

# Factores genéticos de pacientes con ictus asociados al pronóstico.

**Dolors Giralt Casellas**

Máster oficial en Bioinformática y bioestadística

**Consultor: Melchor Sánchez Martínez**

13/01/2016



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## FICHA DEL TRABAJO FINAL

<b>Título del trabajo:</b>	Factores genéticos de pacientes con ictus asociados al pronóstico.
<b>Nombre del autor:</b>	Dolors Giralt Casellas
<b>Nombre del consultor:</b>	Melchor Sanchez Martinez
<b>Fecha de entrega (mm/aaaa):</b>	01/2016
<b>Área del Trabajo Final:</b>	Bioinformática farmaceutica
<b>Titulación:</b>	<i>Máster oficial en Bioinformática y bioestadística</i>
<b>Resumen del Trabajo (máximo 250 palabras):</b>	
<p>El objetivo de este trabajo es utilizando una base de datos de pacientes que han padecido un ictus, encontrar tendencias entre pronóstico al alta y medicamentos que toman los pacientes previamente al ictus. Seguidamente cruzar los factores genéticos asociados anteriormente con bases de datos del genoma humano para ver los genes asociados a las tendencias encontradas y ver la drogabilidad de los genes.</p> <p>Partiendo de 3 bases de datos distintas, se fusionaron en una sola base, contando con 2561 pacientes con ictus.</p> <p>Se hizo un descriptivo y un random forest para cada tipo de pronóstico (mejoría, empeoramiento, mortalidad y dependencia) para encontrar factores clínicos o medicamentos asociados a pronóstico.</p> <p>Los factores más relevantes para el pronóstico fueron la edad, NIHSS y la categoría clínica. Concretamente para mejoría, los factores asociados fueron la edad, la categoría clínica y el tpa.</p> <p>Se buscaron en la literatura las proteínas asociadas al tpa y a mejoría y se encontraron el PLAT y la CRP. De éstas, mediante cytoscape y reactome, se obtubieron las redes de proteínas y se vió que tenían una en común, la fibronectina.</p> <p>Se miró la drogabilidad de la CRP y la fibronectina y se vió que la CRP tenía un porcentaje de drogabilidad medio y el porcentaje de drogabilidad de la fibronectina era nulo.</p>	

**Abstract (in English, 250 words or less):**

The objective of this study is, using a stroke data base, find any trend between outcome at discharge and drugs that patients took before stroke. Then, cross the genetic factors found previously with genome database to find the associated genes of the trends were found previously and finally show the druggability of these genes.

Three data base from stroke patients were joined in one, achieving 2561 stroke patients.

A descriptive analysis and a random forest were performed for each prognostic type (improvement, worsening, mortality and dependency) to find clinic factors or drugs associated to outcome.

The most relevant factors for outcome were age, NIHSS and stroke type. In particular, for improvement at discharge, the most relevant factors were age, stroke type and tpa.

Proteins associated to tpa and improvement were searched in the literature and PLAT and CRP were found. Using cytoscape and reactome, network for these two proteins were performed. Moreover, these networks showed they had fibronectin in common.

Also, the druggability of CRP and fibronectin was analyzed. CRP showed a medium percentage of druggability, but fibronectin did not.

**Palabras clave (entre 4 y 8):**

Ictus, random forest, farmacoabilidad, crp, fibronectina, plat

# Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo.....	2
1.3 Enfoque y método seguido .....	2
1.4 Planificación del trabajo.....	3
1.5 Breve resumen de productos obtenidos .....	3
1.6 Breve descripción de los otros capítulos de la memoria.....	3
2 Métodos .....	4
3 Resultados .....	6
3.1 Descriptivo y Random forest .....	6
3.2 Búsqueda de genes .....	18
3.3 Drug target.....	21
4. Discusión.....	23
5. Conclusiones.....	24
6. Glosario .....	25
7. Bibliografía .....	26

## Lista de figuras

Fig 1: Planificación del trabajo .....	pag 3
Fig 2: Importancia de los factores clínicos y medicamentos según el random forest para mejoría .....	pag 7
Fig 3: Gráfico de errores del random forest para mejoría dependiendo del número de árboles utilizados .....	pag 8
Fig 4: Curva ROC del modelo del random forest para mejoría en la base del test .....	pag 8
Fig 5 Importancia de los factores clínicos y medicamentos según el random forest para empeoramiento.....	pag 10
Fig 6: Gráfico de errores del random forest para empeoramiento dependiendo del número de árboles utilizados .....	pag 11
Fig 7: Curva ROC del modelo del random forest para empeoramiento en la base del test.....	pag 11
Fig 8 Importancia de los factores clínicos y medicamentos según el random forest para mortalidad .....	pag 13
Fig 9: Gráfico de errores del random forest para mortalidad dependiendo del número de árboles utilizados .....	pag 14
Fig 10: Curva ROC del modelo del random forest para mortalidad en la base del test.....	pag 14
Fig 11 Importancia de los factores clínicos y medicamentos según el random forest para dependencia .....	pag 16
Fig 12: Gráfico de errores del random forest para dependencia según el número de árboles utilizados .....	pag 17
Fig 13: Curva ROC del modelo del random forest para dependencia en la base del test.....	pag 17
Fig 14: Red de proteínas para PLAT .....	pag 18
Fig 15: Red de proteínas para CRP .....	pag 19
Fig 16: Red de proteínas para CRP y PLAT .....	pag 20
Fig 17: Drogabilidad para la fibronectina .....	pag 21
Fig 18: Drogabilidad para la CRP .....	pag 22

# 1. Introducción

## 1.1 Contexto y justificación del Trabajo

El ictus es una de las principales causas de mortalidad y discapacidad en los países desarrollados.

El ictus es una obstrucción en un vaso sanguíneo del cerebro (isquémicos) o una rotura del vaso sanguíneo, entonces se trata de una hemorragia.

Encontramos 3 tipos de ictus, los ictus isquémicos establecidos, los accidentes isquémicos transitorios (AIT) y las hemorragias. Se define como AIT como un ictus isquémico en el cual los síntomas han desaparecido dentro de las 24 primeras horas [1]. Además, estos pacientes tienen la particularidad de que la National Institute of Health Stroke Scale (NIHSS) es cero [2].

Existen varias escalas para medir la gravedad del paciente. En este trabajo se utilizarán dos de ellas. La escala NIHSS (National Institute of Health Stroke Scale) que consta de 43 puntos, considerando 0 que el paciente está perfectamente y 42 cuando está muy grave. Otra escala es la de ranking modificado (mRS), se trata de una escala que mide la capacidad que tiene el paciente de valerse por sí mismo. Tiene una escala de valores de 0 a 6, siendo 0-2 una persona independiente capaz de hacer las tareas la vida diaria por sí solo, de 3-5 ese paciente necesita ayuda de otra persona para realizar las tareas y 6 es fallecimiento del paciente.

A fecha de hoy, en la fase aguda del ictus, solo se pueden beneficiar de tratamiento los pacientes que han sufrido un ictus isquémico dentro de las 4.5 primeras horas desde el inicio de los síntomas. Este medicamento se denomina activador tisular del plasminogeno (tpa). Se sabe que la mayoría de los pacientes que se les administra tpa tienen mejor pronóstico [3,4], pero el problema está que solo una pequeña parte de la población se puede beneficiar de ello. Según un estudio, entre un 10% y un 15% de los pacientes se les puede administrar tpa [5]. En cambio, los pacientes que han sufrido hemorragias no se pueden beneficiar de ningún medicamento, solo se les puede bajar la presión arterial. Con lo cual, es bastante necesario encontrar algún medicamento con el que poder tratar al resto de pacientes.

En la literatura se ha descrito que hay medicamentos que tienen una diana terapéutica pero es posible darles un nuevo uso. Un ejemplo de ello son las estatinas, las cuales se utilizan para bajar los niveles de

colesterol. En un estudio encuentran que las estatinas se asocian al pronóstico funcional en pacientes con ictus [6].

Por este motivo es interesante encontrar algún medicamento ya sea nuevo o que ya exista pero con otro uso que ayude a los pacientes a tener mejor pronóstico.

## 1.2 Objetivos del Trabajo

Objetivos generales:

- Encontrar tendencias entre medicamentos y pronóstico al alta.
- Buscar factores genéticos asociados a las tendencias predichas con bases de datos de genomas para encontrar genes implicados.

Objetivos específicos:

- Encontrar alguna tendencia entre algún medicamento y el pronóstico al alta (mejoría o empeoramiento) mediante random forest.
- Buscar factores genéticos en bases de datos de genomas para ver que genes están implicados en las tendencias encontradas anteriormente.
- Extraer posibles biomarcadores o dianas terapéuticas a partir de los genes encontrados.

## 1.3 Enfoque y método seguido

Este trabajo consistirá en fusionar 3 bases de datos de pacientes que han sufrido un ictus (tanto isquémico como hemorrágico). De estas tres bases de datos se dispone de factores de riesgo clásicos del ictus, como son edad, género, diabetes, dislipemia e hipertensión, también se dispone de datos sobre el ingreso en el hospital (tipo de ictus, etiología, etc), medicamentos que tomaban previamente al ictus (antihipertensivos, antidiabéticos, iecas, estatinas, etc) y el pronóstico al alta para cada paciente (mejoría y empeoramiento).

Con estas bases se pretende analizar mediante random forests lo que aporta la clínica y los medicamentos que toman los pacientes al pronóstico de éstos.

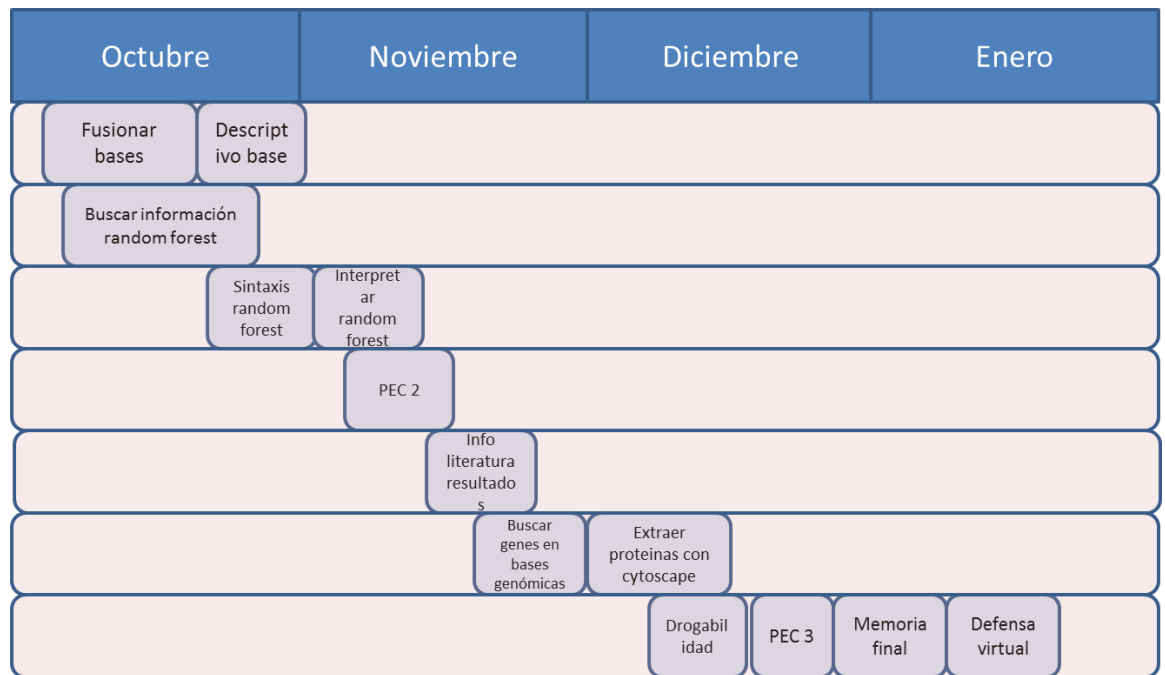
Además, de los medicamentos que se encuentren relacionados con el pronóstico se buscarán los genes asociados a los medicamentos y además se buscará en la literatura genes asociados al pronóstico.

Una vez encontrados los genes, se mirará su drogabilidad.



## 1.4 Planificación del Trabajo

Fig 1: planificación del trabajo



## 1.5 Breve resumen de productos obtenidos

- Relación de los factores clínicos y los medicamentos respecto al pronóstico.
- Obtención de los genes relacionados con los resultados más relevantes del random forest.
- Análisis de la drogabilidad de los genes encontrados

## 1.6 Breve descripción de los otros capítulos de la memoria

En el segundo capítulo se encontrarán los métodos utilizados en este trabajo, los resultados y una pequeña discusión.

## 2. Métodos

### Descripción de la base de datos

Partimos de 3 bases de datos con pacientes de ictus del hospital Vall d'Hebrón. Estas bases fueron fusionadas y los pacientes con algún missing fueron eliminados. Algunas de las variables también fueron eliminadas debido a que tenían gran número de missings y era preferible eliminar dicha variable que no perder más pacientes.

Después de depurar la base de datos final, nos quedamos con 2561 pacientes.

### Definición de las variables resultado:

Se entiende como mejoría al alta, una disminución de al menos 4 puntos en la escala NIHSS al alta respecto la de ingreso. El empeoramiento al alta es un aumento de como mínimo 4 puntos en la escala NIHSS al alta respecto la misma al ingreso [7].

La mortalidad se obtuvo de aquellos pacientes que fallecieron al alta o al 3r mes.

La dependencia al 3r mes se calculó a partir de la escala de ranking modificada (mRS), que consta de 7 puntos. Si el paciente tiene 2 o menos puntos se establece que es independiente y si tiene 3 o más puntos será dependiente para realizar las actividades de la vida diaria [8].

### Análisis estadístico

- Estudio de comparación univariante:

Se hizo un análisis comparando las variables resultado (mejoría y empeoramiento al alta, mortalidad y dependencia al tercer mes) con las características basales de los pacientes, incluyendo los medicamentos que tomaban antes de tener el ictus y el tpa, medicamento que se administra en la fase hiper aguda (<4.5h) del ictus isquémico.

En el análisis descriptivo o univariante, se compararon las variables categóricas con una chi-cuadrado, t-test para la edad y mann-whitney U test para la escala NIHSS. Presentando la información de los porcentajes, media o mediana en una tabla para la correspondiente variable respuesta.

- Random forest:

Para el análisis del random forest se utilizó un training set y un test set. Para crear el training set se utilizó una muestra aleatoria con reemplazo del 80% de la base de datos (n=2048). Para testar los modelos creados por la base de datos del training se utilizó una base

“test” creada a partir del 20% restante del total de la base aleatorizada (n=517).

Para hacer el análisis del random forest se utilizó el paquete RandomForest y para crear las curvas ROC se utilizó el paquete ROCR del repositorio de R.

La importancia de las variables utilizadas en el modelo de los random forest se calculó mediante la reducción promedio del criterio de separación, en este caso, por el método de Gini.

Además se obtuvo la curva ROC y su correspondiente área bajo la curva para la base test para ver el poder de predicción.

- Proteínas asociadas a los resultados del random forest:

Sé buscó en la literatura la proteína sobre la que actúa el tpa y si algún estudio asociaba alguna proteína a mejoría.

Una vez encontradas estas dos proteínas, una para el tpa y la otra asociada a mejoría, la proteína c-reactiva (CRP), se utilizó la página web [www.reactome.org](http://www.reactome.org) para ver redes de interacciones entre proteínas en cada una de las dos encontradas.

Se utilizó el programa cytoscape con la aplicación Reactome FI para juntar las dos redes creadas anteriormente pudiendo apreciar si estas dos redes compartían alguna proteína.

- Farmacabilidad de las proteínas:

Para medir la farmacabilidad de las proteínas encontradas, se utilizaron las webs <https://www.ebi.ac.uk/chembl/drugability/structure> y <http://pockdrug.rpbs.univ-paris-diderot.fr/cgi-bin/index.py?page=Druggability>. Se miró la farmacabilidad de la CRP y de la fibronectina, la proteína PLAT no se miró ya que proviene de un medicamento.

### 3. Resultados

#### 3.1 Descriptivo y Random forest

##### Mejoría al alta

##### Univariante para mejoría:

Del total de pacientes, 862 (33.66%) presentaron alguna mejoría al alta. En esta tabla se encuentran los porcentajes, media o mediana correspondiente a mejoría según si se presenta la característica en cuestión o no.

	No	Si	p-valor
<b>Características basales</b>			
-Sexo (hombre)	31.5%	36.5%	0.007
-Edad	70.43±13.42	70.31±12.86	0.127
-Tipo paciente (isquémico)	26.4%	34.4%	0.013
-Ictus previo	34.1%	28.4%	0.075
-Fibrilación auricular	31.8%	41.5%	<0.001
-Cardiopatía isquémica	32.6%	41.1%	0.003
-Hipertensión	35.3%	32.5%	0.155
-Diabetes	34.5%	31%	0.112
-Tabaco	33.1%	35.6%	0.261
-Alcohol	32.5%	37.1%	0.035
-Dislipemia	34.2%	32.3%	0.328
-Categoría clínica			P<0.001
-Cardioembólico		44.2%	
-Aterotrombótico		29.6%	
-Lacunar		27.5%	
-Indeterminado		31.8%	
-Hemorrágico		26.3%	
-NIHSS basal	4 (1-9)	11 (8-17)	<0.001
<b>Medicamentos</b>			
-tpa	23.2%	63.4%	<0.001
-Clopidogrel	34.2%	30.6%	0.162
-AAS	35.1%	31%	0.035
-Estatinas	35.1%	31.2%	0.044
-Ara II	33.3%	36.6%	0.32
-IECAs	35.3%	29.2%	0.004
-Diureticos	32.7%	38.3%	0.024
-Anti diabéticos	34.3%	30.7%	0.115

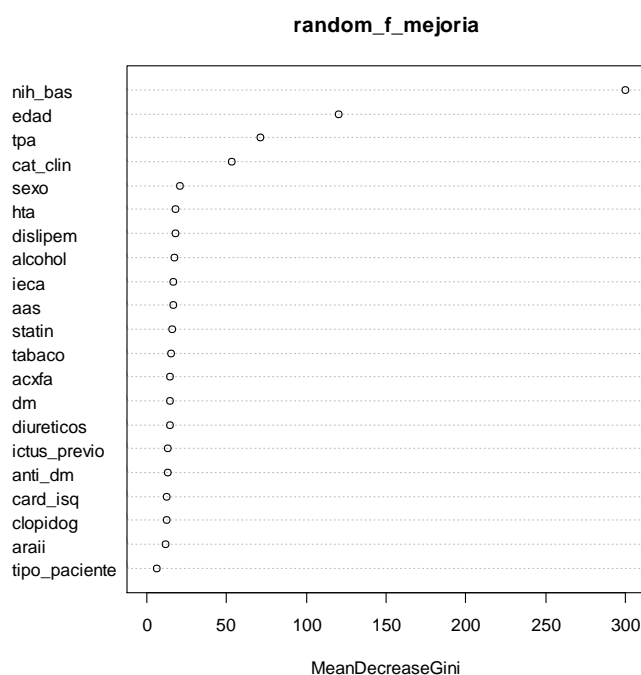
En la tabla se observa que hay diferencias significativas en el sexo, los hombres mejoran más, los ictus isquémicos, los pacientes que tienen fibrilación auricular, cardiopatía isquémica, tomar alcohol, los que tienen un ictus isquémico de etiología cardioembólica, la NIHSS al ingreso alta, el hecho de

habérseles administrado tpa o tomado aspirinas, estatinas, IECAs o diuréticos también son más propensos a mejorar.

### Random forest

Según el gráfico de la importancia de las variables incluidas en el modelo del random forest vemos que la edad, el tpa y la categoría clínica son las variables que más aportan al modelo. La importancia se calcula mediante la reducción promedio del criterio de separación, en este caso, por el método de Gini. En este caso, la edad tiene un valor de 150.91, el tpa un valor de 112.2 y la categoría clínica un valor de 65.63. Las otras variables tienen valores por debajo de 25.

Fig 2: Importancia de los factores clínicos y medicamentos según el random forest para mejoría.



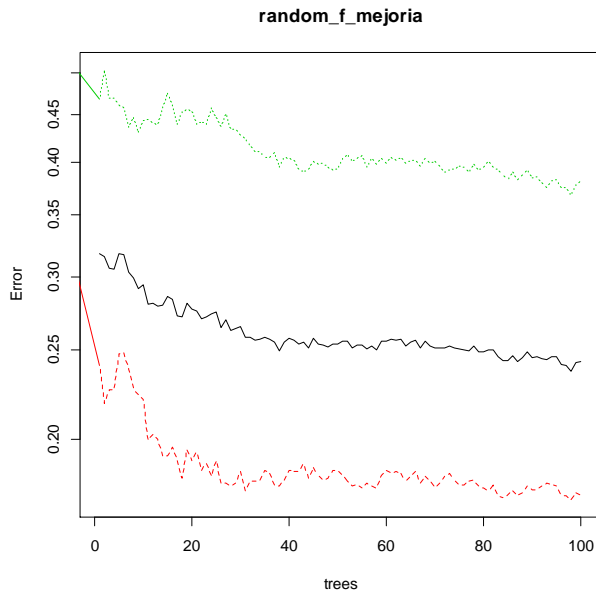
En la figura donde se muestra la importancia de las variables incluidas en el análisis del random forest se observa que las variables con más relevancia son NIHSS al ingreso, edad, tpa y categoría clínica.

Este modelo también nos enseña el error de clasificación para la base test, que en este caso es de un 17.4% para los que no mejoran y un 38.12% para los que mejoran. Por lo tanto, podríamos decir que este modelo clasifica mejor los pacientes que no mejoran ya que el error de clasificación es mucho menor en la no mejoría (17.4%) que en la mejoría (38.12%).

	No mejoría	Mejoría	Error de clasificación
No mejoría	1139	240	17.4%
Mejoría	260	422	38.12%

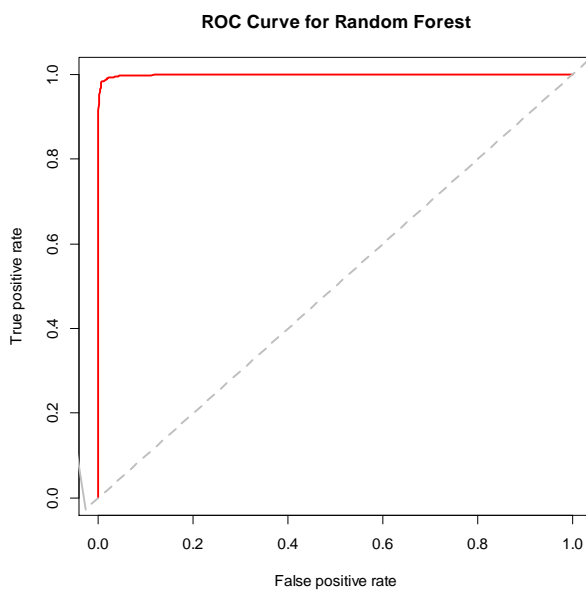
En el gráfico de los errores según el número de árboles, vemos que el error medio es de alrededor de 0.3. Si miramos el gráfico se puede apreciar que a partir de 20 árboles, el error se estabiliza a 0.3. En este gráfico también podemos observar que el intervalo del error es más pequeño cuantos menos árboles haya.

Fig 3: Gráfico de errores del random forest para mejoría dependiendo del número de árboles utilizados.



El área bajo la curva del modelo en la base del test es de 0.999.

Fig 4: Curva ROC del modelo del random forest para mejoría en la base del test.



## Empeoramiento al alta

### Univariante

Del total de pacientes, 133 (6.7%) presentaron alguna mejoría al alta  
En esta tabla se encuentran los porcentajes, media o mediana correspondiente a mejoría según si se presenta la característica en cuestión o no.

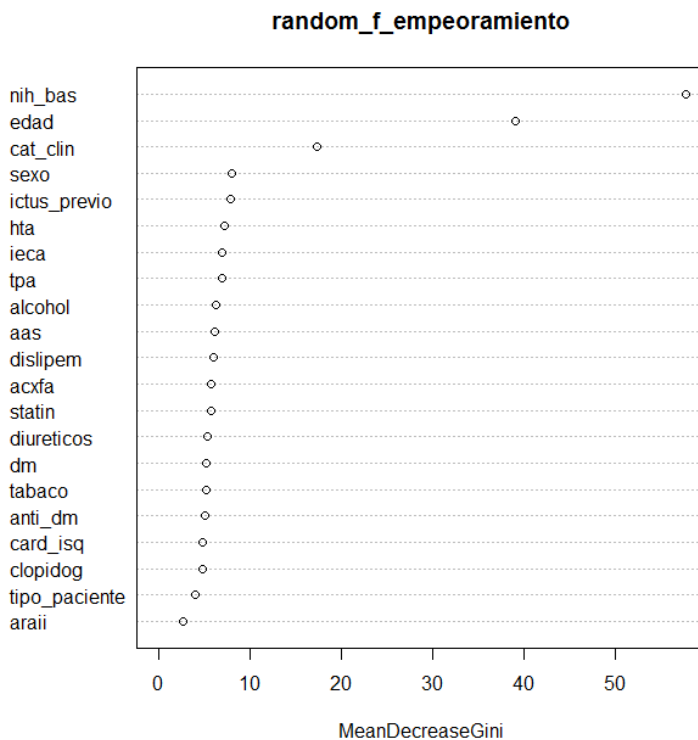
	No	Si	p-valor
<b>Características basales</b>			
-Sexo (hombre)	7.5%	6.2%	0.197
-Edad	70±13.3	75.75±11.05	<0.001
-Tipo paciente (isquémico)	16.1%	5.8%	<0.001
-Ictus previo	5.9%	15.3%	<0.001
-Fibrilación auricular	5.5%	12.2%	<0.001
-Cardiopatía isquémica	6.6%	8.1%	0.314
-Hipertensión	6.6%	6.9%	0.782
-Diabetes	6.6%	7.1%	0.664
-Tabaco	7.3%	4.6%	0.023
-Alcohol	7.3%	5%	0.4
-Dislipemia	6.9%	6.5%	0.714
-Categoría clínica			<0.001
-Cardioembólico		9.7%	
-Aterotrombótico		7.8%	
-Lacunar		3.2%	
-Indeterminado		2.1%	
-Hemorrágico		16.3%	
-NIHSS basal	6 (2-12)	18 (9-21)	<0.001
<b>Medicamentos</b>			
-tpa	5.8%	9.5%	0.001
-Clopidogrel	6.9%	5.8%	0.419
-AAS	7.9%	4.7%	0.002
-Estatinas	8.5%	3.9%	<0.001
-Ara II	7%	3.6%	0.047
-IECAs	6.7%	6.8%	0.937
-Diureticos	6.8%	6.5%	0.855
-Anti diabéticos	6.5%	7.8%	0.282

En la tabla de comparación se observa que para empeoramiento existen diferencias significativas en los pacientes mayores, los pacientes con ictus hemorrágico, los que han sufrido un ictus anterior, o los que tienen fibrilación auricular. También los que tienen NIHSS al ingreso elevadas, tomaron tpa, no tomaban aspirina o estatinas.

## Random forest

Según el gráfico de la importancia de las variables incluidas en el modelo del random forest vemos que la edad, NIHSS al ingreso y la categoría clínica son las variables que más aportan al modelo. En este caso, la edad tiene un valor de 39.02, la NIHSS al ingreso 57.64 y la categoría clínica un valor de 17.39. Las otras variables tienen valores por debajo de 10.

Fig 5 Importancia de los factores clínicos y medicamentos según el random forest para empeoramiento.



En el gráfico del random forest se observa que las variables con más importancia son la NIHSS al ingreso, la edad y la categoría clínica.

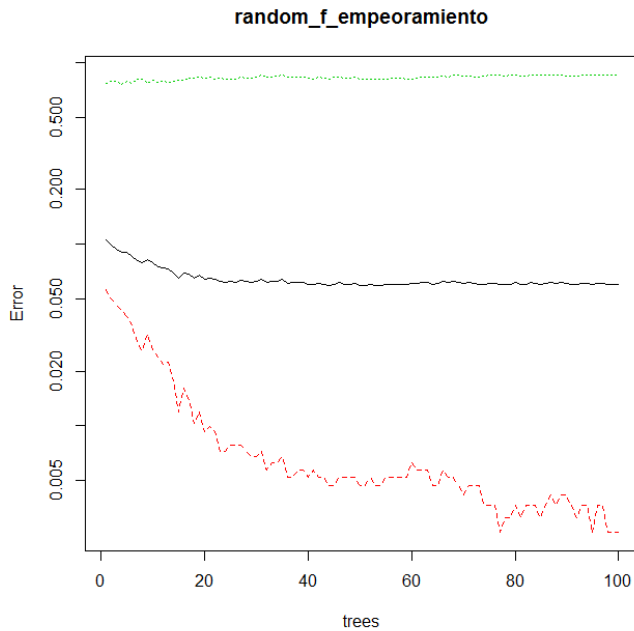
Este modelo también nos enseña el error de clasificación para la base test, que en este caso es de un 0.26% para los que no empeoran y un 85% para los pacientes que empeoran ya que el error de clasificación es mucho menor en la no empeoramiento (0.26%) que en empeoramiento (85%).

	No empeoramiento	Empeoramiento	Error de clasificación
No empeoramiento	1926	5	0.26%
Empeoramiento	119	21	85%



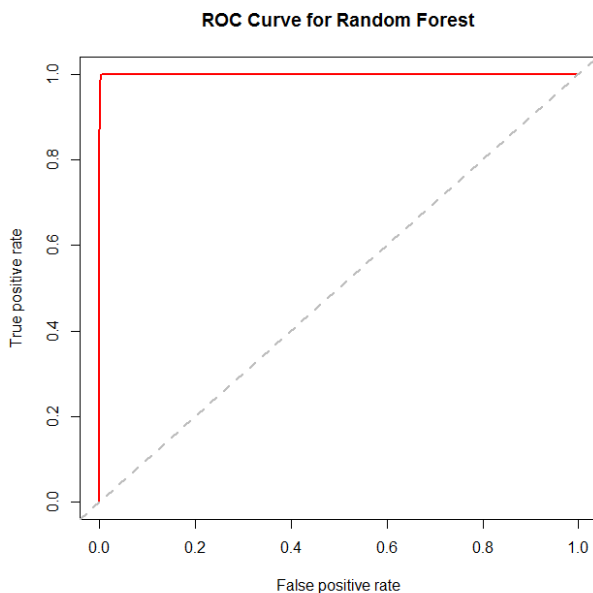
En el gráfico de los errores según el número de árboles, vemos que el error medio es de alrededor de 0.1. Si miramos el gráfico se puede apreciar que a partir de 20 árboles, el error se estabiliza. En este gráfico también podemos observar que el intervalo del error es más pequeño cuantos menos árboles haya.

Fig 6: Gráfico de errores del random forest para empeoramiento dependiendo del número de árboles utilizados.



El área bajo la curva del modelo en la base del test es de 0.99.

Fig 7: Curva ROC del modelo del random forest para empeoramiento en la base del test.



## **Mortalidad**

### **Univariante**

Del total de pacientes, 162 (6.3%) presentaron alguna mejoría al alta  
En esta tabla se encuentran los porcentajes, media o mediana correspondiente a mejoría según si se presenta la característica en cuestión o no.

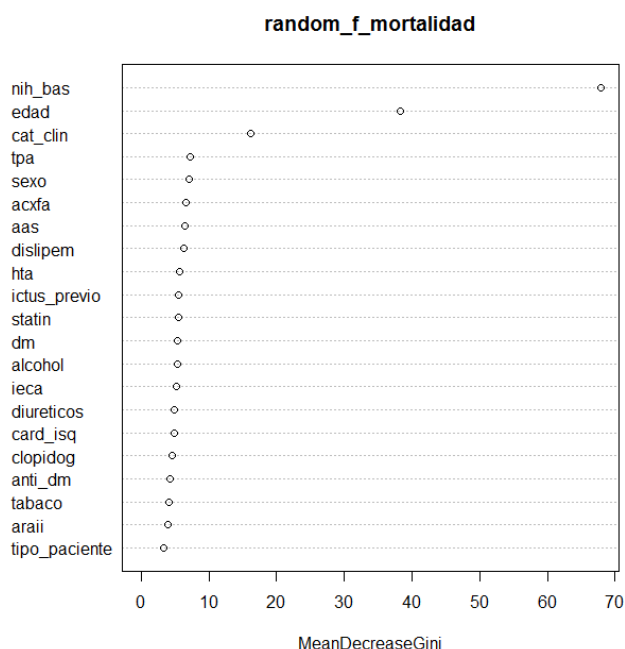
	No	Si	p-valor
<b>Características basales</b>			
-Sexo (hombre)	7.1%	5.7%	0.151
-Edad	69.93±13.29	77.2±10.23	<0.001
-Tipo paciente (isquémico)	5.6%	13.6%	<0.001
-Ictus previo	5.7%	12.3%	<0.001
-Fibrilación auricular	4.7%	13.3%	<0.001
-Cardiopatía isquémica	5.8%	10%	0.004
-Hipertensión	5.1%	7.1%	0.039
-Diabetes	6.4%	6%	0.718
-Tabaco	6.9%	4.1%	0.014
-Alcohol	6.7%	5.1%	0.154
-Dislipemia	6.7%	5.5%	0.24
-Categoría clínica			<0.001
-Cardioembólico		10.6%	
-Aterotrombótico		6.4%	
-Lacunar		2.4%	
-Indeterminado		2.1%	
-Hemorrágico		13.8%	
-NIHSS basal	6 (2-12)	20 (13-22)	<0.001
<b>Medicamentos</b>			
-tpa	5.1%	9.7%	<0.001
-Clopidogrel	6.5%	5.6%	0.499
-AAS	7.2%	4.8%	0.017
-Estatinas	8%	3.4%	<0.001
-Ara II	6.3%	6.7%	0.806
-IECAs	6.5%	5.8%	0.521
-Diuréticos	5.7%	9.3%	0.005
-Anti diabéticos	6.4%	6.1%	0.786

En la tabla anterior se observa que la mortalidad está relacionada con la edad, los ictus isquémicos, los pacientes que tienen fibrilación auricular, ictus previo, cardiopatía isquémica, hipertensión o no son fumadores. También los valores altos de NIHSS al ingreso, la toma de tpa o diuréticos y no tomar aspirina o estatinas.

## Random forest

Según el gráfico de la importancia de las variables incluidas en el modelo del random forest vemos que la edad, NIHSS al ingreso y la categoría clínica son las variables que más aportan al modelo. En este caso, la edad tiene un valor de 38.33, la NIHSS al ingreso 67.87 y la categoría clínica un valor de 16.13. Las otras variables tienen valores por debajo de 10.

Fig 8 Importancia de los factores clínicos y medicamentos según el random forest para mortalidad.



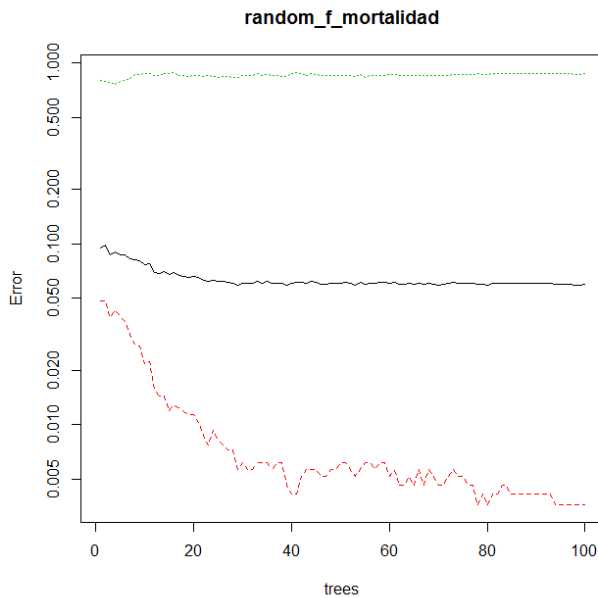
En el gráfico de la importancia del random forest se observa que las variables con más relevancia son la NIHSS al ingreso, la edad y la categoría clínica.

Este modelo también nos enseña el error de clasificación para la base test, que en este caso es de un 0.36% para los pacientes que sobreviven y un 87.3% para los pacientes que fallecen, lo cual indica que puede clasificar mucho mejor la no mortalidad.

	No mortalidad	Mortalidad	Error de clasificación
No mortalidad	1929	7	0.36%
Mortalidad	117	17	87.3%

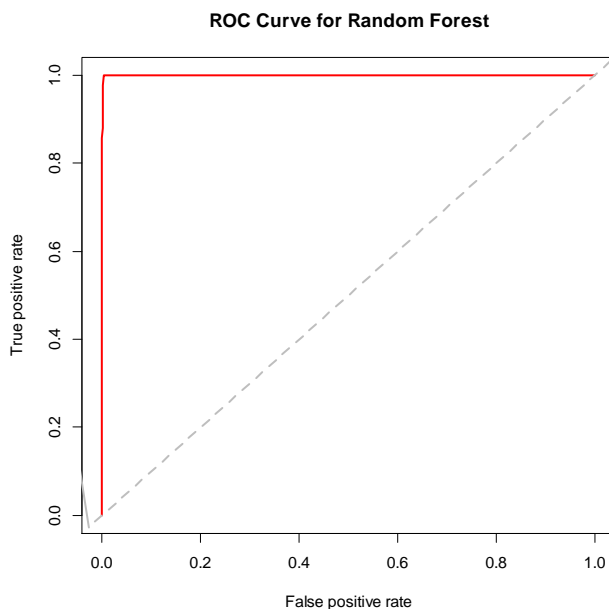
En el gráfico de los errores según el número de árboles, vemos que el error medio es de alrededor de 0.1. Si miramos el gráfico se puede apreciar que a partir de 20 árboles, el error se estabiliza. En este gráfico también podemos observar que el intervalo del error es más pequeño cuantos menos árboles haya.

Fig 9: Gráfico de errores del random forest para mortalidad dependiendo del número de árboles utilizados.



El área bajo la curva del modelo en la base del test es de 0.999.

Fig 10: Curva ROC del modelo del random forest para mortalidad en la base del test.



## Dependencia al 3r mes

### Univariante

Del total de pacientes, 791 (30.8%) presentaron alguna mejoría al alta  
En esta tabla se encuentran los porcentajes, media o mediana correspondiente a mejoría según si se presenta la característica en cuestión o no.

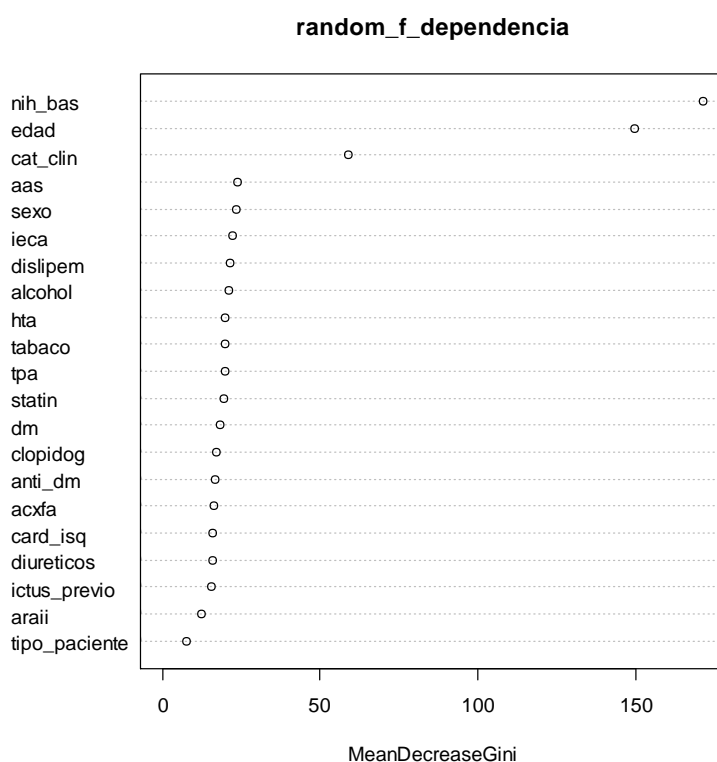
	No	Si	p-valor
<b>Características basales</b>			
-Sexo (hombre)	33%	29.2%	0.041
-Edad	69.57±13.3	72.23±12.91	<0.001
-Tipo paciente (isquémico)	34.7%	34.7%	0.17
-Ictus previo	30.6%	33.1%	0.44
-Fibrilación auricular	28.2%	42.3%	<0.001
-Cardiopatía isquémica	29.7%	39.2%	0.001
-Hipertensión	29.9%	31.5%	0.4
-Diabetes	30.3%	32.6%	0.271
-Tabaco	31.5%	28.5%	0.169
-Alcohol	30.3%	32.6%	0.276
-Dislipemia	31.2%	30%	0.541
-Categoría clínica			<0.001
-Cardioembólico		37.7%	
-Aterotrombótico		32.3%	
-Lacunar		26.2%	
-Indeterminado		24.7%	
-Hemorrágico		34.6%	
-NIHSS basal	5 (2-11)	11 (5-19)	<0.001
<b>Medicamentos</b>			
-tpa	26.6%	42.9%	<0.001
-Clopidogrel	31.2%	28.8%	0.337
-AAS	31.2%	30.2%	0.611
-Estatinas	34.2%	25.1%	<0.001
-Ara II	31%	29.5%	0.641
-IECAs	31.2%	29.8%	0.478
-Diureticos	29.5%	37.4%	0.001
-Anti diabéticos	30.6%	31.6%	0.66

El univariante sobre la dependencia muestra diferencias en el sexo, la edad, la fibrilación auricular, la cardiopatía isquémica, o la etiología cardioembólica. También los valores elevados de NIHSS al ingreso, no tomar estatinas o tomar diuréticos.

## Random forest

Según el gráfico de la importancia de las variables incluidas en el modelo del random forest vemos que la edad, NIHSS al ingreso y la categoría clínica son las variables que más aportan al modelo. En este caso, la edad tiene un valor de 149.57, la NIHSS al ingreso 171.27 y la categoría clínica un valor de 59.01. Las otras variables tienen valores por debajo de 25.

Fig 11 Importancia de los factores clínicos y medicamentos según el random forest para dependencia



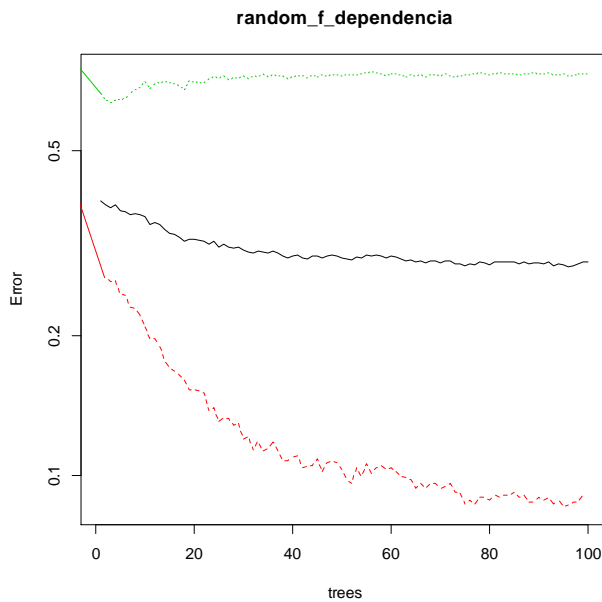
En el gráfico del random forest se observa que las variables con más importancia son la NIHSS al ingreso, la edad y la categoría clínica.

Este modelo también nos enseña el error de clasificación para la base test, que en este caso es de un 9.19% para los que son independientes y un 73.2% para los dependientes, indicando que el modelo clasificará mejor a los pacientes independientes.

	Independencia	Dependencia	Error de clasificación
Independencia	1295	131	9.19%
Dependencia	462	169	73.2%

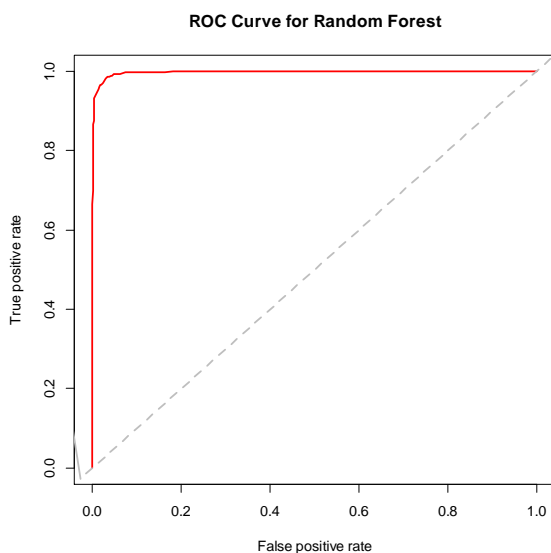
En el gráfico de los errores según el número de árboles, vemos que el error medio es de alrededor de 0.3. Si miramos el gráfico se puede apreciar que a partir de 20 árboles, el error se estabiliza. En este gráfico también podemos observar que el intervalo del error es más pequeño cuantos menos árboles haya.

Fig 12: Gráfico de errores del random forest para dependencia según el número de árboles utilizados.



El área bajo la curva del modelo en la base del test es de 0.997.

Fig 13: Curva ROC del modelo del random forest para dependencia en la base del test.



Los resultados fueron similares cuando se hizo el análisis considerando el training set el 50% de la muestra y el test set, el otro 50%.

### 3.2 Búsqueda de genes

Para encontrar los genes asociados a mejoría se hizo una búsqueda de la literatura en pubmed.

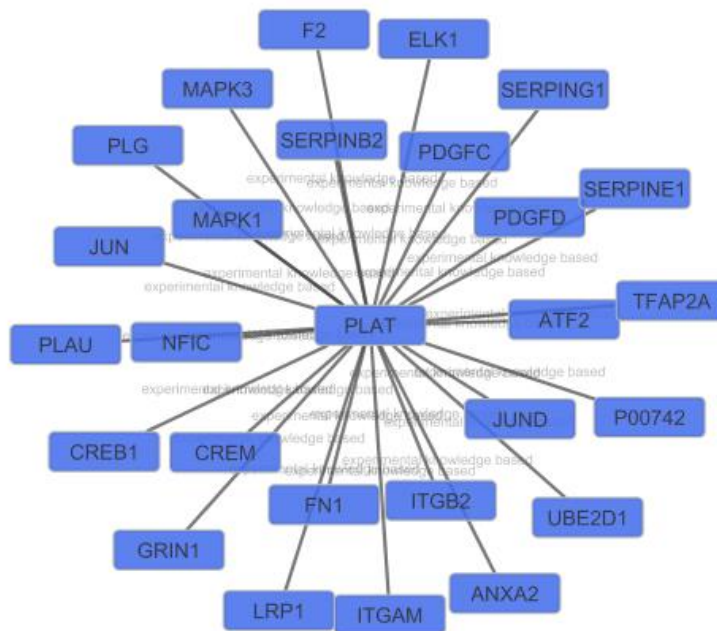
“Prognostic stroke snp”, “Prognostic stroke genetic”, “improvement stroke snp” y “improvement stroke genetic. No se consideraron estudios provenientes de países asiáticos ya que la población estudiada en este trabajo es de origen español.

Un estudio habla de la asociación del polimorfismo crp-717 (rs2794521) de la proteína c-reactiva con mejoría [9]. Para el análisis se consideró la proteína C-reactiva en su totalidad. Su código en el UNIPROT es P02741.

También se encontró la proteína sobre la que actúa el fármaco del tpa, PLAT. Su código en el UNIPROT es P00750.

#### **Red para PLAT**

Fig 14: Red de proteínas para PLAT.

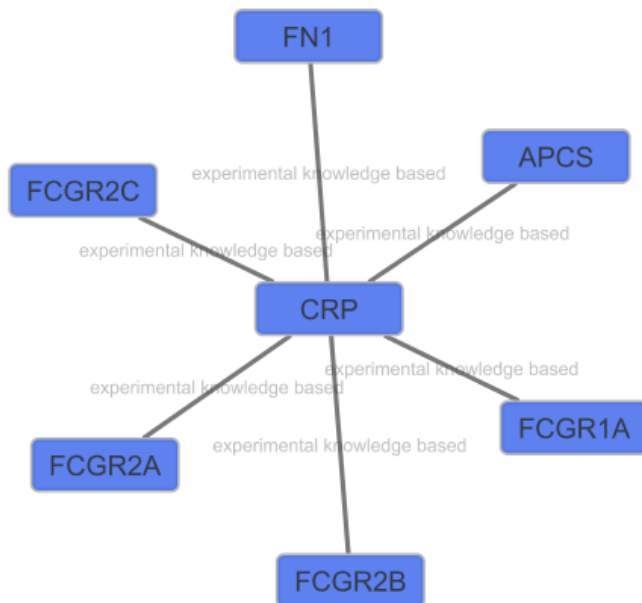


Los nombres de las proteínas se encuentran en la figura anterior. El número de nodos es de 27, el diámetro de la red es de 2, el número medio de los vecinos es de 1.926, la densidad es de 0.074 y la heterogeneidad es de 2.451.



## Red para CRP

Fig 15: Red de proteínas para CRP



LA red de la CRP tiene 7 nodos con sus respectivas proteínas. En este caso la proteína FCG Receptor 1 y receptor 2 (a, b y c), APCS, y la fibronectina (FN1). Esta red tiene diámetro 2, el número medio de vecinos es de 1.714, la densidad de esta red es de 0.286 y su heterogeneidad es de 1.021

## Red conjunta de PLAT y CRP

Fig 16: Red de proteínas para CRP y PLAT



Juntando las dos redes se puede observar que la red de la CRP y la red del PLAT tienen en común la FN1 (fibronectina 1). Su código en el uniprot es p02751.

Esta es una red de diámetro 4 que contiene 33 nodos, con una centralización de 0.8, el número medio de vecinos es de 1.939, su densidad es de 0.061 y la heterogeneidad es de 2.238.

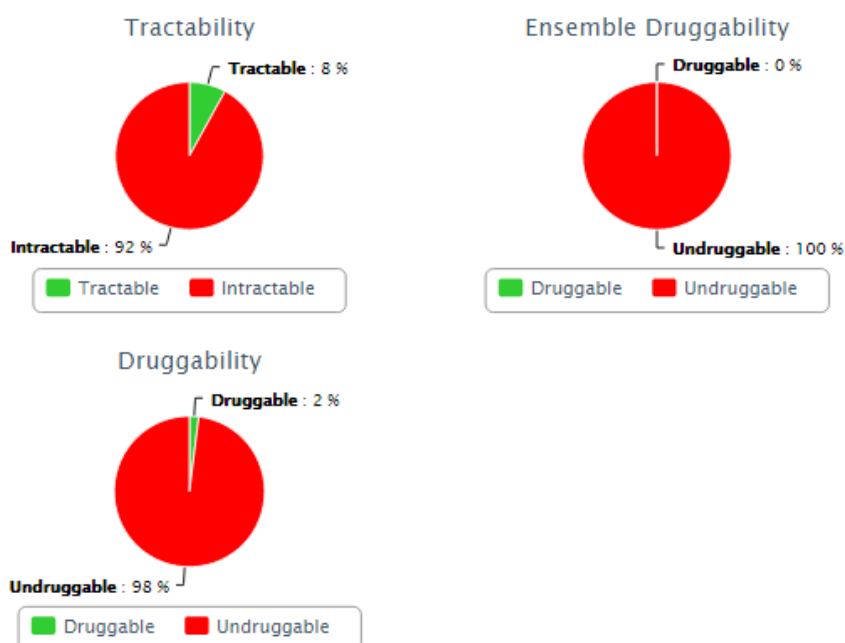
### 3.3 Drug target

Utilizando DrugEBI de la web EMBL-EBI, se buscaron la fibronectina y la crp.

La fibronectina, es una proteina relacionada con la coagulación y no tiene ninguna diana farmacológica aprobada.

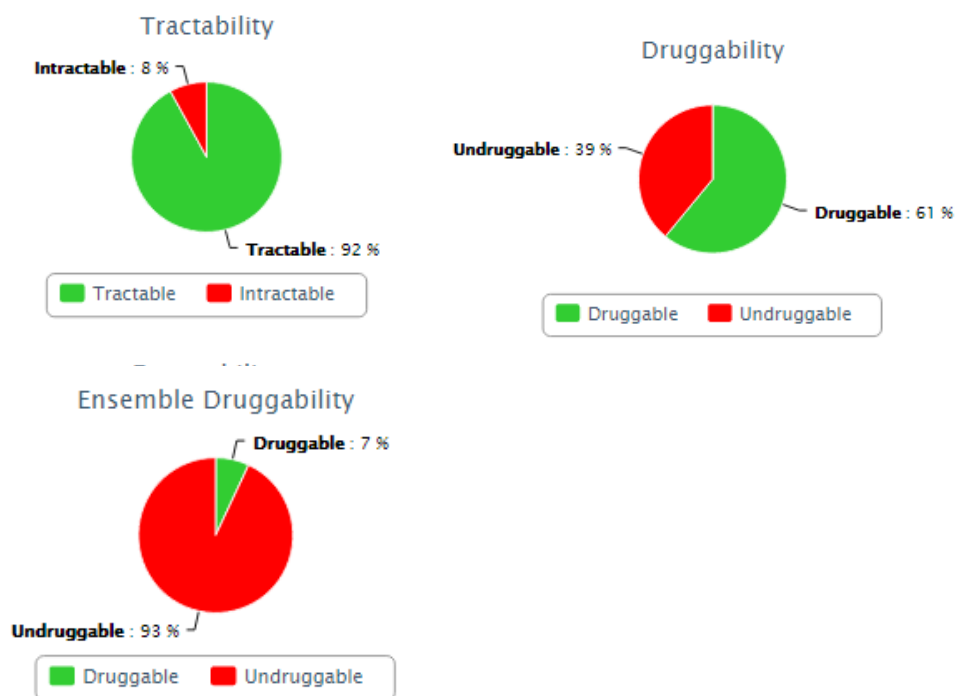
Según su estructura, el 8% de esta proteína es tratable (9 estructuras tienen cavidades farmacológicas, 7 a través del score de tratabilidad y 2 a través del score de drogabilidad). No se encontró ninguna estructura a través del ensamblado.

Fig 17: Drogabilidad para la fibronectina.



La CRP no tiene ninguna diana farmacológica aprobada. Según su estructura, el 92% de esta proteína es tratable (42 estructuras tienen cavidades farmacológicas, 24 a través del score de tratabilidad y 16 a través del score de drogabilidad y 2 por ensamblaje). A parte, un 61% tiene farmacabilidad, pero si lo miramos por ensamblaje, solo sería un 7%.

Fig 18: Drogabilidad para la CRP.



La proteína PLAT no se miró su farmacabilidad dado que esta proteína se encontró a través del tpa.

## 4. Discusión

Los análisis hechos utilizando random forest muestran resultados coherentes. Respecto a mejoría, se sabe que los pacientes jóvenes y los que reciben tpa y dependiendo del tipo de ictus, tienden a mejorar [10]. El tpa está asociado a mejoría, en el univariante también se asocia a mal pronóstico aunque con porcentajes mucho más bajos. Esto puede ser debido a las complicaciones dadas por este medicamento, hay un porcentaje de pacientes (alrededor de un 7%) que, como consecuencia de haberles administrado tpa, tienen como complicación hemorragias cerebrales [11].

Los resultados del random forest para mal pronóstico (empeoramiento, dependencia y mortalidad) también son resultados creíbles ya que la edad y la NIHSS están asociados a ello [12]. También la categoría clínica influye en el mal pronóstico. Los ictus hemorrágicos tienen peor pronóstico y de los isquémicos, los lacunares (de pequeño vaso) tienen un mejor pronóstico ya que son menos severos.

Se ha visto que las redes de la CRP y el PLAT comparten la fibronectina, así se ha construido una sola red teniendo en cuenta estas tres proteínas. También se miró la drogabilidad de la CRP y la fibronectina. Solo la CRP tenía un porcentaje medio de farmacabilidad.

Según A.L. Hopkins y C.R. Groom entre un 10% y un 14% de los genes (aproximadamente unos 3000) pueden ser objeto de la farmacabilidad, también hay alrededor de unos 3000 genes modificados por alguna enfermedad. De la intersección de estos dos grupos de genes se obtendría los medicamentos objetivo (entre 600 y 1500 genes) [13]. Por esta razón es complicado encontrar nuevas dianas terapéuticas, como ha sido el caso de este estudio.

## 5. Conclusiones

En este trabajo se ha visto que los factores mas importantes que influyen en el pronóstico son la edad, la NIHSS y el tipo de ictus, en el caso de mejoría también influye el tpa.

En este trabajo, la intención era encontrar algún medicamento asociado a pronóstico, conseqüentemente, para la búsqueda de genes y la drogabilidad se ha centrado en el apartado de mejoría.

Se hicieron redes de proteínas para la CRP y el PLAT y se vió que compartían la fibronectina.

Posteriormente se miró la drogabilidad de estas proteínas y se encontró que CRP tenía un porcentaje medio de farmacabilidad.

En este trabajo se esperaba encontrar más medicamentos asociados a pronóstico. Por ejemplo, las estatinas, se sabe que están relacionadas con dependencia y aunque en el descriptivo si se relacionaban con el pronóstico, haciendo el random forest no lo han mostrado. Seguramente esto pasa porque tanto la edad como la NIHSS son predictores muy potentes y es complicado encontrar otros factores con una capacidad similar.

También la parte de hacer las redes con las proteínas encontradas y mirar la drogabilidad ha resultado costosa teniendole que dedicar mas tiempo del esperado.

## 6. Glosario

-tpa: activador tisular del plasminógeno (medicamento usado en las 4.5 primeras horas de padecer un ictus isquémico)

-Categoría clínica: clasificación de los ictus según si es ictus hemorrágico o isquémico. Este último, clasificado según sean cardioembólicos, aterotrombóticos, lacunares o indeterminados.

-NIHSS: National institute of Health Stroke Scale. Escala para calcular la gravedad del ictus.

-mRS: Escala de ranking modificada.

-IECAs: inhibidores de la enzima convertidora de angiotensina

-AAS: Acido AcetilSalicilico (aspirina)

-ARA II: antagonistas de los receptores de la angiotensina II

## 7. Bibliografía

1. Tejada J., Maestre J., Laracochea J., Gallego J., Casado I. Guía para el diagnóstico y tratamiento del ictus. Proas Science. Barcelona 2006, pag 65-95.
2. Montaner J. , Alvarez-Sabin J. Escala de Ictus del National Institute of Health (NIHSS) y su adaptación al Español. Neurologia 2006 May;21(4)192-202.
3. Berlet M.H., Stambo G.W., Kelley M., Van Epps K., Woeste T., Steffen D. Does modern ischemic stroke therapy in a large community-based dedicated stroke center improve clinical outcomes? A two-year retrospective study. J Stroke Cerebrovasc Dis. 2014 May-Jun 23 (5)869-78
4. Kablau M., Alonso A., Hennerici M.G., Fatar M. Treatment with tPA predicts better outcome even if MCA occlusion persists. Int J Stroke 2013 Oct; 8(7);496-502.
5. Reuter B., Gumbinger C., Sauer T., Wiethölter H., Bruder I., Rode S., Ringleb P.A., Kern R., Hacke R., Hennerici M.G. Intravenous thrombolysis for acute ischaemic stroke in the elderly: data from the Baden-Wuerttemberg Stroke Registry. Eur. J. Neurol. Volume 23, Issue 1, 13-20.
6. Alvarez-Sabin J., Huertas R., Quintana M., Rubiera M., Delgado P., Ribó M., Molina C.A., Montaner J. Prior statin use may be associated with improved stroke outcome after tissue plasminogen activator. Stroke 2007 Mar; 38 (3) 1076-8.
7. Leigh R., Zaidat O.O., Suri M.F., Lynch G., Sundararajan S., Sunshine J.L., Tarr R., Selman W., Landis D.M., Suarez J.I. Predictors of hyperacute clinical worsening in ischemic stroke receiving thrombolytic therapy. Stroke 2004 Aug; 35 (8); 1903-7
8. Takada T., Yasaka M., Minematsu K., Naritomi H., Yamaguchi T. Predictors of clinical outcome in patients receiving local intra-arterial thrombolysis without subsequent symptomatic intracranial hemorrhage against acute middle cerebral artery occlusion. AJNR Am J Neuroradiol. 2004 Nov-Dec; 25 (10);1796-801.
9. Kotlega D., Bialecka M., Kurzawski M., Drozdik M., Ciercwiez S., Golab-Janowska M., Nowacki P. Risk factors of stroke and -717A>G (rs2794521) CRP gene polymorphism among stroke patients in West Pomerania province of Poland. Neurol Neurochir Pol. 2014 Jan-Feb;48(1)30-4.
10. Naess H., Gjerde G., Waje-Andreassen U. Ischemic stroke in patients older and younger than 80 years. Acta Neurol Scand 2014 Jun; 129(6); 399-404.
11. Montaner J. Stroke biomarkers: Can they help us to guide stroke thrombolysis? Drug news Perspect 2006 Nov; 19(9) 523-32.



12. Appelros P., Terént A. Characteristics of the National Institute of Health Stroke Scale. Results from a population-based stroke cohort at baseline and after one year. *Cerebrovascular disease* 2004; 17(1)21-7.
13. Hopkins A.L., Groom C.R. The druggable genome. *Nat Reviews Drug Disc* 2002 Sep;1(9)727-30.