scientific reports

OPEN

Identifying risk factors associated with the health-related quality of life for coronary heart diseases elderly using association rule mining

Chunzi Wang[1,4], Yunwei Zhang[2,4], Shuangcheng Wang[1], Yi Wang[3], Yifei Shen[2] & Hansheng Ding[2]✉

Health-related quality of life (HRQoL) is a crucial outcome measure in the care of elderly patients with coronary heart disease (CHD). This research aims to identify risk factors associated with HRQoL in elderly patients with CHD. A cross-sectional study was carried out in Shanghai, China, from January to May 2023. Data on demographics and general symptoms and signs of CHD among elderly patients were collected by a structured questionnaire. HRQoL was measured by 21 items including both physical and psychological symptoms. The association rule mining (ARM) technique was performed to identify significant rules (support>10% and confidence >85%) for the sick and healthy conditions. In total, 179141 individuals were enrolled. 3,583 sick individuals (with CHD only, 34.4% male and 65.6% female) and 10,790 healthy individuals (free of any chronic disease, 39.5% male and 60.5% female) were included in our study. Among the significant rules for the sick condition, the most frequently occurring factors were "MedicalConstipation=1", "MotorFunction=1", "Sleep=1", "MasticatoryFunction=1" and "Gender=2". In contrast, for the healthy condition, the frequently occurring factors were "MotorFunction=0", "EducationLevel=3", "Sleep=0", "MedicalConstipation=0", and "MasticatoryFunction=0". ARM is effective in identifying the important risk factors. Impairments in medical constipation, sleep, motor function, and masticatory function are significant risk factors associated with the HRQoL in elderly patients with CHD. Early detection and management of these four symptoms could be crucial in reducing the disease burden and improving outcomes. Additionally, gender and education level may also influence the risk of developing CHD.

Keywords Coronary heart disease, Association rule mining, Health-related quality of life, Risk factor

Coronary heart disease (CHD) is a type of Cardiovascular Disease (CVD), and CVDs are the leading cause of death globally according to the report by the World Health Organization (https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).). The mortality rate of elderly people suffering from CHD is higher because they are usually accompanied by high blood pressure, high blood lipids and high cholesterol. These factors aggravate the condition of CHD. In China, approximately 11.4 million individuals were affected by CHD as of 2019[1]. With the challenge of an aging population, there is an increasing demand for CHD prevention, treatment, and the efficient allocation of medical resources. So, it is essential to gain a clearer understanding of the risk factors for CHD.

Existing literature primarily focuses on investigating clinical risk factors for heart disease[2–6]. For instance,[7] explored factors contributing to heart disease in males and females using UCI heart disease dataset. This work found that resting electrocardiographic (ECG) being either normal or hyper and slope being flat are high risk

[1]Department of Fintech, Shanghai Normal University Tianhua College, No. 1661 Shengxin North Road, Jiading District, Shanghai 201815, P.R. China. [2]Shanghai Health Development Research Center (Shanghai Medical Information Center), No. 602 West Jianguo Road, Shanghai 200031, P.R. China. [3]Dahua Hospital, No. 901 Lao Humin Road, Shanghai, Xuhui District, Shanghai 200237, P.R. China. [4]These authors contributed equally: Chunzi Wang and Yunwei Zhang. ✉email: dinghansheng@shdrc.org

factors for women only and resting ECG being hyper is a high risk factor for man only.[8] utilized eight machine learning (ML) classifiers to identify crucial clinical features that enhance the accuracy of heart disease prediction.

On the other hand, risk factors associated with health-related quality of life (HRQoL) in elderly patients with CHD remain underexplored. HRQoL, a concept that is often subjective and multifaceted, given their specific health limitations[9]. HRQoL encompasses perceptions of both physical and mental health and subjective management of their health condition[10]. CHD is a prevalent condition that significantly impacts patients' HRQoL globally. Individuals with CHD typically experience a markedly lower HRQoL compared to the healthy population due to worsening symptoms, increased disability, and frequent hospitalizations[11]. Symptoms such as shortness of breath, pain, low mood, limited activity, sleeplessness, and constipation hinder their ability to maintain social interactions and perform daily activities. A key objective in managing CHD is to help patients achieve the highest possible level of HRQoL. HRQoL has become a crucial outcome measure in the care of elderly patients with chronic conditions[12]. Poor HRQoL is frequently associated with adverse clinical outcomes such as mortality and rehospitalization. Identifying factors related to HRQoL can lead to the development of more effective strategies to enhance the quality of life and overall outcomes for patients with CHD.

In contemporary medical research, identifying risk factors for CHD and developing diagnostic or predictive models are typically achieved through multivariate statistical methods like logistic regression[13,14]. However, the vast amount of data produced by healthcare systems often contains hidden patterns that traditional approaches cannot reveal. As a result, data mining techniques are better suited for uncovering this hidden knowledge in medical studies. Association rule mining (ARM) is one of the most widely used data mining techniques for uncovering hidden relationships or patterns between variables in large datasets. In the medical field, ARM has numerous applications, including mining significant association rules from medical images[15], identifying adverse drug effects[16], detecting risk factors for heart disease and diabetes[7,17], extracting positive and negative association rules from medical blogs[18], and identifying herb combinations for treating uremic pruritus[19], among others.

In this study, we aimed to apply ARM for identifying risk factors associated with HRQoL in elderly patients with CHD, utilizing data gathered from a structured questionnaire.

Methods

Data collection

This study was designed based on a cross-sectional investigation conducted in Shanghai, China, from January to May 2023. The data was collected using a questionnaire titled the Unified Needs Assessment Form for Elderly Care. The inclusion criteria were: (1) older adults aged 60 and above; (2) applying for long-term care insurance in Shanghai and voluntarily participating in this study. Disease assessment covered 11 chronic diseases, including chronic obstructive pulmonary disease, diabetes, chronic pneumonia, lower limb fracture, Parkinson's disease, intracerebral hemorrhage, hypertension, advanced tumor, cerebral infarction, CHD, and Alzheimer's Disease.

The survey was carried out in accordance with the guidelines proposed in the Declaration of Helsinki, and the study protocol was approved by the Ethics Committee of the Shanghai Health Development Research Center (Approval Number 2024002). All participants signed the written informed consent form prior to commencing the study, and all methods were performed in accordance with relevant guidelines. To ensure the quality of collected data, a training program was given to all investigators before conducting the survey.

This survey was conducted by two uniformly trained assessors at the elderly's home, including a family doctor well-acquainted with the elderly person's circumstances. The two assessors independently evaluated the daily living and health status of the elderly individuals. Subsequently, the data underwent verification by professionals before being uploaded.

Study population and data preprocessing

A total of 179,141 older adults were enrolled. In our study, we included participants from two distinct groups: the sick population, consisting of individuals with only CHD, and the healthy population, comprising individuals free from any chronic diseases. After excluding records with missing values, we obtained 3,583 sick individuals and 10,790 healthy individuals. This imbalance in the dataset may lead to fewer rules and lower support for the minority class (i.e., the sick condition). To address this issue, we proposed randomly partitioning the healthy population into three subsets (To validate this approach, we tested 50 random seeds to randomly divide the healthy population into three subsets and repeated the entire process. The results, provided in the Supplementary materials, consistently revealed similar risk factors for CHD. Statistically, each subset is representative of the overall healthy population due to the sufficiently large sample size.), with 3,597, 3,597, and 3,596 individuals each. Each of these healthy subsets was then combined with the entire sick population to create three mixed datasets, labeled fold1, fold2, and fold3, containing 7180, 7180, and 7179 mixed individuals, respectively. Furthermore, for validation, each fold of the data was split into two parts: 90% for the training set and 10% for the validation set. The significant rules identified in the training set were then tested on the validation set to determine whether the validation data conformed to the rules. Figure 1 illustrates the process of data selection, splitting, and merging. The association rules generated from each fold were gathered to provide a comprehensive analysis.

Measurements of HRQoL

In our research, HRQoL was assessed using 21 items derived from internal medicine, surgery, and diagnosis[20]. These items provided a comprehensive evaluation of elderly individuals' HRQoL, covering dimensions such as symptoms and signs, physical function, emotions, and the risk of adverse events. Two demographic attributes, Gender and EducationLevel, were also included. Table 1 elaborated the explicit explanation of each symptom and attribute. There were three responses to the question about each symptom, with responses a, b, c being the

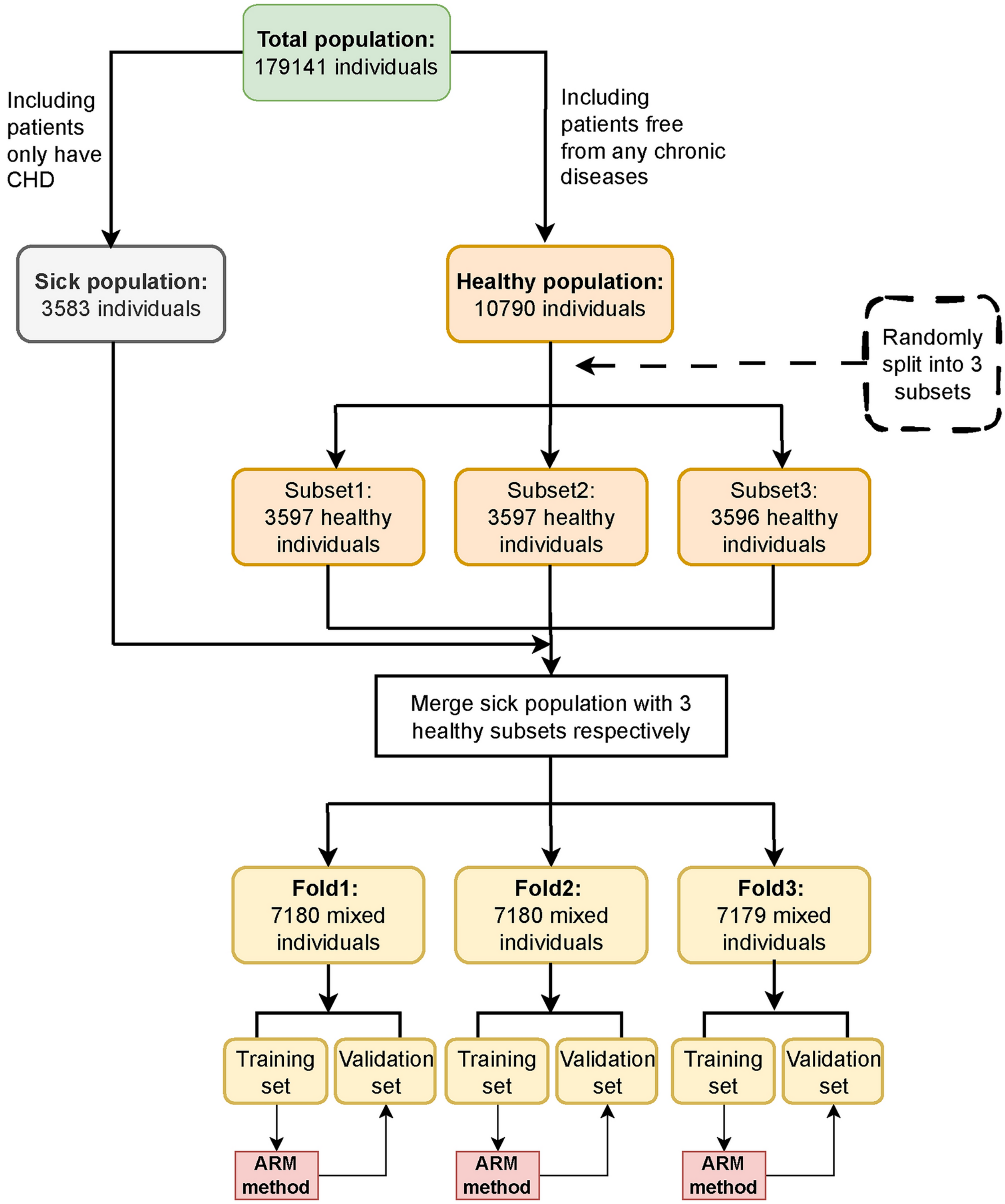

Fig. 1. Flowchart of the general framework of the study.

conditions in order of severity. For analysis, these responses were compressed into binary variables: 0 indicates the normal condition (response a), while 1 represents the abnormal conditions (responses b or c).

Association rule mining

Association rule mining (ARM) is an unsupervised data mining technique designed for discovering interesting relationships, associations, or patterns among a set of items in large datasets[21]. An association rule takes the form $X => Y$, where X and Y are distinct itemsets with no intersection. X and Y are also called left hand side (LHS) and right hand side (RHS) of the association rule, respectively. The support of a rule is calculated by the percentage of samples that contain X and Y to the total number of samples in the dataset, i.e.

$$Support(X => Y) = P(X \cap Y) = \frac{\text{number of samples containing } X \text{ and } Y}{\text{total number of samples}}. \quad (1)$$

Variable Name	Description
Gender	1 = male, 2 = female
EducationLevel	1 = illiteracy, 2 = primary school, 3 = junior high school, 4 = high school, 5 = university level or above
Consciousness	a) awake or basically awake, b) fuzzy consciousness, c) coma
HeartBeat	a) normal or basically normal, b) 40-60 beats/min or 100-120 beats/min, c) $<$ 40 beats/min or $>$ 120 beats/min
Breath	a) normal or basically normal, b) shortness of breath or effort, c) tremor, embarrassment, or sense of suffocation
BloodPressure	a) normal or basically normal, b) systolic blood pressure 141-160 mmHg or diastolic blood pressure 91-110 mmHg, c) diastolic blood pressure $<$ 60 mmHg or $>$ 110 mmHg or systolic blood pressure $>$ 160 mmHg
Nutrition	a) normal or basically normal, b) significant muscle wasting, reduced triceps skinfold thickness, or obvious ankle edema, c) severe muscle wasting, severe loss of triceps skinfold thickness, or significant ankle edema
Pain	a) no pain or basically no pain, b) intermittent or persistent pain affecting rest, requiring medication, c) persistent severe pain that can only be relieved by medication
Mood	a) stable or basically stable, b) anxiety, negativity, pessimism, sadness, misanthropy, selfishness, or suspicion, c) crying, depression, irritability, or easy anger
MedicalConstipation	a) normal or almost normal, b) occasional constipation, c) severe constipation
MuscleForce	a) normal or basically normal, b) severe limitation of limb activity, c) loss of limb activity
Sleep	a) normal or basically normal, b) difficulty sleeping or frequent awakening, c) non-restorative sleep
Dehydration	a) normal or minimal, b) severe thirst, headache, urination, or significantly reduced skin elasticity, c) significant weight loss or loss of skin elasticity
MasticatoryFunction	a) normal or basically normal, b) can chew soft food, c) cannot chew
SensoryFunction	a) normal or basically normal, b) impaired pain, touch, or temperature sensation, c) loss of pain, touch, temperature, motion, or taste sensation
CognitiveFunction	a) normal or basically normal, b) simple cognitive impairment, c) loss of cognitive function
MotorFunction	a) normal or basically normal, b) requires tools or assistance, c) loss of motor function
SocialFunction	a) normal or basically normal, b) severe limitation, c) loss of function
SwallowingFunction	a) normal or basically normal, b) severe limitation, c) loss of function
LanguageFunction	a) normal or basically normal, b) limited conversation, c) aphasia
MemoryFunction	a) normal or basically normal, b) simple memory issues or incoherence, c) loss of function
EmotionalFunction	a) normal or almost normal, b) indifferent or mildly manic, c) lack of desire, severe depression, or mania
RiskOfFall	a) no falls in the last 90 days, b) no falls in the last 30 days but falls in the last 31-90 days, c) falls in the last 30 days

Table 1. Variable names and explanations of the study dataset.

Therefore, support measures how often the rule occurs in the dataset. The confidence of a rule is an conditional probability of Y occurs given that X occurs, and is defined as the fraction of samples that contain both X and Y, relative to the total number of samples that contain X. Confidence, on the other hand, indicates the strength of association rules.

$$Confidence(X => Y) = P(Y|X) = \frac{Support(X => Y)}{Support(X)}. \tag{2}$$

For instance, if the confidence of an association rule $X => Y$ is 80%, it means that 80% of the samples containing X also contain Y[22]. Lift is a measure of rule using both support and confidence. Lift indicates the likelihood of item Y occurring when item X occurs, while accounting for the individual popularity of items Y and X. It quantifies how many times more frequently X and Y occur together than would be expected if they were statistically independent.

$$Lift(X => Y) = \frac{P(X \cap Y)}{P(X) \times P(Y)} = \frac{Support(X => Y)}{Support(X) \times Support(Y)}. \tag{3}$$

Lift = 1 indicates no association between X and Y. A lift value less than 1 suggests that Y is unlikely to occur if X occurs, while a lift value greater than 1 indicates that Y is likely to occur if X occurs.

The widely-used Apriori algorithm[23] is employed for ARM to identify significant rules. The Apriori algorithm needs a minimum support to filter frequent itemsets and minimum confidence to choose more credible rules. Thresholds for support and confidence vary across different studies. In line with[24] and[25], the minimum support was set to 10% and minimum confidence was set to 85% (An exhaustive combination of support values $\{5\%, 8\%, 10\%\}$ and confidence levels $\{85\%, 88\%, 90\%\}$ was tested. The most significant risk factors affecting HRQoL in CHD patients remained consistent across these thresholds. Therefore, we selected standard thresholds based on existing literature.). The maximum number of LHS was set to 5 items[26]. Rules that satisfied these criteria were termed significant rules. Experiments was implemented via R programming.

Results
Characteristics of the study population
The distributions of each variable were displayed in Table 2, with the last column showing the p values from the χ^2 test. The occurrence of CHD was 34.4% (1231 cases) among males and 65.6% (2352 cases) among females.

Variable	Healthy(%) ($n = 10790$)	Sick(%) ($n = 3583$)	p	Variable	Healthy(%) ($n = 10790$)	Sick(%) ($n = 3583$)	p
Gender			<0.001	Sleep			<0.001
1	4267 (39.5)	1231 (34.4)		0	9681 (89.7)	1653 (46.1)	
2	6523 (60.5)	2352 (65.6)		1	1109 (10.3)	1930 (53.9)	
EducationLevel			<0.001	Dehydration			<0.001
1	3741 (34.7)	1577 (44.0)		0	10770 (99.8)	3488 (97.3)	
2	1923 (17.8)	744 (20.8)		1	20 (0.2)	95 (2.7)	
3	2960 (27.4)	616 (17.2)		MasticatoryFunction			<0.001
4	1478 (13.7)	361 (10.1)		0	9524 (88.3)	1637 (45.7)	
5	688 (6.4)	285 (8.0)		1	1266 (11.7)	1946 (54.3)	
Consciousness			<0.001	SensoryFunction			<0.001
0	10767 (99.8)	3541 (98.8)		0	10691 (99.1)	3363 (93.9)	
1	23 (0.2)	42 (1.2)		1	99 (0.9)	220 (6.1)	
HeartBeat			<0.001	CognitiveFunction			<0.001
0	10739 (99.5)	3232 (90.2)		0	10195 (94.5)	2747 (76.7)	
1	51 (0.5)	351 (9.8)		1	595 (5.5)	836 (23.3)	
Breath			<0.001	MotorFunction			<0.001
0	10735 (99.5)	3316 (92.5)		0	8903 (82.5)	814 (22.7)	
1	55 (0.5)	267 (7.5)		1	1887 (17.5)	2769 (77.3)	
BloodPressure			<0.001	SocialFunction			<0.001
0	10296 (95.4)	2954 (82.4)		0	10252 (95.0)	2506 (69.9)	
1	494 (4.6)	629 (17.6)		1	538 (5.0)	1077 (30.1)	
Nutrition			<0.001	SwallowingFunction			<0.001
0	10619 (98.4)	3261 (91.0)		0	10695 (99.1)	3402 (94.9)	
1	171 (1.6)	322 (9.0)		1	95 (0.9)	181 (5.1)	
Pain			<0.001	LanguageFunction			<0.001
0	10159 (94.2)	2612 (72.9)		0	9937 (92.1)	2454 (68.5)	
1	631 (5.8)	971 (27.1)		1	853 (7.9)	1129 (31.5)	
Mood			<0.001	MemoryFunction			<0.001
0	10644 (98.6)	3316 (92.5)		0	10112 (93.7)	2703 (75.4)	
1	146 (1.4)	267 (7.5)		1	678 (6.3)	880 (24.6)	
MedicalConstipation			<0.001	EmotionalFunction			<0.001
0	9734 (90.2)	1628 (45.4)		0	10660 (98.8)	3351 (93.5)	
1	1056 (9.8)	1955 (54.6)		1	130 (1.2)	232 (6.5)	
MuscleForce			<0.001	RiskOfFall			<0.001
0	10341 (95.8)	2780 (77.6)		0	10457 (96.9)	3102 (86.6)	
1	449 (4.2)	803 (22.4)		1	333 (3.1)	481 (13.4)	

Table 2. Characteristics of the study dataset.

The healthy population included 39.5% males (4267 cases) and 60.5% females (6523 cases). The χ^2 test indicated that relationship between all symptoms and CHD was not independent. To further identify the most important variables, the ARM method was conducted across two categories: the sick condition and the healthy condition. The sick condition was characterized by the RHS being "CHD=1", while the healthy condition was characterized by the RHS being "CHD=0".

Association rules for the sick condition

For the sick condition, the Apriori algorithm extracted 4107,2457, and 3793 significant rules for training data of Fold1, Fold2, and Fold3, respectively. First, all significant rules, sorted by confidence, were listed in Table 3. The last three columns of the table give the support, confidence, and lift, respectively. For example, the LHS of the first rule for Fold1 was {MedicalConstipation=1, Sleep=1, MasticatoryFunction=1, RiskOfFall=0}, the RHS was {CHD=1}, with the support, confidence, and lift values of 11.19%, 90.60%, and 1.81, respectively. This rule indicated that 11.19% of individuals with malfunctions in medical constipation, sleep, and masticatory function did not have a simultaneous risk of falling. The lift value greater than 1 suggested that if an individual exhibited these symptoms, they were more likely to develop CHD, with a probability of 90.60%. The focus was solely on the abnormal symptoms for the sick condition. To that end, we summarized the frequency of occurrence of each symptom factor and demographic factor through the entire significant rules, as illustrated in Table 4. The most frequently occurring factor across the three folds was "MedicalConstipation=1". The second and third most frequent factors were "Sleep=1" and "MotorFunction=1", though their order varies across the folds. The

Folds	Rules	Support %	Confidence %	Lift
1	{MedicalConstipation=1, Sleep=1, MasticatoryFunction=1, RiskOfFall=0} => CHD=1	11.19	90.60	1.81
	{MedicalConstipation=1, Sleep=1, MasticatoryFunction=1, MotorFunction=1, RiskOfFall=0} => CHD=1	10.37	90.54	1.81
	{Breath=0, Mood=0,MedicalConstipation=1, Sleep=1, MasticatoryFunction=1} => CHD=1	10.74	90.48	1.81
	{Consciousness=0,MedicalConstipation=1, Sleep=1, MasticatoryFunction=1, RiskOfFall=0} => CHD=1	10.83	90.44	1.81
	{MedicalConstipation=1, Sleep=1,Dehydration=0, MasticatoryFunction=1,RiskOfFall=0} => CHD=1	10.54	90.44	1.81
	⋮	⋮	⋮	⋮
	{Consciousness=0,HeartBeat=0,BloodPressure=0, Sleep=1, MotorFunction=1} => CHD=1	15.78	85	1.70
2	{MedicalConstipation=1, Sleep=1, MotorFunction=1, LanguageFunction=0, EmotionalFunction=0} => CHD=1	10.34	89.19	1.78
	{MedicalConstipation=1,Sleep=1, MotorFunction=1, LanguageFunction=0} => CHD=1	10.74	88.97	1.78
	{Consciousness=0, MedicalConstipation=1, Sleep=1, MotorFunction=1, LanguageFunction=0} => CHD=1	10.69	88.93	1.78
	{MedicalConstipation=1, Sleep=1, SensoryFunction=0, MotorFunction=1, LanguageFunction=0} => CHD=1	10.17	88.90	1.77
	{MedicalConstipation=1, Sleep=1, Dehydration=0, MotorFunction=1, LanguageFunction=0} => CHD=1	10.54	88.90	1.77
	⋮	⋮	⋮	⋮
	{Consciousness=0,MasticatoryFunction=1, CognitiveFunction=0, MotorFunction=1, RiskOfFall=0, } => CHD=1	11.84	85	1.70
3	{MedicalConstipation=1, Sleep=1, MasticatoryFunction=1, MotorFunction=1, RiskOfFall=0} => CHD=1	10.37	90.17	1.80
	{MedicalConstipation=1, Sleep=1, MasticatoryFunction=1, RiskOfFall=0} => CHD=1	11.19	89.93	1.79
	{Consciousness=0, MedicalConstipation=1, Sleep=1, MasticatoryFunction=1, RiskOfFall=0} => CHD=1	10.83	89.74	1.79
	{MedicalConstipation=1, Sleep=1, MasticatoryFunction=1, EmotionalFunction=0} => CHD=1	10.83	89.51	1.79
	{MedicalConstipation=1, Sleep=1, Dehydration=0, MasticatoryFunction=1, RiskOfFall=0} => CHD=1	10.54	89.49	1.79
	⋮	⋮	⋮	⋮
	{Consciousness=0, Breath=0, Sleep=1, Dehydration=0, MotorFunction=1 } => CHD=1	19.47	85	1.70

Table 3. List of all significant rules (ordered by confidence) for the sick condition.

fourth most frequently occurring factor was "MasticatoryFunction=1", which appeared consistently in all three folds. This demonstrated that there was a substantial possibility that impairments in medical constipation, sleep, motor function, and masticatory function serve as risk factors for CHD. Furthermore, all these rules had lifts significantly greater than 1, indicating that individuals were more likely to develop CHD if they experienced malfunctions in these four areas.

Additionally, "Gender=2" was presented in all three folds, though with relatively low frequencies (11.05%, 13.15%, and 14.92%). This suggested that females in our dataset may have a higher likelihood of developing CHD compared to males. However, this could also be due to the fact that there were more females than males in the dataset.

From the experience of[26], the relation network graphs of top 20 rules, sorted by confidence, for each fold were displayed in Figure 2-4 to exhibit the rules more intuitively. Each circle represented a rule, with larger circles indicating higher confidence levels. The intensity of the red color reflected the lift, with darker shades signifying greater lift values. Factors positioned closer to the center were more frequently occurring, while those located further from the center were less common. Among the top 20 confident rules, malfunctions in Medical Constipation, Sleep, Motor Function, and Masticatory Function were most strongly associated with having CHD, as evidenced by their positions being relatively close to the center.

Finally, the values of support, confidence, and lift were calculated on the validation data set for every discovered significant rules in Table 3. The validation results, displayed in Figure 5, reveal that all association rules satisfying the thresholds of support ≥ 0.1, confidence ≥ 0.85, and lift ≥ 1.7 demonstrate robust predictive power for identifying key risk factors related to HRQoL in CHD patients. These thresholds of support and confidence align with those used during the training phase, ensuring consistency in rule evaluation. The lift values exceeding 1.7 indicate that the rules significantly outperform random association, highlighting their practical relevance.

Fold1 (4107 rules)		Fold2 (2457 rules)		Fold3 (3793 rules)	
MedicalConstipation=1	2681(65.28%)	MedicalConstipation=1	1627(66.22%)	MedicalConstipation=1	2640(69.60%)
Sleep=1	1724(41.98%)	MotorFunction=1	1322(53.81%)	Sleep=1	1530(40.34%)
MotorFunction=1	1562(38.03%)	Sleep=1	1033(42.04%)	MotorFunction=1	1414(37.28%)
MasticatoryFunction=1	1034(25.18%)	MasticatoryFunction=1	625(25.44%)	MasticatoryFunction=1	929(24.49%)
RiskOfFall=0	1033(25.15%)	Consciousness=0	585(23.81%)	Consciousness=0	915(24.12%)
Consciousness=0	1024(24.93%)	EmotionalFunction=0	565(23.00%)	EmotionalFunction=0	886(23.36%)
Dehydration=0	973(23.69%)	LanguageFunction=0	549(22.34%)	Dehydration=0	878(23.15%)
SwallowingFunction=0	931(22.67%)	Dehydration=0	531(21.61%)	SwallowingFunction=0	849(22.38%)
Mood=0	888(21.62%)	Nurition=0	497(20.23%)	Mood=0	815(21.49%)
EmotionalFunction=0	867(21.11%)	SwallowingFunction=0	495(20.15%)	Nurition=0	798(21.04%)
SensoryFunction=0	831(20.23%)	RiskOfFall=0	434(17.66%)	SensoryFunction=0	724(19.09%)
Nurition=0	771(18.77%)	SensoryFunction=0	421(17.13%)	BloodPressure=0	718(18.93%)
Breath=0	735(17.90%)	Mood=0	410(16.69%)	RiskOfFall=0	704(18.56%)
BloodPressure=0	629(15.32%)	CognitiveFunction=0	395(16.08%)	Breath=0	618(16.29%)
HeartBeat=0	464(11.30%)	Breath=0	365(14.86%)	Pain=0	576(15.19%)
Gender=2	454(11.05%)	Gender=2	323(13.15%)	Gender=2	566(14.92%)
LanguageFunction=0	450(10.96%)	MemoryFunction=0	313(12.74%)	LanguageFunction=0	487(12.84%)
Pain=0	426(10.37%)	SocialFunction=1	262(10.66%)	CognitiveFunction=0	348(9.17%)
SocialFunction=1	401(9.76%)	MuscleForce=0	260(10.58%)	SocialFunction=1	332(8.75%)
MuscleForce=0	358(8,72%)	BloodPressure=0	181(7.37%)	HeartBeat=0	325(8.57%)
MemoryFunction=0	319(7.77%)	HeartBeat=0	137(5.58%)	MemoryFunction=0	243(6.41%)
CognitiveFunction=0	318(7.74%)	Pain=0	76(3.09%)	MuscleForce=0	207(5.46%)
SocialFunction=0	43(1.05%)	SocialFunction=0	38(1.55%)	SocialFunction=0	48(1.27%)
MemoryFunction=1	7(0.17%)			MemoryFunction=1	4(0.11%)
CognitiveFunction=1	4(0.10%)			MuscleForce=1	4(0.11%)
MuscleForce=1	2(0.05%)				

Table 4. Frequency of factors appearing in rules under the sick condition.

Association rules for the healthy condition

For the healthy condition, a total of 5789, 5180, and 5454 significant rules were identified for the training data of Fold1, Fold2, and Fold3, respectively. Table 5 listed significant rules. It is evident that the healthy condition was predominately associated with normal factors as well as the EducationLevel. The occurrence frequencies of factors were presented in Table 6. The top five most frequently occurring factors were "MotorFunction=0", "Sleep=0", "EducationLevel=3", "MedicalConstipation=0", and "MasticatoryFunction=0", although their order varied across the three folds. However, "MasticatoryFunction=0" was positioned outside the center in the relation network graphs of the top 20 rules, sorted by confidence, as displayed in Figure 6-8.

Consistent with the rules for the sick condition, normal status for medical constipation, sleep, motor function, and masticatory function were strongly correlated with a lower risk of CHD. Additionally, "EducationLevel=3" frequently appeared in the significant rules in Table 6. This indicated that elderly individuals with a junior high school education were more likely to be free from CHD. According to the dataset, the most common education level was junior high school among individuals aged 60-80, while the most common education level was illiteracy among those aged > 80. Younger individuals had a lower proportion of CHD, whereas older individuals had a higher proportion. This resulted in people with a junior high school education being more likely to be free from CHD.

Finally, the validation results, presented in Figure 9, show that all association rules meeting the thresholds of support ≥ 0.1, confidence ≥ 0.85, and lift ≥ 1.7 exhibit strong predictive power in identifying key risk factors associated with HRQoL in the healthy condition.

We also conducted a sensitivity analysis by dividing the data into subgroups based on gender and age. The identified risk factors for CHD were consistent across both females and males. There were minimal variations among the age-based subgroups. Further details are provided in the Supplementary Materials.

Discussion

In this study, the data mining techique, ARM, was used to uncover hidden patterns and network associations between psychological and physical disorders affecting the HRQoL of individuals with CHD. Unlike most of the existing literature, which predominantly focuses on clinical factors such as biomarkers and medical history, our analysis shifts the emphasis to HRQoL factors. By analyzing real-world data collected through comprehensive questionnaire surveys, we aimed to identify non-clinical, HRQoL factors that influence patient outcomes, offering a more comprehensive view of how chronic conditions like CHD affect overall well-being, beyond traditional clinical measures.

Fig. 2. Visulization of top 20 rules (sorted by confidence) for the sick condition in Fold 1. Each circle in the plot corresponds to a specific rule, with its size representing the confidence level-larger circles indicate higher confidence. The color intensity, ranging from light to dark red, reflects the lift value, with darker shades indicating a higher lift. Factors positioned near the center of the plot are more commonly observed, while those further from the center are less frequently occurring.

Fig. 3. Visulization of top 20 rules (sorted by confidence) for the sick condition in Fold 2.

Fig. 4. Visulization of top 20 rules (sorted by confidence) for the sick condition in Fold 3.

The results of the ARM technique applied to both the sick and healthy conditions provide significant insights into the factors associated with CHD in the dataset. The findings revealed that specific abnormal symptoms of medical constipation, sleep, motor function, and masticatory function consistently act as strong risk factors for CHD. These abnormal symptoms were prevalent across the significant rules for the sick condition, with high confidence and lift values. For the healthy condition, the results emphasize the protective role of normal functioning in these four factors. Individuals exhibiting normal levels in these areas were consistently found to be at a lower risk of CHD. These results together indicated a clear association between malfunctions in these areas and an increased likelihood of developing CHD.

Individuals with CHD often experience fatigue, which can affect their physical activity levels and overall motor function[27]. Symptoms of CHD, such as chest pain or discomfort, can interfere with a person's ability to fall asleep or stay asleep. A prospective cohort study revealed that both short sleep duration and poor sleep quality are associated with the risk of CHD[28]. People with CHD may alter their diet to manage their condition, potentially leading to changes in bowel habits. For example, a diet low in fiber might contribute to constipation. Reduced physical activity due to fatigue or other symptoms of CHD can also impact gastrointestinal function and contribute to constipation[29]. Although the correlation between abnormal masticatory function and CHD is relatively lower than other three factors, it remains significant. Despite limited literature on this association, recent findings from the Suita study suggested that reduced maximum bite force-a key indicator of masticatory function-may be a risk factor for cardiovascular disease[30]. This highlights the importance of masticatory function as a potential but under-explored risk factor for CHD, warranting further investigation.

Based on these findings, clinicians may consider incorporating targeted interventions to address these risk factors in individuals at risk of or diagnosed with CHD. Firstly, bowel movement frequency may represent a simple quantifiable indicator of adequate colonic function[31]. Especially, elderly people often face with increased stiffness and reduced sensation in the colon and rectum, as well as long term medication[32]. Some health strategies like scientific dietary structure, intervention-specific in gut microbiota, medication management, exercise therapy (Qigong, walking, physical movement, etc) can be considered[33–35]. Secondly, a large prospective study provided a healthy sleep pattern, included early chronotype, sleep 7-8h per day, never or rarely insomnia, no snoring, and no frequent excessive daytime sleepiness[36]. Early identification and treatment of some specific disease, like obstructive sleep apnea (OSA) and central sleep apnea (CSA), are also important in the intervention for CHD related sleep disorders[37]. Thirdly, a clustering study revealed that middle-aged individuals with poor oral health and severely impaired masticatory capacity have more than twice the risk of incident CHD than those with optimal oral health and preserved masticatory capacity[38]. A similar study by[39] found that $\geqslant 1$ tooth brushing per day or $\geqslant 1$ regular dental visit for professional cleaning per year reduced cardiovascular risk by 9% and 14%, respectively.[40] conducted a program included preparatory oral exercises, mouth-opening training, tongue pressure training, prosodic training, and masticatory training, and observed significant improvements in the intervention group in terms of articulatory oral motor skill. These findings suggest that adherence to daily oral hygiene, and carry out targeted oral training interventions may be an effective way to reduce the risk of cardiovascular disease.

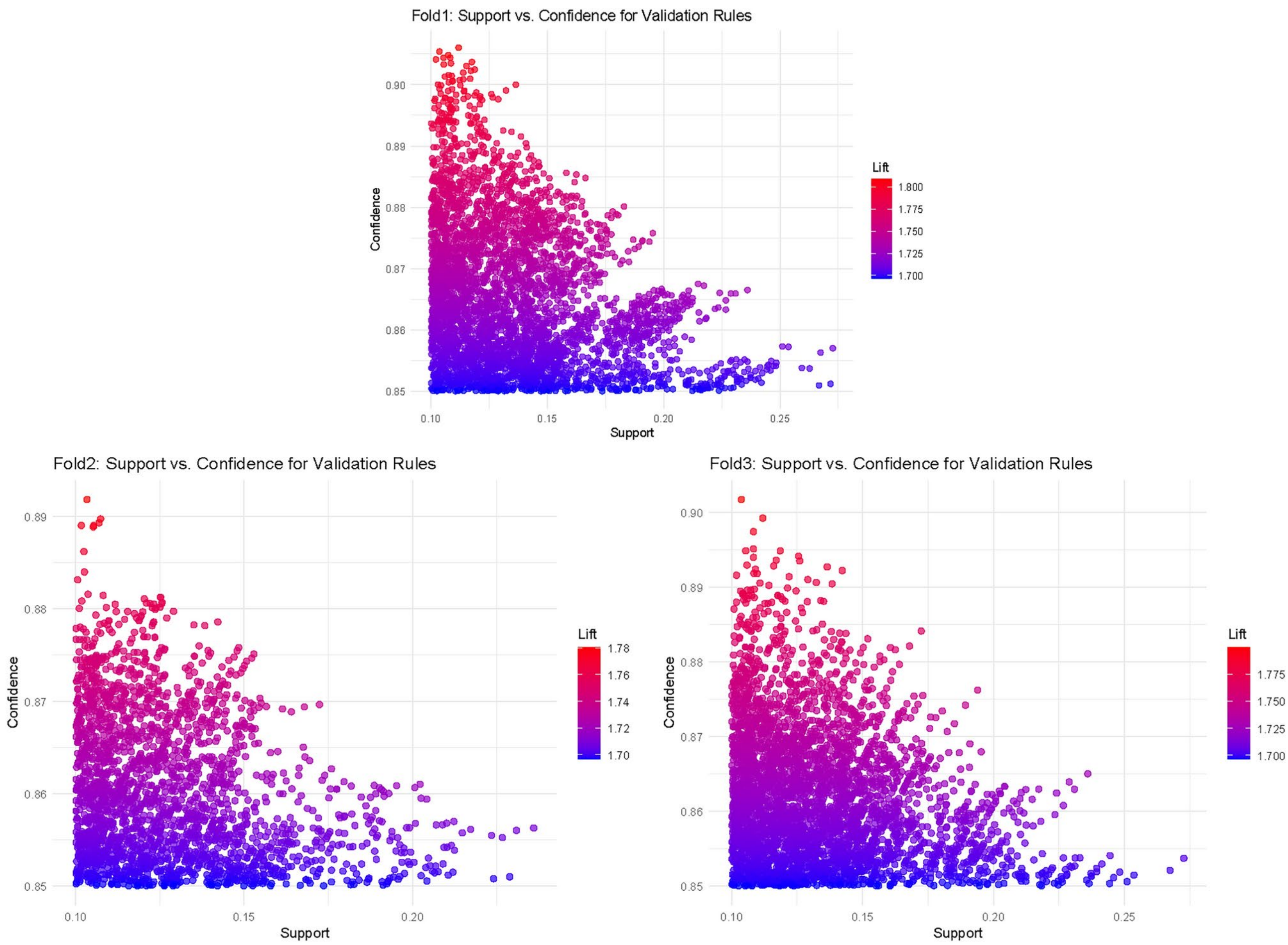

Fig. 5. The validation results for the sick condition, with the x-axis representing support and the y-axis representing confidence. The intensity of the color corresponds to the lift value.

Policy recommendations could include the development of public health programs that raise awareness about the importance of sleep, gastrointestinal health, and motor function, particularly among elderly populations. Policymakers could also consider implementing policies that support access to preventive care, such as nutrition counseling, sleep clinics, and rehabilitation services, to reduce the burden of chronic diseases like CHD.

The analysis also revealed that gender plays a role in the occurrence of CHD. Females appeared more likely to develop CHD compared to males, although this result may be skewed by the gender distribution in the dataset, where females were more prevalent. Previous studies suggested potential mechanisms for a higher risk of CHD in women.[41] found that women have more of the stress-related behavioral profile that has been linked to cardiovascular disease than men. For example, depression and posttraumatic stress disorder (PTSD). Women also performed worse in secondary prevention of CHD, such as lifestyle changes and standard medication treatment. In addition, risk factor management is less in women than men, and that sex differences varied by region[42,43].

Additionally, education level was a key demographic factor, with individuals holding a junior high school education more likely to be free from CHD. This correlation was likely influenced by age: younger individuals (60-80 years old) with this education level had lower CHD rates, while older individuals (over 80) often had lower education and higher CHD rates. This suggested that higher education may be linked to better health outcomes, potentially through increased health literacy and access to resources. A mendelian randomisation study found that longer education was additionally associated with less smoking, lower body mass index, and a favourable blood lipid profile, indicated that higher education level helps to develop a healthy lifestyle and is also beneficial for self-management. These potential effects reduce the risk of CHD occurrence[44]. Notably, this phenomenon may not be restricted to CHD; elderly individuals with a relatively higher education level may have a lower risk of various diseases.

This study has two main limitations. First, the cross-sectional observational design does not permit evaluation of how symptoms related to HRQoL change over time. Second, while combining the entire sick population with subsets of the healthy population creates more meaningful and diverse association rules for both classes, it may result in the loss of some original context or patterns present in the data before partitioning.

Fold	Rule	Support %	Confidence %	Lift
1	{EducationLevel=3, MedicalConstipation=0, Sleep=0, MasticatoryFunction=0, MotorFunction=0} => CHD=0	12.77	95.82	1.92
	{EducationLevel=3, Pain=0, MedicalConstipation=0, Sleep=0, MotorFunction=0} => CHD=0	12.74	95.48	1.91
	{EducationLevel=3, HeartBeat=0, MedicalConstipation=0, Sleep=0, MotorFunction=0} => CHD=0	12.80	95.39	1.91
	{EducationLevel=3, MedicalConstipation=0, Sleep=0, MotorFunction=0, MemoryFunction=0} => CHD=0	12.78	95.27	1.91
	{EducationLevel=3, MedicalConstipation=0, Sleep=0, MotorFunction=0, RiskOfFall=0} => CHD=0	12.75	95.26	1.91
	⋮	⋮	⋮	⋮
	{Consciousness=0, HeartBeat=0, MasticatoryFunction=0, MotorFunction=0, SocialFunction=0 } => CHD=0	40.08	85.00	1.70
2	{EducationLevel=3, MedicalConstipation=0, Sleep=0, MasticatoryFunction=0, MotorFunction=0} => CHD=0	12.04	95.58	1.92
	{EducationLevel=3, Pain=0, MedicalConstipation=0, Sleep=0, MotorFunction=0} => CHD=0	12.04	95.23	1.91
	{EducationLevel=3, HeartBeat=0, MedicalConstipation=0, Sleep=0, MotorFunction=0} => CHD=0	12.07	95.12	1.91
	{EducationLevel=3, MedicalConstipation=0, Sleep=0, MotorFunction=0, RiskOfFall=0} => CHD=0	12.07	95.01	1.90
	{EducationLevel=3, MedicalConstipation=0, Sleep=0, MotorFunction=0, MemoryFunction=0} => CHD=0	12.06	95.00	1.90
	⋮	⋮	⋮	⋮
	{Breath=0, Mood=0, MedicalConstipation=0, MotorFunction=0, SocialFunction=0} => CHD=0	39.46	85.00	1.70
3	{EducationLevel=3, MedicalConstipation=0, Sleep=0, MasticatoryFunction=0, MotorFunction=0} => CHD=0	12.23	95.64	1.92
	{EducationLevel=3, Pain=0, MedicalConstipation=0, Sleep=0, MotorFunction=0} => CHD=0	12.27	95.31	1.91
	{EducationLevel=3, HeartBeat=0, MedicalConstipation=0, Sleep=0, MotorFunction=0} => CHD=0	12.27	95.20	1.91
	{EducationLevel=3, MedicalConstipation=0, Sleep=0, MotorFunction=0, MemoryFunction=0} => CHD=0	12.27	95.08	1.91
	{EducationLevel=3, MedicalConstipation=0, Sleep=0, MotorFunction=0, RiskOfFall=0} => CHD=0	12.27	95.08	1.91
	⋮	⋮	⋮	⋮
	{EducationLevel=3, MedicalConstipation=0, SocialFunction=0, EmotionalFunction=0, RiskOfFall=0} => CHD=0	12.63	85.00	1.70

Table 5. List of significant rules (ordered by confidence) for the healthy condition.

Fold1 (5789 rules)		Fold2 (5180 rules)		Fold3 (5454 rules)	
MotorFunction=0	4334(74.87%)	MotorFunction=0	3965(76.54%)	MotorFunction=0	4150(76.09%)
EducationLevel=3	2802(48.40%)	Sleep=0	2601(50.21%)	Sleep=0	2670(48.95%)
Sleep=0	2725(47.07%)	EducationLevel=3	2446(47.22%)	EducationLevel=3	2575(47.21%)
MedicalConstipation=0	2004(34.62%)	MedicalConstipation=0	1734(33.47%)	MedicalConstipation=0	1856(34.03%)
MasticatoryFunction=0	1393(24.06%)	MasticatoryFunction=0	1295(25.00%)	MasticatoryFunction=0	1338(24.53%)
Pain=0	1064(18.38%)	Pain=0	1036(20.00%)	Pain=0	1054(19.33%)
BloodPressure=0	942(16.27%)	BloodPressure=0	890(17.18%)	BloodPressure=0	912(16.72%)
SocialFunction=0	907(15.67%)	MemoryFunction=0	853(16.47%)	MemoryFunction=0	868(15.91%)
HeartBeat=0	902 (15.58%)	RiskOfFall=0	816(15.75%)	SocialFunction=0	851(15.60%)
RiskOfFall=0	888(15.34%)	SocialFunction=0	801(15.46%)	RiskOfFall=0	843(15.46%)
MemoryFunction=0	884(15.27%)	HeartBeat=0	7791(15.27%)	HeartBeat=0	834(15.29%)
CognitiveFunction=0	858(14.82%)	CognitiveFunction=0	723(13.96%)	CognitiveFunction=0	786(14.41%)
LanguageFunction=0	837(14.46%)	Mood=0	722(13.94%)	Mood=0	770(14.12%)
MuscleForce=0	824(14.23%)	MuscleForce=0	711(13.73%)	LanguageFunction=0	755(13.84%)
Breath=0	789(13.63%)	Breath=0	710(13.71%)	MuscleForce=0	754(13.82%)
Mood=0	789(13.63%)	LanguageFunction=0	700(13.51%)	Nurition=0	742(13.60%)
Nurition=0	780(13.47%)	Nurition=0	678(13.09%)	Breath=0	741(13.59%)
SensoryFunction=0	761(13.15%)	SensoryFunction=0	638(12.32%)	EmotionalFunction=0	686(12.58%)
EmotionalFunction=0	751(12.97%)	EmotionalFunction=0	633(12.22%)	SensoryFunction=0	686(12.58%)
SwallowingFunction=0	736(12.71%)	Dehydration=0	610(11.78%)	Dehydration=0	658(12.06%)
Dehydration=0	733(12.66%)	SwallowingFunction=0	602(11.62%)	SwallowingFunction=0	651(11.94%)
Consciousness=0	723(12.49%)	Consciousness=0	591(11.41%)	Consciousness=0	639(11.72%)
Gender=1	518(8.95%)	Gender=1	487(9.40%)	Gender=1	513(9.41%)
Gender=2	167(2.88%)	Gender=2	197(3.80%)	Gender=2	213(3.91%)

Table 6. Frequency of factors appearing in rules under the healthy condition.

Fig. 6. Visulization of top 20 rules for the healthy condition in Fold 1.

Fig. 7. Visulization of top 20 rules for the healthy condition in Fold 2.

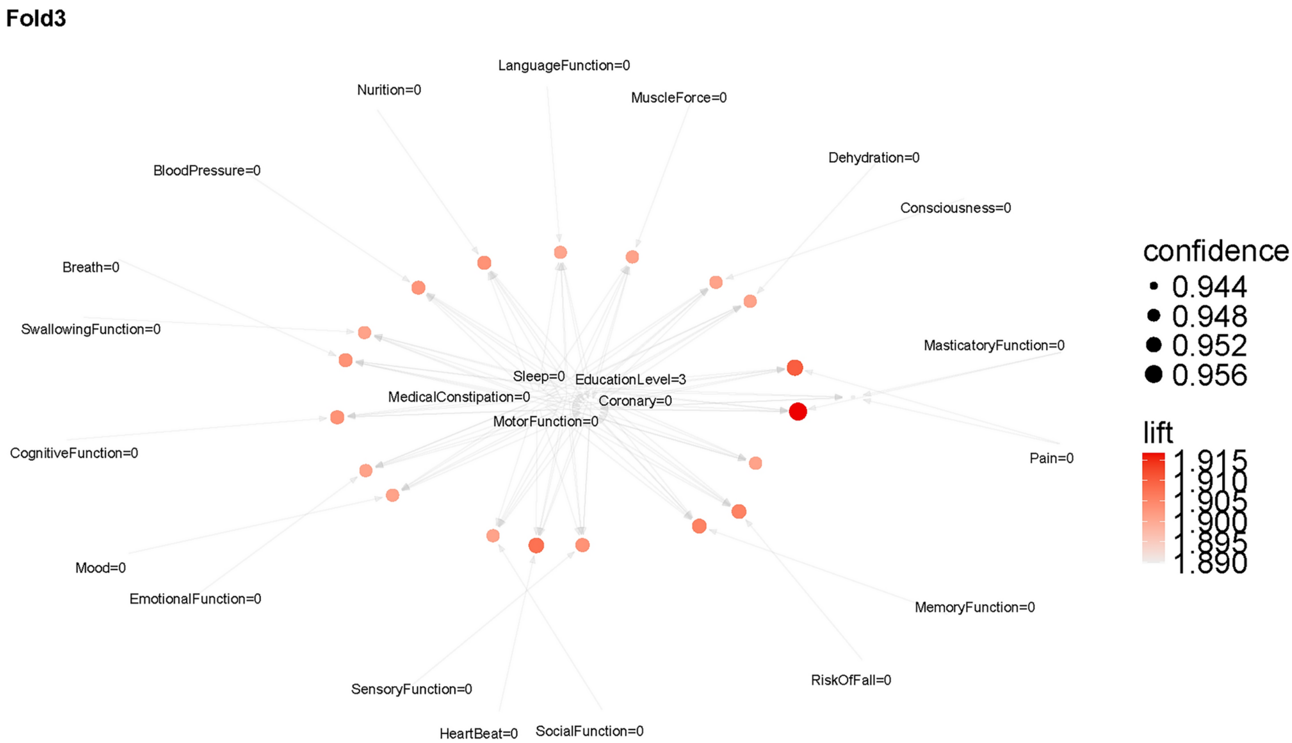

Fig. 8. Visulization of top 20 rules for the healthy condition in Fold 3.

Conclusions

Abnormal symptoms of medical constipation, sleep disturbances, motor function impairments, and masticatory function are identified as strong risk factors related to HRQoL for elderly patients with CHD. Gender and education level may also affect the likelihood of developing CHD in our dataset. Association rule analysis is effective in identifying the important HRQoL risk factors in elderly patients with CHD. The results suggest that early detection and management of these four symptoms could be crucial in reducing the disease burden and improving outcomes.

On the one hand, medical constipation, sleep issues, and motor function impairments are frequently discussed independently in existing literature. By applying ARM and validating the results across multiple subsets, we were able to uncover a clearer, more nuanced understanding of how these factors interact and contribute to the HRQoL of elderly individuals with CHD. On the other hand, the relationship between masticatory function and CHD remains less explored. Our study encourages further research into this area to better understand the potential impact of masticatory function on CHD risk, particularly in elderly populations.

Overall, this research offers a step toward refining the risk profile for CHD-related HRQoL and can guide targeted interventions aimed at improving the quality of life for this vulnerable group. Future studies could further explore the causal effect of these risk factors using panel data and investigate interventions aimed at mitigating the identified risks.

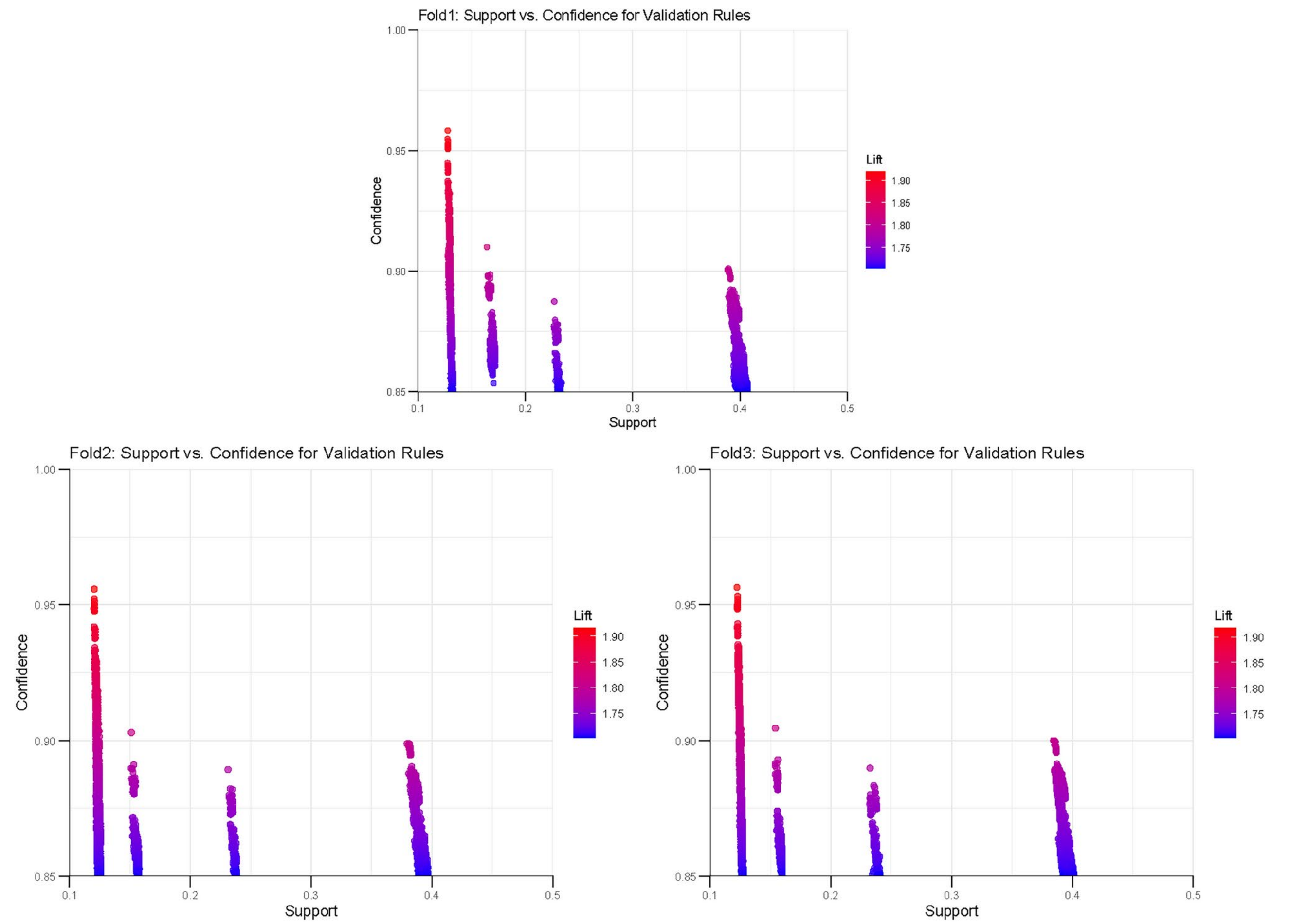

Fig. 9. The validation results for the healthy condition.

Data availability

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Received: 16 October 2024; Accepted: 18 March 2025
Published online: 03 April 2025

References

1. Sheng-Shou, H. U. et al. Report on cardiovascular health and diseases in china 2021: an updated summary. *Journal of Geriatric Cardiology* **20**(6), 399–430 (2023).
2. Ciumărnean, L. et al. Cardiovascular risk factors and physical activity for the prevention of cardiovascular diseases in the elderly. *International Journal of Environmental Research and Public Health* **19**(1), 207 (2021).
3. Shariatnia, S. et al. Modeling the diagnosis of coronary artery disease by discriminant analysis and logistic regression: a cross-sectional study. *BMC medical informatics and decision making* **22**(1), 85 (2022).
4. Ayon, S. I., Islam, M. M. & Hossain, M. R. Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. *IETE Journal of Research* **68**(4), 2488–2507 (2022).
5. Muhammad, L. J. & Algehyne, E. A. Fuzzy based expert system for diagnosis of coronary artery disease in nigeria. *Health and technology* **11**(2), 319–329 (2021).
6. Amarbayasgalan, T., Pham, V., Theera-Umpon, N., Piao, Y. & Ryu, K. H. An efficient prediction method for coronary heart disease risk based on two deep neural networks trained on well-ordered training datasets. *IEEE Access* **9**, 135210–135223 (2021).
7. Nahar, J., Imam, T., Tickle, K. S. & Chen, Y. P. Association rule mining to detect factors which contribute to heart disease in males and females. *Expert systems with applications* **40**(4), 1086–1093 (2013).
8. Srinivasan, S. et al. An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database. *Scientific Reports* **13**(1), 13588 (2023).
9. Juenger, J. et al. Health related quality of life in patients with congestive heart failure: comparison with other chronic diseases and relation to functional variables. *Heart* **87**(3), 235–241 (2002).
10. Le, J., Dorstyn, D. S., Mpfou, E., Prior, E. & Tully, P. J. Health-related quality of life in coronary heart disease: a systematic review and meta-analysis mapped against the international classification of functioning, disability and health. *Quality of life research* **27**, 2491–2503 (2018).
11. Mei, Y. et al. Health-related quality of life and its related factors in coronary heart disease patients: results from the henan rural cohort study. *Scientific reports* **11**(1), 5011 (2021).

12. Tušek-Bunc, K. & Petek, D. Comorbidities and characteristics of coronary heart disease patients: their impact on health-related quality of life. *Health and quality of life outcomes* **14**, 1–12 (2016).
13. Nusinovici, S. et al. Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of clinical epidemiology* **122**, 56–69 (2020).
14. Babiker, S. et al. Logit model in prospective coronary heart disease (CHD) risk factors prediction in saudi population. *Saudi Journal of Biological Sciences* **28**(12), 7027–7036 (2021).
15. Pan, H., Li, J. & Wei, Z. Mining interesting association rules in medical images. In *Advanced Data Mining and Applications: First International Conference. Proceedings 1*, pages 598–609. Springer, (2005).
16. Wang, C. et al. Exploration of the association rules mining technique for the signal detection of adverse drug events in spontaneous reporting systems. *PloS one* **7**(7), e40561 (2012).
17. Simon, G. Y., Schrom, J., Castro, M. R., Li, P. W. & Caraballo, P. J. Survival association rule mining towards type 2 diabetes risk assessment. In *AMIA annual symposium proceedings*, volume 2013, page 1293. American Medical Informatics Association, (2013).
18. Mahmood, S., Shahbaz, M. & Guergachi, A. Negative and positive association rules mining from text using frequent and infrequent itemsets. *The Scientific World Journal* **2014**(1), 973750 (2014).
19. Ping-Hsun, L. et al. An apriori algorithm-based association rule analysis to identify herb combinations for treating uremic pruritus using chinese herbal bath therapy. *Evidence-Based Complementary and Alternative Medicine* **2020**(1), 8854772 (2020).
20. Cuerda, M. C. et al. Development and validation of a specific questionnaire to assess health-related quality of life in patients with home enteral nutrition: Nutriqol® development. *Patient preference and adherence*, pages 2289–2296, (2016).
21. Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, (2009).
22. Kumbhare, T. A. & Chobe, S. V. An overview of association rule mining algorithms. *International Journal of Computer Science and Information Technologies* **5**(1), 927–930 (2014).
23. Altaf, W., Shahbaz, M. & Guergachi, A. Applications of association rule mining in health informatics: a survey. *Artificial Intelligence Review* **47**, 313–340 (2017).
24. Payus, C., Sulaiman, N., Shahani, M. & Bakar, A. A. Association rules of data mining application for respiratory illness by air pollution database. *Int J Basic Appl Sci* **13**(3), 11–16 (2013).
25. Sharma, N. & Om, H. Significant patterns for oral cancer detection: association rule on clinical examination and history data. *Network Modeling Analysis in Health Informatics and Bioinformatics* **3**, 1–13 (2014).
26. Huang, Y., Yaqi, S., Byun, Y., Lee, Y. & Kim, S. Analysis of multiple chronic disease characteristics in middle-aged and elderly south koreans by exercise habits based on association rules mining algorithm. *BMC Public Health* **23**(1), 1232 (2023).
27. Burkauskas, J. et al. Exploring cognitive concomitants of mental fatigue in patients with coronary artery disease. *Neuropsychobiology* **76**(3), 151–160 (2018).
28. Lao, X. Q. et al. Sleep quality, sleep duration, and the risk of coronary heart disease: a prospective cohort study with 60,586 adults. *Journal of Clinical Sleep Medicine* **14**(1), 109–117 (2018).
29. Sundbøll, J. et al. Constipation and risk of cardiovascular diseases: a danish population-based matched cohort study. *BMJ open* **10**(9), e037080 (2020).
30. Hashimoto, S. et al. A lower maximum bite force is a risk factor for developing cardiovascular disease: the suita study. *Scientific Reports* **11**(1), 7671 (2021).
31. Vermorken, A. J. M., Andrès, E. & Cui, Y. Bowel movement frequency, oxidative stress and disease prevention. *Molecular and clinical oncology* **5**(4), 339–342 (2016).
32. Deb, B., Prichard, D. O. & Bharucha, A. E. Constipation and fecal incontinence in the elderly. *Current gastroenterology reports* **22**, 1–16 (2020).
33. Pavy, B. et al. Therapeutic education in coronary heart disease: position paper from the working group of exercise rehabilitation and sport (gers) and the therapeutic education commission of the french society of cardiology. *Archives of Cardiovascular Diseases* **106**(12), 680–689 (2013).
34. Kim, Y. S., Song, B. K., Ji Sun, O. & Woo, S. S. Aerobic exercise improves gastrointestinal motility in psychiatric inpatients. *World Journal of Gastroenterology: WJG* **20**(30), 10577 (2014).
35. Lai, H. et al. Effects of dietary fibers or probiotics on functional constipation symptoms and roles of gut microbiota: a double-blinded randomized placebo trial. *Gut Microbes* **15**(1), 2197837 (2023).
36. Jaspan, V. N. et al. The role of sleep in cardiovascular disease. *Current atherosclerosis reports*, pages 1–14, (2024).
37. Mehra, R. Sleep apnea and the heart. *Cleve Clin J Med* **86**(9 Suppl 1), 10–18 (2019).
38. Deraz, O. et al. Oral condition and incident coronary heart disease: A clustering analysis. *Journal of Dental Research* **101**(5), 526–533 (2022).
39. Hopkins, S. et al. Oral health and cardiovascular disease. *The American Journal of Medicine*, (2023).
40. Shirobe, M. et al. Effect of an oral frailty measures program on community-dwelling elderly people: a cluster-randomized controlled trial. *Gerontology* **68**(4), 377–386 (2022).
41. Vaccarino, V. & Douglas Bremner, J. Behavioral, emotional and neurobiological determinants of coronary heart disease risk in women. *Neuroscience & Biobehavioral Reviews* **74**, 297–309 (2017).
42. Cífková. R. Gender differences in secondary prevention of coronary heart disease: Far from closing the gap, (2022).
43. AlBadri, A. et al. Sex differences in coronary heart disease risk factors: rename it ischaemic heart disease!, (2017).
44. Tillmann, T. et al. Education and coronary heart disease: mendelian randomisation study. *bmj*, 358, (2017).
45. Schiborn, C. et al. A newly developed and externally validated non-clinical score accurately predicts 10-year cardiovascular disease risk in the general adult population. *Scientific Reports* **11**(1), 19609 (2021).
46. Yan, J. et al. A clinical decision support system for predicting coronary artery stenosis in patients with suspected coronary heart disease. *Computers in Biology and Medicine* **151**, 106300 (2022).
47. Wang, L. et al. Profile and correlates of health-related quality of life in chinese patients with coronary heart disease. *Chinese medical journal* **128**(14), 1853–1861 (2015).
48. Yung, L. M. et al. Exercise, vascular wall and cardiovascular diseases: an update (part 2). *Sports Medicine* **39**, 45–63 (2009).

Author contributions

Chunzi Wang and Yunwei Zhang wrote the manuscript and contributed equally to this work. Shuangcheng Wang analyzed the data. Yi Wang validated the clinical application of research results. Yifei Shen edited the manuscript. Hansheng Ding designed the study and reviewed the manuscript.

Funding

This work was funded by National Natural Science Foundation of China (General Program, Study on the Development of Automatic Generation Model of Personalized Long-term Care Plan Based on Care Needs Grade and Performance Evaluation and Popularization, No. 72074152) and Xuhui health system peak discipline construction funding project (No. SHXHZDXK202307).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-95046-2.

Correspondence and requests for materials should be addressed to H.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.