برآورد مدت اقامت بیماران ایسکمی قلبی و کشف عوامل مؤثر بر آن با استفاده از داده‌کاوی

مجید ضرابیان[1]، مسعود عابسی[2]

مقاله پژوهشی

چکیده

مقدمه: بیماری‌های ایسکمیک قلبی IHD (Ischemic Heart Diseases) از موضوعات هزینه‌بر و قابل بحث در حیطه بهداشت و درمان کشور می‌باشد. علاوه بر این، بیمارستان‌ها به شدت دارای محدودیت منابع برای مراقبت از بیماران هستند. از این‌رو، تعیین مدت اقامت LOS (Length of Stay) بیماران از لحاظ مدیریت هزینه بیمارستان، بسیار مهم است. در تحقیق حاضر مدل‌هایی جهت برآورد LOS بیماران ایسکمی قلبی و عوامل مؤثر بر آن ارایه شد.

روش بررسی: این پژوهش در مطالعه‌ای کاربردی، ۶۵۲۴ بیمار بستری ایسکمی قلبی را که طی شش ماه دوم سال ۱۳۹۲ به ۱۶ مرکز درمانی شهر تهران مراجعه کرده بودند و اقدامات درمانی صورت گرفته در پرونده آن‌ها توسط کدهای تعرفه خدمات درمانی ثبت شده بود، بررسی کرده است. پس از جمع‌آوری داده‌ها با استفاده از فرمی طراحی شده و پاکسازی آن‌ها، مدل‌های برآورد LOS با کمک الگوریتم‌های داده‌کاوی شبکه عصبی مصنوعی ANN (Artificial Neural Networks)، ماشین‌های بردار پشتیبان SVM (Support Vector Machines)، اکتشاف تعاملی اتوماتیک CHAID (Chi-squared Automatic Interaction Detection)، درخت‌های دسته‌بندی و رگرسیون CART (Classification and Regression Trees) و مدل ترکیبی در نرم‌افزار Clementine ساخته شدند.

یافته‌ها: میانگین و انحراف معیار LOS بیماران، ۹/۶۰۸ ± ۷/۷۲۷ روز بود. همبستگی خطی با LOS واقعی و خطای نسبی تمام مدل‌ها به ترتیب بیشتر از ۰/۷ و کمتر از ۰/۵ شد. مدل ترکیبی و SVM نیز دقیق‌ترین مدل‌های به دست آمده بود.

نتیجه‌گیری: با توجه به نتایج مدل‌ها، بیماران ایسکمی که نیازمند توان‌بخشی، مشاوره، رادیوتراپی و سی‌تی‌اسکن هستند، دارای LOS بیشتری می‌باشند. همچنین، نوع بیماری ایسکمی و به ویژه کدهای درمانی کالیفرنیا نقش مهمی در برآورد LOS بیماران دارند.

واژه‌های کلیدی: مدت اقامت؛ بیماری‌های ایسکمیک قلبی؛ داده‌کاوی

دریافت مقاله: ۱۳۹۵/۹/۱۳ پذیرش مقاله: ۱۳۹۶/۱/۲۷

ارجاع: ضرابیان مجید، عابسی مسعود. **برآورد مدت اقامت بیماران ایسکمی قلبی و کشف عوامل مؤثر بر آن با استفاده از داده‌کاوی.** مدیریت اطلاعات سلامت ۱۳۹۶؛ ۱۴ (۱): ۱۶-۲۵

مقدمه

بیمارستان‌ها در مواجهه با چالش‌هایی مانند افزایش بیماران، محدودیت تخت بستری و کمبود پرستار، با محدودیت‌های منابع متعددی مواجه هستند که باعث افزایش هزینه و فشار کاری در بخش‌های مختلف بیمارستان می‌شود. افزایش مدت اقامت LOS (Length of Stay) بیماران، از عوامل اصلی افزایش هزینه‌های بیمارستان‌ها به شمار می‌رود. بنابراین، اگر بیمارانی که دارای اقامت طولانی مدت در بیمارستان هستند، بلافاصله پس از پذیرش شناسایی شوند و منابع مناسب جهت تلاش برای سرعت بخشیدن به مراقبت‌های بهداشتی در زمان اولیه فراهم گردد، می‌توان از مشکلات قریب‌الوقوع عملیاتی که منجر به فشارهای مالی و عملکردی در اورژانس یا بیمارستان می‌شود، جلوگیری کرد (۱).

بیماری‌های قلبی با بستری شدن مکرر و طولانی بودن LOS شناخته شده‌اند. بنابراین، توجه قابل ملاحظه‌ای در رابطه با کنترل هزینه‌های بیمارستانی بیماری‌های قلبی- عروقی وجود دارد (۳، ۲). از جمله بیماری‌های مهم قلبی، بیماری‌های ایسکمیک قلبی IHD (Ischemic Heart Diseases) می‌باشد که با افزایش میزان مرگ و میر ناشی از آن، به یکی از موضوعات هزینه‌بر و قابل بحث در حیطه بهداشت و درمان کشور تبدیل شده‌اند (۴). مطالعات محدودی به بررسی LOS بیماران قلبی پرداخته‌اند. رضایی هاچه‌سو و همکاران LOS تعداد ۴۹۴۸ بیمار عروق کرونری مرکز قلب رجایی تهران را به سه گروه ۱ تا ۵، ۶ تا ۹ و بالای ۱۰ روز تقسیم نمودند و با نرم‌افزار SPSS Clementine 12.0 و الگوریتم‌های مختلف داده‌کاوی، به برآورد LOS بیماران در این سه گروه پرداختند (۲). Laurenti و همکاران نیز LOS بیش از ۶۸۹ هزار بیمار IHD را در برزیل بررسی و LOS و هزینه بیماران را با استفاده از خصوصیات دموگرافیک آن‌ها مقایسه نمودند (۵). در تحقیق دیگری محققان با استفاده از نمونه داده‌های فرضی موجود در نرم‌افزار SPSS با نام Patient_los.sav و الگوریتم شبکه عصبی آن نرم‌افزار هزینه و LOS بیماران سکته قلبی را بررسی کردند که به دلیل فرضی بودن داده‌ها، مدل‌های به دست آمده در دنیای واقعی کاربردی ندارد (۶).

مقاله حاصل پایان‌نامه کارشناسی ارشد با شماره ۲۳۵۸۶۱۱ می‌باشد که با حمایت مالی سازمان بیمه سلامت ایران انجام شده است.

۱- دانشجوی کارشناسی ارشد، مهندسی صنایع، گروه مهندسی صنایع، دانشکده فنی مهندسی، دانشگاه یزد، یزد، ایران (نویسنده مسؤول)

Email: zarabian_majid90@yahoo.com

۲- استادیار، مدیریت صنعتی، گروه مهندسی صنایع، دانشکده فنی مهندسی، دانشگاه یزد، یزد، ایران

www.mui.ac.ir

Archive of SID

www.SID.ir

حال از آن‌جا که برآورد LOS هر بیمار و مدیریت بهره‌وری اشغال تخت در بیمارستان، یکی از راه‌های برنامه‌ریزی کاهش هزینه محسوب می‌شود و این امر با استفاده از تکنولوژی‌های مکمل مانند داده‌کاوی میسر می‌باشد (۲)، تحقیق حاضر در یک مطالعه موردی، LOS بیماران IHD را با استفاده از تکنیک‌های داده‌کاوی تعیین و مدل‌ها و قواعد پیش‌بینی آن را استخراج نموده است.

روش بررسی

در این پژوهش کاربردی، LOS تعداد ۶۵۲۴ بیمار بستری IHD را که طی شش ماه دوم سال ۱۳۹۲ به ۱۶ مرکز درمانی شهر تهران مراجعه کرده بودند و اقدامات درمانی صورت گرفته در پرونده آن‌ها توسط کدهای تعرفه خدمات درمانی ثبت شده بود، بررسی گردید. جمع‌آوری اطلاعات به صورت گذشته‌نگر از پرونده بیماران با استفاده از فرم طراحی شده انجام گرفت. در ضمن، جمع‌آوری اطلاعات با رعایت حفظ اسرار و حقوق مربوط آن صورت پذیرفت. بنابراین، از ذکر نام مراکز درمانی در این مطالعه خودداری می‌شود.

بیماری‌های IHD مطابق دسته‌بندی بین‌المللی بیماری‌ها ICD-۱۰ (International Classification of Diseases) در پرونده بیماران ثبت می‌شوند و شامل آنژین‌های صدری با کد I۲۰، انفارکتوس میوکارد با یا بدون بالا رفتن قطعه ST با کد I۲۱، سکته بعدی با یا بدون بالا رفتن قطعه ST با کد I۲۲، عوارض فعلی مشخص ناشی از انفارکتوس میوکارد با یا بدون بالا رفتن قطعه ST با کد I۲۳، دیگر بیماری‌های ایسکمی حاد قلبی با کد I۲۴ و بیماری مزمن ایسکمی قلبی با کد I۲۵ می‌باشد (۷).

داده‌های جمع‌آوری شده ممکن است دارای داده پرت، اطلاعات مفقود و متناقض باشد. بنابراین، قبل از هر نوع تحلیل و داده‌کاوی، پیش‌پردازش بر روی داده‌ها صورت می‌گیرد؛ چرا که داده‌های بی‌کیفیت منجر به بروز نتایج بی‌کیفیت خواهند شد (۸). از پیش‌پردازش‌های صورت گرفته می‌توان به پاکسازی (حذف داده‌های پرت و متناقض) و تقلیل داده (حذف داده‌های تکراری و گروه‌بندی سن بیماران) اشاره کرد. پس از انجام پیش‌پردازش‌های لازم، ۵۵۰ عدد از داده‌ها حذف و در نتیجه، ۵۹۷۴ رکورد وارد مطالعه شد. جدول ۱ ویژگی‌های استفاده شده در مدل را به همراه برخی اطلاعات توصیفی آن‌ها نشان می‌دهد.

جدول ۱: ویژگی‌های استفاده شده در مدل

نام ویژگی	مقادیر ویژگی	فراوانی (تعداد بیمار)	توضیحات
LOS	۱ تا ۱۵۱	–	روز
ICD-10 code	I۲۰	۱۳۳۸	کد ICD-۱۰ پرونده بیماران
	I۲۱	۲۲۵	
	I۲۲	۱۸	
	I۲۳	۱۹	
	I۲۴	۲۳۹۸	
	I۲۵	۱۹۷۶	
Gender	۱	۳۸۰۸	۱ مرد و ۰ زن
	۰	۲۱۶۶	
Age ≥ 70	۱ (بله)	۱۴۰۴	آیا سن بیمار ۷۰ سال به بالا است؟
	۰ (خیر)	۴۵۷۰	
Age	۰–۱۲۹	-	سال
Age group	کمتر از ۱ سال	۱۵	-
	۱–۴	۱۵	
	۵–۱۴	۲۵	
	۱۵–۲۴	۱۰۱	
	۲۵–۳۴	۱۷۰	
	۳۵–۴۴	۳۹۰	
	۴۵–۵۴	۱۱۴۲	
	۵۵–۶۴	۱۹۴۷	
	۶۵–۷۴	۱۳۷۳	
	بیشتر از ۷۵	۷۹۶	

Archive of SID

www.mui.ac.ir
www.SID.ir

جدول ۱: ویژگی‌های استفاده شده در مدل (ادامه)

نام ویژگی	مقادیر ویژگی	فراوانی (تعداد بیمار)	توضیحات
Dialysis patients*	۱	۷۰	آیا بیمار دیالیزی است؟
	۰	۵۹۰۴	
Radiation therapy*	۱	۴۷۲۵	آیا بیمار نیازمند رادیوتراپی است؟
	۰	۱۲۴۹	
Non-invasive medical service*	۱	۱۳۱۵	آیا بیمار نیازمند خدمات پزشکی غیر جراحی است؟
	۰	۴۶۵۹	
Ultrasonography*	۱	۱۱۶۶	آیا بیمار نیاز به سونوگرافی (اکوکاردیوگرافی) دارد؟
	۰	۴۸۰۸	
CT Scan*	۱	۳۷۷	آیا بیمار نیاز به سی‌تی اسکن دارد؟
	۰	۵۵۹۷	
Rehabilitation*	۱	۱۹۷۹	آیا بیمار نیازمند توان‌بخشی است؟
	۰	۳۹۹۵	
Consultation*	۱	۱۹۶۴	آیا برای درمان نیاز به مشاوره است؟
	۰	۴۰۱۰	
Recording tape*	۱	۴۴۶۳	آیا بیمار نیاز به نوارنگاری دارد؟
	۰	۱۵۱۱	
MRI*	۱	۳۱	آیا بیمار نیاز به MRI دارد؟
	۰	۵۹۴۳	
Nuclear medicine*	۱	۵۹	آیا بیمار نیازمند پزشکی هسته‌ای است؟
	۰	۵۹۱۵	
California Code	کدهای تعرفه خدمات	۵۹۷۴	اولین کد خدمت ثبت شده
California Along 1**	کدهای تعرفه خدمات	۲۳۸۹	دومین کد خدمت ثبت شده
California Along 2**	کدهای تعرفه خدمات	۱۴۵۷	سومین کد خدمت ثبت شده
California Along 3**	کدهای تعرفه خدمات	۱۰۸۰	چهارمین کد خدمت ثبت شده
California Along 4**	کدهای تعرفه خدمات	۳۰۴	پنجمین کد خدمت ثبت شده
California Along 5**	کدهای تعرفه خدمات	۱۰۵	ششمین کد خدمت ثبت شده

ICD-10: International Classification of Diseases-10; LOS: Length of Stay; MRI: Magnetic Resonance Imaging

*در این ویژگی‌ها، ۱ به معنی ارایه آن خدمت به بیمار و ۰ به معنی عدم ارایه خدمت است؛ **این ویژگی‌ها مقداری به نام Null دارند که نشان می‌دهد بیمار در آن ویژگی‌ها و ویژگی‌های بعدی California Along فاقد کد تعرفه است. به طور مثال، اگر ویژگی California Along 3 دارای مقدار Null باشد، ویژگی‌های California Along 4 و California Along 5 نیز مقدار Null می‌گیرند.

در جدول ۱ ویژگی Age ≥ ۷۰ با توجه به توصیه Fetter و Thompson (۹) در ساخت گروه‌های مربوط به تشخیص و ویژگی Age group بر اساس تقسیم‌بندی سن توصیه شده توسط سازمان جهانی بهداشت برای ICD-۱۰ ایجاد شده است (۱۰). ویژگی‌های ردیف ۷ تا ۱۶ نیز برخی اقدامات تشخیصی و درمانی صورت گرفته بر روی بیماران را نشان می‌دهد. علاوه بر این، ویژگی‌های ردیف ۱۷ تا ۲۲ کدهای سایر خدمات تشخیصی و درمانی انجام شده در مراکز درمانی می‌باشد که وزارت بهداشت آن را تدوین کرده است. این کدها که از سیستم استاندارد آمریکا یا کتاب کالیفرنیا به منظور برآورد ارزش نسبی خدمات استفاده می‌کرد، تا شهریور سال ۱۳۹۳ مبنای پرداخت هزینه‌های درمانی در ایران بود (۱۱). در رابطه با ویژگی‌های ردیف ۱۷ تا ۲۲ باید گفت که کدهای این ویژگی‌ها نسبت به هم تقدمی ندارند و تنها تا ۶ کد خدمت ارایه شده به بیمار را نشان می‌دهند.

پس از آماده‌سازی داده‌ها، نوبت به فرایند داده‌کاوی می‌رسد. داده‌کاوی بارها در موضوعات کاربردی بهداشت و درمان همچون پیش‌بینی بیماری و مدیریت بیماران استفاده شده است (۲). بنابراین، در تحقیق حاضر به منظور برآورد LOS بیماران، از برخی رایج‌ترین الگوریتم‌های داده‌کاوی موجود در

www.mui.ac.ir
www.SID.ir
Archive of SID

نرم‌افزار SPSS Clementine 12.0 شامل شبکه عصبی مصنوعی ANN (Artificial Neural Networks)، ماشین‌های بردار پشتیبان SVM (Support Vector Machines)، اکتشاف تعاملی اتوماتیک CHAID (Chi-squared Automatic Interaction Detection)، درخت‌های دسته‌بندی و رگرسیون CART (Classification and Regression Trees) و مدل ترکیبی (که با میانگین‌گیری نتایج چهار الگوریتم نامبرده را ترکیب می‌کند) استفاده گردید.

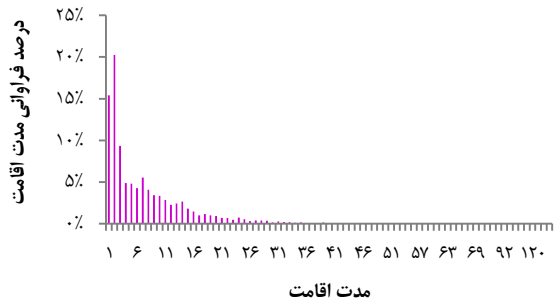
در فرایند داده‌کاوی لازم است به منظور ارزیابی مدل‌ها، داده‌ها را به مجموعه‌های آموزشی و تست تقسیم کرد. از مجموعه داده‌های آموزشی به منظور ساخت مدل‌های برآورد LOS و از مجموعه داده‌های تست برای ارزیابی مدل‌ها استفاده می‌گردد. در مطالعه حاضر حدود ۷۰ درصد داده‌ها برای ساخت مدل‌ها و حدود ۳۰ درصد باقی‌مانده جهت ارزیابی آن‌ها استفاده شد. همچنین، در هنگام ساخت مدل‌ها، مقادیر LOS بالای ۳ انحراف معیار (۳۶/۵۴۲ روز) از فرایند ساخت مدل حذف گردید.

یافته‌ها

میانگین و انحراف معیار سن بیماران بررسی شده ۱۴/۳۰ ± ۵۹/۴۱ سال بود و بیشتر بیماران در بازه سنی ۵۵ تا ۶۴ سال قرار داشتند (۳۲/۵۹ درصد). در بیماری‌های IHD، کد I۲۴ بیشترین تکرار (۴۰/۱۴ درصد) را داشت و سپس کدهای I۲۵ (۳۳/۰۸ درصد)، I۲۰ (۲۲/۴۰ درصد)، I۲۱ (۳/۷۷ درصد)، I۲۳ (۰/۳۱ درصد) و I۲۲ (۰/۳۰ درصد) به ترتیب بیشترین میزان وقوع را به خود اختصاص دادند. میانگین و انحراف معیار LOS بیماران نیز در این بررسی ۹/۶۰۸ ± ۷/۷۲۷ روز بود.

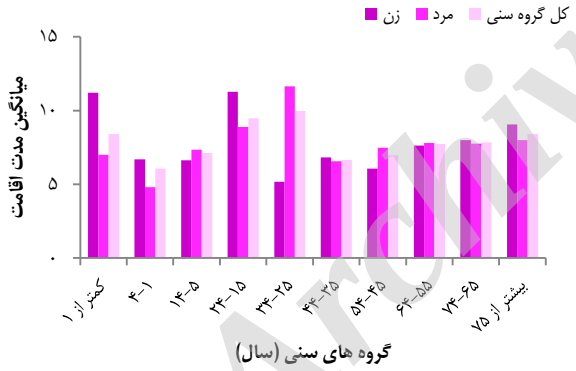
شکل ۱ فراوانی LOS را در داده‌ها نشان می‌دهد. با توجه به این شکل، حدود ۹۵ درصد از بیماران IHD بین ۱ تا ۲۴ روز بستری شده بودند که مدت اقامت حدود ۵۰ درصد آن‌ها، ۱ تا ۴ روز بود. شکل ۲ نیز میانگین LOS بیماران IHD را به تفکیک جنسیت و کل در هر گروه سنی نشان داده است که به ترتیب گروه‌های ۱۵ تا ۳۵ سال، ۷۵ سال به بالا و زیر ۱ سال بیشترین مقدار را داشتند.

از آن‌جایی که ویژگی هدف (LOS) از نوع پیوسته است، نرم‌افزار Clementine 12.0 دقت مدل‌های به دست آمده را با استفاده از دو مقدار همبستگی خطی و خطای نسبی نمایش می‌دهد. میانگین همبستگی خطی و خطای نسبی برای مدل‌های مختلف پس از ده بار تکرار تقسیم داده‌ها به آموزش و تست در جدول ۲ ارایه شده است.



شکل ۱: فراوانی مدت اقامت بیماران

خطای نسبی برای ویژگی هدف برابر است با نسبت مجموع مربعات خطای ویژگی هدف به مجموع مربعات خطا برای مدل خنثی که در آن مقدار میانگین ویژگی هدف به عنوان مقادیر پیش‌بینی شده برای هر مورد محاسبه می‌شود (۶). در نتیجه، هرچه این مقدار از ۱ کمتر و به صفر نزدیک‌تر باشد، مدل بهتری خواهیم داشت. همچنین، در جدول ۲ همبستگی خطی و خطای نسبی مدل ترکیبی توسط آزمون t و در سطح معنی‌داری ۰/۰۵ با سایر روش‌ها مقایسه گردید که طبق آن، مدل ترکیبی از سایر روش‌ها به جز SVM دقت بیشتر و خطای معنی‌دار کمتری داشت، اما در کل مدل ترکیبی با میانگین‌گیری میان نتایج، دقت بیشتری را نسبت به مدل‌های انفرادی به ویژه در مواجهه با داده‌های جدید نشان داد و برآورد بهتری از سایر مدل‌ها حتی SVM خواهد داشت.



شکل ۲: متوسط مدت اقامت بیماران به تفکیک گروه سنی و جنسیت

جدول ۲: میانگین همبستگی خطی و خطای نسبی مدل‌های برآورد مدت اقامت بر مبنای داده‌های تست

مدل‌های برآورد LOS	همبستگی خطی	آماره t همبستگی	P	خطای نسبی	آماره t خطا	P
ANN	۰/۷۶۵	۳/۰۲۸	۰/۰۰۷	۰/۴۰۱	-۳/۱۹۴	۰/۰۰۵
SVM	۰/۷۷۳	۱/۶۰۹	۰/۱۲۵	۰/۳۸۶	-۱/۴۸۵	۰/۱۵۵
CHAID	۰/۷۵۸	۴/۹۳۲	< ۰/۰۰۱	۰/۴۱۸	-۵/۰۴۹	< ۰/۰۰۱
CART	۰/۷۳۲	۱۲/۸۱۵	< ۰/۰۰۱	۰/۴۶۴	-۱۳/۰۱۰	< ۰/۰۰۱
مدل ترکیبی	۰/۷۷۹	-	-	۰/۳۷۳	-	-

LOS: Length of Stay; ANN: Artificial Neural Networks; SVM: Support Vector Machines; CHAID: Chi-squared Automatic Interaction Detection; CART: Classification and Regression Trees

www.mui.ac.ir
Archive of SID
www.SID.ir

در ادامه، از ویژگی‌های مهم مدل CHAID برای استخراج شاخص‌های مهم در برآورد LOS استفاده شد. ویژگی‌های مهم، اهمیت نسبی هر ویژگی در ارزیابی مدل را نشان می‌دهد و ارتباطی با دقت مدل ندارد (۲). جدول ۳ ویژگی‌هایی را که به ترتیب بیشترین اهمیت را در برآورد LOS بیماران IHD دارند، نشان می‌دهد.

با توجه به داده‌های جدول ۳، ویژگی توان‌بخشی و کدهای خدمات درمانی با بیش از ۷۰ درصد وزن اهمیت نسبی، نقش مهمی در برآورد LOS بیماران IHD با استفاده از مدل CHAID داشتند.

جدول ۳: ویژگی‌های مهم مدت اقامت بیماران

ویژگی‌ها	اهمیت نسبی
Rehabilitation	۰/۴۹۴
California Code	۰/۲۲۲
Consultation	۰/۱۸۶
Radiation therapy	۰/۰۴۱
ICD-10 Code	۰/۰۲۷
California along 1	۰/۰۲۲
California along 2	۰/۰۰۵
CT Scan	۰/۰۰۳

برخی الگوریتم‌های داده‌کاوی با وجود ساخت مدل‌های برآورد خوب، قواعد و دانشی را در اختیار محقق قرار نمی‌دهند. الگوریتم‌های ANN و SVM از آن دسته هستند. بنابراین، از نتایج الگوریتم‌های CHAID و CART که از نوع درخت تصمیم می‌باشند، برای استخراج قواعد استفاده گردید. شکل‌های ۳ و ۴ به ترتیب درخت تصمیم حاصل از الگوریتم‌های CHAID و CART را نشان می‌دهد.

همان‌گونه که از شکل‌های ۳ و ۴ مشخص است، الگوریتم CART درخت تصمیم ساده‌تری دارد. بنابراین، در ادامه تنها قواعد الگوریتم CART در جدول ۴ ارایه شده است.

بحث

میانگین LOS بیماران در مطالعه حاضر، ۷/۷۲۷ روز بود که با دیگر تحقیقات مشابه تفاوت داشت. به طور مثال، مطالعه‌ای در ژاپن میانگین LOS بیماران IHD را ۱۲/۹ روز (۱۲) و مطالعه‌ای دیگر در برزیل ۶/۴ روز (۵) گزارش نمود.

در تحقیق حاضر مدل ترکیبی و SVM با بیشترین همبستگی و کمترین خطا، دقیق‌ترین مدل‌ها بودند. نتایج مطالعات دیگر نیز نشان داده است که SVM بالاترین قدرت پیش‌بینی را در میان دیگر الگوریتم‌ها دارد و در حال تبدیل شدن به یک الگوریتم رایج در زمینه پزشکی و سلامت است (۱۳، ۲) که با نتایج پژوهش حاضر همسو بود.

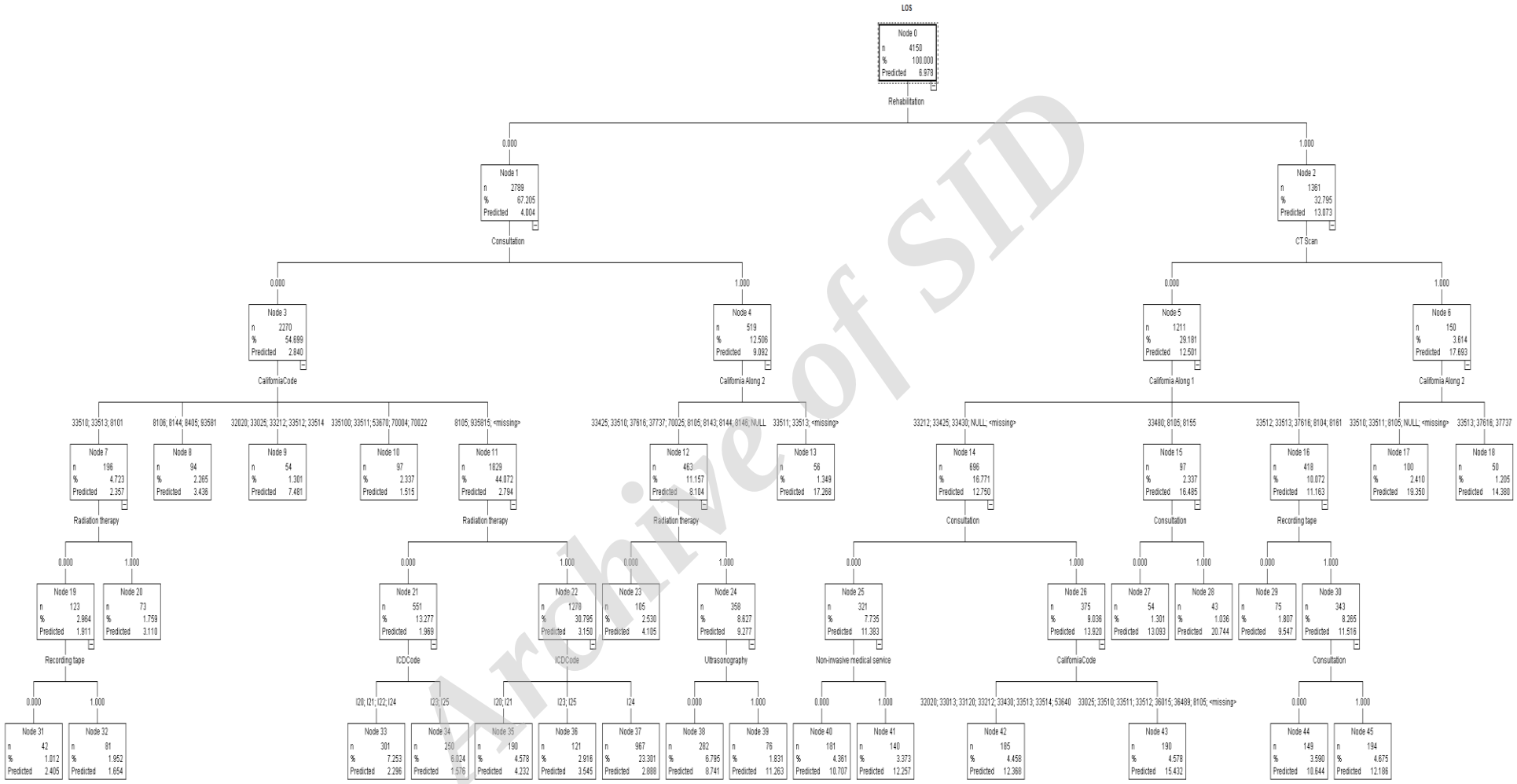
در مدل‌های برآورد مطالعه حاضر، ویژگی‌های توان‌بخشی، مشاوره و کدهای کالیفرنیا ثبت شده برای بیمار تأثیرگذارترین ویژگی‌ها در LOS بیماران IHD بودند.

جدول ۴: قواعد برآورد مدت اقامت بیماران

مقدم (شرط) قاعده	تالی قاعده (روز)
اگر توان‌بخشی = ۰ و مشاوره = ۰ و کالیفرنیا کد = ۳۲۰۲۰ یا ۳۳۰۲۵ یا ۳۳۲۱۲ یا ۳۳۵۱۲ یا ۳۳۵۱۴	آنگاه LOS = ۷/۵۶۹
اگر توان‌بخشی = ۰ و مشاوره = ۰ و کالیفرنیا کد = ۳۳۵۱۰ یا ۳۳۵۱۰۰ یا ۳۳۵۱۱ یا ۳۳۵۱۳ یا ۵۳۶۷۰ یا ۷۰۰۲۲ یا ۸۱۰۱ یا ۸۱۰۵ یا ۸۱۰۶ یا ۸۱۴۴ یا ۸۴۰۵ یا ۹۳۵۸۱۵ و رادیوتراپی = ۰	آنگاه LOS = ۱/۹۵۲
اگر توان‌بخشی = ۰ و مشاوره = ۰ و کالیفرنیا کد = ۳۳۵۱۰ یا ۳۳۵۱۰۰ یا ۳۳۵۱۱ یا ۳۳۵۱۳ یا ۵۳۶۷۰ یا ۷۰۰۲۲ یا ۸۱۰۱ یا ۸۱۰۵ یا ۸۱۰۶ یا ۸۱۴۴ یا ۸۴۰۵ یا ۹۳۵۸۱۵ و رادیوتراپی = ۱ و کد ICD–۱۰ = I۲۰ یا I۲۱	آنگاه LOS = ۴/۰۳۳
اگر توان‌بخشی = ۰ و مشاوره = ۰ و کالیفرنیا کد = ۳۳۵۱۰ یا ۳۳۵۱۰۰ یا ۳۳۵۱۱ یا ۳۳۵۱۳ یا ۵۳۶۷۰ یا ۷۰۰۲۲ یا ۸۱۰۱ یا ۸۱۰۵ یا ۸۱۰۶ یا ۸۱۴۴ یا ۸۴۰۵ یا ۹۳۵۸۱۵ و رادیوتراپی = ۱ و کد ICD–۱۰ = I۲۳ یا I۲۴ یا I۲۵	آنگاه LOS = ۲/۹۳۶
اگر توان‌بخشی = ۰ و مشاوره = ۱	آنگاه LOS = ۹/۰۹۲
اگر توان‌بخشی = ۱ و سی‌تی اسکن = ۰ کالیفرنیا همراه ۱ = ۳۳۴۸۰ یا ۸۱۰۵ یا ۸۱۵۵ و مشاوره = ۰	آنگاه LOS = ۱۳/۰۹۳
اگر توان‌بخشی = ۱ و سی‌تی اسکن = ۰ کالیفرنیا همراه ۱ = ۳۳۴۸۰ یا ۸۱۰۵ یا ۸۱۵۵ و مشاوره = ۱	آنگاه LOS = ۲۰/۷۴۴
اگر توان‌بخشی = ۱ و سی‌تی اسکن = ۱	آنگاه LOS = ۱۷/۶۹۳
اگر توان‌بخشی = ۱ و سی‌تی اسکن = ۰ کالیفرنیا همراه ۱ = ۳۳۲۱۲ یا ۳۳۴۲۵ یا ۳۳۴۳۰ یا ۳۳۵۱۲ یا ۳۳۵۱۳ یا ۳۷۶۱۶ یا ۸۱۰۴ یا ۸۱۶۱ یا Null و کالیفرنیا همراه ۴ = ۳۳۵۶۲ یا ۳۳۶۴۱ یا ۳۳۶۸۲ یا ۳۳۸۶۰ یا ۳۳۸۷۰ یا ۳۳۹۷۰ یا ۳۳۹۷۲ یا ۳۵۱۴۱ یا ۳۷۷۳۱ یا ۳۷۷۳۵ یا ۸۱۰۶ یا ۸۱۶۱	آنگاه LOS = ۱۶/۴۵۲
اگر توان‌بخشی = ۱ و سی‌تی اسکن = ۰ کالیفرنیا همراه ۱ = ۳۳۲۱۲ یا ۳۳۴۲۵ یا ۳۳۴۳۰ یا ۳۳۵۱۲ یا ۳۳۵۱۳ یا ۳۷۶۱۶ یا ۸۱۰۴ یا ۸۱۶۱ یا Null و کالیفرنیا همراه ۴ = ۳۳۴۷۴ یا ۳۷۶۱۶ یا ۳۷۷۳۷ یا ۸۱۰۵ یا ۸۱۰۸ یا ۸۱۲۷ یا Null و مشاوره = ۰	آنگاه LOS = ۱۰/۹۴۹
اگر توان‌بخشی = ۱ و سی‌تی اسکن = ۰ کالیفرنیا همراه ۱ = ۳۳۲۱۲ یا ۳۳۴۲۵ یا ۳۳۴۳۰ یا ۳۳۵۱۲ یا ۳۳۵۱۳ یا ۳۷۶۱۶ یا ۸۱۰۴ یا ۸۱۶۱ یا Null و کالیفرنیا همراه ۴ = ۳۳۴۷۴ یا ۳۷۶۱۶ یا ۳۷۷۳۷ یا ۸۱۰۵ یا ۸۱۰۸ یا ۸۱۲۷ یا Null و مشاوره = ۱	آنگاه LOS = ۱۲/۶۷۰

LOS: Length of Stay

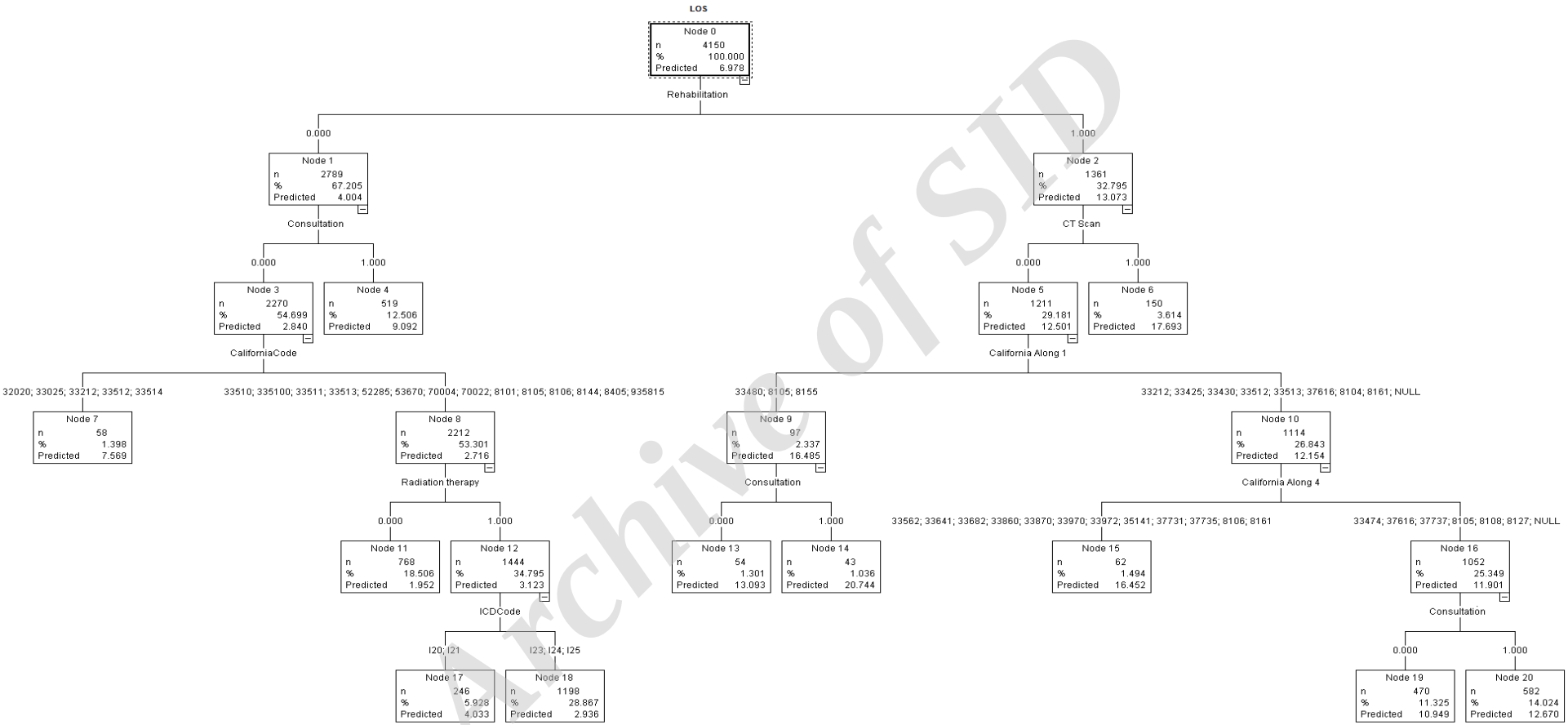
Archive of SID
www.SID.ir



شکل ۳: درخت تصمیم حاصل از الگوریتم CHAID (Chi-squared Automatic Interaction Detection)

Archive of SID

www.SID.ir



شکل ۴: درخت تصمیم حاصل از الگوریتم CART (Classification and Regression Trees)

Archive of SID

www.mui.ac.ir
www.SID.ir

طبق نظر کارشناسان، بیماران نیازمند توان‌بخشی، LOS بیشتری نسبت به سایر بیماران خواهند داشت که قواعد به دست آمده نیز همین مسأله را تأیید می‌کند. در مطالعه Appelros نیز به این نکته اشاره شد که در دسترس بودن تسهیلات توان‌بخشی، بر LOS بیماران اثرگذار است (۱۴). همچنین، در طول فرایند درمان به دلایلی مانند وجود بیماری‌های همراه، ممکن است پزشک نیازمند مشاوره با پزشک (پزشکان) دیگری باشد. در این شرایط اغلب بیمار LOS بیشتری دارد که قواعد جدول ۴ و یافته‌های مطالعه‌ی رضایی و همکاران نیز وجود چنین شرایطی را با بستری شدن بالای بیمار قلبی مرتبط دانستند (۲).

در رابطه با کدهای درمانی کالیفرنیا نیز قواعد جدول ۴ به خوبی تأثیرگذارترین کدها در LOS بیماران IHD را با توجه به شرایط مختلف نشان می‌دهد. هر کدام از این کدها حاکی از اقدامات درمانی مختلفی می‌باشد. به عنوان مثال، کدهای شروع شونده با ۳۳۵ برای جراحی‌های بای‌پس عروق کرونر و ۳۳۴ برای اعمال دریچه قلبی هستند. طبق نتایج تحقیقات Lazar و همکاران، بیماران دارای دو جراحی بای‌پس و دریچه قلبی، LOS معنی‌دار بالایی دارند. همچنین، LOS در بیماران آنژین ناپایدار و سکته قلبی دارای جراحی بای‌پس، طولانی‌تری می‌باشد (۱۵) که سه قاعده آخر و سوم جدول ۴ نیز این مسأله را نشان می‌دهند.

به جز سه ویژگی نامبرده، گروه‌های سنی نیز در برآورد LOS بیماران IHD اهمیت دارد و الگوریتم‌های ANN و SVM آن را از ویژگی‌های مؤثر در برآورد LOS دانسته‌اند. سن بیماران در برخی مطالعات به عنوان یک ویژگی تأثیرگذار شناسایی شده است؛ به طوری که بیان می‌شود مفیدترین ویژگی بیماران برای برآورد LOS، سن آن‌ها می‌باشد (۱۶، ۲).

همچنین، در تحقیق حاضر خدماتی همچون رادیوتراپی و سی‌تی اسکن در برآورد LOS بیماران IHD تأثیرگذار بود. نتایج تحقیقات نشان داده است که رادیوتراپی در دراز مدت یا کوتاه مدت تأثیرات نامطلوبی بر قلب دارد و این دسته از بیماران در بیشتر موارد نیازمند مداخله جراحی برای درمان مؤثر هستند (۱۷). از این‌رو، LOS این بیماران اغلب طولانی‌تر خواهد شد. با توجه به بررسی داده‌ها، سی‌تی اسکن اغلب برای بیماران آنژین صدری و ایسکمی مزمن دارای اعمال جراحی بای‌پس، دریچه قلبی، ترمیم و بستن سینه و ورید و یا کارگذاری (تعویض) مولد پیام‌ساز انجام می‌گیرد که نیازمندی این بیماران به توان‌بخشی، افزایش LOS را در قاعده هشتم جدول ۴ به همراه داشته است.

نتیجه‌گیری

نتایج پژوهش حاضر نشان داد که نوع خدمات تشخیصی و درمانی ارایه شده به بیماران IHD، به خوبی می‌تواند LOS آن‌ها را تحت تأثیر قرار دهد؛ به طوری که بیماران نیازمند توان‌بخشی، مشاوره، رادیوتراپی و سی‌تی‌اسکن، دارای LOS بیشتری می‌باشند. همچنین، از قواعد به دست آمده مشخص شد که کدهای درمانی کالیفرنیا نقش مهمی در برآورد LOS بیماران دارند. علاوه بر این، بررسی LOS بیماران از لحاظ مدیریت هزینه بیمارستان بسیار حایز اهمیت است. از این‌رو، یافته‌های این تحقیق می‌تواند در مدیریت بهتر LOS بیماران IHD و تخصیص مناسب منابع به آن‌ها مفید باشد.

پیشنهادها

اورژانس‌ها یا بیمارستان‌ها می‌توانند از مدل‌ها و قواعد موجود در تحقیق حاضر استفاده کنند و منابع مناسبی را جهت سرعت بخشیدن به مراقبت‌های بهداشتی و کاهش LOS بیماران IHD تخصیص دهند. همچنین، پیشنهاد می‌گردد با ترکیب اطلاعاتی مانند بیماری‌های همراه و عوارض درمان، نتایج آزمایشگاهی و سابقه بیماری‌های قلبی بیماران با اقدامات تشخیصی درمانی، مدل‌های دقیق‌تری برای برآورد و کشف عوامل مؤثر بر LOS بیماران طراحی شود.

تشکر و قدردانی

مطالعه حاضر از نوع پژوهشی اصیل و برگرفته از پایان‌نامه کارشناسی ارشد می‌باشد که تحت حمایت مالی سازمان بیمه سلامت ایران انجام شد. بدین وسیله از آقای دکتر احسان بیطرف و تمامی مسؤولان بیمارستانی و سازمان بیمه سلامت ایران که کمک و حمایت شایانی در انجام این تحقیق نمودند، تشکر و قدردانی به عمل می‌آید.

References

1. Karim H, Tara M, Etminani K. Factors associated with length of stay based on admission data in Emam-Reza and Ghaem Hospitals of Mashhad. Health Inf Manage 2015; 12(4): 405-15. [In Persian].
2. Rezaei Hachesu P, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. Healthc Inform Res 2013; 19(2): 121-9. [In Persian].
3. Wright SP, Verouhis D, Gamble G, Swedberg K, Sharpe N, Doughty RN. Factors influencing the length of hospital stay of patients with heart failure. Eur J Heart Fail 2003; 5(2): 201-9.
4. Talebizadeh N, Haghdoost A, Mirzazadeh A. Age at natural menopause, An epidemiological model (Markov Chain) of cardiovascular disease in Iran. Payesh Health Monit 2009; 8(2): 163-70. [In Persian].
5. Laurenti R, Buchalla CM, de Sousa Caratin CV. Ischemic heart disease. hospitalization, length of stay and expenses in Brazil from 1993 to 1997. Arq Bras Cardiol 2000; 74(6): 488-92.
6. Samadi S, Nazifi M, Abbaspour S. Estimating the duration of treatment and hospitalization costs using neural network approach. Health Inf Manage 2011; 8(7): 948-85. [In Persian].
7. National Center for Health Statistics, Centers for Disease Control and Prevention. ICD-10-CM tabular list of diseases and injuries [Online]. [cited 2010]; Available from: URL: https://www.cms.gov/medicare/coding/icd10/downloads/6_i10tab2010.pdf
8. Han J, Pei J, Kamber M. Data Mining: Concepts and techniques. Philadelphia, PA: Elsevier; 2011.
9. Fetter RB. The new ICD-9-CM diagnosis-related groups classification scheme. 1st ed. Washington, DC: Government Printing Office; 1983. p. 9-67.
10. World Health Organization. International statistical classification of diseases and related health problems. Geneva,

www.mui.ac.ir
www.SID.ir
Archive of SID

Switzerland: WHO; 2016.
11. Monajemi F, Safdari R, Ghorbani V. Necessity of national DRG system for the cardiovascular diseases in Iran. J Health Adm 2009; 12(37): 65-9. [In Persian].
12. Kuwabara K, Imanaka Y, Matsuda S, Fushimi K, Hashimoto H, Ishikawa KB, et al. Impact of age and procedure on resource use for patients with ischemic heart disease. Health Policy 2008; 85(2): 196-206.
13. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: Current issues and guidelines. Int J Med Inform 2008; 77(2): 81-97.
14. Appelros P. Prediction of length of stay for stroke patients. Acta Neurol Scand 2007; 116(1): 15-9.
15. Lazar HL, Fitzgerald C, Gross S, Heeren T, Aldea GS, Shemin RJ. Determinants of length of stay after coronary artery bypass graft surgery. Circulation 1995; 92(9 Suppl): II20-II24.
16. Jiang X, Qu X, Davis LB. Using data mining to analyze patient discharge data for an urban hospital. Proceedings of the 2010 International Conference on Data Mining, DMIN; 2010 July 12-15; Las Vegas, Nevada, NV.
17. Wu W, Masri A, Popovic ZB, Smedira NG, Lytle BW, Marwick TH, et al. Long-term survival of patients with radiation heart disease undergoing cardiac surgery: A cohort study. Circulation 2013; 127(14): 1476-85.

Archive of SID

www.SID.ir

Ischemic Heart Patients' Length of Stay Estimation and Identification of Its Influencing Factors Using Data Mining

Majid Zarabian[1], Masoud Abessi[2]

Original Article

Abstract

Introduction: Ischemic Heart Disease (IHD) is one of the costly and controversial topics in the field of healthcare in Iran. Due to limitation in hospital resources for patient care, studying patient's length of stay (LOS) is very important in hospital management. This study presents suitable models for estimating the LOS of IHD patients and its influencing factors.

Methods: In this applied research, LOS of 6524 IHD in-patients admitted to 16 hospitals in Tehran, Iran, between October 2013 and March 2014 and the remedial proceedings provided for them were recorded using treatment services tariff codes. After data collection using a predesigned form and data cleaning, LOS fort models were created using data mining algorithms of artificial neural network (ANN), support vector machines (SVM), chi-squared automatic interaction detection (CHAID), classification and regression trees (CART), and ensemble model in SPSS Clementine software.

Results: The average and standard deviation of LOS in this study was 7.727 ± 9.608 days. The linear correlation of models with actual LOS and their relative error were above 0.7 and below 0.5, respectively. The most accurate models were the ensemble model and SVM.

Conclusion: According to the proposed models, ischemic patients who required rehabilitation, consultation, radiation therapy, and computerized tomography (CT) scan have longer LOS. Moreover, type of IHD disease and especially California remedial codes are important factors in estimating patients' LOS.

Keywords: Length of Stay; Ischemic Heart Disease; Data Mining

Received: 03 Dec., 2016 Accepted: 16 Apr., 2017

Citation: Zarabian M, Abessi M. **Ischemic Heart Patients' Length of Stay Estimation and Identification of Its Influencing Factors Using Data Mining.** Health Inf Manage 2017; 14(1): 16-25.

Archive of SID

Article resulted from MSc thesis No. 2358611 funded by Iran Health Insurance Organization.

1- MSc Student, Industrial Engineering, Department of Industrial Engineering, School of Engineering, University of Yazd, Yazd, Iran (Corresponding Author) Email: zarabian_majid90@yahoo.com

2- Assistant Professor, Industrial Management, Department of Industrial Engineering, School of Engineering, University of Yazd, Yazd, Iran

www.SID.ir