

[Centro de Información de COVID \(CIC\): Charlas científicas de relámpago](#)



Transcripción de una presentación de Dominique Duncan (Universidad del Sur de California), enero de 2021

Título: [COVID-ARC \(Archivo de Datos COVID-19\)](#)

[Perfil de Dominique Duncan en la base de datos de CIC](#)

Subvención de La Fundación Nacional de Ciencias (NSF, por sus siglas en inglés) #: [2027456](#)

[Grabación de YouTube con diapositivas](#)

[Información del seminario web del CIC de enero 2022](#)

Editora de la Transcripción: Saanya Subasinghe

Editora de la Traducción: Isabella Graham Martínez

Transcripción

Dominique Duncan:

Diapositiva 1

Bien, gracias a Florence y Lauren y a todos los que están detrás del equipo de COVID Info Commons. Primero di una charla hace un año y medio y al final de esta charla daré algunas actualizaciones sobre lo que resultó de eso y una nueva colaboración que se formó. Pero quería hablar de nuestro Archivo de Datos COVID-19, COVID-ARC para abreviar. Soy profesor asistente en la Universidad del Sur de California en la Escuela de Medicina Keck en el Laboratorio de Neuroimágenes.

Diapositiva 2

Por lo tanto, en nuestro laboratorio tenemos mucha experiencia en la construcción de archivos de datos multimodales a gran escala, principalmente para datos cerebrales. Pero al comienzo de la pandemia, pensamos que podríamos usar nuestra experiencia y recursos con todos esos archivos de datos y desarrollar un archivo de datos COVID-19. Así que recibimos - fuimos galardonados con un premio NSF RAPID para desarrollar este archivo de datos llamado COVID-ARC. Así que lo que hacemos es agregar diferentes tipos de datos COVID-19 así como recursos y hemos construido una plataforma de archivos centralizados y en red que almacenan, curan, visualizan y difunden datos COVID-19 multimodales. Tenemos muchos conjuntos de datos de todo el mundo. Muchos de ellos están disponibles públicamente, algunos son privados, por lo que hemos trabajado con esos proveedores de datos para

llegar a acuerdos de uso de datos que se adaptan a sus necesidades. Y tenemos los metadatos disponibles en el sitio web, por lo que si los usuarios desean solicitar acceso a ellos, facilitamos ese proceso. Pero luego los proveedores de datos toman la decisión final. Y muchos de los datos se almacenan en nuestro sitio en USC, pero para algunos de los conjuntos de datos se almacenan en el sitio donde se recopilaron los datos y tenemos los metadatos disponibles para que la gente pueda ver eso. Y mucho de nuestro trabajo ha sido sobre la armonización de los metadatos y esto es para facilitar la investigación sobre cohortes agrupadas para que sea más fácil para las personas hacer diferentes tipos de análisis a través de diferentes sitios en lugar de centrarse en un sitio. En un par de diapositivas, hablaré sobre algunos de esos desafíos y por qué trabajamos en armonizarlo. Y también hemos integrado herramientas de visualización y control de calidad y herramientas analíticas. Una vez más, para ayudar a los investigadores - solo para acelerar la investigación sobre COVID-19. Y además de todo el trabajo en torno al archivo y armonización de datos, también estamos haciendo diferentes tipos de análisis sobre los datos que tenemos y estamos utilizando los principios de retroalimentación y la ciencia de datos para estudiar varios aspectos de COVID-19.

Diapositiva 3

Por lo tanto, en términos de los datos que tenemos diferentes tipos de datos - una gran cantidad se centra en solo imágenes de TC, así como de rayos X, pero tenemos datos clínicos que incluye síntomas, signos vitales, comorbilidades, demografía, historia del paciente, geolocalización. También tenemos, para imágenes, ultrasonido y resonancia magnética, así como algunos datos de electroencefalograma. Y también hemos proporcionado máscaras pulmonares, máscaras de infección y anotaciones del radiólogo. Y aquí pueden ver que usamos el sistema de transferencia de archivos encriptado HIPAA de IBM llamado ASPERA y esta es una manera muy fácil para que los proveedores de datos transfieran datos a COVID-ARC, así como para que los usuarios descarguen datos de COVID-ARC a sus computadoras.

Diapositiva 4

Y ahora mismo tenemos 28 conjuntos de datos de todo el mundo. Como puede imaginar, hay algunos nombres de archivos inconsistentes en estos diferentes conjuntos de datos, formato de metadatos inconsistente, diferencias en el almacenamiento de infraestructuras, así como otras diferencias entre esos conjuntos de datos. Así que lo que hemos hecho es poner todo esto junto en un archivo de datos centralizado y nos hemos asegurado de que haya un nombre y una organización de archivos consistentes, un formato de metadatos consistente y facilidad para descargar varios conjuntos de datos desde una ubicación usando ASPERA.

Diapositiva 5

Y aquí hay solo una captura de pantalla de parte de los datos que tenemos. No pude encajar todo en una diapositiva, pero puedes ver si vas a covid-arc.loni.usc.edu, puedes encontrar qué datos tenemos. Y esto está organizado podemos ver que tenemos el número de sitio, la ubicación donde se recogieron los datos, las modalidades, los formatos de archivo, cualquier metadato que tengamos además de eso, y luego cuántas imágenes hay. Y algunos de ellos se dividen entre COVID y no COVID. Y también tenemos información sobre si está o no disponible públicamente.

Diapositiva 6

Así que ahora, solo quería destacar algunos de los proyectos en los que mis estudiantes han estado trabajando, han sido muy productivos y han estado haciendo una investigación realmente emocionante. Tengo algunas experiencias de investigación de la NSF para estudiantes y Aksh Garg es una de ellas. Empezó cuando estaba en el instituto y ahora es estudiante un primer año de licenciatura en Stanford y apenas la semana pasada su documento fue aceptado en *Expert Systems with Applications* y aquí hizo una comparación de 40 arquitecturas de red neuronal convolucional para distinguir COVID versus no COVID y encontró que el mejor modelo EfficientNet-B5 produjo sensibilidad y especificidad de precisión extremadamente alta. Y el modelo también se basó en características clínicamente relevantes como opacidades y consolidaciones de vidrio molido, que a menudo se ven en pacientes con COVID 19.

Diapositiva 7

Otro proyecto - Alex Bruckhaus es otro galardonado de REU, por lo que este trabajo fue publicado en el *Journal of Immigrant and Minority Health* y aquí él y otros estudiantes estaban mirando la dinámica de vacunación en California. Así que miraron algo llamado el Índice de Vulnerabilidad Social y el SVI tiene - miraron cuatro temas del SVI incluyendo el estatus socioeconómico, la composición del hogar y la discapacidad, el tipo de vivienda y el transporte y el estatus y el idioma de las minorías. Y encontraron que la cobertura de vacunación más baja estaba en los grupos de alta vulnerabilidad. La situación de las minorías y el idioma produjeron la mayor disparidad en la cobertura entre los condados de baja y alta vulnerabilidad. Así que creo que este es un trabajo realmente importante, especialmente porque estamos tratando de vacunar a más de la población.

Diapositiva 8

Otro, otro artículo que Alex Bruckhaus y otros estudiantes publicaron estaba analizando las tasas de infección posteriores al bloqueo tras las reaberturas. Así que miraron 83 condados en todo Estados Unidos con altos recuentos de casos COVID-19 el año pasado. Y estaban mirando diferentes tipos de negocios y se separaron entre una reapertura completa o una reapertura parcial y estaban mirando los cambios en la tasa de infección antes y después de esas reaberturas y viendo qué negocios tuvieron el mayor efecto en un aumento de la infección tasas. Y así bares y gimnasios jugaron un importante - un gran papel en eso.

Diapositiva 9

Yujia Zhang, que es mi asistente de proyecto en este proyecto, hizo un trabajo de revisión el año pasado mirando la asociación del tipo de sangre con COVID-19. Así que miró a 23 estudios que tenían una visión general del tipo de sangre como factores de riesgo y de protección, cómo - cómo ciertos tipos de sangre personas con ciertos tipos de sangre son susceptibles para las pruebas positivas y los resultados clínicos de gravedad. Y también pasó por las asociaciones genéticas y los potenciales mecanismos moleculares subyacentes allí.

Diapositiva 10

Y luego Azrin Khan, que ha sido becaria de REU los últimos dos veranos, está trabajando en este proyecto que analiza la segmentación pulmonar basada en el umbral. Así que este es el método de umbrales de múltiples pasos para cuantificar las anomalías pulmonares con un mejor rendimiento que los métodos existentes.

Diapositiva 11

Y como me estoy quedando sin tiempo voy a apresurarme un poco con esto pero quería hablar de una colaboración que comenzó debido a la primera charla de COVID Info Commons que di y Michael Pazzani y Albert Hsiao de UCSD, de San Diego, también dieron charlas y empezamos una colaboración después de eso. También presentamos una propuesta de salud inteligente que no fue financiada, pero volvimos a presentarla en noviembre pasado, así que estamos esperando por eso. Pero esto es solo una captura de pantalla de uno de los seminarios web de alcance que tuvimos para estudiantes de secundaria en el sur de California, y colectivamente nuestros estudiantes dieron charlas relámpago sobre su proyecto y eso fue, eso fue muy exitoso. Y solo quería darte las gracias. Este es el sitio web del laboratorio el sitio web de COVID-ARC, por favor envíeme un correo electrónico [duncand@usc.edu] si tiene alguna pregunta Y gracias tanto a NSF y NIH por la financiación.