https://doi.org/10.22034/aa.2026.2081276.1022

کاربرد رگرسیون چندنمونه‌ای بیزی مبتنی بر جستجوی تصادفی شاتگان در انتخاب همزمان نمونه و متغیر: مطالعه موردی داده‌های سوختگی

فریبا کرمی[۱] و سید روح‌الله روزگار[۲]*

[۱،۲] گروه ریاضی، دانشکده علوم پایه، دانشگاه یاسوج، یاسوج، ایران

تاریخ دریافت: ۱۴۰۴/۱۰/۰۱
تاریخ پذیرش: ۱۴۰۵/۰۴/۱۶

چکیده

در بسیاری از مطالعات پزشکی، داده‌ها دارای ساختار چندنمونه‌ای هستند؛ به‌گونه‌ای که هر بیمار شامل مجموعه‌ای از نمونه‌ها بوده، در حالی که متغیر پاسخ تنها در سطح بیمار ثبت می‌شود. انتخاب همزمان نمونه‌های مؤثر درون هر بسته و متغیرهای مرتبط با پاسخ، یکی از چالش‌های مهم رگرسیون چندنمونه‌ای است. در این پژوهش، چارچوب رگرسیون چندنمونه‌ای بیزی مبتنی بر جستجوی تصادفی شاتگان (MIR-SSS) به کار گرفته و عملکرد آن با رگرسیون وزنی تجمیعی (APW) و لاسو مقایسه شد. انتخاب متغیرها با پیشین اسپایک–اسلب و انتخاب نمونه‌ها از طریق مدل لوژستیک انجام می‌شود. روش‌ها در چهار حالت شبیه‌سازی و سپس روی داده‌های واقعی ۲۰۲۴ بیمار بستری در بیمارستان سوختگی امیرالمؤمنین شیراز ارزیابی شدند. نتایج نشان داد مدل MIR-SSS در حالت‌های پایه کمترین خطای پیش‌بینی و دقت بالایی در انتخاب متغیرهای مؤثر دارد، هرچند در شرایط پراکندگی بالا و داده‌های پرت، APW عملکرد پیش‌بینی بهتری داشت. مقادیر سطح زیر منحنی ROC نیز توانایی مدل در شناسایی نمونه‌های مؤثر را تأیید کرد. در داده‌های واقعی، MIR-SSS عملکرد پیش‌بینی بهتری نسبت به دو روش مرجع داشت و عوامل بالینی مهم را شناسایی کرد. نتایج نشان می‌دهد این چارچوب امکان انتخاب همزمان نمونه‌ها و متغیرها و کمی‌سازی عدم‌قطعیت را برای تحلیل داده‌های پزشکی فراهم می‌کند.

واژه‌های کلیدی: یادگیری چندنمونه‌ای، مدل‌های بیزی، انتخاب نمونه، انتخاب متغیر، الگوریتم جستجوی تصادفی شاتگان.
رده‌بندی ریاضی: ۶۲F۱۵، ۶۲J۰۷.

۱ مقدمه

تحلیل داده‌های زیستی و پزشکی در سال‌های اخیر با رشد چشمگیر روش‌های یادگیری آماری همراه بوده است. یکی از چالش‌های مهم در این حوزه، وجود داده‌هایی با ساختار گروه‌بندی‌شده است که در آن، اطلاعات در سطح نمونه‌ها قابل مشاهده است اما پاسخ یا برچسب تنها در سطح گروه یا بسته در دسترس قرار دارد. این نوع داده‌ها در حوزه‌های مختلفی از جمله تصاویر پزشکی، تحلیل ژنتیکی، طراحی دارو و داده‌های بالینی مشاهده می‌شوند.

*نویسنده مسئول مقاله، roozegar@yu.ac.ir

یکی از چارچوب‌های مناسب برای تحلیل چنین داده‌هایی، یادگیری چندنمونه‌ای[1] است. در این رویکرد، هر بسته شامل مجموعه‌ای از نمونه‌ها است که هر یک با برداری از ویژگی‌های توضیحی توصیف می‌شوند، اما تنها یک پاسخ کلی برای کل بسته وجود دارد. این ساختار به‌ویژه در داده‌های پزشکی اهمیت دارد، زیرا برچسب‌گذاری دقیق هر نمونه معمولاً دشوار، پرهزینه یا حتی غیرممکن است، در حالی که برچسب کلی بیمار یا وضعیت نهایی بیماری در دسترس قرار دارد. در پژوهش حاضر، داده‌های بیماران سوختگی در قالب چنین ساختاری مورد بررسی قرار گرفته‌اند؛ به‌طوری‌که هر بیمار به‌عنوان یک بسته در نظر گرفته می‌شود و اندازه‌گیری‌های بالینی و آزمایشگاهی انجام‌شده در طول دوره بستری، نمونه‌های درون آن بسته را تشکیل می‌دهند، در حالی که پاسخ نهایی تنها در سطح بیمار ثبت شده است.

با وجود کاربردهای گسترده یادگیری چندنمونه‌ای، بسیاری از روش‌های موجود از نظر تفسیرپذیری با محدودیت مواجه‌اند. برای مثال، مدل‌های مبتنی بر شبکه‌های عصبی عمیق و سایر روش‌های جعبه‌سیاه اگرچه در بسیاری از مسائل از دقت پیش‌بینی بالایی برخوردارند، اما در محیط‌های بالینی که تصمیم‌گیری پزشکی نیازمند شفافیت و قابلیت توضیح است، همواره مطلوب نیستند. از این‌رو، توسعه و به‌کارگیری مدل‌هایی که علاوه بر دقت پیش‌بینی مناسب، امکان تفسیر نتایج و کمّی‌سازی عدم‌قطعیت را نیز فراهم کنند، از اهمیت ویژه‌ای برخوردار است.

در سال‌های اخیر، چارچوب‌های بیزی توجه قابل ملاحظه‌ای را در مسائل یادگیری چندنمونه‌ای به خود جلب کرده‌اند. یکی از جدیدترین پژوهش‌ها در این زمینه، مطالعه پارک و همکاران [۱۹] است که در آن یک چارچوب رگرسیون چندنمونه‌ای بیزی برای انتخاب همزمان نمونه‌های کلیدی و متغیرهای مؤثر ارائه شده است. در این چارچوب، با استفاده از متغیرهای نشانگر پنهان، نمونه‌های اثرگذار در هر بسته شناسایی شده و به‌صورت همزمان انتخاب متغیرها نیز انجام می‌شود. همچنین برای جستجو در فضای بزرگ مدل‌های ممکن، از الگوریتم جستجوی تصادفی شاتگان استفاده شده است که کارایی مناسبی در مسائل انتخاب مدل با ابعاد بالا دارد.

پژوهش حاضر با بهره‌گیری از چارچوب ارائه‌شده توسط پارک و همکاران [۱۹]، به بررسی عملکرد این روش در داده‌های بیماران سوختگی می‌پردازد. هدف اصلی، ارزیابی قابلیت این چارچوب در شناسایی متغیرهای مؤثر و نمونه‌های کلیدی در یک مسئله واقعی پزشکی است. افزون بر این، ساختار بیزی مدل امکان کمّی‌سازی عدم‌قطعیت در فرایند انتخاب نمونه‌ها و متغیرها را فراهم می‌سازد که می‌تواند در تفسیر نتایج و تصمیم‌گیری‌های پزشکی نقش مهمی ایفا کند.

از منظر کاربردی، استفاده از این چارچوب در داده‌های بیماران سوختگی می‌تواند به شناسایی عوامل بالینی و آزمایشگاهی مؤثر بر پیامد بیماران کمک کند. همچنین نتایج حاصل از این مطالعه می‌تواند شواهد تجربی ارزشمندی در خصوص قابلیت استفاده از روش‌های رگرسیون چندنمونه‌ای بیزی در مسائل پزشکی و سلامت فراهم آورد.

مدل MIL نخستین‌بار در زمینه پیش‌بینی فعالیت داروها مطرح شد. مطالعه کلاسیک دیتریش و همکاران [۴] نشان داد که در بسیاری از مسائل یادگیری، هر شیء می‌تواند با چندین بردار ویژگی توصیف شود، اما تنها بخشی از این بردارها مسئول برچسب نهایی هستند.

در ادامه، مفهوم رگرسیون چندنمونه‌ای توسط پیج و ری [۲۱] معرفی شد. این رویکرد به جای تمرکز بر طبقه‌بندی، امکان پیش‌بینی مقادیر پیوسته را فراهم کرد. کاربرد اصلی آن در طراحی دارو بود، جایی که هر مولکول به‌عنوان یک بسته شامل چندین ساختار سه‌بعدی در نظر گرفته می‌شود و تنها برخی از این ساختارها مسئول فعالیت زیستی هستند.

ژانگ و ژو [۲۵] با تمرکز بر مسئله انتخاب ویژگی‌ها در شبکه‌های عصبی چندنمونه‌ای، روش‌هایی برای بهبود عملکرد مدل‌های MIL ارائه کردند. این مطالعه نقطه عطفی در ترکیب چارچوب مدل MIL با روش‌های کاهش ابعاد و انتخاب ویژگی محسوب می‌شود.

مطالعه ژائو و یو [۲۶] اهمیت انتخاب ویژگی‌ها را در چارچوب بیزی و با استفاده از روش لاسو بررسی کرد.

ژائو و همکاران [۲۷] رویکردی برای پیش‌بینی فعالیت داروها با استفاده از انتخاب مشترک نمونه و ویژگی ارائه کردند.

زافرا و همکاران [۲۴] الگوریتم ترکیبی HyDR-MI را معرفی کردند که از ادغام یک روش فیلتر مبتنی بر امتیازدهی ویژگی‌ها (روش ReliefF) و یک الگوریتم ژنتیک برای انتخاب ویژگی‌ها در چارچوب یادگیری چندنمونه‌ای بهره می‌برد.

در سال‌های اخیر، پارک و همکاران [۱۹] چارچوبی بیزی برای رگرسیون چندنمونه‌ای ارائه کردند که امکان انتخاب همزمان نمونه‌های کلیدی و متغیرهای مؤثر را فراهم می‌سازد. در این روش، با استفاده از متغیرهای نشانگر پنهان، نمونه‌های اثرگذار در هر بسته شناسایی شده و انتخاب متغیرها نیز به‌طور همزمان انجام می‌شود. همچنین برای جستجوی مؤثر در فضای مدل‌ها از الگوریتم جستجوی تصادفی شاتگان استفاده شده است. نتایج این پژوهش نشان داد که چارچوب

[1] Multiple Instance Learning (MIL)

پیشنهادی در شناسایی نمونه‌های مؤثر و متغیرهای مرتبط عملکرد مطلوبی دارد.

جمع‌بندی مطالعات انجام‌شده نشان می‌دهد که انتخاب همزمان نمونه‌ها و متغیرها همچنان یکی از موضوعات مهم در حوزه رگرسیون چندنمونه‌ای است. با وجود پیشرفت‌های صورت‌گرفته، هنوز ارزیابی عملکرد روش‌های جدید در مسائل واقعی پزشکی و بالینی اهمیت فراوانی دارد. در این راستا، پژوهش حاضر با استفاده از چارچوب بیزی ارائه‌شده توسط پارک و همکاران [۱۹]، عملکرد این روش را در داده‌های شبیه‌سازی‌شده و داده‌های واقعی بیماران سوختگی بررسی می‌کند. تا آنجا که بررسی ادبیات نشان می‌دهد، کاربرد این چارچوب در تحلیل داده‌های سوختگی تاکنون گزارش نشده است. از این‌رو، سهم اصلی پژوهش حاضر در ارزیابی عملی این روش در یک مسئله واقعی پزشکی و بررسی قابلیت آن در شناسایی عوامل مؤثر بر پیامد بیماران سوختگی است.

۲ مدل‌های یادگیری چندنمونه‌ای و مدل بیزی مبتنی بر جستجوی شاتگان

مدل مورد استفاده در این پژوهش بر پایه چارچوب رگرسیون چندنمونه‌ای بیزی ارائه‌شده توسط پارک و همکاران [۱۹] است. این چارچوب امکان انتخاب همزمان نمونه‌های کلیدی در هر بسته و متغیرهای مؤثر در مدل رگرسیونی را فراهم می‌کند و برای جستجو در فضای بزرگ مدل‌ها از الگوریتم جستجوی تصادفی شاتگان استفاده می‌کند. در پژوهش حاضر، این چارچوب برای تحلیل داده‌های بیماران سوختگی و ارزیابی عملکرد آن در یک مسئله واقعی پزشکی به‌کار گرفته شده است. در ادامه، اجزای اصلی مدل و نحوه عملکرد آن تشریح می‌شود.

۱.۲ یادگیری چند نمونه‌ای (MIL)

در یادگیری چندنمونه‌ای، برخلاف یادگیری نظارت‌شده کلاسیک که هر مشاهده دارای یک برچسب است، داده‌ها به‌صورت «بسته»هایی[۲] از نمونه‌ها سازمان‌دهی می‌شوند. هر بسته شامل مجموعه‌ای از نمونه‌ها است که هر نمونه دارای یک بردار ویژگی $x_{ij} \in \mathbb{R}^p$ می‌باشد، اما تنها یک پاسخ Y_i در سطح بسته مشاهده می‌شود. بنابراین ساختار داده‌ها به‌صورت زیر است:

$$\mathcal{D} = \{(Y_i, X_i) : i = ۱, \dots, n\}, \qquad X_i = \{x_{i۱}, \dots, x_{im_i}\}.$$

هدف اصلی در مدل MIL، مدل‌سازی رابطه بین پاسخ سطح بسته و مجموعه نمونه‌های درون آن است (هِرِرا و همکاران [۹]). این مسئله درکاربردهایی مانند شیمی‌محاسباتی، تشخیص سرطان، تحلیل تصاویر پزشکی و داده‌های طولی تکرارشونده بسیار رایج است (آمورِس [۱]؛ ژو و ژانگ[۲۵]؛ وانگ و همکاران [۲۳]).

۲.۲ رگرسیون چندنمونه‌ای[۳]

در چارچوب MIR، پاسخ Y_i یک متغیر پیوسته است و هدف، مدل‌سازی رابطه آن با نمونه‌های درون بسته است. یکی از ساده‌ترین فرم‌ها، مدل خطی تجمیعی است که به‌صورت زیر تعریف می‌شود:

$$Y_i = \beta_۰ + \sum_{j=۱}^{m_i} w_{ij} x_{ij}^\top \beta + \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{N}(۰, \sigma^۲),$$

که در آن w_{ij} وزن نمونه j در بسته i است. انتخاب یا برآورد این وزن‌ها، تفاوت اصلی میان روش‌های مختلف MIR را ایجاد می‌کند.

مدل وزن‌دهی احتمالی افزوده[۴] (روش وزن‌دهی احتمالاتی نمونه‌ها) توسط پاپاس [۱۸] معرفی شد، یکی از روش‌های مهم در MIR است. این مدل از یک رویکرد دو مرحله‌ای استفاده می‌کند که مراحل آن به فرم زیر است:

مرحله اول: برآورد احتمال «نمونه اصلی» بودن. برای هر نمونه، یک مدل لوژستیک برازش داده می‌شود یعنی:

$$\Pr(\delta_{ij} = ۱ \mid x_{ij}) = \frac{۱}{۱ + \exp\big(-(\eta_۰ + \eta^\top x_{ij})\big)}.$$

در اینجا δ_{ij} نشان می‌دهد که آیا نمونه j در بسته i در تولید پاسخ نقش دارد یا خیر.

مرحله دوم: ساخت نمایش سطح بسته. وزن نمونه‌ها برابر با احتمال پیش‌بینی‌شده مرحله اول است:

$$w_{ij} = \Pr(\delta_{ij} = ۱ \mid x_{ij}).$$

سپس ویژگی‌های سطح بسته به‌صورت زیر ساخته می‌شوند:

$$Z_i = \sum_{j=۱}^{m_i} w_{ij} x_{ij}.$$

[2] Bags
[3] Multiple Instance Regression (MIR)
[4] Augmented Probability Weighting (APW)

در نهایت مدل رگرسیون خطی کلاسیک برازش داده می‌شود:

$$Y_i = \alpha + Z_i^\top \beta + \varepsilon_i,\ i = ۱, \cdots, n.$$

محدودیت‌های مدل APW

- انتخاب متغیرها انجام نمی‌شود.
- عدم‌قطعیت در انتخاب نمونه‌ها لحاظ نمی‌شود.
- ساختار سلسله‌مراتبی داده‌ها به‌طور کامل مدل‌سازی نمی‌شود.

این محدودیت‌ها انگیزه توسعه چارچوب رگرسیون چندنمونه‌ای بیزی ارائه‌شده توسط پارک و همکاران [۱۹] را فراهم کرده است؛ چارچوبی که دو مسئله اساسی را به‌طور همزمان مورد توجه قرار می‌دهد:

۱. انتخاب نمونه‌های اصلی در هر بسته،

۲. انتخاب متغیرهای مؤثر در مدل رگرسیون.

این مدل بر پایه یک ساختار دو سطحی بنا شده است: سطح نمونه و سطح بسته.

سطح بسته: مدل رگرسیون بیزی

برای بسته i، مدل پاسخ به‌صورت زیر تعریف می‌شود:

$$Y_i \mid \beta_۰, \beta, \delta_i, \sigma^۲ \sim \mathcal{N}\left(\beta_۰ + \sum_{j=۱}^{m_i} \delta_{ij} x_{ij}^\top \beta,\ \sigma^۲\right).$$

پیشین اسپایک‌ـاسلب[5] برای انتخاب متغیرها: در انتخاب متغیرهای بیزی، پیشین اِسپایک‌ـاِسلَب یکی از روش‌های کلاسیک و مهم است. این پیشین ساختاری دوگانه دارد: مؤلفه اِسپایک ضرایب را به صفر نزدیک می‌کند و متغیرهای غیرضروری را حذف می‌سازد؛ مؤلفه اِسلَب دامنه‌ای پهن‌تر دارد و اجازه می‌دهد ضرایب مهم آزادانه برآورد شوند. این ترکیب هم تنک‌سازی[6] را ممکن می‌کند و هم تفسیرپذیری و کارایی مدل را افزایش می‌دهد (میتچل و بائوچمپ [۱۶]؛ جورج و مک‌کلاچ [۷]؛ ایشواران و رائو [۱۲]).

برای هر ضریب β_k داریم:

$$\beta_k \mid \gamma_k \sim \gamma_k \mathcal{N}(۰, \tau_{\text{in}}^۲) + (۱ - \gamma_k)\mathcal{N}(۰, \tau_{\text{out}}^۲),$$

که در آن $\gamma_k \in \{۰, ۱\}$ شاخص شمول متغیر k است. نسبت واریانس‌ها به گونه‌ای انتخاب می‌شود که

$$\tau_{\text{in}}^۲ \gg \tau_{\text{out}}^۲,$$

تا مؤلفه اِسپایک (با واریانس کوچک) ضریب را به صفر نزدیک کند و مؤلفه اِسلَب (با واریانس بزرگ) آزادی لازم برای ضرایب مهم را فراهم آورد. همچنین:

$$\gamma_k \sim \text{Ber}(w_k),$$

که در آن w_k احتمال پیشینی حضور متغیر k در مدل است. در نسخه‌های سلسله‌مراتبی‌تر می‌توان برای w_k یک توزیع پیشین بتا به صورت

$$w_k \sim \text{Beta}(a, b)$$

در نظر گرفت تا درجه تنک‌سازی به‌صورت داده‌محور تنظیم شود. معمولا این توزیع پیشین به یکی از سه توزیعی که در ادامه ارائه می‌شود، در نظر گرفته می‌شود:

- اسپایک نقطه‌ای: $\beta_k \mid \gamma_k = ۰ \sim \delta_۰$، یعنی جرم نقطه‌ای در صفر.
- اسپایک نرمال با واریانس کوچک: $\mathcal{N}(۰, \tau_{\text{out}}^۲)$.
- اسلَب پهن: معمولاً نرمال با واریانس بزرگ $\mathcal{N}(۰, \tau_{\text{in}}^۲)$، یا توزیع‌های دم‌سنگین مانند کوشی و لاپلاس برای مقاومت در برابر داده‌های پرت.

این پیشین با ترکیب در مدل احتمال، توزیع پسین ضرایب و شاخص‌های شمول را می‌سازد. نتیجه آن است که برای هر متغیر می‌توان احتمال شمول پسین محاسبه کرد:

$$\text{PIP}_k = \Pr(\gamma_k = ۱ \mid y),$$

که معیار مستقیمی از اهمیت متغیرها در مدل است.

سطح نمونه: مدل لوژستیک برای انتخاب نمونه‌ها

در مدل بیزی MIR-SSS، هدف در سطح نمونه آن است که مشخص شود کدام نمونه‌ها در هر بسته واقعاً در تولید پاسخ Y_i نقش دارند. برای این منظور، برای هر نمونه j در بسته i یک متغیر نهفته دودویی به‌صورت زیر تعریف می‌شود:

$$\delta_{ij} = \begin{cases} ۱, & \text{اگر } j \text{ یک نمونه اصلی باشد}, \\ ۰, & \text{در غیر این صورت}. \end{cases}$$

برای مدل‌سازی احتمال نمونه اصلی بودن، یک مدل رگرسیون لوژستیک در سطح نمونه تعریف می‌شود. به‌طور دقیق:

$$\delta_{ij} \mid \eta_۰, \eta \sim \text{Ber}(\pi_{ij}), \qquad \pi_{ij} = \frac{\exp(\eta_۰ + \eta^\top x_{ij})}{۱ + \exp(\eta_۰ + \eta^\top x_{ij})}.$$

[4] Spike-Slab
[6] Sparsity

در این رابطه:

- δ_{ij}: متغیر دودویی نشان‌دهنده انتخاب یا عدم‌انتخاب نمونه j در بسته i.
- $x_{ij} \in \mathbb{R}^p$: بردار ویژگی‌های نمونه j در بسته i.
- η_0: عرض از مبدأ مدل لوژستیک در سطح نمونه.
- $\eta = (\eta_1, \dots, \eta_p)^\top$: ضرایب رگرسیون لوژستیک که اثر هر ویژگی را بر احتمال انتخاب نمونه تعیین می‌کنند.
- π_{ij}: احتمال اینکه نمونه j در بسته i یک «نمونه اصلی» باشد.

به بیان دیگر، مدل لوژستیک نقش هر نمونه را بر اساس ویژگی‌های آن تعیین می‌کند. اگر مقدار $\eta_0 + \eta^\top x_{ij}$ بزرگ باشد، آنگاه π_{ij} به ۱ نزدیک می‌شود و نمونه مربوطه با احتمال بزرگ‌تر به‌عنوان نمونه اصلی انتخاب می‌شود. برعکس، اگر این مقدار کوچک باشد، احتمال انتخاب نمونه کاهش می‌یابد.

این ساختار دو مزیت مهم دارد:

۱. **تفسیرپذیری**: ضرایب η نشان می‌دهند کدام ویژگی‌ها احتمال «نمونه اصلی» بودن را افزایش یا کاهش می‌دهند.

۲. **انعطاف‌پذیری**: انتخاب نمونه‌ها به‌صورت احتمالاتی انجام می‌شود و عدم‌قطعیت در انتخاب نمونه‌ها به‌طور کامل در چارچوب بیزی لحاظ می‌گردد.

در نهایت، این مدل لوژستیک در کنار پیشین‌های بیزی و الگوریتم SSS، امکان انتخاب همزمان نمونه‌ها و متغیرها را فراهم می‌کند و یکی از اجزای کلیدی مدل MIR-SSS محسوب می‌شود.

۳.۲ جستجوی تصادفی شاتگان (SSS)

الگوریتم جستجوی تصادفی شاتگان یکی از روش‌های کارآمد برای جستجو در فضاهای مدل با ابعاد بالا است. ایده اصلی این الگوریتم، بررسی همزمان مجموعه‌ای از مدل‌های همسایه در هر مرحله جستجو و انتخاب مدل بعدی بر اساس شواهد آماری موجود در داده‌ها است. برخلاف بسیاری از روش‌های متداول مبتنی بر زنجیره مارکوف مونت‌کارلو که در هر تکرار تنها یک حرکت محلی انجام می‌دهند، الگوریتم شاتگان چندین مدل رقیب را به‌صورت همزمان ارزیابی کرده و احتمال حرکت به هر مدل را متناسب با مطلوبیت آن تعیین می‌کند. این ویژگی موجب افزایش سرعت پیمایش فضای مدل و بهبود شناسایی نواحی با احتمال پسین بالا می‌شود (پارک و همکاران [۱۹]).

در مدل MIR-SSS، انتخاب متغیرها و نمونه‌ها از طریق متغیرهای دودویی نهفته انجام می‌شود. برای متغیرها، یک بردار انتخاب داریم:

$$\gamma = (\gamma_1, \dots, \gamma_p)^\top \in \{0,1\}^p,$$

که در آن $\gamma_k = 1$ به‌معنای «فعال بودن» متغیر k در مدل رگرسیون سطح بسته و $\gamma_k = 0$ به‌معنای حذف آن متغیر است. به‌طور مشابه، برای نمونه‌ها در بسته i یک بردار انتخاب داریم:

$$\delta_i = (\delta_{i1}, \dots, \delta_{im_i})^\top \in \{0,1\}^{m_i},$$

که در آن $\delta_{ij} = 1$ نشان می‌دهد نمونه j در بسته i به‌عنوان «نمونه اصلی» در مدل سطح بسته مشارکت دارد.

فضای تمام مدل‌های ممکن، ضرب دکارتی این فضاهای گسسته است و به‌شدت پرحجم می‌شود:

$$\mathcal{M} = \{0,1\}^p \times \prod_{i=1}^{n} \{0,1\}^{m_i},$$

به‌طوری‌که پیمایش مستقیم این فضا (مثلاً با جستجوی سراسری) عملاً غیرممکن است. برای جستجوی کارآمد در این فضای گسسته عظیم، الگوریتم SSS به‌کار گرفته می‌شود (پارک و همکاران [۱۹]).

۱.۳.۲ تعریف تابع امتیاز $Q(\cdot)$ و پسین حاشیه‌ای

در چارچوب حاضر، الگوریتم SSS یک مدل آماری مستقل محسوب نمی‌شود، بلکه یک ابزار محاسباتی برای جستجو در فضای مدل‌های بیزی است. معیار مقایسه مدل‌ها در این الگوریتم بر پایه توزیع‌های پسین تعریف می‌شود و تابع امتیاز $Q(\cdot)$ نمایش‌دهنده پسین حاشیه‌ای مدل‌ها (تا یک ثابت نرمال‌سازی) است. بنابراین، انتخاب متغیرها و نمونه‌ها در تمامی مراحل جستجو تحت چارچوب استنباط بیزی انجام می‌شود و الگوریتم SSS نقش موتور محاسباتی انتخاب مدل را ایفا می‌کند. در SSS، برای مقایسه مدل‌های مختلف لازم است به هر مدل یک امتیاز اختصاص داده شود که متناسب با احتمال پسین آن مدل باشد. برای این منظور، ابتدا فرض کنید θ مجموعه سایر پارامترهای مدل (از جمله ضرایب رگرسیون، واریانس خطا، ضرایب رگرسیون لوژستیک و سایر پارامترهای کمکی) را نشان دهد:

$$\theta = (\beta_0, \beta, \sigma^2, \eta_0, \eta, \dots).$$

اگر Y بردار پاسخ‌ها و X و $\{x_{ij}\}$ مجموعه کامل متغیرهای توضیحی در سطح بسته و نمونه باشند، آنگاه پسین کامل به‌صورت

زیر است:

$$\pi(\gamma,\{\delta_i\},\theta \mid Y,X) \propto L(Y \mid X,\gamma,\{\delta_i\},\theta)\cdot$$
$$\pi(\theta \mid \gamma,\{\delta_i\})\ \pi(\gamma)\ \prod_{i=1}^{n}\pi(\delta_i),$$

که در آن:

- $L(Y \mid X,\gamma,\{\delta_i\},\theta)$: تابع درستنمایی داده‌ها تحت مدل معین‌شده توسط $(\gamma,\{\delta_i\})$ و پارامترها θ،
- $\pi(\theta \mid \gamma,\{\delta_i\})$: پیشین پارامترها (مثلاً پیشین اِسپایک–اِسلِب برای β و پیشین‌های نرمال برای η)،
- $\pi(\gamma)$: پیشین روی بردار انتخاب متغیرها (مثلاً ضربی از توزیع‌های برنولی)،
- $\pi(\delta_i)$: پیشین روی بردار انتخاب نمونه‌ها در هر بسته (که می‌تواند از مدل لوژستیک ناشی شود).

برای مقایسه مدل‌های مختلف، تابع امتیاز $Q(\cdot)$ به‌صورت پسین حاشیه‌ای (تا یک ثابت نرمال‌سازی مشترک) روی θ تعریف می‌شود:

$$Q(\gamma,\{\delta_i\}) = \int L(Y \mid X,\gamma,\{\delta_i\},\theta)$$
$$\cdot\ \pi(\theta \mid \gamma,\{\delta_i\})d\theta\times$$
$$\pi(\gamma)\prod_{i=1}^{n}\pi(\delta_i).$$

در عمل، این انتگرال معمولاً به‌صورت بسته قابل‌محاسبه نیست و یکی از روش‌های زیر به‌کار می‌رود:

- استفاده از شکل‌های هم‌خانواده پیشین–درستنمایی برای بخشی از پارامترها (مثلاً β_0,β) و انتگرال‌گیری تحلیلی،
- تقریب‌های لاپلاس یا تقریب مبتنی بر BIC،
- استفاده از مقدار تخمینی تابع پسین در نقطه برآورد بیشینه پسین MAP.

در الگوریتم SSS، آنچه نیاز داریم، مقادیر $Q(\cdot)$ برای مدل‌های موجود در همسایگی یک مدل معین است، نه مقدار نرمال‌سازی‌شده کامل. بنابراین، کافی است $Q(\cdot)$ را تا یک ضریب ثابت مشترک بین مدل‌های همسایه محاسبه کنیم.

۲.۳.۲ همسایگی مدل و نوع حرکت‌ها

برای تعریف حرکت روی فضای مدل‌ها، یک همسایگی برای هر مدل تعریف می‌شود. برای مثال، در مورد بردار γ:

$$\gamma \in \{0,1\}^p, \qquad N(\gamma) = N^{+}(\gamma)\cup N^{-}(\gamma)\cup N^{\text{swap}}(\gamma),$$

که در آن:

- $N^{+}(\gamma)$: مجموعه مدل‌هایی که با γ فقط در یک مؤلفه تفاوت دارند و آن مؤلفه از ۰ به ۱ تغییر یافته است (حرکت افزودن متغیر).
- $N^{-}(\gamma)$: مجموعه مدل‌هایی که در آن‌ها یک مؤلفه از ۱ به ۰ تغییر کرده است (حرکت حذف متغیر).
- $N^{\text{swap}}(\gamma)$: مجموعه مدل‌هایی که در آن‌ها وضعیت دو مؤلفه (یکی صفر و دیگری یک) با هم جابه‌جا شده است (حرکت جابه‌جایی).

به‌طور مشابه، برای بردارهای δ_i در هر بسته i می‌توان یک همسایگی $N(\delta_i)$ تعریف کرد که شامل افزودن، حذف یا جابه‌جایی وضعیت انتخاب نمونه‌ها باشد. در پیاده‌سازی پارک و همکاران [۱۹]، برای کاهش هزینه محاسباتی، جستجو معمولاً به افزودن و حذف محدود می‌شود و حرکت‌های جابجایی یا حذف می‌شوند یا به‌صورت محدود استفاده می‌گردند.

۳.۳.۲ انتخاب مدل بعدی در SSS

در هر گام الگوریتم، یک مدل جاری (مثلاً $(\gamma,\{\delta_i\})$) در اختیار داریم. ابتدا همسایگی مناسب آن را محاسبه می‌کنیم (مثلاً فقط در مؤلفه γ یا فقط در یکی از δ_iها). سپس برای هر مدل همسایه $m' \in N(m)$ که می‌توان آن را به‌طور خلاصه با $m' = (\gamma',\{\delta_i'\})$ نشان داد، مقدار $Q(m')$ را محاسبه یا تقریب می‌کنیم.

احتمال انتقال از مدل m به مدل m' در SSS به‌صورت زیر تعریف می‌شود:

$$T(m \to m') = \frac{Q(m')}{\sum_{m''\in N(m)} Q(m'')}, \qquad m' \in N(m).$$

در چارچوب انتخاب متغیر، اگر فقط روی γ حرکت کنیم، این رابطه به‌صورت زیر نوشته می‌شود:

$$T(\gamma \to \gamma^*) = \frac{Q(\gamma^*,\{\delta_i\})}{\sum_{\gamma'\in N(\gamma)} Q(\gamma',\{\delta_i\})},$$

که در آن:

- γ^*: یک مدل کاندید در همسایگی $N(\gamma)$،
- $Q(\gamma^*,\{\delta_i\})$: امتیاز پسین حاشیه‌ای مدل حاصل از انتخاب متغیرها طبق γ^* و انتخاب نمونه‌ها طبق $\{\delta_i\}$.

به‌طور مشابه، در مرحله انتخاب نمونه‌ها برای یک بسته مشخص i، همسایگی $N(\delta_i)$ تعریف می‌شود و احتمال انتقال به هر $\delta_i^* \in N(\delta_i)$

متناسب با $Q(\gamma, \delta_1, \ldots, \delta_i^*, \ldots, \delta_n)$ تعیین می‌گردد، در حالی که سایر اجزاء مدل ثابت نگه داشته می‌شوند.

۴.۳.۲ ارتباط SSS با MCMC و مزیت‌های آن

الگوریتم SSS را می‌توان به‌عنوان یک روش MCMC روی فضای مدل‌ها در نظر گرفت که در آن:

- به‌جای پیشنهاد یک مدل منفرد در هر گام، مجموعه‌ای از مدل‌های همسایه ارزیابی می‌شوند،
- احتمال انتخاب هر مدل همسایه متناسب با امتیاز پسین حاشیه‌ای آن است،
- در نسخه‌های کامل‌تر، یک گام متروپلیس–هیستینگز[7] نیز می‌تواند روی مدل انتخاب‌شده اعمال شود، اما در بسیاری از پیاده‌سازی‌ها، تعریف مناسب $Q(\cdot)$ و $N(\cdot)$ به‌گونه‌ای است که انتقال‌ها ذاتاً با توزیع هدف سازگار هستند.

مزیت‌های اصلی به‌کارگیری SSS در مدل MIR-SSS عبارت است از:

۱. **پیمایش کارآمد فضای مدل:** به‌جای حرکت‌های بسیار محلی و کند (مانند تغییر تک‌مؤلفه‌ای ساده در نمونه‌گیری گیبز، روی γ)، SSS با مقایسه همزمان مجموعه‌ای از همسایه‌ها، مدل‌های «خوب» بیشتری را در زمان کوتاه‌تر کشف می‌کند.

۲. **توان تفکیک بالاتر در انتخاب:** به‌دلیل استفاده از امتیاز پسین حاشیه‌ای $Q(\cdot)$، فرایند جستجو به‌شدت به شواهد داده‌ای و ساختار پیشین حساس است و مدل‌های با پشتیبانی قوی‌تر را ترجیح می‌دهد.

۳. **سازگاری طبیعی با چارچوب بیزی:** چون $Q(\cdot)$ بر پایه پسین حاشیه‌ای تعریف می‌شود، الگوریتم SSS به‌طور طبیعی با سایر اجزاء MCMC (مانند به‌روزرسانی پارامترها θ) ترکیب می‌شود و در نهایت یک زنجیره واحد در فضای $(\gamma, \{\delta_i\}, \theta)$ ایجاد می‌کند.

در مدل MIR-SSS، SSS نقش موتور جستجو در فضای گسسته انتخاب متغیر و نمونه را ایفا می‌کند؛ تابع $Q(\cdot)$ عملاً پسین حاشیه‌ای مدل را (تا یک ضریب ثابت) کد می‌کند و همسایگی $N(\cdot)$ ساختاری را فراهم می‌کند که میان مدل‌های رقیب با پیچیدگی متفاوت، حرکت‌های معنادار و کارآمد انجام شود.

۳ مطالعه شبیه‌سازی

مطالعه شبیه‌سازی به منظور ارزیابی عملکرد مدل پیشنهادی در شرایط کنترل‌شده طراحی شد. در داده‌های واقعی، نمونه‌ها و متغیرهای مؤثر ناشناخته هستند؛ بنابراین برای بررسی دقت انتخاب نمونه‌ها و متغیرها لازم است داده‌هایی مصنوعی تولید شود که ساختار و مقادیر واقعی پارامترهای آن‌ها مشخص باشد. این رویکرد امکان مقایسه مستقیم مدل پیشنهادی با روش‌های مرجع و همچنین بررسی پایداری آن در شرایط مختلف از نظر میزان تنکی، سطح پراکندگی و حضور داده‌های پرت را فراهم می‌سازد.

در این مطالعه، عملکرد مدل پیشنهادی با دو روش مرجع شامل روش رگرسیون وزنی تجمیعی [8] و روش لاسو (رگرسیون با جریمه ℓ_1) بر روی نمایش سطح بسته [9] مقایسه شد.

۱.۳ طراحی حالت‌ها

برای ارزیابی عملکرد روش‌های مورد بررسی در شرایط مختلف، چهار حالت شبیه‌سازی طراحی شد. در تمامی حالت‌ها تعداد بسته‌ها برابر با ۱۰۰، تعداد متغیرها برابر با ۵۰ و تعداد نمونه‌ها در هر بسته برابر با ۱۰ در نظر گرفته شد. تفاوت حالت‌ها در میزان تنکی مدل، سطح پراکندگی و وجود داده‌های پرت است.

- **حالت اول:** نسبت متغیرهای مؤثر برابر با $K = ۰/۲$ بوده و بنابراین از میان ۵۰ متغیر، تعداد ۱۰ متغیر دارای ضریب غیرصفر هستند. خطای تصادفی از توزیع نرمال با واریانس کم تولید شده و داده پرت وجود ندارد.
- **حالت دوم:** نسبت متغیرهای مؤثر برابر با $K = ۰/۵$ بوده و تعداد ۲۵ متغیر دارای اثر واقعی هستند. همانند حالت اول، داده‌ها بدون داده پرت و با سطح پراکندگی پایین تولید شدند.
- **حالت سوم:** ساختار تنکی مشابه حالت اول $(K = ۰/۲)$ در نظر گرفته شد، اما برای افزایش پیچیدگی مسئله، انحراف معیار خطا به مقدار ۵ افزایش یافت و به ۵% از مشاهدات نیز داده‌های پرت افزوده شد.

[7] Metropolis-Hastings
[8] Aggregated Probability Weighted Regression (APW)
[9] Bag-Level

- **حالت چهارم:** ساختار تنکی مشابه حالت دوم ($K = ۰.۵$) در نظر گرفته شد، با این تفاوت که همانند حالت سوم، داده‌ها در حضور پراکندگی خیلی زیاد ($\sigma = ۵$) و ۵% داده پرت تولید شدند.

پاسخ هر بسته مطابق رابطه زیر تولید شد:

$$Y_i = \beta_۰ + \sum_{j=۱}^{m} \delta_{ij} x_{ij}^{\top} \beta + \varepsilon_i,$$

که در آن δ_{ij} شاخص انتخاب نمونه، β بردار ضرایب رگرسیونی و ε_i جمله خطا است. در حالت‌های اول و دوم فرض شد

$$\varepsilon_i \sim N(۰, ۱),$$

در حالی که در حالت‌های سوم و چهارم برای ایجاد شرایط دشوارتر، خطاها از توزیع

$$\varepsilon_i \sim N(۰, ۲۵)$$

تولید شدند و علاوه بر آن، به ۵٪ از پاسخ‌ها اغتشاش اضافی جهت ایجاد داده‌های پرت افزوده شد.

داده‌های آزمون

برای هر یک از چهار حالت شبیه‌سازی، یک مجموعه داده آزمون مستقل شامل ۵۰۰ بسته تولید شد. استفاده از حجم نمونه بزرگ‌تر در داده‌های آزمون موجب می‌شود ارزیابی عملکرد مدل‌ها از پایداری بیشتری برخوردار بوده و خطای برآورد معیارهای پیش‌بینی کاهش یابد. داده‌های آزمون برای محاسبه خطای پیش‌بینی و همچنین ارزیابی توانایی شناسایی نمونه‌های مؤثر مورد استفاده قرار گرفتند.

مدل‌های مقایسه‌شده

عملکرد مدل پیشنهادی با دو روش مرجع مقایسه شد:

۱. مدل پیشنهادی: رگرسیون چندنمونه‌ای بیزی مبتنی بر الگوریتم جستجوی تصادفی شاتگان (MIR-SSS).

۲. روش مرجع اول: رگرسیون وزنی تجمیعی (APW).

۳. روش مرجع دوم: رگرسیون لاسو بر روی نمایش سطح بسته (لاسو).

معیارهای ارزیابی

برای مقایسه عملکرد روش‌ها از معیارهای زیر استفاده شد:

- میانگین مربع خطای پیش‌بینی (MSE) بر روی داده‌های آزمون.
- تعداد متغیرهای صحیح شناسایی‌شده (Overlap) که بیانگر میزان هم‌پوشانی بین متغیرهای انتخاب‌شده و متغیرهای واقعاً مؤثر است.
- دقت انتخاب متغیرها در میان متغیرهای برتر (Precision@K).
- مساحت زیر منحنی مشخصه عملکرد گیرنده (AUC) برای ارزیابی توانایی مدل پیشنهادی در شناسایی نمونه‌های مؤثر داخل بسته‌ها.

۲.۳ تنظیمات مدل بیزی MIR-SSS و بررسی همگرایی

در پیاده‌سازی مدل بیزی MIR-SSS، برای ضرایب رگرسیونی پیشین اسپایک و اسلب در نظر گرفته شد تا انتخاب متغیرها به‌صورت صریح و در چارچوب بیزی انجام گیرد. پاسخ‌ها با توزیع نرمال مدل شدند و برای واریانس خطا $\sigma^۲$ پیشین اینورس–گاما در نظر گرفته شد. همچنین برای انتخاب نمونه‌های مؤثر در هر بسته، از سازوکار توجه مبتنی بر تابع Softmax استفاده شد تا احتمال مشارکت هر نمونه در تولید پاسخ بسته برآورد شود.

الگوریتم SSS به‌عنوان موتور جستجوی فضای مدل‌ها عمل کرد. در هر تکرار، مجموعه متغیرهای فعال به‌روزرسانی شد، همسایگی مدل جاری مورد ارزیابی قرار گرفت و بر اساس امتیاز پسین مدل‌ها، حرکت در فضای مدل انجام شد. به‌منظور دستیابی به برآوردهای دقیق‌تر پسین، در فواصل زمانی مشخص نمونه‌گیری MCMC در محیط Stan اجرا گردید. در نهایت، احتمال انتخاب پسینی متغیرها و نمونه‌ها و همچنین برآوردهای میانگین‌گیری‌شده مدل گزارش شدند.

برای بررسی همگرایی زنجیره‌های نمونه‌گیری، شاخص‌های استاندارد تشخیصی شامل R-hat و اندازه نمونه مؤثر (ESS) محاسبه شدند. مقادیر R-hat برای تمامی پارامترها نزدیک به ۱ و مقادیر ESS بزرگ‌تر از ۴۰۰ بودند که بیانگر همگرایی مناسب زنجیره‌هاست. علاوه بر این، نمودارهای ردیابی زنجیره‌ها، چگالی‌های پسین و نمودارهای خودهمبستگی نیز برای ارزیابی پایداری نمونه‌گیری مورد استفاده قرار گرفتند.

در شکل‌های ۱ و ۲ مجموعه‌ای از نمودارهای تشخیصی برای دو حالت پایه شبیه‌سازی ارائه شده است. نمودارهای ردیابی زنجیره‌ها نشان می‌دهند که پس از دوره ابتدایی، زنجیره‌های نمونه‌گیری به ناحیه پایدار رسیده‌اند و رفتار نوسانی منظمی پیرامون میانگین پسین دارند. همچنین همپوشانی مناسب چگالی‌های پسین میان زنجیره‌های مختلف، بیانگر اختلاط مطلوب زنجیره‌ها و نبود شواهدی از عدم همگرایی است.

نمودارهای خودهمبستگی نیز کاهش سریع وابستگی نمونه‌ها را در وقفه‌های ابتدایی نشان می‌دهند که کارایی نمونه‌گیری MCMC را تائید می‌کند. علاوه بر این، تمامی مقادیر R-hat کمتر از ۱٫۰۵ بوده‌اند که مطابق معیارهای رایج، نشان‌دهنده همگرایی مناسب زنجیره‌ها است.

از آنجا که هدف اصلی حالت‌های سوم و چهارم بررسی پایداری مدل در شرایط دشوارتر شامل افزایش پراکندگی و وجود داده‌های پرت بود، برای جلوگیری از افزایش حجم مقاله، تنها نتایج عملکردی و منحنی‌های مشخصه گیرنده (ROC) این حالت‌ها گزارش شده‌اند و نمودارهای تشخیصی تکمیلی در صورت نیاز قابل ارائه خواهند بود.

۳.۳ نتایج

جدول ۱ نتایج مقایسه مدل‌های مورد بررسی را در چهار حالت شبیه‌سازی نشان می‌دهد. دو حالت نخست شرایط پایه مطالعه را شامل می‌شوند، در حالی که حالت‌های سوم و چهارم با هدف بررسی پایداری روش‌ها در شرایط دشوارتر شامل افزایش واریانس خطا و وجود داده‌های پرت طراحی شده‌اند.

نتایج جدول ۱ نشان می‌دهد که مدل پیشنهادی در حالت‌های پایه عملکرد بسیار مطلوبی داشته است. در حالت اول، مدل MIR-SSS کمترین خطای پیش‌بینی را به دست آورده و تمامی متغیرهای مؤثر را به درستی شناسایی کرده است. در حالت دوم نیز با وجود افزایش تعداد متغیرهای مؤثر، عملکرد مدل همچنان پایدار باقی مانده و نتایجی مشابه بهترین روش رقیب ارائه کرده است.

در حالت‌های سوم و چهارم، با افزایش واریانس خطا و افزودن داده‌های پرت، عملکرد تمامی روش‌ها با افت مواجه شد. با این حال، مدل MIR-SSS همچنان توانست دقت انتخاب متغیرهای مؤثر را در سطح بالایی حفظ کند. به طور مشخص، دقت انتخاب متغیرها در حالت سوم برابر با ۰٫۹۰ و در حالت چهارم برابر با ۰٫۸۴ بود. این نتایج بیانگر آن است که چارچوب پیشنهادی در شناسایی متغیرهای مؤثر از پایداری مناسبی در برابر افزایش پراکندگی و وجود داده‌های پرت برخوردار است.

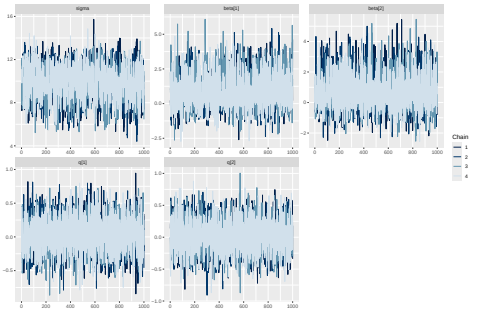
جدول ۱: مقایسه عملکرد مدل‌ها در چهار حالت شبیه‌سازی

حالت	مدل	MSE	همپوشانی	دقت انتخاب متغیرها
اول	لاسو	۱۸۲٫۷۶	۶	۰٫۶۰
اول	APW	۵۲٫۸۳	۱۰	۱٫۰۰
اول	MIR-SSS	۴۸٫۷۵	۱۰	۱٫۰۰
دوم	لاسو	۴۳۶٫۶۱	۱۶	۰٫۶۴
دوم	APW	۱۲۱٫۹۲	۲۴	۰٫۹۶
دوم	MIR-SSS	۱۱۷٫۵۱	۲۴	۰٫۹۶
سوم	لاسو	۱۷۵٫۷۵	۷	۰٫۷۰
سوم	APW	۱۰۱٫۰۵	۹	۰٫۹۰
سوم	MIR-SSS	۱۴۳٫۹۱	۹	۰٫۹۰
چهارم	لاسو	۴۱۶٫۱۱	۱۷	۰٫۶۸
چهارم	APW	۱۶۳٫۲۱	۲۱	۰٫۸۴
چهارم	MIR-SSS	۲۸۷٫۴۲	۲۱	۰٫۸۴

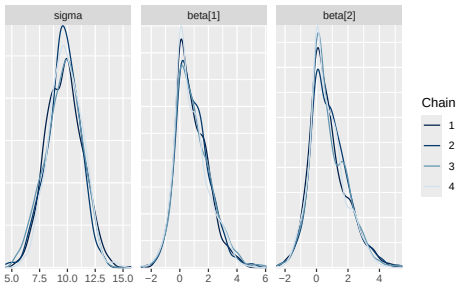
جدول ۲ میانگین و انحراف معیار برآوردهای پسین مربوط به پارامترهای انتخاب نمونه و انتخاب متغیر را در مدل MIR-SSS نشان می‌دهد.

جدول ۲: برآورد میانگین و انحراف معیار پارامترها در مدل MIR-SSS

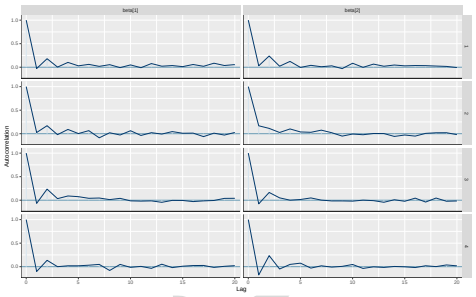
حالت	پارامتر	میانگین پسین	انحراف معیار
اول	δ	۰٫۴۶	۰٫۳۲
اول	γ	۰٫۸۰	۰٫۳۱
دوم	δ	۰٫۵۳	۰٫۳۴
دوم	γ	۰٫۸۷	۰٫۲۶
سوم	δ	۰٫۵۳	۰٫۳۳
سوم	γ	۰٫۷۲	۰٫۱۸
چهارم	δ	۰٫۵۸	۰٫۳۳
چهارم	γ	۰٫۹۳	۰٫۱۸



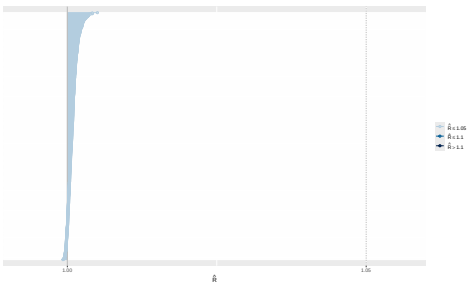
(الف): زنجیره‌های نمونه‌گیری MCMC برای σ، β_1، β_2 و q_1, q_2.



(ب): چگالی پسین پارامترهای σ، β_1 و β_2.

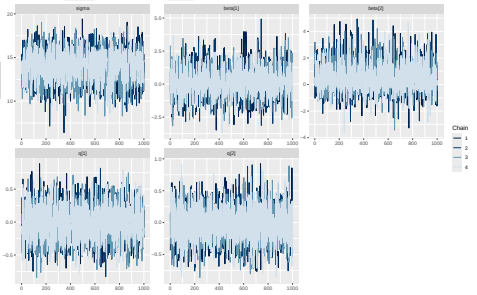


(ج): نمودار ACF.

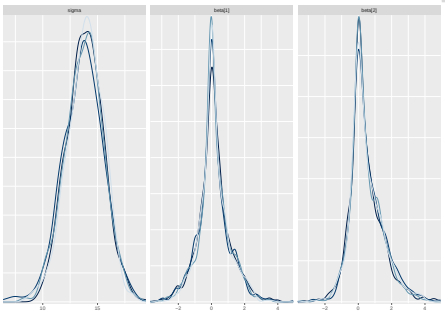


(د): مقادیر R-hat.

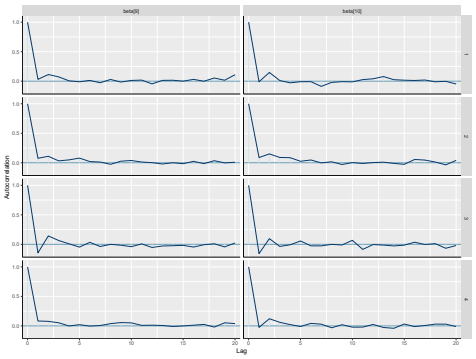
شکل ۱: مجموعه نمودارهای تشخیصی برای مدل MIR-SSS در حالت اول.



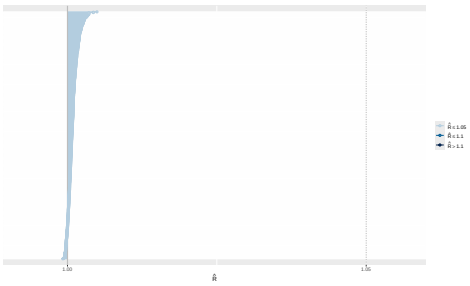
(الف): زنجیره‌های نمونه‌گیری MCMC برای σ، β_1، β_2 و q_1, q_2.



(ب): چگالی پسین پارامترهای σ، β_9 و β_{10}.

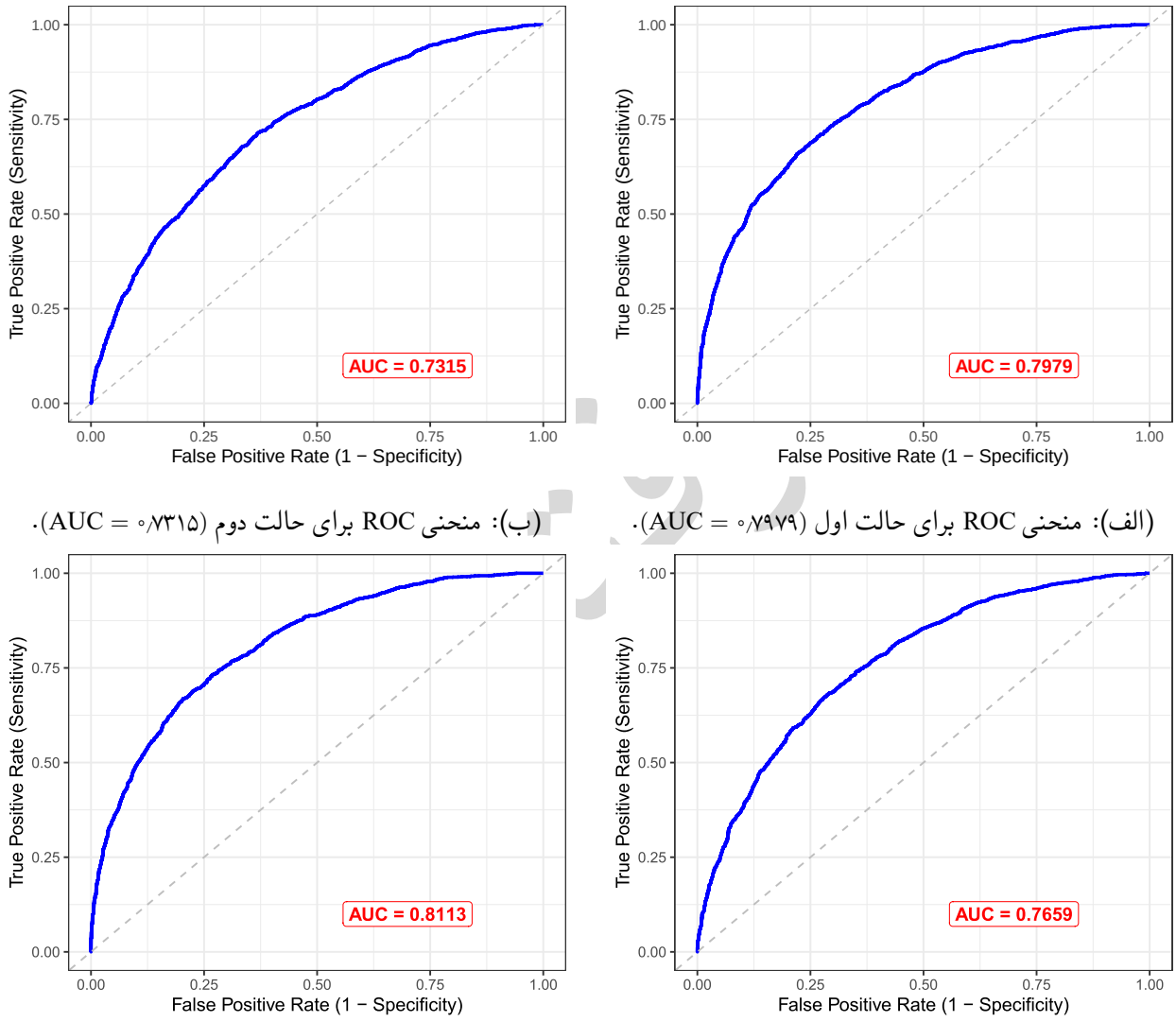


(ج): نمودار ACF.



(د): مقادیر R-hat.

شکل ۲: مجموعه نمودارهای تشخیصی برای مدل MIR-SSS در حالت دوم.



(الف): منحنی ROC برای حالت اول (AUC = ۰٫۷۹۷۹).

(ب): منحنی ROC برای حالت دوم (AUC = ۰٫۷۳۱۵).

(ج): منحنی ROC برای حالت سوم (AUC = ۰٫۷۶۵۹).

(د): منحنی ROC برای حالت چهارم (AUC = ۰٫۸۱۱۳).

شکل ۳: منحنی‌های ROC برای ارزیابی انتخاب نمونه‌ها در چهار حالت شبیه‌سازی.

نتایج جدول ۲ نشان می‌دهد که احتمال‌های انتخاب پسینی مربوط به متغیرها (γ) در تمامی حالت‌ها مقادیر نسبتاً بزرگی داشته‌اند و در مقایسه با احتمال‌های انتخاب نمونه‌ها (δ) از پایداری بیشتری برخوردار بوده‌اند. این موضوع نشان می‌دهد که شناسایی متغیرهای مؤثر نسبت به شناسایی نمونه‌های مؤثر با عدم قطعیت کمتری همراه است.

شکل ۳ عملکرد مدل پیشنهادی را در شناسایی نمونه‌های مؤثر نشان می‌دهد. مقادیر AUC در تمامی حالت‌ها بزرگ‌تر از ۰/۷۳ هستند که بیانگر توانایی مناسب مدل در تفکیک نمونه‌های مؤثر از نمونه‌های غیرمؤثر است. در حالت‌های سوم و چهارم، علی‌رغم افزایش پراکندگی و وجود داده‌های پرت، مقادیر AUC همچنان در محدوده قابل قبول باقی مانده‌اند و حتی در حالت چهارم به مقدار ۰/۸۱۱۳ رسیده‌اند که نشان‌دهنده پایداری مناسب روش پیشنهادی در شرایط دشوارتر است.

به طور کلی، نتایج مطالعه شبیه‌سازی نشان می‌دهد که مدل MIR-SSS قادر است انتخاب متغیرها و انتخاب نمونه‌ها را به صورت همزمان و با دقت مطلوب انجام دهد. همچنین عملکرد مدل در حالت‌های دارای پراکندگی بالا و داده‌های پرت نشان داد که چارچوب پیشنهادی از پایداری مناسبی برخوردار است و می‌تواند در مسائل واقعی با ساختارهای پیچیده نیز مورد استفاده قرار گیرد.

۴ داده‌های واقعی

داده‌های مورد استفاده مربوط به بیماران بستری در بیمارستان سوانح سوختگی امیرالمؤمنین (ع) شهر شیراز است. این بیمارستان بزرگ‌ترین و مجهزترین مرکز تخصصی درمان بیماران دچار سوختگی در جنوب کشور محسوب می‌شود. داده‌های مورد بررسی حاصل از یک سرشماری کامل بیماران بستری شده در سال ۱۴۰۰ هجری شمسی می‌باشد.

مجموعه داده شامل ۲۰۴۲ بیمار است که اطلاعات دموگرافیک، بالینی و آزمایشگاهی آن‌ها در طول دوره بستری ثبت گردیده است. متغیرهای موجود در این مجموعه داده ابعاد مختلفی از وضعیت بیماران را پوشش می‌دهند؛ از جمله مدت زمان بستری و پیامد نهایی (مرگ یا ترخیص)، سن و جنسیت، درصد سطح سوختگی، علت سوختگی، و شاخص‌های بیوشیمیایی خون. نکته مهم آن است که متغیرهای آزمایشگاهی (مانند آلبومین، اوره نیتروژن، کراتینین، سدیم، پتاسیم و شمارش گلبول‌های سفید) برای هر بیمار در سه نوبت مختلف طی دوره بستری اندازه‌گیری شده‌اند. این تکرار اندازه‌گیری‌ها امکان بررسی تغییرات دینامیکی شاخص‌های آزمایشگاهی در طول زمان بستری و ارتباط آن‌ها با پیامد نهایی بیماران را فراهم می‌سازد.

جدول ۴ شامل متغیر پاسخ و متغیرهای توضیحی مورد استفاده در تحلیل داده‌های بیماران سوختگی است.

متغیر پاسخ در این مطالعه، مدت زمان بستری بیمار (Time) در بیمارستان می‌باشد.

شاخص‌های ارزیابی که برای داده های واقعی این مطالعه به‌کار گرفته شده است، شامل میانگین مربعات خطا برای داده‌های آزمون (مقدار MSE داده‌های آزمون)، شیب منحنی کالیبراسیون: میزان انطباق تغییرات پیش‌بینی با تغییرات واقعی و همبستگی رتبه‌ای اسپیرمن: رابطه یکنواخت بین پیش‌بینی‌ها و مقادیر مشاهده شده براساس رتبه مقادیر، می‌باشد.

همان‌گونه که در جدول ۴ ارائه شده است، متغیر «درصد سوختگی» یک متغیر کیفی و ترتیبی محسوب می‌شود. به منظور سهولت در فرایند مدل‌سازی و با توجه به ماهیت ترتیبی سطوح این متغیر، آن به صورت یک متغیر کمی در مدل‌ها وارد شده است. این متغیر در ادبیات مرتبط با سوختگی با عنوان TBSA[10] شناخته می‌شود.

متغیر دیگری که در این مطالعه ساختار کیفی دارد، دلیل سوختگی (Cause_Burn) است. با توجه به این‌که این متغیر از نوع کیفی و اسمی می‌باشد، لازم است در مدل‌سازی به صورت متغیرهای نشانگر وارد شود. در این راستا، سطح اول این متغیر (سوختگی ناشی از حرارت) به عنوان سطح مرجع در نظر گرفته شده و سایر سطوح به شکل متغیرهای نشانگر در مدل لحاظ گردیده‌اند. نحوه کدگذاری این متغیر در جدول ۳ نمایش داده شده است.

جدول ۳: نحوه کدگذاری متغیر دلیل سوختگی در مدل نهایی (سطح ۱ به عنوان مرجع)

سطح اصلی	توضیح	نام متغیر در مدل
۱	حرارت (مانند آتش‌سوزی‌ها)	سطح مرجع (بدون متغیر مجزا)
۲	جسم داغ (مانند اتو یا قابلمه داغ)	CB
۳	مایعات یا مواد جوشان (مانند آب جوش یا روغن داغ)	SB
۴	مواد شیمیایی (مانند اسید)	CHB
۵	الکتریسیته (برق‌گرفتگی)	EB

[10] Total Body Surface Area

جدول ۴: مشخصات متغیرهای موجود در مجموعه داده

نام متغیر	توضیحات
Time	متغیر پاسخ (پیامد نهایی مطالعه)؛ مدت زمان از بستری شدن بیمار تا ترخیص (واحد: روز).
Age	سن بیمار در زمان مطالعه (واحد: سال)
Sex	جنسیت بیمار با سطوح: ۰: مرد (Male) ۱: زن (Female)
Percent	درصد سطح سوختگی بیمار با سطوح: ۱: ۰ تا ۹% ۲: ۱۰ تا ۱۹% ۳: ۲۰ تا ۲۹% ۴: ۳۰ تا ۳۹% ۵: ۴۰ تا ۴۹% ۶: ۵۰ تا ۵۹% ۷: ۶۰ تا ۶۹% ۸: ۷۰ تا ۷۹% ۹: بالای ۸۰%
Cause_Burn	علت سوختگی با سطوح: ۱: حرارت (مانند آتش‌سوزی‌ها) ۲: جسم داغ (مانند اتو یا قابلمه داغ) ۳: مایعات یا مواد جوشان (مانند آب جوش یا روغن داغ) ۴: مواد شیمیایی (مانند اسید) ۵: الکتریسیته (برق‌گرفتگی)
ALB	میزان آلبومین خون (واحد: گرم بر دسی‌لیتر)، اندازه‌گیری در ۳ نوبت طی بستری
BUN	میزان اوره نیتروژن خون (واحد: میلی‌گرم بر دسی‌لیتر)، اندازه‌گیری در ۳ نوبت طی بستری
CR	میزان کراتینین خون (واحد: میلی‌گرم بر دسی‌لیتر)، اندازه‌گیری در ۳ نوبت طی بستری
WBC	شمارش گلبول‌های سفید خون (واحد: هزار عدد در هر میکرولیتر)، اندازه‌گیری در ۳ نوبت طی بستری
NA	میزان سدیم خون (واحد: میلی‌اکی‌والان بر لیتر)، اندازه‌گیری در ۳ نوبت طی بستری
K	میزان پتاسیم خون (واحد: میلی‌مول بر لیتر)، اندازه‌گیری در ۳ نوبت طی بستری

در فرایند مدل‌سازی، به منظور ارزیابی کارایی و قابلیت تعمیم مدل‌ها، داده‌ها به دو بخش آموزشی و آزمون تقسیم گردیدند. بدین ترتیب، ٪۷۰ از کل داده‌ها برای آموزش مدل‌ها مورد استفاده قرار گرفت و ٪۳۰ باقی‌مانده به عنوان داده‌های آزمون به کار گرفته شد. این تقسیم‌بندی امکان برآورد پارامترهای مدل بر اساس داده‌های آموزشی و سپس سنجش دقت و پایداری مدل بر روی داده‌های مستقل آزمون را فراهم ساخت.

در تمامی تحلیل‌ها از نرم‌افزار R 4.5.1 و بسته cmdstanr برای برازش مدل‌های بیزی استفاده شد [۲۹، ۳۰]. محاسبات بر روی رایانه‌ای مجهز به پردازنده AMD Ryzen نسل هفتم و حافظه اصلی ۱۶ گیگابایت انجام گرفت. از نظر هزینه محاسباتی، مدل‌های لاسو و APW بر روی داده‌های واقعی در مدت کمتر از ۱۰ دقیقه برازش شدند، در حالی که زمان اجرای هر بار برازش کامل مدل MIR-SSS حدود ۱۶ ساعت بود. این افزایش زمان محاسباتی ناشی از جستجو در فضای بزرگ مدل‌ها و انجام استنباط بیزی برای انتخاب همزمان نمونه‌ها و متغیرها است. افزون بر این، مدل MIR-SSS امکان کمی‌سازی عدم‌قطعیت در فرایند انتخاب را نیز فراهم می‌کند که در روش‌های مرجع به‌صورت مستقیم در دسترس نیست. همچنین، به دلیل نیاز به توسعه، تنظیم و ارزیابی چندین پیاده‌سازی مبتنی بر Stan به منظور دستیابی به همگرایی و عملکرد مطلوب، مجموع زمان صرف‌شده برای پیاده‌سازی و تحلیل مدل MIR-SSS در این پژوهش حدود ۵۰۰ ساعت پردازش برآورد شد.

شاخص‌هایی که برای ارزیابی عملکرد مدل‌ها در این بخش به کار رفته‌اند، پیش‌تر بیان شدند، این شاخص‌ها معیارهای استانداردی برای سنجش کیفیت برازش و توان پیش‌بینی مدل‌ها محسوب می‌شوند.

در نمودار ۴ فرایند اجرای مدل MIR-SSS به‌صورت گام‌به‌گام نمایش داده شده است. در گام نخست، برای هر ویژیت احتمال مسئولیت با استفاده از یک مدل لوژستیک محاسبه می‌شود و سپس انتخاب ویژیت‌ها از طریق الگوریتم SSS انجام می‌گیرد. پس از آن، ویژگی‌های ویژیت‌های منتخب برای هر بیمار تجمیع شده و پاسخ بیمار در سطح بیمار با یک مدل بیزی رگرسیونی مدل‌سازی می‌شود. در ادامه، انتخاب متغیرها نیز با بهره‌گیری از الگوریتم SSS صورت می‌پذیرد. خروجی نهایی شامل ویژیت‌های انتخاب‌شده، متغیرهای منتخب بر اساس PIP و مجموعه‌ای از شاخص‌های ارزیابی از جمله MSE، همبستگی اسپیرمن و شیب منحنی کالیبراسیون است.

همان‌طور که در بخش ۲ ذکر شده مدل MIR-SSS دارای دو بخش است، بخش پیاده‌سازی بیزی مدل و بخش جستجوی تصادفی شاتگان، به‌طور گرافیکی پیاده‌سازی مدل برای داده‌های واقعی در دیاگرام ۴ توصیف شده است. برای اطمینان از همگرایی در بخش بیزی مدل و درستی الگوریتم پایه برای زنجیره‌های MCMC چند نمودار آورده شده است.

مجموعه‌ی نمودارهای ماتریسی ارائه‌شده در شکل ۵، سه دسته‌ی اصلی از شاخص‌های تشخیصی الگوریتم MCMC برای ضرایب β_1 تا β_{13} در مدل بیزی است.

در **ردیف اول** (شکل ۵، (الف)، (ب) و (پ))، نمودارهای مسیر زنجیره‌ها نشان می‌دهند که پس از دوره‌ی گرم‌سازی، تمامی زنجیره‌ها به حالت پایدار رسیده‌اند. نوسانات بدون روند مشخص و اختلاط مناسب بین زنجیره‌ها بیانگر همگرایی کامل الگوریتم است. این ویژگی برای ضرایب β_1 تا β_{13} که به متغیرهای بالینی و آزمایشگاهی بیماران (سن، جنسیت، انواع علت سوختگی، درصد سوختگی، آلبومین، اوره، کراتینین، گلبول‌های سفید، سدیم، پتاسیم) مربوط هستند مشاهده می‌شود.

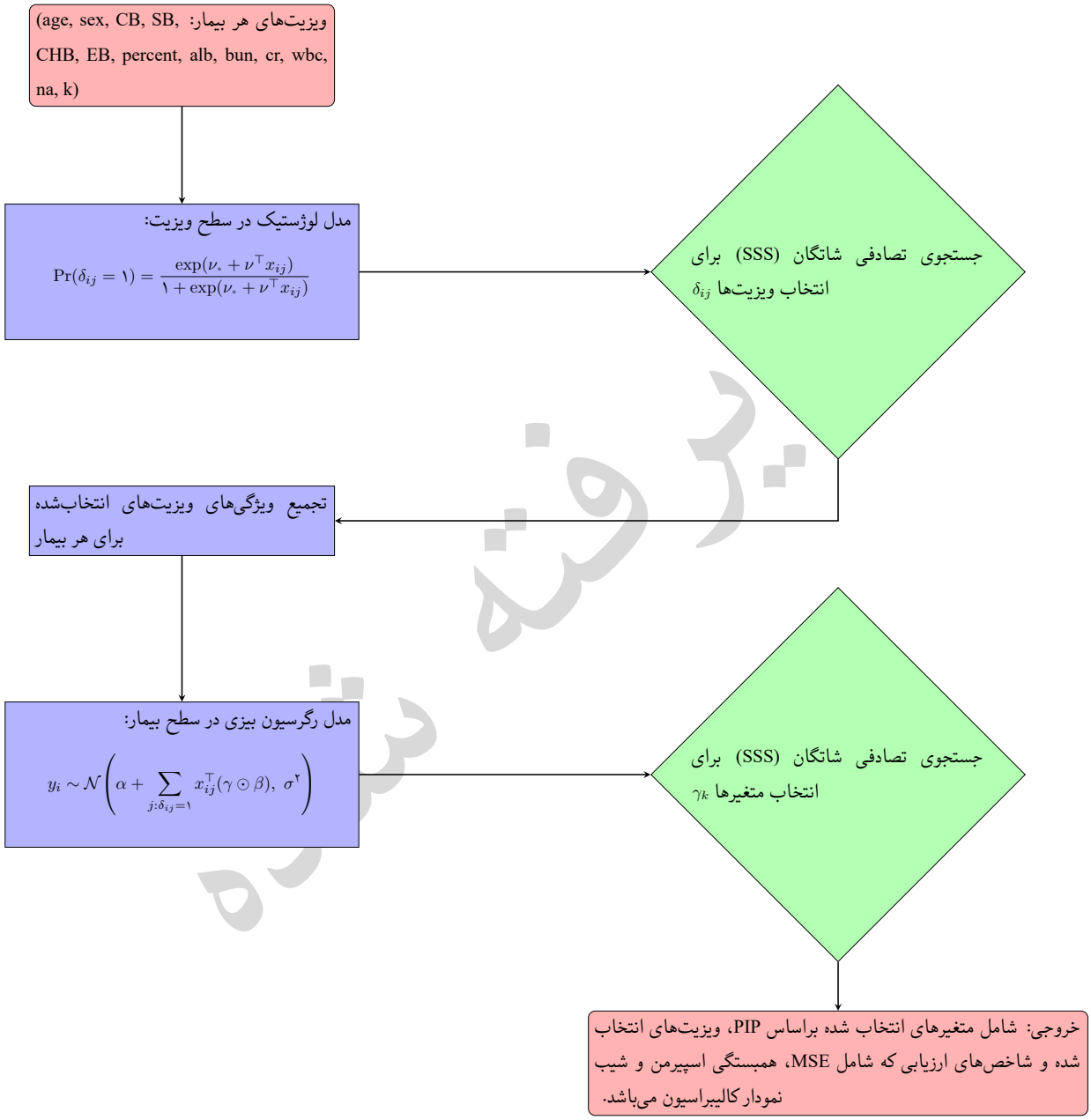
در **ردیف دوم** (شکل ۵، (ت)، (ث) و (ج))، نمودارهای چگالی پسین ضرایب نشان‌دهنده‌ی برآوردهای پایدار و منطبق هستند. همپوشانی مناسب توزیع‌ها و نبود انحراف یا چندمدی بودن، نشان می‌دهد که الگوریتم MCMC توانسته است فضای پارامتر را به خوبی پوشش دهد و تخمین‌های قابل اعتماد تولید کند.

در **ردیف سوم** (شکل ۵، (چ) و (ح))، شاخص‌های $\widehat{R}$ برای تمامی ضرایب کمتر از ۱٫۰۱ هستند. این مقدار نزدیک به ۱ یکی از معیارهای استاندارد برای ارزیابی همگرایی زنجیره‌ها در روش‌های MCMC محسوب می‌شود و نشان‌دهنده‌ی اختلاط مناسب و همگرایی مطلوب است (ژلمن و همکاران[۶]).

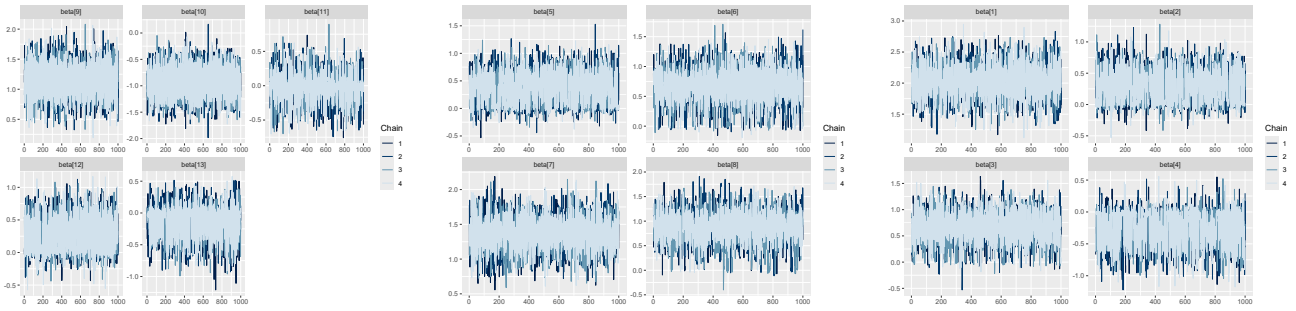
این مجموعه‌ی نمودارها تصویری جامع از پایداری الگوریتم، همگرایی زنجیره‌ها، و اعتبار نتایج بیزی ارائه می‌دهد. بنابراین می‌توان با اطمینان از نتایج استنباطی مدل استفاده کرد و آن‌ها را به عنوان شواهد معتبر در تحلیل داده‌های بالینی بیماران سوختگی به کار گرفت.

۱.۴ نتایج

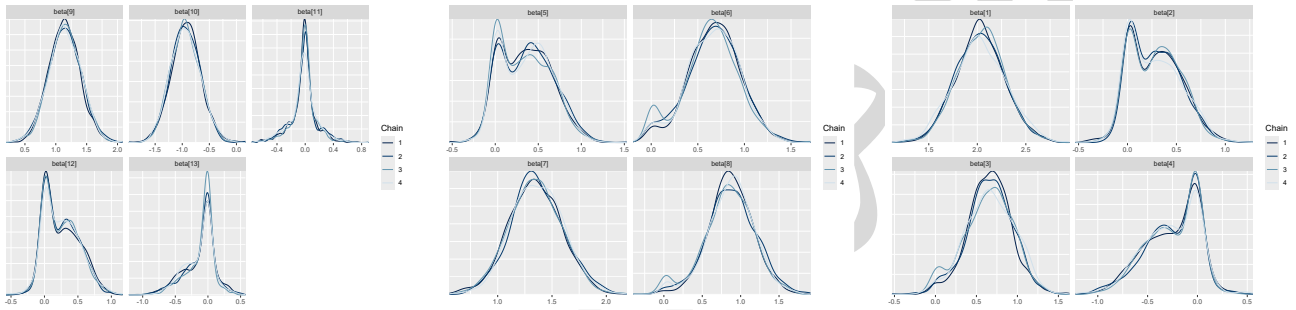
بر اساس نتایج ارائه‌شده در جدول ۵، متغیر «سن» در هر سه مدل لاسو، APW و MIR-SSS دارای ضریب مثبت و معنادار بوده است. این پایداری میان مدل‌های مختلف نشان می‌دهد که اثر سن بر مدت بستری یک رابطه ساختاری و تکرارشونده است و نه وابسته به نوع مدل‌سازی.



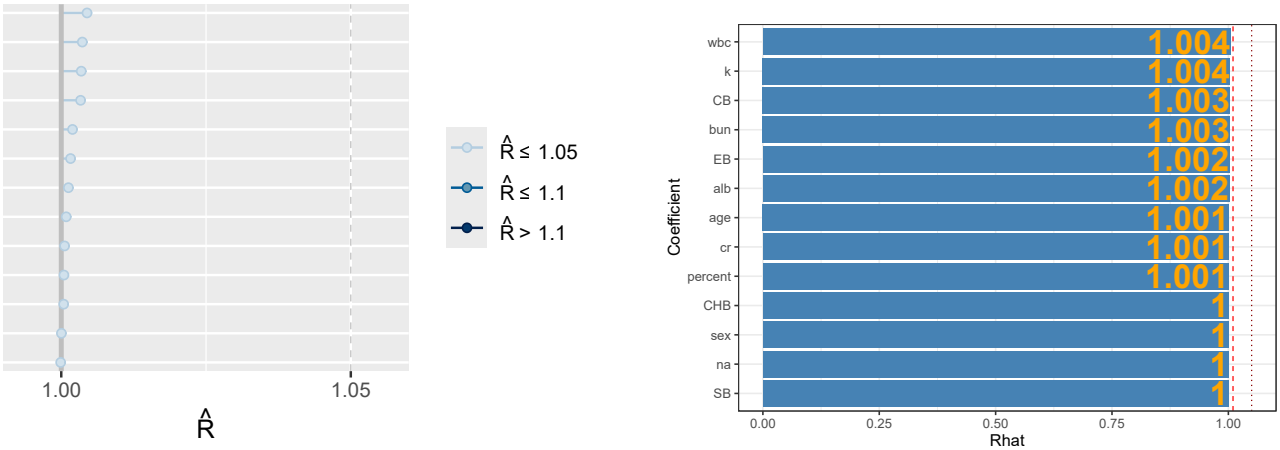
شکل ۴: نمایش الگوریتمی مدل MIR-SSS



(الف): β_1 تا β_4 (ب): β_5 تا β_8 (پ): β_9 تا β_{13}



(ت): چگالی β_1 تا β_4 (ث): چگالی β_5 تا β_8 (ج): چگالی β_9 تا β_{13}



(چ): نمودار میله‌ای مقادیر $\widehat{R}$ برای ضرایب مختلف (ح): نموداری تشخیصی $\widehat{R}$ براساس مقادیر مرزی

شکل ۵: مجموعه‌ی کامل نمودارهای تشخیصی الگوریتم MCMC برای ضرایب β_1 تا β_{13}.

در واقع، افزایش سن بیماران معمولاً با کاهش توان ترمیم بافتی، افزایش ریسک عفونت، و بروز بیماری‌های همراه[11] توأم است که همگی می‌توانند به طولانی‌تر شدن دوره بستری منجر شوند (هرندون [۱۰]). بنابراین، مشاهده ضرایب مثبت و معنادار کاملاً با ادبیات بالینی سازگار است.

شدت سوختگی (درصد سطح سوختگی) نیز در هر سه مدل یکی از قوی‌ترین متغیرهای پیش‌بینی‌کننده بوده است. ضرایب مثبت و معنادار این متغیر در مدل‌های لاسو و APW و همچنین میانگین پسین مثبت آن در مدل MIR-SSS نشان می‌دهد که افزایش سطح سوختگی به‌طور پایدار با افزایش مدت بستری ارتباط دارد. این یافته مطابق مطالعات قبلی است که درصد سوختگی (TBSA) را مهم‌ترین عامل تعیین‌کننده شدت آسیب، نیاز به مراقبت ویژه، و طول دوره ترمیم می‌دانند. (بروسلیرز و همکاران [۳]).

در بخش متغیرهای آزمایشگاهی، الگوی منسجمی دیده می‌شود. متغیر «آلبومین» در هر سه مدل ضریب مثبت و معنادار داشته است. کاهش آلبومین معمولاً با سوءتغذیه، التهاب سیستمیک و اختلال عملکرد کبدی همراه است و با پیامدهای نامطلوب و بستری طولانی‌تر ارتباط دارد (رمضانیان و همکاران [۲۰]). بنابراین، ضرایب مثبت مشاهده‌شده نشان می‌دهند که مقدار کمتر آلبومین (وضعیت تغذیه‌ای ضعیف‌تر) با افزایش مدت بستری مرتبط است.

متغیر «اوره» نیز در هر سه مدل اثر مثبت و معناداری نشان داده است. افزایش اوره بیانگر اختلال عملکرد کلیه، دهیدراتاسیون، یا وضعیت کاتابولیک شدید است که در بیماران سوختگی به‌ویژه در فاز حاد پس از آسیب شایع است (لگارند [۱۵]). بنابراین، اثر مثبت اوره بر افزایش مدت بستری از نظر فیزیولوژیک کاملاً منطقی است.

در مقابل، «کراتینین» در هر سه روش ضریب منفی و معنادار داشته است. این یافته در نگاه اول متناقض به‌نظر می‌رسد؛ با این حال، باید توجه داشت که مقدار کراتینین در بیماران سوختگی شدید به دلیل کاهش توده عضلانی، همولیز، یا افزایش فیلتراسیون گلومرولی ممکن است به‌طور مصنوعی کاهش یابد و بنابراین، لزوماً شاخص دقیقی از عملکرد کلیه نیست (ژو و همکاران [۲۸]). مطالعه امامی و همکاران [۵] اشاره کرده است که «اوره» معمولاً شاخص حساس‌تری برای اختلال کلیه در بیماران سوختگی است و کراتینین ممکن است در مراحل اولیه تغییر معناداری نداشته باشد. بنابراین، یافته‌های این مطالعه با مباحث فیزیولوژیک مرتبط است.

جدول ۵: مقایسه ضرایب سه مدل لاسو، APW و MIR-SSS

متغیر	لاسو			APW			MIR-SSS			
	ضریب	SE	معناداری	ضریب	SE	معناداری	میانگین	SE	%۹۵ CI	معناداری
سن	۲٫۱۷۳۰	۰٫۲۶۰۰	***	۰٫۱۰۷۵	۰٫۰۱۲۹	***	۰٫۰۸۸	۰٫۰۲۰	(۰٫۰۴۳, ۰٫۱۲۱)	**
جنسیت	۰٫۴۱۸۱	۰٫۲۵۱۵		۰٫۸۶۸۳	۰٫۵۱۰۵		۰٫۲۹۸	۰٫۵۶۹	(−۰٫۶۴۱, ۱٫۶۵۸)	
جسم داغ	۰٫۶۸۰۰	۰٫۲۷۴۹	*	۱٫۹۷۶۲	۰٫۷۹۹۸	*	۰٫۶۹۸	۰٫۸۲۰	(−۰٫۱۰۹, ۲٫۵۰۸)	
مایعات جوشان	−۰٫۳۱۹۸	۰٫۲۷۲۳		−۰٫۷۴۱۸	۰٫۶۳۲۰		−۰٫۲۲۷	۰٫۵۲۸	(−۱٫۵۹۲, ۰٫۵۷۷)	
مواد شیمیایی	۰٫۴۵۸۱	۰٫۲۷۹۴		۱٫۵۲۱۷	۰٫۹۶۰۹		−۰٫۰۱۷	۰٫۴۴۶	(−۱٫۰۸۴, ۱٫۰۵۵)	
الکتریسیته	۰٫۷۴۱۳	۰٫۲۷۸۲	**	۲٫۹۸۸۵	۱٫۱۵۰۱	**	۰٫۱۷۲	۰٫۶۰۹	(−۰٫۹۲۷, ۱٫۹۱۹)	
درصد سوختگی	۱٫۴۲۸۶	۰٫۲۵۲۴	***	۰٫۵۷۸۷	۰٫۱۰۳۳	***	۱٫۱۱۰	۱٫۷۱۰	(−۲٫۲۲۸, ۲٫۶۶۸)	
آلبومین	۱٫۰۰۲۵	۰٫۳۰۳۲	***	۰٫۷۲۷۶	۰٫۲۱۸۱	***	۰٫۳۹۲	۰٫۶۰۰	(−۰٫۰۷۵, ۱٫۷۶۲)	
اوره	۱٫۱۷۹۴	۰٫۲۶۴۶	***	۰٫۰۹۵۹	۰٫۰۲۱۷	***	۰٫۰۸۲	۰٫۰۷۰	(۰٫۰۰۱, ۰٫۲۴۳)	*
کراتینین	−۰٫۹۶۶۲	۰٫۲۷۸۰	***	−۲٫۶۱۷۳	۰٫۷۵۶۴	***	−۰٫۸۹۸	۰٫۹۴۷	(−۲٫۹۴۳, ۰٫۰۸۸)	
گلبول سفید	−۰٫۰۲۶۴	۰٫۲۵۱۵		−۰٫۰۰۶۱	۰٫۰۵۲۴		−۰٫۰۰۴	۰٫۰۴۶	(−۰٫۱۰۳, ۰٫۰۷۹)	
سدیم	۰٫۳۵۲۳	۰٫۲۵۵۲		۰٫۰۱۷۱	۰٫۰۱۲۴		۰٫۰۱۸	۰٫۰۳۹	(−۰٫۰۲۴, ۰٫۱۱۰)	
پتاسیم	−۰٫۱۵۸۵	۰٫۲۶۲۳		−۰٫۲۳۵۹	۰٫۴۰۳۸		−۰٫۱۳۸	۰٫۳۸۷	(−۱٫۲۷۳, ۰٫۳۴۳)	

$^{*}p < ۰٫۰۵$؛ $^{**}p < ۰٫۰۱$؛ $^{***}p < ۰٫۰۰۱$.

[11] Comorbidities

در ارتباط با علل سوختگی، نتایج مدل‌ها نشان می‌دهند که «جسم داغ» و «الکتریسیته» اثر مثبت و معناداری در مدل‌های لاسو و APW داشته‌اند و در مدل MIR-SSS نیز میانگین پسین آن‌ها مثبت بوده است، هرچند فاصله اعتباری در حالت الکتریسیته اندکی وسیع است. سوختگی با الکتریسیته معمولاً با آسیب‌های عمقی‌تر، نکروز بافتی، و عوارض سیستمیک مانند آریتمی همراه است که منجر به دوره‌های بستری طولانی‌تر می‌شود (کریم‌زاده و همکاران [۱۳]). همچنین، سوختگی با جسم داغ غالباً از نوع تماس طولانی بوده و درگیری عمقی بیشتری ایجاد می‌کند.

بر اساس جدول ۶ و نمودارهای کالیبراسیون در شکل ۶، می‌توان چند نکته کلیدی را استخراج کرد:

- **دقت پیش‌بینی:** مقدار MSE آزمون برای مدل MIR-SSS برابر با ۸۲٫۱۰ است که کمتر از مقادیر ۸۴٫۴۷ در لاسو و ۸۴٫۵۴ در APW می‌باشد. این برتری در نمودار کالیبراسیون (شکل ۶(ج)) نیز قابل مشاهده است، زیرا نقاط پیش‌بینی‌شده نزدیک‌تر به خط ایده‌آل قرار گرفته‌اند.
- **کالیبراسیون:** شیب کالیبراسیون در مدل MIR-SSS برابر با ۱٫۰۲۰۰ است که نزدیک‌ترین مقدار به شیب ایده‌آل ۱ محسوب می‌شود. در نمودارهای لاسو و APW (شکل‌های ۶(الف) و ۶(ب)) انحراف بیشتری از خط ایده‌آل دیده می‌شود، که با مقادیر شیب ۱٫۰۷۱۰ و ۱٫۰۴۳۳ سازگار است.
- **عرض از مبدأ:** مقدار عرض از مبدأ در مدل MIR-SSS برابر با ۱٫۲۰۰۰− است که نسبت به لاسو (۱٫۶۶۰۲−) و APW (۱٫۱۳۶۱−) کمترین میزان انحراف را نشان می‌دهد. این موضوع نیز در نمودار کالیبراسیون مدل MIR-SSS قابل مشاهده است.
- **همبستگی رتبه‌ای:** ضریب اسپیرمن در مدل MIR-SSS برابر با ۰٫۴۱۰۰ است که بالاتر از مقادیر ۰٫۳۷۴۱ و ۰٫۳۷۲۰ در دو مدل دیگر است. این نشان می‌دهد که مدل MIR-SSS بهتر می‌تواند ترتیب نسبی بیماران از نظر مدت بستری را حفظ کند.

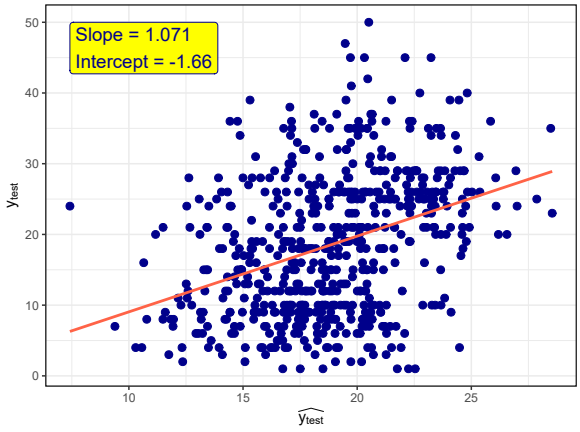
جدول ۶ و نمودارهای کالیبراسیون در شکل ۶ نشان می‌دهند که مدل MIR-SSS از نظر دقت پیش‌بینی، کیفیت کالیبراسیون و همبستگی رتبه‌ای عملکرد برتری نسبت به مدل‌های لاسو و APW دارد.

جدول ۶: مقایسه معیارهای ارزیابی سه مدل

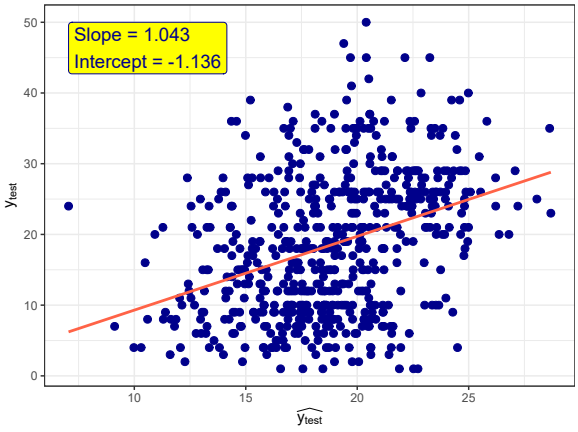
مدل	MSE آزمون	MSE آموزش	شیب کالیبراسیون	عرض از مبدأ	اسپیرمن
لاسو	۸۴٫۴۷	۸۶٫۷۴	۱٫۰۷۱۰	−۱٫۶۶۰۲	۰٫۳۷۴۱
APW	۸۴٫۵۴	۸۶٫۶۹	۱٫۰۴۳۳	−۱٫۱۳۶۱	۰٫۳۷۲۰
MIR-SSS	۸۲٫۱۰	۸۵٫۲۰	۱٫۰۲۰۰	−۱٫۲۰۰۰	۰٫۴۱۰۰

۵ بحث و نتیجه گیری

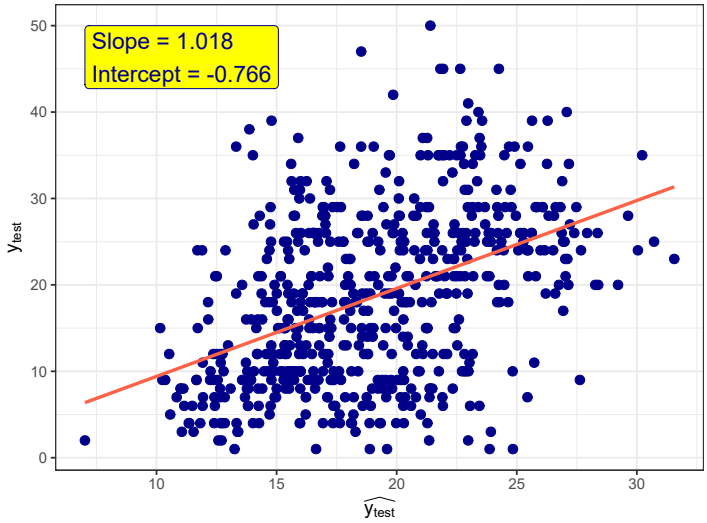
بر اساس نتایج جدول ۶، می‌توان جمع‌بندی دقیقی از عملکرد سه مدل لاسو، APW و MIR-SSS ارائه داد. این جدول نشان می‌دهد که مدل MIR-SSS در تمامی معیارهای کلیدی ارزیابی پیش‌بینی و کالیبراسیون نسبت به دو مدل دیگر برتری دارد. نخست، مقدار میانگین مربعات خطا در داده‌های آزمون برای مدل MIR-SSS برابر با ۷۸٫۱۳۹۹ است که کمتر از مقادیر ۸۴٫۴۷۲۰ در مدل لاسو و ۸۴٫۵۴۲۲ در مدل APW می‌باشد. همچنین در داده‌های آموزش، مقدار MSE برای مدل MIR-SSS برابر با ۷۵٫۹۹۳۲ است که نسبت به مقادیر ۸۶٫۷۳۹۶ و ۸۶٫۶۹۴۴ در دو مدل دیگر کاهش قابل توجهی دارد. این نتایج نشان‌دهنده‌ی دقت بالاتر پیش‌بینی و عدم بیش‌برازش در مدل پیشنهادی است. از منظر کالیبراسیون، شیب برآوردشده در مدل MIR-SSS برابر با ۱٫۰۱۷۷ است که نزدیک‌ترین مقدار به شیب ایده‌آل ۱ محسوب می‌شود. در مقابل، مدل‌های لاسو و APW به ترتیب مقادیر ۱٫۰۷۱۰ و ۱٫۰۴۳۳ را نشان می‌دهند. همچنین عرض از مبدأ در مدل MIR-SSS برابر با ۰٫۷۶۶۲− است که نسبت به مقادیر ۱٫۶۶۰۲− و ۱٫۱۳۶۱− در دو مدل دیگر، نشان‌دهنده‌ی تطابق بهتر بین پیش‌بینی‌ها و مقادیر واقعی می‌باشد. در نهایت، ضریب همبستگی رتبه‌ای اسپیرمن در مدل MIR-SSS برابر با ۰٫۴۴۹۳ گزارش شده است که بالاتر از مقادیر ۰٫۳۷۴۱ و ۰٫۳۷۲۰ در مدل‌های لاسو و APW است. این یافته نشان می‌دهد که مدل پیشنهادی توانسته است ارتباط رتبه‌ای قوی‌تری میان پیش‌بینی‌ها و مقادیر واقعی برقرار کند.



(الف): کالیبراسیون مدل لاسو



(ب): کالیبراسیون مدل APW



(ج): کالیبراسیون مدل MIR-SSS

شکل ۶: مقایسه نمودارهای کالیبراسیون سه مدل: لاسو، APW و MIR-SSS.

جدول ۷: مقایسه جهت و معناداری ضرایب در سه مدل لاسو، APW و MIR-SSS

ویژگی	لاسو	APW	MIR-SSS
سن	↑	↑	↑
جنسیت	–	–	–
سوختگی جسم داغ	↑	↑	–
سوختگی مایعات جوشان	–	–	–
سوختگی مواد شیمیایی	–	–	–
سوختگی الکتریسیته	↑	↑	–
درصد سوختگی	↑	↑	–
آلبومین	↑	↑	–
اوره	↑	↑	↑
کراتینین	↓	↓	–
گلبول‌های سفید	–	–	–
سدیم	–	–	–
پتاسیم	–	–	–

نمادها: ↑ نشان‌دهنده ضریب مثبت و معنادار، ↓ نشان‌دهنده ضریب منفی و معنادار، و – نشان‌دهنده عدم معناداری آماری است. معناداری در مدل‌های کلاسیک (لاسو و APW) بر اساس مقدار $p < ۰/۰۵$ تعیین شده است. در مدل بیزی MIR-SSS، معناداری بر اساس خارج بودن فاصله اعتباری ۹۵٪ از صفر در نظر گرفته شده است.

جدول ۷ تصویری فشرده از جهت و معناداری اثرات متغیرهای کلیدی در سه چارچوب مدل‌سازی (لاسو، APW و MIR-SSS) ارائه می‌دهد. از منظر روش‌شناختی، الگوی کلی نتایج با انتظارهای بالینی هم‌راستاست: سن و اوره در هر سه مدل اثر مثبت و معنادار دارند (↑)، در حالی‌که کراتینین در دو مدل کلاسیک اثر منفی و معنادار نشان می‌دهد (↓) و در چارچوب بیزی فاصله اعتباری آن شامل صفر است (–). این تفاوت‌ها با ماهیت سنجه معناداری در چارچوب بیزی (اتکا به فاصله‌های اعتباری) و تأکید آن بر کنترل عدم‌قطعیت سازگار است.

در سازگاری با نتایج مطالعات پیشین در مورد سوختگی می‌توان به موارد زیر اشاره کرد:

- سن در هر سه مدل مثبت و معنادار است (↑). این هم‌راستایی با ادبیات کلاسیک پیش‌بینی پیامد سوختگی همخوان است که سن را یکی از قوی‌ترین پیش‌بینی‌کننده‌ها معرفی می‌کند (برای نمونه، چارچوب‌های Baux[۱۲] و ABSI[۱۳] به‌طور مستقیم سن را وارد می‌کنند).
- اوره نیز در هر سه مدل مثبت و معنادار است (↑). این یافته

[۱۲] امتیاز Baux یک شاخص کلاسیک برای پیش‌بینی احتمال مرگ‌ومیر در بیماران سوختگی است که بر اساس جمع سن بیمار و درصد سطح سوختگی بدن (TBSA) تعریف می‌شود. نسخه اصلاح‌شده این شاخص، وجود آسیب استنشاقی را نیز با افزودن مقدار ۱۷ در نظر می‌گیرد. این امتیاز یکی از ساده‌ترین و در عین حال معتبرترین ابزارهای پیش‌بینی پیامد در بیماران سوختگی است (اُسلِر [۱۷]).

[۱۳] شاخص ABSI یا Abbreviated Burn Severity Index یک سیستم امتیازدهی پیشرفته‌تر برای پیش‌بینی مرگ‌ومیر بیماران سوختگی است که علاوه بر سن و درصد سطح سوختگی، جنسیت، وجود آسیب استنشاقی و عمق سوختگی را نیز در نظر می‌گیرد. این شاخص نسبت به امتیاز Baux دقت بیشتری در طبقه‌بندی شدت سوختگی و پیش‌بینی پیامد دارد (توبیاسن [۲۲]). این سیستم معمولاً امتیازهایی بین ۱ تا ۵ برای هر مؤلفه اختصاص می‌دهد و مجموع آن‌ها احتمال مرگ‌ومیر را تخمین می‌زند.

با نقش شاخص‌های کلیوی در وخامت بالینی و طول درمان سازگار بوده و بیانگر ارتباط وضعیت کلیه با طول مدت بستری است.

- کراتینین در لاسو و APW منفی و معنادار است (↓)، اما در MIR-SSS غیرمعنادار گزارش شده (−). این اختلاف محتمل است ناشی از هم‌خطی، ناهمگنی بین‌ویزیتی، یا انتخاب نمونه‌ها در چارچوب چندنمونه‌ای باشد که در مدل بیزی با PIP کمتر و فاصله اعتباری شامل صفر بازتاب یافته است.

در ویژگی‌های علّی سوختگی و شدت سوختگی:

- سوختگی ناشی از جسم داغ و الکتریسیته در دو مدل کلاسیک مثبت و معنادار هستند (↑)، اما در MIR-SSS غیرمعنادار (−). این الگو نشان می‌دهد که چارچوب بیزی به‌واسطه لحاظ دقیق عدم‌قطعیت، محافظه‌کارانه‌تر گزارش می‌دهد و صرفاً اثرهایی را قطعی اعلام می‌کند که فاصله اعتباری ۹۵٪ از صفر عبور نکند.
- درصد سطح سوختگی (TBSA) در دو مدل کلاسیک مثبت و معنادار است (↑)، اما در MIR-SSS غیرمعنادار (−). با توجه به نقش غالب TBSA در ادبیات، این عدم‌معناداری بیزی می‌تواند نشانه ناهمگنی زمانی اندازه‌گیری‌ها، هم‌خطی با سایر متغیرهای شدت، یا حساسیت انتخاب نمونه‌ها در چارچوب چندنمونه‌ای باشد و به تحلیل حساسیت بیشتر نیاز دارد.
- آلبومین در دو مدل کلاسیک مثبت و معنادار است (↑) اما در MIR-SSS غیرمعنادار (−). این می‌تواند بازتاب تفاوت در تجمیع اطلاعات و وزن‌دهی ویزیت‌ها در مدل بیزی باشد.

ثبات عدم‌معناداری برای جنسیت، سوختگی ناشی از مایعات جوشان و مواد شیمیایی، گلبول‌های سفید، سدیم و پتاسیم در هر سه مدل (−) نشان می‌دهد این متغیرها در حضور سنجش‌های شدت و شاخص‌های کلیوی، نقشی محدود در تبیین مدت بستری دارند.

نتایج این مطالعه هم‌سو با گزارش‌هایی است که سن و TBSA را پیش‌بینی‌کننده‌های قدرتمند پیامد بیماران سوختگی معرفی کرده‌اند و همچنین نقش شاخص‌های کلیوی/متابولیک (مانند اوره و کراتینین) را در وخامت و طول درمان تأیید کرده‌اند. افزون بر این، تفاوت‌های مشاهده‌شده میان مدل‌های کلاسیک و بیزی، به‌ویژه در معناداری TBSA و آلبومین، با رویکرد MIR-SSS در کنترل عدم‌قطعیت و انتخاب همزمان نمونه‌ها و متغیرها سازگار است و می‌تواند ناشی از ساختار چندنمونه‌ای و تجمیع مبتنی بر وزن باشد.

- **هم‌راستایی:** نتایج این مطالعه با شواهد گزارش‌شده در پایگاه‌های ملی ثبت سوختگی و پژوهش‌های مراکز تخصصی درمان سوختگی هم‌خوان است. این مطالعات نشان داده‌اند که سن و درصد سطح سوختگی بدن از مهم‌ترین عوامل پیش‌بینی‌کننده پیامدهای بیماران هستند و شاخص‌های آزمایشگاهی مرتبط با عملکرد کلیه، از جمله اوره، نقش قابل توجهی در بهبود دقت مدل‌های پیش‌بینی دارند. یافته‌های ما درباره اثرگذاری سن و اوره با نتایج گزارش‌شده در گوتیرز و آنتونی[۸]، باجوا و همکاران [۲]، لی و همکاران [۱۴] و هولت و گرین‌هال [۱۱] کاملاً هم‌سو است.
- تفاوت‌ها: معناداری بیزی برای TBSA و آلبومین در این مطالعه مشاهده نشده است (−)، در حالی‌که بسیاری از مدل‌های کلاسیک آن‌ها را معنادار گزارش می‌کنند. این اختلاف می‌تواند ناشی از تجمیع چندویزیتی، هم‌خطی با سایر شاخص‌های شدت، یا حساسیت انتخاب نمونه‌ها در MIR-SSS باشد و توصیه می‌شود با تحلیل حساسیت و ساختارهای جایگزین تجمیع بررسی شود.

۶ محدودیت‌های مطالعه و پیشنهادهایی برای پژوهش‌های آینده

محدودیت‌های مطالعه

با وجود عملکرد مناسب مدل پیشنهادی در انتخاب همزمان نمونه‌ها و متغیرها، نتایج این پژوهش باید با در نظر گرفتن برخی محدودیت‌ها تفسیر شوند. نخست، چارچوب رگرسیون چندنمونه‌ای بیزی مورد استفاده در این مطالعه مبتنی بر فرض استقلال بسته‌ها (بیماران) از یکدیگر است. در عمل، ممکن است وابستگی‌هایی ناشی از شرایط محیطی، پروتکل‌های درمانی مشترک یا سایر عوامل ثبت‌نشده میان بیماران وجود داشته باشد که در مدل حاضر لحاظ نشده‌اند. توسعه مدل‌های چندنمونه‌ای بیزی با در نظر گرفتن ساختارهای وابستگی میان بسته‌ها می‌تواند زمینه‌ای مناسب برای پژوهش‌های آینده باشد.

دوم، بخش رگرسیونی مدل بر پایه توزیع نرمال برای خطاها بنا شده است. هرچند این فرض در بسیاری از مطالعات آماری رایج است، اما در داده‌های پزشکی و زیستی ممکن است توزیع خطاها دارای چولگی، دنباله‌های سنگین یا مشاهدات پرت باشد. نتایج مطالعه شبیه‌سازی

نیز نشان داد که در حالت‌های دارای پراکندگی بالا و داده‌های پرت، عملکرد مدل پیشنهادی از نظر خطای پیش‌بینی تا حدودی کاهش یافته و در برخی موارد روش APW مقادیر MSE کوچک‌تری ارائه کرده است. این موضوع نشان می‌دهد که حساسیت مدل به ساختار توزیع خطاها می‌تواند بر عملکرد پیش‌بینی اثرگذار باشد. سوم، اطلاعات دقیقی درباره فاصله زمانی میان نوبت‌های مختلف اندازه‌گیری متغیرهای آزمایشگاهی در دسترس نبود. بنابراین اثر احتمالی زمان اندازه‌گیری بر متغیر پاسخ در این مطالعه مورد بررسی قرار نگرفت. استفاده از داده‌های طولی همراه با ثبت دقیق زمان اندازه‌گیری‌ها می‌تواند امکان توسعه مدل‌های پویاتر و غنی‌تر را فراهم سازد.

چهارم، اگرچه هدف اصلی پژوهش حاضر انتخاب همزمان نمونه‌ها و متغیرها بود، بررسی جامع اثرات تعاملی میان متغیرهای توضیحی و همچنین مطالعه ساختار هم‌خطی میان متغیرها خارج از دامنه این مطالعه قرار داشت. بررسی این جنبه‌ها می‌تواند به تفسیر بهتر نتایج و توسعه مدل‌های انعطاف‌پذیرتر منجر شود. در نهایت، داده‌های واقعی مورد استفاده در این پژوهش از یک مرکز درمانی گردآوری شده‌اند. ازاین‌رو، تعمیم نتایج به سایر جمعیت‌ها و مراکز درمانی باید با احتیاط انجام شود. ارزیابی عملکرد مدل در مجموعه‌داده‌های بزرگ‌تر، چندمرکزی و ناهمگن می‌تواند شواهد قوی‌تری در خصوص قابلیت تعمیم آن فراهم آورد.

پیشنهادهایی برای پژوهش‌های آینده

از جمله مسیرهای قابل بررسی در مطالعات آینده می‌توان به توسعه مدل‌های چندنمونه‌ای بیزی مقاوم در برابر داده‌های پرت و پراکندگی بالا اشاره کرد. استفاده از توزیع‌های انعطاف‌پذیرتر برای مدل‌سازی خطاها، نظیر توزیع Student-t، توزیع Skew-t و سایر خانواده‌های توزیع‌های چوله و دنباله‌سنگین، می‌تواند پایداری مدل را در شرایط واقعی افزایش دهد. همچنین بررسی نسخه‌های مقاوم الگوریتم‌های انتخاب متغیر و انتخاب نمونه، توسعه مدل‌های دارای ساختار وابستگی میان بسته‌ها، مطالعه اثرات تعاملی میان متغیرها و گسترش چارچوب حاضر به داده‌های طولی و چندمرکزی، از دیگر زمینه‌های پژوهشی ارزشمند در این حوزه به شمار می‌روند.

مراجع

[1] Amores, J. (2013). Multiple instance classification: Review, taxonomy and comparative study. *Artificial Intelligence*, **201**, 81–105.

[2] Bajwa, M. S., et al. (2022). Predicting thermal injury patient outcomes in a tertiary-care burn center, Pakistan. *Journal of Surgical Research*, **279**, 575–585.

[3] Brusselaers, N., Monstrey, S., Vogelaers, D., Hoste, E., and Blot, S. (2010). Quality of life after burn injuries: A systematic review. *Burns*, **36**(6), 767–776.

[4] Dietterich, T. G., Lathrop, R. H., and Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, **89**(1-2), 31–71.

[5] Emami, A., Javanmardi, F., Rajaee, M., Pirbonyeh, N., Keshavarzi, A., Fotouhi, M., and Hosseini, S. M. (2019). Predictive biomarkers for acute kidney injury in burn patients. *Journal of Burn Care & Research*, **40**(5), 601–605.

[6] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). Chapman & Hall/CRC.

[7] George, E. I., and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, **88**(423), 881–889.

[8] Gutierrez Salazar, M., and Papp, A. (2025). A machine learning model for estimating burn outcome: Analyzing American Burn Association National Burn Repository. *Journal of Burn Care & Research*, **46**(Suppl 1), S69.

[9] Herrera, F., Ventura, S., Bello, R., Cornelis, C., Zafra, A., Sánchez-Tarragó, D., and Vluymans, S. (2016). *Multiple Instance Learning: Foundations and Algorithms*. Cham: Springer.

[10] Herndon, D. N. (2018). *Total Burn Care* (5th ed.). Philadelphia: Elsevier.

[11] Holt, R., and Greenhalgh, D. (2012). Serum albumin and renal function markers as predictors of length of stay and outcomes in burn patients. *Burns*, **38**(5), 607–615.

[12] Ishwaran, H., and Rao, J. S. (2005). Spike and slab priors for Bayesian variable selection in regression models. *Annals of Statistics*, **33**(2), 730–773.

[13] Karimzadeh, A., Mobayen, M., and Zarei, R. (2024). Epidemiology and clinical outcomes of electrical burns in a burn center in the north of Iran: A 10-year retrospective study. *Trauma Monthly*, **29**(4), 1152–1159.

[14] Lee, J., Kim, H., et al. (2022). Prediction of mortality after burn surgery in critically ill burn patients: A comparison of machine learning models. *Journal of Personalized Medicine*, **12**(8), 1293.

[15] Legrand, M., Clark, A. T., Neyra, J. A., and Ostermann, M. (2023). Acute kidney injury in patients with burns. *Nature Reviews Nephrology*, **19**(2), 123–135.

[16] Mitchell, T. J., and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, **83**(404), 1023–1032.

[17] Osler, T., Glance, L. G., and Hosmer, D. W. (2010). Simplified estimates of the probability of death after burn injuries: Extending and updating the Baux score. *Journal of Trauma*, **68**(3), 690–697.

[18] Pappas, N., and Popescu-Belis, A. (2014). Explaining multiple instance learning by identifying primary instances. *Pattern Recognition*, **47**(9), 3293–3301.

[19] Park, S., Kim, J., Wang, X., and Lim, J. (2024). Variable selection in Bayesian multiple instance regression using shotgun stochastic search. *Computational Statistics & Data Analysis*, **196**, 1–18.

[20] Ramezanian, M., Bagheri Toolaroud, P., Emiralavi, C., Farzin, M., Mobayen, M., Moghaddam Ahmadi, M., Tolouei, M., Rimaz, S., Karimian, M., Eftekhari, H., Baghi, K., and Shabbak, A. (2025). Association between serum albumin levels and clinical outcomes in burn patients: A single-center retrospective analysis. *Bulletin of Emergency and Trauma*, **13**(1), 47–52.

[21] Ray, S., and Page, D. (2001). Multiple instance regression. In *Proceedings of the 18th International Conference on Machine Learning* (pp. 425–432).

[22] Tobiasen, J., Hiebert, J., and Edlich, R. (1982). The abbreviated burn severity index. *Annals of Emergency Medicine*, **11**(5), 260–262.

[23] Wang, X., Yu, Q., et al. (2008). Multiple-instance regression for remote sensing applications. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(7), 2024–2037.

[24] Zafra, A., Pechenizkiy, M., and Ventura, S. (2013). HyDR-MI: A hybrid algorithm to reduce dimensionality in multiple instance learning. *Information Sciences*, **222**, 282–301.

[25] Zhang, M. L., and Zhou, Z. H. (2004). Improve multi-instance neural networks through feature selection. *Neural Processing Letters*, **19**(1), 1–10.

[26] Zhao, P., and Yu, B. (2006). On model selection consistency of Lasso. *Journal of Machine Learning Research*, **7**, 2541–2563.

[27] Zhao, Z., Fu, G., Liu, S., Elokely, K. M., Doerksen, R. J., Chen, Y., and Wilkins, D. E. (2013). Drug activity prediction using multiple-instance learning via joint instance and feature selection. *BMC Bioinformatics*, **14**(Suppl 14), S16.

[28] Zhu, W., Xie, X., Shu, Z., Li, L., Hu, G., and Song, H. (2025). Prognostic significance of the serum creatinine level during the shock stage in severe burn patients: A 10-year retrospective study. *Mediators of Inflammation*, **2025**, 1–9.

[29] R Core Team (2025). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

[30] Gabry, J., Češnovar, R., Johnson, A., and Bronder, S. (2025). *cmdstanr: R Interface to 'CmdStan'* (R package version 0.9.0). Available at: https://mc-stan.org/cmdstanr/.

Andisheye Amari, 2024, Vol. 29, No. 2
https://doi.org/10.22034/aa.2026.2081276.1022

Application of Bayesian Multiple-Instance Regression Based on Shotgun Stochastic Search for Joint Instance and Variable Selection: A Case Study on Burn-Injury Data

Fariba Karami[1], **Roohollah Roozegar**[2*]

[1,2] *Departement of Mathematics, College of Science, Yasouj University, Yasouj, Iran*

Received: 2025/12/22
Accepted: 2026/07/07

Abstract

In many medical studies, data exhibit a multiple-instance structure, where each patient is represented by a set of instances while the response variable is observed only at the patient level. Simultaneous selection of influential instances within each bag and response-related variables is one of the major challenges in multiple-instance regression. In this study, a Bayesian multiple-instance regression framework based on Shotgun Stochastic Search (MIR-SSS) was applied and compared with Aggregate Predictive Weighting (APW) regression and the LASSO using the set-level representation. Variable selection was performed using a spike-and-slab prior, whereas instance selection was achieved through a logistic model. The methods were evaluated under four simulation scenarios and subsequently applied to real data from 2024 patients admitted to Amir Al-Momenin Burn Hospital in Shiraz. Simulation results showed that MIR-SSS achieved the lowest prediction error and high variable selection accuracy under the baseline scenarios. However, under high variability and in the presence of outliers, APW yielded better predictive performance. The area under the ROC curve also confirmed the model's ability to identify influential instances. In the real-data analysis, MIR-SSS outperformed the two competing methods in predictive performance and identified important clinical factors associated with patient outcomes. These findings demonstrate that the proposed framework provides an interpretable approach for simultaneous instance and variable selection while enabling uncertainty quantification in the analysis of medical data.

Keywords: Multiple instance learning, Bayesian models, instance selection, variable selection, Shotgun Stochastic Search (SSS).

Mathematics Subject Classification: $62F15$, $62J07$.

*Corresponding author, *roozegar@yu.ac.ir*