

طراحی مدل تشخیصی HTLV1 با استفاده از شمارش کامل سلول‌های خونی

معصومه سرباز*، امید پورنیک، لیلا قالیچی، خلیل کیمیافر، امیررضا رضوی

*دانشجوی دکترای انفورماتیک پزشکی عضو هیئت علمی دانشکده علوم پیراپزشکی مشهد

sarbazm@mums.ac.ir

چکیده

عفونت با ویروس HTLV-I که در بعضی از نقاط ایران به صورت اندمیک دیده می‌شود در اکثر موارد بدون علامت بوده و در عده کمی از بیماران به انواع بدخیمی‌ها و بیماری‌های عصبی منتهی می‌شود. طراحی مدلی برای غربالگری افراد بویژه در مناطق اندمیک می‌تواند به کشف زود هنگام موارد ابتلا و کاهش بار بیماری منجر شود. برای این کار نتایج ۵۹۹ آزمایش شمارش کامل سلول‌های خونی افراد سالم، افراد مبتلا به انواع لوکمی و HTLV-1 مورد بررسی قرار گرفت. مدل‌سازی با استفاده از متد CHAID انجام شد. مدل نهایی بر اساس متغیرهای تعداد گلبول‌های سفید، تعداد پلاکت‌ها و درصد ائوزینوفیل به صحت ۹۱٪ دست یافت. با اجرای این مدل بر روی نتایج CBC افراد بدون علامت یا بیماران متفرقه در مناطق اندمیک کشور می‌توان افراد ناقل احتمالی را کشف و برای آزمایش‌های تکمیلی معرفی کرد. با توجه به احتمال بروز عوارض متعدد در افراد مبتلا، کشف این موارد می‌تواند به کاهش بار بیماری و ارتقای سطح سلامت افراد به ویژه در مناطق اندمیک منجر شود.

واژه‌های کلیدی: HTLV-I، داده‌کاوی، CHAID، درخت تصمیم (Decision Tree)

مقدمه

ممکن است تظاهرات عصبی و خونی مشاهده شود. (۳) با توجه به اینکه عوارض این بیماری دیر هنگام ظاهر می‌شود با توجه به نمونه‌های خون می‌توان بررسی‌های بیشتری در این زمینه انجام داد. تشخیص اولیه بیماری در نمونه خون با بهره‌گیری از تست‌های الایزا و تایید تشخیص با کمک تست Western Blot صورت می‌گیرد. شیوع نسبتاً پایین این بیماری در بسیاری از موارد تشخیص آن را به تعویق می‌اندازد و هزینه‌های بسیاری را برای بیمار و نظام سلامت ایجاد می‌کند.

محیط مراقبت سلامت غنی از اطلاعات و ضعیف از دانش است و در سیستم‌های مراقبت سلامت و آزمایشگاه‌های تشخیصی، داده‌های ارزشمندی در دسترس می‌باشد. (۴) داده‌کاوی از طریق روش‌های مختلف مانند تفسیر تست‌های تشخیصی پیچیده، ترکیب اطلاعات از منابع مختلف (تصاویر، داده‌های بالینی، تصاویر متحرک از نمونه‌ها)، فراهم کردن حمایت‌های لازم برای تشخیص افتراقی و پیش‌آگهی بیمار به پزشکان کمک می‌کند تا روابط پنهان میان داده‌ها و الگوهای مختلف آن پی‌ببرد.

با توجه به اهمیت و بار بیماری‌های ناشی از عفونت با HTLV-1 و محدودیت‌های موجود در مسیر تشخیص آن، داده‌کاوی داده‌های آزمایشگاهی پایه در مبتلایان به این بیماری می‌تواند نتایج ارزنده‌ای را برای نظام سلامت کشور فراهم کند. در این مطالعه

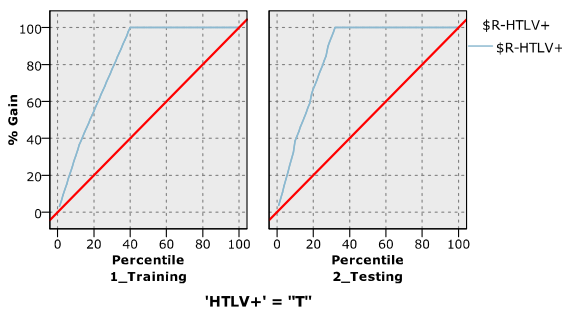
HTLV-I یک رتروویروس است که با انواع مختلفی از بیماری‌ها از جمله بعضی از انواع لنفوم، لوکمی و اختلالات عصبی دیده می‌شود. این ویروس در بعضی از نقاط جهان از جمله جنوب غرب ژاپن اندمیک است. عفونت با این ویروس در ایران نیز در بعضی از نواحی استان خراسان به صورت اندمیک دیده می‌شود. (۱،۲) عفونت با این ویروس منجر به رشد کلونال لنفوسیت CD4 همراه با تجمع مونوکلونال پرو ویروس HTLV در سلول‌های توموری می‌شود. شایع‌ترین یافته‌های بالینی این بیماری لنفادنوپاتی، هیپاتواسپلنومگالی، ضایعات پوستی، هیپرکلسمی و ضایعات لیتیک استخوانی است. از آن جا که این ویروس همانند سایر رتروویروس‌های هم‌خانواده، لکوسیت‌ها را مورد تهاجم قرار می‌دهد و لکوسیت‌ها هستند که با داشتن اطلاعات ژنتیک ویروس می‌توانند در انتشار و انتقال آلودگی نقش داشته باشند، انتشار HTLV-I بیشتر از طریق فرآورده‌های سلولی خون آماده، سوزن‌های آلوده به ویروس، تماس جنسی و از مادر به فرزند از طریق شیر مادر است. در این بیماری احتمال ابتلای به لوسمی - لنفوم سلول T در بین افرادی که از نظر سرولوژی مثبت هستند حدود یک درصد می‌باشد بنابراین بیشتر افراد آلوده به صورت ناقل باقی می‌مانند. عمده ناقلین HTLV1 بدون علامت هستند و تنها در عده کمی از بیماران

^۱The Human T-lymphotropic virus

جدول ۱- میانگین متغیرها در گروه‌های مختلف و سطح معنی‌داری تفاوت بین گروه‌های مختلف بر اساس آنالیز واریانس

	HT LV1	Leuk emia	Hea lthy	P value
WBC(10 000)	13.7	66.2	8.7	0.698
RBC	4.0	3.6	4.6	<0.001
Hgb	11.6	9.8	13.2	<0.001
MCV	88.7	89.7	85.1	0.002
MCH	29.2	42.3	28.9	0.366
MCHC	34.0	32.9	33.9	0.797
Platelet	178	1009	292	0.286
Lymphocyte	29.3	27.8	32.4	0.142
Monocyte	4.7	5.4	6.2	0.504
Eosinophil	3.6	4.8	2.7	0.074

مدل‌سازی با روش CHAID و با $\text{Alpha for Splitting} = 0.05$ و $\text{Alpha for Merging} = 0.05$ و محاسبه کای‌دو به کمک تکنیک پیرسون برای متغیر هدف طبقه‌بندی شده انجام گردید. مراحل تفکیک گروه‌ها یا رشد درخت تصمیم تا رسیدن به زمانی که حداقل یک درصد موارد در شاخه والد باقی بماند ادامه یافت. در مدل نهایی که دارای صحت ۹۱ درصد بود (شکل ۱).



شکل ۱- Gain Chart ارزیابی عملکرد مدل

متغیرهای تعداد گلبول‌های سفید، تعداد پلاکت‌ها و درصد آنوزینوفیل‌ها به کار گرفته شده‌اند و درخت تصمیم نهایی دارای ۴ سطح (Tree depth = ۴) بود (شکل ۲).

قصد داریم با داده‌کاوی بر روی نتایج آزمایش کامل خون^۲ مدلی برای پیشنهاد ابتلا به عفونت با این ویروس مطرح کنیم که بتواند در غربالگری مبتلایان و انتخاب موارد محتمل برای انجام آزمایش‌های تکمیلی به کار رود.

روش کار

برای انجام این مطالعه اطلاعات مربوط به نتایج ۵۹۹ مورد آزمایش CBC شامل ۲۴۶ آزمایش از بیماران مبتلا به HTLV1، ۱۴۲ آزمایش از بیماران مبتلا به ALL، ۱۱۰ آزمایش از بیماران مبتلا به ATL و ۱۰۱ آزمایش از افراد سالم جمع‌آوری گردید. این مجموعه آزمایش مربوط به ۱۰۷ نفر بیمار مبتلا به HTLV1، ۵۰ بیمار مبتلا به ALL، ۴۴ بیمار مبتلا به ATL و ۱۰۱ فرد سالم بود. نتایج آزمایش‌های افراد سالم از یکی از آزمایشگاه‌های تشخیص طبی و نتایج آزمایش‌های بیماران از سیستم اطلاعات بیمارستانی بیمارستان‌های قائم و امام رضا(ع) مشهد استخراج شد. داده‌ها از نظر کیفیت مورد بررسی قرار گرفت و موارد خارج از محدوده قابل قبول بازبینی شد.

متغیر هدف در مدل‌سازی ابتلا به HTLV1 بود. مدل‌سازی بر روی این داده‌ها با استفاده از متد CHAID و به کمک نرم‌افزار Clementine نسخه ۱۲ انجام شد. روش CHAID که مختصر شده عبارت Chi-squared Automatic Interaction Detector می‌باشد یکی از انواع درخت تصمیم است که بر پایه (Bonferroni testing) عمل می‌نماید.^(۵)

این داده‌ها با حفظ نسبت میان خوشه‌ها به دو نیمه تقسیم شد و نیمی از داده‌ها برای آموزش و نیمه دوم برای آزمون مدل به کار رفت. مدل طراحی شده بهینه‌سازی شد و از نظر کیفیت با مدل‌های دیگر مقایسه گردید.

یافته‌ها

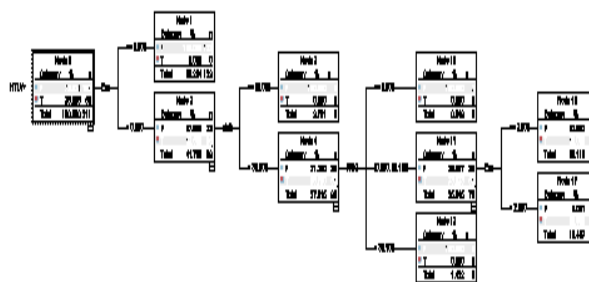
بانک داده‌های مورد استفاده در این مطالعه محتوی نتایج آزمایش CBC ۳۰۲ نفر شامل ۹۴ بیمار مبتلا به لوکمی، ۱۰۱ فرد نرمال و ۱۰۷ بیمار مبتلا به HTLV1 بود. متغیرهای مورد بررسی شامل تعداد گلبول‌های سفید، تعداد گلبول‌های قرمز، هموگلوبین، هماتوکریت، MCV، MCH، MCHC، پلاکت، درصد پلی مورفونوکلرها، درصد لنفوسیت‌ها، درصد مونوسیت‌ها و درصد آنوزینوفیل‌ها بود. نتایج آنالیز اولیه داده‌ها با تکنیک ANOVA و توزیع متغیرهای مذکور در گروه‌های مختلف در جدول شماره ۱ ذکر شده است.

^۲Complete Blood Count

همچنین پیش نهاد می‌شود با استفاده از داده های بیشتر امکان توسعه روش و کیفیت مدل فراهم شود.

نتیجه گیری

مدل های تصمیم گیری مبتنی بر داده کاوی و به طور خاص تکنیک به کار گرفته شده در این مقاله و درخت تصمیم به دست آمده می‌توانند در شناسایی بیماران مبتلا به بیماری HTLV1 کمک کننده بوده و به پزشکان و متخصصین در شناسایی سریع تر و آسان تر این افراد کمک نماید.



شکل ۲- مدل نهایی به دست آمده در قالب درخت تصمیم

بحث

مدل نهایی به دست آمده در این مطالعه دارای صحت ۹۱ درصد در تشخیص افراد مبتلا به HTLV1 بود و این امر تنها با داشتن اطلاعات موجود در یک آزمایش ساده CBC-diff محقق می‌شود. به این ترتیب این مدل می‌تواند در شناسایی این بیماران با سهولت به کار گرفته شود و موارد مناسب برای انجام آزمایش‌های تکمیلی را معرفی کند.

این مدل بر اساس روش CHAID ساخته شده که یکی از انواع روش‌های مبتنی بر درخت تصمیم می‌باشد. مزیت کلی این روش در سادگی نمایش کلیه مراحل طبقه‌بندی و امکان تفسیر روش کار با استخراج قوانین ساده عمل‌کردی آن است. همین امر بهره‌گیری از مدل به دست آمده برای کار بر بالین بیمار و حتی در تشخیص سرپایی مراجعین ممکن می‌سازد.

با اجرای این مدل بر روی نتایج CBC افراد بدون علامت یا بیماران متفرقه در مناطق اندمیک کشور می‌توان افراد ناقل احتمالی را کشف و برای آزمایش‌های تکمیلی معرفی کرد. با توجه به احتمال بروز انواع بدخیمی‌ها و اختلالات سیستم اعصاب در افراد مبتلا، کشف این موارد می‌تواند به کاهش بار بیماری‌های مذکور و ارتقای سطح سلامت افراد به ویژه در مناطق اندمیک منجر شود.

References

- [1] Boostani, R., Ardakani, A. M., and Ashrafi, H., "Khorasan Disease: Prevalence of HTLV-I Associated Myelopathy/Tropical Spastic Paraparesis (HAM/TSP) In West Azarbaijan from 2004 to 2007," Iranian Red Crescent Medical Journal, vol. 13, pp. 428-430, 2011
- [2] Abbaszadegan, M. R., Gholamin, M., Tabatabaee, A., Farid, R., Houshmand, M., and Abbaszadegan, M., "Prevalence of human T-lymphotropic virus type 1 among blood donors from Mashhad, Iran," Journal of clinical microbiology, vol. 41, p. 2593, 2003.
- [3] Iwanaga, M., Watanabe, T., Utsunomiya, A., Okayama, A., Uchimar, K., Koh, K. R., Ogata, M., Kikuchi, H., Sagara, Y., and Uozumi, K., "Human T-cell leukemia virus type I (HTLV-1) proviral load and disease progression in asymptomatic HTLV-1 carriers: a nationwide prospective study in Japan," Blood, vol. 116, p. 1211, 2010.
- [4] Jyoti, S., Ujma, A., Dipesh, S., and Sunita, S., "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction," International Journal of Computer Applications, vol. 17, pp. 43-48, 2011.
- [5] Berka, P., Rauch, J., and Zighed, D. A., Data Mining and Medical Knowledge Management: Cases and Applications: Information Science Reference-Imprint of: IGI Publishing, 2009.