



دانشگاه علم و فناوری گیلان

نشریه پژوهش در نشخوارکنندگان

جلد هفتم، شماره اول، ۱۳۹۸

<http://ejrr.gau.ac.ir>

ارزیابی مقایسه‌ای روش‌های بوستینگ و بیز A در چالش‌های مختلف معماری ژنومی

صفات گسسته و پیوسته

*یوسف نادری

^۱استادیار ژنتیک و اصلاح دام، باشگاه پژوهشگران جوان و نخبگان، دانشگاه آزاد اسلامی، آستارا

تاریخ دریافت: ۹۷/۹/۶؛ تاریخ پذیرش: ۹۸/۲/۱۱

چکیده

سابقه و هدف: انتخاب ژنومی با بهره‌گیری از روش‌های مختلف آماری نقش مهمی در افزایش بهره‌وری اقتصادی و پیشرفت ژنتیکی صفات گسسته و پیوسته ایفا کرد. در این پژوهش، عملکرد روش‌های بوستینگ و بیز A در برآورد ارزش‌های اصلاحی ژنومی صفات آستانه‌ای دودویی و پیوسته در تراکم مختلف نشانگری با استفاده از معماری‌های مختلف ژنومی مورد بررسی قرار گرفت.

مواد و روش‌ها: داده‌های ژنومی از طریق نرم‌افزار QMSim با سطوح متفاوت وراثت‌پذیری (۰/۱ و ۰/۳)، سطوح مختلف LD (کم و زیاد)، تراکم‌های متفاوت جایگاه‌های صفات کمی (۱۵۰ و ۴۵۰) و تراکم مختلف نشانگری (۱۰K و ۵۰k) برای تعداد ۳۰ کروموزم شبیه‌سازی شدند. جهت ایجاد فنوتیپ آستانه‌ای دودویی در مجموعه مرجع، افراد هر نسل بر اساس فنوتیپ پیوسته در خروجی QMSim رتبه بندی شدند، سپس فنوتیپ آستانه‌ای افراد، وابسته به میانگین جمعیت شبیه‌سازی شده به ترتیب کد صفر (پایین‌تر از میانگین صفت) و کد یک (بالا‌تر از میانگین صفت) در نظر گرفته شد. در نهایت، ارزش‌های اصلاحی ژنومی با استفاده از روش‌های بوستینگ و بیز A محاسبه و جهت ارزیابی صحت ژنومی صفات آستانه‌ای و پیوسته مورد استفاده قرار گرفتند.

نتایج: روش بوستینگ دامنه گسترده‌ای از صحت ژنومی در مقایسه با روش بیز A با تغییرات تراکم نشانگرها نشان داد. روش بوستینگ در مقایسه با روش بیز آستانه‌ای A به ترتیب افزایشی ۶/۳ و ۷/۳ درصدی در صحت ژنومی صفات آستانه‌ای برای تراکم‌های نشانگری ۱۰k و ۵۰k نشان داد. عملکرد بیز A برای صفات با توزیع فنوتیپی پیوسته بطور قابل توجهی بیشتر از روش بوستینگ بود، خصوصاً هنگامی که سناریوهای با تراکم نشانگری پایین استفاده شدند. ساختار معماری ژنومی از جمله وراثت‌پذیری، تعداد QTL و LD از فاکتورهای موثر بر صحت ژنومی روش‌های بیز و بوستینگ بودند. در این راستا نقش وراثت‌پذیری بر عملکرد هر یک از این روش‌ها مشهودتر بود. در مجموع، صحت‌های ژنومی روش بیز برای نوسانات تعداد QTL و روش بوستینگ برای نوسانات سطوح LD، حساسیت بیشتری نشان دادند. در تراکم بالای نشانگرها و برای صفات با فنوتیپ آستانه‌ای، بیشترین و کمترین میزان صحت ژنومی به ترتیب برای روش بوستینگ (۰/۵۹۸) و بیز آستانه‌ای A (۰/۵۱۰) هنگامی بود که تعداد بالای QTL وجود داشت. برای صفات پیوسته، بیشترین و کمترین میزان صحت ژنومی به ترتیب برای روش بیز A (۰/۷۰۲) و بوستینگ (۰/۵۶۹) در تعداد QTL پایین مشاهده شد. اثر مثبت افزایش LD بر صحت ژنومی روش‌های بوستینگ و بیز A در سناریوهای با تراکم نشانگر پایین نسبت به سناریوهای با تراکم نشانگری بالا مشهودتر بود.

*نویسنده مسئول: yousefnaderi@gmail.com

نتیجه‌گیری: روند کلی نتایج این تحقیق نشان داد که روش بوستینگ در ارزیابی ژنومی صفات آستانه‌ای و روش بیز A در ارزیابی صفات پیوسته بهترین عملکرد را نشان می‌دهند.

واژه‌های کلیدی: صفات آستانه‌ای، پیش‌بینی ژنومی، وراثت‌پذیری، یادگیری ماشین، عدم تعادل پیوستگی

مقدمه

تکامل ژنتیکی صفات با ماهیت پیوسته (کمی) در حیوانات اهلی به گونه‌ای بوده است که منجر به ایجاد سیستم توارثی پیچیده‌ای در این صفات شده است. در آغاز، مطالعه و کمی‌سازی اساس ژنتیکی این صفات توسط متخصصین ژنتیک منجر به طراحی این سوال شد بود که آیا انتخاب در اشکال مختلف آن منجر به افزایش سود و پیشرفت ژنتیکی پایدار در این صفات خواهند شد یا خیر؟ در این راستا تحقیقات زیادی موفقیت انتخاب ژنومی در بهبود پیشرفت ژنتیکی این صفات و دستیابی به سود اقتصادی را گزارش نموده‌اند (۱۶ و ۳۶). با این حال بسیاری از صفاتی که اهمیت زیست‌شناختی و اقتصادی دارند به شیوه‌ای ناپیوسته تغییر می‌کنند. از دیدگاه اصلاح‌گران و همچنین پرورش‌دهندگان در برنامه‌های امروزی اصلاح نژادی، پیشرفت‌های قابل توجه در بهبود حیوانات اهلی و افزایش سود اقتصادی، مستلزم ورود مستقیم به صفات رفتاری، سلامت حیوان و مقاومت به بیماری‌ها است که دارای ماهیتی گسسته هستند (۸). در نتیجه به خاطر عدم رابطه خطی بین متغیرها و صفت وابسته، عدم توزیع نرمال متغیر وابسته و توزیع متفاوت واریانس این صفات بدیهی است که روش‌های انتخاب ژنومی برای صفات پیوسته نمی‌تواند راه حل مناسبی برای دستیابی به پیشرفت ژنتیکی در این گونه صفات باشد. بنابراین، استفاده بهتر از انتخاب ژنومی و کاربرد آن در اصلاح نژاد دام هم گام با توسعه روش‌های مولکولی در این صفات، مستلزم به کارگیری مدل‌های آماری مناسب به‌منظور

ارزیابی دقیق‌تر است (۳۴).

فهم بهتر آنچه در انتخاب ژنومیک پیش‌بینی می‌شود به روش آنالیز آماری وابسته است (۲۴). روش‌های آماری مختلفی برای تخمین اثرات نشانگرها و پیش‌بینی ارزش‌های اصلاحی ژنومی ارائه شده است. تفاوت عمده این روش‌ها، به فرضیات در نظر گرفته شده برای مدل ژنتیکی پشت صحنه آن‌ها وابسته است (۶). موویسن و همکاران (۲۰۰۱) چگونگی استفاده از ابزار ریاضی و به خصوص تئوری بیز را در پیش‌بینی اثر نشانگرها جهت برآورد ارزش‌های اصلاحی ژنومی تشریح نمودند (۲۰) و بعد از آن محققین دیگری چگونگی استفاده از آن را در برنامه‌های مختلف ژنومی اصلاح نژادی صفات پیوسته مورد ارزیابی قرار دادند (۷ و ۲۸). بیز A یکی از روش‌های قدرتمند بیزی بوده که در مطالعات مختلف جهت پیش‌بینی ژنومی صفات پیوسته مورد استفاده قرار گرفته است (۱۰). در ادامه دلوس کامپوس (۲۰۰۹)، گنزالس ریکاردو و فورنی (۲۰۱۱) و وانگ و همکاران (۲۰۱۲) مدل‌های آستانه‌ای بیز A را برای صفات گسسته گسترش دادند (۷، ۱۴ و ۳۳).

اخیرا روش‌های یادگیری ماشین به مباحث انتخاب ژنومی وارد شده‌اند. این روش‌ها از اساس با روش‌های رایج از جمله بیز در ارزیابی ژنومی متفاوت بوده و فرضیات کمتری درباره توزیع داده‌ها نیاز داشته و انعطاف‌پذیری بالایی در آنالیز داده‌ها دارند (۱۳). بوستینگ^۱ یک روش بازنمونه‌گیری و از نوع ناپارامتری و به عنوان یکی از قدرتمندترین

1. Boosting

کل ژنوم روی تعداد بسیار زیادی از افراد، سخت و هزینه بر است، انجام مطالعات با استفاده از شبیه‌سازی روشی منطقی برای شناسایی هر یک از عوامل ذکر شده خواهد بود. لذا پژوهش حاضر با هدف بررسی صحت ژنومی روش‌های بوستینگ و بیز A در برآورد ارزش‌های اصلاحی ژنومی صفات آستانه‌ای دودویی و پیوسته با استفاده از معماری‌های مختلف ژنومی شامل تعداد متفاوت QTL، سطوح متفاوت وراثت‌پذیری و LD انجام شده است.

مواد و روش‌ها

برای شبیه‌سازی جمعیت‌ها از نرم افزار QMSim استفاده شد (۲۷). در مرحله اول، برای شبیه‌سازی جمعیتی با LD پایین، یک جمعیت پایه با اندازه مؤثر ۱۰۰۰ راسی طی ۱۰۰۰ نسل شبیه‌سازی شد. برای تولید جمعیتی با LD بالا، پس از شبیه‌سازی جمعیت پایه، تعداد افراد جمعیت از طریق ایجاد یک گلوگاه ژنتیکی (Bottleneck) به ۱۰۰ راس در نسل ۱۱۰۰ کاهش یافت. سپس در آخرین جمعیت پایه، بعد از ۱۰۰ نسل (در نسل ۱۲۰۰) تعداد افراد جمعیت به ۱۰۰۰ راس افزایش داده شد. در گام دوم، برای ایجاد جمعیت مرجع و تایید، همه افراد (۱۰۰۰ راس) آخرین نسل جمعیت پایه برای تولید مثل در جمعیت حاضر مورد استفاده قرار گرفتند که در این بین ۳۰ راس نر (اندازه مؤثر جمعیت ۱۱۶) در نظر گرفته شد تا منعکس‌کننده‌ی نسبت نر به ماده‌ی موجود در گله‌های گاو شیری باشد تا بتوان اثر روش تلقیح مصنوعی بر نسبت نر به ماده را تقلید کرد (۳۸). نوع سیستم تلاقی تصادفی بود و برای ۱۰ نسل دیگر (تا نسل ۱۲۱۰) جمعیت تکثیر شد. شانس تلاقی در همه‌ی حیوانات برابر (در هر دو جنس) و یک فرزند برای هر زایش در نظر گرفته شد. درصد جایگزینی برای نر و ماده به ترتیب ۵۰ و ۲۰ درصد در نظر

روش‌های یادگیری ماشینی جهت بهبود عملکرد روش‌های طبقه‌بندی توسعه داده شد (۹). بعدها از آن برای آنالیز ژنومی صفات آستانه‌ای (۱۴) به دلیل توانایی بالای آن در آنالیز حجم زیاد داده‌های ژنومی، توانایی در تشخیص اثرات متقابل ژن-ژن، ژن-محیط، قدرت بالای مدل‌سازی و ارتباط بین ترکیبات مختلف نشانگرها، تشخیص ژنهای مرتبط با بیماری، مشکلات مربوط به نداشتن نمونه کافی و بهبود صحت پیش‌بینی ژنومی استفاده شد (۳۷). تحقیقات شبیه‌سازی با استفاده از روش‌های بیزی و یادگیری ماشینی نشان داد که روش بوستینگ برتری قابل ملاحظه‌ای نسبت به روش‌های بیزی در ارزیابی ژنومی صفات آستانه‌ای دارد (۱۴). همچنین در یک مطالعه با داده‌های شبیه‌سازی شده نشان داده شد که روش‌های بوستینگ و GBLUP قدرت تقریباً یکسانی در برآورد صحت پیش‌بینی ژنومی صفات پیوسته دارند (۱۰). با این حال در اکثر مطالعات انجام شده نشان داده شد که عملکرد صحت ژنومی روش‌های یادگیری ماشینی و دیگر روش‌های آماری به شدت تحت تاثیر معماری ژنتیکی صفت مورد مطالعه قرار می‌گیرد (۲۳، ۲۴ و ۲۶).

امروزه تلاش پژوهشگران بر آن است تا با شناسایی عوامل مؤثر بر صحت برآوردهای ارزش‌های اصلاحی ژنومی، بهترین بهره‌برداری از داده‌های موجود انجام شود. در این راستا عوامل مختلفی می‌توانند صحت ارزش‌های اصلاحی ژنومی و ارزیابی ژنومی را تحت تأثیر قرار دهد، این عوامل شامل توزیع اثرات QTL، مقدار عدم تعادل پیوستگی، نوع و تراکم نشانگرها، وراثت‌پذیری، نحوه رکوردگیری، تعداد داده‌های فوتویی در جمعیت مرجع، فاصله زمانی (تعداد نسل) بین جمعیت مرجع و جمعیت تأیید و نوع صفت مورد مطالعه (پیوسته و گسسته) می‌باشند (۲۱ و ۳۲). در نتیجه، از آن جا که توالی یابی

۰/۴) فرض شد. دو سطح مختلف وراثت پذیری (۰/۱ و ۰/۳) برای هر صفت در نظر گرفته شد. در طراحی جمعیت نهایی، افراد آخرین نسل (نسل ۱۲۱۰) به عنوان جمعیت تایید (۱۰۰۰ راس) در نظر گرفته شد که این افراد اطلاعات ژنوتیپی داشته اما فاقد اطلاعات فنوتیپی بودند. همچنین افراد ۴ نسل ما قبل جمعیت تایید (نسل ۱۲۰۶ تا ۱۲۰۹) در گروه جمعیت های مرجع (۴۰۰۰ راس) که این افراد هم اطلاعات ژنوتیپی داشته و هم ارزش های اصلاحی ژنومی آنها مشخص بود طبقه بندی شدند. خلاصه جمعیت شبیه سازی شده و پارامترهای به کار رفته در جدول ۱ نشان داده شده است.

گرفته شد. در جمعیت اخیر، انتخاب حیوانات برتر برای نسل بعد بر اساس ارزش اصلاحی بالا و معیار حذف بر اساس ارزش اصلاحی پایین صورت گرفت. نشانگرها به صورت دو آلی و به صورت تصادفی در بین ۳۰ کروموزوم به طول ۱۰۰ سانتی مورگان توزیع شدند. به ازای هر کروموزوم ۳۳۳ و ۱۶۶۶ نشانگر برای پنل های ۱۰K و ۵۰K شبیه سازی شد. دو سطح مختلف QTL (۱۵۰ و ۴۵۰) شبیه سازی شد که به صورت تصادفی در طول کروموزوم ها توزیع شدند. نرخ جهش برای نشانگرها و QTL ها در هر جایگاه و در هر نسل $2/5 \times 10^{-6}$ فرض شد (۳۰). فراوانی آلی اولیه برای نشانگرها ۰/۵ و توزیع اثرات QTL ها، گاما

جدول ۱: پارامترهای فرآیند شبیه سازی

Table 1. Parameters of the simulation process

عدم تعادل پیوستگی پایین Low linkage disequilibrium	عدم تعادل پیوستگی بالا Low linkage disequilibrium	پارامتر Parameter
		جمعیت اولیه Historical population
1,000 (1,000)	1,000 (1,000)	تعداد نسل (تعداد افراد) در فاز اول No. of generations (population size) in phase 1
خیر No	بله Yes	گلوگاه Bottleneck
100 (1,000)	100 (100)	تعداد نسل (تعداد افراد) در فاز دوم No. of generations (population size) in phase 2
100 (1,000)	200 (1,000)	تعداد نسل (تعداد افراد) در فاز سوم No. of generations (population size) in phase 3
	10000	تعداد حیوانات در نسل آخر No. of animals in the last generation
		جمعیت اخیر Recent population
	30 (970)	تعداد نرها (ماده ها) No. of founder sires (dams)
	10	تعداد نسل No. of generations
	1	تعداد نتاج به ازای هر زایش No. of offspring per dam
تصادفی Random		سیستم تلاقی Mating system
0.5 (0.2)		نرخ جایگزینی برای نرها (ماده ها) Replacement ratio for males (females)
ارزش اصلاحی برآوردی/سن EBV/age		معیار انتخاب / حذف Criteria for selection/culling
0.5		احتمال جنسیت در نتاج

	Sex probability for offspring
30	تعداد کروموزوم
	No. of chromosomes
3000	طول کل کروموزومها (سانتی مورگان)
	Total length of chromosomes (cM)
0.4	اثر آلل‌های QTL (گاما)
	Effects of QTL alleles(gamma)
150 or 450	تعداد QTL
	No. of QTL
4990 or 49980	تعداد نشانگر
	No. of markers
$2.5 \cdot 10^{-5}$	نرخ جهش در نشانگر و QTLها
	Marker and QTL mutation rate
0.1 or 0.3	وراثت پذیری
	Heritability

همه حیوانات موجود در آخرین نسل مورد استفاده قرار گرفت (۲۵).

مدل آماری

بیز A: در این تحقیق از دو روش بوسستینگ و بیز A جهت پیش بینی ارزش های اصلاحی ژنومی استفاده شد. مدل کلی بیز A به صورت زیر می باشد.

$$\lambda = \mu 1 + Xb + e$$

در این مدل λ ماتریس وکتوری متغیر y می باشد. μ میانگین جمعیت، 1 ماتریسی با ابعاد $n \times 1$ به صورت یک می باشد. b وکتوری برای برآوردهای ضرایب رگرسیونی از تعداد p نشانگر با فرض توزیع نرمال $(N(0, \sigma_j^2))$ و مستقل از توزیع اولیه می باشد. σ_j^2 به صورت واریانس ناشناخته در ارتباط با نشانگرها مفروض شده است و دارای توزیع کای مربع مقیاس دار معکوس $(\chi^2_{j-1} \sim \nu_j \sigma_j^2)$ با $\nu_j = 4$ و $s_j^2 = 0.002$ بود. عناصر ماتریس X به ابعاد $n \times p$ دربرگیرنده اثرات افزایشی می باشد. در بیز A باقی مانده ها (e) با فرض میانگین صفر و واریانس یک در نظر گرفته شدند. این روش از طریق نمونه گیری گیس انجام گرفت. بیز A از طریق بسته BGLR در قالب پیوسته و آستانه ای در نرم افزار R مورد آنالیز قرار گرفت.

در الگوریتم بوسستینگ توابع پایه مورد استفاده

در مجموع ۴ جمعیت (جمعیت اول: وراثت پذیری صفت ۰/۳، QTL ۶۰۰ و سطح پایین LD؛ جمعیت دوم: وراثت پذیری صفت ۰/۳، QTL ۱۵۰ و سطح پایین LD؛ جمعیت سوم: وراثت پذیری صفت ۰/۱، QTL ۱۵۰ و سطح پایین LD؛ جمعیت چهارم: وراثت پذیری صفت ۰/۱، تعداد QTL ۱۵۰ و سطح بالای LD) در تحقیق حاضر شبیه سازی شد. پس از رونوشت برداری از این جمعیت ها (با فنوتیپ پیوسته)، برای شبیه سازی فنوتیپ آستانه ای دودویی، کد صفر برای دام های پایین تر از میانگین صفت و کد یک برای دام های با فنوتیپ بالاتر از میانگین صفت در نظر گرفته شد. نشانگرهای با فراوانی آلی کمیاب (MAF) کمتر از ۰/۰۱ حذف شدند. برای ارزیابی مدل ۱۰ تکرار از هر سناریو در نظر گرفته شد. سطح LD برای سناریوهای مختلف شبیه سازی شده با استفاده از محاسبه ی توان دوم ضریب همبستگی (r^2) بین همه جفت نشانگرهای ممکن ارزیابی گردد (۱۷).

$$r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)}$$

در این فرمول، $D = f(AB) - f(A)f(B)$ بوده و $f(A), f(a), f(B), f(b), f(AB)$ به ترتیب، فراوانی های مشاهده شده هاپلوتایپ AB و آلل های A، a، B و b می باشند. نرم افزار PLINK 1.9 برای برآورد LD بین جفت نشانگرهای مختلف در ژنوم

داده‌های با فنوتیپ گسسته دودویی با تراکم ۵۰k و ۱۰k ترتیب ۲۵۰۰، ۸ و ۰/۰۲ و ۱۰۰۰، ۵ و ۰/۰۲ در نظر گرفته شد. این الگوریتم از طریق بسته gbm در نرم افزار R مورد آنالیز قرار گرفت.

در نهایت نتایج حاصل از داده‌ها شبیه‌سازی شده با استفاده از روش بوستینگ و بیز A در جمعیت تایید برآورد شد و صحت پیش بینی ژنومی از طریق همبستگی پیرسون بین ارزش‌های اصلاحی پیش‌بینی شده و ارزش‌های اصلاحی واقعی محاسبه شد. فرمول ضریب همبستگی پیرسون به صورت زیر می باشد.

$$r_{EBV;TBV} = \frac{\sum_{i=1}^n (\bar{EBV} - EBV)(TBV - \bar{TBV})}{\sqrt{(\bar{EBV} - EBV)^2} \sqrt{(TBV - \bar{TBV})^2}}$$

در اینجا $r_{EBV;TBV}$ ضریب همبستگی بین ارزش اصلاحی واقعی و پیش بینی شده، n تعداد مشاهدات، EBV برارزش اصلاحی پیش‌بینی ژنومی و TBV ارزش اصلاحی واقعی (شبیه سازی شده) می باشد.

نتایج و بحث

اثر تراکم نشانگرها بر عملکرد روش‌های بوستینگ و بیز A: جدول ۲، صحت پیش‌بینی ارزش‌های اصلاحی ژنومی روش‌های بوستینگ و بیز A را در هریک از سناریوهای شبیه‌سازی شده برای تراکم نشانگری ۱۰k و ۵۰k در صفات کمی و آستانه‌ای نشان می‌دهد. به طور کلی استفاده از تراکم نشانگری ۵۰k نسبت به ۱۰k به ترتیب ۱۱/۸ و ۸/۷ درصد افزایش صحت را در مدل‌های بوستینگ و بیز آستانه‌ای A به همراه داشت. همچنین برای روش بوستینگ دامنه گسترده‌ای از صحت در مقایسه با روش بیز A با تغییرات تراکم نشانگری مشاهده شد. به طوریکه با افزایش تراکم نشانگری صحت پیش‌بینی ژنومی در همه سناریوها برای بوستینگ و بیز A به ترتیب افزایشی برابر با ۰/۰۵۵ و ۰/۰۴۵ را در پی داشت. در صفات آستانه‌ای با افزایش تراکم نشانگری

شامل یادگیرنده‌های ضعیف مانند درخت رگرسیونی می‌باشند. در این الگوریتم سعی می‌شود تعدادی یادگیر پایه ضعیف (یادگیرنده‌های بهتر از حالت تصادفی) که مکمل همدیگر هستند تولید شود و از طریق آموزش با استفاده از یادگیرنده‌های قبلی، یادگیرنده‌های جدید قوی‌تری ایجاد شود، در این الگوریتم توابع پایه مانند درختان رگرسیونی به صورت سریالی هریک روی باقی مانده درخت قبلی اضافه می‌شوند در نتیجه اشتباه دسته‌بندی در درخت قبلی باعث کاهش مقدار خطا در درخت بعدی می‌شود و این الگوریتم تا زمانی ادامه می‌یابد که خطای آخرین درختبه حداقل برسد. مدل کلی الگوریتم بوستینگ به صورت زیر می‌باشد.

$$f(x) = \sum_{m=1}^m \beta_m b(x; \gamma_m)$$

در اینجا β_m ($m=1, 2, \dots, M$) ضرائب توزیع و $b(x; \gamma_m)$ توابع ساده با توزیع چند متغیره x به‌همراه یک سری از پارامترها γ می‌باشند. در این روش درختان رگرسیونی به‌عنوان یادگیرنده‌های پایه استفاده می‌شوند. همچنین برای توزیع داده‌های پیوسته و گسسته به ترتیب از توزیع گوسن و برنولی برای حداقل کردن تابع خطا استفاده شد. نتیجه گیری نهایی از طریق میانگین تصحیح شده (به هر درخت یک وزنه داده می‌شود) مجموعه درخت‌ها انجام شد. در این روش پارامترهای بهینه سازی شامل تعداد درخت (ntree)، عمق درخت (tree complexity) و نرخ یادگیری (learning rate) می‌باشند. در این تحقیق جهت دستیابی به بیشینه عملکرد و بهترین ترکیب این پارامترها مقدار خطای اعتبار سنجی محاسبه شد. در این تحقیق تعداد درخت، عمق درخت و نرخ یادگیری برای داده‌های با فنوتیپ پیوسته با تراکم ۵۰k و ۱۰k ترتیب ۵۰۰۰، ۵، ۰/۰۲ و ۱۵۰۰، ۵، ۰/۰۲ و برای

عملکرد روش بوستینگ همواره بیشتر از روش بیز A بود. این درحالی بود که برای صفات پیوسته اثر مثبت تراکم نشانگری بر صحت پیش‌بینی ژنومی روش بیز A مشهودتر بود.

جدول ۲: میانگین و انحراف استاندارد صحت پیش‌بینی ژنومی صفات آستانه‌ای و پیوسته با استفاده از روش‌های بوستینگ و بیز A
Table 2. Mean and standard deviation of genomic prediction accuracy for threshold and continuous traits by Boosting and Bayes A methods

صفات پیوسته		صفات آستانه‌ای		سناریو Scenarios	تراکم Density
Continue traits		Threshold traits			
بیز A	بوستینگ	بیز آستانه‌ای A	بوستینگ		
BayesA	Boosting	Threshold BayesA	Boosting		
0.633±0.06	0.574±0.05	0.51±0.06	0.598±0.04	I ($h^2 = 0.3$, LD = low and 450 QTL)	50k
0.702±0.07	0.569±0.04	0.565±0.06	0.584±0.02	II ($h^2 = 0.3$, LD = low and 150 QTL)	
0.562±0.05	0.438±0.04	0.432±0.05	0.447±0.04	III ($h^2 = 0.1$, LD = low and 150 QTL)	
0.698±0.05	0.485±0.03	0.461±0.05	0.493±0.03	IV ($h^2 = 0.1$, LD = high and 150 QTL)	
0.594±0.05	0.508±0.05	0.457±0.05	0.531±0.03	I ($h^2 = 0.3$, LD = low and 450 QTL)	10k
0.666±0.06	0.532±0.03	0.535±0.05	0.544±0.03	II ($h^2 = 0.3$, LD = low and 150 QTL)	
0.513±0.07	0.377±0.04	0.376±0.04	0.382±0.04	III ($h^2 = 0.1$, LD = low and 150 QTL)	
0.550±0.04	0.425±0.03	0.409±0.03	0.432±0.03	IV ($h^2 = 0.1$, LD = high and 150 QTL)	

بین نشانگرها کمتر باشد، صحت برآورد ژنومی نیز بیشتر خواهد شد (۲۱). با توجه به اینکه در گزینش ژنومی تمام واریانس ژنتیکی توسط نشانگرهای توزیع شده در ژنوم توجیح می‌شود، در نتیجه توانایی پیش‌بینی GEBVs عمیقاً وابسته به تراکم نشانگری می‌باشد (۴ و ۳۵). از طرفی به خاطر اینکه افزایش تراکم نشانگری سطح LD بین QTL و نشانگرها را افزایش می‌دهد در نتیجه بهبود صحت پیش‌بینی ارزش‌های اصلاحی ژنومی را به همراه خواهد داشت. در این راستا تحقیقات به بررسی اثر تعداد نشانگر بر عملکرد دو روش GBLUP و جنگل تصادفی پرداختند و نشان دادند که عملکرد این دو روش به شدت متأثر از تراکم نشانگرها بوده و روش‌های یادگیری ماشین حساسیت بیشتری نسبت به این تغییرات نشان می‌دهند (۲۲). یافته‌های این تحقیق نشان داد که افزایش در تراکم نشانگری بیشتر صحت مدل‌های بوستینگ را تحت تاثیر قرار خواهد داد. بنابراین به کارگیری پنل‌های ۵۰K با نزدیک بودن فاصله SNPها با QTLها همراه خواهد بود و در نتیجه شانس نمونه‌گیری در روش‌های باز نمونه‌گیری (روش‌های یادگیری ماشین) را افزایش خواهد داد که

گزنالس ریکاردو و فورنی (۲۰۱۱) با شبیه‌سازی جمعیت‌های ژنومی با تراکم ۱۰K، دامنه بالایی از صحت را در صفات آستانه‌ای برای روش بوستینگ (۰/۲۴ - ۰/۳۷) نسبت به بیز آستانه‌ای A (۰/۲۶ - ۰/۳۲) گزارش کردند (۱۴). در مطالعه‌ای، سولبرگ و همکاران (۲۰۰۶)، از سه تراکم نشانگری برای نشانگرهای میکروساتلایت و دو تراکم نشانگری برای نشانگرهای SNP استفاده کردند. نتایج آنها نشان داد که برای هر دو نوع نشانگری، صحت برآورد ارزش‌های اصلاحی ژنومی با افزایش تراکم نشانگری افزایش می‌یابد. به طوریکه صحت ارزیابی ژنومی با افزایش تعداد نشانگرها از ۱۰۰ به ۸۰۰ نشانگر حدود ۱۷٪ افزایش یافت (۲۹). هبیر و همکاران (۲۰۰۹) افزایش تعداد نشانگرها را دلیلی بر افزایش عدم تعادل پیوستگی بین ژن‌ها و نشانگرها عنوان نموده و بدین ترتیب صحت ارزیابی ژنومی افزایش خواهد یافت (۱۵). مویر (۲۰۰۷) نیز اثر تراکم نشانگری را با ثابت نگه داشتن تعداد نشانگرها و تغییر اندازه ژنوم از طریق تغییر احتمال نوترکیبی بین نشانگرها (فاصله بین نشانگرها) در یک مطالعه شبیه‌سازی بررسی کرد و نشان داد که هرچه تراکم نشانگری بیشتر و فاصله

بیز A اذعان داشتند (۱۴). همچنین وانگ و همکاران (۲۰۱۷) و بزی و همکاران (۲۰۱۷) نشان دادند که روش‌های بیزی برای صفات با توزیع فنوتیپ پیوسته نسبت به صفات آستانه‌ای همواره عملکرد بهتری خواهند داشت (۳ و ۳۴). در مطالعه‌ای دیگر به منظور بررسی صحت پیش‌بینی ژنومی صفات با توزیع فنوتیپی پیوسته، روش‌های GBLUP و یادگیری ماشین (بوستینگ، SVM و جنگل تصادفی) مورد ارزیابی قرار گرفتند. در مجموع نشان داده شد که روش‌های روش یادگیری ماشینی به دلیل ماهیت ناپارامتری خود قابل مقایسه با روش پارامتری GBLUP نیستند، هرچند که در این میان بوستینگ مقادیر صحت قابل قبولی از خود نشان داد (۱۰). بانه و همکاران (۲۰۱۷) به بررسی عملکرد روش‌های مختلف بیزی در صفات آستانه‌ای پرداختند و صحت‌های قابل قبولی (بالای ۰/۶) در ارزیابی ژنومی بدست آوردند. با این حال نتایج آنها نشان دادند که با افزایش تعداد آستانه‌ی صفت و سیر به سمت صفات با توزیع فنوتیپی پیوسته، عملکرد این روش‌ها به علت ماهیت قدرتمند آنها در برآورد صفات پارامتریکی بهبود می‌یابد (۲).

اثر معماری ژنتیکی بر عملکرد روش‌های بوستینگ

و بیز A

تعداد QTL: برای ارزیابی اثر تعداد QTL بر صحت ژنومی، جمعیت ۱ (QTL ۴۵۰) و ۲ (QTL ۱۵۰) با هم مقایسه شدند (جدول ۲). شکل ۱، اثر کاهش تعداد QTL بر صحت ژنومی روش‌های بیز A و بوستینگ در سری داده‌های با توزیع متفاوت فنوتیپی را نشان می‌دهد. به طور کلی نوسانات روش بیزی نسبت به تغییرات QTL بیشتر از روش بوستینگ بود و این حساسیت با کاهش تعداد نشانگرها بیشتر بود. در تراکم بالای نشانگرها، برای صفات با فنوتیپ گسسته دودویی، بیشترین و کمترین میزان صحت

نتیجه‌ی این اثر مثبت در روش بوستینگ مشهود خواهد بود.

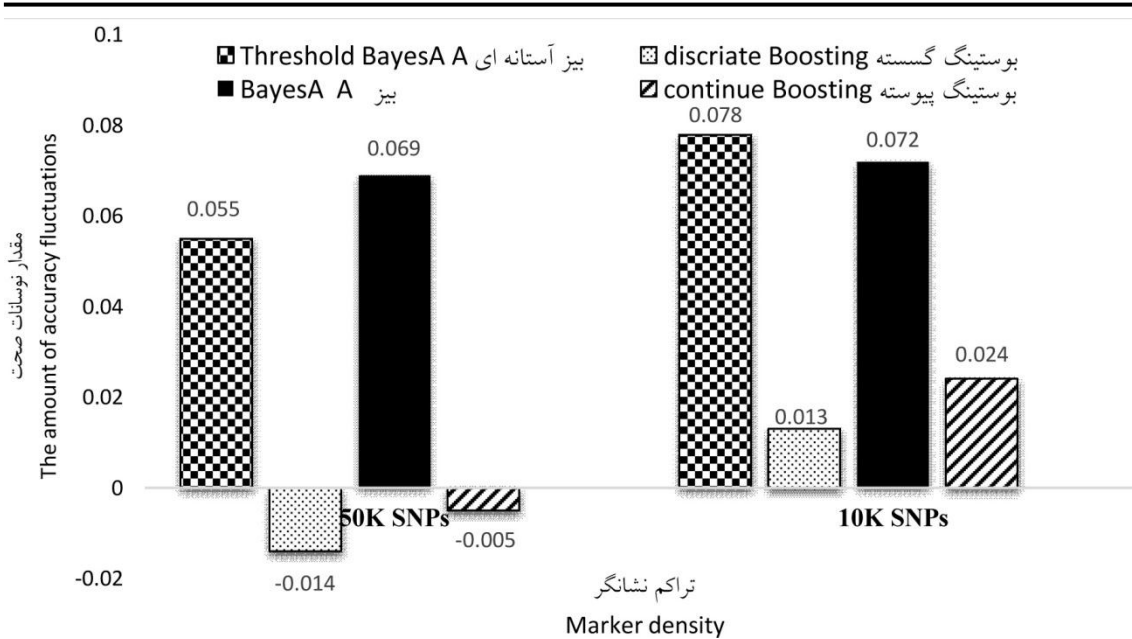
اثر نوع فنوتیپ (پیوسته و گسسته) بر عملکرد روش‌های بوستینگ و بیز A: جدول ۲ میانگین عملکرد روش‌های بوستینگ و بیز A در هریک از سناریوهای شبیه‌سازی شده را برای صفات با فنوتیپ‌های پیوسته و گسسته نشان می‌دهد. به طور کلی عملکرد روش بوستینگ برای صفات با توزیع فنوتیپی گسسته کمی بهتر از روش بیز آستانه‌ای A بود، خصوصاً هنگامی که سناریوهای با تراکم نشانگری بالا استفاده شدند. به طوریکه این میزان برتری در روش بوستینگ نسبت به بیز آستانه‌ای A برای تراکم نشانگری ۱۰k و ۵۰k به ترتیب ۶/۳ و ۷/۳ درصد بود. این درحالی بود که عملکرد بیز A برای صفات با توزیع فنوتیپی پیوسته به طور قابل توجهی بهتر از روش بوستینگ بود، خصوصاً هنگامی که سناریوهای با تراکم نشانگری پایین استفاده شدند. به طوریکه این میزان برتری در روش بیز A نسبت به بوستینگ برای تراکم نشانگری ۱۰k و ۵۰k به ترتیب ۲۸ و ۲۲/۶ درصد بود.

نتایج نشان داد که روش بوستینگ صحت پیش‌بینی ژنومی بالاتری برای صفات گسسته دودویی در مقایسه با روش آستانه‌ای بیز A دارد که این به ماهیت ناپارامتری قوی‌تر آن نسبت به روش‌های ناپارامتری بیزی برمی‌گردد. از سوی دیگر طبیعت ناپارامتریکی آن منجر به افت عملکرد آن نسبت به روش‌های پارامتریکی بیزی شد. در این راستا تحقیقات نشان داد که استفاده از روش‌های یادگیری ماشین از جمله جنگل تصادفی برای صفات آستانه‌ای نسبت به صفات کمی افزایش ۱۴ تا ۱۹ درصدی در صحت را به همراه خواهد داشت (۲۳). همچنین گنزالس ریکاردو و فورنی (۲۰۱۱) در بررسی صحت ژنومی صفات آستانه‌ای، به عملکرد بهتر روش‌های یادگیری ماشین از جمله بوستینگ و جنگل تصادفی نسبت به روش

اثرات نشانگرها در این تحقیق گاما بود و این مطلب در تحقیقات مختلف اذعان شده است (که این توزیع به پیش فرض‌های روش بیزی سازگاری بیشتری دارد) در نتیجه به خاطر این که روش بیزی A از برخی خصوصیات نظیر انتخاب متغیر سود می‌برند و مفروضات آن با تعداد QTL کم سازگارتر است در نتیجه در تعداد اندک QTL، روش بیزی A بهتر عمل کرد (۲۳ و ۲۴). با این حال، در تراکم بالای نشانگری و تعداد بالای QTL (نسبت به تعداد کم QTL) روش بوستینگ عملکرد بهتری نشان داد. در این راستا، هبیر و همکاران (۲۰۰۹) گزارش کردند هنگامی که تعداد QTL افزایش می‌یابد به تعداد بیشتری نشانگر جهت به دام انداختن اثرات همه QTLها احتیاج خواهد بود. در نتیجه افزایش تعداد QTL می‌تواند منجر به افزایش صحت ژنومی شود هنگامی که به طور موازی با افزایش تعداد QTLها، تعداد نشانگرها نیز افزایش یابد (۱۵). همچنین در روش‌های باز نمونه‌گیری و یادگیری ماشین، افزایش تعداد QTL در تراشه‌های با تراکم بالا، منجر به تولید عدم پیوستگی قوی بین برخی نشانگرها با QTLهای کنترل‌کننده‌ی صفت، نزدیک‌تر شدن فاصله نشانگرها با QTLها و افزایش شانس نمونه‌گیری شده، در نتیجه افزایش صحت ژنومی را به همراه خواهد داشت (۲۲، ۲۳ و ۲۴)، که این اثر مثبت برای روش بوستینگ در نتایج تحقیق حاضر صادق بود.

ژنومی به ترتیب در روش بوستینگ (۰/۵۹۸) و بیز آستانه‌ای A (۰/۵۱۰) در تعداد QTL بالا مشاهده شد. این در حالی بود که برای صفات به توزیع پیوسته فنوتیپ، بیشترین و کمترین میزان صحت ژنومی به ترتیب در روش بیز A (۰/۷۰۲) بوستینگ (۰/۵۶۹) در تعداد QTL پایین مشاهده شد. در تراکم پایین نشانگرها، روش‌های بوستینگ و بیز A بهترین عملکرد خود را به ترتیب در صفات با توزیع فنوتیپی گسسته دودویی و پیوسته با تعداد پایین QTL نشان دادند.

تحقیقات در زمینه اثر تعداد QTL بر صحت ژنومی نشان داد که صحت روش‌های آماری به شدت متأثر از تعداد QTL است (۲۳ و ۲۴). در این تحقیق، صحت پیش‌بینی ژنومی ناشی از روش بیز A در هر دو تراکم نشانگرها و بوستینگ در تراکم پایین نشانگرها، با افزایش تعداد QTL کاهش یافت. زیرا با افزایش تعداد QTLها سهم هر QTL در ارزش ژنتیکی کل کاهش یافته و به نوعی قدرت مدل‌ها در پیش‌بینی این اثرات کوچک کاهش خواهد یافت. از طرف دیگر در این خصوص عبدالهی و همکاران (۲۰۱۳) نتایج مشابهی را گزارش کردند و بیان کردند که در تعداد QTL پایین، احتمال شکل‌گیری توزیع اثرات ژنی کم بوده و توزیع آماری مورد نظر با تعداد ژن‌های بزرگ اثر و کوچک اثر به خوبی بیان و نمایان نمی‌شود که می‌تواند محتمل‌ترین دلیل برای نتایج به دست آمده باشد (۱). همچنین به خاطر اینکه توزیع



شکل ۱: اثر کاهش تعداد QTL (از ۴۵۰ به ۱۵۰) بر صحت پیش بینی ژنومی روش‌های Bیز A و بوستینگ در سری داده‌های با توزیع متفاوت فنوتیپی

وراثت‌پذیری پایین مشاهده شد. این روند برتری در تراکم پایین نشانگرها (۱۰k) با شدت کمتری ادامه یافت. همچنین در تراکم بالای نشانگرها (۵۰k)، برای سطوح مختلف وراثت‌پذیری در صفات با توزیع فنوتیپی پیوسته، همواره عملکرد بیز A بیشتر از روش بوستینگ بود. این روند برتری در تراکم پایین نشانگرها (۱۰k) با شدت بیشتری ادامه یافت.

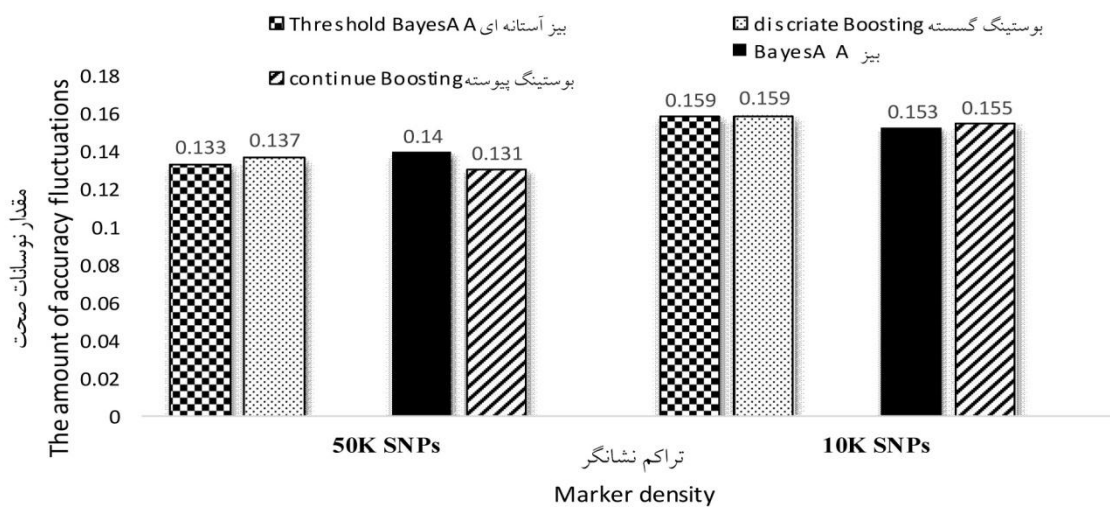
در روش‌های رایج انتخاب، روند بهبود صفات با وراثت‌پذیری پایین به آرامی صورت می‌گیرد، به دلیل این که صحت در این روش‌ها به میزان زیادی به وراثت‌پذیری صفت بستگی دارد. در حالی که در انتخاب ژنومی خصوصاً در حالت تراکم‌های بالای نشانگرها، صحت برای صفات با توارث پایین اثر مثبت تری نسبت به صفات با وراثت‌پذیری بالا دارد (۲۱). در نتیجه برتری حاصل از تراکم بالای نشانگرها بر صفات با توارث‌پذیری پایین بالا بوده و اختلاف صحت کاهش یافته است. به طور کلی دلیل افزایش صحت ژنومی با افزایش وراثت‌پذیری را

اثر وراثت‌پذیری: برای بررسی اثر وراثت‌پذیری بر صحت ژنومی، جمعیت ۲ ($h^2=0.3$) و ۳ ($h^2=0.1$) مورد ارزیابی قرار گرفت. همچنین شکل ۲ اثر افزایش وراثت‌پذیری بر صحت ژنومی روش‌های Bیز A و بوستینگ در سری داده‌های با توزیع متفاوت فنوتیپی را نشان می‌دهد. به‌طورکلی نوسانات روش بیزی و بوستینگ نسبت به تغییرات وراثت‌پذیری با کاهش تعداد نشانگرها بیشتر بود. با این حال برای سناریوهای با تراکم بالای نشانگرها، حساسیت روش بوستینگ و بیز A به ترتیب برای صفات با توزیع فنوتیپ گسسته و پیوسته بیشترین مقدار را نشان داد. این درحالی بود که برای تراکم پایین نشانگرها صحت پیش‌بینی ژنومی به میزان یکسانی تحت تاثیر تغییرات وراثت‌پذیری قرار گرفت. در تراکم بالای نشانگرها، برای صفات با فنوتیپ گسسته دودویی، بیشترین میزان صحت ژنومی برای روش بوستینگ (۰/۵۸۴) در سناریو با وراثت‌پذیری بالا و کمترین میزان صحت ژنومی برای بیز آستانه‌ای A (۰/۴۳۲) در سناریو با

در مورد ارتباط مستقیم بین وراثت‌پذیری و صحت پیش‌بینی ارزش‌های اصلاحی ژنومی بود (۴). در این تحقیق به طور کلی نوسانات روش بیزی A (با توزیع پیوسته فنوتیپ) به تغییرات وراثت‌پذیری بیش از بوستینگ بود. مطالعات جهت بررسی صحت ژنومی نشان داد که با افزایش وراثت‌پذیری (از ۰/۱ به ۰/۳)، روش بیزی (با افزایش ۰/۱۰ در صحت) نسبت به روش جنگل تصادفی (با افزایش ۰/۰۸) از نوسانات بیشتری برخوردار بود (۲۴). هم‌چنین نادری و همکاران (۲۰۱۶) به بررسی اثر وراثت‌پذیری بر سطح زیر منحنی مشخصه عملکرد (AUROC) با روش‌های GBLUP و جنگل تصادفی پرداختند و این نتیجه رسیدند که نوسانات جنگل تصادفی به تغییرات وراثت‌پذیری کم‌تر از روش GBLUP است (۲۲). در چندین مطالعه اثر مطلوب افزایش وراثت‌پذیری بر صحت پیش‌بینی ارزش‌های اصلاحی ژنومی ناشی از مدل‌های بیزی به اثبات رسیده است. این تاثیر مثبت و مطلوب وراثت‌پذیری منجر به تغییرات بالای ژنتیکی و در نتیجه کمک به پیش‌بینی بهتر اثرات نشانگری شد (۳، ۲۴، ۳۴).

می‌توان این‌گونه عنوان کرد که بالا بودن وراثت‌پذیری یعنی بیشتر بودن نسبت واریانس ژنتیکی افزایشی به واریانس فنوتیپی و به معنای بیشتر بودن نقش ژن‌های با بیان افزایشی در ایجاد پراکنش در صفت است که باعث برآورد صحیح‌تر تاثیرات نشانگرها می‌شود. در نتیجه هرچه وراثت‌پذیری صفت بیشتر باشد، فنوتیپ فرد به ارزش ژنتیکی فرد نزدیک‌تر بوده و در نتیجه اثر نشانگرها و به دنبال آن ارزش‌های اصلاحی ژنومی افراد به طور صحیح‌تر پیش‌بینی می‌شود (۲۴). هم‌چنین افزایش واریانس نمونه‌گیری برآورد اثرات نشانگرها به دلیل افزایش واریانس محیطی از مهمترین دلایل کاهش صحت ژنومی در صفات با توارث پذیری پایین می‌باشند.

مطالعات نشان داده‌اند که وراثت‌پذیری صفت مورد نظر به عنوان یک عامل مهم در معماری ژنتیکی صفت، می‌تواند بر صحت پیش‌بینی ژنومی اثرگذار باشد. در وراثت‌پذیری بالاتر، ارزش‌های اصلاحی افراد سهم بیشتری در بروز رکوردهای فنوتیپ داشته و لذا در برآورد اثرات نشانگرها با استفاده از اطلاعات فنوتیپی صحت بیشتری مشاهده خواهد شد (۲۰). نتایج این تحقیق مطابق با تئوریو و همکاران (۲۰۱۷)

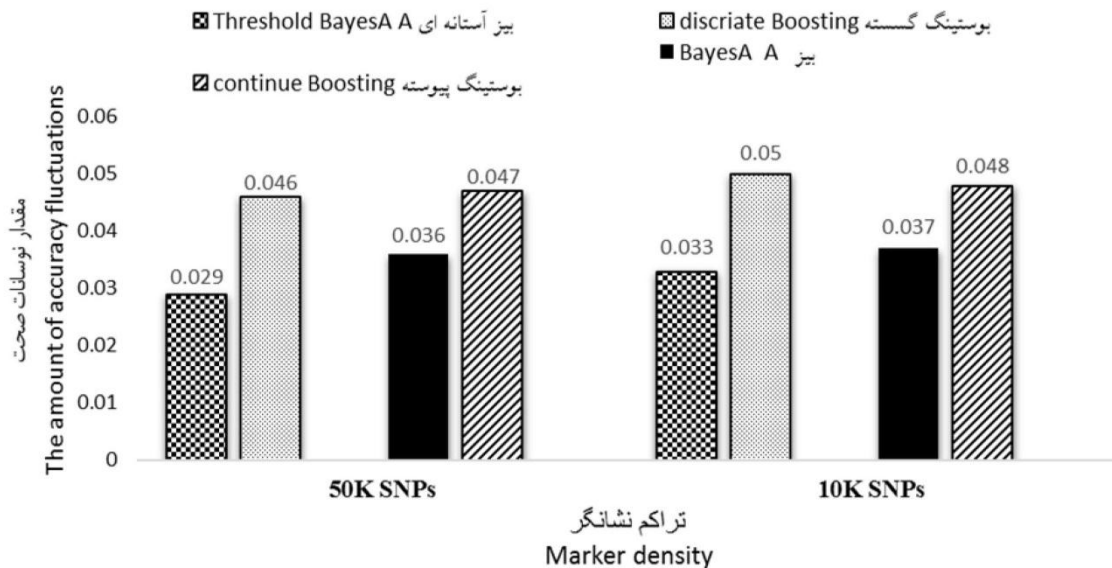


شکل ۲: اثر افزایش وراثت‌پذیری (از ۰/۳ به ۰/۱) بر صحت پیش‌بینی ژنومی روش‌های بیزی A و بوستینگ در سری داده‌های با توزیع متفاوت فنوتیپی

قوی بین نشانگرها مجاور در بررسی مطالعات ژنوم انسان را عامل اساسی در مطالعات گستره ژنوم عنوان کردند (۱۹). کالوس و همکاران (۲۰۰۸) گزارش کردند که اگر سطح LD بین نشانگرها از ۱۰ درصد به ۲۶ درصد افزایش یابد، صحت برآورد ارزش‌های اصلاحی ژنومی از ۰/۶۸ به ۰/۸۲ افزایش خواهد یافت (۵). نتایج این تحقیق نشان داد که در صفات آستانه‌ای برای سطوح مختلف LD، روش بوستینگ نسبت به روش بیز آستانه‌ای عملکرد بهتری دارد و این اثر مثبت با افزایش تراکم نشانگری بیشتر شد. سطح بالای LD میان QTL ها و نشانگرها از یک طرف (۱۲ و ۲۴) و استفاده از تراشه متراکم از طرف دیگر، باعث افزایش احتمال نمونه‌گیری نشانگرهای مجاور و دارای LD بالا در روش‌های باز نمونه‌گیری شد، که این امر عملکرد مثبت بوستینگرا به همراه داشت (۲۳ و ۲۴).

اثر عدم تعادل پیوستگی: برای ارزیابی اثر LD بر صحت پیش‌بینی ژنومی، جمعیت‌های ۳ (LD پایین) و ۴ (LD بالا) با هم مقایسه شدند (جدول ۲). میانگین LD برای جمعیت‌های ۳ و ۴ در فاصله ۰/۰۵ سانتی مورگان به ترتیب ۰/۲۴۴ و ۰/۴۵۷ بود. به طور کلی روش بیز A برای صفات پیوسته و روش بوستینگ برای صفات آستانه‌ای بهترین عملکرد را نشان دادند. شکل ۳، اثر افزایش LD بر صحت ژنومی روش‌های بیز A و بوستینگ را در سری مختلف داده‌ها (با توزیع مختلف فنوتیپ) نشان می‌دهد. با کاهش تعداد نشانگرها، میزان حساسیت روش‌های بیز و بوستینگ به نوسانات LD بیشتر شد. به طور کلی در هر دو سری داده، روش بوستینگ نسبت بیز A حساسیت بیشتری به تغییرات LD نشان داد.

به‌عنوان یک اصل کلی، وجود LD بین نشانگرها و QTL منبع اصلی اطلاعات است و نقش عمده‌ای در صحت پیش‌بینی ارزش‌های اصلاحی ژنومی ایفا می‌کنند (۱۸ و ۳۱). کی و همکاران (۲۰۰۴) وجود LD



شکل ۳: اثر افزایش LD (از ۰/۲۴۴ به ۰/۴۵۷) بر صحت پیش‌بینی ژنومی روش‌های بیز A و بوستینگ در سری داده‌های با توزیع متفاوت فنوتیپی

در این تحقیق سناریوهای مختلف ژنومی جهت

نتیجه‌گیری کلی

- Rokoui, M. 2017. Accuracy of Bayesian methods in genomic evaluation threshold traits with different genetic architecture. *Journal of Ruminant Research*. 5(2): 129-43. (In Persian).
4. Bo, Z., Zhang, J.-J., Hong, N., Long, G., Peng, G., Xu, L.-Y., Yan, C., Zhang, L.-P., Gao, H.-J. and Xue, G. 2017. Effects of marker density and minor allele frequency on genomic prediction for growth traits in Chinese Simmental beef cattle. *Journal of Integrative Agriculture*. 16(4):911-20.
 5. Calus, M., De Roos, A. and Veerkamp, R. 2008. Accuracy of genomic selection using different methods to define haplotypes. *Genetics*. 178(1): 553-61.
 6. Chen, L., Li, C., Sargolzaei, M. and Schenkel, F. 2014. Impact of genotype imputation on the performance of GBLUP and Bayesian methods for genomic prediction. *PloS One*. 9(7): 101544.
 7. De Los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., Weigel, K. and Cotes, J.M. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*. 182(1): 375-85.
 8. Egger-Danner, C., Cole, J., Pryce, J., Gengler, N., Heringstad, B., Bradley, A. and Stock, K.F. 2015. Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal*. 9(2):191-207.
 9. Freund, Y. and Schapire, R.E. 1996. Experiments with a new boosting algorithm. *Icml*. 96: 148-56.
 10. Ghafouri-Kesbi, F., Rahimi-Mianji, G., Honarvar, M. and Nejati-Javaremi, A. 2017. Predictive ability of Random Forests, Boosting, Support Vector Machines and Genomic Best Linear Unbiased Prediction in different scenarios of genomic evaluation. *Animal Production Science*. 57(2): 229-36.
 11. Gianola, D. 2013. Priors in whole-genome regression: the Bayesian alphabet returns. *Journal of Genetics*. 194(3): 573-96.
 12. Goddard, M. 2009. Genomic selection:

بررسی عملکرد الگوریتم بوستینگ و بیز A برای صفات پیوسته و آستانه‌ای مورد ارزیابی قرار گرفتند. به طور کلی تراکم نشانگرها از عوامل مؤثر بر عملکرد هریک از روش‌های مورد مطالعه بود، و در این میان عملکرد الگوریتم بوستینگ از حساسیت بیشتری برخوردار بود. علی‌رغم برتری جزئی روش بوستینگ (به عنوان یک روش ناپارامتری با قدرت لحاظ کردن اثرات متقابل بین SNPs) نسبت به بیز آستانه‌ای A در صفات گسسته دودویی، برای صفات پیوسته روش بیز برتری مطلقی نسبت به روش بوستینگ نشان داد. ارزیابی صحت ژنومی نشان داد که ساختار معماری ژنومی از جمله وراثت‌پذیری، تعداد QTL و LD از فاکتورهای اصلی و مؤثر بر صحت ژنومی دو روش بیز و بوستینگ می‌باشند. در این راستا نقش وراثت‌پذیری بر عملکرد هریک از این روش‌ها مشهودتر بود. در مجموع روش بیز و بوستینگ به ترتیب از حساسیت بیشتری نسبت به نوسانات تعداد QTL و LD نشان دادند. به عنوان یک نتیجه‌گیری کلی روش بوستینگ به عنوان روشی قدرتمند در ارزیابی ژنومی صفات آستانه‌ای و روش بیز A به عنوان روشی قدرتمند در ارزیابی صفات پیوسته پیشنهاد می‌شود.

منابع

1. Abdollahi-Arpanahi, R., Pakdel, A., Nejati-Javaremi, A. and Shahrabak, M.M. 2013. Comparison of genomic evaluation methods in complex traits with different genetic architecture. *Journal of Animal Production*. 15: 65-77.
2. Baneh, H., Nejati Javaremi, A., Rahimi-Mianji, G. and Honarva, M. 2017. Genomic evaluation of threshold traits with different genetic architecture using bayesian approaches. *Research on Animal Production*. 8(15): 149-54. (In Persian).
3. Bazzi, H., Tahmoorespour, M. and

- accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics*. 124(6): 342-55.
22. Naderi, S., Yin, T. and König, S. 2016. Random forest estimation of genomic breeding values for disease susceptibility over different disease incidences and genomic architectures in simulated cow calibration groups. *Journal of Dairy Science*. 99(9): 7261-73.
23. Naderi, Y. 2018. Evaluation of genomic prediction accuracy in different genomic architectures of quantitative and threshold traits with the imputation of simulated genomic data using random forest method. *Research on Animal Production*. 9(20): 129-38. (In Persian).
24. Naderi, Y. 2018. Impact of genotype imputation and different genomic architectures on the performance of random forest and threshold Bayes A methods for genomic prediction. *Iranian Journal of Animal Science*. 49(1): 145-57. (In Persian).
25. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I. and Daly, M.J. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 81(3): 559-75.
26. Sadeghi, S., Rafat, s.A. and Alijani, S. 2018. Evaluation of imputed genomic data in discrete traits using Random forest and Bayesian threshold methods. *Acta Scientiarum Animal Sciences*. 40: 39007.
27. Sargolzaei, M. and Schenkel, F.S. 2009. QMSim: a large-scale genome simulator for livestock. *Journal of Bioinformatics*. 25(5): 680-1.
28. Schaeffer, L. 2006. Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 123(4): 218-223.
29. Solberg, T., Sonesson, A. and Wooliams, J., editors. 2006. Genomic selection using different markers and density. *Proceedings of the 8th World Congress on Genetics Applied to* prediction of accuracy and maximisation of long term response. *Journal of Genetics*. 136(2): 245-57.
13. Goldstein, B.A., Hubbard, A.E., Cutler, A. and Barcellos, L.F. 2010. An application of Random Forests to a genome-wide association dataset: methodological considerations & new findings. *Journal of BMC Genetics*. 11(1): 49.
14. González-Recio, O. and Forni, S. 2011. Genome-wide prediction of discrete traits using Bayesian regressions and machine learning. *Journal of Genetics Selection Evolution*. 43(1): 7.
15. Habier, D., Fernando, R.L. and Dekkers, J.C. 2009. Genomic selection using low-density marker panels. *Journal of Genetics*. 182(1): 343-53.
16. Hayes, B.J., Bowman, P.J., Chamberlain, A. and Goddard, M. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*. 92(2): 433-43.
17. Hill, W. and Robertson, A. 1968. Linkage disequilibrium in finite populations. *TAG Theoretical and Applied Genetics*. 38(6): 226-31.
18. Jónás, D., Ducrocq, V. and Croiseau, P. 2017. The combined use of linkage disequilibrium-based haploblocks and allele frequency-based haplotype selection methods enhances genomic evaluation accuracy in dairy cattle. *Journal of Dairy Science*. 100(4): 2905-8.
19. Ke, X., Hunt, S., Tapper, W., Lawrence, R., Stavrides, G., Ghorri, J., Whittaker, P., Collins, A., Morris, A.P. and Bentley, D. 2004. The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Journal of Human Molecular Genetics*. 13(6): 577-88.
20. Meuwissen, T., Hayes, B. and Goddard, M. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Journal of Genetics*. 157(4): 1819-29.
21. Muir, W. 2007. Comparison of genomic and traditional BLUP-estimated breeding value

35. Wang, Q., Yu, Y., Yuan, J., Zhang, X., Huang, H., Li, F. and Xiang, J. 2017. Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp *Litopenaeus vannamei*. *Journal of BMC Genetics*. 18(1): 45.
36. Wiggans, G., VanRaden, P. and Cooper, T. 2011. The genomic evaluation system in the United States: Past, present, future. *Journal of Dairy Science*. 94(6): 3202-11.
37. Yang, P., Hwa Yang, Y., B Zhou, B. and Y Zomaya, A. 2010. A review of ensemble methods in bioinformatics. *Current Bioinformatics*. 5(4): 296-308.
38. Yin, T., Pimentel, E., Borstel, U.K.v. and König, S. 2014. Strategy for the simulation and analysis of longitudinal phenotypic and genomic data in the context of a temperature× humidity-dependent covariate. *Journal of Dairy Science*. 97(4): 2444-2454.
- Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, Instituto Prociência.
30. Solberg, T., Sonesson, A. and Woolliams, J. 2008. Genomic selection using different marker types and densities. *Journal of Animal Science*. 86(10): 2447-54.
31. Sun, X., Fernando, R. and Dekkers, J. 2016. Contributions of linkage disequilibrium and co-segregation information to the accuracy of genomic prediction. *Journal of Genetics Selection Evolution*. 48(1): 77.
32. Villumsen, T., Janss, L. and Lund, M. 2009. The importance of haplotype length and heritability using genomic selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 126(1): 3-13.
33. Wang, C., Ding, X., Wang, J., Liu, J., Fu, W., Zhang, Z., Yin, Z. and Zhang, Q. 2013. Bayesian methods for estimating GEBVs of threshold traits. *Heredity*. 110(3): 213-9.
34. Wang, C., Li, X., Qian, R., Su, G., Zhang, Q. and Ding, X. 2017. Bayesian methods for jointly estimating genomic breeding values of one continuous and one threshold trait. *PloS One*. 12(4): 175448.



Performance evaluation of Boosting and Bayes A methods by different challenges of genomic architectures in discrete and continue traits

*Y. Naderi

Assistant Prof., Dept of Animal Science and Young Researchers club,
Islamic Azad University, Astara

Received: 27/11/2018; Accepted: 01/05/2019

Abstract

Background and objectives: Genomic selection using different statistical methods played an important role in increasing economic efficiency and the genetic improvement of the discrete and continuous traits. In this study, performance of Boosting and Bayes A methods were compared to evaluate genomic breeding values for binary threshold and continuous traits for different marker densities using different genomic architectures.

Materials and methods: Genomic data were simulated by QMSim software to reflect variations in heritability ($h^2 = 0.1$ and 0.3), linkage disequilibrium (LD=low and high), number of QTL (QTL=150 and 450) and marker densities (10k and 50k) for 30 chromosomes. To create discrete threshold phenotypes in training set, individuals per generation were ranked in ascending order according to continuous phenotypes of QMSim output. Then, depending on average simulated population, the threshold phenotype of individuals was defined as code 0 (higher than average trait) and code 1 (lower than average trait). Eventually, genomic estimated breeding values were calculated using Bayes A and Boosting methods to evaluate accuracy of genomic prediction for threshold and continuous traits.

Results: Compare to Bayes A, Boosting algorithm showed a wide range of genomic accuracies to changes in marker density. Compare to threshold Bayes A, Boosting algorithm demonstrated an increase of 6.3 and 7.3% on genomic accuracy of threshold traits for 10k and 50k SNPs panels, respectively. For traits with continuous phenotypic distribution, performance of Bayes A was much more than Boosting, especially when the sparse panels were used. The structure of genomic architecture including heritability, number of QTLs and LD were the most important factors affecting the accuracy of genomic prediction using Bayes A and Boosting methods. In this regard, impact of heritability on performance of each of these models was more evident. Overall, genomic accuracies of Bayes A and Boosting methods were more sensitive to QTL and LD fluctuations, respectively. For threshold traits with high density marker panels, the highest and the lowest genomic accuracy were obtained using Boosting (0.598) and Bayes A (0.510) methods, respectively, when the data set containing a lot of QTLs was applied. For continuous traits, the highest and the lowest genomic accuracy were obtained using Bayes A (0.702) and Boosting (0.569) methods, respectively, when the data set containing a few QTLs was used. Positive effect of increase LD on accuracies of genomic prediction of Boosting and Bayes A for the sparse panels was much more noticeable than high density panels.

Conclusion: Generally, this study indicated that Boosting and Bayes A methods showed their best performance for threshold and continuous traits, respectively.

Keywords: Threshold traits, Genomic prediction, Heritability, Machine learning, Linkage disequilibrium

*Corresponding author; yousefnaderi@gmail.com