



Original Article

Polycyclic Benzenoids HOMO-LUMO Gap Prediction Using Explainable Artificial Neural Network

Elmira Yektadoust¹, Amin Janghorbani^{1*}, Zohre Bahrami²

¹Biotechnology Department, Faculty of New Sciences and Technologies, Semnan University, Semnan, Iran

²Nanotechnology Department, Faculty of New Sciences and Technologies, Semnan University, Semnan, Iran

Received: 2025-05-04

Revised: 2025-05-23

Accepted: 2025-06-03

Abstract: One of the major challenges in pharmaceutical chemistry is the accurate prediction of molecular electronic properties that influence drug behavior. The HOMO-LUMO energy gap, as a key descriptor of chemical reactivity and stability, plays a crucial role in both pharmacodynamic and pharmacokinetic properties. Although methods such as Density Functional Theory (DFT) are highly accurate, their high computational cost makes them unsuitable for large-scale screening. In this study, a machine learning-based model was developed to rapidly predict this energy gap. A dataset of 8,676 cata-condensed polyaromatic compounds containing 3 to 10 fused benzene rings was compiled. Nine chemically meaningful features were extracted and used to train multilayer perceptron (MLP) and recurrent neural network (RNN) models. The best-performing MLP model achieved an R^2 score of 0.9750 and a test loss of 0.0037. Saliency map analysis revealed that features such as ionization potential, electron affinity, and total energy had the most significant impact. These findings demonstrate that AI-based models can not only achieve high predictive accuracy but also provide valuable insights for drug design

Keywords: Deep Learning, Explainable AI, Drug Discovery, HOMO-LUMO Gap, Polycyclic Benzenoids

*Corresponding Author: a.janghorbani@semnan.ac.ir

How to Cite This Article:

Yektadoust, E., Janghorbani, A., Bahrami, Z., Polycyclic Benzenoids HOMO-LUMO Gap Prediction Using Explainable Artificial Neural Network. Nanomeghyas. 2025;12(1): 75-86. DOI: [10.22034/ns.2025.725012](https://doi.org/10.22034/ns.2025.725012)





پیش بینی فاصله HOMO-LUMO در بنزنویدهای پلی سیکلیک با استفاده از شبکه های عصبی مصنوعی تفسیر پذیر

المیرا یکتادوست^۱، امین جانقربانی^{۱*}، زهره بهرامی^۲

^۱دانشکده بیوفناوری، پردیس علوم و نوین، دانشگاه سمنان، ایران
^۲دانشکده نانوفناوری، پردیس علوم و فناوری های نوین فناوری های، دانشگاه سمنان، ایران

تاریخ پذیرش: ۱۴۰۳/۱۱/۰۳

تاریخ بازنگری: ۱۴۰۳/۱۰/۱۹

تاریخ دریافت: ۱۴۰۳/۰۸/۲۶

چکیده: یکی از چالش های مهم در شیمی دارویی، پیش بینی دقیق ویژگی های الکترونیکی مولکول هاست که بر رفتار دارو تأثیر می گذارند. فاصله انرژی HOMO-LUMO به عنوان یک توصیف گر کلیدی در واکنش پذیری و پایداری شیمیایی، نقش مهمی در خواص فارماکودینامیکی و فارماکوکینتیکی دارد. هرچند روش هایی مانند نظریه تابع چگالی دقیق هستند، اما به دلیل هزینه محاسباتی بالا، برای غربالگری های گسترده مناسب نیستند. در این پژوهش، با بهره گیری از شبکه های عصبی مصنوعی، مدلی برای پیش بینی سریع این فاصله انرژی توسعه داده شد. مجموعه داده ای شامل ۸۶۷۶ ترکیب پلی آروماتیک کاتا-تراکم یافته شامل ۳ تا ۱۰ حلقه بنزن تهیه شد. نه ویژگی شیمیایی استخراج و برای آموزش مدل های پرسپترون چندلایه و شبکه عصبی بازگشتی استفاده شدند. بهترین مدل MLP به R^2 برابر ۰/۹۷۵۰ و خطای آزمون ۰/۰۰۳۷ دست یافت. تحلیل نقشه برجستگی نشان داد ویژگی هایی مانند پتانسیل یونش، میل الکترونی و انرژی کل بیشترین نقش را دارند. این نتایج نشان می دهند که مدل های مبتنی بر شبکه های عصبی مصنوعی می توانند علاوه بر دقت بالا، بینش ارزشمندی در طراحی دارو فراهم کنند.

واژگان کلیدی: یادگیری عمیق، هوش مصنوعی قابل تفسیر، کشف دارو، فاصله HOMO-LUMO، بنزن های چند حلقه ای

a.janghorbani@semnan.ac.ir

نحوه استناد به این مقاله:

یکتادوست، المیرا، جانقربانی، امین، بهرامی، زهره. پیش بینی فاصله HOMO-LUMO در بنزنویدهای پلی سیکلیک با استفاده از شبکه های عصبی مصنوعی تفسیر پذیر. *نانومقیاس*، ۱۲(۱)، ۷۵-۸۶.

DOI: 10.22034/ns.2025.725012



۱- مقدمه

یکی از چالش‌های اصلی در زیست‌فناوری و شیمی دارویی، طراحی و توسعه داروهای جدید است. کشف داروهایی مؤثر و با سمیت کم نیازمند درک عمیقی از ویژگی‌های فیزیکی، شیمیایی و الکترونیکی مولکول‌ها و نحوه تأثیر این ویژگی‌ها بر تعاملات زیستی است [۱]. یکی از پارامترهای کلیدی که بر خواص دارویی یک ترکیب تأثیر می‌گذارد، فاصله انرژی بین بالاترین اوربیتال مولکولی اشغال‌شده (HOMO) و پایین‌ترین اوربیتال مولکولی خالی (LUMO) است که به‌طور معمول با عنوان فاصله HOMO-LUMO شناخته می‌شود [۲].

این فاصله انرژی، اختلاف انرژی بین دو حالت کوانتومی حیاتی در یک مولکول را نشان می‌دهد و تأثیر مستقیمی بر پایداری شیمیایی، واکنش‌پذیری و تعاملات الکترواستاتیکی آن دارد. در طراحی دارو، فاصله HOMO-LUMO نقش مهمی در توانایی یک مولکول برای عبور از غشاهای زیستی، تعامل با آنزیم‌ها و گیرنده‌های سلولی، و در نهایت تعیین اثربخشی درمانی آن ایفا می‌کند [۳]. بنابراین، محاسبه و پیش‌بینی دقیق این فاصله انرژی در فرآیند کشف و توسعه دارو از اهمیت بالایی برخوردار است. ترکیباتی که فاصله انرژی کمتری دارند معمولاً از نظر الکترونیکی ناپایدارتر بوده و می‌توانند تعاملات فعال‌تری با پروتئین‌های زیستی داشته باشند. در مقابل، ترکیباتی با فاصله HOMO-LUMO بزرگ‌تر پایداری بیشتری دارند و ممکن است در محیط‌های زیستی دوام بیشتری داشته باشند [۴]. از این‌رو، تنظیم این پارامتر در طراحی دارو می‌تواند خواص فارماکوکینتیکی مانند جذب، توزیع، متابولیسم، کاهش سمیت و دفع دارو (ADME) را بهینه کند [۵ و ۶].

یکی از دسته‌های مهم ترکیبات آلی که به دلیل ویژگی‌های الکترونیکی منحصر به فرد و خواص زیستی مطلوب، توجه زیادی را به خود جلب کرده‌اند، ترکیبات بنزن‌دار چندحلقه‌ای (BPHs) هستند. این

ترکیبات از حلقه‌های بنزن به هم پیوسته تشکیل شده‌اند و پایداری بالایی در سیستم π -الکترونی از خود نشان می‌دهند. آن‌ها در کاربردهای مختلفی از جمله طراحی داروهای ضدسرطان، ضدویروس و ضدالتهاب، همچنین در سامانه‌های هدفمند انتقال دارو مورد استفاده قرار گرفته‌اند [۷]. از همه مهم‌تر، BPHها را می‌توان به‌عنوان قطعاتی بریده‌شده از ساختار گرافن در نظر گرفت که نقش حیاتی در صنعت دارویی و سایر صنایع ایفا می‌کند [۱۶].

مطالعات نشان داده‌اند که ویژگی‌های ساختاری و الکترونیکی BPHها مانند تعداد حلقه‌های آروماتیک، نوع و موقعیت گروه‌های عاملی، توزیع بار و قطبیت مولکولی تأثیر قابل توجهی بر فاصله HOMO-LUMO دارند. تغییر در این ویژگی‌های ساختاری می‌تواند خواص فارماکودینامیکی و فارماکوکینتیکی مشتقات دارویی حاصل از این ترکیبات را بهبود بخشد [۷].

با وجود نقش حیاتی فاصله انرژی HOMO-LUMO در طراحی دارو، محاسبه دقیق این پارامتر با استفاده از روش‌های سنتی شیمی کوانتومی و مکانیک مولکولی نیازمند زمان محاسباتی بالا و محاسبات پیچیده است. نظریه تابع چگالی (DFT) یکی از رایج‌ترین روش‌های شیمی محاسباتی برای محاسبه فاصله HOMO-LUMO است. اگرچه این روش دقیق است، اما به دلیل مقیاس‌پذیری ضعیف برای مولکول‌های بزرگ، کارایی آن در کاربردهایی با توان عملیاتی بالا مانند طراحی دارو محدود می‌شود [۴]. علاوه بر این، مطالعات سیستماتیک کمی در مورد تحلیل تأثیر ویژگی‌های ساختاری مختلف بر فاصله HOMO-LUMO در BPHها و مدل‌سازی این روابط انجام شده است. این خلأ پژوهشی مانعی برای بهینه‌سازی طراحی داروهای نوین مبتنی بر BPH به شمار می‌رود و نیاز به رویکردهای محاسباتی مؤثرتر را برجسته می‌کند. در نتیجه، توسعه روش‌های جایگزین که قادر به پیش‌بینی سریع و دقیق این پارامتر باشند، بسیار مطلوب است.

یادگیری ماشین (ML) و هوش مصنوعی (AI)

روش‌های مبتنی بر یادگیری ماشین، وابستگی آن‌ها به استخراج ویژگی‌های دستی است که می‌تواند منجر به از دست رفتن اطلاعات مهم و وابستگی شدید به انتخاب صحیح ویژگی‌ها شود.

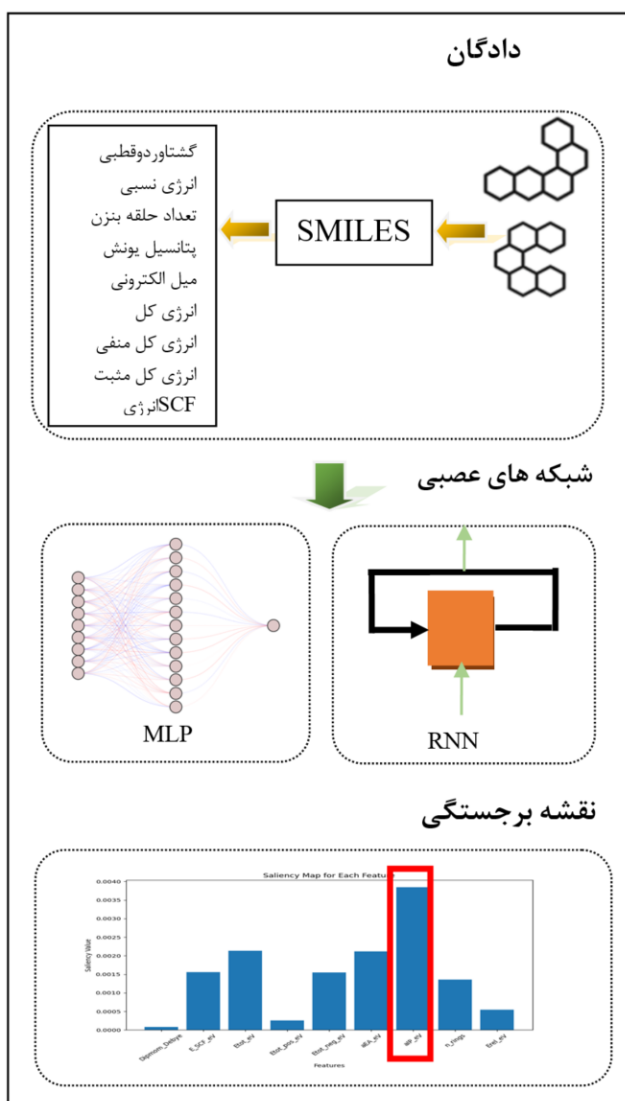
برای غلبه بر این محدودیت‌ها، از یادگیری عمیق (DL) استفاده شده است که توانایی استخراج خودکار ویژگی‌ها از ساختارهای مولکولی را دارد. شبکه‌های عصبی عمیق می‌توانند الگوهای پیچیده و غیرخطی را مستقیماً از داده‌ها یاد بگیرند، بدون نیاز به مهندسی دستی ویژگی‌ها. با این حال، «ماهیت جعبه‌سیاه» این مدل‌ها، تفسیرپذیری نتایج را کاهش داده و چالشی مهم در کاربرد آن‌ها در طراحی دارو به وجود آورده است.

برای رفع این چالش، تکنیک‌های مختلف تفسیرپذیری در کاربردهای پزشکی و سلامت مورد استفاده قرار گرفته‌اند تا به پژوهشگران در درک بهتر تصمیمات مدل‌های یادگیری عمیق کمک کنند. برای مثال، روش‌های نسبت‌دهی ویژگی مانند نقشه‌های برجستگی (Saliency Maps) و نقشه‌های فعال‌سازی کلاس (CAM) برای تحلیل شبکه‌های عصبی کانولوشنی (CNNs) به کار رفته‌اند. همچنین استفاده از نگاشت ویژگی‌های CNN روشی ساده برای تفسیر عملکرد مدل ارائه می‌دهد که نشان می‌دهد تفسیرپذیری یادگیری عمیق می‌تواند اعتماد به پژوهش‌های زیست‌پزشکی مبتنی بر هوش مصنوعی را افزایش دهد [۱۴].

در این مطالعه از نقشه‌های برجستگی و روش‌های تفسیرپذیر یادگیری ماشین برای شناسایی ویژگی‌های مؤثر در پیش‌بینی فاصله HOMO-LUMO استفاده شده است. این رویکرد بینش عمیق‌تری نسبت به عوامل مولکولی کلیدی تأثیرگذار بر ویژگی‌های الکترونیکی در BPH فراهم می‌کند و طراحی منطقی داروهای جدید را تسهیل می‌کند. یک مدل مبتنی بر شبکه‌های گراف کانولوشنی (GCN) برای پیش‌بینی فتوسمیت توسعه یافته که با یادگیری چندوظیفه‌ای و یادگیری فدرال ترکیب شده تا کاربرد آن در شیمی اطلاعات

پتانسیل چشمگیری در حوزه‌های مختلف سلامت، از جمله تحلیل سری‌های زمانی از سیگنال‌های فیزیولوژیکی مانند فشار خون و علائم حیاتی، و نیز پردازش تصاویر پزشکی نشان داده‌اند [۸]، [۹]. این رویکردها امکان پیش‌بینی مؤثر و دقیق در کاربردهای پیچیده زیست‌پزشکی را فراهم کرده و نقش تحول‌آفرین آن‌ها را بیشتر نمایان می‌سازند. تکنیک‌های یادگیری ماشین به‌عنوان جایگزینی کارآمد برای حل این مسئله توسعه یافته‌اند. مدل‌های رگرسیون برای مدل‌سازی رابطه بین ویژگی‌های ساختاری مولکول و فاصله HOMO-LUMO به کار رفته‌اند. مطالعات مختلفی از روش‌های یادگیری ماشین و محاسبات کوانتومی برای پیش‌بینی این فاصله استفاده کرده‌اند. در یک مطالعه، از توصیف‌گر SOAP (همپوشانی نرم موقعیت‌های اتمی) و رگرسیون خطی با اعتبارسنجی متقابل ۱۰- تایی برای برآورد انرژی اوربیتال‌های مرزی استفاده شد [۱۰]. در مطالعه‌ای دیگر، الگوریتم‌های جنگل تصادفی (RF)، ماشین بردار پشتیبان (SVM)، درخت‌های مدل M5P و شبکه‌های عصبی مصنوعی (ANN) برای توسعه چارچوبی پیش‌بینی‌کننده بر اساس ویژگی‌های الکترونیکی، توپولوژیکی و ساختاری به کار گرفته شدند [۱۱]. همچنین، ترکیبی از تحلیل همبستگی خطی و شبکه عصبی آموزش‌دیده با روش شبه-نیوتنی برای شناسایی توصیف‌گرهای کلیدی مانند HOMO، LUMO و تعداد اتم‌های اکسیژن و لیتیم به‌منظور آموزش مدل مورد استفاده قرار گرفت [۱۲]. در مطالعه‌ای دیگر، رویکردی ترکیبی از DFT و یادگیرماشین برای ارزیابی اثر دوپینگ کربن در ساختارهای دوبعدی بورن نیتريد شش‌ضلعی (h-BN) به کار گرفته شد. در این تحقیق، از محاسبات DFT با تابعی PBE۰ و توابع تقارنی Behler-Parrinello برای استخراج توصیف‌گرهای الکترونیکی و ساختاری، و از الگوریتم‌های RF و گرادیان بوستینگ (GB) برای پیش‌بینی فاصله HOMO-LUMO استفاده شد [۱۳].

با این حال، یکی از چالش‌های اصلی در



شکل ۱: نمایش شماتیک از چارچوب کلی این مطالعه. مجموعه داده شامل ساختارهای مولکولی به صورت رمزگذاری شده با نگارش SMILES، ورودی مدل محسوب می‌شود. توصیف‌گرهای مولکولی مرتبط استخراج شده و از طریق یک شبکه عصبی برای مدلسازی پیش‌بینانه پردازش می‌شوند. برای افزایش قابلیت تفسیر مدل، از نقشه برجستگی استفاده می‌شود تا تأثیرگذارترین ویژگی‌هایی که بر پیش‌بینی‌ها اثر می‌گذارند، مشخص شوند.

بر حلقه و ترکیب آن با ترنسفورمر SE (۳) امکان پیش‌بینی دقیق ویژگی‌های الکترونیکی را فراهم کرده و در عین حال هزینه‌های محاسباتی را به طور چشمگیری کاهش داده است. روش GradRAM برای برجسته‌سازی زیرساخت‌های شیمیایی معنادار و SHAP برای تفسیر ویژگی‌های کلیدی مولکولی مؤثر بر فاصله HOMO-LUMO مورد استفاده قرار گرفته‌اند. یافته‌ها نشان داد که مدل‌های تفسیرپذیر یادگیری ماشین نه تنها با نتایج DFT تطابق خوبی دارند، بلکه بینش‌های مکانیکی

تقویت شود. همچنین، روش گرادیان‌های یکپارچه (IG) برای شناسایی زیرساخت‌های مولکولی کلیدی در فتوسمیت به کار رفته است، در حالی که محاسبات شیمی کوانتومی در سطح ۳۱G-۶/B3LYP به منظور افزایش دقت پیش‌بینی‌ها انجام شده‌اند. در مطالعات مربوط به BPH، نمایش مولکولی جدیدی به نام LALAS (دنباله الحاق محلی خطی-زاویه‌ای) و مدل CUSTODI معرفی شدند که نیاز به انتخاب دستی ویژگی‌ها را حذف کرده و تفسیرپذیری را بهبود می‌بخشند. افزون بر این، نمای گرافیکی مبتنی

دست آمده و پس از مکاتبه با نویسنده اصلی مقاله مربوطه در دسترس قرار گرفت [۱۶].

این مجموعه شامل ۸۶۷۶ ترکیب BPH با ساختار کاتا-تراکم یافته است که دارای سه تا ده حلقه شش ضلعی هستند (بنزن و نفتالین حذف شدند زیرا بسیار کوچک اند و ویژگی های معناداری ارائه نمی دهند). فاصله های HOMO-LUMO این ترکیبات با استفاده از روش DFT در سطح نظری B3LYP-D3BJ/def2-SVP محاسبه شده اند.

هر مولکول BPH با ۹ ویژگی مولکولی توصیف شد. این ویژگی ها بر اساس ارتباطشان با تعیین فاصله HOMO-LUMO انتخاب شدند. از جمله مهم ترین آن ها می توان به گشتاور دوقطبی، که توزیع بار الکتریکی مولکول را تعریف می کند، انرژی محاسبه شده با روش میدان خودسازگار (E_SCF_eV) که شاخصی کلیدی برای پایداری الکترونیکی است، و انرژی کل مولکول (E_tot_eV) که بازتابی از سطح کلی انرژی سیستم است، اشاره کرد.

افزون بر این، اجزای مثبت و منفی انرژی کل (E_tot_pos_eV و E_tot_neg_eV) برای درک دقیق تری از ساختار الکترونیکی مورد تجزیه و تحلیل قرار گرفتند. میل الکترونی (aEA_eV)، که تمایل مولکول برای جذب یک الکترون را اندازه گیری می کند، و پتانسیل یونش (aIP_eV)، که انرژی مورد نیاز برای جدا کردن یک الکترون از مولکول را مشخص می کند، نیز بررسی شدند. همچنین تعداد حلقه های بنزن (n_rings)، که ساختار آروماتیک مولکول را تعیین می کند، و انرژی نسبی (Erel_eV) که مقیاسی مقایسه ای از انرژی مولکول در برابر ترکیبات مشابه ارائه می دهد، به عنوان توصیف گرهای مهم در این مطالعه گنجانده شدند.

برای آموزش و ارزیابی مدل های شبکه عصبی، مجموعه داده به سه بخش تقسیم شد: به منظور آموزش مدل و یادگیری ارتباط میان ویژگی های مولکولی و فاصله انرژی HOMO-LUMO، ۷۰٪ از داده ها معادل ۶۰۷۳ نمونه اختصاص یافت. برای جلوگیری از بیش برآش و تنظیم پارامترهای مدل،

ارزشمندی در زمینه پایداری مولکولی، تعاملات $\pi-\pi$ و اثرات کشش مولکولی ارائه می دهند و طراحی منطقی مواد عملکردی و ترکیبات دارویی را در الکترونیک آلی و پژوهش های دارویی تسهیل می کنند.

در این مطالعه، از تکنیک های تفسیرپذیر یادگیری ماشین برای پیش بینی فاصله HOMO-LUMO در هیدروکربن های بنزن دار چندحلقه ای (BPHs) استفاده شده است. مجموعه ای از مدل های رگرسیون، شامل رگرسیون جنگل تصادفی (RFR)، شبکه عصبی مصنوعی (ANN)، رگرسیون بردار پشتیبان (SVR)، درختان تصادفی شدید (ETR) و رگرسیون گرادیان بوستینگ (GBR) برای بررسی رابطه بین ساختار مولکولی و ویژگی های الکترونیکی به کار رفته اند. برای بهبود تفسیرپذیری مدل ها، از روش SHAP استفاده شده تا بینش هایی نسبت به ویژگی های کلیدی مولکولی مؤثر بر فاصله HOMO-LUMO ارائه دهد.

از این رو، در مطالعه حاضر، دو مدل شبکه عصبی پرسپترون چندلایه (MLP) و شبکه عصبی بازگشتی (RNN) برای پیش بینی فاصله HOMO-LUMO در BPH ها استفاده شده اند. مجموعه ای از ویژگی های فیزیکی و شیمیایی مولکولی به عنوان ورودی این مدل ها در نظر گرفته شده اند. این ویژگی ها برای استخراج الگوهای رابطه ای و یادگیری توابع پیچیده بین خواص مولکولی و فاصله HOMO-LUMO به کار رفته اند. شکل ۱ نمایشی شماتیک از چارچوب کلی تحقیق را در قالب یک دیاگرام سطح بالا نشان می دهد.

۲- روش تحقیق

۲-۱- جمع آوری و آماده سازی داده ها

در این مطالعه، از یک مجموعه داده شامل ترکیبات BPH استفاده شد که ویژگی های مولکولی و مشخصه های مربوط به انرژی این ترکیبات را ارائه می دهد. این مجموعه داده از منبعی علمی و معتبر به

۲-۳- پرسپترون چندلایه

یکی از مدل‌های پرکاربرد شبکه عصبی که دقت بالایی در داده‌های جدولی دارد، پرسپترون چندلایه (MLP) است. این شبکه عصبی پیش‌خور، یک خروجی را از چندین ورودی عددی واقعی محاسبه می‌کند. ابتدا ترکیبی خطی از مقادیر ورودی ایجاد می‌شود و سپس خروجی از طریق یک تابع تصمیم‌گیری غیرخطی پیش‌بینی می‌گردد MLP. ها از رایج‌ترین مدل‌های شبکه عصبی برای مسائل رگرسیون مانند تخمین فاصله HOMO-LUMO هستند، زیرا قادرند روابط پیچیده بین ویژگی‌های ورودی و متغیر هدف را یاد بگیرند [۱۵].

۲-۴- معماری شبکه‌ها و ابرپارامترها

برای تخمین فاصله HOMO-LUMO، مدل‌های MLP و RNN با معماری‌های متفاوت طراحی شدند. ساختار این مدل‌ها به گونه‌ای طراحی شده که روابط پیچیده بین ویژگی‌های مولکولی را استخراج کنند. لایه ورودی شامل ۹ نورون بود که معادل تعداد ویژگی‌های مولکولی در هر نمونه است. پیکربندی معماری مدل‌ها در جدول ۱ ارائه شده است که جزئیات تعداد لایه‌ها و نورون‌های استفاده شده برای بهینه‌سازی استخراج ویژگی و یادگیری نمایش‌های سطح بالا را شرح می‌دهد.

۱۰٪ داده‌ها، معادل ۱۷۳۵ نمونه، به عنوان مجموعه اعتبارسنجی در نظر گرفته شد. در نهایت، برای ارزیابی نهایی عملکرد مدل، ۲۰٪ از داده‌ها، معادل ۸۶۸ نمونه، به عنوان مجموعه آزمون مستقل استفاده شد.

۲-۲- شبکه عصبی بازگشتی

شبکه‌های عصبی بازگشتی (RNN) به دلیل توانایی در حفظ وابستگی‌های زمانی از طریق اتصالات بازگشتی، برای مدل‌سازی داده‌های ترتیبی بسیار مناسب‌اند. برخلاف شبکه‌های عصبی پیش‌خور، RNNها از یک حالت پنهان استفاده می‌کنند که در طول زمان تکامل می‌یابد و امکان حفظ و پردازش اطلاعات زمینه‌ای از ورودی‌های قبلی را فراهم می‌کند.

در این مطالعه، هر نمونه معرف یک ترکیب BPH است که توصیف‌گرهای مولکولی آن می‌توانند به صورت ورودی ترتیبی سازماندهی شوند. با در نظر گرفتن این توصیف‌گرها به عنوان دنباله‌ای مرتب شده، RNNها قادر به درک وابستگی‌های پیچیده بین ویژگی‌های مولکولی بوده و نمایشی جامع‌تر از ساختار الکترونیکی ارائه می‌دهند. این رویکرد امکان استخراج الگوهای پنهان در ویژگی‌های مولکولی را فراهم می‌کند و دقت پیش‌بینی فاصله HOMO-LUMO را افزایش می‌دهد، در حالی که روابط درونی بین ویژگی‌های اتمی و الکترونیکی را نیز حفظ می‌کند.

جدول ۱: پارامترهای مختلف آموزشی استفاده شده در مدل‌های گوناگون

مدل	مدل ۱	مدل ۲	مدل ۳	مدل ۴	مدل ۵	مدل ۶	مدل ۷	مدل ۸
نوع شبکه عصبی	MLP	MLP	MLP	MLP	RNN	RNN	RNN	RNN
تعداد پارامترها	۲۷۵۳	۷۰۵	۷۰۵	۳۵۳	۱۴۰۹	۱۱۲۱	۹۶۹۸۳	۵۶۳۳
تعداد	۱	۱	۱	۱	۱	۱	۳	۳
تعداد نورون در هر لایه	۳۲-۶۴	۶۴	۶۴	۳۲	۳۲	۲۸	۱۲۸	۳۲
الگوریتم یادگیری	SGD	Adam	SGD	SGD	SGD	SGD	SGD	Adam
نرخ یادگیری	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱	۰/۰۰۰۰۱	۰/۰۰۰۱

$$S_c(X) \approx w^T X + b = \sum_{i=1}^{|X|} w_i x_i + b \quad (1)$$

$$\text{Saliency Score} = w = \left| \frac{\partial S_c}{\partial X} \right|_{X=X_0} \quad (2)$$

۶-۲- معیارهای ارزیابی

۶-۲-۱- میانگین مربعات خطا

میانگین مربعات خطا که با نام انحراف مربع میانگین نیز شناخته می‌شود، نشان‌دهنده تفاوت‌های مربعی بین مقادیر واقعی و مقادیر پیش‌بینی شده است. این معیار میزان نزدیکی خط رگرسیون (خط برازش‌یافته) به مجموعه داده‌ها را اندازه‌گیری می‌کند. مقدار MSE همواره مقداری مثبت دارد، زیرا با استفاده از توان دوم اختلاف‌ها، اثر مقادیر منفی حذف می‌شود. مقدار پایین‌تر MSE نشان‌دهنده دقت بالاتر پیش‌بینی مدل است [۱۷].

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_1)^2 \quad (3)$$

۶-۲-۲- امتیاز R²

این معیار اندازه‌گیری کمی از میزان تطابق نقاط داده با خط رگرسیون برازش‌یافته را فراهم می‌کند. این معیار که معمولاً به عنوان ضریب تعیین یا ضریب تعیین چندگانه در رگرسیون چندمتغیره شناخته می‌شود، نشان‌دهنده نسبت واریانس موجود در متغیر پاسخ است که توسط مدل توضیح داده شده است. مقدار R² بین ۰٪ تا ۱۰۰٪ متغیر است، به طوری که: ۰٪ نشان می‌دهد که مدل هیچ‌گونه تغییری در داده‌های پاسخ را نسبت به میانگین آن توضیح نمی‌دهد. ۱۰۰٪ نشان می‌دهد که مدل تمام تغییرات موجود در داده‌های پاسخ را به‌طور کامل توضیح

در تمام لایه‌های پنهان مدل‌های MLP از تابع فعال‌سازی ReLU استفاده شد، در حالی که برای مدل‌های RNN تابع tanh به کار گرفته شد تا کارایی بهبود یافته و از مشکل ناپدید شدن گرادیان جلوگیری شود. در نهایت، یک لایه خروجی با یک نورون برای پیش‌بینی فاصله HOMO-LUMO طراحی شد.

برای ارزیابی اختلاف بین مقادیر پیش‌بینی‌شده و واقعی، از تابع خطای میانگین مربعات (MSE یا MSE Loss) استفاده گردید. به‌منظور بهینه‌سازی وزن‌ها، از روش‌های گرادیان کاهشی تصادفی (SGD) یا Adam با نرخ یادگیری متفاوت و مومنتوم ۰/۹ استفاده شد.

۲-۵- تحلیل نقشه برجستگی

یکی از روش‌های مبتنی بر گرادیان برای تفسیرپذیری مدل‌های شبکه عصبی، تکنیک نقشه برجستگی (Saliency Map) است. در این روش، گرادیان پیش‌بینی فاصله HOMO-LUMO نسبت به هر ویژگی ورودی محاسبه می‌شود. مقادیر گرادیان بالاتر نشان‌دهنده تأثیر بیشتر آن ویژگی بر خروجی مدل است.

در این مطالعه، تحلیل رگرسیونی برای تخمین فاصله HOMO-LUMO بر اساس چندین توصیف‌گر مولکولی از جمله گشتاور دوقطبی، انرژی کل، اجزای مثبت و منفی انرژی کل، میل الکترونی، پتانسیل یونش، تعداد حلقه‌ها و انرژی نسبی انجام شد. برای تحلیل دقیق‌تر تأثیر هر ویژگی بر مدل پیش‌بینی، از روش نقشه برجستگی استفاده شد. این تکنیک امکان تفسیر کمی اهمیت ویژگی‌های مختلف را فراهم می‌سازد و دید روشنی نسبت به اینکه کدام توصیف‌گرهای مولکولی در تعیین فاصله HOMO-LUMO نقش کلیدی دارند ارائه می‌دهد.

امتیازهای برجستگی برای هر ویژگی مؤثر با استفاده از معادلات (۱) و (۲) محاسبه شدند که به تفسیرپذیری و اعتبارسنجی مدل کمک می‌کنند [۱۶].

می‌دهد [۱۷].

فرمول ریاضی R^2 (ضریب تعیین) در معادله (۴) ارائه شده است.

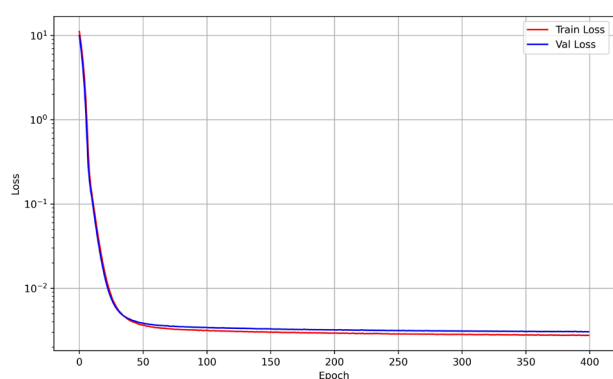
$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (۴)$$

۲

برابر با ۰/۹۷۵۰ را کسب کردند و همچنین کمترین میزان خطای آزمون (۰/۰۰۳۷ تا ۰/۰۰۳۹) را داشتند که نشان‌دهنده پایداری بالای آن‌هاست. جالب آن‌که مدل ۲ که از معماری ساده‌تری با پارامترهای کمتر استفاده کرده بود، عملکردی برابر با سایر مدل‌های پیچیده‌تر داشت، که این موضوع نشان می‌دهد پیچیدگی بیشتر مدل الزاماً برای شناسایی ویژگی‌های کلیدی مؤثر بر فاصله HOMO-LUMO ضروری نیست.

نتایج ارائه‌شده در جدول ۲ همچنین نشان می‌دهند که مدل‌های مبتنی بر RNN نه تنها عملکرد قابل قبولی داشتند، بلکه نسبت به مطالعه Zheng و همکاران [۱۶] بهبود قابل توجهی نشان دادند، که این امر قابلیت استفاده آن‌ها در پیش‌بینی خواص مولکولی را تأیید می‌کند.

برای ارزیابی تأثیر پیچیدگی مدل بر عملکرد پیش‌بینی، معماری‌هایی با تعداد پارامترهای مختلف مورد بررسی قرار گرفتند. به‌ویژه، مدل ۲ که تعداد پارامترهای کمتری داشت، یکی از بهترین عملکردها را با خطای آزمون ۰/۰۰۳۷ و امتیاز R^2 آزمون برابر با ۰/۹۷۵۰ ارائه داد. منحنی‌های خطا در مراحل آموزش و اعتبارسنجی برای مدل ۲ در شکل ۲ نشان داده شده‌اند. این یافته نشان می‌دهد که حتی مدل‌هایی با پیچیدگی پایین نیز می‌توانند ویژگی‌های ضروری مولکولی را بدون کاهش دقت شناسایی کرده و گزینه‌ای محاسباتی بهینه‌تر فراهم کنند [۱۶].



شکل ۲. خطای آموزش و اعتبارسنجی بهترین مدل

۲-۶-۳- تنظیمات تجربی

تمام نتایج تجربی بر روی پلتفرم Google Colab اجرا شده‌اند. تمامی پردازش‌ها با استفاده از نسخه ۳/۱۰ زبان برنامه‌نویسی پایتون انجام شده‌اند. برای آموزش مدل‌ها از یک نرم‌افزار متن‌باز مربوط به کتابخانه‌های شبکه عصبی در زبان پایتون استفاده شد. همچنین کتابخانه Scikit-Learn به همراه دیگر نرم‌افزارهای متن‌باز برای محاسبه معیارهای ارزیابی پیاده‌سازی گردید.

۳- نتایج و بحث

پیش‌بینی دقیق فاصله HOMO-LUMO در شیمی محاسباتی از اهمیت بنیادی برخوردار است، چرا که ویژگی‌های الکترونیکی مولکولی، رفتار نوری و واکنش‌پذیری شیمیایی را تعیین می‌کند. در این مطالعه، مدل‌های شبکه عصبی مبتنی بر MLP و RNN برای پیش‌بینی فاصله HOMO-LUMO بر اساس مجموعه‌ای از توصیف‌گرهای مولکولی توسعه داده شدند. معیارهای ارزیابی ارائه‌شده در جدول ۲، از جمله خطای آموزش، خطای اعتبارسنجی، خطای آزمون و امتیاز R^2 نشان می‌دهند که مدل‌های پیشنهادی عملکرد بهتری نسبت به مدل پایه گزارش شده با امتیاز R^2 برابر با ۰/۹۳۴ و خطای ۰/۰۱ داشته‌اند [۱۶].

در میان مدل‌ها، مدل‌های مبتنی بر MLP شامل مدل‌های ۲، ۳ و ۴ بالاترین امتیاز R^2 آزمون

جدول ۲. خطای آموزش، اعتبارسنجی و آزمون و ضریب تعیین مدل‌ها [۱۹]

شماره مدل	مدل ۱	مدل ۲	مدل ۳	مدل ۴	مدل ۵	مدل ۶	مدل ۷	مدل ۸
خطای آموزش	۰/۰۰۲۴	۰/۰۰۲۳	۰/۰۰۲۸	۰/۰۰۲۸	۰/۰۰۳۸	۰/۰۰۳۶	۰/۱۲۲۰	۰/۰۰۶۱
خطای اعتبارسنجی	۰/۰۰۲۷	۰/۰۰۲۷	۰/۰۰۳۱	۰/۰۰۳۰	۰/۰۰۴۵	۰/۰۰۴۳	۰/۱۸۲۳	۰/۰۰۹۲
خطای آزمون	۰/۰۰۳۹	۰/۰۰۳۷	۰/۰۰۳۸	۰/۰۰۳۸	۰/۰۰۸۶	۰/۰۰۷۶	۰/۳۴۵۰	۰/۱۴۷۹
ضریب تعیین اعتبارسنجی	۰/۹۸۲۳	۰/۹۸۲۰	۰/۹۷۹۷	۰/۹۸۰۲	۰/۹۷۱۵	۰/۹۷۲۹	—	۰/۹۴۶۶
ضریب تعیین آزمون	۰/۹۷۴۴	۰/۹۷۵۰	۰/۹۷۵۰	۰/۸۷۵۰	۰/۹۵۰۵	۰/۹۵۵۴	—	۰/۲۷۰۰

به‌عنوان توصیف‌گرهای بنیادی برای پیش‌بینی فاصله HOMO-LUMO شناخته شده‌اند، زیرا مستقیماً به انرژی‌های اربیتال مولکولی مرزی مرتبط هستند [۱۷]، تحلیل نقشه‌های برجستگی تفاوت جالبی بین مدل‌های MLP و RNN نشان داد.

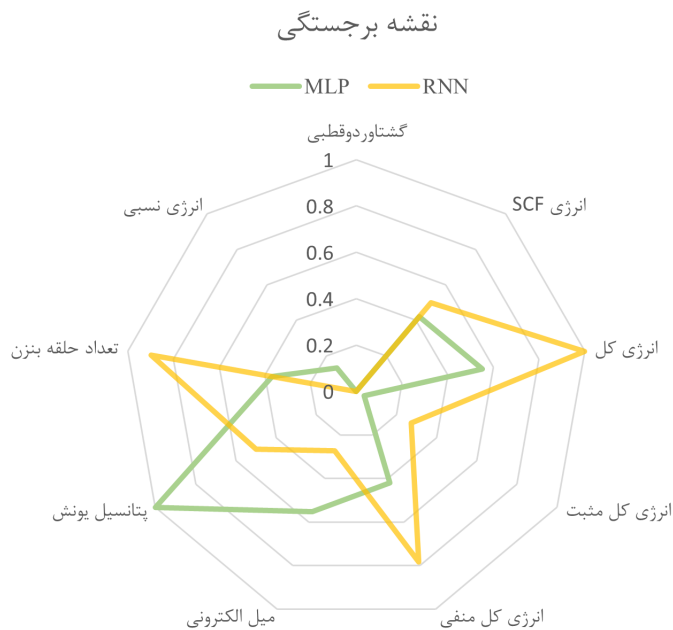
در مورد مدل MLP، IP و EA به درستی به‌عنوان ویژگی‌های مهم شناسایی شدند که با اهمیت نظری آن‌ها مطابقت دارد. طبق تقریب کوپمانز، انرژی HOMO تقریباً برابر منفی IP و انرژی LUMO تقریباً برابر منفی EA است؛ بنابراین فاصله HOMO-LUMO می‌تواند تقریباً با تفاوت بین EA و IP برآورد شود. این مقادیر نمایانگر تمایل مولکول به از دست دادن یا جذب الکترون هستند و مستقیماً با ساختار الکترونیکی و واکنش‌پذیری مرتبط‌اند. تطابق مقادیر برجستگی با دانش حوزه‌ای در مدل MLP نشان می‌دهد که مدل توانسته این اصول فیزیکوشیمیایی را به‌درستی در طول یادگیری درونی‌سازی و بازتاب دهد.

اما در مدل RNN، با وجود عملکرد قابل قبول از نظر دقت پیش‌بینی و امتیاز R^2 بالا، نقشه‌های برجستگی موفق به شناسایی IP و EA به‌عنوان ویژگی‌های غالب نشدند. یکی از دلایل احتمالی این مسئله می‌تواند پدیده محو شدن گرادیان باشد که در معماری‌های RNN شایع است. از آن‌جا که نقشه‌های برجستگی به گرادیان‌های بازپراکندگی برای تخمین اهمیت ویژگی‌ها متکی‌اند، تضعیف یا ناپدید شدن گرادیان‌ها در حین آموزش می‌تواند

به‌منظور اطمینان از این که پیش‌بینی‌ها صرفاً متکی بر روش‌های اکتشافی یادگیری ماشینی نیستند، بلکه بر اصول بنیادین شیمی نیز استوارند، از تکنیک‌های تبیین‌پذیری در تحلیل استفاده شد. شکل ۳، نقشه‌های برجستگی بهترین مدل‌های MLP و RNN یعنی مدل ۲ و مدل ۶ را نشان می‌دهد. این روش‌ها بینش عمیق‌تری نسبت به توصیف‌گرهای مولکولی مؤثر در پیش‌بینی فاصله HOMO-LUMO ارائه دادند و تفسیرپذیری و اعتبار علمی نتایج را افزایش دادند. هم‌راستایی پیش‌بینی‌های محاسباتی با شهود شیمیایی، قابلیت اطمینان این مدل‌ها را برای کاربردهای عملی مانند طراحی مولکولی، علم مواد، و شیمی کوانتومی تقویت می‌کند.

برای تحلیل میزان تأثیر توصیف‌گرهای مولکولی بر پیش‌بینی فاصله HOMO-LUMO، از نقشه‌های برجستگی استفاده شد. نتایج این تحلیل نشان داد که ویژگی‌های “ $E_{tot} - eV$ ” (انرژی کل)، “ aE_{a} ” (الفاای eV) (الفاای میل الکترونی)، و “ aIP_{eV} ” (الفاای پتانسیل یونش) در دقیق‌ترین مدل‌ها (مدل‌های ۲، ۳ و ۴) به‌صورت مکرر به‌عنوان تأثیرگذارترین ویژگی‌ها ظاهر شدند. این یافته‌ها با نظریه‌های شیمی کوانتومی مبتنی بر اصول اولیه که تأکید دارند انرژی‌های اربیتال مرزی، پتانسیل یونش و میل الکترونی از عوامل اصلی تعیین‌کننده فاصله HOMO-LUMO هستند، همخوانی قوی دارند [۴ و ۵].

اگرچه پتانسیل یونش (IP) و میل الکترونی (EA)



شکل ۳: نقشه برجستگی مدل MLP و RNN

بیش‌برازش و کاهش تعمیم‌پذیری شوند. برای بهبود بیشتر دقت و پایداری پیش‌بینی، پژوهش‌های آتی باید بر گسترش مجموعه داده‌ها، پیاده‌سازی تکنیک‌های منظم‌سازی پیشرفته و بررسی معماری‌های نوآورانه مانند شبکه‌های عصبی گرافی (GNN) و مدل‌های مبتنی بر فیزیک متمرکز شوند. این پیشرفت‌ها می‌توانند به توسعه مدل‌هایی قابل اعتمادتر و تفسیرپذیرتر از نظر فیزیکی منجر شوند و کاربردهای شیمی محاسباتی را در طراحی مواد و مولکول‌ها گسترش دهند.

در نهایت، پیش‌بینی دقیق فاصله HOMO-LUMO در هیدروکربن‌های بنزنوئیدی پلی‌سیکلیک (BPHs) مسیر را برای طراحی منطقی و تولید تجاری نانومواد کربنی مبتنی بر گرافن هموار می‌سازد؛ موادی که در داروسازی، الکترونیک و سایر صنایع پیشرفته پتانسیل چشمگیری دارند.

تقدیر و تشکر

نویسندگان بدین‌وسیله مراتب تشکر صمیمانه خود را از دکتر امید ترخانه بابت راهنمایی‌های ارزشمندشان در طول انجام این پژوهش اعلام می‌دارند.

باعث کاهش دقت و قابلیت اعتماد این روش در مدل‌های RNN شود. بنابراین، اگرچه مدل ممکن است عملکرد پیش‌بینی بالایی از خود نشان دهد، اما تکنیک‌های تفسیرپذیری مبتنی بر گرادینان مانند نقشه‌های برجستگی ممکن است نتوانند سهم واقعی توصیف‌گرهای معنادار فیزیکی را در فرآیند تصمیم‌گیری مدل به‌درستی شناسایی کنند.

۴- نتیجه‌گیری

این مطالعه اثربخشی مدل‌های شبکه عصبی، به‌ویژه معماری‌های MLP و RNN، را در پیش‌بینی فاصله HOMO-LUMO بر اساس توصیف‌گرهای مولکولی نشان داد. عملکرد برتر MLP نسبت به RNN اهمیت انتخاب معماری مناسب برای داده‌های غیرتوالی را برجسته می‌کند. همچنین، تحلیل تفسیرپذیری تأیید کرد که ویژگی‌های کلیدی شیمی کوانتومی مانند انرژی کل، میل الکترونی و پتانسیل یونش نقش غالبی در پیش‌بینی فاصله دارند. یافته‌ها همچنین به محدودیت‌های مدل‌های پیش‌ازحد پیچیده اشاره دارند که می‌توانند منجر به

تعارض منافع

نویسندگان این مقاله بدین وسیله اعلام می‌دارند که در ارتباط با انجام، تحلیل، نگارش و انتشار این پژوهش، هیچ‌گونه تعارض منافع مالی، شخصی، علمی یا سازمانی وجود ندارد.

مراجع

- [8] J. Qiu, Q. Wu, G. Ding, Y. Xu, S. Feng, "A survey of machine learning for big data processing," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, 2016. <https://doi.org/10.1186/s13634-016-0355-x>
- [9] Y. Xu, Q. Chu, D. Chen, A. Fuentes, "HOMO-LUMO gaps and molecular structures of polycyclic aromatic hydrocarbons in soot formation," *Front. Mech. Eng.*, vol. 7, 2021. <https://doi.org/10.3389/fmech.2021.744001>
- [10] F. Pereira, K. Xiao, D. A. R. S. Latino, C. Wu, Q. Zhang, J. Aires-De-Sousa, "Machine learning methods to predict density functional theory B3LYP energies of HOMO and LUMO orbitals," *J. Chem. Inf. Model.*, vol. 57, no. 1, pp. 11-21, 2017. <https://doi.org/10.1021/acs.jcim.6b00340>
- [11] "Application of DFT-based machine learning for developing molecular electrode materials in i-ion batteries." [Online]. Available: Google Search. Accessed: Mar. 17, 2025.
- [12] "AI-powered energy gap prediction in carbon-doped 2D materials: harnessing Behler-Parrinello descriptors for high-density systems." [Online]. Available: Google Search. Accessed: Mar. 17, 2025.
- [13] E. Yektadoust, A. Janghorbani, A. F. Talebi, "XCNN-SC: explainable CNN for SARS-CoV-2 variants classification and mutation detection," *Comput. Biol. Med.*, vol. 167, 2023. <https://doi.org/10.1016/j.compbiomed.2023.107606>
- [14] S. A. Hosseini Khorasani, E. Borhani, M. Yousefieh, A. Janghorbani, "Towards tailored thermoelectric materials: an artificial intelligence-powered approach to material design," *Physica B Condens. Matter*, vol. 685, 2024. <https://doi.org/10.1016/j.physb.2024.415946>
- [15] B. Hashemi, A. F. Talebi, A. Janghorbani, "Generative adversarial networks based synthetic biology: a promising approach to SARS-CoV-2 mutations prediction," *Eng. Appl. Artif. Intell.*, vol. 148, 2025. <https://doi.org/10.1016/j.engappai.2025.110395>
- [16] F. Zheng, Z. Zhu, J. Lu, Y. Yan, H. Jiang, Q. Sun, "Predicting the HOMO-LUMO gap of benzenoid polycyclic hydrocarbons via interpretable machine learning," *Chem. Phys. Lett.*, vol. 814, 2023. <https://doi.org/10.1016/j.cplett.2023.140358>
- [17] A. Modelli, L. Mussoni, "Rapid quantitative prediction of ionization energies and electron affinities of polycyclic aromatic hydrocarbons," *Chem. Phys.*, vol. 332, no. 2-3, pp. 367-374, 2007. <https://doi.org/10.1016/j.chemphys.2007.01.004>
- [1] K.-K. Mak, Y.-H. Wong, M. R. Pichika, "Artificial intelligence in drug discovery and development," *Drug Discovery and Evaluation: Safety and Pharmacokinetic Assays*, pp. 1-38, 2023. https://doi.org/10.1007/978-3-030-73317-9_92-1
- [2] G. Zhang, C. B. Musgrave, "Comparison of DFT methods for molecular orbital eigenvalue calculations," *Journal of Physical Chemistry A*, vol. 111, no. 8, pp. 1554-1561, 2007. <https://doi.org/10.1021/jp061633o>
- [3] "Computational quantum chemical study, drug-likeness and in silico cytotoxicity evaluation of some steroidal anti-inflammatory drugs." [Online]. Available: Google Search. Accessed: Mar. 14, 2025.
- [4] M. Miar, A. Shiroudi, K. Pourshamsian, A. R. Oliay, F. Hatamjafari, "Theoretical investigations on the HOMO-LUMO gap and global reactivity descriptor studies, natural bond orbital, and nucleus-independent chemical shifts analyses of 3-phenylbenzo[d]thiazole-2(3H)-imine and its para-substituted derivatives: solvent and substituent effects," *J. Chem. Res.*, vol. 45, no. 1-2, pp. 147-158, 2021. <https://doi.org/10.1177/1747519820932091>
- [5] E. F. Silva-Júnior, T. M. Aquino, J. X. Araújo-Júnior, "Quantum mechanical (QM) calculations applied to ADMET drug prediction: a review," *Curr. Drug Metab.*, vol. 18, no. 6, 2017. <https://doi.org/10.2174/1389200218666170316094514>
- [6] "The impact of topology on the open-shell characters of benzenoid polycyclic hydrocarbons: a data-driven study." [Online]. Available: Google Search. Accessed: Mar. 17, 2025.
- [7] "Application of local linear neuro-fuzzy model in prediction of mean arterial blood pressure time series." [Online]. Available: Google Search. Accessed: Mar. 17, 2025.