

[COVID Information Commons \(CIC\) Research Lightning Talk](#)

Transcript of a Presentation by Dominique Duncan (University of Southern California), September 16, 2020



Title: [COVID-ARC \(COVID-19 Data Archive\)](#)

[Dominique Duncan CIC Database Profile](#)

NSF Award #: [2027456](#)

[YouTube Recording with Slides](#)

[September 2020 CIC Webinar Information](#)

Transcript Editor: Julie Meunier

---

Transcript

Katie Naum:

Nous allons entendre Dominique Duncan de l'Université de Californie du Sud. Dominique, nous sommes prêts quand vous voulez.

Dominique Duncan:

*Slide 1:*

Je vous remercie. Je suis Dominique Duncan, de l'USC, du Nerve Imaging and Informatics Institute, et je vais vous parler de nos archives de données COVID-19. En bref, il s'agit de COVID ARC.

*Slide 2:*

Notre institut a donc une grande expérience des référentiels de données multimodales à grande échelle. Nous avons donc décidé d'appliquer notre expérience et les outils que nous avons développés dans le cadre de ces autres projets et de les étendre à ces archives de données COVID-19. Vous pouvez donc voir ici, en bas à gauche, une capture d'écran de notre page d'accueil. Si vous souhaitez consulter le site web, j'encourage tout le monde à aller sur covid-arc.loni.usc.edu pour en savoir plus sur le projet, les différents ensembles de données dont nous disposons, nos outils analytiques et pour en savoir plus sur le chargement des données ou le téléchargement des données existantes dont nous disposons. Nous disposons donc d'ensembles de données accessibles au public que nous avons conservés et organisés. Nous voulons encourager les chercheurs à être en mesure d'effectuer des analyses non seulement sur des ensembles de données individuels, mais aussi sur des ensembles de données provenant de différents sites, et nous voulons donc leur faciliter la tâche. C'est pourquoi nous fournissons également divers outils, tels que des outils de contrôle de la qualité, que les chercheurs peuvent utiliser pour

évaluer les images et leur qualité à l'aide de diverses mesures. Nous disposons de divers outils de visualisation pour les données d'imagerie et d'autres types de données, ainsi que d'une grande variété d'outils d'analyse que les utilisateurs peuvent utiliser. Pour les ensembles de données qui ne sont pas accessibles au public, les fournisseurs de données peuvent décider s'ils veulent stocker leurs données sur notre serveur ou s'ils veulent les stocker localement sur leur site et nous fournir les métadonnées afin que nous puissions indiquer aux utilisateurs quelles sont les données disponibles. Ainsi, si les fournisseurs de données nous communiquent leurs données, celles-ci sont anonymisées et les fournisseurs de données conservent le contrôle total de l'accès et décident qui peut y accéder. S'ils veulent attendre de publier certains résultats et rendre ensuite leurs ensembles de données accessibles au public, ils peuvent également le faire. Nous utilisons ensuite ASPERA, le système de transfert de fichiers crypté à grande vitesse d'IBM, conforme à la loi HIPAA, pour télécharger les données sur notre serveur ou pour les personnes qui souhaitent y avoir accès. Ils peuvent également les télécharger de cette manière.

*Slide 3:*

Je sais que vous ne pouvez pas voir le texte, mais cela vous donne un aperçu de ce que nous avons sur notre serveur. Si vous demandez l'accès, il s'agit de l'arborescence du projet COVID ARC, qui comprend les données et l'analyse sur laquelle nous avons travaillé, car nous sommes également très impliqués dans l'analyse de ce projet. Les données seront donc séparées en fonction des différents sites et il y aura plus d'informations sur chacun d'entre eux.

*Slide 4:*

Voici donc un tableau des ensembles de données que nous avons actuellement sur le serveur afin que vous puissiez voir l'emplacement, où les données ont été acquises pour chacun de ces ensembles de données, et ensuite le format de données de ces images. Pour l'instant, nous nous concentrons principalement sur la tomodensitométrie thoracique, mais nous nous intéressons également à d'autres types de données et nous disposerons bientôt de données cérébrales, c'est-à-dire l'IRM cérébrale des patients COVID-19, ainsi que l'EEG. Vous pouvez donc voir ici combien d'images COVID et d'images non COVID et de masques proviennent de chacun de ces sites.

*Slide 5:*

Vous pouvez voir ici quelques cartes d'activation de classe. Comme nous le savons, les patients atteints de COVID-19 présentent de nombreuses caractéristiques tomodensitométriques, telles que l'opacité en verre dépoli, la consolidation, *le motif de pavage fou*, le motif réticulaire, etc. Cela met en évidence les régions les plus importantes pour la classification que nous effectuons.

*Slide 6:*

L'un des problèmes que nous découvrons est que la qualité de l'image a tendance à entraîner une mauvaise classification. Nous évaluons donc la qualité des images, mais nous envisageons également divers moyens d'améliorer la qualité des images, en examinant différentes méthodes de filtrage, ce qui est actuellement en cours.

*Slide 7:*

Nous effectuons également un seuillage d'image sur les masques pulmonaires afin de trouver le meilleur type d'image traitée possible dans les masques pulmonaires pour améliorer le taux de prédiction de nos réseaux neuronaux.

*Slide 8:*

Voici un tableau de l'un des ensembles de données. Le premier, provenant du Brésil, contenait environ 1 200 patients COVID et 1 200 images non COVID. Nous avons comparé différentes méthodes pour comparer les réseaux neuronaux convolutifs et leur précision. Nous avons constaté que ResNet-18 était le plus performant dans ce domaine.

*Slide 9:*

Et pour résumer, parce que je pense que je suis à la fin du temps imparti, voici les personnes qui travaillent sur le projet, y compris les étudiants REU. J'aimerais remercier la NSF pour le financement de ce projet, ainsi que Katie et Florence pour l'avoir organisé et m'avoir invité à parler. N'hésitez pas à visiter le site web ou à m'envoyer un email si vous avez des questions, si vous souhaitez accéder aux données ou si vous avez des données que vous aimeriez ajouter au site web. Je vous remercie.