Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Age-Stratified Multimorbidity Architectures at the Physiological System Level: An Integrated Network

Si Tang
Hunan University of Technology and Business
Guoxi Tan
Hunan University of Technology and Business
Yifei Ouyang
Hunan University of Technology and Business
Chenxi Wang
Hunan University of Technology and Business
Xuyuan Kuang
kxy700@126.com

Xiangya Hospital Central South University
Lin Jiang
Hunan University of Technology and Business

Article

Keywords:

Posted Date: October 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10590605/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background Multimorbidity—the coexistence of two or more chronic conditions—poses a defining challenge for primary care systems worldwide. Yet clinical guidelines remain predominantly single-disease-oriented, and systematic, age-stratified frameworks for understanding how diseases cluster across physiological systems are lacking.

Methods We conducted a retrospective cross-sectional analysis of 73,023 electronic health records from Xiangya Hospital (Changsha, China). Diagnoses were mapped to 16 physiological systems per ICD-10. The cohort was stratified into young (18–44 years), middle-aged (45–59 years), and older (≥ 60 years) groups. We deployed association rule mining (Apriori; support ≥ 0.01, confidence ≥ 0.60), Louvain community detection, and nine link prediction algorithms with 10-fold cross-validation. External validation was performed in 1,500 patients from two Jiangxi Province hospitals.

Findings

Comorbidity network density increased monotonically with age (0.089→0.153→0.178; $P < 0.001$). Network architecture shifted from four dispersed modules in young adults to a metabolic-cardiovascular hub in middle age, culminating in a tightly integrated circulatory–metabolic–neural core axis in older adults. The Adamic-Adar (AA) index demonstrated superior performance in both derivation (F1: 0.84/0.88/0.94) and external validation (F1: 0.80/0.85/0.91) cohorts, with strong topological concordance (NMI: 0.78–0.85). Of 89 high-confidence predictions, musculoskeletal→injury (AA = 5.066) in middle-aged and genitourinary→infectious disease (AA = 4.089) in older adults were top-ranked(Table 2).

Interpretation This integrated framework provides a validated, age-stratified approach to system-level multimorbidity assessment suitable for EHR-embedded clinical decision support. External validation confirms generalizability across geographic regions and hospital tiers.

Research in Context

Evidence before this study

We searched PubMed, Web of Science, and CNKI from inception to May 2026 using terms ("multimorbidity" OR "comorbidity") AND ("network analysis" OR "association rule mining" OR "link prediction") AND ("age-stratified" OR "age-dependent"). Prior studies have applied individual methods to characterize disease co-occurrence patterns, but none have integrated association rule mining, community detection, and link prediction within a unified framework. Age-stratified analyses exist in descriptive forms [8,9] but lack predictive components. No prior system-level multimorbidity network study has included independent external validation across geographic regions or hospital tiers.

Added value of this study

To our knowledge, this is the first study to: (1) systematically characterize how multimorbidity network architecture is remodeled across the adult lifespan using a unified three-stage analytical pipeline with network visualization (Figs. 1–6); (2) deploy nine link prediction algorithms with rigorous internal and external validation; (3) translate computational outputs into a structured clinical risk-stratification framework (Fig. 7); and (4) demonstrate generalizability through independent external validation across two hospitals of different tiers in a geographically distinct province.

Implications of all the available evidence

The framework enables a shift from reactive, single-disease management to proactive, system-oriented care. By identifying "bridge conditions" that connect multiple physiological systems, it offers a prioritization strategy for intervention. The robust external validation supports readiness for prospective implementation studies embedded within EHR systems.

Summary

Multimorbidity—the coexistence of two or more chronic conditions—poses a defining challenge for primary care systems worldwide. Yet clinical guidelines remain predominantly single-disease-oriented, and systematic, age-stratified frameworks for understanding how diseases cluster across physiological systems are lacking.

We conducted a retrospective cross-sectional analysis of 73,023 electronic health records from Xiangya Hospital (Changsha, China). Diagnoses were mapped to 16 physiological systems per ICD-10. The cohort was stratified into young (18–44 years), middle-aged (45–59 years), and older (≥ 60 years) groups. We deployed association rule mining (Apriori; support ≥ 0.01, confidence ≥ 0.60), Louvain community detection, and nine link prediction algorithms with 10-fold cross-validation. External validation was performed in 1,500 patients from two Jiangxi Province hospitals.

Findings

Comorbidity network density increased monotonically with age (0.089→0.153→0.178; $P < 0.001$). Network architecture shifted from four dispersed modules in young adults to a metabolic-cardiovascular hub in middle age, culminating in a tightly integrated circulatory–metabolic–neural core axis in older adults. The Adamic-Adar (AA) index demonstrated superior performance in both derivation (F1: 0.84/0.88/0.94) and external validation (F1: 0.80/0.85/0.91) cohorts, with strong topological concordance (NMI: 0.78–0.85). Of 89 high-confidence predictions, musculoskeletal→injury (AA = 5.066) in middle-aged and genitourinary→infectious disease (AA = 4.089) in older adults were top-ranked(Table 2).

This integrated framework provides a validated, age-stratified approach to system-level multimorbidity assessment suitable for EHR-embedded clinical decision support. External validation confirms generalizability across geographic regions and hospital tiers.

Research in Context

Evidence before this study

We searched PubMed, Web of Science, and CNKI from inception to May 2026 using terms ("multimorbidity" OR "comorbidity") AND ("network analysis" OR "association rule mining" OR "link prediction") AND ("age-stratified" OR "age-dependent"). Prior studies have applied individual methods to characterize disease co-occurrence patterns, but none have integrated association rule mining, community detection, and link prediction within a unified framework. Age-stratified analyses exist in descriptive forms [8,9] but lack predictive components. No prior system-level multimorbidity network study has included independent external validation across geographic regions or hospital tiers.

Added value of this study

To our knowledge, this is the first study to: (1) systematically characterize how multimorbidity network architecture is remodeled across the adult lifespan using a unified three-stage analytical pipeline with network visualization (Figs. 1–6); (2) deploy nine link prediction algorithms with rigorous internal and external validation; (3) translate computational outputs into a structured clinical risk-stratification framework (Fig. 7); and (4) demonstrate generalizability through independent external validation across two hospitals of different tiers in a geographically distinct province.

Implications of all the available evidence

The framework enables a shift from reactive, single-disease management to proactive, system-oriented care. By identifying "bridge conditions" that connect multiple physiological systems, it offers a prioritization strategy for intervention. The robust external validation supports readiness for prospective implementation studies embedded within EHR systems.

Introduction

Multimorbidity affects over one-third of the global adult population and more than two-thirds of individuals aged 65 years or older [1, 2]. In primary care settings—where the majority of chronic disease management occurs—clinicians routinely encounter patients with four or more coexisting conditions, yet must rely on single-disease clinical practice guidelines that were never designed to address the complex interplay of multiple morbidities [3, 4]. This mismatch has been identified as one of the most significant threats to healthcare sustainability [5].

A paradigm shift is underway—from viewing diseases as independent entities toward conceptualizing them as interconnected nodes within complex biological and clinical networks [6, 7]. Recent advances in computational methods have provided tools to operationalize this perspective at scale. Association rule mining has been used to quantify disease co-occurrence from large-scale clinical datasets. Bao et al. constructed cross-national disease co-occurrence networks from Chinese and UK inpatient data, revealing stable comorbidity patterns among metabolic diseases that transcend healthcare system

boundaries [8]. Zhang et al. applied similar approaches to identify distinct comorbidity clusters in type 2 diabetes populations. Community detection algorithms have been adapted to identify clinically meaningful disease modules. Most ambitiously, Dervic et al. used multilayer networks to trace life-course disease trajectories from birth to death, demonstrating that comorbidity architecture is fundamentally age-dependent [9].

Despite these advances, three critical gaps impede translation to clinical utility. First, the absence of systematic age-stratified comparisons using a unified predictive-analytical framework; no existing study integrates association mining, community detection, and link prediction within a single pipeline applied identically across age strata. Second, network analysis findings remain predominantly academic—they have not been operationalized into actionable, age-stratified guidance embeddable within EHR systems. Third, the absence of independent external validation across geographic regions and hospital tiers has left the generalizability of network-based comorbidity predictions unaddressed [10, 11].

To address these gaps, we shift the analytical focus from individual diseases to physiological systems—a resolution that balances clinical interpretability with dimensional tractability—and introduce an integrated, three-stage analytical framework. Specifically, we aim to: (1) characterize and visualize system-level comorbidity network architecture across three adult age strata; (2) identify age-specific network hubs and community structures; (3) generate testable comorbidity predictions using nine link prediction algorithms with internal and external validation; and (4) translate predictions into a clinically actionable risk-stratification framework for primary care.

Methods

Study design and data source

This retrospective cross-sectional study utilized de-identified electronic health record data from the Xiangya Hospital Clinical Data Center (Changsha, Hunan Province, China), a tertiary academic medical center serving Hunan Province (population ~ 66 million) and adjacent regions. The study was approved by the Institutional Review Board of Xiangya Hospital (approval number: 202203074). The requirement for informed consent was waived due to the retrospective, de-identified nature of the analysis. This study is reported in accordance with the STROBE guidelines [12].

Study population and age stratification

Medical records were eligible if they contained ≥ 1 chronic disease diagnosis with a corresponding ICD-10 code, complete demographic data (age, sex), and a unique patient identifier enabling deduplication across admissions. After data integrity verification and logical consistency checks, 73,023 records met the inclusion criteria. The cohort was stratified into young (18–44 years, n = 1,803), middle-aged (45–59 years, n = 3,575), and older (≥ 60 years, n = 67,645) groups(Supplementary Figure S1). Stratification cutoffs were informed by immunosenescence thresholds, epidemiological chronic disease incidence

inflection points, and clinical screening guideline boundaries(Supplementary Figure S2). Sensitivity analyses confirmed the robustness of the monotonic increase in network density to alternative cutoff specifications(Supplementary Fig. 3).

Diagnosis-to-system mapping

Individual ICD-10 diagnoses were mapped to 16 physiological systems according to the ICD-10 chapter structure (Supplementary Table S1). This system-level aggregation was chosen because it substantially reduces dimensionality (from thousands of distinct ICD-10 codes to 16 clinically coherent units), preserves clinical interpretability, and attenuates sensitivity to local coding variability, thereby enhancing generalizability.

Association rule mining

To identify statistically significant comorbidity associations between physiological systems, we applied the Apriori algorithm within each age group. Rules take the form X→Y, interpreted as "the presence of disease system X implies the co-occurrence of disease system Y." The minimum support threshold was set at 0.01 (comorbidity pair must appear in ≥ 1% of records within a stratum) and minimum confidence at 0.60 (conditional probability of Y given X ≥ 60%). For each rule, support, confidence, and lift were computed. Rules with lift > 1 were retained. Results are compiled in Supplementary Table S2.

Network construction, visualization, and community detection

For each age stratum, a weighted, undirected comorbidity network was constructed in which nodes represent the 16 physiological systems and edges represent significant association rules. Edge weight was defined as the lift value. Networks were visualized using the Fruchterman-Reingold force-directed layout algorithm, which positions highly connected nodes in proximity. Node size was scaled proportionally to system prevalence (support) within each age stratum. Node colors indicate prevalence hierarchy: the three most frequent systems are shown in soft red, ranks 4–7 in soft orange, and remaining systems in soft blue. Edge thickness is proportional to lift, with dark red edges highlighting strong associations (lift ≥ 2.0) and gray edges indicating weaker associations (lift < 2.0). Community structure was identified using the Louvain algorithm (resolution parameter = 1.0; 100 independent runs with random initialization; partition with highest modularity retained). Communities are visually distinguished by node color in Figs. 4–6.

Link prediction and internal validation

To identify yet-unobserved inter-system comorbidity associations—representing testable clinical hypotheses—we evaluated nine similarity-based link prediction algorithms: Common Neighbors (CN), Jaccard Coefficient, Adamic-Adar (AA), Resource Allocation (RA), Preferential Attachment (PA), Cosine

Similarity, Hub Promoted Index (HPI), Hub Depressed Index (HDI), and Leicht-Holme-Newman (LHN) index. For each algorithm, a similarity score was computed for every unconnected node pair. The top 20% of node pairs by similarity score were designated as predicted edges. Internal validation was performed using 10-fold cross-validation: in each fold, 10% of existing edges were randomly removed (test set), the remaining 90% were used for training, and the algorithm's ability to recover removed edges was assessed using AUC, precision, recall, F1-score, and accuracy. The AA index was selected as the primary algorithm based on consistently superior F1-scores across all age strata (Supplementary Table S3).

External validation

To assess generalizability beyond the derivation cohort, we conducted independent external validation using data from two hospitals in Jiangxi Province, a geographically distinct region: (1) Xiangya Hospital Jiangxi Branch (tertiary academic center) and (2) Jiangxi Provincial People's Hospital (provincial-level general hospital). These sites represent distinct hospital tiers, enabling assessment of generalizability across institutional levels. De-identified EHR data were extracted using identical inclusion criteria (January 2022–December 2024). A target sample of 1,500 records (500 per age stratum after stratified sampling) was prespecified for ≥ 80% power to detect F1 ≥ 0.75. The identical four-step pipeline was applied. Topological concordance between derivation and validation community partitions was assessed via normalized mutual information (NMI; ≥0.70 indicates strong concordance), adjusted Rand index (ARI), and variation of information (VI). Sensitivity analyses were stratified by hospital source (Supplementary Tables S4-S6).

Statistical analysis

All analyses were performed using Python (version 3.10) with Pandas, NetworkX, mlxtend, scikit-learn, and Matplotlib. Network metrics (density, average clustering coefficient, average path length) were computed for each age stratum and compared using permutation tests (10,000 iterations). Continuous variables are summarized as mean ± SD or median (IQR); categorical variables as frequencies and percentages.

Role of the funding source

The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Study population

The final analytical sample comprised 73,023 patients: young (n = 1,803, 2.5%), middle-aged (n = 3,575, 4.9%), and older (n = 67,645, 92.6%). The skewed distribution toward older adults reflects the

demographics of a tertiary hospital population with a high burden of chronic multimorbidity. The mean number of affected physiological systems per patient increased significantly with age: 3.2 ± 1.8 (young), 4.7 ± 2.3 (middle-aged), and 6.1 ± 2.9 (older) ($P < 0.001$ for all pairwise comparisons). Female patients comprised 47.2%, 48.9%, and 46.1% of the young, middle-aged, and older strata, respectively.

Age-dependent remodeling of comorbidity network architecture

Network visualization revealed a striking, monotonic age-dependent increase in comorbidity network complexity (Figs. 1–3). In the young cohort (Fig. 1), the network was sparse (density = 0.089), with thin, predominantly gray edges (lift < 2.0), indicating relatively weak inter-system associations. Only three edges exceeded the lift ≥ 2.0 threshold, and the endocrine–circulatory association—the dominant link in older strata—was notably absent. The network exhibited a dispersed topology characteristic of early-adulthood disease patterns, where comorbidities tend to be disease-specific rather than systemic.

In the middle-aged cohort (Fig. 2), network density increased substantially (density = 0.153). The endocrine, nutritional, and metabolic diseases system and the circulatory system emerged as a prominent dyadic hub, connected by the strongest edge in the network. Multiple dark-red (lift ≥ 2.0) edges radiated from this hub, connecting metabolic and circulatory diseases to the nervous, digestive, and genitourinary systems. This pattern reflects the initial clustering of metabolic syndrome and its cardiovascular sequelae characteristic of middle age. The emergence of this hub marks a critical transition point in the lifespan comorbidity trajectory.

In the older cohort (Fig. 3), network density peaked (density = 0.178). Nearly all node pairs were interconnected by thick edges, and the dark-red edges coalesced into a tightly integrated circulatory–metabolic–neural "core axis." The genitourinary, musculoskeletal, and digestive systems were drawn into this core, reflecting the widespread multisystem involvement that defines geriatric multimorbidity. Quantitative network metrics confirmed these visual patterns: the average clustering coefficient rose from 0.31 (young) to 0.62 (older), and the average path length decreased from 2.14 to 1.48, consistent with the emergence of a small-world topology in which any two physiological systems become reachable through a small number of intermediary steps.

Community structure: hub reconfiguration across the lifespan

Louvain community detection identified age-specific modular organization (Figs. 4–6). In young adults (Fig. 4), four dispersed, low-connectivity communities were identified. The largest community centered on the endocrine–nervous system axis, reflecting the prominence of metabolic and neurological conditions in this age group (e.g., thyroid disorders, headache syndromes). The separation of systems into distinct modules suggests that disease co-occurrence in young adulthood is driven by relatively specific pathophysiological mechanisms rather than diffuse systemic processes.

In middle-aged adults (Fig. 5), community structure consolidated into three larger modules anchored by a dominant circulatory–endocrine–digestive community. The separation of nervous system diseases into an independent module in this age group is notable—it suggests a transition point where neurological conditions in middle age may represent a pathophysiologically distinct pathway from the metabolic-cardiovascular cluster that dominates the older-age phenotype. This finding has implications for the design of age-appropriate integrated care pathways.

In older adults (Fig. 6), community boundaries dissolved into two interconnected mega-communities—one centered on the circulatory–endocrine–nervous axis (with respiratory and genitourinary integration) and the other incorporating the digestive, musculoskeletal, and neoplastic systems. The modularity score (Q) decreased monotonically from 0.36 (young) to 0.21 (older), quantitatively confirming that age-driven multimorbidity progressively erodes the modular organization of disease networks. This erosion has direct implications for healthcare system design: as patients age, single-organ-specialty care models become progressively less appropriate.

Association rule mining

A total of 56 significant association rules were identified across the three age strata (Supplementary Table S2). Lift values were generally modest (range: 1.00–1.03), reflecting broad, diffuse comorbidity patterns at the system level. However, age-specific structural differences were evident. In the young cohort, the nervous system appeared as the consequent in all top-ranked rules—suggesting that neurological symptoms may represent a common downstream manifestation of diverse conditions in young adults. In the middle-aged and older cohorts, the circulatory system assumed this role, appearing as the consequent in 75% and 92% of top rules, respectively—consistent with its emergence as the dominant comorbidity hub visible in the network visualizations (Figs. 2–3).

Link prediction performance and external validation

Comparative evaluation of nine similarity-based algorithms revealed the AA index as the top-performing method across all three age strata (Supplementary Table S3). In the derivation cohort, AA index F1-scores were 0.84 (young), 0.88 (middle-aged), and 0.94 (older). The Cosine similarity index demonstrated competitive precision but lower recall, while PA and LHN tended toward overprediction. The AA index's balanced precision–recall profile, rooted in its mathematical property of down-weighting common neighbors in favor of rare, informative neighbors, makes it particularly suited for systems-level comorbidity prediction where shared physiological pathways—not spurious correlations—should drive predictions.

Building on this internal validation, we next assessed the AA index in an independent external validation cohort from Jiangxi Province. The external cohort comprised 1,500 patients (500 per age stratum) from Xiangya Hospital Jiangxi Branch (n = 800, 53.3%) and Jiangxi Provincial People's Hospital (n = 700, 46.7%). The AA index maintained robust performance: F1-scores of 0.80 (young), 0.85 (middle-aged), and 0.91 (older), representing modest decrements from derivation (Δ: −0.04, − 0.03, − 0.03; Table 1). The monotonic age-dependent improvement in prediction accuracy observed in the derivation cohort was

fully replicated. Community structures were strongly concordant between cohorts (NMI: 0.81 young / 0.85 middle-aged / 0.78 older; Supplementary Table S5). Of 89 high-confidence predictions generated from the derivation cohort, top-ranked predictions (AA ≥ 5.0) showed the highest replication rate in the external cohort. Performance was comparable between the two validation sites (ΔF1 ≤ 0.03 across strata; Supplementary Table S6), suggesting robustness to variations in patient case-mix, coding practices, and healthcare delivery models.

A total of 89 age-specific, high-confidence comorbidity predictions were generated (AA ≥ 3.9; Supplementary Table S7). Top predictions included musculoskeletal→injury/poisoning (AA = 5.066) and nervous system→health factors (AA = 5.052) in middle-aged adults, and respiratory→health factors (AA = 4.870) and musculoskeletal→injury/poisoning (AA = 4.385) in older adults. The persistence of the musculoskeletal→injury prediction across middle-aged and older strata underscores the cross-age clinical relevance of fall risk in patients with chronic musculoskeletal conditions.

Table 1
External Validation Performance of the Adamic-Adar Index

Age Stratum	AUC	Precision	Recall	F1-Score	Accuracy
Young (18–44)	0.75	0.79	0.81	0.80	0.82
Middle-aged (45–59)	0.74	0.83	0.87	0.85	0.85
Older (≥ 60)	0.80	0.89	0.93	0.91	0.90

Table 2

Top Age-Specific Comorbidity Predictions by Adamic-Adar Index Selected from 89 high-confidence predictions (AA ≥ 3.9); see Supplementary Table S7 for complete list.

Age Group	Disease System Pair(A-B)	AA Score	Clinical Implications
Young	Health factors – Nervous system	4.742	Neurological screening in frequent healthcare utilizers
Young	Respiratory – Genitourinary	4.597	GU assessment in chronic respiratory patients
Middle-aged	Musculoskeletal – Injury/poisoning	5.066	Fall risk assessment in osteoarthritis patients
Middle-aged	Nervous – Health factors	5.052	Neurological evaluation in high-utilization patients
Middle-aged	Endocrine-metabolic – Genitourinary	4.649	Renal screening in metabolic syndrome
Elderly	Respiratory – Health factors	4.870	Comprehensive assessment in COPD patients
Elderly	Musculoskeletal – Injury/poisoning	4.385	Geriatric fall prevention program
Elderly	Genitourinary – Infectious diseases	4.089	Systematic UTI screening q6–12mo

Clinical translation: a risk-stratification framework

We translated AA-index predictions into a proof-of-concept, three-tier risk-stratification framework (Fig. 7). This framework maps top predictions to specific, age-stratified clinical actions feasible in primary care settings. High priority (AA ≥ 5.0): in middle-aged adults, all patients with musculoskeletal disorders should receive annual fall risk assessment and balance training; in young adults, frequent healthcare utilizers warrant targeted neurological symptom screening. Moderate priority (AA 4.0–4.9): in older adults, patients with genitourinary conditions should receive systematic infectious disease screening; in middle-aged adults, patients with metabolic syndrome should be assessed for eye complications (retinopathy screening). Watchful waiting (AA 3.0–3.9): pairs in this tier represent lower-urgency associations that could be incorporated into periodic comprehensive geriatric assessment. The framework is designed for potential embedding within EHR systems to provide real-time, age-stratified clinical decision support at the point of care.

Discussion

This study provides the most comprehensive characterization to date of how multimorbidity network architecture is remodeled across the adult lifespan, with four principal findings. First, comorbidity networks undergo profound, monotonic remodeling with advancing age—from sparse, modular

structures in young adulthood to densely interconnected architectures centered on a circulatory–metabolic–neural core axis (Figs. 1–6). The quantitative network metrics—density increasing from 0.089 to 0.178 and modularity decreasing from 0.36 to 0.21—provide precise, replicable benchmarks for future studies. Second, the AA index achieved robust predictive performance (F1 = 0.84–0.94) and identified 89 age-specific comorbidity predictions, with external validation confirming generalizability (F1 = 0.80–0.91). Third, we demonstrate a pathway from computational prediction to clinical action through a structured three-tier risk-stratification framework (Fig. 7). Fourth, the independent external validation across two hospitals of different tiers in a geographically distinct province establishes the framework's robustness to variations in patient populations and healthcare delivery models.

Our age-stratified findings align with and extend several lines of evidence. The shift from respiratory/infectious patterns in younger populations to metabolic/cardiovascular patterns in middle-aged and older adults—observed by Bao et al. in their analysis of 393,170 working-age inpatients [8]—is recapitulated in our system-level networks as the emergence of the endocrine–circulatory hub visible in Fig. 2. The Ningbo study's finding of age-dependent hub-disease transitions [10] is mirrored in our observation that the nervous system serves as the primary consequent in young adults while the circulatory system assumes this role in older age. Our integration of these observations within a predictive-external-validation framework—rather than retrospective description alone—represents a methodological advance. The visual progression from four distinct communities (Fig. 4) to two merged mega-communities (Fig. 6) provides an intuitive demonstration of how disease modularity erodes with age, a finding with direct implications for the design of integrated care models.

To our knowledge, this is the first system-level multimorbidity network study that includes a formal external validation step. Prior studies [8, 10, 11] have relied exclusively on internal validation (cross-validation or split-sample), limiting the evidence for generalizability. Our external validation—across two hospitals of different tiers in a different province, with strong topological concordance (NMI: 0.78–0.85)—addresses a critical evidence gap and provides a template for future multi-center network medicine studies. The comparable F1-scores between validation sites ($\Delta \leq 0.03$; Supplementary Table S6) suggest that the AA-index framework is robust to variations in institutional characteristics.

The clinical implications are fourfold. First, the system-level framework provides primary care clinicians with a more intuitive mental model for multimorbidity: a middle-aged patient with hypertension (circulatory) and diabetes (endocrine) can trigger anticipatory screening for digestive (NAFLD), nervous (neuropathy), and genitourinary (nephropathy) involvement, guided by the network topology visualized in Figs. 1–3. Second, "bridge conditions"—notably hypertension (connecting circulatory, endocrine, nervous, and genitourinary systems) and non-alcoholic fatty liver disease (connecting digestive, endocrine, and circulatory systems)—offer high-yield intervention targets with disproportionate cross-system benefits. Third, the risk-stratification framework (Fig. 7) provides a template for EHR-embedded clinical decision support that could be implemented with minimal additional infrastructure. Fourth, the erosion of modularity with age (Q: 0.36→0.21; Figs. 4–6) supports the case for geriatric comanagement and

integrated care pathways in older populations, consistent with recommendations from the JA-CHRODIS consensus [13].

Strengths of this study include the large derivation dataset (73,023 records), the three-stage integrated framework, the visual-network approach that makes complex comorbidity patterns intuitively interpretable (Figs. 1–6), rigorous internal validation, and—critically—independent external validation across two hospitals of different tiers. The comparable performance across validation sites suggests robustness to variations in patient case-mix, coding practices, and healthcare delivery models.

Several limitations warrant consideration. First, the cross-sectional design precludes determination of temporal order or causality; our use of "age-dependent" refers descriptively to cross-sectional age-group differences. Longitudinal studies with repeated measures within individuals are required to establish intra-individual comorbidity trajectories. Second, the derivation cohort's predominance of older adults (92.6%) reflects tertiary hospital referral patterns, which may limit generalizability to community-based primary care populations with different multimorbidity profiles. Third, while external validation was conducted across two Jiangxi Province hospitals, the geographic scope remains limited to central-southern China; multi-regional, multi-national validation with more diverse populations is needed. Fourth, the validation sample (N = 1,500), while adequate for the primary objective, precluded robust subgroup analyses by sex, socioeconomic status, or disease severity. Fifth, medication exposure, lifestyle factors (smoking, alcohol, physical activity), and disease severity were unavailable for adjustment. Sixth, system-level aggregation may obscure clinically important within-system heterogeneity; future multi-resolution approaches that balance system-level tractability with disease-level specificity are warranted.

Conclusion

This study provides a validated, integrated analytical framework for age-stratified, system-level multimorbidity assessment that bridges the gap between network medicine and primary care decision support. Comorbidity networks become denser and less modular with advancing age—a progression captured both quantitatively and visually in Figs. 1–6. The AA index demonstrates robust predictive performance (F1 = 0.84–0.94) in both derivation and external validation cohorts, enabling the generation of age-specific comorbidity predictions translatable into a structured clinical risk-stratification framework (Fig. 7). External validation in two Jiangxi Province hospitals confirms generalizability across geographic regions and institutional tiers. Prospective implementation studies—embedding the AA-index framework within EHR systems and evaluating its impact on clinical decision-making and patient outcomes—represent the necessary next step toward real-world deployment.

Declarations

Funding Declaration

This research was supported by Decision-Making Consultation Project of the Changsha Association for Science and Technology (Grant No. JY202620), Project of Jiangxi Provincial Administration of Traditional Chinese Medicine (Grant No. 2024B0206), Hunan Provincial Natural Science Foundation

(Grant No. 2024JJ5584), Health Research Project of Jiangxi Provincial Health Commission (Grant No. 202610220) and National College Students' innovation and Entrepreneurship Training Program (Grant No. 202510554033).

Contributors

ST, GT, YO, and CW contributed to data collection, analysis, and manuscript drafting. XK and LJ conceived the study, supervised the analysis, and critically revised the manuscript. All authors approved the final version.

Declaration of interests

The authors declare no competing interests.

Author Contribution

Tang S, Tan G, Ouyang Y, Wang C wrote the main manuscript and acquired funding. Kuang X, Jiang L edited the manuscript. All authors reviewed the manuscript.

Acknowledgement

This research was supported by Decision-Making Consultation Project of the Changsha Association for Science and Technology (Grant No. JY202620), Project of Jiangxi Provincial Administration of Traditional Chinese Medicine (Grant No. 2024B0206), Hunan Provincial Natural Science Foundation (Grant No. 2024JJ5584), Health Research Project of Jiangxi Provincial Health Commission (Grant No. 202610220) and National College Students' innovation and Entrepreneurship Training Program (Grant No. 202510554033).

References

1. Barnett K, Mercer SW, Norbury M, et al. Epidemiology of multimorbidity and implications for health care: a cross-sectional study. Lancet 2012; 380: 37–43.
2. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries, 1990–2019. Lancet 2020; 396: 1204–22.
3. Guthrie B, Payne K, Alderson P, McMurdo MET, Mercer SW. Adapting clinical guidelines to take account of multimorbidity. BMJ 2012; 345: e6341.
4. Tinetti ME, Fried TR, Boyd CM. Designing health care for the most common chronic condition—multimorbidity. JAMA 2012; 307: 2493–94.
5. Whitty CJM, MacEwen C, Goddard A, et al. Rising to the challenge of multimorbidity. BMJ 2020; 368: l6964.

6. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011; 12: 56–68.
7. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. Proc Natl Acad Sci USA 2007; 104: 8685–90.
8. Bao Y, Jia P, Li T, et al. Age-stratified comorbidity pattern transitions in 393,170 working-age inpatients. Lancet Reg Health West Pac 2023; 34: 100717.
9. Dervic E, Sorger J, Kautzky-Willer A, Klimek P. Life-course disease trajectories from birth to death. Nat Commun 2023; 14: 7347.
10. Ningbo Li Huili Hospital Research Group. Age-dependent hub disease transitions in multimorbidity networks. Chin Med J 2024; 137: 958–65.
11. Bao Y, Jia P, Li T, et al. Cross-national comparison of disease co-occurrence networks. J Am Med Inform Assoc 2022; 29: 853–63.
12. von Elm E, Altman DG, Egger M, et al. STROBE statement. Lancet 2007; 370: 1453–57.
13. Palmer K, Marengoni A, Forjaz MJ, et al. Multimorbidity care model: recommendations from JA-CHRODIS. Health Policy 2018; 122: 4–11.
14. Lü L, Zhou T. Link prediction in complex networks: a survey. Physica A 2011; 390: 1150–70.
15. Adamic LA, Adar E. Friends and neighbors on the Web. Soc Networks 2003; 25: 211–30.
16. Agrawal R, Srikant R. Fast algorithms for mining association rules. Proc VLDB 1994: 487–99.
17. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008; P10008.
18. Franceschi C, Garagnani P, Morsiani C, et al. The continuum of aging and age-related diseases. Front Med 2018; 5: 61.
19. Wang L, Chen Z, Zhang M, et al. Prevalence of multimorbidity in China. Lancet 2023; 402: 2041–53.
20. American Diabetes Association. Standards of medical care in diabetes—2024. Diabetes Care 2024; 47: S20–42.
21. Krumholz HM. Big data and new knowledge in medicine. Health Aff 2014; 33: 1163–70.

Figures

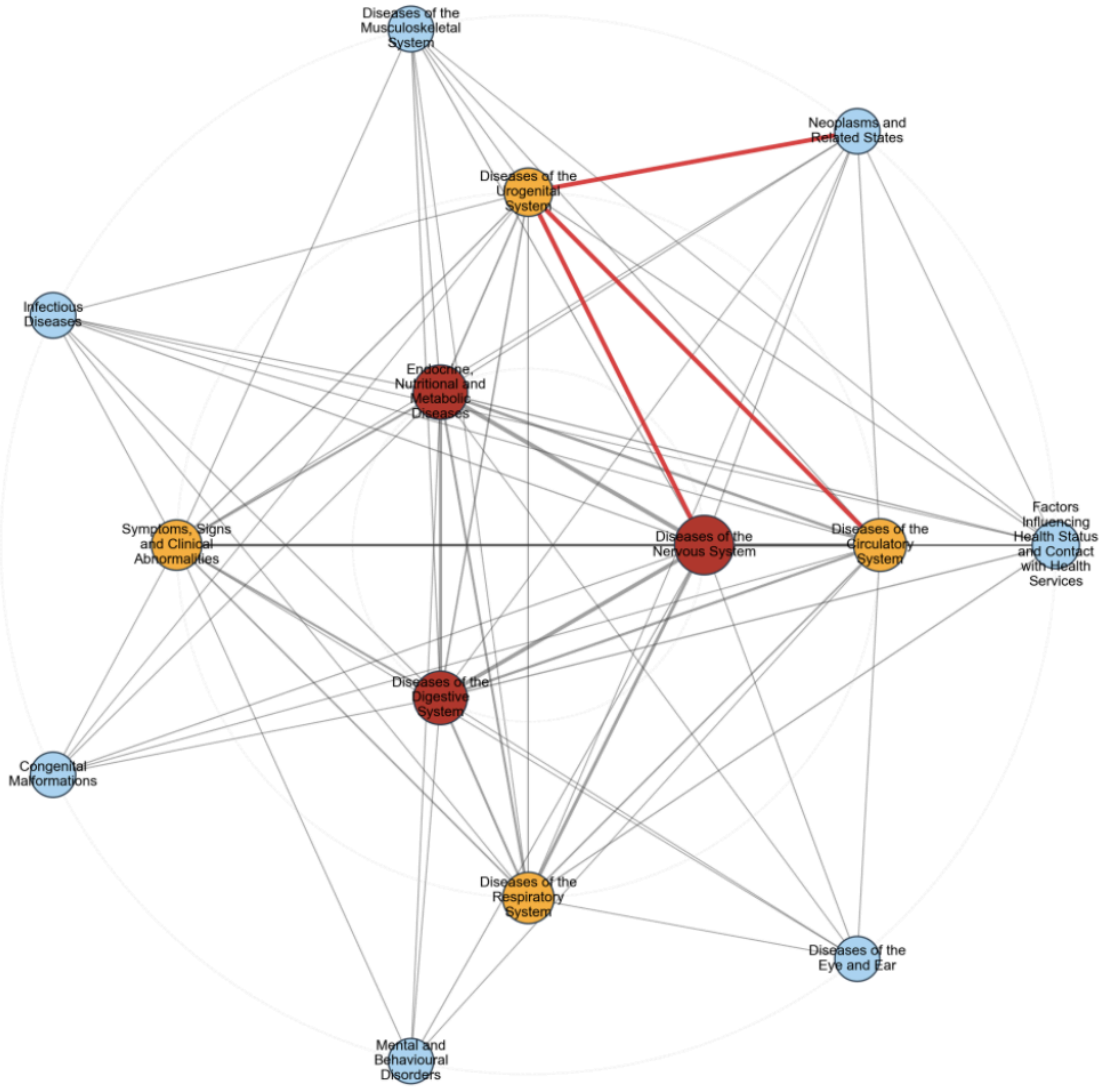


Figure 1

System-level disease association network in the young cohort (aged 18–44 years). Nodes represent 16 physiological systems; node size proportional to system prevalence; edge thickness proportional to lift value. Dark red edges: lift ≥ 2.0; gray edges: lift < 2.0.

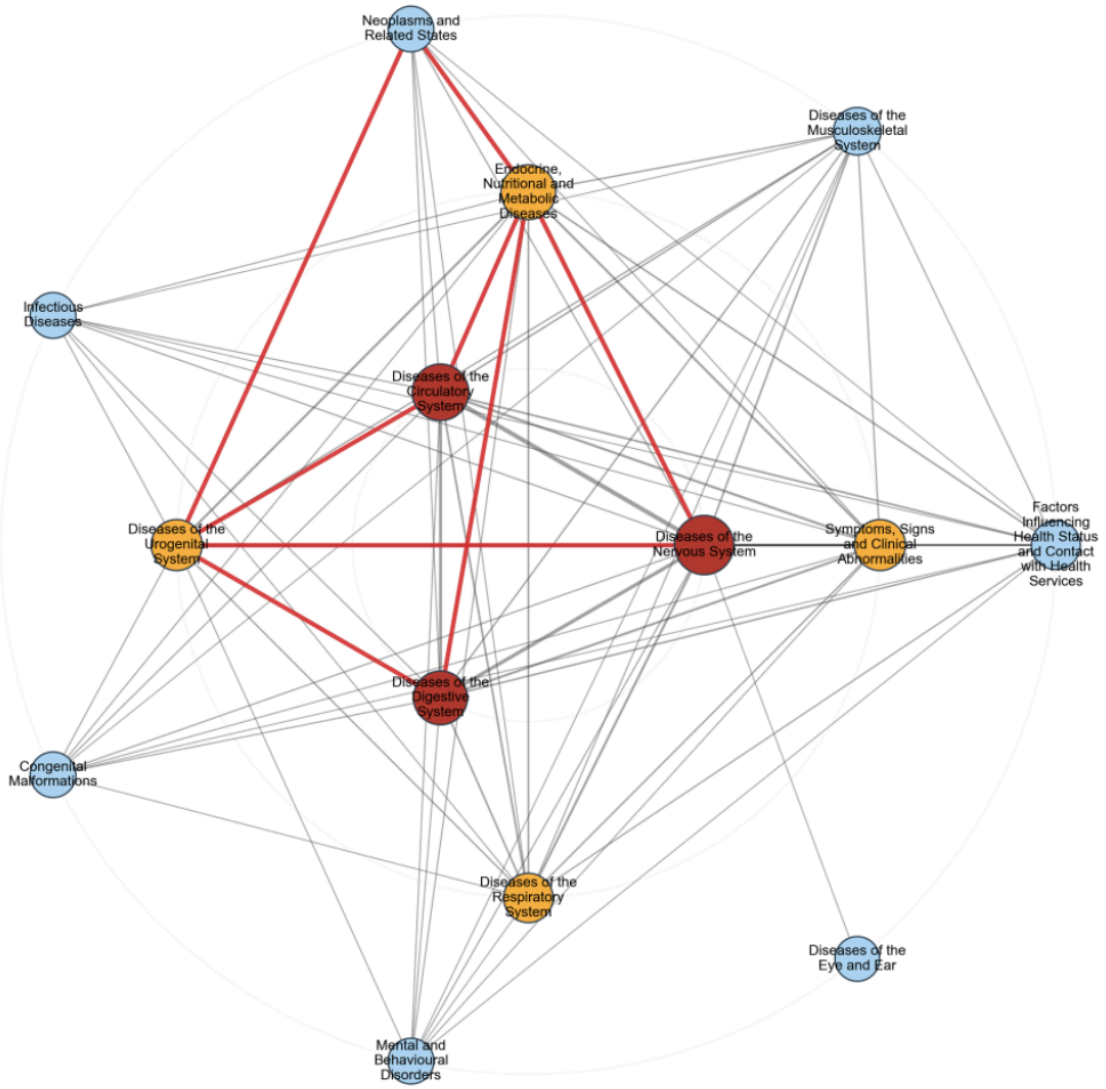


Figure 2

System-level disease association network in the middle-aged cohort (aged 45–59 years). The endocrine–circulatory hub emerges as the dominant structural feature.

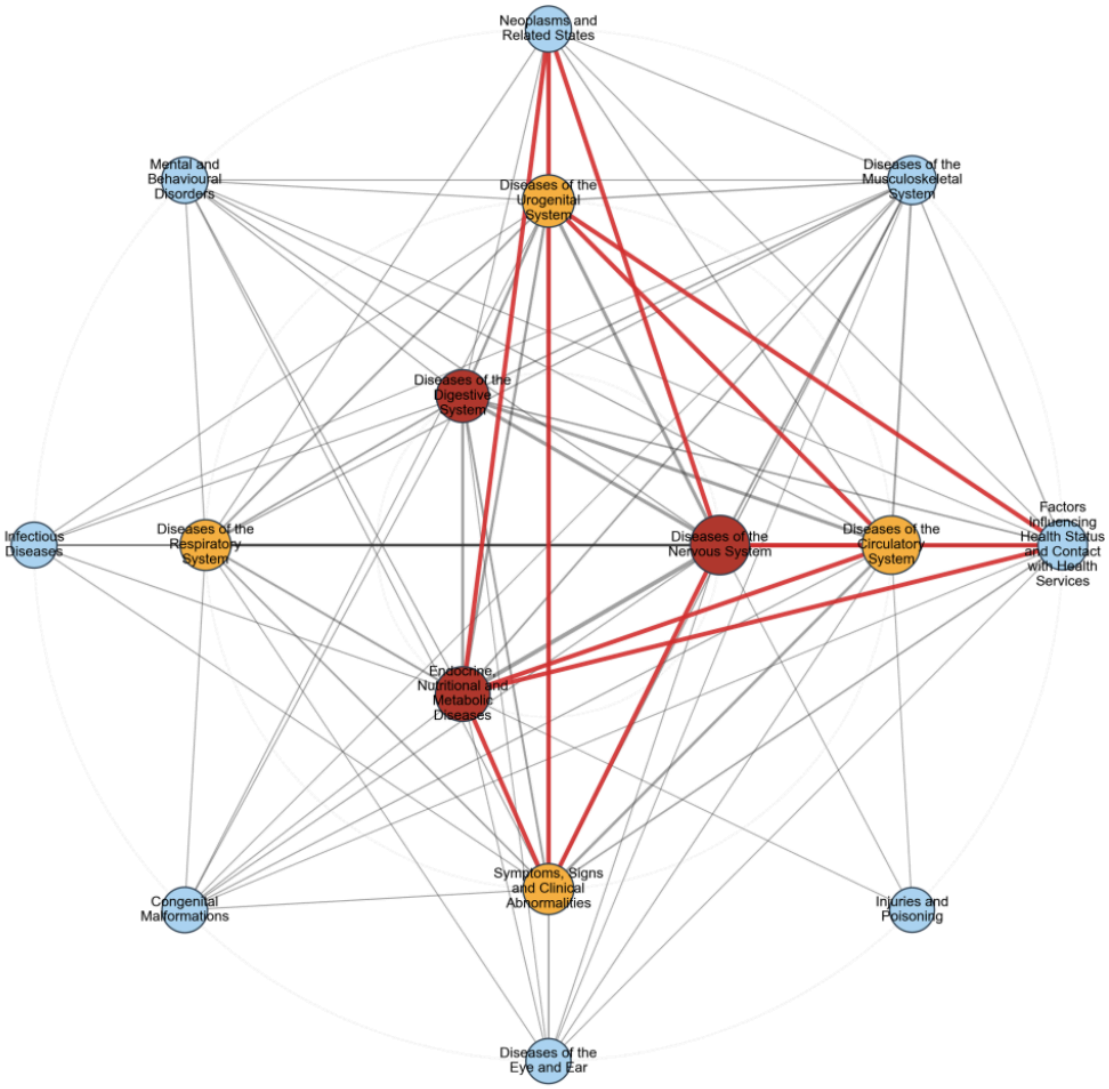


Figure 3

System-level disease association network in the elderly cohort (aged ≥60 years). A tightly integrated circulatory–metabolic–neural core axis dominates the densely interconnected network.

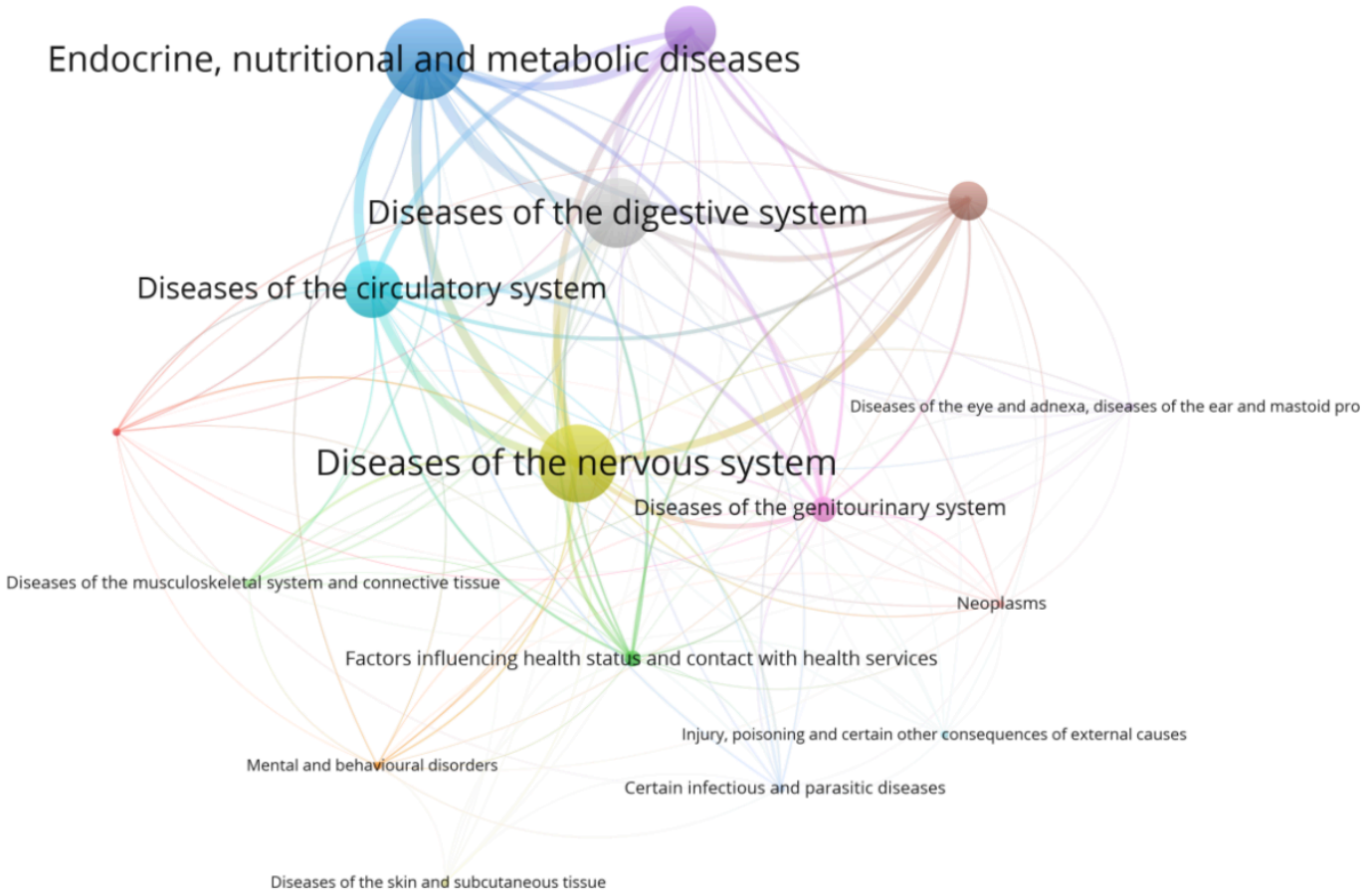


Figure 4

Community structure in the young cohort (Louvain algorithm). Four dispersed, low-connectivity communities identified, with the endocrine–nervous system axis as the largest module.

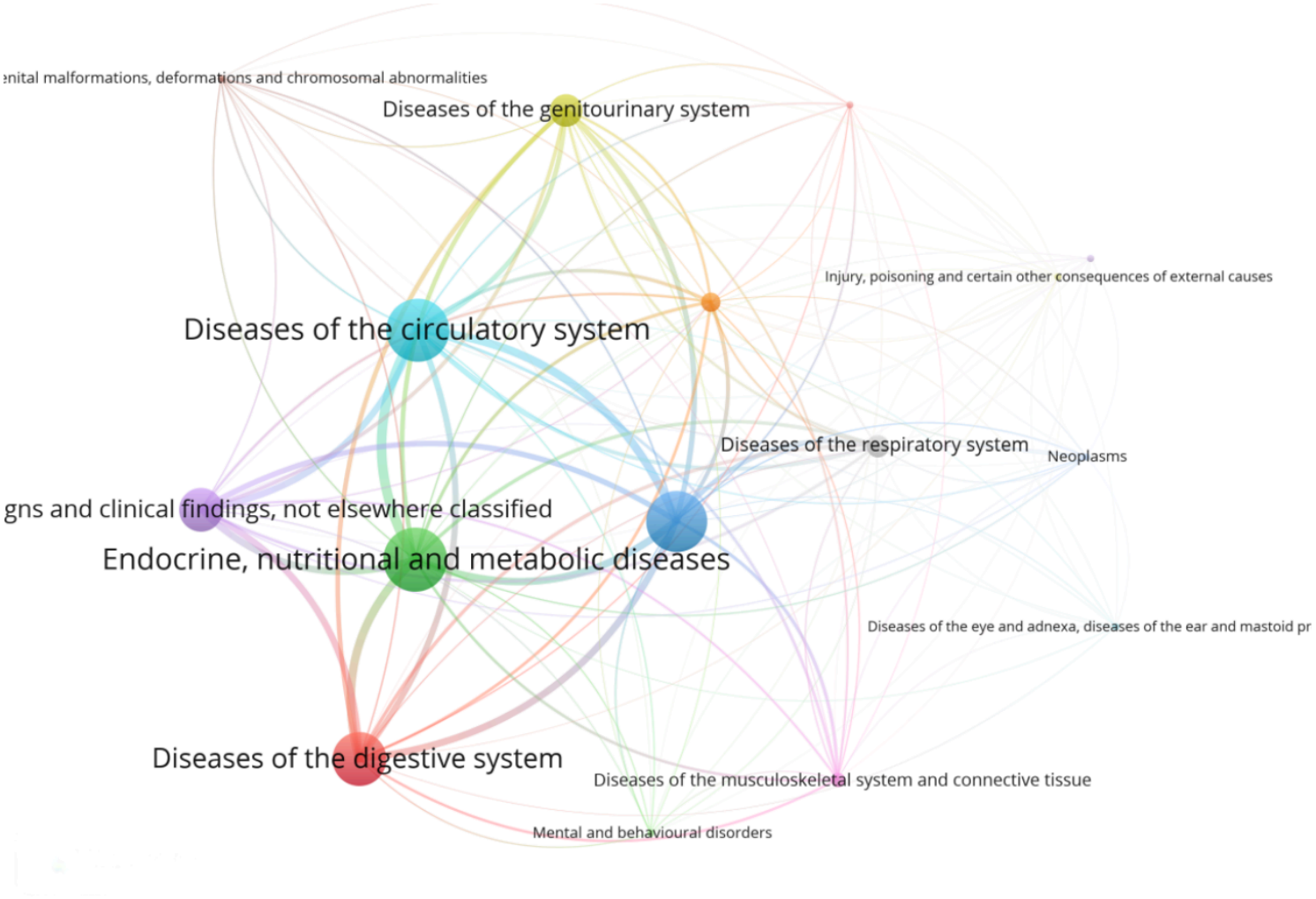


Figure 5

Community structure in the middle-aged cohort. Three larger modules anchored by a dominant circulatory–endocrine–digestive community.

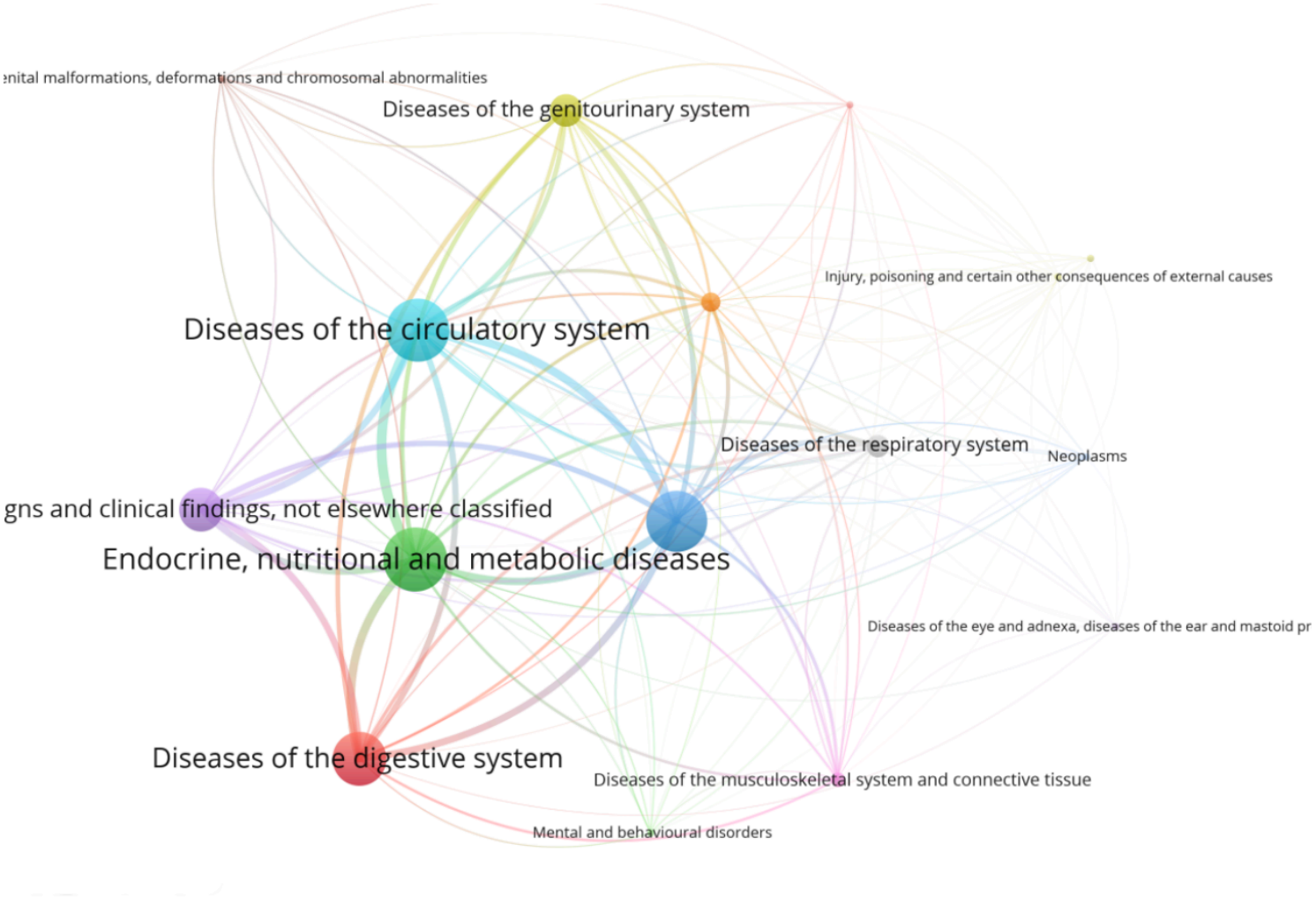


Figure 6

Community structure in the elderly cohort. Two interconnected mega-communities; nearly all systems integrated into the circulatory–endocrine–nervous core. Modularity (Q) decreases monotonically with age.

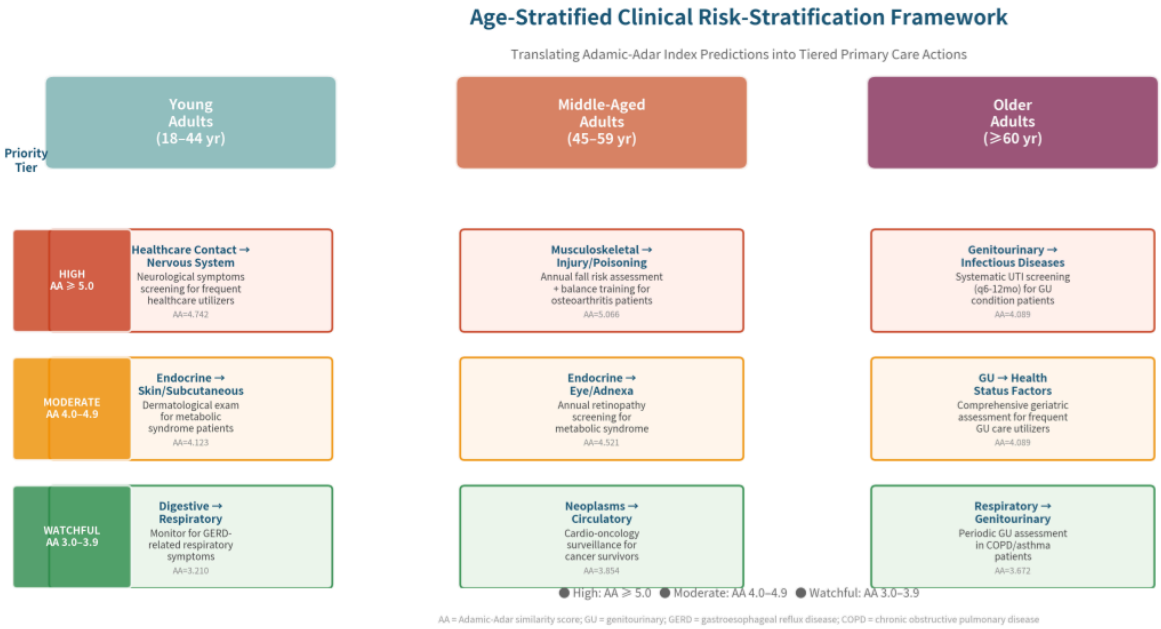


Figure 7

Clinical risk-stratification framework. AA-index predictions mapped to age-stratified, three-tier primary care actions. Priority: High (AA ≥ 5.0, red), Moderate (AA 4.0–4.9, orange), Watchful (AA 3.0–3.9, green). AA = Adamic-Adar similarity score.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials.docx