

Forecasting the effluent COD of refinery wastewater treatment plant using artificial intelligence

Fatemeh Barzamini¹ , Amin Bagheri^{1*} , Reza Saeedi¹ , Taha Hossein Hejazy² ,

Shookoh sadat Khaloo¹ 

1- Department of Health, Safety and Environment, Faculty of Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2- Department of Industrial and Civil Engineering, Garmsar University Branch, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran

Abstract

Background and Aims: Wastewater treatment plants are systems that can contribute to the health of the industry and the environment provided that they function properly. Mathematical and statistical simulators coupled with various models can be used to alleviate the costs of monitoring and managing wastewater treatment systems.

Materials and Methods: This is a retrospective descriptive-analytical study. Considering the complexity of biological processes along with the advances in data processing methods, algorithms (ANN) and M5 model tree were used in this study in order to make a model to estimate the effluent COD in the sanitary wastewater treatment plant of one of the country's refineries. ANN and M5 were developed through learning and testing stages based on daily data of five consecutive years (2015-2020). Various statistical indicators such as MSE and R were used to evaluate the models. Ethical considerations were observed in all stages of the study.

Results: In the ANN model with 100 hidden layers, through the training step, the amount of MSE and R respectively decreased and increased, compared to the two modes of 10 and 30 hidden layers. This is why 100 layers were chosen for this model. Also, in the M5 tree model, $R-S_{qOptimal} = 0.6147$ was calculated by selecting independent data as input, which is significant at the 0.05 level. In other words, with 95% confidence, the developed model is considered sufficiently logical for predicting COD_{out} . The results showed that the ANN model outperformed M5 model tree for predicting effluent COD with a coefficient of determination equal to 0.90 and 0.61, respectively.

Conclusion: Both models have high robustness, reliability and generalizability. Therefore, ANN and M5 model tree data mining techniques can be successfully used for environmental decision-making and estimation of lost data in wastewater treatment plants.

Keywords: Wastewater treatment plant, Modeling, Artificial neural networks, M5 model tree, Quality parameters of wastewater

Please Cite this article as: Barzamini F, Bagheri A, Saeedi R, Hejazy TH, Khaloo ShS. Forecasting the effluent COD of refinery wastewater treatment plant using artificial intelligence. Journal of Health in the Field 2023; 11(3):19-30.

Corresponding Author: Department of Health, Safety and Environment, Faculty of Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Email: bagheri.sbmu@gmail.com

DOI: <https://doi.org/10.22037/jhf.v11i3.42836>

Received: 29 July 2023

Accepted: 20 January 2024

پیش‌بینی میزان COD پساب تصفیه‌خانه فاضلاب پالایشگاه با استفاده از هوش مصنوعی

فاطمه برزمینی^۱، امین باقری^{۱*}، رضا سعیدی^۱، طه حسین حجازی^۲، شکوه السادات خالو^۱

۱- گروه سلامت، ایمنی و محیط زیست، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

۲- گروه آموزشی مهندسی صنایع و عمران، واحد دانشگاهی گرمسار، دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)، تهران، ایران

چکیده

زمینه و اهداف: تصفیه‌خانه‌های فاضلاب سیستم‌هایی هستند که در صورت عملکرد صحیح می‌توانند به سلامت صنعت و محیط زیست کمک کنند. برای کاهش هزینه‌های نظارت و مدیریت سیستم‌های تصفیه فاضلاب، می‌توان از شبیه‌سازهای ریاضی، آماری و انواع مدل‌ها استفاده نمود.

مواد و روش‌ها: مطالعه توصیفی-تحلیلی گذشته‌نگر می‌باشد. با توجه به پیچیدگی فرآیندهای بیولوژیکی و پیشرفت روش‌های پردازش داده‌گان، در این پژوهش از الگوریتم‌های (ANN) و درخت مدل M5 با هدف مدل‌سازی تخمین COD_{out} در تصفیه‌خانه فاضلاب بهداشتی یکی از پالایشگاه‌های کشور استفاده شده است. ANN و M5 از طریق مراحل یادگیری و آزمایش بر پایه داده‌های روزانه ۵ سال (۱۳۹۵-۱۳۹۹) توسعه داده شد و جهت ارزیابی مدل‌ها از شاخص‌های مختلف آماری مانند MSE و R استفاده گردید. رعایت ملاحظات اخلاقی در تمام مراحل اجرای پژوهش در نظر گرفته شد.

یافته‌ها: در مدل ANN در حالت ۱۰۰ لایه پنهان، در گام آموزش، میزان MSE و R نسبت به دو حالت ۱۰ و ۳۰ لایه پنهان به ترتیب کاهش و افزایش داشت که به همین دلیل ۱۰۰ لایه برای انجام این مدل انتخاب گشت. همچنین در مدل درخت M5 با انتخاب داده‌های مستقل به‌عنوان ورود $R-Sq_{Optimal} = 0/6147$ محاسبه گشت که در سطح $0/05$ معنی‌دار است. به عبارت دیگر با ۹۵ درصد اطمینان، مدل منطقی برای پیش‌بینی COD_{out} می‌باشد. نتایج نشان داد که مدل ANN با ضریب تعیین معادل $0/90$ ، برای COD عملکرد بهتری نسبت به درخت مدل M5 با ضریب تعیین $0/61$ دارد.

نتیجه‌گیری: هر دو مدل، استحکام، قابلیت اطمینان و قابلیت تعمیم بالایی را دارا می‌باشند، از این رو، تکنیک‌های داده‌کاوی ANN و درخت مدل M5 را می‌توان با موفقیت برای تصمیم‌گیری‌های زیست محیطی و تخمین داده‌های از دست رفته در تصفیه‌خانه‌های فاضلاب استفاده کرد.

کلید واژه‌ها: تصفیه‌خانه فاضلاب، مدل‌سازی، شبکه‌های عصبی مصنوعی، درخت مدل M5، پارامترهای کیفی فاضلاب

* نویسنده مسئول: گروه سلامت، ایمنی و محیط زیست، دانشکده بهداشت و ایمنی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

Email: bagheri.sbm@gmail.com

تاریخ دریافت مقاله: ۱۴۰۲/۰۵/۰۷

تاریخ پذیرش مقاله: ۱۴۰۲/۱۰/۳۰

مقدمه

افزایش نگرانی در مورد مسائل زیست محیطی متخصصین را تشویق کرده است که توجه خود را روی عملکرد و کنترل درست تصفیه‌خانه‌های فاضلاب متمرکز کنند [۱]. تولید سبز، یک استراتژی پیشگیرانه برای به حداقل رساندن تأثیر محصولات، بر محیط زیست است [۲]. با توجه به اینکه تعداد تصفیه‌خانه‌های فاضلاب و اهمیت آنها رو به افزایش می‌باشد، لذا پیش‌بینی اولیه و سپس تجزیه و تحلیل پارامترهایی مانند Q (دبی)، BOD5 (اکسیژن مورد نیاز بیولوژیکی)، COD (اکسیژن مورد نیاز شیمیایی)، TSS (کل جامدات معلق)، PH (اسیدیته)، TEMP (دما) ورودی و خروجی، ماه و سال براساس روشهای جدید ضروری است [۳]. عملیات و کنترل ایمن یک تصفیه‌خانه فاضلاب می‌تواند با توسعه یک ابزار مدل‌سازی برای پیش‌بینی عملکرد تصفیه‌خانه براساس مشاهدات گذشته از پارامترهای کلیدی خاص کیفیت به دست آید. تصفیه‌خانه فاضلاب شامل چندین فرآیند فیزیکی، بیولوژیکی و شیمیایی پیچیده است. اغلب این فرآیندها رفتار غیرخطی را نشان می‌دهند که توصیف آن با مدل‌های ریاضی خطی دشوار است [۴]. برای ارزیابی و شبیه‌سازی فرآیند می‌توان از روشهای ریاضی، داده‌کاوی و نرم‌افزار استفاده کرد. (Chemical Oxygen Demand: COD) معیاری برای ارزیابی فاضلاب و تایید ایمنی آن برای تخلیه به منابع دریافت کننده هست [۵]. ارزیابی‌های زیست محیطی نیاز به توسعه و تجزیه و تحلیل حجم زیادی از داده‌ها داشته و تصمیم‌گیرندگان باید مقادیر قابل توجهی از اطلاعات را فقط برای به دست آوردن مجموعه‌ای از گزینه‌های قابل دوام مدیریت کنند [۶]. برای رسیدن به این هدف می‌توان از روش‌های مختلف داده‌کاوی استفاده کرد.

در سال‌های اخیر، رویکردهای هوش مصنوعی (Artificial intelligence: AI) مانند شبکه عصبی مصنوعی (Artificial Neural Network: ANN)، سیستم تداخل عصبی- فازی تطبیق (Adaptive Neuro- Fuzzy Inference System: ANFIS)، برنامه‌ریزی بیان ژن (Gene Expression Programming: GEP)، رگرسیون چند جمله‌ای تکاملی، درخت مدل M5 و (Support Vector Machine: SVM) ماشین بردار پشتیبان برای تجزیه و تحلیل جنبه‌های مختلف زیست محیطی فرآیندهای تصفیه‌خانه فاضلاب استفاده شده است [۷]. ANN را می‌توان برای پیش‌بینی بهتر عملکرد فرآیند به دلیل دقت بالا، کفایت و کاربردهای کاملاً امیدوارکننده در مهندسی، علوم آب و زمینه‌های زیست محیطی استفاده کرد [۸]. در مدل M5، معیارهای تقسیم داده‌ها و خروجی مدل قابل مشاهده است، در حالی که در مدل ANN فقط پیش‌بینی بیان می‌شود و نحوه عملکرد آن در سیستم پنهان می‌ماند [۹]. به عبارت دیگر، این ساختار تصمیم‌گیری می‌تواند در قالب تکنیک‌های ریاضیاتی و محاسباتی نیز

معرفی شود که به توصیف، طبقه‌بندی و تعمیم مجموعه‌ای از داده‌ها کمک می‌کند [۱۰].

ون و همکاران در سال ۲۰۱۱، توانایی شبکه عصبی مصنوعی و سیستم استنتاج عصبی فازی در مدل‌سازی مشخصه‌های کیفی COD و SS روزانه تصفیه‌خانه فاضلاب کارخانه کاغذ در چین را بررسی نموده‌اند [۱۱]. جو و همکاران (۲۰۱۵) با استفاده از دو مدل ANN و SVM به پیش‌بینی نیتروژن کل پساب خروجی تصفیه‌خانه فاضلاب اولسان کره با فاصله زمانی یک روزه پرداختند [۱۲]. مانو و تالا (۲۰۱۶)، از دو مدل ANN و SVM برای ارزیابی عملکرد حذف نیتروژن کج‌دال با استفاده از داده‌های pH، TSS، COD، آمونیاک ازاد، نیتروژن آمونیاکی و نیتروژن کج‌دال ورودی به عنوان پارامترهای ورودی به مدل‌ها در تصفیه‌خانه فاضلاب کاوور واقع در منگالور استفاده کرده‌اند [۱۳]. گرانتا و همکاران (۲۰۱۷)، برای ارزیابی شاخص‌های اصلی کیفیت فاضلاب از دو مدل (Support Vector Regression: SVR) رگرسیون بردار پشتیبان و درخت رگرسیون استفاده کردند که هر دو مدل قابلیت اطمینان بالایی داشتند. رگرسیون بردار پشتیبان، در پیش‌بینی TSS، TDS و COD عملکرد بهتری نسبت به درخت رگرسیون نشان داده است. در مورد BOD₅، هر دو مدل عملکرد مشابهی را نشان دادند [۱۴]. تورکمندر و پالا (۲۰۱۷)، از تکنیک‌های شبکه عصبی مصنوعی (ANN) به ارزیابی عملکرد تصفیه‌خانه فاضلاب در ترکیه و به پیش‌بینی BOD پساب خروجی پرداختند [۳]. تحقیقات نشان داده است که در اکثر مطالعات از شبکه عصبی مصنوعی استفاده شده است و استفاده از مدل M5 در نظر گرفته نشده است. بنابراین در تحقیق حاضر از شبکه عصبی مصنوعی که یکی از قدیمی‌ترین روش‌های داده‌کاوی است با مدل M5 که روشی جدید است برای شبیه‌سازی COD و ارزیابی کارایی ANN و M5 استفاده گشته است. هدف اصلی پژوهش حاضر "توسعه، مقایسه و ارزیابی عملکرد ANN و M5 برای ارزیابی عملکرد تصفیه فاضلاب یکی از پالایشگاه‌های کشور و برآورد کیفیت پساب خروجی از آن می‌باشد.

مواد و روش‌ها

منطقه مطالعه

تصفیه‌خانه فاضلاب پالایشگاه مورد نظر که در جنوب ایران واقع می‌باشد، یکی از بزرگ‌ترین پالایشگاه‌های خاورمیانه با ظرفیت پالایشی بیش از ۳۰۰ هزار بشکه در روز است که در حال حاضر و بعد از اجرای طرح افزایش ظرفیت ظرفیت تولید روزانه خود را به ۳۲۰ هزار بشکه در روز رسانیده است. در این سایت روش تصفیه فاضلاب بهداشتی، روش لجن فعال بوده و پساب حاصل از تصفیه‌خانه به دریا ریخته می‌شود.

جمع‌آوری داده‌ها

در این پژوهش ابتدا اطلاعات مربوط به سیستم‌های پایش ورودی و خروجی تصفیه‌خانه یکی از پالایشگاه‌های کشور، نظیر BOD_5 ، COD، TSS، PH، TEMP، Q فاضلاب بین سال‌های ۱۳۹۵ تا ۱۳۹۹ استخراج شد.

تحلیل‌های آماری اولیه

تحلیل‌های آماری اولیه برای داده‌کاوی ضروری است، در این پژوهش بعثت کمی بودن داده‌های مدل M5 نیاز به تبدیل داده‌ها نبوده و فقط پالایش و پردازش اولیه بر روی داده‌ها صورت گرفته است، در حالیکه در مدل ANN، دو نوع داده سال و ماه از نوع کیفی می‌باشند، لذا برای ورود به مدل بایستی کمی گردند و سپس در مدل‌های M5 و ANN مورد پالایش و پردازش قرار گرفت. بر اساس نتایج تحلیل رگرسیون انجام‌شده در طول مدلسازی COD_{out} مشخص شد که پارامترهای COD_{in} ، BOD_{out} ، TSS_{out} ، $YEAR$ ، $MONTH$ به‌صورت مجزا و $COD_{in} * TSS_{out}$ ، $BOD_{out} * YEAR$ ، $BOD_{out} * MONTH$ و $TSS_{out} * YEAR$ ، $TSS_{in} * MONTH$ ، $COD_{in} * MONTH$ ، $TSS_{out} * MONTH$ ، $YEAR * MONTH$ به‌صورت ترکیبی بر COD_{out} مؤثر بوده‌اند؛ سپس وارد مدل گردید.

شبکه عصبی مصنوعی و درخت مدل M5

اولین مرحله پس از تبدیل نمودن نتایج آنالیز ۵ ساله به دیتاشیت، پاک‌سازی داده‌های پرت و مفقود بوده سپس داده‌ها به‌منظور آماده‌سازی جهت ورود کردنشان به الگوریتم‌های داده‌کاوی مورد پالایش و پردازش اولیه قرار گرفت.

در مرحله نهایی، الگوریتم‌های ANN که مدل محاسباتی مبتنی بر شبکه‌های عصبی بیولوژیکی و ابزاری جهت مدل‌سازی غیرخطی است و قادر به مدیریت تعداد زیادی ورودی می‌باشد و مدل درختی M5 که تکنیک یادگیری آماری است و فضای مسئله را به شاخه‌های متعدد

تقسیم می‌کند تا برای هر شاخه یک مدل رگرسیون با استفاده از پارامتر انحراف معیار، برازش کند، به عنوان دو الگوریتم انتخابی به منظور یافتن بهترین مدل جهت پیش‌بینی غلظت COD_{out} در خروجی تصفیه‌خانه فاضلاب پالایشگاه مورد بررسی و مقایسه قرار گرفت؛ در این راستا ۷۰ درصد داده‌ها جهت یادگیری ماشین و ۳۰ درصد مابقی به‌منظور اعتبارسنجی و تست آن از طریق پکیج‌های مربوطه در نرم‌افزارهای MATLAB12 و Minitab16 مورد استفاده قرار گرفت. درنهایت با ملاحظه نتایج برازش مدل‌های فوق‌الذکر و مقایسه شاخص‌های عملکردی نظیر R^2 و R^2_{adj} و MSE و RMSE و سایر پارامترهای مربوطه به انتخاب بهترین مدل که مدل ANN با $R=0/9061$ می‌باشد پرداخته شد.

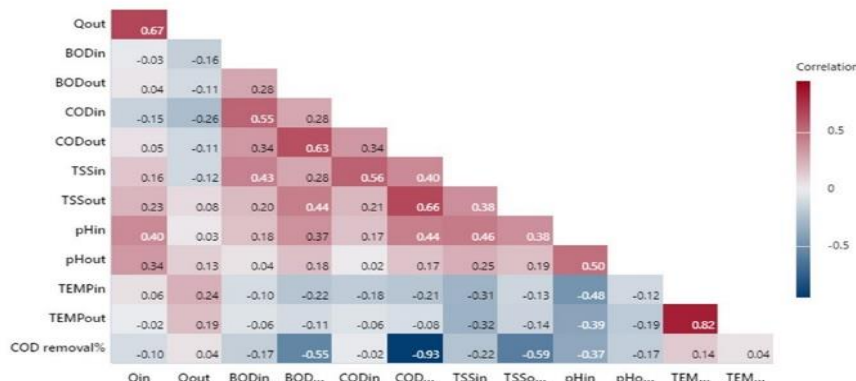
ملاحظات اخلاقی طرح

- کسب مجوز برای استفاده از داده‌ها و انتشار نتایج از پالایشگاه‌های کشور
- تعهد به محرمانه باقی ماندن اطلاعات تصفیه‌خانه فاضلاب
- رعایت اخلاق و بی‌طرفی در انتشار نتایج پژوهش
- رفرنس دهی صحیح در استفاده از منابع علمی و مطالعات پیشین

یافته‌ها

بررسی همبستگی بین داده‌ها

مقدار این ضریب بین ۱- تا ۱ تغییر می‌کند که (۱) به معنای همبستگی مثبت کامل، (۰) به معنی نبود همبستگی، و (۱-) به معنی همبستگی منفی کامل است. کاربرد اصلی ضریب پیرسون زمانی است که متغیرها از نوع پارامتری باشند؛ بدین معنا که توزیع نرمال داشته باشند و در سطح فاصله ای نسبی باشند.



شکل ۱- همبستگی بین متغیرهای Qin، Qout، BODin، BODout، CODin، CODout، TSSin، TSSout، pHin، pHout، TEMPin، TEMPout، با درصد حذف COD در سال ۱۳۹۵

Figure 1- Correlation between variables of Qin, Qout, BODin, BODout, CODin, CODout, TSSin, TSSout, pHin, pHout, TEMPin, TEMPout, with COD removal percentage in 2015

این کار از مجوز Creative Commons Attribution 4.0 International License (CC BY 4.0) تبعیت می‌کند.

($0.5/$ و $TEMP_{out} (-0.8/)$ می‌باشد، همچنین بیشترین و کمترین همبستگی بین متغیرهای مستقل به ترتیب مربوط به ارتباط $TEMP_{out}$ با $TEMP_{in} (0.82/)$ ، و COD_{in} با PH_{out} و $COD_{removal} (0.2/ \pm)$ می‌باشد.

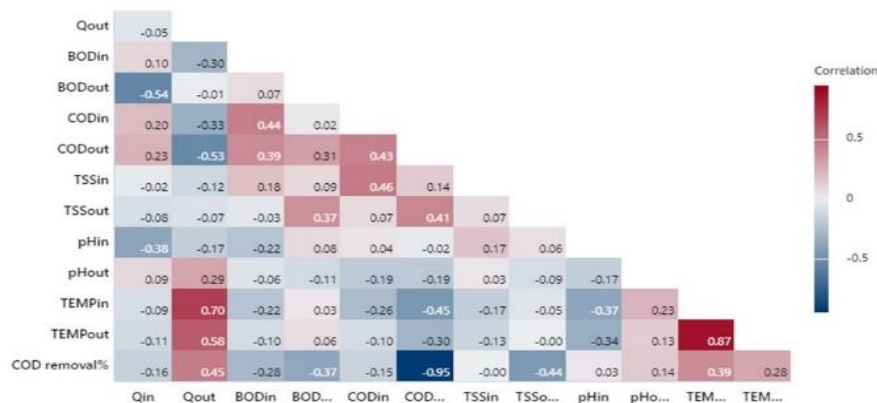


Figure 2- Correlation between variables of Qin, Qout, BODin, BODout, CODin, CODout, TSSin, TSSout, pHin, pHout, TEMPin, TEMPout, with COD removal percentage in 2016

($-0/02$) و TSSin ($0/14$) می‌باشد، همچنین بیشترین و کمترین همبستگی بین متغیرهای مستقل به ترتیب مربوط به ارتباط TEMPout با TEMPin ($0/87$) و TSSin با CODremoval و TSSout با TEMPout ($0/0$) می‌باشد.

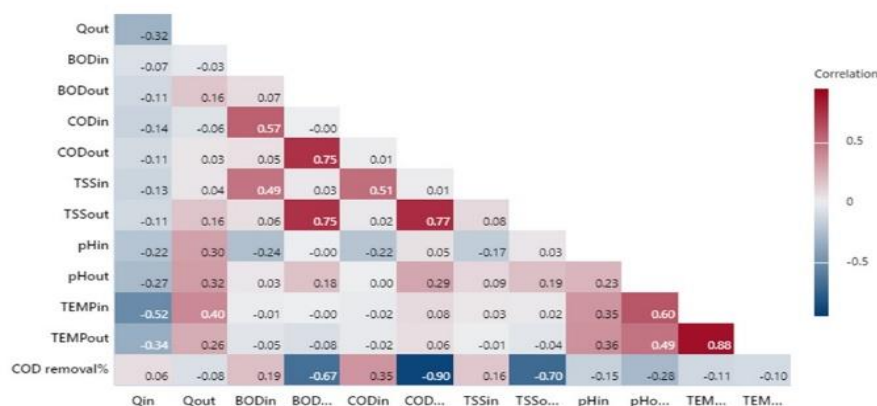
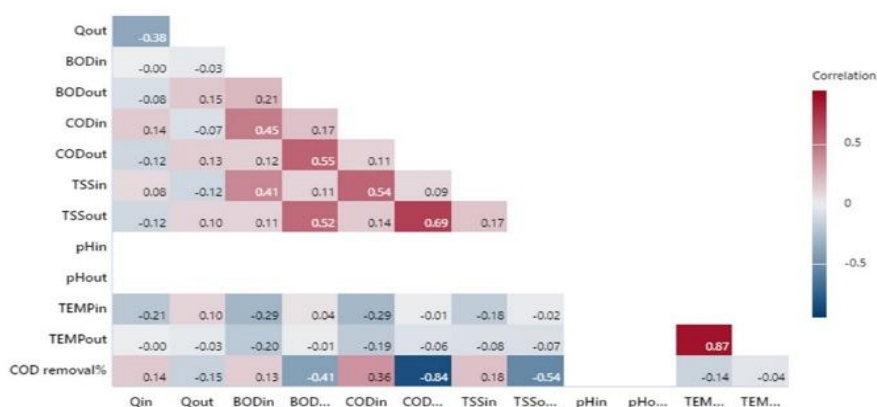


Figure 3- Correlation between variables of Qin, Qout, BODin, BODout, CODin, CODout, TSSin, TSSout, pHin, pHout, TEMPin, TEMPOut, with COD removal percentage in 2017

همبستگی CODout به ترتیب با CODin و TSSin (۰/۰۱) بوده و همچنین بیشترین و کمترین همبستگی، بین متغیرهای مستقل به ترتیب

مربوط به ارتباط TEMPout با TEMPin (۰/۸۸) و CODin با BODout و CODin با PHout (۰/۰۰) می‌باشد.

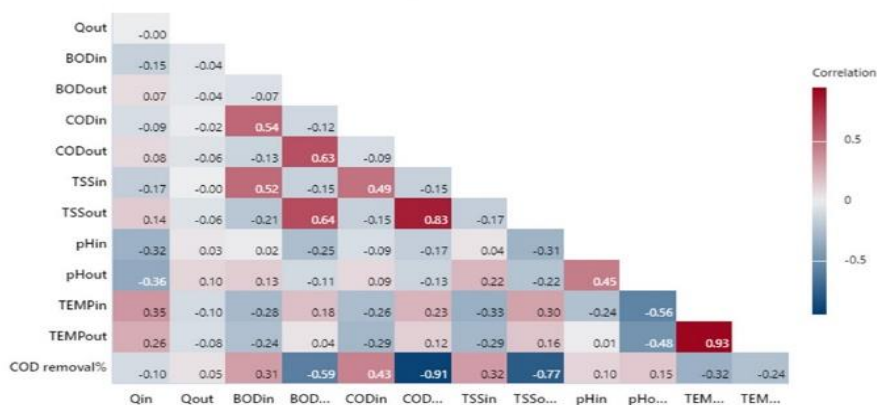


شکل ۴- همبستگی بین متغیرهای Qin، Qout، BODin، BODout، CODin، CODout، TSSin، TSSout، pHin، pHout، TEMPin، TEMPout، با COD حذف در سال ۱۳۹۸

Figure 4- Correlation between variables of Qin, Qout, BODin, BODout, CODin, CODout, TSSin, TSSout, pHin, pHout, TEMPin, TEMPout, with COD removal percentage in 2018

متغیرهای مستقل به ترتیب مربوط به ارتباط TEMPout با TEMPin (۰/۸۷) و BODin با Qout (۰/۰۰) می‌باشد. (لازم به ذکر است ردیفهای خالی در شکل بالا مبین این می‌باشد که داده‌های آن سال اندازه‌گیری نشده است.)

مطابق شکل ۴، بیشترین همبستگی مثبت CODout به ترتیب با TSSout (۰/۶۹) و پارامترهای BODout (۰/۵۵) و بیشترین همبستگی منفی CODout با CODremoval (۰/۸۴) بوده و کمترین همبستگی CODout به ترتیب با TEMPin (۰/۰۱) و TEMPout (۰/۰۶) می‌باشد، و همچنین بیشترین و کمترین همبستگی بین

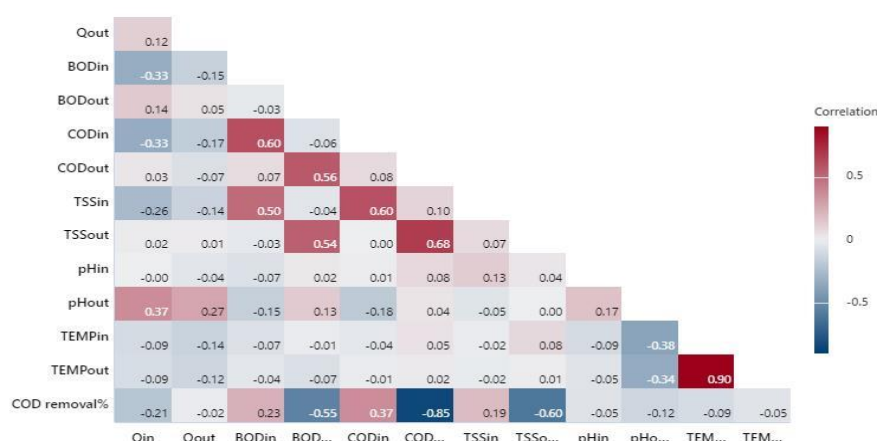


شکل ۵- همبستگی بین متغیرهای Qin، Qout، BODin، BODout، CODin، CODout، TSSin، TSSout، pHin، pHout، TEMPin، TEMPout، با COD حذف در سال ۱۳۹۹ برای ورود به مدل M5

Figure 5- Correlation between variables of Qin, Qout, BODin, BODout, CODin, CODout, TSSin, TSSout, pHin, pHout, TEMPin, TEMPout, with COD removal percentage in 2019 to enter the M5 model

CODout به ترتیب با Qout (۰/۰۶) و Qin (۰/۰۸)، همچنین بیشترین و کمترین همبستگی بین متغیرهای مستقل به ترتیب مربوط به ارتباط TEMPout با TEMPin (۰/۹۳) و Qout با Qin (۰/۰۰) می‌باشند.

مطابق شکل ۵، بیشترین همبستگی مثبت CODout به ترتیب با TSSout (۰/۸۳) و پارامترهای BODout (۰/۶۳) و بیشترین همبستگی منفی CODout با CODremoval (۰/۹۱) بوده و کمترین همبستگی بین



شکل ۶- همبستگی بین متغیرهای Qin, Qout, BODin, BODout, CODin, CODout, TSSin, TSSout, pHin, pHout, TEMPin, TEMPout, با درصد حذف COD برای ورود به مدل ANN

Figure 6- Correlation between variables of Qin, Qout, BODin, BODout, CODin, CODout, TSSin, TSSout, pHin, pHout, TEMPin, TEMPout, with COD removal percentage to enter the ANN model

تعیین پیش‌بینی شده حدود ۷۰٪ تخمین زده شده است. در شکل زیر پراکندگی مقادیر باقیمانده یا همان اختلاف میزان واقعی با پیش‌بینی شده، نشان داده شده است. مطابق نمودارها میزان تراکم مقادیر باقیمانده بین ۳۰ و منفی ۳۰ درصد مشاهده شد.

نتایج مدل ANN

نتایج تحلیل رگرسیون نشان داد که Qin, Qout, BODin, BODout, CODin, CODout, TSSin, TSSout, pHin, pHout, TEMPin, TEMPout, YEAR, MONTH مؤثرترین پارامترها بر روی CODout تصفیه‌خانه فاضلاب بودند و دو فاکتور month و year در این مدل کیفی بوده و برای ورود به نرم‌افزار کمی شده‌اند، لذا تعداد پارامترهای مستقل ۲۸ عدد می‌باشند. پس از مرحله آموزش با ۷۰ درصد داده، مدل با ۲۰ درصد داده‌ها برای معماری CODout اعتبارسنجی و با ۱۰ درصد تست انجام شد. مجموعه داده‌های مورد استفاده در این مدل ۱۸۲۶ داده می‌باشد که ۱۲۷۸ داده برای گام آموزش، ۳۶۵ داده برای گام اعتبارسنجی و ۱۸۳ داده برای گام تست تخصیص داده شده است. تعداد لایه‌های پنهان به صورت پیش‌فرض در نرم‌افزار مورد استفاده ۱۰ بوده که به منظور دستیابی به حالت بهینه، آزمون و خطا انجام شد که در این پژوهش به ترتیب ۳۰ و ۱۰۰ لایه به نرم‌افزار تزریق گشت که در حالت ۱۰۰ لایه، خرجی نرم‌افزار بهینه گردید. همان‌طور که در جدول مشاهده می‌شود در حالت وجود ۱۰۰ لایه پنهان و در گام آموزش، میزان MSE و R نسبت به دو حالت ۱۰ و ۳۰ لایه پنهان به ترتیب کاهش و افزایش داشتند که به همین دلیل ۱۰۰ لایه برای انجام این مدل انتخاب شد، همچنین این نتایج حاکی از آن است که مدل ANN می‌تواند مدل‌سازی موفق و شبیه‌سازی و پیش‌بینی عملکرد تصفیه‌خانه فاضلاب را ارائه دهد.

مطابق شکل ۶، بیشترین همبستگی مثبت CODout به ترتیب با TSSout (۰/۶۸) و پارامترهای BODout (۰/۵۶) و بیشترین همبستگی منفی CODout با CODremoval (۰/۸۵) بوده و کمترین همبستگی CODout به ترتیب با TEMPout (۰/۰۲) و Qin (۰/۰۳) و همچنین بیشترین و کمترین همبستگی بین متغیرهای مستقل به ترتیب مربوط به ارتباط TEMPout با TEMPin (۰/۹۰) و TSSout با PHout و CODin (۰/۰۰) می‌باشند.

تحلیل استنباطی داده‌ها

متغیرها و فاکتورهای مستقل Qin, Qout, BODin, BODout, CODin, TSSin, TSSout, pHin, pHout, TEMPin, TEMPout, YEAR, MONTH تأثیرشان بر روی متغیر وابسته CODout مورد تجزیه و تحلیل قرار گرفته و نتایج رگرسیون حاکی از آن است که متغیرها با $P < 0.05$ معنی‌دار بوده و به عبارتی با درصد اطمینان ۰/۹۵ میتوان نتیجه گرفت که پارامترهای مورد نظر بر روی متغیر وابسته مؤثر می‌باشند. در جدول ۱ تحلیل واریانس کلیه متغیرهای دارای اثر بر روی متغیر وابسته بیان گشته که برخی از پارامترها به صورت مجزا دارای اثر ناچیزی بر روی متغیر وابسته بوده اند لذا با ترکیب یک یا دو پارامتر دیگر، دارای اثر قابل توجهی بر روی پارامتر وابسته شده‌اند. با توجه به نتایج حاصل شده جدول ۱، می‌توان نتیجه گرفت پارامترهای BODout, CODin, TSSout, YEAR, MONTH به صورت مجزا و BODout*MONTH, BODout*YEAR, CODin*TSSout, TSS, TSSout*YEAR, TSSin*MONTH, CODin*MONTH, YEAR*MONTH, out*MONTH به صورت ترکیبی دارای ارتباط معناداری با CODout می‌باشند. جدول ۲ ضریب تعیین تعدیل شده و ضریب تعیین پیش‌بینی شده را نشان می‌دهد. مطابق جدول ۲ که ضریب

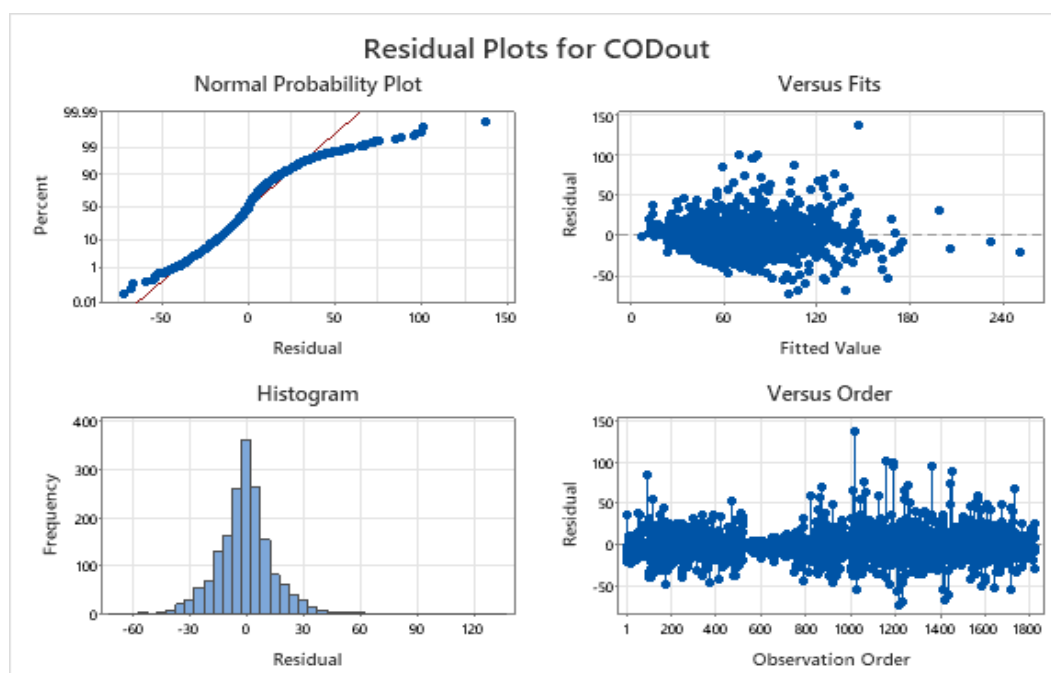
جدول ۱- نتایج تحلیل واریانس متغیرهای تاثیرگذار بر متغیر پاسخ (COD_{out})Table 1- The result of ANOVA for variables affecting on response variable (COD_{out})

P-value	F-value	Adj MS	Adj SS	DF	منابع
۰/۰۰۰	۲۸۵/۱۶	۹۲۴۲۶/۳	۹۲۴۲۶	۱	BODout
۰/۰۴۷	۳/۹۵	۱۲۷۹/۰	۱۲۷۹	۱	CODin
۰/۱۹۶	۱/۶۷	۵۴۱/۵	۵۴۲	۱	TSSin
۰/۰۰۰	۹۰/۶۷	۲۹۳۸۷/۵	۲۹۳۸۸	۱	TSSout
۰/۰۰۰	۱۳/۴۴	۴۳۵۶/۱	۱۷۴۲۴	۴	YEAR
۰/۰۰۱	۲/۹۱	۹۴۲/۰	۱۰۳۶۲	۱۱	MONTH
۰/۰۲۵	۵/۰۳	۱۶۳۱/۰	۱۶۳۱	۱	CODin*TSSout
۰/۰۰۰	۱۹/۲۸	۶۲۴۸/۲	۲۴۹۹۳	۴	BODout*YEAR
۰/۰۰۰	۸/۰۱	۲۵۹۵/۴	۲۸۵۴۹	۱۱	BODout*MONTH
۰/۰۰۱	۲/۹۶	۹۵۸/۹	۱۰۵۴۷	۱۱	CODin*MONTH
۰/۰۰۵	۲/۴۶	۷۶۹/۴	۷۸۶۰	۱۱	TSSin*MONTH
۰/۰۰۰	۳۱/۴۵	۱۰۱۹۲/۵	۴۰۷۷۰	۴	TSSout*YEAR
۰/۰۰۰	۶/۶۲	۲۱۴۵/۷	۲۳۶۰۳	۱۱	TSSout*MONTH
۰/۰۰۰	۱۲/۷۳	۴۱۲۵/۶	۱۸۱۵۲۷	۴۴	YEAR*MONTH
		۳۲۴/۱	۵۵۳۵۹۶	۱۷۰۸	Error
*	*	۳۲۴/۹	۵۵۳۵۹۶	۱۷۰۴	Lack-of-Fit
		۰/۰	۰	۴	Pure Error
			۲۳۰۲۵۵۴	۱۸۲۴	جمع کل

جدول ۲- ضرایب تعیین مدل رگرسیون ارائه شده جهت پیش بینی متغیر پاسخ

Table 2- Coefficients of determination for predictive regression model

واریانس	ضریب تعیین	ضریب تعیین تعدیل شده	ضریب تعیین پیش بینی شده
۱۸/۰۳۳	٪ ۷۵/۹۷	٪ ۷۴/۳۴	٪ ۷۰/۴۹

شکل ۷- نمودار پراکندگی رگرسیون برای COD_{out} Figure 7- Regression scatter plot for COD_{out}

این کار از مجوز Creative Commons Attribution 4.0 International License (CC BY 4.0) تبعیت می کند.

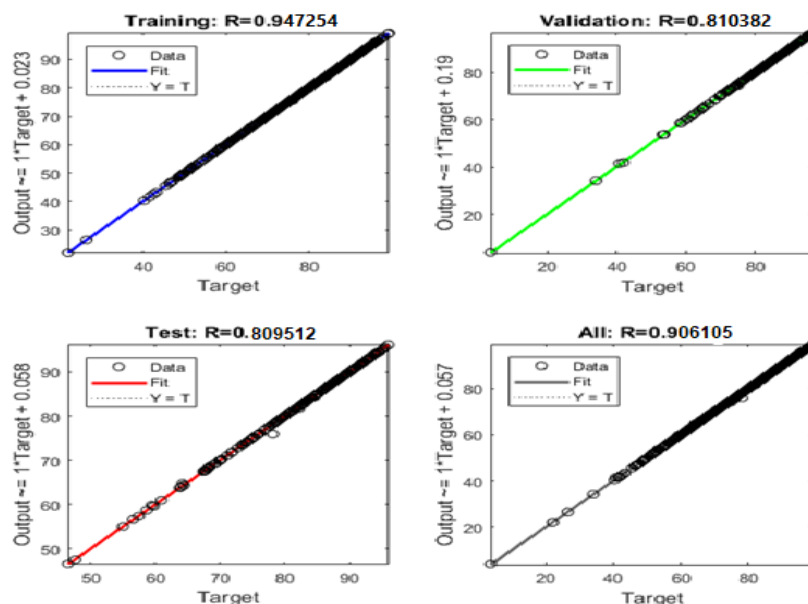
جدول ۳- معیارهای آماری مربوط به توسعه مدل شبکه عصبی مصنوعی

Table 3- Statistical indicator for ANN development

معیارهای آماری با ۱۰ لایه پنهان			
نمونه	میانگین مربعات خطا	ضریب تعیین	
آموزش	۰/۱۸۰۶۹۰	۰/۹۰۸۵۰۵	
اعتبارسنجی	۰/۲۹۴۰۴۷	۰/۸۲۶۱۸۸	
تست	۰/۲۶۵۷۵۳	۰/۸۵۲۳۷۶	
معیارهای آماری با ۳۰ لایه پنهان			
نمونه	میانگین مربعات خطا	ضریب تعیین	
آموزش	۰/۱۷۳۱۵۸	۰/۹۰۹۳۳۰	
اعتبارسنجی	۰/۳۰۵۶۷۴	۰/۸۳۷۰۰۳	
تست	۰/۲۹۸۵۱۵	۰/۸۶۳۶۷۰	
معیارهای آماری با ۱۰۰ لایه پنهان			
نمونه	میانگین مربعات خطا	ضریب تعیین	
آموزش	۰/۱۰۳۷۹۳	۰/۹۴۷۲۵۴	
اعتبارسنجی	۰/۳۶۵۸۴۳	۰/۸۱۰۳۸۲	
تست	۰/۳۶۵۵۵۴	۰/۸۰۹۵۱۲	

شکل زیر نمودارهای خط رگرسیون مربوط به داده‌های یادگیری، تست و اعتبارسنجی مدل ANN بسط داده شده را نشان می‌دهند، همچنین

ضریب تعیین‌های حاصل از مدل در مراحل مختلف نشان داده شده است.



شکل ۸- نتایج رگرسیون در مراحل مختلف مدل ANN

Figure 8- Result of regression in different stages of ANN model

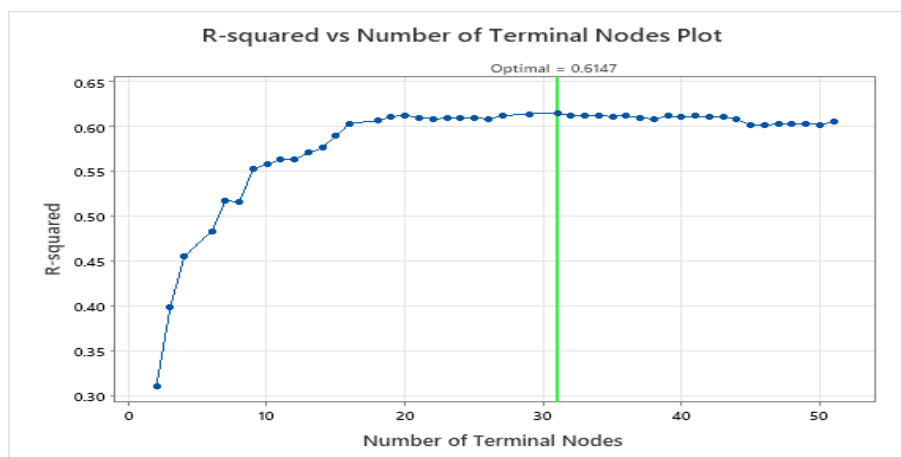
نتایج مدل درخت M5

تمامی داده‌های ورودی و خروجی فاضلاب، به‌عنوان متغیر مستقل به نرم‌افزار وارد شدند، هر داده‌ای به نوعی بر CODout تاثیرگذار می‌باشد و با حذف هر کدام از این پارامترها از دقت مدل کاسته می‌شود همان‌طور

که از TSS برای پیش‌بینی کارایی تصفیه‌خانه استفاده می‌شود حذف بخشی از آن بر روی CODout نیز می‌تواند تاثیرگذار باشد. این مدل در بهترین حالت خود ۳۱ گره تشکیل داده لذا در این مدل مجموعه آموزشی ۱۰ برابر مجموعه تست انتخاب شده، بدین معنا که ۱۸۰ داده

آموزش در مقابل ۱ داده تست قرار می‌گیرد و این کار دهها بار اتفاق می‌افتد تا بهترین داده در دو مجموعه انتخاب گردد.

در مدل درخت M5 با انتخاب داده‌های مستقل به‌عنوان ورودی $R_{-}Sq_{optimal} = 0.6147$ محاسبه شده که در سطح ۰/۰۵ معنی‌دار



شکل ۹- نمودار تعیین تعداد گره بهینه در مدل M5

Figure 9- Determination optimal number of nodes for M5 tree model

در جدول ۴ آمار توصیفی داده‌ها بر اساس درخت M5 آورده شده است.

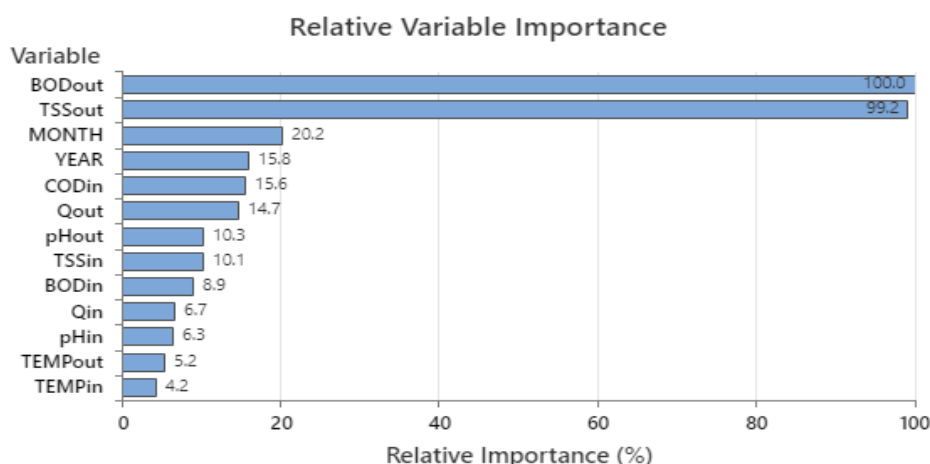
جدول ۴- نتایج آمار توصیفی داده‌ها بر اساس درخت مدل M5

Table 4- Descriptive statistics based on M5 tree model

میان	انحراف از معیار	حداقل	چارک اول	میان	چارک سوم	حداکثر
۸۲/۳۷۹۰	۱۰/۱۰۴۰	۳/۶۶۴۹۲	۷۸/۶۲۹۰	۸۴/۲۷۴۶	۸۹/۴۳۵۰	۹۹/۱۸۰۳

شکل ۱۰ اهمیت نسبی متغیرها را نسبت به BODout نشان می‌دهد، همان‌طور که در شکل مشاهده می‌شود با توجه به همبستگی و ارتباط معنادار BODout با CODout، این متغیر به‌عنوان مؤثرترین پارامتر

تعیین شده و اهمیت سایر متغیرها نسبت به این متغیر تعیین شده‌اند. از بین متغیرها، TSSout و TEMPIn به ترتیب با ۹۹/۲ و ۴/۲، بیشترین و کمترین اهمیت را در تعیین CODout داشتند.



Variable importance measures model improvement when splits are made on a predictor. Relative importance is defined as % improvement with respect to the top predictor.

شکل ۱۰- اهمیت نسبی متغیرهای تأثیرگذار در پیش‌بینی متغیر پاسخ

Figure 10- Effective variable on to predictive the response variables

این کار از مجوز Creative Commons Attribution 4.0 International License (CC BY 4.0) تبعیت می‌کند.

۵۱۰ داده و $Mean=82/76$ تشکیل می‌دهند، لذا گره ۳۳ و گره ۴۷ بیشترین اختلاف میانگین را با نقطه برش TSSout در این طبقه دارند. مدل M5 مجموعه داده اولیه را با استفاده از انحراف استاندارد (SD) اعضای کلاس تقسیم می‌کند و سعی می‌کند این خطا را با آزمایش تمام ویژگی‌های آن گره کاهش دهد و کل پیش‌بینی‌کننده‌ها (متغیرهای تاثیرگذار) در این مدل ۱۳ عدد می‌باشند. در مدل M5 با ۳۱ گره بهینه و ضریب تعیین ۶۱ درصد این کار انجام و خروجی مدل حاصل می‌شود. پس از ساخته شدن درخت مدل M5، بر اساس نمونه‌های هر گره تصمیم‌گیری و گره برگ، یک رگرسیون خطی برای آن گره ایجاد می‌کند، سپس برگ‌های درختی که مقادیر هدف را با دقت پایین پیش‌بینی می‌کنند قطع می‌شوند و شکل نهایی مدل به دست می‌آید.

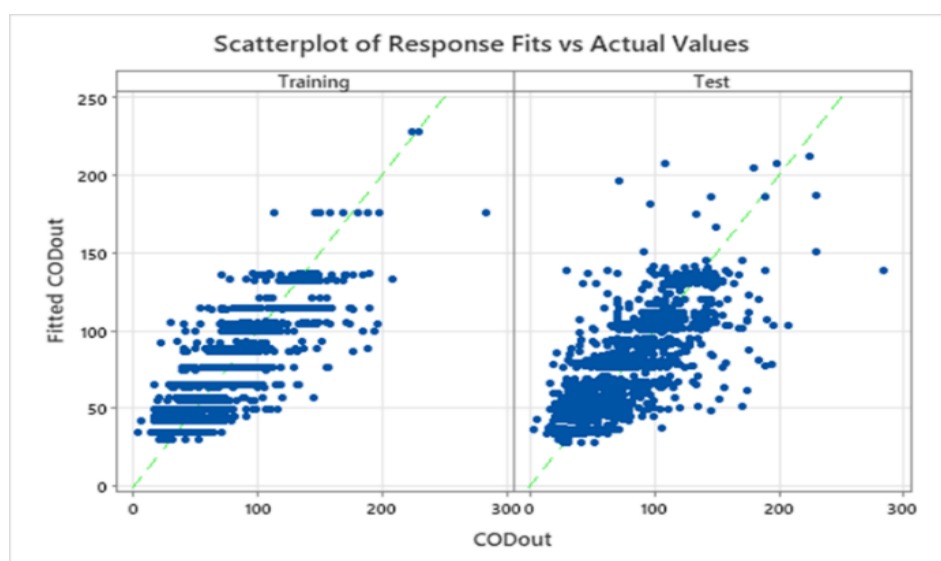
در درخت مدل M5 ریشه درخت (گره ۱) با BODout (به علت بیشترین تاثیر بر CODout) شروع می‌شود. تاثیر BODout اینطور مبین می‌شود که وقتی مقدار BODout کمتر یا مساوی $79/9$ باشد ۱۱۶۲ داده را با $Mean=87/6285$ در یک گره جدید (گره ۲) و زمانی که مقدار BODout بیشتر از $79/9$ باشد ۶۶۳ داده را با $Mean=73/1706$ در گره جدید دیگری (گره ۳۲) طبقه بندی می‌کند. نقطه برش BODout، دو گره با بیشترین اختلاف میانگین را به وجود می‌آورد و هیچ متغیر تاثیر گذار دیگری نمی‌تواند دو گره با بیشترین اختلاف را در این سطح ایجاد کند. حال در گره ۳۲ با ۶۶۳ داده دارای $Mean=73/1706$ و $StDev=10/309$ TSSout به عنوان نقطه برش تعیین می‌شود. TSSout با مقدار کمتر مساوی $354/5$ ، گره ۳۳ را با ۱۵۳ داده دارای $Mean=60/9915$ ، و TSSout با مقدار بیشتر از $354/5$ ، گره ۴۷ را با

جدول ۵- معیارهای آماری مدل توسعه یافته درخت M5
Table 5- Statistical criteria for development M5 tree model

آماره‌ها	آموزش	تست
ضریب تعیین	۰/۷۲۳۰	۰/۶۱۴۷
ریشه میانگین مربعات خطا (RMSE)	۱۸/۶۹۳۸	۲۲/۰۴۹۲
میانگین مربعات خطا (MSE)	۳۴۹/۴۵۷۷	۴۸۶/۱۶۸۱
میانگین انحراف مطلق (MAD)	۱۳/۰۱۶۸	۱۴/۸۸۸۷
میانگین درصد مطلق خطا (MAPE)	۰/۲۳۶۳	۰/۲۶۴۹

و تست به ترتیب ۰/۷۲۳۰ و ۰/۶۱۴۷، مدلی کارآمد برای بررسی کیفیت پساب پالایشگاه می‌باشد.

در جدول ۵ معیارهای آماری ضریب تعیین، ریشه میانگین مربعات خطا و سایر معیارهای آماری در دو مرحله آموزش و تست آورده شده است. نتایج حاکی از آن است که مدل با ۳۱ گره برگ با ضریب تعیین آموزشی



شکل ۱۱- نمودار پراکندگی پاسخ مربوط به مقادیر واقعی در مراحل آموزش و آزمایش

Figure 11- Scatter plot of the response corresponding to the actual values in the training and testing phases

استفاده از الگوریتم‌های هوش مصنوعی مختلف و مجموعه داده‌های مختلف ممکن است به نتایج متفاوتی منجر شود.

ANN

در این پژوهش $R = 0.9061$ برای پیش بینی COD_{out} ارائه گردید. ضریب تعیین بالا و مقدار خطای پایین در مراحل آموزش، اعتبارسنجی و تست در مدل با لایه پنهان ۱۰۰ این امکان را به ما می‌دهد که از توابع بیشتری در لایه پنهان استفاده کرده تا به مقدار بهینه دست یابیم. ضریب تعیین بالا در گام آموزش مبین این است که چند درصد تغییرات متغیر وابسته در یک مدل رگرسیونی با متغیرهای مستقل تبیین می‌شود که این خود نشان‌دهنده انتخاب صحیح پارامترهاست. البته ممکن است با افزودن پارامترهای دیگر مقدار ضریب بهینه افزایش پیدا کند که این، موجود بودن و در دسترس بودن داده را شامل می‌شود. در شکل ۴-۸ رگرسیون در مراحل مختلف با ضریب تعیین بالا منطبق بودن داده‌ها با رگرسیون حاصل را نشان می‌دهد، که داده‌های انتخابی حتی در صورتیکه دارای اثر ناچیز باشند می‌توانند با ترکیب با پارامترهای دیگر اثر افزایشی خود را بر متغیر وابسته وارد سازند. مقایسه ضریب تعیین رگرسیون ساده با ضریب تعیین مدل ANN به ما نشان می‌دهد که در دنیای پیشرفته امروز با کثرت زیادی داده‌ها و تنگی زمان استفاده از یادگیری ماشین در حوزه‌های مختلف امکان صرفه جویی در زمان و هزینه را به ما می‌دهد.

درخت مدل M5

در شکل (۴-۹) دیاگرام درخت بهینه M5 با دادن تعداد گره‌های مختلف به نرم‌افزار و گرفتن ضریب تعیین، تعیین می‌شود که در چه تعداد گره ای، ضریب تعیین بهینه حاصل می‌شود که این خود یکی از برتریهای درخت تصمیم می‌باشد. ضریب تعیین بهینه در مدل 0.61 به دست آمده است. که عدد به نسبت کمی است. یکی از معایب این مدل تکرار در بررسی ویژگیها و اعتبارسنجی مدل می‌باشد که این خود باعث می‌شود ضریب تعیین عدد پایین تری به دست آید. این مدل نسبت به رگرسیون ساده ضریب تعیین پایین تری دارد. اما این به معنای برتری رگرسیون ساده نسبت به درخت تصمیم نیست. در درخت تصمیم با بررسی RMSE میزان خطای مدل محاسبه شده بسیار کوچک است که این دقت مدل درخت تصمیم را به نظر می‌رساند.

نتیجه گیری

با وجود تمام چالش‌های داده کاوی، استفاده از روش‌های مختلف تجزیه و تحلیل داده‌ها اجتناب ناپذیر است. با توجه به نتایج درخت مدل ANN و M5 مشخص شد که این روش‌های ذکر شده بر اساس معیارهای آماری توانایی قابل قبولی برای شبیه‌سازی COD_{out} دارند. به‌طور کلی می‌توان نتیجه گرفت که برای استفاده از روش‌های داده کاوی مانند شبکه‌های عصبی مصنوعی و درخت تصمیم، تعداد داده‌های مورد

در شکل ۱۱ پراکندگی داده‌ها حول محور رگرسیون مشاهده می‌شود که با انجام درخت M5 و برازش لازم برای نتایج به دست آمده مشخص گردید مدل فوق برای پیش‌بینی کیفیت پساب موفق می‌باشد.

مقایسه نتایج مدل شبکه عصبی مصنوعی ANN با مدل درخت مدل M5

از مقایسه دو مدل فوق نتیجه می‌گیریم مدل ANN نسبت به مدل M5 بدلیل ضریب تعیین بالاتر RMSE و MSE پایینتر مدلی با عملکرد بهتر برای پیش‌بینی کیفیت پساب تصفیه‌خانه پالایشگاه می‌باشد. همچنین نتایج تحقیق نشان‌دهنده دقت مناسبتر مدل ANN نسبت به M5 به علت دارا بودن ضریب تعیین بالاتر می‌باشد. در پژوهشی مشخص شد که مدل ANN نسبت به M5 می‌تواند مدل‌سازی موفق و شبیه‌سازی و پیش‌بینی عملکرد تصفیه‌خانه فاضلاب را بهتر ارائه دهد [۸]. مقایسه بین نتایج [۸] و [۱۵] و این تحقیق نشان می‌دهد که دقت مدل‌سازی در این تحقیق نسبت به دو مطالعه دیگر بالاست، که می‌توان آن را به نوع معماری کاربردی و تعداد داده‌های آموزش و اعتبارسنجی مرتبط دانست. با توجه به نتایج درخت مدل M5 و ANN مشخص شد که روش‌های ذکر شده بر اساس معیارهای آماری توانایی قابل قبولی برای شبیه‌سازی COD_{out} دارند. برای استفاده از روش‌های داده کاوی مانند شبکه‌های عصبی مصنوعی و درخت تصمیم، تعداد داده‌های مورد استفاده و پردازش اولیه حائز اهمیت است. در پایان، داده کاوی در نظر گرفته شده مانند درخت مدل M5 و ANN که نتایج قابل قبولی را در تفسیر روابط بین متغیرها و پیش‌بینی کیفیت پساب تصفیه ارائه داده‌اند، می‌توانند برای ارائه تخمینی از مقداری که برای اندازه‌گیری واحدهای تصفیه در نظر گرفته می‌شود در غیاب اقدامات مستقیم توصیه شوند. این روش‌ها می‌توانند برای تصمیم‌گیری‌های زیست محیطی و تخمین داده‌های گمشده در تصفیه‌خانه‌های فاضلاب برای افزایش کارایی سیستم در آینده مفید باشند. نتیجه می‌گیریم که انتخاب یک مدل طبقه‌بندی که برای انواع مختلف و پیچیدگی مجموعه داده‌ها مناسب باشد، مهم است.

بحث

بر اساس نتایج تحلیل رگرسیون انجام شده در طول مدل‌سازی COD_{out} مشخص شد که پارامترهای $YEAR$ ، TSS_{out} ، COD_{in} ، BOD_{out} MONTH به صورت مجزا $COD_{in} * TSS_{out}$ ، $BOD_{out} * YEAR$ ، $TSS_{in} * MONTH$ ، $COD_{in} * MONTH$ ، $BOD_{out} * MONTH$ ، $TSS_{out} * YEAR$ ، $TSS_{out} * MONTH$ ، $YEAR * MONTH$ به صورت ترکیبی بر COD_{out} مؤثر بوده‌اند. نتایج ترکیب ورودیهای مختلف در پژوهشهای مختلف نشان داد که تنها در نظر گرفتن متغیرهایی با بالاترین ضریب همبستگی برای مدل‌سازی دقیق کافی نیست. بنابراین، برای افزایش عملکرد مدل، لازم است متغیرهایی با ضریب همبستگی پایین نیز در نظر گرفته شوند. همچنین، پیش‌بینی با

می‌توان مقادیر خروجی را به دست آورد. علاوه بر این، می‌توان برای کنترل اندازه‌گیری پارامترهای خروجی استفاده کرد. اگر تفاوت زیادی بین مقادیر شبیه‌سازی شده و اندازه‌گیری شده وجود دارد، سیستم باید بررسی شود.

تشکر و قدردانی

نویسندگان این مقاله مراتب تشکر و قدردانی خود را از کلیه کسانی که در انجام این مطالعه یاری نمودند، اعلام می‌دارند. این مقاله حاصل از پایان‌نامه مصوب در دانشگاه علوم پزشکی و خدمات بهداشتی درمانی شهید بهشتی به کد اخلاق IR.UMA.REC.1400.005 می‌باشد.

تضاد منافع

در این مطالعه نویسندگان هیچ تضاد منافی ندارند.

References

- 1- Yel E, Yalpir S. Prediction of primary treatment effluent parameters by fuzzy inference system (FIS) approach. *Procedia Computer Science* 2011; 3659-65.
- 2- Fresner J. Cleaner production as a means for effective environmental management. *Journal of Cleaner Production* 1998; 6(3-4):171-79.
- 3- Türkmenler H, Murat P. Performance assessment of advanced biological wastewater treatment plants using artificial neural networks. *International Journal of Engineering Technologies* 2017; 3(3):151-56.
- 4- Mjalli FS, Al-Asheh S, Alfadala HE. Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. *Journal of Environmental Management* 2007; 83(3):329-38.
- 5- Ward J, Ward P, Saint Ch, Mantzioris E. The urban agriculture revolution. *Water: Journal of the Australian Water Association* 2014; 41(1):69-74.
- 6- Ramírez Y, Kraslawski A, Cisternas LA. Decision-support framework for the environmental assessment of water treatment systems. *Journal of Cleaner Production* 2019; 225:599-609. Doi: 10.1016/j.jclepro.2019.03.319.
- 7- Najafzadeh M, Zeinolabedini M. Prognostication of waste water treatment plant performance using efficient soft computing models: An environmental evaluation. *Measurement* 2019; 138:690-701. Doi: 10.1016/j.measurement.2019.02.014.
- 8- Hamada M, Adel Zaqoot H, Abu Jreiban A. Application of artificial neural networks for the prediction of Gaza wastewater treatment plant performance-Gaza strip. *Journal of Applied Research in Water and Wastewater* 2018; 5(1):399-406.

استفاده و پردازش اولیه حائز اهمیت است. از نتایج این پژوهش می‌توان برای شبیه‌سازی با دقت خوبی استفاده کرد. نتایج تحقیق حاضر نشان داد که مدل ANN نسبت به M5 با توجه به شاخص‌های آماری مانند ضریب تعیین از دقت مناسبتری برخوردار است. در پژوهش حاضر از ۱۲۷۸ داده برای مرحله آموزش و ۳۶۵ داده برای مرحله اعتبار سنجی و ۱۸۳ داده برای تست استفاده گردید. که به نوبه خود دقت این مطالعه را می‌رساند. نتایج درخت مدل ANN و M5 در معماری‌های مختلف بر اساس شاخص‌های آماری هر دو در محدوده قابل قبولی قرار گرفتند و می‌توان نتیجه گرفت که هر دو مدل توانسته اند مدل‌های خوبی برای شبیه‌سازی روزانه COD_{out} از تصفیه‌خانه فاضلاب پالایشگاه ارائه دهند. در کنار ساختار مدل، دقت داده‌ها، طول و ساختار ممکن است تأثیر مؤثری بر عملکرد مدل داشته باشد. در این پژوهش ANN با $R=0.9061$ برای پیش‌بینی COD_{out} پیشنهاد می‌گردد. در صورت مفقود شدن داده‌های خروجی با داشتن ورودی و انجام شبیه‌سازی

- 9- Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* 2018; 4(11):e00938. Doi: 10.1016/j.heliyon.2018.e00938.
- 10- Hand D, Mannila H, Smyth P. Principles of data mining: a comprehensive highly technical look at the math and science behind extracting useful information from large databases. Massachusetts: Massachusetts Institute of Technology Press; 2001.
- 11- Wan J, Huang M, Ma Y, Guo W, Wang Y, Zhang H, et al. Prediction of effluent quality of a paper mill wastewater treatment using an adaptive network-based fuzzy inference system. *Applied Soft Computing* 2011; 11(3):3238-46.
- 12- Guo H, Jeong K, Lim J, Jo J, Kim YM, Park JP, et al. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *Journal of Environmental Sciences* 2015; 32:90-101.
- 13- Manu DS, Thalla AK. Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl Nitrogen from wastewater. *Applied Water Science* 2017; 7:3783-91. Doi: 10.1007/s13201-017-0526-4.
- 14- Granata F, Papirio S, Esposito G, Gargano R., De Marinis G. Machine learning algorithms for the forecasting of wastewater quality indicators. *Water* 2017; 9(2):105. Doi: 10.3390/w9020105.
- 15- Zare Abyaneh H. Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *Journal of Environmental Health Science and Engineering* 2014; 12:1-8. Doi: 10.1186/2052-336X-12-40.