

پیش‌بینی نیاز به مراقبت‌های ویژه برای بیماران کووید ۱۹

با روش‌های یادگیری ماشینی

پیشنهاد پایان‌نامه کارشناسی ارشد

مهنام پدرام

چکیده – با توجه به اطلاعاتی که تاکنون از نحوه‌ی انتشار ویروس SARS-CoV-2 جمع‌آوری شده است، به نظر می‌رسد که ریشه‌کشی جهانی این ویروس و یا دستیابی به ایمنی جمعی (در مقیاس محلی) به زودی امکانپذیر نخواهد بود. شیوع سریع و گسترده‌ی ویروس که در بسیاری موارد نیز بدون علامت‌های معمول رخ می‌دهد، ریشه‌کشی آن را دشوار نموده و تحقق ایده‌ی ایمنی جمعی با فرض توسعه‌ی واکسن‌ها و داروهای مؤثر حداقل چندین سال به طول خواهد انجامید. بنابراین، جان بسیاری از مردم، به ویژه سالمندان، همچنان در خطر بوده و یافتن راه‌های مؤثر برای تشخیص زودهنگام این بیماری در افراد، از اولویت‌های تحقیقاتی سال‌های اخیر به شمار می‌رود. از منظر هزینه‌های مراقبتی و برنامه‌ریزی‌های کلان بیمارستانی، پیش‌بینی نیاز افراد به بستری شدن در بخش مراقبت‌های ویژه‌ی بیمارستان نیز حائز اهمیت بوده و توجه محققان را به خود معطوف نموده است. با توجه به این ضرورت‌ها، در این تحقیق با تکیه بر داده‌گان حاصل از بررسی‌های سوابق پزشکی بیمار، علائم حیاتی و نتایج آزمایش‌های خون و نیز با استفاده از الگوریتم‌های جدید ارائه شده در زمینه‌ی هوش مصنوعی، مدلی ارائه می‌شود که نیاز بیمار به بستری شدن در بخش مراقبت‌های ویژه بیمارستان را به صورت یک متغیر باینری (دو-مقداری) پیش‌بینی می‌کند. الگوریتم مورد نظر براساس داده‌های جمع‌آوری شده توسط بیمارستان سیریولبانز^۱ در سائوپولو برزیل توسعه داده خواهد شد که شامل اطلاعات ۳۸۴ بیمار است.

مقدمه – در زمان همه‌گیری بیماری کووید ۱۹، اغلب کشورهای جهان، از جمله ایران، با مشکل تعداد بالای مبتلایان نیازمند به بستری در بخش مراقبت‌های ویژه بیمارستان و کمبود تعداد این بخش‌ها و امکانات موجود در آن‌ها مواجه بوده‌اند. از آن‌جا که سایر بیمارانی که به عنوان مثال از نارسایی‌های مغزی، قلبی، کلیوی و یا کبدی نیز رنج می‌برند نیز باید در این بخش‌ها بستری شده و تحت مراقبت خاص قرار گیرند، موضوع مدیریت این بخش‌ها و چگونگی تخصیص آن‌ها اهمیت ویژه‌ای در مدیریت مراکز بهداشتی و بیمارستانی پیدا کرده است. با توجه به سرعت و نحوه‌ی انتشار ویروس و همچنین میزان پیشرفت در توسعه‌ی واکسن‌ها و روش‌های درمان دارویی این بیماری، پایان این همه‌گیری در یک یا دو سال آینده چندان تحقق‌پذیر به نظر نمی‌رسد. لذا، با توجه به امکان افزایش تعداد افراد مبتلا، تشخیص سریع این که کدامیک از مبتلایان ممکن است دچار علائم مراحل حاد این بیماری و نیازمند به مراقبت‌های ویژه شوند دارای اهمیت تحقیقاتی و عملی بالایی است. تحقیقاتی که تاکنون در این زمینه انجام شده‌اند نشان می‌دهند با دسترسی به مشخصات بالینی بیماران و به کارگیری الگوریتم‌های یادگیری ماشینی می‌توان با دقت بالایی، آینده‌ی وضعیت بیمار از نظر بهبودی یا تشدید علائم

¹ Sírío-Libanês Hospital

بیماری (به ویژه ناهنجاری‌های شدید ریوی) را پیش‌بینی نمود. در یکی از این پژوهش‌ها و با استفاده از داده‌های جمع‌آوری شده در دو بیمارستان در چین گزارش شده است که با استخراج ویژگی‌هایی مانند میزان آنزیم آلانین آمینوترانسفراز^۲ در کبد، درصد درد ماهیچه^۳ و میزان هموگلوبین خون می‌توان با دقتی حدود ۷۰٪، وقوع موارد حاد را پیش‌بینی نمود [۱]. در تحقیق دیگری که با هدف پیش‌بینی وقوع حالت‌های حاد در افراد مبتلا به کووید ۱۹ با استفاده از الگوریتم XGBoost انجام شده است، ویژگی‌هایی مانند سن، وجود آسیب‌ها و بیماری‌های کلیوی، افزایش LDH، تندنفسی^۴ و هیپرگلیسمی^۵ به عنوان ویژگی‌های اصلی پیش‌بینی کننده معرفی شده‌اند [۲]. نتایجی که در تحقیق [۳] منتشر شده نیز نشان می‌دهد که با تعلیم یک مدل XGBoost و این بار با استفاده از یک مجموعه داده‌گان مربوط به ۳۷۵ بیمار (۲۰۱ بهبود یافته) در بیمارستان دانشگاه تونگجی در ووهان، احتمال مرگ ناشی از این بیماری با دقت ۹۳٪ پیش‌بینی شده است. در این پژوهش نیز ویژگی‌هایی مانند LDH، لنفوسیت^۶، و پروتئین واکنشی سی^۷ به عنوان ویژگی‌های مهم برای پیش‌بینی معرفی شده‌اند. در تحقیقات مشابهی، همین الگوریتم با استفاده از ویژگی‌های دیگری مانند نرخ تنفس، ضربان قلب، شاخص توده بدنی، میزان نیتروژن و کراتینین^۸ خون، بر روی داده‌گان جمع‌آوری شده در بیمارستان‌های مختلف تعلیم داده شده و پیش‌بینی احتمال مرگ با دقت‌های بالاتر از ۹۰٪ گزارش شده است [۴ و ۵ و ۶]. علاوه بر مدل‌های مبتنی بر XGBoost، از مدل ماشین بردار پشتیبان^۹ و رگرسیون لجستیک نیز برای پیش‌بینی حالت‌های حاد و نیز احتمال مرگ مبتلایان استفاده شده است [۷ و ۸ و ۹ و ۱۰]. در این پژوهش‌ها ویژگی‌هایی از بیومارکرهای سرمی (به عنوان مثال، کلسیم، اسید لاکتیک و آلومین، گلوکاتایون^{۱۰}، لنفوسیت‌های T بالغ^{۱۱} و پروتئین تام^{۱۲}) برای تعلیم و یا تخمین پارامترهای مدل استفاده شده است.

تحقیقاتی نیز به منظور تعیین احتمال نیاز به بستری شدن در بخش مراقبت‌های ویژه با استفاده از روش‌های مبتنی بر یادگیری ماشین صورت گرفته است که در آن‌ها علاوه بر الگوریتم‌های XGBoost و رگرسیون لجستیک، مدل‌هایی مانند Random Forest و شبکه‌های عصبی نیز با استفاده از ویژگی‌های بالینی و نتایج آزمایش‌های خون بیماران تعلیم داده شده‌اند [۱۱ و ۱۲ و ۱۳].

با توجه به این پیشینه‌ی تحقیقاتی، هدف از پژوهش پیش رو، مقایسه‌ی این روش‌های یادگیری ماشین در تعیین احتمال نیاز به بستری شدن در بخش مراقبت‌های ویژه است. الگوریتم‌های مذکور با استفاده از ویژگی‌هایی مانند اطلاعات جمعیتی بیمار، سوابق بیماری‌های قبلی، نتایج آزمایش خون، علائم حیاتی و میزان گازهای موجود در خون شریانی که از ۳۸۴ بیمار در بیمارستان سیریولبانز در سائوپولو برزیل جمع‌آوری شده است، تعلیم داده می‌شوند [۱۴]. سپس، عملکرد این الگوریتم‌ها با شاخص‌هایی مانند دقت پیش‌بینی، مساحت زیر منحنی (AUC)، منحنی مشخصه عملکرد سیستم (ROC)، حساسیت و F1-Score مقایسه می‌شوند. مهم‌ترین چالش پیش روی این تحقیق، تعداد محدود نمونه‌ها و همچنین وجود عدم تعادل در تعداد نمونه‌های هر گروه است که عملکرد این الگوریتم‌ها را تحت تاثیر قرار می‌دهد. برای غلبه بر این مشکلات، پیشنهاد می‌شود که از تکنیک‌های وزن‌دار کردن توابع هزینه پیش از تعلیم این مدل‌ها استفاده شود.

² Alanine Aminotransferase ALT

³ Myalgia

⁴ Tachypnea

⁵ Hyperglycemia

⁶ Lymphocyte

⁷ High-sensitivity CRP

⁸ Creatinine

⁹ SVM

¹⁰ GHS

¹¹ CD3 ratio

¹² Total Protein TP

- 1- Jiang, Xiangao, et al. "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity." *Computers, Materials & Continua* 63.1 (2020): 537-551.
- 2- A. Vaid, S. Somani, A.J. Russak, J.K. De Freitas, F.F. Chaudhry, I. Paranjpe, et al. Machine learning to predict mortality and critical events in covid-19 positive New York city patients: a cohort study *J Med Internet Res*, 49 (6) (2020), pp. 1918-1929
- 3- L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jin, M. Zhang, et al. A machine learning-based model for survival prediction in patients with severe covid-19 infection *MedRxiv* (2020)
- 4- E. Rechtman, P. Curtin, E. Navarro, S. Nirenberg, M.K. Horton Vital signs assessed in initial clinical encounters predict covid-19 mortality in an nyc hospital system *Sci Rep*, 10 (2020), pp. 1-6
- 5- D. Bertsimas, G. Lukin, L. Mingardi, O. Nohadani, A. Orfanoudaki, B. Stellato, H. Wiberg, S. Gonzalez-Garcia, C.L. Parra-Calderon, K. Robinson, et al. Covid-19 mortality risk assessment: an international multi-center study *PloS One*, 15 (2020), Article e0243262
- 6- X. Guan, B. Zhang, M. Fu, M. Li, X. Yuan, Y. Zhu, J. Peng, H. Guo, Y. Lu Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized covid-19 patients: results from a retrospective cohort study *Ann Med*, 53 (2021), pp. 257-266,
- 7- A.L. Booth, E. Abels, P. McCaffrey Development of a prognostic model for mortality in covid-19 infection using machine learning *Mod Pathol* (2020), pp. 1-10
- 8- L. Sun, F. Song, N. Shi, F. Liu, S. Li, P. Li, W. Zhang, X. Jiang, Y. Zhang, L. Sun, X. Chen, Y. Shi Combination of four clinical indicators predicts the severe/critical symptom of patients infected covid-19 *J Clin Virol*, 128 (2020), p. 104431
- 9- H. Yao, N. Zhang, R. Zhang, M. Duan, T. Xie, J. Pan, E. Peng, J. Huang, Y. Zhang, X. Xu, et al. Severity detection for the coronavirus disease 2019 (covid-19) patients using a machine learning model based on the blood and urine tests *Frontiers in cell and developmental biology*, 8 (2020), p. 683
- 10- C. Hu, Z. Liu, Y. Jiang, O. Shi, X. Zhang, K. Xu, et al. Early prediction of mortality risk among patients with severe COVID-19, using machine learning *Int J Epidemiol*, 49 (6) (2020), pp. 1918-1929
- 11- Heldt, Frank S., et al. "Early risk assessment for COVID-19 patients from emergency department data using machine learning." *Scientific reports* 11.1 (2021): 1-13.
- 12- Aznar-Gimeno, Rocío, et al. "A Clinical Decision Web to Predict ICU Admission or Death for Patients Hospitalised with COVID-19 Using Machine Learning Algorithms." *International Journal of Environmental Research and Public Health* 18.16 (2021): 8677.
- 13- Ezz, Mohamed, Murtada K. Elbashir, and Hosameldeen Shabana. "Predicting the need for icu admission in covid-19 patients using xgboost." *Computers, Materials and Continua* (2021): 2077-2092.
- 14- <https://www.kaggle.com/S%C3%ADrio-Libanes/covid19>