



کاربرد روش‌های طبقه‌بندی بر اساس نظریه بیزین در پتانسیل‌یابی طلا به کمک داده‌های زمین‌شیمیایی در کانسار طلای اپی‌ترمال ساری‌گونای

حمید گرانیان^{۱*}، سید حسن طباطبائی^۲، هوشنگ اسدی هارونی^۲

۱- دانشجوی دکتری، دانشکده معدن، دانشگاه صنعتی اصفهان، اصفهان، ایران

۲- استادیار دانشکده معدن، دانشگاه صنعتی اصفهان، اصفهان، ایران

* عهده‌دار مکاتبات: h.geranian@mi.iut.ac.ir

دریافت مقاله: ۹۲/۱/۱۵، پذیرش مقاله: ۹۲/۳/۱۸

چکیده

کانسار طلای اپی‌ترمال ساری‌گونای در استان کردستان، یکی از مهم‌ترین کانسارهای کشف‌شده طلای ایران در رده جهانی به‌شمار می‌رود. این کانسار از نوع کانسارهای طلای اپی‌ترمال سولفیداسیون پایین است. در این مقاله با استفاده از داده‌های زمین‌شیمیایی محیط خاکی و به کمک چهار روش طبقه‌بندی بیزین، k- نزدیک‌ترین همسایگی، پنجره پارزن و بیزی ساده که همگی بر اساس تئوری بیزین بنا نهاده شده‌اند، موقعیت کانی‌زایی طلا در این کانسار مدل‌سازی شده است. پارامتر شاخص تولید برای سلول‌های به ابعاد ۲۵×۲۵ متر در سطح زمین تعریف شده است. ۶۵ درصد داده‌ها، به‌عنوان داده‌های آموزشی و ۳۵ درصد باقیمانده به‌عنوان داده‌های آزمایشی به کار رفته‌اند. نتایج نشان داده است که روش بیزی با درستی ۷۲/۶ درصد و پنجره پارزن با درستی ۷۰/۴ درصد عملکرد بهتری نسبت به دو روش دیگر دارند. همچنین این دو روش، شکل و وسعت نواحی امیدبخش کانی‌زایی را بهتر از دو روش دیگر مدل‌سازی کرده است. در این مدل‌ها مناطقی نیز برای حفاری تکمیلی پیشنهاد شده است.

واژه‌های کلیدی: روش بیزی، روش k- نزدیک‌ترین همسایگی، روش پنجره پارزن، روش بیزی ساده، کانسار طلای اپی‌ترمال ساری-گونای

۱- مقدمه

در روش‌های طبقه‌بندی داده‌ها بر اساس تئوری بیزین، تمرکز بر یافتن یک تابع چگالی احتمالی برای هر طبقه از داده‌ها آموزشی و سپس برآورد احتمالی تعلق یک نمونه جدید (از داده‌های آزمایشی) به هر یک از طبقات بر این اساس است. روش‌های توابع تمایز (Discriminant Functions)، بیشترین درست‌نمایی (Maximum Likelihood)، بیزی (Bayes)، k- نزدیک‌ترین همسایگی (k-Nearest Neighbor)، پنجره پارزن (Parzen Window) و بیزی ساده (Naïve-Bayes) جزو این گروه به‌شمار می‌رود. در طی سالیان گذشته بیشتر از روش تحلیل تمایز در طبقه‌بندی داده‌های اکتشافی استفاده شده است، و روش‌های دیگر کمتر مورد کاربرد بوده است. در صورتی که در سایر علوم این روش‌های طبقه‌بندی بیشتر به کار رفته است. برای مثال می‌توان به تحقیقات اسدیانی و سلطانی، (۱۳۸۷) در برآورد میزان بار معلق به مخزن سد، نمازی و برادران، (۱۳۸۷) در تشخیص عیوب شمش‌های فولادی، چرمی و محلول‌جی‌فر، (۱۳۸۹) در طبقه‌بندی تصاویر تانسور انتشار و میرخانی و همکاران، (۱۳۹۱) در توسعه داده کاوی در علوم پزشکی اشاره کرد. اما در زمینه علوم زمین می‌توان به این موارد اشاره کرد. (Dubois et al., 2007) از روش‌های توابع تمایز و روش k- نزدیک‌ترین همسایگی به‌همراه روش فازی برای طبقه‌بندی رخساره‌های سنگی استفاده کرده‌اند. (Cracknell and Reading, 2014) از روش k- نزدیک‌ترین همسایگی به همراه چهار روش دیگر طبقه‌بندی که جزو روش‌های یادگیری ماشین به‌شمار می‌روند، برای تفکیک واحدهای

یکی از نکات کلیدی و مهم در یک پروژه اکتشافی، نحوه ارزیابی نتایج حاصل از هر مرحله اکتشافی برای به‌دست آوردن یک استنباط علمی و اقتصادی به منظور ادامه پروژه و انجام مرحله بعدی اکتشاف است. این مشکل هنگامی که تعداد عملیات‌های اکتشافی افزایش یابد و کیفیت هر عملیات نیز متفاوت باشد، خود را بیشتر نشان می‌دهد. برای مثال، تصمیم‌گیری برای انتقال فاز یک پروژه از مرحله شناسایی به پی‌جویی بسیار ساده‌تر از تصمیم‌گیری برای انتقال فاز همان پروژه از اکتشاف مقدماتی به اکتشاف تفصیلی است. روش‌های گوناگونی مانند روش‌های مقایسه‌ای، تلفیق و مدل‌سازی برای رفع این مشکل پیشنهاد شده است. روش‌های طبقه‌بندی، مهم‌ترین بخش روش مدل‌سازی داده‌های اکتشافی را تشکیل می‌دهند. اصولاً در روش مدل‌سازی، کانسار به دو محدوده آموزشی (Train) و آزمایش (Test) تقسیم می‌شود. در محدوده آموزشی از داده‌های اکتشافی بین دو فاز عملیاتی مدلی ساخته می‌شود. سپس از این مدل و داده‌های اکتشافی محدود آزمایش، نتایج احتمالی مدل برآورد می‌شود. در صورت مثبت‌بودن نتایج برآوردشده، ادامه عملیات اکتشافی پیشنهاد می‌شود. از این روش می‌توان برای دو کانسار مشابه (از نظر ژنز) نیز استفاده کرد. روش‌هایی طبقه‌بندی به سه گروه روش‌هایی براساس نظریه بیزین، روش‌های خطی و غیرخطی تقسیم می‌شوند.

$$P(\omega_i) \quad i = 1, 2, \dots, c \quad (2)$$

در این رابطه m تعداد متغیرها، μ_i و \sum_i به ترتیب بردار میانگین ($m \times 1$ بعدی) و ماتریس واریانس - کواریانس ($m \times m$ بعدی) طبقه i ام هستند، که از روابط زیر به دست می آیند:

$$\mu_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ji} \quad (3)$$

$$\Sigma_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ji} - \mu_i)(x_{ji} - \mu_i)^T \quad (4)$$

با برآورد پارامترها و استفاده از رابطه ۲، تعلق هر نمونه جدید به هر طبقه محاسبه می شود. با توجه به این دیدگاه روش بیزی جزو روش های طبقه بندی پارامتری قرار می گیرد (Theodoridis and Koutroumbas, 2009, Duda et al., 2000).

۲-۲- روش طبقه بندی k -نزدیک ترین همسایگی

در روش k -نزدیک ترین همسایگی از یک تابع چگالی احتمالی ساده تری نسبت به روش بیزی استفاده می شود. برای این منظور احتمال اینکه k نمونه از تعداد کل N ($N = n_1 + n_2 + \dots + n_c$) نمونه در داخل فضای به حجم V و به مرکز نمونه جدید X قرار گیرند از رابطه زیر محاسبه می شود:

$$\hat{P}(x) = \frac{k}{NV} \quad (5)$$

در این روش با ثابت فرض کردن نسبت k/N (یعنی ثابت فرض کردن k)، حجم طوری تغییر می کند تا k نمونه در داخل آن قرار گیرند. حال اگر فرض شود از k نمونه k_i نمونه متعلق به طبقه ω_i باشد، به گونه ای که $k = \sum_{i=1}^c k_i$ باشد، احتمال شرطی از رابطه زیر به دست خواهد آمد:

$$g_i(x) = \hat{P}(x|\omega_i) \times P(\omega_i) = \frac{k_i}{n_i V} \times \frac{n_i}{N} = \frac{k_i}{NV} \quad (6)$$

بنابراین نمونه جدید به طبقه ای نسبت خواهد داشت که دارای تعداد نمونه های بیشتری در داخل حجم V باشد. الگوریتم روش k -نزدیک ترین همسایگی، محاسبه فاصله نمونه جدید از همه نمونه های آموزشی و مرتب کردن آنها به ترتیب صعودی و سپس انتخاب k نمونه اول خواهد بود. در صورتی که $k=1$ فرض شود (به این حالت نزدیک ترین همسایه می گویند)، نمونه جدید به طبقه ای که نزدیک ترین داده آموزشی به آن قرار دارد، تعلق خواهد داشت. در این حالت خواهیم داشت (Duda et al., 2000):

$$g_i(x) = -\min[(x - x_{ji})^T(x - x_{ji})] \quad (7)$$

که i نشان دهنده طبقه و j نشان دهنده شماره نمونه است.

سنگ شناسی توسط تصاویر ماهواره ای در منطقه ای در کشور استرالیا به خوبی بهره برده اند. نیکخواه و همکارانش، ۱۳۹۱ نیز از این روش ها به عنوان روش های شناخت آماری الگو برای تخمین تنش برجا توده سنگ توسط تعیین نقطه اثر کاریز استفاده کرده اند. تخم چی و همکارانش، ۱۳۸۷ از روش پارزن برای طبقه بندی طبقه درزه ها که علاوه بر شیب و امتداد دارای پارامترهای دیگری نیز می باشند، استفاده کرده اند.

در این مقاله ابتدا چهار روش طبقه بندی بیزی، k -نزدیک ترین همسایگی، پنجره پارزن و بیزی ساده معرفی شده است. سپس از این روش ها برای طبقه بندی پتانسیل کانی زایی طلا در کانسار ساری گونای در مجموعه داده های آموزشی که از تلفیق داده های حفاری گمانه و نمونه های زمین شیمیایی محیط خاکی بدست آمده است، استفاده شده است. در نهایت صحت این طبقه بندی در داده های آزمایشی مورد بررسی قرار گرفته است و از مدل بدست آمده برای پیش بینی پتانسیل کانی زایی طلا در نقاط فاقد حفاری گمانه استفاده گردیده است.

۲- طبقه بندی داده ها بر اساس تئوری بیزین

در صورتی که C طبقه $\omega_1, \omega_2, \dots, \omega_c$ داشته باشیم و بخواهیم نمونه جدید x را به یکی از این طبقه ها نسبت دهیم بایستی احتمال های شرطی $P(\omega_i|x)$ ($i=1, 2, \dots, c$) را محاسبه کنیم. بدیهی است که نمونه جدید x به طبقه ای تعلق دارد که دارای بیشترین احتمال شرطی باشد. اما از آنجا که محاسبه احتمال های شرطی به صورت مستقیم میسر نمی باشد از تئوری بیزین به فرمول زیر استفاده می شود:

$$P(\omega_i|x) = \frac{P(x|\omega_i)P(\omega_i)}{P(x)} \quad (1)$$

که $P(x)$ احتمال غیر شرطی و $P(\omega_i)$ احتمال اولیه هر طبقه است. احتمال اولیه هر طبقه، از تقسیم تعداد نمونه های هر طبقه بر تعداد کل نمونه ها به دست می آید. بنابراین نمونه جدید x به طبقه ای تعلق خواهد داشت که $g_i(x) = P(x|\omega_i)P(\omega_i)$ بالاتری داشته باشد (Theodoridis and Koutroumbas, 2009). در این مقاله از دو روش طبقه بندی پارامتری و دو روش طبقه بندی غیر پارامتری بر اساس نظریه یاد شده استفاده شده است، که در ادامه اصول هر روش تشریح می شود.

۲-۱- روش طبقه بندی بیزی

در روش طبقه بندی بیزی، ابتدا یک تابع توزیع احتمال (مثلاً گوسی، بتا و غیره) به داده های هر طبقه نسبت داده می شود. سپس به کمک داده های آموزشی پارامترهای این تابع چگالی احتمال برآورد می شود. برای مثال اگر توزیع داده ها نرمال فرض شود، پارامترهای میانگین و واریانس داده های به عنوان تابع چگالی گوسی برآورد می شود. بنابراین خواهیم داشت:

$$g_i(x) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right] \times$$

۳-۲- روش پنجره پارزن

داده نیاز باشد، برای دستیابی به همین دقت در یک فضای m بعدی به N^m داده نیاز خواهد بود. با توجه به کم‌بودن داده‌ها در پروژه‌های اکتشافی و لزوم صرف هزینه گزاف برای افزایش اطلاعات روش بیزی ساده شده‌ای برای طبقه‌بندی داده‌ها ارائه شده است. در این روش فرض می‌شود که متغیرها مستقل از هم باشند، بنابراین خواهیم داشت:

$$P(x|\omega_i) = \prod_{j=1}^m P(x_j|\omega_i) \quad (11)$$

در این حالت تعداد نمونه مورد نیاز به $m \times N$ عدد کاهش می‌یابد. در نتیجه احتمال شرطی از رابطه زیر به دست خواهد آمد (Theodoridis and Koutroumbas, 2009): در غیر این صورت

$$g_i(x) = \prod_{j=1}^m P(x_j|\omega_i) \times P(\omega_i) \quad (12)$$

که هر متغیر می‌تواند از تابع چگالی احتمال جداگانه‌ای مانند تابع نرمال، لاگ نرمال، گاما یا پیرسون پیروی کند. با محاسبه پارامترهای هر تابع چگالی و احتمال هر متغیر و قراردادن آنها در رابطه ۱۱ احتمال تعلق نمونه جدید به هر طبقه محاسبه می‌شود. بنابراین روش بیزی ساده جزو روش‌های طبقه‌بندی پارامتری به‌شمار می‌رود.

۳- زمین‌شناسی منطقه مطالعاتی

کانسار طلای اپی‌ترمال ساری- گونای در ۴۰ کیلومتری شمال شرقی شهرستان قروه و در جنوب شرقی استان کردستان، حدوداً در مرز بین زون سنندج- سیرجان و کمان ماگمایی ارومیه- دختر، قرار دارد. این کانسار مرتبط با توده نفوذی نیمه‌آتشفشانی بوده که در یک کمپلکس آتشفشانی نیمه‌قلیایی پتاسیم بالا متعلق به میوسن میانی تشکیل شده است. این کانسار با ذخیره کلی ۱۳۰-۱۲۰ میلیون تن با عیار میانگین ۲ گرم بر تن به عنوان یک کانسار طلای بزرگ در رده جهانی به‌شمار می‌رود (Ghorban, 2013). سنگ بستر این منطقه، اغلب سنگ آهک ژوراسیک و اندکی رسوبات آواری و سنگ‌های دگرگونی رسوبی است. این سنگ‌ها به وسیله آتشفشان‌های آندزیتی و آذرآواری‌های الیگومیوسن تا پلیستوسن پوشانده شده‌اند. جنس سنگ‌های تپه ساری‌گونای که حدوداً در مرکز محدوده قرار دارد داسیت و آندزیت‌های پورفیری نیمه‌عمیق و توف‌های دودکشی/دیاتری می‌همراه با لایت پورفیری نیمه‌قلیایی است (شکل ۱). واحدهای سنگ‌های موجود در محدوده به‌وسیله تعداد زیادی رگه سولفیدی آنتیموان و آرسنیک‌دار قطع شده است (Richards et al., 2006).

روش پنجره پارزن همانند روش k - نزدیکترین همسایگی می‌باشد. با این تفاوت که در روش k - نزدیکترین همسایگی k ثابت است و حجم آنقدر تغییر می‌کند که k نمونه در داخل آن قرار گیرند، ولی در روش پنجره پارزن حجم ثابت است و k نشان‌دهنده تعداد نمونه‌های قرار گرفته در داخل آن خواهد بود. اگر حجم V را به صورت یک ابر مکعب با طول لبه‌های h در یک فضای m بعدی در نظر گرفته شود، احتمال افتادن k نمونه در داخل آن برابر است با:

$$\hat{P}(x) = \frac{1}{Nh^m} \sum_{j=1}^N \varphi\left(\frac{x-x_j}{h}\right) \quad (8)$$

$$\varphi(u) = \begin{cases} 1 & |u_i| < \frac{1}{2}h \\ 0 & \text{در غیر این صورت} \end{cases} \quad (9)$$

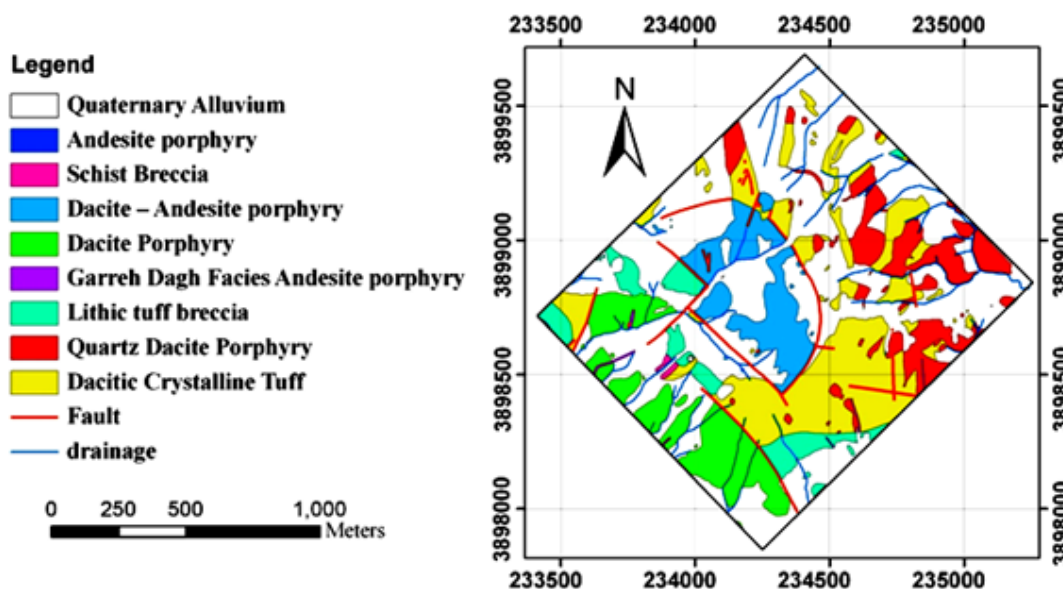
$\varphi(u)$ تابع پنجره و u_i مولفه این تابع در راستای هر یک از لبه‌های ابر مکعب است. از آنجا که تابع پنجره یک تابع چگالی می‌باشد، در نتیجه تابع $\hat{P}(x)$ نیز یک تابع چگالی خواهد بود. این تابع بایستی همواره مثبت باشد و انتگرال آن بر روی x برابر ۱ باشد. توابعی زیادی از قبیل توابع مثلثی، زنگویی و گوسی وجود دارند که ضمن برآورده کردن دو شرط فوق به نقاط نزدیک به مرکز پنجره وزن بیشتری در تابع چگالی نسبت می‌دهند. بنابراین در صورتی که فرض شود، تابع چگالی از نوع گوسی است، احتمال شرطی از رابطه زیر به دست خواهد آمد (Pan et al., 2008):

$$g_i(x) = \frac{1}{N} \sum_{j=1}^N \frac{1}{2\pi^{\frac{m}{2}} h^m} \exp\left(-\frac{(x-x_j)^T(x-x_j)}{2h^2}\right) \times P(\omega_i) \quad (10)$$

انتخاب مناسب پارامترهای k در روش k - نزدیک‌ترین همسایگی و h در روش پنجره پارزن نقش مهمی در دقت طبقه‌بندی داده‌های آزمایشی دارد. این پارامترها معمولاً به روش اعتبار متقابل تعیین می‌شوند. ضمناً این دو روش جزو روش‌های طبقه‌بندی غیر پارامتری محسوب می‌شوند.

۴-۲- روش بیزی ساده

یکی از مشکلات طبقه‌بندی به روش‌های یادشده نیاز داشتن به داده‌های آموزشی زیاد برای دستیابی به دقت مطلوب است. در صورتی که برای رسیدن به یک دقت طبقه‌بندی مشخص در یک فضای یک بعدی به N



شکل ۱- نقشه زمین‌شناسی کانسار طلای اپی‌ترمال ساری- گونای (Wilkinson and Damien, 2005)

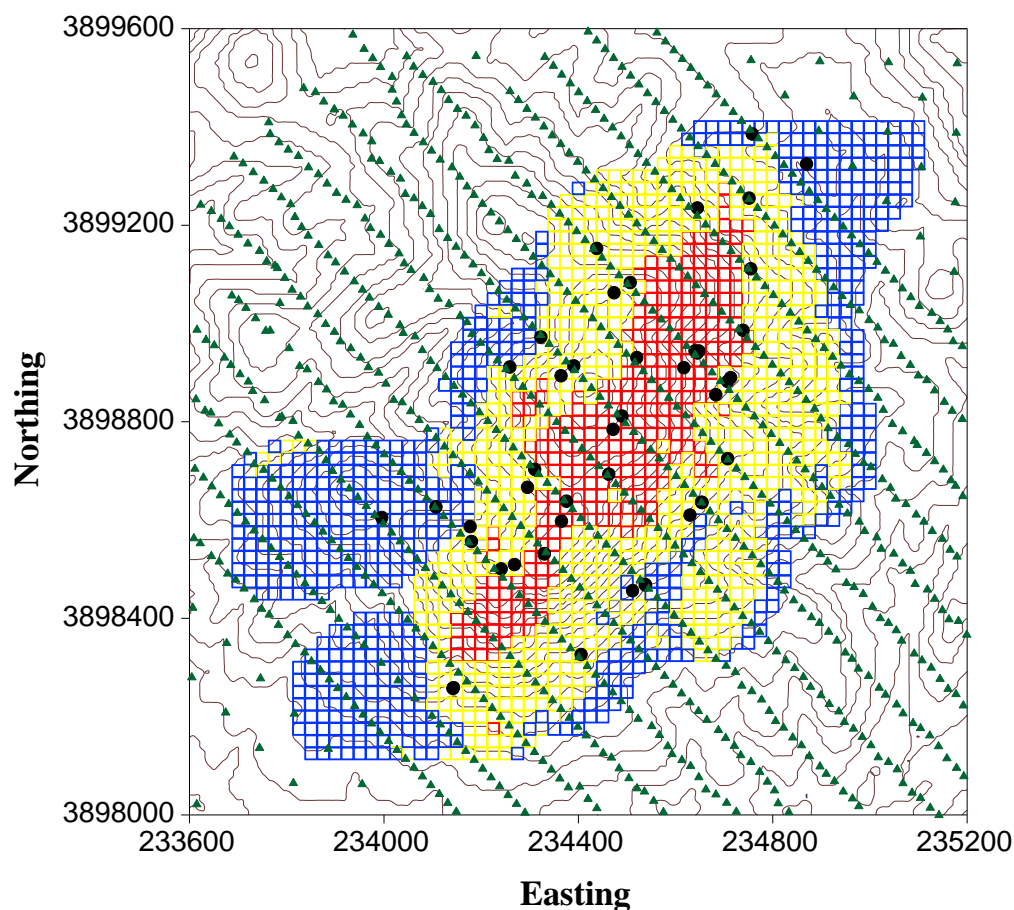
(شکل ۲). طول متوسط مغزه‌های آنالیز شده ۱ متر است. هر مغزه برای ۴۶ عنصر به روش ICP-MS و طلا به روش Fire Assay تجزیه شده است، که در این مقاله فقط از نتایج تجزیه طلای مغزه‌ها برای طبقه‌بندی استفاده شده است. به منظور محاسبه عیار متوسط هر بلوک $25 \times 25 \times 10$ متری (ابعاد بهینه بلوک برای محاسبه ذخیره) در محدوده حفاری‌ها از روش کریجینگ استفاده شده است. سپس برای هر سلول 25×25 متری در سطح زمین (شکل ۲) حاصلضرب عیار بلوک‌های بالاتر از عیار حد (عیار حد 0.5 گرم بر تن) در ارتفاع بلوک‌ها به‌عنوان شاخص تولید آن سلول محاسبه شده است. در جدول ۱ پارامترهای آماری متغیر شاخص تولید نشان داده شده است. از تعداد ۱۷۶۷ سلول، ۶۶۰ سلول فاقد کانی‌زایی اقتصادی طلا بوده که به‌عنوان طبقه زمین در نظر گرفته است. برای تعیین تعداد طبقه‌های (جوامع) بقیه سلول‌ها می‌توان از نقطه عطف در نمودار احتمال استفاده کرد (Sinclair, 1989). شکل ۳ نشان می‌دهد که داده‌های شاخص تولید از دو طبقه تشکیل شده است. در این شکل نقطه عطف (در مقدار $4050 \text{ ppb} \times \text{m}$) به‌عنوان آستانه تفکیک طبقه‌ها در نظر گرفته شده است. بنابراین متغیر شاخص تولید قابل تفکیک به ۳ جامعه A به‌عنوان طبقه‌ای با شاخص تولید صفر، B به‌عنوان طبقه‌ای با شاخص تولید متوسط و C به‌عنوان طبقه‌ای با شاخص تولید بالاست. بدیهی است که بالا بودن شاخص تولید در هر سلول نشان‌دهنده کانی‌زایی مناسب طلا در زیر آن سلول خواهد بود. در شکل ۲ پراکندگی سلول‌های این سه طبقه نشان داده شده است.

در محدوده اکتشافی ساری- گونای، ۱۷۲۴ نمونه زمین‌شیمیایی از محیط خاک برداشت شده است (شکل ۲). این نمونه‌ها نیز برای ۴۷ عنصر تجزیه شده است، که در ابتدا داده‌های سنسورد که تعداد آنها کمتر از ۱۰ درصد است، به روش جایگزینی ساده برآورد شده است. سپس مقادیر خارج از ردیف، به روش دورفل جایگزین شده‌اند.

مجموعه‌ای از ساختارهای بزرگ با روند شمال غربی- جنوب شرقی و مجموعه‌ای از ساختارهای کوچک متقاطع در این ناحیه را می‌توان مؤثر بر توسعه آتشفشانی و کانی‌سازی در منطقه دانست. سامانه ساری- گونای یک سامانه بزرگ اپی‌ترمال با سولفیداسیون پایین همراه با پیوستگی قوی کانی‌سازی در راستای قائم است. این سامانه دارای هسته‌ای غنی از طلا- استیپنیت- جیوه است که به‌وسیله هاله‌ای از آرسنیک و دورتر از آن به‌وسیله هاله‌ای از سرب- نقره- روی- مس محاصره شده است. کانی‌سازی در ساری- گونای شامل دو فاز است. فاز اول کانی‌سازی شامل یک سامانه پورفیری مزوترمال همراه با دگرسانی پتاسیمی، سربیسیت و کوارتز- تورمالین- فلدسپار پتاسیم است. این مرحله منجر به تشکیل کانی‌سازی مس پورفیری کم‌عیار محدود به دایک‌های همراه دگرسانی پتاسیمی و اطراف آنها شده است. فاز دوم کانی‌سازی در دمای پایین‌تر و همراه با دگرسانی محدودتری رخ داده است. در این فاز سیلیسی شدن محلی و همچنین تبدیل سربیسیت تشکیل شده در فاز اول به فنگیت، و تشکیل پیریت آرسنیک‌دار به صورت پراکنده در اطراف ذرات پیریت پورفیروبلاستیکی فاز اول دیده می‌شود. کانی‌سازی طلا در هر دو فاز کانی‌سازی رخ داده است، در فاز اول به‌صورت میانبار در کانی پیریت و در فاز دوم به‌صورت محلول جامد در ترکیب پیریت آرسنیک‌دار یافت می‌شود (Wilkinson and Damien, 2005).

۴- پردازش اولیه داده‌های اکتشافی

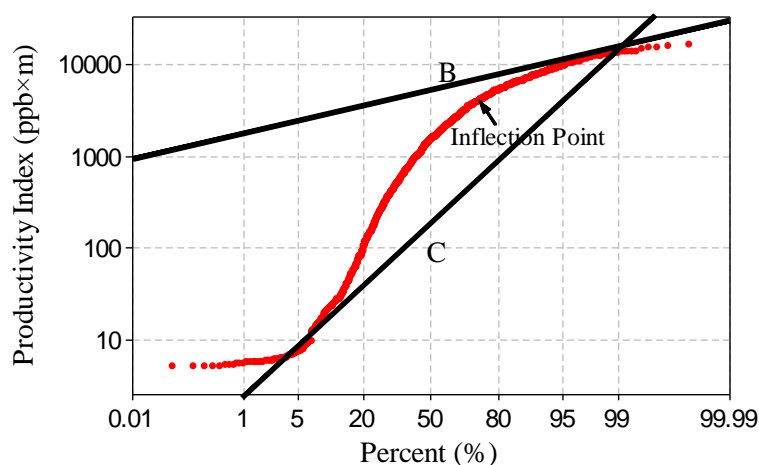
به منظور رسم نقشه پتانسیل‌یابی طلا از داده‌های گمانه‌ای به‌عنوان متغیر مستقل و داده‌های زمین‌شیمیایی محیط خاکی به‌عنوان متغیرهای وابسته به شرح زیر استفاده شده است. در محدوده کانسار ساری- گونای ۴۶ گمانه اکتشافی حفاری شده است



شکل ۲- موقعیت گمانه‌های اکتشافی (دایره‌های سیاه)، سلول‌های دارای شاخص تولید صفر (مربع‌های آبی)، سلول‌های دارای شاخص تولید متوسط (مربع‌های زرد)، سلول‌های دارای شاخص تولید بالا (مربع‌های قرمز) و نمونه‌های زمین‌شیمیایی محیط خاکی (مثلث‌های سبز تیره) بر روی نقشه توپوگرافی محدوده اکتشافی

طبقه‌بندی استفاده شده است. زیرا در این روش علاوه بر نرمال‌سازی داده‌ها، استانداردسازی نیز صورت می‌گیرد. بنابراین میانگین امتیاز نرمالی کل نمونه‌های زمین‌شیمیایی هر سلول برای هر عنصر به عنوان امتیاز نرمالی آن عنصر در آن سلول در نظر گرفته شده است.

در جدول ۱ پارامترهای آماری داده‌های اولیه ۱۶ عنصر مورد استفاده در طبقه‌بندی آورده شده است (از ۴۷ عنصر، ۱۶ عنصر به عنوان متغیرهای موثر در طبقه‌بندی به روش الگوریتم Branch and Bound انتخاب شده‌اند). از داده‌های زمین‌شیمیایی بهنجار شده به روش امتیاز نرمالی برای



شکل ۳- نمودار احتمال متغیر شاخص تولید به همراه موقعیت نقطه عطف و جوامع تفکیک شده

جدول ۱- پارامترهای آماری متغیر وابسته (شاخص تولید) و متغیرهای مستقل (عناصر شیمیایی) در مدل سازی به روش های بیزین

۰.۵۷	نام متغیر (واحد)	میانگین	انحراف معیار	چولگی	کشیدگی
گمانه	شاخص تولید (ppb×m)	۱۷۵۴/۴	۲۹۳۱/۹	۲/۱	۴/۲
داده زمین شیمیایی محیط خاکی	آرسنیک (ppm)	۳۳۱/۷	۵۷۰/۰	۸/۴	۱۰۹/۹
	باریم (ppm)	۴۲۷/۳	۱۹۹/۱	۲/۲	۹/۱
	مس (ppm)	۳۶/۹	۳۹/۸	۵/۸	۴۵/۵
	آهن (/)	۳/۸	۰/۸	۰/۴	۱/۹
	گالیم (ppm)	۵/۴	۲/۷	۲/۸	۲۸/۱
	لیتیم (ppm)	۱۳/۵	۴/۷	۰/۵	۰/۷
	سرب (ppm)	۳۳۷/۵	۶۶۰/۴	۸/۰	۸۹/۲
	گوگرد (/)	۰/۰۸	۰/۱۵	۱۲/۳	۲۵۴/۱
	آنتیمون (ppm)	۱۴۷/۷	۵۱۰/۶	۱۳/۳	۲۲۹/۳
	اسکاندیم (ppm)	۶/۴	۱/۹	۱/۲	۳/۲
	قلع (ppm)	۲/۵	۰/۱	۴۱/۵	۱۷۲۴/۰
	وانادیم (ppm)	۵۴/۸	۱۴/۴	۰/۸	۱/۸
	ایتیم (ppm)	۱۵/۰	۲/۹	۰/۴	۱/۰
	روی (ppm)	۱۵۱/۳	۹۶/۸	۴/۹	۵۳/۷
	زیرکون (ppm)	۳/۶	۱/۹	۲/۱	۷/۸
طلا (ppb)	۱۵۷/۲	۴۸۶/۱	۷/۳	۷۳/۴	

۵- طبقه بندی داده ها

از ۱۷۶۷ سلول که مقدار شاخص تولید در آنها برآورد شده است، ۳۸۷ سلول دارای هم پوشانی با نمونه های زمین شیمیایی محیط خاکی هستند (به دلیل برداشت نمونه های زمین شیمیایی محیط خاکی با فواصل نیم رخ ۱۰۰ متری، برخی سلول ها فاقد نمونه زمین شیمیایی هستند). از این تعداد داده، ۶۵ درصد به عنوان داده های آموزشی (از ۲۵۲ عدد، ۸۸ نمونه متعلق به طبقه A، ۱۲۰ نمونه متعلق به طبقه B و ۴۴ نمونه متعلق به طبقه C) برای ساختن تابع احتمال و از ۳۵ درصد باقیمانده (۱۳۵ عدد، ۴۷ نمونه متعلق به طبقه A، ۶۵ نمونه متعلق به طبقه B و ۲۳ نمونه متعلق به طبقه C) به عنوان داده های آزمایشی برای بررسی درستی مدل به دست آمده و تعیین مقدار خطای طبقه بندی استفاده شده است.

اولین گام در طبقه بندی کاهش تعداد متغیرهای شرکت کننده در تابع احتمال شرطی است، که به آن کاهش ابعاد نیز می گویند. چرا که بعضی از متغیرهای مستقل می توانند ترکیب خطی یا غیر خطی از متغیرهای مستقل دیگر باشند. وجود این متغیرها در تابع احتمال جز افزایش محاسبات، کارایی دیگری نخواهند داشت. ضمناً کاهش تعداد متغیرها منجر به افزایش قابلیت تعمیم پذیری تابع احتمال شرطی و کاهش خطای طبقه بندی داده های آزمایشی نیز می شود. برای این منظور از الگوریتم Branch and Bound با معیار نزدیک ترین همسایگی استفاده شده است. از ۴۷ عنصر موجود، ۱۶ عنصر Au, Fe, Sr, Ga, Zr, Cu, Ba, Pb, Zn, Sb, V, S, Li, موجود، ۱۶ عنصر Y, Sc و As به عنوان متغیرهای انتخاب شده استخراج شده است. حضور

عناصر S, Au, Cu, Pb, Zn مرتبط با عناصر اصلی، As و Sb مرتبط با عناصر فرعی کانی زایی طلای ایبی ترمال سولفیداسیون پایین و عناصر Y, Sc و Zr به عنوان عناصر کم تحرک در منطقه قابل تفسیر است (مقصودی و همکاران، ۱۳۸۴).

تابع احتمال شرطی چهار روش گفته شده به کمک نرم افزار MATLAB با استفاده از عناصر استخراج شده، داده های آموزشی و فرمول های ۲، ۶، ۱۰ و ۱۲ به دست آمده است. سپس به منظور بررسی درستی تابع به دست آمده، داده های آزمایشی توسط این توابع طبقه بندی شده اند، که نتایج آن در جدول ۲ نشان داده شده است. در روش بیزی و بیزی ساده از توابع چگالی احتمال نرمال برای کلیه متغیرها استفاده شده است. همچنین برای محاسبه بهینه پارامترهای k و h در دو روش دیگر نیز از الگوریتم اعتبار متقابل Leave-one-out استفاده شده است. برای این منظور جستجو در فواصل ۱ تا ۱۰ برای پارامتر k و ۰/۵ تا ۲ برای پارامتر h انجام شده است. بهترین نتایج با اعداد k=۱ با خطای ۲۹/۸ درصد و ۰/۹=h با خطای ۲۲/۶ درصد برای داده های آموزشی به دست آمده است. جدول ۲ نشان می دهد که روش بیزی، پنجره پارزن، k- نزدیک ترین همسایگی و بیزی ساده به ترتیب عملکرد بهتری داشته اند. بالاترین درستی تفکیک طبقه A مربوط به روش بیزی و بیزی ساده با مقدار ۷۴/۵ درصد، طبقه B مربوط به روش بیزی با مقدار ۷۵/۴ درصد و طبقه C مربوط به روش بیزی و پنجره پارزن با مقدار ۶۰/۹ درصد است. بنابراین روش بیزی در تفکیک طبقه های مختلف، عملکرد نسبتاً بهتری را با بالاترین درستی داشته است. بعد از روش بیزی، روش پنجره پارزن نیز در جایگاه دوم قرار دارد. روش های k- نزدیک ترین همسایگی و بیزی ساده نیز در رتبه های بعدی قرار دارند.

جدول ۲- نتایج طبقه‌بندی داده‌های آزمایشی به چهار روش نظریه بی‌زین

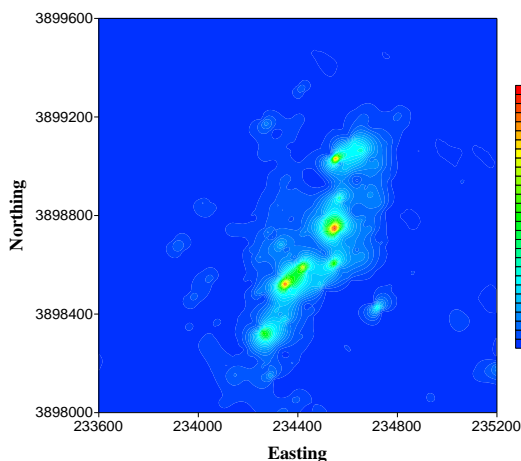
روش بی‌زین			روش پنجره پارزن			روش k- نزدیک‌ترین همسایگی			روش بی‌زین			گروه طبقه‌بندی شده						
گروه اولیه			گروه اولیه			گروه اولیه			گروه اولیه									
C	B	A	C	B	A	C	B	A	C	B	A							
۱	۱۷	۳۵	۰	۱۳	۳۴	۱	۱۴	۳۳	۱	۱۲	۳۵	A						
۱۰	۳۵	۱۰	۹	۴۷	۱۳	۱۰	۴۲	۱۲	۸	۴۹	۱۲	B						
۱۲	۱۳	۲	۱۴	۵	۰	۱۲	۹	۲	۱۴	۴	۰	C						
%۵۲/۲			%۵۳/۸			%۷۴/۵			%۶۰/۹			%۷۵/۴			%۷۴/۵			صحت طبقه‌بندی هر گروه
%۶۰/۱۷			%۷۰/۴			%۶۴/۴			%۷۲/۶									صحت طبقه‌بندی هر روش

بی‌زین به حدود ۷۳-۶۱ درصد افزایش یافته است. بنابراین روش‌های طبقه‌بندی یاد شده می‌تواند در پتانسیل یابی طلا در کانسار ساری‌گونای عملکرد بهتری داشته باشد.

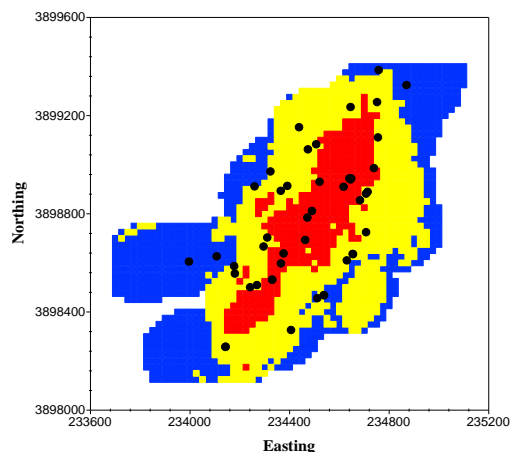
از کل اطلاعات نمونه‌های زمین‌شیمیایی خاکی استفاده کرده و به کمک چهار تابع احتمال شرطی که به روش‌های گفته شده به دست آمده، مدل کانی‌زایی پیش‌بینی شده است. نتایج این مدل‌سازی در شکل ۷ ارائه شده است. مقایسه اولیه بین نتایج مدل‌های پیش‌بینی کانی‌زایی طلا به کمک چهار روش طبقه‌بندی در شکل ۷ با نتایج شکل ۴، نشان می‌دهد که تقریباً هر چهار روش طبقه‌بندی توانسته‌اند محدوده دارای کانی‌زایی بالا (ناحیه قرمز رنگ) و متوسط (ناحیه زرد رنگ) را به کمک داده‌های زمین‌شیمیایی سطحی از زمینه (ناحیه آبی رنگ) تفکیک نمایند. اما به لحاظ شکل و وسعت هر یک از طبقه‌ها، روش بی‌زین و سپس روش پنجره پارزن بهترین تطبیق را با شکل ۴ نشان می‌دهند. این مطلب با نتایج به دست آمده از طبقه‌بندی داده‌های آزمایشی نیز که قبلاً در جدول ۲ ذکر شد، همخوانی دارد. در ضمن دیگر مناطق مستعد کانی‌زایی طلا که در شکل ۷ نشان داده شده است، اما در شکل ۴ دیده نمی‌شود، نیز می‌تواند برای حفاری‌های تکمیلی در فاز نهایی اکتشاف کانسار ساری-گونای پیشنهاد شوند. اولویت اول حفاری‌های تکمیلی در نواحی قرمز رنگ و سپس در نواحی زرد رنگ است.

در شکل ۴ وضعیت پراکندگی سلول‌های متعلق به سه طبقه A، B و C که از اطلاعات روش کریجینگ داده‌های گمانه‌ای به دست آمده است به همراه موقعیت دهانه گمانه‌ها نشان داده شده است. مطابق شکل از ۴۶ گمانه حفاری شده، ۵ گمانه دهانه آن در طبقه A، ۲۸ گمانه در طبقه B و ۱۳ گمانه در ناحیه طبقه C قرار دارند. بنابراین در صورتی که فرض شود الگوی پراکندگی نشان داده شده در شکل ۴ بهترین تطبیق را با واقعیت پراکندگی کانی‌زایی طلا در کانسار ساری-گونای دارد، می‌توان نتیجه گرفت که موقعیت حدود ۸۹ درصد گمانه‌ها به درستی انتخاب شده‌اند.

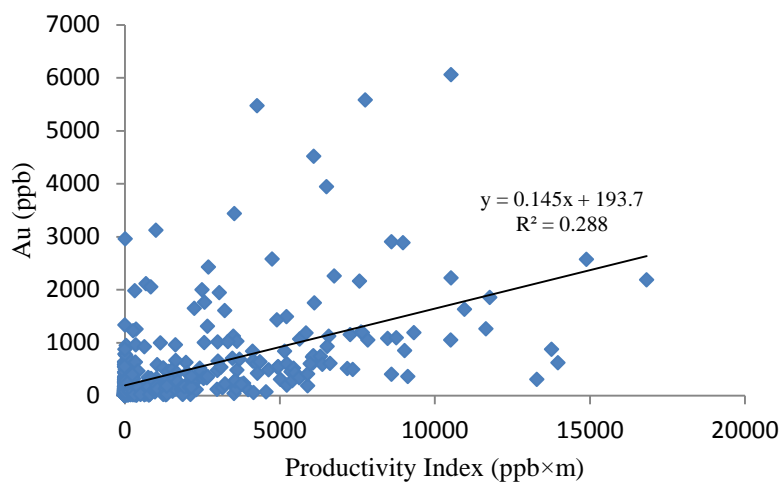
به منظور مقایسه میزان همسنگی و امکان پیش‌بینی شاخص تولید (نتایج به دست آمده در شکل ۴) در کانسار ساری-گونای به کمک مقادیر طلا داده‌های زمین‌شیمیایی محیط خاکی، شکل‌های ۵ و ۶ رسم شده است. در شکل ۵ نقشه کنتوری مقادیر طلا بر حسب ppb نشان داده شده است. مقایسه نتایج شکل‌های ۴ و ۵ نشان می‌دهد که نقشه کنتوری مقادیر طلا فقط توانسته است، ناحیه متعلق به طبقه C همراه با کمی جابه‌جایی به سمت جنوب شرقی شبیه‌سازی کند. نمودار پراکندگی بین مقادیر شاخص تولید و میزان طلا در سلول‌های مشترک نیز در شکل ۶ نشان داده شده است. این نمودار نشان می‌دهد که برآورد مقدار شاخص تولید توسط مقادیر طلای نمونه‌های زمین‌شیمیایی محیط خاکی دارای اعتبار ۲۸/۸ درصد است. در صورتی که این اعتبار توسط روش‌های طبقه‌بندی بر اساس نظریه



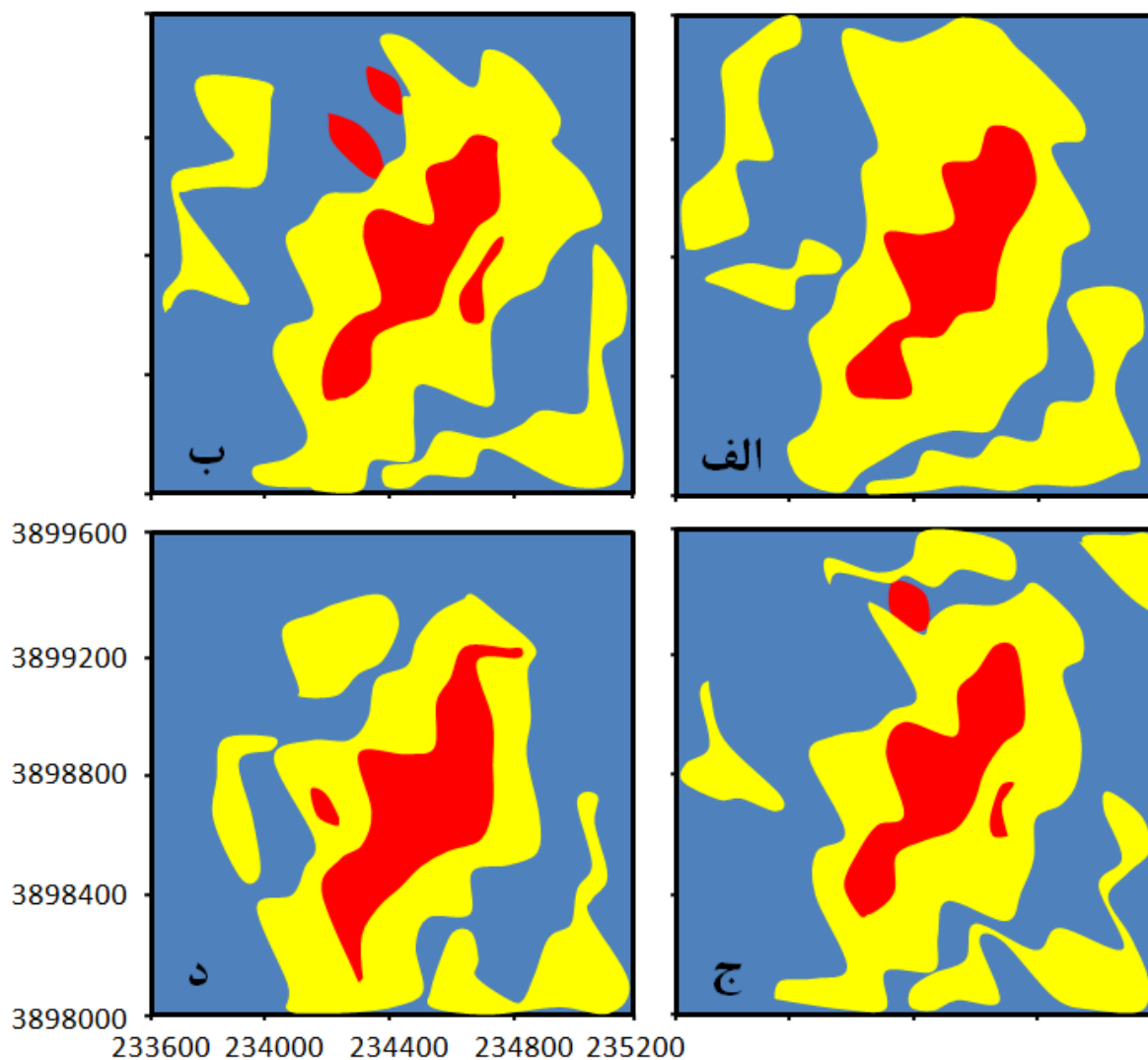
شکل ۵- نقشه کنتوری طلا به دست آمده از داده‌های زمین‌شیمیایی خاک (واحد اندازه‌گیری طلا در مقیاس میلی‌های ppb است)



شکل ۶- نقشه پراکندگی وضعیت کانی‌زایی طلا و موقعیت گمانه‌ها در کانسار ساری-گونای (طبقه A به رنگ قرمز، طبقه B به رنگ زرد و طبقه C به رنگ آبی)



شکل ۶- نمودار پراکندگی مقادیر شاخص تولید برحسب مقادیر طلای نمونه‌های زمین‌شیمیایی خاک



شکل ۵- نقشه پراکندگی مدل‌سازی کانی‌زایی طلا در سه طبقه به روش بی‌زی (الف)، روش k- نزدیک‌ترین همسایگی (ب)، روش پنجره پارزن (ج) و روش بی‌زی ساده (د) در کانساری ساری- گونای (ترکیب رنگ‌های انتخابی مشابه شکل ۴ است)

نتیجه‌گیری

تخم‌چی، ب.؛ معماریان، ح.؛ مشیری، ب.؛ احمدی نویری، ح.، ۱۳۸۷، "بررسی انتقادی روش‌های متداول طبقه‌بندی دسته درزه‌ها" دومین کنفرانس مهندسی معدن ایران، دانشگاه تهران.

چرمی، م.؛ محلوجی فر، ع.، ۱۳۸۹، "مقایسه عملکرد الگوریتم‌های نمو سطح آماری پارامتریک و غیر پارامتریک در بخش‌بندی تصاویر تانسور انتشار" ششمین کنفرانس ماشین بینایی و پردازش تصویر، اصفهان، دانشگاه اصفهان.

مقصودی، ع.؛ رحمانی، م.؛ رشیدی، ب.، ۱۳۸۴، "کانسارها و نشانه‌های معدنی طلا در ایران" انتشارات آری زمین، تهران، ۳۶۴ صفحه.

میرخانی، ع.؛ پورنیک، ا.؛ توفیقی، ش.، شهرابی، ج.، ۱۳۹۱، "کاربرد داده‌کاوی پیش‌بینی کننده در پزشکی: مقایسه الگوهای منتخب درشناسایی زود هنگام بیماری مننژیوم باکتریایی در ایران" هشتمین کنفرانس بین‌المللی مهندسی صنایع، تهران، انجمن مهندسی صنایع ایران، دانشگاه صنعتی امیر کبیر.

نمازی، ن.؛ برادران، ش.، ۱۳۸۷، "تشخیص عیوب شمش‌های فولادی با استفاده از آستانه‌گیری پارزن" پنجمین کنفرانس ماشین بینایی و پردازش تصویر، دانشگاه تبریز.

نیکخواه، م.؛ احمدی، م.، قزوینیان، ع.، ۱۳۹۱، "ارائه روشی جدید برای تعیین نقطه اثر کاریز در سنگ‌ها" نشریه مهندسی معدن، دوره هفتم، شماره هفدهم، صفحه ۶۱ تا ۷۲.

Cracknell, M.J., Reading, A.M., 2014, "Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information", *Computers & Geosciences*, V. 63, P. 22-33.

Dubois, M.K., Bohling, G.C., Chakrabarti, S., 2007, "Comparison of four approaches to a rock facies classification problem", *Computers & Geosciences*, V. 33, P. 599-617.

Duda, R.O., Hart, P.E., Stork, D.G., 2000, "Pattern Classification", second edition, John Wiley & Sons, P.680.

Ghorban, M., 2013, "The economic geology of Iran: Mineral Deposits and Natural Resources", Springer Publication, Netherlands, P.569.

Pan, Z.W., Xiang, D.H., Xiao, Q.W., Zhou, D.X., 2008, "Parzen windows for multi-class classification", *Journal of Complexity*, V. 24, P. 606-618.

Richards, J.P., Wilkinson, D., Ullrich, T., 2006, "Geology of the Sari Gunay epithermal gold deposit northwest Iran", *Economic Geology*, V. 101, P. 1455-1496.

Sinclair, A.J., 1989, "Application of probability graphs in mineral exploration", *The Association of Exploration Geochemists, Special Volume No. 4*, P.95.

Theodoridis, S., Koutroumbas, K., 2009, "Pattern Recognition", fourth edition, Academic Press, P.967.

Wilkinson, L., Damien, A., 2005, "Geology and mineralization of the sari gunay gold deposits, kordestan province Iran", *Rio-Tinto Ltd Technical Repo*.

اگرچه حفر گمانه مهم‌ترین ابزار اکتشافی به‌شمار می‌رود، اما هزینه بالا، زمانبر بودن، پوشش منطقه کوچک و از همه مهم‌تر امکان عدم برخورد به ماده معدنی از معایب این ابزار اکتشافی به‌شمار می‌رود. در این مقاله نشان داده شده که به کمک روش‌های طبقه‌بندی بر روی داده‌های زمین‌شیمیایی سطحی، می‌توان با پیش‌بینی موقعیت کانی‌زایی زیرسطحی از معایب آن کاست. نتایج پیش‌بینی شاخص تولید توسط مقادیر طلای داده‌های زمین‌شیمیایی محیط خاکی نشان می‌دهد که فقط امکان تعیین مقادیر با شاخص تولید بالا همراه با کمی جابه‌جایی و اعتبار ۲۸/۸ درصد ممکن است. در حالی که به کمک روش‌های طبقه‌بندی بیزین افزون بر برآورد همه مناطق کانی‌زایی شده، اعتبار برآورد نیز در حد قابل قبولی افزایش می‌یابد. بنابراین استفاده از گمانه‌های حفرشده و ارتباط دادن آنها با داده‌های زمین‌شیمیایی سطحی به روش‌های طبقه‌بندی می‌تواند در موقعیت‌یابی گمانه‌های تکمیلی بهتر از نتایج تعیین فقط بی‌هنجاری طلا عمل کند.

هر چهار روش طبقه‌بندی مورد استفاده قادر به تفکیک نواحی با کیفیت کانی‌زایی مختلف طلا در کانسار ساری- گونای هستند. اما روش طبقه‌بندی بیزی با درستی ۷۲/۶ درصد و روش پنجره پارزن با درستی ۷۰/۴ درصد در طبقه‌بندی داده‌های آزمایشی و تعیین دقیق‌تر شکل و وسعت نواحی مختلف کانی‌زایی نسبت به دو روش دیگر عملکرد بهتری داشته‌اند.

همچنین با پیش‌بینی وضعیت کانی‌زایی طلا در نواحی فاقد حفاری گمانه در کانسار ساری- گونای، مناطق مستعد حفاری تکمیلی نیز پیشنهاد شده است. یکی دیگر از نکات مهم پیشنهادی این مقاله استفاده از توابع احتمال شرطی به‌دست آمده برای پیش‌بینی کانی‌زایی در سایر کانسارهای طلای اپی‌ترمال سولفیداسیون پایین مانند کانسار آق داغ (که در مجاورت کانسار ساری- گونای قرار دارد) است. با این عمل در وقت و هزینه پروژه صرفه‌جویی قابل ملاحظه‌ای صورت خواهد گرفت.

تقدیر و تشکر

بدین وسیله از شرکت معدنی زرکوه، مالک معدن طلای ساری‌گونای، به دلیل در اختیار گذاشتن داده‌ها اکتشافی تشکر می‌شود.

مراجع

اسدیانی یکتا، ا.؛ سلطانی، ف.، ۱۳۸۷، "مقایسه برآورد میزان بار ورودی به مخزن سد اکباتان با استفاده از مدل استنتاج تطبیقی عصبی- فازی و الگوریتم k- نزدیک‌ترین همسایگی" هفتمین کنفرانس هیدرولیک ایران، دانشگاه صنعت آب و برق، تهران.