

کاربرد یادگیری ماشین در بررسی ویژگی‌های ساختاری و شیمیایی غشاهای پلیمری برای جذب کربن دی‌اکسید

فریده عبداللهی، آرش خسروی⁺*

گروه تحقیقاتی فناوری غشای پایدار (SMTRG)، دانشکده مهندسی نفت، گاز و پتروشیمی، دانشگاه خلیج فارس، بوشهر، ایران

احمد کشاورز

گروه تحقیقاتی پردازش سیگنال و اینترنت اشیا، موسسه تحقیقاتی فناوری اطلاعات و ارتباطات (ICT)، دانشکده مهندسی سیستم‌های هوشمند و علوم داده، دانشگاه خلیج فارس، بوشهر، ایران

سچکین کاراگز

بخش مهندسی شیمی، دانشکده مهندسی، دانشگاه قطر، دوحه، قطر

چکیده: در سال‌های اخیر، کاربرد فناوری‌های غشایی در جداسازی گازهای صنعتی و ذخیره کربن رو به گسترش است و هوش مصنوعی می‌تواند نقش بی‌بدیلی در کاهش هزینه‌ها و رفع موانع اجرایی توسعه این فناوری ایفا نماید. در این پژوهش، به منظور برطرف کردن محدودیت‌های مطالعات آزمایشگاهی و افزایش سرعت شناسایی غشاهای جدید و کارآمدتر در صنعت جداسازی گاز، یک مدل یادگیری ماشین با توجه به خواص فیزیکی و شیمیایی پلیمرها توسعه داده شده است. به طور خاص، الگوریتم جنگل تصادفی برای پیش‌بینی عملکرد غشاء از جمله تراوایی و گزینش پذیری برای جداسازی کربن دی‌اکسید/متان استفاده شده است. سپس روش توضیحات جمعی شیپلی برای تفسیر نتایج مورد استفاده قرار گرفته است. علاوه بر این، در این مطالعه به منظور مدل‌سازی پلیمرها با استفاده از یادگیری ماشین از روش اثر انگشت و توصیف‌گرهای مولکولی استفاده شده است. نتایج نشان داد که مساحت سطح قطبی گروه‌های تشکیل دهنده پلیمر، یکی از اصلی‌ترین پارامترهای مؤثر در عملکرد غشاء است. علاوه بر این، یافته‌ها نشان داد که گروه‌های قطبی در ساختار پلیمر، تأثیر منفی بر تراوایی دارند، در حالی که با گزینش‌پذیری همبستگی مثبتی دارند. تأثیر منفی حلقه‌های آروماتیک بر تراوایی غشاها یکی دیگر از یافته‌های این مطالعه بود.

کلیدواژه‌ها: یادگیری ماشین، جنگل تصادفی، غشاهای پلیمری، جداسازی گاز، توضیحات جمعی شیپلی

KEYWORDS: Machine Learning, Random Forest, Polymeric Membrane, Gas Separation, SHAP

+E-mail: arash.khosravi@pgu.ac.ir

* عهده‌دار مکاتبات

مقدمه

کاربردهای فرآیندهای غشایی با توسعه فناوری‌های بروز جهت حذف گازهای اسیدی از جریان گاز طبیعی و جذب کربن از گاز دودکش، در حال گسترش است [۱]. عملکرد یک غشاء برای جداسازی گاز، با دو فاکتور اصلی تراوایی و گزینش‌پذیری مشخص می‌شود. تراوایی (P_i) توسط قانون نفوذ فیک تعریف می‌شود، $J_i = P_i (\Delta p / l)$ ، که J_i شار نفوذ برای گاز i ، Δp اختلاف فشار در طول غشایی به ضخامت l است، و گزینش‌پذیری به صورت نسبت تراوایی دو گونه گاز تعریف می‌شود، $\alpha_{(A/B)} = P_A / P_B$ ، که در آن P_A و P_B به ترتیب تراوایی گونه‌های A و B هستند [۲، ۳]. از آنجا که یک غشای کاربردی برای جداسازی گاز باید تراوایی و گزینش‌پذیری بالایی داشته باشد، عملکرد غشای ساخته شده براساس دو پارامتر یاد شده با نمودار حد بالای رابسون^۱ نشان داده می‌شود [۴].

در سال‌های اخیر با پیشرفت فن‌آوری‌های طراحی مواد، تعداد زیادی غشاء جدید پیشنهاد شده‌اند که پتانسیل بالایی برای پیشی گرفتن از حد بالای رابسون دارند. به منظور افزایش انتقال گاز در پلیمرها، استراتژی‌های مختلفی مانند افزایش حجم آزاد^۲ پلیمر، اصلاح ساختار پلیمری و گروه‌های عاملی و افزودن بخش‌های مختلف به ستون فقرات پلیمری^۳ در آزمایشگاه‌ها مورد استفاده قرار می‌گیرد [۵، ۶]. با این حال، سنتز و بررسی عملکرد تعداد زیادی از غشاها در آزمایشگاه فرآیندی پرهزینه و زمان‌بر است. هوش مصنوعی و روش‌های یادگیری ماشین پتانسیل بسیار خوبی برای درک خواص و رفتار موادی که در ساخت غشاها، جدید پلیمری مورد استفاده قرار می‌گیرند، ارائه می‌کنند. این روش‌ها با برطرف کردن محدودیت‌های مربوط به هزینه و زمان، طراحی غشاها، جدید پلیمری برای جداسازی گاز را تسریع می‌کنند [۷-۹]. البته همچنان یکی از چالش‌های اصلی برای کاربرد گسترده یادگیری ماشین، دسترسی به مجموعه داده‌های متنوع به منظور ارائه هرچه بیشتر اطلاعات برای آموزش مدل است.

زو و همکاران^۴ با کمک یک مدل یادگیری ماشین، تراوایی گاز را برای ۶ گاز مختلف پیش‌بینی کرده‌اند. آنها از مجموعه داده‌های تجربی جمع‌آوری شده حاوی ۳۱۵ پلیمر مختلف استفاده کردند. در مطالعه آنها، از روش اثر انگشت^۵ برای استخراج ویژگی‌های پلیمر و معرفی پلیمرها به مدل یادگیری ماشین استفاده شد. در نهایت، یک الگوریتم رگرسیون

فرآیند گاوسی^۶ برای پیش‌بینی با دقت بالا استفاده شده‌است [۱۰]. نتایج آن‌ها نشان می‌دهد که تراوایی گاز و گزینش‌پذیری پیش‌بینی‌شده تقریباً مشابه مقادیر تجربی و کران بالایی منحنی رابسون در سال ۲۰۰۹ بود. مطالعه دیگری عملکرد جداسازی گاز در غشاها، پلی (۴-متیل-۱-پنتان) [۷] ترکیب شده با نانوذرات تیتانیوم دی‌اکسید^۸، روی اکسید^۹ و آلومینیوم اکسید^{۱۰} را بررسی کرده‌است. این مدل بر اساس یکی از الگوریتم‌های هوش مصنوعی به نام شبکه تابع پایه شعاعی^{۱۱} توسعه یافته‌است. داده‌های آن پژوهش نشان داد که نانوذرات اثر مثبت در عملکرد غشاء داشته‌اند [۱۱]. همچنین نتایج نشان داد که حداکثر مقادیر ضریب تعیین^{۱۲} برای گزینش‌پذیری کربن دی‌اکسید/نیتروژن^{۱۳} و کربن دی‌اکسید/متان^{۱۴} به ترتیب ۰/۸۶۹۷ و ۰/۷۰۲۸ بود. برنت و همکاران^{۱۵} یک مدل یادگیری ماشین بر پایه الگوریتم رگرسیون فرآیند گاوسی را به منظور پیش‌بینی تراوایی و گزینش‌پذیری ۶ گاز مختلف از جمله کربن دی‌اکسید/متان توسعه دادند. در این مطالعه، مقادیر تجربی اندازه‌گیری شده تراوایی مربوط به ۷۰۰ پلیمر برای ساخت مدل استفاده شد. علاوه بر این، آنها از روش اثر انگشت برای استخراج ویژگی و معرفی واحد تکرار شونده پلیمر به مدل استفاده کردند. در نهایت، برای تست دقت مدل در پیش‌بینی عملکرد غشاء، تعداد دو عدد از غشاها پیشنهاد شده به وسیله مدل در آزمایشگاه سنتز شد و نتایج آنها نشان داد که غشاها، پلیمری مذکور از کران بالایی رابسون برای کربن دی‌اکسید/متان فراتر رفتند [۸]. علاوه بر موارد بررسی شده در مطالعات پیشین، هوش مصنوعی می‌تواند به عنوان یک ابزار قدرتمند در بررسی پارامترهای ساختاری و شیمیایی پلیمرها و همچنین تشریح اثر پارامترهای مذکور بر جداسازی گازی مورد استفاده قرار بگیرد که این موضوع کمتر مورد توجه قرار گرفته است.

در این پژوهش، به منظور ارتقاء کارایی در پیش‌بینی ویژگی‌های تراوایی و گزینش‌پذیری در مخلوط دوجزئی کربن دی‌اکسید/متان در غشاها، پلیمری، از یک رویکرد نوین مبتنی بر یادگیری ماشین بهره گرفته شده‌است. این رویکرد مبتنی بر الگوریتم جنگل تصادفی پیشرفته‌ای است که به دلیل توانایی بالا در تجزیه و تحلیل همبستگی‌های پیچیده بین ورودی‌ها و خروجی‌ها، انتخاب مناسبی برای مطالعه حساسیت‌ها و تأثیرات مختلف متغیرها محسوب می‌شود. علاوه بر این، در این تحقیق، یک رویکرد نوین و مبتکر به منظور شناسایی

(۱) Robeson upper bound

(۳) Polymer backbone

(۵) Fingerprinting

(۷) Poly(4-methyl-1-pentane) (PMP)

(۹) ZnO

(۱۱) Radial basis function (RBF) network

(۱۳) CO₂/N₂

(۱۵) Barnett et al.

(۲) Free Volume

(۴) Zhu et al.

(۶) Gaussian process regression (GPR) algorithm

(۸) TiO₂(۱۰) Al₂O₃

(۱۲) R2

(۱۴) CO₂/CH₄

هایپارامترها برای تعمیم عملکرد مدل برای داده‌های آزمون استفاده می‌شود [۱۴]. مطابق با این روش، سه هایپارامتر از جمله حداکثر تعداد ویژگی‌ها در جستجوی بهترین نقطه تقسیم (ریشه دوم^۸)، تعداد بهینه درختان تصمیم‌گیری در جنگل (۵۰۰) و حداکثر عمق درخت (۲۰) بهینه شده‌است. هایپارامترهای باقیمانده با مقادیر پیش فرض از بسته سایکیت لرن^۹ تنظیم شدند [۱۵].

روش توضیحات جمعی شیپلی

توضیحات جمعی شیپلی روشی است که برای تفسیر نتایج یک الگوریتم یادگیری ماشین استفاده می‌شود. این روش، اهمیت ویژگی‌های ورودی و سهم آن‌ها در پیش‌بینی مدل یادگیری ماشین را با کنار گذاشتن یک به یک آن‌ها از داده‌های آموزش و بررسی تغییر در پیش‌بینی مطابق با معادله ۱ ارزیابی می‌کند [۱۶].

$$\phi_i = \frac{1}{|N|!} \sum_{S \subseteq N} |S|! (|N| - |S| - 1)! [f(S \cup \{i\}) - f(S)] \quad (1)$$

در معادله ۱، سهم نهایی یا مقدار شیپلی ویژگی ϕ_i ، به عنوان متوسط سهم آن در کلیه جایگشت‌های قابل تصور از یک مجموعه ویژگی محاسبه می‌شود. $f(S)$ خروجی مدل یادگیری ماشین را بر اساس مجموعه S از ویژگی‌ها نشان می‌دهد و N تعداد کل ویژگی‌ها است [۱۷].

نتیجه‌ها و بحث

مجموعه داده، ارائه ویژگی‌ها^{۱۰} و انتخاب ویژگی‌ها^{۱۱}

برای ایجاد یک مدل یادگیری ماشین برای پیش‌بینی خصوصیات جداسازی گازی پلیمرها، یک مجموعه داده متشکل از ۵۲۰ پلیمر به صورت دستی جمع‌آوری و با مجموعه داده انجمن غشایی استرالزی^{۱۲} ادغام شد [۱۸]. مجموعه داده مورد استفاده هموپلیمرهای واقعی موجود در مطالعات پیشین از جمله پلیمرهای میکرو متخلخل ذاتی^{۱۳}، با تراوایی گاز اندازه‌گیری شده برای گازهای کربن دی‌اکسید و متان هستند. شکل (ب-۱) توزیع تراوایی داده‌های مربوط به کران بالای رابسون ۲۰۰۸ و ۲۰۱۵ را نشان می‌دهد. همانطور که در شکل (ب-۱) نشان داده شده‌است، تعدادی از پلیمرها دارای تراوایی بالایی برای گاز کربن دی‌اکسید هستند، بیش از Barrers^{۱۴}، که پتانسیل بالایی را برای پیشی گرفتن از حد بالای رابسون

و توصیف دقیق‌تر پلیمرها مورد استفاده قرار گرفته‌است. این رویکرد با استفاده از دو روش به صورت همزمان، یعنی اثر انگشت و توصیف‌گر مولکولی^۱، توانسته‌است به طور دقیق و کامل ویژگی‌های مهم پلیمرها از جمله توپولوژی، بار سطحی، و خواص قطبی و غیره را تحت زبانی قابل فهم برای مدل‌ها تبیین نماید. این تفسیر دقیق می‌تواند به توسعه‌ی بهترین مدل‌ها و مواد پلیمری نوآورانه کمک موثری نماید. در پایان، با استفاده از مقادیر توضیحات جمعی شیپلی^۲، تفسیر نهایی مدل انجام می‌پذیرد و ویژگی‌های ساختاری و شیمیایی پلیمرها به طور دقیق مورد ارزیابی قرار می‌گیرند. این رویکرد نوآورانه این امکان را می‌دهد که اهمیت و تأثیر ویژگی‌های مختلف پلیمرها را به شکل بهتری درک کنیم و این اطلاعات به منظور پیشرفت در طراحی و تولید پلیمرهای مصنوعی و نانوکامپوزیت‌های پیشرفته مورد استفاده قرار گیرد.

روش‌ها

اصول اساسی رگرسیون جنگل تصادفی

الگوریتم جنگل تصادفی یکی از محبوب‌ترین و قدرتمندترین الگوریتم‌هایی است که به طور گسترده در زمینه‌های مختلف برای رگرسیون و طبقه‌بندی استفاده می‌شود [۱۲]. مدل رگرسیون جنگل تصادفی با ترکیب چندین درخت تصمیم‌گیری ساخته شده‌است و هر درخت تصمیم‌گیری با انتخاب نمونه‌های فرعی از مجموعه داده با تکنیک بوت استرپ^۳ ایجاد می‌شود. نمونه‌گیری بوت استرپ راهی برای انتخاب نمونه‌های آموزش تصادفی از یک مجموعه داده اصلی با جایگزینی است، به این معنی که وقتی از داده‌ها برای قرار گرفتن در زیر مجموعه استفاده می‌شود، هنوز هم برای انتخاب شدن در بسیاری از زمان‌های دیگر در دسترس هستند [۱۳]. پیش‌بینی نهایی مدل جنگل تصادفی با در نظر گرفتن میانگین نتایج پیش‌بینی شده درختان تصمیم‌گیری که در زیر مجموعه‌های مختلف و با روش بوت استرپ ایجاد شده‌اند، محاسبه می‌شود.

برای جلوگیری از مشکل بیش‌برازش^۴ در مدل جنگل تصادفی، از یک جستجوی شبکه‌ای^۵ با روش اعتبارسنجی متقابل^۶ استفاده شد. در این روش، ترکیب متفاوت مقادیر هایپارامتر از پیش تعریف شده‌است و یک اعتبارسنجی متقابل پنج فولد^۷ برای ارزیابی بهترین ترکیب

(۱) Molecular descriptors

(۳) Bootstrap

(۵) Grid search

(۷) Fivefold

(۹) scikit-learn

(۱۱) Feature selection

(۱۳) Polymers of intrinsic microporosity (PIMs)

(۲) Shapley Additive explanation (SHAP)

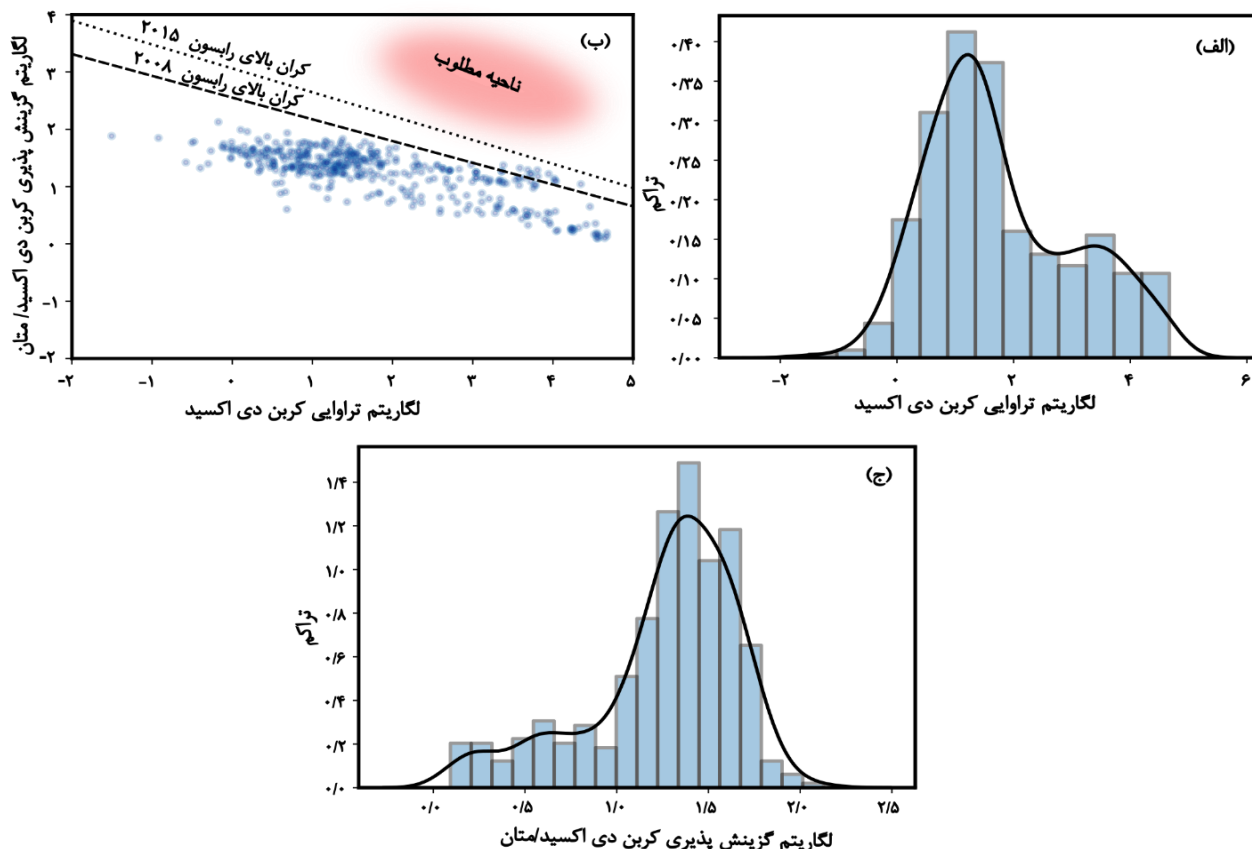
(۴) Overfitting

(۶) Cross-validation

(۸) Square root

(۱۰) Feature representation

(۱۲) Membrane Society of Australasia (MSA) dataset



شکل ۱ - (الف) توزیع تراوایی کربن دی اکسید، نمودار چگالی (ب) توزیع مجموعه داده (ج) توزیع گزینش پذیری کربن دی اکسید/متان، نمودار چگالی.

در این مطالعه، ترکیبی از دو تکنیک مورد اشاره به منظور استخراج هر چه بهتر اطلاعات از پلیمرها و ساخت یک مدل یادگیری ماشین دقیق استفاده شد [۸، ۱۰].

در مورد توصیف‌گرهای مولکولی، از بسته RDKit برای ایجاد توصیف‌کننده‌ها بر مبنای واحد تکرار شونده پلیمر استفاده شده است. این توصیف‌کننده‌های مولکولی می‌توانند اطلاعات فیزیکی و شیمیایی مربوط به یک پلیمر مانند بار، شاخص‌های توپولوژیکی، گروه‌های عاملی و غیره را ارائه دهند، که به طور گسترده‌ای در توسعه QSAR^۴ و مدل‌های یادگیری ماشین استفاده می‌شوند [۹، ۲۰].

در این مطالعه، اثر انگشت روش دیگری است که توصیف‌کننده‌های مولکولی را برای کمک به مدل یادگیری ماشین جهت دسترسی به طیف گسترده‌ای از اطلاعات در مورد یک پلیمر خاص همراه خواهد کرد. در همین راستا، اثر انگشت مورگان^۵ از بسته RDKit با شعاع ۲ و ۱۰۲۴ بیت تولید شده است. با استفاده از این روش مبتنی بر خواص توپولوژیکی، هر زیر ساخت مولکولی حاوی تعداد

و استفاده در کاربردهای صنعتی ارائه می‌دهد. شکل (الف-۱) برای توزیع تراوایی کربن دی اکسید و شکل (ج-۱) برای توزیع گزینش پذیری کربن دی اکسید/متان نشان می‌دهد که توزیع دو-مدلی^۱ در هر شکل وجود دارد به این معنی که دو گروه اصلی تراوایی و گزینش پذیری وجود دارند که در محدوده ۱ تا ۳ برای لگاریتم تراوایی و ۰/۵ تا ۱/۵ برای لگاریتم گزینش پذیری طبقه بندی می‌شوند. این ارقام به وضوح نشان می‌دهند که تعداد قابل توجهی از نقاط داده در مجموعه داده توسعه یافته، پلیمرهای معمولی با تراوایی کمتر از ۱۰۰ Barrers هستند. به منظور معرفی پلیمر به مدل، تمام ساختارهای واحد تکرار کننده پلیمر توسط سیستم اسمایلز^۲ نشان داده شده است، که یک نماد خطی برای توصیف ساختار گونه‌های شیمیایی با استفاده از رشته‌های کوتاه ASCII^۳ [۸، ۱۹] است. یکی از اصلی‌ترین چالش‌ها در پیش بینی تراوایی گازها در غشاهای پلیمری، یافتن روشی مناسب برای ارائه ویژگی‌ها است. تکنیک‌های متنوع بیان ویژگی وجود دارند که برای مثال می‌توان به اثر انگشت و توصیف‌گرهای مولکولی اشاره نمود.

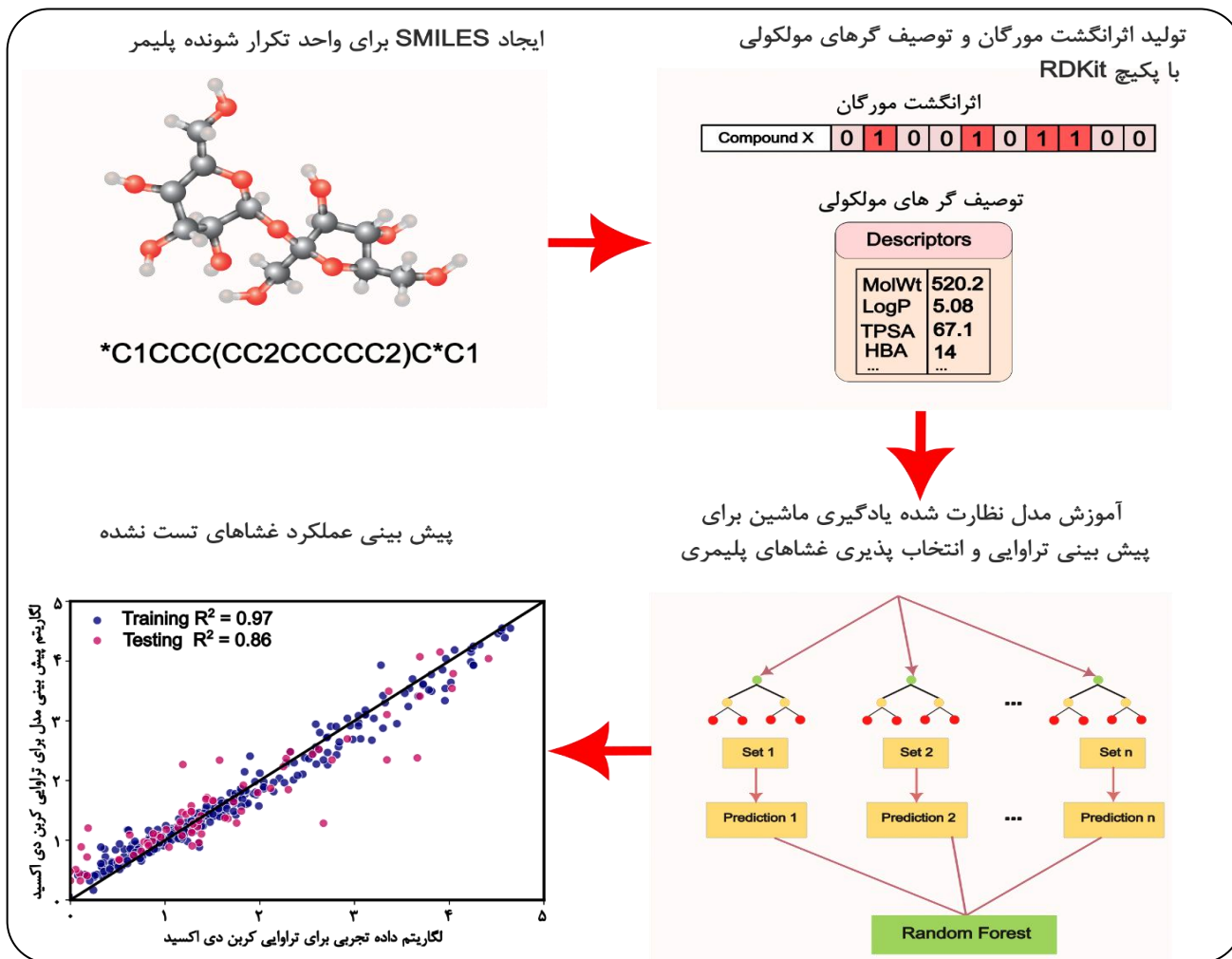
(۱) Bimodal

(۳) American Standard Code for Information Interchange

(۵) Morgan Fingerprint (MF)

(۲) Simplified molecular-input line-entry system (SMILES) notations

(۴) Quantitative structure–property relationship



شکل ۲ - شماتیک فرآیند یادگیری ماشین برای پیش‌بینی تراوایی و گزینش پذیری غشاهای پلیمری.

$$\text{Var}(X) = p(1-p) \quad (2)$$

سرانجام، ۶۷ ویژگی باقیمانده به عنوان مؤثرترین عوامل بر عملکرد غشاء (تراوایی و گزینش‌پذیری) استفاده شدند. شماتیک فرآیند در شکل ۲ نشان داده شده‌است.

عملکرد مدل یادگیری ماشین برای پیش‌بینی تراوایی و گزینش‌پذیری گاز

این مطالعه از الگوریتم جنگلی تصادفی برای کشف عوامل اصلی فیزیکی و ساختارهای شیمیایی مؤثر بر عملکرد غشاء استفاده می‌کند. در اینجا از روش اثرانگشت و توصیف گر مولکولی به عنوان ویژگی‌های ورودی مدل یادگیری ماشین استفاده می‌شود. مجموعه داده به طور تصادفی به مجموعه‌های آموزش ۸۰ درصد و تست ۲۰ درصد تقسیم شد. نتایج مدل به ترتیب در شکل (ب-۳) و (د-۳) برای تراوایی و

مشخصی از پیوندهای شیمیایی (در این حالت ۲) است و سپس در یک بردار ۱۰۲۴ بیتی ریخته می‌شود. هر بیت در این مطالعه یک ویژگی منحصر به فرد در نظر گرفته می‌شود و اطلاعات کافی در مورد تأثیر گروه‌های عاملی و اتصال آنها، به رفتار جداسازی گازی پلیمرها ارائه می‌دهد.

در مورد انتخاب ویژگی‌ها، در این مطالعه ۱۰۴۹ ویژگی از جمله اثرانگشت و توصیف‌گرهای مولکولی ایجاد شده‌است. به منظور کاهش افزونگی، ویژگی‌هایی با واریانس کم با استفاده از روش واریانس متغیرهای تصادفی برنولی (معادله ۲) مشخص شده و سپس حذف شدند [۲۱]. در معادله ۲، X یک ویژگی خاص است، و p احتمال حضور یک ویژگی خاص در تمام نمونه‌ها است. در این مطالعه، مقدار p برابر 0.8 است به این معنی که ویژگی‌هایی با مقادیر ثابت در بیش از ۸۰٪ نمونه‌ها حذف می‌شوند.

(۱) Redundancy

جدول ۱ - لیست توصیف‌گرهای موجود در RDKit و تعاریف آن‌ها.

نام توصیف‌گر	توضیح
TPSA	مساحت سطح قطبی گروه‌های تشکیل دهنده پلیمر ^۱
lipinskiHBA	تعداد پذیرندگان پیوند هیدروژن (H) لپینسکی ^۲ برای یک مولکول
NumRotatableBonds	تعداد پیوندهای قابل چرخش برای یک مولکول
NumHBA	تعداد پذیرندگان پیوند هیدروژن (H) برای یک مولکول
CrippenClogP	محاسبه بر مبنای اتم ضریب تقسیم (log P)
NumHeteroatoms	تعداد هترواتم‌ها برای یک مولکول
NumAromaticRings	تعداد حلقه‌های آروماتیک برای یک مولکول

تراوایی و افزایش گزینش‌پذیری غشاء می‌شود. [۲۴، ۲۵]. از طرفی دیگر، گروه‌های قطبی برای عبور کربن‌دی‌اکسید که مولکولی قطبی است ترجیح بالاتری نسبت به متان یا نیتروژن غیرقطبی ایجاد می‌کنند که بر گزینش‌پذیری اثری مثبت دارد. سایر ویژگی‌های ساختاری شناخته شده که تأثیر بسزایی در جداسازی کربن دی‌اکسید دارند عبارتند از: پیوند هیدروژنی، گروه‌های عاملی قطبی مانند گروه آمینه^۳ و گروه هیدروکسیل^۴، تعداد هترواتم‌ها مانند نیتروژن و اکسیژن، سطح مولکول و سختی زنجیره پلیمری که می‌توان با نگاه کردن به حلقه‌های سخت در ستون فقرات پلیمری تشخیص داده شوند [۲۶-۲۸]. با توجه به سختی پلیمر، معرفی حلقه‌های آروماتیک می‌تواند انعطاف‌پذیری پلیمر را کاهش داده یا سختی آن را افزایش دهد، که منجر به گزینش‌پذیری بالاتر خواهد شد [۲۹، ۳۰].

با توجه به اثر انگشت مورگان، برخی از زیرساختارها می‌توانند اطلاعات جالبی در مورد اثر خصوصیات فیزیکی پلیمرها بر جداسازی کربن دی‌اکسید ارائه دهند. همانطور که اشاره شد، تأثیر زیرساختارهای مهم در شکل ۳ بیان شده‌است. در حالیکه خصوصیات مورد اشاره در شکل ۴ توصیف شده‌اند. همانطور که از شکل (الف-۳) مشاهده می‌شود، مهم‌ترین زیرساختار برای تراوایی کربن دی‌اکسید، زیرساختار شماره ۸۰۰ می‌باشد که نشان دهنده تأثیر حلقه‌های آروماتیک است. بدیهی است که ترکیبات آروماتیک موجود در ستون فقرات پلیمر می‌توانند انعطاف‌پذیری زنجیره اصلی را کاهش دهند و باعث کاهش تراوایی گردند [۳۱]. به طور مشابه، زیرساختار ۲۳۶ نیز که نمادی از حضور حلقه‌های آروماتیک در ستون فقرات پلیمر است در زمره ده پارامتر مهم قرار گرفته است که بر تراوایی غشاء نیز اثر معکوس دارد. شکل ۳ نشان می‌دهد که برخی پیوندهای قطبی مانند اکسیژن‌های با پیوند دوگانه^۵ (زیرساختار ۶۵۱) از آنجا که بر انعطاف‌پذیری زنجیره و مسیرهای عبور اثر می‌گذارند با تراوایی رابطه معکوس دارند. اما، این مساله بدلیل اینکه کربن دی‌اکسید قطر سینتیکی کمتری

گزینش‌پذیری نشان داده شده‌است. یافته‌ها نشان می‌دهد که مدل جنگل تصادفی برای تراوایی می‌تواند R^2 در حدود ۰/۹۷ برای مجموعه آموزش و ۰/۸۶ برای تست و برای گزینش‌پذیری مقداری در حدود ۰/۹۸ برای مجموعه آموزش و ۰/۸۴ برای تست تولید کند.

تفسیر مدل یادگیری ماشین

مدل‌های یادگیری ماشین معمولاً جعبه‌های سیاه محسوب می‌شوند زیرا تفسیر آنها ساده نیست. با این حال، این پژوهش سعی می‌کند تا حد امکان اطلاعات لازم را برای یافتن ویژگی‌های شیمیایی اساسی که تأثیر مهمی در تراوایی و گزینش‌پذیری غشای پلیمری برای جداسازی کربن دی‌اکسید/متان دارند، استخراج کند. بنابراین، اثر عوامل فیزیکی از طریق بررسی دقیق با استفاده از مقادیر توضیحات جمعی شیپلی آشکار می‌شود. اگر ویژگی‌ها دارای مقادیر توضیحات جمعی شیپلی مثبت باشند، آن ویژگی‌ها رابطه مستقیم با تراوایی و گزینش‌پذیری دارند. اما اگر ویژگی‌ها دارای مقادیر توضیحات جمعی شیپلی منفی باشند، رابطه آن‌ها با تراوایی و گزینش‌پذیری معکوس است [۲۲]. شکل (الف-۳) و (ج-۳) ده پارامتر مهم و تأثیرگذار بر روی تراوایی کربن دی‌اکسید و گزینش‌پذیری کربن دی‌اکسید/متان را از بالا به پایین به ترتیب نشان می‌دهند. نام و تعریف کامل توصیف‌گرها در جدول ۱ گزارش شده‌است.

همانطور که از شکل ۳ (الف و ج) مشاهده می‌شود، TPSA دومین عامل مؤثر برای تراوایی کربن دی‌اکسید و مهم‌ترین پارامتر در پیش‌بینی گزینش‌پذیری کربن دی‌اکسید/متان است. TPSA به عنوان مساحت سطح گروه‌های قطبی، در درجه اول اکسیژن و نیتروژن، و اتم‌های هیدروژن متصل به آن‌ها تعریف شده‌است [۲۳]. افزایش تعداد گروه‌های قطبی می‌تواند نیروهای بین مولکولی را که باعث جذب و انسجام بین زنجیره‌ای می‌شوند افزایش دهد. در نتیجه، حجم آزاد پلیمرها محدود می‌شود که منجر به کاهش

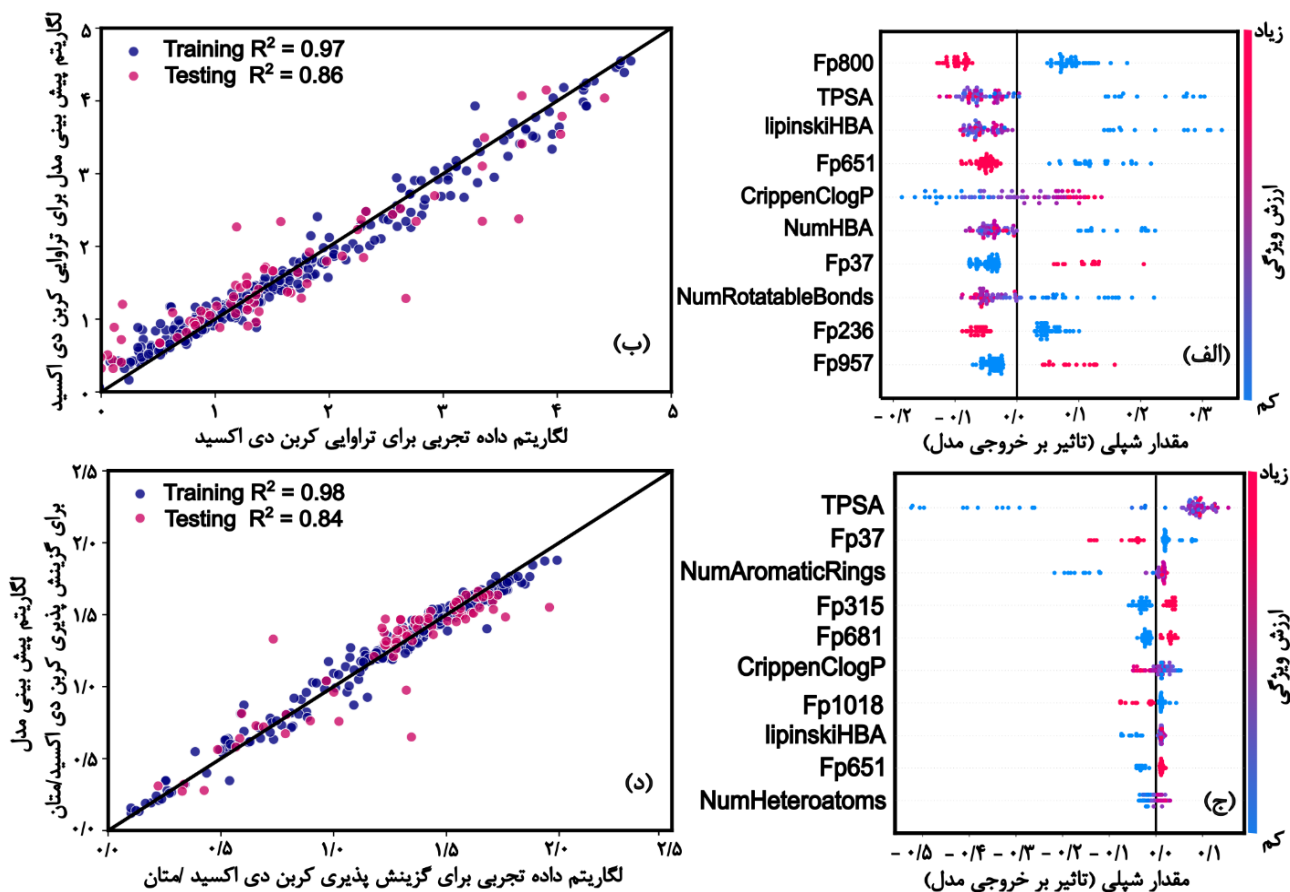
(۱) Topological polar surface area (TPSA)

(۳) NH₂

(۵) Double-bonded oxygens

(۲) LipinskiHBA

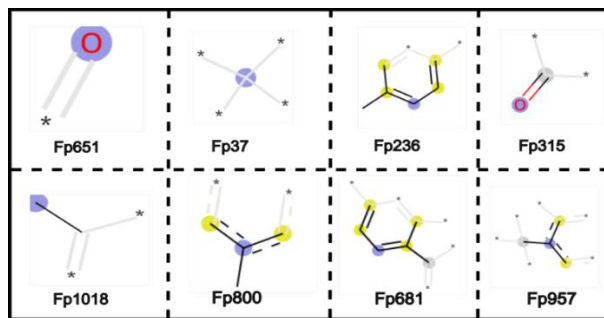
(۴) OH



شکل ۳ - عملکرد مدل‌های یادگیری ماشین برای (الف) ترتیب اهمیت ویژگی برای تراوایی، (ب) نتایج تراوایی کربن دی اکسید، (ج) ترتیب اهمیت ویژگی برای گزینش پذیری کربن دی اکسید/متان، (د) نتایج گزینش پذیری کربن دی اکسید/متان.

زیرساختارهای مورد اشاره به تأثیر آن‌ها بر بهبود انعطاف پذیری زنجیره‌ها بر می‌گردد. در مورد زیرساختار ۹۵۷، اگر چه ترکیب آروماتیک معمولاً باعث کاهش تراوایی می‌گردد اما تعداد اتصالات آن‌ها با کربن‌های چهار پیوندی از آنجا که مشابه مفصل عمل می‌کنند می‌تواند سختی زنجیره را کاهش دهد. در نتیجه، کسر حجم آزاد^۱ افزایش می‌یابد که باعث افزایش ضریب نفوذ و در نهایت تراوایی می‌شود. این الگوی گسترده قبلاً برای تعدادی از ساختارهای پلیمری از جمله پلی آمیدهای^۲ مختلف مشاهده شده است [۳۴-۳۶].

همانطور که از شکل (ج-۳) مشاهده می‌شود، اگرچه وجود زیرساختار شماره ۳۷ در ستون فقرات پلیمر باعث بهبود تراوایی می‌شود، اما این زیرساختار با گزینش پذیری رابطه معکوس دارد. زیرساختار ۱۰۱۸ مانند زیرساختار ۳۷، به خواص پلیمرهای با بیش از دو پیوند اتصال در واحد تکرار شونده اشاره دارد. علاوه بر این، زیر ساختار ۱۰۱۸ اثر معکوس در گزینش پذیری غشاء دارد. نتایج همچنین نشان داد که اکسیژن‌های با پیوند دوگانه،



شکل ۴ - زیرساختارهایی که بیشترین تأثیر را بر تراوایی کربن دی اکسید و گزینش پذیری کربن دی اکسید/متان بر اساس اثر انگشت مورگان دارند.

به نسبت نیتروژن و متان دارد منجر به افزایش گزینش پذیری می‌گردد [۳۲، ۳۳]. در مقابل، نتایج نشان می‌دهد که یک اتم کربن مرکزی با چهار پیوند (زیر ساختار ۳۷) و یک کربن مرکزی متصل به یک حلقه آلیفاتیک و آروماتیک (زیر ساختار ۹۵۷) اثر مثبت بر تراوایی کربن دی اکسید دارند. توضیح احتمالی برای اثرگذاری

(۱) Fractional Free Volume (FFV)

(۲) Polyimides

و تفسیر نتایج بهره گرفته شده است. نتایج نشان داده است که توصیف کننده های شیمیایی مانند بار سطحی و زیرساختارهای پلیمری نقش مهمی در عملکرد جداسازی دارند. نتایج همچنین نشان می دهند که ویژگی های فیزیکی و شیمیایی پلیمر مانند TPSA، تعداد پذیرندگان پیوند هیدروژن (H) لیپینسکی برای یک مولکول، تعداد پیوندهای قابل چرخش، تعداد پذیرندگان پیوند هیدروژن (H)، محاسبه بر مبنای اتم ضریب تقسیم (log P)، تعداد هترواتم ها، تعداد حلقه های آروماتیک و زیرساختارهای پلیمری شماره ۸۰۰، ۶۵۱، ۳۷، ۲۳۶، ۳۱۵، ۱۰۱۸، ۶۸۱ و ۹۵۷ از جمله مهم ترین پارامترهای تاثیرگذار بر پیش بینی تراوایی و گزینش پذیری غشاهای پلیمری برای جداسازی گازی هستند. ویژگی های مذکور عمدتاً به دلیل تاثیر بر انعطاف پذیری زنجیره پلیمری و ایجاد تمایل به سمت جذب گاز کربن دی اکسید نسبت به متان از اهمیت زیادی برخوردار هستند.

زیر ساختارهای (۳۱۵، ۶۵۱) برخلاف اثر منفی که در تراوایی غشاء دارند با گزینش پذیری ارتباط مستقیم دارند. توضیح احتمالی برای این نتایج ممکن است به ترجیح گذاری متقابل^۱ بین تراوایی و گزینش پذیری بسیاری از پلیمرهای مورد استفاده در این مطالعه باشد [۳۲]. در حالی که حلقه آروماتیک در بسیاری از موارد تاثیر منفی بر تراوایی غشاء دارد، زیر ساختار ۶۸۱ نشان داده است که تاثیر آن بر گزینش پذیری غشاء تا حدودی مثبت است [۳۷، ۳۸].

نتیجه گیری

این پژوهش براساس روشی مبتنی بر یادگیری ماشین به شناسایی و مطالعه زیرساختارهای مهم مواد پلیمری و ویژگی های فیزیکی و شیمیایی آن ها در طراحی غشاهای پلیمری جدید با عملکرد بالا در جداسازی گازی می پردازد. مدل های یادگیری ماشین با شاخص عملکرد R^2 بالا برای تراوایی کربن دی اکسید و گزینش پذیری کربن دی اکسید/متان توسعه یافته اند. برای توصیف پلیمرها و استخراج ویژگی های قابل استفاده در مدل های یادگیری ماشین، از دو روش اثرانگشت مورگان و توصیف گرهای مولکولی استفاده شده است. همچنین از روش توضیحات جمعی شیپلی برای تحلیل

تاریخ دریافت: ۱۴۰۳/۰۵/۰۱؛ تاریخ پذیرش: ۱۴۰۳/۰۸/۰۷

مراجع

- [1] Zhang Y., Sunarso J., Liu S., Wang R., [Current status and Development of Membranes for CO₂/CH₄ Separation: A Review](#), *International Journal of Greenhouse Gas Control*, **12**: 84-107 (2013).
- [2] Baker R.W., [Membrane Technology and Applications](#), John Wiley & Sons, (2012).
- [3] Abdollahi F., Hashemifard S., Khosravi A., Matsuura T., [Heat and Mass Transfer Modeling of an Energy Efficient Hybrid Membrane-Based Air Conditioning System for Humid Climates](#), *Journal of Membrane Science*, **625**: 119179 (2021).
- [4] Robeson L.M., [The Upper Bound Revisited](#), *Journal of membrane science*, **320(1-2)**: 390-400 (2008).
- [5] Nobakht D., Abedini R., [Improved gas Separation Performance of Pebax@1657 Membrane Modified by Poly-Alcoholic Compounds](#), *Journal of Environmental Chemical Engineering*, **10(3)**: 107568 (2022).
- [6] Kong J., Liu J., Jia P., Qi N., Chen Z., Xu S., Li N., [Synergistic Effect of Thermal Crosslinking and Thermal Rearrangement on Free Volume and Gas Separation Properties of 6FDA Based Polyimide Membranes Studied by Positron Annihilation](#), *Journal of Membrane Science*, **645**: 120163 (2022).

(۱) Trade-off

- [7] Guan J., Huang T., Liu W., Feng F., Japip S., Li J., Wang X., Zhang S., [Design and Prediction of Metal Organic Framework-Based Mixed Matrix Membranes for CO₂ Capture Via Machine Learning](#), *Cell Reports Physical Science*, **3(5)**: 100864 (2022).
- [8] Barnett J.W., Bilchak C.R., Wang Y., Benicewicz B.C., Murdock L.A., Bereau T., Kumar S.K., [Designing Exceptional Gas-Separation Polymer Membranes Using Machine Learning](#), *Science advances*, **6(20)**: eaaz4301 (2020).
- [9] Tao L., Chen G., Li Y., [Machine Learning Discovery of High-Temperature Polymers](#), *Patterns*, **2(4)**: 100225 (2021).
- [10] Zhu G., Kim C., Chandrasekarn A., Everett J.D., Ramprasad R., Lively R.P., [Polymer Genome-Based Prediction of Gas Permeabilities in Polymers](#), *Journal of Polymer Engineering*, **40(6)**: 451-457 (2020).
- [11] Alihosseini A., Hedayati Moghaddam A., [Permeability and Selectivity Prediction of Poly \(4-Methyl 1-Pentane\) Membrane Modified by Nanoparticles in Gas Separation Through Artificial Intelligent Systems](#), *Polyolefins Journal*, **7(2)**: 91-98 (2020).
- [12] Lipu M.H., Ayob A., Saad M., Hussain A., Hannan M., Faisal M., [State of Charge Estimation for Lithium-ion Battery Based on Random Forests Technique with Gravitational Search Algorithm](#), *2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, *IEEE*, 45-50 (2018).
- [13] Efron B., Tibshirani R., [Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy](#), *Statistical science*, 54-75 (1986).
- [14] Boehmke B., Greenwell B., [Hands-on Machine Learning with R](#), *Chapman and Hall/CRC*, (2019).
- [15] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., [Scikit-Learn: Machine Learning in Python](#), *the Journal of machine Learning research*, **12**: 2825-2830 (2011).
- [16] Meng Y., Yang N., Qian Z., Zhang G., [What Makes an Online Review More Helpful: An Interpretation Framework Using XGBoost and SHAP Values](#), *Journal of Theoretical and Applied Electronic Commerce Research*, **16(3)**: 466-490 (2020).
- [17] Rodríguez-Pérez R., Bajorath J., [Interpretation of Machine Learning Models Using Shapley Values: Application to Compound Potency and Multi-Target Activity Predictions](#), *Journal of computer-aided molecular design*, **34(10)**: 1013-1026 (2020).
- [18] Thornton A., BD F., Robeson L., [Polymer Gas Separation Membrane Database](#), (2012), Online.
- [19] Weininger D., [SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules](#), *Journal of chemical information and computer sciences*, **28(1)**: 31-36 (1988).
- [20] Landrum G., [RDKit: A Software Suite for Cheminformatics, Computational Chemistry, and Predictive Modeling](#), *Greg Landrum*, (2013).

- [21] Nedelman J., Wallenius T., [Bernoulli Trials, Poisson Trials, Surprising Variances, and Jensen's Inequality](#), *The American Statistician*, **40(4)**: 286-289 (1986).
- [22] Lubo-Robles D., Devegowda D., Jayaram V., Bedle H., Marfurt K.J., Pranter M.J., [Machine Learning Model Interpretability Using SHAP Values: Application to a Seismic Facies Classification Task](#), *SEG International Exposition and Annual Meeting, OnePetro*, (2020).
- [23] Fernandes J., Gattass C.R., [Topological Polar Surface Area Defines Substrate Transport by Multidrug Resistance Associated Protein 1 \(MRP1/ABCC1\)](#), *Journal of medicinal chemistry*, **52(4)**: 1214-1218 (2009).
- [24] McKeen L.W., [Permeability Properties of Plastics and Elastomers](#), *William Andrew*, (2016).
- [25] Recio R., Lozano A.E., Prádanos P., Marcos A., Tejerina F., Hernández A., [Effect of Fractional Free Volume and Tg on Gas Separation Through Membranes Made with Different Glassy Polymers](#), *Journal of applied polymer science*, **107(2)**: 1039-1046 (2008).
- [26] Hamouda S.B., Nguyen Q.T., Langevin D., Roudesli S., [Poly\(Vinylalcohol\)/Poly\(Ethylene glycol\) /Poly\(Ethyleneimine\) Blend Membranes - Structure and CO2 Facilitated Transport](#), *Comptes Rendus Chimie*, **13(3)**: 372-379 (2010).
- [27] Bicerano J., [Prediction of Polymer Properties](#), *cRc Press*, (2002).
- [28] Chan D., Chien J.C., Axpe E., Blankemeier L., Baker S.W., Swaminathan S., Piunova V.A., Zubarev D.Y., Maikawa C.L., Grosskopf A.K., [Combinatorial Polyacrylamide Hydrogels for Preventing Biofouling on Implantable Biosensors](#), *Advanced Materials*, 2109764 (2022).
- [29] Mulder M., Mulder J., [Basic Principles of Membrane Technology](#), *Springer science & business media*, (1996).
- [30] Kadirkhan F., Goh P.S., Ismail A.F., Wan Mustapa W.N.F., Halim M.H.M., Soh W.K., Yeo S.Y., [Recent Advances of Polymeric Membranes in Tackling Plasticization and Aging for Practical Industrial CO2/CH4 Applications—A Review](#), *Membranes*, **12(1)**: 71 (2022).
- [31] Guzmán-Lucero D., Froylán Palomeque-Santiago J., Camacho-Zúñiga C., Ruiz-Treviño F.A., Guzmán J., Galicia-Aguilar A., Aguilar-Lugo C., [Gas Permeation Properties of Soluble Aromatic Polyimides Based on 4-Fluoro-4,4'-Diaminotriphenylmethane](#), *Materials*, **8(4)**: 1951-1965 (2015).
- [32] Liu J., Zhang S., Jiang D.-e., Doherty C.M., Hill A.J., Cheng C., Park H.B., Lin H., [Highly Polar But Amorphous Polymers with Robust Membrane CO2/N2 Separation Performance](#), *Joule*, **3(8)**: 1881-1894 (2019).
- [33] Yang J., Tao L., He J., McCutcheon J., Li Y., [Discovery of Innovative Polymers for Next-Generation Gas-Separation Membranes using Interpretable Machine Learning](#), (2021).
- [34] Sanders D.F., Smith Z.P., Guo R., Robeson L.M., McGrath J.E., Paul D.R., Freeman B.D., [Energy-Efficient Polymeric Gas Separation Membranes for a Sustainable Future: A Review](#), *Polymer*, **54(18)**: 4729-4761 (2013).

- [35] Matteucci S., Yampolskii Y., Freeman B.D., Pinnau I., [Transport of Gases and Vapors in Glassy and Rubbery Polymers](#), *Materials science of membranes for gas and vapor separation*, **1**: 1-2 (2006).
- [36] Wu A.X., Drayton J.A., Rodriguez K.M., Qian Q., Lin S., Smith Z.P., [Influence of Aliphatic and Aromatic Fluorine Groups on Gas Permeability and Morphology of Fluorinated Polyimide Films](#), *Macromolecules*, **53(13)**: 5085-5095 (2020).
- [37] Han Y., Ho W.W., [Polymeric Membranes for CO₂ Separation and Capture](#), *Journal of Membrane Science*, **628**: 119244 (2021).
- [38] Hashemifard S.A., Khosravi A., Abdollahi F., Alihemati Z., Rezaee M., [Chapter 11 - Synthetic polymeric Membranes for Gas and Vapor Separations](#), *Synthetic Polymeric Membranes for Advanced Water Treatment, Gas Separation, and Energy Sustainability*, *Elsevier*, 217-272 (2020).