

**“Influencia de la enfermedad COVID-19 y sus
tratamientos en los eventos tromboembólicos: Algoritmo
terapéutico”**

Versión 2.0, de fecha 9 de febrero de 2021

INDICE

EQUIPO INVESTIGADOR	3
RESUMEN EJECUTIVO	4
1. INTRODUCCIÓN	7
Referencias	12
2. HIPÓTESIS	14
3. OBJETIVOS	14
Objetivo principal	14
Objetivos secundarios	14
4. MATERIAL Y MÉTODOS	15
Periodo del estudio	15
Recopilación de los datos	15
Aspectos éticos: Aprobaciones por Comités Éticos locales	16
Análisis de los datos	16
Variables analizadas	17
5. OBJETIVOS CIENTÍFICOS Y TECNOLÓGICOS E	22
INNOVACIONES TECNOLÓGICAS DEL PROYECTO	
6. DESCRIPCIÓN TÉCNICA Y PLAN DE TRABAJO	28
7. CAPACIDAD DEL EQUIPO DE INVESTIGACION	36

EQUIPO INVESTIGADOR

Coordinadores del estudio:

- Joaquín de Haro Miralles. Angiología y Cirugía Vascular. Hospital Universitario de Getafe. Madrid.
- José Antonio Diaz. Director of the Division of Surgical Research. Vanderbilt University Medical Center. Nashville (Tennessee, EEUU)
- Javier Rodríguez Padilla. Angiología y Cirugía Vascular. Hospital HM Montepríncipe y HM Puerta del Sur. Madrid.
- Enrique M. San Norberto. Angiología y Cirugía Vascular. Hospital Clínico Valladolid.

Investigadores Principales:

- Guillermo Moñux (Hospital HM Torrelodones, Madrid)
- Rodrigo Rial (Hospital HM Torrelodones, Madrid)
- Noelia Alonso (Hospital Gómez Ulla, Madrid)
- Fran Alvarez (Hospital Universitario Central Asturias, Oviedo)
- Antonio Romera Villegas (Hospital Universitario Bellvitge, Barcelona)

El presente estudio se encuentra englobado en la red de estudios colaborativos de la SEACV y el CEFyL a través de la RIV (Red de Investigación Vascular).

RESUMEN EJECUTIVO

El COVID-19 supone un importante problema de salud pública a nivel mundial. Los datos recogidos actualmente sugieren que el nuevo coronavirus (SARS-CoV-2) es altamente contagioso (tasa de reproducción estimada, 2-3) y con una tasa estimada de mortalidad de entre el 0.5-5%, afectando esta mortalidad fundamentalmente a personas de ≥ 65 años. Son necesarias intervenciones que disminuyan tanto la incidencia como la severidad del COVID-19. Aunque comparte características con otros virus similares que también surgieron en brotes epidémicos, aún no se conocen por completo los mecanismos fisiopatológicos del virus ni qué respuesta provoca sobre el huésped. Existen indicios que el cuadro clínico de esta enfermedad cursa con un estado procoagulante, con posible aumento de los episodios de enfermedad tromboembólica (ETV). El presente estudio pretende analizar la influencia de COVID-19 sobre la incidencia de trombosis venosa profunda (TVP) en miembros inferiores y superiores, y la variación de la presentación clínica de esta, así como aportar nuevas evidencias aplicables al manejo clínico de estos pacientes y al establecimiento de factores pronósticos que ayuden a la toma precoz de decisiones terapéuticas. En especial, se focalizará en la hipótesis inflamatoria de la etiopatogenia de la enfermedad tromboembólica, basado en un importante *background* bibliográfico en el que han contribuido de forma sustancial los investigadores del grupo de investigación de la Red de Investigación Vascular (RIV) que lideran este proyecto.

Para ello, se llevará a cabo un estudio de cohortes con la información clínica recogida en los registros digitales que HM Hospitales pondrá a disposición de los investigadores y una vez el proyecto sea aprobado, los datos de 2.157 Historias clínicas anonimizadas que recogen las distintas interacciones en el proceso de tratamiento del COVID-19, incluyendo información pormenorizada sobre diagnósticos, tratamientos, ingresos, pasos por UCI y alta o deceso, pruebas diagnósticas de imagen y resultados de laboratorio, englobado en el Proyecto 'COVID DATA SAVE LIVES' promovido por HM Hospitales.

Se creará un Sistema Computacional (TVP-COVID) para la predicción de la influencia de fenómenos inflamatorios y de los tratamientos actuando en vías específicas de

la inflamación sobre la patogenia de enfermedad tromboembólica, así como para la predicción del algoritmo terapéutico más eficaz para evitar y tratar eventos tromboembólicos en estos pacientes. Este Sistema Computacional será una herramienta capaz de realizar predicciones sobre la incidencia de un caso de TVP en urgencias de un hospital e incluso el tratamiento/profilaxis más adecuada específico para ese caso. Este modelo predictivo permitirá, a su vez, trabajar en el desarrollo de nuevas soluciones que faciliten el trabajo de los médicos a la hora de tomar decisiones clínicas que beneficien a los pacientes con esta letal enfermedad.

A través de la aplicación de tecnologías de Inteligencia Artificial (IA) y Big Data se procesarán tanto datos estructurados como no estructurados de la historia clínica de los pacientes de las instituciones sanitarias adheridas al proyecto, con el objetivo de la implementación del sistema computacional complejo para convertirlos en conocimiento de valor para los profesionales del hospital. De esta manera, dispondremos de herramientas más completas de análisis avanzado a la hora de tomar decisiones clínicas. Así mismo se recogerán datos de fuentes externas que serán evaluadas como potenciales factores de riesgo. El modelo de IA debe proporcionar una predicción probabilística de la aparición de un caso de TVP-ETV en un paciente con enfermedad de COVID-19.

La finalidad de este proyecto es avanzar en el conocimiento de este virus, ya que con esos datos clínicos y mediante inteligencia artificial y la investigación de expertos, se pueden obtener modelos predictivos de evolución, modelos epidemiológicos, información sobre la respuesta a los diversos tratamientos aplicados, conocimientos sobre el comportamiento del virus sobre el sistema vascular y datos sociodemográficos sobre el impacto en la población del virus.

Este proyecto tiene carácter nacional, auspiciado por la Sociedad Española de Angiología y Cirugía Vascular (SEACV) y el Capítulo Español de Flebología y Linfología (CEFyL), a través de su Red de Investigación Vascular (RIV), en el que se analizarán variables demográficas, de comorbilidad, de tratamiento concomitantes, estado analítico y de

pruebas complementarias y ecográficas diagnósticas, parámetros de evolución clínica, terapéuticos y de complicaciones y mortalidad a 30 días.

1. INTRODUCCIÓN

El nuevo coronavirus (COVID-19 o SARS-CoV-2) ^[1] supone un importante problema de salud pública. Tras diagnosticarse los primeros casos en diciembre de 2019 en Wuhan, el número de casos ha ido creciendo exponencialmente, estando hasta la fecha registrados más de un millón de casos a nivel mundial con algo más de 200.000 decesos ^[2]. La tasa de mortalidad del SARS-CoV-2 (3,8%) ^[3] es inferior a la del coronavirus del síndrome respiratorio agudo grave (SARS-CoV) (10%) ^[4] o a la del coronavirus del síndrome respiratorio de Oriente Medio (MERS-CoV) (37,1%), pero presenta una capacidad de contagio 10 veces mayor que los anteriores. Esta alta capacidad de contagio unida a que el 20% de los infectados van a desarrollar un cuadro severo de la enfermedad explica el alto número de muertos que se están registrando. Son por tanto necesarias intervenciones que disminuyan tanto la incidencia como la severidad del COVID-19.

Los coronavirus son miembros de la subfamilia *Coronavirinae*, la cual consta de cuatro géneros: Alfacoronavirus, Betacoronavirus, Gammacoronavirus y Deltacoronavirus. Los Alfa y Betacoronavirus infectan únicamente mamíferos y causan enfermedades intestinales y respiratorias en animales y humanos.⁵ No se consideraban altamente patogénicos hasta el brote de 2002 del síndrome respiratorio agudo severo por coronavirus (SARS-CoV), dado que los que circulaban anteriormente en humanos causaban infecciones leves y afectaban principalmente a personas inmunocomprometidas. Al igual que el virus SARS-CoV y el coronavirus que causa el síndrome respiratorio de Oriente Medio (MERS-CoV), surgió 10 años después que el SARS, el SARS-CoV-2 es un virus con envuelta, cuyo material genético es ARN monocatenario de polaridad positiva y pertenece a la familia de los Betacoronavirus. El genoma del SARS-CoV-2 codifica proteínas no estructurales, proteínas estructurales (como la glicoproteína de las espículas) y proteínas accesorias, Las proteínas no estructurales son enzimas clave en el ciclo de vida del virus. Las espículas glicoproteicas, por su parte, son indispensables en la interacción virus-huésped durante la fase de entrada del virus al interior celular.⁶

En comparación con el SARS-CoV, el SARS-CoV-2 tiene una capacidad de transmisión mucho mayor y un periodo de incubación más prolongado, con una media de 5,2 días de incubación y 20 días para la liberación de los viriones en pacientes que superan la infección,⁴⁻⁹, aunque su tasa de mortalidad parece ser inferior.¹⁰ Entre las manifestaciones clínicas más comunes producidas por el SARS-CoV-2 se encuentran: fiebre (98% de los pacientes afectados), tos (76%), disnea (55%), astenia y mialgias (44%), siendo menos comunes la expectoración (28%), cefalea (8%), hemoptisis (5%) y los síntomas gastrointestinales (3%).¹¹ Aunque la mayor parte de los síntomas de la infección por SARS-CoV-2 se asemejan a los que producen SARS-CoV, MERS-CoV e influenza estacional, existen diferencias debidas al tropismo preferente del SARS-CoV-2 por las células del tracto respiratorio inferior, lo que hace menos frecuente la aparición de rinorrea, estornudos o faringitis.^{10,11} A pesar de la posible aparición de complicaciones que pueden comprometer la vida del afectado (síndrome de distrés respiratorio agudo, lesión cardíaca o shock), se ha descrito también el curso asintomático de la infección,^{12,13} lo que supone un factor de riesgo para el contagio, puesto que individuos asintomáticos y portadores podrían diseminar la enfermedad de manera indetectable.

- **Incidencia de eventos tromboembólicos en COVID-19**

Existe una falta de datos fiables sobre la incidencia de ETV en pacientes con COVID-19. Solo se han encontrado datos de dos descripciones de casos y una cohorte retrospectiva pequeña en proceso editorial, sobre todo de episodios de embolia pulmonar (EP). Una de las publicaciones describe un caso de una mujer de 75 con COVID-19 y neumonía bilateral a la que se le encontró una EP bilateral en una exploración con TC ^[14]. La paciente fue hospitalizada después de haber manifestado sintomatología durante 10 días y tener un inicio de disnea reciente. El hallazgo se realizó a pesar de que la paciente no tenía factores de riesgo remarcables y estaba hemodinámicamente estable (no obstante, el estudio describe una leve leucocitosis (11.360/mm²) y valores elevados de proteína C reactiva (180 mg/L), troponinas I 3240,4 ng/mL, y D-dímero (21 microg/mL). En otro estudio de caso único ^[15], se recogen los casos de dos hombres (57 y 70 años) en los que se diagnosticó una EP bilateral en una exploración con angiografía pulmonar por TC sin que se dieran más detalles clínicos sobre los pacientes.

En un manuscrito en prepublicación en *The Lancet Infectious Diseases* se describe una serie de 1008 pacientes de hospitales de Wuhan con una neumonía por COVID-19 de los que se hizo una revisión de sus historias clínicas y entre los que se identificó a 25 pacientes

a los que se realizó una evaluación con angiografía pulmonar por TC por sospecha de EP aguda (media de edad de 65 años [rango 36 a 78 años]) ^[16]. Este estudio describe los principales hallazgos analíticos anormales en estos pacientes con sospecha de ETV. En todos los pacientes con sospecha se detectó una elevación del D-dímero (mediana de 6,06 microgr/L, rango IQ 1,90 a 14,31 microgr/L), niveles que fueron significativamente más grandes que en los 10 pacientes con confirmación de ETV (mediana de 11,07 microgr/L, rango IQ 7,12 a 21,66 microgr/L) que en los 15 que no tuvieron un hallazgo en la evaluación con TC (mediana de 2,44 microgr/L, rango IQ 1,68 a 8,34 microgr/L; $P < 0,05$). El estudio no describe ninguna otra característica analítica que mostrara diferencias significativas entre los pacientes con o sin confirmación de ETV.

Para la resolución de cualquier infección, la respuesta inmune es esencial. No obstante, la propia respuesta inmune podría exacerbar la inmunopatogénesis del virus. Así pues, la progresión hacia síndrome de distrés respiratorio en infecciones por SARS-CoV parece estar asociada a un aumento en los niveles plasmáticos de ciertas citoquinas y quimioquinas pro-inflamatorias, particularmente PCR, IL-1 β , IL-8, IL-6, CXCL10 y CCL2, así como el Dímero-D. Algunos estudios realizados de forma retrospectiva han demostrado que pacientes con SARS-CoV que sucumben a la infección, mantienen niveles elevados de estas citoquinas.⁸ Es más, recientemente, estudios¹⁷ han puesto de manifiesto la importancia de la monitorización de la tormenta de citoquinas secundaria a la infección por SARS-CoV-2, dado que, en ocasiones, puede ser la causa subyacente del empeoramiento y fracaso de estos pacientes. Huang et al.¹³ estudiando 41 pacientes con SARS-CoV-2 también identificaron que, aquellos que necesitaron ser ingresados en las unidades de cuidados intensivos, tenían niveles plasmáticos más elevados de citoquinas pro-inflamatorias y Dímero-D. De hecho, citoquinas pro-inflamatorias, como la IL-6, o supresoras, como la IL-38, podrían ser potenciales dianas terapéuticas para paliar los efectos patológicos derivados de la respuesta inflamatoria.

La irrupción de este virus especialmente contagioso y patogénico está promoviendo el desarrollo de múltiples investigaciones a nivel mundial dirigidas a comprender sus mecanismos fisiopatológicos y buscar posibles tratamientos. No obstante, se desconoce las implicaciones fisiopatológicas que pueda tener en el desarrollo de eventos tromboembólicos en los miembros inferiores y superiores.

- **La interleucina 6 y la trombosis venosa profunda**

La interleucina 6 (IL-6) es una glucoproteína secretada por los macrófagos, células T, células endoteliales y fibroblastos. Es una citocina con actividad antiinflamatoria y proinflamatoria. La IL-6 estimula en la hipófisis la producción de ACTH, interviene en la diferenciación de linfocitos B, activa a los linfocitos T (particularmente los citotóxicos), células plasmáticas, regula la hematopoyesis y es la citocina líder en la síntesis de proteínas de fase aguda hepática, en especial fibrinógeno, acción conjunta con otra citocina, la IL-1. En otras palabras, la IL-6 es un líder inflamatorio.

La trombosis venosa profunda (TVP) es una entidad patológica con gran impacto en nuestra sociedad. Estadísticas de Estados Unidos demuestran que casi 1,000,000 de personas desarrollan TVP en un año y casi 300,000 fallecen en el mismo periodo a causa de embolismo pulmonar (EP) y otras complicaciones del tratamiento. De los sobrevivientes, unos 600,000 pacientes, el 50% de las TVPs se relacionan con un factor de riesgo como cáncer o cirugía por mencionar algunos. Más interesante es que el otro grupo de 50% no se asocia a ningún factor de riesgo y en este grupo el 50% desarrolla síndrome post-trombótico (SPT), una dolencia que deteriora profundamente la calidad de vida y le cuesta al sistema de salud de Estados Unidos 2 billones de dólares por año.

Inflamación y la TVP están relacionadas ¹⁸, esta asociación empieza a ser reconocido por la comunidad científica el instituto nacional de salud de los Estados Unidos (NIH). Nuestro grupo viene estudiando esta relación entre inflamación y DVT en la clínica ¹⁹ y el laboratorio usando modelos animales ^{20, 21}. Particularmente, la relación entre TVP, la IL-6 y su inhibición fue estudiada en detalle y un modelo biológico fue sugerido basado en los resultados experimentales ²². Estos estudios demuestran que la inflamación es un factor determinante en la trombosis venosa profunda. Una representación esquemática de los potenciales eventos que suceden en TVP se muestra en la Fig-1 ²³.

El mecanismo de iniciación de la TVP propuesto en este modelo biológico está basado en hallazgos de trombosis experimental.

Inflamación de inicio: Un proceso inflamatorio inicial, liderado por la IL-6 (entre otras), y la supresión del sistema fibrinolítico liderado por la secreción de PAI-1 como eventos detectados muy temprano ²⁴. La entrada de calcio en la célula endotelial del área donde se inicia el proceso estimula la aproximación de los cuerpos de Weibel-Palade a la membrana celular. Esto determina la exocitosis del factor de von-Willebrand (vWF) y la expresión de Selectina-P. vWF promueve en la luz su interacción con las plaquetas y micropartículas repletas de Factor tisular lo cual favorece la coagulación. La Selectina-P, por su parte, promueve la adherencia e infiltración de los leucocitos polimorfo-nucleares, dentro de la pared venosa, lo que internaliza en la pared venosa la reacción inflamatoria y promueve la formación de Traps extracelulares derivadas de neutrófilos (NETs).

Inflamación de propagación: De aquí en adelante, un proceso inflamatorio secundario contribuye con la cadena secuencial de eventos que terminan la remodelación de la pared venosa que finalizara en la fibrosis de la pared. Esta remodelación incluye la producción de CCL2 y la atracción de los monocitos evidenciado por el cambio de la fase aguda o neutrofílica en crónica o monocítica ²⁵.

La evidencia que se está obteniendo con la pandemia, COVID19, relaciona la IL-6 con la TVP. La tormenta de citosinas liderada por la IL-6 y la secuencia de pacientes iniciando la coagulación en los sistemas arterial y venoso requerirán estudios y confirmaciones, pero no deberíamos descartar lo aprendido. La formulación de un modelo biológico como el expuesto y publicado, justifica un pensamiento en esta dirección. La hipótesis del rol de la IL-6 promoviendo la TVP queda planteada y esta pandemia puede ser motivo de estudio en esta dirección.

Referencias

1. Wu, Y., et al., *SARS-CoV-2 is an appropriate name for the new coronavirus*. Lancet, 2020.
2. Du Toit, A., *Outbreak of a novel coronavirus*. Nat Rev Microbiol, 2020. **18**(3): p. 123.
3. Li, G. and E. De Clercq, *Therapeutic options for the 2019 novel coronavirus (2019-nCoV)*. Nat Rev Drug Discov, 2020. **19**(3): p. 149-150.
4. Zheng, Y.Y., et al., *COVID-19 and the cardiovascular system*. Nat Rev Cardiol, 2020.
5. Cui, J., F. Li, and Z.L. Shi, *Origin and evolution of pathogenic coronaviruses*. Nat Rev Microbiol, 2019. **17**(3): p. 181-192.
6. Zumla, A., et al., *Coronaviruses - drug discovery and therapeutic options*. Nat Rev Drug Discov, 2016. **15**(5): p. 327-47.
7. Turner, A.J., J.A. Hiscox, and N.M. Hooper, *ACE2: from vasopeptidase to SARS virus receptor*. Trends Pharmacol Sci, 2004. **25**(6): p. 291-4.
8. de Wit, E., et al., *SARS and MERS: recent insights into emerging coronaviruses*. Nat Rev Microbiol, 2016. **14**(8): p. 523-34.
9. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein*. Cell, 2020.
10. Li, Q., et al., *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia*. N Engl J Med, 2020.
11. Zhou, F., et al., *Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study*. Lancet, 2020.
12. Guan, W.J., et al., *Clinical Characteristics of Coronavirus Disease 2019 in China*. N Engl J Med, 2020.
13. Huang, C., et al., *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China*. Lancet, 2020. **395**(10223): p. 497-506.
14. Danzi GB, Loffi M, Galeazzi G, Gherbesi E. Acute pulmonary embolism and COVID-19 pneumonia: a random association? Eur Heart J. 2020 Mar 30. pii: ehaa254. doi: 10.1093/eurheartj/ehaa254.
15. Xie Y, Wang X, Yang P, Zhang S. COVID-19 Complicated by Acute Pulmonary Embolism. Radiology: Cardiothoracic Imaging 2020 2:2 <https://doi.org/10.1148/ryct.2020200067>
16. Chen J, Wang X, Zhang S, Liu B, Xiaoqing W, Wang Y, Wang X, Yang M, Jianqing S, Yuanliang X. Findings of Acute Pulmonary Embolism in COVID-19 Patients (3/1/2020). Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3548771>
17. Chen, L., et al., *[Analysis of clinical features of 29 patients with 2019 novel coronavirus pneumonia]*. Zhonghua Jie He He Hu Xi Za Zhi, 2020. **43**(3): p. 203-208.
18. Trombosis venosa profunda | National Heart, Lung, and Blood Institute (NHLBI).
19. Rodriguez AL, Wojcik BM, Wroblewski SK, Myers DD, Wakefield TW, Diaz JA. Statins, inflammation and deep vein thrombosis: A systematic review. J Thromb Thrombolysis 2012; 33.
20. DeRoo EP, Wroblewski SK, Shea EM, Al-Khalil RK, Hawley AE, Henke PK, Myers DD, Wakefield TW, Diaz JA. The role of galectin-3 and galectin-3-binding protein in venous thrombosis. Blood 2015; 125.
21. Diaz JA, Wroblewski SK, Alvarado CM, Hawley AE, Doornbos NK, Lester PA, Lowe SE, Gabriel JE, Roelofs KJ, Henke PK, Schaub RG, Wakefield TW, Myers DD. P-Selectin inhibition therapeutically promotes thrombus resolution and prevents vein wall fibrosis better than enoxaparin and an inhibitor to von willebrand factor. Arterioscler Thromb Vasc Biol Lippincott Williams and Wilkins; 2015; 35: 829–37.

22. Wojcik BM, Wroblewski SK, Hawley AE, Wakefield TW, Myers Jr. DD, Diaz JA. Interleukin-6: A potential target for post-thrombotic syndrome. *Ann Vasc Surg* 2011; 25.
23. Diaz, Jose Antonio; Henke P. Venous Pathophysiology. *Rutherford's Vascular Surgery and Endovascular Therapy*. 2019. p. Section 1, Chapter 9, 97-104.
24. Patterson KA, Zhang X, Wroblewski SK, Hawley AE, Lawrence DA, Wakefield TW, Myers DD, Diaz JA. Rosuvastatin reduced deep vein thrombosis in ApoE gene deleted mice with hyperlipidemia through non-lipid lowering effects. *Thromb Res* 2013; 131.
25. Diaz JA, Alvarado CM, Wroblewski SK, Slack DW, Hawley AE, Farris DM, Henke PK, Wakefield TW, Myers Jr. DD. The electrolytic inferior vena cava model (EIM) to study thrombogenesis and thrombus resolution with continuous blood flow in the mouse. *Thromb Haemost* 2013; 109.

2. HIPÓTESIS

Existe una necesidad urgente de compilar datos clínicos y analíticos de los pacientes infectados con COVID-19 con sospecha de evento tromboembólico en los miembros inferiores y superiores para mejor comprender los mecanismos fisiopatológicos del virus con respecto a esta enfermedad. Obtener datos del mundo real y compartir experiencias nacionales e internacionales ayudará a tomar mejores decisiones sobre el manejo de este complejo grupo de pacientes, tanto en su profilaxis como en su tratamiento, mejorando su atención clínica.

El estudio, de carácter no comercial, no intervencionista es un estudio observacional de bajo nivel de intervención, con un riesgo extremadamente bajo. Este estudio no recopilará ninguna información identificable del paciente (sin incluir fechas) y los datos no se analizarán a nivel hospitalario. Se encuentra auspiciado por la Sociedad Española de Angiología y Cirugía Vascul ar (SEACV) y el Capítulo Español de Flebología y Linfología (CEFyL), englobado dentro de la Red de Investigación Vascul ar (RIV).

3. OBJETIVOS

Objetivo principal

Determinar la incidencia de TVP en miembros superiores e inferiores, factores predictores, asociación con marcadores inflamatorios y sus manifestaciones clínicas y pronósticas propias en pacientes con infección por COVID-19.

Objetivos secundarios

- Mortalidad a los 7 y 30 días.
- Ingreso en la UCI a los 7 y 30 días.
- Régimen de Tratamiento anticoagulante previo al episodio
- Embolia Pulmonar concomitante
- Síndrome de dificultad respiratoria aguda postoperatoria (SDRA).

4. MATERIAL Y MÉTODOS

Criterios de inclusión

- Adultos (edad ≥ 18 años) considerados como caso de enfermedad COVID-19 con sospecha de TVP en miembros inferiores y/o superiores
 - Se considera caso: (i) una prueba de laboratorio que confirme la infección por COVID-19 o (ii) un diagnóstico clínico de infección por COVID-19 (sin realizar ninguna prueba).
- Se considera sospecha de TVP:
 - Pacientes con datos clínicos de sospecha de TVP
 - y/o pacientes con datos analíticos de sospecha de TVP
 - y/ o pacientes con datos ecográficos de sospecha de TVP

Los pacientes que cumplen con los criterios se incluirán independientemente del desenlace clínico.

Periodo de estudio

Estudio retrospectivo con entrada de datos agregados hasta septiembre de 2020; sin embargo, dependiendo de la disponibilidad de la cesión de datos se pueden seleccionar las ventanas de estudio adecuadas, de forma consensuada con Hospitales HM, atendiendo a la implementación del proyecto “COVID DATA SAVE LIVES”, y dependiendo del momento de la epidemia de COVID-19.

Recopilación de datos

Los datos agregados se encuentran en una base de datos absolutamente anonimizada recopilada por HM hospitales, cumpliendo con la normativa vigente del la Ley Orgánica de Protección de Datos - 3/2018 - LOPDGDD y Ley de Protección de Datos y Garantía de Derechos Digitales (LOPDGDD) 2018. El análisis de datos se realizará en línea a través de un servidor seguro en una aplicación web facilitada por la RIV de la Sociedad Española de

Angiología y Cirugía Vascul ar (SEACV) y el Capítulo Español de Flebología y Linfología (CEFyL). Dicha aplicación permite gestionar los datos datos anonimizados agregados en un sistema seguro. Los investigadores recibirán la contraseña de inicio de sesión del servidor del proyecto COVID DATA SAVE LIVES-RIV, lo que les permitirá analizar los datos de forma segura en el sistema.

Aspectos éticos: Aprobaciones por Comités Éticos locales

Este estudio será evaluado para su aprobación por el CEIm

Análisis

Se redactará un plan de análisis estadístico detallado. Los informes incluirán una descripción de los resultados primarios y secundarios en la cohorte. Se realizarán modelos multivariantes para identificar los factores de riesgo, predictivos, pronósticos y mortalidad a los 7 y 30 días. Los análisis se estratificarán según si el paciente con COVID-19 presentaba antecedentes previos de alteración de la coagulación o antecedentes protrombóticos y por niveles de marcadores inflamatorios.

El equipo investigador declara su compromiso de enviar datos para su publicación. En las posibles publicaciones derivadas de la presente investigación el orden de autores, además del coordinador de la investigación y el/los autor/es del manuscrito, vendrá definido por el equipo investigador. Asimismo, se hará constar en todas las publicaciones que el estudio se engloba en el proyecto “COVID DATA SAVE LIVES” promovido por Hospitales HM.

Variables analizadas

Demográficas y de comorbilidad

- 1.- Número de identificación
- 2.- Momento del diagnóstico:
 - Diagnóstico laboratorio COVID 19+
 - Diagnóstico clínico COVID 19+
- 3.- Fecha del diagnóstico o sospecha de TVP
- 4.- Edad
- 5.- Sexo
- 6.- Peso
- 7.- Altura
- 8.- Comorbilidad:
 - HTA
 - DM
 - DL
 - Fumador
 - EPOC
 - Cardiopatía
 - Insuficiencia renal crónica
 - Insuficiencia cardiaca congestiva
 - Cardiopatía isquémica
 - Enfermedad vascular periférica
 - Ictus / AIT
 - Otras (texto libre)
9. Tratamiento previo:
 - IECAs
 - ARA-II
 - Diuréticos
 - Betabloqueantes
 - Antagonistas del Ca
 - Anticoagulantes
 - Antiagregantes

- Estatinas
- Corticoides
- Inmunosupresores (ciclosporina, tacrolimus, sirolimus, azatioprina, mercaptopurina)
- Otros inmunosupresores

10. – Antecedentes procoagulantes:

- Cáncer
- Alteración genética coagulación
- Terapia hormonal
- Encamamiento

Diagnóstico TVP

11.- Clínico

- Diámetro >3cm respecto contralateral
- Aumento trama venosa superficial
- Dolor a la palpación territorio venoso profundo
- Empastamiento gemelar
- Edema con fóvea
- Parálisis, paresia extremidad
- TVP previa
- Diagnostico alternativo
- Total Wells (nº)
- TVP (0=no 1=si)

12.- Analítico: D-Dimero

13.- Ecográfico:

- Afecta Cava
- Afecta iliaca
- Afecta femoral
- Afecta poplítea
- Afecta gemelar
- Afecta TP, TA o peronea

Situación COVID-19

14.- Confusión (estado mental alterado o empeorado): Si/No.

15.- Ventilación (0= no / 1=O2 / 2=VMK 26% /3=VMK 35%/4=VMK 50% /5=AltoFlujo / 6=VMNI/ 7=IOT)

16.- HBPM

17.- Neumonía

18.- TEP

19.- PAFi

20.-Días clínica (0 no clínica / resto ->nº)

21.- Tratamiento COVID-19

- Antibiótico (nombre)
- Antiviral (nombre)
- Quinina o derivados (nombre)
- Corticoesteroides: (nombre y dosis máxima)
- Inmunoglobulinas intravenosas (nombre)
- Interferon (nombre)
- Bloqueantes IL-6 (nombre)
- Soporte respiratorio:
 - o Ninguno
 - o Oxígeno a bajo flujo
 - o Oxígeno a alto flujo
 - o Ventilación no invasiva
 - o Ventilación invasiva
 - o Oxigenación con membrana extracorpórea (ECMO)
- Duración de la ventilación mecánica invasiva:
 - o 1-23 horas
 - o 24-47 horas
 - o 48-71 horas
 - o 72-167 horas
 - o >167 horas
- Prono

22.- Marcadores inflamatorios:

- IL-6
- TNF
- D-dimer HS (ng/mL)
- LDH FG
- Ferritina

- CK
- Linfos x10⁹
- Linfos -> 0 -> x2 (aumento) // 1 -> -2 (descenso)
- Neutros x10⁹
- Índice Neutrófilo/Linfocito (INL)
- Leucos x10⁹
- PCR (mg/L)
- Negativiza PCR en 7 días
- Fibrinógeno (g/L)

23.- Marcadores de coagulación

- Plaquetas x 10E9/L
- Fibrinógeno (g/L)
- Tp
- TTPA

24.- TAC torácico previo a la cirugía:

- No
- Normal
- Consolidación focal
- Infiltrado pulmonar unilateral
- Infiltrado pulmonar bilateral
- Otros hallazgos (texto libre)

25.- qSOFA Score (confusión+Frec respiratoria \geq 22/min+TA_{sis} \leq 100mmHg): 0-3.

26.- CURB-65: Confusión / Urea (mg/dL) / Frec. respiratoria / PA / Edad. Calculadora gratuita en: <http://www.semergencantabria.org/calc/cpcalc2.htm>

Tratamiento TVP

27.- HBPM

- dosis profilácticas
- dosis terapéuticas

28.- Tipo HBPM

29.- NACO:

- Rivaroxaban
- Apixaban
- Dabigatran
- Edoxaban

Resultados

30.- Mortalidad a 7 días

31.- Mortalidad a 30 días

32.- Días ingreso en unidad de cuidados críticos (UCI/REA).

33.- Días de ingreso hospitalario

34.- TEP

35.- Complicaciones hemorrágicas

5. OBJETIVOS CIENTÍFICOS Y TECNOLÓGICOS E INNOVACIONES TECNOLÓGICAS DEL PROYECTO

5.1. Objetivos del proyecto

El objetivo fundamental del proyecto TVP-COVID consiste en proponer un sistema de **predicción de la incidencia de TVP en pacientes COVID y su tratamiento idóneo** haciendo uso de herramientas de IA, procesando grandes volúmenes de datos clínicos y externos. Se pretende conseguir una estimación de la probabilidad global de un caso de TVP en un paciente con enfermedad de COVID-19 y una estimación de la mejor estrategia de tratamiento posible.

Este objetivo funcional se alcanzará mediante la consecución de los siguientes objetivos tecnológicos:

- Desarrollo de un sistema inteligente de **Recuperación de la Información**, que permita extraer contenido selectivamente de la base de datos. En general hablaremos de datos estructurados y datos no estructurados. Para llevar a cabo este sistema necesitaremos aplicar tecnologías de Big Data para la extracción, la securización, el anonimato (en caso de ser necesario), el procesamiento y la indexación de la información.
- Desarrollo de una solución de **Procesamiento de Datos** que realice el análisis, limpieza, y transformación de los datos almacenados de manera distribuida y escalable.
- Desarrollo de un módulo de **Consolidación del Conocimiento** que permita trabajar con toda la información capturada y procesada. Será el encargado de la gestión de las actualizaciones del sistema: actualización y re-entrenamiento de los algoritmos de Aprendizaje Máquina.
- Desarrollo de una solución de **Inteligencia Artificial** que mediante técnicas de Aprendizaje Máquina sea capaz de obtener predicciones de la incidencia de TVP en los pacientes con COVID-19 y conseguir un algoritmo terapéutico adecuado e individualizado para cada paciente, integrando la información proveniente de fuentes

hospitalarias con aquella que provenga de fuentes externas identificadas como factores de interés.

- Integración de las soluciones anteriores para el desarrollo de una Plataforma de **Visualización** que permita gestionar el sistema, obteniendo los resultados de predicción e indicación terapéutica individualizada.

5.2. Principales elementos innovadores del proyecto

La estimación de la predicción de casos de TVP mediante el estudio de fuentes de datos hospitalarias conlleva el desarrollo de una serie de elementos innovadores que podemos resumir como sigue:

- **Recolección e integración** de datos heterogéneos procedentes de distintas fuentes de información en un repositorio integrado que facilite su preparación, análisis y gestión.
- Diseño y validación de un sistema de caracterización de los recursos detectados a partir de la combinación de múltiples factores descriptivos, utilizando técnicas de **Aprendizaje Máquina** para facilitar su análisis.
- Desarrollo de un componente de **Aprendizaje Máquina Dinámico** que permita la clasificación progresiva de los datos a partir de flujos continuos de recolección de información, adaptando dinámicamente los resultados de dicha clasificación para detectar cambios en las relaciones subyacentes que sean detectadas. Se entrenarán modelos basados en **Aprendizaje Profundo** para el desarrollo de estos sistemas.
- Identificación de **eventos y patrones no triviales** asociados a TVP, mediante algoritmos adaptativos, sustentados con tecnología de Big Data y algoritmos de Aprendizaje Máquina.
- Diseño y validación de una **Interfaz Gráfica** que proporcione información, tanto de forma visual como numérica de las predicciones de TVP y tratamiento.

A. Módulo de Recuperación y Procesamiento de la Información

Por un lado, nos enfrentamos al reto de extraer información de bases de datos respetando los aspectos de **privacidad** y **protección** de datos asociados. Por otro lado, nos enfrentamos a la **heterogeneidad** de fuentes de datos externas que podrían estar potencialmente asociadas al dominio de aplicación.

En este módulo los aspectos de innovación más destacables son:

- Desarrollo de técnicas para discernir si un elemento de información es relevante o no para nuestro trabajo, de modo que no se acumulen cantidades de información inservible. Esto es complejo dada la variedad de fuentes de información implicadas. Para llevar a cabo esta tarea se requiere de la actuación de especialistas en el ámbito de aplicación.
- Desarrollo de técnicas para extracción, procesado e indexación de la información adecuada dada la estructura final de nuestra base de datos.
- Desarrollo de técnicas de limpieza, transformación y representación de la información proveniente de fuentes heterogéneas de datos.

A este respecto cabe señalar que para que cualquier técnica de Ciencia de Datos tenga éxito es necesario contar con datos de una mínima calidad, como los aportados por el proyecto “COVID DATA SAVE LIVES”. Se evita de este modo uno de los principales errores que caracteriza a algunos proyectos de Análisis de Datos: pensar que el “Big Data” encuentra la solución a todos los problemas planteados. Ningún método de Ciencia de Datos, o tecnología de Big Data, es capaz de suplir con éxito el lastre que supone la baja calidad de los datos adquiridos.

La **gobernanza de datos** es un conjunto de procesos, funciones, políticas, normas y mediciones que garantizan el uso eficaz y eficiente de la información. Permite fijar una serie de procesos y responsabilidades que aseguran la calidad y la seguridad de los datos que se emplean en un proceso. La gobernanza de datos define quién puede emprender acciones, sobre qué datos, en qué situaciones y mediante qué métodos.

Los factores clave de cada proceso indicarán qué tipo de datos son los que deben controlarse de cerca en su estrategia de gobernanza, así como las ventajas que puede esperar obtener. Esta estrategia constituirá la base de su marco de gobernanza de datos.

Por ejemplo, un factor clave para la estrategia de gobernanza de datos en el proyecto TVP-COVID es **garantizar la confidencialidad de datos sanitarios**, por lo que los datos de los pacientes deberán ser gestionados de forma segura. Se definirán unos requisitos de retención (p. ej., el historial de quién ha modificado información y en qué momento) para asegurar el cumplimiento de la normativa oficial pertinente, como es el caso del RGPD en UE o HIPAA en USA.

Para cumplir con todo ello se desarrollarán **innovadoras herramientas de código abierto escalables y de fácil integración** en el entorno o **red** de las fuentes de información con las que se van a proceder a interconectar el sistema.

En el proyecto TVP-COVID se propone avanzar en el estado del arte sobre **procesamiento** y **análisis** en tiempo real de datos procedentes de fuentes hospitalarias y fuentes de datos externas asociadas a la incidencia de TVP, mediante varios elementos diferenciales:

- Diseño e implementación de sistemas eficientes de **representación** y **análisis** de datos longitudinales, para procesar grandes volúmenes de datos recogidos a lo largo del tiempo.
- Un reto fundamental a tratar en este entorno será el tratamiento y preparación de datos que, potencialmente, pueden contener errores de medida o intervalos de longitud indeterminada sin información. Por tanto, los algoritmos desarrollados deberán considerar la integración de técnicas adecuadas de **imputación** y **limpieza** de datos que puedan mejorar el funcionamiento de los algoritmos diseñados para estos propósitos.
- Integración de estos algoritmos en **tecnologías de Aprendizaje Máquina escalable**, tales como Apache Spark, incluyendo tanto bibliotecas integradas en el proyecto (ML/MLlib) como bibliotecas de terceros proveedores disponibles como software libre (tales como H2O). Precisamente, la utilización de técnicas avanzadas de representación simbólica de series temporales permite un cambio en el dominio de representación de los datos, de forma que sean compatibles con el tipo de información

estructurada y semi-estructurada que se suele utilizar en estas plataformas, ya que normalmente no son compatibles con las técnicas comunes de análisis de series temporales y datos longitudinales.

B. Módulo de Consolidación del Conocimiento

Este módulo está a cargo del almacenamiento de la información extraída y del conocimiento generado para cada una de las fuentes de datos previamente procesadas. Para ello es fundamental asegurar la imposibilidad de **manipulación**, así como asegurar la **trazabilidad** de los datos almacenados. La necesidad de **encriptación** de estos últimos, así como su posible distribución permitiría el desarrollo de una herramienta de mayor calidad.

C. Módulo de Inteligencia Artificial

El desarrollo de un módulo de IA para el sistema TVP-COVID implica los siguientes retos de **Análítica** de datos:

- Aplicación de algoritmos de Aprendizaje Máquina para **datos longitudinales**, que sean capaces de detectar patrones usuales y que permitan identificar y agrupar casos susceptibles de ser anómalos, esto es que se desvíen de dichos patrones típicos. Este es un problema en el que hay que tener en cuenta los avances en métodos de agrupamientos de series temporales, de forma que se utilicen métodos adecuados para la identificación de casos fuera de la norma de comportamiento habitual.
- Tratamiento de datos **fuertemente desequilibrados**. Hay que considerar la baja incidencia de casos de TVP en relación al volumen de datos recogidos. Este desequilibrio entre las clases conlleva dificultades para la aplicación de los métodos clásicos de Aprendizaje Máquina. Será necesario considerar y desarrollar nuevas técnicas basadas en la presencia de “**exceso de ceros**”, suponiendo todo un reto de innovación en el dominio del modelado estadístico.
- La incorporación de tecnologías de **ingeniería de Big Data** que proporcionen soporte para el **análisis dinámico de grandes flujos de datos**, tales como la plataforma Apache Spark, así como sistemas de comunicación de datos y mensajería de alto rendimiento como Apache Kafka, Apache Flume o ZeroMQ permitirán implementar una plataforma de última generación que posibilite abordar estos retos con garantías.

D. Módulo de Visualización

Este módulo permite realizar una visualización de los resultados obtenidos. Se mostrará la evolución temporal de los datos longitudinales. Se presentará la predicción de incidencia de TVP en pacientes con enfermedad de COVID-19. Se emplearán técnicas de reducción de la información para representar las relaciones dinámicas existentes entre los factores de riesgo y la incidencia de casos de TVP.

6. DESCRIPCIÓN TÉCNICA Y PLAN DE TRABAJO

6.1. Descripción y alcance del conjunto de actividades a realizar

Se han realizado estudios retrospectivos de pacientes con COVID-19 y ETV tratando de encontrar los factores de riesgo asociados a la ETV. Sin embargo, estos estudios poseen un número muy escaso de pacientes, dificultando enormemente la detección de factores de riesgo relevantes.

El Sistema Computacional para la predicción de TVP que se desarrollará mediante el proyecto “**COVID DATA SAVE LIVES**”, pretende conseguir predicciones de la incidencia de TVP en pacientes COVID-19 especificando la probabilidad asociada a los diferentes marcadores de riesgo y elucidando una indicación terapéutica adecuada. Para alcanzar dicho objetivo se emplearán técnicas y métodos de Ciencia de Datos, Inteligencia Artificial y Big Data. Se emplearán fuentes de datos hospitalarias. Los datos serán procesados, consolidados y utilizados para convertirlos en conocimiento de valor para los profesionales sanitarios implicados.

- Los datos de los pacientes de las **bases de datos hospitalarias** serán obtenidos mediante procesos y normas de gobernanza de datos, garantizándose de este modo, la privacidad y protección de los datos, asegurándose el uso eficaz y eficiente de la información. Se obtendrá información de **fuentes de datos externas**, ajenas al dominio hospitalario, pero potencialmente de interés en la predicción de casos de TVP. Estas fuentes corresponden a diferentes tipos de dato: datos climatológicos, como la temperatura, la presión atmosférica, etc.; datos socioeconómicos; datos socioculturales; datos geográficos; datos de transporte, como niveles de contaminación, tráfico, etc.

- Los datos obtenidos serán procesados con el objeto de extraer, transformar y limpiar dichos datos mediante técnicas y métodos basados en **Ciencia de Datos**. Todo el conocimiento adquirido durante este proceso será almacenado en una base de datos que será conveniente protegida y asegurada.
- Con toda la información capturada se procede a investigar mediante un módulo de **Inteligencia Artificial**. Se emplearán técnicas basadas en predicción de datos longitudinales, métodos de regresión basados en modelos lineales generalizados con excesos de ceros, técnicas de clasificación de datos extremadamente desequilibrados, con el objetivo de construir modelos predictivos de la aparición de un caso de TVP en un paciente COVID-19 concreto. Así mismo, se pretende conseguir predicciones para el mejor tratamiento posible para ese paciente.
- Finalmente, todo el conocimiento conseguido será presentado a los médicos a través de una **Interfaz Gráfica**.

A. Módulo de Recuperación y Procesamiento de la Información.

Funcionalmente, este módulo será el encargado de recopilar información significativa tanto de datos estructurados, semi-estructurados y no estructurados. Tiene un enorme desafío en la capacidad para recopilar la enorme cantidad de información hospitalaria generada, y tener algoritmos que le hagan ser selectivo en sus capturas de información.

Este módulo deberá contener las siguientes funcionalidades principales:

- I. **Extracción** de Información procedente de Bases de Datos hospitalarias (historias clínicas).
- II. **Extracción** de la Información procedente de páginas web y bases de datos externas. Se utilizarán bases de datos del Instituto Nacional de Estadística, Comunidad de Madrid, y ayuntamientos implicados. Se pretende evaluar la relevancia de factores externos en la incidencia de casos de TVP.

- III. **Indexación** de la Información obtenida para acelerar las consultas que se realicen a la base de datos mediante diferentes términos o criterios de filtrado de resultados.
- IV. En este módulo se emplearán técnicas de **transformación** de las variables necesarias para su posterior análisis y almacenamiento. Para las fuentes de datos de origen no hospitalario se evaluarán y almacenarán diferentes **representaciones** de la información, tratando de recoger la mayor cantidad de características potencialmente relevantes en análisis posteriores. Aquí se adaptarán metodologías que ya están dando buen resultado en otras áreas de aplicación, tales como la representación simbólica de series temporales, y que facilita también aspectos cruciales como la visualización efectiva de datos de esta naturaleza.
- V. Se realizarán tareas de imputación de **datos faltantes** siempre que sean necesarias.
- VI. **Procesamiento** de la **información textual** para su posterior análisis empleando técnicas de Procesamiento de Lenguaje Natural.

Para ello se plantea un esquema general de Extracción-Transformación-Explotación, que se verá particularizado para el caso específico del TVP-COVID, bajo la caracterización específica que se definirá para su perfil asociado. Una vez adquiridos y transformados esos datos, se procede a la extracción de los datos “útiles” y su **privatización** y **anonimización**, siendo encriptados internamente.

B. Módulo de Consolidación del Conocimiento.

Se prevé que el módulo de Consolidación del Conocimiento sea el encargado de almacenar datos estructurados, semi-estructurados y no estructurados simultáneamente. Algunos productos tradicionalmente ligados al ámbito de los modelos relacionales de datos, como PostgreSQL, han expandido recientemente su funcionalidad para soportar también el almacenamiento de datos semiestructurados (en formato JSON y RJSON). Por ahora, los resultados de rendimiento son muy prometedores, superando incluso la eficacia de soluciones NoSQL con cierto recorrido en el mercado como MongoDB.

Este avance mediante la combinación de diferentes técnicas de almacenamiento de datos en un solo producto resulta muy conveniente en casos como el planteado por TVP-COVID. Por otro lado, actualmente ya existen soluciones en el mercado que satisfacen estos requisitos y

sobresalen por su adaptabilidad y eficiencia en múltiples entornos de aplicación. Algunos ejemplos son:

- Redis: un sistema de almacenamiento de datos que combina modelos estructurados tradicionales con la capacidad para almacenar y gestionar datos no estructurados.
- Aerospike: un producto más sofisticado, que combina un motor relacional tradicional con un motor NoSQL adecuado para almacenamiento de datos no estructurados. Adicionalmente, también incluye características como capacidad de procesamiento de flujos de datos y procesamiento en memoria.

Otras ventajas muy interesantes para nuestros propósitos radican en que todas estas soluciones son software libre, poseen una sólida comunidad de desarrollo que las sustenta y han sido ampliamente probadas en entornos de producción de muy diversos dominios de aplicación.

Finalmente, otro aspecto crucial a tener en cuenta es que cualquiera de estas soluciones ofrece múltiples opciones de conectores en los diferentes lenguajes de programación más utilizados para ingeniería y análisis de datos (Python, Java, C/C++, R, etc.). De este modo, se evita que el diseñador vea limitadas sus preferencias por alguno de estos lenguajes.

C. Módulo de Inteligencia Artificial.

El diseño e implementación de un módulo de Inteligencia Artificial para el sistema propuesto involucra numerosos retos de ingeniería y analítica de Big Data. Una solución transversal para el desarrollo de todas las funcionalidades incluidas en este módulo es la selección de una plataforma que proporciona la capacidad de procesar grandes volúmenes de datos de forma dinámica, sin pérdida de rendimiento.

Se propone una doble solución que resuelve sus dos vertientes principales:

- Por un lado, la adopción de Apache Spark como plataforma central de procesamiento de grandes flujos de datos sujetos a continuas modificaciones (a partir de nueva información de entrada que se vaya recolectando).

- La utilización de tecnologías avanzadas para gestión de flujos de datos y comunicación de alto rendimiento entre los distintos componentes de la plataforma, por ejemplo Apache Kafka o ZeroMQ.

En el sistema de Inteligencia Artificial se entrenarán una serie de modelos basados en las siguientes técnicas:

- I. **Clasificación dinámica** de la información.
- II. Clasificación de **datos fuertemente desequilibrados**.
- III. Predicción de **datos longitudinales**.

I. Clasificación dinámica de la información.

En lo referente a los algoritmos de clasificación, la tendencia actual es la de implementar métodos de análisis de datos que permitan actualizar la información y los resultados del sistema **dinámicamente**. En entornos como Spark se dispone actualmente de bibliotecas como Mllib que incluyen una colección de algoritmos adecuados para procesar las abstracciones de flujos de datos que proporciona la plataforma. De esta forma, los algoritmos se integran en el núcleo de la arquitectura de analítica de datos distribuida, y facilita su extensión por medio de la reutilización de componentes ya existentes. Puesto que estos elementos básicos de diseño ya se encuentran integrados dentro de la propia plataforma de Apache Spark, resultan muy convenientes a la hora de combinarlos con otros métodos y algoritmos para aprovechar las capacidades de análisis de grandes flujos de datos dinámicos que proporciona la plataforma. Esta ventaja resulta esencial en nuestro caso, debido al gran dinamismo del contenido que se espera extraer de las distintas fuentes.

II. Clasificación de datos fuertemente desequilibrados.

Los problemas de clasificación de **datos desequilibrados** son aquellos en los que el tamaño de una de las clases y/o la penalización por clasificación errónea son claramente diferentes. Estos problemas tienen mucha relevancia en las aplicaciones prácticas. Los clasificadores más utilizados son sensibles a datos desequilibrados porque los valores de sus parámetros se establecen mediante algoritmos que tratan de optimizar las medidas de rendimiento sin tener

en cuenta los efectos de los desequilibrios. En la actualidad se trabaja en nuevos métodos de **Aprendizaje Máquina** diseñados específicamente para datos desequilibrados.

Uno de los inconvenientes de los modelos para estimar la probabilidad de TVP es que asumen un equilibrio muestral entre el número de casos de TVP y no-casos. Este problema no se ha abordado específicamente en este dominio y para ello se deben desarrollar enfoques metodológicos basados en estrategias de equilibrio de subconjuntos y líneas de aprendizaje automático, que deben incluirse en los modelos de clasificación y predicción asociados.

Se aplicarán técnicas de clasificación basadas en **Modelos Lineales Generalizados (GLMs)**. Un modelo ampliamente utilizado para modelar este tipo de datos es la Regresión de Poisson, debido a la existencia de numerosos ceros en la distribución de la variable respuesta, frente al número de unos (casos de TVP). En este tipo de regresión la variable dependiente es un **recuento** del número total de pacientes con TVP durante un periodo de tiempo. La distribución de Poisson es especialmente útil en este caso pues sirve para modelar la probabilidad de que un determinado número de eventos (casos de TVP) ocurran en un intervalo fijo de tiempo, si estos eventos se producen a una tasa media constante e independientemente del tiempo transcurrido desde el último evento. Desde un punto de vista de **Explicabilidad**, cabe destacar la sencillez en la interpretación de los resultados. Es claro que el proyecto TVP-COVID necesitará abordar uno de los **retos** que aparecen en este tipo de modelos y que más los afectan: **el exceso de ceros**. En esos casos, GLMs basados en la distribución binomial negativa o modelos de “excesos de ceros” son más adecuados y serán entrenados y evaluados mediante las métricas adecuadas.

III. Predicción de datos longitudinales.

Construcción de regresores y clasificadores sobre variables explicativas de carácter **dinámico**. Detección de **anomalías** y valores **atípicos**. Se analizará la relación existente entre la variable respuesta de interés, TVP y las variables explicativas de carácter dinámico, procedentes tanto de fuentes de datos hospitalarias como de fuentes de datos externas. Se llevará a cabo la predicción de la variable de interés mediante técnicas de predicción de datos longitudinales.

D. Módulo de Visualización, Interfaz gráfica.

Se construirá un sistema de visualización que permitirá por una parte realizar la operación del sistema, y por otra su utilización. Dentro de esta última se desarrollará un sistema de inteligencia con indicadores claves, que serán presentados en un cuadro de mando gráfico que permitirá el estudio de los indicadores.

Características de la arquitectura

Todos los desarrollos a efectuar deberán ser tales que el resultado pueda considerarse:

- **Seguro.**

Para evitar el mal uso o manipulación de los datos (personales o no) y estudios contenidos y generados utilizando la plataforma TVP-COVID, esta ha de ser segura en su diseño y concepción. Por ello, se aplicarán las mejores técnicas en desarrollo seguro de software y se realizarán los correspondientes tests de penetración para asegurar la seguridad de la solución. El código fuente se auto-firmará digitalmente para asegurar que este no es modificado por un administrador malicioso con acceso al sistema.

- **Flexible.**

TVP-COVID dispondrá de opciones que lo harán fácilmente adaptable a las necesidades de los usuarios en sí, sobre todo en cuanto a los niveles de seguridad y perfiles de acceso.

- **Abierto.**

La arquitectura abierta de TVP-COVID permitirá la interoperabilidad con sistemas de terceros de forma sencilla y transparente. Se aplicarán las mejores prácticas en diseño de aplicaciones abiertas para orientar el módulo a su uso generalizado.

- **Escalable.**

Por su propia naturaleza, la arquitectura de rAAA será totalmente modular, lo que permitirá ampliar las funcionalidades del sistema si se considerase necesario. Esto facilitaría la

inclusión de nuevas funcionalidades, p.ej. referentes a nuevas fuentes de datos a analizar o algoritmos y/o motores de búsqueda asociados.

- **Auditable.**

La relación de confianza con un software que provea predicciones de casos de TVP es vital para asegurar un buen uso del mismo. Es por ello que el desarrollo tendrá en cuenta criterios para facilitar la auditoría del código fuente por parte de terceros. En este caso, además, se tendrá especialmente en cuenta la trazabilidad del cumplimiento de la GDPR o equivalentes internacionales.

7. CAPACIDAD DEL EQUIPO DE INVESTIGACION

7.1. Perfil profesional del equipo

7.1.1. Red de Investigación Vascular de la Sociedad Española de Angiología y Cirugía Vascular

Joaquín de Haro Miralles. Es Especialista Staff en el Servicio de Angiología y Cirugía Vascular, Hospital Universitario Getafe, Madrid. Profesor Asociado en la Facultad de Ciencias Biomédicas y de la Salud, Departamento Cardiovascular, de la Universidad Europea Madrid. Investigador Titular de la Unidad de Investigación Cardiovascular de la Fundación para la Investigación Biomédica del Hospital Universitario Getafe. Miembro del Comité Científico de Investigación y Comité Ético de Investigación Clínica del Hospital Universitario Getafe. Vocal de Investigación de la Junta Directiva de la Sociedad Española de Angiología y Cirugía Vascular. Miembro del Comité Científico del Capítulo Español de Flebología y Linfología.

Miembro del *Council on Peripheral Vascular Disease* de la American Heart Association/ American Stroke Association (AHA). Miembro y revisor de Cochrane Library, Peripheral Vascular Diseases Group. Miembro del Comité Editorial de *Angiology*, *World Journal of Methodology*, *World Journal of Meta-analysis*, *Journal of Cardiology and Therapy*, *World Journal of Cardiology*, *Clinical Medicine Insights: Cardiology*, *Open Journal of Cardiovascular Surgery*. Editorial Board Member como Review Editor de "*Vascular Surgery*" section of *Frontiers in Surgery* (Nature Publishing Group). Editorial Board Member como Associate Editor of "*Interventional Cardiology*" section of *Frontiers in Cardiovascular Medicine* (Nature Publishing Group).

Revisor de las revistas médicas *British Medical Journal*, *Journal of the American College of Cardiology*(JACC), *Journal of Vascular Surgery*, *British Journal of Surgery*, *British Medical Journal Case Reports*, *Atherosclerosis*, *European Journal of Vascular and Endovascular Surgery*, *Annals of Vascular Surgery*, *European Journal of Pharmacology*, *Vascular Medicine*, *Progress in Health Sciences*, *Diabetes & Vascular Disease Research*

Journal, Progress in Health Sciences, Heart Asia, The Journal of Cardiovascular Surgery, The Journal of Musculoskeletal Pain, Minerva Chirurgica, Vascular Medicine Expert Opinion on Biological Therapy, Histology and Histopathology, Heart and Vessels, Journal of Cardiovascular Computer Tomography.

Evaluador y Miembro del Biomedical and Therapeutic Research Committee of Scottish Government (UK). Evaluador y Miembro del International Biomedical Research Committee of Uruguay's Government.

90+ artículos en publicaciones en revistas internacionales (30+ como primer autor; 85% en Q1).

30+ Premios Nacionales e Internacionales de Investigación.

José Antonio Diaz. Actualmente ocupa el puesto de Research Associate Professor of Surgery, Director of the Division of Surgical Research - Light Surgical, Research & Training Laboratory, Section of Surgical Sciences, Division of Surgical Research, Vanderbilt University, (Nashville, TN, USA). Anteriormente ha desempeñado los cargos de Research Fellow, Vascular Biology, NIH T32 Training Grant Vascular Surgery, Research Training in Vascular Biology, Section of Vascular Surgery, Department of Surgery, University of Michigan, (Ann Arbor, MI, USA; Jun 2006 - May 2008), Research Investigator, Conrad Jobst Vascular Research Laboratories, Section of Vascular Surgery, Department of Surgery, University of Michigan, (Ann Arbor, MI, USA; Jun 2008 – Aug 2013), Research Assistant Professor of Surgery, Conrad Jobst Vascular Research Laboratories, Section of Vascular Surgery, Department of Surgery, University of Michigan, (Ann Arbor, MI, USA; Sept 2013 – April 2019).

Es miembro, Director y Presidente electo durante 2015-2020 de la American Venous Forum (AVF), miembro del Comité Científico y del Consejo de Sanidad de la American Heart Association, Co-Presidente del Comité Científico de la International Society of Thrombosis and Haemostasis (ISTH), miembro del Comité Docente de la International Union of Phlebology (UIP).

Javier Rodríguez Padilla, Facultativo especialista en Hospital Universitario HM Montepríncipe y HM Puerta del Sur, centro de referencia nacional e internacional de

patología venosa compleja, siendo director del Laboratorio de Diagnóstico Vascular y Profesor colaborador en Patología Quirúrgica II en la Universidad CEU San Pablo (desde 2015). Estancia de 3 meses en St. Mary's Hospital (Londres), centro de referencia en patología aórtica compleja. Ejerce también como Facultativo especialista en Angiología y Cirugía Vascular en Hospital Universitario Severo Ochoa y tutor de prácticas de 4º curso de Medicina de la Universidad Alfonso X el Sabio (desde 2018).

Ha presentado más de 20 comunicaciones científicas en diferentes congresos nacionales e internacionales. Ha participado en la publicación de más de 10 artículos científicos, 5 de ellos como primer autor. Ha formado parte de varios proyectos de investigación (Ensayos clínicos en fase 3) (VIRTUS: An Evaluation of the Veniti Vici™ Venous Stent System in Patients with Chronic Iliofemoral Venous Outflow Obstruction (2015), VERNACULAR: The BARD® Venous Stent. A Prospective, Non-Randomized, Multi-Center, Single-Arm Study of the Treatment of Iliofemoral Occlusive Disease an Assessment for Effectiveness and Safety (2016).

Enrique San Norberto. Jefe de Unidad de Angiología y Cirugía Vascular Hospital Clínico Universitario de Valladolid. Profesor Asociado Sanitario Universidad de Valladolid. Doctor por la Universidad de Valladolid, sobresaliente cum laude 2009. Máster en Gestión Sanitaria por la Universidad a Distancia de Madrid 2014. Diplomado especializado en Gestión y Organización Sanitaria por la Universidad de Salamanca 2018. Experto en Gestión de Procesos Asistenciales en Angiología y Cirugía Vascular 2018-2019.

Acreditación como Profesor Ayudante Doctor, Profesor Contratado Doctor y Profesor de Universidad Privado por la ANECA 2017. Miembro de la Sociedad Española de Angiología y Cirugía Vascular, European Society for Vascular Surgery, Sociedad de Cirujanos Vasculares de Habla Hispana. Académico correspondiente de la Real Academia de Medicina y Cirugía de Valladolid. Miembro de la Comisión Nacional de la Especialidad de Angiología y Cirugía Vascular 2006-2008. Secretario del Capítulo de Cirugía Endovascular de la Sociedad Española de Angiología y Cirugía Vascular (2015-2018). Miembro del Comité Científico del Capítulo de Cirugía Endovascular de la Sociedad Española de Angiología y Cirugía Vascular (2011-2014). Miembro del Comité Científico de la Sociedad Española de Angiología y Cirugía Vascular (2015-2019). Representante de la Sociedad Española de Angiología y Cirugía Vascular en la Union Européenne des Médecins spécialistes (UEMS)

(2017-2021). Tesorero de la Sociedad de Cirujanos Vasculares de Habla Hispana (2017-2021). Coordinador del Grupo de Trabajo sobre tratamiento médico de la Trombosis Venosa Profunda del Capítulo de Flebología de la Sociedad Española de Angiología y Cirugía Vascular. Autor de 9 libros publicados, primer autor de 39 capítulos de libros, segundo u otro autor de 142 capítulos de libros. Primer autor de 27 artículos publicados en revistas con factor impacto, segundo u otro autor de 13 artículos publicados en revistas con factor impacto. Primer autor de 54 artículos en otras revistas y segundo u otro autor de 153 artículos en otras revistas. Editor de la revista Angiología, consejo de redacción de la Revista Española de Investigaciones Quirúrgicas y de la Revista Iberoamericana de Cirugía Vascular.

Rodrigo Rial es Médico Especialista del Servicio de Angiología y Cirugía Vascular del Hospital Clínico San Carlos de Madrid. Presidente del Capítulo Español de Flebología y Linfología de la SEACV. Miembro del Comité Científico (2013-2017). Vicepresidente (2015-2019). Miembro de la Junta Directiva del Capítulo de Cirugía Endovascular de la SEACV. Miembro de la Junta Directiva de la Sociedad Española de Angiología y Cirugía Vascular (SEACV). Miembro de la International Atomic Energy Agency (IAEA). Radiation Protection Expert-Lecturer in RPoP. 2012-Presente. Director de Instalaciones de Rayos X con fines de Diagnóstico Médico. Acreditación de Segundo Nivel en Protección Radiológica en Práctica Intervencionista. Profesor Docencia Práctica del Departamento de Cirugía de la Facultad de Medicina. Universidad Complutense. Tutor Curso Anual “Actualización en Patología Venosa. Academia de la vena” de la Sociedad Española de Angiología y Cirugía vascular (SEACV). Tutor Curso Anual “Cómo Escribir y Publicar un artículo Científico” de la SEACV. Tutor Curso Anual “Metodología Científica en la Práctica Clínica” de la SEACV y del Capítulo Español de Flebología de la SEACV.

Guillermo Moñux. Jefe Servicio C. Vascular Hospital HM Torreldones. Profesor asociado Departamento de Cirugía Universidad Complutense de Madrid (Cirugía Vascular). Profesor asociado área cardiovascular en la Universidad Europea de Madrid. 24 Publicaciones en revistas Nacionales. 24 Publicaciones en revistas Internacionales. 9 Capítulos en libros Nacionales e Internacionales. Director de 12 Tesis Doctorales en el departamento de Cirugía de la U. Complutense de Madrid. Responsable de investigación del servicio de Angiología y Cirugía Vascular del Hospital Clínico San Carlos, habiendo dirigido como investigador principal 10 proyectos de investigación en los últimos 15 años. Coordinador responsable e

investigador principal de la red ITACA de la Comunidad de Madrid. Coordinador de la sección de accesos vasculares de la Sociedad Española de Angiología y Cirugía Vascular (2014-2017). Miembro del comité científico del Capítulo Español de Flebología y Linfología.

Noelia Alonso Gómez. Es médico, especialista en Angiología y C. Vascular. Cursó además 2 años de la especialidad de Radiodiagnóstico. Ha trabajado en el ámbito de la medicina privada y pública y dentro de esta última, en un ámbito cívico-militar (actualmente en el HCD GÓMEZ-ULLA/Carabanchel).

Creadora de la app IMAGEN VASCULAR que recibió uno de los galardones “Mejor Idea médica del año 2017”. Su interés investigador se centra en las nuevas TICs y su aplicación al X-learning y más recientemente por el campo de la IA. Otro foco de interés es el campo de la simulación y en especial, la experimentación animal y colaborando puntualmente con diversas fundaciones.

En lo estrictamente curricular ha obtenido la DEA, Tesis doctoral con grado de Doctor, profesor clínico, cursos de experto, publicación y revisión de artículos científicos en Pubmed, capítulos de libros y libros, paneles y comunicaciones a congresos, conocimientos de estadística. Cuenta con varios premios a nivel nacional y con una Cruz del Mérito Militar con Distintivo Blanco. Es miembro de la SEACV y de la ESVS y de varios comités hospitalarios.

Francisco Álvarez Marcos. Es Especialista en Angiología y Cirugía Vascular, actualmente ejerciendo en el Hospital Universitario Central de Asturias (HUCA); previamente en el Complejo Hospitalario Universitario de A Coruña (CHUAC) (2014 – 2018). Es Senior Teaching Assistant en la Harvard T. H. Chan School of Public Health desde 2011.

Posee el grado Doctor Medicina por la Universidad de Oviedo (2017). Máster Oficial en Investigación en Medicina por la Universidad de Oviedo (2010). Master of Science in Biostatistics por la Harvard School of Public Health (2010). 14 publicaciones internacionales (5 en Q1). 7 capítulos de libro (2 internacionales). 40+ comunicaciones a congresos. 2 proyectos financiados en convocatoria pública (FIS).

Miembro del Comité Científico del Capítulo Español del Flebología y Linfología.
Responsable de presencia web de la Sociedad Española de Angiología y Cirugía Vascular
(SEACV).