

چارچوب یادگیری ماشین تفسیرپذیر جهت پیش‌بینی خواص الکتروشیمیایی مواد

الکترودی بر پایه کربن در باتری‌های لیتیوم-یون

سید فردین تقی زاده^{۱*}، الهام زارع^۱

^۱گروه فیزیک، دانشکده علوم پایه، دانشگاه یاسوج، یاسوج، ایران

چکیده

اطلاعات مقاله

در این پژوهش، از روش‌های یادگیری ماشین مبتنی بر توصیف‌گرهای شیمی کوانتومی حاصل از محاسبات نظریه تابعی چگالی برای پیش‌بینی پتانسیل ردوکس مولکول‌های آلی فعال ردوکس استفاده شد. توصیف‌گرهای الکترونی استخراج‌شده از محاسبات نظریه تابعی چگالی به‌عنوان ورودی مدل‌های مختلف یادگیری ماشین به کار گرفته شدند و عملکرد آن‌ها با استفاده از شاخص‌های R^2 ، MAE و RMSE مورد ارزیابی قرار گرفت. نتایج نشان داد که مدل SVR با دستیابی به ضریب تعیین 0.974 و کمترین میزان خطا، بالاترین دقت پیش‌بینی را ارائه می‌دهد، در حالی که مدل Stacking نیز عملکردی بسیار نزدیک و پایدار از خود نشان داد. تحلیل هم‌بستگی، اهمیت ویژگی‌ها و مقادیر SHAP نشان داد که توصیف‌گرهای مرتبط با انتقال الکترون، به‌ویژه الکترون‌خواهی، بیشترین نقش را در پیش‌بینی پتانسیل ردوکس ایفا می‌کنند. هم‌خوانی نتایج یادگیری ماشین با مفاهیم شیمی کوانتومی بیانگر آن است که چارچوب پیشنهادی، ضمن دستیابی به دقت پیش‌بینی بالا، از قابلیت تفسیر مناسبی نیز برخوردار است. این مطالعه نشان می‌دهد که تلفیق محاسبات و یادگیری ماشین می‌تواند رویکردی کارآمد برای غربالگری سریع و طراحی هدفمند مولکول‌های آلی فعال ردوکس با خواص الکتروشیمیایی مطلوب فراهم آورد.

تاریخ دریافت: ۱۴۰۵/۰۵/۰۷

تاریخ پذیرش: ۱۴۰۵/۰۶/۰۴

تاریخ چاپ: ۱۴۰۵/۰۶/۱۰

شاپای چاپی: 2588-493x

شاپای الکترونیکی: 2588-4921

* نویسنده مسئول

Taghizadeh@yu.ac.ir

واژگان کلیدی: نظریه تابعی چگالی، یادگیری ماشین، پتانسیل ردوکس، مولکول‌های آلی

مقدمه

به توسعه سامانه‌های ذخیره‌سازی انرژی با کارایی بالا وابسته کرده است [۲]. در میان فناوری‌های مختلف ذخیره‌سازی انرژی، باتری‌ها به دلیل چگالی انرژی بالا، بازده مناسب و قابلیت استفاده در طیف گسترده‌ای از کاربردها، جایگاه ویژه‌ای یافته‌اند. در این میان، باتری‌های لیتیوم-یون به دلیل چگالی انرژی زیاد، ولتاژ کاری بالا و عمر چرخه‌ای طولانی، به یکی از موفق‌ترین فناوری‌های ذخیره‌سازی انرژی برای

امروزه گذار به سمت منابع انرژی پاک و تجدیدپذیر، به‌ویژه انرژی باد و خورشید، به‌منظور کاهش وابستگی به سوخت‌های فسیلی و مقابله با چالش‌های زیست‌محیطی، به یکی از اولویت‌های اصلی جهان تبدیل شده است [۱]. با این حال، ماهیت متناوب و وابستگی این منابع به شرایط محیطی، بهره‌برداری پایدار از آن‌ها را

تجهیزات الکترونی قابل حمل، خودروهای الکتریکی و سامانه‌های ذخیره‌سازی شبکه تبدیل شده‌اند [۳].

با وجود پیشرفت‌های چشمگیر در توسعه باتری‌های لیتیوم-یون، بیشتر سامانه‌های تجاری همچنان از کاتدهای مبتنی بر اکسید فلزات واسطه استفاده می‌کنند که با محدودیت‌های متعددی همراه هستند. یکی از مهم‌ترین این محدودیت‌ها، کندی سینتیک نفوذ یون‌های لیتیوم در ساختار بلوری این مواد است که در نرخ‌های شارژ و دشارژ بالا، موجب افت عملکرد و کاهش ظرفیت توان می‌شود [۴]. علاوه بر این، هزینه نسبتاً بالا، سمیت بالقوه و محدودیت منابع برخی فلزات واسطه، توسعه مواد کاتدی پایدار، کم‌هزینه و سازگارتر با محیط‌زیست را به ضرورتی اجتناب‌ناپذیر تبدیل کرده است [۵]. در سال‌های اخیر، مواد آلی فعال ردوکس^۱، که قادر به انجام واکنش‌های برگشت‌پذیر اکسایش-کاهش هستند، به‌عنوان جایگزین‌هایی امیدبخش برای کاتدهای نسل آینده مورد توجه قرار گرفته‌اند. در میان این مواد، مشتقات کینونی کربن‌پایه به دلیل جرم مولکولی پایین، انعطاف‌پذیری ساختاری، هزینه تولید کمتر، دسترسی به منابع تجدیدپذیر و سازگاری بیشتر با محیط‌زیست، جایگاه ویژه‌ای پیدا کرده‌اند [۶].

با وجود مزایای قابل توجه مواد آلی فعال ردوکس، تنوع بسیار زیاد ساختارهای مولکولی در این خانواده، شناسایی ترکیباتی با پتانسیل ردوکس مناسب، ظرفیت ویژه بالا و پایداری مطلوب را به چالشی اساسی تبدیل کرده است. از این رو، طراحی و ارزیابی این ترکیبات بر پایه روش‌های سنتی آزمون و خطا، فرآیندی زمان‌بر، پرهزینه و در مقیاس وسیع عملاً ناکارآمد است. در این راستا، محاسبات مکانیک کوانتومی مبتنی بر نظریه تابعی چگالی^۲ به ابزاری قدرتمند برای پیش‌بینی خواص الکتروشیمیایی مولکول‌ها، از جمله پتانسیل ردوکس، انرژی‌های بالاترین اوربیتال مولکولی اشغال‌شده^۳، پایین‌ترین اوربیتال مولکولی اشغال‌نشده^۴ و میل

الکترونی^۵، تبدیل شده‌اند [۷]. با این حال، هزینه محاسباتی بالای DFT، استفاده مستقیم از این روش را برای غربالگری پرفرپیت^۶ پایگاه‌های داده بزرگ مولکولی با محدودیت جدی مواجه می‌سازد [۸].

به‌منظور غلبه بر این محدودیت، ترکیب داده‌های حاصل از محاسبات کوانتومی با الگوریتم‌های یادگیری ماشین، به‌عنوان رویکردی نوین در طراحی و کشف مواد پیشرفته مطرح شده است. این الگوریتم‌ها قادرند روابط پیچیده و غیرخطی میان ویژگی‌های ساختاری و الکترونی مولکول‌ها و خواص هدف را استخراج کرده و در بسیاری از موارد، پیش‌بینی‌هایی با دقت قابل قبول و در زمانی به‌مراتب کوتاه‌تر از محاسبات کوانتومی ارائه دهند [۹]. با این حال، بیشتر مطالعات پیشین بر توسعه مدل‌های منفرد، به‌ویژه شبکه‌های عصبی مصنوعی، متمرکز بوده‌اند و ارزیابی جامع عملکرد الگوریتم‌های مختلف و بهره‌گیری از روش‌های یادگیری جمعی پیشرفته برای افزایش دقت، قابلیت تعمیم و پایداری مدل‌ها، همچنان کمتر مورد توجه قرار گرفته است.

در یک مطالعه مروری، روند طراحی نظری مواد آلی فعال ردوکس برای باتری‌های قابل شارژ را بررسی کردند و نشان دادند که این مواد، به دلیل تنوع ساختاری، قابلیت تنظیم خواص الکتروشیمیایی و بهره‌مندی از عناصر فراوانی مانند C, H, N, O و S، گزینه‌های امیدوارکننده‌ای برای توسعه باتری‌های لیتیوم-یون هستند. آنان بیان کردند که بیشتر مطالعات پیشین بر محاسبات DFT برای پیش‌بینی خواصی مانند پایداری شیمیایی، پایداری الکتروشیمیایی و برگشت‌پذیری واکنش‌های ردوکس متکی بوده‌اند؛ با این حال، هزینه محاسباتی بالای DFT، کاربرد آن را در غربالگری پرفرپیت محدود می‌کند. از این رو، این پژوهشگران یادگیری ماشین را به‌عنوان رویکردی مکمل و آینده‌دار برای تسریع طراحی و بهینه‌سازی مواد آلی فعال

¹ Organic Materials Redox-active

² Density Functional Theory, DFT

³ Highest Occupied Molecular Orbital, HOMO

⁴ Lowest Unoccupied Molecular Orbital, LUMO

⁵ Electron Affinity (EA)

⁶ High-Throughput Screening

۱- روش تحقیق

این مجموعه داده شامل چند خانواده از مواد الکترونی مولکولی آلی و بر پایه کربن، از جمله مشتقات گرافن، کوئینون، آروماتیک‌های آلیش شده با بور و ترکیبات مبتنی بر فنالنیل و آنتراسن است که برای ایجاد یک فضای طراحی شیمیایی متنوع انتخاب شده‌اند [۱۲]. این تنوع ساختاری امکان بررسی اثر تغییرات در گروه‌های عاملی، الگوهای جانشینی و آلیش را بر خواص الکترونیکی و پتانسیل ردوکس فراهم می‌کند. در این پژوهش، از یک رویکرد داده‌محور مبتنی بر یادگیری ماشین برای پیش‌بینی پتانسیل ردوکس محاسبه‌شده با DFT استفاده شد. مجموعه داده شامل ۱۰۶ نمونه بود و متغیر هدف، پتانسیل ردوکس نسبت به Li/Li^+ بود که در مقاله مرجع از طریق چرخه ترمودینامیکی محاسبه شده است [۱۲]. ویژگی‌های ورودی شامل EA، HOMO، LUMO، اختلاف بین ترازهای HOMO و LUMO، پتانسیل شیمیایی (μ)، سختی شیمیایی (η)، شاخص الکتروفیلی (ω)، مجذور پتانسیل شیمیایی (χ^2) و مجذور سختی شیمیایی (η^2) بودند. در این میان، HOMO و LUMO به ترتیب بیانگر توانایی مولکول در واگذاری و پذیرش الکترون هستند و EA نیز تمایل مولکول به پذیرش الکترون را نشان می‌دهد. همچنین، اختلاف انرژی HOMO-LUMO با پایداری و واکنش‌پذیری الکترونی مرتبط است، در حالی که پتانسیل شیمیایی، سختی شیمیایی و شاخص الکتروفیلی به ترتیب گرایش مولکول به تبادل الکترون، مقاومت آن در برابر تغییر تعداد الکترون‌ها و تمایل آن به پذیرش بار الکترونی را توصیف می‌کنند.

۱-۱ پیش‌پردازش و آماده‌سازی داده‌ها

در مرحله آماده‌سازی داده‌ها، مقادیر غیرعددی، بی‌نهایت و داده‌های ناقص شناسایی و حذف شدند. سپس ویژگی‌های ورودی و متغیر هدف به قالب عددی مناسب برای مدل‌سازی

ردوکس معرفی کردند و بر ضرورت ادغام روش‌های کوانتومی با رویکردهای داده‌محور تأکید نمودند [۱۰].

در مطالعه ای دیگر، کاربرد غربالگری مجازی پرطرفیت را در طراحی مشتقات کینونی برای باتری‌های جریان ردوکس آبی-آلی بررسی کردند. آنان نشان دادند که اگرچه کینون‌ها از امیدبخش‌ترین مواد آلی فعال ردوکس به شمار می‌روند، عملکرد آن‌ها از نظر ظرفیت ویژه، چگالی توان و پایداری بلندمدت هنوز با الزامات تجاری‌سازی فاصله دارد. این مطالعه بر بررسی فضای شیمیایی مشتقات کینونی، روش‌های برآورد توصیف‌گرهای عملکردی و روابط ساختار-ویژگی متمرکز بود و در نهایت، بر ضرورت توسعه روش‌های دقیق‌تر و کارآمدتر برای غربالگری و شناسایی مولکول‌های کینونی با عملکرد بالا تأکید کرد [۱۱].

در این پژوهش، چارچوبی جامع مبتنی بر یادگیری ماشین برای پیش‌بینی دقیق پتانسیل ردوکس مواد الکترونی آلی ارائه شده است. در این چارچوب، عملکرد مجموعه‌ای از الگوریتم‌های یادگیری ماشین شامل مدل‌های درخت‌پایه، هسته‌محور و خطی به‌صورت نظام‌مند ارزیابی شده و در ادامه، یک مدل یادگیری جمعی^۱ به‌منظور بهره‌گیری هم‌زمان از نقاط قوت مدل‌های برتر توسعه یافته است. افزون بر این، برخلاف بسیاری از مطالعات پیشین که صرفاً بر بهبود دقت پیش‌بینی متمرکز بوده‌اند، در این پژوهش، در این پژوهش، از روش تفسیر افزایشی مبتنی بر مقادیر شیلی^۲ برای تبیین نقش و میزان تأثیر هر یک از توصیف‌گرهای مولکولی و الکترونی بر پتانسیل ردوکس استفاده شده است این رویکرد، علاوه بر افزایش دقت و قابلیت اطمینان پیش‌بینی‌ها، بینش ارزشمندی درباره عوامل مؤثر بر رفتار الکتروشیمیایی مواد فراهم می‌کند و می‌تواند مسیر طراحی منطقی مواد آلی جدید با عملکرد بالا برای باتری‌های نسل آینده را هموار سازد.

¹ Stacking Ensemble

² SHapley Additive exPlanations (SHAP)

آموزش داده شدند، زیرا این دسته از الگوریتم‌ها معمولاً نسبت به مقیاس ویژگی‌ها حساسیت کمتری دارند.

۳-۱ انتخاب و طبقه‌بندی الگوریتم‌های یادگیری ماشین

برای مدل‌سازی، چندین الگوریتم یادگیری ماشین مورد استفاده قرار گرفت تا بتوان عملکرد آن‌ها را در پیش‌بینی پتانسیل ردوکس نسبت با یکدیگر مقایسه کرد. این مدل‌ها شامل رگرسیون خطی به‌عنوان یک مدل پایه خطی، SVR با کرنل شعاعی برای مدل‌سازی روابط غیرخطی، RF به‌عنوان یک روش مبتنی بر درخت، XGBoost و LightGBM به‌عنوان دو الگوریتم تقویتی قدرتمند، و همچنین یک مدل ترکیبی Stacking بودند. در مدل Stacking، خروجی چند مدل پایه شامل XGBoost، RF، SVR، LightGBM و رگرسیون خطی با یک متامدل ریج^۶ ترکیب شد تا از مزیت مکمل بودن الگوهای یادگیری مختلف استفاده شود. استفاده همزمان از مدل‌های خطی، غیرخطی، درختی و ترکیبی این امکان را فراهم می‌کند که رفتار پیش‌بینی در سطوح مختلف پیچیدگی بررسی شده و مناسب‌ترین مدل برای این دیتاست شناسایی شود. در این فرآیند، تنظیم ابرپارامترها به‌صورت دستی و بر اساس مقادیر از پیش تعریف‌شده انجام شد. به‌طور مشخص، در مدل‌های XGBoost، RF و LightGBM تعداد تخمین‌زن‌ها برابر ۲۰۰ در نظر گرفته شد. در مدل SVR نیز پارامترهای $C=10$ و $\epsilon=0.1$ انتخاب شدند.

۴-۱ معیارهای ارزیابی عملکرد مدل

برای ارزیابی عملکرد مدل‌ها، از چند شاخص مکمل استفاده شد تا دقت پیش‌بینی از زوایای مختلف سنجیده شود. این معیارها شامل ضریب تعیین R^2 ، ریشه میانگین مربعات خطا^۷ و میانگین قدر مطلق خطا^۸ بودند. استفاده همزمان از

تبدیل شده و ردیف‌های دارای مقادیر گمشده یا نامعتبر از فرایند تحلیل کنار گذاشته شدند. پس از پاک‌سازی داده‌ها، مجموعه داده به دو بخش آموزش و آزمون تقسیم شد. سپس روش 5-fold فقط بر روی مجموعه آموزش مورد استفاده قرار گرفت تا فرآیند توسعه و ارزیابی داخلی مدل‌ها انجام شود. مجموعه آزمون مستقل در این مرحله مورد استفاده قرار نگرفت و برای ارزیابی نهایی عملکرد مدل‌ها حفظ شد. همچنین مقدار random_state = 42 برای اطمینان از بازتولیدپذیری فرآیند تقسیم داده‌ها و اعتبارسنجی در نظر گرفته شده است.

در فرآیند اعتبارسنجی متقاطع 5-fold، استانداردسازی به‌صورت مستقل در هر Fold انجام شد؛ به‌گونه‌ای که پارامترهای استانداردسازی تنها از بخش آموزش هر Fold محاسبه و سپس بر بخش اعتبارسنجی همان Fold اعمال شدند. پس از اتمام فرآیند آموزش، پارامترهای استانداردسازی حاصل از کل مجموعه آموزش برای تبدیل مجموعه آزمون مستقل مورد استفاده قرار گرفتند.

۱-۲ راهبرد نرمال‌سازی و مقیاس‌گذاری

در ادامه، به‌منظور بهبود عملکرد مدل‌هایی که به مقیاس ویژگی‌ها حساس هستند، فرایند استانداردسازی داده‌ها انجام شد، به‌گونه‌ای که پارامترهای مقیاس‌دهی تنها بر اساس داده‌های آموزشی محاسبه شدند و سپس همان تبدیل روی داده‌های آزمون اعمال شد. این راهبرد از نشت اطلاعات از مجموعه آزمون به آموزش جلوگیری می‌کند و از نظر آماری یک رویه صحیح در توسعه مدل‌های پیش‌بینی است. لازم به ذکر است که این استانداردسازی برای مدل‌هایی مانند رگرسیون بردار پشتیبان^۱، رگرسیون خطی^۲، و مدل Stacking اعمال شد، در حالی که مدل‌های مبتنی بر درخت تصمیم مانند جنگل تصادفی^۳، الگوریتم تقویت گرادینانی^۴ و الگوریتم تقویت گرادینانی سبک^۵ مستقیماً بر روی داده‌های خام

^۱ Support Vector Regression (SVR)

^۲ Linear Regression

^۳ Random Forest (RF)

^۴ Extreme Gradient Boosting (XGBoost)

^۵ Light Gradient Boosting Machine (LightGBM)

^۶ Ridge

^۷ Root Mean Square Error (RMSE)

^۸ Mean Absolute Error (MAE)



شکل ۱: چارچوب کلی تحلیل داده و مدل‌سازی برای پیش‌بینی پتانسیل ردوکس.

۲- یافته‌ها

۲-۱ تحلیل همبستگی ویژگی‌ها و متغیر هدف

در این بخش، به‌منظور بررسی روابط خطی میان متغیرها، ماتریس همبستگی پیرسون (شکل ۲) از دو منظر تحلیل شد: نخست، رابطه میان توصیفگرهای محاسباتی و متغیر هدف (پتانسیل ردوکس) و دوم، همبستگی میان خود توصیفگرها. نتایج نشان می‌دهد که پتانسیل ردوکس قوی‌ترین همبستگی را با EA دارد، به‌طوری‌که ضریب همبستگی پیرسون برابر با ۰.۹۱۸- است. این رابطه معکوس

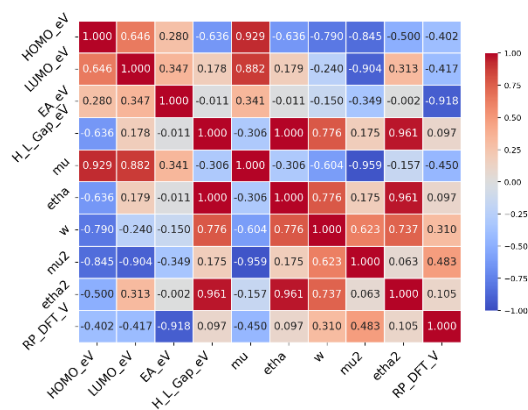
این معیارها ضروری است، زیرا هر یک جنبه متفاوتی از کیفیت پیش‌بینی را نشان می‌دهد.

۵-۱ تحلیل تفسیرپذیری و اهمیت ویژگی‌ها

علاوه بر ارزیابی کمی، در این پژوهش تحلیل تفسیرپذیری مدل‌ها نیز مورد توجه قرار گرفت تا مشخص شود کدام ویژگی‌ها بیشترین سهم را در پیش‌بینی پتانسیل ردوکس داشته‌اند. برای مدل‌های درختی مانند RF، XGBoost و LightGBM از اهمیت ویژگی داخلی مدل^۱ استفاده شد. در مدل رگرسیون خطی، قدر مطلق ضرایب به‌عنوان شاخص اهمیت ویژگی‌ها در نظر گرفته شد و برای SVR از روش اهمیت مبتنی بر جایگشت^۲ بهره گرفته شد. همچنین برای افزایش شفافیت تحلیلی و بررسی سهم هر ویژگی در سطح نمونه و در سطح کل مدل، از روش SHAP نیز استفاده شد. استفاده از SHAP این امکان را فراهم کرد که نه تنها اهمیت نسبی ویژگی‌ها مشخص شود، بلکه جهت و شدت اثر هر توصیفگر بر مقدار پیش‌بینی‌شده نیز تحلیل شود. نمای کلی فرایند تحلیل در شکل ۱ ارائه شده است. در نهایت، این فرایند مجموعه‌ای از خروجی‌های عددی و تصویری، شامل مقایسه عملکرد مدل‌ها، پیش‌بینی‌ها، نتایج مدل stacking، نمودارهای واقعی-پیش‌بینی، نمودارهای باقی‌مانده، نقشه حرارتی همبستگی، اهمیت ویژگی‌ها و نمودارهای SHAP را تولید کرد. این خروجی‌ها امکان ارزیابی کمی و تفسیر علمی نتایج را فراهم کردند.

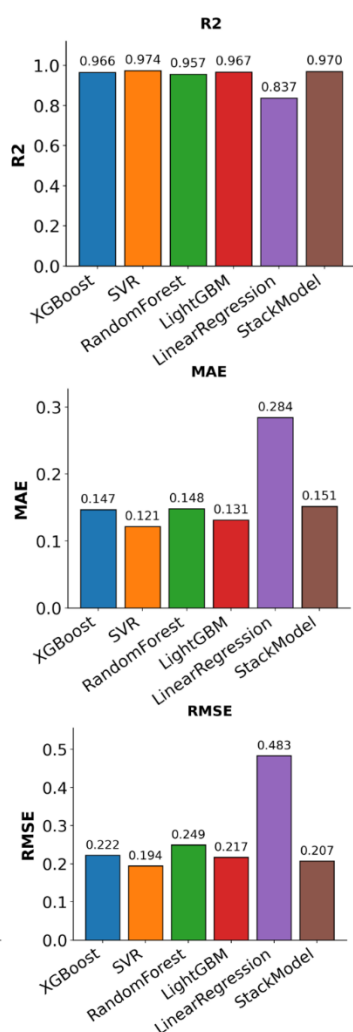
^۱ Feature Importances

^۲ permutation importance



شکل ۲. نقشه‌های حرارتی همبستگی پیرسون میان توصیف‌گرهای مختلف و پتانسیل ردوکس.

۲-۲ ارزیابی عملکرد مدل‌ها و گزینش مدل‌های بهینه



شکل ۳. مقایسه عملکرد مدل‌های یادگیری ماشین در پیش‌بینی

متغیر هدف بر اساس معیارهای R^2 ، RMSE و MAE.

از دیدگاه فیزیکی نیز قابل انتظار است، زیرا افزایش الکترون‌خواهی بیانگر تمایل بیشتر مولکول به پذیرش الکترون و پایداری بیشتر گونه آنیونی است. از آنجا که در این مطالعه پتانسیل ردوکس بر اساس اختلاف انرژی میان حالت خنثی و گونه آنیونی تعریف شده است، افزایش مقدار EA با کاهش مقدار عددی پتانسیل ردوکس همراه خواهد بود؛ بنابراین، روند مشاهده‌شده با تعریف پتانسیل ردوکس و مبانی ترمودینامیکی انتقال الکترون کاملاً سازگار است [۱۲]. در نتیجه، همبستگی بسیار قوی مشاهده‌شده نشان می‌دهد که تغییرات پتانسیل ردوکس در این مجموعه داده ارتباط نزدیکی با ویژگی‌های مرتبط با فرایند انتقال الکترون دارد. همچنین، انرژی‌های اوربیتال‌های HOMO ($r=-0.402$) و LUMO ($r=-0.417$) نیز همبستگی منفی با پتانسیل ردوکس نشان می‌دهند. کاهش انرژی این اوربیتال‌ها معمولاً با افزایش تمایل سیستم به پذیرش الکترون همراه است که می‌تواند به کاهش پتانسیل ردوکس منجر شود. در مقابل، شکاف انرژی تنها همبستگی بسیار ضعیفی با متغیر هدف ($r=0.097$) دارد که بیانگر آن است که این توصیف‌گر، به‌تنهایی، شاخص مناسبی برای پیش‌بینی پتانسیل ردوکس در مجموعه داده حاضر نیست. از سوی دیگر، ماتریس همبستگی وجود وابستگی‌های شدیدی را میان برخی از توصیف‌گرهای شیمی کوانتومی مفهومی آشکار می‌سازد. برای مثال، گاف انرژی و μ همبستگی کامل ($r=1.00$) دارند و بین EA و شاخص w نیز همبستگی مثبت بسیار بالایی ($r=0.94$) مشاهده می‌شود. این روابط قوی ناشی از ماهیت ریاضی این توصیف‌گراست، زیرا بسیاری از آن‌ها مستقیماً از انرژی‌های مرزی HOMO و LUMO استخراج می‌شوند. هرچند این همبستگی‌ها در مدل‌های خطی می‌توانند موجب هم‌خطی شوند، در مدل‌های غیرخطی معمولاً امکان بهره‌گیری از اطلاعات مکمل را فراهم می‌کنند.

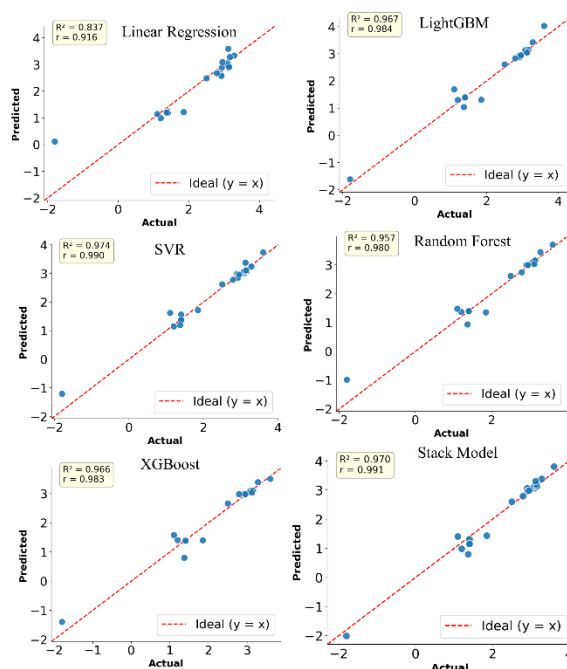
به منظور ارزیابی میزان انطباق مقادیر پیش‌بینی شده با مقادیر واقعی، نمودارهای پیش‌بینی در برابر مقادیر واقعی برای تمامی مدل‌ها بر روی مجموعه آزمون ترسیم شدند (شکل ۴). همان‌گونه که مشاهده می‌شود، مدل‌های رگرسیون خطی پراکندگی قابل توجهی پیرامون خط ایده‌آل ($y=x$) دارند که بیانگر دقت پایین‌تر آن‌ها در پیش‌بینی پتانسیل ردوکس است. در مقابل، در مدل‌های RF، SVR، XGBoost و LightGBM، Stacking، اغلب نقاط در نزدیکی خط ایده‌آل قرار گرفته‌اند که نشان‌دهنده انطباق مناسب مقادیر پیش‌بینی شده با مقادیر واقعی است. در این میان، SVR کمترین میزان پراکندگی را پیرامون خط مرجع نشان می‌دهد، در حالی که Stacking نیز با توزیع یکنواخت نقاط و عدم مشاهده انحراف سیستماتیک، عملکردی بسیار نزدیک به آن ارائه می‌کند. این الگوی پراکندگی، پایداری و توان تعمیم مطلوب این دو مدل را در پیش‌بینی داده‌های آزمون تأیید می‌کند.

برتری مدل‌های غیرخطی با ماهیت الکترونی پتانسیل ردوکس نیز سازگار است. این کمیت حاصل برهم‌کنش هم‌زمان چندین توصیف‌گر شیمی کوانتومی، از جمله انرژی‌های HOMO و LUMO، EA، گاف انرژی و μ است؛ از این رو، وابستگی آن به یک ویژگی منفرد قابل توصیف نیست. توانایی مدل‌های غیرخطی در در نظر گرفتن اثرات متقابل میان این توصیف‌گرا، دلیل معقولی برای دقت بالاتر آن‌ها نسبت به مدل‌های خطی فراهم می‌کند و نشان می‌دهد که برتری این مدل‌ها با مبانی فیزیکی و شیمی کوانتومی مسئله نیز همخوانی دارد.

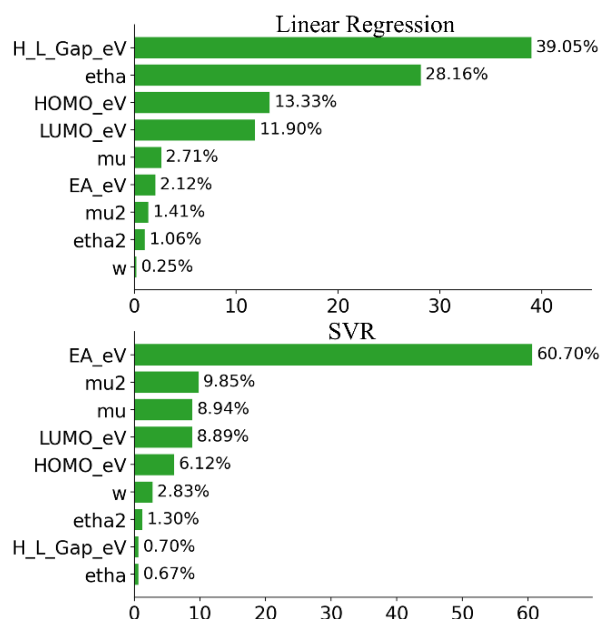
۳-۲ تحلیل اهمیت ویژگی‌ها

به منظور تفسیر نحوه تصمیم‌گیری مدل‌های یادگیری ماشین، تحلیل اهمیت ویژگی‌ها فراتر از یک رتبه‌بندی آماری در نظر گرفته شد و به عنوان ابزاری برای شناسایی توصیف‌گرهای کلیدی مؤثر بر پیش‌بینی پتانسیل ردوکس

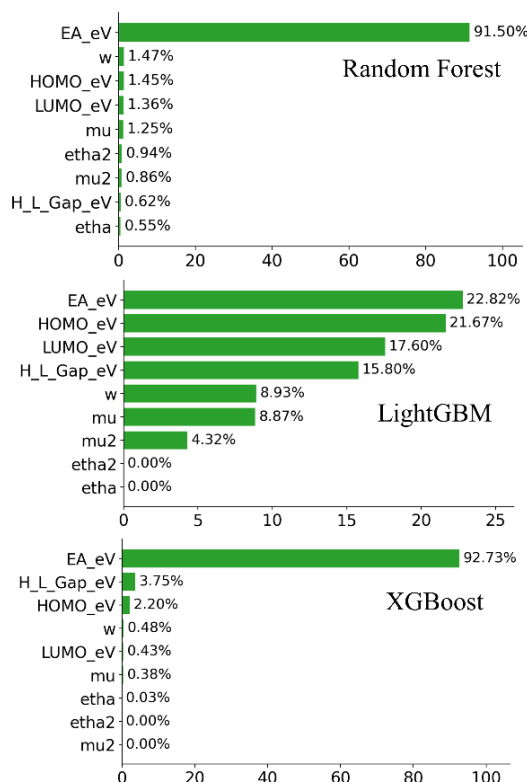
در گام بعد، عملکرد مدل‌های مختلف یادگیری ماشین در پیش‌بینی پتانسیل ردوکس با استفاده از شاخص‌های R^2 ، MAE و RMSE بر روی مجموعه آزمون ارزیابی شد. نتایج این مقایسه در شکل ۳ ارائه شده است. در میان مدل‌های بررسی‌شده، SVR با دستیابی به بالاترین ضریب تعیین ($R^2=0.974$) و کمترین میزان خطا (MAE=0.121، RMSE=0.194) بهترین دقت پیش‌بینی را نشان داد. StackModel نیز با اختلافی اندک ($R^2=0.970$) عملکردی قابل رقابت ارائه کرد که بیانگر پایداری و توان تعمیم مناسب رویکردهای آنسامبلی در این مسئله است. برتری SVR و همچنین مدل‌های مبتنی بر درخت، از جمله XGBoost و LightGBM، را می‌توان به توانایی آن‌ها در مدل‌سازی روابط غیرخطی میان توصیف‌گرهای حاصل از محاسبات DFT و پتانسیل ردوکس نسبت داد. SVR با بهره‌گیری از توابع کرنل، داده‌ها را به فضایی با ابعاد بالاتر نگاشت کرده و امکان یادگیری الگوهای پیچیده را فراهم می‌کند؛ قابلیت‌هایی که در مدل‌های خطی محدودتر است. که نشان می‌دهد که روابط میان توصیف‌گرهای مولکولی و پتانسیل ردوکس عمدتاً ماهیتی غیرخطی دارند.



شکل ۴. ارزیابی عملکرد مدل‌ها با استفاده از نمودارهای پیش‌بینی در مقابل مقادیر واقعی.



شکل ۵. رتبه‌بندی اهمیت ویژگی برتر در مدل رگرسیون خطی و SVR استفاده شده برای پیش‌بینی پتانسیل ردوکس.



شکل ۶. رتبه‌بندی اهمیت ویژگی برتر در مدل RF, XGBoost و LGMB استفاده شده برای پیش‌بینی پتانسیل ردوکس.

مورد استفاده قرار گرفت (شکل ۵ و ۶). اگرچه میزان اهمیت نسبی ویژگی‌ها در مدل‌های مختلف متفاوت است، همه آن‌ها نقش محوری توصیف‌گرهای حاصل از محاسبات DFT را در پیش‌بینی پتانسیل ردوکس تأیید می‌کنند. در مدل رگرسیون خطی، گاف انرژی بیشترین اهمیت را دارد که نشان می‌دهد این مدل تغییرات پتانسیل ردوکس را عمدتاً بر پایه پایداری کلی ساختار الکترونی تفسیر می‌کند. از دیدگاه شیمی کوانتومی، گاف انرژی با پایداری ترمودینامیکی و واکنش‌پذیری الکترونی ارتباط دارد؛ از این‌رو، انتظار می‌رود بر قابلیت انتقال الکترون و در نتیجه پتانسیل ردوکس اثرگذار باشد. در مقابل، تقریباً تمامی مدل‌های غیرخطی SVR، RF، XGBoost و LightGBM را به‌عنوان مهم‌ترین ویژگی شناسایی کرده‌اند. این مشاهده با نقش بنیادی EA در توصیف تمایل گونه‌های شیمیایی به پذیرش الکترون و ارتباط مستقیم آن با فرآیندهای اکسایش-کاهش سازگار است. اختصاص بیش از ۹۰٪ اهمیت به EA در مدل‌های RF و XGBoost نیز نشان می‌دهد که این ویژگی نقش غالبی در تصمیم‌گیری این مدل‌ها ایفا می‌کند.

علاوه بر EA، توصیف‌گرهای HOMO، LUMO، μ ، η و ω نیز در اغلب مدل‌ها سهم قابل‌توجهی دارند. حضور هم‌زمان این ویژگی‌ها نشان می‌دهد که پیش‌بینی پتانسیل ردوکس تنها بر یک توصیف‌گر منفرد متکی نیست، بلکه حاصل ترکیب اطلاعات مربوط به انتقال بار، پایداری الکترونی و واکنش‌پذیری مولکول است. تفاوت در توزیع اهمیت ویژگی‌ها نیز بازتابی از سازوکار متفاوت مدل‌ها در استخراج اطلاعات از فضای ویژگی‌هاست؛ به‌گونه‌ای که مدل‌های غیرخطی نسبت به مدل‌های خطی توانایی بیشتری در بهره‌گیری از روابط پیچیده میان توصیف‌گرهای الکترونی دارند. این نتایج، در کنار عملکرد پیش‌بینی برتر این مدل‌ها، نشان می‌دهد که ویژگی‌های منتخب از نظر آماری و از منظر شیمی کوانتومی، هر دو، اطلاعات معناداری برای پیش‌بینی پتانسیل ردوکس فراهم می‌کنند.

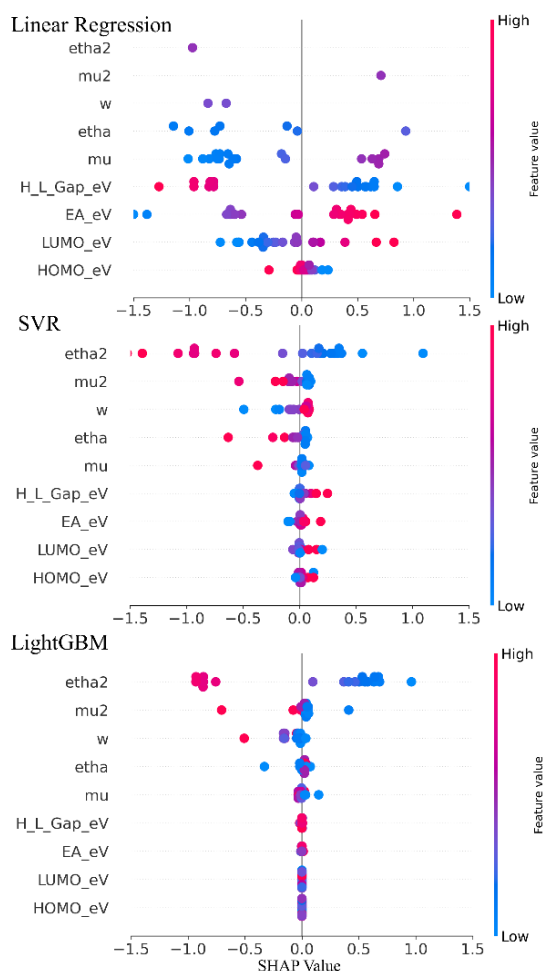
۲-۴ تفسیرپذیری مدل‌های یادگیری ماشین با استفاده از

تحلیل SHAP

به‌منظور تفسیرپذیری مدل‌های یادگیری ماشین و بررسی نحوه تأثیر هر توصیف‌گر بر پیش‌بینی پتانسیل ردوکس، تحلیل مقادیر SHAP برای تمامی مدل‌های توسعه‌یافته انجام شد که نتایج آن در شکل ۶ ارائه شده است. برخلاف تحلیل اهمیت ویژگی‌ها که تنها سهم نسبی هر متغیر را نشان می‌دهد، مقادیر SHAP علاوه بر میزان اهمیت، جهت و شدت اثر هر ویژگی بر خروجی مدل را نیز مشخص می‌کنند. اگرچه الگوی توزیع مقادیر SHAP در مدل‌های مختلف یکسان نیست، تمامی مدل‌ها نقش توصیف‌گرهای الکترونی حاصل از محاسبات DFT را در پیش‌بینی پتانسیل ردوکس تأیید می‌کنند.

در مدل‌های SVR، RF و XGBoost، EA گسترده‌ترین دامنه مقادیر SHAP را نشان می‌دهد که بیانگر تأثیر قابل‌توجه این ویژگی بر خروجی مدل است. همان‌گونه که در شکل ۶ مشاهده می‌شود، مقادیر بالاتر EA (نقاط قرمز) عمدتاً با مقادیر مثبت SHAP همراه هستند؛ در حالی که مقادیر پایین‌تر آن (نقاط آبی) سهم کمتری در افزایش خروجی مدل دارند. این روند نشان می‌دهد که افزایش EA پیش‌بینی مدل را به سمت مقادیر بالاتر پتانسیل ردوکس سوق می‌دهد. همچنین، در مدل LightGBM، ویژگی‌های EA، HOMO و LUMO سهم نسبتاً متوازی در تصمیم‌گیری مدل دارند که بیانگر استفاده هم‌زمان این الگوریتم از چندین توصیف‌گر الکترونی است. ویژگی‌های HOMO و LUMO نیز در اغلب مدل‌ها اثر قابل‌توجهی بر خروجی دارند، اما میزان و جهت این اثر به مقدار ویژگی و نوع مدل وابسته است. چنین رفتاری نشان می‌دهد که پیش‌بینی پتانسیل ردوکس تنها به یک توصیف‌گر منفرد وابسته نیست، بلکه حاصل برهم‌کنش مجموعه‌ای از ویژگی‌های مرتبط با ساختار الکترونی مولکول است. در مقابل، در مدل رگرسیون خطی، گاف انرژی بیشترین سهم را در تغییر خروجی مدل دارد که با فرض خطی بودن روابط در این مدل سازگار است.

در مجموع، نتایج تحلیل SHAP نشان می‌دهد که مدل‌های غیرخطی تصمیم‌گیری خود را بر پایه ترکیبی از توصیف‌گرهای الکترونی انجام می‌دهند و اثر هر ویژگی را متناسب با مقدار آن و در تعامل با سایر ویژگی‌ها در نظر می‌گیرند. هم‌راستایی این الگوها با مفاهیم شناخته‌شده شیمی کوانتومی، قابلیت تفسیر مدل‌های توسعه‌یافته را تقویت کرده و نشان می‌دهد که آن‌ها روابطی سازگار با رفتار الکترونی مولکول‌ها را در فرآیند پیش‌بینی بازتاب می‌کنند.



شکل ۷. نمودارهای SHAP مدل‌های رگرسیون خطی، SVR و LightGBM برای پتانسیل ردوکس.

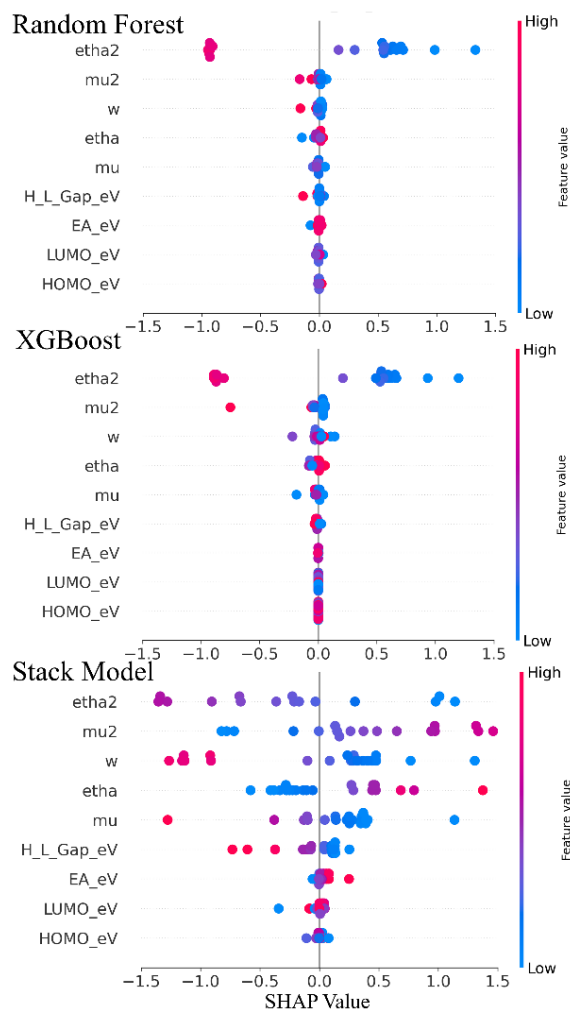
کنار انرژی‌های HOMO و LUMO، مهم‌ترین عوامل مؤثر بر پیش‌بینی پتانسیل ردوکس هستند. همچنین، همخوانی نتایج حاصل از مدل‌های یادگیری ماشین با مبانی شیمی کوانتومی بیانگر آن است که مدل‌های توسعه‌یافته از توصیف‌گرهایی استفاده می‌کنند که از نظر فیزیکی نیز با رفتار اکسایش-کاهش مولکول‌ها سازگار هستند.

اگرچه مدل Stacking در مجموعه داده حاضر عملکردی نزدیک به مدل SVR نشان داد، با توجه به محدود بودن مجموعه داده به ۱۰۶ نمونه، این نتایج نباید به‌عنوان اثبات قابلیت تعمیم گسترده مدل به سایر فضاهای شیمیایی تلقی شوند. بنابراین، استفاده از مجموعه داده‌های بزرگ‌تر و متنوع‌تر، همراه با اعتبارسنجی خارجی و بهینه‌سازی سیستماتیک ابرپارامترها، برای ارزیابی دقیق‌تر پایداری و قابلیت انتقال مدل‌ها در مطالعات آینده ضروری است.

به‌طور کلی، نتایج این مطالعه نشان می‌دهد که تلفیق محاسبات DFT با روش‌های یادگیری ماشین، رویکردی دقیق، تفسیرپذیر و کم‌هزینه برای پیش‌بینی پتانسیل ردوکس فراهم می‌آورد و می‌تواند زمان و هزینه غربالگری محاسباتی ترکیبات جدید را به‌طور قابل‌توجهی کاهش دهد. این چارچوب ظرفیت آن را دارد که در طراحی هدفمند مولکول‌های آلی فعال ردوکس برای باتری‌های جریان اکسایش-کاهش آبی و سایر سامانه‌های ذخیره‌سازی انرژی مورد استفاده قرار گیرد. در ادامه این پژوهش، توسعه مجموعه داده‌های بزرگ‌تر، استفاده از توصیف‌گرهای ساختاری و گراف‌محور، و به‌کارگیری مدل‌های یادگیری عمیق می‌تواند دقت پیش‌بینی و دامنه کاربرد این چارچوب را بیش از پیش افزایش دهد.

منابع

- [1] H. Chen, T. N. Cong, W. Yang, C. Tan, Y. Li, Y. Ding, "Progress in electrical energy storage system: A critical review," Progress in Natural Science, vol. 19, no. 3, pp. 291–312, 2009.



شکل ۸. نمودارهای SHAP مدل‌های خطی، SVR و LightGBM برای پتانسیل ردوکس.

۳- نتیجه گیری

در این پژوهش، قابلیت الگوریتم‌های مختلف یادگیری ماشین در پیش‌بینی پتانسیل ردوکس مولکول‌های آلی فعال ردوکس بر پایه توصیف‌گرهای استخراج‌شده از محاسبات DFT مورد بررسی قرار گرفت. نتایج نشان داد که مدل‌های غیرخطی، به‌ویژه RF، XGBoost و LightGBM، نسبت به مدل‌های خطی عملکرد دقیق‌تری در پیش‌بینی پتانسیل ردوکس دارند که بیانگر ماهیت غیرخطی روابط میان ساختار الکترونی و این ویژگی الکتروشیمیایی است. در میان آن‌ها، SVR بالاترین دقت پیش‌بینی را به دست آورد و مدل Stacking نیز عملکردی رقابتی و پایدار ارائه کرد.

تحلیل‌های هم‌بستگی، اهمیت ویژگی‌ها و مقادیر SHAP نشان دادند که توصیف‌گرهای مرتبط با انتقال الکترون، به‌ویژه EA، در

utilizing the surface redox reactions in folded graphene films,” *Chemistry of Materials*, vol. 27, no. 9, pp. 3291–3298, 2015.

[9] S. Kim, K. C. Kim, S. W. Lee, S. S. Jang, “Thermodynamic and redox properties of graphene oxides for lithium-ion battery applications: a first principles density functional theory modeling approach,” *Physical Chemistry Chemical Physics*, vol. 18, no. 30, pp. 20600–20606, 2016.

[10] T. T. Wu, G. L. Dai, J. J. Xu, F. Cao, X. H. Zhang, Y. Zhao, Y. M. Qian, “Structural design of organic battery electrode materials: from DFT to artificial intelligence,” *Rare Metals*, vol. 42, no. 10, pp. 3269–3303, 2023.

[11] A. Khetan, “High-throughput virtual screening of quinones for aqueous redox flow batteries: Status and perspectives,” *Batteries*, vol. 9, no. 1, Art. no. 24, 2022.

[12] O. Allam, B. W. Cho, K. C. Kim, S. S. Jang, “Application of DFT-based machine learning for developing molecular electrode materials in Li-ion batteries,” *RSC Advances*, vol. 8, no. 69, pp. 39414–39420, 2018.

[2] B. Dunn, H. Kamath, J. M. Tarascon, “Electrical energy storage for the grid: a battery of choices,” *Science*, vol. 334, no. 6058, pp. 928–935, 2011.

[3] G. Jeong, Y. U. Kim, H. Kim, Y. J. Kim, H. J. Sohn, “Prospective materials and applications for Li secondary batteries,” *Energy & Environmental Science*, vol. 4, no. 6, pp. 1986–2002, 2011.

[4] N. Nitta, F. Wu, J. T. Lee, G. Yushin, “Li-ion battery materials: present and future,” *Materials Today*, vol. 18, no. 5, pp. 252–264, 2015.

[5] P. G. Bruce, B. Scrosati, J. M. Tarascon, “Nanomaterials for rechargeable lithium batteries,” *Angewandte Chemie International Edition*, vol. 47, no. 16, pp. 2930–2946, 2008.

[6] J. Hassoun, S. Panero, P. Reale, B. Scrosati, “A new, safe, high-rate and high-energy polymer lithium-ion battery,” *Advanced Materials*, vol. 21, no. 47, pp. 4807–4810, 2009.

[7] M. D. Bhatt, C. O’Dwyer, “Recent progress in theoretical and computational investigations of Li-ion battery materials and electrolytes,” *Physical Chemistry Chemical Physics*, vol. 17, no. 7, pp. 4799–4844, 2015.

[8] T. Liu, K. C. Kim, R. Kaviani, S. S. Jang, S. W. Lee, “High-density lithium-ion energy storage



An Interpretable Machine Learning Framework for Predicting the Electrochemical Properties of Carbon-Based Electrode Materials in Lithium-Ion Batteries

^{1*}Fardin Taghizadeh, ¹Elham Zare

¹Department of Physics, Faculty of Basic Sciences, Yasuj University, Yasuj, Iran

Article details

Received: 2026/07/29
Accepted: 2026/08/26
Published: 2026/09/01

ISSN: 2588-493x
eISSN: 2588-4821

Correspondence email:

Taghizadeh@yu.ac.ir

Abstract

In this study, machine learning methods based on quantum chemical descriptors derived from Density Functional Theory (DFT) calculations were employed to predict the redox potentials of redox-active organic molecules. Electronic descriptors extracted from DFT calculations served as inputs for various machine learning models, whose performances were evaluated using R^2 , MAE, and RMSE metrics. The results demonstrated that the SVR model achieved the highest predictive accuracy, with a coefficient of determination (R^2) of 0.974 and the lowest error rate, while the Stacking model exhibited closely comparable and stable performance. Correlation analysis, feature importance, and SHAP values revealed that electron transfer-related descriptors, particularly electron affinity, play the most significant role in predicting redox potentials. The consistency between ML results and quantum chemical concepts indicates that the proposed framework achieves high predictive accuracy alongside robust interpretability. This study highlights that the integration of DFT calculations and machine learning can provide an efficient approach for the rapid screening and targeted design of redox-active organic molecules with optimized electrochemical properties.

Keywords: Density Functional Theory, Machine Learning, Redox Potential, Organic Molecules