

## Risks and Predictors of Non-Alcoholic Liver Disease Progression Using Association Rules Mining

<https://doi.org/10.3991/ijoe.v16i06.13629>

Tasneem A. Gameel (✉), Sherine Rady, Sanaa Kamal  
Ain Shams University, Cairo, Egypt  
tasneem.gameel@cis.asu.edu.eg

**Abstract**—Non-alcoholic Steatohepatitis Disease (NASH), a progression of Non-alcoholic Fatty Liver Disease (NAFLD), occurs in case of the increase of fat accumulation in the liver. The disease can next progress to fibrosis, cirrhosis or liver cancer. The most accurate way to diagnose NAFLD progression into NASH is through a liver biopsy. This is painful, expensive and difficult to repeat several times to monitor the fibrosis progression. Thus, finding a non-invasive solution through markers can reliably help tracking the disease progression. The objective of this study is to assess the diagnostic and prognostic performance of serum markers to monitor liver disease progression in comparison to findings by liver biopsy. An association rule mining system is proposed using a Frequent Pattern mining algorithm to reach this objective. An Egyptian cohort consisting of 2300 NAFLD and NASH patients is included in an experimental study, where the results showed that the blood tests and serum markers, PIIINP and ELF, can predict the progression of NAFLD into NASH, and can discriminate between the different stages of NASH with confidence value 0.9. The presented results indicate an advantageous promising non-invasive solution in medicine for predicting of the disease and its progression, while avoiding alternative biopsy exposition.

**Keywords**—NAFLD, NASH, Liver Disease, Frequent Pattern Mining, FP Growth Algorithm, Serum Fibrosis Markers

### 1 Introduction

Non-alcoholic fatty liver disease (NAFLD) is a condition characterized by increased deposition of fat in the liver within hepatocytes, which is known as hepatic steatosis. NAFLD happens without having a cause for the fat aggregation such as being alcoholic patient or HCV carrier [1]. It includes hepatic steatosis without inflammation. However, in some patients, NAFLD may progress to non-alcoholic steatohepatitis or NASH, which is associated with fat accumulation, hepatic inflammation, and hepatocyte injury with or without fibrosis or cirrhosis or liver cancer [2, 3]. As in [4], as many as fifteen percent of patients with NASH will progress to cirrhosis cases. NAFLD/NASH is a disease that has a direct complications and relationships to other diseases such as cardiovascular disease, polycystic ovary syndrome, chronic kidney disease and different types of cancers like liver, breast and colon [5].

Currently, NAFLD/NASH is the most common form of liver disease in developed countries and its prevalence is rapidly increasing in those countries. The current worldwide prevalence of NAFLD ranges from 25% to 35% [6]. In the Middle East, certain areas have higher prevalence rates. In United Arab Emirates, Saudi Arabia, and Egypt, the prevalence of NAFLD is about 25% [7,8,9].

The precise diagnosis of NAFLD/NASH is normally performed through a liver biopsy and histopathology to demonstrate the evidence of hepatic steatosis, liver inflammation, and hepatocyte injury. Liver biopsy is an invasive procedure with several adverse events. Therefore, several non-invasive imaging techniques, such as Fibroscan, magnetic resonance imaging, and blood tests, have been developed to monitor NAFLD/NASH and their progression.

The number of patients showing NASH end-stage progression is reported to be rising quickly [10, 11]. Additionally, NASH is becoming the most common reason behind the liver transplantation in a very rapid way [12, 13, 14]. Up to date, the risk factors for the progression of NAFLD into NASH are not clarified [15].

Therefore, this paper aims to present a study to identify the risk factors and predictors of NAFLD progression to NASH and risk factors for each NASH progression stage. The study is based on frequent pattern mining for generating a set of association rules between the data attributes and the disease. The study uses the approach presented in [16], where association rules have successfully managed to predict the liver cancer development using novel single nucleotide polymorphisms in genes that found associations with the liver cancer different stages. This paper extends that work and applies it to the detection of another liver disease, NASH progression, that is to date is considered as the most common liver disease with no clear risk factors.

## **2 Related Work**

NAFLD disease is a common disease with no or few early symptoms that make it hard to diagnose such liver disorder early. To assess the prevalence and incidence of NAFLD, many researches were concerned by analyzing the literature of this disease, such as in [17]. The aim of other researches was to gain more insights on the progression of NAFLD worldwide like in [18].

The most accurate way for assessment of NAFLD is liver biopsies, which is expensive, painful and need to be repeated for assessment of the progression [19]. Because of the previous reasons, it is advantageous to take benefits of the high advances of Information technology in the medical field [20]. Therefore, finding a non-invasive procedure and easily repeatable way of diagnosis was the aim of some researches. In [21], [22], ultrasound elastography techniques were applied on liver tissue to detect fibrosis, but as mentioned in the same reference, those techniques had limitations on the assessment of liver fibrosis.

The limitations of steatosis detection imaging systems were the reason behind the need of identification of factors that would promote increased liver fat, such as body weight, gender, diabetes and abnormal blood tests.

In 2019, a group of researchers in Italy tried to find metabolic predictors of liver fats, which concluded that increased body weight might increase the risk of liver fat accumulation but is not sufficient [23]. Also [24] confirmed that NAFLD occur in lean patients. In 2017, other group of researchers associated the NAFLD with men gender and stated that men are most likely to develop NAFLD than women [25].

Identifying factors associated with the progression of NAFLD to NASH and from NASH to cirrhosis or liver cancer is important as it may provide early prophylactic interference in high risk patients. Computer Aided Systems that help the physician in the task of diagnosis and prediction are becoming highly needed [26].

So here comes the task of knowledge discovery which aims to discover previously unknown understandings to get the best out of the data [27]. Nowadays, many researchers use data mining and knowledge discovery algorithms to solve medical problems. For example, [28], [29] used apriori, naïve Bayes and support vector machine algorithms to predict liver diseases development and progression. Association rules mining, the task of finding relationships between attributes in the data, is specially used in bioinformatics and biomedical field [30]. For example. a rule-based classifier is developed in [31] with accuracy 99.5% to predict the development of hepatocellular carcinoma in hepatitis C virus Egyptian patients.

In this study, an Egyptian cohort of NAFLD patients were prospectively enrolled to identify their different progression rates and define the attributes and risk factors that are associated with the disease progression to assist physicians in the decision-making process to design the optimal timing and type of treatment of NAFLD/NASH.

### **3 Material and Methods**

#### **3.1 Study population**

The study population included are patients visiting Ain Shams University clinical Hospitals between the period 2013 and 2018. The data included in the analysis are aged above 18 years at diagnosis and had medical records available for more than 12 months from registration with the clinic. Exclusion criteria for the data are for those with missing information on age and gender, a record of alcohol abuse at any time prior to diagnosis and a history of liver morbidity within the 12 months period prior to the diagnosis. Some other cases have been also excluded, such as patients with viral hepatitis, excessive alcohol consumption, aspartate aminotransferase (AST) or alanine aminotransferase (ALT) more than 500 and female patients who are pregnant.

After passing the filtering criteria, a training cohort of 2300 Egyptian cases with histologically / Fibroscan confirmed NAFLD patients obtained. Fourteen attributes, displayed in table 1, assessed and measured for the NAFLD patients.

**Table 1.** Used data attributes in the study

Demographic Attributes	Age Gender Body Mass Index (BMI)
Diabetic or not	
Histological factor	Histologic steatosis grade Lobular inflammation Ballooning NAFLD activity score Fibrosis score
Blood Tests	Cholesterol Triglycerides Alanine Transaminase PIIINP ELF

Patients who followed up from the enrollment date until the earliest of occurrence of cirrhosis, or NASH, have been recorded with the NASH disease and their progression until the end of the study period and loss of follow-up. Those progressive cases of interest, who were confirmed as diagnosed or identified cases with cirrhosis, hepatocellular carcinoma or NASH, are important cases for the study.

### 3.2 Methodology

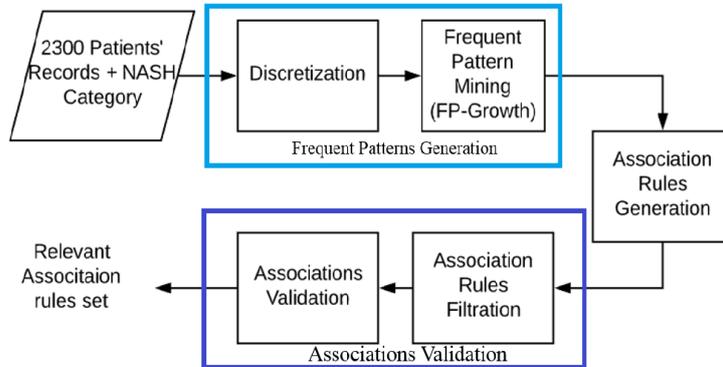
The major work of this paper is to identify NAFLD and NASH patients and to identify the significant features of each NASH stage using association rule mining. To achieve those objectives, a system is designed, whose overview is displayed in Figure 1. This system has been applied two times; First to identify the associations for the NAFLD or NASH case, and second to identify the associations for the progression level in case of NASH case

The proposed system consists of 3 main stages as follows: 1) Frequent patterns generation 2) Associations generation 3) Associations Validation.

The primary stage of the system is the frequent patterns generation. It is mainly responsible for extracting the frequent attributes and includes two steps: 1) Discretization and 2) Frequent attributes generation.

The subsequent stage is the association rule generation, which uses the frequent attributes generated from the previous stage to identify relationship between them and NAFLD/NASH progression in the form of IF/Then principles. The last stage approves the generated rules and consists of two steps: a) Filtration and b) Validation

where the significant associations are filtered depending on a minimum threshold and eventually approved by the domain expert.



**Fig. 1.** System Overview for association rule mining

**Frequent patterns generation**

*Discretization step:* For getting more accurate results, data pre-processing is normally required. Data discretization is a mandatory step before applying the data mining algorithm. In this step, all continuous attributes, Age, Body Mass Index, Cholesterol, Triglycerides, alanine transaminase and histologic steatosis grade, are discretized according to threshold or cut-offs given by the domain expert. Those cut offs shown in table 2.

**Table 2.** Discretization of Continuous Attributes

Attribute Name	Attribute Ranges
Age	$\leq 30$ , $>30 \ \& \ \leq 35$ , $>35 \ \& \ \leq 40$ , $>40 \ \& \ \leq 45$ , $>45 \ \& \ \leq 50$ , $>50 \ \& \ \leq 55$ , $>55 \ \& \ \leq 60$ , $>60$
Body Mass Index (BMI)	$<18$ , $\geq 18.5 \ \& \ <25$ , $\geq 25 \ \& \ <30$ , $\geq 30 \ \& \ <35$ , $\geq 35$
Cholesterol	$\leq 200$ , $>200$
Triglycerides	$<120$ , $\geq 120 \ \& \ <160$ , $\geq 160 \ \& \ <200$ , $\geq 200$
Alanine Transaminase	$\leq 40$ , $>40 \ \& \ \leq 60$ , $>60 \ \& \ \leq 80$ , $>80$
Histologic steatosis grade	$\geq 5 \ \& \ <30$ , $\geq 30 \ \& \ <50$ , $\geq 50 \ \& \ <75$ , $\geq 75$
PIIINP	$\geq 19 \ \& \ <100$ , $\geq 100 \ \& \ <200$ , $\geq 200 \ \& \ <300$ , $\geq 300 \ \& \ <400$ , $\geq 400$

*Extracting frequent patterns:*

For discovering interesting patterns in the data that might be associated with the disease discovery, the Frequent Pattern Algorithm (FP-Growth) is applied on the data as the training set. FP-growth mines to find the mostly appeared set of attributes and their values by first calculating the occurrences of each value and extracting those that exceed a minimum support count. Support count means how many times this attribute appeared through the dataset.

Second, the rearranged records depending on support of their attributes are mapped into a tree. Starting from the null node, a path is followed that contains attributes values

and their frequencies, while incrementing each time this path is visited. Next, a conditional FP Tree is constructed consisting of the common paths for each attribute's value, and which are considered as the frequent attributes. For our case, the resultant frequent patterns represent the frequent features or attributes appearing or associated with the NAFLD/NASH.

For the identification of the NASH progression, the same algorithm is executed but on the NASH training cases, and the resultant patterns mean frequent attributes associated with the NASH stage. For the NASH progression, 3 stages are identified, namely mild, moderate and severe.

**Finding associations:** This phase aims to find the associations between the data attributes and NASH category that are used to predict the occurrence of a NAFLD or NASH disease based on the occurrence of extracted attributes in the data.

*Association rules generation:* The frequent patterns extracted from the previous stage are analyzed to find their associations with the other patterns and find if these associations exceed a minimum threshold, that is to say a minimum confidence set value. This value is normally set to 0.5. Confidence indicates how many times those patterns appeared with given disease stage [32]. The output of this phase is encoded in the format of IF/Then rules for ease of inference, where antecedents are frequent attribute and the consequence is NAFLD or NASH. In the NASH progression, the antecedents are also frequent attributes, but the consequences are NASH stages.

#### **Associations validation**

*Rules filtration:* Since the strong rule implies more accuracy for the disease progression prediction, in the rule filtering phase, the generated rules are assessed, and only the rule set with the highest confidence values is chosen.

*Rules validation:* In the final stage of rule validation, the filtered rule set is validated against the literature results to gain more confidence in the obtained results. In addition, rules are presented to the domain expert to get the logic and approval of the association of an attribute with the disease stages.

## **4 Experimental Study, Results and Discussion**

### **4.1 Dataset**

A dataset of 2300 records of Egyptian NAFLD patients is used for the experimental study. In this dataset, there exist 1907 patients diagnosed with NASH with their NASH progression indicator (mild, moderate and severe). The NASH patients have the following distribution 1016, 586, 206 for with the progression levels mild, moderate and severe respectively. The disease findings have been done through a liver biopsy.

This section provides two experiments. In 4.1.1 an experiment is given to find features associated with the NASH morbidity and in 4.1.2 an experiment is given to find the features associated with different NASH stages.

**Experimenting factors associated with NASH morbidity:** The association rule mining described in section 3.2 is applied on dataset to extract attributes discriminating NASH disease. A minimum support value of 0.5 is used for the FP growth algorithm

and a minimum confidence threshold of 0.5 for the association rules generation. This experiment generated 90 rules at the finding associations step. Those rules are filtered on confidence of 0.7, and a set of 34 rules are obtained.

Significant findings included: variables including histologic steatosis grade, ALT, Triglycerides and lobular inflammation differentiate between the two different NAFLD stages with confidence value of one. Because we are concerned with the blood tests measurements, findings showed direct association between the used blood tests, PIIINP and ELF, and NASH development with confidence 0.87, which can provide us with non-invasive clues. Table 3 summarizes the resultant relevant rules in the form of <attribute, value> that lead to NASH, sorted by highest confidence.

**Table 3.** The resultant associations with NASH development

Attribute Name	Attribute Range	Confidence
Histologic steatosis grade	>=30	1
Alanine transaminase	>=60 and < 80	1
Triglycerides	>=200	1
Lobular inflammation	>5	1
PIIINP	>=200	0.87
ELF	>= 8	0.87

Additional relevant findings also include: The BMI exceeding or equal to 18.5 and less than 25 is associated with NASH development with confidence 0.806; Female gender is found to be more likely to develop NASH disease with approximately the same confidence as BMI; Being a Diabetic NAFLD patient increases the risk of NASH development with confidence 0.78.

**Experimenting factors associated with NASH progression:** 1907 NASH patients with their different progression levels (mild, moderate and severe) are examined in this study to identify the attributes that describe each stage. Results made it clear that histologic steatosis grade is directly associated with the mild stage. ALT and cholesterol can differentiate patients in the controllable moderate stage. PIIINP and ELF directly relates to the severe disease progression level. Table 4 shows the most significant rules obtained for discrimination between the three NASH stages, and sorted by the highest confidence.

Additional relevant findings have also included: Male gender is more likely to be in the mild stage with confidence 0.63; Having no diabetes, patients can be in the controllable stage with confidence 0.55.

**Table 4.** Rules discrimination NASH stages

Attribute Name	Attribute Range	Confidence	NASH stage
Histologic steatosis grade	>= 5 &<30	0.97	Mild
ALT and cholesterol	>60 & <=80, <=200	0.75	Moderate
PIIINP and ELF	>300, >8	0.81	Severe

## 4.2 Discussion

In this study, the risk factors of developing non-alcoholic steatohepatitis into non-alcoholic fatty liver disease are identified using association rule approach applied on 2300 NAFLD and NASH Egyptian patients. Major findings in the study have been identified which can predict the progression of NAFLD into NASH with 3 progression levels. These major findings have been validated through the medical expert showing reinforcement towards them. These findings are as follows:

1. PIIIINP and ELF blood tests are the most relevant dependable markers for the prediction of NAFLD into NASH progression;
2. Though was anticipated to be a major cause, but NASH disease is not directly related with obesity as it occurs to patients with BMI between 18.5 and 20;
3. Female gender is more likely to develop NASH disease and reaching its advanced stages; and finally
4. Being a diabetic patient increases the risk of development of NASH disease.

The presented experimental study case and the results indicate that such mined knowledge set can act as a non-invasive solution and provides an advantageous and promising solution for the predicting of NAFLD/NASH disease and its progression, while avoiding the alternative biopsy exposition.

## 5 Conclusion and Future Work

NASH progression prediction through non-invasive measurements has benefit on disease control, which increases the patients' health condition or survival. This study assessed the ability to predict NASH and its development through blood tests, which are comparatively non-invasive easy markers to measure and considered cheaper than biopsies. A solution approach using association rule mining and frequent patterns is proposed, which was able to identify relevant associations between the data attributes and the disease and its progression stages. The results proved that blood test measurements PIIIINP and ELF can serve as markers for NASH prediction and can differentiate between its progression stages with approximate confidence 0.9.

In future prospects, more blood tests can be included in the study that may be more reliable and may reach the accuracy of liver biopsies. The used methodology can be investigated with other liver disease to predict its progression levels.

## 6 References

- [1] Friedman, S. L., Neuschwander-Tetri, B. A., Rinella, M., & Sanyal, A. J. (2018). Mechanisms of NAFLD development and therapeutic strategies. *Nature medicine*, 24(7), 908-922. doi:10.1038/s41591-018-0104-9. <https://doi.org/10.1038/s41591-018-0104-9>
- [2] Lehman, E.M., 2008. Dynamics of Liver Disease in Egypt: Shifting Paradigms of a Complex Etiology.

- [3] Hsu, C.L., Wu, F.Z., Lin, K.H., Chen, Y.H., Wu, P.C., Chen, Y.H., Chen, C.S., Wang, W.H., Mar, G.Y. and Yu, H.C., 2019. Role of Fatty Liver Index and Metabolic Factors in the Prediction of Nonalcoholic Fatty Liver Disease in a Lean Population Receiving Health Checkup. *Clinical and translational gastroenterology*, 10(5), pp.1-8. <https://doi.org/10.14309/ctg.0000000000000042>
- [4] Bellentani, S., Scaglioni, F., Marino, M. and Bedogni, G., 2010. Epidemiology of non-alcoholic fatty liver disease. *Digestive diseases*, 28(1), pp.155-161. <https://doi.org/10.1159/000282080>
- [5] Chang Y, Jung HS, Cho J, et al. Metabolically healthy obesity and the development of nonalcoholic fatty liver disease. *Am J Gastroenterol*. 2016;111:1133-1140. <https://doi.org/10.1038/ajg.2016.178>
- [6] Byrne C, Targher G. NAFLD: a multisystem disease. *J Hepatol*. 2015;62(1 Suppl):S47-S64. <https://doi.org/10.1016/j.jhep.2014.12.012>
- [7] Vernon G, Baranova A, Younossi ZM. Systematic review: the epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Aliment Pharmacol Ther*. 2011;34:274-285. <https://doi.org/10.1111/j.1365-2036.2011.04724.x>
- [8] Zelber-Sagi S, Nitzan-Kaluski D, Halpern Z, Oren R. Prevalence of primary non-alcoholic fatty liver disease in a population-based study and its association with biochemical and anthropometric measures. *Liver Int*. 2006;26:856-863. <https://doi.org/10.1111/j.1478-3231.2006.01311.x>
- [9] Hagstrom H, Tynelius P, Rasmussen F. High BMI in late adolescence predicts future severe liver disease and hepatocellular carcinoma: a national, population-based cohort study in 1.2 million men. *Gut*. 2017 Mar 20. [https://doi.org/10.1016/S0168-8278\(17\)30403-8](https://doi.org/10.1016/S0168-8278(17)30403-8)
- [10] Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. *Hepatology*. 2018;67:123-133. <https://doi.org/10.1002/hep.29466>
- [11] Wong RJ, Aguilar M, Cheung R, et al. Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. *Gastroenterology*. 2015;148:547-555. <https://doi.org/10.1053/j.gastro.2014.11.039>
- [12] Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology*. 2018;67:328-357. <https://doi.org/10.1002/hep.29367>
- [13] European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD); European Association for the Study of Obesity (EASO). EASL-EASD-EASO clinical practice guidelines for the management of nonalcoholic fatty liver disease. *J Hepatol*. 2016;64:1388-1402. <https://doi.org/10.1016/j.jhep.2015.11.004>
- [14] Bertot LC, Jeffrey GP, Wallace M, et al. Nonalcoholic fatty liver disease-related cirrhosis is commonly unrecognized and associated with hepatocellular carcinoma. *Hepatol Commun*. 2017;1:53-60 <https://doi.org/10.1002/hep4.1018>
- [15] Fierbinteanu-Braticevici, C., Negreanu, L. and Tarantino, G., 2013. Is fatty liver always benign and should not consequently be treated. *J Physiol Pharmacol*, 64(1), pp.3-9.
- [16] Gameel, T.A., Rady, S., ElBahnasy, K.A. and Kamal, S.M., 2019, April. Novel Biomarkers from genes in the apoptotic pathway for Prediction of HCC Progression using Association Rule Mining. In *Proceedings of the 2019 8th International Conference on Software and Information Engineering* (pp. 217-221). ACM. <https://doi.org/10.1145/3328833.3328863>
- [17] Younossi, Z.M., Koenig, A.B., Abdelatif, D., Fazel, Y., Henry, L. and Wymer, M., 2016. Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of

- prevalence, incidence, and outcomes. *Hepatology*, 64(1), pp.73-84 <https://doi.org/10.1002/hep.28431>
- [18] Alexander, M., Loomis, A.K., van der Lei, J., Duarte-Salles, T., Prieto-Alhambra, D., Ansell, D., Pasqua, A., Lapi, F., Rijnbeek, P., Mosseveld, M. and Waterworth, D.M., 2019. Risks and clinical predictors of cirrhosis and hepatocellular carcinoma diagnoses in adults with diagnosed NAFLD: real-world study of 18 million patients in four European cohorts. *BMC medicine*, 17(1), p.95. <https://doi.org/10.1186/s12916-019-1321-x>
- [19] Jain, A. and Mourya, C., 2019. LIVER ELASTOGRAPHY: CLINICAL SIGNIFICANCE AND OUR EXPERIENCE. *International Journal of Scientific Research*, 8(3).
- [20] El-Seoud, S.A., Mady, A. and Rashed, E., 2019. An Interactive Mixed Reality Ray Tracing Rendering Mobile Application of Medical Data in Minimally Invasive Surgeries.
- [21] Sigrist, R.M., Liau, J., El Kaffas, A., Chammas, M.C. and Willmann, J.K., 2017. Ultrasound elastography: review of techniques and clinical applications. *Theranostics*, 7(5), p.1303. <https://doi.org/10.7150/thno.18650>
- [22] Baumeler, S., Jochum, W., Neuweiler, J., Bergamin, I. and Semela, D., 2019. Controlled attenuation parameter for the assessment of liver steatosis in comparison with liver histology: a single-centre real life experience. *Swiss medical weekly*, 149(1718). <https://doi.org/10.4414/smw.2019.20077>
- [23] Di Costanzo, A., Pacifico, L., Chiesa, C., Perla, F.M., Ceci, F., Angeloni, A., D'Erasmus, L., Di Martino, M. and Arca, M., 2019. Genetic and metabolic predictors of hepatic fat content in a cohort of Italian children with obesity. *Pediatric research*, p.1. <https://doi.org/10.1038/s41390-019-0303-1>
- [24] Wang, A.Y., Dhaliwal, J. and Mouzaki, M., 2018. Lean non-alcoholic fatty liver disease. *Clinical Nutrition*. <https://doi.org/10.1016/j.clnu.2018.08.008>
- [25] Ballestri, S., Nascimbeni, F., Baldelli, E., Marrazzo, A., Romagnoli, D. and Lonardo, A., 2017. NAFLD as a sexual dimorphic disease: role of gender and reproductive status in the development and progression of nonalcoholic fatty liver disease and inherent cardiovascular risk. *Advances in therapy*, 34(6), pp.1291-1326 <https://doi.org/10.1007/s12325-017-0556-1>
- [26] Amer, H.M., Abou-Chadi, F.E., Kishk, S.S. and Obayya, M.I., 2019. A CAD System for the Early Detection of Lung Nodules Using Computed Tomography Scan Images. *International Journal of Online Engineering*, 15(4). <https://doi.org/10.3991/ijoe.v15i04.9837>
- [27] Holzinger, A., Dehmer, M. and Jurisica, I., 2014. Knowledge discovery and interactive data mining in bioinformatics-state-of-the-art, future challenges and research directions. *BMC bioinformatics*, 15(6), p.11. <https://doi.org/10.1186/1471-2105-15-S6-11>
- [28] Ilayaraja, M. and Meyyappan, T., 2013, February. Mining medical data to identify frequent diseases using Apriori algorithm. In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (pp. 194-199). IEEE. <https://doi.org/10.1109/ICPRIME.2013.6496471>
- [29] Vijayarani, S. and Dhayanand, S., 2015. Liver disease prediction using SVM and Naïve Bayes algorithms. *International Journal of Science, Engineering and Technology Research (IJSETR)*, 4(4), pp.816-820.
- [30] David, S.K., Saeb, A.T., Rafiullah, M. and Rubeaan, K., 2019. Classification Techniques and Data Mining Tools Used in Medical Bioinformatics. In *Big Data Governance and Perspectives in Knowledge Management* (pp. 105-126). IGI Global. <https://doi.org/10.4018/978-1-5225-7077-6.ch005>
- [31] Gameel, T.A., Rady, S., El-Bahnasy, K.A. and Kamal, S.M., 2017, December. Prediction of liver cancer development risk in genotype 4 Hepatitis C patients using knowledge discovery modeling. In 2017 Eighth International Conference on Intelligent Computing and

Information Systems (ICICIS) (pp. 410-415). IEEE. <https://doi.org/10.1109/INTELCIS.2017.8260080>

- [32] Shaikh, M.R., McNicholas, P.D., Antonie, M.L. and Murphy, T.B., 2018. Standardizing interestingness measures for association rules. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 11(6), pp.282-295. <https://doi.org/10.1002/sam.11394>

## 7 Authors

**Tasneem Gameel** is a Master's student and teaching assistant at Bioinformatics program, Information Systems Department at Faculty of Computer and Information Sciences, Ain Shams University in Egypt. Email: [tasneem.gameel@cis.asu.edu.eg](mailto:tasneem.gameel@cis.asu.edu.eg)

**Sherine Rady** is an Associate Professor at the Faculty of Computer and Information Sciences of Ain Shams University in Cairo, Egypt. She holds a B.Sc in Electrical Engineering (Computer and Systems), Ain Shams University. She got her M.Sc. in Computer and Information Sciences from Ain Shams University and her Ph.D. from University of Mannheim in Germany. Research interests are Data Mining, Medical informatics, and Robotics. Email: [sherine.rady@bue.edu.eg](mailto:sherine.rady@bue.edu.eg)

**Sanaa Kamal** is a Professor at Faculty of Medicine, Ain Shams University. Research interests are Hepatology, Infectious diseases and Gastroenterology research topics. Email: [sanaakamal@gmail.com](mailto:sanaakamal@gmail.com)

Article submitted 2020-02-06. Resubmitted 2020-03-21. Final acceptance 2020-03-25. Final version published as submitted by the authors.