



داده‌کاوی برخی سازه‌های مؤثر بر سخت‌زایی در گاوهای شیری ایران

مریم منتظری نجف‌آبادی^۱، مصطفی قادری زفره‌ای^{۲*}

۱- دانش‌آموخته کارشناسی ارشد، گروه علوم دامی، دانشکده کشاورزی، دانشگاه یاسوج

۲- دانشیار، گروه علوم دامی، دانشکده کشاورزی، دانشگاه یاسوج

(تاریخ دریافت: ۹۹/۰۱/۲۷ - تاریخ پذیرش: ۹۹/۰۴/۰۳)

چکیده

داده‌کاوی، به کشف روابط و الگوهای پنهان و ناپیدا در میان حجم عظیمی از داده‌ها می‌پردازد که بدون استفاده از آن، ممکن است این روابط هیچ‌گاه آشکار نشوند. تعیین سازه‌های مؤثر بر سخت‌زایی در گاوهای شیری می‌تواند به ضعف مدیریتی صنعت گاو شیری و مشکلات ناشی از آن کمک کند. بنابراین ضروری است که از بین سازه‌های ژنتیکی و غیر ژنتیکی مختلف مؤثر بر سخت‌زایی، مهم‌ترین آن‌ها شناسایی و مدیریت شوند. الگوریتم انتخاب ویژگی یکی از روش‌های داده‌کاوی است که می‌تواند در این زمینه مفید باشد. هدف از این پژوهش تعیین مهم‌ترین سازه‌های مؤثر بر سخت‌زایی گاوهای شیری هلشتاین با استفاده از الگوریتم انتخاب ویژگی بود. کل رکوردها برابر با ۴۱۳۲۰۵ به همراه ۱۴ ویژگی مربوط به رکوردهای تولیدمثلی و تولیدی بود. داده با چهار روش مهم الگوریتم انتخاب ویژگی (Ranker و Greedy-Stepwise Genetic-search, Best-First) و نه مدل مختلف (Gain-Ratio-Attribute-Eval, CFS-Subset-Eval و ...) برای تعیین مهم‌ترین عوامل مؤثر بر سخت‌زایی ارزیابی شدند. نتایج نشان داد روش Genetic-search الگوریتم انتخاب ویژگی، مناسب‌ترین روش برای انتخاب عوامل مؤثر بر سخت‌زایی با استفاده از دسته‌بندی Naive Bayes Tree با کم‌ترین میزان خطا (AMS= 0.011, ARAE=0.001) بود. مهم‌ترین سازه‌های مؤثر بر سخت‌زایی به ترتیب تولید شیر، سن مادر هنگام زایش، شکم زایش، طول دوره آبستنی، شیر تصحیح شده بر اساس چهار درصد چربی، فاصله زایش، فصل زایش، نوع زایش، نسبت چربی به پروتئین، جنس گوساله، وزن گوساله و روزهای باز بودند. انتظار می‌رود با در نظر گرفتن سازه‌های مزبور، سخت‌زایی در گاوهای شیری بهتر مدیریت شود.

واژه‌های کلیدی: الگوریتم انتخاب ویژگی، تولیدمثلی، سخت‌زایی، گاو شیری

* نویسنده مسئول: mghaderi@yu.ac.ir

مقدمه

مدل‌های بیزی حد آستانه‌ای انجام شد، وارث‌پذیری صفت سخت‌زایی روی رکوردهای تکرار شده گاوهای هلشتاین در تجزیه تک‌صفتی و چندصفتی به میزان پایین برآورد شد. از این رو، جهت کاهش مشکلات حاصل از سخت‌زایی، ضروری است که مهم‌ترین عوامل مؤثر بر صفت سخت‌زایی شناسایی و اندازه‌گیری شوند. برای تعیین مهم‌ترین عوامل مؤثر بر یک مسأله خاص، راه حل‌ها و الگوریتم‌های فراوانی می‌توانند وجود داشته باشند، که مشکل بعضی از این الگوریتم‌ها بار محاسباتی زیاد آن‌ها است. اگرچه امروزه با ظهور رایانه‌های نسل جدید با قدرت پردازش سریع و منابع ذخیره‌سازی بزرگ، این مشکل تا حدودی مرتفع شده است، اما وجود مجموعه داده بسیار بزرگ برای مسائل جدید که گاه اندازه‌گیری آن‌ها ناممکن و یا حتی زمان‌بر و پرهزینه است سبب شده است که یافتن یک الگوریتم دقیق و سریعی که بتواند با معرفی تعداد محدودی از عوامل اثرگذار به عنوان خروجی بیش‌ترین تغییرات مربوط به یک متغیر خاص را توجیه نماید، مهم باشد. هدف از این پژوهش تعیین مهم‌ترین سازه‌های مؤثر بر سخت‌زایی گاوهای شیری ایران با استفاده از الگوریتم انتخاب ویژگی بود.

مواد و روش‌ها

داده‌ها از تعاونی وحدت گاوداران اصفهان به‌دست آمد (۷۴ گله) که بین سال‌های ۱۳۸۶ تا ۱۳۹۶ جمع‌آوری شده بود. کل رکوردها برابر با ۴۱۳۲۰۵ به همراه ۱۴ رکوردهای تولیدمثلی و تولیدی بود که بعد از ویرایش، ۱۱۹۰۰۰ رکورد برای تحلیل نهایی باقی ماند. داده یاد شده با سه روش مهم الگوریتم انتخاب ویژگی (روش‌های Best-First، Greedy-Stepwise Genetic-search و Ranker) در نرم افزار WEKA (نسخه ۳/۸) واکاوی شد تا مهم‌ترین عوامل مؤثر بر سخت‌زایی شناسایی شوند. خلاصه آماری صفات مورد بررسی در جدول ۱ نشان داده شده است.

امروز پرورش گاو شیری به یک صنعت پیچیده و چند لایه تبدیل شده است که انتظار می‌رود بیش‌تیمی سود در مدیریت صحیح و علمی سازه‌های مؤثر در سطوح مختلف این صنعت حاصل شود. تولیدمثل از جمله سازه‌های کلیدی در تعیین بازده و سودآوری در صنعت پرورش گاوهای شیره محسوب می‌شود (Olson *et al.*, 2009; Ghavi Hossein-Zadeh, 2014). هدف اصلی در مدیریت تولیدمثل گاوهای شیری، بیش‌تیمی نمودن تعداد زایمان‌ها در طول عمر گاوها و افزایش سود حاصل از فعالیت گاو‌داری است. سخت‌زایی یکی از مهم‌ترین صفات تولیدمثلی در گاوهای شیری است، که با استفاده از مدیریت این صفت می‌توان کارآیی پرورش را از راه کاهش هزینه‌ها بهبود داد (Dematawewa and Berger, 1998). سازه‌های مؤثر متعدد ژنتیکی و محیطی می‌توانند این صفت را تحت تأثیر قرار بدهند و لذا با مدیریت هوشمندانه این سازه‌ها، می‌توان مشکلات ناشی از سخت‌زایی را کمتر و ظرفیت سودآوری صنعت پرورش گاوهای شیری را افزایش داد. وقوع سخت‌زایی، هزینه‌های دامپزشکی و نیروی کار، خطر حذف غیر اختیاری و مرگ و میر گوساله و مادر را افزایش می‌دهد و از طرف دیگر با کاهش تولید شیر در دوره شیردهی بعدی، بر عملکرد تولیدمثلی پس از زایش نیز اثر نامطلوب می‌گذارد (De Matuana *et al.*, 2009). پدیده سخت‌زایی در گاو نسبت به دیگر گونه‌های حیوانات مزرعه، بیش‌تر بروز می‌کند و میزان وقوع آن در تلیسه‌ها سه برابر بیش‌تر از گاوهای مسن‌تر است (Meyer *et al.*, 2001; Purohit, 2006). عملکرد سخت‌زایی حاصل از آثار متقابل تغذیه و ژنتیک است و عمده‌ترین علت آن مربوط به نامتناسب بودن اندازه محوطه لگنی مادر با اندازه و وزن گوساله است (Berry *et al.*, 2005; Zaborski *et al.*, 2014). وراثت‌پذیری صفت سخت‌زایی در گاو شیری پایین است. در پژوهش مرتضوی و همکاران (۱۳۹۷) که با استفاده از

جدول ۱- آمار توصیفی متغیرها یا ویژگی‌های مورد استفاده در این پژوهش

Table 1. Descriptive statistics of the variables/features used in this research

| Feature/Variable | Mean | Standard Deviation | Range |
|---|----------|--------------------|--------------------|
| Pregnancy length , day | 277.69 | 5.08 | 223 – 303 |
| Calving age ,day | 1502.64 | 458.31 | 1096 – 3141 |
| Open days , day | 197.93 | 64.24 | 42 – 713 |
| Calving interval ,day | 487.14 | 76.78 | 387 – 742 |
| FCM 4%,kg | 12414.39 | 3002.61 | 2909.88 – 21592.95 |
| Combined fat and protein content in milk, %, % | 1.06 | 0.16 | 0.41 – 1.83 |
| Calving weight, Kg | 42.29 | 7.29 | 15– 70 |
| Lactation peak milk yield, kg | 53.35591 | 8.86 | 13 – 70 |

Qualitative

Year of birth (Number = 24), season of birth (Number = 4), month of birth (Number = 12), herd size (Number = 11-99), dam parity (Number = 6), sex (Number = 2), birth type (Number = 2)

متغیرهای دسته‌ای استفاده شود، درخت تشکیل شده را درخت دسته‌بندی (CT) گویند.

الگوریتم درخت تصمیم برجسته (J48): سعی می‌کند تا به صورت بازگشتی، مجموعه داده را با استفاده از اطلاعات نرمال به زیر مجموعه‌هایی تقسیم کند که از ریشه به سمت برگ حرکت کرده و در نهایت به دسته موردنظر می‌رسد. این الگوریتم در نهایت یک دسته‌بندی را انجام داده که دارای دو گره است. یک گره به صورت برگ که یک دسته را مشخص می‌کند و یک گره تصمیم که آزمون‌هایی روی یک صفت انجام می‌دهد تا یک شاخه به ازای هر خروجی آزمون تولید کند (Qin and Lawry, 2005).

الگوریتم درخت بیز ساده (NB): این الگوریتم برای دسته‌بندی داده از احتمال وقوع یا عدم وقوع پدیده‌ها استفاده می‌کند (Hosseinalizadeh et al., 2019). در واقع، بر اساس ویژگی‌های ذاتی احتمال، با آموزش و ارزیابی داده، در نهایت داده را به زیر مجموعه‌هایی طبقه‌بندی می‌کند (Kohavi, 1996).

الگوریتم درخت جنگل تصادفی (RF): یک الگوریتم گروهی با مجموعه‌ای از درختان تصمیم است. یک ویژگی مهم در ساخت جنگل‌های تصادفی، انتخاب تصادفی در هر گره است (Breiman, 1996).

روش‌های مختلف انتخاب ویژگی تلاش می‌کنند تا زیر مجموعه‌ای با کمینه تعداد ویژگی (سازه‌های اثر گذار) ممکن برای پیش‌بینی هدف مورد نظر پیدا کنند (Hall et al., 2015). در تمام این روش‌ها بر اساس کاربرد و نوع تعریف، زیر مجموعه‌ای به عنوان جواب انتخاب می‌شود که بتواند مقدار یک تابع ارزیابی را بهینه کند. شکل ۱ روال کار پژوهش حاضر را نشان می‌دهد.

فرآیند انتخاب ویژگی: این فرآیند در برگیرنده تابع تولید کننده، تابع ارزیابی، تابع خاتمه و تابع اعتبارسنجی است. تابع تولید کننده، زیر مجموعه‌های کاندید را برای روش مورد نظر پیدا می‌کند. تابع ارزیابی، زیر مجموعه مورد نظر را بر اساس روش داده شده، ارزیابی و یک عدد به عنوان میزان خوبی روش باز می‌گرداند. روش‌های مختلف سعی در یافتن زیر مجموعه‌ای دارند که این مقدار را بهینه کند. شرط خاتمه نیز برای تصمیم‌گیری در مورد زمان توقف الگوریتم به کار می‌رود. در ادامه به کمک تابع اعتبار مشخص می‌شود که آیا زیر مجموعه انتخاب شده معتبر است یا خیر (کارگر و مکرم، ۱۳۹۵). گامه تعیین اعتبار به کمک این تابع در شکل ۲ نشان داده شده است.

درخت‌های تصمیم (DT) روشی برای نمایش یک سری از قوانینی هستند که منتهی به یک رده یا مقدار می‌شوند. با توجه به نوع متغیر هدف، نوع درخت تصمیم نیز متفاوت می‌شود. اگر درخت تصمیم برای پیش‌بینی

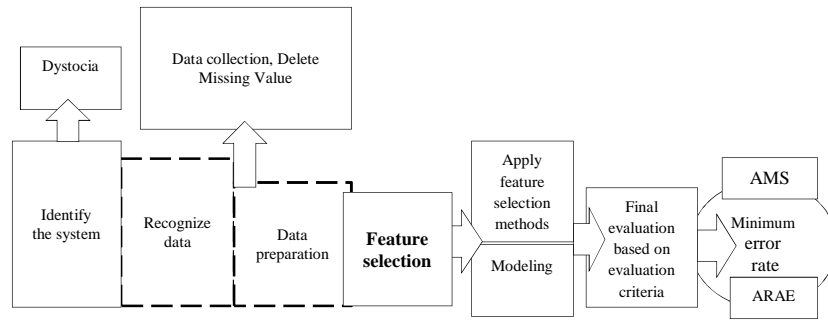


Fig. 1. Procedures and steps undertaken in this study

شکل ۱- روال و مراحل کار انجام شده در این پژوهش

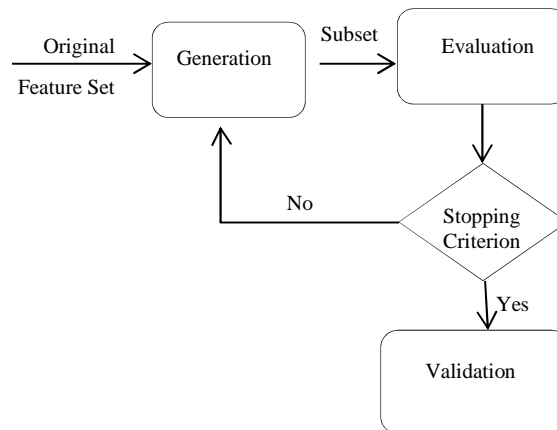


Fig. 2. Feature selection algorithm steps

شکل ۲- گام‌های الگوریتم انتخاب ویژگی

تعیین مؤثرترین عوامل اثر گذار بر صفت سخت‌زایی از دو روش مذکور برای تعیین میزان خطای هر یک از روش‌ها استفاده شده است. در روش AMS، به منظور بررسی دقت مدل از رابطه زیر استفاده می‌شود:

$$AMS = \frac{\sum_{i=1}^n MS_i}{N}$$

که MS_i تعداد تعداد دسته‌های نادرست پیش‌بینی شده به وسیله مدل و N تعداد کل داده را نشان می‌دهد. این روش، نسبت بین میانگین تعداد دسته‌های نادرست پیش‌بینی شده به وسیله مدل به تعداد کل داده است (Dash and Liu, 2003). در روش ARAE از خطای مطلق نسبی به کل داده استفاده شده است (Dash and Liu, 2003).

$$ARAE = \frac{\sum_{i=1}^n RAE_i}{N}$$

انتخاب ویژگی در این پژوهش در دو مرحله انجام شد. در مرحله اول، روش‌های مختلف انتخاب ویژگی برای داده اعمال شد و سپس در مرحله دوم، به منظور تولید درخت تصمیم با یک دسته‌بندی ساده جهت به‌دست آوردن مناسب‌ترین و بهترین روش انتخاب عوامل مؤثر بر صفت سخت‌زایی، از آماره‌های ضریب همبستگی (R)، میانگین خطای مطلق (MAE)، جذر میانگین مربعات خطا (RMSE)، خطای مطلق نسبی (RAE)، جذر مربعات خطای نسبی (RRSE)، تعداد دسته‌های پیش‌بینی شده درست و نادرست (true/false) استفاده شده است. روش‌های مختلفی به منظور بررسی روش‌های انتخاب ویژگی و انتخاب بهترین روش وجود دارد که از معروف‌ترین این روش‌ها می‌توان به AMS و ARAE اشاره کرد که در این پژوهش استفاده شده است (Naseriparsa et al., 2014). بنابراین بعد از اعمال مرحله دوم، در نهایت به منظور اطمینان از انتخاب مناسب‌ترین روش برای

آموزش کل (Full training set) در جدول‌های ۳ تا ۵ ارائه شده است. نتایج نشان داد در دسته‌بندی با استفاده از الگوریتم‌های J48 و RF، روش Best-First و برای الگوریتم NB، روش Genetic Search با بیشترین ضریب همبستگی (بالای ۹۸ درصد) و با داشتن کمترین تعداد دسته‌های پیش‌بینی شده غلط به عنوان مناسب‌ترین روش‌ها برای تعیین مهم‌ترین عوامل اثرگذار بر صفت سخت‌زایی انتخاب شده است. تعداد دسته‌های پیش‌بینی شده غلط در الگوریتم RF در مقایسه با الگوریتم J48 بسیار کمتر است (۲۷۱ در مقابل ۳۱۵). از طرفی برای الگوریتم NB، تعداد دسته‌های پیش‌بینی شده غلط از دو الگوریتم دیگر کمتر است (۲۴۱ عدد).

که RAE_i خطای مطلق نسبی و N تعداد کل داده است.

نتایج و بحث

مؤثرترین عوامل اثرگذار بر صفت سخت‌زایی که با روش‌های مختلف الگوریتم انتخاب ویژگی (در مرحله اول) به دست آمده به تفکیک روش و مدل در جدول ۲ ارائه شده است. آثار گله و سال و ماه زایش گاوها در جدول ۲ در نظر گرفته نشدند زیرا تحلیل ابتدایی نشان داد این‌ها ارتباط بسیار ضعیفی با سخت‌زایی دارند و در مدل برازش نشدند.

در مرحله دوم از انتخاب ویژگی، از الگوریتم‌های J48 (جدول ۳)، NB (جدول ۴) و RF (جدول ۵) استفاده شد که نتایج آن به تفکیک الگوریتم به کار رفته و با در نظر گرفتن حالت آزمون 10 Fold و مدل دسته‌بندی مجموعه

جدول ۲- مهم‌ترین سازه‌های اثرگذار بر صفت سخت‌زایی به دست آمده با روش‌های مختلف انتخاب ویژگی

Table 2. The most important factors affecting dystocia obtained by different feature selection methods

| Method | Model | Factor |
|-----------------|-----------------------------------|--|
| Greedy-Stepwise | CFS-Subset-Eval | Age, FCM4%, F/P, CW |
| Best-First | CFS-Subset-Eval | Age, OD, CI, CW, Peak |
| Genetic Search | Consistency-Subset-Eval | CI, Peak, age, MP, PL, fasl, KP, FCM4%, F/P, CS, CW, OD |
| Ranker | Chi-Squared-Attribute-Eval | Age, FCM4%, OD, sal, CI, PL, MP, fasl, KP, CS, CW, F/P, Peak |
| | Gain-Ratio-Attribute-Eval | FCM4%, OD, Peak, F/P, CW, KP, PL, CI, sal |
| | Info-Gain-Attribute-Eval | FCM4%, OD, F/P, CW, Peak, fasl, KP, MP, PL, CI |
| | OneR-Attribute-Eval | OD, FCM4%, F/P, Peak, CW, fasl, age, CI |
| | Symmetrical-Uncert-Attribute-Eval | Peak, FCM4%, OD, F/P, CW, sal, KP, CI, age, MP, PL, fasl, CS |
| | Principal-Components | Age, Peak, CI, MP, fasl, F/P, PL |
| | Relief-Attribute-Eval | CW, Peak, FCM4%, MP, CI, Age, fasl, KP |

Age: Cow age at calving (months), Peak: Daily milk yield at the peak of lactation (kg), CI: Calving interval (days), MP: Milk period (lactation number), sal: Year of calving, fasl: Season of calving, F/P: The combined fat and protein content in milk (%), PL: Pregnancy length (days), OD: Open days, FCM: 4% fat-corrected milk yield (kg) and calculated according to the formula: $FCM=0.4 \cdot ML+15 \cdot FT$, where ML is real milk yield in lactation (kg) and FT is fat yield (kg), CS: Calf sex at birth, KP: Type of parturition, CW: Calf weight at birth

جدول ۳- ارزیابی دسته‌ها با استفاده از الگوریتم J48

Table 3. Assessment of classes using J48 algorithm

| Method | Model | R | MAE | RMSE | RAE | RRSE | Number of predicted categories | |
|-----------------|-----------------------------------|-------|-------|-------|-------|-------|--------------------------------|-------|
| | | | | | | | True | False |
| Greedy-Stepwise | CFS-Subset-Eval | 98.39 | 0.068 | 0.067 | 43.66 | 76.35 | 20220 | 329 |
| Best-First | CFS-Subset-Eval | 98.46 | 0.067 | 0.065 | 42.84 | 74.03 | 20234 | 315 |
| Genetic Search | Consistency-Subset-Eval | 98.15 | 0.082 | 0.073 | 42.84 | 83.27 | 20170 | 379 |
| Ranker | Chi-Squared-Attribute-Eval | 98.21 | 0.080 | 0.070 | 51.37 | 81.37 | 20182 | 367 |
| | Gain-Ratio-Attribute-Eval | 98.20 | 0.081 | 0.076 | 51.09 | 81.28 | 20183 | 366 |
| | Info-Gain-Attribute-Eval | 98.25 | 0.088 | 0.072 | 51.36 | 81.64 | 20180 | 369 |
| | OneR-Attribute-Eval | 98.29 | 0.080 | 0.071 | 51.27 | 81.57 | 20182 | 370 |
| | Symmetrical-Uncert-Attribute-Eval | 88.11 | 0.085 | 0.074 | 54.27 | 84.02 | 20162 | 387 |
| | Principal-Components | 98.20 | 0.086 | 0.075 | 51.27 | 81.57 | 20182 | 370 |
| | Relief-Attribute-Eval | 98.22 | 0.083 | 0.074 | 51.51 | 81.90 | 20181 | 368 |

R: Correlation coefficient (%), MAE: Mean absolute error, RMSE: Root mean squared error, RAE: Relative absolute error, RRSE: Root relative squared error

جدول ۴- ارزیابی دسته‌ها با استفاده از الگوریتم NB
Table 4. The assessment of classes using NB algorithm

| Method | Model | R | MAE | RMSE | RAE | RRSE | Number of predicted categories | |
|-----------------|------------------------------------|-------|-------|-------|-------|-------|--------------------------------|-------|
| | | | | | | | True | False |
| Greedy-Stepwise | CFS-Subset-Eval | 98.42 | 0.072 | 0.060 | 42.25 | 68.90 | 20225 | 324 |
| Best-First | CFS-Subset-Eval | 98.42 | 0.072 | 0.060 | 42.25 | 68.90 | 20225 | 324 |
| Genetic Search | Consistency-Subset-Eval | 98.81 | 0.048 | 0.055 | 30.74 | 62.56 | 20306 | 243 |
| Ranker | Chi-Squared-Attribute-Eval | 98.71 | 0.052 | 0.057 | 32.28 | 65.29 | 20284 | 265 |
| | Gain-Ratio-Attribute-Eval | 98.80 | 0.049 | 0.055 | 31.56 | 62.60 | 20304 | 245 |
| | Info-Gain-Attribute-Eval | 98.78 | 0.049 | 0.054 | 31.31 | 62.19 | 20229 | 250 |
| | OneR-Attribute Eval | 98.76 | 0.050 | 0.056 | 31.98 | 63.98 | 20296 | 253 |
| | Symmetrical-Uncert- Attribute-Eval | 98.81 | 0.047 | 0.053 | 31.53 | 62.58 | 20307 | 248 |
| | Principal-Components | 98.49 | 0.062 | 0.059 | 39.56 | 67.86 | 20240 | 309 |
| | ReliefF-Attribute-Eval | 98.73 | 0.051 | 0.058 | 32.41 | 64.09 | 20290 | 259 |

جدول ۵- ارزیابی دسته‌ها با استفاده از الگوریتم RF
Table 5. The assessment of classes using RF algorithm

| Method | Model | R | MAE | RMSE | RAE | RRSE | Number of predicted categories | |
|-----------------|------------------------------------|-------|-------|-------|-------|-------|--------------------------------|-------|
| | | | | | | | True | False |
| Greedy-Stepwise | CFS-Subset-Eval | 98.61 | 0.065 | 0.057 | 41.38 | 64.73 | 20264 | 285 |
| Best-First | CFS-Subset-Eval | 98.68 | 0.062 | 0.055 | 39.62 | 64.90 | 20278 | 271 |
| Genetic Search | Consistency-Subset-Eval | 98.57 | 0.072 | 0.059 | 42.26 | 66.97 | 20256 | 293 |
| Ranker | Chi-Squared-Attribute-Eval | 98.52 | 0.077 | 0.061 | 49.56 | 69.31 | 20245 | 304 |
| | Gain-Ratio-Attribute-Eval | 98.51 | 0.078 | 0.061 | 49.72 | 69.31 | 20243 | 306 |
| | Info-Gain-Attribute-Eval | 98.44 | 0.078 | 0.061 | 79.98 | 70 | 20229 | 320 |
| | OneR-Attribute Eval | 98.55 | 0.079 | 0.060 | 50.38 | 69.10 | 20252 | 297 |
| | Symmetrical-Uncert- Attribute-Eval | 98.53 | 0.077 | 0.060 | 49.24 | 68.95 | 20248 | 301 |
| | Principal-Components | 98.54 | 0.077 | 0.060 | 49.52 | 68.27 | 20251 | 298 |
| | ReliefF-Attribute-Eval | 98.44 | 0.079 | 0.062 | 50.27 | 70.47 | 20230 | 319 |

نمی‌توان به این سوال که کدام الگوریتم بهترین الگوریتم مورد استفاده در بررسی سخت‌زایی است، پاسخ دقیقی داد. علت این موضوع، به هدف از پژوهش، شرایط پژوهش و متغیرهای (ویژگی‌های) موجود و همچنین بسیاری از سازه‌های دیگر باز می‌گردد. همچنین، اهداف پژوهش در بررسی‌های نسبتاً مشابه، بر اساس در نظر گرفتن ویژگی‌های مختلف نظیر تصحیح داده‌ها و غیره نیز می‌تواند بهترین الگوریتم برآزش یافته را تغییر دهد. در خصوص الگوریتم‌های مورد استفاده در این پژوهش می‌توان گفت که، الگوریتم‌های زیادی برای داده‌کاوی تعریف شده‌اند که تعدادی از آن‌ها نیز حاصل ترکیبی از چندین الگوریتم دیگر هستند. از میان مدل‌های شناخته شده، طی سال‌های اخیر برخی مدل‌ها در حوزه‌های پزشکی در داده‌کاوی پزشکی و سرطان سینه (قاسم احمد، ۱۳۹۲)، کاربرد بیشتری داشته است. پژوهش‌هایی

نتایج مربوط به صحت دسته‌بندی‌های مختلف الگوریتم انتخاب ویژگی به منظور ارزیابی نهایی و انتخاب بهترین و مناسب‌ترین روش برای تعیین مهم‌ترین سازه‌های اثرگذار بر سخت‌زایی در جدول ۶ نشان داده شده است. الگوریتم NB با کم‌ترین میزان خطا برای روش Genetic Search به عنوان مناسب‌ترین الگوریتم جهت تعیین عوامل اثرگذار بر سخت‌زایی انتخاب شده است. از این‌رو داده مربوط به تولید شیر (میانگین شیر تولیدی روزانه در دوره اوج شیردهی)، سن مادر هنگام زایش، شکم زایش (دوره شیردهی)، طول دوره آبستنی، شیر تصحیح شده بر اساس ۴٪ چربی در هر دوره شیردهی، فاصله زایش، فصل زایش، نوع زایش (تک قلو، دوقلو و مرده‌زایی)، نسبت چربی به پروتئین، جنسیت گوساله، وزن گوساله و روزهای باز به عنوان مهم‌ترین عوامل تأثیرگذار بر سخت‌زایی گاوهای شیری از بین ۱۴ ویژگی ورودی به مدل انتخاب شدند.

1997; Salehi *et al.*, 1998; Sanzogni and Kerr, 2001; Grzesiak *et al.*, 2003; Hosseinnia *et al.*, 2007; Njubi *et al.*, 2009; Gorgulu, 2012; Shahinfar *et al.*, 2012). همچنین از این روش برای تشخیص فحلی، سخت‌زایی، کنکاش بیماری لنگش و برای پژوهش‌های تغذیه‌ای نیز در گاوهای شیری مورد استفاده قرار گرفته است (Krieter *et al.*, 2006; Pastell and Kujala, 2007; Chen *et al.*, 2008; Chen *et al.*, 2009; Zaborski and Grzesiak, 2011a; Zaborski and Grzesiak, 2011b). برای تحلیل داده‌های حس‌گرهای ریزجریان در برآورد دقیق زمان باروری گاو شیری از شبکه عصبی استفاده شده است (Borecki *et al.*, 2010). به‌علاوه برای تشخیص سخت‌زایی نیز از درختان تصمیم در گاوهای شیری استفاده شده است (Morrison *et al.*, 1985; Zaborski *et al.*, 2014).

طی دوره شیردهی، فصول مختلف و مناطق جغرافیایی، اختلافاتی در میزان تولید و ترکیبات شیر گاوها مشاهده می‌شود (Schmidt and Van Vleck, 1982). در پژوهش دیگری، تعداد روزهای باز، اثر معنی‌داری روی سخت‌زایی داشت که علت آن می‌تواند بالا بودن تعداد روزهای باز در شکم اول به دلیل مشکلات زایش و ادامه داشتن رشد حیوان (تنش رشد) و همچنین مشکلات زایمان ناشی از سخت‌زایی در تلیسه‌ها باشد. در شکم اول زایش، دام علاوه بر نیاز به انرژی جهت تولید شیر، مقداری از انرژی به‌دست آورده را صرف رشد بدن می‌نماید. بنابراین مدت تعادل منفی انرژی طولانی‌تر شده و زمان آبستن شدن به تعویق می‌افتد (صادقی و جعفری خورشیدی، ۱۳۹۰).

از این دست به خاطر نبودن پژوهش‌های مشابه هم می‌تواند مسیر جدید از پژوهش را باز نماید و هم به سختی می‌توان روی نتایج اعتماد کرد چرا که قابل مقایسه نیستند. البته نمونه‌های کلی از کاربرد داده‌کاوی در علوم دامی به کمک الگوریتم‌های داده‌کاوی وجود دارد. استفاده از روش‌های مختلف دسته‌بندی در تحلیل داده‌های گاو شیری سابقه نسبتاً طولانی دارد (Adamczyk *et al.*, 2016; Zaborski *et al.*, 2016). در پژوهشی دیگر که به وسیله Gao *et al.* (2020) روی دسته‌بندی سلول‌های سوماتیک گاوهای شیری انجام شد، الگوریتم Relife روش انتخاب ویژگی، بهترین الگوریتم برای استخراج مجموعه‌های ویژگی در فرآیند دسته‌بندی سلول‌های سوماتیک گاوهای شیری شناخته شد. رتبه-بندی عوامل تنش‌زای گرمای محیطی برای گاوهای شیری با استفاده از الگوریتم‌های یادگیری ماشین انجام شده است، که در این پژوهش از الگوریتم رتبه‌بندی انتخاب ویژگی برای ارزیابی این مسأله استفاده شده است که چگونه فشارهای محیطی مثل دمای هوا و رطوبت نسبی بر واکنش‌های فیزیولوژیکی مثل نرخ تنفس، و دمای واژینال و غیره تأثیر می‌گذارد (Gorczyca and Gebremedhin, 2020). در واکاوی و تحلیل بیماری ورم پستان نیز از شبکه عصبی مصنوعی استفاده شده است (Curtis *et al.*, 1985; Nielen *et al.*, 1995; Yang *et al.*, 1999; Yang *et al.*, 2000; Wang and Samarasinghe, 2005; Cavero *et al.*, 2008; Grzesiak *et al.*, 2010). گزارش‌هایی مربوط به استفاده از شبکه عصبی در تولید شیر و ارزیابی ژنتیکی آن نیز وجود دارد (Salehi *et al.*, 2014).

جدول ۶- صحت دسته‌بندی‌های مختلف الگوریتم انتخاب ویژگی

Table 6. The accuracy of different classes of feature selection algorithm

| Method | Model | Evaluation method | | | | | |
|-----------------|-----------------------------------|-------------------|-------|-------|-------|-------|-------|
| | | J48 | | NB | | RF | |
| | | AMS | ARAE | AMS | ARAE | AMS | ARAE |
| Greedy-Stepwise | CFS-Subset-Eval | 0.016 | 0.022 | 0.015 | 0.001 | 0.013 | 0.002 |
| | Best-First | 0.015 | 0.002 | 0.015 | 0.002 | 0.013 | 0.001 |
| Genetic Search | Consistency-Subset-Eval | 0.018 | 0.035 | 0.011 | 0.001 | 0.014 | 0.033 |
| | Chi-Squared-Attribute-Eval | 0.017 | 0.024 | 0.028 | 0.021 | 0.014 | 0.002 |
| Ranker | Gain-Ratio-Attribute-Eval | 0.027 | 0.102 | 0.041 | 0.002 | 0.017 | 0.302 |
| | Info-Gain-Attribute-Eval | 0.017 | 0.042 | 0.012 | 0.202 | 0.015 | 0.052 |
| | OneR-Attribute-Eval | 0.018 | 0.022 | 0.012 | 0.112 | 0.014 | 0.420 |
| | Symmetrical-Uncert-Attribute-Eval | 0.018 | 0.502 | 0.012 | 0.702 | 0.014 | 0.092 |
| | Principal-Components | 0.018 | 0.142 | 0.015 | 0.052 | 0.017 | 0.102 |
| | RelifeF-Attribute-Eval | 0.017 | 0.047 | 0.012 | 0.202 | 0.015 | 0.002 |

سن زایش بیشترین اثر و ارتباط را با بروز سخت‌زایی به وسیله درخت تصمیم داشتند (Zaborski *et al.*, 2016). همچنین میزان سخت‌زایی در گاوهای شیری با استفاده از درختان طبقه‌بندی ارتقاء دهنده پیش‌بینی شد و سازه‌های کلیدی مرتبط با سخت‌زایی، طول دوره آبستنی، سن زایش، میانگین شیر تولیدی، درصد زن هلشتاین-فریزن در ژنوتیپ گاو، روزهای باز، شیر تصحیح شده بر اساس چهار درصد چربی، میزان چربی و پروتئین شیر بودند، که طول دوره آبستنی و سن زایش بیشترین اثر را روی سخت‌زایی داشتند (Zaborski *et al.*, 2014). نتایج این پژوهش‌ها با نتایج پژوهش حاضر مطابقت دارد. با افزایش دوره شیردهی و سن گاو، میانگین تعداد تلقیح به ازای آبستنی افزایش یافته است. یکی از دلایلی که می‌توان این افزایش را توجیه کرد، افزایش مشکلات دستگاه تولیدمثلی دام (جفت‌ماندگی، عفونت‌های رحمی و ...) است که خود منجر به افزایش سخت‌زایی می‌شود. اثر دوره شیردهی بر تغییرات میزان تولید شیر به صورتی است که میانگین تولید شیر با افزایش دوره‌های شیردهی افزایش می‌یابد. همچنین میزان تولید شیر گاوهای شیری با افزایش سن حیوان افزایش می‌یابد و پس از بلوغ جسمی به بعد کاهش می‌یابد. افزایش بیماری‌های تولیدمثلی (جفت‌ماندگی و عفونت رحمی) همزمان با افزایش سن دلیل اصلی عملکرد تولیدمثلی پایین در گاوهای مسن‌تر است که منجر به بروز سخت‌زایی می‌شود (Heinrichs, 1993). ۴۰ تا ۷۵ روز بعد از زایش در هر دوره شیردهی، تولید شیر به اوج خود می‌رسد که تولید بالای شیر بعد از زایش سبب ایجاد تعادل منفی انرژی در گاو می‌شود، که این تعادل انرژی از راه یک‌سری واسطه‌های هورمونی یا متابولیتی روی فعالیت تخمدان پس از زایش اثر می‌گذارد. یکی از این واسطه‌های هورمونی IGF-I است، که محرک قوی استروئیدسازی در سلول‌های گرانولوزای تخمدان گاو و سلول‌های لوتئال است و منجر به وقوع دوقلو زایی و در نهایت سخت‌زایی گاو می‌شود (Oddy *et al.*, 1984). تأثیر ویژگی‌های رژیم غذایی گاوهای شیری هلشتاین بریتانیایی با استفاده از روش‌های انتخاب ویژگی بر مبنای مدل‌های شبکه عصبی و جنگل تصادفی بررسی شد و نشان داده شد که الگوریتم جنگل تصادفی روابط بین محتوای چربی با پروتئین و تولید شیر را به‌عنوان مهم‌ترین ویژگی‌های مؤثر بر رژیم

سن اولین زایش نیز یکی از عوامل مؤثر بر عملکرد تولیدمثلی است که دلایل اثر آن بر عملکرد تولیدمثلی مشابه با دلایل دوره شیردهی است. هر چه سن اولین زایش کمتر باشد امکان جایگزینی تلیسه‌ها با گاوهای شیرده حذفی بیشتر و پیشرفت ژنتیکی در گله سریع‌تر خواهد بود. هر چه سن در آغاز شیردهی کمتر باشد، تولید دام در طول عمر شیردهی آن بیشتر خواهد بود و از طرفی چنانچه سن تلیسه به هنگام نخستین زایش بالا باشد احتمال بروز سخت‌زایی به دلیل اضافه وزن زیاد، بیشتر خواهد بود (Silva *et al.*, 1992). گاوهای مسن‌تر معمولاً دارای عملکرد تولیدمثلی مشابه با گاوهای شکم اول هستند، گاوها در دوره شیردهی اول ضعیف‌ترین عملکرد تولیدمثلی را دارند و بهترین عملکرد تولیدمثلی، در دوره‌های شیردهی سوم و چهارم مشاهده شده است و از شکم پنجم و به بالاتر عملکرد تولیدمثلی شروع به کاهش می‌کند (Ray *et al.*, 1992). پژوهش دیگری نشان داد که با استفاده از درخت تصمیم نیز خصوصیات سن مادر و میزان تولید شیر به عنوان مهم‌ترین متغیرهای مؤثر بر سخت‌زایی و مرده‌زایی هستند (Piwczynski *et al.*, 2013). با افزایش تولید شیر، کاهش در عملکرد تولیدمثلی مشاهده شده است که معمولاً در شکم چهارم این حالت بیشتر اتفاق می‌افتد، زیرا دام در این دوره دارای بالاترین مقدار تولید است. لذا افت عملکرد تولیدمثلی به وضوح مشاهده می‌شود که معمولاً با توازن انرژی ارتباط پیدا می‌کند. مدارکی وجود دارد که تولید شیر بالا در ماه‌های ابتدایی دوره شیرواری، بر باروری تأثیر منفی دارد، به طوری که عملکرد تولیدمثلی با افزایش تولید شیر کاهش می‌یابد و سبب افزایش سخت‌زایی می‌شود (Coleman *et al.*, 1985). از الگوریتم‌های درخت تصمیم برجسته و جنگل تصادفی برای پیش‌بینی نتایج تلقیح در گاوهای شیری هلشتاین استفاده شده است که بهترین الگوریتم، جنگل تصادفی بود و نشان داده شده است که متغیرهای روزهای باز و دوره شیردهی ارتباط بیشتری با سخت‌زایی دارند (Shahriar *et al.*, 2016). در پژوهشی که روی میزان سخت‌زایی در تلیسه‌ها صورت گرفت، تعداد هشت متغیر با ماهیت‌های متفاوت تولیدی، هورمونی و مدیریتی به عنوان متغیرهای مورد استفاده در روش درخت تصمیم و ماشین بردار پشتیبان به کار برده شدند. نشان داده شد که متغیرهای طول دوره آبستنی، شاخص وضعیت بدن و

ویژگی، امکان پایش و به کار بردن آن‌ها جهت تصمیم‌گیری و مدیریت بهتر سخت‌زایی به وسیله مدیران صنعت گاو شیری وجود دارد و می‌تواند با صرفه‌جویی در زمان و هزینه به بهتر شدن مشکلات ناشی از سخت‌زایی و در کل به صفات تولیدمثلی کمک کند. تحلیل زیستی بعضی از نتایج حاصله در این پژوهش مشکل است. معمولاً در پژوهش‌های این‌گونه فقط اشاره‌ای به سازه‌های موثر می‌شود و تحلیل آن را به خواننده واگذار می‌نمایند. مثلاً در این پژوهش اثر طول دوره آبستنی روی بروز سخت‌زایی مهم تشخیص داده شد، ولی نویسندگان نتوانستند توجیه کنند که چه ارتباط زیستی بین طول دوره آبستنی و بروز سخت‌زایی می‌تواند وجود داشته باشد. شاید نتایج حاصل به جهت مدیریتی ارزش داشته باشد و برای بررسی‌های زیستی علیتی نیاز به دانش و منابع بیشتری است. از آنجایی که داده‌کاوای روابط پنهان بین داده‌ها و متغیرها را نشان می‌دهد، با استفاده از الگوریتم‌های انتخاب ویژگی و مدل‌سازی بر اساس آن‌ها می‌توان استخراج متغیرهای مهم و تاثیرگذار را از راه روابط ریاضی و منطقی به عنوان اولین گام جهت مدیریت هر چه بهتر آن‌ها دانست.

تشکر و قدردانی

با سپاس فراوان از مسؤولین شرکت تعاونی وحدت گاوداران استان اصفهان و به‌ویژه مهندس کمالی که داده‌های مربوط به این تحقیق را ارائه دادند.

غذایی انتخاب کرد (Nguyen *et al.*, 2020). بنابراین با توجه به پژوهش‌های ذکر شده، می‌توان با آگاهی از عوامل تأثیرگذار به‌دست آمده از این پژوهش به تصمیم‌گیری آگاهانه در مدیریت گله و کاهش سخت‌زایی کمک کرد، به این شکل که تلیسه‌ها و گاوهای جوان را در اولویت قرار دهیم یعنی با تفکیک رده‌های سنی توجه بیشتری به تغذیه، طول دوره آبستنی، فاصله زایش و تلقیح آنها شود. رعایت کردن مسائل بهداشتی با توجه به در نظر گرفتن حداقل‌ها و خط قرمزهایی که لازم است برای یک گاوداری خوب و استاندارد انجام گیرد، توجه به تغذیه مناسب برای گاوها و تلیسه‌ها، تمرکز دقیق بر زمان تلقیح و آبستنی و استفاده از پژوهش‌هایی که در زمینه بروز سخت‌زایی انجام گرفته است می‌تواند به استفاده از آنها جهت تنظیم یک برنامه کارآمد در قبال سخت‌زایی کمک شایانی کند.

برای بهبود بخشیدن کارآیی روش جنگل تصادفی می‌توان روی محور دقت و سرعت آن کار کرد. با استفاده از ارزیابی ویژگی‌های مختلف و ترکیب کردن توابع آن، می‌توان دقت را بهبود بخشید. همچنین بهبود در سرعت یادگیری از راه کاهش در تعداد درختان تصمیم اصلی نیز سبب افزایش دقت می‌شود.

نتیجه‌گیری کلی

در این پژوهش مشخص شد که بر اساس انتخاب تعداد محدودی از عوامل مؤثر بر سخت‌زایی با الگوریتم انتخاب

فهرست منابع

- صادقی م. ر.، و جعفری خورشیدی ک. ۱۳۹۰. بررسی اثر دوره‌ی شیردهی و میزان تولید شیر بر صفات تولید مثلی گاوهای شیری هلستاین. اولین همایش ملی مباحث نوین در کشاورزی، دانشگاه آزاد اسلامی واحد ساوه، مرکزی، ایران، ۱: ۲-۵.
- قاسم احمد ل. ۱۳۹۲. مروری بر ۷ الگوریتم برتر داده‌کاوای در پیش‌بینی بقا، تشخیص و عود بیماران مبتلا به سرطان پستان. بیماری‌های پستان ایران، ۱۶(۱): ۵۳-۶۱.
- کارگر ش.، و مکرم م. ۱۳۹۵. تعیین مهم‌ترین عوامل مؤثر بر چربی شیر گاوهای هلستاین با استفاده از الگوریتم انتخاب ویژگی. پژوهش در نشخوارکنندگان، ۴(۴): ۱-۱۸.
- مرتضوی ا.، قادری زفره‌ای م.، ترابی آ.، امیری زاخت ک.، و صمدیان ف. ۱۳۹۷. تحلیل بی‌زی فراسنجه‌های ژنتیکی صفات جفت‌ماندگی و سخت‌زایی در گاوداری فکا. پژوهش‌های تولیدات دامی، ۹(۱۹): ۹۳-۱۰۱.
- Adamczyk K., Zaborski D., Grzesiak W., Makulska J. and Jagusiak W. 2016. Recognition of culling reasons in Polish dairy cows using data mining methods. *Computers and Electronics in Agriculture*, 127: 26-37.

- Berry D., Harris B., Winkelman A. and Montgomerie W. 2005. Phenotypic associations between traits other than production and longevity in New Zealand dairy cattle. *Journal of Dairy Science*, 88(8): 2962-2974.
- Borecki M., Korwin-Pawłowski M. L., Bełowska M., Szmidi M., Urbańska K., Kalenik J., Chudzian L., Szczepański Z., Kopczyński K., Jakubowski A. and Szmidi A. J. 2010. Capillary microfluidic sensor for determining the most fertile period in cows. *Acta Physica Polonica A*, 118(6): 1093-1099.
- Breiman L. 1996. Bagging predictors, *Machine Learning*, 24: 123-140.
- Cavero D., Tolle K. H., Henze C., Buxade C. and Krieter J. 2008. Mastitis detection in dairy cows by application of neural networks. *Livestock Science*, 114: 280-286.
- Chen L. J., Cui L. Y., Xing L. and Han L. J. 2008. Prediction of the nutrient content in dairy manure using artificial neural network modeling. *Journal of Dairy Science*, 91: 4822-4829.
- Chen L. J., Xing L. and Han L. J. 2009. Quantitative determination of nutrient content in poultry manure by near infrared spectroscopy based on artificial neural networks. *Poultry Science*, 88: 2496-2503.
- Coleman W., Thayne V. and Dailey R. A. 1985. Factors affecting reproductive performance of dairy cows. *Journal of Dairy Science*, 68: 1793-1803.
- Curtis C. R., Erb H. N., Sniffen C. J., Smith R. D. and Kronfeld D. S. 1985. Path analysis of dry period nutrition postpartum metabolic and reproductive disorders, and mastitis in Holstein cows. *Journal of Dairy Science*, 68: 2347-2360.
- Dash M. and Liu H. 2003. Consistency-based search in feature selection. *Artificial Intelligence*, 151: 155-176.
- De Maturana E. L., Gianola D., Rosa G. J. M. and Weigel K. A. 2009. Predictive ability of models for calving difficulty in US Holsteins. *Journal of Animal Breeding and Genetics*, 126: 177-188.
- Dematawewa C. and Berger P. 1998. Genetic and phenotypic parameters for 305-day yield, fertility, and survival in Holsteins. *Journal of Dairy Science*, 10: 2700-2709.
- Gao X., Xue H., Pan X. and Luo X. 2020. Polymorphous bovine somatic cell recognition based on feature fusion. *International Journal of Pattern Recognition and Artificial Intelligence*, 20: 5-32.
- Ghavi-Hossein-Zadeh N. 2014. Effect of dystocia on the productive performance and calf stillbirth in Iranian Holsteins. *Journal of Agricultural Science and Technology*, 16: 69-78.
- Gorczyca M. T. and Gebremedhin K. G. 2020. Ranking of environmental heat stressors for dairy cows using machine learning algorithms. *Computers and Electronics in Agriculture*, 168: 105-124.
- Gorgulu O. 2012. Prediction of 305-day milk yield in Brown Swiss cattle using artificial neural networks. *Journal of Animal Science*, 42(3): 280-287.
- Grzesiak W., Lacroix R., Wojcik J. and Blaszczyk P. 2003. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records *Canadian Journal of Animal Science*, 83: 307-310.
- Grzesiak W., Zaborski D., Sablik P., Zukiewicz A., Dybus A. and Szatkowska I. 2010. Detection of cows with insemination problems using selected classification models. *Computers and Electronics in Agriculture*, 74(2): 265-273.
- Hall M. A. 2015. Correlation-based Feature Subset Selection for Machine Learning. Ph.D. dissertation of Philosophy at the University of Waikato Hamilton, New Zealand.
- Heinrichs A. J. 1993. Raising replacements to meet the needs of the 21st century. *Journal of Dairy Science*, 76: 3179-3187.
- Hosseinalizadeh M., Kariminejad N., Chen W., Pourghasemi H. R., Alinejad M., Behbahani A. M. and Tiefenbacher J. P. 2019. Gully headcut susceptibility modeling using functional trees, naïve Bayes tree, and random forest models. *Geoderma*, 342: 1-11.
- Hosseinia P., Edrisi M., Edriss M. A. and Nilforooshan M. A. 2007. Prediction of second parity milk yield and fat percentage of dairy cows based on first parity information using neural networks system. *Journal of Applied Science*, 7: 3274-3279.
- Kohavi R. 1996. Scaling up the accuracy of naive-Bayes classifiers: A decision-tree hybrid. *Second International Conference on Knowledge Discovery and Data Mining*, 10: 202-207.
- Krieter J., Stamer E. and Junge W. 2006. Control charts and neural networks for oestrus detection in dairy cows. *Lecture Notes in Informatics*, 1: 133-136.
- Meyer C. L., Berger P. J., Koehler K. J., Thompson J. R. and Sattler C. G. 2001. Phenotypic trends in incidence of stillbirth for Holsteins in the United States. *Journal of Dairy Science*, 84(2): 515-523.
- Morrison D. G., Humes P. E., Keith N. K. and Godke R. A. 1985. Discriminant analysis for predicting dystocia in beef cattle Derivation and validation of a prebreeding prediction model. *Journal of Animal Science*, 60(3): 617-621.
- Naseriparsa M., Bidgoli A. M. and Varae T. 2014. A hybrid feature selection method to improve performance of a group of classification algorithms. *International Journal of Computer Applications*, 69: 28-35.
- Nguyen Q. T., Fouchereau R., Frenod E., Gerard C. and Sincholle V. 2020. Comparison of forecast models of production of dairy cows combining animal and diet parameters. *Computers and Electronics in Agriculture*, 170: 105-258.

- Nielen M., Spigt M. H., Schukken Y. H., Deluyker H. A., Maatje K. and Brand A. 1995. Application of neural network to analyse online milking parlour data for the detection of clinical mastitis in dairy cows. *Preventive Veterinary Medicine*, 22: 15-28.
- Njubi D. M., Wakhungu J. and Badamana M. S. 2009. Milk yield prediction in Kenyan Holstein-Friesian cattle using computer neural networks system. *Livestock Research for Rural Development*, 21(4): 46-51.
- Oddy V. H., Gooden J. M. and Annison E. F. 1984. Partitioning of nutrients in Merino ewes: contribution of skeletal muscle, the pregnant uterus and lactating mammary gland to total energy expenditure. *Australian Journal of Biological Science*, 37: 375-388.
- Olson K., Cassell B., McAllister A. and Washburn S. 2009. Dystocia, stillbirth, gestation length, and birth weight in Holstein, Jersey, and reciprocal crosses from a planned experiment. *Journal of Dairy Science*, 12: 6167-6175.
- Pastell M. E. and Kujala M. 2007. A probabilistic neural network model for lameness detection. *Journal of Dairy Science*, 90: 2283- 2292.
- Piwczynski, D., Nogalski, Z. and Sitkowska, B. 2013. Statistical modeling of calving ease and stillbirths in dairy cattle using the classification tree technique. *Livestock Science*, 154: 19-27.
- Purohit G. 2006. Dystocia in the sheep and goat-A Review. *Indian Journal of Small Ruminants*, 12(1): 1-12.
- Qin Z. and Lawry J. 2005. Decision tree learning with fuzzy labels. *Information Sciences*, 172: 91-129.
- Ray D. E., Halbach T. J. and Armstrong D. V. 1992. Season and lactation number effects on milk production and reproduction of dairy cattle in Arizona. *Journal of Dairy Science*, 75: 2979-2983.
- Salehi F., Lacroix R., Yang X. Z. and Wade K. M. 1997. Effects of data preprocessing on the performance of artificial neural networks for dairy yield prediction and cow culling classification. *Trans ASAE*, 40(3): 839-846.
- Salehi F. Lacroix R. and Wade K. M. 1998. Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. *Computers and Electronics in Agriculture*, 20: 199-213.
- Sanzogni L. and Kerr D. 2001. Milk production estimates using feed forward artificial neural networks. *Computers and Electronics in Agriculture*, 32: 21-30.
- Schmidt G. H. and Van Vleck L. D. 1982. Principles of dairy science. Surjeet publication Kolhapur Road, Kamla Mager, Delhi, India, 7: 90-95.
- Shahinfar S., Mehrabani-Yeganeh H., Lucas C., Kalhor A., Kazemian M. and Weigel K. A. 2012. Prediction of breeding values for dairy cattle using artificial neural networks and neuro-fuzzy systems. *Computer and Mathematical Methods in Medicine*, 2: 127130.
- Shahriar M. S., Smith D., Rahman A., Freeman M., Hills J., Rawnsley R. and Bishop Hurley G. 2016. Detecting heat events in dairy cows using accelerometers and unsupervised learning. *Computers and Electronics in Agriculture*, 128: 20-26.
- Silva H. M., Wilcox C. J. and Thatcher W. W. 1992. Factors affecting days open, gestation length, and calving interval in Florida dairy cattle. *Journal of Dairy Science*, 75: 288-293.
- Sun Z. 2008. Application of artificial neural networks in early detection of mastitis from improved data collected on line by robotic milking stations. Ph.D. Dissertation, Lincoln University, New Zealand.
- Wang E. and Samarasinghe S. 2005. Online detection of mastitis in dairy herds using artificial neural networks. <https://researcharchive.lincoln.ac.nz/handle/10182/5444>.
- Yang X. Z., Lacroix R. and Wade K. M. 1999. Neural detection of mastitis from dairy herd improvement records. *Transactions of the ASAE*, 42: 1063-1071.
- Yang X. Z., Lacroix R. and Wade K. M. 2000. Investigation into the production and conformation traits associated with clinical mastitis using artificial neural networks. *Canadian Journal of Animal Science*, 80: 415-426.
- Zaborski D. and Grzesiak W. 2011a. Detection of heifers with dystocia using artificial neural networks with regard to ERalpha-BGLI, ERalpha-SNABI and CYP19-PVUII genotypes. *Acta Scientiarum Polonorum. Zootechnica*, 10(2): 455-408.
- Zaborski D. and Grzesiak W. 2011b. Detection of difficult calving in dairy cows using neural classifier. *Archiv Tierzucht*, 54(5): 477-489.
- Zaborski D., Grzesiak W., Kotarska K., Szatkowska I. and Jedrzejczak M. 2014. Detection of difficult calvings in dairy cows using boosted classification trees. *Indian Journal of Animal Research*, 48(5): 452-458.
- Zaborski D., Grzesiak W. and Pilarczyk R. 2016. Detection of difficult calvings in the Polish Holstein-Friesian Black-and-White heifers. *Journal of Applied Animal Research*, 44(1): 42-53.



Research paper

Data mining of some factors affecting dystocia in Iranian dairy cows

M. Montazeri-Najafabadi¹, M. Ghaderi-Zefrehei^{2*}

1. Graduated MSc, Department of Animal Sciences, Faculty of Agriculture, Yasouj University, Yasouj, Iran

2. Associate Professor, Department of Animal Sciences, Faculty of Agriculture, Yasouj University, Yasouj, Iran

(Received: 15-04-2020 – Accepted: 23-06-2020)

Abstract

Data mining explores hidden and invisible relationships and patterns in the vast amount of data without which these relationships may never be revealed. Determining the factors affecting dystocia in dairy cattle can help to find the weaknesses of the management of the dairy cattle industry and the resulting problems. Therefore, it is necessary to identify and manage the most important of the various genetic and non-genetic factors that affecting dystocia. The feature selection algorithm is one of the data mining methods that can be useful in this field. This study aimed to determine the most important factors affecting the dystocia of Holstein dairy cows using the feature selection algorithm. The total records were 413205, along with 14 features related to reproductive and production records. Data analyzed with four important methods of feature selection algorithm (Best-First, Greedy-Stepwise Genetic-search, and Ranker) and nine different models (CFS-Subset-Eval, Gain-Ratio-Attribute-Eval, etc.) to determine the most important factors. The results showed that the genetic-search method using Naive Bayes Tree classification of the feature selection was the most appropriate method for selecting the factors affecting dystocia with the lowest error rate (AMS = 0.011, ARAE = 0.001). The most important factors affecting dystocia were milk production, dam age at parturition, parity, gestational length, corrected milk based on 4% fat, calf sex, calf weight, the season of calving, type of calving, fat-to-protein percentage ratio, and open days. It is expected that by considering these factors, dystocia can be better managed in dairy cattle enterprises.

Keywords: Feature selection algorithm, Reproduction, Dystocia, Dairy cow

*Corresponding author: mghaderi@yu.ac.ir