International Journal of Advanced Biotechnology and Research (IJBR)
ISSN 0976-2612, Online ISSN 2278–599X,
Vol-7, Special Issue-Number2-April, 2016, pp1207-1214
http://www.bipublication.com

Review Article

Liver Disease Recognition: A Discrete Hidden Markov Model Approach

Farzan Madadizadeh[1,2], Mohamad Ezati Asar[3] ,
Abbas Bahrampour[*4] and Mitra Montazeri[5]

[1]Noncommunicable Diseases Research Center,
Fasa University of Medical Sciences, Fasa, Iran
[2]Department of Epidemiology and Biostatistics,
School of Public Health, Tehran University of Medical Sciences,
Tehran, Iran.fmadadizadeh@razi.tums.ac.ir
[3] Modeling in Health Research Center , Institute for Futures Studies in Health,
Kerman University of Medical Sciences, Kerman, Iran. aliezati@gmail.com
[4]Department of Epidemiology and Biostatistics,
School of Health, Kerman University of Medical Sciences,
Kerman, Iran. abahrampour@yahoo.com
[5]. Medical Informatics Research Center, Institute for Futures Studies in Health,
Kerman University of Medical Sciences, Kerman, Iran. mmontazeri@yahoo.com
Corresponding author email: abahrampour@yahoo.com

ABSTRACT

The liver alongside the heart and the brain is the largest and the most vital organ within the human body whose absence leads to certain death. In addition, diagnosis of liver diseases takes a long time and requires sufficient expertise of physicians. To this end, statistical methods as automatic prediction systems can help specialists to diagnose liver diseases quickly and accurately. The Discrete Hidden Markov Model (DHMM) is an intelligent and a strong statistical model used to predict the types of liver diseases in patients in this study. The data in this cross-sectional study included information elicited from the records of 1143 patients with 5 different types of liver diseases including cirrhosis of the liver, liver cancer, acute hepatitis, chronic hepatitis, and fatty liver disease admitted to Afzalipour Hospital in the city of Kerman in the time period of 2006-2013. At first, the type of diseases for each patient was identified; however, it was assumed that the type of diseases is unknown and there were attempts to diagnose the type of the disease through the DHMM to examine its accuracy. Therefore, the DHMM was fitted to the data and its performance was evaluated by using the parameters of accuracy, sensitivity, and specificity. Such parameters of the model were separately calculated for the diagnosis of liver diseases. The highest levels of accuracy, sensitivity, and specificity were associated with the diagnosis of cirrhosis of the liver and equal to 0.77, 0.82, 0.96, respectively; and the lowest levels were related to the diagnosis of fatty liver disease with an accuracy level of 0.65 and a sensitivity level of 0.69. As well, the specificity level in the diagnosis of fatty liver disease was 0.94.

The results of this study indicated the potential ability of the DHMM; thus, the use of this model in terms of diagnosing liver diseases was strongly recommended.

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

INTRODUCTION

Weighing approximately 1500 grams, the liver is the largest internal organ and the most important one in the human body beside the heart and the brain and without such a vital organ, life is not viable (1).Liver diseases are placed among the top ten deadliest diseases in the world. They are taken in as reasons for 10% of all death rates in Iran and the fifth cause of death after heart problems, cancer, stroke, and respiratory diseases in Europe (2).Early diagnosis of liver diseases is essential to treat them. Although there have been significant developments in medical sciences, the early diagnosis of liver diseases is hard work and a specialist requires time and practice to gain complete experience in terms of diagnosing liver diseases. Therefore, the presence of a physician assistant system with a precise and quick predictive power to diagnose liver diseases is of utmost importance (3).Statistical methods can help specialists in the form of an intelligent system (4). Recently, the use of statistical techniques and modeling instruments is extensively growing in different fields and there is also an inclination to use such models (5). In this respect, the statistical method of the HMM is one of the concepts which has drawn the attention of numerous researchers. Since this model is based on learning and discovery of the latent relationships in the process of problem-solving, it is included among data-mining, machine learning, and intelligent models(6-9). The given model is used as an instrument in solving problems such as classification and prediction which cannot be fulfilled through traditional methods which are time-consuming for specialists. Such a model can also be employed for differentiation and classification of liver diseases which can lead to reduced laboratory costs, expedition of disease screening in some cases, and assist specialists.

The main purpose of this study was to diagnose liver diseases through the HMM and evaluate the predictive power of the model using the parameters of accuracy, sensitivity, and specificity. In this way, EM algorithm was used to increase the level of accuracy and provide a better performance for this model. According to a review in the present study, most of the investigations using the HMM have been conducted in non-medical fields (e.g. Speech recognition, machine learning, Bioinformatics, etc.) (10-15), and there are few studies on the HMM and its applications in the medical field. Moreover, it should be noted that, this study was among the first studies using the HMM to predict diseases in the medical field.

In a study conducted in 2014 by Vimala et al., the HMM was used to identify and classify Electrocardiography (ECG) signals. The data of this study had been obtained from hospitals under the affiliation of the Massachusetts Institute of Technology (MIT) in the U.S. The data included recorded information for 88 individuals referred to these centers. Among them, 18 people were in a normal state, 22 people suffered from stress, and 48 individuals were affected with heart diseases. The HMM revealed an accuracy level equal to 0.93 following the classification of the ECG data into three groups of normal state, stress, and heart diseases. These results confirmed that the HMM could be used as a powerful tool for grouping ECG signals into three given categories(16).

In another study by Li et al. in 2013, the HMM was used to predict the progression of lung cancer among 508 patients in one of the Chinese hospitals from 2010 to 2012. To develop the model, three states were considered for the disease, and the initial number between zero and one were randomly assigned to the parameters of the model. The progression of cancer was predicted for a new patient and the level of accuracy for the HMM was by 0.81. Finally, the HMM was proposed as a powerful prediction system to help specialists in the medical field (17).The continuous HMM was also employed to classify snoring sounds of 21 patients with sleeping disorders in an investigation by Lee et al. (2013). In this study, patients' snoring sounds were recorded through sensors hanging on their neck and then the data were classified and evaluated. Finally, the level of sensitivity of the model was equal to 0.93 (18).

Furthermore, Wu et al. (2010) conducted a study entitled "Detecting heart diseases through heartbeat sound recognition by using the HMM". The data in this study consisted of 325 recorded heartbeat sounds from ten patients afflicted with heart diseases. To develop the HMM, the number of different states of the disease was tested by considering two to seven states and finally four disease states were considered because the model had the highest level of accuracy equal to 0.95 (19).

In this respect, "Using the HMM to differentiate and detect heartbeat sounds recorded via stethoscope" was another study carried out by Schmidt et al. (2008). The data included 100 audio recordings of heartbeat sounds of patients with various heart diseases. To develop the HMM, 40% (40 individuals) of the data was used in the training phase and 60% (60 individual) of data in the evaluation phase of the model. The number of states of the model was equal to four. At the end, relatively good results such as a sensitivity level by 0.59 and the positive predictive value of 0.55 were obtained which showed that the model was not that undesirable (20).

MATERIALS AND METHODS

The data used in this study were part of records of patients with 5 different types of liver diseases including cirrhosis of the liver, liver cancer, acute hepatitis, chronic hepatitis, and fatty liver disease in the time period of 2006-2013 and admitted to Afzalipour Hospital in the city of Kerman. The total number of cases was 1300 of which about 1143 cases with complete information were collected. It should be also noted that the patient's type of disease was known at first (Gold Standard), but it was assumed to be unidentified to have a precise evaluation of the HMM and diagnose the type of disease through this model.

The statistical analysis

The statistical method used in this study was the HMM which is explained as follows.

Introduced in the late 1960s, the HMM is a statistical method based on Markov chain. This model is a powerful tool for random processing and modeling and it is known as a technique to predict and classify data among other statistical methods(21, 22). An HMM can be established by determining the following parameters:

- **The number of possible states** $\underset{\sim}{S} = \{s_1, ..., s_N\}$: it is the number of disease states. In this study, the liver disease had 5 different states (N=5) so $\underset{\sim}{S} = \{s_1, \ldots, s_5\}$. Fatty liver disease, chronic hepatitis, acute hepatitis, liver cancer, and cirrhosis of the liver were respectively associated with states 1 to 5 in this study.
- **The number of observations in each state** $\underset{\sim}{O} = \{x_1, ..., x_n\}$: there were several observations (attributes) for each of the above-mentioned diseases. In this study, the observations were the laboratory parameters recorded for each person affected with liver disease.
- **Transition probability** a_{ij} : the probability of passing the disease from state i to the state j

$$a_{ij} = p(s_j \mid s_i) \qquad 1 \le i, j \le N \tag{1}$$

$$a_{ij} \ge 0, \quad \sum_{j=1}^{N} a_{ij} = 1 \qquad 1 \le i, j \le N \tag{2}$$

- **Transition probability matrix** $A = [a_{ij}]$: a set of transition probabilities among states. Since the liver disease had five states in this study, the transition probability matrix was presented as follows:

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

- **A Probability distribution of observation in each of the states** b_{jk} : it shows the probability of belonging of each observation (attribute) to each state,

$$b_{jk} = p(O = x_k \mid S = s_j), \quad 1 \le k \le N, \quad 1 \le j \le N \tag{4}$$

$$b_{jk} \ge 0, \quad \sum_{k=1}^{n} b_{jk} = 1, \qquad 1 \le j \le N \tag{5}$$

- **The Probability distribution of observation matrix** $B=[b_{jk}]$: in this study there were 10 observations (laboratory parameters) and five types (five discrete states) of the liver diseases. Therefore, the probability distribution of observation matrix was as follows:

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

- **The initial state distribution vector** $\pi=\{\pi_0(1),.....,\pi_0(N)\}$: there is an initial probability for each state of disease by which the Markov chain begins to work. In this way, an HMM with a discrete probability distribution via specifying 3 parameters $\lambda=(A,B,\pi)$ was obtained.

The most important issue in the use of the HMM is training and evaluation phases of this model. The training phase of the HMM is feasible through Baum-Welch algorithm. In this state, not only the data from patients are given to the model, but also the three required parameters $\lambda=(A,B,\pi)$ for the model are assigned with initial random numbers between zero and one. When the Baum-Welch algorithm is trained, it provides an optimal value for the three parameters. The evaluation phase of the model is also feasible by using forward-backward algorithm. In this method, the model receives a number of observations (attributes) not included in the training phase and shows the belonging value of these attributes for each state of the model through a probability. If each state in the model is considered as a category, then by the Viterbi algorithm, maximum of such probabilities is taken into consideration and the category (state) with maximum probability is predicted as the target one. Then, the accuracy of the model is estimated via comparing the predicted category and the gold standard (main category of the observations). Usually, in the training phase of the model, 75% of the data are used and the remaining 25% of the data are employed to evaluate the model (7, 23). Rather than using Baum-Welch algorithm in this study, EM algorithm was used to train the HMM. The EM algorithm had a continuous action until the estimated value for the three parameters reached an optimal amount which was not much different from the previous iteration values. It should be noted that 75% of the data (for 857 patients) were used for the training phase of the HMM and the remaining 25% of the data (for 286 patients) were employed to evaluate the performance of the model.

Table 1. The clinical variables of liver disease that used in the HMM

ID	Variable name	Type	ID	Variable name	Type
1-	Alkaline Phosphatase	Integer	8-	White Blood Cell	Real Number
2-	Bilirubin, Total	Real Number	9-	Red Blood Cell	Real Number
3-	Bilirubin, Direct	Real Number	10-	Hemoglobin	Real Number
4-	Partial thromboplastic time	Real Number	11-	Hematocrit	Real Number
5-	Mean corpuscular volume	Real Number	12-	Platelet	Real Number
6-	Mean corpuscular hemoglobin	Real Number	13-	Aspartate Aminotransferase	Real Number
7-	Mean corpuscular hemoglobin concentration	Real Number	14-	Alanine Aminotransferase	Real Number

Given that the main purpose of this study was to evaluate the performance of the HMM in predicting and differentiating liver diseases, it was necessary to use statistically appropriate parameters in order to use the HMM with more confidence. To this end, the parameters of accuracy, sensitivity, and specificity of prediction were used. All the analyses in this study were performed via the MATLAB software version 2012.

RESULTS

In this study; the recorded data for 1143 patients with different types of liver diseases, including fatty liver disease (104), chronic hepatitis (235), acute hepatitis (39), liver cancer (152), and cirrhosis of the liver were evaluated. Patients' data included 14 laboratory variables that are described in Table 1. The number of optimal states used in the final model was 5 and 75% of the data and EM algorithms were used to train it. The condition to stop the algorithm was the difference equal to 0.000001 in new estimates of the parameters to the amounts of the previous ones. The algorithm with 100 times was stopped and the optimal values of the parameters of the model were generated. The optimal value obtained for the transition probability matrix between different states of liver disease were as follows (Matrix A) wherein fatty liver disease was associated with state 1, chronic hepatitis was related to state 2, and acute hepatitis was assigned as state 3. As well, liver disease and cirrhosis of the liver were marked with states 4 and 5. For example, the probability that a person with fatty liver disease was affected by liver cancer was approximately 43% ($a_{14} = 0.439$) and the probability that this illness was in a steady state was equal to 41% ($a_{11} = 0.417$).

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

The results of evaluating the HMM are illustrated in Table 2.

Table 2. The diagnostic performance evaluation of DHMM by different liver disease.

Performance Evaluation Measures	Liver Disease				
	Fatty liver	Chronic hepatitis	Acute hepatitis	Liver cancer	Cirrhosis
Sensitivity	0.6985	0.8000	0.7187	0.7100	0.8275
Specificity	0.9468	0.9662	0.9641	0.9653	0.9644
Accuracy	0.6515	0.7362	0.6793	0.7162	0.7782

DISCUSSION

The liver is the largest internal organ and the most vital one in the human body along with the heart and the brain and life is inevitable without such an organ (2). In terms of structure, the liver is an intricate organ with various performances; therefore, specialists are overwhelmed with numerous data to diagnose liver diseases, which is also a time-consuming procedure. As physician assistant systems, statistical methods, including the HMM can help specialists to save time in differentiating and classifying various liver diseases(4).

The structure of the HMM is more complex than other common statistical methods which have been used in different fields of computer science to detect and recognize speech but not in the medical field. To the researchers' knowledge, this study was the first one in which the HMM was used in the medical field particularly for liver disease diagnosis.

Using the HMM to detect ECG signals by Vimala et al. (16), it was stated that this model could be fitted to the data as a powerful prediction system. In this study, the HMM was significantly fitted to the data and revealed accurate results. In general, a desirable performance was shown for the HMM which was consistent with the findings of them.

In this study, five states were included in the model; but such states were considered three in a study by Hui et al.(24), in which the HMM was used to predict the progression of cancer. Moreover, the initial values of the model parameters were assigned to number between zero and one which was in line with the present study. Although the sample size used in their study was lower than that in the present study, acceptable results were obtained in terms of the evaluation of the HMM which confirmed the findings of this study. In the present study, the HMM was of discrete time type while in a study by Lee et al. (18), a continuous HMM was used to classify the snoring sounds leading to a high sensitivity for this model. This indicated the great potential of the HMM in predicting and classifying different types of data. The high sensitivity level obtained in the present study also confirmed this issue and it was in agreement with the findings of them.The highest sensitivity and specificity levels in this study were obtained for predicting cirrhosis of the liver. the high levels of sensitivity and specificity indicated the significant capability of the HMM and the results confirmed each other. The accuracy of the present study was consistent with that of Wu et al.(19), in which a high accuracy level was obtained for the HMM. Both studies selected the HMM states based on the disease states and this resulted in high levels of accuracy of the models. The lowest value obtained for the sensitivity of the model in this study was about 70% which was higher than that found in the investigation by Schmidt et al.(20) ,on the detection of heartbeat sounds. This was probably due to the larger sample size in the present study. This study was in the form of a report and an interpretation of transition probability matrix. Given that the model states were considered as 5 states of the disease, there were appropriate interpretations for the given numbers in the transition probability matrix. For example, $a_{23} = 0.980$ meant that the transition probability from state 2 (chronic hepatitis) to state 3 of the disease (acute hepatitis) among the patients in this study was equal to 98%. Furthermore, the probabilities lower than 0.1 in the Table represented a very weak transition probability such as the transition from liver disease to chronic and acute hepatitis or the probability of transition from cirrhosis of the liver to chronic and acute hepatitis.

To predict the type of diseases for each person, an evaluation was performed through considering the data entered into the model and the results revealed the great potential of the HMM in a way that this model revealed different results for predicting each type of liver disease. The model had the highest level of accuracy in diagnosing cirrhosis of the liver and the lowest accuracy in the diagnosis of fatty liver disease (65 %).

One of the main limitations of this study was the lack of accurate recording of data related to patients with liver diseases which in turn had a significant impact on the accuracy of the final model.

CONCLUSION

In conclusion, it was argued that the diagnosis of diseases through intelligent systems was characterized by quicker, more accurate, and low-cost performance which was of utmost importance in healthcare issues. Therefore, the use of the HMM in the field of diagnosis of liver diseases was recommended.

REFERENCE:

1. Ramana BV, Babu MSP, Venkateswarlu N. (2011), A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis. International Journal of Database Management Systems. 3(2):101-14.
2. Ganji A, Safavi M, Nouraie S, Nasseri-Moghadam S, Merat S, Vahedi H, et al.(2006), Digestive and liver diseases statistics in several referral centers in Tehran, Govaresh. 11(1):33-8.
3. Lin R-H. (2009), An intelligent model for liver disease diagnosis. Artificial Intelligence in Medicine. 47(1):53-62.

4. Madadizadeh F, Bahrampour A, Mousavi SM, Montazeri M.(2015), Using Advanced Statistical Models to Predict the Non-CommunicaČble Diseases. Iranian journal of public health. 44(12):1714-5.
5. Wolfe R, Abramson MJ. Modern statistical methods in respiratory medicine. Respirology. 2014;19(1):9-13.
6. Beal MJ, Ghahramani Z, Rasmussen CE.(2001), The infinite hidden Markov model. Advances in neural information processing systems;.
7. Madadizadeh F, Montazeri M, Bahrampour A. (2015), Predicting the survival in breast cancer using Hidden Markov Model. 10th International Breast Cancer Congress; Shahid Beheshti University of Medical Sciences, Tehran, Iran.
8. Pimentel MA, Santos MD, Springer DB, Clifford GD.(2015), Heart beat detection in multimodal physiological data using a hidden semi-Markov model and signal quality indices. Physiological measurement. 36(8):1717.
9. Zhou C, Huang S, Xiong N, Yang S-H, Li H, Qin Y, et al.(2015), Design and Analysis of Multimodel-Based Anomaly Intrusion Detection Systems in Industrial Process Automation. Systems, Man, and Cybernetics: Systems, IEEE Transactions on. 45(10):1345-60.
10. Bilal S, Akmeliawati R, Shafie AA, Salami MJE.(2013), Hidden Markov model for human to computer interaction: a study on human hand gesture recognition. Artificial Intelligence Review. 40(4):495-516.
11. Cholewa M, Głomb P.(2013), Estimation of the number of states for gesture recognition with Hidden Markov Models based on the number of critical points in time sequence. Pattern Recognition Letters. 34(5):574-9.
12. Farsi H, Saleh R. (2014), Implementation and optimization of a speech recognition system based on hidden Markov model using genetic algorithm. Intelligent Systems (ICIS), 2014 Iranian Conference on; IEEE.
13. Geiger JT, Kneißl M, Schuller BW, Rigoll G, editors.(2014), Acoustic gait-based person identification using hidden Markov models. Proceedings of the 2014 Workshop on Mapping Personality Traits Challenge and Workshop; ACM.
14. Hu M, Shen F, Zhao J.(2014), Hidden Markov models based dynamic hand gesture recognition with incremental learning method. Neural Networks (IJCNN), 2014 International Joint Conference on; IEEE.
15. Jin J, Wang Y, Xu L, Cao L, Han L, Zhou B, et al.(2014), Non-intrusive gesture recognition system combining with face detection based on Hidden Markov Model. SPIE/COS Photonics Asia; International Society for Optics and Photonics.
16. K.Vimala.(2014), Stress causing Arrhythmia Detection from ECG Signal using HMM. International Journal of Innovative Research in Computer and Communication Engineering. 2(10):6079-85.
17. Li H-M FL-Y, Wang P, Yan J-Z.(2013), Hidden Markvo Models Based Research on Lung Cancer Progress Modeling. Research Journal of Applied Sciences, Engineering and Technology. 6(13):2470-3.
18. Lee H-K, Lee J, Kim H, Ha J-Y, Lee K-J.(2013), Snoring detection using a piezo snoring sensor based on hidden Markov models. Physiological measurement. 34(5): 41-45.
19. Wu H, Kim S, Bae K.(2010), Hidden Markov Model with heart sound signals for identification of heart diseases. Proceedings of 20th International Congress on Acoustics (ICA), 23-27 August 2010, Sydney, Australia.
20. Schmidt S, Toft E, Holst-Hansen C, Graff C, Struijk J. (2008), Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov model. Computers in Cardiology, IEEE.
21. Rabiner LR, Juang B-H.(1986), An introduction to hidden Markov models. ASSP Magazine, IEEE. 3(1):4-16.
22. Pezeshk H, Naghizadeh S, Malekpour SA, Eslahchi C, Sadeghi M.(2010), A modified bidirectional hidden Markov model and its application in protein secondary structure prediction. Advanced Computer Control

(ICACC), 2010 2nd International Conference on: IEEE.

23. Rastjoo A, Arabalibeik H.(2008), Evaluation of Hidden Markov Model for p300 detection in EEG signal. Studies in health technology and informatics. 142:265-7.
24. Hui-Min Li L-YF, Pu Wang and Jian-Zhuo Yan.(2013), Hidden Markvo Models Based Research on Lung Cancer Progress Modeling. Research Journal of Applied Sciences, Engineering and Technology. 6(13):2470 - 3.