ژنتیک نوین
دوره چهاردهم، شماره ٤، زمستان ١٣٩٨
صفحه ٣٦٦ - ٣٥٧

ارزیابی مقایسه‌ای روش‌های Boosting، SS-GBLUP و SS-BayesA: با در نظرگرفتن ایمپیوتیشن داده‌های ژنومی

Comparative Evaluation of Boosting, SS-GBLUP, SS-BayesA Methods: Consideration of Genomic Data Imputation

یوسف نادری*[1]، اورنگ استقامت[2]

۱- استادیار، گروه علوم دامی، باشگاه پژوهشگران جوان و نخبگان، دانشگاه آزاد اسلامی، واحد آستارا، آستارا، ایران

۲- استادیار، گروه علوم دامی، دانشگاه آزاد اسلامی، واحد آستارا، آستارا، ایران

Naderi Y*[1], Esteghamat O[2]

1. Assistant Professor, Department of Animal Science, Young Resaerchers and Elite Club, Astara Branch, Islamic Azad University, Astara, Iran

2. Assistant Professor, Department of Animal Science, Astara Branch, Islamic Azad University, Astara, Iran

* نویسنده مسئول مکاتبات، پست الکترونیکی: yousefnaderi@gmail.com

(تاریخ دریافت: ۹۷/۱۲/۱۳ - تاریخ پذیرش: ۹۸/۷/۱۳)

چکیده

امروزه جهت غلبه بر محدودیت ژنوتایپینگ برخی از حیوانات، مدل‌های تک مرحله‌ای پیشنهاد شدند. مدل‌های پیش‌بینی ژنومی تک مرحله‌ای با توجه به عملکرد بالا و برآورد ارزش اصلاحی ژنومی هم‌زمان برای حیوانات ژنوتیپ‌شده و نشده، تبدیل به ابزاری غالب در ارزیابی ژنومی دام‌های اهلی شدند. هدف از تحقیق حاضر، بررسی نقش روابط خویشاوندی ژنومی بین جمعیت مرجع و تأیید و معماری‌های مختلف ژنومی بر عملکرد روش‌های بوستینگ و تک مرحله‌ای بیز A و GBLUP با در نظر گرفتن ایمپیوتیشن (Imputation) داده‌های ژنومی شبیه‌سازی‌شده بود. بدین منظور، جمعیت‌های ژنومی برای سطوح مختلف تعداد جایگاه‌های صفات کمی (۱۰، ۱۰۰ و ۱۰۰۰) بر روی ۲۹ کروموزم شبیه‌سازی شدند. برای شبیه‌سازی شرایط واقعی، به‌طور تصادفی اقدام به حذف (۷۰ درصد) برخی نشانگرها نموده و در مرحله بعد از طریق نرم‌افزار Flmpute اقدام به ایمپیوتیشن و پیش‌بینی نقاط گم شده نموده و صحت ایمپیوتیشن مورد ارزیابی قرار گرفت. در نهایت داده‌های اصلی و ایمپیوتیشن با استفاده از روش‌های بوستینگ و تک مرحله‌ای بیز A و تک مرحله‌ای GBLUP جهت پیش‌بینی ارزش‌های اصلاحی ژنومی طی نسل‌های G1 و G3 مورد ارزیابی قرار گرفتند. بر طبق نتایج، صحت پیش‌بینی ژنومی برای افراد ژنوتیپ‌نشده در مقایسه با افراد ژنوتیپ‌شده با شدت بیشتری تحت تاثیر روابط خویشاوندی بین جمعیت مرجع و تأیید در هر دو سری داده ایمپیوتیشن و اصلی قرار گرفت. کمترین میزان صحت پیش‌بینی ژنومی برای افراد ژنوتیپ‌شده در هر دو سری داده ایمپیوتیشن و اصلی برای روش بوستینگ مشاهده شد. در مقایسه با روش‌های تک‌مرحله‌ای بیز A و بوستینگ، روش تک مرحله‌ای GBLUP عملکرد بالاتری در تعداد بالای QTL نشان داد. به‌طور کلی وجود روابط خویشاوندی بین جمعیت مرجع و تائید نقش مهمی در آنالیز روش‌های تک مرحله‌ای و Boosting ایفا کرد، با این حال سودمندی روش تک مرحله‌ای بیز A هنگامی که صفات تحت تاثیر تعداد کمتر QTL قرار گیرند، مشهودتر بود.

واژه‌های کلیدی

انتخاب ژنومی
روابط خویشاوندی ژنومی
روش تک مرحله‌ای
صحت ژنومی
یادگیری ماشین

[DOR: 20.1001.1.20084439.1398.14.4.9.9]

مقدمه

در اواخر دهه ۸۰ میلادی مطالعات و بررسی‌های به‌عمل آمده روشن نمود که مکانیسم‌های مولکولی در زمره مهم‌ترین فرایندهای ژنتیکی (مشتمل بر همانندسازی DNA، رونویسی، ترجمه و حتی نحوه تنظیم ژن ها) هستند (Mohammadabadi and Tohidinejad 2017) ماده ژنتیکی[1] یک سلول دارای تعداد زیادی ژن می‌باشد که هیچ‌گاه به‌طور هم‌زمان بیان نمی‌شوند و در یک زمان خاص فقط تعداد کمی از آن‌ها بیان شده و پروتئین یا آنزیم مورد نیاز سلول را تولید می‌نمایند (Tohidi nezhad et al. 2015) مقدار محصولات ژن که در همان بافت و نیز در سایر بافت‌هایی که آن محصول را می‌سازند، ساخته شده سبب تنظیم بیان آن ژن می‌شود (Mohammadabadi et al. 2017). یکی از اقدامات اساسی در حیوانات اهلی مطالعه ژن‌ها و پروتئین‌های مرتبط با صفات اقتصادی و مطالعه آن‌ها در سطح سلولی یا کروموزومی است (Jafari Darehdor et al. 2016).

در دهه‌ی ۹۰ میلادی، پژوهش‌های زیادی با استفاده از اطلاعات DNA و انتخاب با کمک نشانگرها انجام گرفته است (Naderi 2018a) که تا حدودی افزایش پیشرفت ژنتیکی را به دنبال داشت، اما با توجه به اینکه نشانگرهای ژنتیکی متداول توجیه بخش کمی از واریانس ژنتیکی را در بر می‌گرفتند، باعث ایجاد محدودیت در این روش شد (Goddard and Hayes 2007). با ظهور انتخاب ژنومی (Nejati-Javaremi et al. 1997) و ارائه روش‌ها و اصول آن (Meuwissen et al. 2001)، تحول چشمگیری در افزایش صحت ارزیابی ارزش‌های اصلاحی و بهره‌وری حیوانات اهلی مشاهده شد (Hayes 2007).

اطلاعات به‌دست آمده از مدل‌های آماری و تحلیل داده‌های زیستی به‌وسیله علم بیوانفورماتیک، در به خط کردن توالی‌ها در بانک‌های اطلاعاتی برای یافتن شباهت‌ها و تفاوت‌های ژنی، پیش‌گویی ساختار و عملکرد محصولات ژن‌ها و یافتن ارتباط فیلوژنتیک میان ژن‌ها و توالی‌های پروتئینی کمک می‌کند (Hadizadeh et al. 2013; Hadizadeh et al. 2014). استفاده بهتر از انتخاب ژنومی و کاربرد آن در ارزیابی ارزش‌های اصلاحی حیوانات همگام با توسعه این روش‌های مولکولی، به فهم بالای روش‌های آماری مورد استفاده در انتخاب ژنومی وابسته است. گسترش روش‌های آماری رایج و متعدد انتخاب ژنومی (روش‌های چند مرحله‌ای) تحول شگرفی در پیشبرد اهداف اصلاحی به‌همراه داشت. در سال‌های اخیر روش‌های یادگیری ماشین[2] به‌طور گسترده‌ای جهت حل چالش‌های ارزیابی ژنومی مطرح شده‌اند (Naderi 2018b; González-Recio and Forni 2011). بوستینگ[3] از جمله روش‌های قدرتمند یادگیری ماشین است (Breiman 2001) که علاوه بر قدرت بالا در برآورد ارزیابی ژنومی، در تشخیص ژن-ژن، پروتئین-پروتئین، اثر متقابل ژن-محیط، ژن‌های مرتبط با بیماری، مدل‌سازی جهت ارتباط میان ترکیب نشانگرها، انتخاب ژن‌های در ارتباط با صفت هدف، شناسایی فاکتورهای تنظیمی در توالی آمینو اسیدها و DNA نقش برجسته‌ای دارند (Yang et al. 2010). تحقیقات در مورد استفاده از روش بوستینگ در ارزیابی ژنومی نشان از برتری این روش‌نسبت به روش‌های بیزی دارد (González-Recio and Forni 2011). مطالعات دیگر از قابل مقایسه بودن قدرت ارزیابی ژنومی روش بوستینگ با GBLUP سخن به میان آورده‌اند (Ghafouri-Kesbi et al. 2017). با این حال، تفاوت عمده این روش‌ها در فرضیات در نظرگرفته شده برای مدل ژنتیکی پشت صحنه آن‌ها می‌باشد (Naderi 2018b). ارزیابی ژنومی روش‌ها رایج یادگیری ماشین و چند مرحله‌ای محدودیت‌هایی از جمله برای افراد ژنوتیپ‌نشده و عدم استفاده همزمان اطلاعات در مدل را به همراه داشت که در دهه اخیر با تحول چشمگیر روش‌های تک مرحله‌ای و با به‌کارگیری همزمان اطلاعات ژنومی و شجره‌ای جهت برآورد ارزش‌های اصلاحی حیوانات ژنوتیپ‌شده و نشده نقش برجسته‌ای در پیشبرد اهداف اصلاحی ایفا کرد (Legarra et al. 2009; Christensen and Lund 2010). مطالعات انجام شده در ارزیابی ژنومی حیوانات اهلی از جمله جوجه گوشتی، گاو شیری و گوشتی نشان داد که روش‌های تک مرحله‌ای عملکرد بالایی در پیش‌بینی صحت ژنومی حیوانات ژنوتیپ‌شده و نشده از خود نشان می‌دهند (Aguilar et al. 2010; Chen et al. 2011;

[1] DNA

[2] Machine learning
[3] Boosting

Liu et al. 2014). با این حال مطالعات محدودی در زمینه ارزیابی صحت پیش‌بینی ژنومی روش‌های تک مرحله‌ای بیزی (Fernando et al. 2014) در مدل‌سازی و برآورد اثر نشانگرها انجام شده‌است. لذا در تحقیق حاضر سعی بر آن شده‌است که با استفاده از یک مطالعه شبیه‌سازی به بررسی عملکرد ژنومی روش‌های تک مرحله‌ای GBLUP (SS-GBLUP) و بیزA (-SS BayesA) و بوستینگ در ترکیبات مختلف معماری ژنومی از جمله تعداد مختلف QTL در جمعیت‌های با روابط خویشاوندی ژنومی در طی چند نسل پرداخته شود. جهت دسترسی به این هدف و با توجه به اهمیت ایمپیوتیشن در پیش‌بینی نشانگرهای تعیین ژنوتیپ‌نشده در جمعیت آزمون با استفاده از الگوی تنوع هاپلوتیپی جمعیت مرجع، به بررسی این عامل (با توجه به نقش محوری آن در کاهش هزینه‌های اقتصادی ارزیابی ژنومی) بر عملکرد مدل‌های فوق الذکر نیز پرداخته خواهد شد.

مواد و روش‌ها

جهت شبیه‌سازی سناریوها از نرم‌افزار QMSim استفاده شد (Sargolzaei and Schenkel 2009). نخست، یک جمعیت پایه ۱۰۰۰ راسی برای ۱۸۰۰ نسل شبیه‌سازی شد. جهت ایجاد عدم تعادل پیوستگی در جمعیت فوق الذکر، پس از شبیه‌سازی جمعیت پایه، تعداد افراد جمعیت از طریق ایجاد یک گلوگاه ژنتیکی (Bottleneck) به ۵۰۰ رأس (در نسل‌های ۱۸۰۱ تا ۱۹۰۰) کاهش یافت. سپس در آخرین جمعیت پایه، بعد از ۱۰۰ نسل (در نسل ۲۰۰۰) تعداد افراد جمعیت به ۱۰۰۰ رأس افزایش داده شد. برای ایجاد جمعیت مرجع و تأیید، همه افراد آخرین نسل جمعیت پایه برای تولید مثل در جمعیت حاضر مورد استفاده قرار گرفتند که در این بین ۵۰ رأس نر در نظر گرفته شد تا منعکس کننده‌ی نسبت نر به ماده‌ی موجود در گله‌های گاو شیری باشد تا بتوان اثر تکنیک تلقیح مصنوعی بر نسبت نر به ماده را تقلید کرد (Yin *et al.* 2014). نوع سیستم تلاقی تصادفی بود و برای ٦ نسل دیگر (تا نسل ۲۰۰٦) 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 به‌ترتیب ۵۰ و ۲۰ درصد در نظر گرفته شد. در جمعیت اخیر، انتخاب حیوانات برتر برای نسل بعد بر اساس ارزش اصلاحی بالا و معیار حذف بر اساس ارزش اصلاحی پایین و سن صورت گرفت. نشانگرها به صورت دو آللی و به‌صورت تصادفی برای هر یک از کروموزوم‌ها (۲۹ کروموزوم در دامنه‌ی ٤۲ تا ۱۵۸ سانتی مورگان) توزیع شدند (jas Barazandeh et al. 2016). با توجه به اندازه کل ژنوم گاو (Barazandeh et al. 2016 Czech) به ازای هر کروموزوم تعداد متفاوت نشانگر (دامنه‌ی ۱۸۰ تا ٦۵۳ با توجه به نوع و اندازه کروموزوم) جهت تولید پنل‌های ۱۰K شبیه سازی شد.

در مجموع سه سطح مختلف QTL (۱۰، ۱۰۰ و ۱۰۰۰) در طول کروموزم‌ها توزیع شدند. نرخ جهش برای نشانگرها و QTLها در هر جایگاه و در هر نسل $۲/۵\times۱۰^{-۵}$ فرض شد (Naderi 2018c; Yin *et al.* 2014). وراثت‌پذیری ۰/۳ در نظر گرفته شد. در طراحی جمعیت نهایی، ۵۰ درصد از افراد نسل‌های ۲۰۰٤ و ۲۰۰٦ (به‌عنوان جمعیت تأیید: دارای اطلاعات ژنوتیپی داشته اما فاقد اطلاعات فنوتیپی) و نسل‌های ۲۰۰۰ تا ۲۰۰۳ (به‌عنوان جمعیت مرجع: دارای اطلاعات فنوتیپی و ژنوتیپی) ژنوتیپ آن‌ها حذف شد و به عنوان افراد ژنوتیپ‌نشده در نظر گرفته شدند. توزیع احتمال QTLها، نرمال فرض شد. همچنین فراوانی آللی اولیه برای نشانگرها ۰/۵ در نظر گرفته شد. در هر نسل و هر جایگاه کل میزان واریانس افزایشی توسط QTL توجیه شد. خلاصه جمعیت شبیه‌سازی شده و پارامترهای به‌کار رفته در جدول ۱ نشان داده شده‌است.

در این تحقیق نشانگرهای با فراوانی آللی کمیاب کمتر از ۰/۰۱ حذف شدند. برای ارزیابی مدل ۱۰ تکرار از هر جمعیت در نظر گرفته شد. بعد از شبیه‌سازی جمعیت با تراکم ۱۰K، با کمک برنامه نویسی در نرم‌افزار R و بطور تصادفی اقدام به حذف ۷۰ درصد نشانگرها در جمعیت‌های تأیید نموده (۵۰ درصد دارای ژنوتیپ) و در مرحله بعدی از طریق برنامه‌ی Flmpute اقدام به ایمپیوتیشن و پیش‌بینی نقاط گم‌شده از طریق روابط فامیلی و الگوریتم‌های برپایه جمعیت شد (Sargolzaei et al. 2011) و در نهایت صحت ایمپیوتیشن از طریق همبستگی داده‌های اصلی و ایمپیوتیشن برای نشانگرها مورد ارزیابی قرار خواهد گرفت.

جدول ۱- پارامترهای فرایند شبیه‌سازی

جمعیت اولیه	
فاز اول تعداد نسل (تعداد افراد)	۱۸۰۰(۱۰۰۰)
گلوگاه	بله
فاز دوم تعداد نسل (تعداد افراد)	۱۹۰۰(۵۰۰)
فاز سوم تعداد نسل (تعداد افراد)	۲۰۰۰(۱۰۰۰)
تعداد حیوانات در نسل آخر	۱۰۰۰
جمعیت اخیر	
تعداد نرهای در نسل اخیر	۵۰
تعداد تکثیری جمعیت اخیر بعد از نسل ۲۰۰۰	٦
جمعیت مرجع	افراد نسل ۲۰۰۱ تا ۲۰۰۳
جمعیت تایید	افراد نسل ۲۰۰٤ و ۲۰۰٦
تعداد نتایج به ازای هر زایش	۱
احتمال نر بودن نتاج	۰/۵
انتخاب و طرح آمیزش	تصادفی
نرخ جایگزینی برای نرها	٪۵۰
نرخ جایگزینی برای ماده‌ها	٪۲۰
معیار حذف	سن بالا/ ارزش اصلاحی برآوردی
ژنوم	
تعداد کروموزوم	۲۹
طول هر کروموزوم (سانتی مورگان)	دامنه‌ی بین ٤۲ تا ۱۵۸ سانتی مورگان
تعداد QTL	۱۰، ۱۰۰ یا ۱۰۰۰
اثر آلل‌های QTL	نرمال
تعداد نشانگر	دامنه‌ی ۱۸۰ تا ٦۵۳ با توجه به نوع و اندازه کروموزوم
نرخ جهش در نشانگر و QTLها	$۲/۵ \times ۱۰^{-۵}$
وراثت‌پذیری	۰/۳

در الگوریتم بوستینگ توابع پایه (شامل یادگیرنده‌های ضعیف) مانند درختان رگرسیونی به‌صورت سریالی هر یک روی باقی‌مانده درخت قبلی اضافه می‌شوند در نتیجه اشتباه دسته‌بندی در درخت قبلی باعث کاهش مقدار خطا در درخت بعدی می‌شود در نتیجه با استفاده از یادگیرنده‌های قبلی، یاد گیرنده‌های جدید قوی‌تری ایجاد می‌شود. این الگوریتم تا زمانی ادامه می‌یابد که خطای آخرین درخت به حداقل برسد. مدل کلی الگوریتم بوستینگ به صورت رابطه ۱ می‌باشد:

رابطه (۱)
$$f(x) = \sum_{m=1}^{M} \beta_m \, b(x; \gamma_m)$$

در این جا β_m (m=۱،۲،... ،M) ضرائب توزیع و $b(x;\gamma_m)$ توابع ساده با توزیع چند متغیره x به‌همراه یک سری از پارامترها y می‌باشد. برای توزیع داده‌ها از توزیع نرمال برای حداقل کردن تابع خطا استفاده شد. در این روش پارامترهای اصلی شامل تعداد درخت[1]، عمق درخت[2] و نرخ یادگیری[3] از طریق محاسبه خطای اعتبارسنجی مشخص شدند. در این تحقیق تعداد درخت، عمق درخت و نرخ یادگیری ترتیب ۲۰۰۰، ۵، ۰/۰۲ در نظر گرفته شد. این الگوریتم از طریق بسته‌ی gbm در نرم‌افزار R مورد آنالیز قرار گرفت.

روش SS-GBLUP اولین بار توسط Legarra et al. (2009) and Christensen and Lund (2010) پیشنهاد و بعدها توسط Fernando et al.(2014) گسترش یافت. مدل کلی آن به‌صورت رابطه ۲ است:

رابطه (۲)
$$\begin{bmatrix} y1 \\ y2 \end{bmatrix} = \begin{bmatrix} X1 \\ X2 \end{bmatrix} \beta + \begin{bmatrix} Z1 & 0 \\ 0 & Z2 \end{bmatrix} \begin{bmatrix} g1 \\ g2 \end{bmatrix} + e$$

در اینجا y1 و y2 به‌ترتیب بردار فنوتیپ برای افراد ژنوتیپ‌نشده و ژنوتیپ‌شده، β ماتریس عوامل ثابت، X1 و X2 به‌ترتیب

[1] Ntree
[2] Tree complexity
[3] Learning rate

ماتریس وقوع آثار ثابت برای افراد ژنوتیپ‌نشده و شده، Z1 و Z2، به ترتیب ماتریس وقوع برای افراد ژنوتیپ نشده و شده و $g1=A_{12}A_{22}^{-1}T_2\hat{\alpha}+\hat{\varepsilon}$ و $g2=T_2\hat{\alpha}$ که در اینجا T_2، ماتریس ژنوتیپ مشاهده شده مقیاس دار و متمرکز افراد ژنوتیپ‌شده است. $\hat{\alpha}$ بردار آثار نشانگرهای برآورد شده و A_{ij} زیر ماتریس از ماتریس خویشاوندی شجره‌ای در ارتباط با g1 و g2. مارتریس واریانس کواریانس g1 و g2 (H) به‌صورت رابطه ۳ تعریف می‌شود (Legarra et al. 2009):

رابطه (۳):

$$\mathbf{H}=\begin{bmatrix} A_{11}+A_{12}A_{22}^{-1}(G_2-A_{22})A_{22}^{-1}A_{12} & A_{12}A_{22}^{-1}G_2 \\ G_2A_{22}^{-1}A_{12} & G2 \end{bmatrix}$$

G2 ماتریس خویشاوندی ژنومی برای افراد ژنوتیپ شده‌است. به‌طور کلی آثار نشانگرهای برآورد شده توزیع نرمال ($N(0, I\sigma_\alpha^2)$) و باقی‌مانده (ε) توزیع نرمال چند متغیره مفروض هستند. معادلات مختلط (MME) برای آثار نشانگرهای روش SS-GBLUP به‌صورت رابطه ٤ است:

رابطه (٤)

$$\begin{bmatrix} X'X & X'W & X'Z_1 \\ W'X & W'W+I\frac{\sigma_\varepsilon^2}{\sigma_\alpha^2} & W'Z_1 \\ Z_1'X & Z_1'W & Z_1'Z_1+A^{11}\frac{\sigma_\varepsilon^2}{\sigma_g^2} \end{bmatrix}\begin{bmatrix}\hat{\beta}\\ \hat{\alpha}\\ \hat{\varepsilon}\end{bmatrix}=\begin{bmatrix} X'y \\ W'y \\ Z_1'y_1 \end{bmatrix}$$

در اینجا σ_ε^2، σ_α^2 و σ_g^2 به ترتیب واریانس باقی‌مانده‌ها، نشانگرها و پلی ژنی هستند.

SS-BayesA

در روش بیز A، واریانس نشانگرها متفاوت فرض می‌شود و عموماً واریانس نشانگرها با توزیع کای مربع مقیاس‌دار معکوس (رابطه ۵) در نظر گرفته می‌شود (Meuwissen et al. 2001).

رابطه (۵)

$$p\left(\sigma_{\alpha_j}^2|v_\alpha,s_\alpha^2\right)=\frac{\frac{v_\alpha s_\alpha^2}{2}}{\Gamma(\frac{v_\alpha}{2})}(\sigma_{\alpha_j}^2)^{-(\frac{v_\alpha}{2}+1)}e^{-\frac{v_\alpha s_\alpha^2}{2}}$$

در اینجا v_α و s_α^2 به ترتیب درجه آزادی و توزیع کای مربع مقیاس‌دار معکوس و j: j مین نشانگر است. معادلات مختلط برای روش SS-BayesA به‌صورت رابطه ٦ است:

رابطه (٦)

$$\begin{bmatrix} X'X & X'W & X'Z_1 \\ W'X & W'W+D^{-1}\sigma_\varepsilon^2 & W'Z_1 \\ Z_1'X & Z_1'W & Z_1'Z_1+A^{11}\frac{\sigma_\varepsilon^2}{\sigma_g^2} \end{bmatrix}\begin{bmatrix}\hat{\beta}\\ \hat{\alpha}\\ \hat{\varepsilon}\end{bmatrix}=\begin{bmatrix} X'y \\ W'y \\ Z_1'y_1 \end{bmatrix}$$

در اینجا D ماتریس قطری برای $\sigma_{\alpha_j}^2$ می‌باشد. در نهایت نتایج حاصل از داده‌ها شبیه‌سازی شده با استفاده از روش‌های بوستینگ، SS-GBLUP و SS-BayesA در جمعیت تأیید برآورد شد و صحت پیش‌بینی ژنومی از طریق همبستگی پیرسون بین ارزش‌های اصلاحی پیش‌بینی شده جمعیت ژنوتیپ شده ($\hat{g}_2=T_2\hat{\alpha}$) و ژنوتیپ نشده ($\hat{g}_1=\hat{T}_1\hat{\alpha}+\hat{\varepsilon}$) با ارزش‌های اصلاحی واقعی با استفاده از نرم‌افزار R محاسبه شد.

نتایج و بحث

جداول ۲ و ۳، صحت پیش‌بینی ژنومی ارزش‌های اصلاحی مدل‌های بوستینگ، SS-GBLUP و SS-BayesA را برای جمعیت تأیید ژنوتیپ‌شده با پنل‌های ۱۰k و ایمپیوتیشن (۳k–۱۰k) و متعاقب آن برای حیوانات ژنوتیپ‌نشده طی دو نسل (G1=۲۰۰٤ و G3=۲۰۰٦) را نشان می‌دهد. میانگین صحت ایمپیوتیشن در سناریو مورد مطالعه ۰/۰۱۳ ± ۰/۹۵٦ بود. به‌طور کلی عملکرد مدل‌های مورد مطالعه هنگام به‌کارگیری جمعیت تأیید ژنوتیپ شده با داده‌های ایمپیوتیشن (۳k–۱۰k) در مقایسه با داده‌های ۱۰k کاهش یافت که این میزان کاهش برای روش بوستینگ مشهودتر بود. به‌طوری که کاهش صحت پیش‌بینی ژنومی برای داده‌های ایمپیوتیشن نسبت به داده‌های اصلی برای بوستینگ، SS-GBLUP و SS-BayesA به‌ترتیب ۰/۰۳۹، ۰/۰۲۵ و ۰/۰۲٦ در نسل G1 و ۰/۰٤٦، ۰/۰٤ و ۰/۰۳۸ در نسل G3 بود. در مجموع افت صحت برای افراد ژنوتیپ‌نشده هنگامی که افراد ژنوتیپ شده از پنل ایمپیوتیشن ۳k در مقایسه با پنل اصلی ۱۰k استفاده کردند بیشتر بود. دلیل این امر را می‌توان به عدم خطای ایمپیوتیشن و در اختیار داشتن اطلاعات ژنومی بیشتر در پنل ۱۰k دانست که در نتیجه آن باعث برآورد بهتر ارزش‌های اصلاحی حیوانات ژنوتیپ نشده از روی اطلاعات خویشاوندان آن‌ها شد.

ارزیابی مدل‌های آماری با استفاده از زیرمجموعه‌های نشانگری در داده‌های با و بدون ایمپیوتیشن به این نتیجه رسیده‌اند که توانایی

پیش‌بینی ژنوتیپ هنگام استفاده از ایمپیوتیشن ژنوتیپ بهبود بخشید و بسیاری از محققان این استراتژی را جهت کاهش هزینه‌ها در برنامه‌های انتخاب ژنومی توصیه نموده‌اند (Mulder et al. 2012; Daetwyler et al. 2011). محققین به بررسی اثر ایمپیوتیشن بر صحت پیش‌بینی ژنومی در تراشه ٦ k در مقایسه با تراشه ۵۰k پرداختند و نشان دادند که در سناریوهای مورد مطالعه کاهش ناچیزی و غیر معنی‌داری در صحت ژنومی روش بیزی برای تراشه ٦k نسبت به ۵۰k مشاهده شد (Chen et al. 2014). نتایج در مورد گاوهای جرسی نشان داد که ایمپیوتیشن یک تراشه با تراکم پایین به تراشه ۵۰ k هنگامی که صحت ایمپیوتیشن بالا باشد منجر به بهبود صحت ژنومی خواهد شد. در نتیجه توجه ویژه به ایمپیوتیشن در مطالعات ژنومی بسیار حائز اهمیت خواهد بود (Weigel et al. 2010). هنگامی که خطای ایمپیوتیشن بالا و به تبع آن صحت ایمپیوتیشن پایین باشد در نتیجه استفاده از تراشه‌های با تراکم بالا توصیه شد (Mulder et al. 2012).

در مطالعه اخیر با توجه به صحت بالای ایمپیوتیشن و به تبع آن عدم اختلاف محسوس در گروه‌های ایمپیوتیشن و اصلی برای افراد ژنوتیپ شده خصوصا در روش‌های تک مرحله‌ای، استفاده از ایمپیوتیشن امری اقتصادی و مقرون به صرفه بود. دلیل نتایج فوق‌الذکر را می‌توان به اثر بخشی بالای ایمپیوتیشن بر صحت پیش‌بینی ارزش‌های اصلاحی ژنومی و مدل طراحی شده (Mulder et al. 2012) که رابطه‌ی خطی بین این دو صحت را در پی داشت مرتبط دانست، به‌طوری که افزایش صحت ایمپیوتیشن، افزایش خطی صحت ژنومی را در پی داشت. در مطالعه‌ای Naderi(2008b) به ارزیابی صحت ژنومی با استفاده از روش‌های بیز آستانه‌ای و جنگل تصادفی در سری داده‌های اصلی و ایمپیوتیشن پرداخت شد و نشان داده است که این دو روش هنگام استفاده از داده‌های ایمپیوتیشن برآورد قابل قبولی از صحت ژنومی را نسبت به داده‌های اصلی از خود نشان داده و این استراتژی را توصیه‌ای اقتصادی در ارزیابی ژنومی عنوان کردند. در این تحقیق صحت پیش‌بینی ژنومی در نسل‌های G1 و G3 به‌عنوان جمعیت تأیید و ٤ نسل ما قبل آن به‌عنوان جمعیت مرجع برای سه روش بوستینگ، SS-GBLUP و SS-BayesA مورد بررسی قرار گرفت (جداول ۲ و ۳). نتایج نشان داد که با افزایش فاصله نسلی از جمعیت مرجع، صحت پیش‌بینی ژنومی برای افراد ژنوتیپ‌شده (برای هر سه مدل) و ژنوتیپ‌نشده (برای مدل‌های تک مرحله‌ای) کاهش یافت. با این حال برای افراد ژنوتیپ شده، کاهش صحت ژنومی ناشی از افزایش فاصله و به دنبال آن کاهش رابطه خویشاوندی بین جمعیت مرجع و تأیید برای مدل بوستینگ نسبت به روش‌های تک مرحله‌ای مشهودتر بود.

جدول ۲- صحت پیش‌بینی ژنومی ارزش‌های اصلاحی مدل‌های بوستینگ، SS-GBLUP و SS-BayesA در داده‌های k ۱۰ و ایمپیوتیشن (۳k–۱۰k) برای حیوانات ژنوتیپ‌شده و نشده طی نسل G1=۲۰۰٤

مدل	*تعداد* QTL	صحت ژنومی			
		حیوانات ژنوتیپ‌شده با $10K$	حیوانات ژنوتیپ‌نشده	حیوانات ژنوتیپ‌شده با (*۳k to ۱۰k*)	حیوانات ژنوتیپ‌نشده
SS-GBLUP	۱۰	۰/٤٥٦ ± ۰/۰۱۵	۰/٤۰۲ ± ۰/۰۲۲	۰/٤۲۷ ± ۰/۰۱۷	۰/۳٦۸ ± ۰/۰۲۵
	۱۰۰	۰/٤۷۳ ± ۰/۰۲	۰/٤۲٤ ± ۰/۰۱۹	۰/٤۵٤ ± ۰/۰۲۲	۰/۳۸۳ ± ۰/۰۲۳
	۱۰۰۰	۰/۵۰٤ ± ۰/۰۲	۰/٤٤۷ ± ۰/۰۲٤	۰/٤۷٦ ± ۰/۰۲۱	۰/۳۹۹ ± ۰/۰۲۲
SS-BayesA	۱۰	۰/۸۷۲ ± ۰/۰۱	۰/٦۲۵ ± ۰/۰۱۲	۰/۸٤۲ ± ۰/۰۱۲	۰/۵۸۸ ± ۰/۰۱۵
	۱۰۰	۰/۵۹۵ ± ۰/۰۲۸	۰/٤۸۰ ± ۰/۰۲	۰/۵٦۳ ± ۰/۰۳۱	۰/٤٤۱ ± ۰/۰۲۱
	۱۰۰۰	۰/٤۸٤ ± ۰/۰۲۱	۰/٤۳۱ ± ۰/۰۲	۰/٤٦۹ ± ۰/۰۲٤	۰/۳۸۷ ± ۰/۰۲۲
بوستینگ	۱۰	۰/۳۰۲ ± ۰/۰۳۲	–	۰/۲۷۱ ± ۰/۰۳۳	–
	۱۰۰	۰/۳۷٦ ± ۰/۰۳۱	–	۰/۳۳۵ ± ۰/۰۳۵	–
	۱۰۰۰	۰/٤٤۷ ± ۰/۰۲۸	–	۰/٤۰۱ ± ۰/۰۳	–

جدول ۳- صحت پیش‌بینی ژنومی ارزش‌های اصلاحی مدل‌های بوستینگ، SS-GBLUP و SS-BayesA در داده‌های ۱۰k و ایمپیوتیشن (۳k–۱۰k) برای حیوانات ژنوتیپ‌شده و نشده طی نسل G3=۲۰۰۶

مدل	تعداد QTL	صحت ژنومی			
		حیوانات ژنوتیپ‌شده با $10K$	حیوانات ژنوتیپ‌نشده	حیوانات ژنوتیپ‌شده با (۳k to ۱۰k)	حیوانات ژنوتیپ‌نشده
SS-GBLUP	۱۰	۰/۳۴۲ ± ۰/۰۱۳	۰/۲۷۵ ± ۰/۰۲۱	۰/۳۰۱ ± ۰/۰۱۵	۰/۲۳٤ ± ۰/۰۲۲
	۱۰۰	۰/۳۵۲ ± ۰/۰۲۲	۰/۲۸۰ ± ۰/۰۲	۰/۳۱۳ ± ۰/۰۲۰	۰/۲۳۸ ± ۰/۰۲۰
	۱۰۰۰	۰/۳۷۲ ± ۰/۰۱۷	۰/۲۵۳ ± ۰/۰۲۱	۰/۳۳۲ ± ۰/۰۱۸	۰/۲۱۰ ± ۰/۰۲۰
SS-BayesA	۱۰	۰/۷٤۳ ± ۰/۰۱۸	۰/۵۳٤ ± ۰/۰۱٤	۰/۷۱۱ ± ۰/۰۱۰	۰/۵۰۱ ± ۰/۰۱۹
	۱۰۰	۰/٤۸۵ ± ۰/۰۲٦	۰/۳۸٦ ± ۰/۰۲٦	۰/٤٤۰ ± ۰/۰۲۰	۰/۳٤۳ ± ۰/۰۲۳
	۱۰۰۰	۰/۳۲٤ ± ۰/۰۲۰	۰/۲٦۱ ± ۰/۰۱۷	۰/۲۸۷ ± ۰/۰۱۸	۰/۲۱٦ ± ۰/۰۱۹
بوستینگ	۱۰	۰/۱٦۱ ± ۰/۰۳	–	۰/۱۲۱ ± ۰/۰۲۷	–
	۱۰۰	۰/۲۱۷ ± ۰/۰۳۳	–	۰/۱٦۳ ± ۰/۰۳۲	–
	۱۰۰۰	۰/۲۵۱ ± ۰/۰۲٦	–	۰/۲۰۸ ± ۰/۰۳٤	–

افت صحت پیش‌بینی ژنومی ناشی از افزایش فاصله نسلی برای افراد بدون ژنوتیپ در روش SS-GBLUP بطور محسوسی بیشتر از روش SS-BayesA بود. همچنین اثر منفی افزایش فاصله نسلی و خویشاوندی بر روی صحت ژنومی داده‌های ایمپیوتیشن نسبت به داده‌های اصلی بیشتر بود.

به‌طور کلی صحت پیش‌بینی ژنومی با افزایش فاصله و شکاف بین جمعیت تأیید و مرجع و همچنین کاهش ارتباط خویشاوندی بر اثر شکسته شدن عدم تعادل پیوستگی بین نشانگرها و QTLها در طی نسل کاهش یافت که مطابق با نتایج مطالعات پیشین بود (Kang et al. 2017; Naderi 2019d). مطالعات (Zhou et al. 2018) نشان دادند که در نسل‌های اولیه جمعیت تأیید، افراد بیشتری با جمعیت مرجع در ارتباط ژنتیکی هستند که منجر به برآورد بهتر ارزش‌های اصلاحی و صحت پیش‌بینی ژنومی در نسل‌های اولیه انتخاب می‌شود. مطالعات در مورد گاوهای هلشتاین آلمان نشان داد که صحت ژنومی برای صفات مختلف اقتصادی با کاهش ارتباط خویشاوندی بین جمعیت مرجع و تأیید کاهش می‌یابد. تحقیقات شبیه‌سازی (Kang et al. 2017) نشان از افت صحت ژنومی در طول نسل برای مدل‌های تک مرحله‌ای و GBLUP دارد. مطالعات جهت آنالیز و پیش‌بینی ژنومی جمعیت‌های مختلف گوسفند نشان داد که مقدار بالای صحت پیش‌بینی ژنومی به‌علت وجود ساختار جمعیت و ارتباط خویشاوندی فامیلی بین جمعیت مرجع و تأیید است. به‌طور کلی به‌دلیل عدم استفاده هم‌زمان از اطلاعات ژنومی و شجره‌ای در ارزیابی روش بوستینگ در مقایسه با روش‌های تک مرحله‌ای، افت بیشتری در عملکرد آن مشاهده شد. جهت پیش‌بینی ژنوتیپ افراد ژنوتیپ نشده از روابط خویشاوندی برپایه شجره استفاده می‌شود. در نتیجه هنگامی که ارتباط خویشاوندی نزدیکتری بین افراد جمعیت مرجع برقرار باشد از طریق روابط خویشاوندی شجره‌ای پیش‌بینی ارزش اصلاحی برای افراد بدون ژنوتیپ امکان‌پذیرتر و از صحت بالاتری برخوردار خواهد بود.

همان‌گونه که در تحقیق حاضر نشان داده شد با افزایش فاصله نسلی و کاهش ارتباط خویشاوندی بین جمعیت مرجع و تأیید از طریق افزایش نسل، عملکرد افراد ژنوتیپ‌نشده در مقایسه با افراد ژنوتیپ شده از افت بیشتری برخوردار بوده که مطابق به نتایج مطالعات پیشین بود (Zhou et al. 2018; Chen et al. 2014). دلیل این امر این است که صحت پیش‌بینی ژنومی افراد ژنوتیپ‌نشده تحت تاثیر دو مسیر، شامل: اثر نشانگرهای برآورد شده و برآورد باقی‌مانده‌های تخمین زده شده‌است که این باقی مانده‌ها به نوع خود وابسته به ماتریس روابط خویشاوندی هستند که در ارزیابی افراد ژنوتیپ‌شده مؤثر نیستند. در نتیجه با افزایش فاصله نسلی، ماتریس روابط خویشاوندی ضعیف‌تر و قدرت مدل برای برآورد ارزش‌های اصلاحی افراد ژنوتیپ‌نشده کمتر می‌شود.

اثر مدل آماری بر صحت پیش‌بینی ژنومی

نتایج نشان داد که برای سناریوهای با تعداد کم (QTL ۱۰) و متوسط (QTL ۱۰۰) همواره عملکرد روش SS-BayesA بیشتر از SS-GBLUP و بوستینگ بود (جدول ۲ و ۳). ولی هنگامی‌که

تعداد QTL بالا بود (۱۰۰۰ QTL) عملکرد روش SS-GBLUP نسبت به SS-BayesA و بوستینگ در هر دو نسل بیشتر بود. به‌طور کلی عملکرد روش بوستینگ نسبت به روش‌های تک مرحله‌ای پایین‌تر و با افزایش نسل این میزان افت مشهودتر یود. با این حال روش بوستینگ در تعداد بالایQTL عملکرد قابل قبول و نزدیک به روش‌های تک مرحله‌ای نشان داد. یافته‌های تحقیق حاضر نشان از برتری روش SS-BayesA نسبت به روش SS-GBLUP و بوستینگ در صفات تحت تاثیر تعداد پایین QTL برای هر دو سری اصلی و ایمپیوتیشن دارد. به‌طورکلی روش SS-BayesA نسبت به روش SS-GBLUP برای افراد ژنوتیپ شده) در مقایسه با ژنوتیپ نشده از برتری محسوس‌تری برخوردار بود. نتایج تحقیق حاضر مطابق با تحقیقات پیشین در زمینه شبیه‌سازی و داده‌های واقعی بود (Habier et al. 2007; Hayes et al. 2009). برتری عملکرد روش SS-BayesA نسبت به GBLUP به‌دلیل پیش فرض توزیع غیر نرمال اثر نشانگرها (توزیع t) بوده که این پیش فرض برای اثر نشانگرها سازگاری بیشتری با ساختار QTL و نشانگرها در مطالعه حاضر بویژه هنگام به‌کار گیری سناریوهای با تعداد پایین QTL دارد. دلیل برتری روش‌های تک مرحله‌ای نسبت به روش بوستینگ را می‌توان بکارگیری هم‌زمان منابع اطلاعات ژنومی و شجره‌ای جهت برآورد ارزش‌های اصلاحی حیوانات دانست. در این زمینه مطالعات مختلفی از برتری SS-GBLUP نسبت به مدل‌های چند مرحله‌ای سخن به میان آوردند که هم‌راستا با مطالعه اخیر می‌باشند (Onogi et al. 2014; Mohammadi et al. 2017). هنگامی که تعداد QTL افزایش یافت (۵۰۰ QTL) عملکرد روش SS-BayesA به‌دلیل توزیع نرمال پراکندگی QTL و عدم تطابق با فرضیات آن کاهش یافت. در نتیجه می‌توان عنوان کرد که سودمند بودن روش‌های SS-GBLUP و SS-BayesA وابسته به تعداد QTL کنترل‌کننده صفت است (Zhou et al. 2018; Kang et al. 2017)

نتایج اثر سطوح مختلف QTL بر صحت پیش‌بینی ژنومی نشان داد که روش SS-BayesA حساسیت بالایی به تغییرات QTL خصوصا برای افراد ژنوتیپ شده (در مقایسه با افراد ژنوتیپ نشده) در هر دو سری داده اصلی و ایمپیوتیشن دارد. مطالعات شبیه‌سازی نشان داد که روش بیز نسبت به روش‌های یادگیری ماشین و GBLUP حساسیت بالایی به تغییرات QTL نشان داده و با افزایش تعداد QTL صحت ژنومی کاهش یافت که دلیل این امر را می‌توان به توزیع محدود واریانس ژنتیکی بر تعداد زیادی QTL دانست که در نتیجه سهم هر QTL در ارزش ژنتیکی کل کاهش یافته، و قدرت مدل بیز (با توجه به پیش فرض‌های اولیه) در پیش‌بینی ارزش‌های اصلاحی کاهش می‌یابد. همچنین روش تک مرحله‌ای بیز 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 بیز A بهتر عمل کرد (Naderi 2018b).

برای روش بوستینگ، صحت پیش‌بینی ژنومی با افزایش تعداد QTL افزایش نسبتا محسوسی داشت. در این راستا، مطالعات نشان دادند هنگامی که تعداد QTL افزایش می‌یابد به تعداد بیشتری نشانگر جهت به دام انداختن اثر همه QTLها احتیاج خواهد بود. در نتیجه افزایش تعداد QTL می‌تواند منجر به افزایش صحت ژنومی شود هنگامی که به‌طور موازی با افزایش تعداد QTLها، تعداد نشانگرها نیز افزایش یابد (Habier et al. 2009). مطالعات در زمینه مدل‌های باز نمونه‌گیری و یادگیری ماشین نشان داد، افزایش تعداد QTL، 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 به‌همراه خواهد داشت (Naderi 2018a; Naderi et al. 2016)، که این اثر مثبت برای روش بوستینگ در نتایج تحقیق حاضر صادق بود. همچنین روش SS-GBLUP حساسیت کمتری نسبت به تغییرات QTL داشت که نشان می‌دهد این مدل یک روش کم نوسان جهت استفاده در سناریوهای با تعداد متفاوت QTL است. دلیل این امر را می‌توان به ماهیت GBLUP در استفاده از روابط خویشاوندی ژنومی میان جمعیت مرجع و تأیید به جای عدم تعادل پیوستگی بین QTL و نشانگرها عنوان کرد. بنابراین هنگامی که تعداد QTL افزایش یا کاهش پیدا می‌کند روابط خویشاوندی ژنومی تغییر نکرده و در نتیجه صحت پیش‌بینی ژنومی نوسانات کمی از خود نشان می‌دهد (Zhuo et al. 2018).

نتیجه‌گیری

به‌طور کلی پیش‌بینی ژنومی حاصل از داده‌های ایمپیوتیشن در ارزیابی افراد تأیید (دارای ژنوتیپ) راه کار مناسبی جهت کاهش هزینه‌های ژنوتایپینگ است. صحت پیش‌بینی ژنومی تحت سناریوهای مختلف ژنومی همواره برای مدل‌های SS-GBLUP و SS-BayesA نسبت به بوستینگ بیشتر بود. بر خلاف روش بوستینگ روش‌های تک مرحله‌ای با استفاده از روابط خویشاوندی ژنومی و شجره‌ای قادر به ارزیابی صحت پیش‌بینی ژنومی برای افراد ژنوتیپ‌نشده بودند. با این حال همواره صحت پیش‌بینی ژنومی برای افراد ژنوتیپ شده بیشتر از افراد ژنوتیپ نشده بود. وجود روابط خویشاوندی بین جمعیت مرجع و تأیید از مهم‌ترین عوامل تاثیر گذار بر صحت ژنومی بود به‌طوری که با افزایش فاصله نسلی (کاهش رابطه خویشاوندی) صحت ژنومی کاهش محسوسی خصوصا برای روش بوستینگ به‌علت عدم استفاده هم‌زمان از اطلاعات ژنومی و شجره‌ای داشت. تعداد QTL از عوامل تاثیرگذار بر صحت پیش‌بینی ژنومی بود. به‌طوری که برای صفات تحت تاثیر تعداد کم و متوسط QTL روش SS-BayesA و برای صفات تحت تاثیر تعداد زیاد QTL روش SS-GBLUP بهترین عملکرد را داشت.

منابع

Aguilar I, Misztal I, Johnson D, Legarra A, Tsuruta S, Lawlor T (2010) Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science 93:743-752.

Barazandeh A, Mohammadabadi M, Ghaderi-Zefrehei M and Nezamabadi-Pour H (2016) Genome-wide analysis of CpG islands in some livestock genomes and their relationship with genomic features. Czech Journal of Animal Science 61: 487-495.

Barazandeh A, Mohammadabadi M, Ghaderi-Zefrehei M and Nezamabadipour H (2016) Predicting CpG islands and their relationship with genomic feature in cattle by hidden markov model algorithm. Iranian Journal of Applied Animal Science 6: 571-579.

Breiman L (2001) Random forests. Machine Learning 45:5-32.

Chen C, Misztal I, Aguilar I, Tsuruta S, Meuwissen T, Aggrey S, Wing T, Muir W (2011) Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. Journal of Animal Science 89:23-28.

Chen L, Li C, Sargolzaei M, Schenkel F (2014) Impact of genotype imputation on the performance of GBLUP and Bayesian methods for genomic prediction. PLoS One 9:e101544.

Christensen OF, Lund MS (2010) Genomic prediction when some animals are not genotyped. Genetics Selection Evolution 42:2.

Daetwyler HD, Wiggans GR, Hayes BJ, Woolliams JA, Goddard ME (2011) Imputation of missing genotypes from sparse to high density using long-range phasing. Genetics 189:317-327.

Fernando RL, Dekkers JC, Garrick DJ (2014) A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genetics Selection Evolution 46:50.

Ghafouri-Kesbi F, Rahimi-Mianji G, Honarvar M, Nejati-Javaremi A (2017) Predictive ability of random forests, boosting, support vector machines and GBLUP in different scenarios of genomic evaluation. Animal Production Science 57:229-236.

Goddard M, Hayes B (2007) Genomic selection. Journal of Animal Breeding and Genetics 124:323-330.

González-Recio O, Forni S (2011) Genome-wide prediction of discrete traits using Bayesian regressions and machine learning. Genetics Selection Evolution 43:7.

Habier D, Fernando RL, Dekkers JC (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389-2397.

Habier D, Fernando RL, Dekkers JC (2009) Genomic selection using low-density marker panels. Genetics 182:343-353.

Hadizadeh M, Mohammad AM, Niazi A, Esmailizadeh KA, Mehdizadeh GY and Molaei MS (2013) Use of bioinformatics tools to study exon 2 of GDF9 gene in Tali and Beetal goats. Modern Genetics Journal 8 283-288 (In Farsi).

Hadizadeh M, Niazi A, Mohammad AM, Esmailizadeh A and Mehdizadeh GY (2014) Bioinformatics analysis of the BMP15 exon 2 in Tali and Beetal goats. Modern Genetics Journal 9: 117-120 (In Farsi).

Hayes B (2007) QTL mapping, MAS, and genomic selection. A short-course. Animal Breeding & Genetics Department of Animal Science. Iowa State University 1:3-4.

Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. Genetics Research 91:47-60.

Jafari Darehdor A, Mohammadabadi M, Esmailizadeh A and Riahi Madvar A (2016) Investigating expression of CIB4 gene in different tissues of Kermani Sheep using

Real Time qPCR. Journal of Ruminant Research 4: 119-132 (In Farsi).
Kang H, Zhou L, Mrode R, Zhang Q, Liu J (2017) Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity 119:459.
Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. Journal of Dairy Science 92:4656-4663.
Liu Z, Goddard M, Reinhardt F, Reents R (2014) A single-step genomic model with direct estimation of marker effects. Journal of Dairy Science 97:5833-5850.
Meuwissen T, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829.
Mohammadabadi M and Tohidinejad F (2017) Characteristics Determination of Rheb Gene and Protein in Raini Cashmere Goat. Iranian Journal of Applied Animal Science 7: 289-295.
Mohammadabadi M, Jafari A and Bordbar F (2017) Molecular analysis of CIB4 gene and protein in Kermani sheep. Brazilian Journal of Medical and Biological Research 50: e6177.
Mohammadi Y, Sattaei Mokhtari M, Razmkabir M, Abdollahi-Arpanahi R (2017) Single Step GBLUP (SS-GBLUP): case study using beef cattle genomic data. Modern Genetics Journal 12:643-647. (In Farsi).
Mulder H, Calus M, Druet T, Schrooten C (2012) Imputation of genotypes with low-density chips and its effect on reliability of direct genomic values in Dutch Holstein cattle. Journal of Dairy Science 95:876-889.
Naderi S, Yin T, 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 99:7261-7273.
Naderi Y (2018a) 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 9:129-138. (In Farsi).
Naderi Y (2018b) 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 49:145-157. (In Farsi).
Naderi Y (2018c) Investigation of genotype× environment interaction with considering imputation in simulated genomic data via different animal models. Animal Production 20:375-387. (In Farsi).
Naderi Y (2019d) The importance of genetic relationships and phenotypic record on genomic accuracy of simulated imputation data via animal models in presence of genotype × environment interactions. Research on Animal Production 9:119-130. (In Farsi).
Nejati-Javaremi A, Smith C and Gibson J (1997) Effect of total allelic relationship on accuracy of evaluation and response to selection. Journal of Animal Science 75: 1738-1745.
Onogi A, Ogino A, Komatsu T, Shoji N, Simizu K, Kurogi K, Yasumori T, Togashi K, Iwata H (2014) Genomic prediction in Japanese Black cattle: application of a single-step approach to beef cattle. Journal of Animal Science 92:1931-1938.
Sargolzaei M, Chesnais J, Schenkel F (2011) FImpute-An efficient imputation algorithm for dairy cattle populations. Journal of Dairy Science 94:421.
Sargolzaei M, Schenkel FS (2009) QMSim: a large-scale genome simulator for livestock. Bioinformatics 25:680-681.
Tohidi NF, Mohammadabadi M, Esmailizadeh A and Najmi NA (2015) Comparison of different levels of Rheb gene expression in different tissues of Raini cashmir goat. Journal of Agricultural Biotechnology 6: 37-49. (In Farsi).
Weigel K, de Los Campos G, Vosa G, Gianola D, Van Tassell C (2010) Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. Journal of Dairy Science 93:5423-5435.
Yang P, Hwa Yang Y, B Zhou B, Y Zomaya A (2010) A review of ensemble methods in bioinformatics. Current Bioinformatics 5:296-308.
Yin T, Pimentel E, Borstel UKv, 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 97:2444-2454.
Zhou L, Mrode R, Zhang S, Zhang Q, Li B, Liu J-F (2018) Factors affecting GEBV accuracy with single-step Bayesian models. Heredity 120:100.