

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

Transcript of a Presentation by Hong Qin (University of Tennessee, Chattanooga), July 26, 2023



Title: [Develop and Evaluate Computational Frameworks to Predict and Prevent Future Coronavirus Pandemics](#)

[Hong Qin CIC Database Profile](#)

NSF Award #: [2200138](#)

[YouTube Recording with Slides](#)

[Summer 2023 CIC Webinar Information](#)

Transcript Editor: Julie Meunier

Transcript

Slide 1

Bonjour à tous. Je m'appelle Hong Qin, et je suis membre du corps enseignant à l'Université du Tennessee à Chattanooga. Aujourd'hui, je suis ravi de partager un projet sur lequel j'ai eu le privilège de travailler avec une équipe diversifiée de collègues de la North Carolina A&T, du Spelman College et de la Catholic University of America. Bien que ce soient les collaborateurs sur papier, la réalité est que nous avons une équipe de plus d'une douzaine de membres travaillant sur ce projet soutenu par la National Science Foundation, appelé PIPP Phase One. L'objectif principal de ce projet est de développer un cadre basé sur l'IA pour prédire et prévenir les futures pandémies de coronavirus, avec la possibilité de généralisation à d'autres pandémies virales.

Slide 2

Explorons maintenant l'aperçu du projet. Bien que nous ayons apporté des modifications, l'idée fondamentale reste la même. Notre approche consiste à générer toutes les variantes possibles du SARS-CoV-X en fonction des motifs de recombinaison ou des changements d'habitat, puis à prédire les interactions potentielles. Le défi réside dans l'identification de celles de ces variantes qui deviendront virulentes ou se propageront dans les populations humaines. C'est là qu'intervient l'IA. Bien que le processus de passage d'un virus d'un hôte à un autre reste une boîte noire, l'IA est un outil puissant pour faire des prédictions. Les séquences d'entrée subissent un génie des fonctionnalités, et le modèle d'IA prédit les résultats. Nous prévoyons d'utiliser des techniques avancées, telles que la simulation dynamique, pour vérifier les gènes candidats aux

changements de virulence. De plus, les connaissances du modèle d'IA peuvent être transférées à d'autres virus tels que la grippe, le VIH ou le VPH.

Les défis comprennent la prédiction de la pathogénicité à partir des séquences, l'établissement de la généralisabilité et la création d'un système d'alerte précoce basé sur l'IA pour le développement de vaccins.

Slide 3

Voici quelques résultats que nous avons développés et accomplis. Il s'agit d'une estimation de la forme physique du virus basée sur les données que nous possédons, spécifiquement pour le sous-variant Omicron. Nous pouvons comparer ce sous-variant Omicron à l'aide d'une méthode de comparaison par paires, puis estimer la différence relative entre ces variants.

Slide 4

Sur la base de ces données, nous pouvons générer un paysage dit de la forme physique du virus. Pour ceux qui sont familiers avec la théorie de l'évolution, il existe un paysage de la forme physique du virus, sujet à de nombreux débats, mais aussi théorisé. Avec ce paysage de la forme physique d'un virus potentiel, nous pourrions théoriquement prédire sa trajectoire évolutive. Nous avons une méthode pour le faire. Ce sont les résultats estimés pour les sous-variants alpha, bêta, delta et omicron aux États-Unis. Nous sommes également en train d'appliquer cela à la grippe ou à la variole. Bien que nous n'ayons pas assez de données pour la variole ou d'autres virus, la grippe semble prometteuse. C'est une découverte significative de notre projet actuel.

Slide 5

Voici quelques résultats du modèle. Sans entrer dans trop de détails sur le modèle d'IA que nous avons développé, il y a beaucoup de paramètres, mais le modèle actuel peut atteindre une précision assez élevée. Bien sûr, cela repose sur ce que nous savons. Le défi réside vraiment dans l'inconnu. Même si nous prédisons l'inconnu, les données sont déjà connues. En apprentissage automatique, nous l'entraînons avec des données d'entraînement et de test. Le défi de la prédiction de l'inconnu reste une question ouverte.

Slide 6

Certaines personnes pourraient demander pourquoi nous devons faire de l'apprentissage en profondeur. Il existe de nombreuses méthodes statistiques, génétiques, génomiques, de traitement quantitatif - pourquoi devrions-nous nous tourner vers l'apprentissage en profondeur? Ici, nous avons des preuves montrant que l'apprentissage en profondeur détecte, au minimum, des signatures différentes par rapport aux études génomiques standard à l'échelle du génome. À gauche se trouve un résultat de notre apprentissage en profondeur basé sur les variants étiquetés par l'OMS, Alpha, Beta, Delta, Omicron, et d'autres. Nous voulons voir quelles signatures dans le génome du SARS-CoV-2 sont importantes pour contribuer à cette augmentation de la virulence. Ces signaux élevés que vous pouvez voir à la fin du génome du SARS-CoV-2. Autour

de cette région que je mets en évidence à environ 20 000 à 25 000, c'est le gène de la protéine spike que l'OMS [a identifié] et à laquelle la plupart des réponses immunitaires réagiront. Sur la base des connaissances conventionnelles, cela devrait être, en fait, c'est ainsi que l'OMS a classé les variants du SARS-CoV-2. Étonnamment, les modèles d'IA captent ces signaux, mais ce ne sont pas les signaux les plus forts, et ils captent de nombreux autres signaux. Certains des signaux les plus forts ne sont pas là. En utilisant des méthodes génomiques statistiques conventionnelles telles que l'étude d'association à l'échelle du génome - cela sélectionne toujours le gène de la protéine spike comme le signal le plus fort. Il y en a d'autres, mais pas très forts. Donc, dans ce cas, l'IA et la méthode statistique conventionnelle sélectionnent, ou au minimum, donnent un poids différent à ces signaux.

C'est surprenant et aussi rassurant, d'une certaine manière. Nous avons choisi la sagesse conventionnelle, la protéine spike, mais nous avons également choisi d'autres signaux qui peuvent ou non être vérifiés par la méthode expérimentale. Mais comment prédisons-nous et comment vérifions-nous cela, n'est-ce pas? C'est assez difficile.

Slide 7

Nous essayons, en général, deux méthodes différentes. L'une consiste à créer un système expérimental modèle pour vérifier les résultats. Notre équipe utilise un modèle biologique, il s'agit du [inaudible], nous avons également un système de lignée cellulaire pour introduire ces gènes dans le cycle de vie et mesurer leurs activités relatives.

Slide 8

Nous avons également une personne spécialisée en informatique pour effectuer des simulations de dynamique moléculaire afin de simuler comment ces mutations affectent [inaudible] de leur activité. Dans ce cas, ACE2 et RBD, mais nous avons également d'autres [inaudible] qui prédisent dans différentes régions comment ils interagissent avec les gènes humains - potentiellement en ciblant les systèmes immunitaires humains.

Slide 9

Nous organisons également des ateliers et des panels d'experts pour discuter de la conception d'une IA digne de confiance afin de promouvoir la confiance sociale dans l'IA - une IA équitable. La pandémie de COVID a montré qu'il existe des disparités importantes dans notre système actuel. Ainsi, si nous utilisons l'IA pour prédire les futures pandémies, cela peut facilement amplifier les disparités cachées dans notre système actuel, probablement également dans nos données actuelles. Si nous reconnaissons cela et organisons un panel assez diversifié, comprenant des personnes d'Afrique du Sud, du Kenya, d'Europe, et d'horizons différents tels que des avocats, des médecins, des décideurs politiques, des agences gouvernementales - NIST - pour avoir toutes sortes de perspectives sur la manière de promouvoir l'IA et de promouvoir la confiance en l'IA dans les communautés historiquement marginalisées.

Slide 10

Nous mettons également l'accent sur le développement de la main-d'œuvre, étant donné nos collaborations avec des institutions desservant des minorités, notamment le Spelman College, la North Carolina A&T et la Catholic University of America, qui comptent de nombreux étudiants hispaniques, et mon institution d'origine, l'Université du Tennessee à Chattanooga. Nous collaborons également avec un système hospitalier du Global South.

Slide 11

Enfin, je tiens à remercier la NSF, l'UTC et l'initiative AI Tennessee pour leur soutien dans ce projet. Et enfin, je n'ai pas inclus dans les diapositives, la NSF a publié un appel à candidatures pour la phase 2 d'une demande de Centre national, qui sera probablement 10 fois plus grand, voire 100 fois plus grand, alors je suis à la recherche de collaborateurs, en particulier dans l'engagement communautaire, la politique publique. J'ai entendu quelqu'un travailler dans la politique. Donc, j'aimerais me connecter après cela. Je vais m'arrêter ici, merci.