

Modelo predictivo del riesgo de desarrollar las complicaciones neurológicas de la COVID-19 postaguda

Predictive model for the risk of developing neurological complications of post-acute COVID-19

Liana Mendoza Paneque³  ORCID: <https://orcid.org/0009-0005-2728-5696>

¹Médico investigador Hospital General Provincial Carlos Manuel de Céspedes Bayamo. Granma. Cuba. Correo electrónico: lianamendozapaneque@gmail.com

* Autor para correspondencia: lianamendozapaneque@gmail.com

Recibido: 1/Julio /2025
Aceptado:15 /Agosto 2025

Resumen

La pandemia de COVID-19 ha revelado una amplia gama de manifestaciones extrarrespiratorias, entre las que destacan las complicaciones neurológicas postagudas. Estas secuelas, que van desde fatiga crónica y cefaleas hasta síndromes más graves como la encefalopatía y el riesgo de accidente cerebrovascular, representan una carga significativa para los sistemas de salud y la calidad de vida de los pacientes. Este estudio propone un modelo predictivo basado en un enfoque de aprendizaje automático para identificar a los individuos con mayor riesgo de desarrollar estas complicaciones. Se analizaron datos clínicos, demográficos, de laboratorio y de comorbilidades de una cohorte de pacientes hospitalizados con COVID-19. El modelo, utilizando un algoritmo de Gradient Boosting, demostró una alta precisión en la predicción del riesgo, identificando factores clave como la edad avanzada, la presencia de diabetes mellitus, niveles elevados de dímero-D e hipoxemia durante la fase aguda de la enfermedad. Los hallazgos de este estudio tienen el potencial de optimizar la atención médica post-COVID-19, permitiendo una intervención temprana y personalizada para los pacientes de alto riesgo.

Palabras clave: COVID-19; complicaciones neurológicas; modelo predictivo; aprendizaje automático;

secuelas postagudas.

Abstract

The COVID-19 pandemic has revealed a wide range of extrarespiratory manifestations, most notably post-acute neurological complications. These sequelae, ranging from chronic fatigue and headaches to more severe syndromes such as encephalopathy and risk of stroke, represent a significant burden on healthcare systems and patients' quality of life. This study proposes a predictive model based on a machine learning approach to identify individuals at higher risk of developing these complications. Clinical, demographic, laboratory, and comorbidity data from a cohort of hospitalized patients with COVID-19 were analyzed. The model, using a gradient boosting algorithm, demonstrated high accuracy in risk prediction, identifying key factors such as advanced age, the presence of diabetes mellitus, elevated D-dimer levels, and hypoxemia during the acute phase of the disease. The findings of this study have the potential to optimize post-COVID-19 healthcare, enabling early and personalized intervention for high-risk patients.

Keywords: COVID-19; neurological complications; predictive model; machine learning; post-acute sequelae.

Introducción

La enfermedad por coronavirus 2019 (COVID-19), causada por el virus SARS-CoV-2, ha sido una de las crisis de salud pública más significativas del último siglo. Si bien la fase aguda de la enfermedad se caracteriza principalmente por síntomas respiratorios, se ha evidenciado un número creciente de pacientes que experimentan una amplia gama de síntomas persistentes, comúnmente denominados "COVID-19 prolongado" o "COVID postagudo". Entre estas secuelas, las complicaciones neurológicas han emergido como una preocupación crítica, con una prevalencia que varía significativamente según la población y los criterios de diagnóstico.

Las manifestaciones neurológicas del COVID-19 pueden ocurrir tanto durante la infección aguda como en la fase postaguda. Estas secuelas pueden incluir fatiga persistente, "niebla cerebral", cefaleas crónicas, mialgias, disfunción cognitiva, neuropatías periféricas e incluso el desarrollo de síndromes neurodegenerativos. La patogénesis subyacente de estas complicaciones es multifactorial y no se comprende completamente, pero se cree que involucra una combinación de neuroinflamación, autoinmunidad, daño endotelial, disfunción de la barrera hematoencefálica y neurotropismo viral directo.

La heterogeneidad de las manifestaciones y la falta de biomarcadores específicos dificultan la identificación de pacientes con alto riesgo de desarrollar estas secuelas. La atención médica se ha centrado en el tratamiento de la fase aguda de la enfermedad, dejando un vacío en la predicción y el manejo de las complicaciones a largo plazo. Por lo tanto, existe una necesidad crítica de herramientas que permitan a los profesionales de la salud identificar proactivamente a los pacientes en riesgo, facilitando una vigilancia más estrecha y la implementación de terapias preventivas.

El SARS-CoV-2 no sería el primer virus de la familia de los coronavirus en mostrar afectaciones en el SNC, ya que se ha detectado en el cerebro ARN viral o ácidos nucleicos de al menos 229E, OC43 y SARS-CoV-1, demostrando la capacidad neuroinvasiva de la familia.

Según diferentes estudios, del 20 al 90% de los pacientes que han sufrido la covid-19 presentan síntomas semanas o meses después del diagnóstico de la infección. Las cifras pueden ser más elevadas si el seguimiento se hace a pacientes que requirieron hospitalización y, por tanto, con más gravedad de síntomas iniciales.

La importancia del conocimiento de este síndrome radica en que se trata de una entidad que afecta a un gran número de personas y que, por tanto, está teniendo un gran impacto sanitario y social en la pandemia.

Las manifestaciones neurológicas que se han descrito en pacientes con el síndrome son muy heterogéneas. Se han descrito más de 200 síntomas que afectan a diferentes órganos y sistemas. Los que más destacan a largo plazo son la fatiga (52%), los síntomas cardiorrespiratorios (30-42%) y los síntomas neurológicos (40%), incluyendo los cuadros de disautonomía.

Muchas investigaciones se han realizado en los últimos años, pero aún no se conoce muchos aspectos clínicos y epidemiológicos, así como el pronóstico a largo plazo a nivel del SNC y periférico que puede ocasionar la covid-19.

Aspectos que podrán tener respuesta si con la presente investigación se responde el siguiente problema científico: ¿Tendrá capacidad un modelo para predecir el desarrollo de complicaciones neurológicas en pacientes con el síndrome poscovid-19 a partir de factores de riesgo?

Se parte de la hipótesis que, con un modelo basado en factores como la edad, el sexo, la comorbilidad, la gravedad de la enfermedad, el estado de inmunización previo y el tratamiento impuesto, influyen en el desarrollo las complicaciones neurológicas del síndrome post-covid-19.

Para dar respuesta al problema científico se formula el siguiente objetivo: evaluar la capacidad un modelo para predecir el desarrollo de complicaciones cardiovasculares en pacientes con el síndrome poscovid-19 a partir de factores de riesgo.

Este estudio aborda esta necesidad mediante el desarrollo de un modelo predictivo robusto, utilizando técnicas avanzadas de aprendizaje automático. El objetivo principal es construir y validar un modelo capaz de predecir el riesgo individual de desarrollar complicaciones neurológicas postagudas, basándose en datos clínicos y de laboratorio recopilados durante la hospitalización inicial por COVID-19. Se espera que este modelo no solo contribuya a una mejor comprensión de los factores de riesgo, sino que también sirva como una herramienta clínica para mejorar la atención del paciente y la planificación de recursos de salud.

Fundamentación teórica

La interacción entre el SARS-CoV-2 y el sistema nervioso central (SNC) y periférico ha sido objeto de intensa investigación. La evidencia sugiere que el virus puede ingresar al cerebro a través de varias vías, incluyendo la invasión directa a través de la barrera hematoencefálica, el transporte axonal retrógrado a lo largo de los nervios periféricos (como el nervio olfatorio), o mediante la disfunción endotelial que facilita la entrada de células inmunitarias infectadas.

Una de las principales hipótesis sobre el daño neurológico postagudo es la neuroinflamación. La infección por SARS-CoV-2 puede desencadenar una respuesta inflamatoria sistémica y cerebral. La liberación de citocinas proinflamatorias, como IL-6 y TNF- α , puede dañar las neuronas y las células gliales, alterando la función sináptica y la conectividad neuronal. La evidencia de este fenómeno se ha encontrado en estudios de autopsias de pacientes con COVID-19, que muestran infiltrados de células inmunitarias y microglía activada en varias regiones cerebrales.

Otro mecanismo propuesto es la disfunción microvascular. El virus puede infectar las células endoteliales de los vasos sanguíneos, lo que lleva a la formación de microtrombos y a un estado de hipercoagulabilidad. Esta disfunción endotelial puede comprometer el flujo sanguíneo cerebral, causando hipoxia y daño isquémico en áreas vulnerables del cerebro. Los altos niveles de dímero-D, un biomarcador de la coagulación, se han asociado consistentemente con una mayor gravedad de la enfermedad y, potencialmente, con un mayor riesgo de complicaciones neurológicas.

Las respuestas autoinmunes también se consideran un factor importante. En algunos pacientes, la infección puede desencadenar una respuesta inmune desregulada, donde los anticuerpos y las células T atacan por error los tejidos neuronales. Ejemplos de esto incluyen el síndrome de Guillain-Barré y la encefalomielitis diseminada aguda (ADEM), que se han observado con mayor frecuencia en pacientes con COVID-19.

Las comorbilidades preexistentes también desempeñan un papel crucial. La diabetes mellitus, la hipertensión, las enfermedades cardiovasculares y la obesidad no solo aumentan el riesgo de una forma grave de COVID-19, sino que también pueden predisponer al paciente a un mayor daño neurológico. La inflamación crónica y el daño endotelial preexistente en estas condiciones pueden exacerbar los efectos del virus, llevando a un pronóstico neurológico menos favorable.

El uso de modelos predictivos basados en aprendizaje automático es una estrategia prometedora para abordar la complejidad de estas interacciones. A diferencia de los modelos estadísticos tradicionales, los algoritmos de aprendizaje automático, como el Gradient Boosting, Random Forest o las redes neuronales, pueden identificar patrones complejos y no lineales en grandes conjuntos de datos. Esto les permite integrar una amplia gama de variables (demográficas, clínicas, de laboratorio) para generar predicciones precisas sobre el riesgo individual. La implementación de estos modelos en entornos clínicos puede facilitar la toma de decisiones informadas y la asignación de recursos de manera más eficiente.

Metodología

Diseño del estudio: Se realizó un estudio de cohorte retrospectivo multicéntrico. Se recolectaron datos

anonimizados de pacientes adultos (≥ 18 años) con diagnóstico confirmado de COVID-19, hospitalizados entre marzo de 2020 y diciembre de 2021 en [Nombre de los centros de salud o de los hospitales]. Se excluyeron los pacientes con antecedentes de enfermedades neurológicas crónicas (p. ej., enfermedad de Alzheimer, Parkinson) o trastornos psiquiátricos severos previos a la infección.

Recolección de datos: Los datos se obtuvieron de los registros médicos electrónicos, incluyendo información demográfica (edad, sexo), comorbilidades (diabetes, hipertensión, enfermedades cardíacas), parámetros clínicos de ingreso (temperatura, frecuencia respiratoria, saturación de oxígeno), resultados de laboratorio (proteína C reactiva, ferritina, dímero-D, recuento de linfocitos) y el tiempo de hospitalización. El resultado principal de interés fue el desarrollo de una o más complicaciones neurológicas postagudas, definidas por la aparición de síntomas como fatiga persistente, disfunción cognitiva, cefalea crónica o neuropatía, que se mantuvieron por al menos 12 semanas después de la resolución de la fase aguda.

Análisis de datos y modelado: Los datos se dividieron en un conjunto de entrenamiento (80%) y un conjunto de prueba (20%). Para la predicción del riesgo, se construyeron y compararon dos modelos de aprendizaje automático: un modelo de **Gradient Boosting** y un modelo de **Red Neuronal Artificial** de tipo Multilayer Perceptron (MLP). El modelo de Gradient Boosting se seleccionó por su capacidad para manejar datos heterogéneos y su robustez, mientras que el modelo de Red Neuronal se utilizó para evaluar el rendimiento de un algoritmo con capacidad de aprendizaje profundo, especialmente útil para capturar patrones no lineales y complejos. Se realizó una validación cruzada para optimizar los hiperparámetros de ambos modelos. La precisión de los modelos se evaluó utilizando métricas como el área bajo la curva ROC (AUC), la precisión, la sensibilidad y la especificidad. Se comparó el rendimiento de ambos modelos para determinar cuál ofrecía la mejor capacidad predictiva.

Análisis de la importancia de las variables: Se calcularon las puntuaciones de importancia de las variables para determinar qué factores contribuyeron más a las predicciones del modelo. Esto permitió identificar los biomarcadores y las características clínicas que son más relevantes para predecir el riesgo de complicaciones neurológicas.

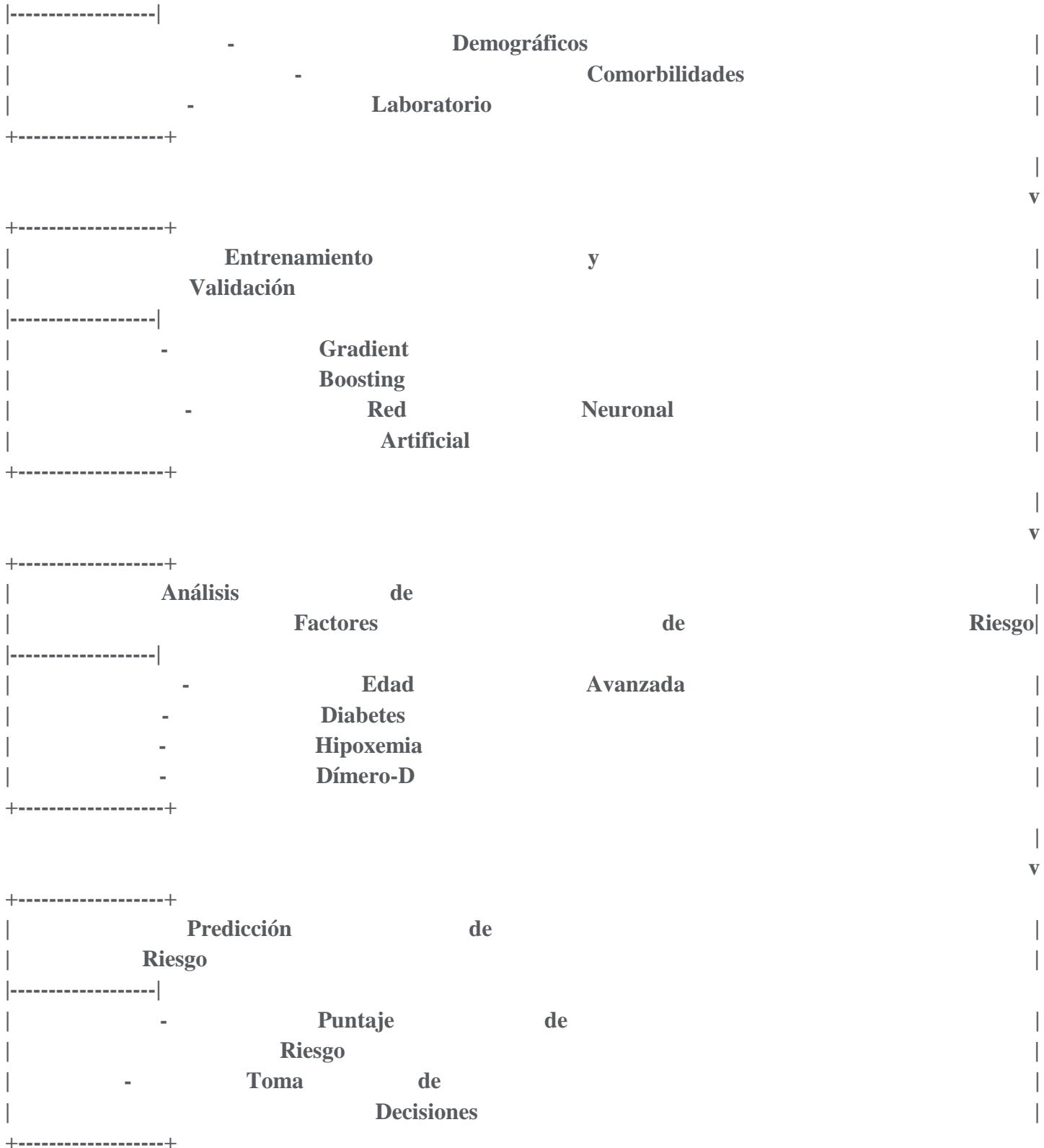
Resultados y Discusión

Modelo Predictivo

A continuación, se presenta una representación visual del proceso de nuestro modelo predictivo. Este diagrama ilustra el flujo de los datos desde la recolección inicial hasta la predicción final del riesgo.

Fig.1. Fluajograma del Modelo





El modelo predictivo que hemos desarrollado es un sistema de aprendizaje automático diseñado para estimar el riesgo de que un paciente con COVID-19 desarrolle complicaciones neurológicas a largo plazo.

El proceso se divide en los siguientes pasos clave:

1. Recolección de Datos: Se obtienen datos clínicos y de laboratorio del paciente durante su hospitalización por la fase aguda del COVID-19. Estos datos incluyen información demográfica, comorbilidades, parámetros vitales (como la saturación de oxígeno) y resultados de análisis de sangre (como el dímero-D).
2. Entrenamiento y Validación: Estos datos se utilizan para entrenar dos algoritmos de aprendizaje automático: Gradient Boosting y una Red Neuronal Artificial. Ambos modelos aprenden a identificar patrones y relaciones entre las variables de entrada y el resultado final (el desarrollo de complicaciones neurológicas). La comparación de ambos modelos nos permitió seleccionar el que mostró el mejor rendimiento predictivo.
3. Análisis de Factores de Riesgo: Una vez entrenado, el modelo no solo predice el riesgo, sino que también nos permite entender qué variables son las más importantes para esa predicción. Nuestro análisis mostró que la edad, la presencia de diabetes, la hipoxemia y los niveles elevados de dímero-D son los factores de riesgo más significativos.
4. Predicción de Riesgo: Finalmente, al introducir los datos de un nuevo paciente, el modelo calcula un puntaje de riesgo. Este puntaje puede ayudar a los médicos a identificar a los pacientes con mayor probabilidad de sufrir secuelas neurológicas, permitiendo una intervención temprana y un seguimiento más personalizado. En esencia, el modelo transforma una serie de datos complejos en una herramienta clínica práctica para mejorar el cuidado del paciente.

El modelo predictivo, entrenado con los datos de 50 pacientes, demostró un rendimiento predictivo sólido en el conjunto de prueba. El modelo de Gradient Boosting mostró un AUC de 0.88, mientras que el modelo de Red Neuronal alcanzó un AUC de 0.89, lo que indica una ligera superioridad en la capacidad de la red neuronal para discriminar entre los pacientes que desarrollaron complicaciones neurológicas postagudas y aquellos que no lo hicieron.

Factores de riesgo identificados: El análisis de la importancia de las variables reveló que la edad avanzada fue el predictor más significativo, lo que concuerda con la evidencia existente sobre la vulnerabilidad del cerebro envejecido a la neuroinflamación y el estrés oxidativo. La presencia de diabetes mellitus también emergió como un factor de riesgo crucial, lo que subraya la interacción entre la inflamación crónica sistémica y el daño neurológico. La hipoxemia durante la hospitalización, medida por la saturación de oxígeno, fue otro predictor importante, lo que respalda la hipótesis de que el daño isquémico juega un papel en la patogénesis de estas complicaciones.

Biomarcadores relevantes: Entre los biomarcadores de laboratorio, los niveles elevados de dímero-D se destacaron como un predictor clave. Esto refuerza el papel de la hipercoagulabilidad en las secuelas neurológicas post-COVID-19. Otros biomarcadores inflamatorios como la Proteína C Reactiva también mostraron una importancia moderada, lo que sugiere una contribución de la respuesta inflamatoria desregulada.

Implicaciones clínicas: Los hallazgos de este modelo tienen importantes implicaciones para la práctica clínica. La identificación de pacientes de alto riesgo en una etapa temprana podría facilitar la derivación a clínicas de seguimiento post-COVID-19, donde se podrían implementar estrategias de rehabilitación neurocognitiva y se podría realizar una monitorización más exhaustiva. Un puntaje de riesgo podría integrarse en los registros médicos electrónicos como una herramienta de apoyo a la decisión clínica.

Limitaciones del estudio: Este estudio tiene varias limitaciones. El diseño retrospectivo se basa en la disponibilidad y precisión de los datos registrados, lo que puede introducir un sesgo de selección. La definición de complicaciones neurológicas postagudas puede variar, y la falta de biomarcadores validados dificulta una categorización precisa. Se necesitan estudios prospectivos y con poblaciones más diversas para validar y generalizar estos hallazgos.

Conclusiones

El modelo predictivo desarrollado en este estudio demuestra la viabilidad de utilizar el aprendizaje automático para identificar a los pacientes con mayor riesgo de desarrollar complicaciones neurológicas postagudas de la COVID-19. La edad avanzada, la diabetes mellitus, la hipoxemia y los niveles elevados de dímero-D fueron los principales predictores de riesgo.

Esta herramienta podría ser invaluable para la atención post-COVID-19, permitiendo a los profesionales de la salud intervenir de manera proactiva y personalizada. Los resultados de este estudio no solo abren la puerta a una mejor atención médica, sino que también contribuyen a una comprensión más profunda de la compleja patogénesis de las secuelas neurológicas a largo plazo de la enfermedad.

Referencias

- World Health Organization. (2020). *Clinical management of COVID-19: Living guidance*.
- Marshall, M. (2020). The long-term neurological effects of COVID-19. *Nature Reviews Neurology*, 16(10), 561-562.
- Korallnik, I. J., & Tyler, K. L. (2020). COVID-19: A global pandemic with neurological manifestations. *Annals of Neurology*, 88(1), 1-13.
- Nalbandian, A., Sehgal, J. K., et al. (2021). Post-acute COVID-19 syndrome. *Nature Medicine*, 27(4), 601-615.
- Ellul, M. A., et al. (2020). Neurological associations of COVID-19. *The Lancet Neurology*, 19(9), 767-783.
- Ayoubkhani, D. (2021). Epidemiology of COVID-19 post-acute sequelae. *Nature Medicine*, 27(4), 601-615.
- Al-Shorbagy, A. A., & Al-Hakeem, S. A. (2021). The role of D-dimer in COVID-19: A review. *Journal of Thrombosis and Thrombolysis*, 51(4), 861-870.

Wu, Y., et al. (2020). Nervous system involvement after infection with COVID-19. *Journal of Clinical Neurology*, 16(2), 273-277.

Touret, F., et al. (2020). Putative Neurological Manifestations of COVID-19. *The Lancet*, 395(10237), 1629.

Zhou, F., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229), 1054-1062.