

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



[Transcripción de una presentación de Samson Qian \(Universidad de California, San Diego\), 18 de agosto de 2021](#)

[Título: \*Generando Explicaciones para las Predicciones de Neumonía en el Escáner Médico de Tórax\*](#)

[Grabación de YouTube con diapositivas](#)

[Información del seminario web del CIC de Agosto 2021](#)

[Editora de la Transcripción:](#) Macy Moujabber

[Editora de la Traducción:](#) Isabella Graham Martínez

---

### Transcripción

Lauren Close:

Me gustaría presentar a nuestro último orador hoy, Samson Qian. Samson es uno de nuestros tres ganadores del desafío inaugural de papel para estudiantes de pregrado del CIC que se llevó a cabo a principios de esta primavera. Así que damos la bienvenida a Samson y estamos muy contentos de compartir su investigación como un erudito emergente. Samson, por favor, llévatelo.

Samson Qian:

*Diapositiva 1*

Gracias a ti Lauren. Así que hoy voy a hacer una presentación que también es ligeramente diferente de los anteriores sobre la generación de explicaciones para las predicciones de aprendizaje automático sobre la neumonía viral y bacteriana.

*Diapositiva 2*

Por lo tanto, el objetivo general de esta investigación es tomar varios pacientes de rayos X de tórax que tienen bacterias y neumonía viral, así como pacientes sanos y construir un modelo de aprendizaje automático para clasificar entre estos pacientes utilizando estas imágenes. Y entonces, ¿es posible construir un modelo de aprendizaje automático que pueda identificar con precisión estas diferentes clases, pero también interpretar este modelo de aprendizaje automático para entender dónde está mirando el modelo cuando está generando predicciones? Y esto introduce la idea de usar algoritmos explicativos, también conocidos como IA explicable, para analizar un modelo y los datos que predice para

entender exactamente dónde está mirando el modelo. Y el objetivo final es ver si este tipo de marco puede ayudar a los radiólogos en su trabajo en el diagnóstico de diferentes pacientes.

### *Diapositiva 3*

Así que una rápida visión general de la IA explicable. Hay varios tipos de algoritmos de IA explicables, como, más como una rama familiar de diferentes tipos de métodos. Para esta investigación, en particular, nos centraremos en algoritmos post-Hoc que significan algoritmos para analizar modelos complicados después de que se entrenan en los datos en lugar de durante el proceso o antes. Y el beneficio de esto es que puedes usar modelos más complejos, como redes neuronales convolucionales más profundas para aprender las características de los rayos X. Y hay un compromiso entre la interpretabilidad y la precisión, a veces. Y así, los modelos más profundos y complejos son más difíciles de interpretar que los más simples, por lo que estos algoritmos proporcionan una manera de entender estos modelos más complejos. Así que el primer tipo de algoritmo se conoce como LRP, Propagación de Relevancia por Capas, y este es un método del que hablaré más adelante, pero esto esencialmente se ve en las capas y pesos del modelo y entiende qué píxeles en la imagen contribuyen más a la activación de la estructura interna de un modelo.

El segundo es algo conocido como LIME y este método es modelo-agnóstico, significando que no depende de qué tipo de modelo usted utilizó. Este método difiere de LRP en que parte de los datos en primer lugar, en lugar del modelo y mira las subsecciones individuales de los datos para encontrar qué regiones en la imagen contribuyen más a la predicción de un modelo.

Y luego el siguiente se llama Grad-CAM, que es similar al tipo de LRP, pero en lugar de mirar cada capa y cada peso, echa un vistazo a las capas de convolución en el modelo y luego toma los gradientes que cuando se ajusta a una imagen, los gradientes de la última capa convolucional, y luego produce una especie de mapa de activación de los gradientes.

Y el último tipo de algoritmo es más bien un algoritmo nuevo conocido como Contrastive LRP que es una modificación de la LRP regular, pero toma las relevancias y luego difiere entre diferentes clases como la neumonía bacteriana viral, para que pueda visualizar la diferencia entre las clases más.

### *Diapositiva 4*

Así que construir una red neuronal convolucional para identificar las características en este tipo de pacientes requiere una estructura de modelo compleja como VGG16 y ResNet50 que son dos modelos de última generación que hacen un muy buen trabajo en la clasificación de imágenes, y como se puede ver en las estructuras del modelo, es una red convolucional muy profunda que es capaz de recoger muchas características en las imágenes. Y así estos dos modelos dan algunos de los mejores resultados en la clasificación entre las tres clases diferentes de pacientes.

### *Diapositiva 5*

Así que las imágenes de rayos X se recogieron de los datos de Mendeley y se requiere un montón de pre-procesamiento de imágenes en las imágenes para encajar en el modelo. Así que asegurarse de que hay una cantidad similar de ejemplos de imágenes de todos los diferentes tipos de clases y luego hacer un poco de pre-procesamiento en las imágenes para encajar en el modelo.

### *Diapositiva 6*

Y aquí hay como una imagen de solo una visión general del rendimiento del modelo durante el entrenamiento. Así que a la izquierda tienes la gráfica de precisión para el conjunto de entrenamiento en el conjunto de validación de imágenes. Y a medida que entrenas el modelo más y más cada iteración, cada época, la precisión crece más y luego la pérdida que es lo que estamos tratando de minimizar en el lado derecho está disminuyendo constantemente. Así que esto implica- esto muestra que nuestro modelo está recogiendo las características de los datos- las imágenes que estamos alimentando y está haciendo un buen trabajo en la clasificación entre las radiografías de tórax.

### *Diapositiva 7*

Así que aquí hay más análisis sobre el modelo construido, y es muy importante tener un modelo preciso con el fin de identificar o examinar lo que el modelo está mirando en cada imagen. Así que se puede ver una matriz de confusión a la izquierda que muestra lo que el modelo es confuso y en general está haciendo un muy buen trabajo entre las tres clases además de distinguir entre neumonía viral y bacteriana que está haciendo- se está confundiendo en unos pocos ejemplos. Por lo tanto, es importante ejecutar estos algoritmos de explicación para entender lo que el modelo está mirando.

### *Diapositiva 8*

Así que aquí hay un breve resumen de LRP, Propagación de Relevancia de Nivel, de nuevo, y esencialmente lo que esto hace es que tienes tu red neuronal con cada capa, una capa oculta dentro. Y propaga tu imagen desde la capa de salida hacia atrás. Así que en lugar de propagarse hacia adelante, se propaga en la imagen hacia atrás y luego se calculan las puntuaciones relevantes en cada neurona, que es estos círculos aquí. Y estas puntuaciones relevantes representan lo importante que es cada píxel en la imagen para una predicción que hace el modelo.

### *Diapositiva 9*

Y aquí puedes echar un vistazo a la derecha es un ejemplo de LRP. Y como se puede ver, produce una especie de mapa de calor basado en los píxeles individuales de la imagen y que muestra lo importante que cada píxel en la imagen está contribuyendo a la predicción del modelo en la clase. Y luego a la izquierda se puede ver una comparación de LIME que es más un método basado en regiones que toma

regiones en la imagen, y luego encuentra qué regiones son más importantes para hacer una predicción modelo.

#### *Diapositiva 10*

Y aquí hay más ejemplos de LIME para las tres clases diferentes. Lo que hace LIME- no va específicamente en cada capa del modelo para examinar los pesos, sino que toma sub-regiones elegidas específicamente en la imagen y se ejecuta en el modelo para determinar cuál de las sub-regiones es más importante para la predicción del modelo y destaca en esas regiones específicas.

#### *Diapositiva 11*

Y luego otro tipo de algoritmo que discutí se llamaba Grad-CAM y Grad-CAM esencialmente es un algoritmo como LRP, pero en su lugar examina los gradientes de las capas convolucionales en su modelo. Y este método también toma una imagen y luego la alimenta al modelo y luego produce un mapa de activación que es un mapa de calor que representa todos los gradientes en una capa convolucional y que recoge exactamente donde el modelo está mirando.

#### *Diapositiva 12*

Así que aquí hay algunos ejemplos de Grad-CAM en su lugar. Es como una combinación entre LRP y LIME. Produce un mapa de calor de una región específica en la imagen, y se puede ver en qué se está enfocando el modelo y en qué regiones no se está enfocando.

#### *Diapositiva 13*

Y la última variación que es la LRP más contrastada es una ligera modificación al método LRP original, pero en su lugar, toma las puntuaciones de relevancia y aplica una modificación que distingue la relevancia entre varios tipos de clases. Y así se puede identificar o distinguir entre la neumonía viral y bacteriana mucho más claramente utilizando estos diferentes tipos de puntuaciones relevantes. Y por lo tanto, es muy importante en general para buscar y comparar estos diversos tipos de métodos de explicación con el fin de entender lo que su modelo está mirando específicamente. Y así hay muchos tipos de algoritmos que se pueden utilizar, pero con el fin de obtener una comprensión completa de lo que su modelo ha aprendido- lo que su modelo está mirando cuando se está haciendo predicciones- es importante comparar estos tipos de métodos y mientras se construye un modelo preciso para ayudar a los pacientes.

#### *Diapositiva 14*

Y solo una investigación en curso en este momento- La IA explicable es un campo en continuo crecimiento. Cada día se están desarrollando más algoritmos y se están recopilando más datos para construir modelos más precisos y diferentes tipos de datos, como las tomografías computarizadas, también pueden ayudar a los radiólogos a diagnosticar a los pacientes. Así que la ejecución de estos algoritmos en diferentes tipos de modelos construidos para examinar qué modelo ha aprendido los datos con mayor precisión que otros.

#### *Diapositiva 15*

Y sí, muchas gracias por escuchar mi presentación y me gustaría agradecer al Dr. Michael Pazzani, quien también está en la llamada ahora mismo por toda su ayuda en la realización de su investigación, la recopilación de datos, y la implementación de estos algoritmos para identificar la neumonía. Y gracias. Este es mi correo electrónico si quieres contactarme.

#### *Diapositiva 16*

Muchas gracias a Sansón. Es realmente emocionante ver tu investigación evolucionar y estaremos, ya sabes, observando tu carrera y tu investigación científica con interés. Realmente apreciamos que comparta su trabajo con nosotros, y felicidades de nuevo por ganar nuestro desafío de papel estudiantil.