Journal of Modeling in Engineering 21 (2023) 231-241

Semnan University

Journal of Modeling in Engineering

Journal homepage: https://modelling.semnan.ac.ir/

Research Article

A Novel Approach Based on Catboost and Explainable Artificial Intelligence for Diagnosis of COVID-19 Cases Using Patients' Symptoms

Samaneh Emami[1,*], Ali SeyyedMomeni[2], Hamid Nasiri[3]

1. Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran.
2. Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran.
3. Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran.

*Corresponding Author: s_emami@semnan.ac.ir

PAPER INFO

Paper history:
Received: 25 December 2022
Revised: 25 April 2023
Accepted: 02 May 2023

Keywords:
CatBoost algorithm,
Corona Virus,
Deep Neural Network,
Covid-19 Disease,
Machine Learning,
SHAP.

ABSTRACT

The COVID-19 virus, which was discovered in December 2019 in the city of Wuhan, China and quickly spread throughout the world, continues to be an important threat to the health of the world. Despite all the strategies used to deal with the spread of COVID-19, more contrivances are still needed to deal with its consequences. In this research, the clinical characteristics of people have been used as input data to diagnose a person with COVID-19, which is the result of collecting information from similar studies. Also, various algorithms including support vector machine, logistic regression, k nearest neighbor (k=9), simple bayes, random forest, LightGBM, XgBoost and CatBoost have been used, among which the CatBoost algorithm, with a sensitivity of 97.97%, accuracy 97.72% and 96.96% accuracy showed the best results. In this algorithm, the trial and error method has been used to adjust hyperparameters as accurately as possible to achieve the desired results, and SHAP is used to interpret the results and determine the impact of features on the output.

DOI: https://doi.org/10.22075/jme.2023.29413.2385

How to cite this article:
Emami, S., Nasiri, H., & SeyyedMomeni, A. (2023). A novel approach based on CatBoost and explainable artificial intelligence for diagnosis of COVID-19 cases using patients' symptoms. Journal of Modeling in Engineering, 21(74), 231-241. doi: 10.22075/jme.2023.29413.2385

یک رویکرد جدید مبتنی بر الگوریتمCatBoost و هوش مصنوعی تفسیرپذیر به منظور تشخیص بیماری کرونا بر اساس علائم بیماری

سمانه امامی[۱،*]، علی سیدمومنی[۲] و حمید نصیری[۳]

اطلاعات مقاله

نوع مقاله: پژوهشی
دریافت مقاله: ۱۴۰۱/۱۰/۰۴
بازنگری مقاله: ۱۴۰۲/۰۲/۰۵
پذیرش مقاله: ۱۴۰۲/۰۲/۱۲

واژگان کلیدی:
الگوریتم CatBoost،
ویروس کرونا،
شبکه عصبی عمیق،
بیماری کووید-۱۹،
یادگیری ماشین،
SHAP.

چکیده

ویروس کرونا که در ماه دسامبر ۲۰۱۹ در شهر ووهان چین دیده شد و به سرعت در سراسر جهان شیوع پیدا کرد، همچنان یک تهدید مهم برای سلامت جهان به شمار می‌آید. علی‌رغم همه استراتژی‌های مورد استفاده برای مقابله با گسترش کووید-۱۹، هنوز به تدابیر بیشتری برای رفع پیامدهای ناشی از آن نیاز است. در این پژوهش برای تشخیص فرد مبتلا به کووید-۱۹ از ویژگی‌های بالینی افراد به عنوان داده‌های ورودی استفاده شده است که حاصل جمع‌آوری اطلاعات از پژوهش‌های مشابه است. همچنین از الگوریتم‌های مختلفی شامل یادگیری ماشین بردار پشتیبان، رگرسیون لجستیک، k نزدیکترین همسایه (k=9)، بیز ساده، جنگل تصادفی، LightGBM، XgBoost و CatBoost استفاده شده که از میان آنها الگوریتم CatBoost، با کسب حساسیت ۹۷/۹۷ درصد، دقت ۹۷/۷۲ درصد و صحت ۹۶/۹۶ درصد بهترین نتایج را از خود نشان داد. در این الگوریتم، برای تنظیم هر چه دقیقتر فوق‌پارامترها به منظور رسیدن به نتایج مطلوب از روش آزمون و خطا استفاده شده و از SHAP برای تفسیر نتایج و مشخص کردن تاثیر ویژگی‌ها بر خروجی الگوریتم استفاده گردیده است.

۱- مقدمه

همه گیری کووید-۱۹ مهمترین مصیبت جهانی قرن و بزرگترین چالشی است که بشر از زمان جنگ جهانی دوم با آن روبه‌رو بوده است. امروزه مکانیسم منحصر به فرد بیماری‌زایی SARS-COV-2 به همراه طیف وسیعی از علائم، موضوع مطالعات و پژوهش‌های متخصصین است. در همین راستا، جمع‌آوری اطلاعات منتشر شده درباره ویژگی‌های بالینی کووید-۱۹ از مطالعات مختلف برای ایجاد یک مجموعه‌داده جامع برای دستیابی به بینشی عمیق‌تر از موارد بیماری‌زایی و ویژگی‌های بالینی کووید-۱۹ مورد نیاز می‌باشد [۱]. امروزه استفاده از الگوریتم‌های مختلف هوش مصنوعی در تشخیص انواع موضوعات راه‌حلی رایج و کاربردی می‌باشد [۲ ، ۳ ، ۴]. توانایی تشخیص مبتلایان به کووید-۱۹ نیز بر اساس متغیرهای بالینی و با استفاده از مدل محاسباتی قابل دسترس، برای مقابله با فقدان قابلیت‌های آزمایش برای کووید-۱۹ در سراسر جهان به خصوص در کشورهایی که قادر به تخصیص منابع آزمایشی کافی و سیستم‌های مراقب بهداشتی نیستند، بسیار مفید خواهد بود [۱،۵].

در این پژوهش، نسبت به پژوهش‌های مشابه، مدل دقیق تری از تشخیص کووید-۱۹بر اساس ویژگی های بالینی بیمار ایجاد شده است. علائم و نتایج آزمایشات بالینی با استفاده از یادگیری‌ماشین برای تجزیه و تحلیل داده‌ها مورد بررسی قرار می‌گیرند و ورودی الگوریتم CatBoost

* پست الکترونیک نویسنده مسئول: s_emami@semnan.ac.ir
۱. استادیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان
۲. کارشناسی، دانشکده مهندسی برق وکامپیوتر، دانشگاه سمنان
۳. دانشجوی دکتری، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر

خواهند بود. هدف این پژوهش بررسی همبستگی بین متغیرهای بالینی و ایجاد یک مدل طبقه‌بندی‌کننده یادگیری ماشین برای تمایز بیماران کووید-۱۹ و بیماران آنفولانزا تنها بر اساس متغیرهای بالینی و همچنین تفسیر خروجی مدل با استفاده از SHAP[1] می‌باشد.

ساختار مقاله به این شرح می‌باشد: در بخش دوم به مرور کارهای انجام شده در زمینه تشخیص ویروس کرونا پرداخته می‌شود. در بخش سوم، به معرفی الگوریتم‌های پایه استفاده شده در روش پیشنهادی پرداخته و در بخش چهارم، روشی برای تشخیص بیماری کووید-۱۹ از روی علائم بیمار ارائه می‌گردد. مقایسه نتایج حاصل از روش‌های پیشنهادی در بخش پنجم بیان شده و نتیجه‌گیری پژوهش در بخش ششم ارائه می‌گردد.

۲- مروری بر پژوهش‌های پیشین

در مدتی که از بروز بیماری کووید-۱۹ گذشته است، پژوهش‌های زیادی در زمینه استفاده از الگوریتم‌های یادگیری ماشین برای تشخیص این بیماری صورت گرفته است.

لی۲و همکارانش [۶] متغیرهای بالینی افراد را مورد مطالعه قرار داده و از الگوریتم XGboost به عنوان طبقه‌بندی‌کننده استفاده کرده‌اند. در این پژوهش، ارتباطات جدیدی بین متغیرهای بالینی کشف شد که دارای حساسیت ٪۹۲/۵ و دقت تشخیص ٪۹۷/۹ تشخیص است.

یازید[3] و همکارانش [۷] در پژوهشی بر روی داده‌های وزارت بهداشت رژیم صهیونیستی که شامل هشت ویژگی دودویی (همچون جنسیت و سن بالای ۶۰ و ارتباط با افراد مبتلا) می‌باشد، با استفاده از الگوریتم Xgboost به حساسیت ٪۸۷/۳۰ و تشخیص ٪۷۱/۹۸ و در آزمایشی دیگر به حساسیت ٪۸۵/۷۶ و تشخیص ٪۷۹/۱۸ رسیده‌اند. استفاده ازویژگی‌هایی که مقادیر بیشتری از دو حالت را شامل شوند ممکن بود بر دقت این پژوهش تاثیر بهتری داشته باشد.

مکرم سوئی[4] و همکارانش [۸] در پژوهشی از الگوریتم ژنتیک مرتب‌سازی غیر مسلط (NSGA-II[5]) برای انتخاب ویژگی‌ها، به‌منظور به حداقل رساندن تعداد ویژگی‌ها و به حداکثر رساندن وزن ویژگی‌های انتخاب شده استفاده کرده و در مرحله طبقه‌بندی از طبقه‌بندی کننده AdaBoost [۸] استفاده نموده‌اند.

بانیک[6] و همکارانش [۹] در پژوهشی با هدف برآورد احتمال ابتلای افراد به کووید-۱۹ از چندین الگوریتم یادگیری ماشین مثل درخت تصمیم[7] (DT) و ماشین بردار پشتیبان[8] (SVM) استفاده کرده‌اند تا یک مدل دقیق برای پیش‌بینی احتمال آلوده شدن بیمار بر اساس علائم بالینی بیماران مشخص گردد. این مدل ویژگی‌های بسیاری را که دارای اهمیت یکسانی هستند در نظر گرفت. با این حال، مجموعه ای از ویژگی‌ها اهمیت بیشتری برای شناسایی موارد کووید-۱۹ دارند.

نان سان۹ و همکارانش [۱۰] در پژوهشی با هدف استخراج عوامل خطرناک در بروز کووید-۱۹ از داده های بالینی بیماران مبتلا به کووید-۱۹ از چهار الگوریتم یادگیری ماشین کلاسیک شامل رگرسیون لجستیک۱۰ (LR)، ماشین بردار پشتیبان، درخت تصمیم، جنگل تصادفی۱۱ (RF) و یک روش مبتنی بر یادگیری عمیق برای تشخیص زودهنگام کووید-۱۹ استفاده کرده‌اند. بر اساس نتایج، مدل پیش‌بینی رگرسیون لجستیک با نرخ تشخیص ٪۹۵ و حساسیت ٪۸۲ را ارائه می‌دهد که آن را برای غربالگری عفونت اولیه کووید-۱۹ بهینه می‌کند.

چانسیک۱۲ و همکارانش [۱۱] در پژوهشی برای تشخیص کووید-۱۹ از اطلاعات اجتماعی، جمعیت شناختی و پزشکی استفاده کردند. در این مطالعه از مدل‌های یادگیری ماشین نظیر حداقل عملگر انقباض و انتخاب مطلق13 (LASSO)، ماشین بردار پشتیبان خطی، جنگل تصادفی و k نزدیکترین همسایه۱۴ استفاده کردند که در این میان LASSO و SVM به ترتیب به نرخ حساسیت ٪۹۰/۷۰ و ٪۹۲ و نرخ تشخیص ٪۹۱/۴۰ و ٪۹۱/۸۰ رسیدند.

کنگ فانگ۱۵ و همکارانش [۱۲] در پژوهشی به منظور شناسایی اولیه بیمار برای بهینه‌سازی استراتژی درمانی یک

[1] Shapley additive explanations
[2] Li
[3] Yazeed
[4] Soui, Makram
[5] Non-dominated Sorting Genetic Algorithm II
[6] Banik
[7] Decision tree
[8] Support Vector Machine
[9] Nan, Sun
[10] Logistic Regression
[11] Random Forest
[12] Chansik
[13] Least Absolute Shrinkage and Selection Operator
[14] k-nearest neighbors
[15] Cong Fang

سیستم هشدار اولیه با تکنیک‌های یادگیری ماشین نظیر شبکه‌های حافظه کوتاه‌مدت-بلندمدت[1] (LSTM) برای مدل سازی اطلاعات زمانی در داده های اسکن CT و حداقل عملگر انقباض و انتخاب مطلق (LASSO) برای انتخاب ویژگی‌های داده‌های بالینی، برای پیش‌بینی پیشرفت بدخیم کووید-۱۹ طراحی کردند که در آن از داده‌های اسکن CT و داده‌های بالینی بیماران سرپایی استفاده کردند و به نرخ حساسیت ٪۸۸/۳ و نرخ تشخیص ٪۸۸/۵ رسیدند.

۳- معرفی الگوریتم‌های پایه استفاده شده

در این بخش به معرفی و شرح مختصری از الگوریتم‌های استفاده شده در روش پیشنهادی پرداخته خواهد شد.

۳-۱- الگوریتم XGBoost[2]

این الگوریتم مبتنی بر تقویت درخت است که در سال ۲۰۱۶ توسط چن و گسترین[3] [۱۳] پیشنهاد شده است و در واقع یک نسخه بهبود یافته از روش تقویت گرادیان درخت تصمیم[4] (GBDT) است [۱۴،۱۵]. الگوریتم XGBoost یک الگوریتم مقیاس‌پذیر است که برای پیدا کردن بهترین مدل درخت از بسط تیلور مرتبه دوم استفاده می‌کند [۱۶]. در این الگوریتم هدف پیدا کردن تابعی مانند $\hat{F}(x)$ است که x را به y نگاشت می‌کند تا خروجی $\hat{y}$ را پیش‌بینی کند. همانطور که در (۱) نشان داده شده است، خروجی مدل $\hat{y}_i$ مجموع تمام امتیازات پیش‌بینی شده توسط k درخت است.

$$y_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in \Gamma \qquad (۱)$$

در رابطه (۱)، x_i بردار ویژگی ورودی، k تعداد درخت‌های رگرسیون، f_k بیانگر امتیاز k امین درخت یا همان امتیاز برگ[5] و Γ فضای تابعی[6] است که شامل تمام درخت‌های رگرسیون ممکن می‌باشد [۱۷،۱۸]. به منظور یادگیری مجموعه توابع به کار رفته در مدل، الگوریتم XGBoost سعی می‌کند تابع هدف تنظیم شده را که در (۲) آمده‌است، کمینه کند.

$$\phi = \sum_{i=1}^{n} l(y_i, y_i) + \sum_{j=1}^{k} \Omega(f_j) \qquad (۲)$$

در رابطه (۲)، $l(y_i, \hat{y}_i)$ اختلاف بین مقدار هدف y_i و مقدار پیش‌بینی شده $\hat{y}_i$ را اندازه‌گیری می‌کند و $\Omega(f)$ عبارت تنظیم‌کننده است که مدل‌های پیچیده را جریمه می‌کند تا از بیش‌برازش[7] جلوگیری کند. $\Omega(f)$ با استفاده از رابطه (۳) محاسبه می‌شود.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \| \omega \| \qquad (۳)$$

در رابطه (۳)، T تعداد گره‌های برگ و ω امتیاز هر برگ است. λ و γ نیز ضرایب ثابتی هستند که به ترتیب میزان تنظیم مدل و هزینه پیچیدگی را معین می‌کند[۱۹،۲۰].

۳-۲- الگوریتم CatBoost

محبوب‌ترین پیاده‌سازی‌های تقویت گرادیان از درخت تصمیم به عنوان پیش‌بینی‌کننده‌ی پایه استفاده می‌کنند که برای کار با ویژگی‌های عددی ساده است. در حالی که در عمل بسیاری از مجموعه‌داده‌ها شامل ویژگی‌های طبقه‌بندی شده هستند که برای پیش‌بینی هم در نظر گرفته می‌شوند. ویژگی‌های طبقه‌بندی شده دارای مجموعه‌ای مجزا از مقادیر است که قابل مقایسه با یکدیگر نیستند. متداول‌ترین روش برای برخورد با ویژگی‌های طبقه‌بندی شده در تقویت گرادیان، تبدیل آنها به اعداد در مراحل پیش‌پردازش داده‌هاست. الگوریتم CatBoost از یک استراتژی به مراتب بهینه‌تر از جایگزینی دسته‌ها با مقدار متوسط برچسب‌ها در کل داده‌ها استفاده می‌کند که هم احتمال وقوع بیش‌برازش را کاهش می‌دهد و هم اجازه استفاده از تمام داده‌ها را برای آموزش مدل می‌دهد. برای این کار باید یک جایگشت تصادفی از داده‌ها ایجاد کرده و برای هر مثال میانگین مقدار برچسب را با همان مقداری که قبل از جایگشت داده شده بود جایگذاری شود [۲۱].

$$\frac{\sum_{j=1}^{p-1}[x_{\sigma_j,k} = x_{\sigma_p,k}]Y_{\sigma_j} + a.p}{\sum_{j=1}^{p-1}[x_{\sigma_j,k} = x_{\sigma_p,k}] + a} \qquad (۴)$$

در رابطه (۴)، $X_i = (x_{i,1}, \dots, x_{i,m})$ آرایه‌ای به طول m از ویژگی‌ها، $Y_i \epsilon R$ برچسب مقادیر و $\sigma = (\sigma_1, \dots, \sigma_n)$

[1] Long Short-Term Memory Networks
[2] Extreme Gradient Boosting
[3] Chen and Guestrin
[4] Gradient-Boosted Decision Tree
[5] Leaf Score
[6] Function Space
[7] Overfitting

جایگشت می‌باشد. اضافه کردن مقدار اولیه p و پارامتر $a > 0$ که یک عمل رایج است، باعث کاهش نویز ناشی از دسته‌های فرکانس پایین می‌شود [۲۲،۲۱].

۳-۳- الگوریتم LightGBM

LightGBM[1] یک الگوریتم متن‌باز مبتنی بر درخت تصمیم است که توسط مایکروسافت توسعه یافته است که یک درخت تصمیم‌گیری مبتنی بر هیستوگرام بهینه را پیاده‌سازی می‌کند و از نظر کارایی و مصرف حافظه عملکرد خوبی دارد. از دیگر مزایای این الگوریتم می‌توان به آموزش موازی[2]، منظم‌سازی[3] و دسته‌بندی[4] اشاره کرد. این الگوریتم از دو تکنیک جدید به نام‌های نمونه‌برداری یک طرفه مبتنی بر گرادیان[5] (GOSS) و دسته‌بندی ویژگی انحصاری[6] (EFB) استفاده می‌کند که به الگوریتم اجازه می‌دهد تا در ضمن حفظ دقت، سریع اجرا شود [۲۳].

نمونه‌برداری یک طرفه مبتنی بر گرادیان

نمونه‌هایی با گرادیان‌های مختلف، نقش‌های متفاوتی در محاسبه اطلاعات دارند. طبق تعریف، به دست آوردن اطلاعات نمونه‌هایی با گرادیان بزرگتر نقش بیشتری در کسب اطلاعات دارند. در نتیجه GOSS نمونه‌هایی با گرادیان‌های بزرگتر را نگه داشته و نمونه‌هایی که شیب‌های کوچکتری دارند، به صورت تصادفی حذف می‌کند. این کار دقت تخمین بالاتری را در مقایسه با نمونه‌برداری تصادفی یکنواخت با نرخ نمونه‌گیری یکسان ارائه می‌کند [۲۴]. از نظر ریاضی، یک مجموعه داده آموزشی با n نمونه $\{x_1, x_2, x_3, ..., x_n\}$ را در نظر بگیرید که در آن x_i یک بردار با بعد s در فضای ورودی x^s است. در هر تکرار تقویت گرادیان، گرادیان منفی تابع ضرر نسبت به خروجی مدل به صورت $\{g_1, g_2, g_3, ..., g_n\}$ نشان داده می‌شود. در روش GOSS ابتدا نمونه‌های آموزشی به ترتیب نزولی بر اساس مقادیر صحیح گرادیان آنها مرتب می‌شوند. سپس a درصد بالا از نمونه‌ها با گرادیان‌های بزرگتر در زیرمجموعه‌ای مانند A ذخیره می‌شوند. پس از آن، برای مجموعه باقیمانده A^C که شامل $(1-a)$ درصد از نمونه‌های با گرادیان‌های کوچکتر است، زیرمجموعه B به اندازه $b * |A^c|$ به صورت تصادفی نمونه‌گیری می‌شود. در نهایت، نمونه‌ها بر روی مجموعه $A \cup B$ باتوجه به افزایش تخمینی واریانس $\tilde{V}_j(d)$ تقسیم می‌شوند که می‌توان با استفاده از رابطه (۵) محاسبه کرد [۲۵،۲۳].

در رابطه (۵)، $A_l = \{x_i \in A : x_{ij} \le d\}$، $A_r = \{x_i \in A : x_{ij} > d\}$، $B_l = \{x_i \in B : x_{ij} \le d\}$، $B_r = \{x_i \in B : x_{ij} > d\}$ و $\frac{1-a}{b}$ برای نرمالسازی[7] گرادیان‌ها استفاده می‌شود [۲۰،۲۲].

$$\tilde{V}_j(d) = \frac{1}{n}\left(\frac{\left(\sum_{x_i \in A_l} g_i + \frac{1-a}{b}\sum_{x_i \in B_l} g_i\right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b}\sum_{x_i \in B_r} g_i\right)^2}{n_r^j(d)} \right) \quad (۵)$$

دسته‌بندی ویژگی انحصاری (EFB)

داده‌ها در یک فضای با ابعاد بالا معمولاً بسیار پراکنده هستند و تجزیه و تحلیل آنها چالش‌های زیادی دارد. پراکندگی فضای ویژگی باعث می‌شود تا با طراحی روشی بدون تلفات، تعداد ویژگی‌ها کاهش یابد. در روش EFB، چندین ویژگی در یک واحد جمع می‌شوند. بنابراین، پیچیدگی محاسباتی LightGBM از $O(\#data \times \#feature)$ به $O(\#data \times \#bundle)$ تغییر می‌کند و تا زمانی که $\#bundle \ll \#features$ سرعت آموزش بدون کاهش دقت، افزایش می‌یابد [۲۵،۲۳]. به طور کلی، الگوریتم LightGBM یک نسخه بهبود یافته و بسیار کارآمدتر از روش درخت تصمیم تقویت شده با گرادیان (GBDT) مانند XGBoost است [۲۵، ۲۶، ۲۷]. این الگوریتم می‌تواند روند آموزشی GBDT را بیش از ۲۰ برابر تسریع بخشد، در حالی که تقریباً به همان دقت دست می‌یابد. خروجی پیش‌بینی شده توسط مدل LightGBM در رابطه (۶) آمده است که در آن M و $h_m(x)$ به ترتیب بیشترین تعداد تکرار و درخت تصمیم پایه را نشان می‌دهند.

$$F_M(x) = \sum_{m=1}^{M} h_m(x) \quad (۶)$$

۴-۳- توضیحات افزودنی شاپلی[8]

[1] Light Gradient Boosting Machine
[2] Parallel training
[3] Regularization
[4] Bagging
[5] Gradient-based One Side Sampling
[6] Exclusive Feature Bundling
[7] Normalize
[8] Shapley Additive Explanations

بر اساس تئوری بازی‌های کلاسیک SHAP می‌تواند خروجی مدل‌های هوش مصنوعی را توصیف کند. به بیان دیگر مقادیر SHAP نشان می‌دهند که چگونه هر نقطه از متغیر می‌تواند به پیش‌بینی مدل کمک کند. SHAP چهارچوبی بر پایه تئوری بازی‌هاست که یک رویکرد واحد را برای تفسیر پیش‌بینی‌های مدل‌های یادگیری ماشین ارائه می‌دهد. همچنین اهمیت ویژگی‌ها را به عنوان متغیر ورودی بیان می‌کند[۲۹،۲۸]. SHAP یک روش انتساب ویژگی افزودنی است در سال ۲۰۱۷ توسط لونبرگ و لی [۳۰] ارائه شده است. $g(z)$ مدل توصیفی SHAP برای پیش‌بینی $f(x)$ است که در رابطه (۷) بیان شده است[۳۱،۲۶].

$$f(x) = g(z') = \Phi_0 + \sum_{i=1}^{M} \Phi_i z_i' \qquad (۷)$$

در رابطه (۷)، m نشان‌دهنده متغیرهای ورودی، z' نمایش‌دهنده ورودی ساده شده و Φ_0 یک مقدار ثابت در غیاب همه ویژگی‌هاست. مقادیر SHAP جهت تاثیر ویژگی‌ها روی خروجی مدل را نمایش می‌دهد [۳۳،۳۲] و می‌تواند به عنوان یک راه‌حل ممکن برای (۷) که سه ویژگی دقت محلی[1]، فقدان[2] و ثبات[3] را برآورده می‌کند نشان داد که رابطه (۸) برقرار می‌باشد.

$$\phi_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!(M-|z'|-1)!}{M!}[f_x(z') - f_x(z'^{\vee})] \qquad (۸)$$

در رابطه (۸)، x' نشان دهنده آرایه‌ای از m متغیر انتخابی ورودی، f مدل آموزشی، $[f_x(z') - f_x(z'^{\vee})]$ نشان دهنده مشارکت iامین متغیر خروجی مدل و $|z'|$ نشان دهنده‌ی تمامی مقادیر غیر صفر z' است[۲۶، ۳۱، ۳۴].

۴-روش پیشنهادی

داده‌های بالینی بیماران به صورت دستی جمع‌آوری شده است و در مجموع ۱۴۳۹ نفر مورد بررسی قرار گرفته‌اند که داده‌های بالینی اولیه از تجزیه و تحلیل حذف شده‌اند و پس از بررسی نهایی مجموعه‌داده‌های بیماران فوق حاصل شده‌اند. مجموعه‌داده‌های نهایی با متغیرهای بالینی برای هر بیمار به همراه منبع مطالعه‌ی هر بیمار در مخزن Github به آدرس https://github.com/yoshihiko1218/COVID19ML یافت می‌شود [۶]. متغیرها شامل سن، جنسیت، سطح سرمی‌های نوتروفیل، لکوسیت، لنفوسیت به صورت پیوسته و معمولی، نتیجه اسکن CT و سایر علائم گزارش شده شامل تب، سرفه، اسهال، زخم‌گلو، حالت تهوع، خستگی، درجه حرارت بدن و زمینه عوامل خطر (بیماری‌های کلیوی و دیابت) است. داده‌های فوق را به دو مجموعه‌ی جدا برای آموزش و آزمایش با نسبت ۸۰ به ۲۰ تقسیم گردید و از الگوریتم‌های یادگیری ماشین بردار پشتیبان، رگرسیون لجستیک، k نزدیکترین همسایه (k=9)، بیز ساده و جنگل تصادفی و CatBoost [۲۱] برای تشخیص بیماری استفاده شده است. در الگوریتم CatBoost، به جای استفاده از بهینه‌ساز Bayesian با تنظیم فوق‌پارامترها به صورت دستی، ورودی‌های الگوریتم تعیین شده است.

۵-ارزیابی آزمایشات

به منظور ارزیابی روش پیشنهادی از جمع‌آوری داده‌های بالینی به دست آمده از مطالعات مختلف برای تشخیص بیماری کووید-۱۹ از آنفولانزا استفاده شده است. خلاصه‌ای از فراداده‌های مورد استفاده در جدول ۱ آمده است.

جدول ۱-بخشی از فراداده متغیرهای بالینی (مجموعه‌داده تسه‌لی)

متغیرهای بالینی	تعداد داده‌ها	میانگین	میانه	انحراف معیار
سن	۳۸۹	۳۸/۹۱	۳۹	۲۱/۸۵
تعداد اعضای آلوده شده خانواده	۵۴	۳/۳۷	۲	۲/۶۳
نوتروفیل	۱۰۳	۶/۸۵	۳/۳۱	۱۲/۶۲
سطح سرمی گلبول های سفید	۱۳۰	۷/۰۳	۵/۹۶۵	۴/۲۵
لنفوسیت ها	۱۳۵	۲/۰۲	۰/۹۸	۴/۲۰
پلاکت ها	۵۰	۲۲۰/۳۲	۱۸۵/۵	۱۴۶/۳۳
سلول های قرمز خون	۴	۴/۲۲۵	۴/۲۰۵	۰/۱۸
هموگلوبین	۲۴	۴۵/۵	۱۴/۵	۴۹/۹۹
تب	بله	۲۶۱		۹۱/۲۵
	خیر	۲۵		۸/۷۴
سرفه	بله	۱۶۴		۸۲/۸۲
	خیر	۳۴		۱۷/۱۷
تنگی نفس	بله	۴۵		۶۰
	خیر	۳۰		۴۰

[1] Local accuracy
[2] Missingness
[3] Consistency

نتایج حاصل از پیاده‌سازی پنج الگوریتم یادگیری ماشین بردار پشتیبان، رگرسیون لجستیک، k نزدیکترین همسایه (k=9)، بیز ساده و جنگل تصادفی در جدول ۲ نشان داده شده است. در این الگوریتم‌ها، نسبت مجموعه داده آموزشی به آزمایشی ۸۰ به ۲۰ در نظر گرفته شده است. نتایج نشان می‌دهد از میان این پنج الگوریتم، ماشین بردار پشتیبان از نظر دقت، صحت و امتیاز F1 [۳۵] بهترین عملکرد را داشته و جنگل تصادفی از نظر بازخوانی نتایج بهتری از خود نشان داده است.

جدول ۲-نتایج پیاده‌سازی پنج الگوریتم پایه

نام الگوریتم	دقت	صحت	بازخوانی	امتیاز **F1**
ماشین بردار پشتیبان	۹۱/۸۱	۹۱/۷۸	۸۴/۸۱	۸۸/۱۵
رگرسیون لجستیک	۸۸/۱۸	۸۵/۳۳	۸۱/۰۱	۸۳/۱۱
k نزدیکترین همسایه	۸۹/۰۹	۹۱/۰۴	۷۷/۲۱	۸۳/۵۶
بیز ساده	۸۱/۸۱	۷۶/۷۱	۷۰/۸۸	۷۳/۶۸
جنگل تصادفی	۸۵/۰	۷۰/۹۰	۹۸/۷۳	۸۲/۵۳

جدول ۳ ویژگی‌های مجموعه داده زوابی۱ [۷] را نشان می‌دهد. پارامترهای الگوریتم CatBoost به صورت دستی و با روش آزمون و خطا تنظیم شده‌اند. مقدار پارامترهای تعیین شده در جدول نشان داده شده است.

برای مقایسه نتایج روش پیشنهادی با روش‌های دیگر، از معیارهای ارزشیابی زیر استفاده شده است:

حساسیت[۲]: نتیجه تقسیم موارد مثبت صحیح بر مجموع مثبت صحیح و منفی ناصحیح است که در رابطه (۹) بیان شده است.

$$Sensitivity = \frac{TP}{TP+TN} \qquad (۹)$$

تشخیص: نتیجه تقسیم موارد منفی صحیح بر مجموع منفی صحیح و مثبت ناصحیح است که در رابطه (۱۰) بیان شده است.

$$Specificity = \frac{TN}{TN+FP} \qquad (۱۰)$$

دقت: نتیجه تقسیم موارد مثبت صحیح بر مجموع مثبت صحیح و منفی ناصحیح است که در رابطه (۱۱) بیان شده است.

جدول ۳-ویژگی‌های مجموعه داده (مجموعه‌داده زوابی)

ویژگی		مجموع =تعداد ۹۹۲۳۲	کووید-۱۹ مثبت =تعداد ۳۹۳	کووید-۱۹ منفی =تعداد ۹۰۸ ۳۹
جنسیت	زن	۴۸۸۸۲	۳۵۸۸	۴۵۲۹۴
	مرد	۵۰۳۵۰	۴۸۰۵	۴۵۵۴۵
سن بالای ۶۰	درست	۱۵۲۷۹	۱۶۶۰	۱۳۶۱۹
	نادرست	۸۳۹۵۳	۶۷۳۳	۷۷۲۲۰
سرفه	درست	۱۴۷۶۸	۴۰۵۳	۱۰۷۱۵
	نادرست	۸۴۲۲۳	۴۳۱۴	۷۹۹۰۹
تب	درست	۸۱۲۲	۳۷۳۵	۴۳۸۷
	نادرست	۹۰۸۶۸	۴۶۳۱	۸۶۲۳۷
گلو درد	درست	۱۲۷۳	۱۱۷۷	۹۶
	نادرست	۹۵۰۶۲	۷۰۰۳	۸۸۰۵۹
تنگی نفس	درست	۹۳۰	۸۵۹	۷۱
	نادرست	۹۵۴۰۵	۷۳۲۱	۸۸۰۸۴
سردرد	درست	۱۷۹۹	۱۷۳۱	۶۸
	نادرست	۹۴۵۳۶	۶۴۴۹	۸۸۰۸۷
تماس با افراد آلوده	درست	۵۵۰۷	۴۰۵۲	۱۴۵۵
	نادرست	۹۳۷۲۵	۴۳۴۱	۸۹۳۸۴

جدول ۴-پارامترهای تنظیم شده برای دسته‌بندی‌کننده CatBoost

پارامتر	تعداد تکرار	نرخ یادگیری	حداکثر عمق	حالت تصادفی
مقدار	۱۰۰	۰/۸۶	۵	۱۲۴

$$Precision = \frac{TP}{TP+FP} \qquad (۱۱)$$

صحت: نتیجه تقسیم مجموع موارد منفی صحیح و مثبت صحیح بر مجموع موارد مثبت صحیح و منفی صحیح و مثبت ناصحیح و منفی ناصحیح (تعداد کل موارد) است که در رابطه (۱۲) بیان شده است.

[1] Zoabi

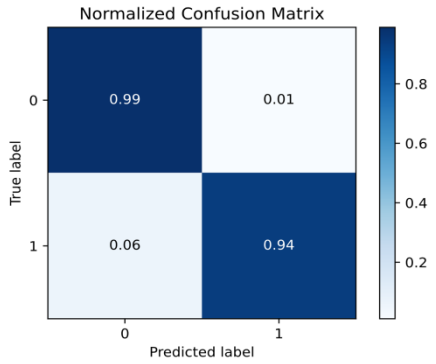
[2] Sensitivity

$$Accuracy = \frac{TP+FP}{TP+FP+TN+FN} \qquad (۱۲)$$

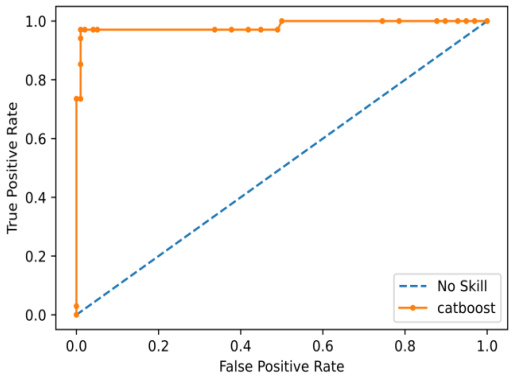
امتیاز-F1: نتیجه تقسیم دو برابر حاصل ضرب دقت در حساسیت بر مجموع دقت و حساسیت است که در رابطه (۱۳) بیان شده است [۳۵].

$$F_1\text{-}Score = \frac{2\times(Precision\times Sensitivity)}{Precision+Sensitivity} \qquad (۱۳)$$

برای مجموعه‌داده تسه‌لی[1] [۶]، ماتریس درهم ریختگی[2] داده‌ها در شکل (۱) نشان داده شده است. مدل پیشنهادی توانسته است به دقت ٪۹۷/۷۲ برسد. نمودارهای ROC3[3] و دقت بازیابی[4] درشکل (۲) و شکل (۳) به ترتیب نشان داده شده است.

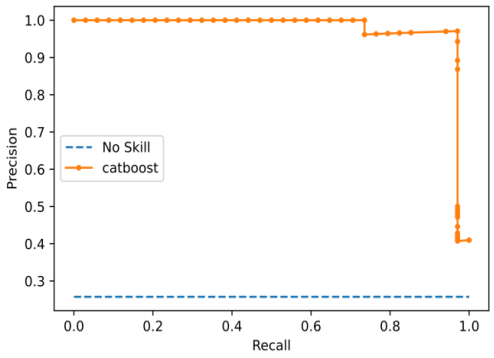


شکل ۱- ماتریس درهم‌ریختگی (مجموعه‌داده تسه‌لی)

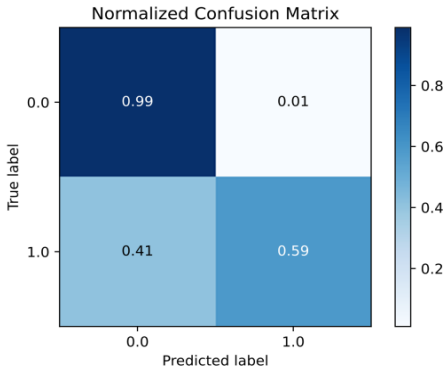


شکل ۲- نمودار ROC (مجموعه‌داده تسه‌لی)

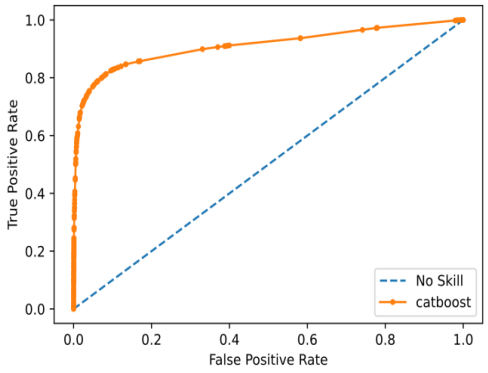
برای مجموعه‌داده زوابی، ماتریس درهم ریختگی داده‌ها در شکل (۴) نشان داده شده و نمودار ROC در شکل (۵) آمده است. در شکل (۶) نمودار AUC[5] مدل CatBoost در کنار مدل‌های LightGBM و XGBoost مشاهده می‌شود که تمامی فوق‌پارامترهای مدل‌های ذکر شده به روش آزمون و خطا به دست آمده‌اند و با توجه به نتایج فوق در این پژوهش از مدل CatBoost به عنوان طبقه‌بندی‌کننده استفاده شده است.



شکل ۳- نمودار دقت-بازخوانی (مجموعه‌داده تسه‌لی)



شکل ۴- ماتریس درهم‌ریختگی (مجموعه‌داده زوابی)

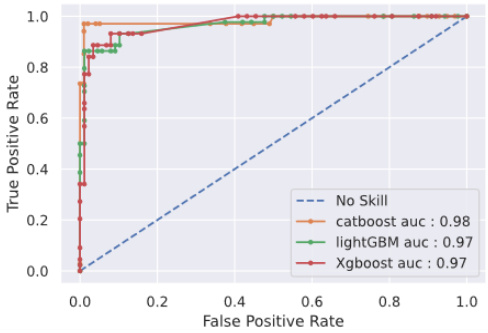


شکل۵- نمودارROC (مجموعه‌داده زوابی)

در جداول ۵ و ۶ مقایسه نتایج به دست آمده با استفاده از روش پیشنهادی با روش ارائه شده توسط تسه‌لی و

[1] Tse Li
[2] Confusion Matrix
[3] Receiver Operating Characteristic
[4] Recall-Precision
[5] Area Under Curve

همکارانش و یازید و همکارانش از نظر معیارهای حساسیت، تشخیص، صحت و امتیاز F1 نشان داده شده است.



شکل ۶- مقایسه AUC به دست آمده از روش‌های Xgboost، LightGBM ،Catboost

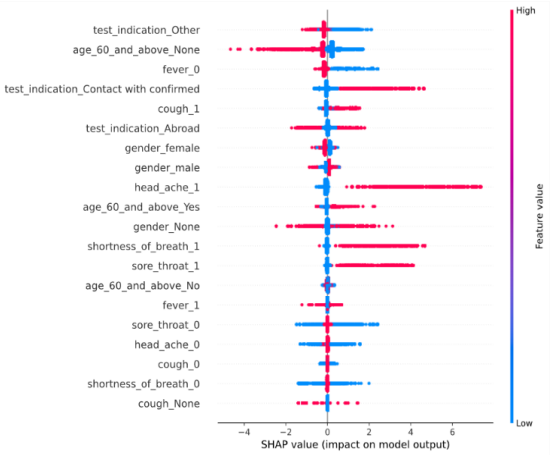
جدول ۵- نتایج مقایسه روش پیشنهادی با روش‌های مشابه (مجموعه‌داده تسه‌لی)

روش	صحت	امتیاز F1	حساسیت	تشخیص	ROC
روش پیشنهادی	۹۷/۷۲	۹۵/۵	۹۷/۹۷	۹۶/۹۶	۹۸/۳
وی تسه و همکاران	۹۵	۹۵/۱۴	۹۲/۵۰	۹۷/۹	۹۹

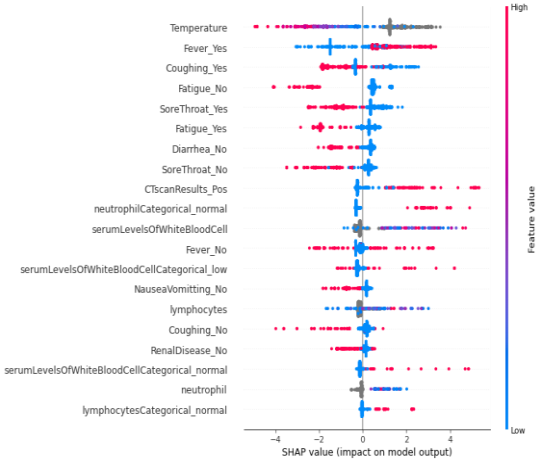
جدول ۶- نتایج مقایسه روش پیشنهادی با روش‌های مشابه (مجموعه‌داده زوابی)

روش	صحت	امتیاز F1	حساسیت	تشخیص	ROC
روش پیشنهادی	۹۶/۹	۶۸	۹۷/۶	۸۰/۷	۹۱
یازید و همکاران	۹۲/۴۲	۹۳	۸۷/۳۰	۷۱/۹۸	۹۰

برای بررسی اجمالی ویژگی‌های مهم‌تر برای یک مدل، می‌توان مقادیر SHAP هر ویژگی را برای هر نمونه ترسیم کرد. شکل (۷) ویژگی‌ها را بر اساس مجموع مقادیر SHAP در همه نمونه‌های مجموعه داده تسه‌لی و شکل (۸) در نمونه‌های مجموعه داده زوابی مرتب می‌کند و از مقادیر SHAP برای نشان دادن توزیع تاثیرات هر ویژگی بر خروجی مدل استفاده می‌کند. همانطور که در این شکل‌ها مشاهده می‌شود تاثیر ویژگی‌های دما، تب و سرفه برای مجموعه داده تسه‌لی از سایر ویژگی‌ها بیشتر است. همچنین ویژگی‌های ارتباط با افراد مبتلا، سن بالای ۶۰ سال و تب، برای مجموعه‌داده زوابی دارای بیشتر تاثیر روی خروجی مدل می‌باشند.



شکل ۷- نمودار Summary Plot برای نمایش خلاصه اطلاعات از چگونگی تاثیر ویژگی‌های برتر بر خروجی مدل (مجموعه‌داده تسه‌لی)



شکل۸- نمودار Summary Plot برای نمایش خلاصه اطلاعات از چگونگی تاثیر ویژگی‌های برتر بر خروجی مدل (مجموعه‌داده زوابی)

۶- نتیجه‌گیری و کارهای آینده

در این پژوهش، یک مدل مبتنی بر یادگیری ماشین برای تشخیص بیماری کووید-۱۹ از روی ویژگی‌های بالینی ارائه شد. در این راستا، ابتدا از پنج الگوریتم پایه یادگیری ماشین بردار پشتیبان، رگرسیون لجستیک، k نزدیکترین همسایه (k=9)، بیز ساده و جنگل تصادفی برای تشخیص بیماری

استفاده شده و سپس سه الگوریتم LightGBM، XgBoost و CatBoost پیاده‌سازی گردید که مطابق پیش‌بینی این سه الگوریتم نتایج بهتری نسبت به الگوریتم‌های پایه از خود نشان دادند. اما از میان این سه الگوریتم نیز پس از تعیین فوق پارامترها و مقایسه نتایج، سطح زیر نمودار (AUC) آنها به ترتیب برابر ۹۷٪، ۹۷٪ و ۹۸٪ به دست آمده است و در نهایت الگوریتم CatBoost به عنوان طبقه‌بندی‌کننده روی داده‌های ورودی مورد استفاده گردیده است. روش فوق موفق به دستیابی به درجه حساسیت ۹۷/۹۷٪، دقت ۹۷/۷۲٪ و صحت ۹۶/۹۶٪ شد که صحت و حساسیت آن نسبت به روش تسه‌لی و همکارانش به ترتیب ۲/۷۲٪ و ۵/۴۷٪ و نسبت به روش یازید و همکارانش ۴/۴۸٪ و ۹/۷٪ بهبود داشته است. این بهبود، ناشی از تعیین دقیقتر فوق پارامترهای الگوریتم در پیاده‌سازی انجام شده می‌باشد.

در ادامه‌ی این پژوهش، می‌توان از ترکیبی از الگوریتم‌های یادگیری ماشین و یا الگوریتم‌های نوین یادگیری عمیق برای تشخیص دقیقتر بیماری از روی علائم آن استفاده کرد.

مراجع

[1] I. Chakraborty and P. Maity, “COVID-19 outbreak: Migration, effects on society, global environment and prevention,” Sci. Total Environ., vol. 728, 2020, pp. 138882.

[۲] علی احمدیان رمکی، عباس رسولزادگان وعباس جوان‌جعفری، "تشخیص نفوذ مبتنی بر مدل‌های مخفی مارکوف: روش‌ها، کاربردها و چالش‌ها"، نشریه مدل‌سازی در مهندسی، دوره ۱۶، شماره ۵۳، تیر ۱۳۹۷، صفحه ۱۸۳- ۲۰۶.

[۳] الهام پارسایی‌مهر، مهدی فرتاش و جواد اکبری ترکستانی، "بهبود استخراج ویژگی با استفاده از یک مدل یادگیری عمیق گروهی برای تشخیص موجودیت"، نشریه مدل‌سازی در مهندسی، دوره ۲۰، شماره ۶۹، تیر ۱۴۰۱، صفحه ۱۰۳- ۱۱۲.

[۴] محمود معلم و علی‌اکبر پویان، "کشف ناهنجاری با استفاده از کدکننده خودکار مبتنی بر LSTM"، نشریه مدل‌سازی در مهندسی، دوره ۱۷، شماره ۵۶، اردیبهشت ۱۳۹۸، صفحه ۱۹۱- ۲۱۱.

[5] M. Ciotti et al., “COVID-19 Outbreak: An Overview,” Chemotherapy, vol. 64, no. 5–6, 2020, pp. 215–223.

[6] W. T. Li et al., “Using machine learning of clinical data to diagnose COVID-19: A systematic review and meta-analysis,” BMC Med. Inform. Decis. Mak., vol. 20, no. 1, Sep. 2020.

[7] Y. Zoabi, S. Deri-Rozov, and N. Shomron, “Machine learning-based prediction of COVID-19 diagnosis based on symptoms,” npj Digit. Med., vol. 4, no. 1, 2021, pp. 1–5.

[8] M. Soui, N. Mansouri, R. Alhamad, M. Kessentini, and K. Ghedira, “NSGA-II as feature selection technique and AdaBoost classifier for COVID-19 prediction using patient’s symptoms,” Nonlinear Dyn., vol. 106, no. 2, 2021, pp. 1453–1475.

[9] S. Banik, S. Banik, A. Ghosh, and A. Mukherjee, “Probabilistic estimation of COVID-19 using patient’s symptoms,” in Data Driven Approach Towards Disruptive Technologies, Springer, 2021, pp. 369–378.

[10] S. N. Nan et al., “A prediction model based on machine learning for diagnosing the early COVID-19 patients,” pp. 1–12, 2020.

[11] A. Chansik et al., “Machine learning prediction for mortality of patients diagnosed with COVID-19: a nationwide Korean cohort study,” Scientific report in nature research, 2020.

[12] C. Fang et al., “Deep learning for predicting COVID-19 malignant progression,” in Medical Image Analysis, vol. 79, 2021.

[13] A. Mariot, S. Sgoifo, and M. Sauli, “I gozzi endotoracici: contributo casistico-clinico (20 casi),” Friuli Med., vol. 19, no. 6, 1964.

[14] Y. Xu, X. Zhao, Y. Chen, and Z. Yang, “Research on a Mixed Gas Classification Algorithm Based on Extreme Random Tree,” Appl. Sci., vol. 9, no. 9, 2019, pp. 1728.

[15] W. Wang, G. Chakraborty, and B. Chakraborty, “Predicting the risk of chronic kidney disease (Ckd) using machine learning algorithm,” Appl. Sci., vol. 11, no. 1, 2021, pp. 1–17.

[16] K. Song, F. Yan, T. Ding, L. Gao, and S. Lu, “A steel property optimization model based on the XGBoost algorithm and improved PSO,” Comput. Mater. Sci., vol. 174, 2020, pp. 109472.

[17] H. Wang, C. Liu, and L. Deng, “Enhanced prediction of hot spots at protein-protein interfaces using extreme gradient boosting,” Sci. Rep., vol. 8, no. 1, 2018, pp. 1–13.

[18] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, “Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization,” Geosci. Front., vol. 12, no. 1, 2021, pp. 469–477.

[19] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2097–2106.

[20] J. Ma, Y. Ding, J. C. P. Cheng, Y. Tan, V. J. L. Gan, and J. Zhang, “Analyzing the leading causes of traffic fatalities using XGBoost and grid-based analysis: a city management perspective,” IEEE Access, vol. 7, 2019, pp. 148059–148072.

[21] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: unbiased boosting with categorical features,” Adv. Neural Inf. Process. Syst., vol. 31, 2018.

[22] A. V. Dorogush, V. Ershov, and A. Gulin, “CatBoost: gradient boosting with categorical features support,” arXiv Prepr. arXiv1810.11363, 2018.

[23] G. Ke et al., “LightGBM: A highly efficient gradient boosting decision tree,” Adv. Neural Inf. Process. Syst., vol. 2017-Decem, no. Nips, 2017, pp. 3147–3155.

[24] M. Ezzoddin, H. Nasiri, and M. Dorrigiv, “Diagnosis of COVID-19 Cases from Chest X-ray Images Using Deep Neural Network and LightGBM,” in 2022 International Conference on Machine Vision and Image Processing (MVIP), 2022, pp. 1–7.

[25] C. Chen, Q. Zhang, Q. Ma, and B. Yu, “LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion,” Chemom. Intell. Lab. Syst., vol. 191, 2019, pp. 54–64.

[26] S. Chehreh Chelgani, H. Nasiri, and A. Tohry, “Modeling of particle sizes for industrial HPGR products by a unique explainable AI tool- A ‘Conscious Lab’ development,” Adv. Powder Technol., vol. 32, no. 11, 2021, pp. 4141–4148.

[27] S. C. Chelgani, H. Nasiri, and M. Alidokht, “Interpretable modeling of metallurgical responses for an industrial coal column flotation circuit by XGBoost and SHAP-A ‘conscious-lab’ development,” Int. J. Min. Sci. Technol., vol. 31, no. 6,2021, pp. 1135–1144.

[28] A. Movsessian, D. G. Cava, and D. Tcherniak, “Interpretable machine learning in damage detection using Shapley Additive Explanations,” 2021.

[29] H. Mao et al., “Driving safety assessment for ride-hailing drivers,” Accid. Anal. \& Prev., vol. 149, 2021, pp. 105574.

[30] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 4765–4774.

[31] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, “Explainable machine learning in credit risk management,” Comput. Econ., vol. 57, no. 1, 2021, pp. 203–216.

[32] S. Mangalathu, S. H. Hwang, and J. S. Jeon, “Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach,” Eng. Struct., vol. 219, no. February,2020, pp. 110927.

[33] K. Zhou, S. Li, X. Zhou, Y. Hu, C. Zhang, and J. Liu, “Data-driven prediction and analysis method for nanoparticle transport behavior in porous media,” Measurement, vol. 172,2021, pp. 108869.

[34] S. Mangalathu, H. Shin, E. Choi, and J.-S. Jeon, “Explainable machine learning models for punching shear strength estimation of flat slabs without transverse reinforcement,” J. Build. Eng., vol. 39, 2021, pp. 102300.

[35] H. Nasiri and S. A. Alavi, “A Novel Framework Based on Deep Learning and ANOVA Feature Selection Method for Diagnosis of COVID-19 Cases from Chest X-Ray Images,” Comput. Intell. Neurosci., vol. 2022, pp. 4694567.