

ارزیابی تطبیقی و بهینه‌سازی مدل‌های یادگیری ماشین در تخمین گرانروی نفت مرده

محمد حاجی‌نوامری^۱، رفعت پارسایی*^۲، مسعود ریاضی*^۳، جعفر قاجار^۲، سلیمان حسن^۳، علی صفایی^۴ و پیام ستوده^۵

۱- مرکز ازدیاد برداشت از مخازن نفت و گاز، دانشکده مهندسی شیمی، نفت و گاز، دانشگاه شیراز، شیراز، ایران

۲- بخش مهندسی نفت، دانشکده مهندسی شیمی، نفت و گاز، دانشگاه شیراز، شیراز، ایران

۳- دانشکده معدن و علوم زمین، دانشگاه نظربایف، آستانه، قزاقستان

۴- دانشکده فنی فومن، دانشگاه تهران، گیلان، ایران

۵- بخش مهندسی شیمی، دانشکده مهندسی شیمی، نفت و گاز، دانشگاه شیراز، شیراز، ایران

تاریخ دریافت: ۱۴۰۴/۰۵/۰۹ تاریخ تغییر: ۱۴۰۴/۰۸/۰۳ تاریخ پذیرش: ۱۴۰۴/۰۸/۱۸

چکیده

درک گرانروی نفت نقش اساسی در حل چالش‌های متنوع مهندسی مخزن دارد. اندازه‌گیری تجربی این ویژگی عموماً پرهزینه و زمان‌بر است و از این‌رو، توسعه مدل‌های پیش‌بینی دقیق و کارآمد اهمیت ویژه‌ای دارد. در این پژوهش، افزون بر توسعه یک شبکه عصبی پرسپترون چندلایه برای برآورد گرانروی نفت مرده و ارزیابی ده تابع بهینه‌سازی، عملکرد مدل‌های یادگیری ماشین دیگر شامل جنگل تصادفی، گرادیان بوستینگ و برازش بردار پشتیبان نیز مقایسه گردید. برای شبکه عصبی پرسپترون چندلایه، تأثیر ابرپارامترهای کلیدی نظیر نرخ یادگیری، تعداد لایه‌های پنهان و پیکربندی نورون‌ها به صورت نظام‌مند بررسی شد تا ساختار بهینه تعیین گردد. نوآوری اصلی پژوهش، به کارگیری هم‌زمان ویژگی‌های فیزیکی-شیمیایی و ویژگی‌های ترکیبی برش‌های مختلف نفت خام (از سبک تا سنگین) در قالب مجموعه داده‌ای جامع شامل ۲۲۹۹ نمونه است که به افزایش دقت و تعمیم‌پذیری مدل‌ها منجر شد. نتایج نشان داد الگوریتم گرادیان بوستینگ با دست‌یابی به تعادل میان دقت و توان تعمیم‌دهی، بهترین عملکرد کلی را ارائه می‌دهد ($R^2=0/8266$ ، $RMSE=2/0012$ و $AAPRE=3/5004$ در داده‌های آزمایش). مدل جنگل تصادفی نیز با ثبت مقدار $R^2=0/8191$ در داده‌های آزمایش، عملکردی رقابتی داشت، هرچند در مرحله آموزش ($R^2=0/9749$) نشانه‌هایی از بیش‌برازش مشاهده شد. مدل MLP-AdamW با $R^2=0/8541$ و $RMSE=1/8064$ در داده‌های آزمایش دقت قابل توجهی ارائه داد و نشان داد شبکه‌های عصبی در شناسایی روابط غیرخطی میان متغیرها مؤثر هستند، هرچند حساسیت بیشتری نسبت به ناهمگنی داده‌ها دارند. در مقابل، مدل SVR به دلیل ماهیت نسبتاً خطی خود، پایین‌ترین دقت را در داده‌های غیرخطی و حجیم نشان داد ($R^2=0/7699$). بر این اساس، می‌توان نتیجه گرفت که الگوریتم گرادیان بوستینگ با بهره‌گیری از یادگیری مرحله‌ای و جلوگیری از بیش‌برازش، گزینه بهینه برای پیش‌بینی گرانروی نفت مرده در داده‌های وسیع محسوب می‌شود.

کلمات کلیدی: گرانروی، یادگیری ماشین، شبکه عصبی پرسپترون چند لایه، جنگل تصادفی، گرادیان بوستینگ، برازش بردار پشتیبان

*مسئول مکاتبات
آدرس الکترونیکی

masoud.riazi@nu.edu.kz
parsaei@shirazu.a.c.ir

شناسه دیجیتال: (DOI: 10.22078/pr.2025.5728.3547)

مقدمه

در میان تمام ویژگی‌های وابسته به فشار-دما-حجم^۱ که بر حرکت سیال در محیط متخلخل در مراحل مختلف تأثیر می‌گذارند، گرانروی نقش بسیار مهمی دارد. اندازه‌گیری دقیق گرانروی برای فرآیندهای اکتشاف و تولید نفت ضروری است [۳-۱]. این پارامتر در ارزیابی راندمان چاه، جریان سیال در محیط متخلخل، آزمایش چاه، طراحی و شبیه‌سازی چاه و خطوط لوله، عملیات تولید و توسعه میدان، و تجهیزات انتقال اهمیت زیادی دارد [۳-۵]. گرانروی نفت خام به عواملی همچون ویژگی‌های فیزیکی و ترمودینامیکی، حلالیت و نوع گاز، گرانرش نفت، دما، فشار نقطه حباب، نوع سیال (نیوتنی یا غیرنیوتنی) [۶ و ۷]، چگالی، ترکیب شیمیایی [۸] و ساختار مولکولی سیال بستگی دارد [۹]. در میان این عوامل، اثر فشار و دما بر گرانروی قابل توجه است. کاهش فشار اشباع به دلیل آزاد شدن گازهای محلول موجب افزایش گرانروی می‌شود، درحالی‌که افزایش دما معمولاً باعث کاهش آن می‌شود [۹]. بنابراین، بسته به شرایط فشار مخزن، گرانروی نفت خام را می‌توان در سه حالت گرانروی نفت مرده (μ_{od})، گرانروی نفت اشباع (μ_{ob}) و گرانروی نفت زیر اشباع (μ_{oa}) بررسی کرد [۱۰-۱۱]. به‌طور سنتی، تعیین گرانروی نفت خام از طریق آزمایش‌های تجربی بر روی نمونه‌های سطحی یا زیرسطحی انجام می‌شود. این فرآیند معمولاً زمان‌بر، پرهزینه و نیازمند تخصص فنی بالا است [۱۳-۱۵]. در چند دهه اخیر، روابط تجربی و نیمه‌تجربی متعددی، اغلب بر پایه روابط حالت، برای پیش‌بینی گرانروی نفت خام ارائه شده‌اند. بیشتر این روابط برای محدوده‌های مشخصی از دما و فشار توسعه یافته‌اند و در صورت استفاده در خارج از آن محدوده، دقت خود را از دست می‌دهند [۴]. در مقیاس میدانی، روابط تجربی برای تخمین گرانروی نفت مرده، اشباع و زیر اشباع به کار می‌روند. با این حال، این روابط اغلب به دلیل

محدود بودن داده‌های توسعه یا وابستگی به یک ناحیه خاص، دقت کافی ندارند و نیازمند بهبود هستند [۱۶]. از مهم‌ترین روابط تجربی می‌توان به روابط ارائه شده توسط گلاسو [۱۷]، بگز و رابینسون [۱۸]، الخفاجی و همکاران [۱۹]، توسط بیل [۲۰]، کای [۲۱]، همتی و همکاران [۱۰]، آلومایر و همکاران [۲۲]، ناصری و همکاران [۲۳]، حسین و همکاران [۲۴]، الشرکوی و علی‌خان [۲۵]، بنیسون [۲۶]، دیگتو و همکاران [۲۷]، کارتواتمودجو و اشمیت [۲۸]، لابدی [۲۹]، اگبوگاه و نگ [۳۰]، پتروسکی [۳۱]، اشاره کرد. برخی پژوهشگران معتقدند که گرانروی به ویژگی‌هایی وابسته است که معمولاً اندازه‌گیری نمی‌شوند، مانند ضریب بی مرکزی^۲، دمای بحرانی و جرم مولی [۳۲-۳۴]. علاوه بر این، در دسترس نبودن نمونه‌های سیال مخزن، می‌تواند مانعی برای اندازه‌گیری دقیق گرانروی باشد. این محدودیت‌ها باعث شده‌اند تا مهندسان مخزن از روابط تجربی موجود برای تخمین گرانروی استفاده کنند. از آنجاکه بسیاری از این روابط براساس داده‌های محدود توسعه یافته‌اند [۱۱]، استفاده از آن‌ها در شرایط متفاوت ممکن است با خطای زیاد همراه باشد [۱۵، ۳۵-۳۷]. با توجه به این موارد، نیاز به توسعه مدل‌هایی وجود دارد که بتوانند گرانروی نفت مرده را در گستره وسیعی از شرایط عملیاتی پیش‌بینی کنند. همچنین در مواردی که نمونه فیزیکی در دسترس نیست یا محدودیت زمانی و اقتصادی وجود دارد، استفاده از روش‌های کارآمدتر ضروری است. برای رفع این چالش‌ها، در سال‌های اخیر از روش‌های هوش مصنوعی و یادگیری ماشین نظیر مدل‌های مدل‌های یادگیری جمعی^۳ [۳۸]، روش گروهی پردازش داده‌ها^۴ [۳۹]، ماشین بردار پشتیبان^۵ [۴۰]، الگوریتم ژنتیک^۶ [۴۱]،

1. PVT

2. Acentric Factor

3. Ensemble Models

4. Group Method of Data Handling (GMDH)

5. Support Vector Machine (SVM)

6. Genetic Algorithm (GA)

بزرگ‌تر از مطالعات پیشین است که موجب افزایش قابلیت تعمیم و اعتبار نتایج می‌شود. برای شبکه عصبی پرسپترون چندلایه، تأثیر توابع بهینه‌سازی مختلف، نرخ‌های یادگیری و ساختار شبکه (شامل تعداد لایه‌ها و نرون‌ها) به صورت نظام‌مند بررسی و بهینه‌سازی شده است. ترکیب این جنبه‌ها، زمینه‌ساز درکی عمیق‌تر از رفتار مدل‌های یادگیری ماشین و بهبود دقت و پایداری در پیش‌بینی گرانروی نفت مرده شده است.

جمع‌آوری و پیش‌پردازش داده‌ها

یکی از بخش‌های مهم برای بررسی دقت و قابلیت اطمینان مدل‌های یادگیری ماشین، جمع‌آوری داده‌ها است. در این مطالعه از داده‌های گزارش شده در پایگاه داده تحلیل نفت خام^۴ استفاده شده است [۵۹]. در مجموعه داده‌های کوچک، اکثر روش‌های یادگیری ماشین فاقد دقت و استحکام لازم بوده و قادر به پیش‌بینی مقدار دقیق پارامتر مورد نظر نیستند. بنابراین به مجموعه‌ای با داده‌های زیاد به منظور آموزش مدل نیاز است. در این مطالعه، تعداد کل داده‌های برداشت‌شده از پایگاه داده تحلیل نفت خام ۹۰۵۶ عدد بوده است. پس از فرآیند پیش‌پردازش شامل حذف داده‌های ناقص، شناسایی و حذف داده‌های خارج از محدوده با استفاده از روش آماری فاصله میان چارک‌ها^۵ و کنترل هم‌خطی میان متغیرها از طریق محاسبه ضرایب همبستگی پیرسون، متغیرهایی با $|r| > 0.9$ به‌عنوان بسیار همبسته شناسایی شدند. در این موارد، یکی از متغیرهای همبسته با توجه به اهمیت فیزیکی و میزان ارتباط آن با خروجی حذف گردید. در نهایت، پس از اجرای این مراحل، تعداد ۲۲۹۹ داده معتبر برای انجام مطالعات مدل‌سازی مورد استفاده قرار گرفت.

شبکه‌های عملکردی^۱ [۱۶ و ۴۲]، شبکه عصبی مصنوعی^۲ [۴۳ و ۴۴]، شبکه عصبی تابع پایه شعاعی^۳ [۴۵]، به‌منظور بهبود پیش‌بینی گرانروی نفت، استفاده شده‌اند. نتایج این مطالعات نشان می‌دهد که مدل‌های هوش مصنوعی، در مقایسه با روابط تجربی، دقت بالاتری در پیش‌بینی گرانروی دارند [۴۳ و ۴۶-۴۸]. در جدول ۱، پژوهش‌های اخیر منتشرشده در زمینه پیش‌بینی گرانروی نفت با استفاده از مدل‌های هوش مصنوعی آورده شده است. هدف این پژوهش، توسعه و مقایسه چهار مدل یادگیری ماشین شامل شبکه عصبی پرسپترون چندلایه، جنگل تصادفی، گرادیان بوستینگ و برآزش بردار پشتیبان برای پیش‌بینی گرانروی نفت مرده به‌عنوان تابعی از پارامترهایی مانند وزن مخصوص نفت خام، محتوای نیتروژن، نقطه ریزش، باقی‌مانده کربن حاصل از تقطیر خلأ، چگالی برش‌های مختلف (بنزین سبک، نفتا، نفت سفید، نفت گاز، روان‌کننده‌ها) و وزن مخصوص باقی‌مانده در 60°F است. در این مطالعه برای نخستین بار از پارامترهای حاصل از فرآیند جداسازی نفت مرده به‌منظور پیش‌بینی گرانروی استفاده شده است. مدل‌های توسعه‌یافته می‌توانند نیاز به روابط تجربی جدید را کاهش دهند و در صورت بهره‌گیری از داده‌های آموزشی متنوع، قابلیت تعمیم بالایی برای انواع نفت خام و شرایط مخزنی داشته باشند.

در این پژوهش، نوآوری اصلی در ترکیب هم‌زمان پارامترهای فیزیکی-شیمیایی نفت خام و ویژگی‌های برش‌های مختلف آن (از برش‌های سبک تا سنگین) به‌عنوان ورودی مدل‌ها و همچنین بررسی تطبیقی چهار الگوریتم یادگیری ماشین در پیش‌بینی گرانروی نفت است. برخلاف مطالعات پیشین که معمولاً از تعداد محدودی ویژگی عمومی مانند دما، درجه API یا وزن مخصوص استفاده کرده‌اند، در این تحقیق مجموعه‌ای جامع‌تر از ویژگی‌های فیزیکی و ترکیبی نفت به‌کار گرفته شده است. علاوه‌براین، تعداد داده‌های مورد استفاده (۲۲۹۹ نمونه) به‌مراتب

1. Functional Networks (FN)

2. Artificial Neural Network (ANN)

3. Radial Basis Function Neural Network (RBFNN)

4. COADB

5. IQR

جدول ۱: پژوهش‌های اخیر منتشر شده در زمینه پیش‌بینی گرانروی نفت با استفاده از مدل‌های هوش مصنوعی.
Table 1. Recent published research on predicting oil viscosity using artificial intelligence models

Author & Year	Data	Summary of Work	Models Investigated	Results	Advantages of Present Study	Ref
Li et al., 2023	combined lab and 243 literature data. Inputs: API, Temp, and oil composition	Application of multiple computational models to estimate crude oil (especially heavy oil) viscosity as a function of API, Temp, and fluid composition; hyperparameter optimization and design space analysis	Decision Tree (DT), Multi-Layer Perceptron (MLP), Generalized Neural Networks (GNN)	RMSE for final DT, MLP, and GNN models were 52.40, 25.08, and 30.83, respectively. R ² values were 0.921, 0.780, and 0.933	Higher accuracy and generalization using 2,299 samples and adding parameters from fraction separation and physico-chemical properties	[49]
Langerouy et al., 2023	empirical data 1,368 points describing Iranian crude oil. Inputs: P, T, API, and GOR	Use of Iranian field datasets and intelligent models for reservoir viscosity determination; focus on black oil parameters and generalized model design	Ensemble Boosting algorithms (Gradient Boosting, XGBoost, CatBoost)	XGBoost was the best method with higher accuracy and lower error. It provided a MAPE of 1.968%, reducing hybrid method error by half and black oil method error by five times	Demonstrates superior generalization beyond regional data using fraction features, physico-chemical parameters, and a larger dataset	[50]
Fotias et al., 2023	Combined lab and field data. Inputs: T, P, API	Providing a workflow for undersaturated oil viscosity prediction using ML models and available field/lab inputs	Random Forest (RF), Gradient Boosting (GB), ANN	Significant improvement in prediction accuracy compared to conventional correlations in undersaturated conditions	First study to simultaneously use various fraction parameters and physico-chemical features, increasing accuracy and generalization	[51]
Hadavi Moghaddam et al., 2021	Multi-source dead oil viscosity measurements; Inputs: API, T, Density	Comparing ML algorithm performance in dead oil viscosity estimation against classical correlations	SuperLearner (Ensemble), XGBoost, RF, ANN, SVR	The best model (SuperLearner ensemble) achieved R ² = 0.96, showing clear improvement over classical correlations	Provides higher accuracy and generalization using 2,299 samples and new input sets (fractions, hydrocarbon content, pour point, residual carbon)	[52]
Bahonar et al., 2022	PVT empirical data 1,950 from Iranian reservoirs; Inputs: API, T, Density	Using data mining techniques to generate new viscosity correlations and comparing them with traditional correlations	SVR, ANN, and Ensemble Boosting methods	Improved accuracy in the studied datasets	Offers greater generalization through more comprehensive inputs and multi-source data compared to improved correlations	[53]
Dolajati et al., 2024	lab data; Inputs: API, 243 T, P, and other physical parameters	Proposing a new mathematical algorithm for estimating viscosity in hydrocarbons and comparing with existing correlations	Nonlinear mathematical methods vs. data-driven techniques	Reported higher accuracy than some conventional correlations for specific samples	A data-driven ML approach providing better adaptability and generalization than pure mathematical models via high sample count and fraction inputs	[54]
Gao et al., 2022	datasets from various 156 heavy oil fields in China. Inputs: T, light oil viscosity, heavy oil viscosity, and dilution ratio	Predicting viscosity of heavy oil diluted with lighter oil using MLP and comparing with existing models	(Multilayer Perceptron (MLP	New model predicts diluted heavy oil viscosity with higher accuracy; MAPE: 10.44%, Std. Dev: 45.8%, R ² : 0.950	Increased generalization and accuracy across wider conditions using 2,299 samples, thermodynamic inputs, and fractions	[55]
Stratov et al., 2023	viscosity data points 158 using an Angler viscosimeter; Parameters: Mass ratio and T	Analyzing viscosity of heavy oil mixtures with diluents and comparing empirical correlations with ANN	Empirical correlators and ANN	Viscosity decreases exponentially with increasing diluent concentration; ANN performed well compared to correlations	Designed for dead crude oil viscosity using direct inputs from fraction separation and chemical properties, whereas others focus on mixtures	[56]
Almudhhi et al., 2025	records; Inputs: T, 371 density, mole fraction, and molecular weight	Modeling heavy oil viscosity using multiple ML models; analyzing impact of T, density, C ₁₊ , and C ₂₊ mole fraction	Comparison of Back-propagation ANN and General Fitting ANN	Back-propagation ANN (15 hidden neurons) achieved 27.1% MAPE on blind test data, outperforming GNN (15% MAPE) and empirical correlations	Uses more comprehensive inputs (fractions, hydrocarbon content, pour point, residual carbon) and a larger dataset for better generalization	[57]
Lei et al., 2025	lab data points (shear 526 rate, T, molecular specs	Predicting apparent viscosity of waxy oils based on shear rate, T, and molecular properties	KNN, CNN, AdaBoost, SVR, DT, RF, MLP, and Ensemble Learning	Strong agreement between predicted and actual values; provides an accurate estimation method without costly/time-consuming experiments	Focuses on dead oil viscosity with fraction and physico-chemical inputs, unlike studies focusing on shear-rate-dependent waxy oils	[58]

جهت مدل‌سازی گرانیروی را به‌صورت خلاصه نشان می‌دهد. همان‌طور که در **جدول ۳** نشان داده شده است، مقادیر متوسط پارامترهای SU100، SGRR، NON_LUBVOL، MED_LUBVOL، VIS_LUBVOL، DGN، KER_DISVOL، AVED، GASOIL_VOL، CRN، POUR_POINT، SGRAV_CRUD به ترتیب ۰/۷۳۹۱۶، ۰/۲۵۵۵۸، ۰/۸۴۴۳۱۵، ۰/۲۳۹۱۰۸، ۰/۹۴۱۳۵۵، ۰/۱۱۷۲۵۳۹، ۰/۱۰۶۲۰۴، ۰/۱۴۷۱۷۶۶، ۰/۸۲۶۰۴۷ و ۰/۷۳۹۱۶ می‌باشد. محدوده گرانیروی و وزن مخصوص نفت خام به ترتیب ۳۱-۵۴ و ۰/۷۷۲-۰/۸۷۷ می‌باشد.

روش‌شناسی

شبکه عصبی پرسپترون چند لایه^۱

پرسپترون چند لایه یکی از انواع شبکه پیش‌خور^۲ است [۶۰ و ۶۱] که در آن انتقال اطلاعات یک طرفه است، یعنی از لایه ورودی به لایه خروجی است. یک مدل MLP از یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی تشکیل شده است.

این داده‌ها شامل وزن مخصوص نفت خام، محتوای نیتروژن در نفت خام، نقطه ریزش، باقی‌مانده کربن باقی‌مانده از تقطیر خلاء، چگالی بنزین سبک، چگالی کسر بنزین و نفتا، چگالی تقطیر نفت سفید، چگالی نفت گاز، درصد حجمی تقطیر نفت سفید، درصد حجمی نفت گاز، درصد حجمی تقطیر روان‌کننده غیر چسبناک، درصد حجمی تقطیر روان‌کننده چسبناک متوسط، درصد حجمی تقطیر روان‌کننده چسبناک، وزن مخصوص باقی‌مانده در ۶۰ °F به‌عنوان ورودی و گرانیروی در °F ۱۰۰ به‌عنوان خروجی بوده است. نسبت داده‌های آموزش، اعتبارسنجی و آزمایش در این مطالعه ۷۵، ۱۵ و ۱۰ می‌باشد. جزئیات داده‌های مذکور در **جدول ۲** ارائه شده است. در این تحقیق از پارامترهای آماری حداقل، حداکثر، واریانس نمونه، انحراف معیار، مد، میانه، خطای استاندارد، میانگین، و دو پارامتر آماری جدید دیگر شامل چولگی و کشیدگی برای پیش‌پردازش داده‌ها استفاده شد. **جدول ۳** پارامترهای آماری برای داده‌های مورد استفاده

جدول ۲ جزئیات داده‌های ورودی و خروجی مورد استفاده جهت پیش‌بینی گرانیروی توسط مدل‌های یادگیری ماشین [۵۹]

Table 2. Details of input and output data used for viscosity prediction by machine learning models [59].

abbreviation	Equivalent phrase	Unit
SGRAV_CRUD	Specific gravity of crude	-
POUR_POINT	Pour point	°F
CRN	Nitrogen in crude	% Weight
CAR_RES_WT	Carbon residue of residuum from vacuum distillation, obtained by Ramsbottom method and converted to Conradson equivalent	% Weight
DLG	Light gasoline density	gr/cm ³
DGN	Gasoline and naphtha fraction density	gr/cm ³
KER_DISVOL	Kerosene distillate	% Volume
DKD	Kerosene distillate density	gr/cm ³
GASOIL_VOL	Gas oil	% Volume
AVED	Gas oil density	gr/cm ³
NON_LUBVOL	No viscous lubricating distillate	% Volume
MED_LUBVOL	Medium viscous lubricating distillate	% Volume
VIS_LUBVOL	Viscous lubricating distillate	% Volume
SGRR	Specific gravity at 60 °F of the residue	-
SU100	Viscosity at 100 °F	SUS (Saybolt Universal Seconds)

1. Multilayer Perceptron (MLP)
2. Feedforward Network (FNN)

جدول ۳ پارامترهای آماری برای داده‌های مورد استفاده جهت مدل‌سازی گرانروی.

Table 3. Statistical parameters for data used for viscosity modeling

Statistical Parameters	SGRAV_CRUD	POUR_POINT	CRN	DGN	KER_DISVOL	GASOIL_VOL	AVED	NON_LUBVOL	MED_LUBVOL	VIS_LUBVOL	SGRR	SU100
Mean	0.826	2.938	0.025	0.739	11.725	14.717	0.844	10.062	5.961	0.239	0.941	39.970
Standard Error	0	0.057	0	0	0.084	0.075	0	0.027	0.029	0.010	0	0.097
Median	0.827	5.00	0	0.738	11.001	14.900	0.844	9.900	6.100	0	0.941	39.00
Mode	0.821	5.00	0	0.736	10.50	14.100	0.842	9.700	6.100	0	0.937	38.00
Standard Deviation	0.017	2.776	0.035	0.010	4.118	3.687	0.005	1.321	1.434	0.503	0.020	4.725
Sample Variance	0	7.711	0.001	0	16.958	13.598	2.84E-05	1.745	2.057	0.253	0	22.327
Kurtosis	-0.216	-0.903	0.654	0.215	0.022	-0.056	0.283	0.433	0.925	2.984	0.179	0.135
Skewness	-0.076	0.228	1.268	0.329	0.411	-0.019	0.140	0.282	-0.656	2.058	0.185	0.671
Range	0.105	10.00	0.150	0.069	22.400	19.600	0.033	8.00	9.600	2.00	0.119	23.00
Minimum	0.772	0	0	0.708	2.500	5.300	0.827	6.00	0.500	0	0.886	31.00
Maximum	0.877	10.00	0.150	0.777	24.900	24.900	0.860	14.00	10.100	2.00	1.005	54.00
Sum	1945.341	6919	60.190	1740.721	27613.300	34660.100	1988.362	23696.100	14040.500	563.100	2216.891	94131
Count	2299	2299	2299	2299	2299	2299	2299	2299	2299	2299	2299	2299

و سپس همین محاسبه برای گره‌های لایه بعدی انجام می‌شود که می‌تواند لایه مخفی بعدی یا لایه خروجی باشد. روابط ۳ و ۴ به شرح زیر هستند:

$$O_k = \sum_{j=1}^h (w_{jk} \cdot S_j) - \theta_k, k = 1, 2, \dots, m \quad (3)$$

$$O_k = \text{sigmoid}(o_k) = \frac{1}{(1 + \exp(-O_k))}, k = 1, 2, \dots, h \quad (4)$$

در این روابط m تعداد گره‌های لایه بعدی را نشان می‌دهد. پارامتر θ_j بایاس گره‌ها است. با استفاده از روابط ۱-۴، مطابقت بین ورودی‌ها و خروجی‌ها را می‌توان ایجاد کرد. همچنین می‌توان مشاهده کرد که عملکرد MLP به مقادیر پارامتر وزن‌ها و بایاس‌ها بستگی دارد که می‌تواند توسط الگوریتم‌های فراابتکاری بهینه شود. در این مورد، متغیرهایی که باید بهینه شوند، وزن‌ها و بایاس‌ها هستند. تابع هدف می‌تواند میانگین، میانگین مربعات خطا (MSE) باشد که به صورت رابطه ۵ تعریف می‌شود:

$$\overline{MSE} = \frac{\sum_{k=1}^m (O_i^k - d_i^k)^2}{S} \quad (5)$$

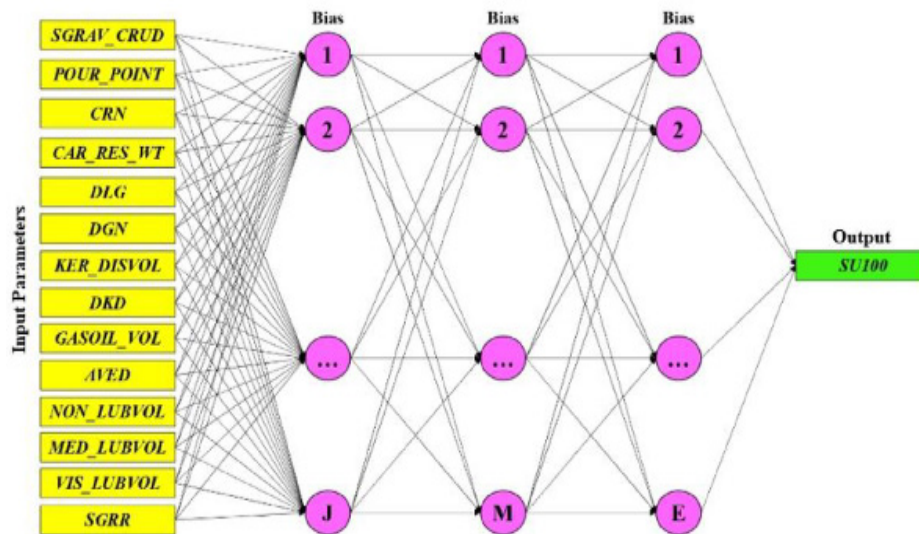
در این رابطه، s تعداد نمونه‌های آموزشی است. پارامتر m تعداد گره‌ها در لایه خروجی است. پارامتر O_m^k خروجی واقعی در گره خروجی نام برای نمونه آموزشی کام است. پارامتر d_k^k خروجی صحیح در گره خروجی نام برای نمونه آموزشی کام است.

شکل ۱، شماتیک شبکه عصبی پرسپترون چند لایه مورد استفاده در این مطالعه را نشان می‌دهد. در هر لایه، چندین گره برای محاسبه سیگنال‌های ارسالی وجود دارد. وزن‌های بین هر دو گره از لایه‌های مختلف و بایاس‌ها برای گره‌ها به جز آنهایی که در لایه ورودی هستند، دو پارامتر کلیدی برای یک مدل MLP معین هستند که باید بهینه شوند [۶۲]. علاوه بر این، ساختارهای مدل MLP را نیز می‌توان برای عملکرد بهتر در مسائل عملی، مانند اصلاح تعداد گره‌ها در هر لایه و تعداد لایه‌ها، تنظیم کرد [۶۳]. روش‌های محاسباتی MLP به شرح زیر ارائه شده است [۶۴ و ۶۵]. اولین مرحله محاسبه مجموع وزنی سیگنال‌های ورودی با استفاده از رابطه ۱ است:

$$S_j = \sum_{i=1}^n (W_{ij} \times X_i) - \theta_{i,j}, j = 1, 2, \dots, h \quad (1)$$

در این رابطه، n و h تعداد سیگنال‌ها و گره‌های ورودی در لایه پنهان را نشان می‌دهد. پارامتر W_{ij} وزن بین گره i ام و گره j ام است. پارامتر X_i نشان دهنده مقدار سیگنال ورودی است. پارامتر θ_j بایاس گره j ام است. مورد بعدی محاسبه تابع فعال‌سازی با استفاده از تابع سیگموئید برای گره j ام است که به شرح رابطه ۲ است:

$$S_i = \text{sigmoid}(S_j) = \frac{1}{(1 + \exp(-S_j))}, j = 1, 2, \dots, h \quad (2)$$



شکل ۱ ساختار شبکه MLP با سه لایه پنهان.

Figure 1. MLP network with three hidden layers

می‌شود.

۳. پیشرفت‌های اخیر

در چند سال اخیر، تحقیقات زیادی در زمینه الگوریتم‌های بهینه‌سازی انجام شده و یافته‌های جدید ممکن است به بهبود عملکرد شبکه‌های عصبی کمک کند.

۴. کاهش حساسیت به فراپارامترها

با به‌کارگیری چندین الگوریتم می‌توان حساسیت مدل را نسبت به انتخاب فراپارامترها کاهش داد و یک راهکار بهینه‌سازی با ثبات بیشتر پیدا کرد.

۵. تنوع در ویژگی‌ها

هر الگوریتم بهینه‌سازی ویژگی‌ها و مزایای خاص خود را دارد. بررسی چندین الگوریتم می‌تواند به شناسایی مناسب‌ترین روش براساس ویژگی‌های خاص داده‌ها کمک کند.

۶. تجربه کاربری بهبود یافته

در عمل، حقایق نشان می‌دهد که استفاده از روش‌های متنوع می‌تواند به شناسایی نقاط ضعف و قوت مدل‌ها کمک کند و در نهایت تجربه کاربری را بهبود بخشد. در ادامه به بررسی و توضیح توابع بهینه‌سازی مختلف که در این مطالعه به کار رفته، پرداخته شده است.

بنابراین، آموزش مدل MLP با استفاده از روش‌های فراابتکاری، یافتن وزن‌ها و بایاس‌های بهینه برای به حداقل رساندن میانگین MSE است. در ادامه به شرح توضیحاتی در رابطه با توابع بهینه‌سازی پرداخته شده است [۶۵].

توابع بهینه‌سازی

در این مطالعه از ده نوع تابع بهینه‌سازی مختلف استفاده شده است. از دلایل انجام این کار می‌توان به مقایسه عملکرد، پوشش شرایط مختلف، بررسی پیشرفت‌های اخیر، کاهش حساسیت به فراپارامترها، تنوع در ویژگی‌ها و تجربه کاربری بهبود یافته اشاره کرد. توضیح موارد مذکور به شرح زیر است.

۱. مقایسه عملکرد

با استفاده از الگوریتم‌های مختلف، می‌توان عملکرد مدل را در سناریوهای مختلف مقایسه کرد. این مقایسه می‌تواند شامل معیارهای دقت، سرعت همگرا شدن و توانایی تعمیم‌پذیری مدل باشد.

۲. پوشش شرایط مختلف

برخی الگوریتم‌ها ممکن است در شرایط خاص یا با ویژگی‌های خاص داده‌ها بهتر عمل کنند. با استفاده از چندین الگوریتم می‌توان اطمینان حاصل کرد که در شرایط مختلف بهترین گزینه انتخاب

بهینه‌ساز AdaGrad

بهینه‌ساز AdaGrad خانواده‌ای از الگوریتم‌های زیر‌گرادیان^۱ برای بهینه‌سازی تصادفی است. الگوریتم‌های متعلق به آن خانواده مشابه نزول‌گرادیان تصادفی مرتبه دوم^۲ با تقریبی برای هسین^۳ تابع بهینه‌شده هستند. نام AdaGrad از Adapta-tive Gradient گرفته شده است. این الگوریتم با رویکردی شهودی، نرخ یادگیری هر ویژگی را متناسب با هندسه تخمینی مسئله تطبیق می‌دهد [۶۶]. به‌طور مشخص، این روش نرخ‌های یادگیری بالاتری را به ویژگی‌های نادر اختصاص می‌دهد تا تضمین کند که به‌روزرسانی پارامترها کمتر تحت تأثیر فرکانس و بیشتر تابع میزان ارتباط باشند. بهینه‌ساز AdaGrad توسط دوچی و همکاران معرفی شد [۶۷]. این الگوریتم یکی از محبوب‌ترین الگوریتم‌ها برای یادگیری ماشین (مخصوصاً برای آموزش شبکه‌های عصبی عمیق) است و بر توسعه الگوریتم Adam تأثیر گذاشته است [۶۸ و ۶۹]. هدف AdaGrad به حداقل رساندن مقدار مورد انتظار یک تابع هدف تصادفی، با توجه به مجموعه‌ای از پارامترها، و با توجه به دنباله‌ای از تحقق تابع است. همانند سایر روش‌های مبتنی بر شیب فرعی^۴، با به‌روزرسانی پارامترها در جهت مخالف زیر‌گرادیان‌ها به این مهم دست می‌یابد. روش‌های زیر‌گرادیان استاندارد از قوانین به‌روزرسانی با اندازه‌های مرحله‌ای استفاده می‌کنند که اطلاعات مشاهدات گذشته را نادیده می‌گیرند، درحالی‌که AdaGrad نرخ یادگیری را برای هر پارامتر به‌صورت جداگانه با استفاده از توالی تخمین‌های گرادیان تطبیق می‌دهد [۶۶]. الگوریتم بهینه‌سازی AdaGrad در بسیاری از الگوریتم‌های یادگیری ماشین مانند مدل‌های خطی و شبکه‌های عصبی عملکرد بسیار خوبی داشته و باعث بهبود دقت و سرعت آنها شده است. این الگوریتم اما مشکلاتی مانند افزایش مجموع مربعات گرادیان‌ها در طول زمان دارد که این امر سبب کاهش نرخ یادگیری می‌شود. برای

رفع این مشکل الگوریتم‌های بهینه‌سازی دیگری مانند Adam و RMSProp پیشنهاد شده‌اند [۷۰ و ۷۱].

بهینه‌ساز Adadelta^۵

الگوریتم Adadelta یک روش بهینه‌سازی تصادفی است که امکان استفاده از روش نرخ یادگیری بعدی را برای SGD فراهم می‌کند. این بهینه‌ساز یک الگوریتم توسعه‌ای از AdaGrad است که به‌دنبال کاهش نرخ یادگیری تهاجمی و یکنواخت در حال کاهش است [۷۲]. به‌جای انباشته کردن تمام گرادیان‌های مربع گذشته، Adadelta پنجره گرادیان‌های گذشته انباشته شده را به‌اندازه ثابت ω محدود می‌کند [۷۳]. به‌جای ذخیره‌سازی ناکارآمد ω گرادیان‌های مجذور قبلی، مجموع گرادیان‌ها به‌صورت بازگشتی به‌عنوان میانگین در حال فروپاشی همه گرادیان‌های مجذور گذشته تعریف می‌شود. مزیت اصلی AdaDelta این است که نیازی به تنظیم نرخ یادگیری پیش فرض ندارد [۷۴].

بهینه‌ساز RMSprop^۶

الگوریتم RMSProp یک الگوریتم بهینه‌سازی نرخ یادگیری تطبیقی است که برای رسیدگی به برخی از مسائلی که در روش نزول‌گرادیان تصادفی (SGD) در آموزش شبکه‌های عصبی عمیق با آن مواجه می‌شوند، طراحی شده است [۷۵]. الگوریتم RMSProp توسط جفری هینتون در کلاس Coursera در شبکه‌های عصبی معرفی شد و به‌طور رسمی منتشر نشده است، اما به‌دلیل کارایی آن در کاربردهای مختلف محبوبیت پیدا کرده است. الگوریتم RMSProp با حفظ میانگین متحرک مجذور گرادیان‌ها برای هر وزن و تقسیم نرخ یادگیری بر این میانگین، به موضوع نرخ یادگیری کلی می‌پردازد [۷۶].

1. Sub-Gradient
2. Second-Order Stochastic Gradient Descend
3. Hessian
4. Ub-Gradient-Based Methods
5. An Adaptive Learning Rate Method
6. Root Mean Square Propagation

بهینه‌سازی یک تابع هدف با ویژگی‌های همواری مناسب^۱ (مانند متمایز شدن یا تفکیک‌پذیری) است. می‌توان آن را به‌عنوان یک تقریب تصادفی از بهینه‌سازی نزول گرادیان در نظر گرفت، زیرا گرادیان واقعی (محاسبه‌شده از کل مجموعه داده‌ها) را با تخمینی از آن (محاسبه‌شده از زیر مجموعه‌ای از داده‌ها که به‌طور تصادفی انتخاب شده‌اند) جایگزین می‌کند. به‌خصوص در مسائل بهینه‌سازی با ابعاد بالا، این امر بار محاسباتی بسیار بالا را کاهش می‌دهد و در ازای نرخ همگرایی کمتر، به تکرارهای سریع‌تر دست می‌یابد [۸۳ و ۸۴]. ایده اصلی پشت تقریب تصادفی را می‌توان به الگوریتم رایینز-مونرو در دهه ۱۹۵۰ ربط داد. امروزه، نزول گرادیان تصادفی به یک روش بهینه‌سازی مهم در یادگیری ماشین تبدیل شده است [۸۵ و ۸۶].

بهینه‌ساز ASGD

الگوریتم ASGD یا نزول گرادیان اندازه دسته‌ای تطبیقی مبتنی بر نمونه^۱، نوعی از روش نزول گرادیان است. این الگوریتم بدون فرآیندهای اضافی، می‌تواند روند بهینه‌سازی شبکه عصبی عمیق را سرعت بخشد و به‌دقت بالاتری نسبت به SGD دست یابد. علاوه‌براین، می‌تواند به کشف گره‌های اضافی در یک شبکه عصبی کمک کند [۸۷].

بهینه‌ساز Adam^{۱۱}

الگوریتم بهینه‌سازی Adam یک بهینه‌سازی مرتبه اول مبتنی بر گرادیان تابع تصادفی است. این الگوریتم یک روش مناسب برای پیاده‌سازی مستقیم برای هر مدلی از نظر مجموعه داده‌ها و پارامترهای بزرگ است.

این الگوریتم تضمین می‌کند که نرخ یادگیری برای هر وزن در مدل تطبیق داده می‌شود و امکان به‌روزرسانی‌های ظریف‌تر را فراهم می‌کند. ایده کلی این است که نوسانات در جهت‌ها را با شیب‌های تند کاهش داد درحالی‌که امکان حرکت سریع‌تر در مناطق مسطح از Loss را فراهم کرد [۴۶ و ۷۷]. به‌روزرسانی RMSProp روش AdaGrad را تنظیم می‌کند تا میزان یادگیری تهاجمی و یکنواخت آن را کاهش دهد. الگوریتم RMSProp به‌جای انباشته کردن تمام گرادیان‌های مربع گذشته، از یک فروپاشی نمایی استفاده می‌کند که تاریخچه را از گذشته بسیار دور می‌اندازد تا بتوان پس از یافتن یک کاسه محدب^۱، به‌سرعت همگرا شود، گویی یک AdaGrad با شروعی تازه است [۷۵ و ۷۸].

بهینه‌ساز Rprop^۲

الگوریتم Rprop، یک روش اکتشافی یادگیری برای یادگیری نظارت شده در شبکه‌های عصبی مصنوعی پیشخور است. این تابع یک الگوریتم بهینه‌سازی مرتبه اول است. این الگوریتم توسط مارتین ریدمیلر و هاینریش براون ایجاد شد [۷۹ و ۸۰]. مشابه قانون به‌روزرسانی منتهن، Rprop فقط علامت مشتق جزئی را روی همه الگوها در نظر می‌گیرد (نه بزرگی)، و به‌طور مستقل روی هر وزن عمل می‌کند. الگوریتم Rprop در صورت بزرگ بودن گرادیان‌ها، می‌تواند منجر به افزایش یا کاهش بسیار زیاد وزن‌ها شود؛ این مسئله هنگام استفاده از تکه‌های کوچک^۳ به‌جای دسته‌های کامل^۴، چالش‌برانگیز خواهد بود [۸۱]. الگوریتم RMSprop این مشکل را با حفظ میانگین متحرک گرادیان‌های مجذور برای هر وزن و تقسیم گرادیان بر جذر مجذور میانگین حل می‌کند. الگوریتم RPROP یک الگوریتم به‌روزرسانی دسته‌ای^۵ است. در کنار الگوریتم همبستگی آبخاری^۶ و الگوریتم لونبرگ-مارکوارت^۷، Rprop یکی از سریع‌ترین مکانیسم‌های به‌روزرسانی وزن است [۸۲].

بهینه‌ساز SGD^۸

نزول گرادیان تصادفی یک روش تکراری برای

1. Convex Bowl
2. Resilient Backpropagation
3. Mini-Batches
4. Full Batches
5. Batch Update Algorithm
6. Cascade Correlation Algorithm
7. Levenberg-Marquardt Algorithm
8. Stochastic Gradient Descent
9. Smoothness Properties
10. Sample-Based Adaptive Batch Size Gradient Descent (ASGD)
11. Adaptive Moment Estimation

می‌کند تا در بهروزرسانی گرادیان طبق رابطه ۷ ظاهر شود:

$$\theta_{t+1} = \theta_{t,d} - \eta \left(\frac{1}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t + \omega_{t,d} \theta_{t,d} \right), \forall t \quad (7)$$

بهینه‌ساز Adamax^۲

بهینه‌ساز Adamax تعمیم Adam از معیار L_2 به معیار L_∞ است [۹۳].

$$u_t = \beta_2^x v_{t-1} + (1 - \beta_2^x) |g_t|^{\infty} = \max(\beta_2 v_{t-1}, |g_t|) \quad (8)$$

می‌توان با جایگزین کردن $\sqrt{\hat{v}_t + \epsilon}$ با u_t به رابطه بهروزرسانی Adam متصل شد تا قانون بهروزرسانی AdaMax به صورت رابطه ۹ به دست آید.

$$\theta_{t+1} = \theta_t - \frac{\eta}{u_t} \hat{m}_t \quad (9)$$

مقادیر پیش فرض $\eta = 0.002$ و $\beta_1 = 0.9$ و $\beta_2 = 0.999$ می‌باشند [۹۴].

بهینه‌ساز NAdam^۴

یک الگوریتم بهینه‌سازی است که روش بهینه‌سازی Adam و مفهوم گرادیان شتابدهی نستروف^۵ (NAG) را ترکیب می‌کند. قانون بهروزرسانی این الگوریتم به شکل رابطه ۱۰ است [۹۵ و ۹۶].

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \left(\beta_1 \hat{m}_t + \frac{(1 - \beta_1) g_t}{1 - \beta_1^t} \right) \quad (10)$$

به منظور بررسی دقیق‌تر عملکرد توابع بهینه‌سازی مورد استفاده، جدول ۴ نقاط قوت و ضعف هر یک از آن‌ها را به صورت مقایسه‌ای نشان می‌دهد. همان‌طور که مشاهده می‌شود، هر بهینه‌ساز ویژگی‌های خاص خود را دارد و انتخاب بهینه‌ساز مناسب به ماهیت داده‌ها، نرخ یادگیری و رفتار تابع هزینه بستگی دارد [۹۷ و ۹۸]. براساس مقایسه انجام‌شده در جدول ۴، الگوریتم‌های مبتنی بر ممنوم مانند Adam و NAdam معمولاً عملکرد پایدارتری نسبت به روش‌های کلاسیک مانند SGD و Rprop دارند. با این حال، انتخاب بهینه‌ساز مناسب باید براساس ویژگی‌های مجموعه داده و هدف مدل تنظیم شود.

از نظر منابع سخت‌افزاری، به حافظه کمتری نیاز دارد و از نظر محاسباتی بسیار کارآمد است [۸۸]. علاوه بر این، برای گرادیان‌های با نویز زیاد و یدکی برای اهداف و مشکلات غیر ثابت به خوبی مناسب است. تنظیم یک عامل مهم برای هر مدل موفق است، به طوری که الگوریتم بهینه‌سازی Adam معمولاً به تنظیم کمی نیاز دارد [۸۹]. برآورد لحظه تطبیقی (Adam) روشی است که نرخ یادگیری تطبیقی را از هر پارامتر جدا نگه می‌دارد و یک نرخ یادگیری واحد را برای تمام بهروزرسانی‌های وزنی که در طول آموزش تغییر نمی‌کنند، حفظ می‌کند [۹۰]. بهینه‌ساز Adam همچنین ترکیبی از RMSprop و نزول گرادیان تصادفی است، زیرا Adam لحظات اول و دوم گرادیان را برای متعادل کردن نرخ یادگیری برای هر وزن از شبکه مدل تخمین می‌زند. از مزیت‌های این بهینه‌ساز می‌توان به نیاز به حافظه کم، محاسبات بهینه، آسانی پیاده‌سازی، عدم تغییر در برابر مقیاس‌دهی مجدد قطری گرادیان‌ها، مناسب بودن برای پارامترها یا داده‌های بزرگ، مناسب بودن برای مسائل دارای گرادیان‌های نامتراکم یا بسیار نویزدار، و نیاز اندک به تنظیم پارامتر اشاره کرد [۹۰ و ۹۱].

بهینه‌ساز AdamW^۱

بهینه‌ساز AdamW یک روش بهینه‌سازی تصادفی است و اغلب نسبت به Adam با تنظیم L_2 برتری دارد زیرا کاهش وزن را از به روز رسانی‌های مبتنی بر گرادیان جدا می‌کند و منجر به منظم‌سازی مؤثرتر، تعمیم بهتر و همگرایی بهبود یافته می‌شود. این باعث می‌شود AdamW بهینه‌ساز قوی‌تر و قابل اعتمادتر برای طیف گسترده‌ای از کاربردهای یادگیری عمیق باشد. برای مشاهده این موضوع، تنظیم L_2 در Adam معمولاً با اصلاح انجام شده در رابطه ۶ اجرا می‌شود که در آن میزان کاهش وزن در زمان است [۹۲].

$$g_t = \nabla f(\theta_t) + \omega_t \theta_t \quad (6)$$

در حالی که AdamW عبارت کاهش وزن را تنظیم

1. Decoupling Weight Decay from the Gradient Update Process
2. L2 Regularization
3. Adaptive Moment Estimation with Infinity Norm
4. Nesterov-Accelerated Adaptive Moment Estimation
5. Nesterov Accelerated Gradient

جدول ۴ مقایسه نقاط قوت و ضعف توابع بهینه‌سازی مورد استفاده در این مطالعه.

Table 4. Comparison of strengths and weaknesses of the optimization functions used in this study.

Optimizer	Strengths	Weaknesses
SGD	Simple, fast, suitable for large datasets.	Sensitive to learning rate, slow convergence in non-convex landscapes.
ASGD	Better stability than SGD, reduced oscillations.	Requires more time for averaging, occasionally slow at the start.
Rprop	Scale-independent of gradient, high convergence speed.	Not suitable for noisy data, limited application in deep networks.
AdaGrad	Automatic learning rate adjustment, suitable for sparse data.	Excessive reduction of learning rate during training.
Adadelta	Resolves AdaGrad's learning rate decay; no initial tuning required.	May get stuck in local optima for some problems.
RMSprop	Suitable for non-stationary problems, dynamic learning rate.	May oscillate around the optimal value.
Adam	Combines momentum and RMSprop; fast and stable convergence.	Sensitive to hyperparameter tuning; poor performance with high noise.
AdamW	Fixes weight normalization issues in Adam; improves generalization.	Requires precise tuning of the weight decay coefficient.
Adamax	More stable than Adam for large gradients.	Slower convergence than Adam in certain problems.
NAdam	Combines Adam with Nesterov Momentum; faster and more accurate.	Higher computational complexity and parameter tuning needs.

و روابط غیرخطی را دارد و به دلیل ساختار ترتیبی خود، دقت نهایی بسیار بالایی ارائه می‌دهد. تنظیم دقیق نرخ یادگیری و تعداد درخت‌ها نقش کلیدی در جلوگیری از بیش‌برازش دارد [۱۰۳ و ۱۰۵].

الگوریتم برآزش بردار پشتیبان^۴

مدل برآزش بردار پشتیبان (SVR) بر پایه نظریه ماشین بردار پشتیبان توسعه یافته و هدف آن یافتن تابعی است که حداکثر حاشیه^۵ با حداقل خطا نسبت به داده‌های آموزشی داشته باشد [۱۰۶ و ۱۰۷]. با استفاده از کرنل‌هایی مانند RBF، قابلیت مدل‌سازی روابط غیرخطی پیچیده فراهم می‌شود. الگوریتم SVR به‌ویژه در مسائل با ابعاد بالا عملکرد پایدار و قابل اعتمادی ارائه می‌دهد، اما تنظیم بهینه پارامترهای آن برای دستیابی به دقت بهینه ضروری است [۱۰۸].

الگوریتم جنگل تصادفی^۱

جنگل تصادفی یک الگوریتم یادگیری جمعی^۲ از نوع bagging است که تعداد زیادی درخت تصمیم را با نمونه‌برداری تصادفی از داده‌ها و ویژگی‌ها آموزش می‌دهد [۹۹ و ۱۰۰]. خروجی نهایی براساس تجمیع (میانگین‌گیری در مسائل برآزش) پیش‌بینی‌های تک‌درخت‌ها به‌دست می‌آید. این ساختار باعث کاهش واریانس و افزایش پایداری مدل می‌شود و نسبت به نویز و داده‌های خارج از محدوده مقاومت بالایی دارد. الگوریتم جنگل تصادفی به‌طور ذاتی امکان ارزیابی اهمیت ویژگی‌ها را فراهم می‌کند [۱۰۱ و ۱۰۲].

الگوریتم گرادیان بوستینگ^۳

گرادیان بوستینگ از رویکرد Boosting استفاده می‌کند؛ به این معنی که به‌صورت مرحله‌ای مدل‌های ضعیف‌تر (معمولاً درخت‌های تصمیم کم‌عمق) را به‌گونه‌ای ایجاد می‌کند که خطای باقی‌مانده مدل قبلی را جبران کنند [۱۰۳ و ۱۰۴]. این روش قابلیت یادگیری الگوهای پیچیده

1. Random Forest (RF)

2. Ensemble

3. Gradient Boosting (GB)

4. Support Vector Regression (SVR)

5. Margin

ارزیابی‌های آماری

تابع اتلاف میانگین مربعات خطا^۱

هنگامی که یک سیگنال ورودی توسط مدل ماشین دریافت می‌شود، از طریق یک تابع تحت پردازش قرار می‌گیرد و پس از یک سری محاسبات، به یک مقدار خروجی تبدیل می‌شود. سپس مدل خروجی تولید شده را با مقدار خروجی واقعی مقایسه کرده و تابع اتلاف را محاسبه می‌کند. تابع اتلاف معیاری است که نشان می‌دهد مدل در یک کار مشخص چقدر خوب عمل می‌کند. یکی از توابع اتلاف محبوب یعنی MSE، مجذور اختلاف بین مقادیر را محاسبه می‌کند. در این تابع، فاصله میان مقدار واقعی داده (y_i) و مقدار پیش‌بینی شده توسط مدل (\hat{y}_i) با استفاده از مربع این فاصله محاسبه شده و در نهایت میانگین این مربعات برای تمام داده‌ها محاسبه می‌گردد. تابع اتلاف میانگین مربعات برای n داده به صورت رابطه ۱۱ تعریف می‌شود [۱۰۹].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (11)$$

تابع MSE پایین‌تر نشان‌دهنده تناسب بهتر بین پیش‌بینی‌های مدل و مقادیر واقعی است. در حالت ایده آل، مدلی با تناسب کامل دارای MSE صفر خواهد بود که نشان می‌دهد تمام پیش‌بینی‌ها دقیقاً با مقادیر واقعی مطابقت دارند. در مقابل، MSE بالاتر نشان می‌دهد که پیش‌بینی‌های مدل به‌طور قابل توجهی از واقعیت انحراف دارند. از مزایای این تابع اتلاف می‌توان به آسان بودن برای محاسبه و تفسیر اشاره کرد. این تابع اتلاف اندازه‌گیری جامعی از دقت مدل را ارائه می‌دهد. در مقابل از محدودیت‌های این تابع می‌توان به حساسیت نسبت به خطاهای بزرگ و کوچک اشاره کرد [۱۱۰].

تابع اتلاف L1^۲

میانگین خطای مطلق، به‌عنوان تابع اتلاف L1 نیز شناخته می‌شود و کارهای برآزشی کاربرد دارد. این تابع میانگین اختلاف مطلق بین مقادیر پیش‌بینی

شده y_i از یک مدل یادگیری ماشین و مقادیر هدف واقعی \hat{y}_i را محاسبه می‌کند. تابع اتلاف L1 برای n داده به صورت رابطه ۱۲ تعریف می‌شود [۱۱۱].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (12)$$

از مزایای این تابع می‌توان به اندازه‌گیری ساده و قابل تفسیر از میانگین اختلاف مطلق بین مقادیر پیش‌بینی شده و واقعی اشاره کرد. این تابع به‌طور گسترده‌ای در مسائل مختلف برآزشی برای ارزیابی دقت مدل‌ها و مقایسه با الگوریتم‌های مختلف استفاده می‌شود. استحکام آن نسبت به داده‌های خارج از محدوده باعث می‌شود که به‌ویژه در شرایطی که مقادیر شدید می‌توانند محاسبه خطا را به‌طور قابل توجهی تحت تأثیر قرار دهند، مفید باشد. از محدودیت‌های تابع اتلاف L1 آن است که با همه خطاها بدون توجه به جهت آنها به‌طور مساوی برخورد می‌کند. این تابع میان برآورد بیش از حد و برآورد کمتر از حد تفاوتی قائل نمی‌شود. در صورتی که جهت خطاها در مسئله بسیار مهم باشد، معیارهای دیگری مانند میانگین درصد مطلق خطا یا معیارهای جهت ممکن است مناسب‌تر باشند. از دیگر معایب این تابع نیز می‌توان به پیچیده‌تر بودن بهینه‌سازی (به‌عنوان مثال مشتق‌ناپذیری در نقطه صفر) اشاره کرد [۱۱۲]. به‌منظور حل برخی از مشکلات مذکور می‌توان از تابع Huber استفاده کرد [۱۱۳]. این تابع در رابطه ۱۳ و شکل ۲ نشان داده شده است.

$$L = \begin{cases} 0.5 \times (y_p - y_i)^2 & |y_p - y_i| < \delta \\ \delta \times (|y_p - y_i| - 0.5 \times \delta) & otherwise \end{cases} \quad (13)$$

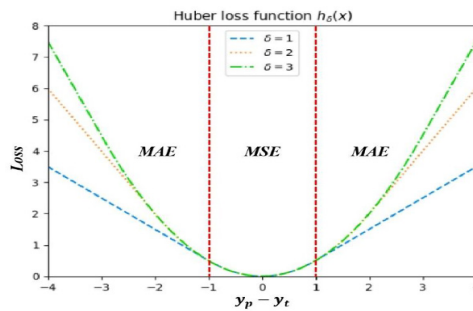
ضریب تعیین^۳

ضریب تعیین یا R^2 میزان دقت در پیش‌بینی یک مدل آماری را اندازه‌گیری می‌کند. کمترین مقدار ممکن برای R^2 صفر و بالاترین مقدار ممکن ۱ است. به زبان ساده، هر چه یک مدل در پیش‌بینی بهتر عمل کند، R^2 آن به ۱ نزدیک‌تر خواهد بود [۱۱۴].

1. Mean Squared Error Loss

2. Mean Absolute Error Loss

3. R2 Score (Coefficient of Determination)



شکل ۲ تابع اتلاف Huber.
Fig. 2 Huber loss function

۵، پارامترهای مدل‌سازی که در این مطالعه مورد بررسی قرار گرفته‌اند آورده شده است. لازم به ذکر است که شبکه MLP2، MLP1 و MLP3 به معنای شبکه عصبی پرسپترون چند لایه به ترتیب با یک، دو و سه لایه پنهان است. همان‌طور که پیش از این گفته شد، در این مطالعه نسبت داده‌های آموزش، اعتبارسنجی و آزمایش به‌منظور مدل‌سازی، ۷۵، ۱۵ و ۱۰٪ می‌باشد. همان‌طور که پیش از این گفته شد، در این مطالعه پارامترهای مختلفی مانند الگوریتم‌های بهینه‌سازی، توابع اتلاف، تعداد شبکه، تعداد نوروها و نرخ یادگیری مورد بررسی قرار گرفته و بهترین حالت ممکن انتخاب شده است. علاوه بر این، به‌منظور ارزیابی اثر تابع فعال‌سازی بر عملکرد مدل، چندین تابع از جمله Sigmoid و Tanh، نیز مورد آزمایش قرار گرفتند. نتایج اولیه نشان داد که توابع Sigmoid و Tanh باعث کندی فرایند یادگیری و بروز مشکل اشباع گرادیان شدند، و پایداری و دقت کمتری نسبت به ReLU داشتند. بر این اساس، تابع ReLU به‌عنوان گزینه نهایی انتخاب گردید، زیرا ضمن حفظ سرعت همگرایی بالا، دقت مدل را به شکل قابل توجهی بهبود داد. برای جلوگیری از طولانی شدن مقاله، نتایج مربوط به سایر توابع فعال‌سازی ارائه نشده‌اند.

محیط و تنظیمات اجرای مدل

در این پژوهش، پیاده‌سازی مدل در محیط نرم‌افزار پایتون^۲ انجام گرفت.

به‌منظور محاسبه ضریب تعیین در تحلیل برازش، ابتدا لازم است مقادیر پیش‌بینی شده مدل را با مقادیر واقعی داده‌ها مقایسه کرد و سپس مجموع مربعات اختلاف میان پیش‌بینی‌های مدل و مقادیر واقعی را با مجموع مربعات اختلاف میان میانگین داده‌های واقعی (به‌عنوان یک خط مبنای^۱ ساده) مقایسه کرد. در نهایت ضریب تعیین عبارت است از یک منهای نسبت مجموع مربعات خطای پیش‌بینی مدل به مجموع مربعات خطای پیش‌بینی مدل ساده با میانگین داده‌های واقعی که در رابطه ۱۴ ارائه شده است.

$$R^2 - \text{Score} = 1 - \left[\frac{MSE(\text{Model})}{MSE(\text{Baseline})} \right] \quad (14)$$

موردی وجود دارد که R^2 می‌تواند مقادیر منفی بدهد. معنای مقدار منفی برای ضریب تعیین آن است که مدل مورد نظر، بهتر از مدل ساده یعنی پیش‌بینی مقدار میانگین نیست. به‌عبارت دیگر، مدل مذکور بهتر از یک حد پایین نبوده و پیش‌بینی‌های آن بدتر از پیش‌بینی با میانگین داده‌های واقعی است. ضریب تعیین می‌تواند به‌طور شهودی نسبت به MAE، MSE و RMSE در ارزیابی تحلیل برازش آموزنده باشد، زیرا اولی را می‌توان به‌صورت درصد بیان کرد، درحالی‌که معیارهای دوم دارای محدوده‌های دلخواه هستند [۱۱۵].

مدل کردن ویسکوزیته توسط شبکه عصبی پرسپترون چند لایه

در این بخش سه مدل شبکه عصبی چند لایه تعریف شده است. تفاوت این سه مدل با یکدیگر، در تعداد لایه‌های پنهان است. در جدول

1. Baseline
2. Python

جدول ۵ پارامترهای مورد استفاده در مدل‌سازی شبکه عصبی پرسپترون چند لایه.
Table 5. Parameters used in multilayer perceptron neural network modeling.

Optimization functions	AdaGrad, Adadelata, RMSprop, Rprop, SGD, ASGD, Adam, AdamW, Adamax, NAdam
Learning rates	0.001, 0.01, 0.1
Network type	MLP1, MLP2, MLP3
Parameters for measuring model accuracy and error	(R ² , MSE, L1 (MAE
Activation function type	ReLU
Number of periods	25-800

مبتنی بر کتابخانه‌های استاندارد پایتون (PyTorch و Scikit-learn) به صورت تخمینی و میانگین چند اجرا اندازه‌گیری شدند. در این چارچوب، زمان آموزش به‌عنوان مدت زمان لازم برای همگرایی مدل بر روی داده‌های آموزشی، و زمان انجام پیش‌بینی به صورت زمان موردنیاز برای انجام ۱۰۰۰ پیش‌بینی متوالی تعریف شده است. **جدول ۶** مقادیر حاصل را ارائه می‌دهد. لازم به تأکید است که این مقادیر با هدف مقایسه نسبی بازده محاسباتی مدل‌ها گزارش شده‌اند و بسته به نوع پیاده‌سازی، تغییر اندازه مجموعه داده یا استفاده از سخت‌افزار قدرتمندتر (به‌ویژه GPU)، قابل تغییر خواهند بود.

نتایج و بحث

ماتریس همبستگی

در این بخش به منظور ارزیابی داده‌ها از جهت استقلال و وابستگی، مجموعه داده‌ها توسط ماتریس همبستگی مورد بررسی قرار گرفته است. لازم به توضیح است که در فرآیند مدل‌سازی برای اهداف این تحقیق می‌بایست پارامترهای ورودی با همدیگر حداقل ارتباط و همبستگی و با پارامتر خروجی همبستگی بالایی داشته باشند. به عبارت دیگر پارامترهای ورودی می‌بایست مستقل از همدیگر باشند ولی نسبت به پارامتر هدف (خروجی) وابستگی داشته باشند.

پایتون به دلیل داشتن کتابخانه‌های قدرتمند در زمینه یادگیری ماشین، تحلیل عددی و مصورسازی داده‌ها، یکی از ابزارهای اصلی برای توسعه مدل‌های هوش مصنوعی در مهندسی نفت محسوب می‌شود. در این راستا، از کتابخانه‌ها NumPy و Pandas برای مدیریت و پردازش داده‌ها، از Matplotlib برای ترسیم نمودارها، و از کتابخانه scikit-learn جهت ساخت، آموزش و ارزیابی مدل شبکه عصبی بهره گرفته شد.

کارایی زمان مدل‌ها^۱

به منظور ارزیابی کارایی محاسباتی مدل‌ها از نظر زمان اجرا، زمان آموزش^۲ و زمان انجام پیش‌بینی^۳ برای هر یک از مدل‌های توسعه‌یافته اندازه‌گیری و گزارش شد. مجموعه داده مورد استفاده شامل ۲۲۹۹ نمونه بود که مطابق با رویه‌های متداول یادگیری ماشین به سه بخش ۷۵٪ برای آموزش، ۱۵٪ برای اعتبارسنجی و ۱۰٪ برای آزمایش تفکیک شد (به ترتیب ۱۷۲۴، ۳۴۵ و ۲۳۰ نمونه). مدل شبکه عصبی پرسپترون چندلایه با سه لایه پنهان و ده تابع بهینه‌سازی مختلف شامل SGD، ASGD، Rprop، Adagrad، Adadelata، RMSprop، Adam، Adam-، W، Adamax و Nadam آموزش داده شد. علاوه بر این، سه مدل یادگیری ماشین شناخته‌شده شامل جنگل تصادفی (Random Forest)، گرادیان بوستینگ (Gradi-Boosting) و برآزش بردار پشتیبان (SVR) نیز برای مقایسه افزوده شدند. تمامی زمان‌ها بر روی سیستم مرجع Intel(R) Core (TM) i5-7200U CPU @ 2.50GHz (بدون GPU اختصاصی) و با پیاده‌سازی

1. Run Time
 2. Training Time
 3. Inference Time

جدول ۶ مدت زمان آموزش و پیش‌بینی مدل‌های مورد بررسی.

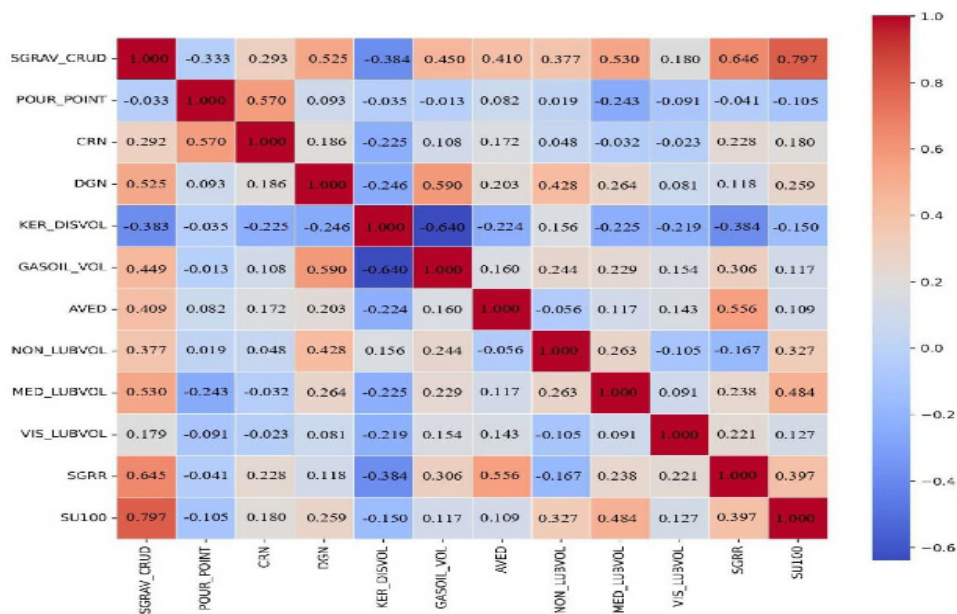
Table 6. Training and prediction times of studied models

Model	Training Time (s)	Prediction Time (ms)
MLP-AdaGrad	68	45
MLP-Adadelta	70	45
MLP-RMSprop	60	44
MLP-Rprop	60	44
MLP-SGD	75	45
MLP-ASGD	80	45
MLP-Adam	40	40
MLP-AdamW	45	40
MLP-Adamax	48	40
MLP-Nadam	42	40
Random Forest (100 trees, scikit-learn)	12	60
Gradient Boosting (100 estimators, scikit-learn)	30	35
SVR (RBF kernel, scikit-learn)	180	150

دارا می‌باشند. لازم به توضیح است که حد کمینه قابل قبول برای وجود همبستگی بین پارامترهای ورودی با پارامتر هدف در این مطالعه ۱۰٪ در نظر گرفته شده است. در این پژوهش، تعیین حدود همبستگی میان پارامترهای ورودی و همچنین بین پارامترهای ورودی و پارامتر هدف (گرانروی) با هدف اطمینان از کفایت آماری داده‌ها و جلوگیری از بروز هم‌خطی در مدل انجام شد. حد بیشینه ۷۰٪ برای میزان همبستگی بین پارامترهای ورودی بر مبنای اصل استقلال نسبی متغیرها در مدل‌های یادگیری ماشین انتخاب شده است؛ چراکه همبستگی‌های بالاتر از این مقدار می‌توانند منجر به هم‌پوشانی اطلاعات و کاهش دقت تعمیم مدل شوند. از سوی دیگر، حد کمینه ۱۰٪ برای همبستگی بین پارامترهای ورودی و پارامتر هدف با هدف حذف متغیرهای کم‌اثر و غیرمؤثر در پیش‌بینی گرانروی تعیین گردید تا تنها متغیرهایی که سهم قابل توجهی در تبیین تغییرات گرانروی دارند در مدل لحاظ شوند. به‌منظور بررسی بهتر پارامتر مورد بررسی در این مطالعه، نمودار هیستوگرام مجموعه داده‌های گرانروی در دمای ۱۰۰ °F در شکل ۵ نشان داده شده است.

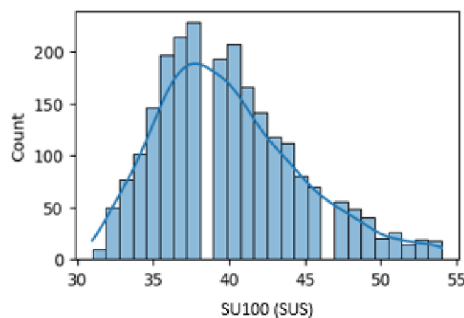
هرچه وابستگی پارامترهای ورودی با پارامتر هدف (در اینجا گرانروی) بالاتر و در مقابل وابستگی پارامترهای ورودی با همدیگر کمتر باشد، نتیجه مدل‌سازی بهتر خواهد بود. ماتریس همبستگی برای مجموعه داده‌های به‌کار رفته در مدل‌سازی در شکل ۳ نشان داده شده است. در نمودار نقشه گرمایی^۱ نمایش داده شده در شکل ۳، هرچه بخش مربوطه به رنگ قرمز پررنگ نزدیک‌تر باشد، به معنای وابستگی بالای داده‌ها بهم بوده و هرچه بخش مورد نظر به رنگ آبی پررنگ نزدیک‌تر باشد، به معنای استقلال دو پارامتر از همدیگر است. خط قطری قرمز رنگ در نمودار نقشه گرمایی، ارتباط مستقیم دو پارامتر یکسان به همدیگر را نشان می‌دهد و مقدار آن برابر یک می‌باشد. همان‌طور که به‌صورت بصری مشاهده می‌شود، پارامترهای ورودی ارتباط زیادی با همدیگر نداشته و مستقل از یکدیگرند. حد بیشینه قابل قبول برای مستقل بودن دو پارامتر ورودی در این مطالعه، ۷۰٪ در نظر گرفته شده است. مقادیر همبستگی پارامترهای ورودی با پارامتر هدف (گرانروی) به‌صورت خاص در شکل ۴ نشان داده شده است. همان‌طور که در شکل ۴ مشاهده می‌شود، وزن مخصوص نفت خام بالاترین میزان همبستگی و نقطه ریزش نفت کمترین میزان همبستگی با گرانروی را

1. Heat Map



شکل ۳ ماتریس همبستگی برای پارامترهای مختلف

Figure 3. Correlation matrix for different parameters



شکل ۵ نمودار هیستوگرام داده‌های گرانیوی در دمای ۱۰۰ °F.

Fig. 5 Histogram plot of viscosity data at 100 °F.



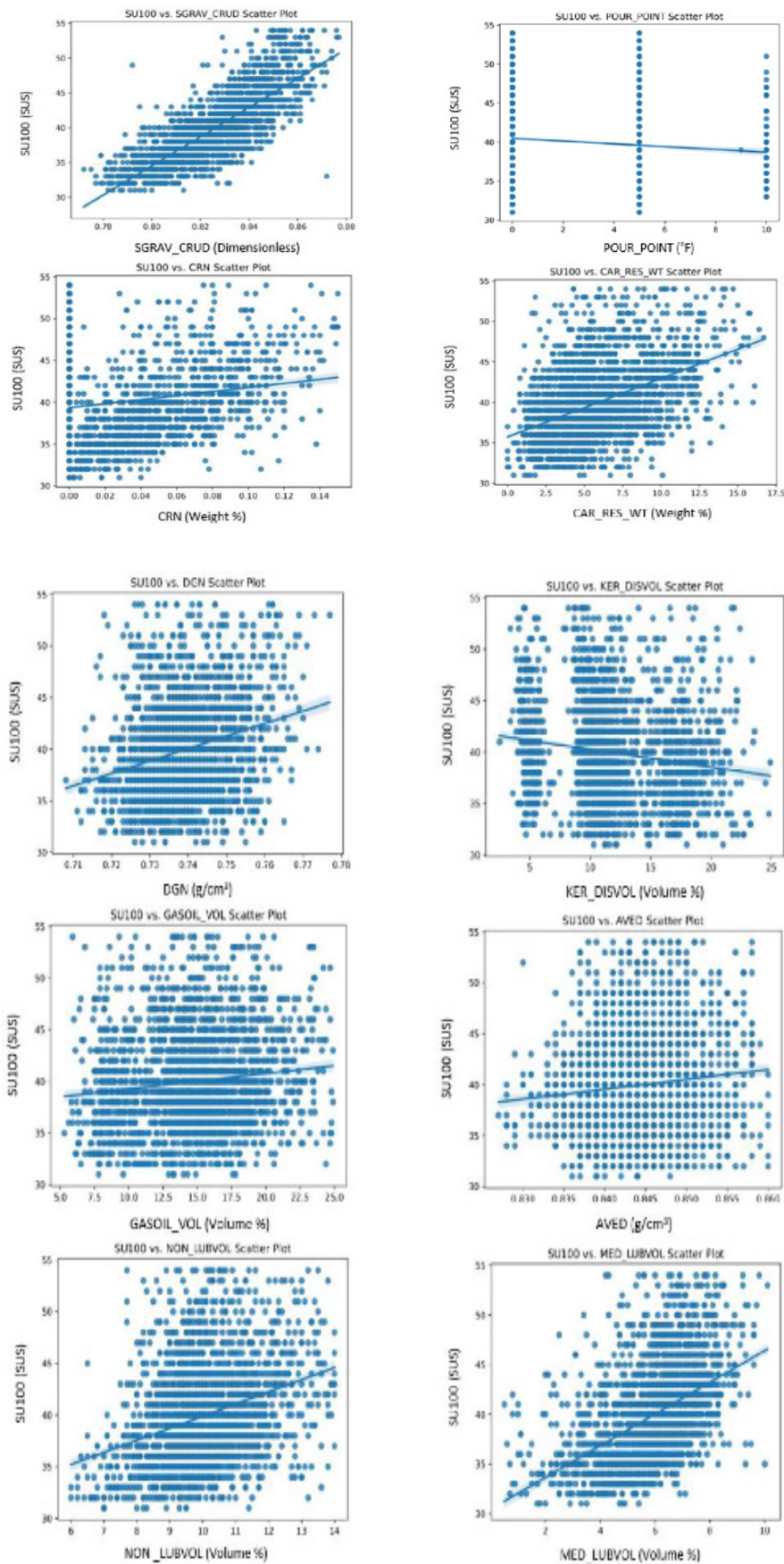
شکل ۴ ماتریس هم بستگی پارامترهای مختلف ورودی نسبت به گرانیوی در دمای ۱۰۰ °F.

Fig. 4 Correlation matrix of various input parameters with respect to viscosity at 100 °F.

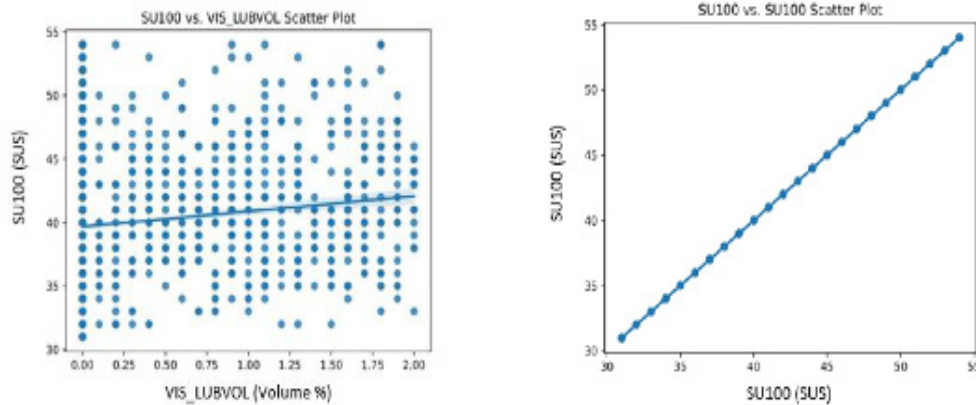
میزان ارتباط خطی میان دو پارامتر گرانیوی و سایر داده‌های ورودی است. در این نمودار همچنین، توزیع نقاط نشان داده شده است و می‌توان تخمینی از دامنه گرانیوی در آزمایش‌ها برای هر داده ورودی به دست آورد. براساس شکل ۶، پارامتر گرانیوی با SGRAV_CRUD بیشترین ارتباط خطی را دارا می‌باشد. این پارامتر وزن مخصوص نفت خام بوده و مقدار همبستگی آن با گرانیوی براساس نتایج قبلی، ۰/۷۹۷ می‌باشد.

نمودار پراکندگی داده‌های مورد استفاده جهت مدل‌سازی گرانیوی نفت مرده

در این بخش نمودار پراکندگی بین گرانیوی نفت مرده و باقی داده‌های ورودی ترسیم شده است. همان‌طور که در شکل ۶ مشاهده می‌شود، هر نقطه از این نمودار نمایانگر یک آزمایش و موقعیت آن براساس گرانیوی و سایر ویژگی‌ها است. خطی که به صورت خاص مشاهده می‌شود، نشان‌دهنده



شکل ۶ نمودار گرانیروی در دمای ۱۰۰ °F برحسب ویژگی‌های ورودی مختلف.
 Fig. 6 Viscosity plot at 100 °F for different input characteristics.



ادامه شکل ۶

توزیع در دو سمت به‌طور متقارن کاهش می‌یابد. این موضوع نشان‌دهنده توزیع متعادل داده‌ها و نبود داده‌های خارج از محدوده در این ویژگی است. در مجموع، تحلیل توزیع چگالی احتمال نشان می‌دهد که داده‌های اصلی از توزیع نسبتاً منظم و آماری پایدار برخوردارند. این ویژگی باعث بهبود پایداری مدل در فرآیند آموزش شبکه عصبی شده و خطر بیش‌برازش را کاهش می‌دهد. همچنین، وجود چولگی اندک در پارامتر گرانشی تأیید می‌کند که استفاده از الگوریتم‌های بهینه‌سازی تطبیقی مانند AdamW در مراحل بعدی مدل‌سازی انتخاب مناسبی بوده است.

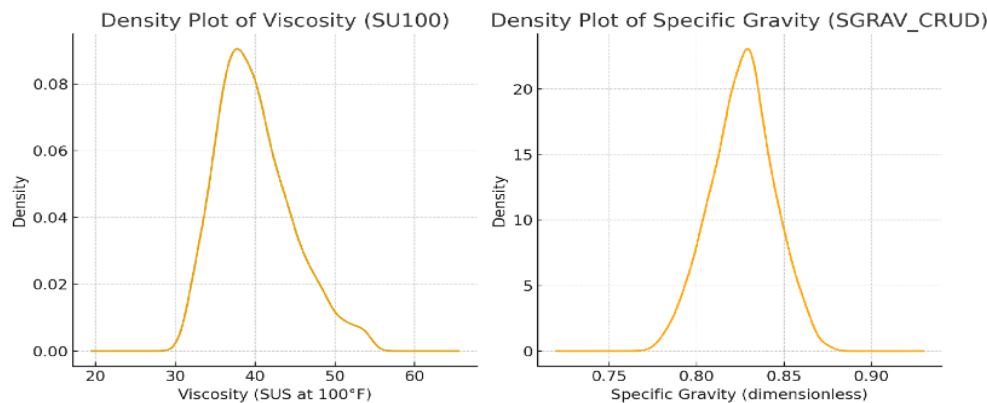
تحلیل اکتشافی داده‌ها^۲

به‌منظور درک بهتر از ویژگی‌های آماری و روابط پنهان میان متغیرهای ورودی و خروجی، تحلیل اکتشافی داده‌ها بر روی مجموعه داده پالایش شده انجام شد. هدف از این بخش، بررسی توزیع داده‌ها، نرمال بودن، وجود داده‌های خارج از محدوده، و شناسایی روندها و همبستگی‌های مؤثر میان پارامترها پیش از مدل‌سازی است. در گام نخست، مقادیر آماری اساسی شامل میانگین، میانه، انحراف معیار، چولگی و کشیدگی برای تمامی متغیرهای ورودی و خروجی محاسبه شد (جدول ۳).

تحلیل توزیع چگالی داده‌ها^۱

به‌منظور بررسی الگوی توزیع داده‌ها و ارزیابی نرمال بودن متغیرهای کلیدی، از نمودار چگالی احتمال استفاده شد. این نمودارها که با روش Kernel Density Estimation (KDE) ترسیم شده‌اند، شکل تقریبی توزیع داده‌ها را در فضای پیوسته نشان می‌دهند. در شکل ۷، توزیع چگالی دو پارامتر اصلی یعنی گرانشی در 100°F و وزن مخصوص نفت خام نمایش داده شده است. نتایج نشان می‌دهد که توزیع گرانشی تقریباً نرمال بوده ولی دارای چولگی مثبت خفیف است؛ به این معنا که اغلب نمونه‌های نفتی دارای گرانشی پایین‌تر از میانگین هستند، درحالی‌که تعداد محدودی از نمونه‌ها با گرانشی بالا وجود دارند. این رفتار با مقدار عددی چولگی ($\text{Skew} = +0/67$) در جدول آماری نیز همخوانی دارد. در نتیجه، منحنی توزیع چگالی داده‌ها گرانشی دارای قله‌ای در حدود ۳۸ تا ۴۰ SUS بوده و دم آن به سمت مقادیر بالاتر گسترش یافته است. چنین توزیعی بیانگر آن است که داده‌های گرانشی، تمایل به تمرکز در محدوده میانی دارند اما نمونه‌های سنگین‌تر با گرانشی بالا نیز در مجموعه داده حضور دارند. در مقابل، توزیع وزن مخصوص نفت خام کاملاً متقارن و نزدیک به نرمال مشاهده می‌شود. منحنی نمودار چگالی احتمال این متغیر دارای یک قله یکنواخت در حدود مقدار میانگین ($0/83 \approx$) است و دم‌های

1. Probability Density Function Analysis (PDF)
2. Exploratory Data Analysis (EDA)



شکل ۷ توزیع چگالی دو پارامتر اصلی این مطالعه شامل گرانیروی در ۱۰۰ °F و وزن مخصوص نفت خام.
Fig. 7 Density distribution of the two main parameters of this study including viscosity at 100 °F and specific gravity of crude oil.

مقادیر بسیار کوچک است. بررسی‌های بیشتر متغیرهای کلیدی نشان داد که ویژگی‌های مربوط به برش‌های روان‌کننده (NON_LUBVOL و MED_LUBVOL) دارای چند داده خارج از محدوده هستند که احتمالاً به ترکیب‌های غیرمعمول نفت‌های سنگین یا نمونه‌های دارای آسفالتن زیاد مربوط می‌شوند. حذف یا تعدیل این داده‌ها در مرحله پیش‌پردازش، به پایداری بهتر مدل در آموزش کمک کرده است. در ادامه، برای تحلیل روابط بین متغیرها، ماتریس همبستگی پیرسون محاسبه شد. نتایج آن شکل ۳ و ۴ نشان داد که پارامتر SGRAV_CRUDE بیشترین همبستگی مثبت را با گرانیروی در دمای ۱۰۰ °F ($r=0.79$) دارد، درحالی‌که POUR_POINT کمترین مقدار همبستگی را نشان می‌دهد. این موضوع تأکید می‌کند که چگالی و ترکیب برش‌های سنگین نقش اصلی را در تغییرات گرانیروی دارند. همچنین، مقدار بیشینه قابل قبول برای همبستگی میان پارامترهای ورودی ۰/۷ و کمینه قابل قبول میان ورودی‌ها و خروجی ۰/۱ در نظر گرفته شده است تا از هم‌خطی و افزونگی اطلاعات جلوگیری شود. برای بررسی شکل توزیع داده‌ها، نمودار چگالی احتمال برای پارامترهای کلیدی از جمله گرانیروی در دمای ۱۰۰ °F و وزن مخصوص نفت خام ترسیم شد.

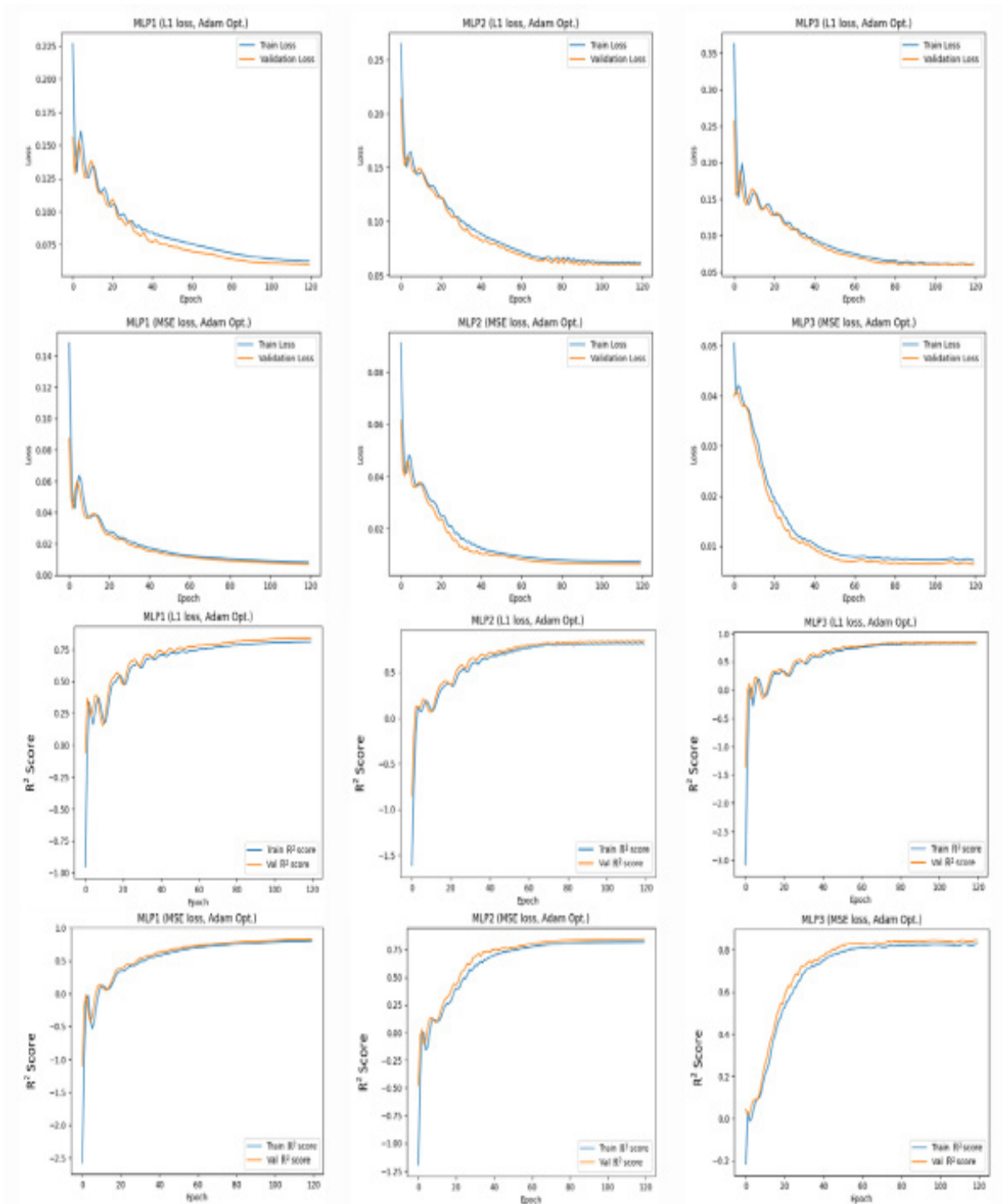
نتایج نشان داد که اغلب متغیرها مانند SGRAV_CRUDE، AVED، GASOIL_VOL دارای چولگی کم و توزیع نسبتاً نرمال هستند، درحالی‌که متغیرهایی مانند CRN و VIS_LUBVOL دارای چولگی قابل توجه و توزیع دم‌بلند می‌باشند. این موضوع بیانگر احتمال وجود داده‌های خارج از محدوده یا مقادیر غیرمعمول در برخی ویژگی‌ها است. نتایج نشان داد که اغلب متغیرها مانند SGRAV_CRUDE، AVED، GASOIL_VOL دارای چولگی کم و توزیع نسبتاً نرمال هستند، درحالی‌که متغیرهایی مانند CRN و VIS_LUBVOL دارای چولگی قابل توجه و توزیع دم‌بلند می‌باشند. این موضوع بیانگر احتمال وجود داده‌های خارج از محدوده یا مقادیر غیرمعمول در برخی ویژگی‌هاست. بررسی تفاوت میان میانگین و میانه در تعدادی از متغیرها نیز نشان داد که توزیع برخی پارامترها متقارن نبوده و از الگوی نرمال انحراف دارند. به‌طور خاص، برای متغیر CRN که نشان‌دهنده میزان نیتروژن در نفت خام است، مقدار چولگی برابر با ۱/۲۶۸ گزارش شده و این مقدار بالا تأییدکننده وجود چند نمونه با مقادیر نیتروژن غیرمعمول است. همچنین، متغیر VIS_LUBVOL دارای چولگی منفی زیاد (۲/۰۵۸) بوده که نشان‌دهنده تمرکز داده‌ها در مقادیر پایین و وجود چند نمونه با

این نتایج برای شبکه‌هایی با یک، دو و سه لایه پنهان (MLP1، MLP2 و MLP3) محاسبه شده است. همان‌طور که پیش از این اشاره شد، تابع اتلاف MSE و تابع اتلاف MAE از پرکاربردترین توابع در یادگیری ماشین هستند. هر دو تابع اتلاف برای محاسبه خطا در پیش‌بینی مقدار یک متغیر پیوسته استفاده می‌شوند. تابع اتلاف MSE خطا را با میانگین مربعات فاصله بین مقدار پیش‌بینی‌شده و مقدار واقعی محاسبه می‌کند. این تابع برای شبکه‌های عصبی کارآمد است زیرا گرادینت‌ها را به‌صورت پیوسته محاسبه می‌کند و یادگیری را تسریع می‌بخشد. تابع اتلاف MAE خطا را با میانگین قدرمطلق اختلاف بین پیش‌بینی و مقدار واقعی محاسبه می‌کند و در مسائل دارای نویز بیشتر عملکرد بهتری از MSE دارد. به‌طور کلی، انتخاب تابع اتلاف به نوع داده و میزان نویز آن بستگی دارد. اگر داده‌ها نویز زیادی داشته باشند، MAE گزینه مناسب‌تری است، اما در داده‌های دقیق‌تر و با نویز کمتر، MSE عملکرد بهتری دارد. با این حال، در بسیاری از مسائل، استفاده هم‌زمان از هر دو تابع نیز متداول است. در این بخش، علاوه‌بر مقایسه توابع بهینه‌سازی مختلف، مقایسه‌ای میان عملکرد توابع اتلاف MSE و MAE نیز انجام شده است. بررسی شکل‌های ۸ و ۹ برای الگوریتم‌های Adam و AdamW نشان می‌دهد که این الگوریتم‌ها عملکرد بسیار مطلوبی داشته و پیش‌بینی‌های دقیقی ارائه داده‌اند. نتایج مربوط به ۱۰ نوع تابع بهینه‌سازی مختلف، در جدول ۷ ذکر شده است. همان‌طور که در جدول ۷ مشاهده می‌شود، الگوریتم‌های بهینه‌سازی Adam، AdamW و Adamax بهترین عملکرد را در بین سایر الگوریتم‌ها از خود نشان داده‌اند، به‌طوری‌که کمترین مقدار تابع اتلاف (Loss) و بالاترین ضریب تعیین (R^2) را ایجاد کرده‌اند. با این حال، الگوریتم AdaGrad نیز بر روی داده‌های آموزشی و اعتبارسنجی عملکرد نسبتاً قابل‌قبولی داشته و توانسته است در مقایسه با توابع کلاسیک‌تر مانند SGD و ASGD به همگرایی مناسب‌تری برسد.

همان‌گونه که پیش از این در شکل ۷ مشاهده شد، منحنی چگالی احتمال گرانشی دارای توزیع تقریباً نرمال با چولگی مثبت خفیف ($Skewness = +0/67$) است. این موضوع بیانگر آن است که اکثر نمونه‌های نفتی دارای گرانشی پایین‌تر از میانگین بوده و تعداد محدودی از نمونه‌ها با گرانشی بالا وجود دارند. در مقابل، توزیع وزن مخصوص نفت خام کاملاً متقارن و نزدیک به نرمال است ($Skewness = -0/76$)، که نشان‌دهنده پایداری و تعادل آماری در داده‌های این متغیر می‌باشد. در مرحله نهایی، داده‌ها برای جلوگیری از سوگیری ناشی از تفاوت در مقیاس، با استفاده از روش Min-Max نرمال‌سازی شدند تا تمامی ویژگی‌ها در بازه pp قرار گیرند. این فرایند موجب هم‌مقیاس‌سازی داده‌ها و بهبود همگرایی مدل شبکه عصبی پرسپترون چندلایه گردید. به‌طور کلی، تحلیل اکتشافی داده‌ها نشان داد که مجموعه داده مورد استفاده از توزیع آماری مناسب و روابط معنی‌دار میان متغیرها برخوردار است. وجود چولگی اندک در متغیر گرانشی و توزیع متعادل سایر ویژگی‌ها، بیانگر کفایت آماری داده‌ها برای آموزش مدل‌های یادگیری ماشین بوده و نشان می‌دهد که ساختار داده‌ها برای مدل‌سازی دقیق گرانشی نفت خام، بهینه و پایدار است.

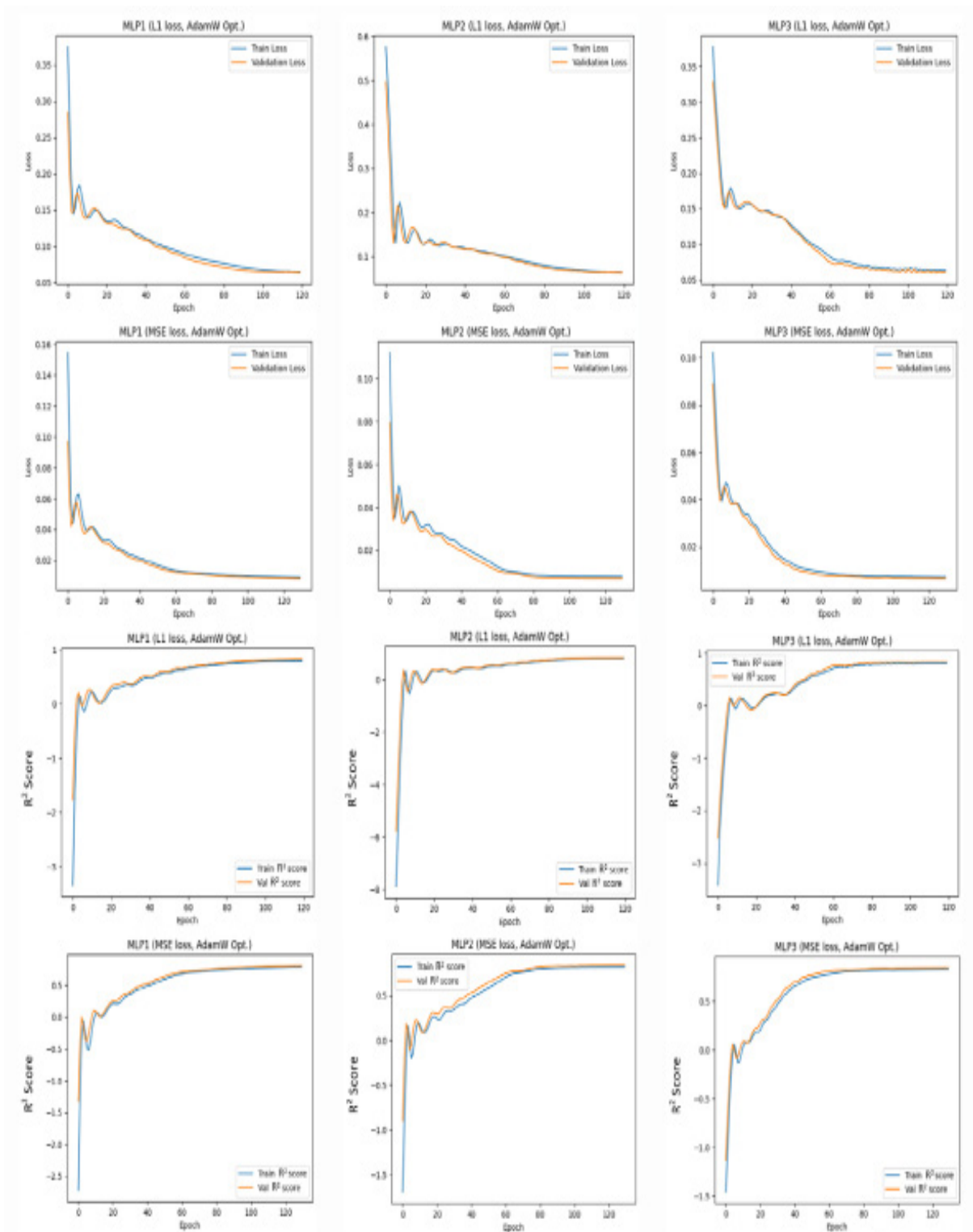
بررسی توابع بهینه‌سازی مختلف در شبکه عصبی پرسپترون چند لایه

الگوریتم‌های بهینه‌سازی از یک روند سیستماتیک برای یافتن پاسخ بهینه به مسئله تعریف‌شده استفاده می‌کنند. این کار از طریق کاوش تکرارشونده در فضای جست‌وجو انجام می‌شود. الگوریتم‌های بهینه‌سازی معمولاً بر پایه تکنیک‌های ریاضی عمل می‌کنند و پاسخ به‌دست‌آمده می‌تواند قطعی یا احتمالی باشد. در ادامه، نتایج حاصل از به‌کارگیری الگوریتم‌های بهینه‌سازی مختلف در این مطالعه بررسی شده است. در شکل‌های ۸ و ۹ نتایج مدل‌سازی داده‌های آموزشی و اعتبارسنجی گرانشی برحسب R^2 و Loss برای دو تابع اتلاف MSE و MAE (L1) برای دو تابع بهینه‌سازی Adam و AdamW نشان داده شده است.



شکل ۸ نتایج مدل‌سازی داده‌های آموزشی و اعتبارسنجی گرانیوی با تابع بهینه‌سازی Adam برحسب Loss و R^2 برای دو تابع اتلاف MSE و MAE.

Fig. 8 Results of training and validation data modeling of viscosity with Adam optimization function in terms of Loss and R^2 for two loss functions including MSE and MAE.



شکل ۹ نتایج مدل‌سازی داده‌های آموزشی و اعتبارسنجی گرانروی با تابع بهینه‌سازی AdamW برحسب Loss و R^2 برای دو تابع اتلاف MAE و MSE.

Fig. 9 Results of training and validation data modeling of viscosity with AdamW optimization function in terms of Loss and R^2 for two loss functions including MSE and MAE.

جدول ۷ نتایج کلی تمام توابع بهینه‌سازی مورد استفاده در مدل‌سازی گرانیروی توسط شبکه عصبی پرسپترون چند لایه.

Table 7 Overall results of all optimization functions used in viscosity modeling via multilayer perceptron neural network.

Optimizer	Loss Function	Network Type	Epoch	R ² -Train	R ² -Test	Loss-Train	Loss-Test
AdaGrad	MSE	MLP1	50	0.4444	0.4352	0.0231	0.0252
		MLP2	70	0.6091	0.6078	0.0162	0.0175
		MLP3	75	0.6734	0.6897	0.0136	0.0138
	MAE	MLP1	50	0.3922	0.3626	0.1214	0.1261
		MLP2	70	0.5969	0.6511	0.0935	0.0926
		MLP3	75	0.6260	0.6506	0.0984	0.0913
Adadelta	MSE	MLP1	650	0.2238	0.1869	0.0322	0.0362
		MLP2	650	0.1529	0.0991	0.0352	0.0402
		MLP3	700	0.0245	0.0232	0.0405	0.0435
	MAE	MLP1	600	0.2392	0.2252	0.1365	0.1413
		MLP2	650	0.2003	0.1279	0.1399	0.1498
		MLP3	800	0.2149	0.1817	0.1395	0.1464
RMSprop	MSE	MLP1	25	0.1921	0.2806	0.0336	0.0321
		MLP2	25	0.1459	0.1215	0.0355	0.0392
		MLP3	25	0.0350	0.0401	0.0401	0.0428
	MAE	MLP1	25	0.2790	0.3429	0.1329	0.1295
		MLP2	25	0.0630	0.1498	0.1533	0.1498
		MLP3	25	0.0829	0.1736	0.1497	0.1454
Rprop	MSE	MLP1	50	0.7257	0.7348	0.0114	0.0118
		MLP2	50	0.7554	0.6923	0.0102	0.0137
		MLP3	55	0.6587	0.7326	0.0142	0.0119
	MAE	MLP1	50	0.7262	0.7295	0.0785	0.0786
		MLP2	50	0.7302	0.7424	0.0776	0.0763
		MLP3	60	0.8023	0.8194	0.0647	0.0627
SGD	MSE	MLP1	30	0.1471	0.1785	0.0354	0.0366
		MLP2	50	-0.1174	-0.1320	0.0464	0.0505
		MLP3	70	-0.0197	-0.0486	0.0423	0.0467
	MAE	MLP1	25	-0.0406	0.0354	0.1555	0.1554
		MLP2	50	-0.0090	-0.0156	0.1553	0.1616
		MLP3	70	-0.0291	-0.0493	0.1596	0.1652
ASGD	MSE	MLP1	50	0.1709	0.1799	0.0344	0.0366
		MLP2	60	0.2081	0.2178	0.0329	0.0349
		MLP3	80	0.0870	0.0758	0.0379	0.0412
	MAE	MLP1	30	0.0092	-0.0174	0.1580	0.1640
		MLP2	50	0.1536	0.1362	0.1442	0.1498
		MLP3	60	-0.0569	-0.0789	0.1619	0.1674
Adam	MSE	MLP1	120	0.8036	0.8262	0.0082	0.0077
		MLP2	120	0.8210	0.8432	0.0074	0.0070
		MLP3	120	0.8279	0.8457	0.0071	0.0069
	MAE	MLP1	120	0.8140	0.8355	0.0626	0.0602
		MLP2	120	0.8135	0.8438	0.0618	0.0588
		MLP3	120	0.8161	0.8409	0.0610	0.0596

ادامه جدول ۷

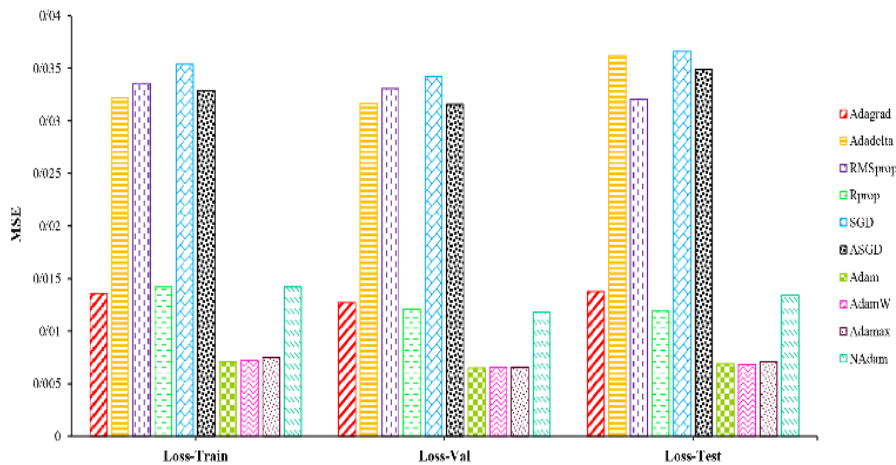
AdamW	MSE	MLP1	130	0.7863	0.8006	0.0089	0.0089
		MLP2	130	0.8159	0.8385	0.0076	0.0072
		MLP3	130	0.8269	0.8480	0.0072	0.0068
	MAE	MLP1	120	0.7991	0.8184	0.0644	0.0633
		MLP2	120	0.8037	0.8266	0.0637	0.0621
		MLP3	120	0.8059	0.8390	0.0633	0.0602
Adamax	MSE	MLP1	140	0.7944	0.8053	0.0085	0.0087
		MLP2	140	0.8140	0.8346	0.0077	0.0074
		MLP3	160	0.8192	0.8401	0.0075	0.0071
	MAE	MLP1	130	0.7937	0.8086	0.0670	0.0650
		MLP2	140	0.8173	0.8356	0.0614	0.0603
		MLP3	140	0.8170	0.8377	0.0612	0.0602
Nadam	MSE	MLP1	60	0.6851	0.6810	0.0131	0.0142
		MLP2	80	0.6584	0.7003	0.0142	0.0134
		MLP3	80	0.5477	0.6316	0.188	0.0164
	MAE	MLP1	60	0.5828	0.6728	0.0950	0.0942
		MLP2	80	0.5625	0.6830	0.0970	0.0942
		MLP3	100	0.5972	0.7028	0.0931	0.0896
Constants	Network type: MLP Learning rate: 0.01 Number of neurons in the first, second and third hidden layers: 30, 20 and 10 Activation function type: ReLu						

نیست. این تابع در این مطالعه و برای تعداد دوره‌های بسیار بالا بین ۶۰۰ تا ۸۰۰ برای داده‌های آموزشی و اعتبارسنجی، قادر است به همگرایی برسد. بررسی نتایج برای الگوریتم RMSprop نیز نشان داد که این تابع در تعداد دوره پایین به همگرایی می‌رسد و از آنجا که مقادیر R^2 برای این تابع پایین است، عملکرد مناسبی ندارد. بررسی الگوریتم Rprop نشان می‌دهد که این تابع از دقت بالاتری برخوردار است. براساس یافته‌های این مطالعه الگوریتم‌های SGD و ASGD عملکرد ضعیف‌تری نسبت به الگوریتم Rprop دارند. در این الگوریتم‌ها مقادیر منفی برای R^2 به دست آمده است و همان‌طور که پیش از این گفته شد، معنای مقدار منفی برای ضریب تعیین آن است که مدل مورد نظر، بهتر از مدل ساده یعنی پیش‌بینی مقدار میانگین نیست. به عبارت دیگر، مدل مذکور بهتر از یک حد پایین نبوده و پیش‌بینی‌های آن بدتر از پیش‌بینی با میانگین داده‌های واقعی است.

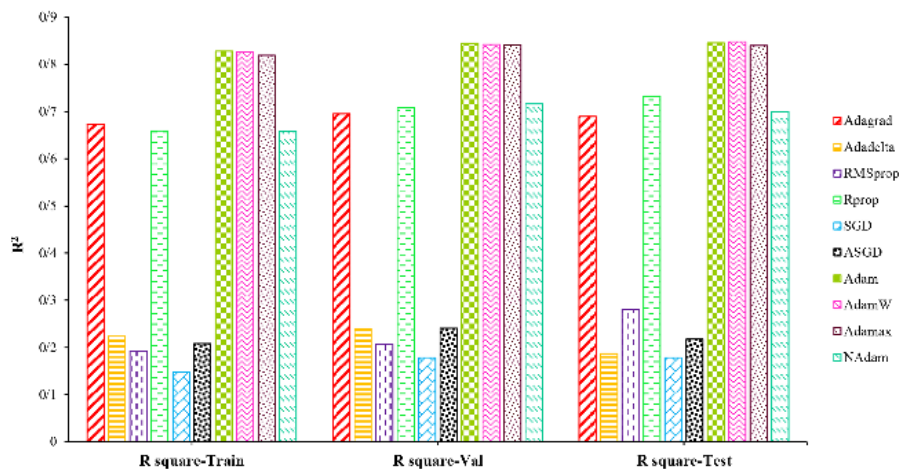
این موضوع نشان می‌دهد که الگوریتم AdaGrad در مسائلی با گرادین‌های پراکنده و تغییرات بزرگ می‌تواند عملکرد مناسبی داشته باشد، هرچند در این مطالعه، الگوریتم‌های خانواده Adam (شامل AdamW، Adamax و Adam) به‌طور کلی نتایج دقیق‌تر و پایاتری ارائه داده‌اند. در شرایطی که داده‌ها پراکنده‌گی زیادی دارند و گرادین‌ها متغیرهای بزرگی را شامل می‌شوند، AdaGrad عملکرد مناسبی دارد. در مقابل، در شرایطی که گرادین‌ها تغییرات کوچک‌تری دارند و تابع هزینه یکنواخت‌تر است، بهینه‌سازهای خانواده Adam همگرایی سریع‌تر و پایدارتری دارند. بنابراین، نتایج این پژوهش نیز تأیید می‌کند که استفاده از الگوریتم‌های مبتنی بر Adam برای داده‌های مربوط به ویژگی‌های فیزیکی نفت‌خام، که الگوهای نسبتاً منظم‌تری دارند، انتخاب بهتری است. بررسی نتایج نشان می‌دهد که، الگوریتم بهینه‌سازی Adadelta، عملکرد موفق‌تری نداشته است و مقادیر R^2 رضایت بخش

در فرآیند آموزش سبب می‌شود که مدل پیش از همگرایی کامل بر روی داده‌های آموزش متوقف گردد، در حالی که ممکن است بر روی داده‌های آزمایش عملکرد اندکی بهتر نشان دهد. در مجموع، اختلاف مقادیر یادشده در محدوده قابل قبول بوده و نشان‌دهنده تعمیم مناسب مدل است، نه بیش‌برازش^۱ یا کم‌برازش^۲. در شکل‌های ۱۰ و ۱۱ می‌توان مقادیر اتلاف و ضریب تعیین برای توابع بهینه‌سازی مختلف را مشاهده کرد. همان‌طور که در شکل ۱۰ مشاهده می‌شود.

همان‌طور که در جدول ۷ مشاهده می‌شود، در برخی از مدل‌ها، مقدار R^2_{Test} اندکی بیشتر از R^2_{Train} گزارش شده است. به‌عنوان مثال در الگوریتم Adam مقدار R^2_{Test} معادل ۰/۸۴۵۷ بوده که بیشتر از R^2_{Train} با مقدار ۰/۸۲۷۹ است. در نگاه دقیق‌تر به داده‌ها، این اختلاف کوچک نشان‌دهنده خطا در مدل‌سازی نیست، بلکه ناشی از ماهیت تصادفی داده‌ها در فرآیند تفکیک آموزش و آزمایش و وجود داده‌های یکنواخت‌تر یا کمتر نویزی در بخش آزمایش است. همچنین، استفاده از توقف زودهنگام



شکل ۱۰ مقایسه مقادیر MSE برای توابع بهینه‌سازی مختلف
Figure 10. Comparison of MSE values for different optimization functions.



شکل ۱۱ مقایسه مقادیر ضریب تعیین (R^2) برای توابع بهینه‌سازی مختلف.
Fig. 11 Comparison of coefficient of determination (R^2) values for different optimization function.

1. Overfitting
2. Underfitting

نیز برای شبکه‌های MLP1 و MLP3 از تابع اتلاف MAE مقادیر R^2 منفی به‌دست آمده است. در تابع بهینه‌سازی Adam بالاترین مقدار R^2 مربوط به شبکه MLP3 با تابع اتلاف MSE برابر با ۰/۸۴۵۷ است. مقدار Loss در این مدل ۰/۰۰۶۹ است و پایین‌ترین مقدار ممکن میان شبکه‌ها و توابع اتلاف دیگر در این تابع بهینه‌سازی است. در تابع AdamW نیز مدل‌های MLP3-MSE با Loss و R^2 برابر با ۰/۰۰۶۸ و ۰/۸۴۸۰ و MLP3-MAE با Loss و R^2 برابر با ۰/۰۶۰۲ و ۰/۸۳۹۰ بهترین نتایج را نسبت به سایر حالات داشته‌اند. با توجه به تفاوت کم میان مقادیر R^2 و اختلاف بسیار زیاد میان مقادیر Loss، شبکه MLP3 با تابع اتلاف MSE به‌عنوان مدل مناسب‌تر انتخاب شده است. در تابع Adamax نیز مدل‌های MLP3-MSE با Loss و R^2 برابر با ۰/۰۰۷۱ و ۰/۸۴۰۱ و MLP3-MAE با Loss و R^2 برابر با ۰/۰۶۰۲ و ۰/۸۳۷۷ بهترین نتایج را نسبت به سایر حالات داشته‌اند. با توجه به تفاوت موجود میان مقادیر R^2 و اختلاف بسیار زیاد میان مقادیر Loss، شبکه MLP3 با تابع اتلاف MSE به‌عنوان مدل مناسب‌تر انتخاب شده است. در تابع NAdam نیز مدل‌های MLP2-MSE با Loss و R^2 برابر با ۰/۰۱۳۴ و ۰/۷۰۰۳ و MLP3-MAE با Loss و R^2 برابر با ۰/۰۸۹۶ و ۰/۷۰۲۸ بهترین نتایج را نسبت به سایر حالات داشته‌اند. با توجه به اینکه مقدار Loss برای شبکه MLP2 با تابع اتلاف MSE بسیار پایین‌تر است، مدل MLP2-MSE به‌عنوان بهترین مدل انتخاب شده است. بهترین نتایج به‌دست آمده براساس توابع بهینه‌سازی مختلف در **جدول ۸** به‌صورت خلاصه نشان داده شده است. همان‌طور که در **جدول ۸** مشاهده می‌شود، بهترین نتایج مربوط به توابع بهینه‌سازی Adam و AdamW با شبکه MLP3 و تابع اتلاف MSE می‌باشد. براساس نتایج به دست آمده تابع اتلاف MSE در هر دو تابع بهینه‌سازی بهترین عملکرد را داشته است. ترتیب دقت در توابع مذکور به‌صورت $AdamW > Adam$ می‌باشد.

بالاترین مقدار MSE مربوط به توابع SGD، Adadelta، ASGD، و RMSprop است. کمترین مقدار MSE نیز مربوط به توابع Adam، AdamW، و Ad-Ad-amax می‌باشد. در **شکل ۱۱** نیز می‌توان مشاهده کرد که بالاترین مقدار ضریب تعیین مربوط به توابع بهینه‌سازی توابع Adam، AdamW و Adamax می‌باشد. کمترین مقدار ضریب تعیین نیز مربوط به توابع SGD، Adadelta، و ASGD می‌باشد. در تابع بهینه‌سازی Adadelta، بهترین نتیجه مربوط به تابع اتلاف MAE و شبکه MLP1 می‌باشد. اگرچه در این مدل نیز میزان R^2 بالا نیست اما به وضوح از مقادیر به‌دست آمده در حالات دیگر بهتر است. نتایج به‌دست آمده در تابع بهینه‌سازی RMSprop نشان می‌دهد که شبکه MLP1 و تابع اتلاف MAE نسبتاً بهتر از سایر مدل‌ها عمل کرده است. اگرچه در این حالت مقدار Loss رضایت‌بخش نیست و شبکه MLP1 با تابع اتلاف MSE کمترین میزان Loss را دارا می‌باشد، ولیکن از آنجا که هر دو پارامتر Loss و R^2 در انتخاب مدل مناسب نقش دارند، از این جهت شبکه MLP1 با تابع اتلاف MAE در اینجا به‌عنوان مدل مناسب‌تر معرفی شده است. در تابع بهینه‌سازی Rprop شبکه MLP3 و تابع اتلاف MAE عملکرد نسبتاً بهتری دارد. در این تابع بهینه‌سازی برای شبکه MLP3 و تابع اتلاف MAE، مقدار R^2 و Loss به‌ترتیب برابر با ۰/۸۱۹۴ و ۰/۰۶۲۷ می‌باشد. در تابع بهینه‌سازی SGD نیز شبکه MLP1 و تابع اتلاف MSE با Loss برابر ۰/۰۳۶۶ و R^2 برابر ۰/۱۷۸۵ نسبت به سایر مدل‌ها بهتر بوده است. مشاهده می‌شود که در این تابع مقادیر R^2 برای سایر شبکه‌ها و برای هر دو تابع اتلاف MSE و MAE منفی است و این نشان می‌دهد که در این شبکه‌ها، مدل به‌دست آمده بهتر از یک حد پایین نبوده و پیش‌بینی‌های آن بدتر از پیش‌بینی با میانگین داده‌های واقعی است. در تابع بهینه‌سازی ASGD، شبکه MLP2 با تابع اتلاف MSE نسبت به سایر مدل‌ها عملکرد بهتری داشته است. در این حالت

جدول ۸ بهترین نتایج به دست آمده براساس توابع بهینه‌سازی مختلف.

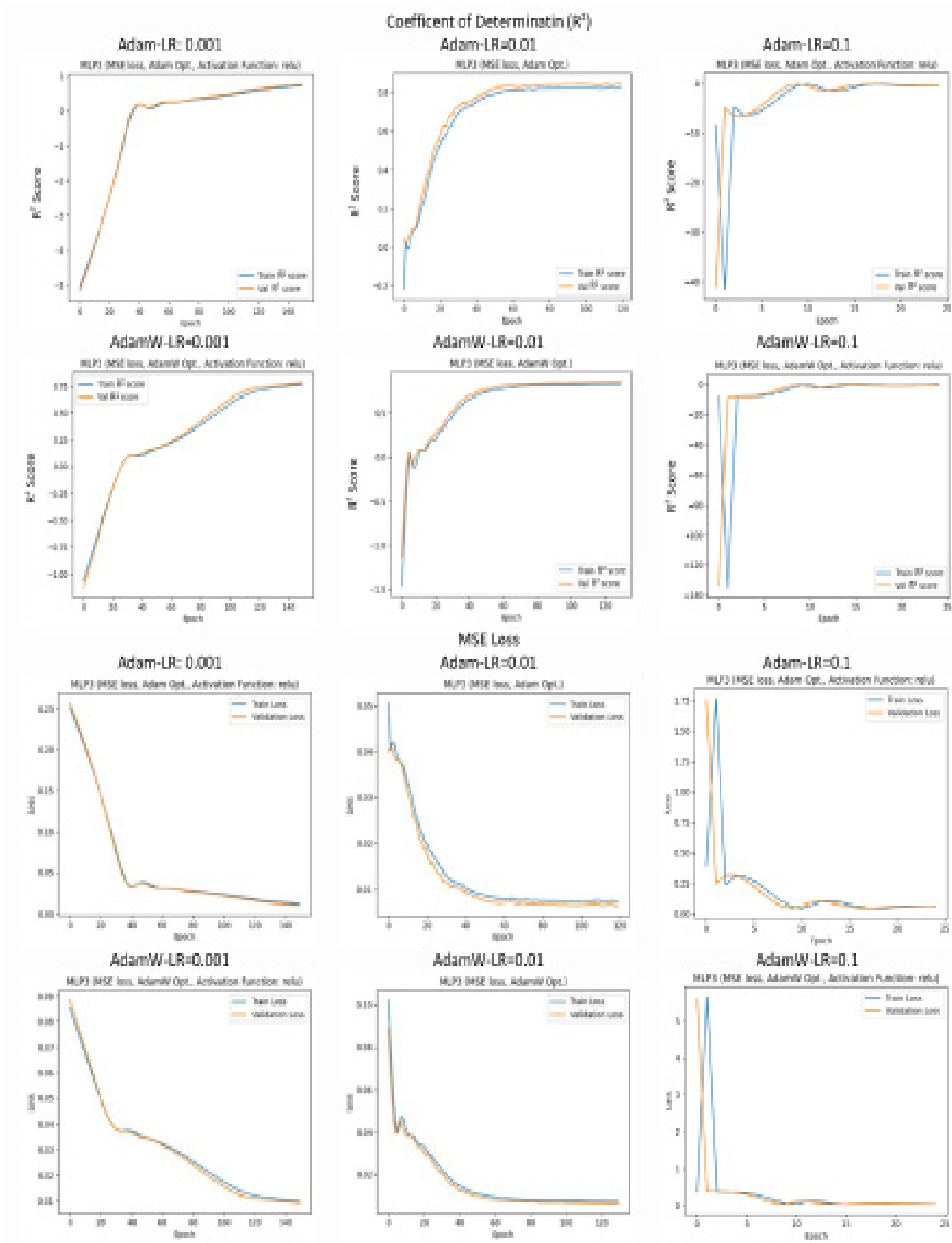
Table 8 Best results obtained based on different optimization functions

Network type	Loss function	Optimization function	Epoch	R ² -Train	R ² -Val	R ² -Test	Loss-Train	Loss-Val	Loss-Test
AdaGrad	MSE	MLP3	75	0.6734	0.6950	0.6897	0.0136	0.0127	0.0138
Adadelta	MAE	MLP1	600	0.2392	0.2511	0.2252	0.1365	0.1361	0.1413
RMSprop	MAE	MLP1	25	0.2790	0.3214	0.3429	0.1329	0.1312	0.1295
Rprop	MAE	MLP3	60	0.8023	0.8242	0.8194	0.0647	0.0622	0.0627
SGD	MSE	MLP1	30	0.1471	0.1785	0.1785	0.0354	0.0342	0.0366
ASGD	MSE	MLP2	60	0.2081	0.2406	0.2178	0.0329	0.0316	0.0349
Adam	MSE	MLP3	120	0.8279	0.8447	0.8457	0.0071	0.0065	0.0069
AdamW	MSE	MLP3	130	0.8269	0.8422	0.8480	0.0079	0.0066	0.0068
Adamax	MSE	MLP3	160	0.8192	0.8415	0.8401	0.0075	0.0066	0.0071
NAdam	MSE	MLP2	80	0.6584	0.7159	0.7003	0.0142	0.0118	0.0134
Constants	Learning rate: 0.01 Number of neurons in the first, second and third hidden layers: 30, 20 and 10 Activation function type: ReLu								

می‌گردد. نتایج مربوط به تغییر نرخ یادگیری بر داده‌های آزمایشی در جدول ۹ ارائه شده است. براساس جدول ۹، افزایش نرخ یادگیری از ۰/۰۰۱ تا ۰/۰۱ موجب بهبود مقادیر تابع اتلاف و ضریب تعیین در هر دو الگوریتم بهینه‌سازی شده است. با این حال، افزایش بیشتر نرخ یادگیری تا ۰/۱ باعث کاهش قابل توجه دقت مدل و ایجاد نوسانات شدید در فرآیند همگرایی گردیده است. از دیدگاه نظری، نرخ یادگیری پارامتری تعیین کننده در اندازه گام به روزرسانی وزن‌ها در فرآیند آموزش است. در الگوریتم‌های تطبیقی مانند Adam و AdamW، این نرخ به صورت پویا و متناسب با میانگین درجه دوم گرادیان‌ها تنظیم می‌شود. زمانی که نرخ یادگیری بیش از حد بزرگ انتخاب شود، مدل ممکن است از ناحیه بهینه عبور کرده و دچار نوسان شود؛ در حالی که نرخ‌های بسیار کوچک نیز منجر به همگرایی کند یا گیر افتادن در مینیمم‌های محلی می‌گردند.

بررسی تاثیر نرخ یادگیری در دو تابع بهینه‌سازی انتخابی در شبکه عصبی پرسپترون چند لایه

به منظور بررسی ثانویه و دست‌یابی احتمالی به نتایج بهتر، نرخ یادگیری در دو الگوریتم بهینه‌سازی منتخب (Adam و AdamW) تغییر داده شد. همان‌طور که در شکل ۱۲ مشاهده می‌شود، با افزایش نرخ یادگیری، اگرچه نتایج بهبود نسبی پیدا می‌کنند، اما در برخی بازه‌ها نوساناتی در مقادیر تابع اتلاف ایجاد می‌شود که البته از نوع واگرایی نیست. توجه به این نکته ضروری است که تفاوت‌های مشاهده شده در شکل‌های ۸، ۹ و ۱۲ در روند تغییرات Loss و R² بین داده‌های آموزش و اعتبارسنجی، ناشی از رفتار متفاوت توابع بهینه‌سازی (Adam و AdamW) و اثر نرخ‌های یادگیری مختلف است. به‌ویژه در مراحل اولیه آموزش، نوسانات جزئی در نمودار Loss به دلیل تنظیمات پویا در ضرایب مومنتوم و وزن‌های تطبیقی این توابع مشاهده می‌شود که بخشی طبیعی از فرآیند همگرایی محسوب



شکل ۱۲ نتایج مربوط به الگوریتم‌های بهینه‌سازی Adam و AdamW برحسب R^2 و Loss برای سه نرخ یادگیری مختلف. Fig. 12 Results of the AdamW and Adam optimization algorithms in terms of Loss and R^2 for three different learning rates.

جدول ۹ تاثیر افزایش نرخ یادگیری بر داده‌های آزمایشی در سه الگوریتم انتخابی.

Table 9 The effect of increasing the learning rate on the experimental data in the three selected algorithms.

Optimization function	Loss function	Network type	Learning rate	Epoch	R ² -Train	R ² -Val	R ² -Test	Loss-Train	Loss-Val	Loss-Test
Adam	MSE	MLP3	0.1	25	-0.0791	-0.2587	-0.3605	0.0448	0.0524	0.0606
			0.01	120	0.8279	0.8447	0.8457	0.0071	0.0065	0.0069
			0.001	150	0.7106	0.7533	0.7048	0.0210	0.0103	0.0132
AdamW	MSE	MLP3	0.1	25	-0.6240	-0.7125	-0.0429	0.0674	0.0714	0.0465
			0.01	130	0.8269	0.8422	0.8480	0.0072	0.0066	0.0068
			0.001	150	0.7683	0.7798	0.7827	0.0096	0.0092	0.0097
constants	Number of neurons in the first, second and third hidden layers: 30, 20 and 10 Activation function type: ReLu									

بهترین الگوریتم به‌منظور پیش‌بینی گرانروی نفت، الگوریتم بهینه‌سازی AdamW با سه لایه پنهان (با تعداد نورون ۶۰، ۴۰، ۲۰ به ترتیب در لایه پنهان اول، دوم و سوم) و نرخ یادگیری ۰/۰۱ با تابع اتلاف MSE می‌باشد.

محاسبه پارامترهای آماری کل برای الگوریتم بهینه‌سازی انتخابی در شبکه عصبی پرسپترون چند لایه

در این بخش محاسبه MAE، MSE و R² برای الگوریتم بهینه‌سازی AdamW و برای کل داده‌ها انجام شده است. مقدار MSE، MAE و R² برای الگوریتم بهینه‌سازی AdamW برای کل داده‌ها به ترتیب برابر ۰/۰۰۶۸، ۰/۰۵۹۳ و ۰/۸۳۹۶ می‌باشد. در شکل ۱۳ نتایج نهایی مربوط به مدل‌سازی الگوریتم بهینه‌سازی AdamW برحسب Loss و R² برای کل داده‌ها آورده شده است.

مقایسه نتایج شبکه عصبی پرسپترون چند لایه انتخابی با سه الگوریتم دیگر

برای ارزیابی عملکرد مدل‌های مختلف در پیش‌بینی گرانروی، چهار الگوریتم MLP-AdamW، Random، Gradient Boosting و Forest Support Vector Regression (SVR) مورد بررسی قرار گرفتند. در جدول ۱۱ شاخص‌های آماری شامل AAPRE (میانگین درصد خطای مطلق نسبی)، RMSE (ریشه میانگین مربعات خطا) و R² (ضریب تعیین) برای داده‌های آموزش و آزمایش محاسبه شدند.

در این پژوهش، مقدار ۰/۰۱ به‌عنوان حد بهینه تجربی در بازه آزموده‌شده انتخاب گردید. لازم به ذکر است که این مقدار یک معیار تجربی برای داده‌ها و ساختار شبکه مورد استفاده در این مطالعه است و لزوماً به‌عنوان یک مقدار تعمیم‌یافته برای سایر داده‌ها یا مدل‌ها محسوب نمی‌شود. به‌طور کلی، مقدار بهینه نرخ یادگیری تابعی از نوع تابع بهینه‌سازی، توزیع گرادیان‌ها و ویژگی‌های مسئله مدل‌سازی است.

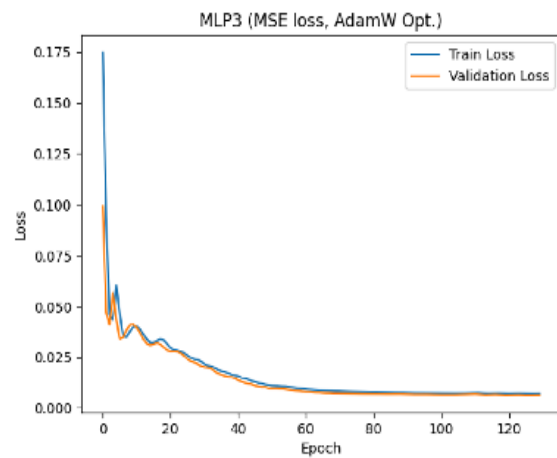
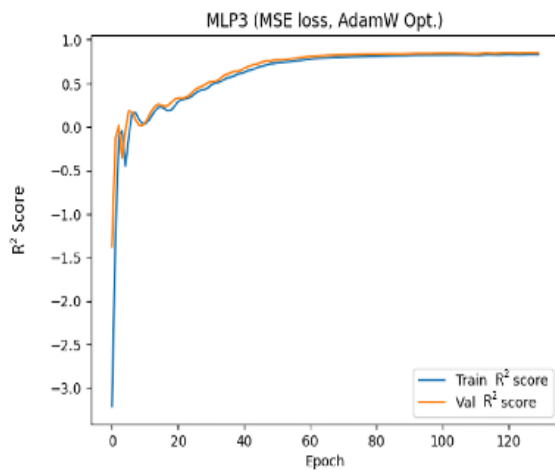
بررسی تاثیر تعداد نورون در سه تابع بهینه‌سازی انتخابی در شبکه عصبی پرسپترون چند لایه

در این بخش تاثیر تغییر تعداد نورون‌ها در دو الگوریتم بهینه‌سازی انتخابی بررسی شده است و در نهایت بهینه‌ترین حالت ممکن معرفی می‌شود. در جدول ۱۰ می‌توان تعداد نورون‌ها در هر لایه و میزان Loss و R² را مشاهده کرد. همان‌طور که در جدول ۱۰ مشاهده می‌شود، در الگوریتم بهینه‌سازی AdamW و درحالتی که تعداد نورون‌ها در لایه پنهان اول، دوم و سوم به ترتیب ۶۰، ۴۰ و ۲۰ باشد، مقدار R² و Loss برای داده‌های آزمایشی به ترتیب ۰/۸۵۴۱ و ۰/۰۰۶۵ می‌باشد که بیشترین مقدار میان حالات قبلی برای این الگوریتم و نیز الگوریتم Adam می‌باشد. این نتیجه پس از بررسی‌های متعدد روی تعداد نورون‌های دیگر و نیز پس از بارها سعی و تلاش برای به‌دست آوردن بهترین جواب ممکن بوده است. بنابراین براساس این جدول و نتایج به‌دست آمده در بررسی‌های پیشین،

جدول ۱۰ تاثیر تعداد نورون در هر لایه بر Loss و R^2 .

Table 10 Effect of the number of neurons in each layer on Loss and R^2 .

Optimization function	Network type	Epoch	Hidden layer			R^2 -Train	R^2 -Test	Loss-Train	Loss-Test
			1 st	2 nd	3 rd				
Adam	MLP3	120	15	10	5	0.7209	0.7318	0.0116	0.0120
			30	20	10	0.8279	0.8457	0.0071	0.0069
			60	40	20	0.8179	0.8311	0.0076	0.0075
AdamW	MLP3	130	15	10	5	0.7784	0.7885	0.0092	0.0094
			30	20	10	0.8269	0.8480	0.0072	0.0068
			60	40	20	0.8321	0.8541	0.0070	0.0065
constants	Learning rate: 0.01 Loss function type: MSE Activation function type: ReLu								



شکل ۱۳ نتایج مربوط به الگوریتم بهینه‌سازی AdamW برحسب Loss و R^2 برای کل داده‌ها.
 Fig. 13 Results of the AdamW optimization algorithms in terms of Loss and R^2 for the entire data.

جدول ۱۱ شاخص‌های آماری محاسبه شده برای چهار الگوریتم مورد بررسی در این مطالعه.

Table 11. Statistical indices calculated for the four algorithms examined in this study.

Model	Data Type	AAPRE	RMSE	R^2
Multi-Layer Perceptron (MLP)-AdamW	Train set	3.1853	1.8031	0.8321
	Test set	3.2448	1.8064	0.8541
Random Forest (RF)	Train set	1.3071	0.7390	0.9749
	Test set	3.6030	2.0442	0.8191
Gradient Boosting (GB)	Train set	2.4410	1.3242	0.9196
	Test set	3.5004	2.0012	0.8266
Support Vector Regression (SVR)	Train set	3.4395	2.0386	0.8095
	Test set	3.8850	2.3059	0.7699

*** مدل MLP-AdamW**

مدل شبکه عصبی چندلایه با بهینه‌ساز AdamW توانست روابط غیرخطی میان متغیرهای ورودی و گرانشی را با دقت مناسبی مدل‌سازی کند. مقدار R^2 داده‌های آزمایش (۰/۸۵۴۱) بیانگر توان تعمیم قابل قبول مدل است. با توجه به حجم زیاد داده‌ها، این مدل در یادگیری الگوهای کلی موفق بوده، اما به دلیل حساسیت بالای شبکه‌های عصبی به توزیع داده‌ها، در برخی نواحی داده‌ای رفتار نوسانی داشته است. این امر نشان می‌دهد که بیش‌برازش در MLP بیشتر ناشی از ناهمگنی داده‌ها بوده تا کمبود آنها.

*** مدل Random Forest**

مدل جنگل تصادفی بالاترین مقدار R^2 را در مرحله آموزش (۰/۹۷۴۹) ثبت کرده است که حاکی از توان بسیار بالا در یادگیری الگوهای موجود در مجموعه داده بزرگ است. با این حال، کاهش R^2 در داده‌های آزمایش (۰/۸۱۹۱) نشان‌دهنده بروز بیش‌برازش جزئی است. این رفتار در داده‌های با حجم زیاد طبیعی است، زیرا Random Forest تمایل دارد الگوهای محلی و جزئی را نیز به خوبی بیاموزد، که در نهایت باعث کاهش اندکی در توان تعمیم‌دهی می‌شود. با وجود این، مقادیر نسبتاً پایین RMSE و AAPRE گویای دقت و پایداری مناسب مدل در پیش‌بینی گرانشی است.

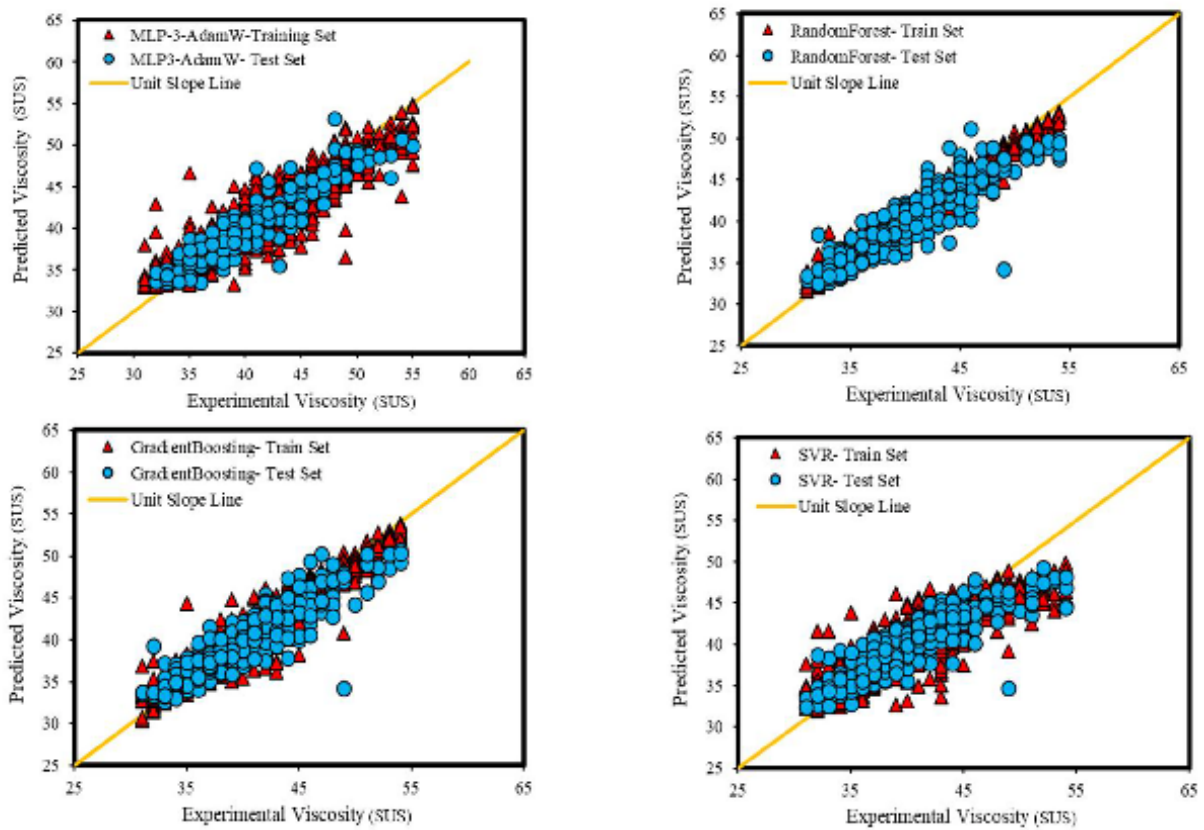
*** مدل Gradient Boosting**

مدل گرادیان بوستینگ با مقادیر R^2 برابر ۰/۹۱۹۶ (آموزش) و ۰/۸۲۶۶ (آزمایش) عملکردی پایدار و متعادل از خود نشان داده است. در حجم بالای داده‌ها، این مدل توانسته است بدون بروز بیش‌برازش محسوس، همگرایی مناسبی نسبت به داده‌ها داشته باشد. ویژگی یادگیری مرحله‌ای این الگوریتم موجب می‌شود خطاهای موضعی به تدریج اصلاح شوند و مدل به‌طور تطبیقی با ساختار داده سازگار گردد. از این‌رو، گرادیان بوستینگ از نظر تعمیم‌دهی نسبت به Random Forest عملکرد بهتری دارد.

*** مدل (SVR) Support Vector Regression**

مدل SVR پایین‌ترین مقادیر R^2 را در میان مدل‌های مورد بررسی دارد (مقادیر ۰/۸۰۹۵ برای آموزش، و ۰/۷۶۹۹ برای آزمایش). این امر نشان می‌دهد که در مجموعه داده بزرگ و غیرخطی، این مدل نتوانسته است الگوهای پیچیده میان متغیرها را به خوبی شناسایی کند. در شکل ۱۴، نمودار کراس پلات مقادیر پیش‌بینی‌شده در برابر مقادیر واقعی گرانشی برای چهار مدل MLP-AdamW، Gradient Boosting، Random Forest و Support Vector Regression (SVR) آورده شده است. نمودار کراس پلات^۱ یکی از ابزارهای متداول در ارزیابی عملکرد مدل‌های پیش‌بینی است که در آن مقادیر پیش‌بینی‌شده در برابر مقادیر واقعی رسم می‌شوند. در حالت ایده‌آل، نقاط داده باید در امتداد خط ۴۵ درجه (خط برابری یک‌به‌یک) قرار گیرند که نشان‌دهنده پیش‌بینی کاملاً دقیق مدل است. فاصله گرفتن نقاط از این خط بیانگر خطا در پیش‌بینی است. تراکم بالای نقاط در نزدیکی خط برابری نشان می‌دهد که مدل توانسته بخش عمده‌ای از داده‌ها را با دقت بالا پیش‌بینی کند. در مقابل، پراکندگی زیاد یا تراکم در نواحی دور از خط برابری بیانگر ضعف مدل در تطابق با مقادیر واقعی یا وجود نوسانات در داده‌ها است. بنابراین، توزیع و تراکم نقاط در نمودار کراس پلات معیاری بصری برای سنجش دقت و پایداری مدل محسوب می‌شود. شکل ۱۴ نشان می‌دهد که مدل‌های Gradient Boosting و m Forest بیشترین تطابق را با داده‌های واقعی دارند، در حالی که SVR از دقت کمتری برخوردار است. مدل MLP-AdamW نیز در داده‌های حجیم عملکرد قابل قبولی داشته است، اما در مقایسه با مدل‌های مبتنی بر درخت، نوسانات بیشتری در نواحی داده‌ای با پراکندگی بالا نشان می‌دهد.

1. Cross Plot



شکل ۱۴ مقادیر پیش‌بینی شده در برابر مقادیر واقعی گرانروی برای چهار مدل MLP، Random Forest، Gradient Boosting و Support Vector Regression.

Fig. 14 Predicted values versus actual viscosity values for the four models including MLP، Random Forest، Gradient Boosting and Support Vector Regression.

در Boosting و Support Vector Regression انجام شد. در این روش، داده‌های کامل به ده بخش تقسیم شده و در هر تکرار، ۹ بخش برای آموزش و یک بخش برای آزمون مورد استفاده قرار گرفت. سپس میانگین و انحراف معیار شاخص‌های آماری R^2 ، RMSE و AAPRE برای هر مدل محاسبه گردید (جدول ۱۲). نتایج جدول نشان می‌دهد که تمامی مدل‌ها از ثبات قابل‌قبولی در تکرارهای مختلف برخوردار هستند و انحراف معیار شاخص‌ها در محدوده‌ای بسیار پایین قرار دارد. این امر بیانگر آن است که مدل‌ها در برابر تغییرات در مجموعه داده‌ها رفتار باثباتی داشته‌اند.

به‌طور کلی می‌توان نتیجه گرفت که حجم زیاد داده‌ها باعث افزایش حساسیت مدل‌های پیچیده مانند Random Forest به الگوهای جزئی و در نتیجه بروز بیش‌برازش خفیف شده است. در مقابل، Gradient Boosting با بهره‌گیری از یادگیری مرحله‌ای و کنترل تدریجی خطاها، توانسته است تعادلی میان دقت و تعمیم‌دهی برقرار کند. بر این اساس، مدل Gradient Boosting به‌عنوان مناسب‌ترین الگوریتم برای پیش‌بینی گرانروی در داده‌های حجیم و غیرخطی معرفی می‌شود.

تحلیل نتایج اعتبارسنجی K-fold cross-validation

به‌منظور ارزیابی پایداری و تعمیم‌پذیری مدل‌ها، فرآیند اعتبارسنجی متقابل ده‌تایی^۱ برای چهار الگوریتم MLP-AdamW، Random Forest، Gradient

1. 10-Fold Cross-Validation

جدول ۱۲ نتایج اعتبارسنجی متقابل ده تایی برای چهار الگوریتم مورد بررسی در این مطالعه.
Table 12 The results of 10-Fold Cross-Validation for the four algorithms examined in this study.

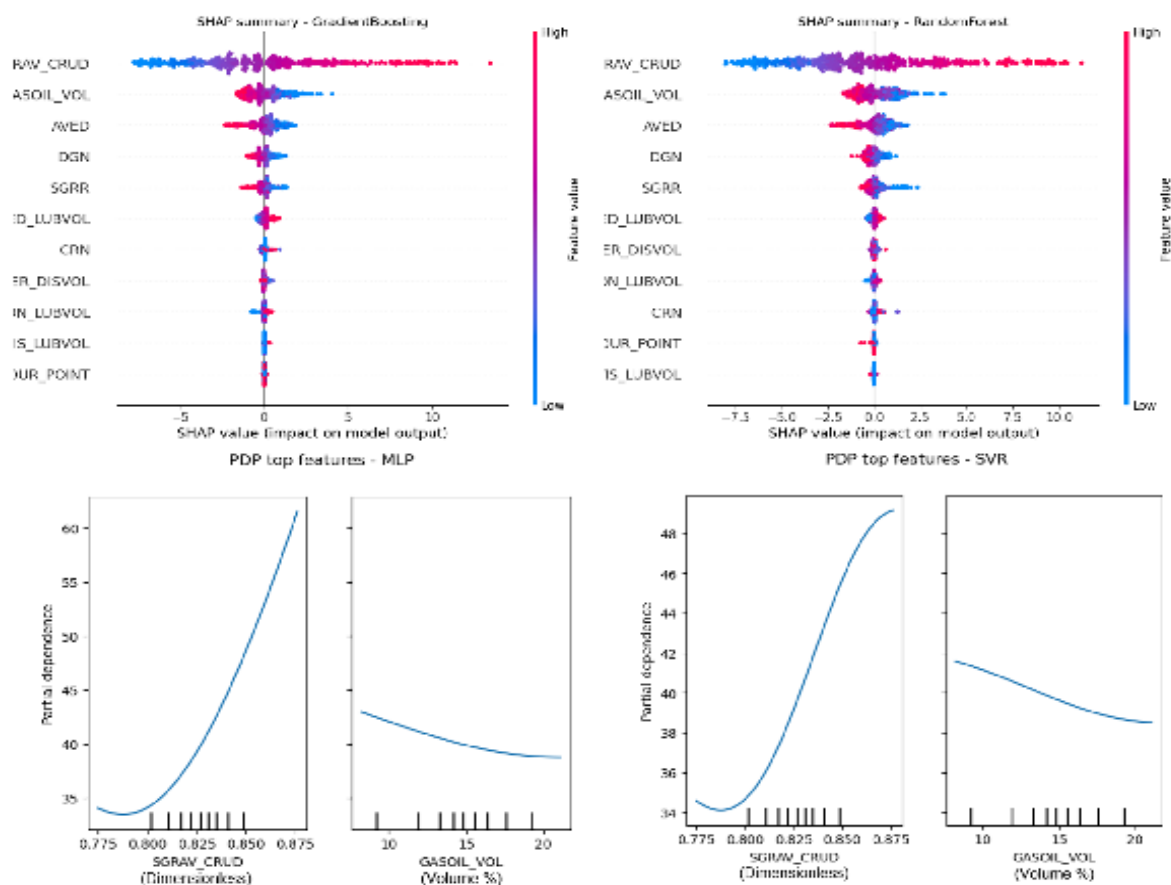
Models	AAPRE (mean \pm SD)	RMSE (mean \pm SD)	R ² (mean \pm SD)
Multi-Layer Perceptron (MLP)	3.26 \pm 0.09	1.82 \pm 0.05	0.851 \pm 0.012
Random Forest (RF)	3.55 \pm 0.11	2.03 \pm 0.07	0.523 \pm 0.018
Gradient Boosting (GB)	3.42 \pm 0.08	1.98 \pm 0.06	0.836 \pm 0.010
Support Vector Regression (SVR)	3.88 \pm 0.13	2.29 \pm 0.10	0.771 \pm 0.021

Multi-Layer Per- و (SVR) Support Vector Regression cepton (MLP)، به دلیل ماهیت غیرخطی و حساس بودن وزن‌ها به آموزش، از روش permutation importance استفاده شد که با جابه‌جایی هر ویژگی و اندازه‌گیری کاهش عملکرد، ارزش آن ویژگی را تخمین می‌زند. جهت مقایسه بین مدل‌ها، مقادیر هر مدل نرمال شد (هر مدل نسبت به بیشینه خود) و سپس مقایسه نسبی بین مدل‌ها برای برترین ویژگی‌ها انجام گرفت. همچنین برای MLP و SVR از انحراف معیار اهمیت (std) برای تحلیل پایداری استفاده شد. نتایج این بررسی در شکل ۱۵ نشان داده شده است. نتایج تحلیل SHAP برای چهار مدل مورد بررسی نشان داد که ویژگی SGRV_CRUD، که بیانگر وزن مخصوص نفت خام است، در تمامی مدل‌ها دارای بالاترین میزان اهمیت می‌باشد. در مدل‌های Random Forest و Gradient Boosting، این ویژگی بالاترین مقدار میانگین قدرمطلق SHAP را داشته است (به ترتیب ۳/۴۶ و ۳/۵۳)، که نشان‌دهنده نقش غالب آن در تعیین مقدار خروجی مدل است. این یافته در مدل‌های مبتنی بر شبکه عصبی (MLP) و برآزش بردار پشتیبان (SVR) نیز تأیید شد، به طوری که در هر دو مدل، SGRV_CRUD بیشترین مقدار اهمیت به دست آمده از روش permutation را داراست. این یافته بیانگر آن است که تغییر در وزن مخصوص نفت خام بیشترین تأثیر را بر پاسخ مدل دارد. از دیدگاه فیزیکی، وزن مخصوص شاخصی از ترکیب کلی نفت خام است؛ نفت‌های سنگین‌تر (با وزن مخصوص بالاتر) حاوی اجزای آروماتیک و آسفالتی بیشتر بوده و رفتار متفاوتی در فرآیندهای پالایشی از خود نشان می‌دهند.

در میان مدل‌ها، Gradient Boosting با میانگین R² برابر ۰/۸۳۶ \pm ۰/۰۱۰، کمترین مقدار (۱/۹۸ \pm ۰/۰۶) RMSE و پایین‌ترین AAPRE (۳/۴۲ \pm ۰/۰۸) بهترین عملکرد را از خود نشان داده است. انحراف معیار پایین در تمام شاخص‌ها نیز گواهی پایداری بالای این مدل در تقسیم‌بندی‌های مختلف داده‌هاست. این موضوع حاکی از آن است که Gradient Boosting ضمن دستیابی به دقت بالا، از بیش‌برآزش قابل توجهی رنج نمی‌برد و توانسته است ساختار کلی داده‌های حجیم را به درستی یاد بگیرد. مدل‌های Random Forest و MLP-AdamW نیز نتایج مناسبی ارائه داده‌اند، اما نوسان اندکی در مقادیر R² آن‌ها مشاهده می‌شود که می‌تواند ناشی از حساسیت این مدل‌ها به ترکیب‌های متفاوت داده در هر فولد باشد. مدل SVR در مقایسه با سایر مدل‌ها ضریب تعیین پایین‌تری داشته و به دلیل ماهیت نسبتاً خطی خود، توانایی کمتری در مدل‌سازی روابط غیرخطی میان پارامترهای ورودی و گراندروی نشان داده است. در مجموع، نتایج حاصل از اعتبارسنجی متقابل ده تایی نشان می‌دهد که مدل‌های توسعه یافته از پایداری آماری و قابلیت تعمیم مناسب برخوردارند و مدل Gradient Boosting به عنوان بهترین گزینه برای پیش‌بینی گراندروی در داده‌های حجیم و پیچیده پیشنهاد می‌شود.

تحلیل نتایج SHAP

اهمیت ویژگی‌ها برای Random Forest (RF) و Gradient Boosting (GB)، با استفاده از تحلیل SHAP محاسبه شد تا سهم متوسط هر ویژگی در پیش‌بینی مدل به صورت محلی و کلی سنجیده شود. برای مدل‌های



شکل ۱۵ نتایج تحلیل SHAP و permutation importance برای چهار مدل مورد مطالعه.
Fig. 15 Results of SHAP and permutation importance analysis for the four models studied.

مثبت بر متغیر هدف دارد، که از دیدگاه فرآیندی قابل تفسیر است؛ زیرا افزایش درصد برش نفت‌گاز اغلب منجر به افزایش ویژگی‌های فیزیکی مرتبط با غلظت و انرژی در محصول نهایی می‌گردد. از سوی دیگر، AVED که معرف چگالی نفت‌گاز است، نقش قابل توجهی در عملکرد مدل داشته و در میان سه ویژگی برتر قرار گرفته است. مقدار میانگین قدرمطلق SHAP برای این متغیر در محدوده $0/7-0/6$ در مدل‌های درختی قرار داشته و نتایج مشابهی در مدل‌های MLP و SVR مشاهده شده است. از دیدگاه فیزیکی، چگالی نفت‌گاز به ترکیب مولکولی آن (نسبت ترکیبات پارافینی، نفتنی و آروماتیک) وابسته است؛ افزایش چگالی معمولاً نشان‌دهنده حضور بیشتر ترکیبات سنگین‌تر و آروماتیک است

بنابراین، تأثیر بالای SGRAV_CRUD در مدل کاملاً با ماهیت فیزیکی سیستم هم‌خوانی دارد و نشان می‌دهد که خواص پایه‌ای خوراک نقش تعیین‌کننده‌ای در متغیر هدف دارند. در رتبه‌های بعدی، ویژگی‌های GASOIL_VOL و AVED نیز اهمیت بالایی نشان داده‌اند. ویژگی GASOIL_VOL که نشان‌دهنده درصد کسر حجمی نفت‌گاز در ترکیب برش‌های نفتی است، از نظر فیزیکی معرف سهم اجزای میان‌تقطیر در ساختار کلی خوراک یا محصول است. مقادیر بالاتر این متغیر به معنای حضور بیشتر برش‌های سنگین‌تر (نظیر نفت‌گاز) بوده که معمولاً منجر به افزایش چگالی، نقطه جوش و محتوای انرژی کل سیستم می‌شود. تحلیل SHAP نشان داد که افزایش GASOIL_VOL اثر

بیش‌برازش یا کم‌برازش منجر شود.
* **تفسیرپذیری محدود:** با وجود استفاده از تحلیل SHAP برای بهبود تفسیرپذیری، برخی مدل‌ها به‌ویژه شبکه‌های عصبی همچنان به‌صورت جعبه‌سیاه عمل می‌کنند و تبیین تصمیمات آن‌ها دشوار است.

* **هزینه محاسباتی بالا:** مدل‌های پیچیده مانند گرادیان بوستینگ و شبکه‌های عصبی نیازمند منابع محاسباتی قابل توجه و زمان آموزش طولانی هستند، به‌ویژه در داده‌های حجیم.

* **وابستگی به حوزه داده:** عملکرد مطلوب مدل‌ها ممکن است تنها در محدوده داده‌های مورد استفاده حفظ شود و در شرایط یا داده‌های خارج از محدوده آموزش، کاهش یابد.

در مجموع، اگرچه این مدل‌ها توانایی بالایی در مدل‌سازی روابط غیرخطی و پیش‌بینی دقیق دارند، اما استفاده از آن‌ها نیازمند داده‌های باکیفیت، تنظیم دقیق پارامترها و تحلیل‌های تفسیرپذیری است. در پژوهش‌های آینده، می‌توان با ترکیب روش‌های یادگیری عمیق با رویکردهای تبیینی‌تر و توسعه مدل‌های هیبریدی، بخشی از این محدودیت‌ها را کاهش داد.

نتیجه‌گیری

در این پژوهش، به‌منظور پیش‌بینی گرانروی نفت مرده، از شبکه عصبی پرسپترون چندلایه و سه الگوریتم مقایسه‌ای شامل جنگل تصادفی، گرادیان بوستینگ، و برآزش بردار پشتیبان استفاده شد. مهم‌ترین یافته‌ها به‌صورت زیر جمع‌بندی می‌شوند:

* نوآوری پژوهش

برای نخستین بار از پارامترهای مرتبط با برش‌های مختلف نفت خام شامل چگالی برش‌های سبک تا سنگین (نظیر بنزین سبک، نفتا، نفت سفید، نفت گاز و روان‌کننده‌ها) و نیز ویژگی‌های فیزیکی-شیمیایی (مانند وزن مخصوص، نقطه ریزش، محتوای

که می‌تواند خواص حرارتی، احتراقی و پایداری فرآیند را تغییر دهد. جهت اثر مثبت AVED در نمودارهای وابستگی SHAP بیانگر آن است که افزایش چگالی با افزایش مقدار متغیر هدف همراه است، که این امر با واقعیت فیزیکی رفتار سوخت‌های سنگین سازگار است. در مجموع، نتایج SHAP نشان داد که سه ویژگی کلیدی SGRAV_CRUD (وزن مخصوص نفت خام)، GASOIL_VOL (درصد کسر حجمی نفت‌گاز) و AVED (چگالی نفت‌گاز) بیشترین اثر را بر پیش‌بینی متغیر هدف دارند. این همگرایی میان مدل‌های مختلف و هم‌خوانی آن با مفاهیم فیزیکی سیستم، اعتبار نتایج مدل را تقویت می‌کند. به بیان دیگر، متغیرهای مرتبط با ترکیب و چگالی برش‌های نفتی بیشترین نقش را در رفتار کلی مدل ایفا می‌کنند و از این رو می‌توانند به‌عنوان شاخص‌های فیزیکی کلیدی در کنترل و بهینه‌سازی فرآیند مطرح شوند. به‌طور کلی، استفاده از SHAP موجب افزایش شفافیت مدل‌ها، قابلیت توضیح فیزیکی نتایج و درک بهتر نقش پارامترهای فرآیندی در رفتار سیستم گردید.

محدودیت‌های مدل‌های هوشمند توسعه‌یافته

اگرچه مدل‌های هوشمند توسعه‌یافته در این پژوهش عملکرد قابل توجهی در پیش‌بینی گرانروی نشان داده‌اند، اما برخی محدودیت‌ها نیز در کاربرد و تفسیر این مدل‌ها وجود دارد که توجه به آن‌ها ضروری است.

* **وابستگی به کیفیت داده‌ها:** دقت مدل‌های یادگیری ماشین به‌شدت به کیفیت، گستره و تنوع داده‌های آموزشی وابسته است. وجود نویز، داده‌های خارج از محدوده یا توزیع نامتوازن می‌تواند موجب کاهش دقت و تعمیم‌پذیری مدل‌ها شود.

* **حساسیت به پارامترهای تنظیمی:** مدل‌هایی مانند شبکه عصبی پرسپترون چندلایه و گرادیان بوستینگ به تنظیم دقیق ابرپارامترها نیاز دارند. انتخاب نامناسب این پارامترها ممکن است به

دقت بالا، از نوسانات کمتر و پایداری بالاتری نسبت به سایر مدل‌ها برخوردار بوده و بنابراین می‌تواند به‌عنوان گزینه پیشنهادی جهت به‌کارگیری در تصمیم‌گیری‌های عملیاتی و مهندسی استفاده گردد.

* کاربردهای مهندسی مدل

مدل توسعه‌یافته می‌تواند به‌عنوان ابزاری سریع، دقیق و اقتصادی برای تخمین گرانی در شرایطی که داده‌های آزمایشگاهی محدود یا در دسترس نیستند مورد استفاده قرار گیرد. این مدل در حوزه‌های زیر کاربرد دارد:

- طراحی و شبیه‌سازی خطوط لوله انتقال نفت خام؛
- مدل‌سازی جریان سیالات در مخازن نفتی؛
- تحلیل فرآیندهای ازدیاد برداشت؛
- پیش‌بینی رفتار رئولوژیکی سیالات در شرایط مختلف دما و فشار.
- مزیت کاربردی و صنعتی استفاده از مدل پیشنهادی باعث کاهش هزینه و زمان آزمایش‌های تجربی و افزایش دقت تصمیم‌گیری مهندسی در صنایع بالادستی نفت می‌شود.

این نتایج نشان می‌دهد که ترکیب مدل‌سازی داده‌محور مبتنی بر یادگیری ماشین و تحلیل عدم قطعیت می‌تواند راهکاری مطمئن برای پیش‌بینی خواص پیچیده نفت خام ارائه دهد و مسیر را برای توسعه ابزارهای هوشمند در صنایع پالایش و بهره‌برداری هموار نماید. [۱۱۵]

نیتروزن، و باقی‌مانده کربن از تقطیر خلا) به‌عنوان ورودی مدل‌ها استفاده شد که دقت پیش‌بینی را نسبت به مطالعات پیشین به‌طور قابل توجهی بهبود داد.

* عملکرد مدل‌ها و مقایسه نهایی

براساس نتایج آماری، الگوریتم گرادیان بوستینگ با دست‌یابی به تعادل میان دقت و تعمیم‌دهی، بهترین عملکرد را در میان مدل‌های مورد بررسی نشان داد. مدل‌های جنگل تصادفی و MLP-AdamW نیز عملکرد رقابتی و قابل قبولی ارائه دادند؛ با این حال SVR به دلیل ماهیت نسبتاً خطی خود دقت پایین‌تری در داده‌های غیرخطی و حجیم داشت.

* تحلیل عدم قطعیت پیش‌بینی

جهت افزایش قابلیت اعتماد مدل، بازه عدم قطعیت (PI) برای خروجی‌ها برآورد شد. نتایج نشان داد که مدل Gradient Boosting کمترین پهنای PI را داشته و بیشترین اطمینان در پیش‌بینی را ارائه می‌دهد؛ در حالی که SVR دارای بیشترین عدم قطعیت بود.

نتیجه‌گیری نهایی و مدل پیشنهادی

با توجه به دقت عددی، توان تعمیم‌دهی، پهنای باریک بازه عدم قطعیت ۹۵٪ و پایداری آماری در اعتبارسنجی K-fold، مدل گرادیان بوستینگ به‌عنوان مناسب‌ترین و قابل اعتمادترین الگوریتم برای پیش‌بینی گرانی نفت مرده در شرایط صنعتی و داده‌های حجیم معرفی می‌شود. این مدل علاوه بر

مراجع

- [1]. Clark, N. J. (1962). Fundamentals of Reservoir Fluids. Journal of Petroleum Technology, 14(01), 11-16. doi.org/10.2118/91-PA.
- [2]. Xu, D. H., & Khurana, A. K. (1996, October). A simple and efficient approach for improving the prediction of reservoir fluid viscosity. In SPE Asia Pacific Oil and Gas Conference and Exhibition (pp. SPE-37011). SPE. doi.org/10.2118/37011-MS.
- [3]. Raupov, I., Rogachev, M., & Sytnik, J. (2025). Overview of modern methods and technologies for the well production of high-and extra-high-viscous oil. Energies, 18(6), 1498.
- [4]. El-Hoshoudy, A. N., Farag, A. B., Ali, O. I. M., El-Batanoney, M. H., Desouky, S. E. M., & Ramzi, M. (2013). New correlations for prediction of viscosity and density of Egyptian oil reservoirs. Fuel, 112, 277-282. doi.org/10.1016/j.fuel.2013.05.045.
- [5]. Shao, M., Shao, Y., Wang, Y., Fu, L., Liao, K., Zhu, T., & Yang, Z. (2026). A review and perspective on vis

- cosity reduction development technologies for heavy oil. *Petroleum Science and Technology*, 44(12), 1835-1851. doi.org/10.1080/10916466.2025.2465835.
- [6]. McCain, W.D., J.P. Spivey, and C.P. Lenn, *Petroleum reservoir fluid property correlations*. 2011: PennWell Books.
- [7]. Zhang, Y., Yang, S., Luo, Y., Han, Y., Tian, L., Liu, Q., Gao, X., Shen, B., Yang, R. and Li, J., 2025. Visual characterization, exergy and mechanism analysis of in-situ nonequilibrium dynamic phase-behavior of CO₂-oil. *Energy*, 328, p.136560. doi.org/10.1016/j.energy.2025.136560.
- [8]. Makinde, F. A., Ako, C. T., Orodu, O. D., & Asuquo, I. U. (2012). Prediction of crude oil viscosity using feed-forward back-propagation neural network (FFBPNN). *Petroleum & Coal*, 54(2). Available online at www.vurup.sk/petroleum-coal.
- [9]. McCain Jr, W. D. (1991). Reservoir-fluid property correlations—state of the art. *SPE Reservoir Engineering*, 6(02), 266-272.
- [10]. Hemmati-Sarapardeh, A., et al., Reservoir oil viscosity determination using a rigorous approach. *Fuel*, 2014, 116: pp. 39-48.
- [11]. Hemmati-Sarapardeh, A., Khishvand, M., Naseri, A., & Mohammadi, A. H. (2013). Toward reservoir oil viscosity correlation. *Chemical Engineering Science*, 90, 53-68. doi.org/10.1016/j.ces.2012.12.009.
- [12]. Dolatyari, A., Ahmady, M., & Kazemi, A. (2024). A novel mathematical model for modeling viscosity and temperature relationship for dead oils. *Scientific Reports*, 14(1), 22836.
- [13]. Ghorbani, B., Ziabasharhagh, M., & Amidpour, M. (2014). A hybrid artificial neural network and genetic algorithm for predicting viscosity of Iranian crude oils. *Journal of Natural Gas Science and Engineering*, 18, 312-323.
- [14]. Al-Gathe, A. A., Abd-El Fattah, K. A., El-Banbi, A. H., & El-Metwally, K. A. (2015). A hybrid neuro-fuzzy approach for black oil viscosity prediction. *International Journal of Innovation and Applied Studies*, 13(4), 946.
- [15]. Soltanbekova, K., Ramazanova, G., & Zhabbasbayev, U. (2025). Potential of Gas-Enhanced Oil Recovery (EOR) Methods for High-Viscosity Oil: A Core Study from a Kazakhstani Reservoir. *Energies*, 18(15), 4182. doi.org/10.3390/en18154182.
- [16]. Al-Marhoun, M. A., Nizamuddin, S., Raheem, A. A., Ali, S. S., & Muhammadain, A. A. (2012). Prediction of crude oil viscosity curve using artificial intelligence techniques. *Journal of Petroleum Science and Engineering*, 86, 111-117. doi: 10.1016/j.petrol.2012.03.029.
- [17]. Glaso, O. (1980). Generalized pressure-volume-temperature correlations. *Journal of Petroleum Technology*, 32(05), 785-795. doi.org/10.2118/8016-PA.
- [18]. Beggs, H. D., & Robinson, J. R. (1975). Estimating the viscosity of crude oil systems. *Journal of Petroleum technology*, 27(09), 1140-1141. doi.org/10.2118/5434-PA.
- [19]. Al-Khafaji, A. H., Abdul-Majeed, G. H., & Hassoon, S. F. (1987). Viscosity correlation for dead, live and undersaturated crude oils. *J. Pet. Res*, 6(2), 1-16.
- [20]. Beal, C. (1946). The viscosity of air, water, natural gas, crude oil and its associated gases at oil field temperatures and pressures. *Transactions of the AIME*, 165(01), 94-115. doi.org/10.2118/946094-G.
- [21]. Kaye, S. E. (1985). Offshore California viscosity correlations. COFRC, TS85000940.
- [22]. Alomair, O., Elsharkawy, A., & Alkandari, H. (2011, December). Viscosity predictions of Kuwaiti heavy crudes at elevated temperatures. In *SPE International Heavy Oil Conference and Exhibition* (pp. SPE-150503). SPE. doi.org/10.2118/150503-MS.
- [23]. Naseri, A., Nikazar, M., & Dehghani, S. M. (2005). A correlation approach for prediction of crude oil viscosities. *Journal of petroleum science and engineering*, 47(3-4), 163-174. doi: 10.1016/j.petrol.2005.03.008.
- [24]. Hossain, M. S., Sarica, C., Zhang, H. Q., Rhyne, L., & Greenhill, K. L. (2005, November). Assessment and development of heavy-oil viscosity correlations. In *SPE International Thermal Operations and Heavy Oil Symposium* (pp. SPE-97907). SPE. doi.org/10.2118/97907-MS.
- [25]. Elsharkawy, A. M., & Alikhan, A. A. (1999). Models for predicting the viscosity of Middle East crude oils. *Fuel*, 78(8), 891-903. doi.org/10.1016/S0016-2361(99)00019-8.
- [26]. Bennison, T. (1998, December). Prediction of heavy oil viscosity. In *IBC Heavy oil field development conference* (Vol. 2, p. 4). Citeseer.
- [27]. De Ghetto, G., Paone, F., & Villa, M. (1994, October). Reliability analysis on PVT correlations. In *SPE Europec featured at EAGE Conference and Exhibition?* (pp. SPE-28904). SPE. doi.org/10.2118/28904-MS.
- [28]. Kartoatmodjo, T., & Schmidt, Z. (1994). Large data bank improves crude physical property correlations. *Oil and Gas Journal*; (United States), 92(27).
- [29]. Labedi, R. (1992). Improved correlations for predicting the viscosity of light crudes. *Journal of Petroleum Science and Engineering*, 8(3), 221-234. doi: 10.1016/0920-4105(92)90035-Y.
- [30]. Egbogah, E. O., & Ng, J. T. (1990). An improved temperature-viscosity correlation for crude oil systems.

- Journal of Petroleum Science and Engineering, 4(3), 197-200. doi: 10.1016/0920-4105(90)90009-R.
- [31]. Petrosky, G. E. (1990). PVT correlations for gulf of mexico crude oils (Doctoral dissertation, University of Southwestern Louisiana).
- [32]. Teja, A. S., & Rice, P. (1981). Generalized corresponding states method for the viscosities of liquid mixtures. *Industrial & engineering chemistry fundamentals*, 20(1), 77-81. doi.org/10.1021/i100001a015. doi.org/10.1021/i100001a015.
- [33]. Mehrotra, A. K. (1991). Generalized one-parameter viscosity equation for light and medium liquid hydrocarbons. *Industrial & engineering chemistry research*, 30(6), 1367-1372. doi.org/10.1021/ie00054a044.
- [34]. Svrcek, W. Y., & Mehrotra, A. K. (1988). One parameter correlation for bitumen viscosity. *Chem. Eng. Res. Des.:(United Kingdom)*, 66(4).
- [35]. Johnson, S. E., Svrcek, W. Y., & Mehrotra, A. K. (1987). Viscosity prediction of Athabasca bitumen using the extended principle of corresponding states. *Industrial & engineering chemistry research*, 26(11), 2290-2298. doi.org/10.1021/ie00071a020.
- [36]. Al-Marhoun, M. A. (2004). Evaluation of empirically derived PVT properties for Middle East crude oils. *Journal of Petroleum Science and Engineering*, 42(2-4), 209-221. doi: 10.1016/j.petrol.2003.12.012.
- [37]. Alboudwarej, H., Svrcek, W. Y., & Yarranton, H. W. (2002, June). PVT investigation of asphaltene precipitation and redissolution from bitumens. In *PETSOC Canadian International Petroleum Conference* (pp. PETSOC-2002). PETSOC. doi.org/10.2118/2002-015-EA.
- [38]. Hadavimoghaddam, F., Ostadhassan, M., Heidaryan, E., Sadri, M.A., Chapanova, I., Popov, E., Cheremisin, A. and Rafieepour, S., (2021). Prediction of dead oil viscosity: Machine learning vs. classical correlations. *Energies*, 14(4), p.930. doi.org/10.3390/en14040930.
- [39]. Oloso, M. A., Hassan, M. G., Bader-El-Den, M. B., & Buick, J. M. (2018). Ensemble SVM for characterisation of crude oil viscosity. *Journal of Petroleum Exploration and Production Technology*, 8(2), 531-546.
- [40]. Ghorbani, B., Hamedi, M., Shirmohammadi, R., Mehrpooya, M., & Hamedi, M. H. (2016). A novel multi-hybrid model for estimating optimal viscosity correlations of Iranian crude oil. *Journal of Petroleum Science and Engineering*, 142, 68-76. doi: 10.1016/j.petrol.2016.01.041.
- [41]. Hemmati-Sarapardeh, A., Aminshahidy, B., Pajouhandeh, A., Yousefi, S. H., & Hosseini-Kaldozakh, S. A. (2016). A soft computing approach for the determination of crude oil viscosity: Light and intermediate crude oil systems. *Journal of the Taiwan Institute of Chemical Engineers*, 59, 1-10. doi.org/10.1016/j.jtice.2015.07.017.
- [42]. Naseri, A., Yousefi, S. H., Sanaei, A., & Ghahreshaikhlu, A. A. (2012). A neural network model and an updated correlation for estimation of dead crude oil viscosity. *Brazilian journal of petroleum and gas*, 6(1).
- [43]. Oloso, M. A., Khoukhi, A., Abdulraheem, A., & Elshafei, M. (2009, October). Prediction of crude oil viscosity and gas/oil ratio curves using recent advances to neural networks. In *SPE/EAGE Reservoir Characterization & Simulation Conference* (pp. cp-170). European Association of Geoscientists & Engineers. doi.org/10.3997/2214-4609-pdb.170.spe125360.
- [44]. Torabi, F., Abedini, A., & Abedini, R. (2011). The development of an artificial neural network model for prediction of crude oil viscosities. *Petroleum Science and Technology*, 29(8), 804-816. doi.org/10.1080/10916460903485876.
- [45]. Gharbi, R. B., & Elsharkawy, A. M. (1997, March). Neural network model for estimating the PVT properties of Middle East crude oils. In *SPE Middle East Oil and Gas Show and Conference* (pp. SPE-37695). SPE. doi.org/10.2118/37695-MS.
- [46]. Mohd Fadzil, M. A., Razak, S. A., Razali, A. A., Salleh, M. N. M., & Zabiri, H. (2025). Calculation of the Viscosity of a Mixture of Heavy Hydrocarbon Oil with Temperature Relation Consideration. *ACS omega*, 10(29), 31274-31297. doi.org/10.1021/acsomega.4c06517.
- [47]. Mohammadi A., Kheirollahi S., Ayatollahi H., Pishvai S. Sh., Seyed Mahmoud Reza. (2025). Using machine learning models to predict acid and oil emulsion formation in static acidizing tests using a hybrid database. *Petroleum Research*, February-March. 34, 2025-6 (139): 73-91. 10.22078/pr.2024.5452.3426.
- [48]. Kheirollahi, H., Zaedi, M., Sobhani, S., Chehardoli, M., Simju, M. (2023). Screening of oil reservoir EOR methods using integration of artificial intelligence methods. *Petroleum Research*, 33, 2023-5 (132): 51-62. doi:10.22078/pr.2023.5151.3284.
- [49]. Li, D., Zhang, X., & Kang, Q. (2023). Machine learning estimation of crude oil viscosity as function of API, temperature, and oil composition: Model optimization and design space. *Plos one*, 18(2), e0282084. doi.org/10.1371/journal.pone.0282084.
- [50]. Langeroudy, K. P. A., Esfahani, P. K., & Movaghar, M. R. K. (2023). Enhanced intelligent approach for determination of crude oil viscosity at reservoir conditions. *Scientific Reports*, 13, 1666. doi: 10.1038/s41598-023-28770-2.
- [51]. Fotias, S. P., & Gaganis, V. (2023). Workflow for predicting undersaturated oil viscosity using machine

- learning. Results in Engineering, 20, 101502. doi.org/10.1016/j.rineng.2023.101502.
- [52]. Hadavimoghaddam, F., Ostadhassan, M., Heidaryan, E., Sadri, M.A., Chapanova, I., Popov, E., Cheremisin, A. and Rafieepour, S., 2021. Prediction of dead oil viscosity: Machine learning vs. classical correlations. Energies, 14(4), p.930. doi.org/10.3390/en14040930.
- [53]. Bahonar, E., Chahardowli, M., Ghalenoei, Y., & Simjoo, M. (2022). New correlations to predict oil viscosity using data mining techniques. Journal of Petroleum Science and Engineering, 208, 109736. doi: 10.1016/j.petrol.2021.109736.
- [54]. Dolatyari, A., Ahmady, M., & Kazemi, A. (2024). A novel mathematical model for modeling viscosity and temperature relationship for dead oils. Scientific Reports, 14(1), 22836.
- [55]. Gao, X., Dong, P., Cui, J., & Gao, Q. (2022). Prediction model for the viscosity of heavy oil diluted with light oil using machine learning techniques. Energies, 15(6), 2297. doi.org/10.3390/en15062297.
- [56]. Stratiev, D., Nenov, S., Shishkova, I., Sotirov, S., Sotirova, E., Dinkov, R., Yordanov, D., Pilev, D., Atanassov, K., Vasilev, S. and Stratiev, D.D. (2023). Prediction of viscosity of blends of heavy oils with diluents by empirical correlations and artificial neural network. Industrial & Engineering Chemistry Research, 62(49), pp.21449-21463. doi.org/10.1021/acs.iecr.3c02472.
- [57]. Almudhhi, S., Lababidi, H. M., & Garrouch, A. A. (2025). Application of machine learning for modeling heavy oil viscosity. Journal of Engineering Research. doi.org/10.1016/j.jer.2025.05.009.
- [58]. Lei, H., Hussein, M.A., Vora, T., Yadav, A., Rajiv, A., Shankhyan, A., Jaidka, S., Manu, M., Kadhim, I.M., Hamoodah, Z.J. and Faez, F., 2025. Accurate machine-learning based modeling of apparent viscosity of waxy oils mixed with polymeric wax inhibitors. Flow Measurement and Instrumentation, p.102968. doi.org/10.1016/j.flowmeasinst.2025.102968.
- [59]. Lillis, P., & DOE, N. , Crude Oil Analysis (COA) database. [Data set]. Open Energy Data Initiative (OEDI). National Renewable Energy Laboratory., 2019. https://data.openei.org/submissions/178.
- [60]. Bebis, G., & Georgiopoulos, M. (2002). Feed-forward neural networks. Ieee Potentials, 13(4), 27-31.
- [61]. Echamsi, S., Belouafi, E., Kassimi, A. E. B., Alali, A., Bourroumi, A., & Guennoun, A. (2025, May). Optimizing Training Hyperparameters for Multilayer Perceptrons in Deep Learning. In 2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET) (pp. 1-8). IEEE.
- [62]. Irmak, B., Karakoyun, M., & Gülcü, Ş. (2023). An improved butterfly optimization algorithm for training the feed-forward artificial neural networks. Soft Computing, 27(7), 3887-3905.
- [63]. Bansal, P., Gupta, S., Kumar, S., Gorelick, S., & Sharma, S. (2019). MLP-LOA: a metaheuristic approach to design an optimal multilayer perceptron: P. Bansal et al. Soft Computing, 23(23), 12331-12345.
- [64]. Mirjalili, S. Z., Saremi, S., & Mirjalili, S. M. (2015). Designing evolutionary feedforward neural networks using social spider optimization algorithm. Neural Computing and Applications, 26(8), 1919-1928.
- [65]. Lv, L., Zhang, C., Fan, J., & Shen, W. (2025). Deep reinforcement learning for job shop scheduling problems: A comprehensive literature review. Knowledge-Based Systems, 321, 113633. doi: 10.1016/j.knosys.2025.113633.
- [66]. Pamadi, V. N., & Rastogi, D. (2025). Optimizing Neural Network Performance Through Adaptive Learning Algorithms. International Journal of Research in all Subjects in Multi Languages [Subject: Computer Science], 13, 2321-2853.
- [67]. Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. Journal of machine learning research, 12(7).
- [68]. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980.
- [69]. Almudhhi, S., Lababidi, H. M., & Garrouch, A. A. (2025). Application of machine learning for modeling heavy oil viscosity. Journal of Engineering Research. doi.org/10.1016/j.jer.2025.05.009.
- [70]. Liu, Y., Pan, R., & Zhang, T. (2025, May). Adagrad under anisotropic smoothness. In International Conference on Learning Representations (Vol. 2025, pp. 19574-19608).
- [71]. Jiang, R., Maladkar, D., & Mokhtari, A. (2024). Provable Complexity Improvement of AdaGrad over SGD: Upper and Lower Bounds in Stochastic Non-Convex Optimization. arXiv preprint arXiv:2406.04592.
- [72]. Liu, Q., Yang, S., Yu, D., & Gai, T. (2024). BP neural network construction cost prediction model based on different optimization algorithms. doi.org/10.21203/rs.3.rs-3602749/v1.
- [73]. Poudel, A., Burns, D., Adhikari, R., Duron, D., Hendrix, J., Gentimis, T., Tubana, B. and Setiyono, T., (2025). Cover Crop Biomass Predictions with Unmanned Aerial Vehicle Remote Sensing and TensorFlow Machine Learning. Drones, 9(2), 131. doi.org/10.3390/drones9020131.
- [74]. Uppal, M., Gupta, D., Juneja, S., Gadekallu, T. R., El Bayoumy, I., Hussain, J., & Lee, S. W. (2023). Enhancing accuracy in brain stroke detection: Multi-layer perceptron with Adadelta, RMSProp and AdaMax optimizers. Frontiers in Bioengineering and Biotechnology, 11, 1257591. doi.org/10.3389/fbioe.2023.1257591.

- [75]. Sakketou, F., & Ampazis, N. (2024, September). A Computationally Efficient Second Order Deep Learning Training Algorithm with Adaptive Momentum. In Proceedings of the 13th Hellenic Conference on Artificial Intelligence (pp. 1-6). doi.org/10.1145/3688671.3688754.
- [76]. Upadhyay, R., Pipersenia, A. K., & Nidhya, M. S. (2024, January). Analyzing Multilayer Perception Architectures for Reinforcement Learning. In 2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC) (pp. 1-7). IEEE. doi: 10.1109/ICOCWC60930.2024.10470491.
- [77]. Elshamy, R., Abu-Elnasr, O., Elhoseny, M., & Elmougy, S. (2023). Improving the efficiency of RMSProp optimizer by utilizing Nesterov in deep learning. *Scientific Reports*, 13(1), 8814.
- [78]. Kurbiel, T., & Khaleghian, S. (2017). Training of deep neural networks based on distance measures using RMSProp. arXiv preprint arXiv:1708.01911. doi.org/10.48550/arXiv.1708.01911.
- [79]. Riedmiller, M., & Braun, H. (1993, March). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In IEEE international conference on neural networks (pp. 586-591). IEEE..
- [80]. Norouzian, M., Orosnjak, M., Kedziora, S. A., Useldinger, R., & Aslan, F. Acoustic Emission Signatures as Predictors of Laser Weld Integrity: A Machine Learning Framework for Real-Time Laser Welding Quality Assessment. Available at SSRN 5556630.
- [81]. Saini, R. (2025). A review on artificial neural networks for structural analysis. *Journal of Vibration Engineering & Technologies*, 13(2), 142.
- [82]. Babu, H., Ganesan, L. K., Kanniappan, J., & Marimuthu, H. R. (2025, March). Prediction of oral cancer with CT scanned images and Dysplasia slides using machine learning algorithm. In AIP Conference Proceedings (Vol. 3175, No. 1, p. 020036). AIP Publishing LLC.
- [83]. Bottou, L., & Bousquet, O. (2007). The tradeoffs of large scale learning. *Advances in neural information processing systems*, 20.
- [84]. Patnala, S. K., BG, S., Sethi, A., Gupta, S., & Asthana, S. (2024, January). SGD-MLP: Structure Generation and Distillation using a graph free MLP. In Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD) (pp. 64-72). doi.org/10.1145/3632410.3632431.
- [85]. Bottou, L. (1998). Online algorithms and stochastic approximations. *Online learning in neural networks*.
- [86]. Talebi, H., Samadianfard, S., & Valizadeh Kamran, K. (2025). Estimation of daily reference evapotranspiration implementing satellite image data and strategy of ensemble optimization algorithm of stochastic gradient descent with multilayer perceptron. *Environment, Development and Sustainability*, 27(2), 3707-3729.
- [87]. Shi, H., Yang, N., Tang, H., & Yang, X. (2022). aSGD: Stochastic gradient descent with adaptive batch size for every parameter. *Mathematics*, 10(6), 863. doi.org/10.3390/math10060863.
- [88]. Kimanzi, R., Kimanga, P., Cherori, D., & Gikunda, P. K. (2024). Deep Learning algorithms used in intrusion detection systems--a review. arXiv preprint arXiv:2402.17020. doi.org/10.48550/arXiv.2402.17020
- Focus to learn more.
- [89]. Bank, D., Koenigstein, N., & Giryas, R. (2023). Autoencoders. *Machine learning for data science handbook: data mining and knowledge discovery handbook*, 353-374.
- [90]. Brownlee, J. (2017). Deep learning for natural language processing: develop deep learning models for your natural language problems. *Machine Learning Mastery*.
- [91]. Hospodarskyy, O., Martsenyuk, V., Kukharska, N., Hospodarskyy, A., & Sverstiuk, S. (2024). Understanding the Adam Optimization Algorithm in Machine Learning. CITI, 2024, 2nd.
- [92]. Tanbhir, G., Shahriyar, M. F., Shahed, K., Chy, A. M. R., & Al Adnan, M. (2024, December). Hybrid machine learning model for detecting bangla smishing text using bert and character-level cnn. In 2024 13th International Conference on Electrical and Computer Engineering (ICECE) (pp. 57-62). IEEE.
- [93]. Madni, S. H. H., A/L Pathmanatan, L., Faheem, M., Shahzad, H. M. F., & Shah, S. (2025). Exploring optimizer efficiency for facial expression recognition with convolutional neural networks. *The Journal of Engineering*, 2025(1), e70060. doi.org/10.1049/tje2.70060Digital Object Identifier (DOI).
- [94]. Rikhari, S., & Lakshmi, K. M. (2025). Tumor extraction system using U-Net and Adamax optimizer. In *Recent Trends in VLSI and Semiconductor Packaging* (pp. 1-6). CRC Press.
- [95]. Sriwahyuni, L., Nurdianti, S., Nugrahani, E. H., & Najib, M. K. (2025). Performance Of Machine Learning For Imputing Missing Daily Rainfall Data In East Java Under Multiple Satellite Data Models. *Geographia Technica*, 20(1).
- [96]. Jirakitpuwapat, W., Dubey, P., Prasertsuk, N., Phanthong, C., Tritham, C., Tritham, C., Chandharakool, S., Tharathep, C. and Soontornpipit, P., (2024). Convergence property of Nesterov-accelerated adaptive moment estimation with safety helmet detection and classification in smart industry application. *Mathematical Methods in the Applied Sciences*, 47(16), pp.12650-12667. doi.org/10.1002/mma.10174.
- [97]. Saber, A., Emara, T., Elbedwehy, S., & Hassan, E. (2025). A novel approach for breast cancer detection

- using a Nesterov accelerated adam optimizer with an attention mechanism. *Scientific Reports*, 15(1), 27065.
- [98]. Khosravi, M., & Zeraati, M. (2019). Predictive modeling of the length of prepared CNT by CVD through ANN-MPSO and GEP. *Journal of Particle Science and Technology*, 5(4), 145-159. doi: 10.22104/jpst.2020.3835.1156.
- [99]. Iranzad, R., & Liu, X. (2025). A review of random forest-based feature selection methods for data science education and applications. *International Journal of Data Science and Analytics*, 20(2), 197-211.
- [100]. Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024, 69-79. doi.org/10.58496/BJML/2024/007.
- [101]. Choudhury, A., Mondal, A., & Sarkar, S. (2024). Searches for the BSM scenarios at the LHC using decision tree-based machine learning algorithms: a comparative study and review of random forest, AdaBoost, XGBoost and LightGBM frameworks. *The European Physical Journal Special Topics*, 233(15), 2425-2463.
- [102]. Haddouchi, M., & Berrado, A. (2024). A survey and taxonomy of methods interpreting random forest models. *arXiv preprint arXiv:2407.12759*. doi.org/10.48550/arXiv.2407.12759.
- [103]. Malashin, I., Tynchenko, V., Gantimurov, A., Nelyub, V., & Borodulin, A. (2025). Boosting-based machine learning applications in polymer science: A review. *Polymers*, 17(4), 499. doi.org/10.3390/polym17040499.
- [104]. Khatir, A., Capozucca, R., Khatir, S., Magagnini, E., Benaissa, B., & Cuong-Le, T. (2024). An efficient improved gradient boosting for strain prediction in near-surface mounted fiber-reinforced polymer strengthened reinforced concrete beam. *Frontiers of structural and civil engineering*, 18(8), 1148-1168.
- [105]. Tepakhan, W., Srisintorn, W., Penglong, T., & Saelue, P. (2025). Machine learning approach for differentiating iron deficiency anemia and thalassemia using random forest and gradient boosting algorithms. *Scientific Reports*, 15(1), 16917.
- [106]. Malashin, I., Tynchenko, V., Gantimurov, A., Nelyub, V., & Borodulin, A. (2025). Support vector machines in polymer science: A review. *Polymers*, 17(4), 491. doi.org/10.3390/polym17040491.
- [107]. Nasim, F., Khatoon, S., Nasiruddin, I., Shahid, M., Urooj, S., & Bilal, B. (2025). Support-Vector-Regression-Based Intelligent Control Strategy for DFIG Wind Turbine Systems. *Machines*, 13(8), 687. doi.org/10.3390/machines13080687.
- [108]. Tanveer, M., Rajani, T., Rastogi, R., Shao, Y. H., & Ganaie, M. A. (2024). Comprehensive review on twin support vector machines. *Annals of Operations Research*, 339(3), 1223-1268.
- [109]. Haque, M.A., Nirob, J.H., Nahin, K.H., Jizat, N.M., Zakariya, M.A., Ananta, R.A., Abdulkawi, W.M., Aljaloud, K. and Al-Bawri, S.S., (2025). Machine learning-based technique for gain prediction of mm-wave miniaturized 5G MIMO slotted antenna array with high isolation characteristics. *scientific reports*, 15(1), 276.
- [110]. Habibi, M., Bernard, S., Wang, J., & Fuge, M. (2025). Mean squared error may lead you astray when optimizing your inverse design methods. *Journal of Mechanical Design*, 147(2), 021701. doi.org/10.1115/1.4066102.
- [111]. Wang, H., Ma, J., Wang, C., Sun, H., Du, S., Wen, H., & Li, Y. (2025). Machine learning-based performance prediction for energy storage medium-deep borehole ground source heat pump systems. *Journal of Building Engineering*, 99, 111471. doi: 10.1016/j.jobbe.2024.111471.
- [112]. Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 2022, 1-10.
- [113]. Meyer, G. P. (2021). An alternative probabilistic interpretation of the huber loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5261-5269).
- [114]. Routray, A., Mistry, K. D., Arya, S. R., & Chittibabu, B. (2025). Applied machine learning in wind speed prediction and loss minimization in unbalanced radial distribution system. *Energy sources, Part A: recovery, utilization, and environmental effects*, 47(1), 3790-3810. doi.org/10.1080/15567036.2020.1859010.
- [115]. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ computer science*, 7, e623.