

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

Transcript of a Presentation by Ioannis Paschalidis (Boston University), April 24, 2023



Title: [Predictive Models of COVID-19 Severity and Patient Outcomes](#)

[Ioannis Paschalidis CIC Database Profile](#)

NSF Award #: [2200052](#)

[YouTube Recording with Slides](#)

[Spring 2023 CIC Webinar Information](#)

Transcript Editor: Julie Meunier

Transcript

Slide 1

Merci beaucoup et merci de m'avoir invité à faire partie de cet après-midi passionnant avec un excellent groupe de conférenciers. Je vais parler des modèles prédictifs de gravité de la COVID-19 et des résultats des patients que nous avons développés à l'Université de Boston.

Slide 2

Pour donner un peu de contexte, nous travaillons depuis assez longtemps avant la pandémie sur divers modèles qui prédisent la maladie et des événements clés importants, par exemple, les hospitalisations, et aussi des modèles qui prescrivent des traitements. Donc, lorsque la pandémie a commencé, nous nous sommes mobilisés, comme le reste de la communauté scientifique, et nous avons également mobilisé un réseau relativement important de collaborateurs, nous avons donc obtenu l'accès à divers ensembles de données différents.

Vous pouvez voir sur cette diapositive les différents ensembles de données que nous avons. Des ensembles de données locaux du Massachusetts - de deux réseaux hospitaliers différents dans le Massachusetts. L'un provenait du réseau Mass General Brigham, de cinq hôpitaux différents, environ 2 500 cas, et puis une autre cohorte relativement importante du Boston Medical Center, d'environ 7 000 cas. Nous avons également obtenu quelques cas de Wuhan, en Chine, qui était l'origine évidente de l'épidémie. Et enfin, nous avons pu avoir accès à certains ensembles de données nationaux importants - un ensemble de données du Brésil et un autre ensemble de données du Mexique.

Dans cette présentation très courte, je me concentrerai davantage sur les travaux les plus récents qui ont pris en compte l'ensemble de données - la série de patients du Boston Medical Center. Le

Boston Medical Center est l'hôpital d'enseignement affilié à la Faculté de médecine de BU et est également un hôpital de sécurité sociale. Comme vous le verrez, cela a des implications intéressantes dans les résultats que nous avons pu obtenir.

Slide 3

Nous avons eu accès à l'ensemble de cohortes du BMC pour l'année 2020, soit environ 7 000 patients. Il s'agissait de patients testés positifs pour la COVID-19. Quelques statistiques approximatives - environ 20 % ont été admis. Parmi les admis, environ 23 % ont été admis en soins intensifs. Parmi ceux admis en soins intensifs, environ 58,7 % ont été intubés. Parmi ceux qui ont été intubés, malheureusement, environ 70 % n'ont pas survécu. Nous avons beaucoup d'informations sur ces patients, y compris leurs données démographiques, leurs signes vitaux tout au long du séjour à l'hôpital, les rapports de radiologie, leurs antécédents médicaux, leurs symptômes, les résultats des analyses, les médicaments et même des informations sur l'état dépressif, le code postal, et nous avons également des informations sur l'occupation de l'hôpital au moment où chacun de ces patients était vu. Nous avons également des informations sur les déterminants sociaux de la santé. Le BMC gère un programme appelé le programme Thrive, où toute personne ayant un contact avec l'hôpital remplit un sondage où nous leur demandons des informations sur leurs besoins dans divers domaines, notamment le logement, la nourriture, le transport, l'aide à la garde, les médicaments, l'accès aux médicaments et le paiement des médicaments, l'éducation et l'emploi. Une grande partie des informations étaient sous forme tabulaire, plus facilement traitable par des méthodes d'apprentissage automatique en IA, mais il y avait aussi pas mal d'informations, en particulier les rapports de radiologie et d'autres rapports pendant le séjour à l'hôpital, qui étaient simplement des récits - des rapports de cliniciens. Nous avons dû utiliser beaucoup de traitement du langage naturel pour extraire les conclusions appropriées du texte. Ce que nous avons fait, c'est que nous avons développé un ensemble de modèles robustes et interprétables qui peuvent prédire l'hospitalisation, l'admission en soins intensifs, la ventilation mécanique et la mort. Je vais vous montrer quelques exemples, je ne serai pas exhaustif, encore une fois dans l'intérêt du temps.

Slide 4

Tout d'abord, ce que nous avons fait, c'est que pour chaque patient, nous avons construit une chronologie à partir du moment où nous avons connaissance du résultat positif du test du patient et des informations sur ce patient jusqu'au point de l'événement d'intérêt, que ce soit, disons, une admission en soins intensifs ou une ventilation mécanique. Voici le résultat d'intérêt, puis en remontant dans le temps, nous avons créé ces compartiments de données. Nous avons supprimé toutes les informations disponibles juste avant le moment de l'événement d'intérêt parce que nous voulions des modèles capables de prédire ce qui se passera à l'avenir. Peut-être même qu'un étudiant de première année en médecine peut identifier un patient se dirigeant vers les soins intensifs, nous voulions donc que les modèles fassent cette prédiction avec des informations plus précoces. Vous verrez que nous avons obtenu différentes versions des modèles qui avaient différentes coupures en termes d'informations que le modèle utilisait pour faire la prédiction. Une autre raison pour laquelle nous avons créé ces compartiments de temps est que nous voulions capturer l'évolution dynamique de la progression du patient pendant son séjour à l'hôpital, par

exemple, l'évolution dynamique des signes vitaux. Nous avons compris que cela était plutôt important. Plutôt que de simplement regarder l'instantané des signes vitaux, disons à un moment spécifique, et d'utiliser cette information pour faire une prédiction vers l'avant, les valeurs exactes sont importantes mais les tendances le sont aussi. Les médecins, lorsqu'ils examinent les patients, regardent les tendances des signes vitaux du patient. Nous avons donc développé un modèle qui utilisait des méthodologies d'apprentissage profond assez sophistiquées, y compris des réseaux de type LSTM et une architecture de transformateur qui prenaient comme entrée les signes vitaux - six signaux vitaux à différents moments - et produisaient un score. Ce score capturait l'évolution dynamique des signes vitaux et ce score vital était ensuite utilisé dans un modèle d'ensemble qui tentait de faire une prédiction pour le résultat d'intérêt.

Slide 5

Par exemple, les prédictions d'hospitalisation. Vous pouvez voir que celles-ci sont assez précises. Certains des meilleurs modèles vous donnent 92 % - c'est en aire sous la courbe. Vous pouvez le considérer comme une mesure de la précision du modèle. Le meilleur est de 100 %, une supposition aléatoire vous donne 50 %. Donc, 92 %, c'est une assez bonne performance. Vous pouvez voir ici que certains des modèles linéaires que nous avons développés ont également trouvé certains des facteurs qui étaient importants dans la prédiction d'hospitalisation. En bleu, vous voyez certaines des variables associées à des conditions de santé antérieures qui sont, par exemple, fortement corrélées à l'hospitalisation. Vous verrez également que l'occupation de l'hôpital, si elle était élevée, réduisait la probabilité que le patient soit hospitalisé. De plus, vous verrez deux déterminants sociaux de la santé : le besoin de nourriture et le besoin de transport. Ces deux éléments contribuaient à la décision d'hospitalisation. Les patients ayant ces besoins étaient plus susceptibles d'être hospitalisés. Je voudrais souligner le rôle de ces déterminants sociaux de la santé. C'est quelque chose que nous avons également observé dans d'autres ensembles de données, en particulier dans l'ensemble de données brésilien, qui était un ensemble de données nationales, et nous avons constaté que les facteurs socio-démographiques avaient un impact sur les décisions d'hospitalisation.

Ce que nous avons également constaté, c'est que le modèle - le modèle naïf que l'on est capable de produire - est en réalité assez biaisé. Vous pouvez voir ici comment le modèle se comporte hors échantillon pour les individus noirs et blancs. Le taux de faux positifs du modèle pour les individus noirs était deux fois plus élevé que le taux de faux positifs pour les individus blancs - même si nous contrôlions la race, nous contrôlions les facteurs socio-démographiques, nous contrôlions les déterminants sociaux de la santé. Malgré cela, le modèle était beaucoup plus enclin à prédire que l'individu noir allait être hospitalisé par rapport à un individu blanc. De manière correspondante, il était plus susceptible de faire une prédiction de faux négatif pour un individu blanc par rapport à un individu noir, ce qui suggère qu'il existe apparemment des caractéristiques cachées dans les données qui ne nous sont pas visibles, reflétant peut-être un biais structurel et d'autres facteurs qui font que le modèle fait cette prédiction biaisée. Il existe des moyens et nous les avons abordés dans un article que nous avons publié sur la manière dont on peut corriger ces facteurs et produire des modèles qui n'ont pas ce genre de biais.

Slide 6

Ce sont quelques résultats sur la prédiction des soins intensifs - prédire une admission en soins intensifs - à peu près l'écart médian entre une admission à l'hôpital et une admission en soins intensifs au moins dans notre ensemble de données était d'environ quatre heures. Nous utiliserons différentes coupures. Si vous utilisez les informations les plus récentes, alors vous obtenez des modèles assez précis avec des aires sous la courbe de l'ordre de 93 à 95 %. Si vous commencez à couper les informations que vous alliez utiliser, donc 12 heures à l'avance, la performance du modèle tombe à environ 86 %. 24 heures à l'avance, la performance du modèle tombe à environ 80 %. Ce que je voudrais souligner, c'est que nous avons comparé ces modèles que nous avons développés à certains modèles standard qui prédisent les admissions en soins intensifs, il y a certains modèles de sepsie bien connus, NEWS, qSOFA, ils sont appelés, et ceux-ci sont assez incohérents dans ce cas, ce qui indique que les modèles standard de prédiction des soins intensifs, du moins dans les cas de COVID, échouent à prédire une admission en soins intensifs. Cela indique une signature assez unique de la maladie. Ici, vous trouvez certaines des variables qui étaient à nouveau fortement corrélées avec le résultat d'une admission en soins intensifs. Ce qui était intéressant pour nous, c'est que le score vital que nous avons produit qui capturait l'évolution dynamique des signes vitaux racontait pratiquement toute l'histoire. Il y a d'autres variables ou certaines variables de laboratoire (LDH, CRP) qui ont été identifiées par d'autres études et qui contribuent également, mais si l'on prend simplement l'évolution dynamique des signes vitaux, cela raconte pratiquement toute l'histoire de l'admission en soins intensifs.

Slide 7

Enfin, nous avons produit un certain nombre de calculateurs que nous avons mis à disposition sur le Web. Celles-ci, je le comprends, ont été utilisées par nos collègues et collaborateurs du Massachusetts General Hospital au début de l'épidémie. Il était très facile d'entrer certaines des variables clés et d'obtenir une prédiction, par exemple, pour une admission en soins intensifs ou pour une intubation mécanique d'un patient, et nous avons constaté que, vous savez, nous avons eu du mal à trouver des cas - qu'est-ce que le modèle me dit que le clinicien très expérimenté ne peut potentiellement pas prédire en regardant simplement le patient. Voici quelques cas, et il y en a beaucoup d'autres, où les patients étaient admis, ils étaient stables pendant quelques jours, rien dans leur perspective clinique ne laissait penser que l'état de ce patient allait se détériorer, mais ensuite le modèle a pu - dès l'admission - basé sur certains résultats de laboratoire spéciaux, prédire que le patient allait avoir besoin de soins intensifs. Ce sont deux cas avec les détails de ces cas.

Slide 8

Cela m'amène à la conclusion de cette présentation. Bien sûr, je ne fais que présenter. Beaucoup de personnes ont contribué à ce travail et je tiens à les remercier, y compris les étudiants de mon groupe, mais aussi nos collaborateurs du Mass General Brigham Boston Medical Center et certaines des autres zones d'où nous avons pu obtenir des données. Merci beaucoup pour votre attention et j'attends avec impatience les questions à la fin de la session.