

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

Transcript of a Presentation by Jayavanth Shenoy (Onai) January 30, 2024



Title: [Scaled Medical Records Analysis](#)

[Guha Jayachandran CIC Database Profile](#)

NSF Award #: [2028008](#)

[YouTube Recording with Slides](#)

[Winter 2024 CIC Webinar Information](#)

Transcript Editor: Lauren Close

Transcript

Slide 1

Je m'appelle Jayavanth et je suis ingénieur à l'Onai. Aujourd'hui, je vais vous parler de notre projet d'analyse à grande échelle des dossiers médicaux, financé par la NSF.

Slide 2

Le problème, c'est que les dossiers médicaux sont des informations très importantes, mais nous n'avons pas de bonne méthode pour les étudier à grande échelle. Traditionnellement, les études se faisaient dans des hôpitaux individuels sur les dossiers de cet hôpital. Même là, c'est un travail laborieux, avec des résidents et des étudiants qui transcrivent manuellement les données dans des feuilles de calcul. Ce processus est aussi vraiment lent. Dans des situations comme celle du COVID-19, cela devient inacceptable – nous ne voulons pas attendre qu'un seul hôpital ait suffisamment de données pour que des études statistiquement significatives puissent être menées. C'est ce qui s'est passé – enfin, c'est ce qui se passe généralement. Mais nous voulons vraiment obtenir des résultats dès les premières étapes de la propagation. Pour cela, nous devons être capables de réaliser des études à travers plusieurs hôpitaux de manière agrégée. Il y a des questions très basiques qui, dans l'absolu, devraient être faciles à répondre de nos jours, mais qui ne l'étaient pas pendant le COVID et ne le sont toujours pas aujourd'hui. Par exemple, les médicaments contre l'hypertension améliorent-ils ou empiront-ils le COVID ? Des dizaines de millions d'Américains prennent ces médicaments contre l'hypertension, donc on pourrait penser que ce serait une question facile à répondre simplement en se basant sur les résultats des patients ayant pris des médicaments contre l'hypertension et qui ont fini à l'hôpital avec le COVID. Cela devrait être aussi simple qu'une requête Google, mais ce n'est pas le cas.

Slide 3

Notre objectif était de permettre une recherche rapide sur les dossiers médicaux à travers plusieurs institutions de santé de manière efficace et sans compromettre la confidentialité. Grâce à nos outils, les institutions de santé et les entreprises pharmaceutiques peuvent atteindre des populations statistiquement significatives beaucoup plus rapidement, et la recherche peut être réalisée immédiatement. C'est également plus sécurisé. Alors, pouvons-nous obtenir le meilleur des deux mondes ? Pouvons-nous avoir un large ensemble de données pour la recherche tout en maximisant la sécurité ?

Slide 4

Oui, nous pouvons le faire grâce à cette technologie cryptographique. La technologie qui permet d'atteindre cet objectif est une technique magique appelée "calcul multipartite". Cela permet à un ensemble de parties de calculer un résultat sur leurs entrées sans révéler leurs entrées à l'autre ou à quiconque d'autre. Un exemple de cela est ce qu'on appelle le "problème du millionnaire de Yao". Andrew Yao a introduit cela en 1982 et il s'agit de nos deux millionnaires Alice et Bob. Ils veulent savoir lequel d'entre eux est le plus riche sans révéler leur richesse réelle. Donc, vous avez deux personnes, Alice et Bob. Elles connaissent leur propre richesse mais ne veulent pas la dire à l'autre. Et supposons qu'elles ne veulent pas impliquer un tiers, pas d'IRS ou quoi que ce soit. Pouvez-vous penser à un moyen pour qu'elles découvrent qui est le plus riche ? Eh bien, croyez-le ou non, il existe une manière de faire cela. Un autre exemple : supposons que nous voulions calculer la moyenne des numéros de sécurité sociale de toutes les personnes de ce webinaire, mais nous ne voulons pas révéler nos numéros de sécurité sociale les uns aux autres. Existe-t-il un moyen de le faire ? Oui, il existe avec le calcul multipartite sécurisé (MPC). Nous avons appliqué le même concept aux dossiers de santé et le résultat d'une requête est calculé à travers plusieurs hôpitaux sans que les hôpitaux aient besoin de transmettre ou de révéler aucun dossier. Nous avons démontré cela à travers quelques sites disparates dans notre projet.

Slide 5

Voici quelques exemples de questions que l'on pourrait souhaiter poser, comme la durée moyenne du séjour à l'hôpital des patients de plus de 70 ans, ou le pourcentage de patients très jeunes qui se retrouvent sous respirateur.

Slide 6

Pour un public plus averti en matière de cryptographie, nous avons utilisé une mascotte pour la technique des PNJ. Je n'entrerai pas dans les détails maintenant, mais vous pourrez me poser des questions à ce sujet plus tard si cela vous intéresse. La seule chose que je mentionnerai ici est que ce protocole est sécurisé contre la majorité malhonnête, ce qui signifie que ce protocole est sécurisé même si une majorité des participants est malveillante.

Slide 7

Je tiens à rappeler qu'avec notre approche, aucune donnée ne quitte jamais l'hôpital. Les données de chaque hôpital sont complètement cloisonnées dans leur service. Les services informatiques n'ont pas à

s'inquiéter de la sécurité, et pas seulement du fait que rien ne peut être désossé. La coordination des calculs se fait via la blockchain ou le grand livre de distribution.

Slide 8

Ce cadre permet également de réaliser des essais cliniques plus équitables. Traditionnellement, les populations de patients et les essais cliniques ne sont pas très diversifiés. L'une des priorités de la FDA et d'autres organismes est d'améliorer cette situation et, grâce à cette technologie, il est possible de localiser des sites où se trouvent des patients intéressants, même s'il ne s'agit pas d'institutions avec lesquelles une entreprise travaille habituellement.

Slide 9

Depuis, nous avons étendu cette approche de la recherche à l'entraînement de modèles d'intelligence artificielle sur des sites disparates, tout en préservant la vie privée. Nous travaillons également avec le NIH pour la découverte de médicaments en préservant la vie privée. Notre objectif final est de disposer de réseaux d'intelligence préservant la vie privée dans l'ensemble de la biomédecine.

Slide 10

Je vous remercie de m'avoir écouté et j'aimerais remercier le CIC pour cette opportunité et la NSF pour le financement de notre projet, nos collaborateurs à travers le pays et mes collègues de l'Onai. N'hésitez pas à me faire part de vos questions à la fin de la session. Je vous remercie à nouveau.