

Research Paper

Assessing the Performance of Ridge Regression Method-6 in Genomic Evaluation of Discrete Threshold Traits with Additive and Dominance Genetic Architecture

Farhad Ghafouri-Kesbi 

Department of Animal Science, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran,
(Corresponding author: f.ghafouri@basu.ac.ir)

Received: 23 March, 2023

Accepted: 31 July, 2023

Extended Abstract

Background: One of the important issues in genomic selection is estimating the effect of markers. In recent years, various methods have been proposed to estimate marker effects, each providing genomic breeding values with different levels of accuracy. One commonly used method in genomic evaluation is ridge regression (rrBLUP), which has been applied in various studies to predict genomic breeding values. Recently, modifications to the parameters of the rrBLUP method have led to the development of a variant known as Ridge Regression method 6 (RR-m6) to address regression problems. However, this method has not yet been utilized in the genomic evaluation of threshold traits with additive and dominant genetic architectures, and its performance in this context remains unknown. Therefore, this research aimed to compare the prediction performance of the RR-m6 method with other common methods of genomic evaluation.

Methods: A genome consisting of 10 chromosomes, each containing 1,000 bi-allelic single nucleotide polymorphisms (SNPs), was simulated at a heritability level of 0.5. All quantitative trait loci (QTLs) were assigned additive genetic effects, modeled by a gamma distribution. Two scenarios for the number of QTLs were considered: 1% and 10% of the total number of SNPs (100 and 1,000 QTLs, respectively). Additionally, in different scenarios, 0%, 50%, and 100% of QTLs were assigned dominance effects. Genomic breeding values were estimated using RR-m6, rrBLUP, GBLUP, BayesA, regression tree (RT), Random Forest (RF), and boosting. Indicators from the LR method, including prediction accuracy, bias, and dispersion (inflation) of genomic breeding values, were used to analyze the breeding values estimated by different methods. Furthermore, the computing time and memory required to execute the codes of each method on the CPU were calculated.

Results: The results indicated that using a purely additive model when genetic dominance effects contributed to the phenotypic variation of the trait led to decreased accuracy and increased bias and dispersion of the genomic breeding values. The extent of these effects depended on the number of QTLs exhibiting dominance. Compared to other methods, RR-m6 demonstrated excellent performance, as it consistently yielded the highest accuracy and the lowest bias and dispersion across all studied scenarios, although differences were not significant in most cases when compared to BayesA. In terms of computational speed, the RR-m6 method was the fastest and required less memory than other methods for analysis.

Conclusion: The findings suggest that the RR-m6 method predicts genomic breeding values with high accuracy and is efficient in terms of computing time and memory requirements, making it a viable option for the genomic evaluation of threshold traits.

Keywords: Dominance genetic effects, Genomic selection, QTL, Single nucleotide polymorphism

How to Cite This Article: Ghafouri-Kesbi, F. (2023). Assessing the Performance of Ridge Regression Method-6 in Genomic Evaluation of Discrete Threshold Traits with Additive and Dominance Genetic Architecture. *Res Anim Prod*, 14(4), 102-113. <https://doi.org/10.61186/rap.14.42.102>





مقاله پژوهشی

بررسی عملکرد روش ششم از رگرسیون ریج در ارزیابی ژنومی
صفات آستانه‌دار با معماری ژنتیکی افزایشی و غالبیت

فرهاد غفوری کسبی

(fghafouri@basu.ac.ir)

تاریخ دریافت: ۱۴۰۲/۰۵/۹ | تاریخ پذیرش: ۱۴۰۲/۰۷/۳

صفحه: ۱۱۳ تا ۱۰۲

چکیده مبسوط

مقدمه و هدف: یکی از مسائل مهم در انتخاب ژنومی، برآورد اثر نشانگرها است. روش‌های مختلفی در طی سال‌های اخیر جهت برآورد اثر نشانگرها پیشنهاد شده‌اند که هر کدام ارزش‌های اصلاحی ژنومی را با صحت متفاوتی برآورد می‌نمایند. یکی از روش‌های مورد استفاده در ارزیابی ژنومی روش رگرسیون ریج می‌باشد (rrBLUP) که در مطالعات مختلف برای پیش‌بینی ارزش‌های اصلاحی ژنومی مورد استفاده قرار گرفته است. اخیراً با اعمال تغییراتی در پارامترهای روش rrBLUP نسخه‌ای از این روش تحت عنوان رگرسیون ریج-روش ششم (RR-m6) برای حل مسائل رگرسیون پیشنهاد شده است. با این حال، تاکنون این روش در ارزیابی ژنومی صفات آستانه‌دار با معماری ژنتیکی افزایشی و غالبیت مورد استفاده قرار نگرفته است و عملکرد آن در این زمینه مشخص نیست. لذا در این تحقیق عملکرد پیش‌بینی این روش با سایر روش‌های رایج در ارزیابی ژنومی مقایسه شد.

مواد و روش‌ها: ژنومی شامل ۱۰ کروموزوم، هر کدام خواهد ۱۰۰ نشانگر تک نوکلئوتیدی دو الی (SNP) (QTL) (اثرات ژنتیکی افزایشی) داده شد که اثرات آنها با توزیع گاما مدل سازی شد. دو ستابیو از تعداد QTL به صورت ۱٪ و ۱۰٪ از تعداد کل SNP‌ها (به ترتیب ۱۰ و ۱۰۰ QTL به ازای هر کروموزوم) در نظر گرفته شد. همچنین در ستاربوهای مختلف (به ۰، ۵۰ و ۱۰۰٪) از QTL‌ها اثر غالبیت داده شد. ارزش‌های اصلاحی ژنومی با استفاده از روش‌های GBLUP، rrBLUP، RR-m6، BayesA، BayesA_G، جنگل تصادفی (Random Forest, RF) و Boosting (Random Forest, RF) مورد ارزیابی شدند. در ضمن، زمان محاسباتی و میزان حافظه مورد نیاز برای پردازش کدهای مربوط به هر روش بر روی CPU محاسبه شد.

یافته‌ها: نتایج نشان داد استفاده از یک مدل صرف‌افزایی زمانی که اثرات ژنتیک غالبیت در تنوع فنوتیپی صفت مشارکت داشته باشد، منجر به کاهش صحت و افزایش ارزیابی و پراکنش ارزش‌های اصلاحی ژنومی خواهد شد که میزان آن با تعداد QTL‌هایی که اثر غالبیت دارند رابطه مستقیم داشت. در مقایسه با سایر روش‌ها، روش RR-m6 عملکرد سیار مطلوبی از خود نشان داد بدصورتی که در تمامی ستاربوهای مطالعه شده، ارزش‌های اصلاحی ژنومی حاصل از آن از بیشترین صحت و کمترین ارزیابی و پراکنش برخوردار بود اگرچه در بیشتر موارد تفاوت آن با روش BayesA معنی دار نبود. از نظر سرعت محاسباتی، روش RR-m6 سریعترین روش‌ها بود و در ضمن در مقایسه با سایر روش‌ها به حافظه کمتری برای انجام تجزیه و تحلیل ها احتیاج داشت.

نتیجه‌گیری: نتایج نشان داد از آن جا که روش RR-m6 ارزش‌های اصلاحی ژنومی را با صحت بالای پیش‌بینی می‌کند و در ضمن از نظر مدت زمان انجام محاسبات و میزان حافظه مورد نیاز نیز سیار کارآمد است، می‌توان از آن برای ارزیابی ژنومی صفات آستانه‌دار استفاده نمود.

واژه‌های کلیدی: انتخاب ژنومی، اثرات ژنتیکی غالبیت، نشانگر تک نوکلئوتیدی دو الی، QTL

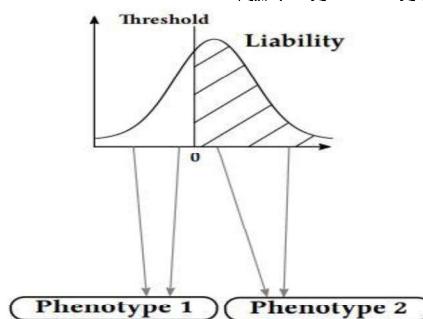
شجره‌ای آنها تخمین زده می‌شد انجام می‌گرفت. این راهبرد پیشرفت قابل توجهی در عملکرد تولیدی حیوانات اهلی ایجاد کرد (Hill, 2008). امروزه با پیشرفت علم اصلاح دام و تکنیک‌های ژنتیک ملکولی و شناسایی تعداد زیادی نشانگر ژنتیکی از نوع نشانگرهای چند شکلی تک نوکلئوتیدی (SNP)، می‌توان ارزش اصلاحی ژنومی مستقیم را از طریق تجمعیم اثرات نشانگرهایی که کل ژنوم را پوشش می‌دهند انجام داد. به این روش، انتخاب ژنومی گفته می‌شود (Meuwissen et al., 2001). هدف از انتخاب ژنومی استفاده همزمان از داده‌های ژنوتیپی در سطح توالی DNA به همراه داده‌های فنوتیپی است تا توان بدون نیاز به صرف زمان و هزینه زیاد دامها را ارزیابی کردد و دامهای برتر را گزینش نمود. در تحقیقی گزارش شد که در مقایسه با روش رایج ارزیابی نتایج، انتخاب ژنومی میانگین پیشرفت ژنتیکی سالانه را در سه نژاد موبیلیارد، نورماندی و هاشتاین کشور فرانسه را به ترتیب به میزان ۵۰، ۷۱ و ۳۳ درصد افزایش داد و به موازات آن، فاصله نسل را در سه نژاد مذبور به میزان ۱/۷ و ۲ سال کاهش داده است (Doublet et al., 2019).

مقدمه تأمین احتیاجات غذایی کشور همواره از اولویت‌های مهم دولت‌ها بوده بوده است چرا که امنیت غذایی سلامتی نیروی کار و توسعه کشور در ابعاد مختلف را در پی خواهد داشت. برای تأمین احتیاجات غذایی با منشاء دامی نمی‌توان صرفأً تعداد حیوانات موجود را افزایش داد، بلکه باستی راههایی که سبب افزایش تولید در واحد دامی شوند را جستجو کرد. این هدف از طریق اصلاح نژاد قابل حصول است (Hill, 2008). اصلاح نژاد علم و هنر تثبیت ژن‌های موثر در صفات تولیدی دامها است و انتخاب یکی از راهبرد سیار مهم در اصلاح نژاد است. در سال‌های دور و قبل از کشف اصول توارث متندلی و چگونگی کنترل ژنتیکی صفات، انتخاب بر مبنای عملکرد فنوتیپی دامها انجام می‌شد که بدلیل دخالت عوامل محیطی روش دقیقی نبود و لزوماً منجر به انتخاب حیوانات با پتانسیل ژنتیکی بالاتر نمی‌شد. از اواسط قرن بیستم به کمک معادلات مختلط ارائه شده توسط هندرسون (1975)، Henderson، انتخاب حیوانات برتر بر اساس ارزش‌های اصلاحی آن‌ها که توسط مدل‌های حیوانی و با ترکیب اطلاعات فنوتیپی و

وجود داشته باشد. به طور کلی اگر n تعداد آستانه باشد، تعداد دسته فنوتیپی برابر $i+1$ خواهد بود.

یکی از مسائل مهم در انتخاب ژنومی، برآورد اثر نشانگرها است. روش‌های مختلفی در طی سال‌های اخیر جهت برآورد نشانگرها نشانگرها پیشنهاد شده‌اند (Howard et al., 2014) که هر کدام ارزش‌های اصلاحی ژنومی را با صحت متفاوتی برآورد می‌نمایند (Baneh et al., 2017; Ghafouri-Kesbi et al., 2016; Teimuria et al., 2016). یک روش خاص همیشه نمی‌تواند روش برتر در ارزیابی ژنومی صفات مختلف با معماری ژنتیکی متفاوت باشد و لازم است روش‌های مختلف در پیش‌بینی ارزش‌های اصلاحی ژنومی حیوانات برای صفت موردنظر مورد بررسی و مقایسه قرار گیرند و بهترین روش با بیشترین صحت پیش‌بینی انتخاب مورد استفاده قرار گیرد (Neves et al., 2012).

بسیاری از صفاتی که اهمیت زیست شناختی و اقتصادی دارند به شیوه‌ای ناپیوسته تغییر می‌کنند، ولی کنترل ژنتیکی آن‌ها به مشابه صفات کمی است و توسط ژن‌های زیادی کنترل می‌شوند. صفاتی مانند حساسیت به بیماری با دسته فنوتیپی مبتلا و سالم، تعداد نوزاد در یک زایمان و درجات مختلف سخت زایی در این دسته قرار می‌گیرند. برای این صفات یک متغیر پنهان (Latent Variable, Viability) متصور است (مانند سطح یک هورمون در خون) که تغییرات آن در جامعه به صورت نرمال است. مادامی که متغیر پنهان پایین‌تر از آستانه قرار گیرد، فرد موردنظر یکی از حالات‌های فنوتیپی مثلاً طبیعی را دارد و زمانی که متغیر پنهان آستانه موردنظر را رد کند، فرد در دسته فنوتیپی دوم (بیمار) قرار می‌گیرد (شکل ۱، از de Villemereuil, 2018). برای برخی صفات دیگر مانند تعداد نتاج در هر زایش ممکن است بیشتر از یک آستانه و در نتیجه تعداد بیشتر، دسته‌های، فنوتیپی،



شکل ۱- تظاهر فنوتیپی یک صفت آستانه‌دار با یک آستانه و دو دسته فنوتیپی (de Villemereuil, 2018)
Figure 1. Phenotypic expression of a threshold trait with one threshold and two phenotypic classes
(de Villemereuil, 2018)

برای یک صفت آستانه‌دار با معماری ژنتیکی افزایشی و غالباً انجام شد.

مواد و روش‌ها

شبیه‌سازی جمعیت و ژنوم

در این تحقیق از بسته نرم‌افزاری Technow, (hyperd) در این تحقیق از بسته نرم‌افزاری (Howard, 2013) جهت شبیه‌سازی ژنوم و جمعیت استفاده شد. فرض شد که تعداد ۱۰ کروموزوم ساختار ژنوم را تشکیل می‌دهند که بر روی هر کدام ۱۰۰۰ نشانگر تک نوکلتوتیدی دو آللی (SNP) به طور یکنواخت پراکنده شده‌اند. برای شبیه‌سازی جمعیت، ابتدا یک جمعیت پایه شامل ۵۰ فرد نر و ۵۰ فرد ماده شبیه‌سازی شد و با استفاده از آمیزش تصادفی برای ۵۰ نسل، سطح قابل قبولی از LD بین نشانگر و QTL ایجاد گردید. میزان LD با استفاده از پارامتر r^2 (Hill and Robertson, 1968) برآورد شد که میزان آن 0.22 بود. اندازه جمعیت در نسل ۵۱ از 100 فرد به 1000 فرد افزایش داده شد. جمعیت حاصل اطلاعات ژنوتیپی و فنوتیپی را داشتند و جمعیت مرجع را تشکیل دادند. فنوتیپ حیوانات نسل ایجاد شد. جمعیت مرجع را توزیع نرمال با میانگین صفر و انحراف استاندارد و از جذر واریانس محیطی نمونه‌گیری شد به دست آمد و فرض شد که نتیجه مقادیر فنوتیپی نیز در جمعیت

یکی از روش‌های مورد استفاده در ارزیابی ژنومی روش Endelman (Ridge regression) می‌باشد (Endelman, 2011) که در مطالعات مختلف برای پیش‌بینی ارزش‌های اصلاحی ژنومی مورد استفاده قرار گرفته است (Baneh et al., 2017; Neves et al., 2012; Ogutu et al., 2011) و Piepho et al. (2012) با اعمال تغییراتی در پارامترهای روش arBLUP (2012) با اعمال تغییراتی در این روش در نتیجه از این روش به نام روش ریج رگرسیون-روش ششم (RR-m6) را پیشنهاد نمودند. با این حال تاکنون این روش در ارزیابی ژنومی صفات آستانه‌دار و با در نظر گرفتن معماري ژنتیکی افزایشی + غالباً مورد استفاده قرار نگرفته است و عملکرد آن در این زمینه مورد بررسی قرار نگرفته است. صحت پیش‌بینی روش‌های ریج ارزیابی ژنومی در بهترین حالت در دامنه 70 تا 80 درصد است (Howard et al., 2014). هرگونه افزایش در صحت پیش‌بینی ارزش‌های اصلاحی، افزایش در پیشرفت ژنتیکی سایانه را در پی خواهد داشت. اگر عملکرد پیش‌بینی روش RR-m6 بالا باشد می‌تواند جایگزین خوبی برای روش‌های ریج ارزیابی ژنومی باشد. بنابراین، پژوهش شبیه‌سازی شده حاضر با هدف مقایسه روش RR-m6 با برخی روش‌های ریج از نظر عملکرد پیش‌بینی (صحت پیش‌بینی، اربیتی و پراکنش ارزش‌های اصلاحی ژنومی)، سرعت محاسباتی و میزان حافظه موردنیاز

مشخص شد. در ادامه نسل ۵۲ از افراد نسل ۵۱ ایجاد شدند که فقط دارای اطلاعات ژنتیکی بودند ولی اطلاعات فنوتیپی نداشتند. نسل ۵۲، جمعیت تأیید (افراد کاندید انتخاب) بودند که ارزش‌های اصلاحی آن‌ها باید برآورده شد (جدول ۱).

مرجع دارای توزیع نرمال است. برای تبدیل فنوتیپ نرمال به فنوتیپ آستانه‌دار، ازتابع پروبیت (Gianola, 1982) استفاده شد. با استفاده از تابع Probit در نرم‌افزار R دسته فنوتیپی هر فرد با توجه به مقدار فنوتیپی هر فرد و آستانه تعیین شده،

جدول ۱- پارامترهای استفاده شده در شبیه‌سازی ژنوم و جمعیت

Table 1. Parameters used for simulation of genome and population

10	تعداد کروموزوم
1	اندازه هر کروموزوم (مورگان)
1000	Chromosome length (Morgan)
.10	تعداد نشانگر به ازاء هر کروموزوم
100	Number of marker per chromosome
.05	تعداد QTL به ازاء هر کروموزوم
Gamma	Number of QTL per chromosome
1.50	توزیع اثرات QTL effect distribution
0.05	(MAF) فرکانس اُل نادر Minor allele frequency (MAF)
51	تعداد نسل‌ها برای ایجاد عدم تعادل بیوستگی (LD) Number of generations for creating LD
52	سل نسل مرجع Reference generation
	نسل تأیید Validation generation

ژنتیکی (δ_u^2) به صورت δ_e^2/δ_u^2 برآورده شود (Ogutu et al., 2011) همچنین $D = (0 \oplus Ip)$ که \oplus نشان‌دهنده جمع مستقیم است.

در نهایت بردار ارزش‌های اصلاحی ژنومی (\hat{g}) به صورت زیر برآورده شوند:

$\hat{g} = Zu$

از بسته نرم‌افزاری (de los Campos, 2022) BGLR جهت آنالیزهای استفاده شد.

روش رگرسیون ریج-روش ششم (RR-m6) (Piepho et al., 2012) در روش RR-m6 اثرات انفرادی SNP‌ها، معادله (۳) در روش ریج رگرسیون به صورت $\hat{u} = \delta_u^2 Z^T V^{-1} (y - \mathbf{1}_n \mu)$ تغییر می‌کند. آنالیزهای RR-m6 نیز با بسته نرم‌افزاری (Schulz-Streck et al., 2015) "rrBLUPmethod6" انجام شد.

روش بهترین پیش‌بینی تاریب خطی ژنومی (GBLUP) مدل آماری استفاده شده در روش GBLUP (Teimurian et al., 2016) به صورت رابطه زیر بود:

$y = \mu + Xb + e$ در رابطه فوق، y بردار حاوی مقادیر فنوتیپی است. μ میانگین کل می‌باشد. عناصر ماتریس X شامل اعداد صفر، یک، و دو است که نشان‌دهنده تعداد آلل‌های خاص هر یک از مکان‌های نشانگری با توجه به ژنتیکی هر فرد است. b بردار ارزش‌های اصلاحی ژنومی است و e نیز بردار اثرات باقی‌مانده است. از بسته نرم‌افزاری BGLR (de los Campos, 2022) جهت آنالیزهای GBLUP استفاده شد.

روش بیز A (BayesA) مدل آماری استفاده شده در این روش به صورت زیر است (Meuwissen et al., 2001):

$$y_i = \mu + \sum x_{ij(mj)} + e_{ijk}$$

در معادله فوق، i اندیس مربوط به تعداد افراد می‌باشد. j نشان‌دهنده مکان‌های نشانگر با توجه به ژنتیکی افراد است.

سناریوهای مورد بررسی

(۱) تعداد QTL: تعداد QTL به صورت درصدی از تعداد نشانگر (۱ و ۱۰ درصد از تعداد SNP‌ها؛ به ترتیب ۱۰ و ۱۰۰ بهزاره هر کروموزوم) در نظر گرفته شد.

(۲) اثرات ژنتیکی غالبیت: برای در نظر گرفتن اثرات غالبیت در مدل، سناریوهای مختلفی در نظر گرفته شد: در سناریوی اول به همه QTL‌ها فقط اثر افزایشی داده شد (سناریو A). در سناریوی دوم به همه QTL‌ها اثر افزایشی داده شد و برای سناریوی سوم به همه QTL‌ها اثر افزایشی و غالبیت داده شد (سناریو A+100%D).

روش‌های ارزیابی ژنومی

روش رگرسیون ریج (rrBLUP)

مدل آماری استفاده شده در روش rrBLUP به صورت رابطه زیر بود:

$$y = I_n \mu + Zu + e \quad (1)$$

در رابطه فوق y بردار مشاهدات فنوتیپی می‌باشد. I_n برداری ستونی حاوی عدد ۱ و n میانگین کل می‌باشد. Z یک ماتریس با ابعاد $n \times p$ است که حاوی ژنتیکی ها می‌باشد. e نیز نشانگر SNP است و u حاوی اثرات باقی‌مانده است. ماتریس کوواریانس واریانس-کوواریانس به صورت زیر بود:

$$V = var(y) = \Gamma \delta_u^2 + R \quad (2)$$

در رابطه فوق $\Gamma = ZZ^T$ و Z^T ترانهاد Z را نشان می‌دهد.

برای برآورده اثرات انفرادی نشانگرهای SNP، روش رگرسیون ریج معادله ۱ را با استفاده از روش حداقل مربعات جریمه‌ای و طبق رابطه زیر حل می‌کند:

$$\hat{u} = (Z^T Z + \lambda^2 D)^{-1} Z^T y \quad (3)$$

در رابطه فوق، λ پارامتر جریمه است که با توجه به برآوردهای REML برای واریانس باقی‌مانده (δ_e^2) و واریانس

این الگوریتم درخت رگرسیونی بود. در **Boosting** توابع پایه مانند درختان رگرسیونی به صورت سریالی هریک روی باقیمانده درخت قبلی (اطلاعاتی که به درستی تقسیم‌بندی شده‌اند) اضافه می‌شوند و انتظار این است که با تمرکز بر داده‌های به اشتباه دسته‌بندی شده در درخت قبلی مقدار خطای در درخت بعدی کاهش پیدا کند و تا زمانی که خطای در حال کاهش است الگوریتم ادامه می‌پاید (Ghofouri-Kesbi et al., 2016). در نهایت درختی که در آن خطای حداقل شود (ممکن است آخرين درخت نباشد) برای تجزیه و تحلیل **Boosting** اطلاعات استفاده می‌شود. از این روش می‌توان گفت **Boosting** روشی نخیه‌گرا است. روش **Boosting** نیز همانند **RF** باید برای پارامترهای تنظیم شود. پارامترهای مهم در **Boosting** عبارتند از تعداد درخت (*ntree*) و نرخ یادگیری (*lr*). برای *ntry* مقداری (۱۰۰، ۲۵۰، ۵۰۰ و ۱۰۰۰) و برای *lr* مقدار (۰/۰۱، ۰/۰۱ و ۰/۰۵) در نظر گرفته شد و بهترین ترکیب (*lr* برابر ۰/۰۱ و *ntry* برابر ۵۰۰) برای آنالیزها مورد استفاده قرار گرفت.

مقایسه روش‌ها

به منظور بررسی ارزش‌های اصلاحی ژنومی برآورد شده از برخی شاخص‌های روش **LR** (Legarra and Reverter, 2018) شامل (۱) صحت پیش‌بینی: که عبارت بود از همبستگی پیرسون بین ارزش‌های اصلاحی پیش‌بینی شده و ارزش‌های اصلاحی واقعی، (۲) اریبی (Bias) که عبارت بود از میانگین ارزش‌های اصلاحی واقعی و (۳) پراکنده‌گی یا تورم (Dispersion)، (Inflation) که عبارت بود از رگرسیون ارزش‌های اصلاحی واقعی روزی ارزش‌های اصلاحی پیش‌بینی شده استفاده شد. هر سناریو مورد مطالعه که ترکیبی از روش مورد مطالعه، تعداد QTL و درصد QTL‌های دارای اثر غالیت بود، ۱۰ بار تکرار شد و میانگین ۱۰ تکرار ذخیره شد.

زمان انجام محاسبات و میزان حافظه مورد استفاده با استفاده از سنته نرم‌افزاری **Pryr** (Wickham, 2018) میزان فضای موردنیاز برای پردازش کدهای هر روش بر روی حافظه اندازه‌گیری شد. زمان محاسبات نیز با استفاده از یک تابع در نرم‌افزار **R** اندازه‌گیری گردید.

نتایج و بحث

صحت پیش‌بینی روش‌های مورد مطالعه در شکل ۱ نشان داده شده است. در تمامی روش‌های مطالعه شده، با افزایش درصد QTL‌های دارای اثر غالیت از صفر (سناریو A) به ۱۰۰٪ (سناریو A+100%D) صحت پیش‌بینی کاهش یافت. برای مثال در روش RRBLUP-m6 RRBLUP-m6 با افزایش درصد QTL‌های دارای اثر غالیت از صفر به ۵۰ و سپس به ۱۰۰٪ صحت پیش‌بینی از ۰/۶۵ به ۰/۶۲ و سپس به ۰/۵۵ کاهش یافت. در سناریو A، در هر دو حالت ۵۰ و ۵۰۰ QTL، روش‌های BayesA و BayesA و BayesA از حداقل صحت پیش‌بینی برخوردار بودند و از طرف دیگر روش RT به طور معنی‌داری از صحت پایین‌تری نسبت به سایر روش‌ها برخوردار بود. نتیجه اخیر در دو سناریو دیگر (A+50%D و A+100%D) نیز

۷) مقادیر فنوتیبی برای هر فرد بوده و μ میانگین کل می‌باشد. ۸) یک عنصر از ماتریس X برای نشانگر زام و فرد نام می‌باشد. m_j اثر تصادفی نشانگر زام بوده و e_i اثر باقیمانده است. از سنته نرم‌افزاری BGLR (de los Campos, 2022) استفاده شد.

درخت رگرسیونی (Regression Tree, RT)

تحلیل RT زمانی انجام می‌شود که متغیر مورد مطالعه یک متغیر کمی با توزیع نرمال باشد (به عنوان مثال تولید روزانه شیر در گاوها شیری) برای صفات با توزیع فنوتیبی (Classification Tree) گستته از درخت کلاس‌بندی (Classification Tree) استفاده می‌شود. فرض کنیم u بردار مشاهدات باشد X ماتریس ژنوتیبی باشد، به صورتیکه $z_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ است که ژنوتیپ هر حیوان را برای تعداد p نشانگر SNP نشان داد RT را می‌توان به صورت (y, X) نشان داد که در آن y درخت و X و u به ترتیب بردار اطلاعات ژنوتیبی و فنوتیبی می‌باشند که وارد درخت می‌شوند. مدل در جمعیت مرجع آموزش دیده و بر جمعیت تایید (حيوانات کاندیدای انتخاب) اعمال می‌شود. (۱) ابتدا نمونه‌های اطلاعات از مجموعه داده‌های آموزشی شامل اطلاعات ژنوتیبی و فنوتیبی انتخاب می‌شود (۲) در اولین گره درخت، گروه کوچکی از SNP‌ها به طور تصادفی انتخاب می‌شوند و u براساس آنکه که تابع خطای حداقل می‌کند انتخاب می‌شود. (۳) گره مورد نظر به دو زیرشاخه تقسیم می‌شود و حیوانات بر اساس آنکه که برای SNP موردنظر حمل می‌کند وارد یکی از شاخه‌ها می‌شوند. (۴) مراحل ۱ تا ۳ تکرار می‌شوند تا یکنواختی در گره‌های پایانی ایجاد شود (حيوانات با ژنوتیپ‌های مشابه در گره‌های جداگانه دسته‌بندی می‌شوند). مقدار پیش‌بینی شده برای x_i برابر است با میانگین فنوتیپ حیوانات دارای ژنوتیپ مشابه در گره پایانی است. آنالیزهای درخت رگرسیونی با استفاده از بسته نرم‌افزاری rpart (Therneau et al., 2022) انجام شد.

جنتک تصادفی (Random Forest, RF)

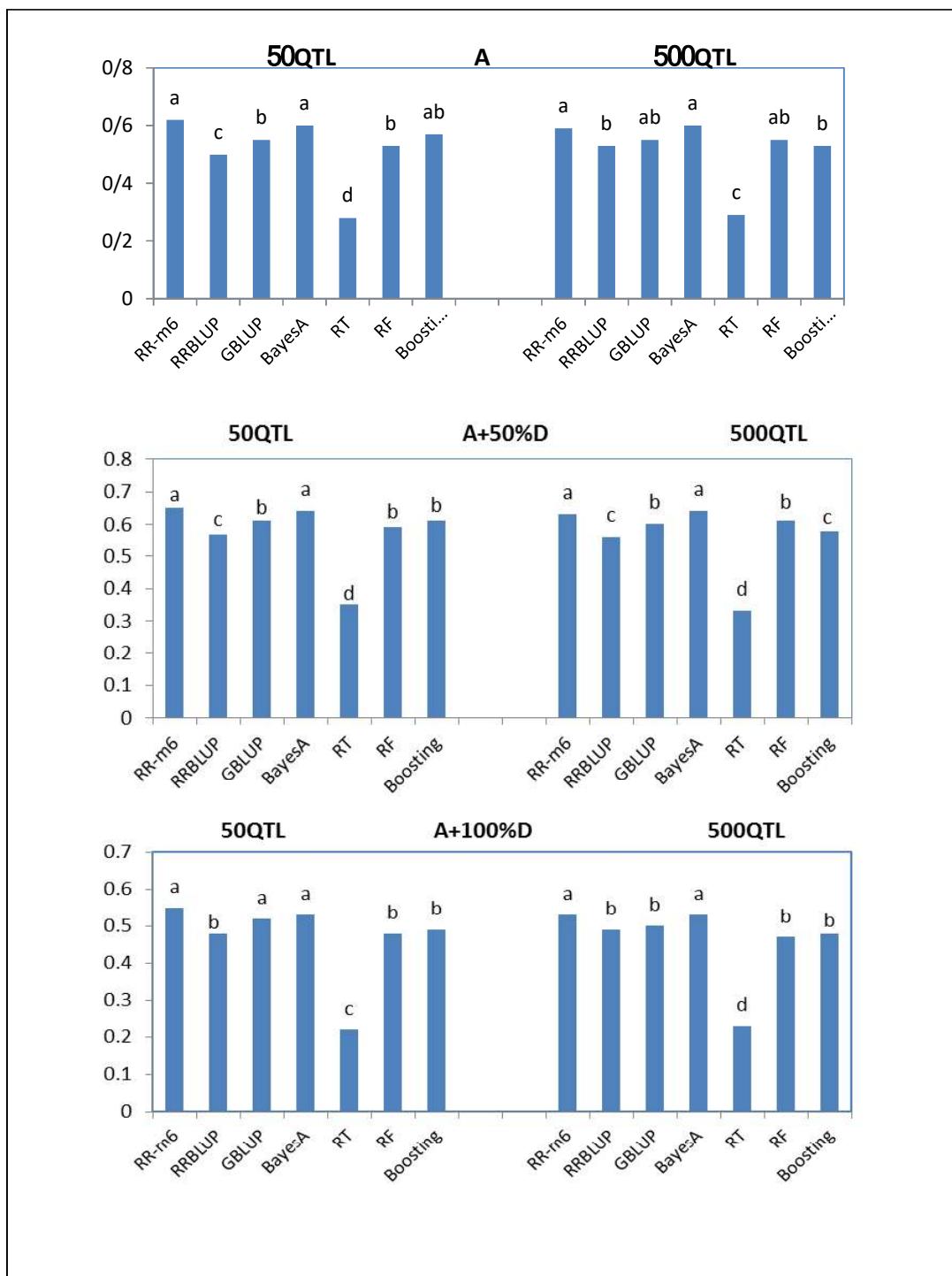
الگوریتم جنگل تصادفی (Breiman, 2001) توسط بسته نرم‌افزاری randomforest (Liaw, 2022) پیش‌بینی ارزش‌های اصلاحی ژنومی استفاده گردید. جنگل تصادفی از مجموعه‌ای از درختان رگرسیونی ایجاد می‌شود که هر کدام شامل n نمونه از اطلاعات ژنوتیبی و فنوتیبی افراد مرجع می‌باشد. در واقع از نتیجه صدها یا هزاران درخت میانگین گیری می‌شود و می‌توان گفت RF بر مبنای رأی اکثریت است. پارامترهای بسیار مهم در RF تعداد متغیر انتخاب شده در هر گره درخت (*ntry*) و تعداد درخت (*ntree*) است که قبل از انجام آنالیزها باید مقدار مناسب آنها تعیین شود. برای یافتن بهینه‌ترین ترکیب از $mtry$ و $ntree$ ترکیبات مختلفی از $mtry$ (۰، ۵۰۰، ۲۵۰، ۱۵۰۰، ۱۰۰۰) و $ntree$ (۰، ۵۰۰، ۱۰۰۰، ۱۵۰۰، ۲۰۰۰) برای RF در نظر گرفته شد. در بهینه‌ترین ترکیب که منجر به حداقل خطای پیش‌بینی و حداقل صحت پیش‌بینی گردید پارامتر $mtry$ برابر ۵۰۰ و $ntree$ برابر ۱۵۰۰ بود.

Boosting

این الگوریتم در قالب بسته نرم‌افزاری gbm (Ridgeway, 2022) مورد استفاده قرار گرفت. توابع پایه مورد استفاده در

این نظر روش RT در رتبه آخر قرار گرفت.

مشاهده شد بهصورتیکه RR-m6 و BayesA با بیشترین صحت ارزش‌های اصلاحی ژنومی را پیش‌بینی نمودند و از

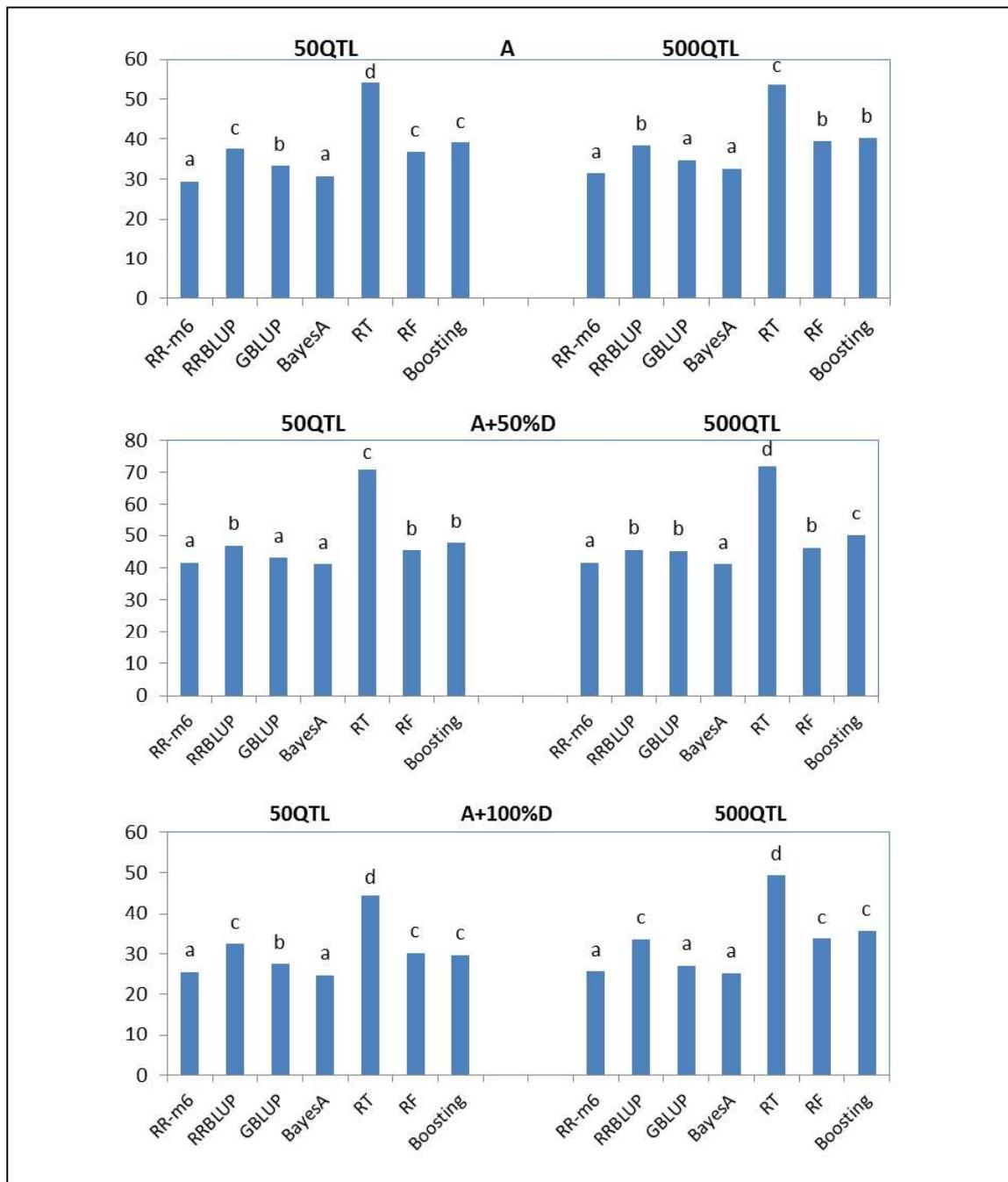


شکل ۲- صحت ارزش‌های اصلاحی ژنومی در سناریوهای مختلف از تعداد QTL و درصد QTL‌های دارای اثر غالبیت.

Figure 2. Accuracy of genomic breeding values in different scenarios of QTL number and percent of QTL with dominance effect.

نموداری بین دو روش RR-m6 و BayesA وجود داشت هر چند تفاوت بین آنها معنی‌دار نبود ($p>0.05$) و از این نظر هر دو روش رتبه اول را کسب کردند. روش GBLUP با اندکی تفاوت نسبت به این دو روش در رتبه دوم قرار گرفت و روش نیز ارزش‌های اصلاحی ژنومی بین RT (RR-m6) تا در رتبه آخر قرار گرفت.

شکل ۳ اربیی ارزش‌های اصلاحی پیش‌بینی شده توسط روش‌های مختلف را نشان می‌دهد. همانگونه که مشاهده می‌شود با افزایش درصد QTL‌های دارای اثر غالبیت از صفر (سناریو A) به ۱۰۰٪ (سناریو A+100%D) میزان اربیی ارزش‌های اصلاحی ژنومی بین ۴۰ درصد (RR-m6) تا ۶۰ درصد (RT) افزایش یافت. از نظر میزان اربیی، رقابت

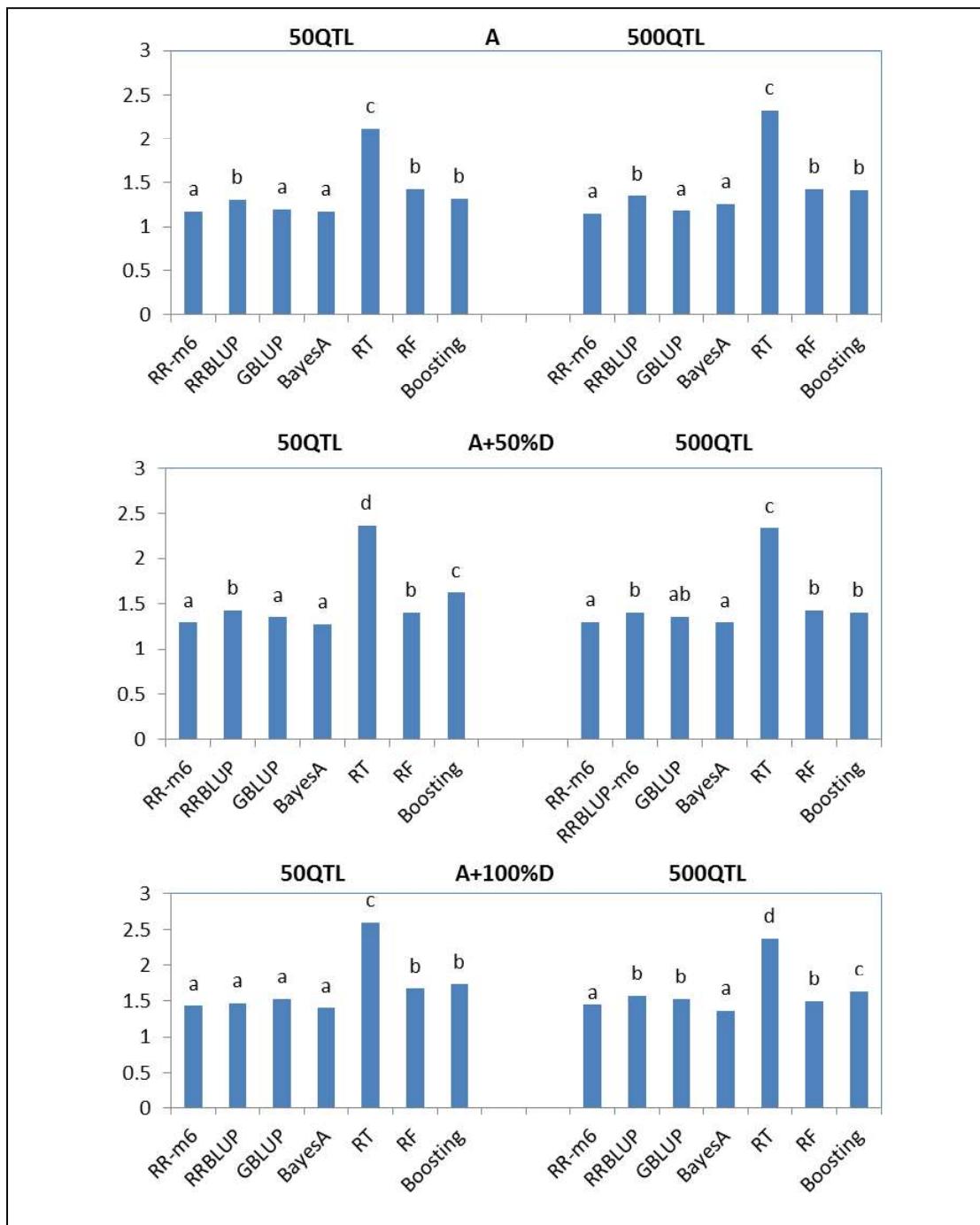


شکل ۳- اربیی ارزش‌های اصلاحی ژنومی در سناریوهای مختلف از تعداد QTL و درصد QTL‌های دارای اثر غالبیت

Figure 3. Bias of genomic breeding values in different scenarios of QTL number and percent of QTL with dominance effect.

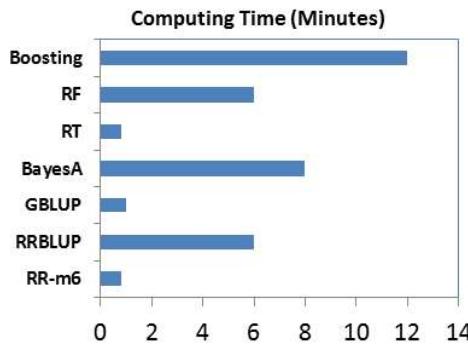
(سناریو A) به ۱۰۰٪ (سناریو A+100%D) اثر منفی بر میزان پراکندگی ارزش‌های اصلاحی گذاشته است. روش‌های GBLUP و BayesA، RR-m6 و RF ارزش‌های اصلاحی ژئومی را با حداقل پراکندگی برآورد نمودند.

در شکل ۳، میزان پراکندگی (تورم) ارزش‌های اصلاحی ژئومی پیش‌بینی شده توسط روش‌های مختلف نشان داده شده است. به طور مشابه با پارامترهای صحت و اریبی، در اینجا نیز افزایش درصد QTL‌های دارای اثر غالبیت از صفر



شکل ۴- پراکنش ارزش‌های اصلاحی حاصل از روش‌های مورد بررسی در سناریوهای مختلف از تعداد QTL و درصد QTL‌های دارای اثر غالبیت
Figure 4. Dispersion of genomic breeding values in different scenarios of QTL number and percent of QTL with dominance effect.

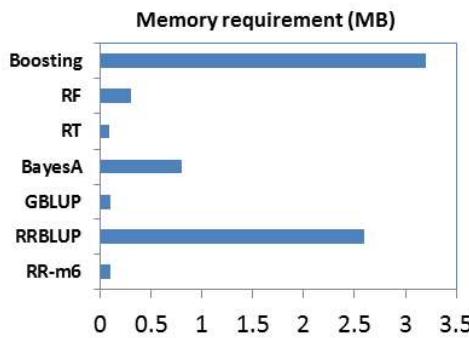
زمان محاسباتی مورد نیاز برای انجام یک آنالیز مشابه با ۱۲ دقیقه زمان مورد نیاز کنترلرین روش بود و در رتبه آخر قرار گرفت. روش های BayesA، RRBLUP و RF نیز با ۸، ۶ و ۴ دقیقه زمان محاسباتی مورد نیاز جزو روش های نسبتاً کند بودند. سریعترین روش ها بودند. روش GBLUP با ۶۰ ثانیه زمان محاسباتی مورد نیاز در رتبه دوم قرار گرفت. روش Boosting میان روش های بودند.



شکل ۵- زمان محاسباتی روش های بررسی شده
Figure 5. Computing time of the studied methods

۲/۶ های Boosting و RRBLUP نیز به ترتیب با $\frac{3}{2}$ و $\frac{1}{2}$ مگابایت رتبه های پایانی را به خود اختصاص دادند. روش بیز A نیز به حافظه نسبتاً بالایی نیاز داشت.

در شکل ۶ حافظه مورد نیاز برای پردازش کدهای روش های مورد مطالعه بر روی CPU نشان داده شده است. از این نظر، روش های RR-m6، RT و GBLUP با $\frac{1}{10}$ مگابایت حافظه مورد نیاز در رتبه اول قرار گرفتند و روش



شکل ۶- حافظه مورد نیاز روش های بررسی شده
Figure 6. Memory requirement of the studied methods

اثرات ژنتیکی غالبیت از اثرات ژنتیکی افزایشی زمانی که اثرات غالبیت وجود داشته باشند، منجر به برآوردهای ارزش های اصلاحی با صحت پیش بینی پایین، اریب و غیر قابل اعتماد خواهد شد. محمدی (Mohammadi, 2018) صفات شیر تولیدی، مقدار چربی و پروتئین را در گاو های شیری هشتادین ایران را ارزیابی ژنومی نمود. برای صفات مذکور اگرچه میزان مشارکت اثرات ژنتیکی غالبیت در تنوع فتوتیپی صفات مذکور کم بود (به ترتیب $\%21$ ، $\%11$ و $\%1$)، اما لحاظ نمودن اثرات غالبیت در مدل، صحت ارزیابی ژنومی برای صفات تولید شیر، مقدار چربی و پروتئین را به ترتیب به میزان $\%3$ ، $\%5$ و $\%6$ در مقایسه با مدلی که فقط شامل اثرات افزایشی بود افزایش داد. همچنین محمدی و ساعتی مختاری (Mohammadi and Sattaei Mokhtari, 2018) با

اثرات غالبیت و ارزیابی ژنومی در اکثر مطالعات انجام شده در ارتباط با ارزیابی ژنومی چه با استفاده از داده های واقعی و چه با استفاده از داده های شبیه سازی شده تنها اثرات ژنتیکی افزایشی ژنها در نظر گرفته شده و از اثرات غیر افزایشی مانند اپیستازی و غالبیت چشم پوشی شده است (Ashoori-Banaei et al., 2021; Ridgeway, 2017; Zhang et al., 2022). مطالعات اخیر نشان میدهد که اثرات غالبیت و اپیستازی در تنوع فتوتیپی صفات اقتصادی دام های اهلی مشارکت دارند و در صورتی که این مشارکت قابل توجه باشد در نظر گرفتن این اثرات در مدل می تواند منجر به افزایش صحت ارزیابی ژنومی شود (Aliloo et al., 2016). نتایج مطالعه ما نشان داد که استفاده از یک مدل صرفاً افزایشی و در نتیجه عدم تفکیک

همچنین ارزش‌های اصلاحی حاصل از این روش جزء نالریب ترین برآوردها بوده و از این نظر با برآوردهای روش BayesA قابل مقایسه بوده و در برخی موارد کمتر هم بود، اگرچه تفاوت‌ها معنی دار نبودند. از یک دیدگاه اصلاحی ترازدی برای پیش‌بینی صحیح پاسخ به انتخاب، برآوردهای نالریب از ارزش‌های اصلاحی مورد نیاز است (Macedo et al., 2020). اریبی برآورد شده در تحقیق حاضر برای همه روش‌ها مثبت بود. به عبارت دیگر برآوردهای ارزش‌های اصلاحی برای افراد کاندید انتخاب بیشتر از مقدار ارزش‌های اصلاحی واقعی (شیوه‌سازی شده) آنها بود. این مسئله منجر به برآوردهای بیش از انتظار از پیشرفت ژنتیکی سالیانه خواهد شد که برای روش‌های RT, RRBLUP, RF و Boosting شد که سایر روش‌های بررسی شده بیشتر خواهد بود. از نظر شاخص پراکندگی نیز برآوردهای روش RR-m6 از حداقل پراکندگی برخوردار بودند و از این نظر نیز در کنار روش‌های GBLUP و BayesA جزو بهترین روش‌ها بود. عملکرد پیش‌بینی برابر یا بهتر روش روش RR-m6 نسبت به دو روش GBLUP و BayesA از این نظر حائز اهمیت است که دو روش اخیر از بهترین روش‌های ارزیابی ژنومی به شمار می‌روند. برای مثال، عبدالهی اربناهی (Abdolahi Arpanahi, 2017) دو روش پارامتری GBLUP و BayesB، روش شبه پارامتری RKHS و روش‌های غیرپارامتری RF و Bagging را برای ارزیابی ژنومی صفات کمی با معماری ژنتیکی افزایشی مقایسه نمودند و گزارش کردند که در کل دو روش GBLUP و BayesB عملکرد بهتری نسبت به روش‌های شبه پارامتری و ناپارامتری داشتند. همچنین صالحی و همکاران (Salehi et al., 2021) در یک مطالعه شیوه سازی، عملکرد پیش‌بینی روش‌های پارامتری GBLUP و BayesB را با دو روش ناپارامتری BagBLUP و RF مورد مقایسه قرار دادند. این محققین گزارش کردند که روش BayesB از حداکثر صحت پیش‌بینی برخوردار بود (۰/۷۳) و روش‌های Random Forest و GBLUP با ۰/۸۳ و ۰/۶۹ در رتبه‌های بعدی قرار گرفتند. از طرف دیگر روش RR-m6 از سرعت محاسباتی سیار بالایی برخوردار بود که باعث می‌شود محاسبات را در مدت زمان کوتاهی به انجام برساند و در ضمن بهطور معنی‌داری به حافظه کمتری برای انجام محاسبات نیازمند بود (شکل‌های ۴ و ۵) که این برتری بر روش‌های دیگر کاملاً مشهود است. حتی این روش بر روش GBLUP که یکی از سریع‌ترین روش‌های پیش‌بینی ارزش‌های اصلاحی است تا حدودی برتری داشت. توجه کنیم که ممکن است در مورد داده‌هایی با حجم کوچک این برتری چندان مهم نباشد اما در زمان استفاده از داده‌های حجمی جایی که نتایج باید به سرعت در دسترس محققین قرار گیرد مسئله متفاوت است. برای مثال در تحقیق حاضر ابعاد ماتریس ژنوتیپی 1000×1000 بود که ماتریس ژنوتیپی کوچکی محسوب می‌شود. اخیراً تراشه‌های SNP که در برگیرنده SNP ۷۷۷۰۰۰ هستند برای گاو شیری توسعه داده شده‌اند نظر گرفتن جمعیت مرجع برابر ۱۰۰۰ فرد ابعاد ماتریس

استفاده از شبیه‌سازی کامپیوتری، نشان دادند که تغییک اثرات غالبیت از اثرات افزایشی، صحت ارزیابی ژنومی را افزایش خواهد داد بهصورتی که در روش BayesA و در روش BayesL صحت ارزیابی ژنومی از ۰/۶۳ به ۰/۶۹ و در روش BayesS صحت ارزیابی ژنومی از ۰/۶۵ به ۰/۶۷ افزایش یافت. همچنین افزایش در اریبی ارزش‌های اصلاحی ژنومی در نتیجه چشم‌پوشی از اثرات ژنتیکی غالیت در مدل ارزیابی ژنومی توسط محمدی و ساعتی مختاری (Mohammadi and Sattaei Mokhtari, 2018) گزارش شده است که با نتایج تحقیق حاضر در تطابق است. علیلو و همکاران (Aliloo et al., 2016) گزارش کردند که یکی از عواقب چشم‌پوشی از اثرات غالبیت، ارزیابی ژنومی اریب افراد کاندید انتخاب است که منجر به رتبه‌بندی اشتباہ افراد بر اساس ارزش‌های اصلاحی خواهد شد. در مطالعه علیلو و همکاران (Aliloo et al., 2016) وارد نمودن اثرات غالبیت در مدل ارزیابی ژنومی منجر به تغییر رتبه حیوانات بر اساس ارزش اصلاحی شد بهصورتیکه همبستگی رتبه‌بندی حیوانات برتر در دو مدل حاوی اثرات افزایشی و مدل حاوی اثرات افزایشی و غالبیت ۹۰٪ بود. در ارتباط با تأثیر وارد نمودن اثرات غالبیت در مدل منجر به برآوردهای کمتر از حد واقع از ارزش‌های اصلاحی ژنومی (deflated breeding values) برای افراد کاندید انتخاب شده است (Liaw, 2022).

مقایسه روش‌ها

تاکنون روش‌های زیادی بر پایه الگوریتم‌های مختلف برای پیش‌بینی ارزش‌های اصلاحی ژنومی مورد استفاده قرار گرفته‌اند. این روش‌ها دامنه وسیعی از روش‌های بسیار ساده مانند روش‌های حداقل مربعات معمولی تا روش‌های پیچیده مبتنی بر هوش مصنوعی را شامل می‌شوند (Howard et al., 2014). در یک دیدگاه کلی، سه پارامتر عملکرد کلی یک سرعت محاسباتی و مقدار حافظه مولود نیاز افزایشی عملکرد کلی یک روش را تحت تاثیر قرار می‌دهد (Ashoori-Banaei et al., 2021). از نظر صحت پیش‌بینی، روشی که صحت آن تزدیک به ۱ باشد، از نظر اریبی روشی که اریبی آن تزدیک به صفر باشد و از نظر پراکندگی روشی که ضریب پراکندگی آن به ۱ تزدیکتر باشد برای ارزیابی ژنومی مطلوب تر است (Macedo et al., 2020). صحت پیش‌بینی روش RR-m6 بالاتر از روش rrBLUP بود و در بیشتر سناریوها عملکردی برابر یا بهتری نسبت به دو روش GBLUP و BayesA داشت که از جمله بهترین روش‌های انتخاب ژنومی هستند. صحت بالا به این معنی است که ارزش‌های اصلاحی ژنومی حیوانات صحیح‌تر برآورده شده و لذا رتبه‌بندی حیوانات بر اساس ارزش‌های اصلاحی ژنومی آنها درست‌تر انجام خواهد شد که در جای خود پیشرفت ژنتیکی بالاتر را باعث می‌شود. از این نظر هر درصد افزایش در صحت ارزش‌های اصلاحی ژنومی حائز اهمیت است چرا که به حداقل‌سازی پیشرفت ژنتیکی کمک خواهد کرد و از این لحاظ روش RR-m6 دارای مزیت است.

ژنتیکی $\times 77000 \times 1000$ خواهد شد که تجزیه چنین داده های بزرگی با روش های مانند روش های بیزی مدت زمان طولانی و نیازمندی های محاسباتی بالایی را می طلبد که ممکن است انجام آن ها با کامپیوتر های معمولی غیر ممکن شود. در آینده نزدیک حتی ممکن است تراشه های بسیار متراکم تری که در برگیرنده اطلاعات چند میلیون SNP باشند جهت استفاده برای اهداف اصلاح نژادی توسعه داده شوند. در چنین شرایطی است که برتری معنی دار روش های مانند RR-m6 بر سایر روش ها مشخص می شود.

نتیجه گیری کلی

نتایج نشان داد که استفاده از یک مدل صرفاً افزایشی زمانی که اثرات غالیت در تنوع ژنتیکی صفت مشارکت داشته باشد منجر به ارزیابی ژنومی نامناسبی خواهد شد. ارزش های اصلاحی با صحت پایین، اربی و غیر قابل اعتماد نتیجه چنین ارزیابی ژنومی خواهد بود. روش RR-m6 را می توان به عنوان یک روش جایگزین برای روش های رایج ارزیابی ژنومی خصوصاً در مواجهه با داده های بزرگ پیشنهاد کرد چرا که از عملکرد پیش بینی بالایی برخوردار بود، بسیار سریع بود و نیازمندی های محاسباتی بسیار کمی داشت.

References

- AbdolahiArpanahi, R. (2017). Comparison of parametric and resampling methods in genetic evaluation of quantitative traits with different genetic structure. *Animal Production*, 19(1), 1-12. 10.22059/jap.2017.206657.623047.
- Aliloo, H., Pryce, J. E., González-Recio, O., Cocks, B. G., & Hayes, B. J. (2016). Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. *Genetics Selection Evolution*, 48(1), 1-11. 10.1186/s12711-016-0186-0.
- Ashoori-Banaei, S., Ghafouri-Kesbi, F., & Ahmadi, A. (2021). Comparison of regression tree-based methods in genomic selection. *Journal of Genetics*, 100, 1-8. 10.1007/s12041-021-01334-x.
- Baneh, H., Nejati-Javaremi, A., Rahimi-Mianji, G. H., & Honarvar, M. (2017). Genomic evaluation of threshold traits with different genetic architecture using bayesian approaches. *Research on Animal Production*, 8(6). 10.29252/rap.8.15.149.
- Breiman, L. (2001). Random Forests. *Machine learning*, 45, 5-32.
- de Villemereuil, P. (2018). Quantitative Genetic Methods Depending on the Nature of the Phenotypic Trait. *Annals of the New York Academy of Sciences*, 1422(1), 29-47. 10.1111/nyas.13571.
- Doublet, A. C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., ... & Restoux, G. (2019). The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. *Genetics Selection Evolution*, 51, 1-13. 10.1186/s12711-019-0495-1.
- Endelman, J.B. (2011). Ridge Regression and Other Kernels for Genomic Selection with R Package rrblup. *The plant genome*, 4(3) 10.3835/plantgenome2011.08.0024.
- Ghafouri-Kesbi, F., Rahimi-Mianji, G., Honarvar, M., & Nejati-Javaremi, A. (2016). 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), 229-236. 10.1071/AN15538.
- Gianola, D. (1982). Theory and Analysis of Threshold Characters. *Journal of animal Science*, 54(5), 1079-1096. 10.2527/jas1982.5451079x.
- de los Campos, P.P.R. (2022). BGLR: Bayesian Generalized Linear Regression. Available At: <https://cran.r-project.org/web/packages/BGLR/BGLR.pdf>
- Henderson, C.R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics*, 423-447. 10.2307/2529430.
- Hill, W. & A. Robertson. (1968). Linkage Disequilibrium in Finite Populations. *Theoretical and applied genetics*, 38, 226-231. 10.1007/BF01245622.
- Hill, W.G. (2008). Estimation, Effectiveness and Opportunities of Long Term Genetic Improvement in Animals and Maize. *Lohmann information*, 43(1), 3-20.
- Howard, R., Carriquiry, A. L., & Beavis, W. D. (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes, Genomes, Genetics*, 4(6), 1027-1046. 10.1534/g3.114.010298.
- Legarra, A., & Reverter, A. (2018). Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. *Genetics Selection Evolution*, 50, 1-18. 10.1186/s12711-018-0426-6.
- Liaw, A. (2022). Breiman and Cutler's Random Forests for Classification and Regression. Available At: <http://cran.r-project.org/web/packages/randomForest/index.html>
- Macedo, F. L., Christensen, O. F., Astruc, J. M., Aguilar, I., Masuda, Y., & Legarra, A. (2020). Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genetics Selection Evolution*, 52(1), 1-10. 10.1186/s12711-020-00567-1.
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *genetics*, 157(4), 1819-1829. 10.1093/genetics/157.4.1819.

- Mohammadi, Y. (2018). Accuracy of Genomic Selection Using Models with Additive Effects for Productive Traits in Iranian Holstein Cows. *The second international conference and the third national conference on agriculture, environment and food security*, Jiroft University, Jiroft, Iran.
- Mohammadi, Y., & Sattaei Mokhtari, M. (2018). Genomic selection accuracy parametric and nonparametric statistical methods with additive and dominance genetic architectures. *Research On Animal Production* (Scientific and Research), 8(18), 161-167. 10.29252/rap.8.18.161.
- Neves, H. H., Carvalheiro, R., & Queiroz, S. A. (2012). A comparison of statistical methods for genomic selection in a mice population. *BMC genetics*, 13(1), 1-17. 10.1186/1471-2156-13-100.
- Ogutu, J. O., Piepho, H. P., & Schulz-Streeck, T. (2011). A comparison of random forests, boosting and support vector machines for genomic selection. In *BMC proceedings* (Vol. 5, No. 3, pp. 1-5). BioMed Central.
- Piepho, H. P., Ogutu, J. O., Schulz-Streeck, T., Estaghvirou, B., Gordillo, A., & Technow, F. (2012). Efficient computation of ridge-regression best linear unbiased prediction in genomic selection in plant breeding. *Crop Science*, 52(3), 1093-1104. 10.2135/cropsci2011.11.0592.
- Ridgeway, G. (2022). gbm: Generalized Boosted Regression Models a Available At <http://cran.r-project.org/packages/gbm/index.html>.
- Salehi, A., Bazrafshan, M., & Abdollahi-Arpanahi, R. (2021). Assessment of parametric and non-parametric methods for prediction of quantitative traits with non-additive genetic architecture. *Annals of Animal Science*, 21(2), 469-484. 10.2478/aoas-2020-0087.
- Schulz-Streeck, T., B. Estaghvirou & F. Technow. (2015). Package ‘Rrblup-Method6’. Available At: <https://mran.microsoft.com/snapshot/2016-10-12/web/packages/rrBlupMethod6/index.html>
- Technow, F. (2013). Hypred: Simulation of Genomic Data in Applied Genetics. Available At: <http://cran.r-project.org/web/packages/hypred/index.html>
- Teimuria, M., Shariati, M. M., & Aslaminejad, A. A. (2016). Comparison of methods for the implementation of genomic selection in Holstein. *Research on Animal Production*, 7(14), 198-203. 7(14), 198-203. 10.29252/rap.7.14.203.
- Therneau, T., B. Atkinson and B. Ripley B. 2022. Rpart: Recursive Partitioning for Classification, Regression and Survival Trees. An Implementation of Most of the Functionality of the 1984 Book by Breiman, Friedman, Olshen and Stone. Available At: <https://cran.r-project.org/web/packages/rpart/index.html>
- Wickham, H. (2018). Pryr: Useful Tools to Pry Back the Covers of R and Understand the Language at a Deeper Level. Available At: <https://cran.r-project.org/web/packages/pryr/index.html>
- Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., ... & Zhang, X. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Frontiers in Plant Science*, 8, 1916. 10.3389/fpls.2017.01916.