Service quality evaluation of integrated health and social care for older Chinese adults in residential settings based on factor analysis and machine learning

Zhihan Liu[*] [iD], Caini Ouyang[*], Nian Gu, Jiaheng Zhang, Xiaojiao He, Qiuping Feng and Chunguyu Chang

Abstract

Objective: To evaluate the service quality of integrated health and social care institutions for older adults in residential settings in China, addressing a critical gap in the theoretical and empirical understanding of service quality assurance in this rapidly expanding sector.

Methods: This study employs three machine learning algorithms—Backpropagation Neural Networks (BPNN), Feedforward Neural Networks (FNN), and Support Vector Machines (SVM)—to train and validate an evaluative item system. Comparative indices such as Mean Squared Error, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and predictive performance metrics were employed to assess the models.

Results: The service quality evaluation model, enhanced by factor analysis and fuzzy BPNN, demonstrated reduced error rates and improved predictive performance metrics. Key factors influencing service quality included daily care, medical attention, recreational activities, rehabilitative services, and psychological well-being, listed in order of their impact.

Conclusion: The BPNN-based model provides a comprehensive and unified framework for assessing service quality in integrated care settings. Given the pressing need to match service supply with the complex demands of older adults, refining the service delivery architecture is essential for enhancing overall service quality.

Keywords

Integrated care, service quality evaluation, machine learning, aged care, healthcare analytics

Introduction

The demographic landscape of China has undergone a transformative shift, gravitating towards a "low birth, low death, low growth" paradigm. This demographic transition has precipitated an accelerated aging of the population, thereby amplifying the exigencies for geriatric care services.[1] In response to this burgeoning demand, the "*Yiyang Jiehe*" service model was inaugurated post-2013, which means "integration of health and social care" in Chinese, is an innovative model that combines health care with aged care.[2,3] After nearly a decade of development,

four primary modes of "*Yiyang Jiehe*" have been created in China: hospital care in nursing homes, nursing care in

School of Public Administration, Central South University, Changsha, Hunan, China

[*]These authors contributed equally to this work.

Corresponding author:
Zhihan Liu, School of Public Administration, Central South University, 22 Shaoshan South Road, Changsha, Hunan 410004, China.
Email: liuzhihan@csu.edu.cn

hospitals, contract cooperation between hospitals and nursing homes, and community-based adult services (CBAS) integrated with medical care.[4] "*Yiyang Jiehe*" provides a range of services for older adults, including medical care, daily living assistance, rehabilitation, and health maintenance, as well as cultural and recreational activities and emotional support in different settings such as homes, communities, nursing homes, and hospitals.[2] In order to promote the integration of health and social care services and alleviate the supply and demand gap in the context of China's aging population, the Chinese government has conducted pivotal national strategies such as the "Healthy China 2030" and the "National Medium- and Long-Term Plan for Active Response to the Aging Population." By the end of 2020, the number of licensed institutions specializing in integrated care for older adults had increased by 59.4% compared to the end of 2017, reaching a total of 5857 institutions. Remarkably, over 90% of geriatric care facilities nationwide have the capability to offer medical and healthcare services in various modalities.[5] Despite these advancements, there exists a discernible incongruity: the empirical and theoretical frameworks for evaluating the quality of integrated care have not kept pace with the rapid proliferation of such institutions.[6,7] The development of service quality evaluation for integrated health and social care for older adults in China began relatively late. Theoretical research on the quality evaluation of integrated aged care services has lagged behind practical applications, often relying on directly adapting theories from foreign studies. Additionally, theoretical research has not kept pace with the rapid progress in practical advancements.

Key challenges in constructing a quantitative model for evaluating integrated care quality

A pivotal component in the effective evaluation of integrated care quality resides in the formulation of a robust quantitative model for the evaluation item system. A crucial aspect of creating such a model is the concept of "item weighting", which refers to the process of assigning weights to different indicators within the item. These weights determine the relative importance of each indicator in calculating the overall evaluation score. Historically, the literature has predominantly employed subjective weighting methodologies, including direct rating, consensus-based approaches like the Delphi method and expert panels,[89–10] as well as the analytic hierarchy process (AHP).[11] While these methods have contributed to the field, they are not without limitations. Specifically, the weights derived from these subjective approaches are susceptible to the idiosyncratic experiences and predilections of the experts consulted, thereby potentially overlooking the inherent patterns and attributes of the data under scrutiny.[12]

Moving away from these traditional methods, recent research has increasingly embraced objective or mixed-method approaches for evaluation. For instance, Goodson et al. employed a Bayesian network to conduct a comprehensive assessment of care quality in nursing homes.[13] Similarly, Li et al.[14] utilized a fuzzy hierarchical entropy method to evaluate the service quality of an aged care station in Beijing, China.

The promise of machine learning in addressing the complexities of evaluation

Traditional weighting and evaluation methodologies often grapple with limitations when faced with the intricate nonlinear relationships and inherent fuzziness that characterize the interplay between evaluation indicators and service quality.[15] These conventional approaches may compromise the accuracy and reliability of the evaluation outcomes. In contrast, machine learning emerges as a cutting-edge technology in objective weighting, capable of ascertaining item weights through iterative training and testing cycles.[16] This computational approach substantially mitigates the subjectivity introduced by human assessors and their inherent biases. Furthermore, the stability and intelligent operations intrinsic to machine learning algorithms augment the reliability and scientific rigor of the evaluation model.

Among the various machine learning algorithms available, neural networks (NNs), also known as artificial neural networks (ANNs), stand out as particularly salient. These algorithms emulate the operational principles of the human brain, endowed with robust memory, learning, and adaptive capabilities. The versatility of NN algorithms allows them to adeptly manage both small and large datasets, rendering them especially well-suited for tackling nonlinear problems that often confound traditional methods.[17]

The underutilization of machine learning algorithms in service quality evaluation

Despite the promising advantages offered by machine learning technologies, their application remains conspicuously underrepresented in the domain of service quality evaluation. Among the myriads of specific algorithms available, the Backpropagation Neural Network (BPNN) has garnered considerable attention. An empirical study in the realm of private goods quality attested to BPNN's superior training accuracy, demonstrating a mean squared error (MSE) of 0.120 and a predictive performance metrics of 90% when juxtaposed with other regression models.[18] Additionally, psychological research has demonstrated that BPNN outperforms Support Vector Machines (SVM) in predictive tasks involving self-reported data.[19] Nevertheless, a substantial research gap persists in the comparative analysis of different machine learning algorithms for the specific purpose of evaluating the service quality of integrated care for older adults.

Objectives and contributions of the present study

This study aims to advance this research by determining the item weights and by simulating and training the item system using three widely recognized machine learning algorithms: BPNN (Backpropagation Neural Network), FNN (fuzzy NN), and SVM (Support Vector Machine).

BPNN, FNN, and SVM represent the most commonly employed machine learning methods in service quality evaluation.[18] BPNNs are extensively utilized for predicting outcomes in scenarios where complex relationships between variables exist, making them particularly suitable for evaluating healthcare service quality, where numerous interrelated factors impact outcomes.[20] BPNNs are adept at modeling complex nonlinear relationships between various health indicators and patient outcomes, enhancing their utility in healthcare service quality evaluations.[21]

FNNs incorporate fuzzy logic principles to address the imprecision and uncertainty often found in qualitative data, such as patient satisfaction and perceptions of care quality. This capability makes FNNs exceptionally valuable for datasets involving subjective human assessments.[22]

SVMs excel in classification tasks with ambiguous boundaries, a common challenge in service quality assessments where service level categorizations can be highly nuanced.[23]

The final evaluation model is established through a comparative analysis of predictive performance metrics and error metrics. This study not only continues the exploration of machine learning applications in service quality but also engages in a detailed discussion of the evaluation outcomes and empirical evidence gathered. Through this, it contributes significantly to both the theoretical frameworks and practical implementations of integrated care quality assessment.

Materials and methods

Materials

In a precursor study, we developed an index system based on the SERVPERF conceptual framework to evaluate the service quality of integrated health and social care for older adults in Chinese residential settings.[24] The index system was designed through extensive literature research and expert consultations to ensure comprehensive coverage of the main characteristics relevant to service quality evaluation in such settings.[25] Experts from related fields were consulted to validate the content and ensure that the measurement items accurately represented the quality dimensions of integrated health and social care services. Following the preliminary design, small-scale interviews were conducted to refine the index system; one item was removed, and the survey methodology was improved to enhance the validity of the system.[26] This iterative process resulted in a final questionnaire, which demonstrated strong psychometric properties, as detailed in Supplemental Appendix 1.

The Cronbach's α coefficient for the overall index system is 0.925, indicating high reliability, while sub-dimension reliability coefficients range from 0.767 to 0.893, reflecting acceptable internal consistency reliability.[24] The validity of the index system was confirmed through Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). For more detailed results on the validity, please refer to Supplemental Tables 1 and 2 and Figure 1.

Building on this foundation, the current study aims to apply advanced machine learning algorithms to further analyze the evaluation index system. While the baseline model was initially established using Structural Equation Modeling (SEM) in our prior work, this study extends that analysis by leveraging machine learning techniques to model the data and determine the weights of different items within the index system. This approach allows us to move beyond traditional linear models and explore nonlinear relationships that may offer more nuanced insights into the factors influencing service quality.

To avoid redundancy and focus on the novel contributions of this study, we provide a summary of the index system's development and refer readers to our earlier work for comprehensive details on the methodologies and results. The data collection for the current analysis took place from 14 January to 10 April 2021, following the workflow shown in Supplemental Figure 2. Using random sampling, we selected 17 integrated health and social care institutions across cities in Hunan Province, including Changsha, Xiangtan, Zhuzhou, Hengyang, and Shaoyang. Cluster sampling was conducted within each institution based on the following inclusion criteria: (1) residence in the institution for over 3 months; (2) age over 60, or 55 if unable to live independently; (3) mental capacity and ability to communicate; (4) understanding of the survey's purpose and willingness to participate. A total of 336 questionnaires were distributed, with 329 retrieved, yielding a response rate of 97.92%. After data cleaning, 301 valid questionnaires were retained, resulting in a valid response rate of 89.58%.

In this study, the service quality was evaluated using a Likert scale ranging from 1 (very dissatisfied) to 5 (very satisfied). Although Likert scales are inherently ordinal, this study treats these ratings as continuous data. This approach is supported by the research objective of quantifying subtle variances in service quality, which may not be as effectively captured using categorical analysis techniques. Treating Likert scale data as continuous is a common practice in research fields such as social norms,[27] investment decision-making,[28] individual health,[29] and government trust,[30] particularly when the scale includes five or more points,[31] as is the case here. This assumption allows for the use of linear

regression techniques, providing more granular insights into the factors influencing service quality ratings and their relative impacts.

Ethics statement

This study was conducted in strict accordance with the ethical standards of the Declaration of Helsinki and was approved by the Clinical Medical Ethics Committee of Xiangya Hospital, Central South University (Ethics Review No. 202011184). Prior to the commencement of the study, all participants were provided with comprehensive information about the objectives, methods, potential benefits, and risks associated with the research. Informed consent was obtained from all participants involved in the study. The consent procedure was designed to ensure that participants were fully aware of their rights, including the right to withdraw from the study at any point without any consequences. To document the process, written consent was obtained from each participant, which they signed after having sufficient time to read the consent form and having the opportunity to ask questions.

Methodological approach for item weighting

To ascertain the weight of the item system, we adopted a hybrid approach that amalgamates factor analysis with machine learning techniques. Factor analysis was employed to decouple the interrelated indicators, thereby extracting five orthogonal common factors that encapsulate the majority of the information inherent in the original item. This analytical step enhances the objectivity and scientific rigor in the determination of item weights. Complementarily, machine learning algorithms, serving as an avant-garde artificial intelligence evaluation methodology, were utilized to refine the weights of secondary indicators. This was achieved through multiple iterative training and testing cycles, predicated on the outcomes derived from the factor analysis. This dual-method approach effectively mitigates the subjectivity potentially introduced by human evaluators, while the stability and intelligent operations intrinsic to machine learning algorithms further bolster the reliability and scientific validity of the weight determination.

Calculation of common factor score. Determining the Weight of Secondary Indicators. The number of secondary indicators associated with each primary indicator (categories A–E) varies based on the factor loadings derived from the exploratory factor analysis. Each primary indicator was found to encompass a different set of dimensions that describe the service quality. Consequently, the number of secondary indicators varies, reflecting the comprehensive nature of each primary indicator's coverage of specific service quality aspects. Equation (1) was employed to

determine the weight of each secondary indicator based on their respective standardized factor loadings, ensuring that each factor's unique contribution to the overall service quality assessment is accurately quantified:

$$\delta_{ij} = \frac{\alpha_{ij}}{\sum_{n=1}^{f} \alpha_{in}} \tag{1}$$

where δ_{ij} denotes the weight of the jth item on the ith common factor; α_{ij} represents the standardized factor load value of the jth item on the ith common factor; α_{in} is the standardized factor load value of the item on the ith common factor; i is the common factor number, with $i = 1, 2, 3, 4, 5$; j is the sequence number of indicators subordinate to the common factor, with $j = 1, 2, 3, 4, 5, 6, 7$; f is the number of indicators under the common factor, with $f = 5, 6, 7$.

Utilizing Equation (1), the weight value for each secondary indicator under the five primary indicators can be computed, as delineated in Supplemental Table 3.

Formula for Common Factor Scores. The scores for each common factor within individual samples can be computed using the weight values of the secondary indicators. The calculation is represented by the following equations:

$$
\begin{aligned}
F_1 &= 0.166a_1 + 0.184a_2 + 0.158a_3 + 0.173a_4 \\
&\quad + 0.162a_5 + 0.157a_6 \\
F_2 &= 0.164b_1 + 0.188b_2 + 0.189b_3 + 0.186b_4 \\
&\quad + 0.138b_5 + 0.135b_6 \\
F_3 &= 0.219c_1 + 0.239c_2 + 0.188c_3 + 0.164c_4 + 0.19c_5 \\
F_4 &= 0.184d_1 + 0.186d_2 + 0.165d_3 + 0.17d_4 \\
&\quad + 0.146d_5 + 0.149d_6 \\
F_5 &= 0.112e_1 + 0.163e_2 + 0.142e_3 + 0.159e_4 \\
&\quad + 0.145e_5 + 0.133e_6 + 0.146e_7
\end{aligned}
$$

$$\tag{2}$$

In Equation (2), $F_1 - F_5$ is the score of five common factors, and $a_1 - a_6$, $b_1 - b_6$, $c_1 - c_5$, $d_1 - d_6$, and $e_1 - e_7$ correspond to the evaluation data for each sample. By substituting these variables into Equation (2), the score for each dimension within each sample can be ascertained. The resultant scores range between [0, 5], with the intervals [0, 1], [1, 2], [2, 3], [3, 4], and [4, 5], respectively, denoting levels of satisfaction: very dissatisfied, dissatisfied, average, fairly satisfied, and very satisfied. These scores will serve as the input data for the BPNN, FNN, and SVM.

Normalized Data Processing. To ensure that each dimension of the secondary indicators operates on a comparable scale, thereby facilitating a more robust comprehensive evaluation, we employ data normalization techniques. Specifically, the mapminmax function in MATLAB is utilized to normalize the dataset, ensuring that the values for each dimension are within the same order of magnitude.

Data Classification for Machine Learning. In the realm of machine learning, a conventional ratio for partitioning training and test datasets is 4:1. Adhering to this standard, the present study designates 251 data points as the training dataset, while the remaining 50 data points are allocated to the test dataset.

Data analysis techniques

BPNN. The BPNN is a well-established artificial neural network architecture comprising three layers: the input layer, hidden layer, and output layer. In this study, we employ a three-layer BPNN with a single hidden layer. This single-layer, multi-neuron hidden layer structure has been demonstrated to effectively enhance training accuracy and is sufficient for most research applications.[32] The input layer comprises 30 neurons corresponding to the 30 secondary indicators identified in the evaluation item system which consists of five common factor score values, and the hidden layer contains 15 neurons, determined through empirical testing and cross-validation to balance model complexity and performance, while the output layer is designed with one node corresponding to the overall evaluation degree of each sample's service quality in integrated health and social care institutions. The BPNN processes the input data through the hidden layer and outputs the predicted service quality evaluation scores for each sample. The study employs a trial-and-error approach to optimize parameters such as the number of nodes, activation functions, training functions, maximum learning rate, target error value, and maximum number of iterations. Performance metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE) are used for model evaluation, supplemented by Mean Absolute Percentage Error (MAPE) and Accuracy.

T-S FNN. The T-S FNN is selected due to its capability to model complex nonlinear relationships and its strong adaptive abilities. The model approximates continuous nonlinear systems with arbitrary accuracy through a set of "If-Then" fuzzy rules. The node design for the input and output layers mirrors that of the BPNN, comprising an input layer of 30 neurons, two hidden layers with 20 and 10 neurons, respectively, and an output layer with one neuron. The output layer generates evaluation results based on the principle of maximum membership degree. After normalization, the initial domain for all input data is set to [0, 1], and linguistic variables are defined across five levels: "A (very poor)," "B (poor)," "C (medium)," "D (good)," and "E (excellent)." The Gaussian function is selected as the fuzzy processing function, with parameters such as center, width, and membership function coefficients initialized randomly.

SVM. For SVM model construction, the selection of an appropriate kernel function and the optimization of parameters, including support vectors and Lagrange multipliers, are crucial. The Radial Basis Function (RBF) kernel is often preferred due to its effective fitting capabilities.[33] In this study, the RBF kernel function was selected due to its effectiveness in handling nonlinear relationships in the data, The penalty parameter (C) and kernel coefficient (gamma) were optimized through grid search cross-validation, testing values in the ranges of $C = [0.1, 1, 10, 100]$ and gamma $= [0.001, 0.01, 0.1, 1]$. The SVMcgForClass function was employed to iteratively refine the penalty and kernel parameters. The optimal penalty parameter was determined to be 2, and the kernel function parameter was 0.17678, yielding the minimum training error and optimal predictive outcomes.

Parameter selection and optimization. The selection of the number of neurons and layers for the neural network models was based on iterative experimentation and performance evaluation using metrics such as Mean Squared Error (MSE) and R-squared (R^2) (see Supplemental Table 4). Various configurations were tested to prevent overfitting and underfitting, ensuring optimal model generalization.

Training and evaluation. All models were trained using the Adam optimizer with a learning rate of 0.001 over 100 epochs. Early stopping mechanisms were employed to prevent overfitting. The performance of each model was evaluated on a separate test dataset comprising 50 samples, using evaluation metrics appropriate for regression tasks, including MSE, Root Mean Squared Error (RMSE), and R^2.

Type of activation functions

Specifically, we utilized the Rectified Linear Unit (ReLU) activation function in the hidden layers of the BPNN and FNN models. ReLU was chosen due to its advantages in mitigating the vanishing gradient problem, which is commonly encountered in deep neural networks. The ReLU function allows for faster convergence during training by enabling nonlinearity while maintaining computational efficiency.[34]

For the output layer, we employed a linear activation function in the BPNN and FNN models, which is appropriate for regression tasks where the output is a continuous variable. The linear activation function ensures that the output can take any real value, which is essential for accurately predicting the continuous service quality scores.

We also experimented with other activation functions, including the Sigmoid and Tanh functions. However, these functions tended to suffer from slower convergence and the vanishing gradient issue, particularly in the deeper layers of the networks, leading to less accurate predictions. The ReLU function consistently outperformed these alternatives in terms of both training speed and prediction accuracy, as evidenced by lower MSE values on the validation set.[35]

Overall, the choice of ReLU as the activation function in the hidden layers significantly contributed to the robustness and efficiency of the models, enabling them to handle the nonlinear relationships inherent in the service quality data more effectively. The use of a linear activation function in the output layer further ensured that the models could accurately predict the continuous target variable.

Output constraints and float handling

To ensure that the predicted values from the regression model remained within the 1 to 5 range of the Likert scale, we incorporated constraints directly in the model output layer by applying activation function to limit predictions to this range, and the impact of different activation functions was also assessed, with ReLU and Sigmoid functions demonstrating superior performance in terms of convergence speed and predictive performance metrics compared to alternatives like Tanh and Linear functions. Additionally, for any rare cases where predictions exceeded these bounds due to model adjustments, we applied post-hoc adjustments, capping all out-of-range values at 1 and 5. This combination of output constraints and post-hoc adjustments ensured the predictions were consistent and aligned with the Likert scale's evaluative standards.

The original Likert scale responses were discrete integer values from 1 to 5. However, the regression model generated continuous (float) predictions, which represent varying degrees of satisfaction or intensity, even if they do not correspond to exact Likert scale points. These float values enable a more nuanced interpretation by capturing subtle differences in predicted satisfaction levels. For practical applications requiring integer values, we rounded the floats to the nearest integer within the 1–5 range to maintain consistency with the original Likert scale format.

Results

Comparative analysis of experimental results

After optimizing a series of model parameters, 251 training samples were fed into both the BPNN and FNN for training and learning processes. The BPNN achieved an optimal network training result with an error rate of 0.11954 after 12 iterations, while the FNN reached its optimal network training result with an error rate of 0.11483 after 527 iterations. Multiple training comparisons confirmed that these results represent the most favorable training outcomes, thereby yielding the BPNN- and FNN-based service quality evaluation models for integrated health and social care institutions.

To validate the predictive performance metrics of these models, 50 test samples were input into the trained networks for simulation testing. Subsequent calculations revealed that the predictive and evaluative accuracies for

the BPNN, FNN, and SVM models are 90%, 86%, and 76%, respectively. These specific results are tabulated in Supplemental Table 5.

Through computational analyses of the three models, it becomes evident that the service quality evaluation model predicated on BPNN demonstrates robust predictive efficacy, minimal error rates, and commendable generalization capabilities. These attributes collectively suggest that the item weight determination methodology, when grounded in the BPNN framework, possesses substantive feasibility for application in evaluating the service quality of integrated health and social care institutions.

Determination of item weight value and evaluation results

According to the service quality evaluation model of the integrated health and social care institutions based on the BPNN, the weights of the secondary indicators within the item system are further determined. To elucidate the specific relationship between input and output data, it is imperative to calculate the weight coefficients between the input layer and the hidden layer, as well as between the hidden layer and the output layer. The mathematical formulations for these calculations are presented as Equations (3) to (5), with the steps delineated as follows:

1. Determine the correlation significance coefficient,
2. Determine the correlation coefficient,
3. Obtain the weight value.

$$r_{ij} = \sum_{k-1}^{n} w_{ki} \frac{(1-e^{-w}jk)}{(1 + e^{-w}jk)} \qquad (3)$$

$$R_{ij} = \left| \frac{1-e^{-r}ij}{1 + e^{-r}ij} \right| \qquad (4)$$

$$T_{ij} = \frac{R_{ij}}{\sum_{i=1}^{p} R_{ij}} \qquad (5)$$

In Equations (3) to (5), i represents the number of input layer nodes of BPNN, $i = 1, 2, 3, \ldots, p$, $p = 30$; j represents the number of nodes in the output layer, $j = 1$; k represents the number of hidden layer nodes, $k = 1, 2, 3, \ldots, n$, $n = 7$; w_{ki} represents the weight coefficient between the hidden layer neuron k and the input layer neuron i; w_{jk} represents the weight coefficient between the output layer neuron j and the hidden layer neuron k.

Taking the secondary indicator "A1: living care facilities and equipment are relatively perfect" as an example, the weight coefficient from the input layer to the hidden layer can be obtained as $W_{ki} = \{0.025, -0.0178, 0.0311, -0.0293, -0.0298, -0.0242, 0.0336\}$, and from the hidden layer to the output layer as $W_{ji} = \{0.1951, -0.1423,$

0.2322, −0.2198, −0.2251, −0.1894, 0.2452}. Substituting these into Equation (3) yields a correlation significance coefficient of 0.0202 (r_{11}). Further substitution into Equation (4) results in a correlation coefficient of 0.0101. Finally, substituting into Equation (5) yields a weight value for A1 of 0.0327 (T_{11}).

Consequently, the weights of the remaining 29 secondary indicators can be calculated in a similar manner, and the weights of the primary indicators can be obtained through Equation (6).

$$\delta_m = \frac{\sum_{i=1}^{q} T_{ig}}{\sum_{i=1}^{p} T_{ij}} \tag{6}$$

In Equation (6), m represents the primary indicator numbers 1, 2, 3, 4, 5; g is the secondary indicator corresponding to the primary indicator, ranging from 1, 2, 3, …, to q, where $q = 7$; T_{ig} is the weight value of the gth secondary indicator on the ith primary indicator; δ_m is the weight value of the mth primary indicator. After calculation, the weight values of the primary indicators are ranked as follows: medical nursing service > life care service > spiritual comfort service > rehabilitation health service > cultural and entertainment service. The specific results are presented in Supplemental Table 6.

Evaluation results

Utilizing the item weight values delineated in the preceding section, the final score, overall score, and comprehensive score across the five dimensions for the 301 samples data were computed. The computational results are tabulated in Supplemental Table 7.

The evaluation results indicate an overall comprehensive score of 3.5859 out of a maximum of 5 points for the 301 samples, situating the service quality in the upper-middle range. This suggests that the older adults surveyed are generally satisfied with the service quality provided by the institutions. While the services essentially meet the basic needs of the older adults, there remains substantial room for improvement. The data also reveal a considerable score gap, with the highest score being 5 and the lowest 1.8903. Although this variance can be attributed to the subjective experiences of the older adults and the disparities among institutions, it also underscores the imperative for integrated care institutions to enhance service quality, elevate client satisfaction, and narrow this gap. Furthermore, the scores across the five dimensions rank as follows: Life Care Services > Medical Care Services > Spiritual Comfort Services > Rehabilitation and Health Services > Cultural and Entertainment Services.

Discussion

The primary objective of this research was to construct a robust service quality evaluation model for institutions that amalgamate health and social care services for older adults. Utilizing three machine learning algorithms—BPNN, FNN, and SVM—the study conducted rigorous training and validation processes. The results indicated that the BPNN-based model was the most efficacious in accurately representing the real-world service quality experienced by older adults. This finding corroborates the extant literature that underscores the superior learning and generalization capabilities of machine learning algorithms in constructing efficient, intelligent, and scientifically rigorous evaluation models.[36]

Upon employing the optimal model, the study revealed that living care services emerged as a pivotal component in the integrated health and social care institutions. The capacity of these institutions to offer comprehensive and nuanced living care services that align with the needs of older adults was identified as a significant determinant of their satisfaction, a finding that is in agreement with prior research.[37] Medical care services also surfaced as a critical quality indicator, underscoring the pressing medical needs of older adults and the imperative of integrating health and social care services.[38]

Moreover, other dimensions such as spiritual comfort, rehabilitation and health services, and cultural and entertainment services were found to influence the quality-of-service satisfaction, albeit to a lesser degree. These are domains where institutional efforts could be intensified. Since the integration of health and social care is still in its early stages domestically, most related research has focused on individual employee performance evaluation,[39,40] older residents' willingness to move in, and their needs for accommodation.[41,42] There is relatively little research specifically addressing service quality, and in the pursuit of scale, quality evaluation is easily overlooked. Therefore, compared to other studies, this study accentuates the necessity of developing sophisticated service quality evaluation models for integrated health and social care institutions to cater to the multifaceted needs of older adults.[43] From the perspective of reliability, machine learning algorithms have strong learning generalization ability, which helps to build more efficient, intelligent, and scientific evaluation models. The service quality evaluation model in this study has multiple advantages such as high accuracy, stability, and low error, which can effectively reflect the true service quality level of integrated health and social care institutions for older adults. In terms of application, the model in this study only needs to input the collected data into a trained BP neural network to obtain the weight values of the indicators. After simple calculation, the evaluation score can be obtained, which helps simplify the evaluation work and improve the quality.

The implications of this research are manifold. To enhance the quality of service delivering to older adults, there is a pressing need to optimize service structures and bridge the service supply–demand gap. This optimization should be congruent with the specific needs of older

adults and the resource constraints that institutions may face. The government should organize experts and practitioners from related disciplines and industries such as medical and health care, and medical insurance to form a specialized research group for evaluating the service quality of health and social care institutions. Based on the needs and characteristics of older adults, a universal evaluation standard for the service quality of integrated health and social care institutions should be jointly developed. At the same time, pilot projects in different regions should be accelerated and actively adjusted to form a mandatory national standard forcing quality improvement model. Institutions should aim to ensure high-quality basic services and also focus on the health needs of older adults by providing stable, long-term medical services, timely interventions, and competent medical staff.[44] Additionally, the introduction of regular rehabilitation and health care programs, along with an emphasis on improving the quality of soft services like spiritual communication, is advisable. In addition, services such as spiritual comfort, cultural and entertainment services need to be improved. After meeting basic living needs, the government and institutions should enhance the sense of achievement of older adults in these service projects through various measures such as strengthening infrastructure construction, and organizing mutual assistance activities.

The contributions of this study are twofold. First, it adds substantive value to the extant literature on service quality evaluation in the context of care for older adults. By employing machine learning algorithms, the study introduces a level of methodological rigor and predictive performance metrics that enhances the reliability and validity of service quality assessments. Second, the study holds practical significance as it provides actionable insights into the areas requiring improvement within integrated health and social care institutions. These insights are particularly valuable for stakeholders, including policymakers and institutional administrators, who are vested in elevating the standard of care provided to older adults.

Limitations and future research directions

This study has some limitations that should be noted. Firstly, the generalizability of the findings may be constrained by the specific context of the Chinese healthcare system and the particular characteristics of the social care settings involved in the study, with potentially significant differences between domestic and international contexts.

Secondly, the reliance on machine learning algorithms necessitates a substantial amount of data, and the quality of the outcomes is inherently dependent on the quality of the data collected. The data for this study were sourced from Hunan Province. Compared to nationwide data, the sample size is relatively small, which may result in some deviations in the results. Even though, based on the test results, these deviations remain within an acceptable range.

Thirdly, while this study primarily focuses on comparing various nonlinear models for evaluating service quality, it does not provide a direct comparison with linear models in the current analysis. In our previous research,[24] we have extensively utilized linear SEM to establish and validate the foundational relationships within our service quality framework. Building on that foundation, this study explores advanced nonlinear models to address the limitations that linear models may face in capturing complex, nonlinear interactions. Future research could expand on this work by directly comparing the performance and suitability of both linear and nonlinear models to offer a more comprehensive understanding of their relative strengths in different contexts of service quality evaluation.

Lastly, while our model for evaluating service quality was designed to capture the complex and interdependent relationships among various service dimensions, we acknowledge that the absence of ablation studies is a limitation. Ablation studies are useful for isolating and understanding the individual contributions of specific model components by systematically removing or modifying them. However, in our study, the focus was on optimizing the overall model performance to reflect the intricate interactions within the service quality data, rather than isolating individual components. Conducting ablation studies in this context could potentially oversimplify these complex relationships and shift the focus away from our goal of understanding the overall model's effectiveness. Nevertheless, future research could incorporate ablation studies to dissect the contributions of different model elements, striking a balance between comprehensive evaluation and the understanding of specific component impacts. This approach could provide more detailed insights and potentially enhance model efficiency.

Furthermore, while this study provides an integrated evaluation of health and social care quality in settings for older adults, future research could benefit from examining these components independently to better understand their distinct impacts. Separate evaluations of health care and social care may highlight unique strengths and areas requiring improvement in each domain. Additionally, analyzing the interaction between health and social care services could uncover synergies that enhance overall service quality. Such insights are crucial for optimizing resource allocation and designing services that comprehensively address the needs of older adults.

Conclusions

The present study employed three machine learning algorithms—BPNN, FNN, and SVM—to develop distinct service quality evaluation models for institutions that integrate health and social care services for older adults.

Simulation software facilitated the training of these models, with the factor analysis-fuzzy BPNN model ultimately being selected based on its error indices and predictive performance metrics. Subsequent to this selection, a comprehensive evaluation of service quality in integrated health and social care settings was conducted through a process of empowerment and training. This evaluation serves to illuminate existing service-related issues in a manner that is both intuitive and empirically grounded.

Looking ahead, future research endeavors could consider expanding the sample size to include a more diverse range of institutions. This would enable a more nuanced understanding of how service quality varies across different types of care settings. Moreover, such an expansion would provide the empirical basis for offering targeted recommendations aimed at improving service quality across various categories of institutions. Overall, this study serves as a foundational step in enriching and advancing the research landscape concerning the quality evaluation of services for older adults.

ORCID iD: Zhihan Liu https://orcid.org/0000-0003-0659-1318

References

1. Zhan G, Huang W, Zhou C, et al. Evolutions of population distribution in China from the perspective of urban agglomeration. *J Geogr* 2018; 73: 1513–1525.
2. Wang Z and Liu Z. Older adults' demand for integrated care and its influencing factors: a scoping review. *Int J Integr Care* 2021; 21: 28.
3. Liu Z. Integrated care for older people: theories and practices. *Int J Integr Care* 2023; 23: 24, 1–23.
4. Wang Z and Liu Z. Latent classes and related predictors of demand for home-and community-based integrated care for older Chinese adults. *Front Public Health* 2023; 11: 1109981.
5. Dong K, Wang Z and Zhang D. China's aging population and construction of pension system. *Comp Econ Soc Syst* 2020; 01: 53–64.
6. Shen S. The ability of integrated health and social care services continued to improve, https://www.gov.cn/xinwen/2021-04/09/content_5598575.htm (2021, accessed 1 October 2022).
7. Chang B, Yang Y, Leon GAB, et al. Effect of collaborative governance on medical and nursing service combination: an evaluation based on Delphi and entropy method. *Healthcare* 2021; 9: 1456.
8. Uittenbroek RJ, Reijneveld SA, Stewart RE, et al. Development and psychometric evaluation of a measure to evaluate the quality of integrated care: the patient assessment of integrated elderly care. *Health Expect* 2016; 19: 962–972.
9. Moxey ED, O'connor JP, White E, et al. Developing a quality measurement tool and reporting format for long-term care. *Jt Comm J Qual Improv* 2002; 28: 180–196.
10. Jones J and Hunter D. Consensus methods for medical and health services research. *Br Med J* 1995; 311: 376–380.
11. Hu J, Zhang Y, Wang L, et al. An evaluation item system of basic elderly care services based on the perspective of accessibility. *Int J Environ Res Public Health* 2022; 19: 4256.
12. Castle NG and Ferguson JC. What is nursing home quality and how is it measured? *Gerontologist* 2010; 50: 426–442.
13. Goodson J and Jang W. Assessing nursing home care quality through Bayesian networks. *Health Care Manag Sci* 2008; 11: 382–392.
14. Li A, Wang D and Zhu M. Service evaluation through FH-entropy method: a framework for the elderly care station. *Concurr Comput Pract Exp* 2022; 34: e6045.

15. Gustafson DH, Fiss CJ, Fryback JC, et al. Measuring the quality of care in nursing homes: a pilot study in Wisconsin. *Public Health Rep* 1980; 95: 336–343.

16. Jordan MI and Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015; 349: 255–260.

17. Cai K, Cai L, Wang W, et al. On the neural network approach in software reliability modeling. *J Syst Softw* 2001; 58: 47–62.

18. Yuan Y and Wang C. IPTV video quality assessment model based on neural network. *J Vis Commun Image Represent* 2019; 64: 102629.

19. Sun X and Yang X. The structure of mental elasticity education for children in plight using deep learning. *Front Psychol* 2022; 12: 766658.

20. Heaton J. An empirical analysis of feature engineering for predictive modeling. In: *SoutheastCon 2016*. Norfolk, VA, USA, 30 March–3 April 2016, pp.1–6. IEEE.

21. Meneganti M, Saviello FS and Tagliaferri R. Fuzzy neural networks for classification and detection of anomalies. *IEEE Trans Neural Netw* 1998; 9: 848–861.

22. Yu W and Li X. Fuzzy identification using fuzzy neural networks with stable learning algorithms. *IEEE Trans Fuzzy Syst* 2004; 12: 411–420.

23. Akram-Ali-Hammouri Z, Fernández-Delgado M, Cernadas E, et al. Fast support vector classification for large-scale problems. *IEEE Trans Pattern Anal Mach Intell* 2021; 44: 6184–6195.

24. Liu Z and Ou C. Establishment of care service quality evaluation index system for the pension institutions combined with medical service. *Chin J Health Policy* 2021; 14: 59–67.

25. Wang L, Feng Y and Wang C. Research on service quality evaluation of elderly care institutions. *Popul Dev* 2017; 6: 96–102.

26. Berdie DR. Reassessing the value of high response rates to mail surveys. *Market Res* 1989; 1: 52–64.

27. Hu Z and Navarrete E. Understanding the impact of foreign language on social norms through lies. *Biling: Lang Cogn* 2024; First View: 1–15.

28. Suresh G. Impact of financial literacy and behavioural biases on investment decision-making. *FIIB Bus Rev* 2024; 13: 72–86.

29. Yu Y, Li J and Yang Z. Grassroots performance and government trust: evidence from community public service facilities construction in "people's city" Shanghai. *J Soc Sci* 2024; 09: 158–169.

30. Zhao M and Huang Y. The influence of adverse experiences on the health of Chinese older adults: mechanism analysis of adverse childhood experiences and moderating analysis of current experiences. *Popul Res* 2024; 48: 114–128.

31. Rhemtulla M, Brosseau-Liard PÉ and Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods* 2012; 17: 354–373.

32. Chen M. *Matlab neural network principles and case studies.* People's Posts and Telecommunications Press, 2014 [cited 2023 6–24]. https://thinker.cnki.net/bookstore/book/bookdetail?bookcode=9787115348685000&type=book.

33. Yang Y and Huang D. Research and applications of artificial neural networks. *J East China Univ Sci Technol* 2002; 5: 551–554.

34. Nair V and Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp.807–814.

35. Glorot X, Bordes A and Bengio Y. Deep sparse rectifier neural networks. In: Proceedings of the 14th international conference on artificial intelligence and statistics (AISTATS-11), 2011, pp.315–323.

36. Jost A. Neural networks: a logical progression in credit and marketing decision system. *Credit World* 1993; 81: 26–33.

37. Hjaltadóttir I, Hallberg IR and Ekwall AK. Thresholds for minimum data set quality indicators developed and applied in Icelandic nursing homes. *J Nurs Care Qual* 2012; 27: 266–276.

38. Ivan F, Hernández PS, Kelienny D, et al. Quality of care in nursing homes in Brazil. *J Am Med Dir Assoc* 2017; 18: 555–638.

39. Wang H. Service efficiency and its influencing factors of integrated health and elder care institutions: based on the data analysis of 226 integrated health and elder care institutions in Shandong Province. *Chongqing Soc Sci* 2020; 306: 131–142.

40. Xiao X, Huang Z, Wu Y, et al. Construction of a service performance evaluation item system for integrated health and elder care institutions. *Chin Gen Pract* 2019; 22: 3233–3237.

41. Han Y and Li H. Analysis of the willingness to reside in integrated health and social care institutions and its influencing factors among older adults in Jinzhou. *Chin Gen Pract* 2018; 21: 1456–1460.

42. Liu X, Wei K, Gu L, et al. Service needs attributes and influencing factors of older residents in integrated health and social care institutions. *Nurs Res* 2020; 34: 3373–3381.

43. Wang J, Peng D, Liu W, et al. Current status of physical restraint among older residents in integrated health and social care institutions and its influencing factors. *Chin Nurs Manag* 2020; 20: 1503–1509.

44. Nakrem S, Vinsnes AG, Harkless GE, et al. Nursing sensitive quality indicators for nursing home care: international review of literature, policy and practice. *Int J Nurs Stud* 2009; 46: 848–857.