

[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: Shikhar Johri

---

केटी भमः

हम दक्षिणी कैलिफोर्निया विश्वविद्यालय में डोमिनिक डंकन से सुनेंगे। जब भी आप होते हैं, हम तैयार होते हैं।

डोमिनिक डंकनः

स्लाइड 1:

धन्यवाद। इसलिए, मैं यूएससी से डोमिनिक डंकन हूँ, नर्व इमेजिंग और इंफॉर्मेटिक्स इंस्टीट्यूट से और मैं अपने कोविड -19 डेटा आर्काइव के बारे में बात करूंगा। संक्षेप में इसे कोविड आर्क कहा जाता है।

स्लाइड 2:

इसलिए, हमारे पास बड़े पैमाने पर मल्टीमॉडल डेटा रिपॉजिटरी के साथ हमारे संस्थान में बहुत अनुभव है, इसलिए हमने अपने अनुभव और उन उपकरणों को लागू करने का फैसला किया, जिन्हें हमने इन अन्य परियोजनाओं में विकसित किया है और उन्हें इस COVID-19 डेटा संग्रह में विस्तारित किया है। तो, आप यहाँ नीचे बाईं ओर देख सकते हैं, यह हमारे होम पेज का एक स्क्रीनशॉट है। यदि आप वेबसाइट पर जाना चाहते हैं, तो मैं सभी को परियोजना के बारे में अधिक जानकारी प्राप्त करने के लिए [covid-arc.loni.usc.edu](https://covid-arc.loni.usc.edu) पर जाने के लिए प्रोत्साहित करता हूँ, विभिन्न डेटा सेट जो हमारे पास हैं, हमारे विश्लेषणात्मक उपकरण और या तो अधिक जानकारी प्राप्त करने के लिए डेटा अपलोड करना या मौजूदा डेटा डाउनलोड करना जो हमारे पास है। इसलिए, हमारे पास कुछ सार्वजनिक रूप से उपलब्ध डेटा सेट हैं जिन्हें हम क्यूरेटिंग और आयोजन कर रहे हैं। हम शोधकर्ताओं को प्रोत्साहित करना चाहते हैं कि वे न केवल व्यक्तिगत डेटा सेटों पर विश्लेषण करने में सक्षम हों, बल्कि विभिन्न साइटों से डेटा सेट पर विश्लेषण करें और इसलिए हम उनके लिए उस प्रक्रिया को आसान बनाना चाहते हैं। और इसलिए हम विभिन्न उपकरण जैसे गुणवत्ता नियंत्रण उपकरण भी प्रदान कर रहे हैं जो लोग छवियों का मूल्यांकन

करने और विभिन्न मैट्रिक्स का उपयोग करके गुणवत्ता का मूल्यांकन करने के लिए उपयोग कर सकते हैं। हमारे पास इमेजिंग डेटा और अन्य प्रकार के डेटा के लिए विभिन्न विजुअलाइज़ेशन टूल हैं और फिर विभिन्न प्रकार के विश्लेषणात्मक उपकरण हैं जिनका उपयोग लोग कर सकते हैं। और फिर डेटा सेट के लिए जो सार्वजनिक रूप से उपलब्ध नहीं हैं, डेटा प्रदाता यह तय कर सकते हैं कि क्या वे अपने डेटा को हमारे सर्वर पर संग्रहीत करना चाहते हैं या यदि वे अपने डेटा को स्थानीय रूप से अपनी साइट पर संग्रहीत रखना चाहते हैं और बस हमें मेटाडेटा प्रदान करना चाहते हैं ताकि हम दे सकें ताकि हम दे सकें उपयोगकर्ताओं को पता है कि क्या डेटा उपलब्ध हैं और फिर हम केवल उस डेटा तक पहुंच का अनुरोध करने की प्रक्रिया को सुविधाजनक बनाते हैं। इसलिए, यदि डेटा प्रदाता हमें अपना डेटा देते हैं, तो यह अनाम है और डेटा प्रदाता एक्सेस का पूर्ण नियंत्रण बनाए रखते हैं और वे यह तय करते हैं कि किसकी पहुंच प्राप्त होती है। यदि वे कुछ निष्कर्षों को प्रकाशित करने के लिए इंतजार करना चाहते हैं, और फिर उनके डेटा सेट को सार्वजनिक रूप से उपलब्ध कराते हैं, तो वे ऐसा करने में सक्षम हैं। और फिर हम Aspera का उपयोग करते हैं जो कि IBM का HIPAA कॉम्प्लेंट एन्क्रिप्टेड हाई-स्पीड फ़ाइल ट्रांसफर सिस्टम या तो हमारे सर्वर पर डेटा अपलोड करने के लिए या उन लोगों के लिए है जो डेटा तक पहुंच चाहते हैं। वे इसे इस तरह से डाउनलोड कर सकते हैं।

स्लाइड 3:

मुझे पता है कि आप इस पाठ में से कोई भी नहीं देख सकते हैं, लेकिन यह आपको हमारे सर्वर पर क्या है, इसका अवलोकन देता है। इसलिए, यदि आप एक्सेस का अनुरोध करते हैं, तो यह कोविड एआरसी प्रोजेक्ट की ट्री स्ट्रक्चर है और इसलिए यह डेटा में अलग हो गया है और फिर हमारा विश्लेषण है कि हम इस पर काम कर रहे हैं क्योंकि हम इस परियोजना के लिए एनालिटिक्स में भी शामिल हैं। तो, डेटा को अलग-अलग साइटों द्वारा अलग किया जाएगा और फिर उनमें से प्रत्येक पर अधिक जानकारी होगी।

स्लाइड 4:

तो, यहां डेटा सेट की एक तालिका है जो वर्तमान में हमारे पास सर्वर पर है ताकि आप उस स्थान को देख सकें, जहां डेटा उन डेटा सेटों में से प्रत्येक के लिए अधिग्रहित किया गया था, और फिर उन छवियों का डेटा प्रारूप। अभी, हम मुख्य रूप से चेस्ट सीटी पर ध्यान केंद्रित कर रहे हैं, लेकिन हम अन्य प्रकार के डेटा में भी रुचि रखते हैं और जल्द ही हमारे पास कुछ मस्तिष्क डेटा होगा, इसलिए COVID-19 रोगियों के मस्तिष्क एमआरआई के साथ-साथ ईईजी भी। तो यहाँ आप बस देख सकते हैं कि उन साइटों में से प्रत्येक से कितने कोविड छवियां और गैर-कोविड इमेज और मास्क हैं।

स्लाइड 5:

यहां आप कुछ वर्ग सक्रियण मानचित्र देख सकते हैं। इसलिए, कई विशेषताएं हैं, जैसा कि हम जानते हैं, COVID-19 मरीजों की सीटी फीचर्स जैसे ग्राउंड ग्लास अपारदर्शिता, समेकन, पागल पाविंग पैटर्न, रेटिकुलर पैटर्न, आदि, इसलिए, यह सिर्फ उन क्षेत्रों को उजागर करता है जो हमारे वर्गीकरण के लिए सबसे महत्वपूर्ण थे कि हम हम हैं कि हम हम हैं। फिर से कर रहे हैं।

स्लाइड 6:

जिन मुद्दों को हम खोज रहे हैं, उनमें से एक यह है कि छवि की गुणवत्ता गर्भपात की ओर ले जाती है और इसलिए हम छवियों की गुणवत्ता का आकलन कर रहे हैं, लेकिन फिर छवियों की गुणवत्ता में सुधार के विभिन्न तरीकों पर भी विचार कर रहे हैं ताकि विभिन्न फिल्टरिंग विधियों को देखते हुए और यह है कुछ ऐसा जो वर्तमान में प्रगति पर है।

स्लाइड 7:

हम फेफड़े के मास्क पर छवि थ्रेसहोल्डिंग भी कर रहे हैं, इसलिए हम अपने तंत्रिका नेटवर्क की भविष्यवाणी दर में सुधार करने के लिए फेफड़े के मास्क में सबसे अच्छा संभव संसाधित छवि प्रकार खोजने के लिए छवि थ्रेशोल्डिंग का उपयोग कर रहे हैं।

स्लाइड 8:

और यह डेटा सेट में से एक पर एक तालिका है। ब्राजील से यह पहला जिसमें लगभग 1,200 कोविड मरीज और 1,200 गैर-कोविड छवियां थीं और हमने कन्व्यूशनल न्यूरल नेटवर्क और उनकी सटीकता की तुलना करने के लिए विभिन्न तरीकों को देखते हुए तुलना की। इसलिए, हमने पाया कि RESNET-18 ने इसके लिए सबसे अच्छा प्रदर्शन किया।

स्लाइड 9:

और सिर्फ संक्षेप में, क्योंकि मुझे लगता है कि मैं उस समय के अंत में हूँ जो वे लोग हैं जो आरईयू छात्रों सहित परियोजना पर काम कर रहे हैं। और मैं इस परियोजना के साथ -साथ केटी और फ्लोरेंस को इस आयोजन के लिए और मुझे बात करने के लिए आमंत्रित करने के लिए एनएसएफ को धन्यवाद देना चाहता हूँ। और कृपया वेबसाइट पर जाएं या मुझे ईमेल करें यदि आपके कोई प्रश्न हैं और यदि आप किसी भी डेटा तक पहुंच चाहते हैं या यदि आपके पास कुछ डेटा है जिसे आप वेबसाइट में योगदान देना चाहते हैं। धन्यवाद।