

Association for Information Systems

AIS Electronic Library (AISeL)

Proceedings of the 2020 Pre-ICIS SIGDSA
Symposium

Special Interest Group on Decision Support and
Analytics (SIGDSA)

Winter 12-12-2020

Analyzing Non-Alcoholic Fatty Liver Disease Risk Using Time-Series Model

Ofir Ben-Assuli

Shani Shenhar-Tsarfaty

Orit Goldman

Arie Jacobi

Ori Rogowski

See next page for additional authors

Follow this and additional works at: <https://aisel.aisnet.org/sigdsa2020>

This material is brought to you by the Special Interest Group on Decision Support and Analytics (SIGDSA) at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the 2020 Pre-ICIS SIGDSA Symposium by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Authors

Ofir Ben-Assuli, Shani Shenhar-Tsarfaty, Orit Goldman, Arie Jacobi, Ori Rogowski, David Zeltser, Itzhak Shapira, Shlomo Berliner, and Shira Zelber-Sagi

Analyzing Non-Alcoholic Fatty Liver Disease Risk Using Time-Series Model

Extended Abstract: Research-in-Progress

Ofir Ben-Assuli

Ono Academic College, Israel
ofir.benassuli@gmail.com

Orit Goldman

Ono Academic College, Israel
oritgol@bezeqint.net

Ori Rogowski

Tel-Aviv University, Israel
orir@tlvmc.gov.il

Itzhak Shapira

Tel-Aviv University, Israel
shapira@tlvmc.gov.il

Shani Shenhar-Tsarfaty

Tel-Aviv University, Israel
shani.shenhar@mail.huji.ac.il

Arie Jacobi

Ono Academic College, Israel
jacobi.arie@gmail.com

David Zeltser

Tel-Aviv University, Israel
davidz@tlvmc.gov.il

Shlomo Berliner

Tel-Aviv University, Israel
berliners@tlvmc.gov.il

Shira Zelber-Sagi

University of Haifa, Israel
zelbersagi@bezeqint.net

Abstract

Non-alcoholic fatty-liver disease (NAFLD) is the most global frequent liver disease, with a prevalence of almost 20% in the overall population. NAFLD may progress to fibrosis and later into cirrhosis in addition to other diseases. Our objective is to stratify patients' risks for NAFLD and advanced fibrosis over time and suggest preventive medical decisions. We used a cohort of individuals from the Tel-Aviv medical center. Time-series clustering machine learning model (Hidden Markov Models (HMM)) was used to profile fibrosis risk by modeling patients' latent medical status and trajectories over time.

The best-fitting model had three latent HMM states. Initial results show that tracking individuals over time and their relative risk for fibrosis in each point of time provides significant clinical insights regarding each state (and its group of individuals). Thus, longitudinal risk stratification can enable early identification of specific individual groups following distinct medical trajectories based on their routine visits.

Keywords

Hidden Markov Models, Time-Series Models Machine Learning, Predictive Analytics.

Introduction, Background and Objective

Non-alcoholic fatty liver disease (NAFLD) is the most global frequent liver disease (Loomba and Sanyal 2013) with a prevalence of almost 20% in the overall population (Vandromme et al. 2020). The NAFLD diagnosis and discovery of its risk factors for steatosis are vital since beyond the hepatic destruction, NAFLD carries an additional independent risk for cardiovascular and diabetes diseases (Stefan et al. 2008). Complications from liver damage (scar tissue) can disturb liver function. When it comes to Hepatitis C, in the initial stage of the virus, it can trigger little to mild liver injury but over time scarring which is hardening, or stiffening tissue progresses in the liver. This phase of liver damage is called fibrosis. In many cases, there are little to no symptoms with fibrosis until liver damage becomes severe enough (cirrhosis) (Ginès et al. 2016). Thus, there is a great potential to learn the individual characteristics over time.

The NAFLD fibrosis 4 calculator (FIB-4) has been externally validated in ethnically different NAFLD populations, with reliable and stable results (European Association for the Study of the et al. 2016). We classified the FIB-4 scores into two accepted cut-off fibrosis levels (low and intermediate-high). Our study applied novel machine learning approach for risk stratification by tracking high risk groups for fibrosis disease incidence and progression over time (in contrast to previous works). The main goal of this project is to enable physicians and clinicians to stratify patients' risks for fibrosis and disease progression and to suggest possible medical junctions optimal for preventive medical decisions. Previous works have mostly been based on narrow data, rather than on individual-level big data derived from rich datasets as we are doing here. We detect and model individuals' latent state representing the risk for fibrosis over time. Hidden Markov Model (HMM) has been chosen here in order to detecting the changes in the risk levels across time, by representing real time latent states.

Data Description and Hidden Markov Model (HMM)

We used a cohort of apparently healthy volunteers from the Tel-Aviv medical center inflammation survey. We modeled the risk of fibrosis using FIB-4 index values over time among individuals in Israel from 2007 to 2017. The individual-level data included demographic, lifestyle properties, diagnoses and laboratory values. Time-series clustering machine learning model (e.g. HMM) are robust approaches to profiling fibrosis risk by modeling patients' latent medical status and trajectories over time. In our HMM (performed using the DepmixS4(Visser 2011)), the observation variable FIB-4 is the sole known variable.

Results

At first, we evaluated three, four and five HMM states and measured their goodness of fit using BIC, AIC and Log Likelihood, identifying three latent HMM states as optimal. When observing the distribution of individuals and the FIB-4 prevalence across the visits, the number of individuals as High-risk state decreases from visit to visit (visits: two (75%) three (22%), four (15%), five (14%) and six (12%)). This is showing that the HMM becoming more discriminating method (discovering fewer individuals prone to fibrosis) as the number of visit are growing and more information are loaded. We share some findings by profiling of individuals in the fifth visit (as an example), according to their risk level for fibrosis. Individuals in the High-risk group have more extreme laboratory test values, demographics and diagnoses¹. Among the three risk states, the Age, Gender, Hypertension and Diabetes were different expressing the worse condition for the high-risk group followed by the medium risk group. BMI, waist and Metabolic Syndrome were higher and worse for the high-risk group not significantly probably due to low sample size. Among the selected blood tests that we show, AST, Total Cholesterol, Glucose, Platelets, Albumin and Bilirubin characteristics were significantly different. Individuals in the High-risk state have generally worse values in most indicators relative to the other two other risk states (14 indicators compare to 9). The picture, however, is less clear between the Low- and Medium-risk states, and further analyses are needed and planned.

Discussion

We used the non-invasive calculator FIB-4 that has been externally validated to learn about fibrosis development. It was used instead of the invasive liver biopsy. Beyond the biopsies' needed medical resources and associated medical costs, many of the individuals that have fibrosis are simply not undertaking such biopsies and remain undiagnosed and unknown. Even prior studies that have performed predictive analytics used classification models (e.g. logistic regressions and decision trees) on a single/first visit level (Eslam et al. 2016; Hashem et al. 2012; Hosseini et al. 2011).

The HMM method highlights that individual risk change over time and each group may change its' risk stratification differently (in the next phase we plan to add more methods as and Covariate Latent Class Models to support the results with robustness checks). Furthermore, our approach classified a small subgroup of patients as High-risk during each of their visits. It can help the health care organization to

¹ Selected values from the descriptive statistics for the high-risk group (MEAN ± STDEV): 84% male, age: 55.5±6.6 Waist: 96.5±10.3, Weight: 79.6±13.7, BMI: 26.1±4.3, 57.9% Hypertension, 16.7% Diabetes, 14.9% Metabolic Syndrome, AST: 29±10.5, ALT: 27.6±12.6, GGT: 23.6±12.7, Cholesterol: 171.8±29.8, HDL Cholesterol: 54.4±15.1, Triglycerides: 105±60.9, Glucose: 92.2±16.5 and hemA1C: 5.7±0.5.

direct targeted resources to diminish the risk of fibrosis, especially, case managers involvements, or enrollment in event warning systems (Gutteridge et al. 2014).

Unlike previous works, we took all points in time and execute longitudinal analysis of the data. Our initial results signified the importance of time. Fibrosis risk increased from visit to visit for the Medium- and High-risk states, but it stays quite steady for the Low-risk state. Additionally, we benefited more data components than what have been used in the previous fibrosis studies (Lackner et al. 2005; Lee et al. 2019; Zhou et al. 2010) and merged laboratory with demographic and other medical data, with all variables being temporal covariates. This kind of comprehensive clinical data is, making our study as exclusive, emphasizing the value of laboratory and diagnostic variables for Machine Learning research. From a feasible viewpoint, we demonstrated that leveraging patients' longitudinal data can advance to conditions that enhance risk stratification. HMM stratification may be beneficial attitude to identifying individuals' latent risk for Fibrosis. Consequently, it may be executed for population specific medical preventable care that can decrease the risk for NAFLD fibrosis. We presume that the method applied in this study can be practical in one or more of the following 1) handling individuals, 2) broad, individual-level, longitudinal data is accessible, and 3) examining fibrosis of the same individuals.

Acknowledgement - This work was funded by The Israel National Institute for Health Policy Research, Grant Number: 2018/52/s.

References

- Eslam, M., Hashem, A. M., Romero-Gomez, M., Berg, T., Dore, G. J., Mangia, A., Chan, H. L. Y., Irving, W. L., Sheridan, D., and Abate, M. L. 2016. "Fibrogene: A Gene-Based Model for Staging Liver Fibrosis," *Journal of hepatology* (64:2), pp. 390-398.
- European Association for the Study of the L., European Association for the Study of, D., and European Association for the Study of, O. 2016. "Easl-Easd-Easo Clinical Practice Guidelines for the Management of Non-Alcoholic Fatty Liver Disease," *J Hepatol* (64:6), pp. 1388-1402.
- Ginès, P., Graupera, I., Lammert, F., Angeli, P., Caballeria, L., Krag, A., Guha, I. N., Murad, S. D., and Castera, L. 2016. "Screening for Liver Fibrosis in the General Population: A Call for Action," *The lancet Gastroenterology & hepatology* (1:3), pp. 256-260.
- Gutteridge, D. L., Genes, N., Hwang, U., Kaplan, B., Shapiro, J. S., and Investigators, G. W. 2014. "Enhancing a Geriatric Emergency Department Care Coordination Intervention Using Automated Health Information Exchange-Based Clinical Event Notifications," *EGEMS* (2:3).
- Hashem, A. M., Rasmy, M. E. M., Wahba, K. M., and Shaker, O. G. 2012. "Single Stage and Multistage Classification Models for the Prediction of Liver Fibrosis Degree in Patients with Chronic Hepatitis C Infection," *Computer methods and programs in biomedicine* (105:3), pp. 194-209.
- Hosseini, S.-M., Mousavi, S., Poursafa, P., and Kelishadi, R. 2011. "Risk Score Model for Predicting Sonographic Non-Alcoholic Fatty Liver Disease in Children and Adolescents," *Iranian journal of pediatrics* (21:2), p. 181.
- Lackner, C., Struber, G., Liegl, B., Leibl, S., Ofner, P., Bankuti, C., Bauer, B., and Stauber, R. E. 2005. "Comparison and Validation of Simple Noninvasive Tests for Prediction of Fibrosis in Chronic Hepatitis C," *Hepatology* (41:6), pp. 1376-1382.
- Lee, J. I., Lee, H. W., and Lee, K. S. 2019. "Value of Controlled Attenuation Parameter in Fibrosis Prediction in Nonalcoholic Steatohepatitis," *World journal of gastroenterology* (25:33), p. 4959.
- Loomba, R., and Sanyal, A. J. 2013. "The Global Nafld Epidemic," *Nat Rev Gastroenterol Hepatol* (10:11), pp. 686-690.
- Stefan, N., Kantartzis, K., and Haring, H. U. 2008. "Causes and Metabolic Consequences of Fatty Liver," *Endocr Rev* (29:7), pp. 939-960.
- Vandromme, M., Jun, T., Perumalswami, P., Dudley, J. T., Branch, A., and Li, L. 2020. "Automated Phenotyping of Patients with Non-Alcoholic Fatty Liver Disease Reveals Clinically Relevant Disease Subtypes," *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*: World Scientific, pp. 91-102.
- Visser, I. 2011. "Seven Things to Remember About Hidden Markov Models: A Tutorial on Markovian Models for Time Series," *Journal of Mathematical Psychology* (55:6), pp. 403-415.
- Zhou, K., Gao, C. F., Zhao, Y. P., Liu, H. L., Zheng, R. D., Xian, J. C., Xu, H. T., Mao, Y. M., Zeng, M. D., and Lu, L. G. 2010. "Simpler Score of Routine Laboratory Tests Predicts Liver Fibrosis in Patients with Chronic Hepatitis B," *Journal of gastroenterology and hepatology* (25:9), pp. 1569-1577.