

## پیش‌بینی بیماری‌های کبدی با استفاده از مدل مارکف پنهان

**فرزان مددی زاده:** دانشجوی دوره دکتری تخصصی آمار زیستی، مرکز تحقیقات بیماری‌های غیرواگیر، دانشگاه علوم پزشکی فسا، فسا، ایران و گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت و انستیتو تحقیقات بهداشتی، دانشگاه علوم پزشکی تهران، تهران، ایران. fmadadizadeh@razi.tums.ac.ir

**میترا منتظری:** کارشناسی ارشد هوش مصنوعی، پژوهشکده آینده پژوهی در سلامت، مرکز تحقیقات انفورماتیک پزشکی، دانشگاه علوم پزشکی کرمان، کرمان، ایران. mmontazeri@yahoo.com

\* **عباس بهرام پور:** استاد آمار زیستی، مرکز تحقیقات مدل‌سازی در سلامت، پژوهشکده آینده پژوهی در سلامت، دانشگاه علوم پزشکی کرمان، کرمان، ایران، و گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی کرمان، کرمان، ایران (\*نویسنده مسئول). abahrampour@yahoo.com

تاریخ پذیرش: ۹۵/۲/۲۵

تاریخ دریافت: ۹۴/۱۱/۱۷

## چکیده

**زمینه و هدف:** کبد بزرگترین اندام درونی و مهم‌ترین عضو پس از قلب و مغز در بدن انسان است و بدون آن ادامه حیات غیر ممکن است. تشخیص بیماری‌های کبدی نیازمند زمان طولانی و تخصص کافی پزشک معالج دارد. روش‌های آماری می‌توانند همانند یک سیستم پیش‌بینی اتوماتیک در جهت تشخیص دقیق و سریع بیماری‌های کبد به پزشکان متخصص کمک کنند. مدل مارکف پنهان از روش‌های هوشمند و قوی آمار است که در راستای این هدف در پژوهش حاضر به کار رفته است.

**روش کار:** داده‌های این پژوهش مقطعی، اطلاعات پرونده بیمارانی است که به ۵ نوع مختلف بیماری‌های کبد از جمله سیروز کبدی، سرطان کبد، هپاتیت حاد، هپاتیت مزمن و کبد چرب مبتلا و در سال‌های ۱۳۸۵ تا ۱۳۹۲ در بیمارستان افضلی پور کرمان بستری بوده‌اند. مدل مارکف پنهان با الگوریتم آموزشی EM به داده‌ها برازش و برای ارزیابی عملکرد آن معیارهای دقت، حساسیت و ویژگی استفاده شد.

**یافته‌ها:** میزان دقت، حساسیت و ویژگی مدل برای تشخیص هر بیماری کبدی جداگانه محاسبه شد و بالاترین مقدار آن‌ها در تشخیص سیروز کبدی با میزان دقت ۰/۸۷، حساسیت ۰/۸۲ و ویژگی ۰/۹۶ بدست آمد. همچنین کمترین میزان هم در تشخیص کبد چرب با دقت ۰/۶۵، حساسیت ۰/۶۹ و ویژگی ۰/۹۴ بود.

**نتیجه‌گیری:** نتایج این پژوهش حاکی از توانمندی بالقوه مدل مارکف پنهان است، لذا به کارگیری این مدل در زمینه تشخیص بیماری‌های کبدی پیشنهاد می‌شود.

**کلیدواژه‌ها:** مدل مارکف پنهان، الگوریتم EM، بیماری‌های کبدی، پیش‌بینی

## مقدمه

کبد به وزن تقریبی ۱۵۰۰ گرم، بزرگترین اندام درونی و مهم‌ترین عضو پس از قلب و مغز در بدن انسان است که بدون آن ادامه حیات غیرممکن است (۱). بیماری‌های کبدی در شمار ده بیماری‌کننده در جهان قرار داشته، در ایران حدود ده درصد کل مرگ‌ها و دراروپا پنجمین عامل مرگ و میر بعد از ناراحتی قلبی، سرطان، سگته و بیماری‌های تنفسی به شمار می‌روند (۲). تشخیص زودهنگام بیماری‌های کبدی برای درمان آن‌ها امری ضروری است. هرچند پیشرفت‌های شگرفی در زمینه علم پزشکی صورت گرفته است اما همچنان تشخیص زودهنگام بیماری‌های کبدی امری دشوار است و به دست

آوردن تجربه کامل در زمینه تشخیص بیماری‌های کبدی برای یک پزشک متخصص نیاز به بازه زمانی طولانی و کار عملی بسیار دارد. وجود یک سیستم کمک یار پزشک که قابلیت پیش‌بینی دقیق و سریع بیماری‌های کبدی را داشته باشد ضروری به نظر می‌رسد (۳).

روش‌های آماری می‌توانند در قالب یک سیستم هوشمند به کمک متخصصان بشتابند. در سال‌های اخیر کاربرد روش‌های آماری و ابزارهای مدل‌بندی در حوزه‌های مختلف به‌طور گسترده در حال افزایش است و گرایش به سمت این روش‌ها افزایش یافته است (۴). در این راستا روش آماری مدل مارکف پنهان از مفاهیمی بوده که نظر بسیاری از پژوهشگران را به خود جلب کرده و به

جهت آنکه بر اساس یادگیری و کشف رابطه نهفته در فرآیند به حل مسئله می‌شتابد در زمره روش‌های داده‌کاوی، یادگیری ماشین و هوشمند نیز محسوب شده است (۵-۷). مدل مذکور به‌عنوان ابزاری در حل مسائلی مانند طبقه‌بندی و پیش‌بینی که توسط روش‌های سنتی گذشته قابل حل نبوده و یا برای پزشکان متخصص امری زمان‌بر بوده است استفاده شده است. می‌توان از این روش در افتراق و طبقه‌بندی بیماری‌های کبدی نیز استفاده کرد که این به‌نوبه خود باعث کاهش در هزینه‌های آزمایشگاهی و در مواردی تسریع در عمل غربالگری بیماری و کمک به متخصصان می‌شود. هدف اصلی این پژوهش، تشخیص بیماری‌های کبد با استفاده از مدل مارکف پنهان و ارزیابی قدرت پیش‌بینی این مدل با استفاده از معیارهای ارزیابی مانند حساسیت، ویژگی و دقت می‌باشد. در این راستا جهت افزایش دقت و عملکرد بهتر این مدل از الگوریتم امید ریاضی بیشینه (Expectation Maximization algorithm) استفاده می‌شود.

بیشتر مطالعات مبنی بر مدل مارکف پنهان در زمینه‌های غیرپزشکی (مثلاً در شناسایی گفتار، یادگیری ماشین، بیوانفورماتیک و...) صورت گرفته است (۸-۱۳) و کاربرد آن در زمینه پزشکی بسیار اندک بوده است. همچنین مطالعه حاضر در زمره اولین مطالعاتی است که از مدل مارکف پنهان جهت پیش‌بینی بیماری‌ها در حوزه پزشکی استفاده می‌کند.

در مطالعه ای که توسط هیو کیلی و همکاران (Hu Kili) در سال ۲۰۱۳ انجام شد مدل مارکف پنهان پیوسته برای طبقه‌بندی صداهای خر و پف ۲۱ بیمار که اختلال خواب داشتند به کار گرفته شد در این مطالعه صدای خر و پف بیماران توسط سنسورهایی که به گردن آن‌ها آویخته شده بود ضبط گردیده و مورد بررسی و طبقه‌بندی قرار گرفت و در نهایت حساسیت مدل ۹۳/۳ درصد به دست آمد (۱۶).

در مطالعه ای که توسط رافعی و همکاران در سال ۲۰۱۲ صورت گرفت از مدل مارکف پنهان جهت کشف حالات سل ریوی در ایران استفاده شد داده‌های این مطالعه اطلاعات بیمارانی بود که از سال ۲۰۰۵ الی ۲۰۱۱ در ایران گردآوری شده بود ۲ مدل مارکف پنهان یکی با گرایش فصلی و دیگری با گرایش غیر فصلی استفاده شد که در نهایت با استفاده از معیارهای برازش مدل  $R^2$  و  $BIC$  آن‌ها مقایسه شد که در مدل فصلی این معیارها به ترتیب ۰/۷۲ و ۱۳۳۶/۶۶ و در مدل غیر فصلی به ترتیب برابر ۰/۵۶ و ۱۳۸۶/۷۵ و همچنین حساسیت و ویژگی مدل فصلی برابر ۰/۹۴ و ۰/۷۳ و در مدل غیر فصلی ۰/۶۸ و ۰/۷۳

در مطالعه ای که توسط ویمالا (Vimala K) و همکاران در سال ۲۰۱۴ صورت گرفت مدل مارکف پنهان جهت تشخیص و طبقه‌بندی سیگنالهای نوار قلب (ECG) به کار رفت. داده‌های مورد استفاده در این مطالعه از بیمارستان‌های وابسته به موسسه فناوری ماساچوست (MIT) آمریکا تهیه شده بود که شامل اطلاعات ثبت شده از ۸۸ فرد مراجعه کننده به مرکز که ۱۸ نفر آن‌ها دارای وضعیت نرمال، ۲۲ نفر داری استرس و ۴۸ نفر دارای بیماری قلبی بودند. مدل مارکف پنهان پس از طبقه‌بندی داده‌های نوار قلب در سه دسته نرمال، دارای استرس و بیمار قلبی، دارای دقت

بیشتر مطالعات مبنی بر مدل مارکف پنهان در زمینه‌های غیرپزشکی (مثلاً در شناسایی گفتار، یادگیری ماشین، بیوانفورماتیک و...) صورت گرفته است (۸-۱۳) و کاربرد آن در زمینه پزشکی بسیار اندک بوده است. همچنین مطالعه حاضر در زمره اولین مطالعاتی است که از مدل مارکف پنهان جهت پیش‌بینی بیماری‌ها در حوزه پزشکی استفاده می‌کند.

در مطالعه ای که توسط ویمالا (Vimala K) و همکاران در سال ۲۰۱۴ صورت گرفت مدل مارکف پنهان جهت تشخیص و طبقه‌بندی سیگنالهای نوار قلب (ECG) به کار رفت. داده‌های مورد استفاده در این مطالعه از بیمارستان‌های وابسته به موسسه فناوری ماساچوست (MIT) آمریکا تهیه شده بود که شامل اطلاعات ثبت شده از ۸۸ فرد مراجعه کننده به مرکز که ۱۸ نفر آن‌ها دارای وضعیت نرمال، ۲۲ نفر داری استرس و ۴۸ نفر دارای بیماری قلبی بودند. مدل مارکف پنهان پس از طبقه‌بندی داده‌های نوار قلب در سه دسته نرمال، دارای استرس و بیمار قلبی، دارای دقت

مدل مارکف پنهان یک روش آماری است که بر پایه زنجیر مارکف استوار است و در اواخر دهه ۱۹۶۰ معرفی شده است. این مدل ابزاری قدرتمند در پردازش و مدل بندی فرآیندهای تصادفی محسوب می شود و به عنوان یک روش در پیش بینی و طبقه بندی داده ها در بین روش های آماری مطرح است (۲۰).

یک مدل مارکف پنهان را می توان با تعیین پارامترهای زیر ایجاد نمود:

**تعداد حالات ممکن**  $S = \{s_1, \dots, s_N\}$ : تعداد حالات (وضعیت های) بیماری است. در این مطالعه بیماری های کبدی دارای ۵ حالت مختلف می باشد پس  $S = \{s_1, \dots, s_5\}$  است.

**تعداد مشاهدات** در هر حالت  $Q = \{x_1, \dots, x_n\}$ : برای هر حالت بیماری ذکر شده در بالا تعدادی مشاهده (ویژگی) داریم. در این مطالعه مشاهدات همان پارامترهای آزمایشگاهی ثبت شده برای هر فرد مبتلا به بیماری کبدی است.

**احتمالات انتقال**  $a_{ij}$ : احتمال عبور از بیماری حالت  $i$  به بیماری حالت  $j$

$$1 \leq i, j \leq N = 5 \\ \sum_{j=1}^N a_{ij} = 1, a_{ij} \geq 0 \quad a_{ij} = p(s_j | s_i)$$

**ماتریس احتمالات انتقال**  $A = [a_{ij}]$ : یک مجموعه از احتمالات انتقال در بین حالتهاست. در این مطالعه چون بیماری های کبدی ۵ حالت دارد ماتریس احتمال انتقال به شکل زیر در می آید:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ 2 & a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ 3 & a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ 4 & a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ 5 & a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix}$$

در مطالعه حاضر حالت ۱ تا ۵ به ترتیب مربوط به بیماری های کبد چرب، هپاتیت مزمن، هپاتیت حاد، سرطان کبد و سیروز کبدی است.

**احتمال توزیع مشاهدات**  $b_{jk}$ : احتمال تعلق

به دست آمد که در نهایت بهترین مدل همان مدل مارکف پنهان فصلی گزارش شد (۱۷).

وو هانگ (Wu, Hang) و همکاران در سال ۲۰۱۰ به مطالعه ای تحت عنوان شناسایی بیماری های قلبی از طریق شناسایی صدا های قلب توسط مدل مارکف پنهان پرداختند. داده های مطالعه شامل ۳۲۵ صدای قلب ضبط شده از ده نوع بیماری قلبی بود. برای ساخت مدل مارکف پنهان تعداد حالات مختلفی بین ۲ تا ۷ حالت امتحان شد و در نهایت تعداد حالات برابر ۴ حالت در نظر گرفته شد زیرا در این وضعیت مدل بالاترین دقت برابر ۰/۹۵۰۸ را دارا بود (۱۸).

در مطالعه ای دیگر توسط اسچمیدت (Schmidt S) و همکارانش در سال ۲۰۰۸ تحت عنوان؛ جداسازی و تشخیص صداهای قلب ضبط شده توسط دستگاه سیتوسکوپ (Stethoscope) با استفاده از مدل مارکف پنهان؛ صورت گرفت. داده ها شامل ۱۰۰ صدای ضبط شده از بیماران قلبی مختلف بود. برای ساخت مدل مارکف پنهان ۴۰ درصد (۴۰ نفر) از داده ها برای مرحله آموزش مدل و ۶۰ درصد (۶۰ نفر) از داده ها برای مرحله ارزیابی مدل استفاده و تعداد حالات برای مدل برابر ۴ حالت در نظر گرفته شد. در پایان نتایج نسبتاً مطلوبی مانند حساسیت برابر ۰/۵۹۵ و ارزش اخباری مثبت برابر ۰/۵۵ به دست آمد که حاکی از آن بود که مدل مربوطه چندان هم بد نیست (۱۹).

## روش کار

داده های استفاده شده در این پژوهش بخشی از اطلاعات پرونده بیمارانی می باشد که به ۵ نوع مختلف بیماری های کبد از جمله سیروز کبدی، سرطان کبد، هپاتیت حاد، هپاتیت مزمن، کبد چرب مبتلا و در سال های ۱۳۸۵ الی ۱۳۹۲ در بیمارستان افضل پور کرمان بستری بوده اند. تعداد کل پرونده ها برابر با ۱۳۰۰ مورد بود که از این تعداد ۱۱۴۳ مورد که دارای اطلاعات کامل بود جمع آوری شدند. روش آماری مورد استفاده مدل مارکف پنهان بوده است.

مدل مارکف پنهان (Hidden Markov model):

مورد نیاز مدل  $(\lambda = (A, B, \pi))$  هم اعداد تصادفی در بازه صفر و یک می‌دهیم الگوریتم پس از آموزش یافتن مقداری بهینه برای سه پارامتر ارائه می‌دهد. ارزیابی مدل از طریق الگوریتم پیش‌رو - پس‌رو (Forward-Backward) امکان پذیر است. در این روش مدل با دریافت تعدادی از مشاهدات یا ویژگی‌ها که در امر آموزش شرکت نداشته‌اند میزان تعلق این ویژگی‌ها به هر حالت مدل را با یک احتمال نشان می‌دهد. اگر هر حالت مدل به‌عنوان یک طبقه در نظر گرفته شود آن‌گاه الگوریتم ویتربی ماکزیمم این احتمالات تعلق را در نظر گرفته و طبقه‌ای (حالت) را که دارای ماکزیمم احتمال است به‌عنوان طبقه هدف پیش‌بینی می‌کند. سپس دقت مدل از مقایسه حالت پیش‌بینی شده و حالت واقعی مشاهدات برآورد می‌شود. معمولاً در قسمت آموزش مدل از ۷۵ درصد داده‌ها استفاده می‌شود و از ۲۵ درصد باقی‌مانده داده‌ها برای ارزیابی مدل استفاده می‌شود (۲۲). در این مطالعه به جای الگوریتم بام ولج از الگوریتم EM جهت آموزش مدل مارکف پنهان استفاده شد، شرط اولیه در مدل مارکف پنهان این است که الگوریتم تا زمانی ادامه یابد که مقدار برآورد شده برای سه پارامتر مدل به میزانی بهینه رسیده و با مقدار برآورد شده در تکرار قبلی الگوریتم تفاوت چندانی نداشته باشد (۲۴). لازم به ذکر است که ۷۵ درصد از داده‌ها (اطلاعات ۸۵۷ بیمار) جهت مرحله آموزش مدل مارکف پنهان و ۲۵ درصد باقی‌مانده داده‌ها (اطلاعات ۲۸۶ بیمار) جهت ارزیابی عملکرد مدل استفاده شد. با توجه به این‌که هدف اصلی در این پژوهش

داشتن هر مشاهده (ویژگی) به هر حالت را نشان می‌دهد:

$$b_{jk} = p(O = x_k | S = s_j) \\ \text{و } 1 \leq k \leq n, \quad 1 \leq j \leq N$$

$$b_{jk} \geq 0, \quad \sum_{k=1}^n b_{jk} = 1, \quad 1 \leq j \leq N$$

ماتریس احتمال توزیع مشاهدات  $B = [b_{jk}]$ : در این مطالعه ۱۰ مشاهده (پارامتر آزمایشگاهی) و ۵ حالت بیماری کبدی داریم بنابراین ماتریس احتمال توزیع مشاهدات به صورت زیر خواهد بود:

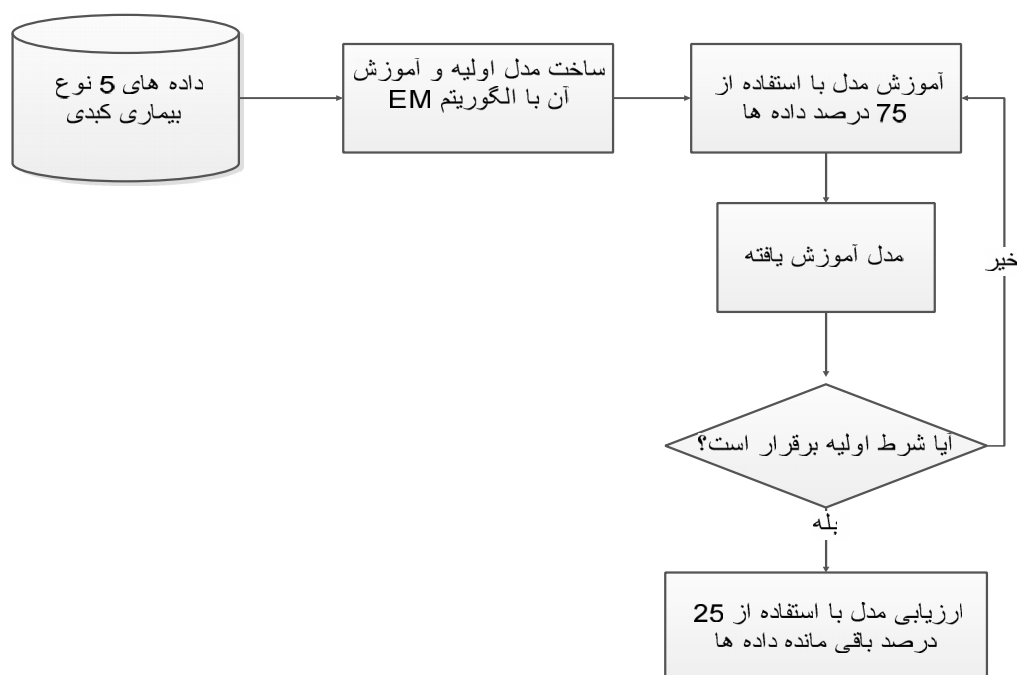
$$B = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \\ 1 & b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} & b_{17} & b_{18} & b_{19} & b_{110} \\ 2 & b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} & b_{27} & b_{28} & b_{29} & b_{210} \\ 3 & b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & b_{36} & b_{37} & b_{38} & b_{39} & b_{310} \\ 4 & b_{41} & b_{42} & b_{43} & b_{44} & b_{45} & b_{46} & b_{47} & b_{48} & b_{49} & b_{410} \\ 5 & b_{51} & b_{52} & b_{53} & b_{54} & b_{55} & b_{56} & b_{57} & b_{58} & b_{59} & b_{510} \end{bmatrix}$$

توزیع احتمال حالت آغازین  $\pi = \{\pi_0(1), \dots, \pi_0(N)\}$ : برای هر حالت بیماری یک احتمال اولیه ابتدا وجود دارد که زنجیر مارکف با آن احتمال اولیه از آن حالت شروع به کار می‌کند. به این ترتیب ما می‌توانیم یک مدل مخفی مارکف با توزیع احتمال گسسته را با مشخص کردن ۳ پارامتر  $\lambda = (A, B, \pi)$  به دست آوریم (۲۱).

مهمترین موضوع در کاربرد مدل مارکف پنهان موضوع آموزش و ارزیابی مدل است آموزش مدل مارکف پنهان از طریق الگوریتم بام ولج (Baum-Welch) امکان پذیر است. در این حالت علاوه بر ارائه اطلاعات بیماریها به مدل، به سه پارامتر

جدول ۱- متغیرهای مورد استفاده در مطالعه

ردیف	نام متغیر	ردیف	نام متغیر
۱-	تعداد گلبول‌های سفید	۸-	آنزیم آلکالین فسفاتاز
۲-	تعداد گلبول‌های قرمز	۹-	بیلی روبین کلی
۳-	هموگلوبین	۱۰-	بیلی روبین مستقیم
۴-	هماتوکریت	۱۱-	زمان نسبی ترومبوپلاستین
۵-	پلاکت	۱۲-	حجم متوسط سلول (MCV)
۶-	آنزیم آسپارات آمینوترانسفراز	۱۳-	متوسط وزن هموگلوبین در گلبول قرمز (MCH)
۷-	آنزیم آلانین آمینوترانسفراز	۱۴-	متوسط غلظت هموگلوبین هر گلبول قرمز (MCHC)



شکل ۱- فرآیند ساخت مدل مارکف پنهان

ریاضی (EM) استفاده و شرط توقف الگوریتم هم وجود تفاوتی به اندازه  $0/000001$  در برآوردهای جدید پارامترها نسبت به مقدار پیشینشان در نظر گرفته شد. این الگوریتم با ۱۰۰ بار تکرار متوقف و مقادیر بهینه برای پارامترهای مدل تولید شد. مقدار بهینه به دست آمده برای ماتریس احتمالات انتقال بین حالات مختلف بیماری‌های کبدی به شرح ذیل است (ماتریس A) که در آن حالت ۱ بیماری کبد چرب، حالت ۲ بیماری هیپاتیت مزمن، حالت ۳ بیماری هیپاتیت حاد، حالت ۴ بیماری سرطان کبد و حالت ۵ بیماری سیروز کبدی را نشان می دهد. برای مثال احتمال اینکه یک فرد دارای بیماری کبد چرب به بیماری سرطان کبد مبتلا شود حدود ۴۳ درصد است ( $a_{14} = 0.439$ ) و احتمال آن که بیماری اش دارای وضعیت ثابت باشد برابر ۴۱ درصد است ( $a_{11} = 0.417$ ).

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & a_{11} = 0.417 & a_{12} = 0.00025 & a_{13} = 0.00075 & a_{14} = 0.439 & a_{15} = 0.143 \\ 2 & a_{21} = 0.0033 & a_{22} = 0.0069 & a_{23} = 0.980 & a_{24} = 0.0043 & a_{25} = 0.0055 \\ 3 & a_{31} = 0.037 & a_{32} = 0.008 & a_{33} = 0.301 & a_{34} = 0.311 & a_{35} = 0.3428 \\ 4 & a_{41} = 0.108 & a_{42} = 0.0006 & a_{43} = 0.0010 & a_{44} = 0.514 & a_{45} = 0.371 \\ 5 & a_{51} = 0.031 & a_{52} = 0.0004 & a_{53} = 0.0008 & a_{54} = 0.335 & a_{55} = 0.6328 \end{bmatrix}$$

ارزیابی عملکرد مدل مارکف پنهان در پیش‌بینی و افتراق بیماری‌های کبدی می‌باشد، لازم است از معیارهای آماری مناسب جهت ارزیابی استفاده گردد تا در صورت لزوم بتوان با اطمینان بیشتر مدل مارکف پنهان را مورد استفاده قرار داد. بدین منظور از شاخص‌های حساسیت، ویژگی و دقت پیش‌بینی استفاده شد. تمامی تحلیل‌ها در نرم افزار MATLAB نسخه ۲۰۱۲ صورت گرفت. فرآیند ساخت مدل مارکف پنهان در شکل ۱ نمایش داده شده است.

### یافته‌ها

در این مطالعه اطلاعات ثبت شده ۱۱۴۳ بیمار مبتلا به انواع مختلف بیماری‌های کبد شامل اطلاعات ۱۰۴ بیمار کبد چرب، ۲۳۵ بیمار هیپاتیت مزمن، ۳۹ بیمار هیپاتیت حاد، ۱۵۲ بیمار سرطان کبد و ۶۱۳ بیمار سیروز کبدی مورد بررسی قرار گرفت. اطلاعات بیماران شامل ۱۴ متغیر آزمایشگاهی بود که در جدول ۱ توصیف شده اند.

تعداد حالات بهینه به کار رفته در مدل نهایی ۵ حالت در نظر گرفته، جهت آموزش مدل از ۷۵ درصد اطلاعات و الگوریتم پیشینه سازی امید

جدول ۲- توصیف نتایج نهایی حاصل از ارزیابی مدل مارکف پنهان

بیماری‌های کبدی					
معیارهای ارزیابی	کبد چرب	هپاتیت مزمن	هپاتیت حاد	سرطان کبد	سیروز
حساسیت	۰/۶۹۸۵	۰/۸۰۰۰	۰/۷۱۸۷	۰/۷۱۰۰	۰/۸۲۷۵
ویژگی	۰/۹۴۶۸	۰/۹۶۶۲	۰/۹۶۴۱	۰/۹۶۵۳	۰/۹۶۴۴
دقت	۰/۶۵۱۵	۰/۷۳۶۲	۰/۶۷۹۳	۰/۷۱۶۲	۰/۷۷۸۲

نتایج حاصل از ارزیابی مدل مارکف پنهان در جدول ۲ ذکر شده است.

### بحث و نتیجه‌گیری

کبد به لحاظ ساختاری عضوی پیچیده با عملکردهای مختلف است از این جهت متخصصین در تشخیص بیماری‌های این عضو بدن با کثرتی از داده‌ها رو به رو هستند و امر تشخیص زمان‌بر است روش‌های آماری مانند مدل مارکف پنهان می‌توانند با افتراق و طبقه‌بندی بیماری‌های مختلف کبد به‌عنوان کمک‌یار پزشک در زمان صرفه‌جویی به عمل آورند.

ساختار مدل مارکف پنهان در مقایسه با مدل‌های آماری مرسوم پیچیده‌تر است و این مدل‌ها در زمینه‌های گوناگون علوم کامپیوتر در تشخیص و بازشناسی گفتار به کار رفته‌اند و به ندرت در زمینه پزشکی به کار گرفته شده‌اند. به جرات می‌توان گفت مطالعه حاضر اولین مطالعه‌ای است که از مدل مارکف پنهان در زمینه پزشکی و در زمینه تشخیص بیماری‌های کبدی استفاده می‌کند.

مدل مارکف پنهان به‌عنوان یک سیستم پیشگوی قوی به خوبی به داده‌ها فیت می‌شود این عبارت توسط ویمالیا و همکاران در سال ۲۰۱۴ (۱۴) در هنگامی که از مدل مارکف پنهان جهت تشخیص سیگنال‌های نوارقلب استفاده کردند ذکر شد. در این مطالعه مدل مارکف پنهان بسیار فوق‌تصور و عالی به داده‌ها فیت شده و نتایج بسیار دقیقی ارائه داد در مطالعه حاضر هم به‌طور کلی عملکرد نسبتاً مطلوبی از مدل مارکف پنهان شاهد بودیم و نتایج ما هم موید نتایج ایشان است.

در مطالعه حاضر تعداد حالات مدل برابر با ۵ حالت در نظر گرفته شد این در حالی است که در

مطالعه هوآی مین لی و همکاران (۱۵) که از مدل مارکف پنهان برای پیش‌بینی روند پیشرفت بیماری سرطان استفاده نموده‌اند تعداد حالات برابر سه حالت در نظر گرفته شده است همچنین مشابه با مطالعه حاضر مقادیر اولیه پارامترهای مدل را عددی بین صفر و یک اتخاذ کرده‌اند در مجموع اگرچه حجم نمونه مطالعه ایشان از حجم نمونه مطالعه حاضر کمتر بوده اما نتایج قابل‌قبولی از ارزیابی مدل مارکف پنهان حاصل شده بود که موید نتایج مطالعه حاضر است.

در مطالعه حاضر مدل مارکف پنهان به کار رفته از نوع گسسته زمان بود در حالی که در مطالعه هیو کیلی و همکاران (۱۶) یک مدل مارکف پنهان پیوسته برای طبقه‌بندی صداهای خر و پف به کار گرفته شد که در پایان حساسیت بالایی برای مدل به دست آمد که حاکی از توانمندی بالای مدل مارکف پنهان در پیش‌بینی و طبقه‌بندی انواع مختلف داده‌هاست. حساسیت به دست آمده در مطالعه حاضر نیز تایید کننده این امر و هم‌سو با آن می‌باشد.

حداکثر میزان حساسیت و ویژگی در مطالعه حاضر در پیش‌بینی بیماری سیروز کبدی اتفاق افتاد اگرچه این میزان کمتر از مقدار حساسیت و ویژگی گزارش شده در مطالعه رافعی و همکاران (۱۷) بود اما به هر حال بالا بودن میزان حساسیت و ویژگی در هر دو مطالعه گویای توانمندی زاید الوصف مدل مارکف پنهان است که این نتایج تایید کننده نتایج مطالعه حاضر نیز می‌باشد.

دقت مطالعه حاضر هم‌سو با دقت مطالعه وو هانگ و همکاران (۱۸) بود و هر دو آن‌ها نشان دهنده دقت بالای مدل مارکف پنهان بود همچنین شباهت دیگر هر دو مطالعه انتخاب حالات مدل پنهان مارکف بر اساس وضعیت بیماری بود شاید

### تقدیر و تشکر

از زحمات پرسنل قسمت بایگانی اطلاعات بیمارستان افضلی پور کرمان و سایر بزرگوارانی که در این مطالعه ما را یاری نمودند سپاسگزاریم.

### منابع

1. Ramana BV, Babu MSP, Venkateswarlu N. A critical study of selected classification algorithms for liver disease diagnosis. *IJDMS*; 2011.3(2):101-114.
2. Ganji A, Safavi M, Nourai S, Nasser-Moghadam S, Merat S, Vahedi H, et al. Digestive and liver diseases statistics in several referral centers in Tehran, 2000-2004. *Govareh*; 2006.11(1):33-38.
3. Lin RH. An intelligent model for liver disease diagnosis. *Artif Intell Med*; 2009.47(1):53-62.
4. Wolfe R, Abramson MJ. Modern statistical methods in respiratory medicine. *Respirology*; 2014.19(1):9-13.
5. Beal MJ, Ghahramani Z, Rasmussen CE, editors. The infinite hidden Markov model. *Adv Neural Inf Process Syst*; 2001. 577-584.
6. Pimentel MA, Santos MD, Springer DB, Clifford GD. Heart beat detection in multimodal physiological data using a hidden semi-Markov model and signal quality indices. *Physiol Meas*; 2015.36(8):1717.
7. Zhou C, Huang S, Xiong N, Yang SH, Li H, Qin Y, et al. Design and Analysis of Multimodel-Based Anomaly Intrusion Detection Systems in Industrial Process Automation. *IEEE Trans Syst Man Cybern B Cybern*; 2015.45(10):1345-1360.
8. Jin J, Wang Y, Xu L, Cao L, Han L, Zhou B, et al. editors. Non-intrusive gesture recognition system combining with face detection based on Hidden Markov Model. *SPIE/COS Photonics Asia*; 2014. 731-735.
9. Hu M, Shen F, Zhao J, editors. Hidden Markov models based dynamic hand gesture recognition with incremental learning method. *Neural Networks (IJCNN)*, 2014 International Joint Conference on IEEE; 2014. pp.3108-3115
10. Geiger JT, Kneißl M, Schuller B, Rigoll G. Acoustic Gait-based Person Identification using Hidden Markov Models. *Workshop on Mapping Personality Traits Challenge and Workshop*. ACM; 2014. pp. 25-30
11. Farsi H, Saleh R, editors. Implementation and optimization of a speech recognition system based on hidden Markov model using genetic algorithm. In *Intelligent Systems (ICIS)*, Iranian Conference on IEEE; 2014. pp. 1-5
12. Cholewa M, Głomb P. Estimation of the number of states for gesture recognition with Hidden Markov Models based on the number of

همین امر موجب دقت بالای دو مدل شده است. کمترین میزان حساسیت مدل در مطالعه حاضر تقریباً ۷۰ درصد است که این میزان از حساسیت مدل مارکف پنهان که در مطالعه اسچمیدت و همکاران برای تشخیص صداهای قلب استفاده شد بیشتر است شاید دلیل آن هم حجم نمونه بیشتر در مطالعه ما باشد (۱۹).

از مزیت های اصلی این مطالعه گزارش و تفسیر ماتریس احتمالات انتقال است با توجه به آنکه تعداد حالات مدل برابر ۵ حالت بیماری در نظر گرفته شد برای اعداد ذکر شده در ماتریس احتمالات انتقال تفسیر مناسبی وجود دارد مثلاً  $a_{23} = 0.980$  بدان معناست که احتمال انتقال از حالت دوم (هپاتیت مزمن) به حالت سوم بیماری (هپاتیت حاد) در بیماران این مطالعه ۹۸ درصد بوده است. همچنین احتمالات کوچکتر از ۰/۱ در جدول بیانگر این موضوع است که احتمال انتقال بسیار ضعیف است مانند انتقال از بیماری کبدی به هپاتیت مزمن و حاد یا احتمال انتقال از بیماری سیروز کبدی به هپاتیت حاد و مزمن.

ارزیابی مدل برای پیش بینی نوع بیماری هر فرد با توجه به اطلاعات داده شده به مدل نیز صورت گرفت و نتایج حاکی از توانمندی بالای مدل مارکف پنهان بود به طوری که این توانمندی برای پیش بینی هر نوع بیماری کبدی متفاوت بود. در تشخیص بیماری سیروز کبدی مدل دارای بیشترین مقدار دقت (۷۸ درصد) و در تشخیص بیماری کبد چرب دارای کمترین میزان دقت (۶۵ درصد) بود.

یکی از محدودیت های اصلی این مطالعه عدم ثبت دقیق داده ها در پرونده بیماران کبدی بود که این به نوبه خود تاثیر چشمگیری بر دقت مدل نهایی داشت.

در نتیجه گیری از پژوهش حاضر می توان گفت تشخیص بیماری ها از طریق سیستم های هوشمند سریع تر، دقیق تر و کم هزینه تر است که این امر در مسائل بهداشتی و درمانی بسیار حائز اهمیت است. لذا به کار گیری مدل مارکف پنهان در زمینه تشخیص بیماری های کبدی پیشنهاد می شود.

critical points in time sequence. *Pattern Recognit Lett*; 2013.34(5):574-9.

13. Bilal S, Akmeliawati R, Shafie AA, Salami MJE. Hidden Markov model for human to computer interaction: a study on human hand gesture recognition. *Artificial Intelligence Review*; 2013.40(4):495-516.

14. Vimala K. Stress causing Arrhythmia Detection from ECG Signal using HMM. *IJIRCCCE*; 2014.2(10):6079-6085.

15. Li HM, Fang LY, Wang P, Yan JZ. Hidden Markov Models Based Research on Lung Cancer Progress Modeling. *Research Journal of Applied Sciences, Engineering and Technology*; 2013 6(13): 2470-2473.

16. Lee HK, Lee J, Kim H, Ha JY, Lee KJ. Snoring detection using a piezo snoring sensor based on hidden Markov models. *Physiol Meas*; 2013.34(5):41.

17. Orak R. Tuberculosis Surveillance Using a Hidden Markov Model. *IJPH*; 2012.41(10):87.

18. Wu H, Kim S, Bae K, editors. Hidden Markov Model with heart sound signals for identification of heart diseases. *Proceedings of 20th International Congress on Acoustics (ICA)*, Sydney, Australia; 23-27 August 2010.

19. Schmidt S, Toft E, Holst-Hansen C, Graff C, Struijk J, editors. Segmentation of heart sound recordings from an electronic stethoscope by a Hidden-Markov model. *Comput Cardiol*; 2008. 35:345-349

20. Rabiner L, Juang B. An introduction to hidden Markov models. *IASSP*; 1986.3(1):4-16.

21. Moridani M. Hidden Markov model and diagnostic cardiac anemia. *Journal of Biomedical Engineering and Laboratory Equipment*; 2010. 108:18.

22. Rastjou ardakani A. AaH. Evaluation of the hidden markov model for detection of P300 in EEG signals. *IJMP*; Fall 2008-Winter 2009.20-21:3.

## Predicting of liver disease using Hidden Markov Model

**Farzan Madadzadeh**, PhD student of Biostatistics, Noncommunicable Diseases Research Center, Fasa University of Medical Sciences, Fasa, Iran, & Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. fmadadzadeh@razi.tums.ac.ir

**Mitra Montazeri**, MSc in Artificial Intelligence, Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. mmontazeri@yahoo.com

\***Abbas Bahrapour**, Professor of Biostatistics, Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran, & Department of Epidemiology and Biostatistics, School of Health, Kerman University of Medical Sciences, Kerman, Iran (\*Corresponding author). abahrapour@yahoo.com

### Abstract

**Background:** The liver is the largest internal organ and the most important organ after heart and brain in the human body without which life is impossible. Diagnosis of liver disease requires a long time and sufficient expertise of the doctor. Statistical methods can be classified as an automated forecasting system and help specialists for quickly and accurately diagnose liver disease. Hidden Markov model is an intelligent and robust statistical method that has been used in present study.

**Methods:** The data used in this cross sectional study collected from records of patients with five different types of liver diseases, including cirrhosis, liver cancer, acute hepatitis, chronic hepatitis, and fatty liver disease. The patients have been admitted to Afzalipour hospital in Kerman, Iran, from 2006 to 2013. Hidden Markov model using EM algorithm for learning was fitted to the data and for evaluating the performance of the model, criteria as accuracy, sensitivity and specificity were used.

**Results:** The decision, sensitivity, and specificity criteria of the model for diagnosis of each liver disease were separately calculated and the highest level criteria in diagnosis of cirrhosis of the liver were 77% decision, 82% sensitivity, and 96% specificity, and also the lowest level of diagnosis for fatty liver disease was 65% decision, 69% sensitivity and 94% specificity.

**Conclusion:** The results of this study indicate the potential capabilities of the Hidden Markov Model. Therefore, using Hidden Markov Model for prediction of diagnosis of liver disease is recommended

**Keywords:** Hidden Markov Model, EM algorithm, Liver disease, Prediction