

# تعیین کل کربن آلی (TOC) و رخساره‌های آلی از داده‌های چاه‌پیمایی به کمک معادلات هوشمندانه مشتق شده و آنالیز خوشه‌ای: مطالعه موردی از میدان نفتی اهواز، جنوب غرب ایران

هدی عبدی‌زاده<sup>۱\*</sup>، علی کدخدائی<sup>۲</sup>، مسعود شایسته<sup>۳</sup>، و محمد حسین حیدری‌فرد<sup>۳</sup>

۱- گروه زمین‌شناسی، دانشکده علوم، دانشگاه سیستان و بلوچستان، زاهدان، ایران

۲- گروه زمین‌شناسی، دانشکده علوم طبیعی، دانشگاه تبریز، ایران

۳- گروه زمین‌شناسی، شرکت ملی مناطق نفت‌خیز جنوب، اهواز، ایران

تاریخ دریافت: ۹۲/۸/۲۶ تاریخ پذیرش: ۹۳/۱۱/۶

## چکیده

کل کربن آلی TOC یکی از پارامترهای مهم ژئوشیمیایی برای ارزیابی پتانسیل هیدروکربن‌زائی سنگ منشاء است. اندازه‌گیری این پارامتر نیاز به انجام آزمایشات ژئوشیمی بر روی کنده‌های حضاری دارد که پرهزینه و وقت‌گیر است. به طور کلی، سنگ‌های غنی از مواد آلی توسط نگاره‌های پتروفیزیکی با تخلخل بالا، زمان عبور صوت بالا، چگالی پایین، پرتو گاما بالاتر و مقاومت ویژه بیشتر از سایر سنگ‌ها مشخص می‌شوند. در این مطالعه، مدل‌های الگوریتم ژنتیک خطی و غیر خطی جهت تخمین کل کربن آلی از داده‌های پتروفیزیکی برای سنگ‌های منشأ کژدمی، گورپی و پابده در میدان نفتی اهواز استفاده شد. از مدل الگوریتم ژنتیک خطی نتایج معتبرتر و مقبول‌تری نسبت به مدل غیر خطی حاصل شد. راه حل‌های الگوریتم ژنتیک با آنالیز رگرسیون به واسطه ضرائب مناسب با معادلات TOC مقایسه شد. اجرای آنالیز براساس میانگین مربع خطا و ضریب همبستگی، کارایی بالاتر معادلات هوشمندانه مشتق شده را در مقایسه با آنالیز رگرسیون آماری نشان می‌دهد. در مرحله بعدی مطالعه، روش آنالیز خوشه‌ای جهت طبقه‌بندی نگار TOC تخمین زده شده و شناسایی زون‌های ژئوشیمیایی مورد استفاده قرار گرفت. نتایج مدل الگوریتم ژنتیک به شکل مقبول، سنگ‌های منشأ را با استفاده از مدل آنالیز خوشه‌ای به رخساره‌های غنی و فقیر از مواد آلی تقسیم کردند.

کلمات کلیدی: کل کربن آلی، الگوریتم ژنتیک، پیرولیز راک-ایول، داده‌های چاه‌پیمایی، رخساره‌های آلی، آنالیز خوشه‌ای

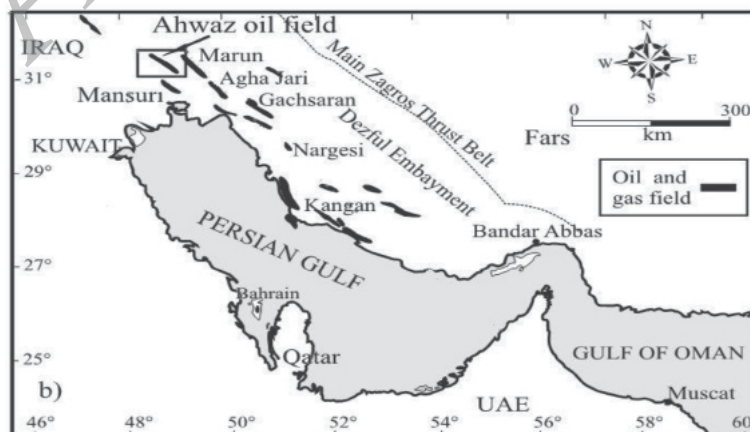
## مقدمه

جستجو الهام گرفته از تکامل بیولوژیکی است [۷]. در بیست سال اخیر الگوریتم ژنتیک به عنوان یک تکنیک بهینه برای توابع پیچیده مورد توجه قرار گرفته است [۸ و ۹].

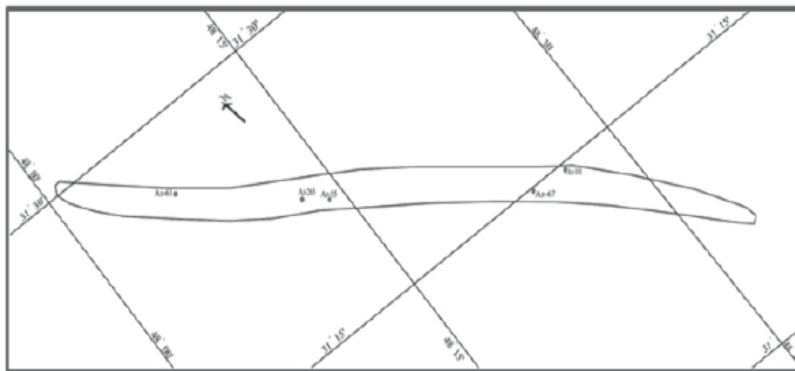
میدان نفتی اهواز (شکل ۱) یکی از مهم‌ترین میادین نفتی بسیار بزرگ ایران است که در سال ۱۹۶۵ کشف شد و اکنون بیش از ۴۵۰ چاه تولیدی دارد. این میدان نفتی دارای ساختاری طاق‌دیزی با طول ۷۲ km و عرض ۶ km است. همچنین در جنوب غرب ایران و بخش مرکزی شمال منطقه دزفول واقع شده است. در این مطالعه داده‌های ژئوشیمیایی و پتروفیزیکی از سه سنگ منشأ مهم شامل سازندهای کژدمی، گورپی و پابده تهیه شده است. در شکل ۲ موقعیت چاه‌های مورد مطالعه در میدان نفتی اهواز نشان داده شده است. مطالعات مشابهی در میدان نفتی اهواز توسط [۱۰] بر روی توزیع سنگ منشأ و بلوغ با استفاده از داده‌های ژئوشیمیایی انجام شده است [۱۱]. سنگ‌های منشأ نفت را در سازند کژدمی میدان نفتی اهواز با استفاده از پارامترهای ژئوشیمیایی ارزیابی کردند [۳]. پتروفیزیک و ژئوشیمی بخش ماسه سنگ سازند آسماری اهواز را با اشاره به منشأ و مجموعه تکنیکی مطالعه کردند [۱۲]. سنگ‌های منشأ هیدروکربن سازندهای گورپی و پابده همان میدان را ارزیابی کردند.

حضور محتوای کل کربن آلی<sup>۱</sup> در سنگ‌های منشأ یکی از پارامترهای ژئوشیمیایی مهمی است که می‌تواند برای ارزیابی پتانسیل تولید باقی‌مانده و خصوصیات ژئوشیمی واحدهای هیدروکربن‌دار استفاده شود. تا کنون محققان زیادی سعی کرده‌اند ارتباط کیفی و کمی بین داده‌های پتروفیزیکی و مقدار کل کربن آلی سنگ‌های مختلف، ایجاد کنند. از جمله [۱] از نگار گاما برای شناسایی سنگ‌های غنی از ماده آلی استفاده نموده و بنیان نهادند [۲]. تکنیک  $\Delta\text{LOGR}$  را اختراع نمودند. در این روش با همپوشانی نگارهای تخلخل (نوترون، چگالی و صوتی) و نگار مقاومت کل کربن آلی شناسایی و محاسبه می‌شود. بتازگی، [۳-۵] با تکنیک‌های هوشمندی مثل شبکه عصبی و سیستم‌های فازی برای تعیین کل کربن آلی از داده‌های چاه‌پیمایی استفاده کردند.

امروزه از تکنیک‌های محاسبات تکاملی در صنعت نفت، به ویژه در زمینه اکتشاف و تولید استفاده می‌شود. محاسبات تکاملی شاخه‌ای از هوش مصنوعی می‌باشند. الگوریتم ژنتیک اولین بار توسط جان هالند در سال ۱۹۶۰ مطرح شد و به علاوه، دانشجویان و همکاران او به کمک هم در سالهای ۱۹۶۰ و ۱۹۷۰ آن را توسعه دادند [۶]. الگوریتم ژنتیک سرآمد روش‌های تکاملی است و الگوریتم‌های



شکل ۱- نقشه موقعیت میدان نفتی اهواز در فروافتادگی دزفول، جنوب غرب ایران [۱۳]



شکل ۲- موقعیت چاه‌های مورد مطالعه در میدان نفتی اهواز

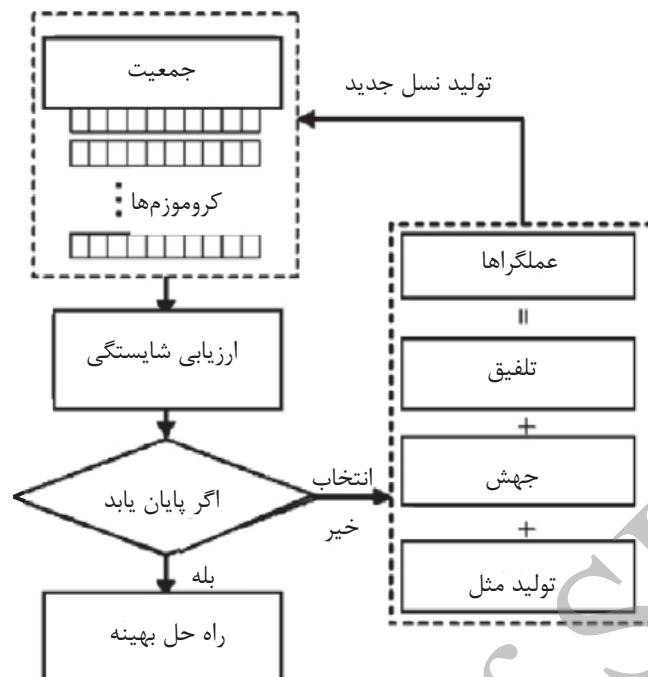
شکل‌گیری یک جمعیت جدید (نسل بعدی) افراد بر مبنای میزان شایستگی‌شان انتخاب می‌شوند. روش‌های مختلفی برای انجام این انتخاب وجود دارند که ساده‌ترین آنها که توسط هالند در دهه هفتاد میلادی (۱۹۶۰) مطرح شده انتخاب متناسب با شایستگی می‌باشد. روال تکامل عبارتست از انتخاب، ترکیب و جهش. یک الگوریتم تکاملی عموماً جمعیت آغازین را به صورت تصادفی تعیین می‌کند. همچنین میزان شایستگی هر یک از افراد را بر مبنای ارزششان در محیط مورد نظر اندازه‌گیری می‌نماید. فرایند انتخاب براساس ارزیابی صلاحیت مدل فردی است. صلاحیت یک مدل فردی مرتبط با هدف تابع می‌باشد. از عملگرهای ژنتیکی الهام گرفته از طبیعت مثل تلفیق و جهش جهت تولید جمعیت جدید استفاده می‌شود. عملیات انتخاب اغلب طی دو مرحله انجام می‌شود. انتخاب والدین و ابقاء. در انتخاب والدین الگوریتم والدین را تعیین کرده و معین می‌سازد که آنها چند فرزند داشته باشند. فرزندان از طریق تلفیق یا همان انتقال اطلاعات بین دو والد و یا جهش خلق می‌شوند. این عملگر نقاط جدیدی را به صورت تصادفی در فضای جستجو ایجاد می‌کند. سپس فرزندان دوباره ارزیابی می‌گردند و در نهایت در گام ابقاء تعیین می‌شود که کدام یک از آنها در جمعیت جدید ابقاء شوند (شکل ۳).

در این مطالعه از طریق مطالعه ژئوشیمی پتروفیزیک میدان نفتی اهواز اهدافی چون دستور سازی داده‌های پتروفیزیک به داده‌های ژئوشیمیایی با استفاده از مدل‌های رگرسیون خطی و رگرسیون چند خطی، دستور سازی داده‌های پتروفیزیک به داده‌های ژئوشیمیایی با استفاده از مدل‌های الگوریتم ژنتیک خطی و غیرخطی، تطابق بین آنالیز رگرسیون و برنامه‌نویسی ژنتیکی برای حل ضرائب مناسب با معادلات کل کربن آلی و زون‌بندی ژئوشیمیایی و توصیف رخساره‌های آلی با استفاده از تکنیک آنالیز خوشه‌ای دنبال می‌شود.

## مواد و روش‌ها

### الگوریتم ژنتیک

الگوریتم ژنتیک یک تکنیک بهینه‌سازی بر اساس فرآیند تکامل طبیعی است. الگوریتم ژنتیک از اصول انتخاب طبیعی داروین برای یافتن فرمول بهینه جهت پیش‌بینی الگو استفاده می‌کند. در الگوریتم ژنتیک ابتدا جمعیتی از افراد تشکیل شده و سپس این افراد نسل به نسل با روال تکراری تکامل پیدا می‌کنند. هر مرحله تکاملی یک نسل نامیده می‌شود، هر یک از افراد جمعیت حاضر از کد خارج شده و بر مبنای یک سری ملاک‌های کیفی از پیش تعریف شده ارزیابی می‌شوند. این کار توسط تابع شایستگی انجام شده و نتیجه کار تحت عنوان میزان شایستگی فرد شناخته می‌شود. برای



شکل ۳- نمودار گردش کلی الگوریتم ژنتیک [۱۴]

### روش آنالیز خوشه‌ای

جهت اعتبارسنجی مدل استفاده شد. محدوده داده‌های ژئوشیمیایی و پتروفیزیکی سنگ‌های منشأ مطالعه شده در جدول ۱ نشان داده شده است. برای آماده‌سازی داده‌ها، داده‌های چاه‌پیمایی کنترل شد و فواصل بد چاه حذف گردید. برای یکی شدن محدوده داده‌ها، همه داده‌های ورودی و خروجی بین ۰ و ۱ نرمالایز شدند. نرمالایز کردن امر مهمی است که می‌تواند جهت تولید بهتر داده، بر روی داده‌ها اعمال شود. این کار به خصوص برای تولید کل کربن آلی مفید است.

### انتخاب ورودی‌ها

این گام از کار نقش مهمی را در ساخت مدل ایفا می‌کند. به طور کلی، ارتباط قوی بین داده‌های ورودی و خروجی، می‌تواند پیشگویی دقیق‌تری را نسبت به داده‌های با تطابق ضعیفتر ایجاد کند. ارتباط داده‌های پتروفیزیکی با مقدار کل کربن آلی در شکل ۴ نشان داده شده است. تطابق نشان می‌دهد که نگار گاما، نوترون و صوتی ارتباط نزدیکی با کل کربن آلی دارند، در حالی که این ارتباط برای داده‌های نگار مقاومت ضعیف‌تر است.

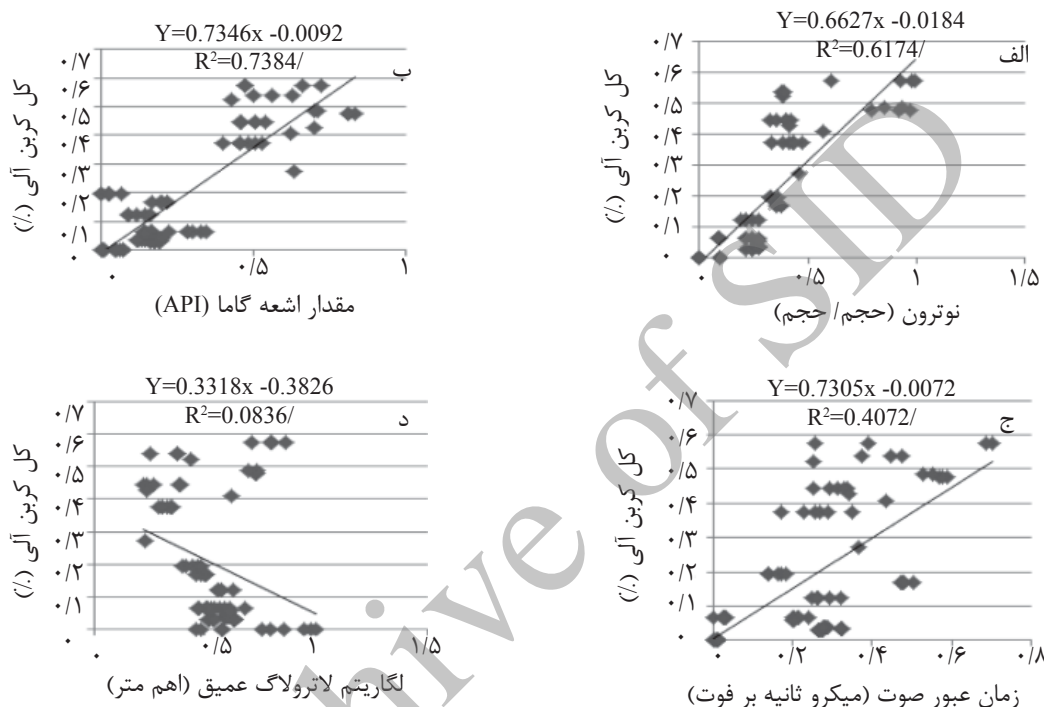
در این مطالعه، تکنیک آنالیز خوشه‌ای برای شناسایی رخساره‌های آلی بر اساس داده‌های کل کربن آلی مشتق شده از مدل الگوریتم ژنتیک به کار رفته است. آنالیز خوشه‌ای رده‌بندی اهداف در گروه‌های مختلف، یا دقیقتر، تفکیک مجموعه داده‌ها به زیر مجموعه‌ها (خوشه‌ها) است. خوشه‌بندی داده‌ها یک تکنیک متداول برای آنالیز آماری داده‌ها است.

### مطالعه موردی

در این مطالعه سیستم هوشمند الگوریتم ژنتیک برای تولید نگار کل کربن آلی از داده‌های پتروفیزیکی گاما، نوترون، صوتی و مقاومت استفاده شده است. مجموعه‌ای از ۱۰۳ داده ژئوشیمیایی اندازه‌گیری شد و تطابق آنها با داده‌های پتروفیزیکی موجود در پنج حلقه چاه در میدان نفتی اهواز به ۸۱ داده آموزشی شامل ۲۲ داده تست تقسیم شد. داده‌های آموزشی شامل داده‌های پتروفیزیکی می‌باشند و تطابقشان با مقادیر کل کربن آلی برای ساخت معادلات کل کربن آلی به کار رفته است. داده‌های تست در این مطالعه

جدول ۱- پارامترهای آماری داده‌های ژئوشیمیایی و پتروفیزیکی سنگ‌های منشأ

نگارها پارامتر	CNL (v/v)	LLD (ohm.m)	DT ( $\mu\text{s}/\text{ft}$ )	GR (API)
حداقل	۰/۰۰۰	۰/۲۷۵	۵۱/۹۵۲	۱۸/۵۲۷
حداکثر	۰/۴۲۷	۱۹۷۷/۰۷۵	۱۱۸/۸۹۱	۷۱/۳۳۱
میانگین	۰/۰۸۴	۱۶۲/۷۴۵	۷۶/۹۸۰	۴۵/۹۴۶
انحراف استاندارد	۰/۰۹۹	۴۳۴/۱۰۰	۱۵/۸۷۹	۱۳/۴۱۶



شکل ۴- نمایش ارتباط بین TOC و داده‌های چاه‌پیمایی متداول شامل (الف) نوترون، (ب) گاما، (ج) زمان عبور صوت و (د) مقاومت

می‌باشد. دو مدل ممکن از معادله ۱ (یعنی الگوریتم ژنتیک خطی و غیرخطی) جهت تولید نگار کل کربن آلی از داده‌های چاه‌پیمایی طراحی شد، که در زیر بحث شده است.

#### مدل خطی

معادله ۲ برای تخمین نهایی کل کربن آلی از طریق الگوریتم ژنتیک به کار گرفته شد:

$$TOC_{GA(lin)} = C_1 \cdot GR + C_2 \cdot CNL + C_3 \cdot DT + C_4 \cdot LLD \quad (2)$$

در این معادله پارامترهای  $C_1$ ،  $C_2$ ،  $C_3$  و  $C_4$  به ترتیب ضرائب مربوط به ورودی‌های GR، CNL، DT و LLD هستند. تابع شایستگی که باید توسط الگوریتم ژنتیک مینیمم تعیین شود، به عنوان MSE از پیش‌گویی داده‌های آموزشی تعریف می‌شود (معادله ۳)

#### استفاده از الگوریتم ژنتیک برای پیش‌گویی کل کربن آلی

نگار کل کربن آلی اطلاعات با ارزشی برای ارزیابی ژئوشیمی سنگ‌های منشأ هیدروکربن فراهم می‌سازد. در این بخش، مدل الگوریتم ژنتیک برای تعیین ضرایب وزنی از داده‌های آموزشی استفاده شد. سپس، آنها برای داده‌های تست به کار گرفته شدند (معادله ۱). در ذیل معادله استفاده شده برای تخمین نهایی کل کربن آلی از داده‌های چاه‌پیمایی توسط الگوریتم ژنتیک آمده است:

$$TOC_{GA} = \sum_{i=1}^n C_i \cdot x_i \quad (1)$$

در این معادله،  $n$  تعداد کل لاگ‌های پتروفیزیکی ورودی،  $C_i$  ضریب وزنی برای هر یک از لاگ‌ها و  $x_i$  مقدار عددی هر یک از لاگ‌های پتروفیزیکی

در این مدل مقدار میانگین مربع خطا و ضریب همبستگی برای داده‌های تست به ترتیب ۰/۰۰۹۹ و ۰/۷۷۵۲ به دست آمد (شکل ۶ الف). در شکل ۷ الف TOC اندازه‌گیری و TOC تطابق گرافیکی بین مقادیر شبیه سازی برای داده‌های تست با استفاده از مدل الگوریتم ژنتیک خطی نشان داده شده است.

#### مدل غیرخطی

معادله ۵ برای تخمین نهایی TOC از طریق الگوریتم ژنتیک به کار گرفته شد:

$$TOC_{GA(nlin)} = C_1 \cdot GR^{C_2} + C_3 \cdot CNL^{C_4} + C_5 \cdot DT^{C_6} + C_7 \cdot LLD^{C_8} \quad (5)$$

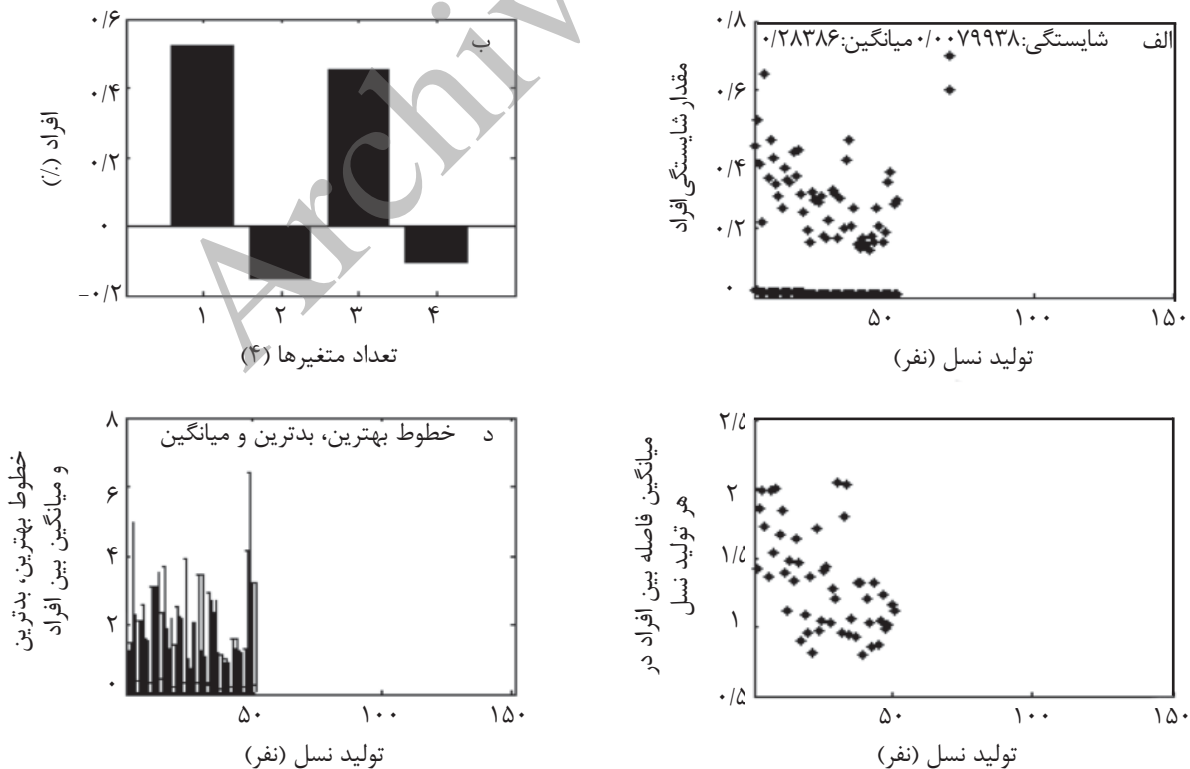
در این معادله پارامترهای  $C_1$  و  $C_2$ ،  $C_3$  و  $C_4$ ،  $C_5$  و  $C_6$ ،  $C_7$  و  $C_8$  به ترتیب ضرایب مربوط به ورودی‌های GR، DT، CNL و LLD هستند. تابع برازش، تابع هدفی است که به وسیله الگوریتم ژنتیک بهینه‌سازی می‌شود و در زیر تعریف شده است:

$$MSE = \sum_{i=1}^N \frac{1}{N} (TOC_{GA(nlin)} - TOC_{real})^2 \quad (6)$$

$$MSE = \sum_{i=1}^N \frac{1}{N} [TOCGA(lin) - TOC_{real}]^2 \quad (3)$$

N تعداد داده‌های آموزشی (۸۱ داده)، MSE میانگین مربع خطا،  $TOC_{real}$  تابع هدف (اندازه‌گیری‌های حاصل از آنالیز راک ایول)،  $TOC_{GA(lin)}$  مقدار کل کربن آلی محاسبه شده از الگوریتم ژنتیک می‌باشند. بعد از اجرای الگوریتم ژنتیک (شکل ۵)، ضرایب وزنی بهینه شده برای تولید آخرین خروجی (TOC) استفاده شدند. بهترین افراد، میانگین فاصله بین افراد و بهترین، بدترین و میانگین بین ۱۵۰ نسل در شکل ۵ نشان داده شده است. مقادیر مشتق شده از GA برای  $C_1$ ،  $C_2$ ،  $C_3$  و  $C_4$  مربوط به ورودی‌های GR، CNL، DT و LLD به ترتیب (۰/۴۵۵)، (۰/۵۲)، (۰/۱۵۷) و (۰/۱۰۸) محاسبه شد. الگوریتم ژنتیک پارامتر TOC را برای داده‌های تست (۲۲ داده) طبق معادله زیر محاسبه نمود:

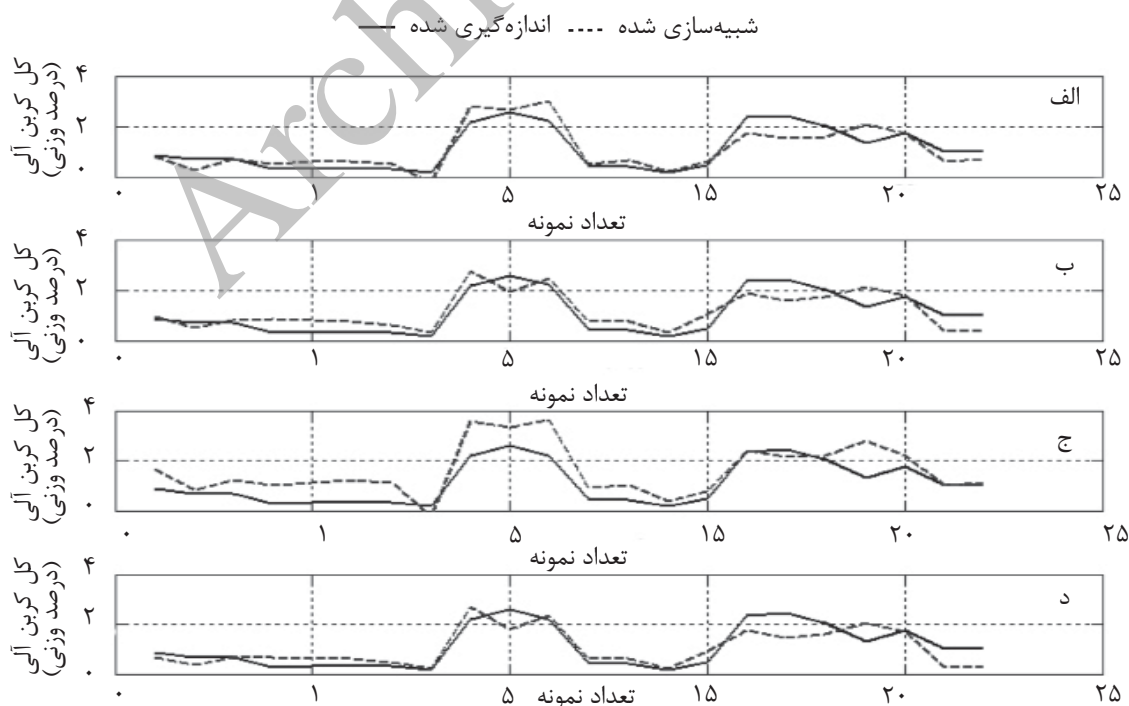
$$TOC_{GALIN} = 0.455 \cdot GR + 0.52 \cdot CNL + (-0.157) \cdot DT + (-0.108) \cdot LLD \quad (4)$$



شکل ۵- نمایش پلات (الف) مقادیر میانگین و شایسته برای تابع شایستگی، (ب) نمایش بهترین افراد، (ج) نمایش میانگین فاصله بین افراد و (د) نمایش خطوط بهترین، بدترین و میانگین بین ۱۲۰ نسل



شکل ۶- نمایش ارتباط ضرایب بین مقادیر TOC اندازه‌گیری شده و TOC ساخته شده برای داده‌های تست: (الف) با استفاده از مدل GA خطی، (ب) با استفاده از مدل GA غیرخطی، (ج) با استفاده از رگرسیون خطی و (د) با استفاده از رگرسیون چند خطی



شکل ۷- تطابق گرافیکی بین مقادیر TOC اندازه‌گیری و TOC شبیه‌سازی برای داده‌های تست: (الف) با استفاده از مدل GA خطی، (ب) با استفاده از مدل GA غیرخطی، (ج) با استفاده از رگرسیون خطی و (د) با استفاده از رگرسیون چند خطی

گاما و مقاومت استفاده شد. این تکنیک مرحله به مرحله با انتخاب رو به جلو یا حذف رو به عقب اجرا می‌شود و متغیرهای حاصل از مدل برای تشخیص زیر مجموعه مفیدی از پیشگویی کننده‌ها اضافه شده یا حذف می‌شوند. معادله رگرسیون چند خطی به شکل زیر بیان شده است:

$$TOC_{MLR} = 0.462*CNL - 0.222*LLD + 0.22*DT + 0.440*GR + 0.1005 \quad (9)$$

ضریب تطابق و تطابق گرافیکی بین مقادیر TOC اندازه‌گیری شده و TOC پیشگویی شده با استفاده از معادله ۹ در شکل‌های ۶ - د و ۷ - د نشان داده شده است. کارایی معادلات TOC به کمک ضرایب مشتق شده از نظر هوش و از نظر آمار در جدول ۲ مقایسه شده‌اند. همان‌گونه که نشان داده شده، مدل الگوریتم ژنتیک خطی نتایج معتبرتر و مقبول‌تری نسبت به مدل غیر خطی فراهم می‌سازد. مدل رگرسیون چند خطی با مقدار میانگین مربع خطا  $0.123$  بهتر از مدل رگرسیون تکی ( $MSE=0.268$ ) اجرا می‌شود. مقایسه بین نتایج نشان می‌دهد که دقت مدل الگوریتم ژنتیک ( $MSE=0.099$ ) بیشتر از مدل رگرسیون چند خطی است. ضریب همبستگی رگرسیون چند خطی تا حد قابل قبول به مدل الگوریتم ژنتیک خطی نزدیک است. به هر حال، همان‌گونه که در شکل ۶ د (نتایج رگرسیون چند خطی) نشان داده شده، شایسته‌ترین خط (خط قرمز)، از خط با زاویه ۴۵ درجه گذر کرده از مبدأ (خط چین) منحرف می‌شود. میانگین مربع خطا مدل الگوریتم ژنتیک خطی کوچک‌تر از رگرسیون چند خطی است. نتایج این مطالعه نشان می‌دهد هنگامی که ارتباط خطی قوی بین داده‌های ورودی و خروجی وجود داشته باشد، هم برنامه‌نویسی ژنتیکی و هم آنالیز رگرسیون تا حد قابل قبول به نتایج نزدیکی خواهند رسید. بنابراین، در مدل الگوریتم ژنتیک خطی داده‌های پتروفیزیکی چاه‌های  $61/55$ ،  $101/67$  و  $265$  به عنوان ورودی بکار رفتند و نگار TOC از آنها ساخته شد (شکل ۸ الف).

N تعداد داده‌های آموزشی (۸۱ داده)، MSE میانگین مربع خطا،  $TOC_{real}$  تابع هدف (اندازه‌گیری‌های حاصل از آنالیز راک-ایول)،  $TOC_{GA(nlin)}$  مقدار کل کربن آلی محاسبه شده از GA می‌باشند. بعد از ۱۵۰ نسل مقادیر مشتق شده از GA برای  $C_1$ ،  $C_2$  و  $C_3$  و  $C_4$ ،  $C_5$ ،  $C_6$ ،  $C_7$  و  $C_8$  به ترتیب  $0.704$ ،  $0.917$ ،  $0.704$ ،  $0.008$ ،  $0.011$ ،  $0.914$ ،  $0.13$  و  $0.913$  محاسبه شد. الگوریتم ژنتیک پارامتر TOC را برای داده‌های تست (۲۲ نقطه) طبق معادله زیر محاسبه نمود:

$$TOC_{GANLIN} = 0.704*GR^{(0.917)} + 0.008*CNL^{(0.949)} + 0.011*DT^{(0.914)} + 0.013*LLD^{(0.913)} \quad (7)$$

بر اساس شکل ۶ ب در مدل الگوریتم ژنتیک غیرخطی مقدار میانگین مربع خطا و ضریب همبستگی برای داده‌های تست به ترتیب  $0.122$  و  $0.6968$  حاصل شد. تطابق گرافیکی بین مقادیر TOC اندازه‌گیری و TOC شبیه‌سازی برای داده‌های تست با استفاده از مدل الگوریتم ژنتیک غیرخطی در شکل ۷ ب نشان داده شده است.

#### مقایسه نتایج الگوریتم ژنتیک با آنالیز رگرسیون

در این بخش، راه حل‌های GA به واسطه ضرایب مناسب با معادلات TOC با روش‌های رگرسیون خطی و رگرسیون چند خطی مقایسه شد. نتایج اجرای آنالیز رگرسیون تکی نشان می‌دهد که TOC می‌تواند از داده‌های پرتو گاما با استفاده از معادله ۸ تخمین زده شود:

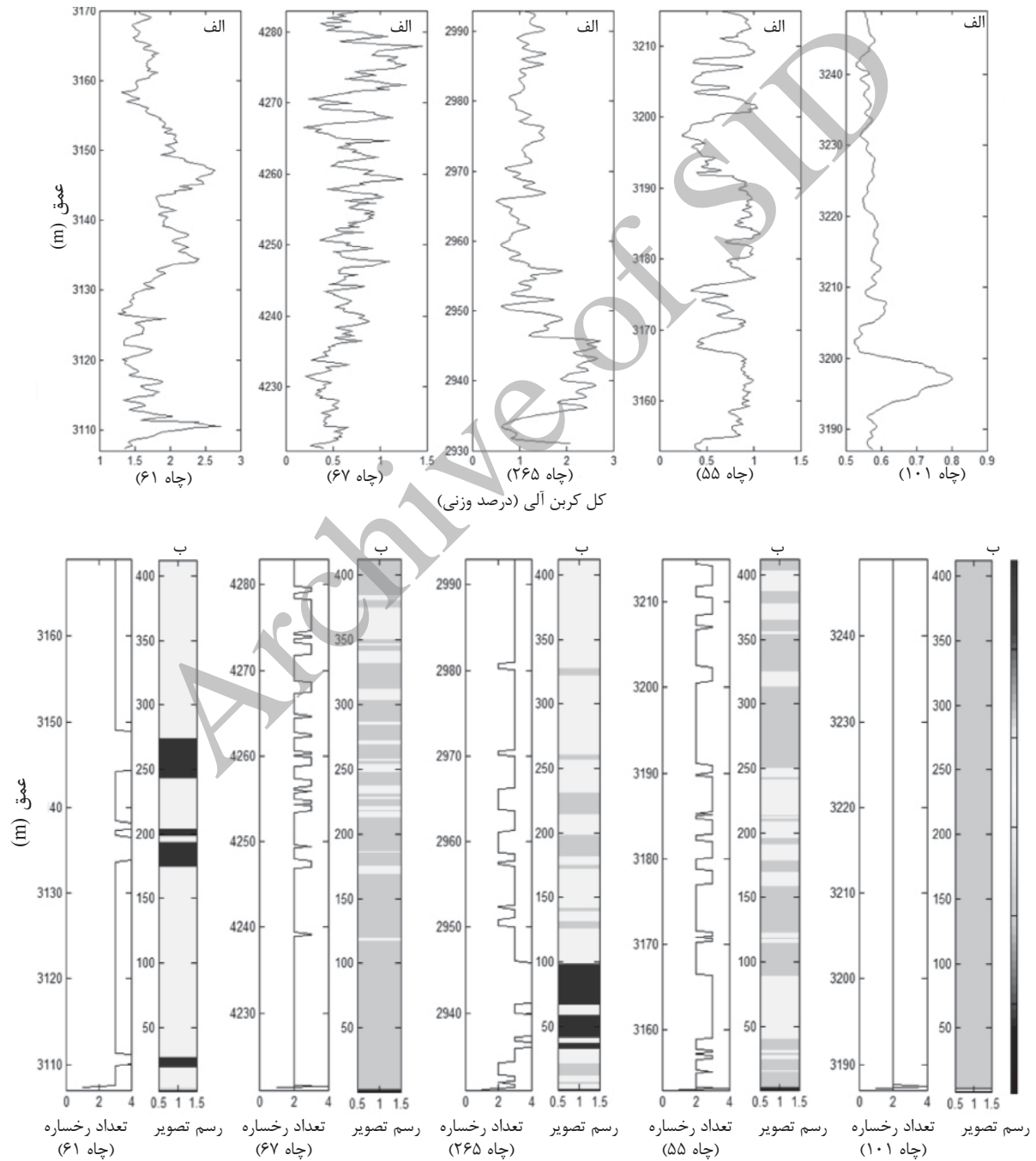
$$TOC_{LR} = 0.735*GR - 0.00921 \quad (8)$$

مقدار میانگین مربع خطا برای تخمین TOC در داده‌های تست با استفاده از معادله ۸ و مقدار ضریب همبستگی به ترتیب  $0.268$  و  $0.76$  می‌باشد. ضریب تطابق و تطابق گرافیکی بین مقادیر TOC اندازه‌گیری شده و TOC پیشگویی شده با استفاده از معادله ۸ در شکل‌های ۶ - ج و ۷ - ج نشان داده شده است. در روش رگرسیون چند خطی، از یک رگرسیون مرحله‌ای برای تشخیص معادله بین TOC (متغیر واکنش) و داده‌های چاه‌پیمایی متداول (پیشگویی کننده‌ها) شامل نوترون، صوتی، پرتو



جدول ۲- نمایش معادلات TOC با استفاده از ضرائب مشتق شده از نظر هوشی و آماری

R <sup>2</sup>		MSE		معادله	مدل
داده‌های تست	داده‌های آموزشی	داده‌های تست	داده‌های آموزشی		
۰/۷۷۵	۰/۸۰۲	۰/۰۰۹۹	۰/۰۲۱۸	$TOC_{GALIN} = 0.455 * GR + 0.52 * CNL + (-0.157) * DT + (-0.108) * LLD$	الگوریتم ژنتیک خطی
۰/۶۹۶	۰/۷۳۴	۰/۰۱۲۲	۰/۰۱۱۱	$TOC_{GANLIN} = 0.704 * GR^{(0.917)} + 0.008 * CNL^{(0.949)} + 0.011 * DT^{(0.914)} + 0.013 * LLD^{(0.913)}$	الگوریتم ژنتیک غیرخطی
۰/۷۶۰	۰/۷۸۱	۰/۰۲۶۸	۰/۰۱۰۱	$TOC_{LR} = 0.735 * GR - 0.00921$	رگرسیون خطی
۰/۶۹۴	۰/۷۳۸	۰/۰۱۲۳	۰/۰۱۰۰	$TOC_{MLR} = 0.462 * CNL - 0.222 * LLD + 0.22 * DT + 0.440 * GR + 0.1005$	رگرسیون چند خطی



شکل ۸- (الف) پلات‌های TOC شبیه‌سازی شده را در چاه‌های ۵۵، ۶۱، ۶۷، ۱۰۱ و ۲۶۵ نشان می‌دهند، (ب) تصاویر رخساره‌های آلی مشتق شده از آنالیز خوشه‌ای را در چاه‌های ۵۵، ۶۱، ۶۷، ۱۰۱ و ۲۶۵ نشان می‌دهند.

از خطوط U شکل مرتبط با اهداف زیاد در یک درخت سلسله‌ای تشکیل شده است. ارتفاع هر U فاصله بین دو هدف متصل شده را نشان می‌دهد. آخرین مرحله آنالیز خوشه‌ای تولید خوشه داده‌ها با استفاده از مجموعه‌ای از مقدار حد برش<sup>۲</sup> است. مقدار حد برش برای فواصل گروه‌بندی شده توسط تابع اتصال بکار رفته است. درخت خوشه‌ای در مقدار حد برش قطع شده است و خوشه‌ها تولید شده‌اند. در این مطالعه، چندین گروه از خوشه‌ها از طریق مقادیر مختلف حد برش به کار رفته یا آستانه‌ها تولید شده‌اند. یک آزمایش مبنی بر آزمون و خطا نشان داد که استخراج چهار خوشه سازگارتر با سنگهای منشأ ناهمگن است. در زمان میانگین، شناسایی شاخه‌های بیشتر، آنالیز کل کربن آلی و مدل‌سازی آنها را پیچیده‌تر خواهند کرد و خوشه‌های کمتر نمی‌توانند تمام حوضه‌ها را پوشش دهند (انعکاس سنگ منشأ ناهمگن). بنابراین تابع تولید خوشه اجرا شد و چهار خوشه بر اساس نگارهای کل کربن آلی ساخته شده حاصل شد. هر خوشه تعیین شده در این روش به یک رخساره آلی نسبت داده شده است و توسط مجموعه‌ای از پارامترهای آماری شامل میانگین، متوسط، انحراف معیار، حداقل و حداکثر داده کل کربن آلی مشخص شده است (جدول ۳). همانگونه که نشان داده شده، رخساره ۴ یک سنگ منشأ غنی از ماده آلی با میانگین کل کربن آلی ۲/۳۵۳٪ وزنی است، در حالی که رخساره آلی ۱ با میانگین کل کربن آلی ۰/۰۰۱٪ وزنی یک سنگ مادر فقیر است. کل کربن آلی ساخته شده توسط مدل الگوریتم ژنتیک و تطابق رخساره‌های آلی برای چاه‌های ۵۵، ۶۱، ۶۷، ۱۰۱ و ۲۶۵ در شکل ۸ الف و ب نشان داده شده است.

### تشخیص رخساره‌های آلی با استفاده از تکنیک آنالیز خوشه‌ای

در این بخش از تحقیق، نگار TOC تولید شده توسط تکنیک الگوریتم ژنتیک برای شناسایی رخساره‌های آلی در چاه‌های ۵۵، ۶۱، ۶۷، ۱۰۱ و ۲۶۵ میدان نفتی اهواز استفاده شد. به همین منظور، تکنیک آنالیز خوشه‌ای که براساس مفهوم فاصله است، به کار برده شد. مهم‌ترین مرحله خوشه‌بندی انتخاب اندازه فواصل است، که تعیین خواهد کرد چگونه همانندی دو گروه محاسبه شده است. این انتخاب بر شکل خوشه مؤثر خواهد بود، به طوری که چند گروه طبق انتخاب یک فاصله نزدیک به هم و طبق انتخاب فاصله دیگر از یکدیگر دور باشند. از مقیاس بلوک شهر<sup>۱</sup> برای محاسبه فاصله داده‌ها بین مقادیر TOC استفاده شد، که به شکل معادله ۱۰ بیان شده است:

$$d(a,b) = \sum_{j=1}^n (|b_j - a_j|) \quad (10)$$

(a,b) فاصله بین نقاط a و b در ماتریس TOC است. سطرهای ماتریس TOC مطابق با متغیرها است و d(a,b) بردار سطری طول n(n-1)/2 مطابق با جفت متغیرها در ماتریس TOC است. بعد از محاسبه فاصله بین مشاهدات در ماتریس TOC، یک درخت خوشه‌ای سلسله‌ای از فاصله داده‌ها با استفاده از تابع اتصال ایجاد شد. تابع اتصال فاصله داده‌ها را گروه‌بندی کرده و ماتریس (m-1)-by-3 را محاسبه می‌کند. m تعداد متغیرها در داده‌های اصلی است. ستون‌های ۱ و ۲ ماتریس اتصال در بر دارنده شاخص‌های خوشه‌ای در جفت‌های بهم پیوسته هستند که یک درخت دوتایی تشکیل می‌دهند. گره‌های برگ از یک تا m شماره‌گذاری شده‌اند. در این مطالعه تابع اتصال کامل، برای گروه‌بندی فاصله داده‌ها استفاده شد. این تابع برای بیشترین فاصله بین هدف‌ها در دو نقطه به کار می‌رود:

$$d(a,b) = \max [\text{dist}(a_i, b_j)], i \in (1, \dots, n_a), j \in (1, \dots, n_b) \quad (11)$$

برای نمایش نتایج تابع اتصال به کار رفته، یک دندروگرام (درخت خوشه‌ای) برای سنگ‌های منشأ مطالعه شده استخراج شده است. یک دندروگرام

1. City Block  
2. Cut off

جدول ۳- پارامترهای آماری برای رخساره‌های آلی مشتق شده از آنالیز خوشه‌ای

انحراف استاندارد	میانگین	حداکثر	حداقل	پارامترها رخساره‌ها
۰/۰۸۸	۰/۰۰۱	۰/۱۴۸	-۰/۱۸۸	رخساره ۱
۰/۰۹۳	۰/۵۶۱	۰/۸۵۳	۰/۱۵۹	رخساره ۲
۰/۳۱۲	۱/۳۵۰	۲/۱۱۴	۰/۸۵۵	رخساره ۳
۰/۱۶۰	۲/۳۵۳	۲/۷۹۸	۲/۱۲۰	رخساره ۴

### نتیجه‌گیری

آمد. راه حل‌های الگوریتم ژنتیک به واسطه ضرائب مناسب معادلات کل کربن آلی با آنالیز رگرسیون مقایسه شد. ضریب همبستگی رگرسیون چند خطی تا حد قابل قبول به مدل الگوریتم ژنتیک خطی نزدیک است. با این وجود، میانگین مربع خطا مدل الگوریتم ژنتیک خطی کوچک‌تر از رگرسیون چند خطی است. چهار رخساره آلی از طریق اجرای آنالیز خوشه‌ای شناسایی شد. نتایج گویای این است که نگار کل کربن آلی می‌تواند سنگ منشأ ناهمگون را بهتر نشان دهد. این یک روش سریع، دقیق و مؤثر برای زون‌بندی و طبقه‌بندی سنگ منشأ است.

### مراتب تشکر و قدردانی

نویسندگان از شرکت مناطق نفت‌خیز جنوب به دلیل حمایت مالی و تهیه داده و اجازه برای نشر نتایج این تحقیق و همچنین از دانشگاه تبریز به دلیل حمایت مالی جهت اجرای این تحقیق کمال سپاس‌گزاری و تشکر را دارند.

در این مطالعه از مدل بهینه‌سازی الگوریتم ژنتیک جهت تخمین پارامتر کل کربن آلی از داده‌های چاه‌پیمایی در میدان نفتی اهواز استفاده شد. داده‌های چاه‌پیمایی (مقاومت، زمان عبور صوت، نوترون و گاما) به عنوان ورودی و TOC حاصل از آنالیز به عنوان خروجی مطلوب شبکه تعیین گردید. در اجرای این روش انتخاب ورودی مناسب بسیار حائز اهمیت است. دو نوع مدل الگوریتم ژنتیک جهت تخمین کل کربن آلی طراحی شد: خطی و غیر خطی. نتایج نشان می‌دهد که از مدل خطی نسبت به مدل غیرخطی نتایج قابل قبول‌تری حاصل می‌شود. هر یک از داده‌های چاه‌پیمایی ورودی یک ضریب وزنی برای تخمین کل کربن آلی دارند که توسط الگوریتم ژنتیک بدست آمده است. ضرائب وزنی مشتق شده از الگوریتم ژنتیک برای  $C_1$ (GR)،  $C_2$ (CNL)،  $C_3$ (DT) و  $C_4$ (LLD) به ترتیب ۰/۴۵۵، ۰/۵۲، -۰/۱۵۷ و -۰/۱۰۸ محاسبه شد و مقدار میانگین مربع خطا و ضریب همبستگی برای داده‌های تست به ترتیب ۰/۷۶۳۵ و ۰/۰۰۹۶ بدست

### مراجع

- [1]. Schmoker J. W., "Determination of organic-matter content of Appalachian Devonian shales from gamma-ray logs", American Association of Petroleum Geologists Bulletin, Vol. 65, pp. 2165-2174, 1981.
- [2]. Passey O. R., Moretti F. U., and Stroud J. D., "A practical modal for organic richness from porosity and resistivity logs", American Association of Petroleum Geologists Bulletin., Vol. 74, pp. 1777-1794, 1990.
- [3]. Kadkhodaie-Ilkhchi A., Rahimpour-Bonab H., and Rezaee M. R., A., "Committee machine with intelligent systems for estimation of total organic carbon content from petrophysical data: an example from Kangan and Dalan reservoirs in South Pars Gas Field", Iran, Journal of Comput. Geosci., Vol. 35, pp. 459-474, 2008.

- [4] Kadkhodaie A., Rezaee M. R., and Rahimpour-Bonab H., A., "Committee Neural Network for Prediction of Normalized Oil Content from Well Log Data: an Example from South Pars Gas Field", Persian Gulf, Journal of Petroleum Science and Engineering, Elsevier Publications., Vol. 65, pp. 23-39, 2009.
- [5]. Kamali M. R. and Mirshady A. A., "Total organic carbon content determined from well logs using  $\Delta\log R$  and neuro fuzzy techniques", Journal of Petroleum Science and Engineering., Vol. 45, pp. 141-148, 2004.
- [6]. Mitchell M., *An Introduction to genetic algorithms*, MIT Press, Cambridge (MA)., 1996.
- [7]. Holland J. H., *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor., 1975.
- [8]. Goldberg D. E., *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, Reading, MA., 1989.
- [9]. Michalewicz Z., *Genetic algorithms +data structures=evolution programs*, Springer Verlag, Berlin., 1992.
- [10]. Bordenave M. L. and Burwood R., "Source rock distribution and maturation in the Zagros orogenic belt: provenance of the asmari and bangestan reservoir oil accumulations", Organic Geochemistry, Vol. 16, pp. 369-387, 1990.
- [11]. Alizadeh B. and Sepahvand S., "Petroleum source rock evaluation of Kazhdumi formation", Southwestern Iran, GeoArabia., Vol. 6, No. 2, 282, 2001.
- [12]. Rabbani A. R. and Bagheri Tirtashi R., "Hydrocarbon source rock evaluation of the super giant ahwaz oil field, SW Iran", Australian Journal of Basic and Applied Sciences, Vol. 4, No. 5, pp. 673-686, 2010.
- [13]. Jafarzadeh M. and Hosseini-Barzi M., "Petrography and geochemistry of Ahwaz sandstone member of asmari formation, Zagros, Iran: implications on provenance and tectonic setting", Revista Mexicana de Ciencias Geológicas., No. 25(2), pp. 247-260, 2008.
- [14]. Chen C. H. and Lin Z. S., "A committee machine with empirical formulas for permeability prediction", Comput., Geosci., Vol. 32, pp. 485-496, 2006.