cols. reportaron que la mejoría en la ERC-R durante las primeras cuatro semanas en rehabilitación, predice una mejor recuperación al alta después de una lesión cerebral grave [18]. Aunque los estudios psicométricos han demostrado consistentemente una alta confiabilidad entre evaluadores y evaluaciones [19], las puntuaciones de ERC están sujetas a inexactitudes atribuibles al error del examinador y otros factores de confusión que pueden conducir a una mala interpretación de los resultados cuando se utilizan como factor pronóstico aislado [20], [21]. Los profesionales que administran la ERC deben entonces conocer las ventajas y limitaciones de seleccionar diferentes puntos de corte score total [22].

Algunos exámenes complementarios vinculadas al procesamiento cerebral de la información como Neuroimágenes Funcionales [23], [24], el F-FDG PET [25], algunos paradigmas de Electroencefalograma (EEG) [26] y Potenciales Evocados [27], [28] también han sido reportados como factores pronóstico. En este sentido, la Guía sobre el Diagnóstico de Coma y otros Trastornos de la Conciencia de la Academia Europea de Neurología [29] sugiere que además de la evaluación clínica estandarizada, las técnicas basadas en EEG y la neuroimagen funcional deben integrarse para la evaluación multimodal de pacientes con DPC. No obstante, estos estudios pueden tener alto costo y no siempre están disponibles en unidades de

cuidados intensivos [30] o en entornos de rehabilitación. En consecuencia, los equipos de rehabilitación suelen valerse de las características clínicas y demográficas de los pacientes, fáciles de recopilar en la fase subaguda/crónica, para planificar el programa de rehabilitación y tratamiento de pacientes con DPC, especialmente cuando los marcadores pronósticos exámenes complementarios no están disponibles. Sin embargo, aunque todos los factores pronósticos clínicos mencionados previamente son relevantes, utilizados por separado, contienen información limitada para la práctica cotidiana. Esto se debe principalmente a que el paciente típico presenta un conjunto altamente individual y una combinación de predictores positivos y negativos, que pueden interactuar de varias maneras. Por lo tanto, estimar el pronóstico de los pacientes continúa siendo un desafío para los equipos de neurorrehabilitación.

En las últimas décadas, las herramientas de aprendizaje automático (AA) se han vuelto cada vez más populares entre los investigadores en medicina. El AA es un campo de la informática y un tipo de inteligencia artificial que se dedica al estudio de los programas que aprenden a realizar una tarea en base a la experiencia (datos) sin la necesidad de ser programadas explícitamente [31]. Los modelos predictivos que permite desarrollar el AA, tienen entonces como ventaja, sobre los modelos de la estadística tradicional, que no requieren una hipótesis previa y que pueden aprender de nuevos datos mejorando su poder predictivo. Con el tiempo han demostrado una capacidad predictiva precisa y exhaustiva, y se utilizan cada vez más en el diagnóstico y pronóstico de diferentes enfermedades [32], inclusive en trauma [34]. En este contexto, nos preguntamos si la tecnología del AA permitiría desarrollar modelos de predicción individualizada del pronóstico de pacientes con DPC utilizando variables clínicas recolectadas durante el primer mes de internación; que faciliten la planificación de estrategias de abordaje y permitan proporcionar a las familias la información pronóstica más precisa posible para guiar la toma de decisiones.

2 Objetivo

El presente estudio tiene como objetivo explorar la viabilidad del uso de la tecnología de AA para predecir la EECM de pacientes con DPC utilizando variables habitualmente recolectadas en la práctica cotidiana en los Centros de Rehabilitación, que no consumen recursos extra ni tienen alto costo.

3 Materiales y métodos

3.1 Diseño y población

Estudio observacional de cohorte retrospectiva de pacientes con DPC admitidos al programa de estimulación sensitiva y motriz, en FLENI Escobar desde Enero de 2004 a Septiembre de 2018.

3.2 Criterios de inclusión

1. Adultos con diagnóstico de alteración prolongada de la conciencia (más de 4 semanas de evolución) en SVSR y ECM
2. Etiología traumática (TEC: traumatismo craneoencefálico), anoxia, accidente cerebrovascular (ACV) y neoplásica (tumor).
3. Edad igual o mayor a 18 años

3.3 Criterios de exclusión

1. Pacientes en coma.
2. Emergencia del Estado de Conciencia Mínima antes de haber cumplido una semana en el programa de rehabilitación.
3. Enfermedad neurodegenerativa previa.
4. Trastorno neuropsiquiátrico previo.
5. Falta de datos de seguimiento o internación menor a un mes en el programa

3.4 Conjunto de datos

Los datos fueron extraídos del registro de datos generales de pacientes y del registro de evaluaciones semanales ERC-R del equipo de Traumatismo de Cráneo y Trastornos de Conciencia del Instituto FLENI. Los datos faltantes se recopilaron, cuando fue posible, de las historias clínicas de los pacientes.

3.5 Materiales

- Escala de Recuperación de Coma Revisada (ERC-R) [36]: Consiste en 6 subescalas que evalúan jerárquicamente la función auditiva (0 a 4 puntos), visual (0 a 5), motora (0 a 6), verbal/oromotor (0 a 3), comunicación (0 a 2) y vigilia (0 a 3).

- Método de detección de combinaciones imposibles o improbables de las subescalas ERC-R: Se utilizó el método empírico descrito por Chatelle & cols (2016) [17] para identificar combinaciones imposibles e improbables entre las puntuaciones de las subescalas. Existen 9 combinaciones imposibles y 36 improbables las cuales podrían indicar patologías subyacentes.

3.6 Definición de grupos

Se dicotomizó el resultado de de la recuperación de la conciencia (EECM: Emergencia del Estado de Conciencia Mínima) con las etiquetas de SI (1) o NO (0) y se definieron dos grupos para realizar un modelo predictivo de clasificación binaria:

- Grupo 0: Pacientes que NO EECM durante el tiempo de seguimiento.
- Grupo 1: Pacientes que SI EECM durante el tiempo de seguimiento.

3.7 Análisis estadístico y tecnología utilizada

Se utilizaron las librerías de software libre NumPy, Pandas, SciPy y Scikit-Learn, entre otras, con el lenguaje de programación Python. Link al repositorio: https://github.com/fermarquez2019/DOC_model

El procedimiento de análisis exploratorio contempló:

1. Variables categóricas:
 - a. Descripción con frecuencias absolutas y porcentuales.
 - b. Comparación de grupos: para evaluar la independencia se realizó test de Chi-Cuadrado y Kruskal Wallis.
2. Variables numéricas:
 - a. Se valoró la distribución mediante el Test de Shapiro-Wilk.
 - i. Variables con distribución normal: se describieron mediante la media y el desvío estándar.
 - ii. Variables sin distribución normal: se describieron con la mediana, cuartil inferior (Q1), superior (Q3), límite inferior (LI) y límite superior (LS).
 - b. Comparación de grupos: para evaluar la independencia se realizó test de Mann-Whitney.

3.8 Preprocesamiento

Durante el período de estudio, un total de 134 pacientes ingresaron en el Programa de Rehabilitación Multisensorial de Fleni Escobar. Entre ellos, 25 pacientes fueron eliminados por falta de datos respecto de las características generales, 3 porque

recuperaron la conciencia durante la primera semana y 16 porque estuvieron menos de un mes en el programa de Estimulación Multisensorial. Entre el total de evaluaciones de ERC durante el seguimiento de los pacientes incluidos, se eliminaron 30 registros que presentaban combinaciones imposibles entre las subescalas, considerándose un error de medición o registro. Al final del preprocesado de todos los registros de evaluaciones ERC de cada paciente, sólo 90 cumplieron los criterios de inclusión y contaban con una recopilación completa de datos dentro de las 4 semanas de seguimiento.

3.9 Ingeniería de variables

La mayoría de los algoritmos de aprendizaje automático no pueden manejar variables categóricas a menos que sean convertidas en valores numéricos por lo cual las variables nominales (género, diagnóstico del estado de conciencia al ingreso y presencia de una combinación improbable) fueron preprocesadas a tipo binario o “*dummie*”. Por ejemplo para “Etiología de la lesión”, registrada como TEC, ACV, Anoxia y Tumor se agregan al dataset una columna TEC (recodificada a 0 y 1 de acuerdo a la presencia o no de este diagnóstico), al igual que para ACV, Anoxia y Tumor. La función permite también automáticamente eliminar una de las columnas generadas y evitar así la colinealidad y que los algoritmos no presenten un sobre ajuste final a los datos.

Teniendo en cuenta que el número de columnas depende del número de categorías distintas, este método puede producir un gran número de nuevas variables al procesar las subescalas de la ERC, ralentizando el aprendizaje significativamente dado que el número de las categorías sería muy alto para la función, y pudiendo afectar el rendimiento final por dispersión de los puntos conforme aumenta la dimensionalidad del espacio (la “*maldición de la dimensionalidad*”). Por este motivo para incluir las subescalas de la ERC de forma más eficiente y asegurando que se conserve la naturaleza de estas variables se realizó una codificación ordinal utilizando los valores que se le asignan habitualmente para calcular la ERC-T.

En estudios previos Portaccio y colaboradores [35] reportaron que una mejora en la puntuación total de ERC-R y en diferentes subescalas durante las primeras cuatro semanas de rehabilitación hospitalaria discrimina a los pacientes con mejor pronóstico. En base a esto y las observaciones de la práctica cotidiana se definieron dos nuevas variables para investigar su correlación con EMCS y eventualmente incluirlas en los modelos.

- Variación de la ERC-R: diferencia entre la mejor puntuación ERC-R total (y subescalas) entre el día 8 al 30 de seguimiento, respecto de la puntuación inicial.
- Teniendo en cuenta que se seleccionó la mejor puntuación de la ERC-T y sus respectivas subescalas dentro las primeras 4 semanas del seguimiento, el registro no fue siempre exactamente en el mismo tiempo de seguimiento y por lo tanto se decidió desarrollar una nueva variable que ajuste la

puntuación de acuerdo al momento en el cual fue realizada la evaluación: Tasa de cambio de ERC-R (“*Velocidad de variación de la ERC*”). Se la definió operativamente como el coeficiente que surge de la variación de la ERC dividido en la cantidad de días de seguimiento.

3.10 Selección de variables

- Se exploraron las diferencias de las variables (categóricas y numéricas) entre el Grupo EMCS y No EMCS y su independencia estadística ($p < 0.05$ en test Chi Cuadrado o Mann Mann-Whitney y Kruskall-Wallis)
- Se establecieron dos puntos de corte para la selección inicial de variables: al ingreso al Centro de Rehabilitación y al mes seguimiento.
 - a. Ingreso (16 variables): edad, edad dicotomizada 1 (mayores y menores de 50 años), edad dicotomizada 2 (mayores y menores de 30 años), género, etiología, tiempo de evolución desde la lesión al ingreso, diagnóstico del estado de conciencia al ingreso (SVSR-ECM), puntuación en ERC-Total y en cada una de las subescalas.
 - b. Al mes (75 variables): Todas las variables anteriores y además, la mejor ERC-Total durante el seguimiento con sus respectivas subescalas, sus respectivas mejoría (variación respecto del ingreso) y sus respectivas relaciones entre esta mejoría y el tiempo total de seguimiento en el cual fue observada (tasa de variación). También se incluyó la presencia de al menos una de las combinaciones improbables en el total de evaluaciones y la presencia específica de las combinaciones improbables (36 variables) que describe Chatelle, Giacino y cols [17].
- Se redujo la cantidad de variables con un modelo de Eliminación Recursiva de Variables con un Bosque Aleatorio en las etapas que criterio médico fuese conveniente. Este método asigna un peso a cada una de las variables y las de pesos absolutos más pequeños se eliminan del conjunto. El procedimiento se repite de forma recursiva hasta obtener la variables más adecuadas [37] para el modelo predictivo, pero no nos permite saber si la exhaustividad es mejor con 5 o 25 variables en total. Por este motivo realizamos una eliminación recursiva de variables con validación cruzada [38] para obtener las mejores variables y el número más óptimo.
- Se analizó el coeficiente de correlación entre variables para valorar las interacción con la recuperación de la conciencia y las interacciones entre todas las variables del siguiente modo:
 - a. Ordinales o numéricas:
 - i. Sin distribución normal: se utilizó coeficiente de correlación de Spearman
 - ii. Con distribución normal: se utilizó coeficiente de correlación de Pearson
- Se realizó la eliminación manual paso a paso de las variables con menor importancia de las que tuviesen una correlación casi perfecta (coeficiente de

correlación >0.9), en un segundo paso con correlación fuerte (>0.7 y <0.9) y moderada (>0.5 y <0.7).

3.11 Modelos

Con las distintas opciones de selección de variables, se entrenaron 7 modelos de aprendizaje automático supervisado dividiendo el total de datos en un conjunto de entrenamiento (70% de la muestra) y uno de test o validación (30% de la muestra). El verdadero resultado o etiquetas (EECM o no EECM) es conocido por el algoritmo en el conjunto de datos utilizado para el entrenamiento, no así en el conjunto de test. La estimación de parámetros es lograda por el algoritmo de forma iterativa, sujeto a minimizar la función de error, la función de la clase real y prevista. Los algoritmos de clasificación utilizados fueron:

1. Perceptrón multicapa: MLPClassifier (MLP)
2. Árbol de decisión: DecisionTreeClassifier (DT)
3. Regresión logística: LogisticRegression (LR)
4. Bosque aleatorio: RandomForestClassifier (RF)
5. Máquina de vectores soporte: Support Vector Classification (SVC)
6. K-vecinos más cercanos: KNeighborsClassifier (KNN)
7. Clasificador Bayesiano ingenuo: GaussianNB (GNB)

3.12 Optimización de hiper-parámetros y validación cruzada:

Se buscó la mejor combinación de hiper-parámetros [39] definiendo un espacio de parámetros posibles y tomando muestras mediante una “búsqueda de grilla” o *Grid-search*, la combinatoria de pasos equidistantes para cada parámetro posible. Se dividió el conjunto de datos en 5 partes para realizar validación cruzada, y se eligió el mejor candidato mediante la métrica exactitud balanceada [40] para minimizar el sesgo del desbalance en la cantidad de datos de los subgrupos del conjunto de entrenamiento.

4 Resultados

4.1 Características generales

La población total estudiada incluyó 62 hombres y 28 mujeres que permanecieron en el programa durante 138 días (mediana, CI: 78 - CS: 209), de los cuales el 67% del total habían ingresado en SVSR y el 33% en ECM. Las características generales de la población se presentan en el *Anexo* en la *Tabla 1*.

El 26% (23 pacientes) recuperaron la conciencia y, en comparación con el grupo que no la recuperó, al ingreso se encontraban más frecuentemente en ECM (16/23 vs 14/67, $p < 0.001$), presentaban menor tiempo de evolución desde la lesión (54 días vs 86 días, $p = 0.027$ por Mann Whitney) y registraron puntuaciones más altas en la ERC-R Vigilia, ERC-R Motora, ERC-R Comunicación y ERC-R Total al ingreso. No se encontraron diferencias significativas en edad, género y etiología del DPC entre los grupos. Al mes de seguimiento, los pacientes que EMCS registraban puntuaciones significativamente mayores en todas las subescalas de la CRS (excepto ERC-Vigilia) y en la ERC-R Total, observándose también una mayor mejoría (variación) y mejoría en función del tiempo de seguimiento (tasa de variación) en las mismas y una mayor frecuencia de presencia combinaciones improbables entre la subescalas de la ERC-R respecto del grupo que no recuperó la conciencia. No se encontraron diferencias estadísticamente significativas al analizar cada combinación improbable por separado. Se presentan más detalles en el *Anexo, Tabla 2*.

4.3 Selección de variables y optimización de hiper-parámetros

La selección de variables se realizó a partir de dos puntos de corte: al ingreso (Opción 1: 16 variables) y al mes (Opción 2: 75 variables). Según juicio médico se aplicaron criterios de exclusión (correlaciones altas o moderadas, eliminación recursiva de variables mediante un bosque aleatorio con validación cruzada y análisis multivariable) e inclusión de variables a partir de ambos puntos de corte (*Anexo, Tabla 3*). La *Figura 2* muestra la exactitud de la Validación Cruzada función de la Eliminación Recursiva de variables con el Bosque Aleatorio utilizando la cantidad total de variables al mes (Opción 2 $n=75$). La importancia de las variables en función de este método se observa en la *Figura 3*. En el análisis multivariable la única independiente del total explorado ($n=75$) fue la mejor puntuación Total de la Escala de Recuperación de Coma Revisada (ERC-R Total) durante el seguimiento (Razón de productos cruzados (*Odds Ratio*)=1.414, $p < 0.001$ Intervalo de Confianza 95%(2.5%):1.22 - Intervalo de Confianza 95%(97.5%):1.64).

Figura 2: Exactitud de la Validación Cruzada en función de cantidad de variables seleccionadas en Bosque Aleatorio

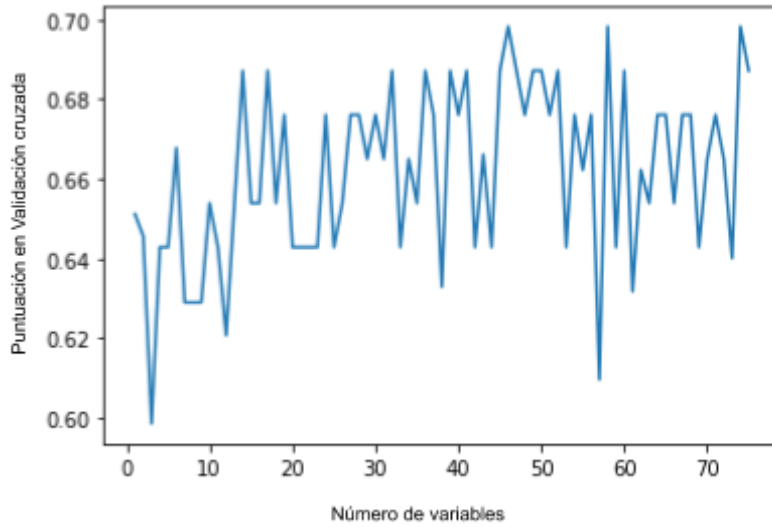
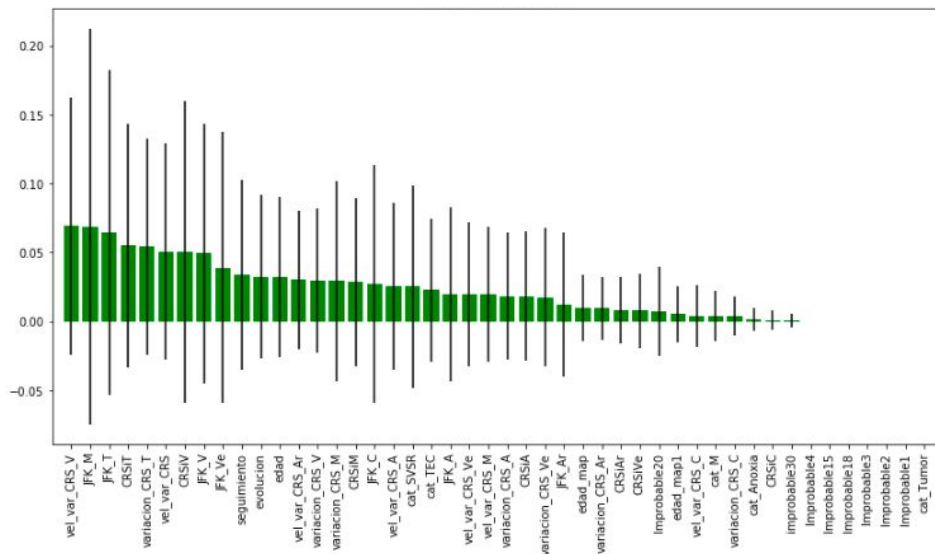


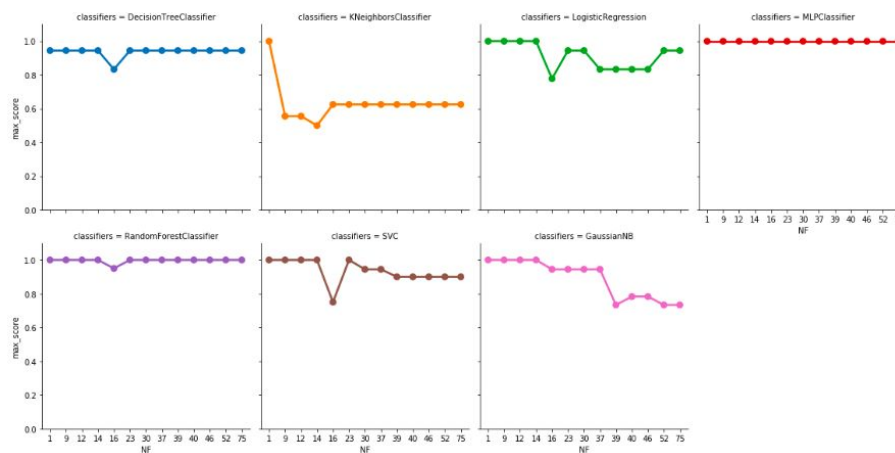
Figura 3: Importancia de las variables en la Eliminación Recursiva de Variables con Validación Cruzada



Luego de buscar los mejores hiper-parámetros para cada algoritmo en el conjunto de entrenamiento de los 13 dataset obtenidos en la selección de variables, se

entrenaron los modelos. La *Figura 4* representa la máxima exactitud alcanzada de los modelos en función de la cantidad de variables. Se observa una tendencia a descender la exactitud al utilizar únicamente las variables registradas al ingreso ($n=16$). El Perceptrón Multicapa presentó una exactitud alta y estable en relación a la cantidad de variables, al igual que el Árbol de Decisión y el Bosque Aleatorio, excepto cuando se utilizaron únicamente las variables al ingreso. Del total de modelos entrenados en se preseleccionaron los que presentaron mejores exactitudes máximas, medias y mínimas en cada dataset de entrenamiento.

Figura 4: Máxima exactitud alcanzada por los modelos en función de la cantidad de variables en el conjunto de entrenamiento



Referencias. max_score: máxima exactitud balanceada alcanzada en el conjunto de entrenamiento; classifiers: clasificadores; MLPClassifier: Perceptrón multicapa; DecisionTreeClassifier: Árbol de decisión; LogisticRegression: Regresión logística; RandomForestClassifier: Bosque aleatorio; SVC: Máquina de vectores soporte; KNeighborsClassifier: K-vecinos más cercanos; GaussianNB: Clasificador Bayesiano ingenuo.

4.4 Validación en el conjunto de test

Considerando el alto riesgo de sobreajuste a los datos en conjuntos pequeños (especialmente RF y MLP10) [41], se evaluó el desempeño de los modelos en el conjunto de test (27 pacientes) seleccionando los 7 modelos definitivos, que habían logrado una exactitud balanceada máxima perfecta en el conjunto de entrenamiento (excepto GNB 0.94 y DT 0.94), ver *Tabla 4*. En el conjunto de test, el Perceptrón Multicapa entrenado con 37 variables logró la mejor exactitud balanceada (EB): 0.95 y un Área Bajo la Curva (ABC): 0.96 (*Figura 5*), Sensibilidad: 100%, Especificidad: 90%. El Clasificador Bayesiano Ingenuo entrenado con 23 variables obtuvo el mismo desempeño (EB, sensibilidad, especificidad) excepto por una ABC más baja (0.95). La Regresión Logística y la Máquina de Vectores Soporte presentaron mayor especificidad (95%) que los modelos anteriores utilizando como única variable la

mejor puntuación Total de la Escala de Recuperación de Coma Revisada (ERC-R) durante el seguimiento, aunque el resto de las métricas fueron más bajas (EB: 0.84, ABC: 0.94, Sensibilidad: 83%). El Árbol de Decisión presentó mejores resultados con 75 variables (EB: 0.87) y el Bosque Aleatorio con 46 variables (EB: 0.81) pero con baja sensibilidad (0.67). El algoritmo de K-Vecinos más cercanos fue el de peor desempeño. Los siete algoritmos se comparan en la *Tabla 4*.

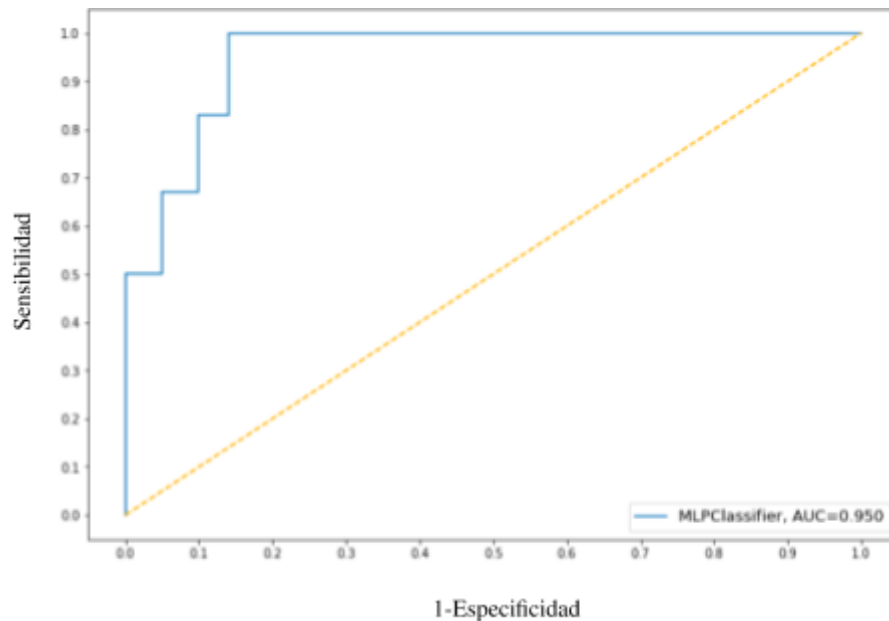
Tabla 4. Métricas de los mejores modelos en el conjunto de Entrenamiento y Testeo

NV	Clasificadores	Conjunto entrenamiento (Exactitud balanceada)			Conjunto de test				
		Score máximo	Score medio	Score mínimo	Exactitud Balanceada	ABC	Sensibilidad	Especificidad	f_1 Score
37	Perceptron Multicapa	1	0,71	0,51	0,95	0,96	1	0,9	0,95
23	Clasificador Bayesiano	0,94	0,66	0,46	0,95	0,95	1	0,9	0,95
1	Regresion Logistica	1	0,76	0,62	0,89	0,94	0,83	0,95	0,89
1	Maquina Vectores Soporte	1	0,76	0,62	0,89	0,94	0,83	0,95	0,89
75	Arbol de Decisión	0,94	0,69	0,51	0,87	0,86	0,83	0,9	0,87
46	Bosque Aleatorio	1	0,71	0,57	0,81	0,9	0,67	0,95	0,78
1	K Vecinos Más ceranos	1	0,71	0,5	0,73	0,9	0,5	0,95	0,66

Referencias. NV: Número de Variables utilizadas; max_score: máxima exactitud balanceada alcanzada en el conjunto de entrenamiento; media_score: exactitud balanceada media alcanzada en el conjunto de entrenamiento; min_score: mínima exactitud balanceada alcanzada en el conjunto de entrenamiento; ABC: Área Bajo la Curva; VP: Verdaderos Positivos; VN: Verdaderos Negativos; FP: Falsos Positivos; FN: Falsos Negativos.

5 Discusión

Los modelos desarrollados demostraron en esta primera validación generalizar de forma adecuada en el conjunto de test, especialmente en el grupo de pacientes que no recuperaron la conciencia. Esto probablemente está relacionado a que el conjunto de datos utilizado está desbalanceado y presenta más cantidad de pacientes que no recuperaron la conciencia (21 de un total de 27 en el test), como habitualmente sucede en los estudios de este tipo de pacientes. El valor clínico del desarrollo de modelos predictivos para este tipo de pacientes radica en la necesidad de transmitir el pronóstico a los familiares, tomar decisiones anticipadas y guiar las conductas por ejemplo al solicitar equipamiento definitivo, definir nuevas intervenciones o cambiar la modalidad de rehabilitación en los casos que se considere necesario.

Figura 5: Curva ROC Perceptrón Multicapa

Referencias. MLPClassifier: Clasificador Perceptrón Multicapa; AUC: Área Bajo la Curva

La utilización de cada modelo dependerá de la pregunta que deseamos responder. Si el objetivo es identificar los pacientes que recuperarán la conciencia y contamos con un seguimiento completo el MLP y GNB serían los modelos ideales ya que obtuvieron la mejor BA y sensibilidad. La LR, SVC y DT fueron los modelos con mejor especificidad y una sensibilidad mayor a KNN y RF. Esto los convierte en los modelos indicados si el objetivo es identificar los pacientes que no recuperarán la conciencia, si solo contamos con la ERC-R Total durante el seguimiento (LR y SVC).

Es necesario validar estos resultados en conjunto de validación nuevo ya que los resultados de los modelos seleccionados podrían estar sobre-ajustados a los datos del test, lo cual determina que las limitaciones del uso en la práctica clínica. Por lo tanto, los resultados obtenidos deben ser reproducidos en conjuntos de datos de mayor cantidad de pacientes. En este sentido el código de programación es de fuente abierta, para fomentar el trabajo colaborativo entre distintos centros y entrenar modelos más robustos.

5 Perspectivas a futuro

Creemos que con la transformación digital, el cambio de paradigma orientado a la colaboración internacional [43] y el crecimiento de registros de datos en general

salud [44] y en lesiones cerebral en particular [45]; los posibilidades de entrenar más y mejores modelos será solo una cuestión de tiempo y fortalecimiento del trabajo colaborativo. Un desafío a futuro será entonces poner en producción los modelos mediante una interfase gráfica para poder ingresar datos de nuevos pacientes que automáticamente re-entrenen los modelos en la nube. Asimismo, creemos relevante estudiar la exactitud de las predicciones automatizadas valorando la concordancia con predicciones realizadas únicamente bajo juicio humano y a la vez, comparar la eficacia de utilizar la combinación de ambas, en contextos de diferentes cantidades de variables como input. Nuestra hipótesis es que la predicción de la recuperación de la conciencia en pacientes con DOC puede mejorar si se combinan las predicciones humanas y automatizadas. Esto permitiría determinar si es posible tomar decisiones de manera más eficiente, combinando la anticipación de la predicción automatizada con el juicio humano en la toma de decisiones para optimizar procesos que consumen tiempo, costos y recursos.

6 Conclusiones

Con la tecnología del aprendizaje automático fue posible predecir con buena exactitud, sensibilidad y especificidad la recuperación de la conciencia de pacientes con DPC independientemente que se encuentren en SVSR o ECM. La mejor exactitud se obtuvo al incluir como variables las mejores puntuaciones de la ERC-R durante el primer mes, sus respectivas mejorías (variación respecto de las ERC-R al ingreso) y la relación entre esta mejoría y el momento en el cual es observada (tasa de variación). Estas conclusiones deben ser validadas en poblaciones de una mayor cantidad de pacientes para lo cual el código de programación es de fuente abierta.

7 Anexo

7.1 Tabla 1

Tabla 1. Características generales de la población.

	Total (n=90)
Edad, mediana [Q1,Q3]	36.0 [27.0,54.5]
Género, n (%)	F 28 (31.1)
	M 62 (68.9)
Etiología, n (%)	ACV 17 (20.2)
	Anoxia 20 (23.8)
	TEC 46 (54.8)
	Tumor 1 (1.2)
Evolución (días), mediana [Q1,Q3]	79.5 [44.2,132.0]
Diagnóstico al ingreso, n (%)	MC 30 (33.3)

	SVSR	60 (66.7)
Tiempo desde el ingreso al alta del prote (días), mediana [Q1,Q3]		138.5 [78.0,209.8]
ERC-R Auditiva inicial, mediana [Q1,Q3]		1.0 [0.0,1.0]
ERC-R Visual inicial, mediana [Q1,Q3]		0.0 [0.0,2.8]
ERC-R Motor inicial, mediana [Q1,Q3]		2.0 [1.0,2.0]
ERC-R Verbal-Oromotor inicial, mediana [Q1,Q3]		1.0 [0.2,1.0]
ERC-R Comunicación inicial, mediana [Q1,Q3]		0.0 [0.0,0.0]
ERC-R Vigilia inicial, mediana [Q1,Q3]		1.0 [1.0,2.0]
ERC-R Total inicial, mediana [Q1,Q3]		5.0 [4.0,7.0]
ERC-R Auditiva al mes, mediana [Q1,Q3]		1 [1,2]
ERC-R Visual al mes, mediana [Q1,Q3]		2 [0,3]
ERC-R Motor al mes, mediana [Q1,Q3]		2 [2,4]
ERC-R Verbal-Oromotor al mes, mediana [Q1,Q3]		1 [1,1]
ERC-R Comunicación al mes, mediana [Q1,Q3]		0 [0,0]
ERC-R Vigilia al mes, mediana [Q1,Q3]		2 [2,2]
ERC-R Total al mes, mediana [Q1,Q3]		8 [6,12]
Diagnóstico al egreso, n (%)	EMCS	23 (25.8)
	MC	27 (30.3)
	SVSR	39 (43.8)
Combinación improbable a los 30 días, n (%)	No	82 (91.1)
	Si	8 (8.9)
Recuperación de la conciencia, n (%)	No	67 (74.4)
	Si	23 (25.6)

Referencias. Q1: primer cuartil, Q3: tercer cuartil, F: femenino, M: masculino, ACV: Accidente Cerebrovascular, TEC: Traumatismo Craneoencefálico, MC: Mínima Conciencia, SVSR: Síndrome de Vigilia sin Respuesta, ERC-R: Escala de Recuperación de Coma Revisada, EMCS: Emergencia del Estado de Conciencia Mínima.

7.2 Tabla 2

Tabla 2. Características de los grupos.

		No EMCS	EMCS	Valor de P	Test
n		67	23		
Edad, mediana [Q1,Q3]		37.0 [27.0,60.0]	35.0 [25.5,43.5]	0,16	Kruskal-Wallis
Edad mayor a 30 años, n (%)	No	19 (28.4)	9 (39.1)	0,483	Chi-squared
	Si	48 (71.6)	14 (60.9)		
Edad mayor a 50 años, n (%)	No	42 (62.7)	19 (82.6)	0,132	Chi-squared
	Si	25 (37.3)	4 (17.4)		
Género, n (%)	F	22 (32.8)	6 (26.1)	0,732	Chi-squared
	M	45 (67.2)	17 (73.9)		
Etiología, n (%)	ACV	16 (25.8)	2 (8.7)	0,075	Chi-squared
	Anoxia	16 (25.8)	4 (17.4)		
	TEC	30 (48.4)	16 (69.6)		
	Tumor		1 (4.3)		
Evolución (días), mediana [Q1,Q3]		86.0 [51.5,144.0]	54.0 [38.0,101.0]	0,054	Kruskal-Wallis

Diagnóstico al ingreso, n (%)	MC	14 (20.9)	16 (69.6)	<0.001	Chi-squared
	SVSR	53 (79.1)	7 (30.4)		
ERC-R Auditiva inicial, mediana [Q1,Q3]		1.0 [0.0,1.0]	1.0 [0.5,1.0]	0,297	Kruskal-Wallis
ERC-R Visual inicial, mediana [Q1,Q3]		0.0 [0.0,1.0]	3.0 [1.0,3.0]	<0.001	Kruskal-Wallis
ERC-R Motor inicial, mediana [Q1,Q3]		2.0 [1.0,2.0]	2.0 [2.0,2.0]	0,030	Kruskal-Wallis
ERC-R Verbal-Oromotor inicial, mediana [Q1,Q3]		1.0 [0.0,1.0]	1.0 [1.0,1.0]	0,276	Kruskal-Wallis
ERC-R Comunicación inicial, mediana [Q1,Q3]		0.0 [0.0,0.0]	0.0 [0.0,0.0]	0,021	Kruskal-Wallis
ERC-R Vigilia inicial, mediana [Q1,Q3]		1.0 [1.0,2.0]	1.0 [1.0,2.0]	0,732	Kruskal-Wallis
ERC-R Total inicial, mediana [Q1,Q3]		5.0 [4.0,6.0]	7.0 [6.0,9.0]	<0.001	Kruskal-Wallis
Seguimiento (días), media (DE)		23.7 (5.7)	23.9 (5.4)	0,892	Two Sample T-test
ERC-R Auditiva al mes, mediana [Q1,Q3]		1 [1,2]	4 [2,4]	<0.001	Kruskal-Wallis
ERC-R Visual al mes, mediana [Q1,Q3]		1 [0,3]	4 [3,5]	<0.001	Kruskal-Wallis
ERC-R Motor al mes, mediana [Q1,Q3]		2 [2,2]	5 [2,6]	<0.001	Kruskal-Wallis
ERC-R Verbal-Oromotor al mes, mediana [Q1,Q3]		1 [1,1]	2 [1,2]	<0.001	Kruskal-Wallis
ERC-R Comunicación al mes, mediana [Q1,Q3]		0 [0,0]	1 [0,2]	<0.001	Kruskal-Wallis
ERC-R Vigilia al mes, mediana [Q1,Q3]		2 [2,2]	2 [2,2]	0,066	Kruskal-Wallis
ERC-R Total al mes, median [Q1,Q3]		7 [6,9]	15 [11,20]	<0.001	Kruskal-Wallis
Combinación improbable a los 30 días, n (%)	No	64 (95.5)	18 (78.3)	0,024	Fisher's exact
	Si	3 (4.5)	5 (21.7)		
Variación ERC-R Total (final-inicial), mediana [Q1,Q3]		2.0 [1.0,4.0]	7.0 [3.0,13.0]	<0.001	Kruskal-Wallis
Variación ERC-R Auditiva, mediana [Q1,Q3]		0.0 [0.0,1.0]	1.0 [0.0,3.0]	0,002	Kruskal-Wallis
Variación ERC-R Visual, mediana [Q1,Q3]		0.0 [0.0,1.0]	2.0 [1.0,2.0]	0,001	Kruskal-Wallis
Variación ERC-R Comunicación, mediana [Q1,Q3]		0.0 [0.0,0.0]	0.0 [0.0,2.0]	<0.001	Kruskal-Wallis
Variación ERC-R Vigilia, mediana [Q1,Q3]		0.0 [0.0,1.0]	1.0 [0.0,1.0]	0,209	Kruskal-Wallis
Variación ERC-R Motor, mediana [Q1,Q3]		0.0 [0.0,1.0]	1.0 [0.5,3.5]	<0.001	Kruskal-Wallis
Variación ERC-R Verbal-Oromotor, mediana [Q1,Q3]		0.0 [0.0,1.0]	0.0 [0.0,1.0]	0,019	Kruskal-Wallis
Tasa de variación ERC-R, mediana [Q1,Q3]		0.0 [0.0,0.0]	0.0 [0.0,0.1]	0,002	Kruskal-Wallis
Tasa de variación ERC-R Visual, mediana [Q1,Q3]		0.0 [0.0,0.0]	0.1 [0.0,0.1]	<0.001	Kruskal-Wallis
Tasa de variación ERC-R Comunicación, mediana [Q1,Q3]		0.0 [0.0,0.0]	0.0 [0.0,0.1]	<0.001	Kruskal-Wallis
Tasa de variación ERC-R Vigilia, mediana [Q1,Q3]		0.0 [0.0,0.0]	0.0 [0.0,0.0]	0,278	Kruskal-Wallis
Tasa de variación ERC-R Motor, mediana [Q1,Q3]		0.0 [0.0,0.0]	0.1 [0.0,0.1]	<0.001	Kruskal-Wallis
Tasa de variación ERC-R Verbal-Oromotor, mediana [Q1,Q3]		0.0 [0.0,0.0]	0.0 [0.0,0.0]	0,049	Kruskal-Wallis
Tasa de variación ERC-R Total, mediana [Q1,Q3]		0.1 [0.0,0.1]	0.3 [0.1,0.5]	<0.001	Kruskal-Wallis
Tiempo desde el ingreso al alta del		156.0	77.0	0,003	Kruskal-Wallis

programa (días), mediana [Q1,Q3]		[94.0,213.5]	[40.5,144.5]		
Diagnóstico al egreso, n (%)	EMCS	0	23 (100.0)	<0.001	Chi-squared
	MC	27 (40.9)	0		
	SVSR	39 (59.1)	0		

Referencias. Q1: primer cuartil, Q3: tercer cuartil, F: femenino, M: masculino, ACV: Accidente Cerebrovascular, TEC: Traumatismo Craneoencefálico, MC: Mínima Conciencia, SVSR: Síndrome de Vigilia sin Respuesta, ERC-R: Escala de Recuperación de Coma Revisada, EMCS: Emergencia del Estado de Conciencia Mínima, DE: Desvío Estándar.

7.3 Tabla 3

Tabla 3: Selección de variables

Opción	Cantidad de variables	Criterio de selección
1	16	<i>Variables únicamente recolectadas al ingreso</i>
1.A	23	Opción 1 + ERC-R Total y subescalas dentro del primer mes
1.B	30	Opción 1.A + variación ERC-R Total y subescalas respecto del ingreso.
1.C	37	Opción 1.B + variación en función de tiempo total de seguimiento de ERC-R Total y subescalas respecto del ingreso.
2	75	<i>Total de variables</i>
2.A	52	Exclusión de combinaciones improbables con 0 observaciones
2.A.1	39	Exclusión de todas las combinaciones improbables
2.A.1.a	14	Eliminación recursiva incluyendo las 39 variables
2.A.1.a.1	12	Exclusión de variables con correlación alta o casi perfecta
2.A.1.a.2	9	Exclusión de variables con correlación moderada
2.B	46	Eliminación recursiva incluyendo las 75 variables
2.B.1	40	Exclusión de combinaciones improbables con 0 observaciones
2.C	1	Exclusión de variables dependientes en análisis multivariable

Referencias

- Maas AIR, Stocchetti N, Bullock R. Moderate and severe traumatic brain injury in adults. *Lancet Neurol.* 2008;7:728–41.
- Laureys S, Celesia GG, Cohadon F, Lavrijsen J, León-Carrión J, Sannita WG, Szabon L, Schmutzhard E, von Wild KR, Zeman A. Unresponsive wakefulness syndrome: A new name for the vegetative state or apallic syndrome. *BMC Med.* 2010;8:68.
- Giacino JT, Ashwal S, Childs N, Cranford R, Jennett B, Katz DI, Kelly JP, Rosenberg JH, Whyte J, Zafonte RD, et al. The minimally conscious state: definition and diagnostic criteria. *Neurology.* 2002;58:349–53.
- Medical aspects of the persistent vegetative state (2). The multi-society task force on PVS. *N Engl J Med.* 1994;330:1572–79.
- Estraneo A, Moretta P, Loreto V, Lanzillo B, Santoro L, Trojano L. Late recovery after traumatic, anoxic, or hemorrhagic long-lasting vegetative state. *Neurology.* 2010;75:239–45.
- Katz DI, Polyak M, Coughlan D, Nichols M, Roche A. Natural history of recovery from brain injury after prolonged disorders of consciousness: outcome of patients admitted to inpatient rehabilitation with 1-4 year follow-up. *Prog Brain Res.* 2009;177:73–88.
- Turgeon AF, Lauzier F, Simard J-F, Scales DC, Burns KEA, Moore L, Zygun DA, Bernard F, Meade MO, Dung TC. Mortality associated with withdrawal of life-sustaining therapy for patients with severe traumatic brain injury: a Canadian multicentre cohort study. *Can Med Assoc J.* 2011;183:1581–88.
- Jain S, Dharap SB, Gore MA. Early prediction of outcome in very severe closed head

- injury. *Injury*. 2008;39:598–603. doi:10.1016/j.injury.2007.06.003.
9. Rovlias A, Kotsou S. Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables. *J Neurotrauma*. 2004;21:886–93. doi:10.1089/0897715041526249.
 10. Lin Q1, Xie QY, He YB, Chen Y, Ni XX, Guo YQ, Shen Y, Yu RH. Nan Fang Yi Ke Da Xue Xue Bao. 2017 Mar 20;37(3):337-341. Factors affecting recovery of consciousness in patients with disorders of consciousness following brain trauma: a logistic regression analysis
 11. Kochanek PM, Bell MJ. Making an IMPACT in traumatic brain injury research. *Lancet Neurol*. 2013;12:1132–33.
 12. Voss A. Standards der neurologischen-neurochirurgischen Frührehabilitation. Ein Konzept der Arbeitsgemeinschaft neurologisch-neurochirurgische Frührehabilitation. In: Wild KJHH, editor. *Spectrum der Neurorehabilitation: frührehabilitation; Rehabilitation von Kindern und Jugendlichen*. Bern, Switzerland: Zuckerschwerdt; 1993. p. 112–20.
 13. Hamilton JA, Perrin PB, Campbell TA, Danish SJ, Goldstein AL. Predicting emergence from a disorder of consciousness using the Coma Recovery Scale–Revised. *Neuropsychol Rehab*. 2018;16:1–15.
 14. Portaccio E, Morrocchesi A, Romoli AM, Hakiki B, Taglioli MP, Lippi E, Di Renzone M, Grippo A, Macchi C. Intensive Rehabilitation Unit Study Group of the IRCCS Don Gnocchi Foundation, Italy. Improvement on the Coma Recovery Scale-Revised during the first four weeks of hospital stay predicts outcome at discharge in intensive rehabilitation after severe brain injury. *Arch Phys Med Rehabil*. 2018;99:914–19.
 15. Estraneo A, Moretta P, Loreto V, Lanzillo B, Cozzolino A, Saltalamacchia A, Lullo F, Santoro L, Trojano L. Predictors of recovery of responsiveness in prolonged anoxic vegetative state. *Neurology*. 2013;80:464–70.
 16. Giacino, Joseph T et al. “Practice guideline update recommendations summary: Disorders of consciousness: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology; the American Congress of Rehabilitation Medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research.” *Neurology* vol. 91,10 (2018): 450-460.
 17. Chatelle, Camille et al. “Detection and Interpretation of Impossible and Improbable Coma Recovery Scale-Revised Scores.” *Archives of physical medicine and rehabilitation* vol. 97,8 (2016): 1295-1300.e4.
 18. Portaccio E, Morrocchesi A, Romoli AM, Hakiki B, Taglioli MP, Lippi E, Di Renzone M, Grippo A, Macchi C, for the Intensive Rehabilitation Unit Study Group of the IRCCS Don Gnocchi Foundation, Italy, Improvement on the Coma Recovery Scale-Revised during the first four weeks of hospital stay predicts outcome at discharge in intensive rehabilitation after severe brain injury, *Arch Phys Med Rehabil*. 2018.
 19. Seel RT, et al. Assessment scales for disorders of consciousness: evidence-based recommendations for clinical practice and research. *Arch Phys Med Rehabil*. 2010; 91(12):1795–813.
 20. La Porta F, et al. Can we scientifically and reliably measure the level of consciousness in vegetative and minimally conscious States? Rasch analysis of the coma recovery scale-revised. *Arch Phys Med Rehabil*. 2013; 94(3):527–535e1.
 21. Gerrard P, Zafonte R, Giacino JT. Coma Recovery Scale-Revised: evidentiary support for hierarchical grading of level of consciousness. *Arch Phys Med Rehabil*. 2014; 95(12):2335–41.
 22. Bodien YG, Carlowicz CA, Chatelle C, Giacino JT. Sensitivity and Specificity of the Coma Recovery Scale--Revised Total Score in Detection of Conscious Awareness. *Arch Phys Med Rehabil*. 2016;97(3):490–492.e1.
 23. Di HB, Yu SM, Weng XC, Laureys S, Yu D, Li JQ, Qin PM, Zhu YH, Zhang SZ, Chen YZ. Cerebral response to patient’s own name in the vegetative and minimally conscious

- states. *Neurology*. 2007;68:895–99.
24. Coleman MR, Davis MH, Rodd JM, Robson T, Ali A, Owen AM, Pickard JD. Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. *Brain*. 2009;132:2541–52.
 25. Stender J, Gosseries O, Bruno M-A, Charland-Verville V, Vanhaudenhuyse A, Demertzi A et al. . Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. *Lancet* 2014; 384: 514–22.
 26. Engemann DA, Raimondo F, King J-R, et al. Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain* 2018; 141: 3179–3192.
 27. Robinson LR, Micklesen PJ, Tirschwell DL, Lew HL. Predictive value of somatosensory evoked potentials for awakening from coma. *Crit Care Med*. 2003;31:960–67. Review.
 28. Estraneo A, Moretta P, Loreto V, Lanzillo B, Cozzolino A, Saltalamacchia A, Lullo F, Santoro L, Trojano L. Predictors of recovery of responsiveness in prolonged anoxic vegetative state. *Neurology*. 2013;80:464–70. doi:10.1212/WNL.0b013e31827f0f31.
 29. Kondziella, D., Bender, A., Diserens, K., van Erp, W., Estraneo, A., Formisano, R., Laureys, S., Naccache, L., Ozturk, S., Rohaut, B., Sitt, J.D., Stender, J., Tiainen, M., Rossetti, A.O., Gosseries, O., Chatelle, C. and (2020), European Academy of Neurology guideline on the diagnosis of coma and other disorders of consciousness. *Eur J Neurol*, 27: 741-756.
 30. Robinson LR, Chapman M, Schwartz M, Bethune AJ, Potapova E, Strauss R, Scales DC. Patterns of use of somatosensory-evoked potentials for comatose patients in Canada. *J Crit Care*. 2016;36:130–33.
 31. Arthur Samuel, 1959
 32. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. & Fotiadis, D. I. Machine Learning Applications in Cancer Prognosis and Prediction. *Comput. Struct. Biotech*. 13, 8–17 (2015).
 33. Pan, L. et al. Machine Learning Applications for Prediction of Relapse in Childhood Acute Lymphoblastic Leukemia. *Sci. Rep.-UK*. 7, 7402–7409 (2017).
 34. N.T. Liu, J. Salinas. Machine learning for predicting outcomes in trauma. *Shock*, 48 (2017), pp. 504-510
 35. Portaccio EI, Morrocchesi A, Romoli AM, Hakiki B, Taglioli MP, Lippi E, Di Renzone M, Grippo A, Macchi C. Improvement on the Coma Recovery Scale-Revised During the First Four Weeks of Hospital Stay Predicts Outcome at Discharge in Intensive Rehabilitation After Severe Brain Injury. *Arch Phys Med Rehabil*. 2018 May;99(5):914-919.
 36. Giacino, J & Kalmar, K. (2006). Coma Recovery Scale-Revised. The Center for Outcome Measurement, Brain Injury.
 37. http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html
 38. http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFECV.html
 39. Probst, A.-L. Boulesteix, B. Bischl. Tunability: importance of hyperparameters of machine learning algorithms. *J Mach Learn Res*, 20 (2019), pp. 1-32
 40. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.balanced_accuracy_score.html
 41. E.W. Steyerberg, A.J. Vickers, N.R. Cook, T. Gerds, M. Gonen, N. Obuchowski, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*, 21 (2010), pp. 128-138
 42. A. Marmarou, J. Lu, I. Butcher, G.S. McHugh, N.A. Mushkudiani, G.D. Murray, et al. IMPACT database of traumatic brain injury: design and description. *J Neurotrauma*, 24 (2007), pp. 239-250
 43. <https://commonfund.nih.gov/bd2k>
 44. https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf

45. T.L. Neurology. The changing landscape of traumatic brain injury research. *Lancet Neurol*, 11 (2012), p. 651