Journal of Universal Computer Science, vol. 30, no. 8 (2024), 1068-1088
submitted: 13/11/2023, accepted: 17/2/2024, appeared: 28/8/2024

Enhancing Health Risk Prediction in Internet of Medical Things: Leveraging Association Rule Mining

Anwar Ahmed Khan
(Millennium Institute of Technology & Entrepreneurship, Karachi, Pakistan
https://orcid.org/0000-0002-2237-5124, yrawna@yahoo.com)

Shama Siddiqui
(DHA Suffa University, Karachi, Pakistan
https://orcid.org/0000-0002-3547-0307, shamasid@hotmail.com)

Indrakshi Dey
(Walton Institute for Information and Communication Systems Science, Ireland
https://orcid.org/0000-0001-9669-6417, indrakshi.dey@waltoninstitute.ie)

Abstract: Due to rapid advancements in the field of Internet of Medical Things (IoMT), a continuous influx of health data is being generated at a large scale. The primary objective of IoMT solutions is to transmit critical health data from patients to remote locations in real-time. Apart from remote patient monitoring, the extensive collection of health data offers opportunities for uncovering noteworthy patterns and potential risks associated with future diseases. This study introduces a novel risk prediction approach, namely Association Rule Mining for Risk Prediction (ARMR), which integrates an IoMT framework with the emerging machine learning technique known as Association Rule Mining (ARM). The proposed scheme employs a dataset obtained from various hospitals. The findings demonstrate that ARMR effectively extracts rules to identify a patient's risk of heart disease by considering demographic, physiological, and lifestyle data. Moreover, intriguing, and unexpected patterns and associations in the disease data can be identified, aiding medical professionals in guiding diagnosis and treatment decisions more efficiently.

Keywords: predictive analytics; diabetes; heart diseases; data mining; remote monitoring
Categories: L.4, I.2.1, I.2.6, J.3
DOI: 10.3897/jucs.115261

1 Introduction

Internet of Medical Things (IoMT) has been regarded as an emerging technology that holds a potential to revolutionize the healthcare sector. The technology deals with remote collection of data from patients' wearable or implantable sensors and its transmission to the local/handheld and far off remote devices [Iqbal, 2021]. Using the sporadically collected data, it becomes possible to timely diagnose the medical conditions and take appropriate preventive measures. Moreover, customized and effective treatment strategies for each patient may be developed with the help of IoMT data. Various solutions for monitoring diverse diseases and medical conditions have been introduced lately, such as detecting seizure [Sayeed, 2019], measuring blood glucose [Joshi, 2020], measuring vital parameters [Sethuraman, 2021], monitoring stress via food consumption patterns [Rachakonda, 2020], monitoring stress via cortisol levels [Nath, 2021], monitoring sleep quality [Rachakonda, 2021], and many more. Interestingly, the data generated by IoMT does not only support the individual patients, but it provides an insight into interesting health parameters about entire populations[Dey, 2017].

Data mining techniques can add a significant value to the IoMT by modelling, clustering, and categorizing a large volume of data [Guo, 2020]. Generally, data mining process deals with analysing data from various perspectives and angles. For example, the data collected by IoMT without using mining algorithm can just be helpful for continuous remote monitoring of patients, but the integration of data mining algorithms with IoMT enhances the scope and introduces smart decision support based on the extraction of unique features and patterns [Jagadeeswari, 2018]. As a result, delivery of a high-level personalized care becomes possible. Hence, data mining and IoMT integrated with each other may serve as a major milestone for digital transformation of healthcare.

Along with the other smart healthcare technologies, use of advanced data mining and analytics algorithms for analysing the streams of data generated by IoMT offers opportunities for discovering new information and predicting risks for individuals and populations which could in turn play a vital role to improve the quality of life [Jagadeeswari, 2018] [Tuli, 2020]. Association Rule Mining (ARM) refers to the use of machine learning models for analysing data to infer patterns and co-occurrences in a database. The focus of ARM is to identify the features with most occurrences in the data and also to filter out the attributes which are not common [Altaf, 2017]. Hence, the algorithm has often been used for extracting the rules from huge volumes of data for the purpose of prediction. Although ARM has been used for various IoT solutions [Chen, 2020], to the best of our knowledge, the model has not yet been integrated with IoT solutions for risk assessment using population disease data.

In our study, we present the utilization of Association Rule Mining (ARM) to classify patients' health risks by considering a range of demographic, lifestyle, and physiological parameters. ARM is employed to identify the key factors that significantly influence a patient's susceptibility to diseases like diabetes or heart conditions. To the best of our understanding, no prior research has explored the integration of ARM with IoMT to achieve risk categorization for patients. The major contributions of this work are as below:

- To propose a novel Association Rule Mining enabled IoMT architecture for health risk categorization.
- To develop a high-level implementation for extracting rules from the population health data to identify the possible presence or absence of diseases.
- To illustrate real-world scenarios by applying the proposed technique on existing datasets of heart diseases to study the impact of physiological parameters and lifestyle attributes on probability of disease occurrence.

The rest of this paper is organized as follows. In Section 2, we present a brief review of related work. In Section 3, the proposed architecture integrating IoMT and ARM has been presented. The preliminary high-level implementation of ARM has been discussed in Section 4. Section 5 presents the simulation results obtained using heart disease datasets. Section 6 sheds light on the ethical aspect of this study, and finally, Section 6 concludes the work and identifies the directions for future research.

2 Related Work

Several Fuzzy Association Rule Mining based algorithms have been proposed in literature for various cutting-edge applications of IoT and Wireless Sensor Networks (WSN); some of them have been summarized in Table 1. In [Khedr, 2020], association rule has been used to mine data from the sensor nodes before forwarding it to the cluster head nodes. Using the proposed algorithm, the sensors perform local processing and only forward the crucial high-level statistics to the cluster heads which not only improves the network performance by quickly processing the massive data generated by the sensors but also reduces the computation overhead and hence the network lifetime is improved. Behaviour pattern mining has also been regarded as an interesting enhancement to IoT where data collected from nodes is proposed to be used for monitoring behaviours in the industrial, healthcare, and smart city settings [Rashid, 2020].

Due to the increased automation, huge volumes of agricultural data are being generated which increase the probability of accurate predictions about environmental and other risks [Colizzi, 2020]. For example, Fuzzy Association Rule Mining (FARM) has been proposed for plants in [Wedashwara, 2019] to extract rules from IoT data about the environment. The environmental data required for maintaining plant health including soil and air humidity, light intensity, rainfall, and temperature was collected and changes in the environment were detected using FARM. Another interesting and crucial application for large-scale agriculture is pesticide management. In [Divya, 2014], dynamic data from a WSN mounted on a thrip of groundnut pest was collected. The data was fed into ARM rule extraction system to identify the risks based on crop, weather, and pesticide information. Such schemes promise to improve the quality production of crops via performing efficient data analytics. Environment pollution monitoring has also been facilitated by using ARM and IoT. In a system proposed in [Tuysuzoglu, 2020], the concentration of different air pollutants and their features is regularly monitored, and a dataset generated at 21 stations in Turkey has been used. The data is fed to ARM system for extracting the relationships between different pollutants and their features. The rules are used for generating alerts about the rising air pollution levels at some locations.

Similar technique, Evolutionary Fuzzy Association Rule Mining (EFARM) has been used in [Wedashwara, 2019] where the electrical energy audit using IoT data has been conducted. Data was collected from five rooms for a period of one week and Fuzzy Rules and Tree Based Evolutionary Computation (EC) are interpreted. The system successfully captured electricity consumption patterns, compared the consumption patterns based on location, identified similarities and dissimilarities between the consumption patterns of different locations and compared the power consumption patterns in association with other measurable parameters such as temperature and light intensity.

Food supply chain monitoring has been an interesting area for IoT in general. In [Wang, 2017], authors developed an ARM based scheme to detect food safety risks. The system is designed to generate timely alerts for the managers so that they could make the decisions crucial for ensuring quality of food products. Expert analysis is conducted whenever a risk is detected in order to identify if a warning must be generated. ARM has also been used for smart building management, particularly for the purpose of energy conservation. The modern buildings deploy various systems such as variable refrigerant flow (VRF) system for cooling [Qian, 2020] which require energy conservation and regular maintenance. Therefore, the massive data generated by building operations has been integrated by big data analytics for guaranteeing efficient long-term operation. In [Fan, 2018], big data about building operations has been collected using sensors to optimize various processes such as lightning, heating, air conditioning, energy consumption, etc. The authors presented a case study for educational building and the extracted rules provided an insight about operation patterns of the building, operational deficiencies, and energy conservation opportunities.

ARM has also been used for detecting the presence of malware in IoT networks. In [Ozawa, 2020], big stream data collected on a large-scale darknet was analysed using ARM to identify the similarities between the parameters associated with attack. The rules from sensor darknet were extracted for the parameters including type of service, destination ports and TCP window sizes. The technique was shown to be efficient for early detection of malware by analysing the source code that is received just before the attack.

Ambient Assisted Living (AAL) is an emerging concept in the field of smart homes, where data about physical activity level and other lifestyle parameters is captured to analyse the health state of a person [Marques, 2021]. In such settings, the healthcare applications are facilitated due to the ability of algorithm to learn and discover the human activity patterns. Frequent pattern mining and ARM has been proposed in [Yassine, 2017] for identifying the electricity consumption pattern (at appliance level) for users in smart home. The electricity consumption pattern is subsequently used to identify any behaviour change such as not using the shower or not cooking.

Similarly, ARM has been deployed for various healthcare scenarios in the recent past: [Saha, 2023] applied ARM on pharmacy's database to identify association between the medicines prescribed to hospitalized patients using Apriori approach; the rules identified are expected to help the pharmacies to manage the demand and supply of medicines. Clustered ARM was applied for assessment of psychiatric disorders in [Bertl, 2023], where comorbidity extraction and indicator disease mining have been performed based on health insurance billing data. Liver diseases dataset was used for rule extraction through ARM in [Patel, 2022]. Moreover, impact of hypertension and diabetes on the risk of myocardial infarction has been assessed using ARM in [Singh, 2022].

A brief survey of medical literature reveals that there is a strong correlation between the fitness level, physical activity level and occurrence of diseases such as diabetes and heart diseases. It has been found by [Amidei, 2022] lower risk of cardiovascular outcomes was reported for the men who engaged in a physical activity for 20 minutes a day. It has even been reported by [Ho, 2022] that the probability of heart failure reduced for physically active patients. Similarly, in the summary document released by the American Society for Preventive Cardiology (ASPC) high blood sugar, high blood pressure, dyslipidaemia, and increased body fat have been regarded as cardiometabolic risk factors [Bays, 2022]. It has also been found by that obese patients have a higher risk for cardiovascular diseases, where obesity has been measured using BMI [Valenzuela, 2022]. Thus, in the light of these associations, we have identified the antecedents and consequents for our datasets.

Reference	Method	Gaps
[Saha, 2023]	Rules are mined from the Pharmacy database using Apriori algorithm	The associations between medicines are identified, not the health risks
[Bertl, 2023]	Rules are mined for identifying psychiatric risks	Associations were identified for the psychiatric disorders, not for heart diseases
[Singh, 2022]	Rules are mined for identifying impact of hypertension on CVD	Rules are mined for associations between smoking habits, hypertension, and CVD. Physical activity and fitness level not studied.
[Khedr, 2020]	Rules are mined at nodes before forwarding data to the cluster heads to preserve the network resources.	Focus is on improving network lifetime, rather than health risk prediction.
[Rashid, 2020]	Rules are mined for behavioural pattern identification in the smart city, healthcare and industrial IoT networks.	Behaviours of individuals or machines are identified, not the health risks
[Wedashwara, 2019]	Rules are identified for monitoring plant health	Rules about optimal humidity, soil and light intensity, rainfall conditions have been mined, not about humans
[Divya, 2014]	Rules were mined to optimize the groundnut production	Optimal relationships between groundnut crop, weather and pests were identified, the same could be developed for human health data
[Tuysuzoglu, 2020]	Rules were identified for concentration of different air pollutants and their features	Relationships were extracted to predict the pollution levels at different regions, and the focus is not on healthcare
[Wedashwara, 2019]	Rules were extracted for identifying the electricity consumption patterns	Relationships between electricity consumption, temperature and light intensity were extracted. Similar study could be conducted for healthcare
[Wang, 2017]	Rules were mined to identify the food safety risks while moving through the supply chain	Relationships between key variables such as type, quantity, temperature related movement, etc. were identified, and the focus is not on health risk prediction
[Fan, 2018]	Rules were mined for automated building management and resource optimization	Relationship between building operations, operational deficiencies and energy conservation

		opportunities were identified, and healthcare aspects were not included
[Ozawa, 2020]	Rules were extracted for detecting malware in IoT networks	Parameters associated with network attacks were identified, not the healthcare
[Yassine, 2017]	Rules were mined for identifying the electricity consumption pattern for smart homes	Focus of extracted relationships was to identify the electricity consumption pattern of people, not their health risks

Table 1:Existing Association Rule Mining Schemes for IoT and Sensor Networks

3 Proposed Design of ARMR

The proposed system integrates conventional IoMT architecture with ARM. IoMT would provide the patients with an opportunity for being continuously monitored without disturbing their routine activities. The proposed risk prediction architecture, ARMR has been illustrated in Figure 1.

As shown in Figure 1, health data is collected from patients using wearable, ambient or implanted sensors. Data from multiple users is transmitted to cloud via their handheld devices and appropriate wireless communication infrastructure. With the span of time, the volume of data collected at the remote database shall increase facilitating the mining process. The interesting patterns shall be identified for individual patients as well as for the population dataset. Based on the population dataset, the extracted rules shall guide the physicians to identify the conditional probability for occurrence of a certain diseases/condition in the patients. Subsequently, the machine learning models can adapt the new rules from the updated datasets.

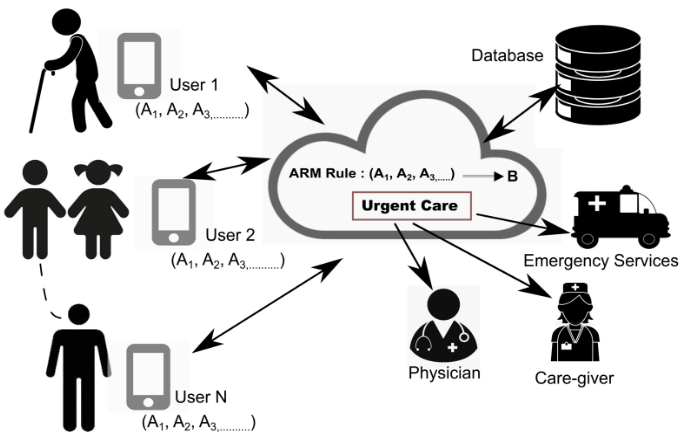


Figure 1: ARMR- ARM enabled IoMT architecture for risk prediction

4 Implementation of ARMR

The major concepts of ARM inherited for the proposed system are summarized below:

- Antecedent- it is an attribute/variable that is found in the dataset.
- Consequent- it is an attribute which is found in combination with antecedent. The probability of occurrence of consequent in association with occurrence of antecedent is mainly studied in ARM.
- Support- it measures the frequency with which a certain if/then (antecedent/consequent) relationship is found in the dataset.
- Confidence- this indicates the frequency with which each if/then (antecedent/consequent) relationship holds true for the dataset.
- Lift- this factor provides a ratio between the observed frequency of antecedent/consequent and the expected frequency.

To formulate the proposed algorithm mathematically, we consider three independent variables (antecedents) for our IoMT data, such as fitness level (may take value fit or unfit), activity level (may take value active or non-active) and risk due to blood pressure (may take value low or high). Let us consider the three antecedents A1, A2, and A3 corresponding to fitness level: unfit, activity level: non-active, and risk due to blood pressure: high. Here, ARM can be used to predict the occurrence of dependent variable (consequent) which is assumed as risk for the patient to get a heart disease and may take values yes or no; we assume B as Yes. Based on these assumptions, the expressions for support, confidence, and lift are presented in the info-graphic representation in Figure 2, to predict the occurrence of B when all A1, A2, and A3 occur.

It is evident that the rules can be extracted for occurrence of disease in association with the physiological and lifestyle factors using the existing datasets. The rules can then serve as a guideline for risk assessment for the patients in future. The algorithm developed for extracting rules from IoMT dataset using ARM has been presented in Algorithm 1.

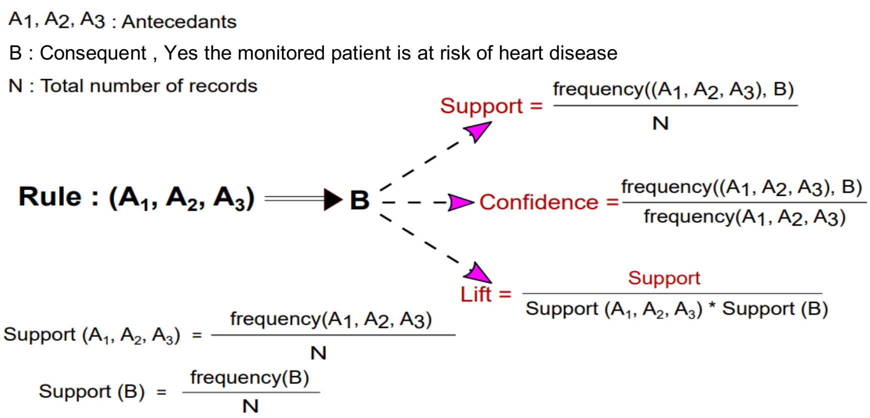


Figure 2: Info-graphic representation of the steps in ARM

Algorithm 1 Association Rule Mining for Health Monitoring Database *HD*

Require: Health Monitoring Database *(HD)*
Ensure: Discovery of frequent records along with occurrence of disease. Reports in a separate database, Mined Rules *(MR)*

Determine the database size *S* for *HD*.
for All records in *HD*
  Assign numeric values to antecedents.
  Mine unique strings (combination of antecedents in numeric form) from *HD* to *MR*.
  Count occurrences of unique strings from *HD* and report in *MR* as *String_Count*.
  Count occurrences of each consequent against unique strings of antecedents from *HD* and report in *MR* as *Frequency*.
end for
for All records in *MR*
  Compute and store *Support*, *Confidence* and *Limit*.
end for
Sort according to *String_Count* which would provide the rules repeated most times.

5 Simulation Results and Analysis

For validating the efficiency of proposed ARMR and demonstrating the rule extraction from real-world data, we conducted simulation using publicly available clinical data collected from a general Hospital based at Karachi and from Cleveland database of heart disease [Janosi, 1988]. The clinical datasets, generated using conventional vital parameter monitoring methods, have been used to demonstrate the working of ARMR. In future, the data collection will be performed using IoMT instead of using data recorded at clinics. We present the rule extraction and results for each dataset in this section.

5.1 Heart Disease Dataset Obtained from Cardiac Clinic

The dataset was obtained from a general hospital at Karachi from Cardiac care clinic. The patients were between age range 40-60 and the data collection period was December 2020 to July 2021. On total, dataset comprised of 2000 records and 5 attributes. The attributes (antecedents) of fitness level, physical activity level, blood pressure, and heart rate were assigned as independent variable/antecedent. The correlation of these attributes was found with the occurrence of heart disease (heart attack or stroke) which has been taken as a dependent variable/consequent. Based on the analysis of physicians, the scales for each antecedent were defined qualitatively as below:

- Fitness level (defined based on Body Mass Index (BMI)): Extremely fit, fit, moderately fit, slightly unfit, unfit, and extremely unfit.
- Physical Activity Level (defined based on whether or not the patient was practising recommended physical activity for at least 30 minutes a day): Active, non-active.
- Risk due to Blood Pressure (defined for each patient based on their age, fitness level and BP reading): Normal, at risk (pre-hypertension), high (hypertension).

- Risk due to heart rate (defined for each patient based on their age, fitness level, heart health, ECG measures and pulse rate reading): Normal, at risk, high.

These attributes have been summarized in Table 2. The rule extraction algorithm used for applying ARM to the datasets is summarized in Algorithm 1. Python was used for categorizing the data and extracting rules to predict the consequent (Diagnosis of Heart Disease). On total, 108 unique combinations were found for 2000 records. The maximum repetitions for unique string were 59, and least repetition was found 2 times. Top 30 rules have been presented in Table 3 along with the unique string count, frequency, support, confidence, and support for each association. The interesting patterns depicting the association between antecedents and consequent have been plotted in Figure 3.

Attributes	**Categorization**	
Fitness Level	1: Extremely fit 2: Fit 3: Moderately fit	4: Slightly unfit 5: Unfit 6: Extremely unfit
Physical Activity Level	1: Active 2: Non-Active	
Blood Pressure in mmHg	1: Normal 2: At risk	3: High
Heart Rate (bpm)	1: Normal 2: At risk	3: High

Table 2: Attributes in Clinical Dataset

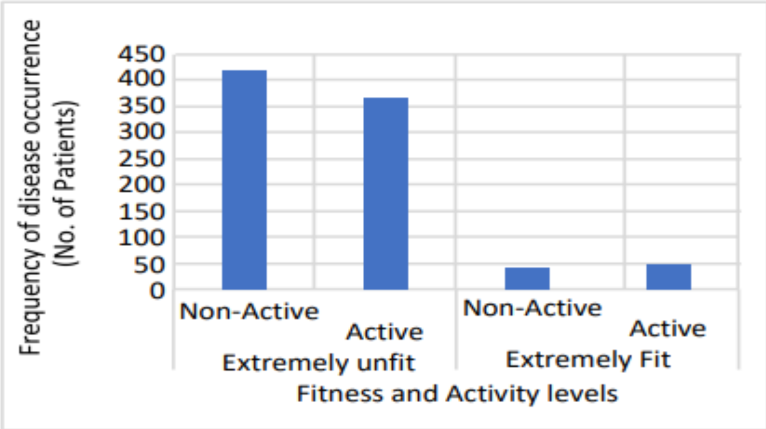
S.No	**Rule**	**String Count**	**Frequency**	**Support**	**Confidence (%)**	**Lift**
1	Extremely unfit, Non-Active, BP High, HR At Risk → Yes	59	52	0.026	88.14	1.34
2	Extremely unfit, Non-Active, BP High, HR High → Yes	58	42	0.021	72.41	1.10
3	Unfit, Active, BP High, HR At Risk → Yes	55	41	0.0205	74.55	1.13
4	Extremely unfit, Active, BP At Risk, HR High → Yes	54	44	0.022	81.48	1.24
5	Slightly Fit, Non-Active, BP High, HR Normal → Yes	53	49	0.0245	92.45	1.41
6	Unfit, Non-Active, BP At Risk, HR High → Yes	52	42	0.021	80.77	1.23
7	Extremely unfit, Active, BP High, HR High → Yes	49	30	0.015	61.22	0.93
8	Unfit, Non-Active, BP High, HR At Risk → Yes	49	31	0.0155	63.27	0.96
9	Slightly Fit, Non-Active, BP Normal, HR High → Yes	47	36	0.018	76.6	1.17

10	Extremely unfit, Non-Active, BP At Risk, HR At Risk → Yes	47	34	0.017	72.34	1.10
11	Unfit, Active, BP High, HR High → Yes	47	39	0.0195	82.98	1.26
12	Unfit, Non-Active, BP Normal, HR Normal → Yes	46	46	0.023	100	1.52
13	Extremely unfit, Active, BP High, HR Normal → Yes	46	29	0.0145	63.04	0.96
14	Unfit, Non-Active, BP High, HR High → Yes	46	40	0.02	86.96	1.32
15	Unfit, Non-Active, BP Normal, HR High → Yes	44	38	0.019	86.36	1.38
16	Extremely unfit, Non-Active, BP At Risk, HR Normal → Yes	44	40	0.02	90.91	1.31
17	Unfit, Active, BP At Risk, HR High → Yes	43	32	0.016	74.42	1.13
18	Extremely unfit, Non-Active, BP High, HR Normal → Yes	43	37	0.0185	86.05	1.31
19	Slightly Fit, Non-Active, BP At Risk, HR High → Yes	43	34	0.017	79.07	1.20
20	Extremely unfit, Non-Active, BP At Risk, HR High → Yes	43	37	0.0185	86.05	1.31
21	Slightly Fit, Non-Active, BP High, HR At Risk → Yes	43	30	0.015	69.77	1.06
22	Extremely unfit, Active, BP At Risk, HR Normal → Yes	43	43	0.0215	100	1.52
23	Extremely unfit, Non-Active, BP Normal, HR High → Yes	43	26	0.013	60.47	0.92
24	Extremely unfit, Active, BP Normal, HR High → Yes	42	34	0.017	80.95	1.23
25	Slightly Fit, Non-Active, BP At Risk, HR At Risk → Yes	42	41	0.0205	97.62	1.49
26	Extremely unfit, Non-Active, BP Normal, HR At Risk → Yes	41	30	0.015	73.17	1.11
27	Unfit, Non-Active, BP At Risk, HR At Risk → Yes	41	28	0.014	68.29	1.04
28	Extremely unfit, Active, BP At Risk, HR At Risk → Yes	41	37	0.0185	90.24	1.37
29	Unfit, Non-Active, BP Normal, HR At Risk → Yes	39	24	0.012	61.54	0.94
30	Unfit, Non-Active, BP High, HR Normal → Yes	38	34	0.017	89.47	1.36

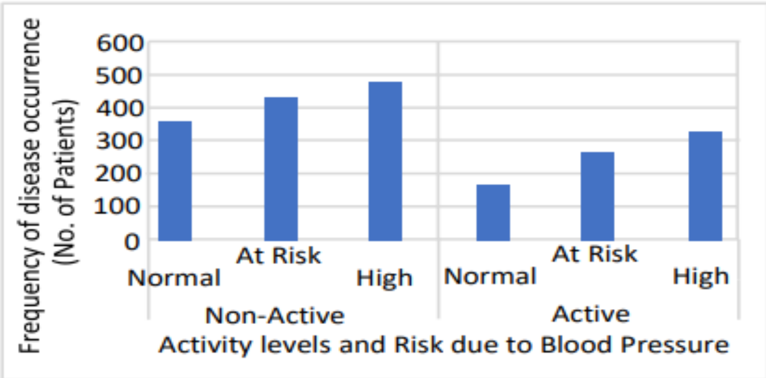
Table 3: Rule Extraction for Clinical Dataset

Figure 3(a) shows the number of patients against the association between fitness and activity levels. As previously mentioned, the fitness level in this dataset refers to the physical health determined via measuring BMI at the clinic. Each patient had been provided a guideline for engaging in physical activity to reduce the risk of obesity and maintain heart health. First observation inferred from the figure is that there was a significant difference for the frequency of heart disease occurrence for fit and unfit patients. Moreover, it can be observed from the figure that the patients who were extremely unfit and non-active suffered from heart disease the most. Similarly, the patients who were active from the extremely unfit category, also had a high occurrence of heart disease. However, there is a slight difference between active and non-active patients. This difference indicates that if the patients belong to the extremely unfit category, and they begin exercising, their risk of heart disease decreases. In contrast, for the group of patients who were categorized as extremely fit, there is not much difference between the risks for active and non-active patients.

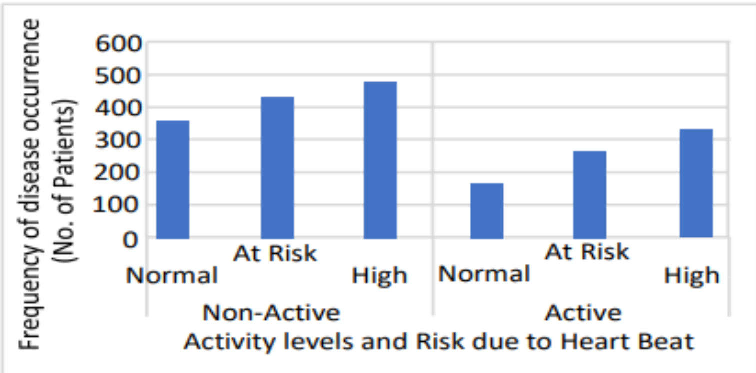
In Figures 3(b) and 3(c), the frequency of heart disease occurrence has been presented against association between the antecedents of activity levels, blood pressure (BP) and heart rate, respectively. As previously shown in the Table 2, patients were sub-categorized into 3 groups "Normal", "At Risk" and "High Risk". In Figures 3(b) and 3(c), we attempted to visualize the association between activity levels and categories of the health risk patients belong to. In both Figures 3(b) and (c), a clear difference in the heart diseases risk has been seen for the non-active and active patients. Figure 3(b) shows that for the patients who are non-active, the occurrence of heart diseases was most common for the patients who had their BP falling in the high range. The frequency of heart disease occurrence showed reduction if the value of BP was at risk or normal. Thus, the patients with lower risk of BP within non-active group suffered from lesser heart disease occurrences. Similar pattern has been observed for the active group of patients. Moreover, figure 3(c) also showed similar associations between activity levels, risk due to heart rate, and occurrence of heart diseases.



(a)



(b)



(c)

Figure 3: Patterns Identified from Clinical Dataset for impact of antecedents on the occurrence of heart disease. (a) Impact of fitness and activity levels, (b) Impact of Activity levels and Blood Pressure, (c) Impact of Activity Levels and Heart Rate

5.2 Heart Disease Dataset Obtained from Cleveland Database

The second heart disease dataset was obtained from UCI repository. The dataset comprised of 303 records, all of which were used for rule extraction in this work. For this dataset, physiological parameters were recorded for the patients while they were at rest and also as they exercised to identify the impact of exercise on the risk of heart disease. The heart disease has been defined by the degree of narrowing of arteries; if the narrowing was less than 50 percent, the heart disease was considered to be absent (No) and was assumed to be present (Yes), otherwise. On total, the dataset comprises of 76 attributes, but we only used 7 relevant attributes. Selected attributes represent demographic, and health status data. In future, we plan to extend the proposed methodology to test the ARMR by using the data collected from IoMT solutions. Brief description of each attribute in the dataset has been presented in Table 4. Also, the categories assigned for each attribute in the original dataset have been presented. During the pre-processing stage, we categorized some of the attributes differently in order to optimize the execution of ARMR. We present our categorization for each attribute in column 3 of Table 4. Whether exercise was efficient or not was decided based on the relationship given by American Heart Rate Association, *Maximum heart rate = 220 - age of the person* [American Heart Association, 2021].

Attributes	**Dataset Categorization/Values**	**ARMR Categorization/Values**
Demographic		
Age	29 -77	1: < 40 2: >40
Gender	1: Male 2: Female	1: Male (M) 2: Female (F)
Resting Health Stats		
Chest Pain Type (CP)	1: typical angina 2: atypical angina	0: angina (T) or atypical angina (AT)
	3: non-anginal pain 4: asymptomatic	1: non-anginal pain (NP) 2: asymptomatic (AS)
Resting Blood Pressure in mmHg (trestbps)	94 - 200	1: <120 (Normal) 0: >= 120 (High Risk)
Blood Cholesterol in mg/dl (chol)	126 - 564	1: <220 (Normal) 0: >= 220 (High Risk)
Fasting blood sugar 120 mg/dl (fbs)	1: True 2: False	1: True (Normal) 2: False (High Risk)
Exercise Related Health Stats		
Maximum Heart Rate Achieved (thalach)	71 – 202	1: Inefficient Exercise 0: Efficient Exercise
Diagnosis of Heart Disease (Angiographic Disease Status)	0: < 50% diameter narrowing 1: > 50% diameter narrowing	0: < 50% diameter narrowing (A) 1: > 50% diameter narrowing (P)

Table 4: Attributes in Cleveland Dataset [Janosi, 1988]

If the heart rate achieved during exercise is 50% – 85% of the maximum heart rate, the exercise is considered efficient. We set the threshold to be 75% of the maximum heart rate and if the person reached this value, he was considered to performing efficient exercise.

Algorithm 1 was used to extract the rules from Cleveland dataset. Since the dataset only comprised of 303 records, 64 unique combinations were obtained based on 7 antecedents, which were used to define rules. We identified the number of times each combination of attributes was repeated and the rules for predicting the heart disease risk were extracted accordingly. The 24 top rules have been shown in Table 5, where 30 is the highest number of repetition and 4 is the lowest.

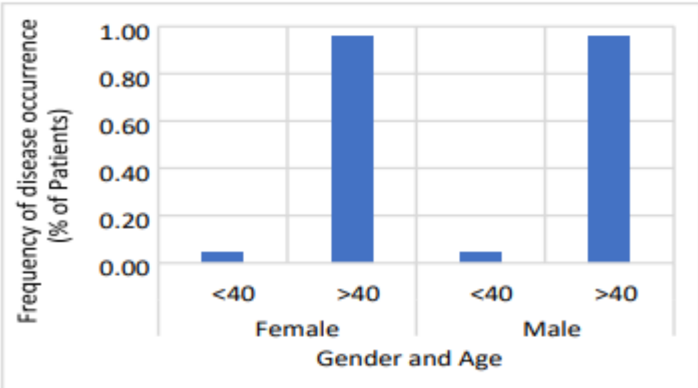
The most interesting patterns identified for impacting the presence of heart diseases are visually presented in Figures 4 and 5.

S.No	Rule	String Count	Frequency	Support	Confidence (%)	Lift
1	>40, M, Asymptomatic, BP High, Chol High, FBS_Normal, Inefficient, Exercise → Yes	30	21	0.069	70	0.504
2	>40, M, Typical/Atypical Angina, BP High, Chol High, FBS_Normal, Inefficient, Exercise → Yes	25	8	0.026	32	0.23
3	>40, F, Asymptomatic, BP High, Chol High, FBS_Normal, Inefficient, Exercise → Yes	21	11	0.036	52.4	0.377
4	>40, M, Non-anginal Pain, BP High, Chol High, FBS_Normal, Inefficient, Exercise → Yes	18	7	0.023	38.9	0.28
5	>40, M, Asymptomatic, BP High, Chol High, FBS_Normal, Inefficient, Exercise → Yes	16	14	0.046	87.5	0.629
6	>40, M, Asymptomatic, BP High, Chol Normal, FBS_Normal, Inefficient, Exercise → Yes	14	11	0.036	78.6	0.565
7	>40, F, Typical/Atypical Angina, BP High, Chol High, FBS_Normal, Inefficient, Exercise → Yes	12	1	0.003	8.3	0.06
8	>40, F, Non-anginal Pain, BP High, Chol High, FBS_Normal, Inefficient, Exercise → Yes	11	0	0	0	0
9	>40, M, Asymptomatic, BP Normal, Chol High, FBS_Normal, Inefficient, Exercise → Yes	11	9	0.03	81.8	0.589

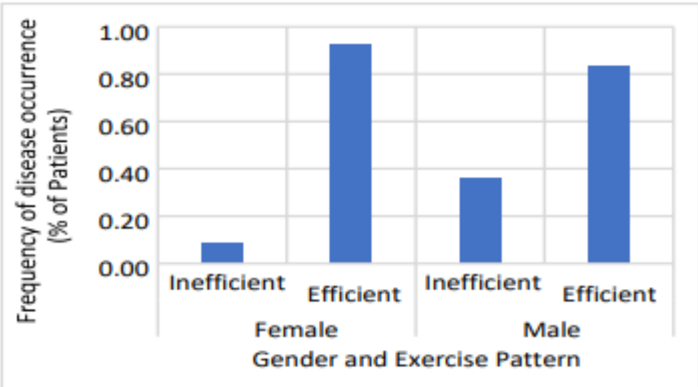
10	>40, F, Non-anginal Pain, BP High, Chol Normal, FBS_Normal, Inefficient, Exercise → Yes	8	0	0	0	0
11	>40, M, Non-anginal Pain, BP High, Chol High, FBS_High, Inefficient, Exercise → Yes	7	2	0.007	28.6	0.206
12	>40, M, Non-anginal Pain, BP High, Chol Normal, FBS_Normal, Inefficient, Exercise → Yes	7	3	0.01	42.9	0.308
13	>40, M, Asymptomatic, BP High, Chol Normal, FBS_Normal, Inefficient, Exercise → Yes	7	7	0.023	100	0.719
14	>40, M, Non-anginal Pain, BP Normal, Chol High, FBS_Normal, Inefficient, Exercise → Yes	7	2	0.007	28.6	0.206
15	>40, M, Asymptomatic, BP Normal, Chol Normal, FBS_Normal, Inefficient, Exercise → Yes	7	3	0.01	42.9	0.308
16	>40, M, Non-anginal Pain, BP High, Chol Normal, FBS_High, Inefficient, Exercise → Yes	6	1	0.003	16.7	0.12
17	>40, M, Typical/Atypical Angina, BP High, Chol Normal, FBS_Normal, Inefficient, Exercise → Yes	6	1	0.003	16.7	0.12
18	>40, M, Asymptomatic, BP High, Chol High, FBS_High, Inefficient, Exercise → Yes	6	6	0.02	100	0.719
19	>40, F, Asymptomatic, BP High, Chol Normal, FBS_Normal, Inefficient, Exercise → Yes	5	3	0.01	60	0.432
20	>40, F, Non-anginal Pain, BP Normal, Chol High, FBS_Normal, Inefficient, Exercise → Yes	5	0	0	0	0
21	>40, M, Typical/Atypical Angina, BP High, Chol High, FBS_High, Inefficient, Exercise → Yes	4	1	0.003	25	0.18
22	>40, M, Typical/Atypical Angina, BP Normal, Chol High, FBS_Normal, Inefficient, Exercise → Yes	4	2	0.007	50	0.36
23	>40, F, Asymptomatic, BP High, Chol High, FBS_High, Inefficient, Exercise → Yes	4	4	0.013	100	0.719
24	>40, F, Asymptomatic, BP Normal, Chol High, FBS_Normal, Inefficient, Exercise → Yes	4	1	0.003	25	0.18

Table 5: Extraction of Rules from Cleveland Dataset

Figures 4(a) shows the impact of gender and age over occurrence of heart diseases. For both males and females, it has been shown that the risk for heart diseases was higher for the patients older than 40 years. In Figure 4(b), the impact of gender and exercise patterns over the occurrence of heart disease has been visualized. As described previously, we have defined the exercise patterns as efficient or inefficient based on the percentage of maximum heart rate achieved by the patients during the activity. It has been observed in the figure that the patients from both genders were at a higher risk of occurrence of diseases if they were performing efficient exercise. The identified pattern is contradictory to the common belief that there is a direct association between the quality of exercise and the heart health.



(a)



(b)

Figure 4: Patterns Identified from Cleveland Dataset for impact of antecedents on the occurrence of heart disease. (a) Impact of Gender and Age, (b) Impact of Gender and Exercise Pattern

Figure 5 shows the association between gender, risk due to health status, and occurrence of heart disease. No significant difference in the heart disease risk has been found due to gender. In Figure 5(a), it has been found that the patients who were at a high risk of heart diseases due to BP suffered from lower degree of heart diseases and similar pattern has been observed in Figure 5(c). for Fasting blood sugar. Clearly, this pattern is not intuitive as it may be believed that the patients having the values of BP and FBS in high range must be having a high probability for occurrence of heart disease, however, this is not the case. In contrast, figure 5(b) shows that for the patients with values of Cholesterol falling in high-risk range suffered with heart diseases in a majority. Therefore, the physicians may need to focus more on maintaining Cholesterol as compared to the BP and blood sugar for the cardiac patients. Hence, Figure 5 has shed light on the hidden trends in the Cleveland dataset.

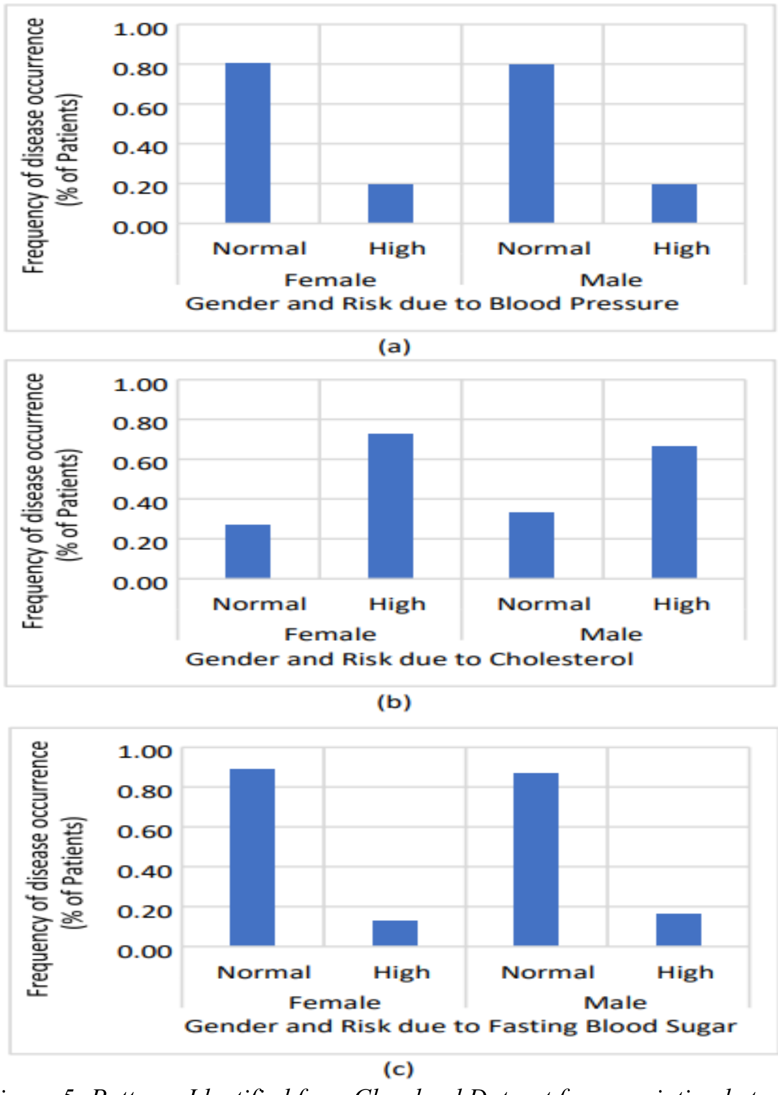


Figure 5: Patterns Identified from Cleveland Dataset for association between gender and health risks impacting the occurrence of heart disease. (a) Impact of Blood Pressure, (b) Impact of Cholesterol, (c) Impact of Fasting Blood Sugar.

6 Ethical Considerations

This work has proposed to integrate IoMT and ARMR, which would require real-time data collection from the wearable devices, and would hence raise several ethical concerns. First and foremost, the patients will be informed about the storage and usage of their data for the rule mining purpose. In this regard, the clinical facilities/hospitals shall assume the responsibility of maintaining anonymity and confidentiality of the patients' data. There will be a crucial need to implement security mechanisms such that no breaches could occur. Moreover, the compliance with the ethical regulations such as HIPAA will need to be ensured.

7 Conclusion and Future Work

This paper introduces a novel Association Rule Mining scheme called ARMR, which leverages data acquired from IoMT solutions to predict patient risks. ARMR aims to provide physicians with efficient diagnosis and preventive measures by utilizing the vast collection of population health data. The impact of health status and exercise patterns on the risk of heart disease has been thoroughly investigated using datasets from two distinct populations. Noteworthy association rules and patterns have been analysed for each dataset. The proposed methodology enables physicians to identify relevant physical parameters and mine behavioural patterns to predict risks associated with critical diseases such as diabetes, hypertension, and obesity.

Moving forward, there are several interesting directions to enhance the proposed technique. Firstly, real-world data collection from patients using wearable devices will be implemented to validate the solution. Addressing technical and confidentiality challenges will be crucial in this endeavour. Secondly, the effectiveness of ARM in predicting disease risks based on diverse datasets will be evaluated through rigorous training and validation processes. Thirdly, additional lifestyle and health-related attributes will be explored to gain insights into disease risks across diverse populations. Furthermore, the integration of other machine learning techniques, such as clustering, with IoMT will be considered for comparison with ARM's performance.

References

[Altaf, 2017] Altaf, W., Shahbaz, M., & Guergachi, A.: Applications of association rule mining in health informatics: a survey. *Artificial Intelligence Review*, *47*(3). https://doi.org/10.1007/s10462-016-9483-9, 2017.

[Amidei, 2022] Amidei, C. B., Trevisan, C., Dotto, M., Ferroni, E., Noale, M., Maggi, S., Sergi., G.: Association of physical activity trajectories with major cardiovascular diseases in elderly people. *Heart*, 360-366, 2022.

[American Heart Association, 2021] American Heart Association.: *Target heart rates chart.* https://www.heart.org/en/healthy-living/fitness/fitness-basics/target-heart-rates, 2021.

[Bays, 2022] Bays, H. E., Agarwala, A., German, C., Satish, P., Iluyomade, A., Dudum, R., & al., A. T.: Ten things to know about ten cardiovascular disease risk factors–2022. *American Journal of Preventive Cardiology*, 2022.

[Bertl, 2023] Bertl, M., Shahin, M., Ross, P., & Draheim, D.: Finding Indicator Diseases of Psychiatric Disorders in BigData using Clustered Association Rule Mining. *ACM/SIGAPP Symposium on Applied Computing*, (pp. 826-833), 2023.

[Chen, 2020] Chen, Y.: Mining of instant messaging data in the Internet of Things based on support vector machine. *Computer Communications*, *154*. https://doi.org/10.1016/j.comcom.2020.02.080, 2020.

[Colizzi, 2020] Colizzi, L., Caivano, D., Ardito, C., Desolda, G., Castrignanò, A., Matera, M., Khosla, R., Moshou, D., Hou, K. M., Pinet, F., Chanet, J. P., Hui, G., & Shi, H.: Introduction to agricultural IoT. *Agricultural Internet of Things and Decision Support for Precision Smart Farming*, 1–33. https://doi.org/10.1016/B978-0-12-818373-1.00001-9, 2020.

[Dey, 2017] Dey, N., Ashour, A. S., & Bhatt, C.: Internet of Things Driven Connected Healthcare. In *Studies in Big Data* (Vol. 23). https://doi.org/10.1007/978-3-319-49736-5_1, 2017.

[Divya, 2014] Divya, M., Manjunath, T. N., & Hegadi, R. S.: A study on developing analytical model for groundnut pest management using data mining techniques. *Proceedings - 2014 6th International Conference on Computational Intelligence and Communication Networks, CICN 2014*. https://doi.org/10.1109/CICN.2014.152, 2014.

[Fan, 2018] Fan, C., & Xiao, F.: Mining big building operational data for improving building energy efficiency: A case study. *Building Services Engineering Research and Technology*, *39*(1). https://doi.org/10.1177/0143624417704977, 2018.

[Ho, 2022] Ho, F. K., Zhou, Z., Petermann-Rocha, F., Para-Soto, S., Boonpor, J., Welsh, P., & al., J. M.: Association between device-measured physical activity and incident heart failure: a Prospective Cohort Study of 94 739 Uk Biobank participants. *Circulation*, 883-891, 2022.

[Guo, 2020] Guo, Y., Wang, N., Xu, Z. Y., & Wu, K.: The internet of things-based decision support system for information processing in intelligent manufacturing using data mining technology. *Mechanical Systems and Signal Processing*, *142*. https://doi.org/10.1016/j.ymssp.2020.106630, 2020.

[Iqbal, 2021] Iqbal, M. A., Hussain, S., Xing, H., & Imran, M. A.: Enabling the Internet of Things: fundamentals, design, and applications. In *Nucl. Phys*, 2021.

[Jagadeeswari, 2018] Jagadeeswari, V., Subramaniyaswamy, V., Logesh, R., & Vijayakumar, V.: A study on medical Internet of Things and Big Data in personalized healthcare system. *Health Information Science and Systems*, *6*(1). https://doi.org/10.1007/s13755-018-0049-x, 2018.

[Janosi, 1988] Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R.: UCI Machine Learning Repository: Heart Disease Data Set. *Uci*. https://archive.ics.uci.edu/dataset/45/heart+disease, 1988.

[Joshi, 2020] Joshi, A. M., Jain, P., Mohanty, S. P., & Agrawal, N.: IGLU 2.0: A New Wearable for Accurate Non-Invasive Continuous Serum Glucose Measurement in IoMT Framework. *IEEE Transactions on Consumer Electronics*, *66*(4). https://doi.org/10.1109/TCE.2020.3011966, 2020.

[Khedr, 2020] Khedr, A. M., Osamy, W., Salim, A., & Abbas, S.: A Novel Association Rule-Based Data Mining Approach for Internet of Things Based Wireless Sensor Networks. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3017488, 2020.

[Marques, 2021] Marques, G., Bhoi, A. Kumar., de Albuquerque, V. H. C., & K. S., Hareesha.: IoT in Healthcare and Ambient Assisted Living. In *Studies in Computational Intelligence* (Vol. 933), 2021.

[Nath, 2021] Nath, R. K., & Thapliyal, H.: Smart Wristband-Based Stress Detection Framework for Older Adults with Cortisol as Stress Biomarker. *IEEE Transactions on Consumer Electronics*, *67*(1). https://doi.org/10.1109/TCE.2021.3057806, 2021.

[Ozawa, 2020] Ozawa, S., Ban, T., Hashimoto, N., Nakazato, J., & Shimamura, J.: A study of IoT malware activities using association rule learning for darknet sensor data. *International Journal of Information Security*, *19*(1). https://doi.org/10.1007/s10207-019-00439-w, 2020.

[Patel, 2022] Patel, P., Sivaiah, B., & Patel., R.: Relevance of Frequent Pattern (FP)-Growth-Based Association Rules on Liver Diseases. *Intelligent Systems: Proceedings of ICMIB*, 2022.

[Qian, 2020] Qian, M., Yan, D., Liu, H., Berardi, U., & Liu, Y.: Power consumption and energy efficiency of VRF system based on large scale monitoring virtual sensors. *Building Simulation*, *13*(5). https://doi.org/10.1007/s12273-020-0670-x, 2020.

[Rachakonda, 2021] Rachakonda, L., Bapatla, A. K., Mohanty, S. P., & Kougianos, E.: SaYoPillow: Blockchain-Integrated Privacy-Assured IoMT Framework for Stress Management Considering Sleeping Habits. *IEEE Transactions on Consumer Electronics*, *67*(1). https://doi.org/10.1109/TCE.2020.304368321]

[Rachakonda, 2020] Rachakonda, L., Mohanty, S. P., & Kougianos, E.: ILog: An Intelligent Device for Automatic Food Intake Monitoring and Stress Detection in the IoMT. *IEEE Transactions on Consumer Electronics*, *66*(2). https://doi.org/10.1109/TCE.2020.297600, 2020.

[Rashid, 2020] Rashid, M. M., Kamruzzaman, J., Hassan, M. M., Shahriar Shafin, S., & Bhuiyan, M. Z. A.: A Survey on Behavioral Pattern Mining from Sensor Data in Internet of Things. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.2974035, 2020.

[Saha, 2023] Saha, E., & Rathore, P.: Discovering hidden patterns among medicines prescribed to patients using Association Rule Mining Technique. *International Journal of Healthcare Management*, 277-286, 2023.

[Sayeed, 2019] Sayeed, M. A., Mohanty, S. P., Kougianos, E., & Zaveri, H. P.: Neuro-Detect: A Machine Learning-Based Fast and Accurate Seizure Detection System in the IoMT. *IEEE Transactions on Consumer Electronics*, *65*(3). https://doi.org/10.1109/TCE.2019.2917895, 2019.

[Sethuraman, 2021] Sethuraman, S. C., Kompally, P., Mohanty, S. P., & Choppali, U.: MyWear: A Novel Smart Garment for Automatic Continuous Vital Monitoring. *IEEE Transactions on Consumer Electronics*, *67*(3). https://doi.org/10.1109/TCE.2021.3085888, 2021.

[Singh, 2022] Singh, A., Singh, D., Sharma, S., Upreti, K., Maheshwari, M., Mehta, V., Dabla, P. K.: Discovering patterns of cardiovascular disease and diabetes in myocardial infarction patients using association rule mining. *Folia Medica Indonesiana*, 242-250, 2022.

[Tuli, 2020] Tuli, S., Tuli, S., Wander, G., Wander, P., Gill, S. S., Dustdar, S., Sakellariou, R., & Rana, O.: Next generation technologies for smart healthcare: challenges, vision, model, trends, and future directions. In *Internet Technology Letters* (Vol. 3, Issue 2). https://doi.org/10.1002/itl2.145, 2020.

[Tuysuzoglu, 2020] Tuysuzoglu, G., & Birant, D.: Air pollution monitoring in intelligent cities using weighted association rule mining. In *Developing and Monitoring Smart Environments for Intelligent Cities*. https://doi.org/10.4018/978-1-7998-5062-5.ch007, 2020.

[Valenzuela, 2022] Valenzuela, P. L., Santos-Lozano, A., Barrán, A. T., Fernández-Navarro, P., Castillo-García, A., Ruilope, L. M., Lucia, A.: Joint association of physical activity and body mass index with cardiovascular risk: a nationwide population-based cross-sectional study. *European Journal of Preventive Cardiology*, e50-e52, 2022.

[Wang, 2017] Wang, J., & Yue, H.: Food safety pre-warning system based on data mining for a sustainable food supply chain. *Food Control*, *73*. https://doi.org/10.1016/j.foodcont.2016.09.048, 2017.

[Wedashwara, 2019] Wedashwara, W., Ahmadi, C., & Arimbawa, I. W. A.: Sequential fuzzy association rule mining algorithm for plants environment classification using internet of things. *AIP Conference Proceedings*. https://doi.org/10.1063/1.5141287, 2019.

[Wedashwara, 2019] Wedashwara, W., Ahmadi, C., Wayan Agus Arimbawa, I., & Gede Eka Wiantara Putra, I.: Internet of Things based Smart Energy Audit using Evolutionary Fuzzy Association Rule Mining. *Proceedings of 2019 4th International Conference on Sustainable Information Engineering and Technology, SIET 2019*. https://doi.org/10.1109/SIET48054.2019.8986148, 2019.

[Yassine, 2017] Yassine, A., Singh, S., & Alamri, A.: Mining Human Activity Patterns from Smart Home Big Data for Health Care Applications. *IEEE Access*, *5*. https://doi.org/10.1109/ACCESS.2017.2719921, 2017.