





دانشکده مهندسی معدن، نفت و ژئوفیزیک

گروه نفت و ژئوفیزیک

کارشناسی ارشد ژئوفیزیک-گرایش لرزه‌شناسی

کاربرد روش آدابوست در تفسیر داده‌های لرزه‌ای بازتابی

پدرام ابراهیمی

اساتید راهنما:

دکتر امین روشندل کاهو

دکتر سید رضا قوامی ریابی

استاد مشاور:

دکتر بهزاد تخم‌چی

تیرماه ۱۳۹۵

تقدیم به

پدرم، که عالمانه به من آموخت تا چگونه در عرصه زندگی، ایستادگی را تجربه نمایم

مادرم، دریای بی کران فداکاری و عشق که وجودم برایش همه نچ بود و وجودش برایم همه مهر

و اعضای خانواده ام که همیشه و همه جا برایم دلگرمی به همراه داشته اند...

تعهد نامه

اینجانب **پدرام ابراهیمی** دانشجوی دوره کارشناسی ارشد رشته ژئوفیزیک - گرایش لرزه‌شناسی دانشکده مهندسی معدن، نفت و ژئوفیزیک دانشگاه صنعتی شاهرود نویسنده پایان‌نامه کاربرد روش آدابوست در تفسیر داده‌های لرزه‌ای بازتابی تحت راهنمایی دکتر امین روشندل کاهو و دکتر سید رضا قوامی ریابی و مشاوره دکتر بهزاد تخم‌چی متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه شاهرود می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

چکیده

تعیین ویژگی‌های پتروفیزیکی و مخزنی از اهداف مهم اکتشاف نفت است. برای تخمین ویژگی‌های پتروفیزیکی و مخزنی با اطمینان بالا، استفاده از انواع داده‌های لرزه‌ای و چاه‌نگاری به صورت مکمل ضروری است. با استفاده از نشانگرهای لرزه‌ای، اطلاعات پنهان شده در داده‌های خام لرزه‌ای استخراج می‌شوند و با یافتن روابطی میان نشانگرهای لرزه‌ای و پارامترهای پتروفیزیکی مخزن در مکان چاه‌ها، می‌توان توزیع این پارامترها را در سایر نقاطی که داده‌های حاصل از نگارهای چاه وجود ندارند با دقت بهتری به دست آورد.

مشکل اساسی که در راه تلفیق داده‌های لرزه‌ای و داده‌های چاه وجود دارد، روش ادغام این داده‌ها با یکدیگر به منظور تخمین خصوصیات مخزن می‌باشد. تحقیقات زیادی در این زمینه انجام شده است که امکان برآورد خصوصیات پتروفیزیکی مخزن را با استفاده از داده‌های لرزه‌ای (خام و نشانگرها) و نگارهای چاه‌نگاری نشان می‌دهد. این مطالعات که تحت عنوان کلی تحلیل چند نشانگری انجام شده‌اند، اکثراً با استفاده از روش‌های متداول رگرسیون و یا شبکه‌های عصبی به تلفیق داده‌های لرزه‌نگاری و چاه‌نگاری پرداخته‌اند و به نتایج قابل ملاحظه‌ای در توصیف و تخمین توزیع پارامترهای پتروفیزیکی مخزن رسیده‌اند. از دیگر روش‌های تلفیق داده‌های لرزه‌ای و چاه‌نگاری روش‌های طبقه‌بندی چند نشانگری می‌باشند. روش طبقه‌بندی آدابوست که در سال‌های اخیر توجه بیشتری به آن شده است و امروزه جزو الگوریتم‌های برتر داده‌کاوی محسوب می‌شود. ویژگی برجسته آدابوست تنظیم طبقه‌بندی کننده‌ها در مرحله یادگیری است. این الگوریتم که نسبتاً ساده است و سرعت بالایی دارد تا حد زیادی مشکل بیش‌برازش سایر روش‌ها را حل می‌کند. در این پایان‌نامه از روش آدابوست به منظور طبقه‌بندی چند نشانگری در مطالعات لرزه‌نگاری بازتابی جهت طبقه‌بندی یک افق بر اساس اطلاعات پتروفیزیکی و شناسایی ساختارهایی نظیر گنبد نمکی استفاده می‌شوند. به این ترتیب دقت طبقه‌بندی افق لرزه‌ای بر مبنای تخلخل ۷۸ درصد بوده است. همچنین در بخش تعیین مزره‌های گنبد نمکی از اطراف و کف گنبد نمکی، دقت ۹۴ درصد بوده است. در مقایسه با روش طبقه‌بندی

SVM که دارای دقت ۹۱ درصد است، طبقه‌بندی با روش آدابوست دقت بالاتر دارد.

کلمات کلیدی: طبقه بندی، آدابوست، گنبد نمکی، نشانگرهای لرزه ای.

مقالات مستخرج از پایان نامه

ابراهیمی پ، روشندل ا، قوامی ریابی ر، تخم‌چی ب، (۱۳۹۴)، " تعیین محدوده گنبد نمکی با استفاده از طبقه‌بندی نشانگرهای بافتی لرزه‌ای به روش آدابوست"، سی و چهارمین گردهمایی و دومین کنگره تخصصی بین‌المللی علوم زمین، تهران.

فهرست مطالب

- فصل اول: کلیات..... ۱
- ۱-۱ تعریف مساله ۲
- ۲-۱ اهداف و ضرورت انجام تحقیق ۲
- ۳-۱ سابقه و مروری بر کارهای انجام شده ۳
- ۴-۱ روش انجام تحقیق ۶
- فصل دوم: تخلخل و نشانگرهای لرزه‌ای..... ۹
- ۱-۲ تخلخل..... ۱۰
- ۲-۲ عوامل تعیین کننده میزان تخلخل..... ۱۱
- ۳-۲ تقسیم بندی تخلخل..... ۱۲
- ۱-۳-۲ تخلخل مطلق ϕ_i ۱۲
- ۲-۳-۲ تخلخل مؤثر (ϕ_e) ۱۳
- ۴-۲ نشانگرهای لرزه‌ای..... ۱۳
- ۱-۴-۲ طبقه بندی نشانگرها ۱۵
- ۲-۴-۲ محاسبه و بررسی نشانگرهای عمومی ۱۸
- ۱-۲-۴-۲ فاز لحظه‌ای..... ۱۸
- ۲-۲-۴-۲ فرکانس لحظه‌ای..... ۱۹
- ۳-۲-۴-۲ مقاومت ظاهری لرزه‌ای نسبی..... ۲۰

- ۲۱.....فاکتور کیفیت لحظه‌ای.....۴-۲-۴-۲
- ۲۲.....پوش ردلرزه.....۵-۲-۴-۲
- ۲۳.....مشتق اول پوش ردلرزه.....۶-۲-۴-۲
- ۲۴.....مشتق دوم پوش ردلرزه.....۷-۲-۴-۲
- ۲۶.....نشانگرهای بافتی.....۳-۴-۲
- ۲۶.....نشانگر لرزه‌ای بافتی مبتنی بر GLCM.....۱-۳-۴-۲
- ۳۳.....فصل سوم: روشهای ترکیب نشانگرها.....
- ۳۴.....۱-۳ مقدمه.....
- ۳۴.....۲-۳ رگرسیون خطی چند نشانگری.....
- ۳۴.....۱-۲-۳ انجام تحلیل یک نشانگری.....
- ۳۷.....۲-۲-۳ استفاده از منحنی تقاطعی در چند نشانگرها.....
- ۳۷.....۳-۲-۳ رگرسیون چند نشانگری با ضرایب وزنی هم‌میختی.....
- ۳۹.....۴-۲-۳ تعیین بهترین نشانگرها با استفاده از رگرسیون مرحله‌ای.....
- ۴۰.....۳-۳ شبکه‌های عصبی.....
- ۴۲.....۱-۳-۳ شبکه عصبی چندلایه‌ای پیشرو.....
- ۴۳.....۲-۳-۳ شبکه عصبی احتمالاتی.....
- ۴۶.....۳-۳-۳ اعتبار سنجی.....
- ۴۹.....۴-۳ مبانی و اصول طبقه‌بندی و خوشه‌بندی.....
- ۵۰.....۵-۳ خوشه‌بندی.....

- ۵۱.....۳-۶ طبقه‌بندی.....
- ۵۲.....۳-۶-۱ طبقه‌بند k نزدیک‌ترین همسایه.....
- ۵۴.....۳-۶-۲ ماشین بردار پشتیبان.....
- ۵۵.....۳-۶-۱-۲ روش طبقه‌بندی SVM.....
- ۵۵.....۳-۶-۲-۲ نحوه تشخیص ابر سطح جدا کننده توسط SVM.....
- ۵۷.....۳-۶-۳ روش آداپوست.....
- ۵۷.....۳-۶-۱-۳ طبقه‌بندی کننده‌های بهبود یافته.....
- ۵۸.....۳-۶-۴ آداپوست.....
- ۶۳.....۳-۷-۷ مقایسه دقت روش‌های knn, SVM و آداپوست.....
- ۶۴.....۳-۷-۱ مدل مصنوعی دوبعدی با دو کلاس.....
- ۶۶.....۳-۷-۲ مدل مصنوعی دو بعدی با سه کلاس.....
- ۶۷.....۳-۷-۳ مدل مصنوعی سه بعدی با سه کلاس.....
- ۷۱.....فصل چهارم: اعمال روش بر روی داده‌های واقعی.....
- ۷۲.....۴-۱ مقدمه.....
- ۷۲.....۴-۱-۱ داده‌های میدان بلک فوت.....
- ۷۶.....۴-۲ تخمین تخلخل به کمک نشانگرهای لرزه‌ای.....
- ۷۷.....۴-۲-۱ تحلیل به کمک یک نشانگر (رگرسیون معمولی).....
- ۷۸.....۴-۲-۲ تحلیل به کمک چند نشانگر (رگرسیون چند متغیره).....
- ۸۳.....۴-۲-۳ تحلیل به کمک شبکه‌های عصبی.....

۳-۴ استفاده از روش آدابوست در طبقه‌بندی تخلخل..... ۸۶

۱-۳-۴ مرحله اول آماده‌سازی داده‌ها..... ۸۶

۴-۴ تعیین محدوده گنبد نمکی با استفاده از روش آدابوست..... ۸۹

فصل پنجم: نتیجه‌گیری و پیشنهادات ۹۵

۱-۵ نتیجه‌گیری ۹۷

۲-۵ پیشنهادات ۹۷

فهرست اشکال

- شکل ۱-۲. تخلخل‌های مختلف و کاربردی در کارهای نفتی..... ۱۳
- شکل ۲-۲. توسعه نشانگرهای لرزه‌ای و ارتباط نشانگرها با پیشرفت‌های اساسی در اکتشافات لرزه‌ای
..... ۱۴
- شکل ۳-۲. طبقه‌بندی نشانگرهای لرزه‌ای بر مبنای خواص کینماتیک/ دینامیک موج با تغییر ۱۸
- شکل ۴-۲. چهار نشانگر مختلف: الف) دامنه لرزه‌ای. ب) امپدانس صوتی ج) فرکانس لحظه‌ای د) فاز
لحظه‌ای..... ۲۱
- شکل ۵-۲. چهار نوع نشانگر لرزه‌ای: الف) پوش ردلرزه، ب) فرکانس لحظه‌ای، ج) مشتق اول پوش
ردلرزه، د) مشتق دوم پوش ردلرزه (والنتی ۲۰۰۹)..... ۲۵
- شکل ۶-۲: همسایگی‌های مختلف از عنصر مورد نظر ۲۷
- شکل ۷-۲. نمایش شماتیک محاسبه ماتریس GLCM در راستاهای مختلف با گام یک ۲۹
- شکل ۱-۳. نگار هدف در سمت چپ و سه نشانگر لرزه‌ای در سمت راست قرار دارند..... ۳۵
- شکل ۲-۳. یک ماتریس $4*4$ از منحنی‌های متقاطع که در آن نگار هدف و تمام نشانگرهای نمایش
داده شده در شکل شماره ۱ در مقابل هم رسم شده‌اند..... ۳۶
- شکل ۳-۳. مقایسه بین الف) نگار هدف و ب) نشانگر لرزه‌ای با تأکید بر اختلاف در محتوای
فرکانسی..... ۳۸
- شکل ۴-۳. یک مثال شماتیک از تفاوت بین الف) وزن‌های تک نقطه‌ای به وسیله رابطه ۱ و ب)

وزن‌های همامیختی ایجاد شده توسط رابطه (۳-۳) (همپسون ۲۰۰۱)..... ۳۸

شکل ۳-۵. منحنی متقاطع نگار هدف در مقابل نشانگر لرزه‌ای. خط رگرسیون به‌وسیله رابطه کمینه

کننده () برازش..... ۴۱

شکل ۳-۶. ساختار یک شبکه عصبی ساده با یک لایه پنهان ۴۲

شکل ۳-۷. منحنی تخمین به دست آمده از روش MLFN با ۵ گره در لایه پنهان..... ۴۳

شکل ۳-۸. منحنی متقاطع شبکه عصبی PNN. نمودار برازش داده شده به داده‌ها در بازه پایینی

داده‌ها پایدارتر است..... ۴۶

شکل ۳-۹. مثال اعتبارسنجی متقاطع. نشانگر لرزه‌ای در مقابل نگار چاه..... ۴۷

شکل ۳-۱۰. بررسی خطای اعتبارسنجی با استفاده از روش رگرسیون..... ۴۹

شکل ۳-۱۱. روش تشخیص ابر سطح جدا کننده توسط روش SVM..... ۵۶

شکل ۳-۱۲. وزن‌دهی یکسان اولیه به تمام نمونه‌ها..... ۶۰

شکل ۳-۱۳. مرحله یک بهبود. در این مرحله بهترین فرضیه ضعیف h_1 می‌باشد که دارای خطای

$\varepsilon = 0.30$ و وزن $\alpha_1 = 0.42$ می‌باشد..... ۶۱

شکل ۳-۱۴. مرحله دو بهبود. در این مرحله بهترین فرضیه ضعیف h_2 به دست می‌آید. خطای

طبقه‌بندی این فرضیه ضعیف $\varepsilon_2 = 0.21$ و وزن آن $\alpha_2 = 0.65$ می‌باشد..... ۶۲

شکل ۳-۱۵. مرحله سه بهبود. بهترین فرضیه با خطای $\varepsilon_3 = 0.14$ و وزن $\varepsilon_3 = 0.92$ به دست

می‌آید..... ۶۲

شکل ۳-۱۶. نحوه ترکیب فرضیه‌های ضعیف به همراه وزن‌های مربوط به آن‌ها که طی مراحل

مختلف بهبود به دست می‌آیند..... ۶۳

شکل ۳-۱۷. نحوه ترکیب فرضیه‌های ضعیف به همراه وزن‌های مربوط به آن‌ها که طی مراحل

- مختلف بهبود به دست می‌آیند.....۶۳
- شکل ۳-۱۷. کل داده‌های مورد استفاده برای مدل مصنوعی دوبعدی با دو کلاس الف) کل داده های مورد استفاده از مثال ب) داده های مورد استفاده در مرحله آموزش ج) داده‌های مورد استفاده در مرحله آزمایش.....۶۴
- شکل ۳-۱۸. نحوه طبقه بندی داده‌های آموزش توسط روشهای الف) KNN. ب) SVM و ج) آدابوست.....۶۵
- شکل ۳-۱۹. داده‌های مربوط به مدل مصنوعی دو کلاسه و سه بعدی. داده‌های مورد استفاده در بخش آموزش الف) و داده‌های مورد استفاده برای آزمایش ب).....۶۶
- شکل ۳-۲۰. نحوه طبقه بندی داده‌های آموزش سه کلاسه دوبعدی توسط روشهای الف) KNN. ب) SVM و ج) آدابوست.....۶۷
- شکل ۳-۲۱. داده‌های مربوط به مدل مصنوعی سه بعدی و سه کلاسه. الف) کل داده‌ها، ب) داده‌های آموزشی، ج) داده‌های آزمایشی.....۶۹
- شکل ۳-۲۲. نحوه طبقه بندی داده‌های آموزش سه کلاسه سه بعدی توسط روشهای الف) KNN، ب) SVM و ج) آدابوست.....۷۰
- شکل ۴-۱: موقعیت جغرافیایی میدان بلک فوت، کانادا.....۷۲
- شکل ۴-۲: موقعیت چاه‌های حفر شده نسبت به محدوده برداشت داده‌های لرزه‌نگاری سه بعدی.....۷۳
- شکل ۴-۳. نگارهای موجود در چاه ۰۸-۰۸. به ترتیب از سمت چپ: نگار چگالی، سرعت موج p، نگار تخلخل و نگار امپدانس موج p.....۷۵
- شکل ۴-۴. الف) موقعیت خطوط لرزه‌نگاری و چاه‌های حفر شده در منطقه. ب) مکعب داده‌های لرزه‌ای.....۷۶

- شکل ۴-۵. نمایش داده‌های تخلخل، مقاومت ظاهری لرزه‌ای و ردلرزه در محل چاه‌ها.....۷۷
- شکل ۴-۶. نتایج اعتبار سنجی متقابل چند نشانگر لرزه‌ای با استفاده از روش رگرسیون چند متغیره.....۷۹
- شکل ۴-۷. نمودار مقاطع داده‌های تخلخل واقعی و تخلخل به دست آمده از روش چند نشانگری در محل چاه‌ها.....۸۰
- شکل ۴-۸. مقایسه نگارهای تخلخل واقعی و تخمین زده شده توسط روش رگرسیون چند متغیره در چاه‌ها.....۸۱
- شکل ۴-۹. مقطعی از تخلخل به دست آمده در محدوده چاه ۰۸-۱۶.....۸۲
- شکل ۴-۱۰. نمودار مقاطع داده‌های تخلخل واقعی و تخلخل به دست آمده به وسیله شبکه عصبی احتمالاتی در محل چاه‌ها.....۸۴
- شکل ۴-۱۱. نمایش تطابق نگار تخلخل واقعی و محاسبه شده توسط شبکه عصبی احتمالاتی در محل چاه.....۸۵
- شکل ۴-۱۲. نتایج تخمین تخلخل در محدوده مورد مطالعه با اعمال شبکه عصبی PNN.....۸۶
- شکل ۴-۱۳. نمونه‌هایی از نشانگرهای به دست آمده از روی داده‌های لرزه‌ای: (الف) دامنه لرزه‌ای، (ب) کسینوس فاز، (ج) مشتق اول، (د) پوش دامنه، (ه) فرکانس لحظه‌ای و (و) مشتق دوم در زمان ۱۰۶۰ میلی ثانیه می‌باشد.....۸۷
- شکل ۴-۱۴. نمودار تجمعی خطا به ازای تکرار الگوریتم در مرحله آزمایش. در تکرار ۶۶ام کمترین مقدار خطا حاصل می‌شود.....۸۸
- شکل ۴-۱۵. کلاسهای بدست آمده حاصل از استفاده از روش طبقه بندی آدابوست. برش زمانی در زمان (ms) ۱۰۶۰.....۸۹
- شکل ۴-۱۶. الف) مقطع فاز لحظه ای مربوط به یک گنبد نمکی، ب) مقطع لرزه‌ای واقعی مربوط به

گنبد نمکی مشابه ۹۰.....

شکل ۴-۱۷. نشانگرهای بافتی. (الف) سایه خوشه، (ب) برتری خوشه، (ج) اینرسی، (د)

واریانس. ۹۱.

شکل ۴-۱۸. مقایسه نتیجه به دست آمده از روش آدابوست و ماشین بردار پشتیبان..... ۹۳.....

فهرست جداول

- جدول ۱-۲. پارامترهای مورد استفاده در محاسبه نشانگرهای GLCM..... ۲۹
- جدول ۲-۲. خواص آماری قابل استخراج از ماتریس GLCM و روابط حاکم بر آنها..... ۳۰
- جدول ۱-۴. رأس سازندهای مشاهده شده در چاههای میدان بلک فوت..... ۷۴
- جدول ۲-۴. نتایج به کمک یک نشانگر (رگرسیون معمولی)..... ۷۸
- جدول ۳-۴. نشانگرهای مورد استفاده در روش رگرسیون چند متغیره و نتایج آنها..... ۸۰
- جدول ۴-۴. ماتریس درهم ریختگی مربوط به کلاسهای تخلخل..... ۸۹
- جدول ۵-۴. نتایج به دست آمده از روش آدابوست..... ۹۱
- جدول ۶-۴. نتایج به دست آمده از روش ماشین بردار پشتیبان..... ۹۲

فصل اول

کلیات

۱-۱ تعریف مساله

تعیین ویژگی‌های پتروفیزیکی و مخزنی از اهداف مهم اکتشاف نفت است. از مهم‌ترین و قابل‌اتکاترین روش‌های کسب این اطلاعات و شناسایی اهداف، حفاری چاه و مطالعه اطلاعات حاصله است. به دلیل گسترده بودن میدان‌های نفتی و هزینه بالای مربوط به بخش حفاری، استفاده از تعداد بالای چاه برای پوشش تمام منطقه امکان‌پذیر نیست. لذا برای تخمین ویژگی‌های پتروفیزیکی و مخزنی با اطمینان بالا، به منظور شبیه‌سازی‌های مخزنی استفاده از انواع داده‌ها و به‌صورت مکمل ضروری است.

از منابع مهم تأمین اطلاعات ویژه و همه‌جانبه، می‌توان مطالعات لرزه‌ای و چاه‌نگاری را برشمرد. علی‌رغم اینکه داده‌های حاصل از نگار چاه دارای تفکیک‌پذیری عمقی مناسب می‌باشند؛ توزیع آن‌ها پراکنده بوده و با توجه به تغییرات جانبی خصوصیات در محدوده مخزن، تفکیک‌پذیر جانبی آن‌ها مناسب نیست. از طرفی داده‌های لرزه‌نگاری دارای تفکیک‌پذیری عمقی نامناسبی هستند؛ اما تفکیک‌پذیری جانبی بهتری دارند و یک پوشش سه‌بعدی از منطقه را در اختیار مفسر قرار می‌دهند. این دو نوع داده با توجه به ویژگی‌های ذکرشده، با رفع نواقص هر یک توسط دیگری، می‌توانند به‌صورت مکمل به کار برده شوند [۱].

۱-۲ اهداف و ضرورت انجام تحقیق

استفاده از داده‌های لرزه‌ای خام پردازش‌شده به‌تنهایی کافی نیست و می‌بایست با استفاده از نشانگرهای لرزه‌ای، اطلاعات پنهان‌شده در داده‌های خام لرزه‌ای استخراج شوند و با یافتن روابطی میان نشانگرهای لرزه‌ای و پارامترهای پتروفیزیکی مخزن در مکان چاه‌ها، توزیع این پارامترها را در سایر نقاطی که داده‌های حاصل از نگارهای چاه وجود ندارند با دقت بهتری به دست آورد.

مشکل اساسی که در راه تلفیق داده‌های لرزه‌ای و داده‌های چاه وجود دارد، روش ادغام این داده‌ها با یکدیگر به‌منظور تخمین خصوصیات مخزن می‌باشد. تحقیقات زیادی در این زمینه انجام شده است که امکان برآورد خصوصیات پتروفیزیکی مخزن را با استفاده از داده‌های لرزه‌ای (خام و نشانگرها) و

نگارهای چاه‌نگاری نشان می‌دهد. این مطالعات که تحت عنوان کلی تحلیل چند نشانگری انجام شده‌اند، اکثراً با استفاده از روش‌های متداول رگرسیون و یا شبکه‌های عصبی به تلفیق داده‌های لرزه‌نگاری و چاه‌نگاری پرداخته‌اند و به نتایج قابل ملاحظه‌ای در توصیف و تخمین توزیع پارامترهای پتروفیزیکی مخزن رسیده‌اند.

۳-۱ سابقه و مروری بر کارهای انجام شده

یکی از اولین روش‌های تلفیق داده‌های لرزه‌ای و چاه‌نگاری روش مدل‌سازی معکوس داده‌های لرزه‌ای (وارون‌سازی لرزه‌ای) به‌منظور محاسبه نگر مقاومت ظاهری لرزه‌ای می‌باشد که در اوایل دهه ۱۹۷۰ مطرح شد. لیندست [۲]، کک و اشنایدر (۱۹۸۳)، آلدن برگ و همکاران [۳] و چی و همکاران (۱۹۸۴) از پیشگامان این روش بودند [۴].

در سال‌های بعد، برای تلفیق داده‌های حاصل از نگارهای چاه و داده‌های لرزه‌ای، روش‌های زمین‌آماری مورد استفاده قرار گرفت. دوچ [۵] روش سنتی کوکریجینگ و انواع آن را برای تلفیق داده‌های لرزه‌ای و داده‌های حاصل از نگارهای چاه به‌کاربرد. به دلیل استفاده از تنها یک نشانگر جهت برآورد نگارهای چاه این روش‌ها تا اواخر دهه ۸۰ میلادی از دقت مطلوبی برخوردار نبودند؛ بنابراین در اوایل دهه‌ی ۹۰ میلادی، شالتز و همکاران [۶،۷] و رون و همکاران [۸] برای اولین بار ایده استفاده از بیش از یک نشانگر (تحلیل چند نشانگری) را مطرح کردند. بدین ترتیب قدرت برآورد نگارهای چاه افزایش یافت.

در ابتدا به دلیل خطی در نظر گرفتن روابط میان نشانگرهای لرزه‌ای و خصوصیات پتروفیزیکی، از روش‌های آماری خطی برای یافتن این ارتباط استفاده می‌شد. به دلیل ماهیت غیرخطی روابط موجود بین نشانگرها و ویژگی‌های مخزنی، روش‌های آماری خطی قادر به تخمین این روابط با دقت بالایی نیستند؛ بنابراین به‌منظور ارتقاء همبستگی نشانگرها و داده‌های پتروفیزیکی و در نتیجه افزایش دقت پیش‌بینی‌ها، رون و همکاران، شالکه و همکاران [۹] و شالکه و کورین [۱۰] از شبکه عصبی استفاده

کردند. آن‌ها شبکه عصبی را به شیوه‌ای جدید به کار گرفتند تا رابطه بین نشانگرهای لرزه‌ای، مقاومت صوتی و تخمین تخلخل از روی نگارهای چاه را پیدا کند.

پس از آن دانشمندان مختلفی همچون لی‌فارت و هارت [۱۱]، پرامانیک و همکاران [۱۲]، عدالت و همکاران [۱۳]، از انواع شبکه عصبی به منظور تعیین روابط بین نشانگرهای لرزه‌ای و پارامترهای پتروفیزیکی استفاده کردند.

منطق فازی نیز علاوه بر زمین‌آمار و شبکه‌های عصبی، در روش‌های جدیدتر برای شبیه‌سازی و تخمین پارامترهای پتروفیزیکی مخازن نفتی استفاده شده است. کدخدایی ایلخچی و همکاران [۱۴] با ساخت یک سیستم استنتاج فازی که از الگوریتم ژنتیک بهره می‌برد، باعث بهینه‌سازی مدل‌های ساخته‌شده از نشانگرهای لرزه‌ای توسط روش‌های فازی شدند. آن‌ها سیستم استنتاج فازی را بهترین الگوریتم و پس از آن شبکه عصبی احتمال‌پذیر را بهترین شبکه عصبی برای تخمین تخلخل از روی نشانگرهای لرزه‌ای معرفی کردند.

از دیگر روش‌های تلفیق داده‌های لرزه‌ای و چاه‌نگاری روش‌های طبقه‌بندی چند نشانگری می‌باشند. بسته به اینکه آیا نمونه‌های آموزشی در طبقه‌بندی استفاده می‌شود یا نه، روش‌های رده‌بندی لرزه‌ای موجود را می‌توان به روش نظارت‌شده و نظارت‌نشده تقسیم کرد. روش بدون نظارت می‌تواند بدون دانش قبلی داده‌ها را به دسته‌های مختلف تقسیم کند، درحالی‌که روش نظارت‌شده برای ساخت یک طبقه‌بند از نمونه‌های آموزشی استفاده می‌کند. نمونه‌های آموزشی از داده چاه و یا اطلاعات زمین‌شناسی معلوم، تعیین می‌شوند. روش تحت نظارت به دلیل استفاده از دانش قبلی پتانسیل این را دارند که پاسخ‌های با دقت بالاتر ارائه دهند. عموماً به روش طبقه‌بندی با نظارت "طبقه‌بندی" و به روش طبقه‌بندی بدون نظارت "خوشه‌بندی" می‌گویند.

امروزه تکنولوژی طبقه‌بندی و خوشه‌بندی چند نشانگری به صورت وسیعی در تحلیل رخساره‌های لرزه‌ای، تحلیل‌های سنگ‌شناسی، اکتشاف هیدروکربن و تخمین پارامترهای مخزن استفاده می‌شود. از

روش‌های خوشه‌بندی، نقشه خودسازمان‌دهی کوهنن^۱ و k-means به‌طور گسترده‌ای برای این کاربردها استفاده می‌شوند. در سال‌های اخیر نیز روش‌های خوشه‌بندی جدیدی معرفی شده‌اند که شامل: الگوریتم یادگیری رقابتی^۲ [۱۵] و ازدحام ذرات بهینه‌سازی^۳ [۱۶] می‌باشند. از روش‌های طبقه‌بندی k- نزدیک‌ترین هم‌سایه و BP دو روش پرکاربرد هستند. در سال‌های اخیر نیز روش‌های طبقه‌بندی جدیدی شامل ماشین بردار پشتیبان^۴ و شبکه عصبی تابع پایه شعاعی^۵ گسترش یافته‌اند [۱۷].

یکی از مشکلات بسیاری از الگوریتم‌های طبقه‌بندی مشکل بیش‌برازش است. از آنجایی که پیچیدگی الگوریتم طبقه‌بندی کننده قابل تنظیم است، ایجاد یک مدل بیش از حد پیچیده و حساس ممکن است منجر به طبقه‌بندی مناسب نمونه‌های آموزش شود، اما بعید است که منجر به ایجاد یک طبقه‌بندی خوب برای الگوهای جدید شود، این وضعیت به‌عنوان بیش‌برازش شناخته شده است. یکی از مهم‌ترین بخش‌های پژوهش در طبقه‌بندی الگوی آماری، تعیین نحوه پیچیدگی مدل است به‌طوری که نه چندان ساده باشند که نتواند تفاوت بین این طبقات مختلف را توضیح دهد. همچنین نه‌چندان پیچیده که در الگوهای جدید، طبقه‌بندی ضعیفی را ارائه دهد [۱۸]. این مشکل تا حد زیادی در روش آداپوست برطرف شده است.

الگوریتم بهبود^۶ برای اولین بار توسط شاپیر [۱۹] به کار گرفته شد. آداپوست (به معنای بهبود تطبیقی) یکی از معروف‌ترین الگوریتم‌های خانواده بهبود می‌باشد که در سال ۱۹۹۵ توسط فروند و شاپیر به‌عنوان یک روش کلی برای ایجاد یک طبقه‌بند قوی از مجموعه‌ای از طبقه‌بند ضعیف با دقت بالا ارائه شده است [۲۰].

^۱ Kohonen Self Organizing Map

^۲ Penalized competitive learning algorithm

^۳ Particle swarm optimization

^۴ Support Vector Machine

^۵ Radial basis function neural networks

^۶ boosting

روش طبقه‌بندی آدابوست که در سال‌های اخیر توجه بیشتری به آن شده است و امروزه جزو الگوریتم‌های برتر داده‌کاوی محسوب می‌شود، با اختصاص جریمه به چولگی^۱ (انحراف از توزیع نرمال به سمت چپ یا راست) توزیع داده‌ها در فرآیند آموزش، این مشکل را برطرف می‌کند.

ویژگی برجسته آدابوست تنظیم طبقه‌بندی‌کننده‌ها در مرحله یادگیری است. این الگوریتم که نسبتاً ساده است و سرعت بالایی دارد تا حد زیادی مشکل بیش‌برازش سایر روش‌ها را حل می‌کند و همچنین می‌تواند با روش‌های دیگر ترکیب شود تا یک الگوریتم طبقه‌بندی سریع را تشکیل دهد. در حال حاضر، الگوریتم آدابوست در سایر رشته‌ها در تشخیص چهره، سنجش از راه دور تصاویر، شناسایی بارکد، تشخیص تعداد اسکناس استفاده می‌شود. از این روش به‌منظور طبقه‌بندی سنگ‌شناسی با استفاده از داده‌های چاه‌نگاری استفاده شد. ژیونگ و همکارانش [۲۱] از این ابزار با هدف طبقه‌بندی در لرزه‌شناسی استفاده کرده‌اند.

۴-۱ روش انجام تحقیق

در این پایان‌نامه از روش آدابوست به‌منظور طبقه‌بندی چندشانگری در مطالعات لرزه‌نگاری بازتابی جهت طبقه‌بندی یک افق بر اساس اطلاعات پتروفیزیکی و شناسایی ساختارهایی نظیر گنبد نمکی استفاده می‌شوند. این پایان‌نامه در پنج فصل تهیه و تنظیم شده است. فصل اول شامل مقدمه، اهمیت و ضرورت مطالعه و اهداف موردنظر است. در فصل دوم به تعریف تخلخل، معرفی و طبقه‌بندی نشانگرهای لرزه‌ای مورد استفاده در این پایان‌نامه پرداخته می‌شود. در فصل سوم ابتدا روش‌های مورد استفاده در نرم‌افزار همپسون-راسل برای تخمین تخلخل با استفاده از نشانگرهای لرزه‌ای مورد بررسی قرار می‌گیرد. در بخش بعد مفاهیم طبقه‌بندی و خوشه‌بندی مورد بررسی قرار گرفته و چند روش معروف طبقه‌بندی با یکدیگر مقایسه می‌شوند. فصل چهارم شامل دو بخش است. بخش اول، شامل تخمین تخلخل با استفاده از نرم‌افزار همپسون-راسل، طبقه‌بندی افق لرزه‌ای بر مبنای تخلخل با

^۱ skewness

استفاده از روش آدابوست و مقایسه این دو با یکدیگر است. در بخش دوم از روش آدابوست برای تشخیص مرزهای جانبی و کف گنبد نمکی از محیط پیرامون استفاده شده است. در فصل پنجم نتایج نهایی این پایان نامه پیشنهادها و کارهایی که در آینده قابل طرح و انجام است، ارائه شده است.

فصل دوم

تخلخل و نشانگرهای لرزه‌ای

۱-۲ تخلخل

نسبت حجم فضای موجود یک نمونه سنگ به حجم کل نمونه سنگ را تخلخل می‌گویند. مقدار تخلخل اهمیت زیادی در اقتصادی بودن و قابلیت بهره‌برداری یک مخزن دارد. نماد یونانی ϕ بیانگر تخلخل است و از نظر فیزیکی، تخلخل نسبت حجم فضای خالی سنگ به حجم کل سنگ را شامل می‌شود [۲۲]:

$$\phi = \frac{V_p}{V_b} = 1 - \frac{V_{ma}}{V_b} \quad (1-2)$$

که در آن V_p ، V_{ma} ، V_b به ترتیب حجم کل سنگ، حجم ماتریکس سنگ و حجم فضای خالی سنگ مورد نظر می‌باشند. تعیین تخلخل یکی از پارامترهای مهم و اساسی در تعیین ویژگی‌های یک میدان نفتی است و از اهمیت خاصی برخوردار است. برای تعیین تخلخل مخزن عموماً از اندازه‌گیری مستقیم تخلخل (آزمایش بر روی مغزه) و یا نگارهای چاه‌نگاری (اندازه‌گیری غیرمستقیم تخلخل) استفاده می‌شود. در اندازه‌گیری تخلخل با استفاده از نگارهای چاه‌نگاری به منظور ایجاد ارتباط بین متغیرهای مختلفی که توسط نگارهای چاه مذکور اندازه‌گیری می‌شوند، از روابط تجربی و ریاضی خاصی استفاده و مقدار تخلخل سنگ مورد ارزیابی قرار می‌گیرد. به دلیل مشکلات مربوط به تعیین تخلخل از طریق مستقیم و انجام آزمایش‌های مغزه، استفاده از نگارهای پتروفیزیکی برای ارزیابی تخلخل مخازن نفتی از گذشته مورد توجه بوده و امروزه در بسیاری موارد این کار انجام می‌گیرد. به منظور برقراری ارتباط بین داده‌های مربوط به قرائت‌های درون‌چاهی (مثل زمان عبور صوت و چگالی) با مقدار تخلخل تا کنون چندین رابطه مختلف ارائه شده است.

دانستن بسیاری از مشخصه‌ها از قبیل لیتولوژی، محتوای رس، چگالی دانه‌ها و نوع سیال منفذی، لازمه تبدیل این پارامترها به یکدیگر می‌باشند که در حالت کلی ناشناخته هستند. در تعیین حجم ذخیره مخازن، تخلخل از مهم‌ترین عوامل است که در موارد طبیعی بازه تغییرات آن از کمتر از یک درصد در بعضی سنگ آهک‌های کریستالی تا ۸۰ درصد برای رسوبات نرم کف اقیانوس‌ها می‌باشد.

تخلخل در مخازن بین ۵ تا ۴۰ درصد و در سایر موارد بین ۱۰ تا ۲۰ درصد متغیر است. می‌توان تخلخل یک نمونه سنگ را از اندازه‌گیری چگالی حجمی آن تعیین کرد [۲۳]. یکی از معادلات اساسی برای ارتباط بین چگالی سنگ و تخلخل به صورت رابطه (۲-۲) است که در سال ۲۰۰۲ توسط المقبل ارائه گردیده است [۲۴]:

$$\rho_{rock} = \rho_{grain} (1 - \phi) + \rho_{fluid} \phi \quad (2-2)$$

که در آن، ρ_{rock} چگالی سنگ، ρ_{grain} چگالی ذرات تشکیل دهنده سنگ، ρ_{fluid} چگالی سیال موجود در خلل و فرج و ϕ تخلخل می‌باشد.

۲-۲ عوامل تعیین‌کننده میزان تخلخل

عوامل مؤثر بر میزان تخلخل به دو گروه اصلی عوامل اولیه و عوامل ثانویه تقسیم می‌شوند؛ که این عوامل تأثیر متفاوتی بر سنگ‌های مختلف دارند. در سنگ‌های تخریبی، عوامل اولیه نقش اصلی را ایفا می‌کنند؛ اما در سنگ‌های کربناته، عوامل ثانویه مؤثرتر هستند. از مهم‌ترین عوامل اولیه، می‌توان جور شدگی و شکل ذرات را نام برد. با افزایش جور شدگی میزان تخلخل کاهش می‌یابد. شکل ذرات، نحوه قرارگیری ذرات در مجاورت یکدیگر را تحت تأثیر قرار می‌دهد. گرد شدگی و کرویت پارامترهای کلیدی هستند که با افزایش آن‌ها تخلخل نیز افزایش می‌یابد.

عوامل ثانویه پس از مرحله رسوب‌گذاری و در طی فرآیندهای فشردگی و دیاژنز بر رسوبات تأثیر گذاشته تا جایی که سنگ با تخلخل بالا و یا سنگی متراکم حاصل شود. از عوامل ثانویه می‌توان به فشردگی، سیمان شدگی، شکستگی و انحلال اشاره کرد. فشردگی و سیمان شدگی باعث کاهش تخلخل و شکستگی و انحلال باعث افزایش تخلخل می‌گردند.

۳-۲ تقسیم‌بندی تخلخل

۱-۳-۲ تخلخل مطلق ϕ_t

نسبت کل حفره‌هایی که توسط اجسام جامد پر نشده (مثل روزنه‌ها، درزها، شکستگی‌ها، حفره‌ها) را بر کل حجم سنگ، تخلخل مطلق می‌گویند [۲۵].

$$\phi_t = \frac{V_t - V_s}{V_t} = \frac{V_p}{V_t} \quad (۳-۲)$$

بطوریکه:

V_p = حجم فضاهای خالی که معمولاً توسط سیالات (آب، نفت و گاز) پر می‌شود.

V_s = حجمی که توسط اجسام جامد اشغال شده است (قسمت جامد تشکیل دهنده سنگ).

V_t = حجم کل سنگ.

از یک طرف تخلخل اولیه^۲ (ϕ_1) و از نوع بین دانه‌ای^۳ و بین بلوری است و به مقدار زیادی به شکل و اندازه دانه و جور شدگی موادی که بخش جامد سنگ را می‌سازند، بستگی دارد و بیشتر در سنگ‌های تخریبی یافت می‌شود.

از طرف دیگر، تخلخل‌های حفره‌ای که در اثر حلال‌ها به وجود می‌آید و تخلخل ناشی از شکستگی که منشاء مکانیکی دارند، تشکیل تخلخل ثانویه (ϕ_2) را می‌دهند و بیشتر همراه با سنگ‌هایی که منشاء شیمیایی یا بیوشیمیایی دارند دیده می‌شود. تخلخل مطلق عبارت است از:

$$\phi_t = \phi_1 + \phi_2 \quad (۴-۲)$$

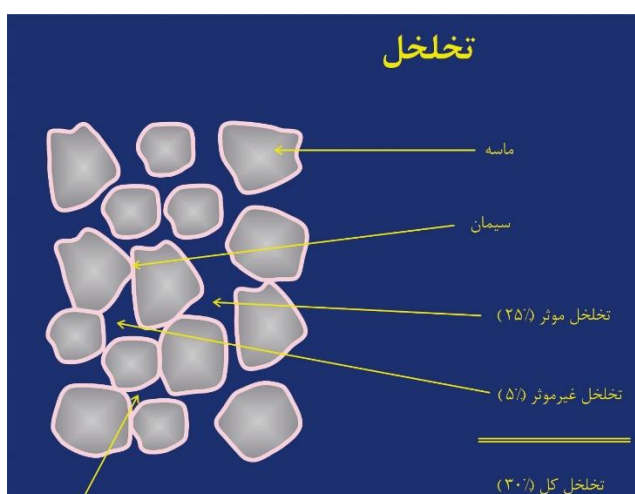
^۱ Total porosity

^۲ Primery porosity

^۳ Inter granular

۲-۳-۲ تخلخل مؤثر^۱ (ϕ_e)

به بخشی از تخلخل که قابلیت جانشینی توسط سیال را داشته باشد، تخلخل مفید اطلاق می‌گردد. این تخلخل شامل تخلخل‌های به هم ناپیوسته و تخلخل‌های اشغال شده توسط آب جذبی و آب محصور در شیل‌ها نشده و این تعریف صرفاً در بررسی نمودارها مورد استفاده قرار می‌گیرد. شکل ۲-۱، تخلخل کل، تخلخل مؤثر و تخلخل غیر مؤثر را نشان می‌دهد.



شکل ۲-۱. تخلخل‌های مختلف و کاربردی در کارهای نفتی.

۴-۲ نشانگرهای لرزه‌ای

هدف تحقیقات لرزه‌ای به نقشه درآوردن ویژگی‌های مرتبط با شواهد وجود، تولید، مهاجرت و به تله افتادن هیدروکربن یا به عبارت دیگر مشخص کردن ویژگی‌های ایستا^۲ و پویای^۳ مخازن زیرسطحی است. کاستینو^۴ [۲۶] این پارامترها را به صورت ساختار (عمق افق‌ها، ضخامت مخزن، گسل‌ها و ...)، ناهمسانی^۵، ویژگی‌های پتروفیزیکی (تخلخل، تراوایی، ...) و ویژگی‌های هیدروکربن (تولید، ترمودینامیک، ...) به لیست در آورد.

^۱ Effective porosity

^۲ static

^۳ dynamic

^۴ Cosentini

^۵ heterogeneity

پارامترهای چاه‌پیمایی مرسوم به‌صورت یک بعدی (قائم) نمونه‌برداری می‌شوند. بسیاری از ویژگی‌های بالا در تمام چاه‌ها اندازه‌گیری نمی‌شوند؛ اما نیاز است که تخمین زده شوند. یک نشانگر لرزه‌ای خوب یا مستقیماً به ویژگی زمین‌شناسی و مخزنی مورد نظر حساس است یا از طریق تعیین محیط ساختاری و محیط رسوبی، موجب استنتاج برخی از ویژگی‌های مورد نظر می‌شود. نقاط روشن^۱ (بازتاب قوی) مثال آشکاری از نشانگرها هستند که مستقیماً به پارامتر مورد نظر ارتباط دارند. استنتاج ساختار یا لایه‌بندی با اولین رکوردهای لرزه‌شناسی در دهه‌ی ۱۹۳۰ آغاز شد. اولین نشانگر به‌صورت ساده‌ای زمان سیر دوطرفه‌ی رخدادهای بازتابی را انتخاب می‌کرد. در شکل ۲-۲ تاریخچه نشانگرهای لرزه‌ای به تصویر در آمده است [۲۷].



^۱ Bright spot

شکل ۲-۲. توسعه نشانگرهای لرزه‌ای و ارتباط نشانگرها با پیشرفت‌های اساسی در اکتشافات لرزه‌ای [۲۸].

مفسر ماهر و متخصص پس از بررسی کردن داده‌ها با توجه به مجموعه تجربیاتی که در کارهای مختلف کسب کرده است و همچنین علم زمین‌شناسی خود، یک یا تعداد بیشتری فرضیه زمین‌شناسی مطرح می‌کند و به‌منظور یافتن الگوها آن را دستور کار خود قرار می‌دهد. عده زیادی تفسیر لرزه‌ای را ترکیبی از هنر و علم می‌دانند. زمانی که مفسر یک ویژگی یا الگوی لرزه‌ای را در محدوده چاه‌ها شناسایی می‌کند، می‌تواند سریعاً تعداد بیشتری از همان را در سایر قسمت‌ها پیدا کند. این نحوه شناسایی الگو توسط مفسرین با تجربه انجام می‌گیرد، برای ژئوفیزیک‌دانان جوان‌تر که اغلب دارای پایه ریاضیاتی هستند، گیج‌کننده است.

یکی از اهداف نشانگرهای لرزه‌ای این است که از طریق کمی کردن دامنه و ویژگی‌های ریخت‌شناسی دیده شده در داده‌های لرزه‌ای که به‌وسیله محاسبات کامپیوتری انجام می‌گیرد؛ به مفسران کم تجربه کمک کند و به نحوی این کمبود تجربه را برای آن‌ها جبران کند. برای مثال، نشانگر همبستگی که در اواسط دهه‌ی ۱۹۹۰ ایجاد شد، همان ناپیوستگی‌هایی را که توسط مفسران در حدود ۴۰ سال پیش در داده‌های لرزه‌ای مشاهده و به‌عنوان گسل تفسیر شده بود را نشان می‌داد.

روش‌های جدید زیادی برای استفاده چند نشانگری به‌صورت هم‌زمان ایجاد شدند. بارنز^۱ [۲۹] پیشنهاد داد برای انتخاب نشانگرهای چندگانه مناسب از نشانگرهایی که مستقل از یکدیگرند، استفاده شود. کالکومی^۲ [۳۰] هشدار داد که به‌منظور دوری از خطای همبستگی مثبت باید فقط از نشانگرهایی که با ویژگی‌های فیزیکی و ویژگی‌های مورد نظر در مخزن در ارتباط هستند، استفاده شود.

^۱ Barnes

^۲ Kalkomey

۲-۴-۱ طبقه‌بندی نشانگرها

همان‌طور که در سه دهه‌ی اخیر نشانگرهای لرزه‌ای در تعداد و تنوع رشد کردند، بسیاری از محققان برای طبقه‌بندی آن‌ها به خانواده‌هایی، با هدف نهایی فهم و کاربرد بهتر تلاش کردند.

تانر^۱ و همکاران [۳۱] نشانگرها را به دو دسته کلی هندسی و فیزیکی تقسیم کردند. هدف نشانگرهای هندسی، افزایش دید ویژگی‌های هندسی داده‌های لرزه‌ای شامل شیب، آزمون و پیوستگی می‌باشند. براون^۲ [۳۲] با استفاده از ساختار درختی نشانگرها را به چهار شاخه اصلی زمانی، دامنه، فرکانس و میرایی دسته‌بندی کرد که هر شاخه اصلی به دو زیرشاخه پیش از برانبارش و پس از برانبارش تقسیم می‌گردد. نشانگرهای زمانی اطلاعاتی درباره ساختارها به دست می‌آورند در حالی که نشانگرهای دامنه داده‌هایی از لایه‌بندی و مخزن تولید می‌کنند.

بارنز [۳۳] بر اساس ارتباط بین نشانگرهای مختلف با داده لرزه‌ای یک طبقه‌بندی از نشانگرهای ردلرزه مختلط ارائه داد. نشانگرهای دامنه و فاز به‌عنوان نشانگرهای پایه که تمام نشانگرهای دیگر از آن‌ها مشتق شده‌اند، تعیین شدند. نشانگرها به سه صورت دسته‌بندی شدند:

۱. به‌صورت یک بعدی، دو یا سه بعدی

۲. به‌صورت زمان یا عمق

۳. به‌صورت لحظه‌ای یا محلی

چنین طبقه‌بندی‌هایی برای ایجاد دید اولیه‌ای از نشانگرهای مختلف انجام می‌شود و این موضوع به کاربرد یا ترکیب نشانگرها در تعیین ویژگی‌های زیرسطحی کمک می‌کنند.

لینر^۳ و همکاران [۳۴] نشانگرها را به دو کلاس نشانگرهای عمومی و نشانگرهای خاص تقسیم کردند. نشانگرهای عمومی اندازه‌گیری‌های هندسی، کینماتیک، دینامیک یا ویژگی‌های آماری مشتق

^۱ Taner

^۲ Brown

^۳ Liner

شده از داده لرزه‌ای است. آن‌ها شامل دامنه بازتابنده، زمان بازتابنده، شیب و آزیموت بازتابنده، دامنه و فرکانس مختلط، نشانگرهای هیلبرت تعمیم یافته، AVO و تجزیه طیفی هستند. نشانگرهای عمومی همچنین بر پایه‌ی خصوصیت فیزیکی یا ریخت‌شناسی داده‌ها هستند که به سنگ‌شناسی یا زمین‌شناسی گره خورده است و بنابراین عموماً در سرتاسر جهان قابل کاربرد است. در مقابل، نشانگرهای خاص تعریف نامشخصی بر روی حوزه فیزیک یا زمین‌شناسی دارند. بسیاری از این نشانگرهای خاص مجموع، ضرب، یا ترکیبات دیگری از نشانگرهای پایه عمومی هستند. یک نشانگر خاص ممکن است در یک حوزه همبستگی خوبی با یک ویژگی زمین‌شناسی یا با تولید مخزن داده شده داشته باشد، اما این همبستگی‌ها عموماً برای یک حوزه دیگر جواب‌گو نیست.

مارفورت^۱ و چوپرا^۲ رده سوم را به طبقه‌بندی لینر و همکارانش به نام نشانگرهای ترکیبی (به‌وسیله ملداهل^۳ [۳۵] متا-نشانگر نامیده شد) اضافه کردند. این نشانگرها از تعدادی نشانگر لرزه‌ای و دید مفسر برای استفاده از آن‌ها در روش‌های هوشمند (زمین‌آمار، شبکه عصبی، خوشه‌بندی، طبقه‌بندی، رگرسیون و ...) به منظور کشف ویژگی‌های منحصر به فرد است. این نشانگرها موجب به دست آوردن اطلاعات جامع و یک پارچه‌تری از خصوصیات زمین‌شناسی، پتروفیزیکی و لرزه‌ای می‌گردند و برای تخمین رخساره‌ها، سنگ‌شناسی، احتمال حضور هیدروکربن، کشف گسل و کانال، دودکش‌های گازی و گنبد‌های نمکی مورد استفاده قرار می‌گیرند.

چن^۴ و سیدنی^۵ [۳۶] نمونه دیگری از طبقه‌بندی نشانگرهای لرزه‌ای بر مبنای خواص کینماتیک و دینامیک موج ارائه کردند. بر اساس این طبقه‌بندی نشانگرهای لرزه‌ای در ۸ طبقه قرار می‌گیرند که عبارتند از: دامنه، شکل موج، فرکانس، میرایی، فاز، همبستگی، انرژی و نسبت‌ها. در شکل ۲-۳ این نوع طبقه‌بندی به نمایش درآمده است.

^۱ Marfurt

^۲ Chopra

^۳ Meldahl

^۴ Chen

^۵ Sidney

دسته بندی نشانگرهای لرزه ای بر مبنای کینماتیک / دینامیک



شکل ۲-۳. طبقه بندی نشانگرهای لرزه ای بر مبنای خواص کینماتیک/ دینامیک موج با تغییر از (چن و سیدنی،

۱۹۹۷).

۲-۴-۲ محاسبه و بررسی نشانگرهای عمومی

۲-۴-۲-۱ فاز لحظه ای

از آنجایی که جنبه موج به عنوان خطوط با فاز ثابت تعریف می شوند و مستقل از دامنه می باشند، نشانگر فاز هم یک نشانگر فیزیکی است و می تواند به صورت مؤثری در تشخیص اشکال هندسی به کار گرفته شود.

ردلرزه $T(t)$ و تبدیل هیلبرت آن $H(t)$ از طریق رابطه های (۲-۵) و (۲-۶) به پوش ردلرزه $E(t)$ و فاز لحظه ای $\phi(t)$ مرتبط هستند [۳۷]:

$$T(t) = E(t) \cos(\phi(t)) \quad (۲-۵)$$

$$H(t) = E(t) \sin(\phi(t)) \quad (۲-۶)$$

برای نشان دادن فاز لحظه ای از مقیاس درجه $(-\pi, \pi)$ استفاده می شود. (تانر، ۲۰۰۰).

از کاربردهای این نشانگر می‌توان به موارد زیر اشاره نمود:

- بهترین مشخص کننده پیوستگی جانبی.
- تجسم با جزئیات از وضعیت لایه‌بندی.
- وابسته به اجزاء تشکیل دهنده فاز انتشار موج.
- نمایش همه رویدادها، به دلیل نداشتن اطلاعات دامنه.
- استفاده در محاسبه فرکانس و شتاب لحظه‌ای.

۲-۲-۴-۲ فرکانس لحظه‌ای

فرکانس لحظه‌ای، مشتق زمانی فاز لحظه‌ای (نرخ تغییرات فاز) است با که به صورت رابطه (۷-۲) تعریف می‌شود [۳۳]:

$$F(t) = d(\phi(t)) / dt \quad (7-2)$$

بعد پاسخ برحسب دور بر ثانیه می‌باشد. فرکانس لحظه‌ای به انتشار موج و نهشته‌های محیطی وابسته می‌باشد بنابراین از نوع نشانگرهای فیزیکی بوده و در تمیز دادن موارد زیر مؤثر است:

- آشکارسازی لبه‌های لایه‌های نازک با مقاومت لرزه‌ای ظاهری کم.
- فرکانس لحظه‌ای می‌تواند نشانگر ضخامت لایه و همچنین پارامترهای سنگ‌شناسی باشد.
- شاخص نواحی دارای شکستگی. این نواحی با فرکانس پایین ظاهر می‌شوند.
- شاخص هیدروکربن با استفاده از ناهنجاری فرکانس پایین. این اثر گاهی به وسیله نفت موجود در خلل و فرج ماسه تحکیم نیافته، برجسته می‌شود.
- شاخص نواحی با بازتاب‌های آشفته.
- شاخص نسبت ماسه به شیل.

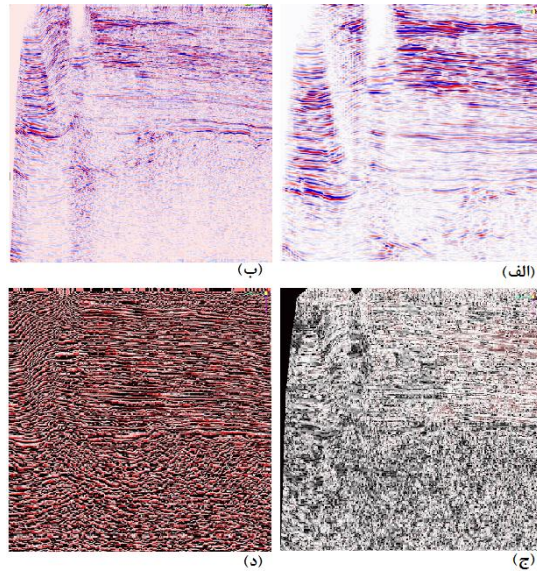
۲-۴-۲-۳ مقاومت ظاهری لرزه‌ای نسبی

مقاومت ظاهری لرزه‌ای، حاصل ضرب سرعت موج در چگالی محیط می‌باشد. به عبارتی مقاومت لرزه‌ای، مقاومت سنگ‌ها در برابر انتشار امواج الاستیک می‌باشد. تباین مقاومت ظاهری لرزه‌ای در فصل مشترک لایه‌ها باعث بازتاب امواج لرزه‌ای می‌گردد.

سنگ‌های مختلف، مقادیر مقاومت ظاهری لرزه‌ای متفاوتی دارند. البته این مقادیر برای سنگ‌های مختلف همپوشانی دارند، به این معنی که دو سنگ متفاوت، ممکن است یک مقدار مقاومت ظاهری لرزه‌ای داشته باشند. این تشابه می‌تواند علل مختلفی داشته باشد که عبارت از عمق قرارگیری، فشردگی تکتونیکی، تخلخل بین دانه‌ای، تخلخل ناشی از شکستگی، نوع شکستگی، نوع سیمان و میزان اشباع خلل و فرج سنگ می‌باشند. در نتیجه از مشاهدات تغییرات مقاومت ظاهری در داخل یک لایه با ویژگی‌های سنگ‌شناسی مشخص می‌توان به تغییرات رخساره سنگی در داخل لایه پی برد. مقاومت ظاهری لرزه‌ای را می‌توان به سایر مشخصات توده سنگ نظیر تخلخل، میزان اشباع، شکستگی و ... ربط داد؛ بنابراین تغییرات مقاومت ظاهری لرزه‌ای در محل مورد مطالعه، تغییرات این پارامترها را نیز می‌تواند مشخص سازد [۱].

- با تخلخل ارتباط دارد.
- تباین زیاد این نشانگر، شاخصی از لایه‌های متوالی موجود است.
- ناپیوستگی‌ها را آشکار می‌سازند.
- شاخصی از دگرشیبی سطوح است.

در شکل ۲-۴ نشانگرهای دامنه لرزه‌ای، امپدانس صوتی، فرکانس لحظه‌ای و فاز لحظه‌ای برای یک داده لرزه‌ای (شکل ۲-۴-الف) نشان داده شده است.



شکل ۲-۴. چهار مقطع از نشانگرهای مختلف: الف) دامنه لرزه‌ای. ب) امپدانس صوتی ج) فرکانس لحظه‌ای د)

فاز لحظه‌ای [۳۷].

۲-۴-۲-۴ فاکتور کیفیت لحظه‌ای

فاکتور کیفیت، نسبت انرژی ساطع شده بین دو قله با فاصله زمانی 2π به کل انرژی ساطع شده در یک دوره می‌باشد. در لرزه شناسی فاکتور کیفیت (Q) برای سنگ‌ها بین ۵۰ تا ۳۰۰ می‌باشد. این عامل متناسب با عوامل جذب می‌باشد و به صورت رابطه (۲-۸) نشان داده می‌شود [۳۴]:

$$\frac{1}{Q} = \frac{\alpha v}{\pi f} = \frac{\alpha \lambda}{\pi} \quad (۲-۸)$$

که در آن، v سرعت، f فرکانس، λ طول موج و α نماد جذب می‌باشد که نشان دهنده کاهش نمایی دامنه با فاصله به علت پدیده جذب می‌باشد. این نشانگر از نوع نشانگرهای عبوری می‌باشد و همچنین یک نشانگر فیزیکی نیز می‌باشد.

از کاربردهای این نشانگر می‌توان به موارد زیر اشاره نمود:

- شاخص ویژگی‌های جذب نسبی لایه‌ها.
- تعیین حجم مایع، به وسیله تعیین نسبت فشار به فاکتور کیفیت لحظه‌ای مقطع موج برشی.
- ارتباط شدید با تخلخل، نفوذپذیری و شکستگی.

۲-۴-۲-۵ پوش رد لرزه

رد لرزه مختلط را می‌توان به صورت رابطه (۲-۹) تعریف نمود [۲۹]:

$$F(t) = T(t) + H(t) \quad (۲-۹)$$

که $T(t)$ مقدار حقیقی رد لرزه مختلط یا همان رد لرزه یا داده لرزه‌ای ثبت شده و $g(t)$ قسمت موهومی رد لرزه مختلط است که تبدیل هیلبرت تابع $T(t)$ می‌باشد. پوش رد لرزه را می‌توان به صورت رابطه (۲-۱۰) تعریف نمود:

$$E(t) = \sqrt{T(t)^2 + H(t)^2} \quad (۲-۱۰)$$

در واقع $E(t)$ میزان انرژی لحظه‌ای کل را نشان می‌دهد و بزرگی آن با بزرگی رد لرزه ورودی برابر است. مقدار آن تقریباً بین صفر و بیشترین مقدار دامنه رد لرزه تغییر می‌کند. همان‌طور که رابطه ۲-۱۰ نشان می‌دهد، پوش رد لرزه مستقل از فاز بوده و به صورت مستقیم به اختلاف مقاومت ظاهری لرزه‌ای بستگی دارد. این نشانگر ممکن است، نشان دهنده یک بازتابنده و یا نشان دهنده پاسخ ترکیب شده چندین بازتابنده باشد که این موضوع بستگی به پهنای باند لرزه‌ای دارد. پوش رد لرزه یک نشانگر فیزیکی می‌باشد و می‌تواند به عنوان یک عامل تمیز دهنده مؤثر در مشخصه‌های زیر استفاده گردد:

- نمایان کردن اختلاف مقاومت ظاهری لرزه‌ای.
- تعیین محل لکه‌های روشن.
- تعیین محل تجمع گاز.
- تشخیص توالی لایه‌ها.
- تعیین اثر کاهش ضخامت لایه‌ها.
- مشخص کردن ناپیوستگی.
- تغییرات اساسی سنگ‌شناسی.
- تغییرات اساسی در نهشته‌های محیطی.

- تغییرات جانبی نشان دهنده گسلش.
- همبستگی فضایی یا تخلخل و سایر تغییرات سنگ شناسی.

در شکل شماره ۲-۵-الف این نشانگر مشاهده می شود.

۲-۴-۲-۶ مشتق اول پوش^۱ ردلرزه

مشتق اول پوش ردلرزه (نرخ زمانی تغییرات پوش) تغییرات انرژی رویدادهای بازتابیده را نشان می دهد. دامنه رویدادهایی که با سرعت نسبی زیادی و به صورت ناگهانی افزایش می یابد، نشان دهنده یک پهنای باند گسترده تر و بنابراین اثر جذب کمتر هستند. برای اجرای این مشتق گیری از یک فیلتر در حوزه زمان استفاده می شود [۲۹].

$$RE(t) = dE(t) / dt = E(t) * diff(t) \quad (۱۱-۲)$$

که RE مشتق اول پوش ردلرزه، * نشان دهنده عملگر هممیخت و $diff(t)$ عملگر مشتق گیری می باشد.

به علاوه این نشانگر، یک نشانگر فیزیکی است و می تواند برای شناسایی شکستگی و اثرات جذب مورد استفاده قرار گیرد. از این نشانگر همچنین برای محاسبه راستای انتشار گروه استفاده می شود. زمانی که با راستای انتشار فاز مقایسه می گردد، می تواند نشان دهنده امواج پاششی باشد.

از کاربردهای این نشانگر می توان به موارد زیر اشاره نمود:

- بارز کردن تغییرات در بازتاب ها که همچنین به اثرات جذب مربوط است. افزایش جذب باعث کاهش تندی و تیزی پوش ردلرزه می شود.
- تغییرات جانبی در این نشانگر نشان دهنده ناپیوستگی ها می باشد.
- شناسایی شکستگی ها.

^۱ First Derivation of the Envelope

- کمک به شناسایی امواج پاششی.

در شکل ۲-۵-ج) این نشانگر برای ردلرزه موجود در شکل ۲-۵-الف) نشان داده شده است.

۲-۴-۲-۷ مشتق دوم پوش ردلرزه

مشتق دوم پوش ردلرزه، اندازه‌ای از تیزی قله پوش می‌باشد که ممکن است به عنوان نمایش یک نشانگر اصلی مفیدتر باشد. این نشانگر به صورت رابطه (۲-۱۲) محاسبه می‌شود:

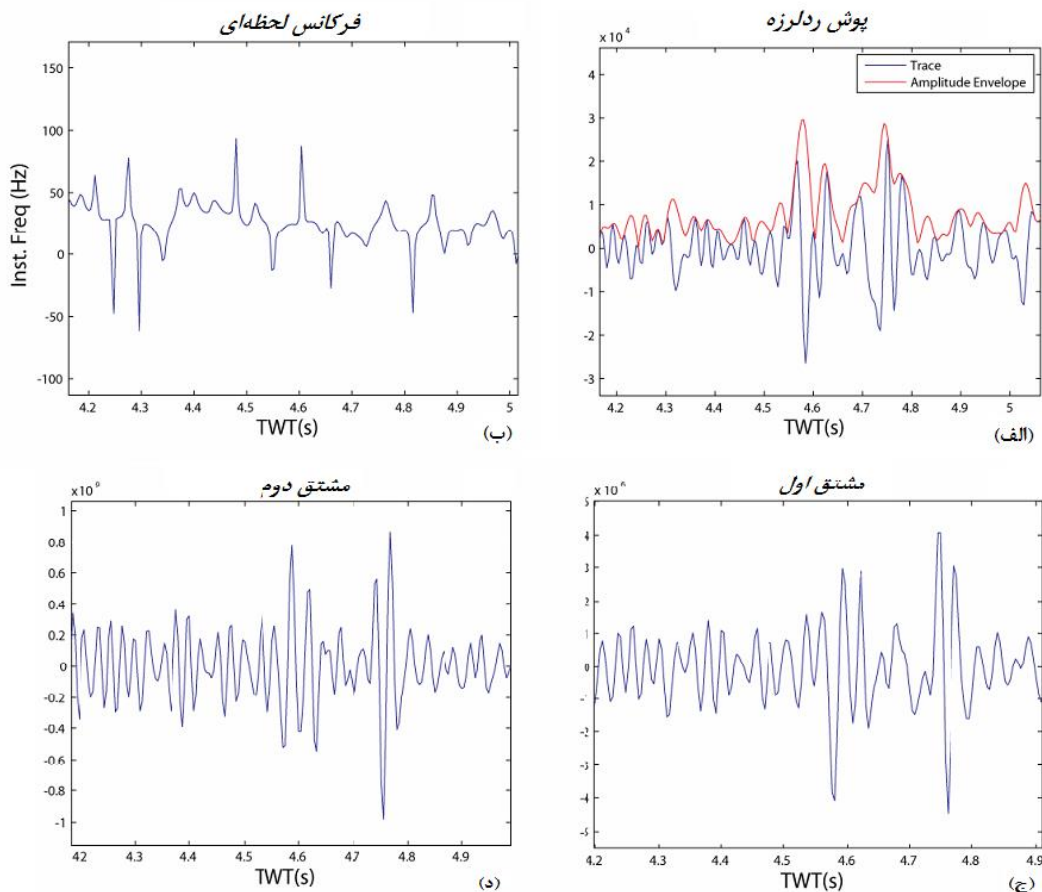
$$DDE(t) = d^2 E(t) / dt^2 \quad (۲-۱۲)$$

از کاربردهای این نشانگرها می‌توان به موارد زیر اشاره کرد.

- ارائه تصویر زیرسطحی خیلی خوب در محدوده پهنای باند لرزه‌ای.
- نشان دهنده فصل مشترک بازتابنده که در پهنای باند لرزه‌ای دیده می‌شوند.
- نمایش تغییرات ناگهانی و تند سنگ‌شناسی.
- نشان دهنده تیزی رویدادها.

در شکل ۲-۵-د) نشانگر مشتق دوم پوش ردلرزه محاسبه شده برای ردلرزه آبی رنگ شکل

۲-۵-الف) نشان داده شده است.



شکل ۲-۵. چهار نوع نشانگر لرزه‌ای: الف) پوش ردلرزه، ب) فرکانس لحظه‌ای، ج) مشتق اول پوش ردلرزه، د) مشتق دوم پوش ردلرزه [۳۸].

این نشانگرها و بسیاری دیگر از نشانگرهای معمول، اندازه‌گیری‌های آماری مرتبه اول از دامنه هستند. چنین نشانگرهایی برای شناسایی دقیق و با جزئیات مشخصه‌های ساختاری، چینه‌شناسی و مخزنی کارآمد نیستند. چراکه ممکن است دو پاسخ لرزه‌ای مربوط به رخساره‌های لرزه‌ای متفاوت، دامنه میانگین یا فرکانس غالب مشابه داشته باشند؛ بنابراین برای مشخص کردن چنین رخساره‌هایی به نشانگرهای حساس‌تر که از آمار مرتبه دوم استفاده می‌کنند، احتیاج است [۳۹].

نشانگرهای لرزه‌ای رایج، از روابط ریاضی بر اساس خواص فیزیکی و هندسی زیرسطحی برای آشکار سازی ویژگی‌های مورد علاقه استفاده می‌کنند. نشانگرهای بافتی از قبیل نشانگرهای مبتنی بر

ماتریس هم رخداد سطح خاکستری^۱، می‌توانند وابستگی فضایی رخساره‌های لرزه‌ای را توصیف کنند. از آنجا که ویژگی‌های زمین‌شناسی عموماً تخت نیستند، ضروری است که شیب ساختاری در محاسبات نشانگرها وارد شود. نشانگرهای بافتی لرزه‌ای، اندازه‌گیری‌های آماری از دامنه‌های استخراج‌شده در راستای شیب ساختاری هستند [۴۰].

۲-۴-۳ نشانگرهای بافتی

بافت یکی از ویژگی‌های اصلی است که در تصاویر برای شناسایی اشیاء و مناطق مورد نظر استفاده می‌شود. در اصطلاح روزمره، بافت به صورت مفاهیمی مانند زبری، ابریشمی، ناهموار و غیره تعریف می‌شود. بافت زبر، حس تفاوت ارتفاعی شدیدی را به انگشتان انتقال می‌دهد؛ در صورتی که بافت ابریشمی ناشی از اختلاف کم ارتفاعی می‌باشد. این تعاریف به شیوه مشابهی به علم لرزه‌شناسی وارد شده است. در لرزه‌شناسی تغییرات دامنه نماینده تغییرات ارتفاع است و پنجره تحلیل مستطیل و یا بیضی شکل در راستای ساختار جایگزین انگشتان کاوشگر می‌شود.

در تصاویر غیر رنگی برای بیان شدت روشنایی از اصطلاح سطح خاکستری استفاده می‌شود. در بررسی ویژگی‌های بافتی وابستگی فضایی تغییرات سطوح روشنایی مشخص می‌شود.

نشانگرهای بافتی در سه گروه عمده تقسیم‌بندی می‌شوند [۴۱].

۱. نشانگرهای بر مبنای ماتریس GLCM

۲. نشانگرهای بر مبنای فرکانس

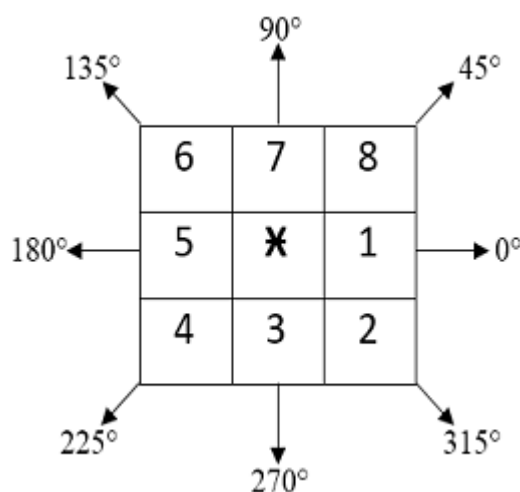
۳. نشانگرهای بر مبنای شیب و شباهت

به دلیل اینکه در این پایان‌نامه به شناسایی مرز گنبد‌های نمکی با استفاده از نشانگرهای بافتی مبتنی بر GLCM پرداخته می‌شود، در این قسمت فقط به توضیحات مربوط به این دسته از نشانگرهای بافتی پرداخته می‌شود.

^۱. Gray level co-occurrence matrix (GLCM)

۱-۳-۴-۲ نشانگر لوزه‌ای بافتی مبتنی بر GLCM

در یک تصویر I ، هر پیکسل می‌تواند دارای ۸ همسایگی در راستاهای مختلف باشد. این موضوع در شکل ۶-۲ نشان داده شده است.



شکل ۶-۲: همسایگی‌های مختلف از عنصر مورد نظر (با تغییر از [۴۲]).

در تصویر I ، می‌توان رابطه فضایی سطح‌های خاکستری را به وسیله ماتریس فراوانی نسبی رخ دادن عناصر همسایه، یکی با سطح خاکستری i و دیگری با سطح خاکستری j که در فاصله d از یکدیگر قرار دارند $G(i, j)$ ، بیان کرد.

اگر تصویر I دارای ابعاد $k_1 \times k_2$ و n سطح رو شنایی (سطح خاکستری) باشد، ماتریس GLCM ماتریسی مربعی $G_{n \times n}$ است که هر عنصر آن نشان دهنده تعداد رخداد جفت سطوح خاکستری در پیکسل‌های تصویر در جهت $\theta = \{0, 45, 90, 135, 180, 225, 270, 315\}$ و با فاصله افقی d_x و فاصله قائم d_y از یکدیگر قرار گرفته‌اند و به صورت رابطه (۱۲-۲) تعریف می‌شود:

(۱۲-۲)

$$G(i, j) = \sum_{x=1}^{K_1} \sum_{y=1}^{K_2} \begin{cases} 1, & I(x, y) = i \text{ AND } I(x + dx, y + dy) = j \\ 0, & I(x, y) \neq i \text{ OR } I(x + dx, y + dy) \neq j \end{cases}, i, j = 1, 2, \dots, n$$

با توجه به اینکه ماتریس‌های GLCM، متقارن هستند، بین ماتریس‌های GLCM در جهت‌های

مختلف رابطه‌های (۲-۱۳) تا (۲-۱۶) برقرار خواهند بود [۴۳].

اگر $G^T(d, \theta^\circ)$ ترانهاده ماتریس $G(d, \theta^\circ)$ باشد:

$$G(d, 0^\circ) = G^T(d, 180^\circ) \quad (۲-۱۳)$$

$$G(d, 45^\circ) = G^T(d, 225^\circ) \quad (۲-۱۴)$$

$$G(d, 90^\circ) = G^T(d, 270^\circ) \quad (۲-۱۵)$$

$$G(d, 135^\circ) = G^T(d, 315^\circ) \quad (۲-۱۶)$$

در نتیجه اطلاعاتی که ماتریس‌های $G(d, 180^\circ)$ ، $G(d, 225^\circ)$ ، $G(d, 270^\circ)$ و $G(d, 315^\circ)$ در مورد بافت ایجاد می‌کنند تکراری است و این راستاها در محاسبات مورد استفاده قرار نمی‌گیرند.

در شکل ۲-۷، نحوه محاسبه ماتریس GLCM در راستاهای مختلف و دورافت ۱ به صورت شماتیک نشان داده شده است. برای طبقه‌بندی تصویر با استفاده از بافت می‌توان تعداد ۱۴ خاصیت آماری را از ماتریس GLCM استخراج و استفاده نمود.

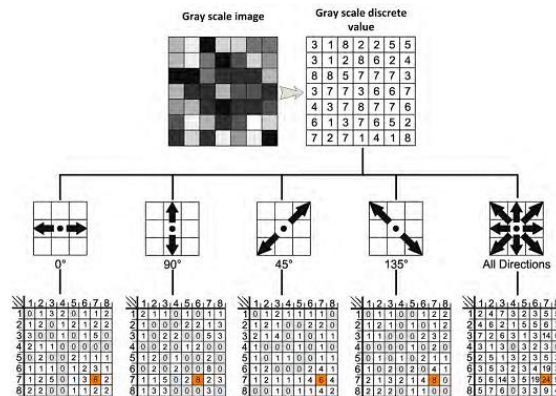
برای محاسبه نشانگر موردنظر در داده لرزه‌ای می‌توان یک پنجره محلی متحرک در راستای شیب بر روی داده انتخاب کرد، خصوصیات آماری را به صورت محلی محاسبه و به مرکز پنجره نسبت داد. ابعاد پنجره بایستی به گونه‌ای باشد که به تغییرات ساختاری هدف حساس باشد و نباید کمتر از طول موجک باشد. هر کدام از ۱۴ خاصیت آماری که از ماتریس GLCM قابل استخراج می‌باشد، می‌تواند به عنوان یک نشانگر لرزه‌ای در نظر گرفته شود.

به منظور محاسبه نشانگرهای بافتی، GLCM باید با استفاده از رابطه (۲-۱۷) به هنجار شود.

$$p(i, j) = \frac{G(i, j)}{R} \quad (۲-۱۷)$$

که $G(i, j)$ ماتریس GLCM، $p(i, j)$ به هنجار شده ماتریس GLCM و R مجموع تمام مقادیر ماتریس

GLCM می‌باشد. در محاسبه نشانگرهای بافتی، پارامترهای متعددی که بر پایه‌ی $p(i, j)$ محاسبه شده‌اند، مورد استفاده قرار می‌گیرند. این روابط در جدول ۱-۲ آورده شده‌اند. همچنین ۱۴ خاصیت آماری به همراه روابطی که برای محاسبه آنها مورد نیاز است، در قالب جدول ۲-۲ ارائه شده است.



شکل ۲-۷. نمایش شماتیک محاسبه ماتریس GLCM در راستاهای مختلف با گام یک.

جدول ۱-۲. پارامترهای مورد استفاده در محاسبه نشانگرهای GLCM.

$p_x(i) = \sum_{j=1}^{N_g} p(i, j)$	(۱۸-۲)
$p_y(j) = \sum_{i=1}^{N_g} p(i, j)$	(۱۹-۲)
$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j), \quad i + j = k, \quad k = 2, 3, \dots, 2N_g$	(۲۰-۲)
$p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j), \quad i - j = k, \quad k = 0, 1, \dots, N_g - 1$	(۲۱-۲)
$\mu_x = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_g} p(i, j) = \sum_{i=1}^{N_g} i p_x(i, j)$	(۲۲-۲)
$\mu_y = \sum_{j=1}^{N_g} j \sum_{i=1}^{N_g} p(i, j) = \sum_{j=1}^{N_g} j p_y(i, j)$	(۲۳-۲)

$\sigma_x^2 = \sum_{i=1}^{N_g} (i - \mu_x)^2 \sum_{j=1}^{N_g} p(i, j) = \sum_{j=1}^{N_g} (p_x(i) - \mu_x(i))^2$	(۲۴-۲)
$\sigma_y^2 = \sum_{j=1}^{N_g} (j - \mu_y)^2 \sum_{i=1}^{N_g} p(i, j) = \sum_{j=1}^{N_g} (p_y(j) - \mu_y(j))^2$	(۲۵-۲)

جدول ۲-۲. خواص آماری قابل استخراج از ماتریس GLCM و روابط حاکم بر آنها.

نشانهگر	رابطه
نشانهگر BGD	$Big\ Grads\ Dominance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} j^2 p(i, j)$
نشانهگر اینرسی	$Inertia = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i, j)$
نشانهگر شدت	$Intensity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij p(i, j)$
نشانهگر میانگین	$Mean = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ip(i, j) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} jp(i, j)$
نشانهگر SGD	$Small\ Grads\ Dominance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i, j)}{j^2}$
نشانهگر برتری خوشه	$Cluster\ Prominence = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^4 p(i, j)$
نشانهگر همبستگی	$Correlation = \left(\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (ij) p(i, j) - \mu_x \mu_y \right) / \sigma_x \sigma_y$

نشانگر عدم شباهت	$Dissimilarity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) i-j $
نشانگر انرژی	$Energy = \left(\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (p(i,j))^2 \right)^{1/2}$
نشانگر آنتروپی	$Entropy = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \log(p(i,j))$
نشانگر همگنی	$Homogeneity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} p(i,j)$
نشانگر سایه خوشه	$Cluster Shade = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i+j - \mu_x - \mu_y)^3 p(i,j)$
نشانگر شباهت	$Similarity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) i+j $
نشانگر رد	$Trace = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) , i = j$
نشانگر واریانس	$Variance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 p(i,j)$

فصل سوم

روش‌های ترکیب نشانگرها

به دلیل ضعف نشانگرها در مشخص کردن روابط میان اطلاعات لرزه‌ای با اطلاعات پتروفیزیکی و زمین‌شناسی تاکنون تعداد زیادی نشانگر لرزه‌ای معرفی شده است. تحلیل چند نشانگری این امکان را برای ژئوفیزیک‌دانان فراهم می‌کند تا به طور هم‌زمان بر روی چند نشانگر کار کنند و با ترکیب چندین نوع اطلاعات متفاوت اما حیاتی با یکدیگر، بتوانند از طریق داده‌های لرزه‌ای راحت‌تر به اطلاعات مورد نیاز خود دست یابند. تا کنون روش‌های متفاوتی از قبیل رگرسیون (خطی و غیرخطی)، شبکه عصبی، طبقه‌بندی و خوشه‌بندی برای ترکیب نشانگرهای مختلف ارائه شده است. در ادامه تعدادی از این روش‌ها شرح داده خواهد شد [۱].

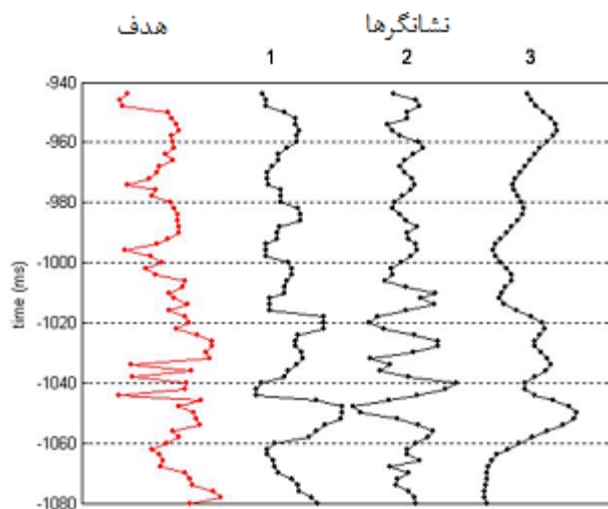
۳-۲ رگرسیون خطی چند نشانگری

در روش رگرسیون خطی چند نشانگری هدف یافتن عملگری خطی است که بتواند اطلاعات نگارهای چاه پیمایی را با استفاده از نزدیک‌ترین داده‌های لرزه‌ای تخمین بزند. به دلیل اینکه بسیاری از ویژگی‌های لرزه‌ای در درون ردلرزه خام نهفته است، نمی‌توان تنها با داده‌های لرزه‌ای خام تمام خصوصیات مخزنی را دید. بنابراین با تجزیه داده‌های ورودی به اجزای تشکیل دهنده آن (نشانگرهای لرزه‌ای) و استفاده از آن‌ها به عنوان ورودی روش رگرسیون خطی چند نشانگری قدرت پیش‌بینی روش افزایش می‌یابد.

۳-۲-۱ انجام تحلیل یک نشانگری

به‌طور ساده هرگاه یک نشانگر لرزه‌ای به صورت یکنواخت به ازای هر نمونه از داده‌های لرزه‌ای مقدار داشته باشد، با برقراری یک ارتباط خطی یا غیرخطی بین مقادیر آن و مقادیری از نگارهای چاه (تخلخل) در محل چاه و سپس تعمیم رابطه تبدیل در طول مقطع لرزه‌ای می‌توان مقادیر نشانگر لرزه‌ای را به مقادیر تخلخل تبدیل کرد [۴۴].

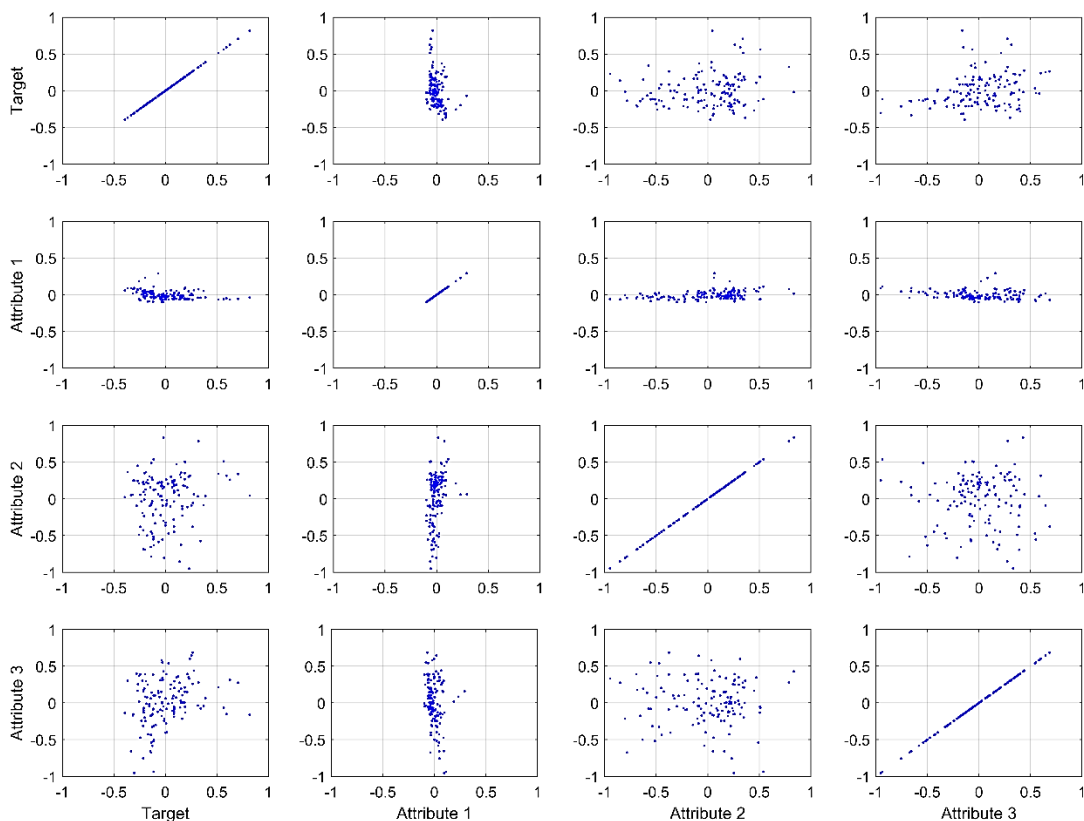
ساده‌ترین روش برای برقراری ارتباط بین داده‌های لرزه‌ای هدف و نشانگرهای لرزه‌ای رسم منحنی تقاطع میان آن‌ها می‌باشد. منحنی تقاطعی نشان دهنده همبستگی خطی میان دو متغیر است. برای درک بهتر مفهوم منحنی تقاطعی بهتر است این موضوع توسط یک مثال ساده توضیح داده شود. در شکل ۱-۳ چهار نمودار رسم شده است.



شکل ۱-۳. نگار هدف در سمت چپ و سه نشانگر لرزه‌ای در سمت راست قرار دارند [۴۵].

نمودار سمت چپ نمودار هدف (نگار صوتی) و سه نمودار سمت راست نشانگرهای لرزه‌ای می‌باشند. رابطه‌ای یک به یک میان نمونه‌های روی نگار صوتی و نمونه‌های روی نشانگرهای لرزه‌ای وجود دارد. قرار است همبستگی بین تک تک این متغیرها (نشانگرها و هدف) با یکدیگر بررسی شود. در اینجا از یک نوع رسم به نام پراکندگی p بعدی استفاده می‌شود [۴۵] که در شکل ۲-۳ قابل مشاهده است.

سه نوع منحنی متقاطع در شکل ۲-۳ مشاهده می‌شود. نوع اول از منحنی متقاطع بر روی قطر اصلی شکل مشاهده می‌شود که خود همبستگی هدف با خودش و هر نشانگر منحصر به فرد با خودش را نمایش می‌دهد. این بهترین همبستگی خطی است و سایر همبستگی‌ها کیفیتی پایین‌تر از این خواهند داشت. این نوع همبستگی متقابل همواره خطی مستقیم را ایجاد خواهد نمود و توزیع پراکندگی نقاط همواره بر روی این خط خواهد بود و نمونه‌ها به صورت پیوسته و یا خوشه‌هایی مجزا دیده خواهند شد.



شکل ۳-۲. یک ماتریس ۴*۴ از منحنی‌های مقاطع که در آن نگار هدف و تمام نشانگرهای نمایش داده شده در شکل شماره ۱ در مقابل هم رسم شده‌اند.

نوع دوم رسم در ردیف بالایی و همچنین ستون سمت چپ ماتریس مشاهده می‌شود. این‌ها منحنی مقاطع نمودار هدف با هر یک از نشانگرها است. این نوع از ترسیم‌ها بعداً در رگرسیون خطی مورد استفاده قرار خواهند گرفت.

این شکل نشان دهنده وجود همبستگی مثبت میان هدف و نشانگر شماره یک می‌باشد که در آن با افزایش مقادیر نشانگر، مقادیر هدف نیز افزایش می‌یابد؛ اما همچنین مشاهده می‌شود که تعداد زیادی نقاط پراکنده نیز وجود دارد که از خط بهینه (خطی با زاویه ۴۵ درجه) دور هستند.

سومین نوع منحنی مقاطع، رسم نشانگرهای لرزه‌ای در مقابل یکدیگر است و در روش‌های طبقه‌بندی کاربرد دارد. به منظور استفاده از نشانگرهای لرزه‌ای برای تخمین پارامترهای مخزنی ابتدا

باید روابط میان آن‌ها تعیین گردد و نشانگرهایی که با هدف همبستگی بیشتری دارند انتخاب شوند. یکی از راه‌های تعیین ارتباط بین دو متغیر استفاده از منحنی تقاطع است.

۳-۲-۲ استفاده از منحنی تقاطعی در چند نشانگرها

هنگامی که نشانگرها با هدف مورد نظر همبستگی خطی قابل قبولی نداشته باشند، استفاده از آن‌ها نتیجه مطلوبی ایجاد نمی‌کند. حال هدف استفاده از روش رگرسیون خطی برای افزایش کیفیت برازش می‌باشد. اگر L نگار هدف و M تعداد نشانگر باشد و هرکدام دارای N نمونه باشند، می‌توان رابطه پایه‌ای برای پیش‌بینی خطی را به صورت رابطه (۱-۳) نوشت [۴۴]:

$$L = Aw \quad (1-3)$$

$$\text{که در آن } w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_M \end{bmatrix} \text{ است و } A = \begin{bmatrix} 1 & a_{11} & \cdots & a_{1M} \\ 1 & a_{21} & \cdots & a_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{N1} & \cdots & a_{NM} \end{bmatrix}$$

اختصاص داده شده به هر نشانگر A_i می‌باشند.

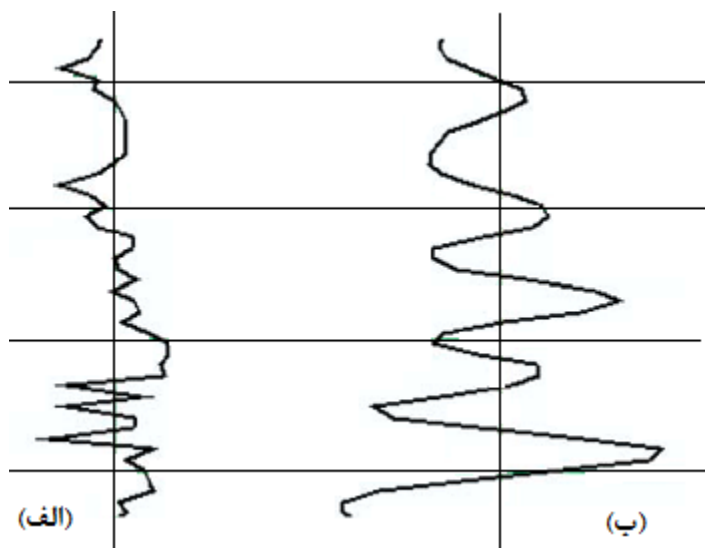
ضرایب وزنی در رابطه (۱-۳) می‌توانند از طریق کمینه کردن میانگین مربعات خطا استخراج شوند. با بسط دادن معادله میانگین مربعات خطا به صورت رابطه (۲-۳) به دست می‌آید.

$$E^2 = \frac{1}{N} \sum_{i=1}^N (L_i - w_0 - w_1 A_{1i} - w_2 A_{2i} - \cdots - w_M A_{Mi})^2 \quad (2-3)$$

۳-۲-۳ رگرسیون چند نشانگری با ضرایب وزنی هم‌امیختی

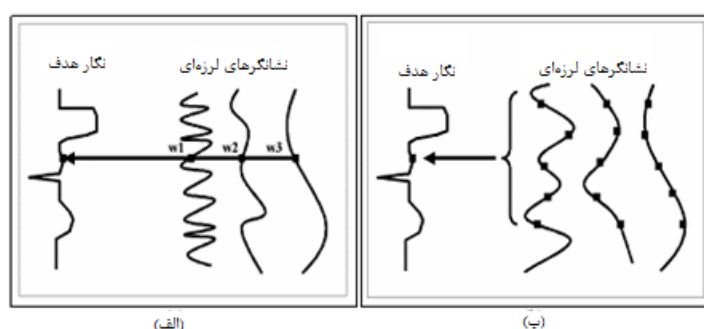
زمین یک فیلتر پایین‌گذر است و در حین عبور موج از لایه‌های مختلف، یک سری از فرکانس‌ها حذف می‌شوند؛ بنابراین محتوای فرکانسی نگار هدف و نشانگرهای لرزه‌ای با یکدیگر برابر نیستند (محتوای فرکانسی نگار هدف بسیار بیشتر از نشانگرهای لرزه‌ای می‌باشد). در نتیجه همبسته کردن نشانگرهای لرزه‌ای با نگار هدف به صورت بهینه انجام نمی‌پذیرد. شکل ۳-۳ نشان دهنده اختلاف میان

محتوای فرکانسی نگار هدف و نشانگر لرزه‌ای می‌باشد [۱].



شکل ۳-۳. مقایسه بین الف) نگار هدف و ب) نشانگر لرزه‌ای با تأکید بر اختلاف در محتوای فرکانسی.

ردلرزه ثبت شده توسط گیرنده‌های لرزه‌ای حاصل هم‌میخت موجک با سری بازتاب هستند؛ بنابراین مقدار مربوط به هر نمونه از نشانگر از مقادیر نمونه‌های بالایی و پایینی خود تأثیر پذیرفته است. بنابراین بهتر است با استفاده از عملگرهایی مانند عملگر هم‌میخت هر نمونه از نگار هدف به گروهی از نمونه‌های موجود در مجاورت نمونه نظیر در نشانگر لرزه‌ای نسبت داده شود. این موضوع در شکل ۳-۴ نشان داده شده است.



شکل ۳-۴. یک مثال شماتیک از تفاوت بین الف) وزن‌های تک نقطه‌ای به وسیله رابطه (۳-۱) و ب) وزن‌های

هم‌میختی ایجاد شده توسط رابطه (۳-۳) (همپسون ۲۰۰۱).

بعد از این توضیحات کاملاً قابل درک است که به جای رابطه (۳-۱)، رابطه (۳-۳) که عملگر

همامیخت در آن وارد شده است برای برقراری ارتباط به کار برده شود.

$$L = w_0 + w_1 * a_1 + w_2 * a_2 + w_3 * a_3 \quad (3-3)$$

که در آن علامت *، نشان دهنده همامیخت می باشد. یک بار دیگر ضرایب عملگر، یعنی وزن های مربوطه از طریق کمینه کردن خطای تخمین میانگین مربعات به صورت رابطه (3-4) به دست می آیند:

$$E^2 = \frac{1}{N} \sum_{i=1}^N (L_i - w_0 + w_1 A_{1i} + w_2 A_{2i} + \dots + w_M A_{Mi})^2 \quad (3-4)$$

3-2-4 تعیین بهترین نشانگرها با استفاده از رگرسیون مرحله ای

در بخش های گذشته در مورد انواع مختلف نشانگر که می توانند برای پیش بینی ویژگی های مخزنی مورد استفاده قرار گیرند، بحث گردید. با توجه به روش های ترکیب و الگوریتم های موجود، نشانگرهای مورد بحث تنها بخشی از نشانگرهای موجود در دنیای لرزه می باشند؛ بنابراین مسئله انتخاب نشانگرها و تعداد بهینه آنها برای استفاده در کارهای لرزه ای به وجود می آید. روشی که در اینجا شرح داده خواهد شد، ترکیبات مختلفی از نشانگرها را به منظور کمینه کردن حداقل مربعات خطا بین نمونه های آموزشی (که بر روی نشانگرها قرار دارند) و مقادیری که باید پیش بینی شوند (بر روی نگارهای هدف قرار دارند) مورد بررسی قرار می دهد [44].

$$E^2 = \frac{1}{N} \sum_{j=1}^N (L_j - w_0 - w_1 A_{j1} - w_2 A_{j2} - \dots - w_M A_{jM})^2 \quad (3-5)$$

در این روش از میان تمام نشانگرها (مجموعه ای با تعداد T نشانگر) دسته ای با تعداد M نشانگر جدا می شود. با استفاده از روشی که به آن روش «جست و جوی کامل» گفته می شود، مقدار حداقل مربعات برای تمام ترکیب های ممکن از نشانگرها محاسبه می شود و مجموعه ای که دارای کمترین مقدار خطاست انتخاب می شود. این فرآیند هزینه بسیاری را در بر خواهد داشت؛ بنابراین روش دیگری به نام روش رگرسیون مرحله ای پیشنهاد گردید [45] که از مراحل زیر تشکیل شده است.

۱. ابتدا با استفاده از روش جست و جوی کامل بهترین تک نشانگر (نشانگر دارای کمترین خطای

تخمین) انتخاب می‌گردد و نشانگر (۱) نامیده می‌شود.

۲. از طریق ترکیب نشانگر (۱) و سایر نشانگرها جفت نشانگری که دارای کمترین خطای تخمین میانگین باشند انتخاب می‌شوند.

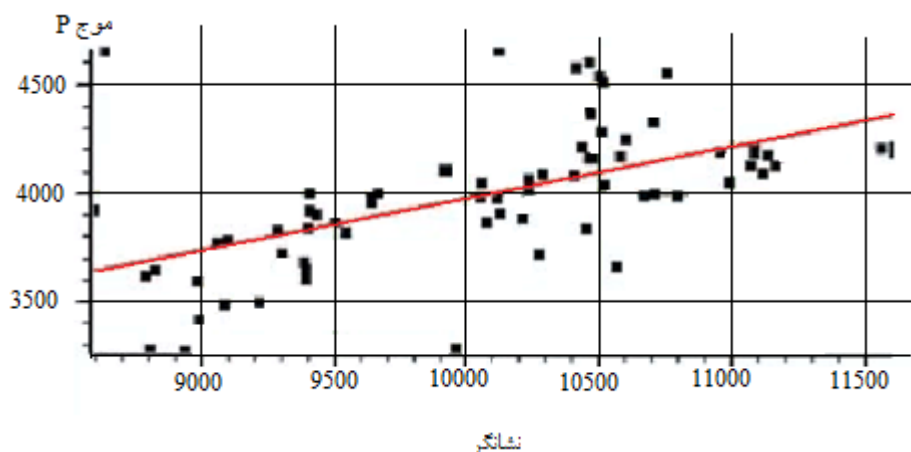
۳. با استفاده از زوج نشانگر انتخاب شده در مرحله ۲ و ترکیب آن‌ها با سایر تک نشانگرها، سه نشانگری که ترکیب آن‌ها کمترین خطای تخمین را داشته باشند تعیین می‌گردند.

با ادامه این فرآیند بهترین مجموعه با تعداد M نشانگر توسط روش رگرسیون مرحله‌ای به دست می‌آید. روش رگرسیون خطی نسبت به روش جست‌وجوی کامل سرعت بالاتری دارد و علاوه بر آن می‌توان از استقلال نشانگرهای مورد مطالعه مطمئن بود؛ زیرا رگرسیون مرحله‌ای به‌طور خودکار نشانگرهای غیر مستقل را که سهمی در بهبود خطای میانگین تخمین ندارند و تنها باعث طولانی شدن زمان محاسبات می‌شوند، حذف می‌کند؛ اما روش رگرسیون خطی از نظر به دست آمدن پاسخ بهینه قابل اعتماد نیست. به طور مثال اگر ۴ نشانگر در این روش انتخاب شوند، ممکن است همان بهترین ۴ نشانگر انتخاب شده در روش جست‌وجوی کامل نباشد. در این روش خطای تخمین هر نشانگر اضافی که انتخاب می‌شود، کمتر یا مساوی با خطای تخمین ترکیبات کوچک‌تر قبلی است. اما در صورتی که خطا بزرگ‌تر باشد، رگرسیون مرحله‌ای تمام وزن‌ها را برای این نشانگر جدید برابر صفر قرار داده و در نتیجه خطای تخمین برابر با خطای تخمین محاسبه شده برای ترکیب قبلی می‌شود.

۳-۳ شبکه‌های عصبی

تحلیل‌هایی که تا کنون در این بخش مطرح شدند از نوع خطی می‌باشند. محدودیت این روش‌ها می‌تواند با یک مثال بهتر درک شود. در شکل ۳-۵ موج P به عنوان نگار هدف در مقابل یک تک نشانگر رسم شده است. همان‌طور که پیش‌تر مطرح گردید، رگرسیون خطی به‌وسیله کمینه کردن خطای حداقل مربعات تخمین بین نشانگر لرزه‌ای نگار هدف محاسبه می‌شود. چندین گزینه برای

محاسبه این منحنی وجود دارد. یک گزینه اعمال یک تبدیل چند نشانگری به یکی یا هر دو متغیر و برازش یک خط مستقیم به داده‌های تبدیل شده است. گزینه دوم برازش یک خط با درجه بالاتر (برازش یک منحنی) به داده‌ها است. راه سوم استفاده از شبکه عصبی برای پیدا کردن رابطه بین نشانگرها و هدف می‌باشد.

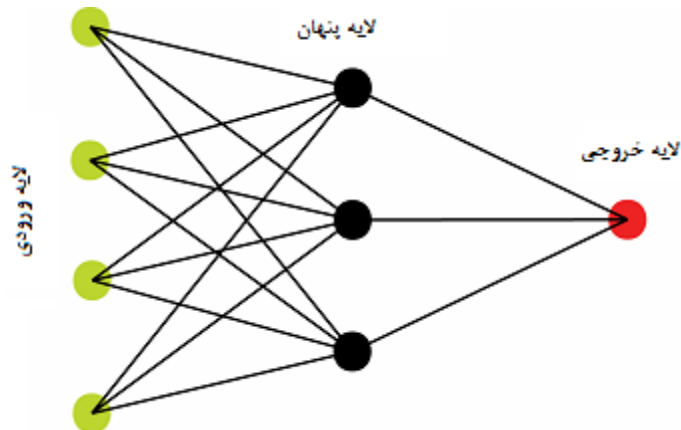


شکل ۳-۵. نمودار مقاطع نگار هدف در مقابل نشانگر لرزه‌ای. خط رگرسیون به وسیله رابطه کمینه کننده (۳-۵) برازش داده شده است.

شبکه عصبی عملکرد بهتری در تعیین رابطه مناسب بین نشانگرها و نگار هدف دارد. هدف اصلی این روش طراحی مدلی ریاضی از روی مسائل زیستی است به طوری که خصوصیات و ساختارهای شبکه‌های عصبی زیستی را به منظور ایجاد یک سیستم هوشمند پردازش اطلاعات تقلید کند. شبکه‌های عصبی با الگوریتم و ساختارهای مختلف وجود دارند. یکی از شبکه‌های عصبی پرکاربرد در تخمین پارامترهای پتروفیزیکی مخزن با استفاده از نشانگرهای لرزه‌ای و نگارهای چاه‌نگاری، شبکه عصبی احتمال پذیر می‌باشد که موفقیت زیادی در برآورد تخلخل با استفاده از نشانگرهای لرزه‌ای دارد [۴۶].

۳-۳-۱ شبکه عصبی چند لایه‌ای پیشرو

شبکه عصبی سال‌هاست که در ژئوفیزیک مورد استفاده قرار می‌گیرد [۴۷]. لیو [۴۸] عملکرد شبکه عصبی چند لایه پیشرو (MLFN)^۱ را برای تخمین نگار از روی داده‌های لرزه‌ای بکار برد. MLFN شبکه مرسوم است که به صورت شماتیک در شکل ۳-۶ نشان داده شده است.



شکل ۳-۶. ساختار یک شبکه عصبی ساده با یک لایه پنهان.

شبکه شامل یک لایه ورودی، یک لایه خروجی و یک یا تعداد بیشتری لایه پنهان است. هر لایه شامل گره‌هایی است و گره‌ها به وسیله وزن‌هایی به هم مرتبط هستند. در کارهای لرزه‌ای تعداد گره‌های لایه ورودی برابر با تعداد نشانگرها است. لایه خروجی یک گره دارد، چون تنها یک نگار مورد تخمین قرار گرفته است.

فرآیند آموزش شامل پیدا کردن وزن‌های بهینه بین گره‌ها است. آموزش توسط ارائه نمونه‌های آموزشی به شبکه صورت می‌گیرد. هر مثال شامل داده مربوط به یک نمونه زمانی $\{A_1, A_2, A_3, L\}$ است که A_i نشانگرها و L مقدار اندازه‌گیری شده نگار هدف است. به تعداد مجموع تمام نمونه‌های لرزه‌ای در پنجره‌ی چاه‌های در دسترس نمونه آموزشی وجود دارد.

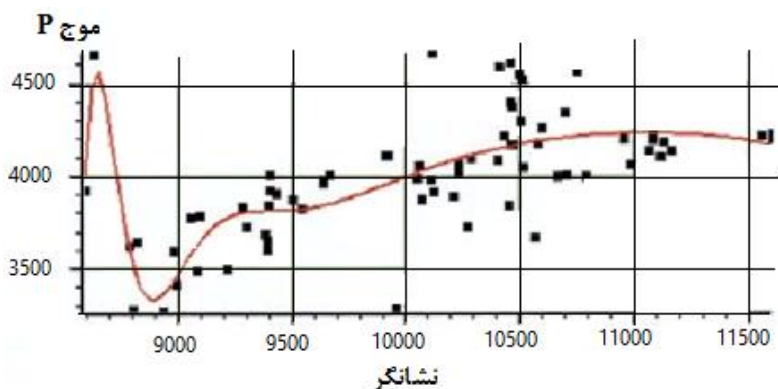
مسئله تخمین وزن‌ها می‌تواند به عنوان یک مسئله بهینه‌سازی غیرخطی در نظر گرفته شود که

^۱ . Multi-Layer Feedforward Neural network

هدف کمینه کردن خطای میانگین مربعات بین مقادیر واقعی نگار هدف و مقادیر تخمین زده شده نگار هدف به وسیله شبکه عصبی می باشد. این مسئله به طور سنتی از طریق پس انتشار که یک حالت از گرادیان نزولی است، حل می شود. امروزه برای سرعت بخشیدن به همگرایی و دوری جستن از کمینه های محلی از روش های جدید گرادیان مزدوج و شبیه سازی ذوب^۱ استفاده می کنند [۴۶].

به عنوان یک مثال شکل ۳-۷ منحنی تخمین برای نمونه های مشابهی که در شکل ۳-۱۵ استفاده شده است را با استفاده از MLFN با پنج گره در لایه پنهان نشان می دهد. در این مورد، چون فقط یک نشانگر وجود دارد، تنها یک تک گره در لایه ورودی وجود دارد.

شکل ۳-۷ دو جنبه از رفتارهای MLFN را نشان می دهد. جنبه مثبت این است که در روش MLFN نسبت به روش رگرسیون خطی، مقادیر داده ها بر روی بازه بزرگتری از نشانگر با دقت بیشتری مدل می شود. جنبه منفی ناپایداری در مقادیر پایین نشانگر است. چون که شبکه تلاش می کند تا داده ها را بسیار نزدیک به هم مدل کند. این نمونه ای از شرایطی است که تحت عنوان بیش برازش شناخته می شود.



شکل ۳-۷. منحنی تخمین به دست آمده از روش MLFN با ۵ گره در لایه پنهان.

^۱ simulated annealing

۳-۳-۲ شبکه عصبی احتمالاتی^۱

شبکه عصبی احتمالاتی نوع خاصی شبکه عصبی است [۴۹]. شبکه عصبی احتمالاتی (PNN) در واقع یک طرح درون‌یابی ریاضیاتی است که از یک شبکه عصبی برای مقاصدش استفاده می‌کند. این یک مزیت است، زیرا به وسیله مطالعه روابط ریاضیاتی اغلب بهتر می‌توان نسبت به MLFN رفتار را متوجه شد.

داده‌هایی که توسط PNN مورد استفاده قرار می‌گیرند، مشابه همان داده‌های آموزشی مورد استفاده در MLFN است. آن‌ها شامل یک سری از مثال‌های آموزشی هستند، یکی برای هر نمونه لرنزهای در پنجره‌هایی از تمام چاه‌ها:

$$\begin{aligned} & \{A_{11}, A_{21}, A_{31}, L_1\} \\ & \{A_{12}, A_{22}, A_{32}, L_2\} \\ & \{A_{13}, A_{23}, A_{33}, L_3\} \\ & \vdots \\ & \{A_{1n}, A_{2n}, A_{3n}, L_n\} \end{aligned}$$

که n نمونه آموزشی و سه نشانگر وجود دارد. مقادیر L_i مقادیر اندازه‌گیری شده هدف برای هر نمونه هستند. با وارد کردن داده آموزش به PNN، هر مقدار جدید از نگار خروجی به صورت ترکیب خطی از مقادیر نگار موجود در داده‌های آموزشی در نظر می‌گیرد. برای یک نمونه داده جدید با مقادیر نشانگر

$$x = \{A_{1j}, A_{2j}, A_{3j}\}, \quad (۶-۳)$$

مقدار نگار جدید به صورت رابطه (۷-۳) محاسبه می‌شود.

$$\hat{L}(x) = \frac{\sum_{i=1}^n L_i \exp(-D(x, x_i))}{\sum_{i=1}^n \exp(-D(x, x_i))} \quad (۷-۳)$$

^۱. Probabilistic Neural Network

که

$$D(x, x_i) = \sum_{j=1}^3 \left(\frac{x_j - x_{ij}}{\sigma_j} \right)^2 \quad (8-3)$$

کمیت $D(x, x_i)$ فاصله بین نقطه ورودی و هر یک از نقاط آموزشی x_i است. این فاصله در فضای چند بعدی به کار گرفته شده توسط نشانگرها و مقیاس شده با استفاده از کمیت σ_j اندازه گیری می شود که برای هر نشانگر تفاوت زیادی دارد. روابط (3-7) و (3-8) عملکرد شبکه PNN را شرح می دهند. آموزش شبکه شامل تعیین مجموعه ای بهینه از پارامترهای هموارسازی σ_j می باشد. معیار تعیین این پارامترها این است که پاسخ شبکه باید کمترین خطای ارزیابی را داشته باشد.

تعریف پاسخ ارزیابی برای m-امین نمونه هدف به صورت رابطه (3-9) می باشد.

$$\hat{L}_m(x_m) = \frac{\sum_{i \neq 1} L_i \exp(-D(x_m, x_i))}{\sum_{i \neq 1} \exp(-D(x_m, x_i))} \quad (9-3)$$

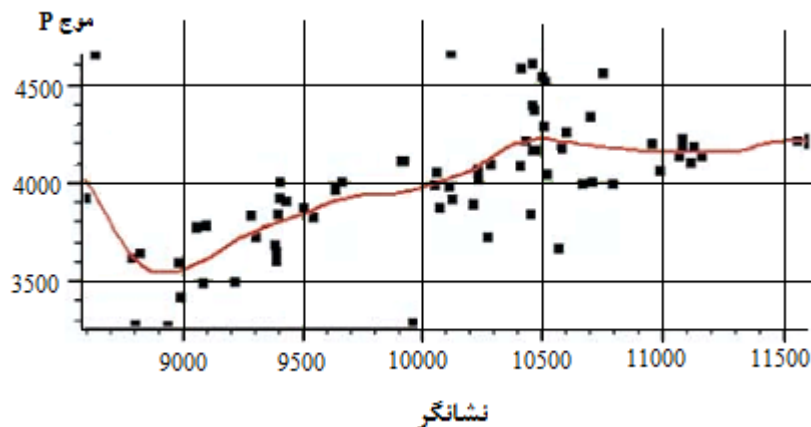
این مقدار محاسبه شده ی m-امین نمونه هدف است، زمانی که آن نمونه از داده های آموزشی خارج شده باشد. از آنجایی که مقدار واقعی این نمونه مشخص است، می توان خطای تخمین برای آن نمونه را محاسبه کرد. این فرآیند برای هر کدام از نمونه های آموزشی تکرار می شود و می توان خطای تخمین کلی را به صورت رابطه (3-10) تعریف نمود:

$$E_V(\sigma_1, \sigma_2, \sigma_3) = \sum_{i=1}^N \left(L_i - \hat{L}_i \right)^2 \quad (10-3)$$

بایستی توجه نمود که مقدار خطای تخمین به انتخاب پارامتر σ_j بستگی دارد. این کمیت به وسیله استفاده از الگوریتم گرادیان مزدوج غیرخطی که توسط مسترز [49] معرفی شد، کمینه می شود. شبکه حاصله این ویژگی را دارد که خطای ارزیابی در آن کمینه شده است.

عملکرد PNN بر روی یک منحنی متقاطع ساده در شکل 3-8 نشان داده شده است. می توان مشاهده

کرد که PNN در داده‌های موجود عملکردی مطلوب و نزدیک به عملکرد MLFN دارد، آن ناپایداری که در مورد محدودیت بازه نشانگرها در MLFN وجود داشت را ندارد. بزرگ‌ترین مشکل PNN آن است که دلیل اینکه چون تمام داده‌های خروجی را با تک تک داده‌های آموزشی مقایسه می‌کند، می‌تواند بسیار زمان‌بر شود.



شکل ۳-۸. منحنی تقاطع شبکه عصبی PNN. نمودار برازش داده شده به داده‌ها در بازه پایینی داده‌ها پایدارتر است.

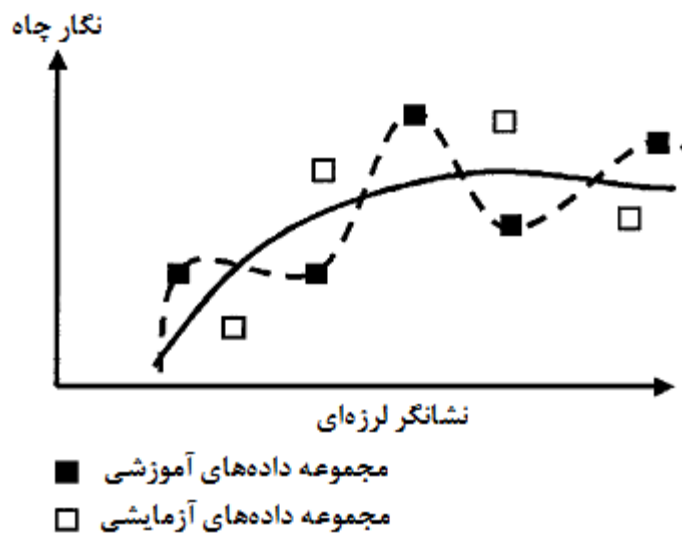
۳-۳-۳ اعتبار سنجی

تعیین تعداد بهینه نشانگرها در روش چند نشانگری از اهمیت بالایی برخوردار است. در روش رگرسیون مرحله‌ای دسته‌ای از نشانگرها که موجب کاهش خطا در مرحله آموزش می‌شوند، تعیین می‌شود. به دلیل اینکه همواره خطای تخمین تبدیل چند نشانگری با $M+1$ باید کمتر و یا مساوی با خطای تخمین تبدیل نشانگر چندگانه با M نشانگر باشد؛ با افزایش تعداد نشانگرها مقدار خطای تخمین کاهش می‌یابد. این کاهش خطا که به دلیل افزایش تعداد نشانگرهای استفاده شده در روش رگرسیون مرحله‌ای می‌باشد، ممکن است موجب بهبود نتیجه نهایی نگردد و حتی در مواردی نتایج را بدتر کند که به این پدیده اصطلاحاً بیش آموزش گفته می‌شود.

استفاده کردن از تعداد بیشتری از نشانگرها در روش رگرسیون چند نشانگری مشابه برازش دادن منحنی تقاطع با چند جمله‌ای درجه بالاتر است. تا کنون تکنیک‌های آماری متعددی برای اندازه‌گیری

میزان اطمینان و درستی برازش منحنی‌های درجه بالاتر با نشانگرها معرفی شده‌اند که برخی تنها برای تخمین‌های خطی کارایی دارند و از آن‌ها نمی‌توان در تخمین غیرخطی که توسط شبکه عصبی انجام می‌گیرد استفاده شوند. روش اعتبار سنجی متقابل که در ادامه شرح داده خواهد شد برای استفاده در هر دو نوع تخمین کاربرد دارد [۵۰].

در این روش کل مجموعه داده‌های آموزشی به دو زیرمجموعه (مجموعه داده‌های آموزشی و مجموعه داده‌های اعتبار سنجی) تقسیم می‌گردد. از داده‌های آموزشی برای به دست آوردن تبدیل و از داده‌های اعتبار سنجی به منظور اندازه‌گیری خطای تخمین استفاده می‌شود. در شکل ۳-۹ مفهوم بیش آموزش نشان داده شده است. همان‌طور که مشاهده می‌شود آموزش بیش از حد موجب می‌شود تا برازش ضعیف‌تری به مجموعه داده‌های اعتبار سنجی صورت گیرد.



شکل ۳-۹. مثال اعتبارسنجی متقاطع. نشانگر لرزه‌ای در مقابل نگار چاه.

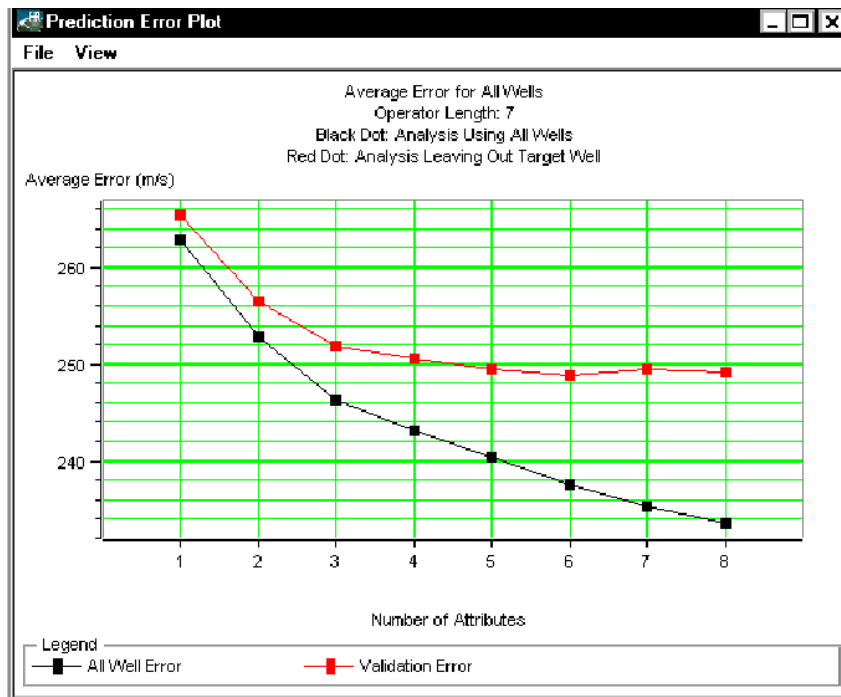
با توجه به شکل (۳-۹) مشاهده می‌گردد که با استفاده از منحنی‌های درجه بالاتر برازش با داده‌های آموزشی افزایش می‌یابد اما این کار همیشه به ایجاد جواب بهینه ختم نمی‌شود. در این شکل منحنی خط چین تطابق بسیار بالایی با داده‌های آموزشی دارد اما بر خلاف آن تطابق بسیار ضعیفی با داده‌های اعتبار سنجی دارد. اینجاست که اهمیت و لزوم انجام اعتبار سنجی متقابل نمایان می‌گردد. در روش اعتبار سنجی متقابل در هر بار آموزش، نمونه‌ای را از کل نمونه‌های آموزشی جدا می‌شود و

شبکه آموزش می‌یابد. سپس با استفاده از شبکه آموزش یافته، مقدار نمونه‌ای که کنار گذاشته شده بود تخمین زده می‌شود. این فرآیند به تعداد تمام نمونه‌های آموزشی تکرار می‌شود و در نهایت خطای اعتبارسنجی کل محاسبه می‌گردد. در لرزه‌شناسی هر بار نمونه‌های مربوط به یک چاه کنار گذاشته شده، با استفاده از نمونه چاه‌های دیگر آموزش انجام گرفته و در نهایت نمونه‌های مربوط به چاه کنار گذاشته شده تخمین زده می‌شود. این فرآیند به تعداد چاه‌ها تکرار می‌شود. خطای اعتبارسنجی کل را می‌توان از رابطه (۱۱-۳) محاسبه کرد.

$$E_v^2 = \frac{1}{N} \sum_{i=1}^N e_{vi}^2 \quad (11-3)$$

در این معادله e_{vi} خطای اعتبارسنجی برای چاه i ام و N تعداد چاه‌های موجود در تحلیل است. در هر مرحله از رگرسیون مرحله‌ای خطای اعتبارسنجی محاسبه می‌گردد تا از اثربخش بودن نشانگر اضافه شده اطمینان حاصل گردد. زمانی که با افزایش تعداد نشانگر خطا افزایش یابد، رگرسیون مرحله‌ای متوقف و تعداد بهینه نشانگر مشخص می‌شود.

شکل ۳-۱۰ یک منحنی اعتبارسنجی متقابل را نشان می‌دهد که در آن تأثیر افزایش تعداد نشانگر بر خطای اعتبارسنجی کل بررسی شده است. نمودار سیاه رنگ خطای تخمین کل با استفاده از تمام داده را نشان می‌دهد. مشاهده می‌گردد زمانی که از تمام داده‌ها برای برازش منحنی استفاده می‌شود با افزایش تعداد نشانگرها خطا کاهش می‌یابد که در اینجا بحث بیش آموزش پدید می‌آید؛ اما برای تعیین تعداد بهینه نشانگر باید از روی میزان خطای اعتبارسنجی کل تصمیم‌گیری کرد. نمودار قرمز رنگ خطای اعتبارسنجی کل است و مشاهده می‌گردد که از نشانگر شماره ۶ به بعد خطا رو به افزایش است، بنابراین عدد ۶ نماینده تعداد بهینه نشانگر است.



شکل ۳-۱۰. بررسی خطای اعتبار سنجی با استفاده از روش رگرسیون.

۴-۳ مبانی و اصول طبقه‌بندی و خوشه‌بندی

بازشناسی الگو یکی از خصوصیات است که در انسان و سایر موجودات وجود دارد. توصیف انسان از اشیاء پیرامون خود را الگو می‌نامند. الگو را می‌توان به دو دسته الگوهای زمانی و الگوهای مکانی تقسیم کرد. از الگوهای مکانی می‌توان به حروف، اثر انگشت و نقشه‌های هواشناسی و از الگوهای زمانی می‌توان به سیگنال گفتار، سیگنال لرزه‌ای و حرکت اشاره کرد. امروزه در زندگی روزمره باز شناسی الگو کاربردهای فراوان و مهمی دارند که تصدیق ام‌ضاء، تصدیق اثر انگشت، شناسایی سکانس DNA و شناسایی رخساره‌های لرزه‌ای نمونه‌هایی از آنها هستند [۵۱].

در باز شناسی الگو عمدتاً یک مجموعه به زیرمجموعه‌هایی تقسیم می‌شود به نحوی که اعضای هر زیرمجموعه با یکدیگر دارای بیشترین شباهت و با اعضای سایر زیرمجموعه‌ها بیشترین اختلاف را دارند که این فرآیند به رسته بندی معروف است. از آنجایی که بازشناسی الگو معادل محاسبه شباهت یا مجاورت می‌باشد، در عمل رسته بندی یک ملاک شباهت سنجی مورد نیاز است که به وسیله آن بین نمونه ورودی مجهول و نمونه معلوم مقایسه انجام گیرد و میزان شباهت آنها به یکدیگر بدست

برای بررسی شباهت میان اشیاء مختلف، باید ویژگی‌هایی برای آن‌ها تعریف شود و عمل رسته بندی بر روی این ویژگی‌ها صورت گیرد. این ویژگی‌ها به صورت یک بردار (یک سری از اعداد) هستند.

روش‌های رسته بندی به دو گروه اصلی روش‌های طبقه‌بندی با سرپرستی و روش‌های طبقه‌بندی بدون سرپرستی تقسیم‌بندی می‌شوند. در روش طبقه‌بندی با سرپرستی بر اساس دانش قبلی تعداد طبقات و طبقه یا کلاس مربوط به هر نمونه مشخص است و به این داده‌ها، داده‌های برچسب خورده می‌گویند. این داده‌ها به دو دسته داده‌های آموزش و داده‌های آزمایش تقسیم می‌شوند. در مرحله اول با استفاده از داده‌های آموزش الگوی میان نشانگرها و برچسب مربوطه تعیین می‌گردد و در مرحله بعد به منظور ارزیابی الگوی به دست آمده از داده‌های آزمایش استفاده می‌گردد. در صورت مطلوب بودن نتیجه ارزیابی، الگو مورد تأیید قرار می‌گیرد و بر روی سایر داده‌های بدون برچسب اعمال می‌گردد.

از آنجایی که داده‌های لرزه‌بازتابی نمونه‌های زمانی هستند که طبقه مربوط به آن‌ها معلوم نیست، این داده‌ها جزو دسته داده‌های برچسب نخورده هستند؛ اما در محل چاه می‌توان عمل برچسب گذاری داده‌های لرزه‌بازتابی را با استفاده از داده‌های چاه‌نگاری (دانش قبلی) را انجام داد.

در روش‌های طبقه‌بندی بدون سرپرست که خوشه‌بندی نیز نامیده می‌شود، نظم و الگوی بین داده‌هایی که هیچ‌گونه دانش قبلی (تمام داده‌ها بدون برچسب هستند) در مورد آن‌ها وجود ندارد، کشف می‌شود.

۳-۵ خوشه‌بندی

در بسیاری از مسائل طبقه‌بندی که هیچ‌گونه اطلاعاتی در مورد ساختار داده‌ها در دست نیست، روش‌های طبقه‌بندی با سرپرستی قابل استفاده نیستند. در چنین شرایطی مهندسان از روش‌های طبقه‌بندی بدون سرپرست یا خوشه‌بندی بهره می‌گیرند. هدف اصلی روش‌های خوشه‌بندی استفاده از ملاک‌های شباهت به منظور آشکار کردن سازمان‌دهی الگوها به گروه‌های معنادار می‌باشد به طوری که

بتوان با استفاده از شباهت‌ها و تضادهای موجود در داده‌ها تصمیم‌گیری مناسبی انجام داد. فرآیند خوشه‌بندی از مراحل زیر تشکیل می‌شود [۵۱].

۱. انتخاب ویژگی: ویژگی‌هایی که در خوشه‌بندی استفاده می‌گردد باید به نحوی انتخاب گردند که حداکثر اطلاعات ممکن را برای دستیابی به هدف مورد نظر فراهم آورند. به‌طور مثال در زمینه تشخیص رخساره‌های لرزه‌ای، ویژگی‌های استخراج شده مورد استفاده در روش خوشه‌بندی نشانگرهای لرزه‌ای می‌باشند.

۲. الگوریتم خوشه‌یابی:

در این مرحله الگوریتمی انتخاب می‌شود که بتواند برای یک مجموعه داده، خوشه‌بندی بهینه‌ای را انجام دهد. با استفاده از یک معیار مجاورت مناسب و یک معیار خوشه‌بندی خوب، می‌توان یک الگوریتم خوشه‌یابی بهینه را ایجاد کرد.

۳. انتخاب تعداد خوشه: در داده‌های لرزه‌ای بازتابی، انتخاب تعداد خوشه‌ها به تجربه مفسر بستگی دارد. منحنی زمان دوام می‌تواند یکی از ملاک‌های خوشه‌بندی باشد.

۴. تفسیر نتایج: با توجه به اینکه روش‌های خوشه‌یابی از روش‌های بدون سرپرست می‌باشند در بخش تفسیر نتایج تا ارتباط رخساره‌های به دست آمده با رخساره‌های زمین‌شناسی تعیین گردد.

۳-۶ طبقه‌بندی

از روش‌های متداول طبقه‌بندی می‌توان به روش‌های نزدیک‌ترین همسایه (knn)، روش ساده بیزی، درخت تصمیم‌گیری، شبکه‌های عصبی شعاعی و ماشین بردار پشتیبان اشاره کرد. با توجه به کاربرد گسترده روش ماشین بردار پشتیبان و نتایج مطلوب حاصل از آن، در این پایان‌نامه این روش برای مقایسه با الگوریتم آدابوست استفاده گردید.

۳-۶-۱ طبقه‌بند k نزدیک‌ترین همسایه

یکی از بهترین طبقه‌بندها، طبقه‌بند نزدیک‌ترین هم‌سایه است. این طبقه‌بند یک روش یادگیری بر پایه فاصله است که در بخش‌هایی مانند داده‌کاوی، شناسایی الگوی آماری و پردازش تصویر کاربرد دارد.

در مجموعه داده‌ها، هر نمونه دارای n نشانگر است که از آن‌ها برای تشکیل بردار n بعدی X استفاده می‌شود:

$$X = (x_1, x_2, \dots, x_n) \quad (3-12)$$

این n نشانگر مستقل از هم در نظر گرفته می‌شوند. همچنین هر نمونه یک نشانگر دیگر هم دارد که با y نشان داده می‌شود (متغیر وابسته) که مقدار آن به n نشانگر دیگر (X) وابسته است. فرض می‌شود که:

۱. y متغیر طبقه‌بندی است و یک رابطه خطی $y = f(X)$ وجود دارد که کلاس‌های هر برداری را به آن مرتبط می‌کند.

۲. یک دسته از چنین بردارهایی (T) همراه با کلاس‌های مربوطه داده شده‌اند:

$$(X^i, y^i) \quad \text{for } i=1,2,\dots,T \quad (3-13)$$

این مجموعه همان داده‌های آموزش هستند.

فرض می‌شود که یک نمونه جدید وجود دارد که در آن $X = u$ است. هدف تعیین کلاسی است که نمونه جدید به آن تعلق دارد. اگر تابع f مشخص باشد، به راحتی می‌توان $v = f(u)$ را محاسبه و نمونه جدید را طبقه‌بندی کرد.

ایده روش k -نزدیک‌ترین همسایه، شناسایی k نمونه در مجموعه آموزش است که در آن‌ها متغیرهای مستقل X شبیه به u باشد. از این k نمونه برای طبقه‌بندی این نمونه جدید استفاده می‌شود.

گردد و v را از مقادیر y برای این نمونه‌ها به دست می‌آورد.

وقتی که در مورد همسایه‌ها صحبت می‌شود منظور یک اندازه فاصله یا ناهماهنگی است که می‌توان بین نمونه‌ها بر پایه متغیرهای مستقل محاسبه کرد. مشهورترین روش برای اندازه‌گیری فاصله، فاصله اقلیدسی می‌باشد. فاصله اقلیدسی بین نقطه‌های X و u به صورت رابطه (۱۴-۳) تعریف می‌شود.

$$d(X, u) = \sqrt{\sum_{i=1}^n (x_i - u_i)^2}. \quad (14-3)$$

ساده‌ترین حالت در روش k -نزدیک‌ترین همسایه برای $k=1$ است. در این حالت در مجموعه آموزش نمونه‌ای را که نزدیک‌ترین فاصله را به u دارد تعیین و $v=y$ قرار داده می‌شود (y کلاس نزدیک‌ترین نمونه همسایه است).

این یک واقعیت جالب توجه است که زمانی که تعداد زیادی نمونه در دسته آموزشی وجود دارد، ایده ساده استفاده از نزدیک‌ترین همسایه برای طبقه‌بندی نمونه‌ها می‌تواند بسیار قدرتمند باشد. می‌توان ثابت کرد که اگر تعداد زیادی داده وجود داشته باشد و از یک قانون طبقه‌بندی پیچیده استفاده شود، در بهترین حالت می‌توان خطای طبقه‌بندی را تا نصف قانون ساده ۱-نزدیک‌ترین همسایه کاهش داد. در روش k -نزدیک‌ترین همسایه، تعداد k عدد از نزدیک‌ترین همسایه‌های u را پیدا کرده و از قانون اکثریت آرا برای طبقه‌بندی داده جدید استفاده می‌شود.

مزیت مقادیر بالاتر k این است که ریسک بیش‌برازش ناشی از وجود نوفه در داده‌ها را کاهش می‌دهد. در کاربردهای متداول k یک واحد یا ده‌تایی است. توجه کنید که اگر $k=n$ باشد، تعداد نمونه‌ها در مجموعه داده آموزش، برای تمام نمونه‌ها صرف نظر از u فقط کلاسی را که در داده‌های آموزش اکثریت دارد، پیش‌بینی می‌کند. این مشخصاً یک مورد از بیش‌همواری است، جز در حالتی که تمام متغیرهای مستقل هیچ گونه اطلاعاتی نسبت به متغیر وابسته وجود نداشته باشد.

۳-۶-۲ ماشین بردار پشتیبان^۱

واپنیک^۲ [۵۳] اولین الگوریتم ماشین بردار پشتیبان را ارائه کرد. این الگوریتم توسط بوسر و همکاران [۵۴] برای حالت غیرخطی تعمیم داده شد که از آن می‌توان برای طبقه‌بندی و رگرسیون استفاده کرد. این روش نسبتاً جدید به دلیل قابلیت استفاده در حل مسائل گوناگون و کارایی بهتر نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی از محبوبیت بالایی برخوردار گردیده است. روش SVM دارای مزیت‌هایی همچون آموزش نسبتاً سریع، هوشمند بودن و به دام نیفتادن در بیشینه‌های محلی، کارایی مناسب در داده‌های با حجم بالا و عدم نیاز به داده‌های آموزشی فراوان می‌باشد.

طبقه‌بند SVM بر پایه طبقه‌بندی خطی داده استوار است و برای یافتن خطی که بین داده‌های مربوط به دو طبقه مختلف، دارای بیشترین حاشیه اطمینان باشد تلاش می‌کند. این طبقه‌بند به منظور حل معادله‌ی مربوط به پیدا کردن خط بهینه برای داده‌ها، از روش‌های شناخته شده در حل مسائل محدودیت‌دار، مانند روش‌های برنامه‌ریزی درجه دوم^۳ استفاده می‌کند.

در SVM غیرخطی برای ایجاد امکان جداسازی داده‌ها از یکدیگر، ابتدا داده‌ها به کمک تبدیل‌های پیچیده نگاشت می‌یابند و در مرحله بعد چگونگی جداسازی داده‌ها بر اساس برجسب‌های تعریف شده مشخص می‌شود.

SVM کاربردهای فراوانی در سیستم راهنمایی خودکار اتومبیل، سیستم تحلیل ریسک، سیستم‌های بازرسی کیفیت، کنترل هواپیمای بدون خلبان، کنترل مسیر ربات، جراثیل و سیستم‌های بصری و غیره در مباحث غیر ژئوفیزیکی و تفکیک بافت، تفکیک سنگ‌شناسی و غیره در بحث ژئوفیزیکی دارد [۵۵].

۳-۶-۲-۱ روش طبقه‌بندی SVM

هدف روش‌های طبقه‌بندی خطی تفکیک داده‌ها از هم به وسیله ساختن یک ابر سطح (یک معادله

^۱. Support Vector Machine

^۲. Vapnik

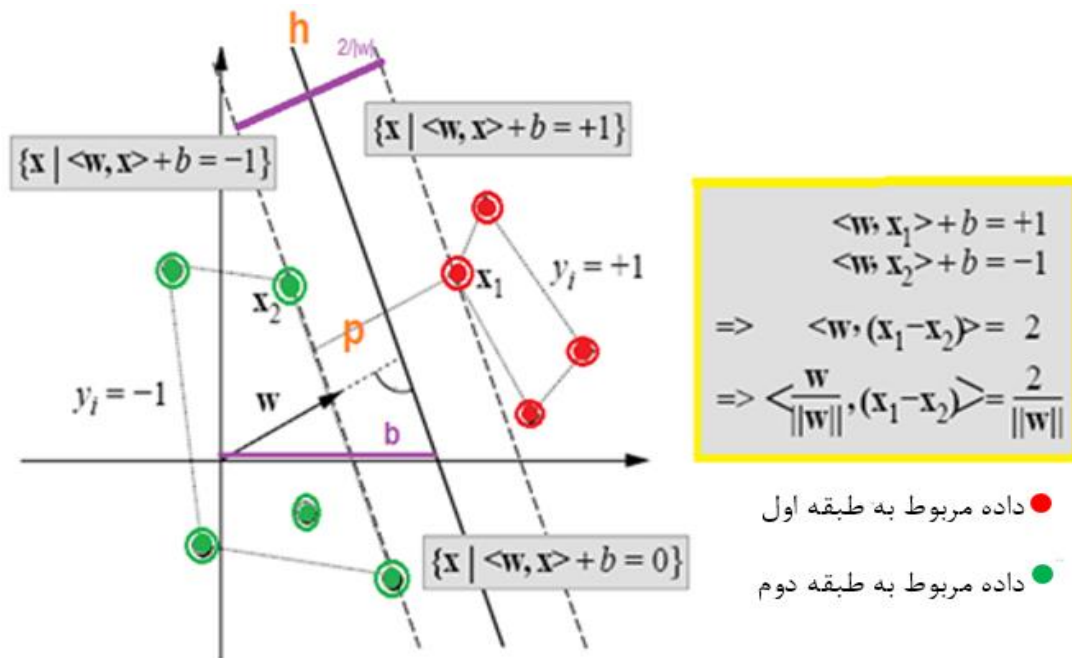
^۳. Quadratic Programming

خطی است) می‌باشد. روش SVM نیز که یک روش طبقه‌بندی خطی است، از این قاعده مستثنا نمی‌باشد و تلاش می‌کند تا ابرسطحی را تعیین کند که با حداکثر فاصله، نمونه‌های مربوط به دو کلاس مختلف را از یکدیگر تفکیک نماید [۵۶].

طبقه‌بندی‌های SVM به دو دسته SVM خطی و SVM مبتنی بر کرنل تقسیم می‌شوند. در روش SVM خطی با فرض اینکه داده‌ها دارای خاصیت تفکیک‌پذیری خطی می‌باشند، ابر سطح‌هایی ایجاد می‌شوند که داده‌ها را با حداکثر حاشیه ممکن از یکدیگر جدا می‌کند. روش SVM مبتنی بر کرنل زمانی مورد استفاده قرار می‌گیرد که داده‌ها قابلیت تفکیک‌پذیری خطی را نداشته باشند، بنابراین از طریق داده‌ها به فضای با ابعاد بیشتر نگاشت می‌شوند و در این فضای جدید داده‌ها به صورت خطی تفکیک می‌شوند.

۳-۲-۲-۲ نحوه تشخیص ابر سطح جدا کننده توسط SVM

در این بخش، نحوه ساخت ابرسطح جداکننده به تفصیل بیان شده است. در شکل ۳-۱۱ نحوه تشکیل ابرسطح جدا کننده نشان داده شده است. مجموعه نقاط داده $\{(x_1, c_1), (x_2, c_2), \dots, (x_n, c_n)\}$ مفروض است و هدف، طبقه‌بندی آن‌ها به دو طبقه‌ی $c_i = \{-1, 1\}$ می‌باشد. در شکل (۱۳)، داده‌ها دو بعدی هستند، یعنی هر داده تنها با دو متغیر تعریف می‌شود و عملگر $\langle * \rangle$ به معنی ضرب داخلی است [۵۳].



شکل ۳-۱۱. روش تشخیص ابر سطح جدا کننده توسط روش SVM.

در روش SVM مبتنی بر کرنل، بردارهای ورودی به یک فضای چندبعدی نگاشت می‌شوند. پس از آن، یک ابرسطح ساخته خواهد شد که با حداکثر فاصله ممکن، بردارهای ورودی را از هم جدا کند. به این ابرسطح، ابرسطح با حداکثر مرز جداکننده، گفته می‌شود. همان‌گونه که در شکل ۳-۱۱ نشان داده شده است، دو ابرسطح موازی در دو سمت ابرسطح با حداکثر مرز جداکننده ساخته خواهد شد که داده‌های مربوط به دو طبقه را به‌گونه‌ای از هم مجزا می‌کنند که هیچ داده‌ای در مرز بین این دو ابرسطح قرار نمی‌گیرد. ابرسطح با حداکثر مرز جداکننده، ابرسطحی است که فاصله بین دو ابرسطح موازی را به حداکثر می‌رساند. فرض بر این است که هر قدر مرز جداکننده یا در واقع، فاصله بین دو ابرسطح موازی بیشتر باشد، خطای طبقه‌بندی هم کمتر خواهد بود [۴۳].

ابتدا یک پوسته محدب^۱ در اطراف هر کدام از طبقه‌ها در نظر گرفته می‌شود. در شکل ۳-۱۱، در اطراف نقاط مربوط به طبقه ۱- و نقاط مربوط به طبقه +۱، پوسته محدب رسم شده است. خط p خطی است که نزدیک‌ترین فاصله بین دو پوسته محدب را نشان می‌دهد. h که در واقع همان ابرسطح

^۱. Convex

جداکننده است، عمود منصف p است. b ، عرض از مبدأ برای ابرسطح با حداکثر مرز جداکننده است. اگر از b صرف نظر شود، تنها ابرسطح‌هایی که از مبدأ می‌گذرند، پاسخ می‌باشند. فاصله عمودی ابرسطح تا مبدأ با تقسیم قدر مطلق b بر طول w به دست می‌آید [۵۶].

رابطه ریاضی برای این دو ابرسطح موازی که مرز جداکننده را تشکیل می‌دهند با (۳-۱۵) و (۳-۱۶) تعریف شده است.

$$w \cdot x - b = 1 \quad (۳-۱۵)$$

$$w \cdot x - b = -1 \quad (۳-۱۶)$$

نکته قابل توجه این است که اگر داده‌های تعلیمی به صورت خطی تفکیک پذیر باشند، می‌توان دو ابرسطح مرزی را به گونه‌ای انتخاب کرد که هیچ داده‌ای بین آن‌ها نباشد و سپس فاصله بین این دو ابرسطح موازی را به حداکثر رساند. با به کار گیری قضایای هندسی، فاصله این دو ابرسطح عبارت است از $2/|w|$. پس باید $|w|$ را به حداقل رساند. همچنین باید از قرار گرفتن نقاط داده در ناحیه درون مرز جلوگیری کرد. برای این کار یک محدودیت ریاضی به تعریف اضافه می‌شود که با رابطه‌های (۳-۱۷) و (۳-۱۸) تعریف می‌شوند، اطمینان حاصل می‌شود که هیچ نقطه‌ای در مرز قرار نمی‌گیرد.

$$w \cdot x_i - b \geq 1 \quad (۳-۱۷)$$

$$w \cdot x_i - b \leq -1 \quad (۳-۱۸)$$

می‌توان صورت کلی این محدودیت‌ها را به صورت رابطه (۳-۱۹) نشان داد [۵۳].

$$c_i(w \cdot x_i - b) \geq 1, \quad 1 \leq i \leq n \quad (۳-۱۹)$$

۱-۱-۳ روش آدابوست

۳-۳-۱ طبقه‌بندی کننده‌های بهبود یافته

طبقه‌بندی کننده بهبود یافته از طریق ترکیب طبقه بندی کننده ای ضعیف با یکدیگر یک طبقه

بندی کننده قوی را ایجاد می کند. اگر h_1, h_2, \dots, h_T یک دسته از فرضیات^۱ به دست آمده از طبقه بندی کننده های ضعیف باشد، ترکیب این نتایج در یک فرضیه قوی به صورت رابطه (۳-۲۱) ارائه می شود.

$$H(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (۳-۲۱)$$

که α_t ، وزن فرضیه های ضعیف را نشان می دهد و این در طول توالی بهبود منسوب می شود. آدابوست (بهبود تطبیقی) اولین و عمومی ترین الگوریتم استفاده کننده از طبقه بندی کننده های بهبود یافته برای طبقه بندی دوتایی است. آدابوست برای تشخیص چهره، شناسایی انسان و بسیاری از کاربردهای دیگر مورد استفاده قرار گرفته است [۲۱].

۳-۶-۴ آدابوست

آدابوست یک الگوریتم تکراری است. ایده اصلی این است که طبقه بندی کننده های مختلف ضعیف با استفاده از یک مجموعه داده یکسان آموزش داده می شوند و سپس برای ایجاد یک طبقه بندی کننده قوی تر ترکیب می شوند. در طبقه بندی آدابوست، زمانی که داده های آموزش در حال طبقه بندی شدن هستند، بر اساس نرخ صحت به هر طبقه بندی ضعیف وزنی منسوب می شود و به علاوه، به هر نمونه آموزشی وزنی منسوب می شود که نشان می دهد آیا این به وسیله طبقه بندی کننده کنونی درست طبقه بندی شده است یا نه. در طول توالی اضافه شدن یک طبقه بندی کننده ضعیف، اگر هر یک از نمونه های آموزش به درستی طبقه بندی شود وزنش کاهش می یابد. در غیر این صورت، وزن آن افزایش خواهد یافت، بنابراین طبقه بندی کننده در تکرار بعد می تواند بر روی نمونه اشتباه طبقه بندی شده تمرکز کند [۵۷].

الگوریتم روش طبقه بندی آدابوست به صورت زیر است:

^۱ hypothesis

وارد کردن داده‌های آموزش $D = \{(x_1, y_1), \dots, (x_m, y_m)\}$ که بردار ویژگی‌ها، y_i کلاس داده‌ها و m تعداد داده‌های آموزش است.

$$w_1(x_i) = 1/m \text{ مقداردهی اولیه به بردار وزن}$$

پیدا کردن T تا طبقه‌بندی کننده ضعیف $h_t (t = 1, 2, \dots, T)$ و مقداردهی اولیه $t=0$

$$P_t(x_i) = \frac{w_t(x_i)}{\sum_{i=1}^m w_t(x_i)} \text{ الف) تعیین}$$

ب) محاسبه نرخ خطای طبقه‌بندی کننده h_t :

$$\varepsilon_t = \sum P_t(x_i) [h_t(x_i) \neq y_i]$$

اگر $\varepsilon_t > 0.5$ بود $T = t - 1$ و از چرخه خارج شو.

$$\alpha_t = \log\left(\frac{\varepsilon_t}{1 - \varepsilon_t}\right) : h_t \text{ محاسبه وزن}$$

$$w_{t+1}(x_i) = w_t(x_i) \times \begin{cases} e^{-\alpha_t} \\ e^{\alpha_t} \end{cases} : \text{د) اصلاح کردن وزن تمام داده‌های آموزش از } i \text{ تا } m$$

$e^{-\alpha_t}$ برای داده‌هایی که صحیح طبقه‌بندی شده‌اند

e^{α_t} برای داده‌هایی که اشتباه طبقه‌بندی شده‌اند.

ه) اگر $\varepsilon_t > 0.001$ و $t < T$ ، بازگشت به مرحله الف

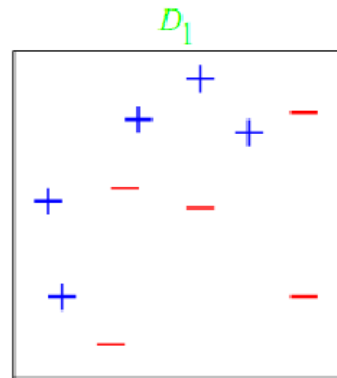
خروجی یک طبقه‌بندی کننده قوی است که از ترکیب طبقه‌بندی کنندگان ضعیف وزن دار تشکیل

$$H(x) = \arg \max_{1 \leq j \leq J} \sum_{t=1}^T a_t [h_t(x) = y] \text{ می‌شود:}$$

به منظور درک بهتر مفهوم روش کار طبقه‌بندی کننده آدابوست، در ادامه مثالی در قالب شکل

آورده شده است. در شکل ۳-۱۲ مجموعه کل داده‌ها به نمایش در آمده است. این داده‌ها که هر کدام

از آن‌ها دارای دو ویژگی می‌باشد، شامل دو کلاس هستند. یک کلاس توسط + و کلاس دیگر توسط - نشان داده شده است. در ابتدا تمامی این داده‌ها دارای وزنی یکسان D_1 هستند.



شکل ۳-۱۲. وزن‌دهی یکسان اولیه به تمام نمونه‌ها.

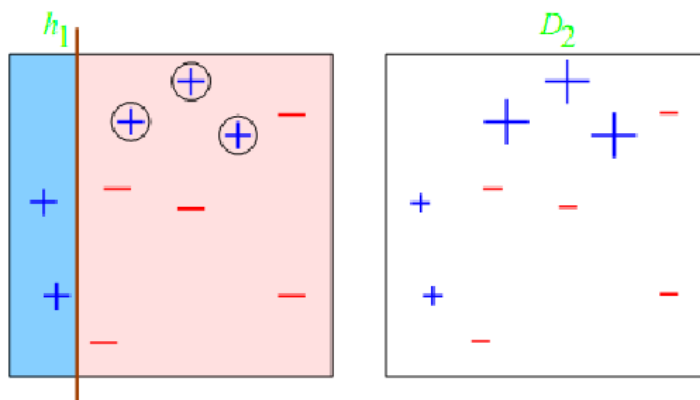
هر مرحله از بهبود شامل دوگام می‌باشد، گام اول پیدا کردن بهترین فرضیه ضعیف بر پایه خطای کمتر است. گام دوم، به روز رسانی وزن نمونه در هر مرحله بهبود است.

مرحله اول بهبود

- در مرحله ۱، اولین طبقه‌بندی کننده ضعیف مورد استفاده قرار می‌گیرد و داده‌ها را به دو کلاس تقسیم می‌کند. در شکل ۳-۱۳ نحوه تقسیم‌بندی کلاس‌ها نشان داده شده است. خط h_1 فرضیه‌ای است که حاصل طبقه‌بندی کننده ضعیف شماره ۱ است. نمونه‌هایی که در سمت راست خط h_1 قرار می‌گیرند، به عنوان کلاس + شناخته می‌شوند. همان‌طور که مشاهده می‌شود، این طبقه‌بندی کننده ضعیف در مورد سه نمونه که متعلق به کلاس + هستند، اشتباه عمل کرده و آن‌ها را به عنوان کلاس - تشخیص داده است. این نمونه‌ها توسط دایره‌هایی که دورشان کشیده شده است بر روی شکل نشان داده شده‌اند. مقدار خطای این فرضیه $\alpha_1 = \ln \frac{(1-\varepsilon_1)}{\varepsilon_1} = 0.42$ است. سپس وزن فرضیه ضعیف به صورت

می‌شود. در انتهای این مرحله، وزن‌های مربوط به نمونه‌ها به روز رسانی می‌شوند. وزن

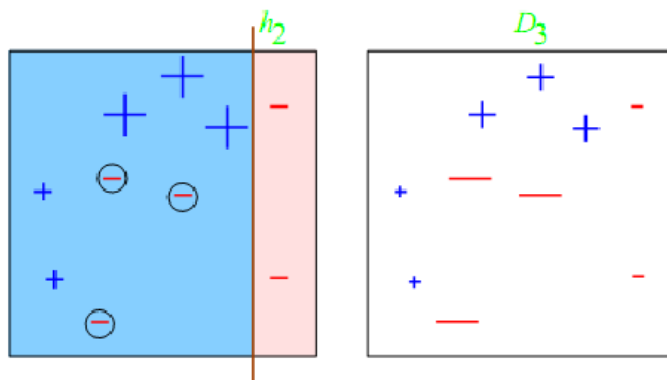
نمونه‌هایی که درست تشخیص داده شده‌اند نسبت به حالت اولیه خود کاهش می‌یابد و وزن نمونه‌هایی که اشتباه تشخیص داده شده‌اند، افزایش می‌یابد. در شکل ۳-۱۳-ب) وزن‌های به روز رسانی شده D_2 نشان داده شده است. اندازه علامت‌ها نشان‌دهنده بزرگی وزن آنهاست. هر چه علامت بزرگتر باشد دارای وزن بیشتری است.



شکل ۳-۱۳. مرحله یک بهبود. در این مرحله بهترین فرضیه ضعیف h_1 می‌باشد که دارای خطای $\varepsilon = 0.30$ و وزن $\alpha_1 = 0.42$ می‌باشد.

مرحله دوم بهبود

- در مرحله دوم، فرضیه ضعیف h_2 انتخاب می‌شود. همان‌طور که در شکل ۳-۱۴ نشان داده شده است، بر اساس این فرضیه، نمونه‌های سمت راست خط h_2 متعلق به کلاس - و نمونه‌های سمت چپ آن متعلق به کلاس + می‌باشند. با اعمال فرضیه h_2 بر روی داده‌ها، خطای آن $\varepsilon_2 = 0.21$ محاسبه می‌شود. وزن فرضیه ضعیف h_2 به صورت $\alpha_2 = 0.65$ محاسبه شده است. در پایان این مرحله، وزن داده‌ها برای ورود به مرحله بعد به روز رسانی می‌شود و به صورت D_3 نشان داده می‌شود.

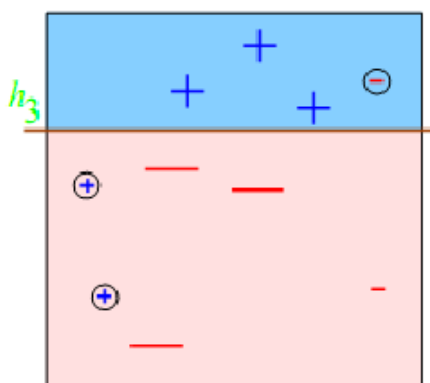


شکل ۳-۱۴. مرحله دو بهبود. در این مرحله بهترین فرضیه ضعیف h_2 به دست می‌آید. خطای طبقه‌بندی این

فرضیه ضعیف $\epsilon_2 = 0.21$ و وزن آن $\alpha_2 = 0.65$ می‌باشد.

بهبود مرحله سوم

در این مرحله فرضیه ضعیف h_3 که یک خط افقی است به منظور طبقه‌بندی مورد استفاده قرار گرفته است. فرضیه ضعیف دارای خطای $\epsilon_3 = 0.14$ است. وزن مربوط به این فرضیه محاسبه نیز به صورت $\epsilon_3 = 0.92$ محاسبه می‌شود.



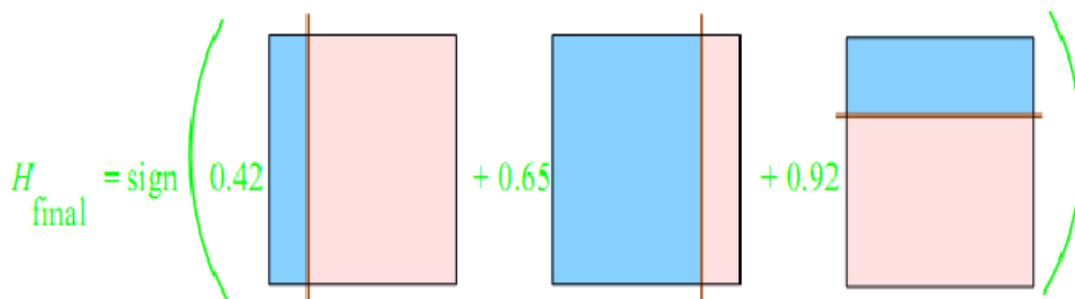
شکل ۳-۱۵. مرحله سه بهبود. بهترین فرضیه با خطای $\epsilon_3 = 0.14$ و وزن $\epsilon_3 = 0.92$ به دست می‌آید.

فرضیه قوی نهایی

در نهایت بهترین فرضیه‌های ضعیف h_i و وزن‌های مربوط به آن‌ها مشخص می‌گردند. فرضیه نهایی قوی ترکیبی از فرضیه‌های ضعیف با وزن آن‌هاست؛ که به صورت رابطه (۳-۲۲) محاسبه می‌شود و در

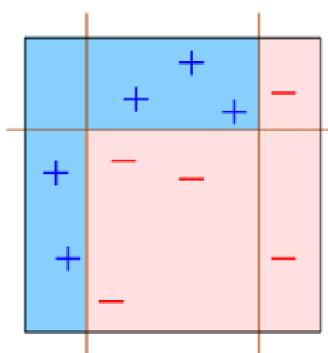
شکل ۳-۱۶ به صورت شماتیک نشان داده شده است.

$$H_{final} = \text{sign}(0.42 * h_1 + 0.65 * h_2 + 0.92 * h_3) \quad (3-22)$$



شکل ۳-۱۶. نحوه ترکیب فرضیه‌های ضعیف به همراه وزن‌های مربوط به آن‌ها که طی مراحل مختلف بهبود به دست می‌آیند.

در شکل ۳-۱۷ نحوه تقسیم‌بندی نهایی با H_{final} نشان داده شده است. ناحیه‌ای که با رنگ آبی نمایش داده شده است مربوط به کلاس + و ناحیه صورتی رنگ مربوط به کلاس - می‌باشد. بنابراین هر نمونه‌ای که در ناحیه آبی قرار گیرد به عنوان کلاس + و هر نمونه‌ای که در ناحیه صورتی رنگ قرار گیرد به عنوان کلاس - شناخته خواهد شد.



شکل ۳-۱۷. نحوه ترکیب فرضیه‌های ضعیف به همراه وزن‌های مربوط به آن‌ها که طی مراحل مختلف بهبود به دست می‌آیند.

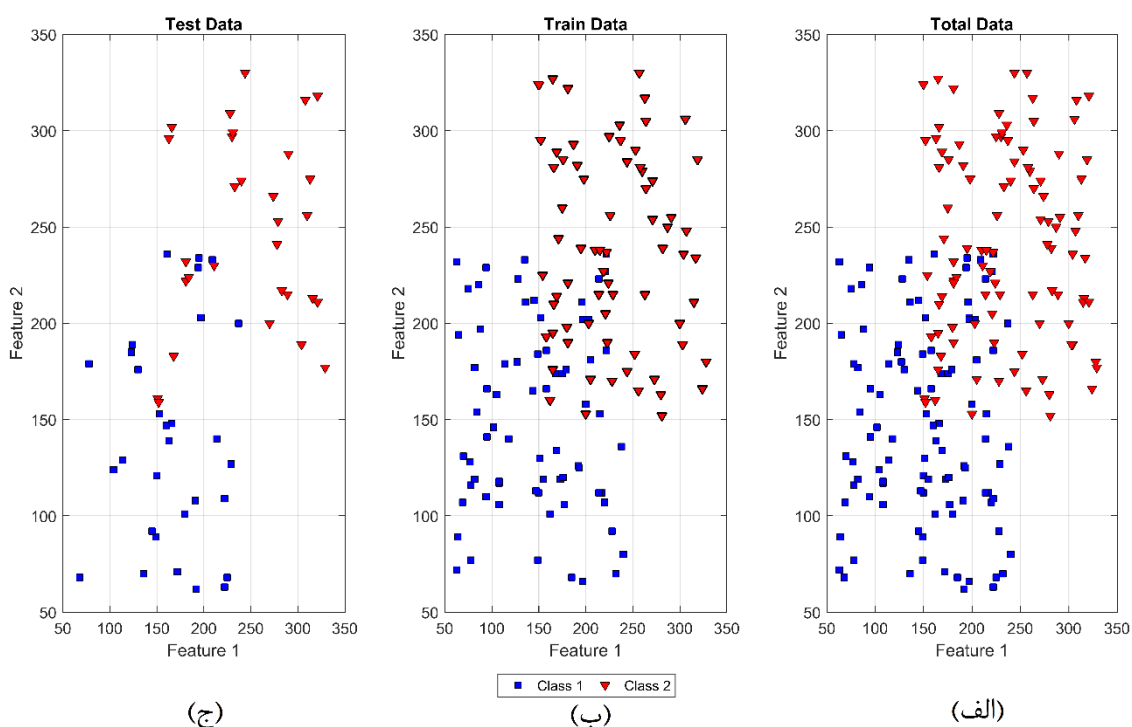
۷-۳ مقایسه دقت روش‌های knn, SVM و آدابوست

به منظور مقایسه دقت سه روش طبقه‌بندی معروف knn, SVM و آدابوست، نتایج آن‌ها برای

چندین مثال مصنوعی مورد بررسی و مقایسه قرار می‌گیرد.

۳-۷-۱ مدل مصنوعی دوبعدی با دو کلاس

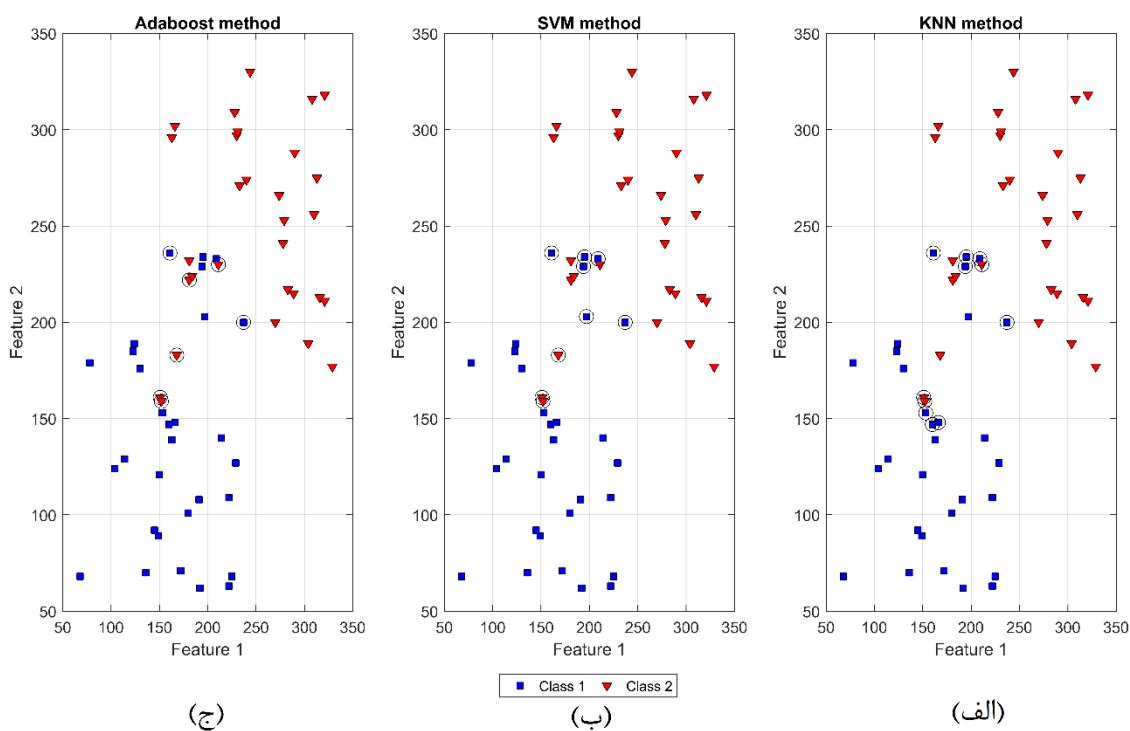
این مثال شامل دو کلاس و در فضای دوبعدی می‌باشد. بردار ویژگی کلاس یک شامل ۱۰۰ نمونه است که در هر دو راستای افقی و عمودی به صورت تصادفی از بازه (50,250) انتخاب شده‌اند و در شکل ۳-۱۷ بوسیله مربع‌های آبی رنگ نشان داده شده‌اند. بردار ویژگی کلاس دو نیز شامل ۱۰۰ نمونه است که در هر دو راستای افقی و عمودی به صورت تصادفی از بازه (150,330) انتخاب شده‌اند. این کلاس در شکل ۳-۱۷ با مثلث‌های قرمز رنگ نشان داده شده‌اند. در واقع در شکل ۳-۱۷ کل داده‌های مورد استفاده در این مثال نشان داده شده است.



شکل ۳-۱۷. کل داده‌های مورد استفاده برای مدل مصنوعی دوبعدی با دو کلاس (الف) کل داده‌های مورد استفاده از

مثال (ب) داده‌های مورد استفاده در مرحله آموزش (ج) داده‌های مورد استفاده در مرحله آزمایش.

پس از آموزش روش‌های SVM، knn و آدابوست با استفاده از داده‌های آموزش، الگوریتم‌های یاد شده بر روی داده‌های آزمایش اعمال می‌شوند و کلاس‌های تعیین شده توسط این روش‌ها با کلاس‌های واقعی آن‌ها مقایسه می‌گردد. با شمارش تعداد کلاس‌هایی که درست تعیین شده‌اند و تقسیم تعداد آن‌ها بر تعداد کل داده‌های آزمایش مقدار دقت این روش‌ها به دست می‌آید. در شکل ۱۸-۳ نحوه عملکرد طبقه‌بندی کننده‌های SVM، knn و آدابوست نشان داده شده است. نمونه‌هایی که کلاس آن‌ها اشتباه تشخیص داده شده است با دایره‌هایی مشخص شده‌اند.

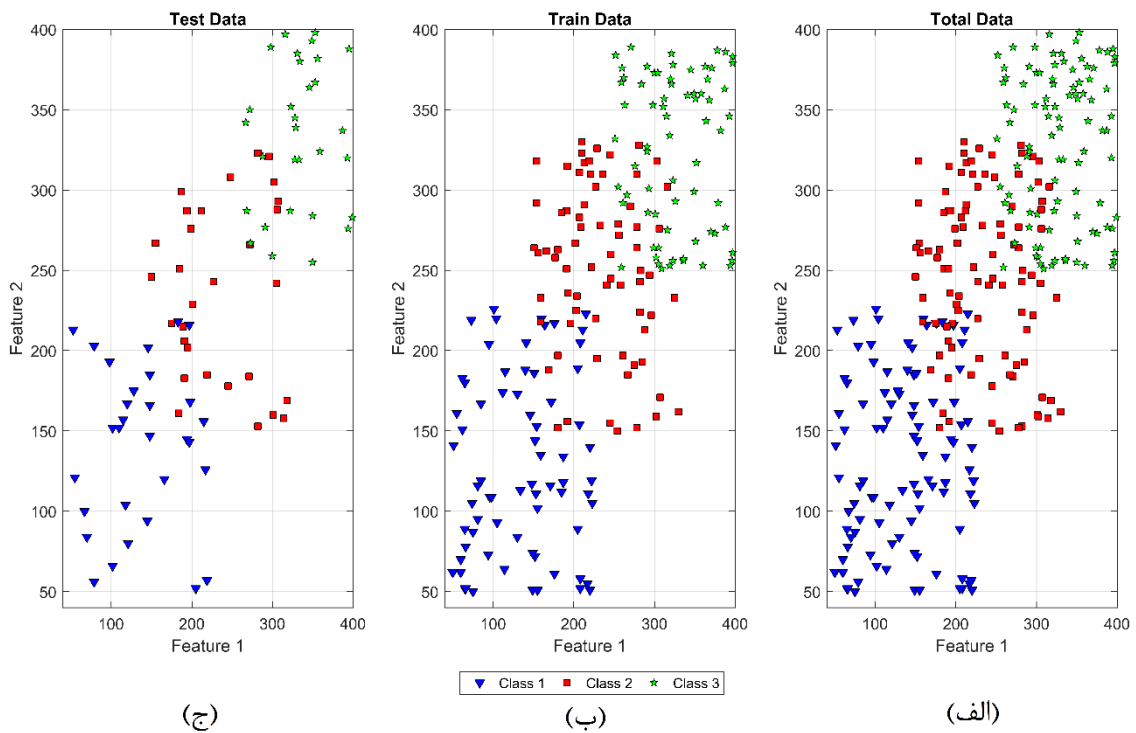


شکل ۱۸-۳. نحوه طبقه‌بندی داده‌های آموزش توسط روش‌های الف) KNN، ب) SVM و ج) آدابوست.

همانطور که مشاهده می‌شود تعداد ۱۱ نمونه توسط روش knn، تعداد ۹ نمونه توسط روش SVM و تعداد ۷ نمونه توسط روش آدابوست خطا طبقه‌بندی شده‌اند. نتایج به دست آمده از روش‌های ذکر شده بدین صورت می‌باشد: دقت روش knn برابر ۸۱ درصد، دقت روش SVM برابر ۸۴ درصد و دقت روش آدابوست برابر با ۹۱/۲ درصد می‌باشد. همان‌طور که از نتایج پیدا ست روش آدابوست دارای بالاترین مقدار دقت است. پس آن روش SVM در رده دوم و روش knn در آخر قرار می‌گیرد.

۳-۷-۲ مدل مصنوعی دو بعدی با سه کلاس

این مثال شامل دو کلاس با سه ویژگی است. کلاس یک دارای ۱۰۰ نمونه است و ویژگی‌های آن در راستای هر سه بعد در بازه (70,230) می‌باشد. در شکل ۳-۱۹ این کلاس بو سیله مثلث‌های آبی رنگ نشان داده شده است. کلاس دو نیز دارای ۱۰۰ نمونه است و ویژگی‌های آن در راستای هر سه بعد در بازه (150,330) می‌باشد. در شکل ۳-۱۹ این کلاس بوسیله مربع‌های قرمز رنگ نشان داده شده است. کلاس سه نیز دارای ۱۰۰ نمونه است و ویژگی‌های آن در راستای هر سه بعد در بازه (250,400) می‌باشد. در شکل ۳-۱۹ این کلاس بو سیله ستاره‌های سبز رنگ نشان داده شده است. در این مثال نیز ۷۰ درصد از داده‌های کل به صورت تصادفی برداشته می‌شود و با استفاده از آن آموزش انجام گرفته و سپس بر روی داده‌های آزمایش (۳۰ درصد باقیمانده) اعمال می‌گردد.



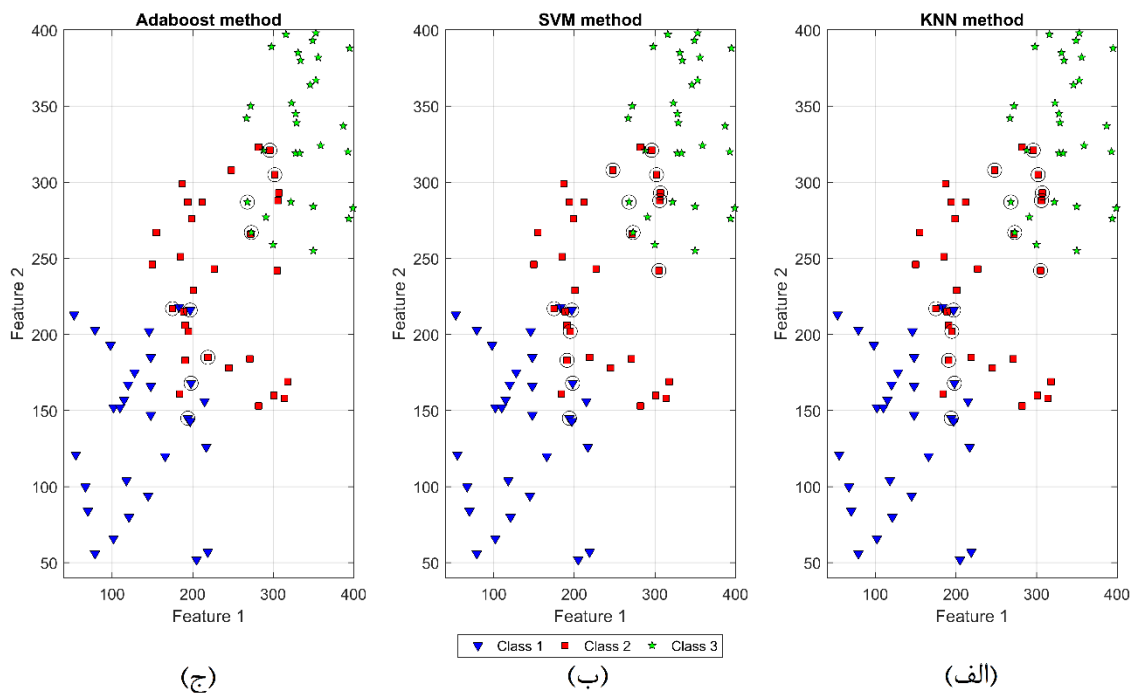
شکل ۳-۱۹. داده‌های مربوط به مدل مصنوعی دو کلاسه و سه بعدی. داده‌های مورد استفاده در بخش آموزش

(الف) و داده‌های مورد استفاده برای آزمایش (ب).

در شکل ۳-۱۹ نحوه عملکرد طبقه بندی کننده های knn، SVM و آدابوست نشان داده شده است.

نمونه هایی که کلاس آنها اشتباه تشخیص داده شده است با دایره هایی مشخص شده اند. همانطور که مشاهده می شود تعداد ۱۳ نمونه توسط روش knn، تعداد ۱۳ نمونه توسط روش SVM و تعداد ۹ نمونه توسط روش آدابوست خطا طبقه بندی شده اند.

دقت روش knn و SVM برابر ۸۵ درصد و دقت روش آدابوست برابر با ۹۰ درصد می باشد.

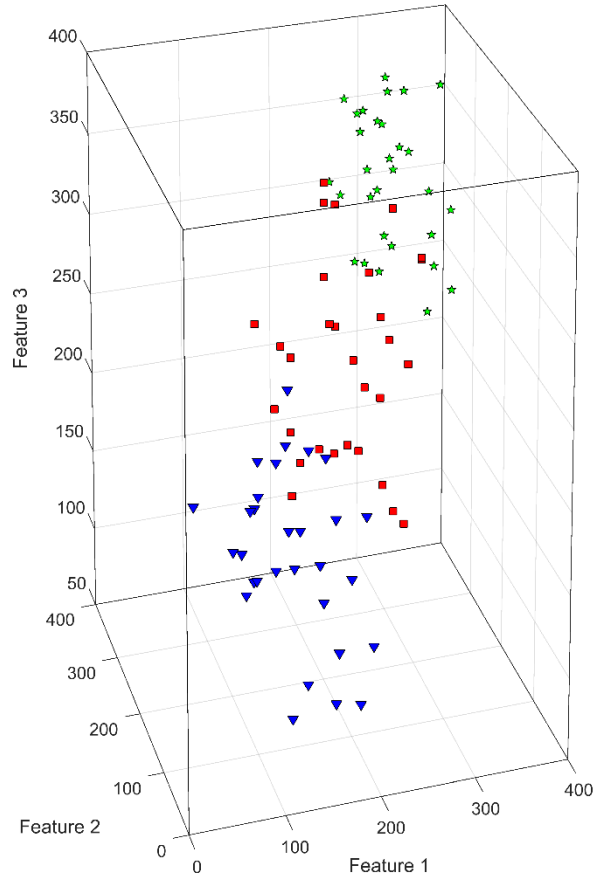


شکل ۳-۲۰. نحوه طبقه بندی داده های آموزش سه کلاسه دویبعی توسط روش های الف) KNN، ب) SVM و ج) آدابوست.

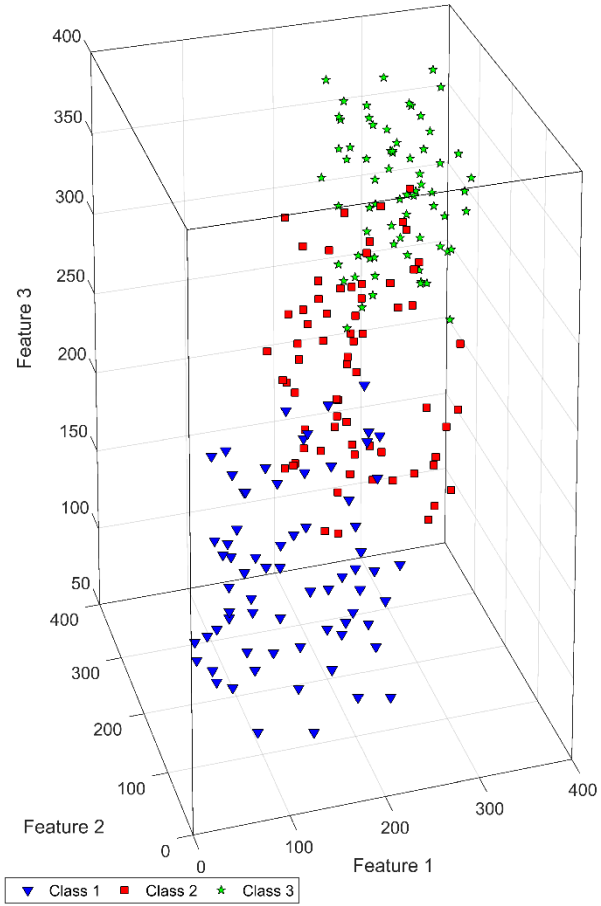
۳-۷-۳ مدل مصنوعی سه بعدی با سه کلاس

این مثال شامل سه کلاس با سه ویژگی است. کلاس یک دارای ۱۰۰ نمونه است و ویژگی های آن در راستای هر سه بعد در بازه (60,240) می باشد. در شکل ۳-۲۰ این کلاس بوسیله مثلث های آبی رنگ نشان داده شده است. کلاس دو نیز دارای ۱۰۰ نمونه است و ویژگی های آن در راستای هر سه بعد در بازه (150,330) می باشد. در شکل ۳-۲۰ این کلاس بصورت مربع های قرمز رنگ نشان داده شده است. کلاس سه دارای ۱۰۰ نمونه است و ویژگی های آن در راستای هر سه بعد در بازه

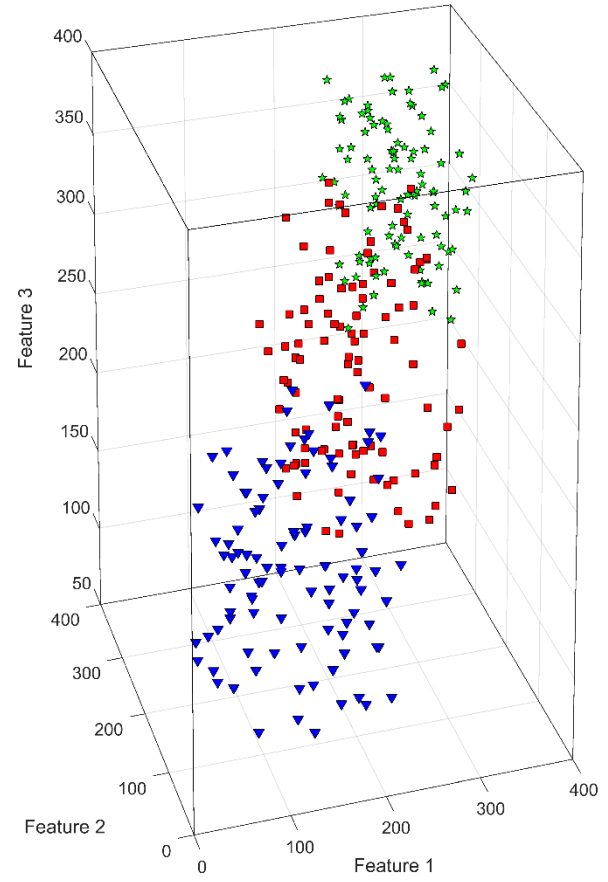
(240,380) می‌باشد. در شکل ۳-۲۰ این کلاس بصورت ستاره های سبز رنگ نشان داده شده است. در این مثال ۷۰ درصد از داده‌های کل به صورت تصادفی برای آموزش الگوریتم و ۳۰ درصد باقیمانده برای آزمایش استفاده می‌شود. در شکل ۳-۲۱ نحوه عملکرد طبقه بندی کننده های SVM, knn و آدابوست نشان داده شده است. نمونه هایی که کلاس آنها اشتباه تشخیص داده شده است با دایره هایی مشخص شده اند. همانطور که مشاهده می شود تعداد ۱۸ نمونه توسط روش knn، تعداد ۱۴ نمونه توسط روش SVM و تعداد ۸ نمونه توسط روش آدابوست خطا طبقه بندی شده اند. دقت روش knn برابر ۸۰ درصد، دقت روش SVM برابر ۸۵ درصد و دقت روش آدابوست برابر با ۹۱ درصد می‌باشد.



(ج)

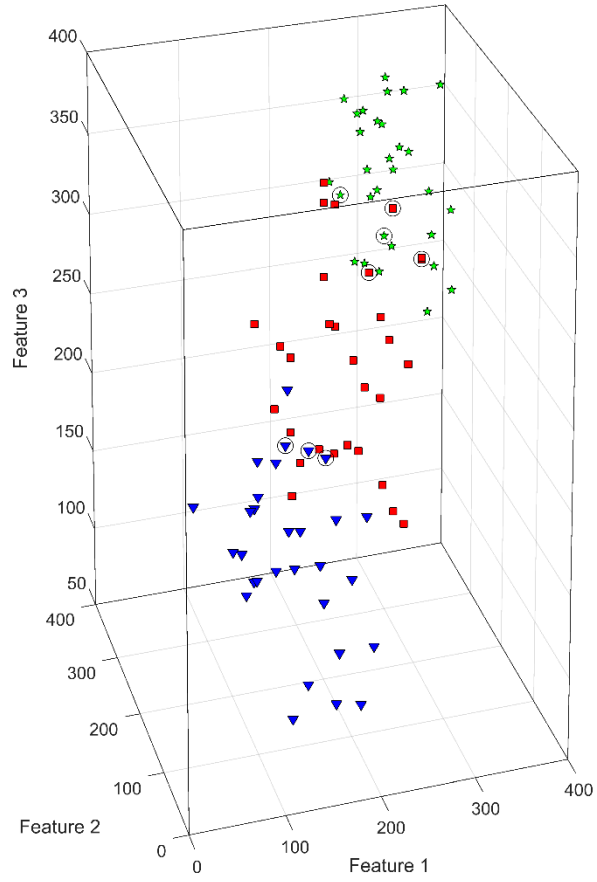


(ب)

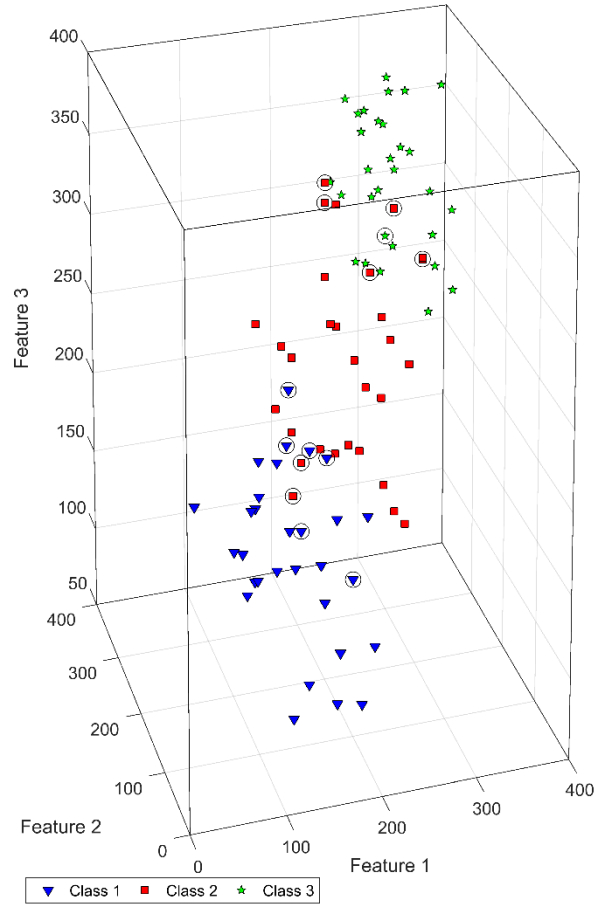


(الف)

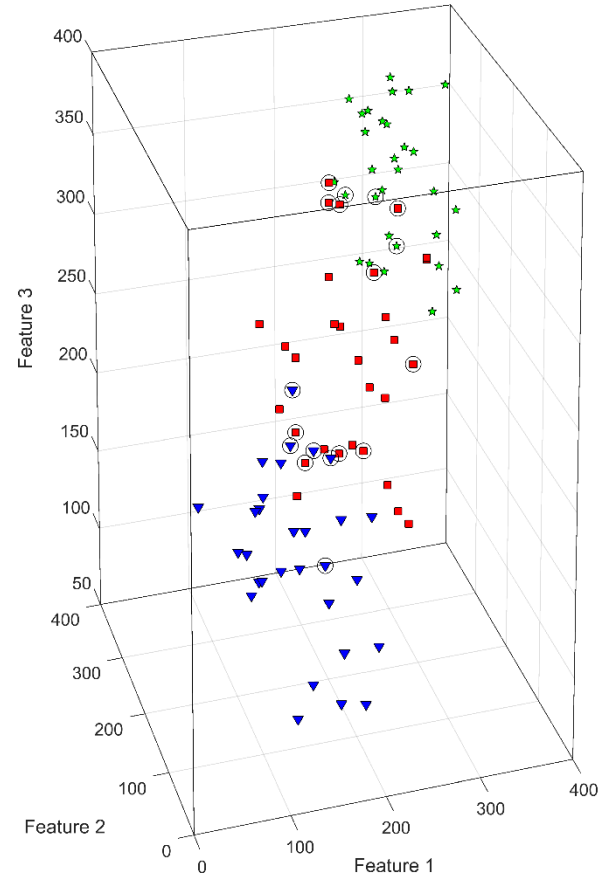
شکل ۳-۲۱. داده‌های مربوط به مدل مصنوعی سه بعدی و سه کلاسه. الف) کل داده‌ها، ب) داده‌های آموزشی، ج) داده‌های آزمایشی.



(ج)



(ب)



(الف)

شکل ۳-۲۲. نحوه طبقه بندی داده‌های آموزشی سه کلاسه سه بعدی توسط روشهای الف) KNN، ب) SVM و ج) آداپوست.

فصل چهارم

اعمال روش بر روی داده‌های واقعی

۱-۴ مقدمه

در این پایان نامه از داده‌های میدان بلک فوت^۱ استفاده شده است. داده‌های این میدان همراه نرم‌افزار همپسون-راسل در اختیار کاربر قرار می‌گیرد.

میدان نفتی بلک فوت، در جنوب شرقی شهر کوچک استراتمور^۲ در آلبرتا^۳ واقع شده است. مخزن این میدان یک ماسه سنگ گلوکونیتی رودخانه‌ای/دلتایی است که متعلق به کرتاسه پایینی بوده و در عمق حدود ۱۵۵۰ متری قرار گرفته است. نفت در تله‌های چینه‌ای متشکل از ماسه سنگ‌های کوارتز دار در درون کانال‌های عضو گلوکونیتی در منطقه بلک فوت به دام افتاده‌اند. این تله‌های چینه‌ای معمولاً توسط ماسه سنگ‌های ناتراوا یا شیل‌ها پوشیده شده‌اند و می‌توانند ضخامتی تا ۴۵ متر داشته باشند. در شکل ۱-۴ موقعیت جغرافیایی میدان بلک فوت نشان داده شده است.



شکل ۱-۴: موقعیت جغرافیایی میدان بلک فوت، کانادا.

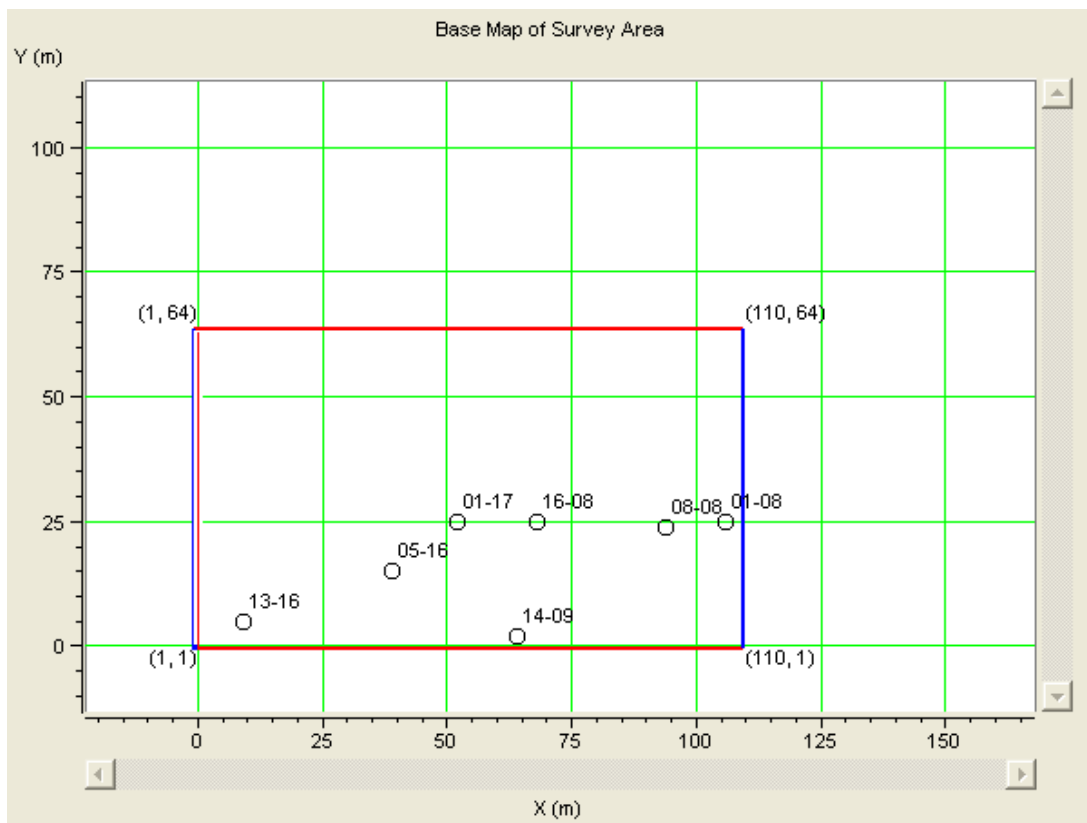
^۱ Blackfoot

^۲ Strathmore

^۳ Alberta

۱-۱-۴ داده‌های میدان بلک فوت

از میان چاه‌های حفر شده به منظور مطالعات مختلف بر روی میدان بلک فوت، پایگاه داده‌ی پیش ساخته‌ی بخش EMEARGE نرم‌افزار همپسون-راسل برای تخلخل مورد استفاده قرار گرفته است. در شکل ۲-۴ موقعیت چاه‌های حفر شده نسبت به محدوده برداشت داده‌های لرزه‌نگاری سه بعدی نشان داده شده است.



شکل ۲-۴: موقعیت چاه‌های حفر شده نسبت به محدوده برداشت داده‌های لرزه‌نگاری سه بعدی.

داده‌های این چاه شامل اطلاعات نگار چگالی، تخلخل به دست آمده از چگالی، موج P (با قدرت تفکیک ۲۰ سانتی‌متر) و نقاط کنترلی^۱ هستند. در شکل ۳-۴ این نگارها نشان داده شده اند. در جدول ۱-۴ می‌توان رأس سازندهای مختلف را در هر چاه مشاهده کرد. داده‌های سه‌بعدی برداشت شده در این میدان، مورد نیاز در روند پایان‌نامه نیز، همراه نرم‌افزار همپسون-راسل وجود دارد. این

^۱Check shot

داده‌ها در سال ۱۹۹۵ و به منظور شناسایی عضو گلوکونیتی گروه منویل^۱ برداشت شده است (Russell, 2004). در شکل ۴-۴ داده‌های لرزه‌نگاری سه بعدی مورد نظر نشان داده شده است. این داده‌ها شامل ۱۱۰ خط گیرنده و ۶۴ خط چشمه می‌باشند و با گام ۲ میلی‌ثانیه نمونه‌برداری زمانی شده‌اند.

جدول ۴-۱ رأس سازندهای مشاهده شده در چاه‌های میدان بلک فوت.

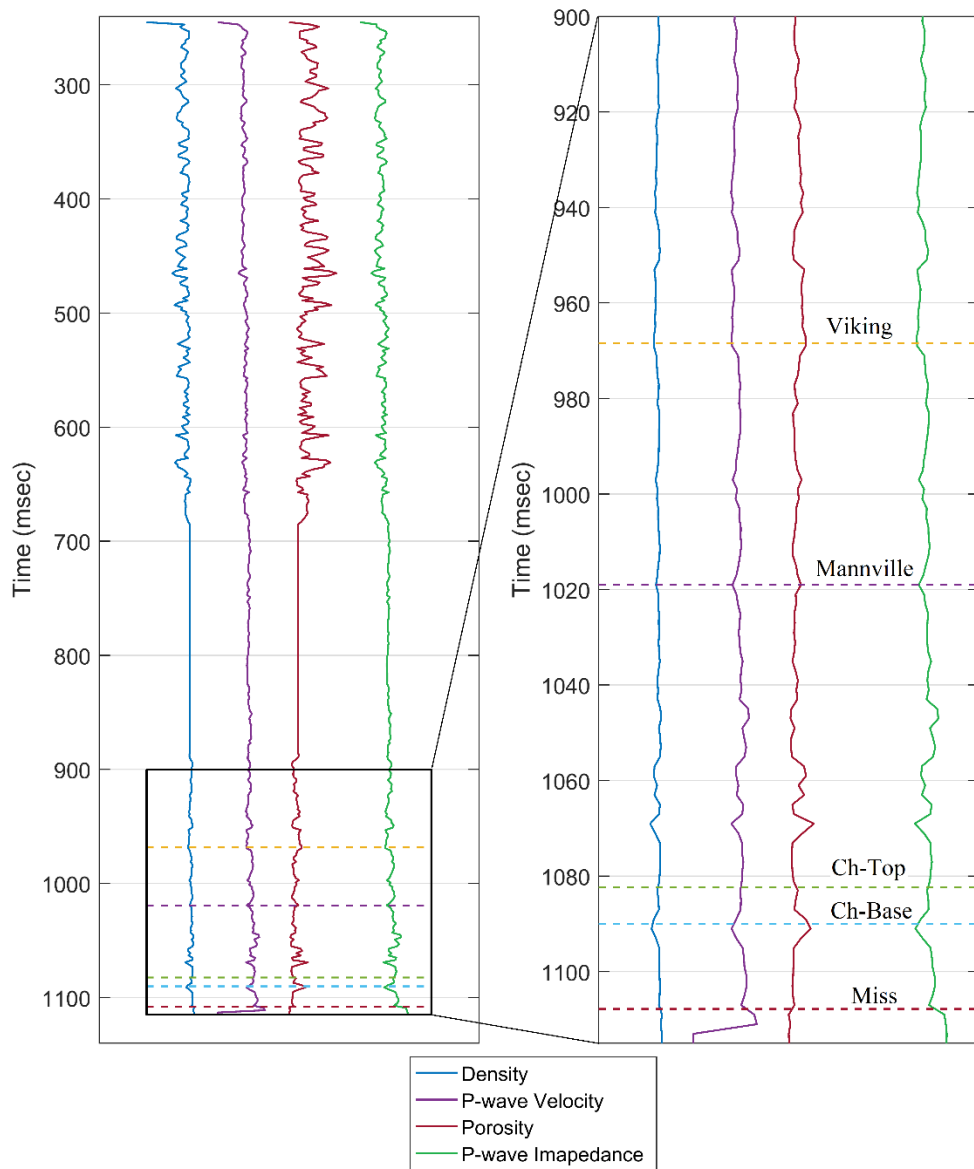
رأس سازند میس ^۴ (m)	رأس سازند ch-top (m)	رأس سازند منویل ^۳ (m)	رأس سازند وایکینگ ^۲ (m)	نام چاه
۱۶۰۰	۱۵۴۵	۱۴۲۴	۱۳۲۸	۰۱-۰۸
۱۶۱۰	۱۵۵۷	۱۴۳۵	۱۳۴۱	۰۱-۱۷
۱۶۳۹	۱۵۶۶	۱۴۴۷	۱۳۵۲	۰۵-۱۶
۱۶۱۲	۱۵۵۲	--	۱۳۳۵	۰۸-۰۸
۱۶۳۴	۱۵۸۶	۱۴۶۷	۱۳۷۵	۱۳-۱۶
۱۶۰۰	--	--	۱۳۴۰	۱۴-۰۹
۱۶۰۰	۱۵۵۱	۱۴۳۳	۱۳۴۰	۱۶-۰۸

^۱ mannville

^۲ Viking

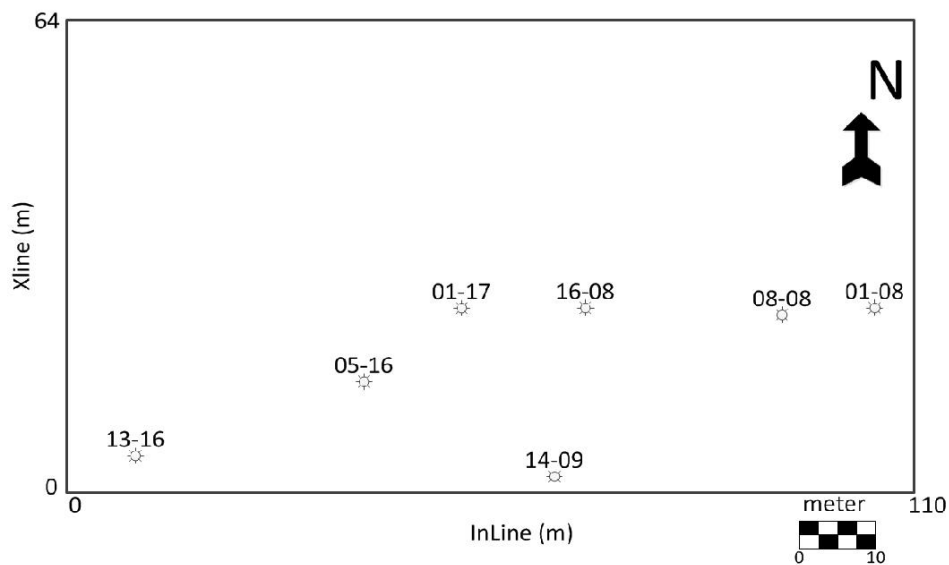
^۳ Mannville

^۴ Miss

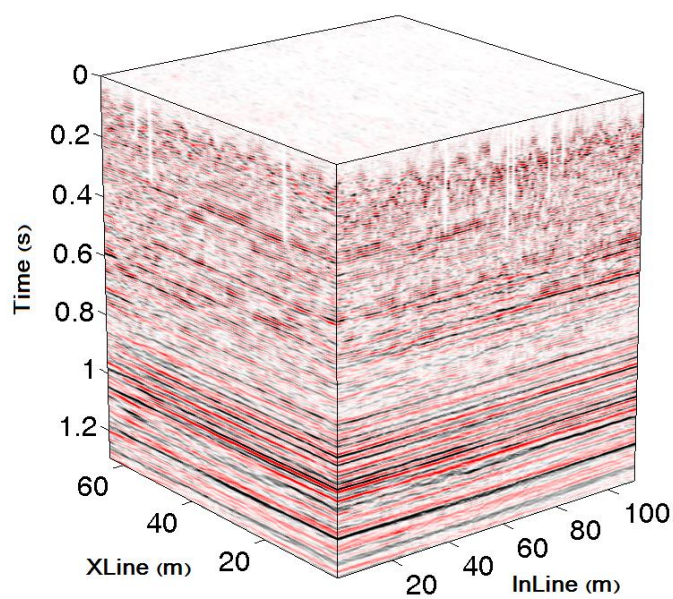


شکل ۴-۳. نگارهای موجود در چاه ۰۸-۰۸. به ترتیب از سمت چپ: نگار چگالی، سرعت موج p، نگار تخلخل و نگار

امپدانس موج p.



(الف)



(ب)

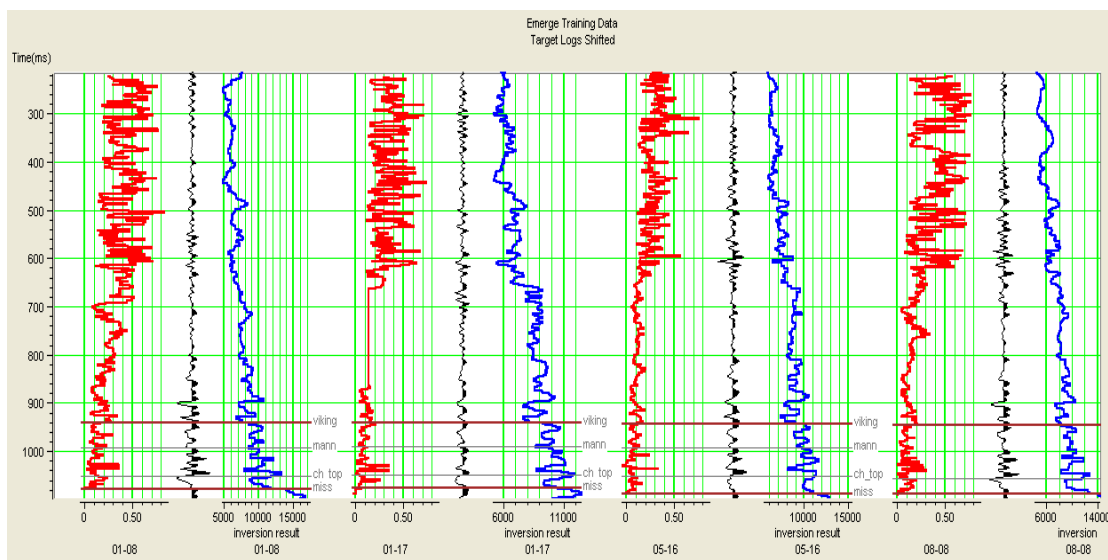
شکل ۴-۴. (الف) موقعیت خطوط لرزه‌نگاری و چاه‌های حفر شده در منطقه. (ب) مکعب داده‌های لرزه‌ای.

۴-۲ تخمین تخلخل به کمک نشانگرهای لرزه‌ای

در نرم‌افزار همپسون-را سل بخشی به نام EMERGE وجود دارد که با تلفیق داده‌های چاه‌نگاری و

نشانه‌های لرزه‌ای، تخلخل را به دست می‌آورد. به منظور مطالعه مخزن با استفاده از نشانه‌های لرزه‌ای، ابتدا داده‌های سه‌بعدی لرزه‌ای منطقه مورد مطالعه به همراه نتیجه وارون سازی به عنوان یک نشانهگر خارجی، در نرم‌افزار فراخوانی می‌شود. سپس باید با استفاده از روش اعتبار سنجی متقابل بهترین نشانه‌های لرزه‌ای و تعداد بهینه آن‌ها، برای تخمین تخلخل تعیین گردد.

در شکل ۴-۵ اطلاعات مورد استفاده در محل چاه‌ها برای این پایان‌نامه نمایش داده شده است. نمودارهای قرمز رنگ نگار تخلخل، نمودارهای سیاه رنگ داده‌های خام لرزه‌ای و نمودار آبی رنگ نشانهگر مقاومت صوتی است که به عنوان نشانهگر خارجی مورد استفاده قرار گرفته است.



شکل ۴-۵. نمایش داده‌های تخلخل، مقاومت ظاهری لرزه‌ای و ردلرزه در محل چاه‌ها.

در گام بعدی در محل چاه‌ها یک یا چند نشانهگر لرزه‌ای از داده‌های لرزه‌ای استخراج شده و نشانه‌گرهایی که در همان محل با پارامترهای مورد نظر تطابق بیشتری دارند، مشخص می‌شوند.

۴-۲-۱ تحلیل به کمک یک نشانهگر (رگرسیون معمولی)

در روش رگرسیون معمولی پس از ارزیابی تمامی نشانه‌گرها (اعم از داخلی و خارجی)، نشانهگری انتخاب می‌شود که در محل چاه بیشترین همبستگی را با پارامتر هدف (تخلخل) داشته باشد.

نتیجه بررسی و نتایج روش رگرسیون معمولی (یک نشانهگری) در جدول ۴-۲ نمایش داده شده

است. در این روش نشانگرهای مختلف مورد بررسی قرار گرفته و پنج تا از بهترین نتایج حاصله به صورت ارائه شده می‌باشد. همان‌طور که مشخص است از بین پنج نشانگر نشان داده شده، بهترین نتیجه مربوط به همبستگی بین نگار تخلخل و نشانگر خارجی (معکوس مقاومت ظاهری لرزه‌ای) می‌باشد؛ که مقدار آن ۴۳ درصد می‌باشد.

جدول ۴-۲. نتایج به کمک یک نشانگر (رگرسیون معمولی).

Target	Attribute	Error	Correlation
Porosity	1/Accoustic Impedanc	0.060357	0.435455
porosity	Accoustic Impedance	0.061049	-0.413454
porosity	Instantaneous phase	0.065669	-0.201794
porosity	Quadrature trace	0.065669	-0.201748
porosity	Amplitude Weighted Phase	0.065384	-0.221384

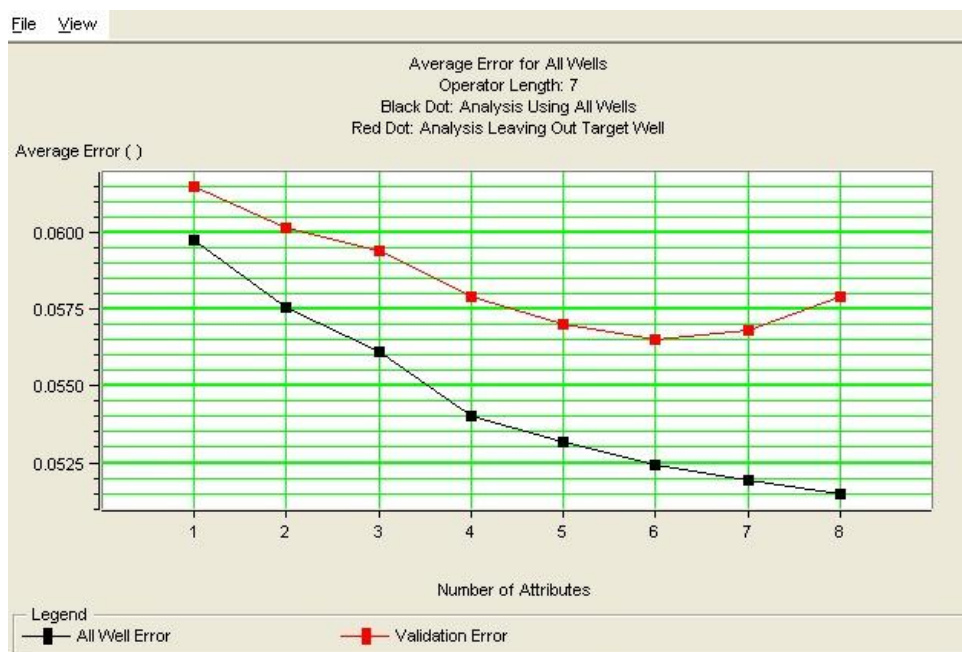
نتایج روش رگرسیون معمولی مناسب نمی‌باشد. لذا باید این روش کنار گذاشته می‌شود و از روش رگرسیون چند متغیره استفاده می‌گردد.

۴-۲-۲ تحلیل به کمک چند نشانگر (رگرسیون چند متغیره)

در روش رگرسیون چند متغیره (چند نشانگره) تعداد نشانگرهایی که مورد استفاده قرار می‌گیرند، دارای اهمیت است. در نرم‌افزار همپسون-راسل از روش‌های اعتبار سنجی متقابل و رگرسیون مرحله‌ای برای تعیین تعداد بهینه نشانگرها، استفاده می‌شود.

بنابراین نشانگرها به نرم‌افزار داده شده و به منظور بهینه کردن زمان محاسبات، روش اعتبار سنجی متقابل برای حداکثر ۸ نشانگر انجام می‌گردد که در شکل ۴-۶ نمایش داده شده است. در تصویر محور افقی نشان دهنده تعداد نشانگرها و محور عمودی نشان دهنده خطای میانگین است. نمودار سیاه رنگ خطای تخمین به ازای تمام چاه‌ها و نمودار قرمز رنگ (به عنوان نمودار اعتبار سنجی) نشان دهنده خطای تخمین برای تمام چاه‌ها به جز چاه هدف را نشان می‌دهد.

همان‌طور که از شکل پیداست، با افزایش تعداد نشانگرها میزان خطای تخمین و اعتبار سنجی هر دو کاهش می‌یابد. این روند تا رسیدن به تعداد ۶ نشانگر ادامه می‌یابد. اما استفاده از تعداد بیشتری از نشانگرها در تخمین تخلخل موجب افزایش خطای میانگین نمودار اعتبار سنجی می‌گردد، در حالی که خطای آموزشی همچنان کاهش می‌یابد. بنابراین در تعداد ۶ نشانگر فرآیند تخمین متوقف شده و عدد ۶ نشان دهنده تعداد بهینه می‌باشد. نشانگرهای مورد استفاده و نتایج تحلیل آن‌ها در جدول ۴-۳ آورده شده است.



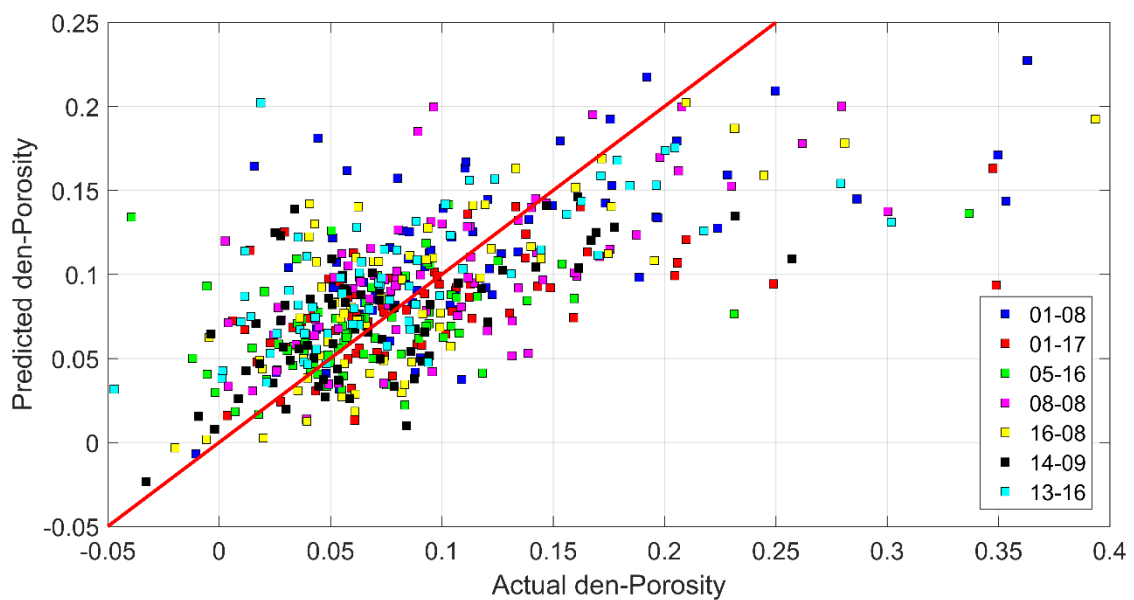
شکل ۴-۶. نتایج اعتبار سنجی متقابل چند نشانگر لرزه‌ای با استفاده از روش رگرسیون چند متغیره.

در شکل ۴-۷ نمودار مقاطع تخلخل واقعی و تخلخل محاسبه شده در محل چاه را نشان می‌دهد. در این شکل داده‌های مربوط به ۷ چاه مورد استفاده قرار گرفته است که داده‌های مربوط به هر چاه با رنگ متفاوتی نشان داده شده است. محور عمودی شکل تخلخل‌های محاسبه شده با استفاده از روش چند نشانگری، محور افقی مربوط به تخلخل واقعی مربوط به هر نمونه بر روی نگار هدف و خط قرمز رنگ نیز خط رگرسیون می‌باشد. همچنین در شکل ۴-۸ نگار تخلخل واقعی و محاسبه شده در محل چاه در کنار یکدیگر رسم شده‌اند. این شکل نشان می‌دهد که مقدار تطابق به دست آمده به کمک این روش ۶۲ درصد و خطایی معادل ۵ درصد دارد. نتیجه اعمال رابطه به دست آمده از طریق

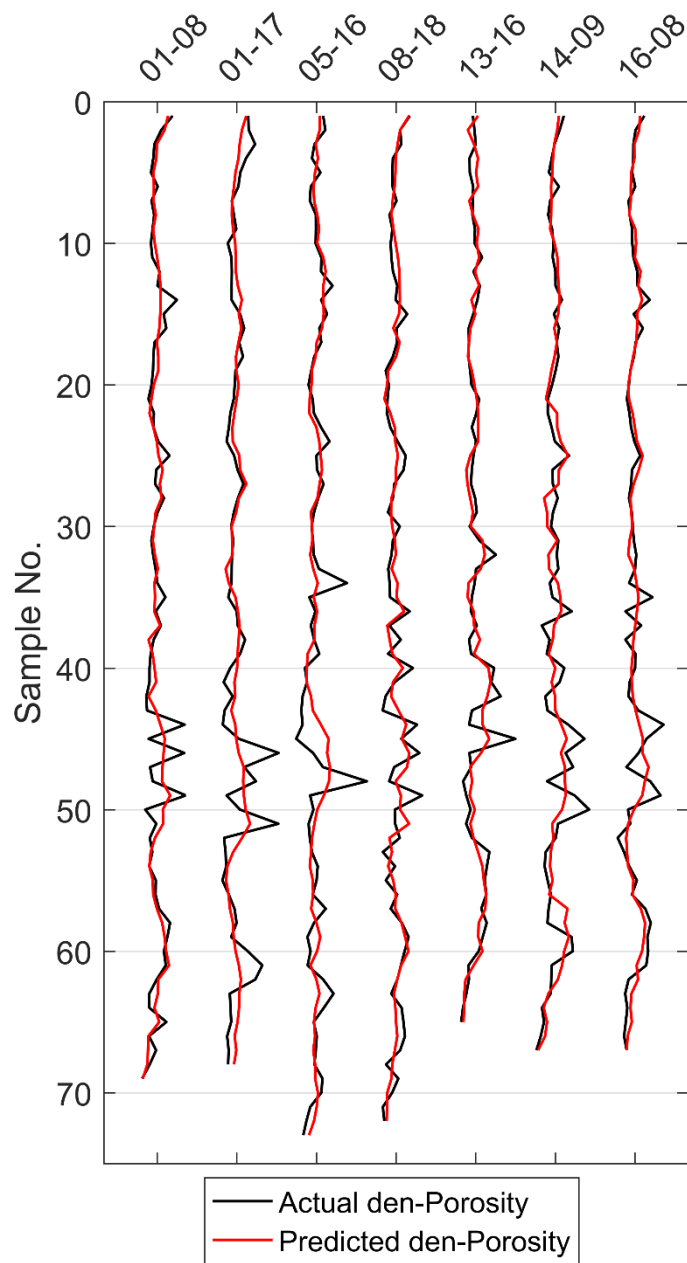
رگرسیون چند متغیره بر روی داده‌های لرزه‌ای و نشانگرهای آن در شکل ۴-۹ برای خط گیرنده ۹۵ رسم شده است.

جدول ۴-۳. نشانگرهای مورد استفاده در روش رگرسیون چند متغیره و نتایج آن‌ها.

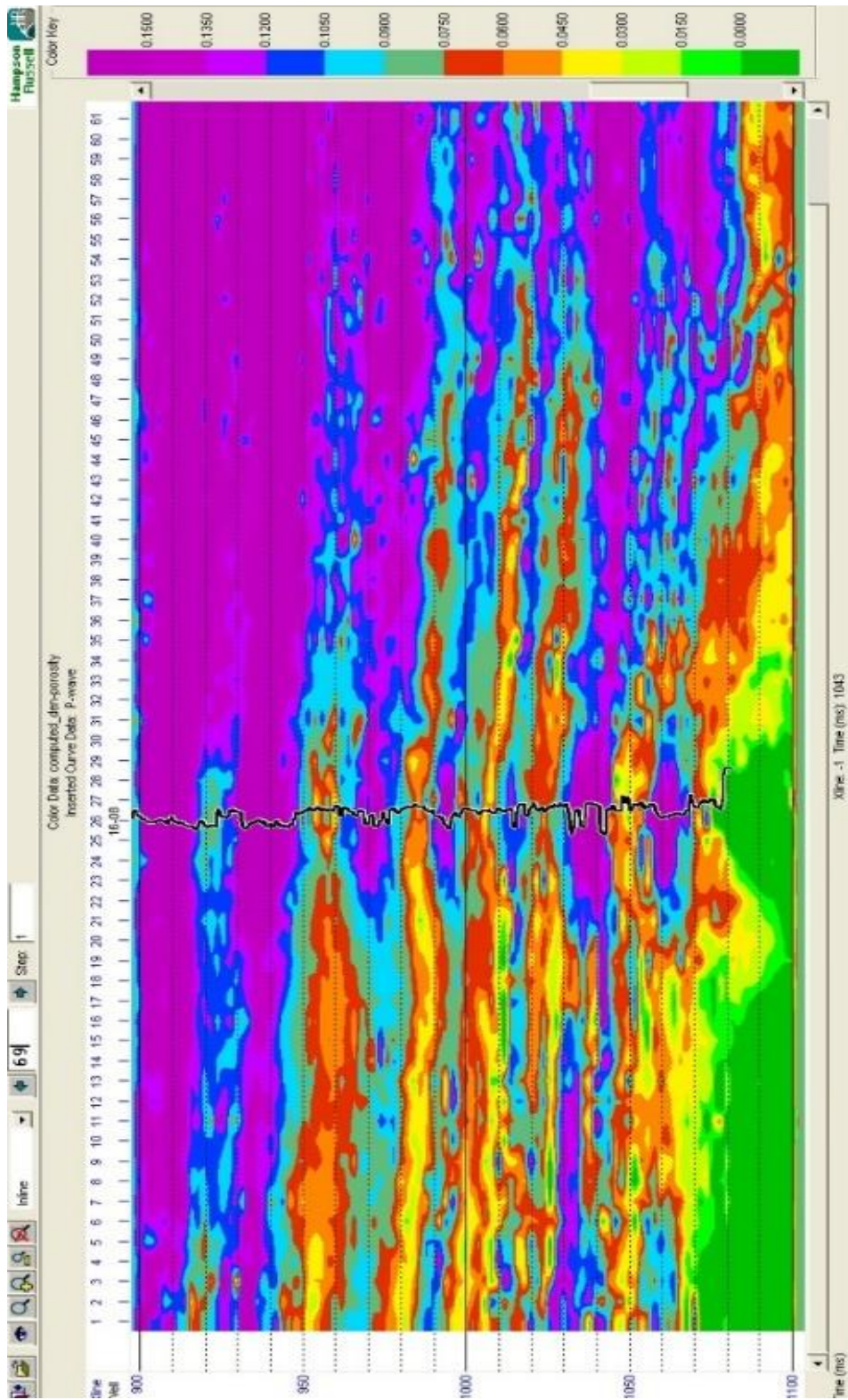
	Target	Final Attribute	Training Error	Validation Error
1	Porosity	1/(Inversion Result)	0.059749	0.061488
2	porosity	Amplitude Weighted frequency	0.057566	0.060147
3	porosity	Cosine Instantaneous phase	0.054103	0.59397
4	porosity	Integtate	0.054038	0.057885
5	porosity	Y-Coordinate	0.053162	0.056985
6	porosity	Integtated Absolute Amplitude	0.052437	0.056512
7	porosity	Second Derivative	0.051910	0.056814
8	porosity	Instantaneous Phase	0.051483	0.057913



شکل ۴-۷. نمودار متقاطع داده‌های تخلخل واقعی و تخلخل به دست آمده از روش چند نشانگری در محل چاه‌ها.



شکل ۴-۸. مقایسه نگارهای تخلخل واقعی و تخمین زده شده توسط روش رگرسیون چند متغیره در چاهها.



شکل ۹-۴. مقطعی از تخلخل به دست آمده در محدوده چاه ۰۸-۱۶.

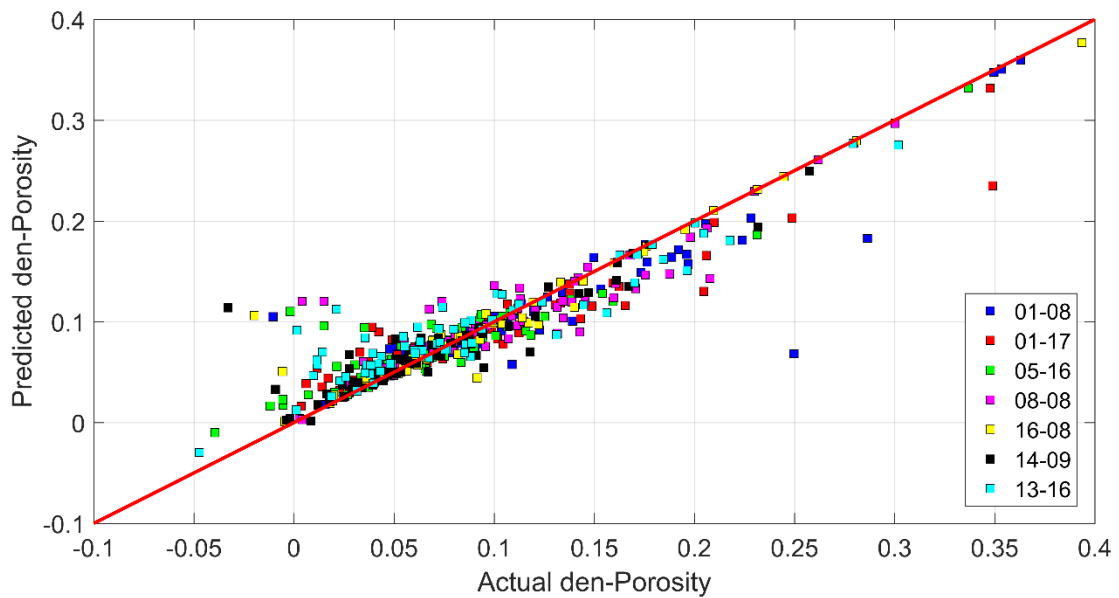
۳-۲-۴ تحلیل به کمک شبکه‌های عصبی

در شبکه عصبی از دو مرحله یادگیری و مرحله آزمایش استفاده می‌شود. بنابراین داده‌ها به دو دسته تقسیم می‌شوند. بخش بیشتر داده‌ها (۷۰ درصد) در مرحله یادگیری و باقی داده‌ها (۳۰ درصد) در مرحله آزمایش مورد استفاده قرار می‌گیرند. در مرحله یادگیری نمونه‌های آموزشی که شامل داده‌های ورودی و خروجی هستند به شبکه عصبی داده می‌شوند و ارتباط بین آن‌ها تشخیص داده می‌شود. در مرحله آزمایش روابط ایجاد شده بر روی داده‌های باقیمانده اعمال شده و میزان خطای شبکه به دست می‌آید. در این پایان‌نامه داده‌های ورودی به شبکه عصبی برای یادگیری و آزمایش، همان نشانگرهایی است که توسط روش رگرسیون چند نشانگری در موقعیت چاه‌ها تعیین (ورودی شبکه عصبی) گردید و تخلخل‌های مربوط به آن‌ها (خروجی شبکه عصبی) هستند. در نهایت رابطه به دست آمده در شبکه عصبی بر روی داده‌های لرزه‌ای خارج از چاه اعمال و تخلخل تخمین زده خواهد شد. با توجه به اینکه در اغلب بررسی‌ها شبکه عصبی احتمالاتی نتایج بهتری از سایر شبکه‌های عصبی می‌دهد، در این مطالعه از شبکه عصبی احتمالاتی استفاده گردیده است.

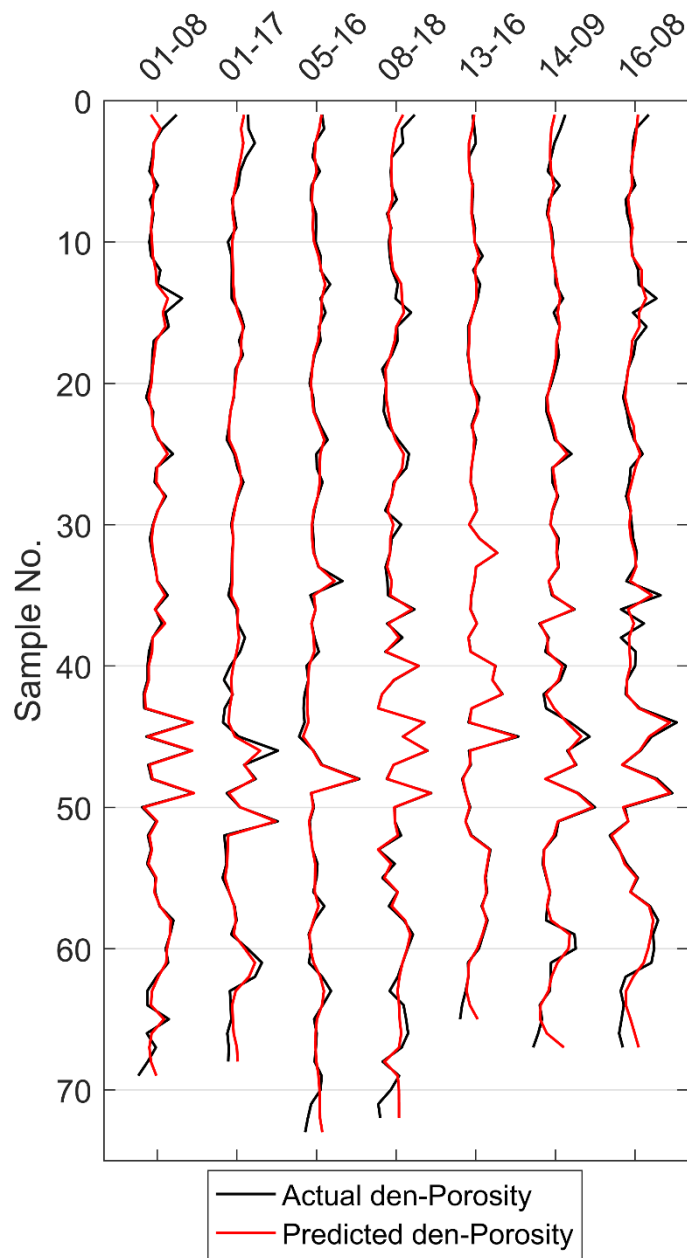
در شکل‌های ۴-۱۰، ۴-۱۱، ۴-۱۲، نتایج و تخمین تخلخل با شبکه عصبی احتمالاتی نشان داده شده است. در شکل ۴-۱۰ نمودار متقاطع تخلخل واقعی و تخلخل محاسبه شده توسط شبکه عصبی احتمالاتی در محل چاه‌ها به نمایش درآمده است. خط قرمز رنگ در این تصویر نماینده حالتی است که داده‌های تخمین زده شده و واقعی کاملاً (۱۰۰٪) تطابق دارند. در این شکل مشاهده می‌شود که اغلب داده‌های تخمین زده شده بر روی این خطوط یا بسیار نزدیک به آن قرار گرفته‌اند که نشان دهنده دقت بالای شبکه عصبی در تخمین تخلخل است.

همچنین در شکل ۴-۱۱ نگار تخلخل واقعی و محاسبه شده (برای هر ۷ چاه) در کنار یکدیگر رسم شده‌اند. نمودار سیاه رنگ مربوط به نگار تخلخل حقیقی و نمودار قرمز رنگ مربوط به نگار تخلخل تخمین زده شده است. شکل‌های ۴-۱۰ و ۴-۱۱ نشان دهنده تطابق بالا (۹۲٪) و خطای کم (۲٪) روش شبکه عصبی احتمالاتی است. در نهایت شبکه عصبی احتمال پذیر که در این مرحله نشان داده

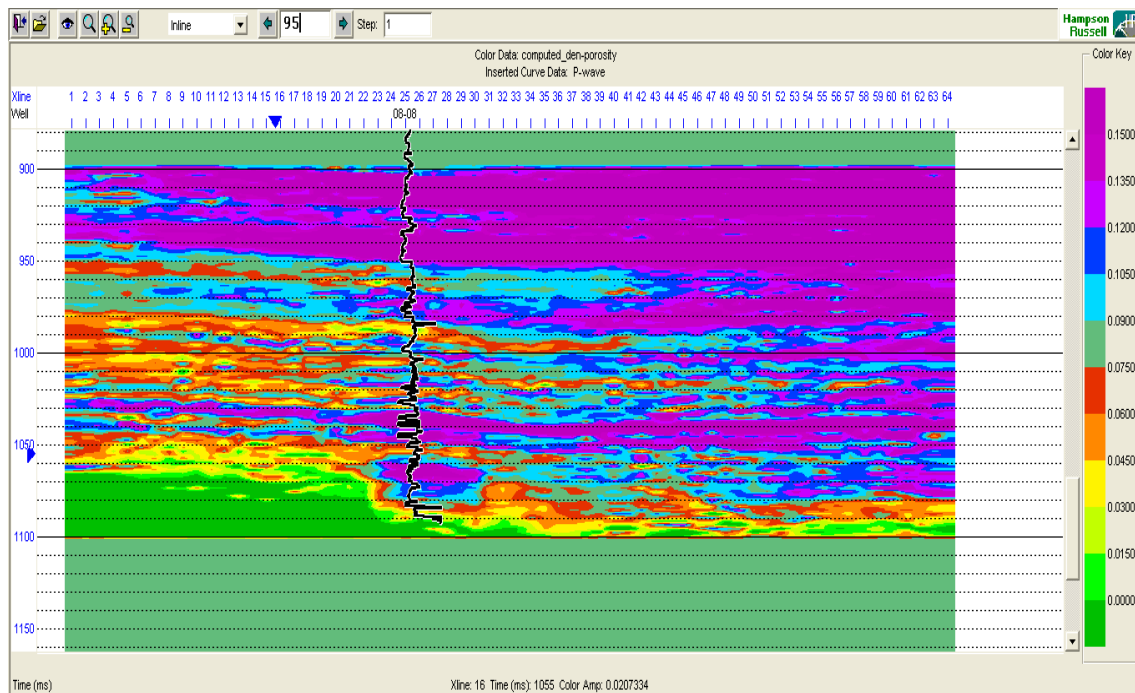
شد که دقت مناسبی دارد به منظور تخمین تخلخل در مناطق خارج از چاه مورد استفاده قرار گرفت. در شکل ۴-۱۲ نتایج تخمین تخلخل حاصل از شبکه عصبی به صورت مقطعی در راستای $\text{inline}=95$ در بازه زمانی ۹۰۰ تا ۱۱۰۰ میلی ثانیه نمایش داده شده است.



شکل ۴-۱۰. نمودار متقاطع داده‌های تخلخل واقعی و تخلخل به دست آمده به وسیله شبکه عصبی احتمالاتی در محل چاه‌ها.



شکل ۴-۱۱. نمایش تطابق نگار تخلخل واقعی و محاسبه شده توسط شبکه عصبی احتمالاتی در محل چاه.



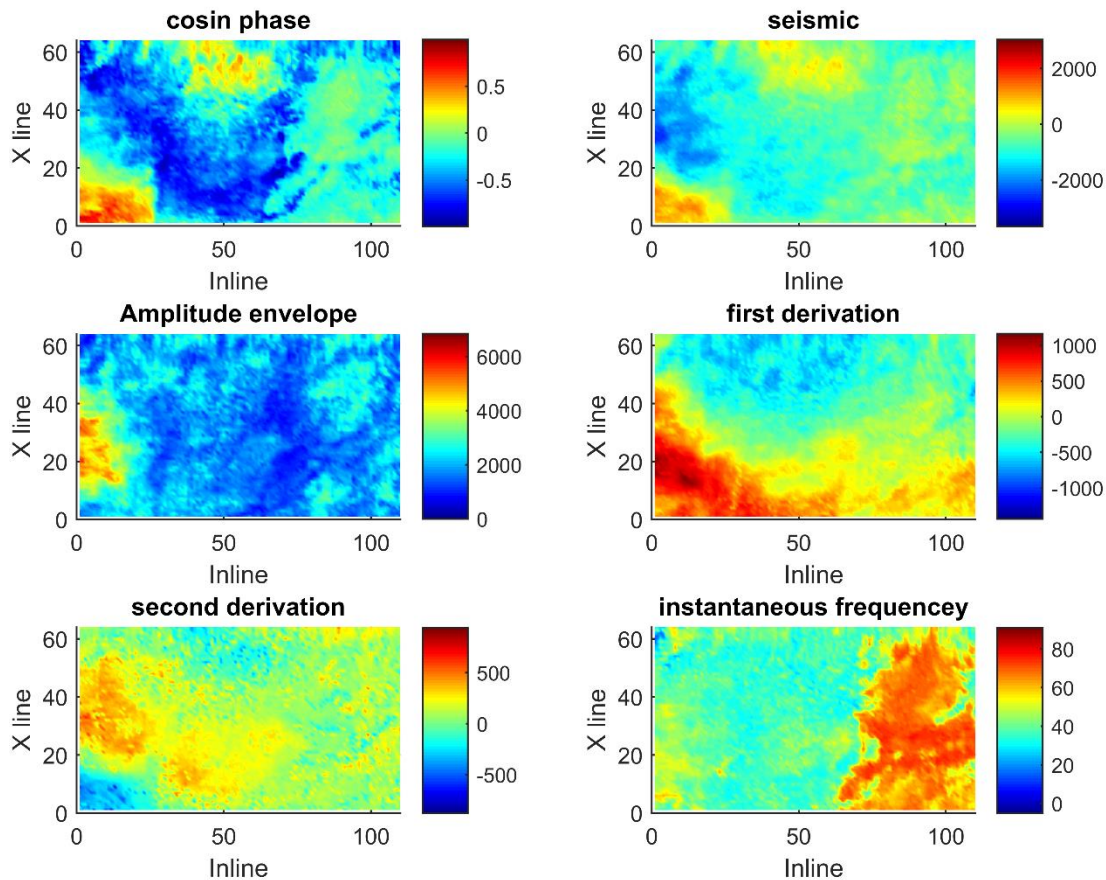
شکل ۴-۱۲. نتایج تخمین تخلخل در محدوده مورد مطالعه با اعمال شبکه عصبی PNN.

۴-۳ استفاده از روش آدابوست در طبقه‌بندی تخلخل

۴-۳-۱ مرحله اول آماده‌سازی داده‌ها:

ابتدا داده‌های نگارهای تخلخل مربوط به چاه‌ها را از نرم‌افزار همپسون-راسل استخراج می‌کنیم. نرخ نمونه‌برداری داده‌های لرزه‌ای ۲ میلی‌ثانیه و نرخ نمونه‌برداری نگار تخلخل ۱ میلی‌ثانیه است. بنابراین این دو نوع داده باید قبل از انجام هر کاری هم مقیاس شوند. بدین منظور پیش از استخراج نگار تخلخل از نرم‌افزار، ابتدا نگار چاه افزایش مقیاس داده می‌شوند. افزایش مقیاس در نرم‌افزار همپسون-راسل و با استفاده از روش میانگین‌گیری صورت می‌گیرد. در گام بعدی داده‌های لرزه‌ای به نرم‌افزار Petrel وارد کرده و داده‌ها در بازه ۹۰۰ تا ۱۱۰۰ میلی‌ثانیه جدا می‌شوند و در مرحله بعد نشانگرهای مورد نیاز استخراج می‌گردند که عبارتند از: نشانگرهای پوش ردلرزه، فرکانس لحظه‌ای، فاز لحظه‌ای، مشتق اول و دوم و ... نشانگر وارون لرزه‌ای نیز به صورت محاسبه شده در دسترس قرار دارد. شکل ۴-۱۳ تعدادی از نشانگرهای استخراج شده (نشانگرهای کسینوس فاز، دامنه لرزه‌ای، پوش دامنه، مشتق اول، مشتق دوم و فرکانس لحظه‌ای) از داده‌های لرزه‌ای مربوط به زمان ۱۰۶۰ میلی‌ثانیه را

نشان می دهد.



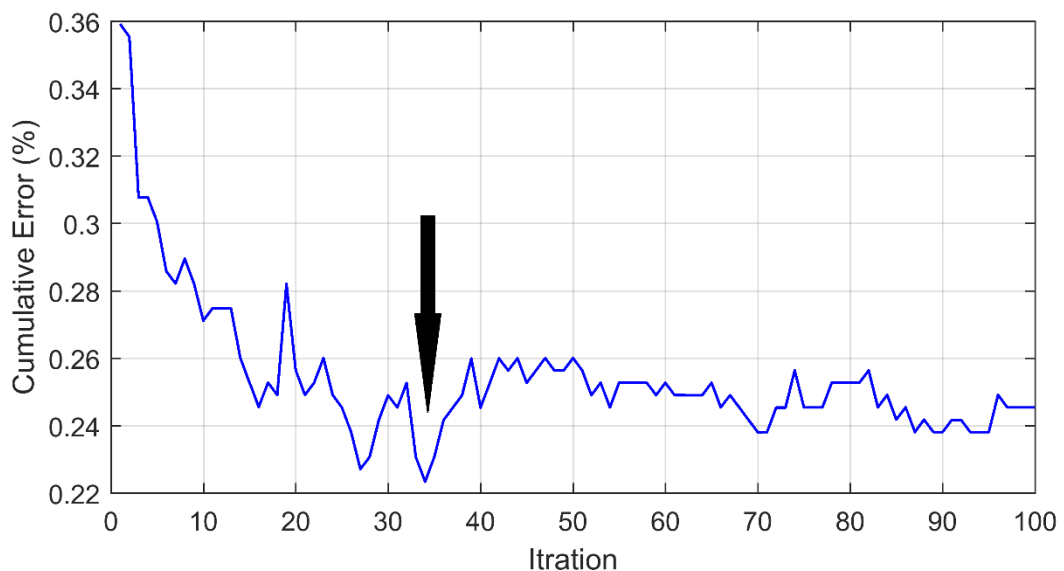
شکل ۴-۱۳. نمونه‌هایی از نشانگرهای به دست آمده از روی داده‌های لرزه‌ای: (الف) دامنه لرزه‌ای، (ب) کسینوس فاز، (ج) مشتق اول، (د) پوش دامنه، (ه) فرکانس لحظه‌ای و (و) مشتق دوم در زمان ۱۰۶۰ میلی ثانیه می‌باشد.

با توجه به اینکه داده‌های لرزه‌ای در هر نقطه از محیط اطراف آن نیز تأثیر می‌پذیرد؛ از داده لرزه‌ای و سایر نشانگرها را میانگین‌گیری انجام می‌گیرد. میانگین‌گیری با پنجره‌ای با طول ۵ نمونه در راستای قائم و دو میانگین‌گیری دیگر با پنجره‌های به طول سه نمونه در راستای خطوط برداشت صورت گرفته و نتیجه حاصله به مرکز پنجره اختصاص می‌یابد.

سپس در مکان چاه‌ها نشانگرها استخراج می‌شوند. با استفاده از این نشانگرها ماتریسی ایجاد می‌شود که هر ستون آن یک نشانگر مجزا است و هر سطر آن مربوط به یک داده تخلخل است. این ماتریس، همان داده‌های ورودی به روش آدابوست است. داده‌های تخلخل توسط روش knn به سه کلاس طبقه‌بندی می‌شوند و به عنوان خروجی روش آدابوست مورد استفاده قرار می‌گیرند. داده‌ها به صورت

تصادفی به دو بخش تقسیم می شوند. ۷۰ درصد برای آموزش و ۳۰ درصد برای تست الگوریتم مورد استفاده قرار می گیرند. ۷۰ درصد داده آموزش به الگوریتم داده و الگوریتم آموزش می یابد. در مرحله آزمایش با وارد کردن ۳۰ درصد باقیمانده به الگوریتم، خطای آن محاسب می شود.

الگوریتم آدابوست ۱۰۰ مرتبه تکرار می شود و خطای آن در هر بار تکرار محاسبه می گردد. شکل ۴-۱۴ خطای روش آدابوست در هر مرحله تکرار را نشان می دهد. مشاهده می شود که در ۳۴امین تکرار میزان خطای الگوریتم به کمترین مقدار خود می رسد و از آن به بعد شبکه دچار بیش برآزش می شود. بنابراین تعداد بهینه تکرار ۳۴ در نظر گرفته می شود و بار دیگر الگوریتم را با ۳۴ تکرار آموزش داده می شود.



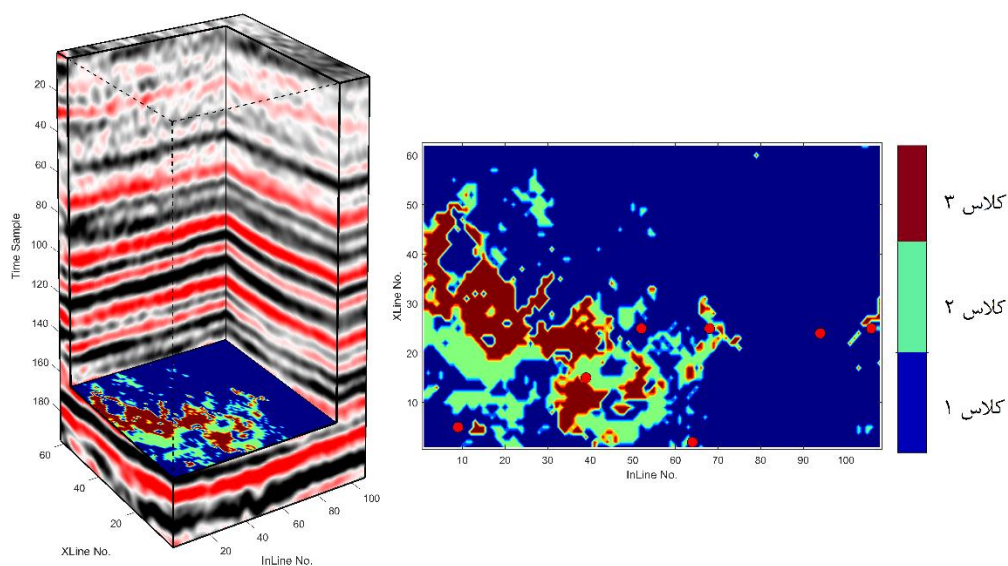
شکل ۴-۱۴. نمودار تجمعی خطا به ازای تکرار الگوریتم در مرحله آزمایش. در تکرار ۳۴ام کمترین مقدار خطا حاصل می شود.

مقدار خطای نهایی، با استفاده از ماتریس در هم ریختگی محاسبه می گردد. ماتریس درهم ریختگی برای الگوریتم آدابوست با تعداد تکرار بهینه به صورت جدول ۴-۴ نشان داده شده است. همان طور که مشاهده می شود ۸۹/۷٪ از داده های مربوط به کلاس ۱، به درستی تعیین شده اند. همچنین ۷۱ درصد داده های کلاس ۲ و ۶۸/۴ درصد از داده های کلاس ۳ به درستی تعیین شده اند.

جدول ۴-۴. ماتریس درهم ریختگی مربوط به کلاس های تداخل.

	کلاس واقعی ۱	کلاس واقعی ۲	کلاس واقعی ۳
کلاس تعیین شده ۱	۸۹/۷	۲۶/۴	۱۰
کلاس تعیین شده ۲	۹/۹	۷۱	۲۱/۶
کلاس تعیین شده ۳	۰/۴	۲/۵	۶۸/۴

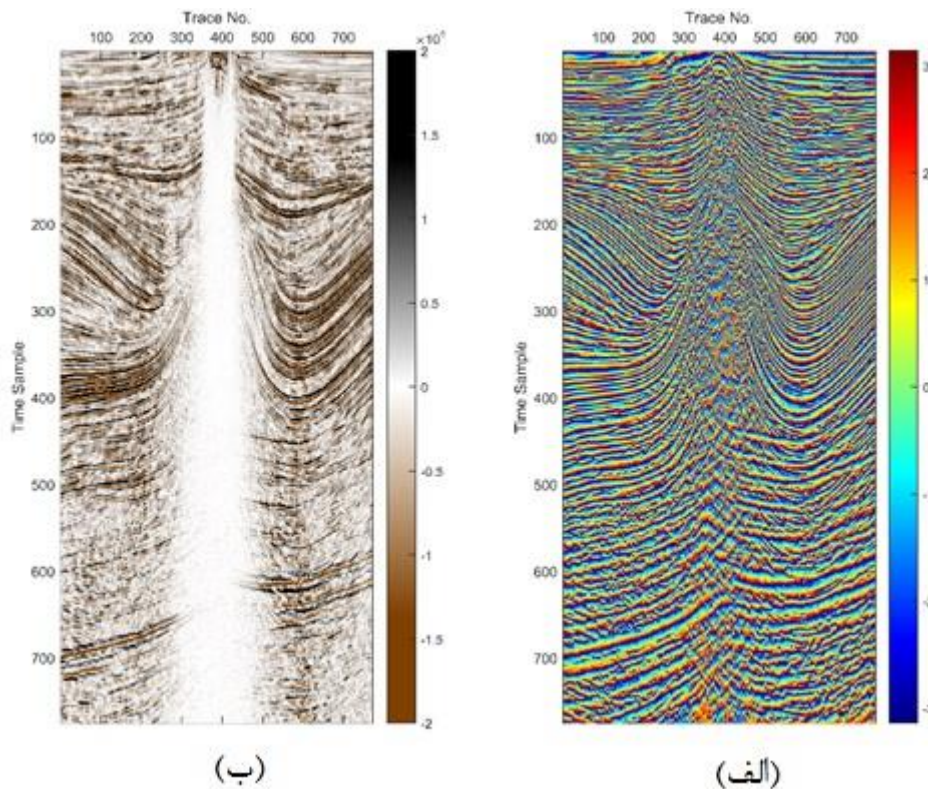
الگوریتم آموزش دیده با ۳۴ تکرار بهینه می‌باشد و از آن برای تعیین تخلخل کلاس مربوط به داده‌های خارج از چاه استفاده می‌شود. شکل ۴-۱۵ از اعمال الگوریتم آدابوست بر روی حجم لرزه‌ای به دست آمده است. تصویری که مشاهده می‌شود مربوط به برش زمانی ۱۰۶۰ میلی ثانیه می‌باشد.



شکل ۴-۱۵. کلاسهای بدست آمده حاصل از استفاده از روش طبقه بندی آدابوست. برش زمانی در زمان ۱۰۶۰ (ms).

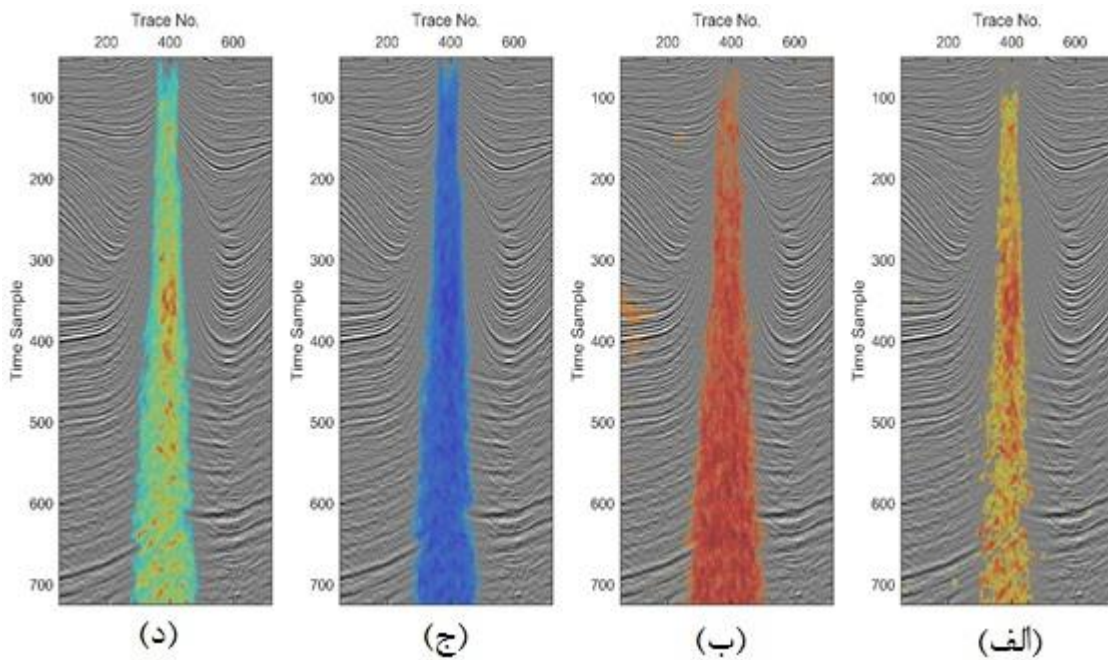
۴-۴ تعیین محدوده گنبد نمکی با استفاده از روش آدابوست

داده واقعی مورد استفاده بخشی از یک مقطع لرزه‌ای دوبعدی مربوط به داده‌های دریایی خلیج فارس می‌باشد که با گام نمونه برداری ۴ میلی ثانیه برداشت شده است. در شکل ۴-۱۶ داده مورد نظر نشان داده شده است. مقطع نشان داده شده حاوی یک گنبد نمکی می‌باشد.



شکل ۴-۱۶. الف) مقطع فاز لحظه ای مربوط به یک گنبد نمکی، ب) مقطع لرزه‌ای واقعی مربوط به گنبد نمکی مشابه.

در این پایان‌نامه ماتریس GLCM داده‌های لرزه‌ای درون یک پنجره متحرک بر روی داده‌های لرزه‌ای محاسبه شده است و از روی آن، نشانگرهای بافتی برتری خوشه، اینرسی، انرژی، آنتروپی، شباهت، واریانس، میانگین، سایه، شدت، همبستگی و همگنی در راستای ۹ زاویه مختلف محاسبه شد؛ که منجر به تولید ۹۹ نشانگر برای هر نمونه زمانی گردید. در شکل ۴-۱۷، چهار نمونه از نشانگرهای ایجادشده نمایش داده شده است که حاصل عمل ماتریس GLCM با زاویه صفر درجه بر روی نشانگرهای برتری خوشه، اینرسی، واریانس و سایه خوشه هستند.



شکل ۴-۱۷. نشانگرهای بافتی. (الف) سایه خوشه، (ب) برتری خوشه، (ج) اینرسی، (د) واریانس.

همان‌طور که مشاهده می‌شود؛ برخی از این نشانگرهای بافتی صرفاً در شناسایی محدوده جانبی گنبد نمکی موفق عمل می‌کنند و می‌توانند در افق‌های لرزه‌ای مورد استفاده قرار گیرند. در حالی که برخی دیگر از این نشانگرها علاوه بر اینکه از لحاظ جانبی تا حدودی موفق‌اند، تفاوت‌هایی هر چند اندک، میان نمک و زیر نمک قائل می‌شوند. بنابراین به منظور ایجاد یک طبقه‌بندی کننده قوی، تمامی این نشانگرها مورد استفاده قرار می‌گیرند.

این ۹۹ نشانگر به همراه کلاس (سه کلاس شامل گنبد نمکی، محیط پیرامون گنبد نمکی و زیر گنبد نمکی) مربوط به هر داده برای آموزش و آزمایش الگوریتم آدابوست مورد استفاده قرار گرفت که نتایج آن در جدول ۴-۵ ارائه شده است.

جدول ۴-۵. نتایج به دست آمده از روش آدابوست.

	کلاس ۱ واقعی	کلاس ۲ واقعی	کلاس ۳ واقعی
کلاس ۱ پیش بینی شده	۹۵	۱۰	۰
کلاس ۲ پیش بینی شده	۵	۹۰	۰
کلاس ۳ پیش بینی شده	۰	۰	۱۰۰

با توجه به جدول ۴-۵ هیچ یک از کلاس‌های ۱ و ۲ (گنبد نمکی و زیر آن) به عنوان کلاس ۳

(بخش‌های در بر گیرنده گنبد نمکی) طبقه‌بندی نشده‌اند. بنابراین در تشخیص مرزهای جانبی گنبد نمکی هیچ‌گونه خطایی وجود ندارد. در بخش تشخیص کف گنبد نمکی مشاهده می‌شود که قدری خطا وجود دارد و به‌طور متوسط ۷/۵ درصد خطا وجود دارد.

با اعمال روش SVM بر روی داده‌های موجود، طبقه‌بندی گنبد نمکی از محدوده پیرامون آن و محدوده زیر آن مجدداً تکرار می‌گردد. نتایج این طبقه‌بندی در جدول ۴-۶ ارائه شده است.

جدول ۴-۶. نتایج به دست آمده از روش ماشین بردار پشتیبان.

	کلاس ۱ واقعی	کلاس ۲ واقعی	کلاس ۳ واقعی
کلاس ۱ پیش بینی شده	۹۱/۶	۱۱/۷	۰
کلاس ۲ پیش بینی شده	۸/۴	۸۶/۷	۰
کلاس ۳ پیش بینی شده	۰	۱/۶	۱۰۰

با توجه به جدول ۴-۶ هیچ یک از کلاس‌های ۱ و ۲ (گنبد نمکی و زیر آن) به‌عنوان کلاس ۳ (بخش‌های در بر گیرنده گنبد نمکی) طبقه‌بندی نشده‌اند. بنابراین در تشخیص مرزهای جانبی گنبد نمکی هیچ‌گونه خطایی وجود ندارد. در بخش تشخیص کف گنبد نمکی مشاهده می‌شود که قدری خطا وجود دارد و به‌طور متوسط ۱۰/۵ درصد خطا وجود دارد.

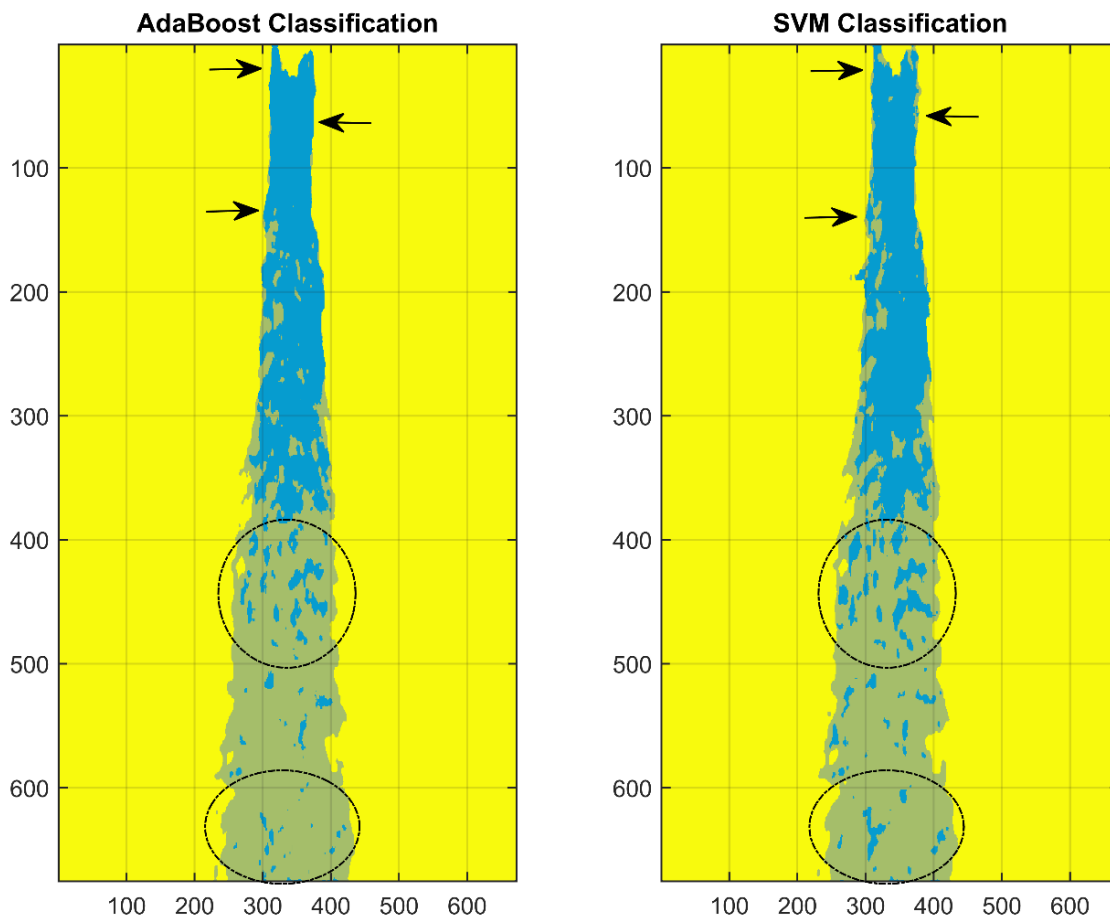
با مقایسه نتایج بیان شده به‌صورت جداول ۴-۶ و ۴-۷ دقت روش آدابوست ۹۴ درصد و دقت روش SVM ۹۲ درصد می‌باشد. بنابراین روش آدابوست نسبت به روش قدرتمند SVM از دقت بالاتری برخوردار می‌باشد.

همان‌طور که در شکل ۴-۱۶ نمایان است، در نشانگر فاز لحظه‌ای بازتاب‌های لرزه‌ای در عمق زمانی حدودی ۴۰۰ نمونه زمانی نمایان شده‌اند و دچار کشیدگی به سمت بالا^۱ شده‌اند.

شکل ۴-۱۸، شامل نتیجه نهایی روش طبقه‌بندی آدابوست (الف) و SVM (ب) است. همان‌طور که در شکل مشاهده می‌شود مقاطع لرزه‌ای به خوبی به سه کلاس گنبد نمکی (رنگ آبی)، لایه‌بندی‌های

^۱. Pull up

اطراف گنبد نمکی (رنگ زرد) و زیر گنبد نمکی (رنگ پسته‌ای) تفکیک شده‌اند که تطابق خوبی با مقطع فاز لحظه‌ای نشان می‌دهد.



شکل ۴-۱۸. مقایسه نتیجه به دست آمده از روش آدا بوست و ماشین بردار پشتیبان.

مشاهده می‌شود که هر دو روش SVM و آدا بوست حدود عمقی گنبد نمکی در محدوده نمونه زمانی ۴۰۰ تشخیص داده‌اند. با مقایسه این نتایج با نشانگر فاز لحظه‌ای و مشاهده تطابق بالای آن‌ها با یکدیگر، نتیجه به دست آمده توسط SVM و آدا بوست تا حد زیادی اعتبار می‌یابند.

در شکل ۴-۱۸ مشاهده می‌شود که روش آدا بوست توانسته است تا در قسمت‌های بالایی گنبد نمکی مرزها را با دقت بالاتری از محیط پیرامون خود تفکیک کند. همچنین در قسمت پایین گنبد نمکی بر اثر خطای این روش‌ها بخش‌هایی از کلاس ۳ به عنوان کلاس ۲ تشخیص داده شده‌اند که این تشخیص اشتباه در روش آدا بوست کمتر است.

فصل پنجم

نتیجه گیری و پیشنهاد

۵-۱ نتیجه‌گیری

در این پایان‌نامه از روش جدید آدابوست به منظور طبقه‌بندی افق لرزه‌ای بر مبنای تخلخل و تعیین مرزهای جانبی و کف گنبد نمکی از محیط پیرامون خود استفاده شده است. نشان داده شده است که طبقه‌بندی کننده آدابوست از روش‌های متداول طبقه‌بندی مانند KNN و SVM دارای دقت بالاتر می‌باشد. بنابراین با این پیش فرض در مرحله اول به منظور طبقه‌بندی افق لرزه‌ای بر مبنای تخلخل از طبقه‌بندی کننده آدابوست استفاده گردید. دقت این طبقه‌بندی کننده ۷۸ درصد می‌باشد. به منظور بررسی بیشتر از روش‌های رگرسیون معمولی، رگرسیون چند متغیره و شبکه عصبی احتمال پذیر برای تخمین مقادیر تخلخل استفاده گردید. روش طبقه‌بندی آدابوست نتایج بهتری نسبت به روش‌های تخمینی رگرسیون معمولی و چند متغیره به دست می‌دهد. اما در مقایسه با روش تخمینی شبکه عصبی احتمال پذیر که دارای ۹۲ درصد دقت در تخمین است، ضعیف‌تر عمل می‌کند. از مهم‌ترین دلایل این موضوع می‌توان به تفاوت تعداد نشانگر استفاده شده در روش شبکه عصبی احتمال پذیر اشاره کرد که به دلیل تفاوت در مبنای این دو روش می‌باشد. در روش شبکه عصبی احتمال پذیر در مرحله انتخاب نشانگرها و تعداد بهینه آن‌ها ۷ نشانگر تعیین و مورد استفاده قرار می‌گیرند. اما در روش آدابوست که انتخاب نشانگرها به صورت خودکار انجام می‌گیرد، از مجموع ۱۵ نشانگر ورودی تنها ۴ نشانگر به عنوان نشانگرهای مؤثر مورد استفاده قرار می‌گیرند. از طرف دیگر برخی از نشانگرها مانند موقعیت قرارگیری نمونه بر روی inline که برای استفاده در روش شبکه عصبی احتمالی هستند و موجب افزایش دقت آن می‌شود، برای استفاده در روش طبقه‌بندی آدابوست مناسب نیستند و موجب افزایش خطا می‌شوند.

در مرحله بعد از روش طبقه‌بندی آدابوست برای تعیین مرزهای جانبی و کف گنبد نمکی مورد استفاده قرار گرفت؛ که طی آن سه کلاس (گنبد نمکی، محیط پیرامون آن، لایه‌های موجود زیر گنبد نمکی) را با دقت بالای ۹۵ درصد از یکدیگر تفکیک گردید. به منظور بررسی قدرت روش طبقه‌بندی آدابوست، این روش با روش طبقه‌بندی SVM که قدرتمندترین روش طبقه‌بندی گنبد نمکی تاکنون

است، مقایسه گردید. روش SVM کلاس‌های عنوان شده را با دقت ۹۲ درصد از یکدیگر تفکیک کرد. بنابراین دقت روش آدابوست از روش SVM بیشتر می‌باشد و در حال حاضر روش طبقه‌بندی آدابوست قوی‌ترین طبقه‌بندی کننده برای تفکیک مرزهای گنبد نمکی و کف آن از محیط پیرامون خود می‌باشد. از نظر سرعت کار، روش آدابوست نسبت به روش شبکه عصبی دارای به صورت محسوسی دارای سرعت بالاتری می‌باشد و نسبت به روش SVM نیز به میزان اندکی سریع‌تر است.

۲-۵ پیشنهادات

در الگوریتم آدابوست به منظور انتخاب نشانگرهای بهینه، هر نشانگری که بتواند بیش از ۵۰ درصد از نمونه‌هایی که در مرحله قبل اشتباه طبقه‌بندی شده است را درست تشخیص دهد انتخاب می‌شود و به آن وزنی اختصاص داده می‌شود و در صورت عدم توانایی انجام این موضوع حذف می‌شود و از نشانگر دیگری استفاده می‌شود. پیشنهاد می‌گردد به منظور افزایش تعداد نشانگرهای مورد استفاده حد آستانه برای انتخاب نشانگر را که در الگوریتم اصلی ۵۰ درصد است تا مرز ۳۰ درصد کاهش داده و نتایج اعمال این تغییرات مورد بررسی قرار گیرد.

- [۱] قریشی ف.، ۱۳۹۳، پایان نامه ارشد، " بهبود روش افزایش مقیاس نگارهای پتروفیزیکی در نرم افزار همپسون راسل مبتنی بر موجک، مطالعه موردی: یکی از میادین نفتی جنوب ایران"، دانشکده مهندسی معدن، نفت و ژئوفیزیک، دانشگاه شاهرود.
- [2] Lindseth, R. O. (1979). Synthetic sonic logs-a process for stratigraphic interpretation. *Geophysics*, 44(1), 3-26.
- [3] Oldenburg, D. W., Scheuer, T., & Levy, S. (1983). Recovery of the acoustic impedance from reflection seismograms. *Geophysics*, 48(10), 1318-1337.
- [4] Hampson, Daniel P., James S. Schuelke, and John A. Quirein. "Use of multiattribute transforms to predict log properties from seismic data." *Geophysics* 66.1 (2001): 220-236.
- [5] Deutsch, C. (1989). Calculating effective absolute permeability in sandstone/shale sequences. *SPE Formation Evaluation*, 4(03), 343-348.
- [6] Schultz, P. S., Ronen, S., Hattori, M., & Corbett, C. (1994). Seismic-guided estimation of log properties (Part 1: A data-driven interpretation methodology). *The Leading Edge*, 13(5), 305-310.
- [7] Schultz, P. S., Ronen, S., Hattori, M., Mantran, P., & Corbett, C. (1994). Seismic-guided estimation of log properties (Part 3: A controlled study). *The Leading Edge*, 13(7), 770-776.
- [8] Ronen, S., Schultz, P. S., Hattori, M., & Corbett, C. (1994). Seismic-guided estimation of log properties (Part 2: Using artificial neural networks for nonlinear attribute calibration). *The Leading Edge*, 13(6), 674-678.
- [9] Schuelke, J. S., Quirein, J. A., & Fredrick, J. (1998). Reservoir architecture and porosity distribution, Pegasus Field, West Texas—an integrated sequence stratigraphic–seismic attribute study using neural networks.
- [10] Schuelke, J. S., & Quirein, J. A. (1998, January). Validation: A technique for selecting seismic attributes and verifying results. In *1998 SEG Annual Meeting*. Society of Exploration Geophysicists.
- [11] Leiphart, D. J., & Hart, B. S. (2001). Comparison of linear regression and a

probabilistic neural network to predict porosity from 3-D seismic attributes in Lower Brushy Canyon channeled sandstones, southeast New Mexico. *Geophysics*, 66(5), 1349-1358.

[12] Pramanik, A. G., Singh, V., Vig, R., Srivastava, A. K., & Tiwary, D. N. (2004). Estimation of effective porosity using geostatistics and multiattribute transforms: A case study. *Geophysics*, 69(2), 352-372.

[۱۳] عدالت ع.، سیاه کوهی ح. و توکلی مقدم ر.، ۱۳۸۸، "برآورد تخلخل موثر مخزن با استفاده از تحلیل چند نشانگری"، مجله ژئوفیزیک ایران، جلد ۳، شماره ۱، صفحه ۱-۱۸.

[۱۴] کدخدایی ایلخچی ع.، پور اسلامی اردکانی ا.، اسفندیاری م. و برگریزان م.، "تخمین نگارهای چاه پیمایی با استفاده از نشانگرهای لرزه ای چندگانه در سازند آسماری، میدان نفتی هندیجان"، مجله علوم دانشگاه تهران، دوره ۳۴، شماره ۴، صفحه ۳۵-۴۳

[15] Zhan S., L. Li, W. Xiong and Z. Wan, 2011, *Automatic geological body identification using the modified rival penalized competitive learning clustering algorithm*: 81st annual Meeting Expanded Abstracts, SEG

[16] Liu X., X. Zheng, G. Xu, H. Yang and J. Song, 2011, *PSO-based multi-attribute dynamic clustering technology and its application*: 81st annual Meeting Expanded Abstracts, SEG, 1913-1917

[17] Li L., W. Xiong, S. Zhan and Z. Wan, 2011, *Reservoir property prediction using the dynamic radial basis function network*: 81st annual Meeting Expanded Abstracts, SEG, 1754-1758

[18] Duda R. O., P. E. Hart, D. E. Stoek, 2000, *Pattern Classification*, Vol. 2, Wiley, pp.6804

[19] Schapire R. E., 1990, **The strength of weak learnability**, *Mach. Learn.*, vol.5, pp. 197-227

[20] Freund Y. and R. Schapire, 1995, *A decision-theoretic generalization of on-line learning and an application to boosting*: The Second European Conference on Computational Learning Theory, P23-37

[21] Xiong W., Z. H. Wan, X. L. Bai, H. G. Xing, H. G. Zuo, K. H. Zhu and S. G. Yang, 2014, *AdaBoost-based Multi-attribute Classification Technology and Its Application: 76th Conference & Exhibition, EAGE, Amsterdam RAI, The Netherlands*

[۲۲] ذاکری نژاد ز.، ۱۳۹۰، پایان نامه ارشد، "تهیه نقشه شباهت و تحلیل رگرسیون داده های لرزه ای و نگارهای چاه در یک مخزن و کاربرد آن در تعیین نقاط با پتانسیل بالای اکتشافی"، دانشکده مهندسی معدن، نفت و ژئوفیزیک، دانشگاه شاهرود.

[23] Bradley, H. B. (1987). *Petroleum engineering handbook*.

[24] Al-Moqbel, A. M. S. (2002). *Reservoir characterization using seismic reflectivity and attributes* (Doctoral dissertation, Massachusetts Institute of Technology).

[۲۵] مراد زاده ع.، قوامی ریایی ر.، ۱۳۸۰، "چاه پیمایی برای مهندسی"، دانشگاه شاهرود، ۲۴۵.

[26] Cosentino, L., 2001, *Integrated reservoir studies: Editions Technip*. Dalley, R. M., E. C. A. Gevers, G. M. Stampfli, D. J. Davies, C. N.

[27] Chopra, S., & Marfurt, K. J. (2005). Seismic attributes—A historical perspective. *Geophysics*, 70(5), 3SO-28SO.

[28] Barnes, A. E. (1994, January). Theory of two-dimensional complex seismic trace analysis. In *1994 SEG Annual Meeting*. Society of Exploration Geophysicists.

[29] Barnes, A. E. (2000). Attributes for automated seismic facies analysis: 70th Annual International Meeting, SEG, Expanded Abstracts, 553–556.

[30] Kalkomey, C. T. (1997). Potential risks when using seismic attributes as predictors of reservoir properties. *The Leading Edge*, 16(3), 247-251.

[31] Taner, M. T., Koehler, F., & Sheriff, R. E. (1979). Complex seismic trace analysis. *Geophysics*, 44(6), 1041-1063.

[32] Brown, A. R. (2001). Understanding seismic attributes. *Geophysics*, 66(1), 47-48.

[33] Barnes, A. E., 1997, Genetic classification of complex seismic trace attributes: 67th Annual International Meeting, SEG, Expanded Abstracts, 1151–1154

[34] Liner, C., Li, C. F., Gersztenkorn, A., & Smythe, J. (2004, January). SPICE: A new general seismic attribute. In *2004 SEG Annual Meeting*. Society of Exploration

Geophysicists.

[35] Meldahl, P., Heggland, R., Bril, B., & de Groot, P. (2001). Identifying faults and gas chimneys using multiattributes and neural networks. *The Leading Edge*, 20(5), 474-482.

[36] Chen, Q., & Sidney, S. (1997). Seismic attribute technology for reservoir forecasting and monitoring. *The Leading Edge*, 16(5), 445-448.

[37] Sheriff, R. E., comp., 2002, Encyclopedic dictionary of applied geophysics, 4th ed.: SEG, Geophysical References Series 13.

[38] Subrahmanyam, D., & Rao, P. H. (2008). Seismic attributes-a review. In *7th International Conference and Exposition on Petroleum Geophysics, Hyderabad, India*.

[39] Valenti, J. C. A. F. (2009). *Porosity prediction from seismic data using multiattribute transformations, N Sand, Auger Field, Gulf of Mexico* (Doctoral dissertation, The Pennsylvania State University).

[40] Gao, D. (2003). Volume texture extraction for 3D seismic visualization and interpretation. *Geophysics*, 68(4), 1294-1302.

[41] Yenugu, M., Marfurt, K. J., & Matson, S. (2010). Seismic texture analysis for reservoir prediction and characterization. *The Leading Edge*, 29(9), 1116-1121.

[42] Berthelot, A., Solberg, A.H.S., Morisbak, E., Gelius, L.-J., (2011) Salt diapirs without well defined boundaries — a feasibility study of semi-automatic detection: *Geophysical Prospecting* 59, 682–696.

[43] Haralick, R. M., Shanmugam, K., and Dinstein, I., (1973), Textural features for image classification: *IEEE Tran. Systems, Man, and Cybernetics*, SMC-3, 610–621.

[۴۴] صیادی ع.، ۱۳۹۳، پایان نامه ارشد، " کاربرد نشانگرهای بافتی بر مبنای ماتریس هم رخداد سطح خاکستری در تفسیر داده های لرزه ای بازتابی"، دانشکده مهندسی معدن، نفت و ژئوفیزیک، دانشگاه شاهرود.

[45] Russell, B. H. (2004). *The application of multivariate statistics and neural networks to the prediction of reservoir parameters using seismic attributes*.

[46] Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (Vol. 5, No. 8). Upper Saddle River, NJ: Prentice hall.

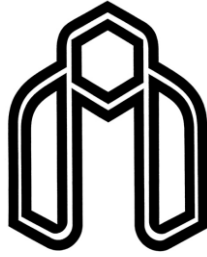
[47] Masters, T. (1993). *Practical neural network recipes in C++*. Morgan Kaufmann.

- [48] McCormack, M. D. (1991). Neural computing in geophysics. *The Leading Edge*, 10(1), 11-15.
- [49] Liu, Z., & Liu, J. (1998). Seismic-controlled nonlinear extrapolation of well parameters using neural networks. *Geophysics*, 63(6), 2035-2041.
- [50] Masters, T. (1995). *Advanced algorithms for neural networks: a C++ sourcebook*. John Wiley & Sons, Inc..
- [51] Kalkomey, C. T. (1997). Potential risks when using seismic attributes as predictors of reservoir properties. *The Leading Edge*, 16(3), 247-251.
- [52] Guha, S., Rastogi, R., & Shim, K. (1998, June). CURE: an efficient clustering algorithm for large databases. In *ACM SIGMOD Record* (Vol. 27, No. 2, pp. 73-84). ACM.
- [۵۳] ثابتی ح، ۱۳۸۶، پایان نامه ارشد، "تحلیل رخساره لرزه‌های براساس طبقه‌بندی نشانگرهای لرزه‌های دو و سه‌بعدی"، موسسه ژئوفیزیک، دانشگاه تهران
- [54] Vapnik, V. (1963). Pattern recognition using generalized portrait method. *Automation and remote control*, 24, 774-780.
- [55] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152). ACM.
- [56] Burges, C. J. C., and Scholkopf, B., (1997), Improving the accuracy and speed of support vector learning machines. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, pages 375–381, Cambridge, MA. MIT Press.
- [57] Cortes, C., and Vapnik, V., (1995), Support vector networks: *Machine Learning*, 20:273–297.
- [58] Burges, C. J. C., and Scholkopf, B., (1997), Improving the accuracy and speed of support vector learning machines. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, pages 375–381, Cambridge, MA. MIT Press.
- [59] Changrampadi, M. H. (2011). A fusion-based multiclass AdaBoost for classifying object poses using visual and IR images.

Abstract

Petrophysics and reservoir characterization is the main purposes of oil exploration. For Petrophysics and reservoir characteristics estimation with high confidence, to early reservoir simulations using seismic and well log data is essential. Using seismic attributes, hidden information extract from seismic raw data and finding relationships between seismic attributes and reservoir petrophysical parameters in place wells, to achive the distribution of these parameters with better accuracy where the well log data do not exist. The fundamental problem in the way of integration of seismic data and well data, is an integration method of this data with each other in order to estimate reservoir properties. Much research has been conducted in this area that make the estimation of petrophysical properties of the reservoir using seismic data (raw and indicators) and well log possible. The studies that have been conducted under the general title of multi attributes analysis, often using conventional regression or neural networks methods to intedrate seismic and well log data and have reached significant results in description and estimation of the distribution of petrophysical parameters of the reservoir. Multi attributes classification is another method to integrate seismic and well logs data. Adaboost classification method which achive more attention in recent years and today it is one of the 10 superior data-mining algorithm. Adaboost outstanding feature is setting classification in the learning phase. The algorithm is fairly simple and fast solves overfitting pooblem there is in other methods and can also be combined with other techniques to form a fast classification algorithm. In this thesis the Adaboost method has used to multi attribute classification in a reflection seismology studies for classification of a horizon based on petrophysical information and identify structures such as salt are used. Adaboost method has 78% accuracy in seismic horizon classification based on porosity. Also it has 94% accuracy in the determining of salt dome boundaries from the surrounding media and the salt dome base which. Compared with SVM classification method that is 91 percent accurate, the Adaboost methods accuracy is higher.

Key words: classification, Adaboost, salt dome, seismic attributes



Shahrood University of Technology
Faculty of mine engineering, oil and geophysics

**Application of AdaBoost method in interpretation of
reflection seismic data**

Pedramebrahimi

Supervisor(s):

Amin Roshandel Kahoo

Seyed-Reza Ghavami-Riabi

Advisor:

Behzad Tokhmechi

Date:7/2016