

Sumarización de Textos Médicos Relacionados a COVID-19

Entrega Final

Edwin Ricardo Mahecha Parra, Jimmy Alexander Pulido Arias

I. INTRODUCCIÓN

Una de las tareas que más se destacan en el procesamiento de lenguaje natural es la sumarización de texto. Esta tarea no resulta sencilla puesto a que existen varios elementos que influyen en la generación de resúmenes como lo es la estructura del documento al cual se le está realizando el resumen, las temáticas que aborda, etc. Dicho esto no solo es necesario buscar formas adecuadas de filtrar la información de importancia del un documento, si no también que el resumen resultante tenga coherencia y cohesión.

II. OBJETIVOS Y ALCANCE

Recientemente se han generado grandes volúmenes de datos e información relacionados a la pandemia del COVID-19 desde diversas fuentes como el campo de la medicina, noticias, etc. por lo que hay variedad de documentos que siguen diferentes tipos de estructura según el público al que están dirigidos.

Dicho esto nuestro el objetivo principal de nuestro proyecto es realizar resúmenes a textos de carácter científico(principalmente papers) de tal forma que estos puedan ser utilizados por los actores interesados como lo son investigadores, personal de salud, etc. para identificar de forma eficaz cuales de estos documentos pueden ser de utilidad para sus actividades.

II-A. *Objetivos Específicos*

- Aplicar técnicas de sumarización de textos orientados en áreas de la salud.
- Realizar una revisión de los métodos y tipos de sumarización existentes en la actualidad para determinar qué tipo de sumarización se ajusta mejor a nuestras necesidades y restricciones de tiempo y capacidad computacional.
- Crear API que permita acceder a los resúmenes según ID.

II-B. *Alcance*

Como se mencionó anteriormente los documentos suelen tener diferentes formas y estructuras lo que puede

afectar la aplicación de algún método de sumarización. Dicho esto la idea es centrarse en un solo tipo de documento que mantenga cierto tipo de estructura. Idealmente nos centraremos en documentos de carácter científico(papers) que tienen cierta estructura definida. Por estas razones seleccionamos PubMed Central como repositorio principal de los papers a resumir ya que como se explicará más adelante estos documentos se pueden descargar en un formato estándar que nos facilita su procesamiento y limpieza.

También es importante recordar que existen varias técnicas de sumarización por lo que exploraremos algunas de estas, sin embargo nos quedaremos con la que muestre mejor desempeño y sus resultados sean aceptables.

III. BRIEFING DEL ESTADO DEL ARTE

La sumarización de textos se resume en sistemas capaces de extraer información importante de un texto y comunicarla de forma fluida. Para ello existen varias formas para realizar resúmenes, entre las cuales las más importantes son los sistemas de **sumarización abstractiva** y **sumarización extractiva**.

III-A. *Métodos Extractivos*

Los métodos extractivos funcionan identificando las sentencias más importantes que se den como entrada, que puede ser un documento o una serie de documentos relacionados. La decisión se toma principalmente por la entrada del sistema de sumarización. Usualmente la operación de los sistemas resumidores se divide en 3 tareas independientes: crear una representación intermedia de la entrada que solo captura los aspectos importantes del texto, scoring de frases basado en esa representación y la selección del resumen consistiendo en varias sentencias.

- **Representación Intermedia:** Incluso los sistemas simples realizan alguna representación intermedia del texto que necesitan resumir e identificar el contenido importante basado en esta representación. Uno de los enfoques más utilizados en esta tarea es el **Topic Representation** que convierte el texto

a una representación intermedia interpretada como el tema discutido en el texto.

- **Scoring de frases:** Una vez que la representación intermedia ha sido derivada, a cada frase se le ha asignada un score que indica su importancia. En enfoques como topic representation, este score está usualmente relacionado a que tan bien una frase expresa uno de los temas más importantes en el documento y a que extensión combina información de varios temas.
- **Selección de frases para el resumen:** Finalmente, el resumidor tiene que seleccionar la mejor combinación de frases importante para formar un resumen de tamaño de un párrafo. En el enfoque *best n approaches*, las n frases más importantes que combinadas forman el resumen esperado son seleccionadas para formar el resumen. También existen otros enfoques en esta etapa como: *n maximal marginal relevance* y *global selection*.

Hay muchos factores que toman juego para la categorización de las frases para el resumen, si tenemos información acerca del contexto en el que el resumen es generado, esto puede ayudar en determinar importancia. El Contexto puede tener la forma de información acerca de las necesidades del usuario, usualmente representada mediante una query. Contexto puede incluir el ambiente el cual un documento de entrada es situado, como los links que apuntan a una página web. otro factor importante es el género de un documento. Ya sea que el documento de entrada sea un artículo, un hilo de correos, una página web o un artículo de prensa esto influye la estrategia utilizada para seleccionar las frases.

III-B. Métodos Abstractivos

A diferencia de los métodos extractivos que extrae las sentencias principales del texto y las combina, los métodos basados en abstracción generan nuevas sentencias a partir de la información dada por el texto, pero pueden lograr una mayor cohesión entre las frases generadas ya que puede incluso introducir nuevas palabras que no se encontraban originalmente en el documento.

Muchos de los métodos para preprocesar los textos antes de aplicar un algoritmo abstractivo como por ejemplo representación intermedia son los mismos métodos o similares a los que se utilizan en extracción, por lo que no se ahondará en ello.

En los últimos años los métodos abstractivos se han decantado más hacia métodos basados en redes neuronales

y deep learning. A continuación presentamos una lista con algunas de estas técnicas y métodos relacionados:

III-C. Enfoques para Sumarización de textos

A continuación veremos dos tablas con las técnicas y enfoques más utilizados para la generación de resúmenes de textos:

Cuadro I
ENFOQUES Y TÉCNICAS EXTRACTIVAS

Enfoques y Técnicas Extractivas	
Topic Representation	Topic Words
	Frequency-driven Approaches
	Latent Semantic Analysis
	Bayesian Topic Models
	Sentence Clustering and Domain-dependent Topics
Influence of Context	Web Summarization
	Summarization of Scientific Articles
	Query-focused Summarization
	Email Summarization
Indicator Representations and Machine Learning for Summarization	Graph Methods for Sentence Importance Machine Learning for Summarization
Summarization Selecting Summary	Greedy Approaches:
Sentences	Maximal Marginal Relevance
	Global Summary Selection

Cuadro II
ENFOQUES Y TÉCNICAS ABSTRACTIVAS

Enfoques y técnicas abstractivas		
Neural Attention Model	RNN encoder-decoder	LSTM
		GRU
		NAS
Attention models and Transformers		
Large Vocabulary Trick and Feature-rich Encoder		
Hierarchical Attention		
Pointer Generator Network		
Coverage Mechanism (avoid repetition)		
Intra-Attention on Decoder outputs		
Reinforcement Learning	Policy Gradient-based Reinforcement Learning	

IV. SUMARIZACIÓN BASADA EN MODELOS DE ATENCIÓN

Como pudimos apreciar existen gran variedad de modelos tanto extractivos como abstractivos. En un principio se penso en realizar una sumarización extractiva puesto a que en comparación a los métodos abstractivos basados en redes neuronales, estos requieren menos recursos computacionales, sin embargo actualmente acceder a una GPU y a modelos preentrenados ya no es un problema por lo que se decidió tomar este camino ya que los resultados pueden ser más satisfactorios.

IV-A. RNN - Recurrent Neural Networks

Las redes neuronales recurrentes (RNN) son modelos 'conexionistas' que capturan la dinámica de secuencias por medio de ciclos en una red de nodos. A diferencia de las redes 'feedforward' estándar, las redes neuronales recurrentes mantienen un estado que puede representar información de una ventana de contexto arbitrariamente grande. [1]

IV-A1. LSTM: LSTM es una de las arquitecturas RNN más conocidas. Esta introduce el concepto de 'memory cell' lo que permite a las redes solucionar problemas que se presentaban anteriormente con las RNN como lo es *vanishing gradient* y *exploding gradient*

Durante varios años se consideraron a las RNN y sus variantes como el estado del arte para modelar secuencias y problemas de transducción, sobre todo orientadas al modelos de lenguaje y arquitecturas encoder-decoder. Si bien su desempeño ha sido bueno para cierto tipo de problemas, hay un problema inherente a este tipo de arquitecturas, y es que de cierta forma su comportamiento es secuencial. En una RNN, el estado oculto h_t depende directamente del estado anterior h_{t-1} por lo que no hace uso adecuado de los recursos computacionales como por ejemplo GPUs, ya que solo se puede hacer paralelismo entre ejemplos de entrenamiento, lo cual es crítico para secuencias largas.[2]

Debido a esto se plantean los modelos de atención con Transformers de los cuales trataremos a continuación y los cuales utilizaremos para realizar nuestra tarea de sumarización.

IV-B. Transformers y Modelos de Atención

Como se dijo anteriormente los modelos basados en RNN no son tan paralelizables en comparación a una red 'feedforward'. Debido a esto, la idea principal de los transformers es utilizar una arquitectura tradicional con ciertas consideraciones.

En la figura 1 se puede apreciar la arquitectura de Transformers para una tarea que requiere encoding y decoding, como por ejemplo para tareas del tipo seq2seq. A continuación explicaremos estos elementos de forma muy general para entender como estos modelos pueden modelar secuencias efectivamente[2].

- **Input/Output embeddings:** Esta capa consiste en simplemente realizar embeddings (usualmente word2vec) de tal forma que se obtenga una representación vectorial de nuestro texto.
- **Positional encoding:** Debido a que el modelo no contiene recurrencias, para que el modelo tenga

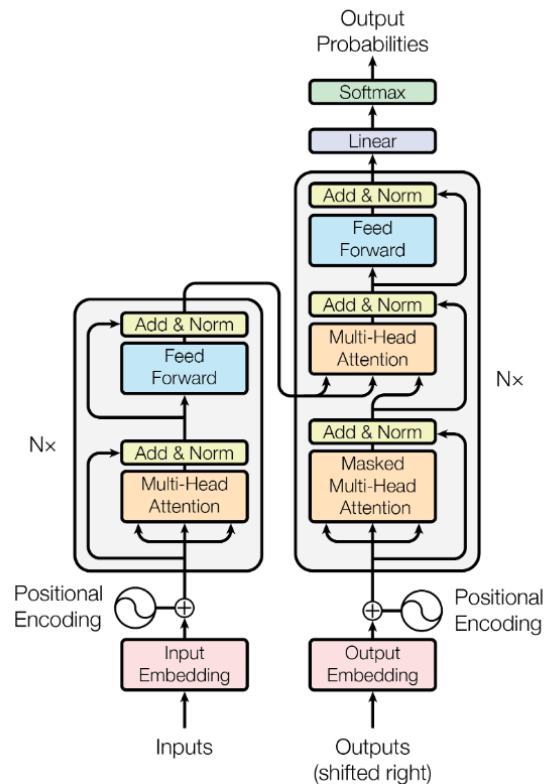


Figura 1. Arquitectura de un Transformer para encoding-decoding. [2]

información de orden, es decir, que sepa si una palabra se ubica antes o después respecto a otras, la idea es realizar un 'encoding posicional', de tal forma que se integra la posición de cada palabra dentro de su encoding. No discutiremos como funciona, pero la idea principal es hacer uso de senos y cosenos en conjunto con el desplazamiento de la frecuencia.

- **Multi-Head Attention:** Una función de atención toma n inputs y retorna n outputs. La idea es identificar la relevancia de cada palabra respecto a una en particular. Esto se realiza para todas las palabras. Cuando hablamos de *multi-head attention* la idea es principalmente prestar atención a diferentes elementos del lenguaje como gramática, vocabulario, etc.

Los elementos mencionados anteriormente están presentes tanto en el encoder como en el decoder. Los recuadros grises de la figura 1 son elementos que se pueden anidar varias veces (N).

La arquitectura que presentamos anteriormente ha inspi-

rado gran cantidad de modelos como BERT, GPT, etc. que hacen uso de transformers. Para el desarrollo del proyecto se decidió hacer uso de uno de estos modelos, en específico el modelo de Google T5.

IV-B1. T5 - Text-To-Text Transfer Transformer: Modelos como BERT solo pueden dar como resultado una etiqueta, o segmentos del input. La idea principal de T5 es que con un único modelo y los mismos pesos se puede crear una función que pueda procesar 'cualquier' tarea de NLP, incluyendo machine translation, summarización de documentos, etc. Este modelo fue entrenado sobre el dataset C4 - Colossal Clean Crawled Corpus cuyo tamaño es de alrededor de 803.19GiB[10].

Respecto al modelo preentrenado que se utilizó, T5 se encuentra en la librería de transformers de hugging face y puede utilizarse de forma sencilla, e incluso realizar fine tuning para la tarea necesaria[3].

V. OBTENCIÓN DE DATASETS

Actualmente existen gran cantidad de fuentes de datos sobre los cuales se puede realizar procesamiento de lenguaje natural. Para nuestro caso de estudio que va orientado principalmente a los papers relacionados con covid-19 encontramos las siguientes fuentes de datos:

V-A. NCBI

La NCBI (National Center for Biotechnology Information) provee acceso a diversas fuentes de datos, entre las cuales se encuentran PubChem, Pubmed, PMC(PubMed Central) entre otras. Para nuestro caso de estudio el servicio dispone de una colección para minería de texto en la cual se encuentran los siguientes conjuntos de datos[5]:

1. **Open Access (OA) Subset:** Son artículos disponibles para ser descargados bajo Creative Commons o una licencia similar que permite una mayor facilidad en la distribución y reuso de trabajo con derechos de autor. La descarga de artículos de OA dispone del texto completo en formatos como XML, PDF y txt además de las imágenes y material suplementario a los artículos.
2. **Author Manuscript Collection:** Es una colección que contiene el texto completo en XML o txt de los manuscritos que se encuentran disponibles en PMC en cumplimiento con las políticas de acceso público de la NIH. La colección contiene todos los manuscritos publicados en PMC desde Julio del 2008.
3. **Historical OCR Collection:** Esta colección contiene más de 2 décadas de investigación biomédica

de un subconjunto de revistas que participaron en el NLM's Back Issue Digitization Project (2014-) y el Journal Backfiles Digitization Project (2004-2010).

4. **LitCovid:** Es un dataset curado y actualizado sobre el novel 2019 coronavirus. En este dataset contiene información tal como etiquetas por categorías para cada artículo, abstract, etc. y además, si el artículo se encuentra en el *Open Access Subset* podemos encontrar el texto del artículo en un formato más sencillo de procesar, en comparación a formatos como XML o PDF[7][8][9].

Para nuestro caso de estudio nos interesan *OA subset*, *Author Manuscript Collection* y *LitCovid* ya que se puede extraer información fácilmente es estos conjuntos de datos. *LitCovid* tal vez es la fuente de datos más importante de las que mencionamos anteriormente ya que la mayoría de textos del *OA subset* se encuentran procesados en este y nos evita la tarea de preprocesamiento y extracción de información, sin embargo, si en un futuro se quisiera extender el API para otros temas diferentes al coronavirus nos limitaría. Debido a esto se consideran los otros dos datasets.

Para acceder a los otros datasets la NCBI provee los siguientes métodos:

1. **Entrez Programming Utilities (E-utilities):** Es un conjunto de nueve utilidades que proveen una interface estable para acceder al sistema de peticiones y bases de datos de la NCBI(Entrez). E-utilities es la interfaz estructurada para acceder al sistema Entrez que incluye 38 bases de datos sobre variedad de datos biomédicos, incluyendo secuencias de proteínas y nucleótidos, registros genéticos, literatura biomédica entre otros[4].
2. **FTP Service** Permite el acceso a las tres colecciones de texto mencionadas anteriormente a través de un servidor FTP.
3. **OAI-PMH Service** Provee el acceso a los artículos disponibles haciendo uso del formato *OAI-PMH version 2.0*
4. **OA Web Service** Provee archivos en formato tgz o PDF tomados del FTP. Los archivos en formato tgz contienen tanto el artículo en PDF, XML y recursos relacionados como imágenes.

VI. PREPROCESAMIENTO¹

Para realizar esta tarea fue necesario determinar que fuente de datos resultaba más conveniente ya que nos

¹Does not apply for LitCovid

concentramos principalmente en la tarea de sumarización de textos y no nos interesa realizar tareas grandes de extracción y limpieza de los textos. Por ello consideramos que realizar crawling de PDFs no resulta viable ya que requeriría preprocesar los documentos para obtener solamente el texto relevante (sin incluir las etiquetas a figuras, formulas matemáticas, imágenes, etc).

Después de revisar cuidadosamente los métodos para acceder a los conjuntos de datos decidimos hacer uso del servicio OAI-PMH de la NCBI con lo que podemos acceder fácilmente a las bases de datos de PubMed y PMC y extraer los artículos en formato XML bajo la especificación JATS (Journal Archiving and Interchange Tag Set)[6] lo que nos facilita la tarea de limpieza de los textos ya que bajo esta especificación se define un tag para cada elemento que compone al texto como lo son figuras, formulas matemáticas, citas, etc.

VI-A. Limpieza de Datos

Puesto a que estamos usando archivos XML bajo la especificación JATS, es posible determinar que elementos del texto no son relevantes y eliminarlos de forma recursiva del árbol XML. Algunos de los tags que decidimos eliminar fueron:

- <fig>
- <code>
- <graphic>
- <text-math>
- entre otros ...

Para mayor información de los tags ver: <https://jats.nlm.nih.gov/archiving/tag-library/1.3d1/index.html>

Después de realizar esta limpieza de tags quedamos con un texto sin elementos como formulas matemáticas, figuras y demás por lo que ya está listo para ser procesado y realizar la sumarización.

VII. OBTENCIÓN DE RESUMENES

Se obtuvieron alrededor de 6700 resúmenes de artículos médicos relacionados al COVID-19 del repositorio de PubMed, estos resúmenes fueron generados utilizando el Transformer Text-To-Text T5 desarrollado por Google. [10] A continuación veremos algunos ejemplos de los resúmenes generados:

VII-1. Ejemplos de Resumen: Los siguientes ejemplos cuentan con el título, la cita del artículo original y el resumen generado.

1. **Surgical Strategy During the COVID-19 Pandemic in a University Metropolitan Hospital in Milan, Italy** [11]:

The global pandemic of coronavirus disease 2019 started as an atypical pneumonia. the disease leads to severe acute respiratory syndrome and resulted in thousands of deaths. in Italy, the hardest hit region has been Lombardy, where the first case was reported on February 20. important changes in surgical activities have been introduced by hospitals in response to COVID-19-related challenges, says dr. san pellegrini.

2. **COVID-19 critical illness pathophysiology driven by diffuse pulmonary thrombi and pulmonary endothelial dysfunction responsive to thrombolysis** [12]:

Patients with severe COVID-19-induced respiratory failure demonstrate gas exchange abnormalities including shunt and dead-space ventilation. their syndrome is atypical in that the majority have relatively well-preserved lung mechanics. high rate of venous thromboembolism (VTE) in critically ill COVID-19 patients. d-dimer levels have also been noted to be elevated. rapid rises presage cardiopulmonary decompensation.

3. **The potential impact of COVID-19-related disruption on tuberculosis burden** [13]:

Before the COVID-19 pandemic, over 4000 people were dying from tuberculosis (TB) every day. the impact of COVID-19 on TB outcomes is a serious cause for concern but is currently unknown. physical distancing interventions could also limit Mycobacterium tuberculosis transmission outside of households, where most transmission occurs. this has not been adequately explored in existing work. cdc: impact of COVID-19 on TB outcomes is a

4. **What's New With the Old Coronaviruses?** [14]:

A coronavirus was first isolated as a causative agent of bronchitis in birds in 1937. it was originally discovered in humans during studies that evaluated the common cold. the novel coronavirus SARS-CoV-2 is spreading worldwide, causing anxiety, disease, and mortality. limited pediatric data are available regarding HCoV infections in children in the pre-coronavirus disease 2019 (COVID-19) era.

VIII. API

Uno de los objetivos de este trabajo es la creación de una API que permita acceder a los resúmenes de los artículos en cualquier momento. Para esto se crearon los

siguientes dos endpoints: ²

- **api_summaries:** Este endpoint es de tipo POST, recibe un json con el pmcid del artículo de PubMed y retorna el resumen de ese artículo. ³
- **api_summaries_ids:** Este endpoint es de tipo GET, no recibe parámetros y retorna todos los ids de los papers que tienen resumen disponible. ⁴

IX. RESULTADOS

Al observar los resúmenes obtenidos de los artículos podemos ver que se obtienen muy buenos resultados. Por ejemplo el resumen del artículo No. 1 obtiene contenido muy relevante, identifica el lugar donde ocurren los hechos el cual es Italia, además el tema principal que trata sobre una nueva estrategia quirúrgica introducida en hospitales en respuesta a la pandemia del COVID-19. Otro buen ejemplo es el resumen del artículo No. 3 que aborda el impacto potencial que puede tener el COVID-19 sobre pacientes con antecedentes de Tuberculosis, el modelo identifica muy bien el tópico ya que se ve reflejado en el resumen. El resumen 3 también es un ejemplo donde el modelo puede llegar a quedar corto, ya que en el final "impact of COVID-19 on TB outcomes is a..." se puede ver que faltó completar esta última frase, una causa de esto pudo ser las restricciones que se le pusieron al modelo en cuanto a la longitud máximo del resumen.

X. CONCLUSIONES

La obtención de datos de COVID-19 fue muy conveniente, debido a que existen varias organizaciones que prepararon repositorios exclusivos para la exploración de artículos médicos relacionados al COVID-9, de esta manera nos fue posible extraer los artículos y formar un conjunto de datos lo suficiente grande para el desarrollo de este proyecto.

Los enfoques de modelos de atención son en el estado de arte para la generación de resúmenes abstractivos de texto, por lo cual tienen un desempeño muy destacable. Al utilizar transfer-learning con el modelo transformador de transferencia de text-to-text T5 creado por google en febrero del 2020, los resultados fueron los esperados, además nos facilitó la tarea de generar estos resúmenes al ser un modelo pre entrenado y de multipropósito.

²Para una descripción más detallada de esta API, por favor consultar el repositorio de Github <https://github.com/PLN-Collaborative-Project/Summaries>

³api_summaries: https://us-central1-nlp-final-281200.cloudfunctions.net/api_summaries

⁴api_summaries_ids: https://us-central1-nlp-final-281200.cloudfunctions.net/api_summaries_ids

Cabe destacar que el modelo T5 fue entrenado sobre un corpus de 805 GiB, el cual se compone de documentos extraídos de internet. Es posible mejorar nuestros resultados si se realizara fine tuning del modelo sobre un conjunto de datos relacionado a áreas de la salud, y en específico COVID-19.

REFERENCIAS

- [1] L. Zachary & Berkowitz, J., 2015. *A Critical Review Of Recurrent Neural Networks For Sequence Learning*. [online] Arxiv.org. Available at: <https://arxiv.org/pdf/1506.00019.pdf>.
- [2] A. Vaswani, Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. & Kaiser L. *Attention is All you Need* [online] Arxiv.org. Available at: <https://arxiv.org/pdf/1706.03762.pdf>.
- [3] Huggingface.co. n.d. Transformers — Transformers 2.11.0 Documentation. [online] Available at: <https://huggingface.co/transformers/>.
- [4] *Entrez Programming Utilities Help* [online]. Bethesda (MD): National Center for Biotechnology Information (US); 2010-. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK25501/>
- [5] *Text Mining Collections* [online]. Bethesda (MD): National Center for Biotechnology Information (US); 2010-. Available at: <https://www.ncbi.nlm.nih.gov/pmc/tools/textmining/>
- [6] *JATS - Journal Archiving and Interchange Tag Set*. [online] <https://jats.nlm.nih.gov/archiving/>
- [7] Chen, Q., Allot, A., & Lu, Z. (2020). *Keep up with the latest coronavirus research*. *Nature*, 579(7798), 193. doi: 10.1038/d41586-020-00694-1
- [8] Comeau, D. C., Wei, C. H., Islamaj, R., & Lu, Z. (2019). *PMC text mining subset in BioC: about three million full-text articles and growing*. *Bioinformatics*, 35(18), 3533-3535. btz070.
- [9] Wei, C. H., Allot, A., Leaman, R., & Lu, Z. (2019). *PubTator central: automated concept annotation for biomedical full text articles*. *Nucleic acids research*, 47(W1), W587-W593. doi: 10.1093/nar/gkz389
- [10] *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*; Colin Raffel and Noam Shazeer and Adam Roberts and Katherine Lee and Sharan Narang and Michael Matena and Yanqi Zhou and Wei Li and Peter J. Liu, 2019, 1910.10683, arXiv.cs.LG
- [11] Mariani, N. M., Pisani Ceretti, A., Fedele, V., Barabino, M., Nicastro, V., Giovenzana, M., Scifo, G., De Nicola, E., & Opocher, E. (2020). *Surgical Strategy During the COVID-19 Pandemic in a University Metropolitan Hospital in Milan, Italy*. *World journal of surgery*, 1–6. Advance online publication. <https://doi.org/10.1007/s00268-020-05595-y>
- [12] Poor, H. D., Ventetuo, C. E., Tolbert, T., Chun, G., Serrao, G., Zeidman, A., Dangayach, N. S., Olin, J., Kohli-Seth, R., & Powell, C. A. (2020). *COVID-19 critical illness pathophysiology driven by diffuse pulmonary thrombi and pulmonary endothelial dysfunction responsive to thrombolysis*. *Clinical and translational medicine*, 10.1002/ctm2.44. Advance online publication. <https://doi.org/10.1002/ctm2.44>

- [13] McQuaid, C. F., McCreesh, N., Read, J. M., Sumner, T., Houben, R., White, R. G., Harris, R. C., & CMMID COVID-19 Working Group (2020). The potential impact of COVID-19-related disruption on tuberculosis burden. *The European respiratory journal*, 2001718. Advance online publication. <https://doi.org/10.1183/13993003.01718-2020>
- [14] Ogimi, C., Kim, Y. J., Martin, E. T., Huh, H. J., Chiu, C. H., & Englund, J. A. (2020). What's New With the Old Coronaviruses?. *Journal of the Pediatric Infectious Diseases Society*, 9(2), 210–217. <https://doi.org/10.1093/jpids/piaa037>