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.

Exploring the Potential of Machine Learning for Mapping FACT-L to EQ-5D-5L in Lung Cancer

Hongli Tao
West China Tianfu Hospital, Sichuan University
Longlin Jiang
jianglonglin@stu.cdutcm.edu.cn
Chengdu University of Traditional Chinese Medicine
Pengwei Yang
University of Electronic Science and Technology of China

Research Article

Keywords: Lung Cancer, Machine Learning, Mapping, Health Utility, EQ-5D-5L, FACT-L, Health-Related Quality of Life

Posted Date: March 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8990430/v1

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

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Quality of Life Research on November 22nd, 2026. See the published version at https://doi.org/10.1007/s11136-026-04437-z.

Exploring the Potential of Machine Learning for Mapping FACT-L to EQ-5D-5L in Lung Cancer

Hongli Tao[1], Longlin Jiang[2*], Pengwei Yang[3]

[1]Department of General Medicine, West China Tianfu Hospital, Sichuan University, Chengdu, Sichuan, China.
[2]College of Nursing, Chengdu University of Traditional Chinese Medicine, Chengdu, China.
[3]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China.

*Corresponding author(s). E-mail(s): jianglonglin@stu.cdutcm.edu.cn;
Contributing authors: hongli.tao@hotmail.com;
pengwei.yang@std.uestc.edu.cn;

Abstract

Background: To evaluate machine learning approaches for mapping FACT-L to EQ-5D-5L utilities in lung cancer patients, with a particular focus on exploring neural network models and their predictive performance.
Methods: We enrolled 347 hospitalized lung cancer patients at West China Tianfu Hospital of Sichuan University and collected sociodemographic, clinical, FACT-L, and EQ-5D-5L data. Random forest (RF), extreme gradient boosting (XGBoost), support vector regression (SVR), and neural networks (NN) were evaluated using five-fold cross-validation. Performance was assessed with mean squared error (MSE) and mean absolute error (MAE), and we examined input dimensionality and neural network architecture.
Results: Using all FACT-L items, SVR achieved the best performance (MSE $\approx$ **0.033**, MAE $\approx$ **0.136**), with RF and XGBoost close behind. Model performance improved with increasing input dimensionality, with the best results obtained using all FACT-L items. Among neural networks, the simple architecture (one hidden layer, 32 neurons) achieved competitive performance.
Conclusion: Machine learning improves FACT-L to EQ-5D-5L mapping predictive performance; SVR performed best under small-to-medium dataset conditions, and appropriately designed neural networks show promise for utility estimation in lung cancer.

Keywords: Lung Cancer; Machine Learning; Mapping; Health Utility; EQ-5D-5L; FACT-L; Health-Related Quality of Life

1 Introduction

Lung cancer remains one of the most prevalent and fatal malignancies worldwide and represents the leading cause of cancer-related mortality in China [1, 2]. Beyond its substantial impact on patients' health-related quality of life (HRQoL), lung cancer also places a considerable economic burden on patients across different treatment modalities. HRQoL instruments such as the Functional Assessment of Cancer Therapy–Lung (FACT-L) are used to measure lung cancer–specific quality of life and have been validated in this population [3, 4]. FACT-L is a non-preference-based measure and therefore cannot directly generate utilities for cost-utility analyses (CUAs), limiting its use in resource allocation and treatment decision-making. CUAs are the most common economic evaluation approach in health care, and quality-adjusted life years (QALYs) are the standard outcome [5, 6]. QALYs are calculated by combining survival time with health-state utilities derived from preference-based measures (PBMs) [6].

The EuroQol five-dimension questionnaire (EQ-5D) is the most widely used PBM in clinical trials and health services research and is the preferred instrument for cost-utility assessment in many health technology assessments [7–10]. However, EQ-5D-5L is not disease-specific and can be less sensitive to clinically meaningful changes in lung cancer over time [11].

When PBMs are not collected, health utilities can be estimated by mapping from other HRQoL instruments [11–13]. Mapping algorithms are commonly developed using regression models and are endorsed in best-practice guidance [12]. Traditional regression mapping often overestimates utilities in poorer health and underestimates utilities in better health because of limited flexibility and inadequate modeling of non-linear relationships and interactions among subscales [14]. Modern machine learning methods offer more flexible function approximation and may improve predictive accuracy compared with traditional regression approaches, yet they remain underused in utility mapping [14–16]. Among these approaches, neural networks are particularly attractive because they can model complex non-linear relationships and interactions between HRQoL dimensions, although their performance may depend strongly on data size and model architecture [17]. Therefore, the aim of this study was to develop machine learning-based mapping algorithms to predict EQ-5D-5L utility in postoperative lung cancer patients based on FACT-L scores for accurate economic evaluation without PBMs.

To date, several studies conducted in the United Kingdom, the United States, and China have explored mapping from FACT-L to EQ-5D-3L among patients with lung cancer [18–20]. However, previous evidence indicates that EQ-5D-5L demonstrates improved measurement properties compared with EQ-5D-3L, highlighting the potential value of developing mapping algorithms from FACT-L to EQ-5D-5L for quality-of-life assessment and utility estimation [21, 22]. Although studies have begun to focus on EQ-5D-5L, the application of modern machine learning approaches, such

as neural networks, remains limited, and further exploration is needed to improve model performance and generalizability [23]. Therefore, we aimed to further investigate machine learning-based mapping algorithms from FACT-L to EQ-5D-5L index scores based on a Chinese lung cancer population, in accordance with the recommendations of the "Mapping onto Preference-based Measures Reporting Standards Reporting Statement" [24] and the "ISPOR Good Practices Task Force Report on Mapping" [13].

2 Methods

2.1 Participants

We enrolled 347 hospitalized lung cancer patients from the Departments of Thoracic Surgery, the Lung Cancer Center, and the Department of Comprehensive Medicine at West China Tianfu Hospital of Sichuan University. Inclusion criteria were: (1) pathologic or clinical diagnosis of lung cancer; (2) age 18 years or older; (3) no cognitive impairment; and (4) ability to provide informed consent. Exclusion criteria included refusal or inability to provide consent, severe comorbid physical or psychiatric illness, and age under 18 years. All participants provided written informed consent. The study was approved by the Ethics Committee of West China Tianfu Hospital of Sichuan University. The study was endorsed by the Ethics Committee of West China Tianfu Hospital, Sichuan University with approval number MR-51-25-042281.

2.2 Instruments

FACT-L is a disease-specific HRQoL instrument for lung cancer, consisting of the FACT-G core domains (physical, social/family, emotional, and functional well-being) plus the Lung Cancer Subscale (LCS). The FACT-L (v4.0) includes 36 items scored on a 0–4 Likert scale, with a total score of 0–144; higher scores indicate better HRQoL [3]. It comprises 36 items across five dimensions: 7 items assessing physical well-being (PWB), 7 items assessing social/family well-being (SWB), 6 items assessing emotional well-being (EWB), and 7 items assessing functional well-being (FWB), while the LCS includes 9 items (2 of the 9 items are not included in the score) [3]. FACT-L has demonstrated strong reliability and validity in lung cancer populations and is sensitive to symptoms such as cough, dyspnea, and chest pain. The Trial Outcome Index (TOI) is commonly reported as a condensed indicator emphasizing physical and functional outcomes [25].

EQ-5D-5L is a widely used generic preference-based measure with five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) and the EQ-VAS (0–100). EQ-5D includes the 3-level (EQ-5D-3L) and 5-level (EQ-5D-5L) versions. EQ-5D-5L assesses mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has five response levels, yielding 3,125 health states. The Chinese EQ-5D-5L value set ranges from -0.391 to 1, where 1 indicates full health, 0 is equivalent to death, and negative values represent states worse than death [26].

2.3 Correlation analysis

We used Spearman's rank correlation coefficient (ρ) to assess the strength of monotonic associations between variables. Because there is no universally accepted standard for interpreting correlation magnitude, we followed commonly used empirical thresholds in medical statistics: absolute values < 0.20 indicate very weak correlation, 0.20–0.39 weak correlation, 0.40–0.59 moderate correlation, 0.60–0.79 strong correlation, and absolute values ≥ 0.80 very strong correlation [27, 28]. Notably, stronger correlations between source and target instruments indicate greater shared variance in the constructs being measured, which facilitates more accurate mapping algorithm development [29]. Specifically, when correlations between FACT-L domains and EQ-5D-5L dimensions exceed moderate levels ($\rho > 0.40$), the predictability of utilities improves substantially, enabling robust regression or machine learning mapping functions.

2.4 Modeling approaches

Given the bounded range, skewed distribution, and potential ceiling effects of EQ-5D-5L utilities, and the likely non-linear and interactive relationships between FACT-L dimensions and health utility, we developed mapping models using four machine learning methods: RF, XGBoost, SVR, and NN. These methods have been increasingly used in utility mapping and are considered well suited to capturing complex relationships.

Random forest (RF). Random forest is an ensemble learning method based on bootstrap sampling and random feature selection. It builds multiple regression trees and averages their predictions to obtain final estimates [30]. The method does not rely on linear assumptions and can capture non-linear relationships and interactions. In the context of utility mapping, where FACT-L domains may be highly interrelated, RF can help identify influential predictors and improve stability through model averaging. However, the piecewise nature of tree-based predictions may limit performance in extreme utility ranges.

Extreme Gradient Boosting (XGBoost). XGBoost is a gradient boosting framework that fits regression trees sequentially to residuals from prior iterations, thereby refining predictions. Unlike RF's parallel trees, XGBoost uses a stage-wise strategy and incorporates regularization to control complexity. Gradient boosting has been applied in multiple utility-mapping studies [15]. Given the potential heterogeneity in EQ-5D-5L distributions, XGBoost's iterative optimization may improve overall fit, although performance depends on hyperparameter tuning and thus requires systematic cross-validation.

Support vector regression (SVR). SVR is grounded in structural risk minimization and seeks a maximum-margin regression function. By using kernel functions, SVR maps inputs into a higher-dimensional feature space to model non-linear relationships [31]. SVR has shown strong generalization in studies with limited sample sizes. With the radial basis function (RBF) kernel, SVR can flexibly approximate the relationship between FACT-L and EQ-5D-5L. A limitation is that SVR does not explicitly enforce bounded utility values, which may affect predictions near the upper bound.

Neural networks (NN). Artificial neural networks transform inputs through multiple non-linear layers to approximate complex functions. In utility mapping, NNs

can capture higher-order interactions among questionnaire domains [32]. However, performance is sensitive to sample size and architecture. Overly complex networks may reduce stability when data are limited. We therefore compared multiple network structures to assess suitability under the current sample size.

3 Model optimization and evaluation

Machine learning models are sensitive to hyperparameters that must be optimized during model development [33]. To maximize model performance, we used grid search to identify optimal hyperparameter combinations for the first three algorithms, and we conducted a series of experiments to explore neural network architectures and their optimal hyperparameters [34]. Specifically, we used grid search to exhaustively evaluate predefined hyperparameter combinations and selected configurations that maximized model performance under cross-validation, thereby supporting generalizability to unseen data. In addition, evidence from machine learning suggests that richer feature sets can improve performance when features are relevant [35]. We therefore evaluated three input configurations: all FACT-L items, the five domain scores, and the total score alone.

3.1 Model optimization

We identified key hyperparameters based on their relevance to each algorithm and their expected impact on predictive accuracy, generalization, and computational cost. For random forest, we tuned the number of trees, maximum depth, minimum samples required to split a node, and minimum samples per leaf to balance accuracy and overfitting [30]. For XGBoost, we tuned the number of estimators and maximum depth, the learning rate controlling each tree's contribution, subsampling and feature-subsampling ratios, and the regularization term γ that sets the minimum loss reduction for a split, as well as L1 and L2 penalties to improve robustness [15]. For SVR, we tuned the penalty parameter C, kernel type (linear and radial basis function), the kernel width γ for non-linear kernels, and the ϵ-insensitive loss margin [31].

For neural networks, we designed shallow architectures with varying depth and neuron counts (16–128 per layer) to evaluate model complexity [36]. We used SmoothL1Loss to balance MAE- and MSE-sensitive errors and the AdamW optimizer with weight decay for stable updates. Key hyperparameters (batch size, learning rate, weight decay, and number of epochs) were tuned via grid search [34, 37, 38].

All models included the FACT-L total score, domain scores, and all FACT-L items as inputs (see Table 2). To avoid overfitting and ensure interpretability, we did not add higher-order terms or interaction effects; instead, we relied on the flexibility of the machine learning models to capture non-linearities. Age and sex were included as covariates given their established association with HRQoL in prior mapping studies [39].

3.2 Neural networks architecture search

Figure 1 summarizes the neural network architecture search. As aforementioned, we explored the effects of different input features, network depth, and neuron counts on model performance. Table 1 lists the neural network depth and neuron counts, corresponding to the exploration directions shown in Figure 1. Table 2 describes the input feature dimensions used in the mapping models.

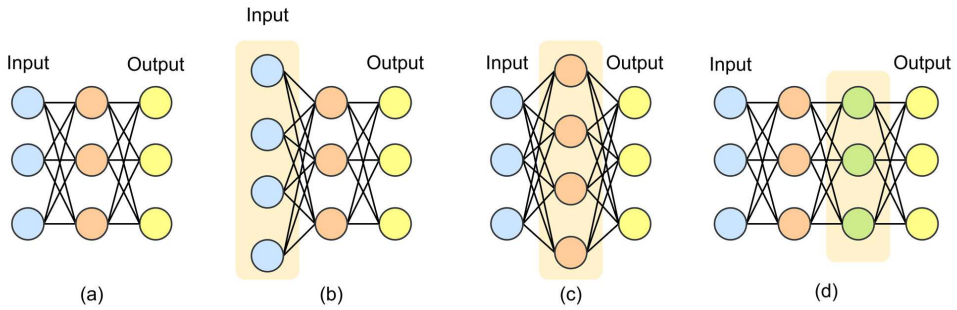


Fig. 1 Neural network architecture search schematic. Panel a shows the initial neural network. Panel b varies the input-layer feature count. Panel c varies the number of neurons in the hidden layers without changing the number of layers. Panel d varies the number of hidden layers.

Table 1 Neural network architecture variants (hidden layers only; input and output layers excluded)

Model	Neurons per layer	Description	Fig. panel
NN	128–64	Original baseline model	a
NN1	64–32	Half the neurons per layer vs. NN	c
NN2	64	Fewer layers (single hidden layer)	d
NN3	32	Fewer layers and half the neurons vs. NN2	d+c
NN4	16	Further halves neurons vs. NN3	d+c
NN5	32–16	Two hidden layers with fewer neurons	c

Table 2 Input feature sets used for model evaluation (see Figure 1 panel b)

Feature set	Dimensions	Description
FACT-L total score	1	Total score only
FACT-L 5 domains	5	PWB, SWB, EWB, FWB, and LCS
FACT-L items	36	All FACT-L item-level dimensions included

3.3 Model evaluation

In the absence of external validation data, we used five-fold cross-validation. The data were split into 80% training and 20% testing in each fold, ensuring that each observation served as test data once. Performance was summarized as the average MSE and MAE across folds. The metrics are defined as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \qquad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|.$$

We chose MSE and MAE because they are complementary: MSE is more sensitive to large errors and highlights severe deviations, while MAE measures overall bias in absolute terms and is more robust and interpretable. Combining them balances penalizing large errors with maintaining overall stability. We optimized a composite error objective:

$$\text{Error} = \alpha \cdot \text{MSE} + (1 - \alpha) \cdot \text{MAE}, \tag{1}$$

with $\alpha = 0.5$ in this study. Future work may examine the sensitivity of model performance to alternative weights.

4 Results

4.1 Score distribution and sample characteristics

We included 347 patients with lung cancer. Demographic and clinical characteristics are summarized in Table 3. Most participants were female (67.4%), and the largest age group was 55–64 years (34.0%). Approximately half had a bachelor's degree or higher (49.0%), and most were married (87.9%). Clinically, stage I disease predominated (76.9%), and non-small cell lung cancer accounted for 99.4% of cases. An ECOG performance status of 2 was most common (57.1%). The majority were non-smokers (85.3%), and BMI was mainly within the normal range (59.4%). Overall, the sample reflects a clinically stable lung cancer population.

Descriptive statistics for EQ-5D-5L and FACT-L are reported in Table 4. The mean EQ-5D-5L utility was 0.842 (SD = 0.215), ranging from -0.163 to 1, with a median of 0.942. The distribution was markedly skewed (skewness = -1.590) with a pronounced ceiling effect (47.56%), indicating that a substantial proportion reported full health. In contrast, the FACT-L total score averaged 111.291 (SD = 16.531), ranged from 63.771 to 142, and had a median of 113. No obvious floor or ceiling effects were observed, and skewness was modest (-0.306). These results suggest that FACT-L provides a wider measurement range and better discrimination in this sample.

FACT-L total and subscale scores are summarized in Table 5. The mean total score was 111.29 (SD = 16.53). Among subscales, SWB was highest (23.61 ± 4.32), FWB was lowest (16.50 ± 5.12), and the LCS was relatively high (27.75 ± 4.66). P10–P90 ranges indicate variability across subscales, highlighting heterogeneity in HRQoL dimensions.

Table 3 Demographic and clinical characteristics of the participants ($n = 347$)

Variables	Grouping	Number of people	Composition ratio (%)
Gender	Male	113	32.6
	Female	234	67.4
Age at time of survey	< 45	99	28.5
	45–54	75	21.6
	55–64	118	34.0
	≥ 65	55	15.9
Educational attainment	Elementary and below	27	7.8
	Junior high school	84	24.2
	Senior high school	64	18.4
	Undergraduate or over	170	49.0
Marital status	Single	18	5.2
	Married	305	87.9
	Divorced/separated	10	2.9
	Widowed	14	4.0
Number of children	One child	255	73.5
	Two children	50	14.4
	Three children	11	3.2
TNM stage	I	267	76.9
	II	22	6.3
	III	0	0.0
	IV	0	0.0
Pathological pattern	Small cell lung cancer	2	0.6
	Non-small cell lung cancer	345	99.4
ECOG score	0	4	1.2
	1	39	11.2
	2	198	57.1
	3	105	30.3
	4	1	0.3
Smoking history	Yes	51	14.7
	No	296	85.3
BMI index	< 18.5	12	3.5
	18.5–23.9	206	59.4
	≥ 24	129	37.2

Expressed as n (%) unless otherwise specified.

4.2 Correlations

Spearman rank correlations between EQ-5D-5L and FACT-L are shown in Table 6. Significant associations were observed, supporting construct consistency between the instruments. Overall, EQ-5D-5L dimensions were negatively correlated with FACT-L total and subscale scores, indicating that greater reported health problems on EQ-5D-5L corresponded to lower HRQoL on FACT-L. Stronger correlations were mainly observed between EQ-5D-5L dimensions and the PWB and FWB subscales, whereas correlations with SWB were weaker.

Additionally, EQ-5D-5L index values showed moderate to strong positive correlations with FACT-L total and subscale scores. The overall pattern suggests higher

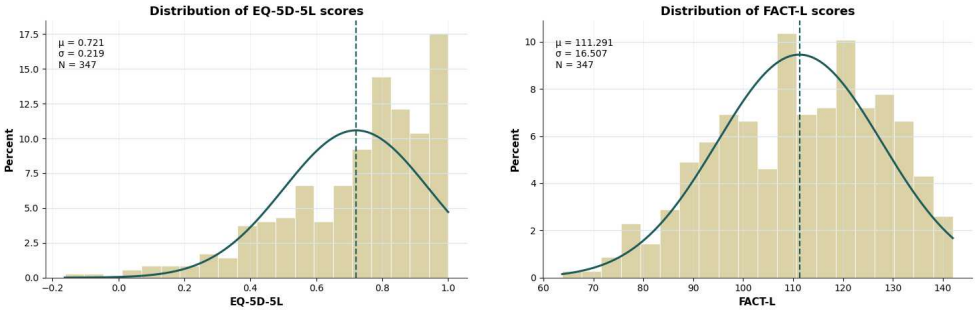


Fig. 2 Distribution of EQ-5D-5L scores (left) and FACT-L scores (right).

Table 4 Description of EQ-5D-5L and FACT-L

Variable	Mean	SD	Minimum	Median	Maximum
	Flooring (%)	Ceiling (%)	Skewness	Kurtosis	Missing data (%)
EQ-5D-5L	0.842	0.215	-0.163	0.942	1.000
	0.16	47.56	-1.590	2.339	< 1
FACT-L	111.291	16.531	63.771	113.000	142.000
	0.00	0.00	-0.306	-0.629	< 1

SD standard deviation

Table 5 Total score and dimension score of FACT-L

Variable	Mean	SD	Minimum	P10	P50	P90	Maximum
FACT-L total	111.29	16.53	63.77	88.94	113.00	132.35	142.00
PWB	21.14	4.48	4.00	15.00	22.00	27.00	28.00
SWB	23.61	4.32	7.00	18.00	24.50	28.00	28.00
EWB	20.25	3.36	5.00	16.00	21.00	24.00	24.00
FWB	16.50	5.12	1.00	9.00	17.00	23.00	24.00
LCS	27.75	4.66	13.50	21.75	28.12	33.75	36.00

alignment in physical and functional constructs, with some divergence in social domains, indicating complementary measurement content.

4.3 Model performance

We report model selection performance using MSE and MAE across input configurations and neural network variants. MSE and MAE are ranked separately (lower is better), and the total rank summarizes each metric set. The overall order follows the total rank.

Table 6 Correlation between EQ-5D-5L scale and FACT-L scale scores

Dimension	FACT-L total score	PWB	SWB	EWB	FWB	LCS
Mobility	-0.30*	-0.32*	-0.08	-0.18*	-0.22*	-0.27*
Self-care	-0.33*	-0.36*	-0.12	-0.23*	-0.24*	-0.18*
Usual activities	-0.38*	-0.42*	-0.14	-0.21*	-0.31*	-0.24*
Pain/Discomfort	-0.39*	-0.63*	-0.18*	-0.33*	-0.32*	-0.26*
Anxiety/Depression	-0.39*	-0.36*	-0.16	-0.41*	-0.29*	-0.19*
EQ-5D-5L total score	0.46*	–	–	–	–	–

* $p < 0.001$

Table 7 shows that SVR achieves the best overall MSE ranking across input settings. Several neural network variants also perform well, with relatively shallow or simpler architectures (e.g., NN3 and NN1) ranking near the top under MSE.

Table 7 MSE-based model ranking across input settings

Model	MSE-36*	rank	MSE-5*	rank	MSE-total*	rank	Total rank
SVR	0.0333	1	0.0372	1	0.0410	4	1
NN3	0.0355	4	0.0380	4	0.0399	2	2
XGBoost	0.0333	1	0.0379	3	0.0417	7	3
NN1	0.0358	5	0.0393	6	0.0405	3	4
NN2	0.0383	9	0.0375	2	0.0410	4	5
NN4	0.0377	8	0.0397	9	0.0394	1	6
RF	0.0346	3	0.0393	6	0.0467	9	6
NN5	0.0367	7	0.0386	5	0.0415	6	6
NN	0.0365	6	0.0394	8	0.0426	8	9

Bold indicates top three models.
*MSE-36, MSE-5, and MSE-total correspond to the MSE results of 36-item, 5-domain, and total-score inputs. See Table 2 for details.

Table 8 provides the MAE-based comparison. SVR remains the most stable performer, while multiple neural network variants (particularly the shallower configurations such as NN4/NN3/NN1) achieve competitive ranks on specific input settings, indicating that simpler network structures tend to generalize better in this dataset.

To provide an overall comparison, we computed a composite ranking by averaging within-setting ranks across six metrics (MSE and MAE for 36 items, 5 domains, and total score), where lower values receive better ranks and ties are averaged.

First, we compare machine learning algorithms for mapping FACT-L to EQ-5D-5L and find that, with a sample size of 347, SVR achieves the best overall performance while neural networks still show clear potential. Second, we identify a best-fitting hyperparameter pair for each algorithm via grid search; because our objective balances

Table 8 MAE-based model ranking across input settings

Model	MAE-36*	rank	MAE-5*	rank	MAE-total*	rank	Total rank
SVR	0.1357	1	0.1410	1	0.1488	2	1
NN4	0.1406	5	0.1474	3	0.1480	1	2
NN3	0.1407	6	0.1450	2	0.1511	4	3
XGBoost	0.1357	1	0.1486	6	0.1561	7	4
NN1	0.1391	4	0.1476	4	0.1559	6	4
NN2	0.1429	8	0.1481	5	0.1490	3	6
NN	0.1421	7	0.1488	7	0.1536	5	7
RF	0.1385	3	0.1492	8	0.1669	9	8
NN5	0.1436	9	0.1543	9	0.1598	8	9

Bold indicates top three models.
*MAE-36, MAE-5, and MAE-total correspond to the MAE results of 36-item, 5-domain, and total-score inputs. See Table 2 for details.

Table 9 Overall performance ranking

Model	Mean MSE†	rank	Mean MAE†	rank	Avg. rank	Overall rank
SVR	0.0372	1	0.1418	1	1.92	1
NN3	0.0378	3	0.1456	3	3.67	2
XGBoost	0.0376	2	0.1468	5	4.33	3
NN4	0.0389	5	0.1453	2	4.50	4
NN1	0.0385	4	0.1475	6	4.75	5
NN2	0.0389	6	0.1467	4	5.25	6
Random forest	0.0402	9	0.1515	8	6.42	7
NN	0.0395	8	0.1482	7	6.83	8
NN5	0.0389	6	0.1526	9	7.33	9

Bold indicates top three models.
†Mean values are computed by averaging results from 36-, 5-, and total-feature input settings.

MSE and MAE, we do not further optimize on a single metric, and a deeper study of hyperparameter sensitivity is left for future work. Third, we evaluate how the number of input dimensions affects performance and observe that higher-dimensional inputs generally improve accuracy; a possible next step is to analyze inter-dimension correlations to determine when added dimensions are most beneficial. Fourth, we study the impact of neural network depth and width (model complexity) and conclude that an appropriately complex architecture is required for a given dataset; when data are limited, shallower networks outperform deeper ones.

To provide a unified visual comparison across metrics and model families, Fig. 3 presents radar plots summarizing model performance after metric normalization. The left panel contrasts traditional machine learning models, showing that SVR consistently dominates most axes, while XGBoost remains competitive and RF performs relatively weaker on several error metrics. The middle panel compares neural network configurations, illustrating that model performance varies substantially with

architecture depth and width, supporting the observation that network complexity must be matched to dataset size. The right panel highlights the two best-performing models from each family (SVR, XGBoost, NN3, and NN1). This comparison shows that although SVR maintains the strongest overall performance, the best neural network variants reach comparable performance on multiple metrics, especially under the total-score setting. These findings indicate that neural networks already approach the performance of traditional machine learning models under appropriate architectural choices, despite limited sample size.

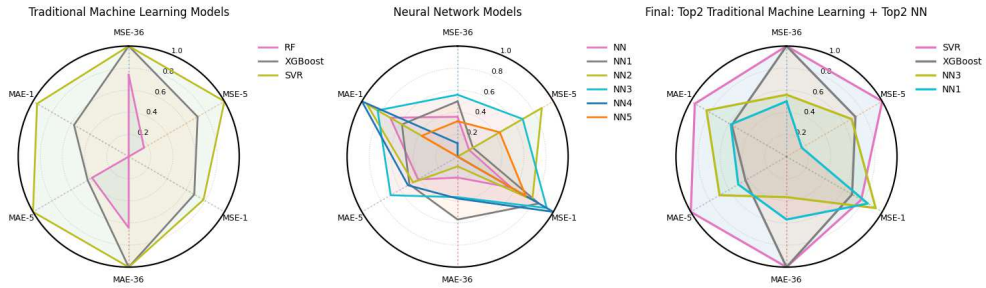


Fig. 3 Radar plots comparing model performance. The left panel contrasts traditional models (RF, XGBoost, SVR) across MSE and MAE using 36 items (MSE-36, MAE-36), five dimensions (MSE-5, MAE-5), and total score (MSE-Total, MAE-Total). The middle panel compares neural network variants (NN, NN1–NN5) on the same metrics. The right panel highlights the best two traditional models (SVR, XGBoost) and the best two neural networks (NN3, NN1), showing that SVR consistently performs strongest overall while the top neural networks are competitive on several metrics. For visualization, all error metrics were transformed into normalized scores using global min–max scaling (across all models) with inversion so that higher values indicate better performance.

5 Discussion

This study provides a systematic comparison between traditional machine learning methods and neural network models for mapping FACT-L scores to EQ-5D-5L utilities. The use of mapping algorithms has supported the conduct of CUAs [40]. We makes four contributions: (1) compares the performance of machine learning algorithms for mapping FACT-L to EQ-5D-5L; (2) identifies hyperparameter settings that best fit our dataset; (3) examines the impact of input dimensionality on model performance; and (4) compares neural network architectures for utility mapping in a moderate-sized clinical dataset.

The results suggest that traditional machine learning models, particularly SVR, remain strong baselines under moderate sample sizes, mainly due to their robustness and ability to generalize well in relatively low-data regimes [41]. In contrast, neural networks exhibit higher sensitivity to architectural design and training settings, which explains the larger performance variability observed across different NN configurations [42].

The radar-based analysis offers additional insight beyond single-metric comparisons. While SVR model achieves the most balanced performance across all evaluation dimensions in the moderate dataset, the best neural networks demonstrate competitive performance on several metrics, especially when using aggregated or total-score inputs. Previous studies have explored the application of neural networks in utility mapping; however, our work extends these efforts by providing a more systematic evaluation of neural network architectures and model complexity under different input configurations [14]. This suggests that neural networks may better capture nonlinear interactions among questionnaire dimensions when the architecture is appropriately constrained [42]. However, overly deep or overly wide networks tend to underperform, likely due to overfitting and insufficient training data, highlighting the importance of aligning model complexity with dataset scale [43].

At present, there is one retrospective study on mapping FACT-L to EQ-5D-5L and one study on mapping FACT-L to EQ-5D-3L among Chinese patients with lung cancer [20, 23]. However, previous studies lack exploration of modern machine learning approaches. From a practical perspective, our findings imply that traditional models may currently be preferable in clinical utility mapping tasks with limited sample sizes, owing to their stability and reduced tuning requirements. Nonetheless, the rapid improvement observed among optimized neural networks indicates strong potential for future applications as larger datasets become available.

We also investigated the correlation between EQ-5D-5L dimensions and FACT-L domains, and multiple correlation coefficients were found to be statistically significant (P ¡ 0.001). These findings further support the validity of the mapping approach, as previous mapping studies have suggested that significant correlations between scales provide evidence for valid mapping between different measurement systems [44, 45].

Several limitations should be acknowledged. First, the sample size remains relatively small for fully exploiting deep learning approaches. Second, hyperparameter tuning focused on a balanced MSE–MAE objective, and alternative optimization targets may yield different model rankings. Third, the analysis did not explicitly examine feature interactions or dimensional correlations, which may influence both model selection and interpretability [11, 23, 46].

This study was based on a single-center cross-sectional dataset. Future work should investigate larger and more diverse datasets, explore regularization strategies and architecture search for neural networks, and analyze inter-dimension relationships within FACT-L to better understand when increased input dimensionality provides meaningful performance gains. In addition, incorporating uncertainty estimation and external validation could further strengthen model applicability in real-world clinical settings.

6 Conclusion

Using data obtained from patients undergoing lung cancer surgery, we systematically compared traditional machine learning models and neural networks for mapping FACT-L to EQ-5D-5L utility scores. In health economic evaluations, the results show

that traditional methods, particularly SVR, achieve the most stable overall performance under a moderate sample size, while neural networks demonstrate competitive results when model complexity is properly controlled. Our findings suggest that model performance depends strongly on both input dimensionality and architectural design, highlighting the need to balance model capacity with available data. Overall, traditional models currently provide the most reliable solution for small-to-medium datasets, whereas neural networks remain a promising direction as larger datasets and improved optimization strategies become available.

Abbreviations

CUAs	cost-utility analyses
EQ-5D	EuroQol five-dimension questionnaire
EWB	emotional well-being
FACT-L	Functional Assessment of Cancer Therapy–Lung
FWB	functional well-being
HRQoL	health-related quality of life
LCS	Lung Cancer Subscale
MAE	mean absolute error
MSE	mean squared error
NN	neural networks
PBMs	preference-based measures
PWB	physical well-being
QALYs	quality-adjusted life years
RBF	radial basis function
RF	Random forest
SD	standard deviation
SVR	support vector regression
SWB	social/family well-being
TOI	Trial Outcome Index
XGBoost	extreme gradient boosting

Acknowledgements. The authors thank all participants for their time and effort.

Author contributions. HLT, LLJ, PWY conceived and designed this study; HLT participated in the acquisition of the data; HLT analyzed the data, interpreted the results, and wrote the first draft of the manuscript;and LLJ supervised the statistical analysis, gave feedback on the manuscript, and revised it critically for important intellectual content. All authors have read and approved the final version of the manuscript.

Data availability. The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest. The authors declare no conflicts of interest.

Ethical approval. This study was approved by the Ethics Committee of West China Tianfu Hospital, Sichuan University (Ethics approval No. MR-51-25-042281). The study followed the principle of the Declaration of Helsinki.

Informed consent. Informed consent was obtained from all individual participants included in the study.

References

[1] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians **71**(3), 209–249 (2021) https://doi.org/10.3322/caac.21660

[2] Xia, C., Dong, X., Li, H., Cao, M., Sun, D., He, S., Yang, F., Yan, X., Zhang, S., Li, N., *et al.*: Cancer statistics in china and united states, 2022: profiles, trends, and determinants. Chinese medical journal **135**(05), 584–590 (2022) https://doi.org/10.1097/CM9.0000000000002108

[3] Cella, D.F., Bonomi, A.E., Lloyd, S.R., Tulsky, D.S., Kaplan, E., Bonomi, P.: Reliability and validity of the functional assessment of cancer therapy—lung (fact-l) quality of life instrument. Lung Cancer **12**(3), 199–220 (1995) https://doi.org/10.1016/0169-5002(95)00450-F

[4] Wan, C., Zhang, C., Cai, L., Tu, X., Feng, C., Luo, J.-h., Zhang, X.: Psychometric properties of the chinese version of the fact-l for measuring quality of life in patients with lung cancer. Lung cancer **56 3**, 415–21 (2007) https://doi.org/10.1016/j.lungcan.2007.01.004

[5] Vilhena, E., Pais-Ribeiro, J., Silva, I., Pedro, L., Meneses, R., Cardoso, H., Silva, A., Mendonça, D.: Psychosocial factors as predictors of quality of life in chronic portuguese patients. Health and quality of life outcomes **12**, 3 (2014) https://doi.org/10.1186/1477-7525-12-3

[6] Whitehead, S.J., Ali, S.: Health outcomes in economic evaluation: the qaly and utilities. British medical bulletin **96**, 5–21 (2010) https://doi.org/10.1093/bmb/ldq033

[7] Herdman, M., Gudex, C., Lloyd, A., Janssen, M.F., Kind, P., Parkin, D., Bonsel, G.J., Badia, X.: Development and preliminary testing of the new five-level version of eq-5d (eq-5d-5l). Quality of Life Research **20**, 1727–1736 (2011) https://doi.org/10.1007/s11136-011-9903-x

[8] Su, Q., Peng, X., Zhang, Y., Sun, L.: Mapping to estimate health state utility: a systematic review of development and limitations. Health and Quality of Life Outcomes **23** (2025) https://doi.org/10.1186/s12955-025-02426-3

[9] Teckle, P., McTaggart-Cowan, H., Hoek, K., Chia, S.K.L., Melosky, B.L., Gelmon, K.A., Peacock, S.J.: Mapping the fact-g cancer-specific quality of life instrument to the eq-5d and sf-6d. Health and Quality of Life Outcomes **11**, 203–203 (2013) https://doi.org/10.1186/1477-7525-11-203

[10] Meregaglia, M., Borsoi, L., Cairns, J.A., Tarricone, R.: Mapping health-related quality of life scores from fact-g, faact, and facit-f onto preference-based eq-5d-5l utilities in non-small cell lung cancer cachexia. The European Journal of Health Economics **20**, 181–193 (2017) https://doi.org/10.1007/s10198-017-0930-6

[11] Longworth, L., Rowen, D.: Mapping to obtain eq-5d utility values for use in nice health technology assessments. Value in Health **16**(1), 202–210 (2013) https://doi.org/10.1016/j.jval.2012.10.010

[12] Brazier, J., Yang, Y., Tsuchiya, A., Rowen, D.: A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. The European Journal of Health Economics **11**(2), 215–225 (2010) https://doi.org/10.1007/s10198-009-0168-z

[13] Wailoo, A.J., Hernandez-Alava, M., Manca, A., Mejia, A., Ray, J., Crawford, B., Botteman, M., Busschbach, J.: Mapping to estimate health-state utility from non–preference-based outcome measures: An ispor good practices for outcomes research task force report. Value in Health **20**(1), 18–27 (2017) https://doi.org/10.1016/j.jval.2016.11.006

[14] Gao, L., Luo, W., Tonmukayakul, U., Moodie, M., Chen, G.: Mapping macnew heart disease quality of life questionnaire onto country-specific eq-5d-5l utility scores: a comparison of traditional regression models with a machine learning technique. The European Journal of Health Economics **22**, 341–350 (2021) https://doi.org/10.1007/s10198-020-01259-9

[15] Hagiwara, Y., Shiroiwa, T., Taira, N., Kawahara, T., Konomura, K., Noto, S., Fukuda, T., Shimozuma, K.: Gradient boosted tree approaches for mapping european organization for research and treatment of cancer quality of life questionnaire core 30 onto 5-level version of eq-5d index for patients with cancer. Value in Health **26**(2), 269–279 (2023) https://doi.org/10.1016/j.jval.2022.07.020

[16] Aghdaee, M., Parkinson, B., Sinha, K., Gu, Y., Sharma, R., Olin, E., Cutler, H.: An examination of machine learning to map nonreference based patient reported outcome measures to health state utility values. Health Economics **31**(8), 1525–1557 (2022) https://doi.org/10.1002/hec.4503

[17] Schmidhuber, J.: Deep learning in neural networks: An overview. Neural Networks **61**, 85–117 (2015) https://doi.org/10.1016/j.neunet.2014.09.003

[18] Kind, P., Macran, S.: Eliciting social preference weights for functional assessment of cancer therapy-lung health states. PharmacoEconomics **23**, 1143–1153 (2012)

https://doi.org/10.2165/00019053-200523110-00006

[19] Pickard, S., Ray, S., Ganguli, A., Cella, D.: Comparison of fact- and eq-5d-based utility scores in cancer. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research **15 2**, 305–11 (2012) https://doi.org/10.1016/j.jval.2011.11.029

[20] Xi-Jing, F.U., Min-Hong, L., Mao, S.: Analysis on applying mapping method in measuring health utility values. Chinese Health Economics **33**(10), 32–25 (2014)

[21] Zhu, J., Yan, X., Liu, C., Wang, H., Wang, L., Cao, S.-m., Liao, X.-Z., Xi, Y., Ji, Y., Lei, L., Xiao, H., Guan, H.J., Wei, W., Dai, M., Chen, W., Shi, J.: Comparing eq-5d-3l and eq-5d-5l performance in common cancers: suggestions for instrument choosing. Quality of Life Research **30**, 841–854 (2020) https://doi.org/10.1007/s11136-020-02636-w

[22] van Hout, B., Janssen, M.F., Feng, Y.-S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P., Pickard, A.S.: Interim scoring for the eq-5d-5l: Mapping the eq-5d-5l to eq-5d-3l value sets. Value in Health **15**(5), 708–715 (2012) https://doi.org/10.1016/j.jval.2012.02.008

[23] Jiang, L., Zhou, H., Yang, Q., Luo, X., Huang, D.: Development of algorithms to estimate the eq-5d-5l from the fact-l in patients with lung cancer: a mapping study. Quality of Life Research **33**, 805–816 (2023) https://doi.org/10.1007/s11136-023-03567-y

[24] Petrou, S., Rivero-Arias, O., Dakin, H.A., Longworth, L., Oppe, M., Froud, R., Gray, A.M.: The maps reporting statement for studies mapping onto generic preference-based outcome measures: Explanation and elaboration. PharmacoEconomics **33**, 993–1011 (2015) https://doi.org/10.1007/s40273-015-0312-9

[25] Cella, D., Eton, D.T., Fairclough, D.L., Bonomi, P., Heyes, A.E., Silberman, C., Wolf, M.K., Johnson, D.H.: What is a clinically meaningful change on the functional assessment of cancer therapy–lung (fact-l) questionnaire?: Results from eastern cooperative oncology group (ecog) study 5592. Journal of Clinical Epidemiology **55**(3), 285–295 (2002) https://doi.org/10.1016/S0895-4356(01)00477-2

[26] Luo, N., Liu, G., Li, M., Guan, H., Jin, X., Rand-Hendriksen, K.: Estimating an eq-5d-5l value set for china. Value in health **20**(4), 662–669 (2017) https://doi.org/10.1016/j.jval.2016.11.016

[27] Hagiwara, Y., Shiroiwa, T., Taira, N., Kawahara, T., Konomura, K., Noto, S., Fukuda, T., Shimozuma, K.: Mapping eortc qlq-c30 and fact-g onto eq-5d-5l index for patients with cancer. Health and Quality of Life Outcomes **18**(354) (2020) https://doi.org/10.1186/s12955-020-01611-w

[28] Wuensch, K.L., Evans, J.D.: Straightforward statistics for the behavioral sciences. Journal of the American Statistical Association **91**(436) (1996) https://doi.org/10.2307/2291607

[29] Brazier, J., Rowen, D., Mavranezouli, I., Tsuchiya, A., Young, T., Yang, Y., Barkham, M., Ibbotson, R.: Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health Technology Assessment **16**(32) (2012) https://doi.org/10.2307/2291607

[30] Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001) https://doi.org/10.1023/A:1010933404324

[31] Smola, A., Scholkopf, B.: A tutorial on support vector regression. Statistics and Computing **14**, 199–222 (2004) https://doi.org/10.1023/B:STCO.0000035301.49549.88

[32] Orhan, A.E., Ma, W., Ma, W.: Efficient probabilistic inference in generic neural networks trained with non-probabilistic feedback. Nature Communications **8** (2016) https://doi.org/10.1038/s41467-017-00181-8

[33] Feurer, M., Hutter, F.: Hyperparameter optimization. In: Automated Machine Learning (2019). https://doi.org/10.1007/978-3-030-05318-5_1

[34] Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. Journal of Machine Learning Research **13**(10), 281–305 (2012) https://doi.org/10.5555/2188385.2188395

[35] Nasrabadi, N.M.: Pattern recognition and machine learning. Journal of Electronic Imaging **16**(4), 049901 (2007) https://doi.org/10.1117/1.2819119

[36] Elsken, T., Metzen, J.H., Hutter, F.: Neural architecture search: A survey. Journal of Machine Learning Research **20**(55), 1–21 (2019) https://doi.org/10.1007/978-3-030-05318-5_3

[37] Girshick, R.: Fast r-cnn. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). ICCV '15, pp. 1440–1448. IEEE Computer Society, USA (2015). https://doi.org/10.1109/ICCV.2015.169

[38] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2017). https://doi.org/10.48550/arXiv.1711.05101

[39] Chen, P., Wan, G., Zhu, B.: Incidence and risk factors of symptomatic thrombosis related to peripherally inserted central catheter in patients with lung cancer. Journal of Advanced Nursing **77**(3), 1284–1292 (2021) https://doi.org/10.1111/jan.14666

[40] Hagiwara, Y., Kawahara, T., Shiroiwa, T.: What is a valid mapping algorithm in cost-utility analyses? a response from a missing data perspective. Value in Health **23**(9), 1218–1224 (2020) https://doi.org/10.1016/j.jval.2020.03.020

[41] Roy, A., Manna, R., Chakraborty, S.: Support vector regression based metamodeling for structural reliability analysis. Probabilistic Engineering Mechanics **55**, 78–89 (2019) https://doi.org/10.1016/j.probengmech.2018.11.001

[42] Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, ??? (2016). http://www.deeplearningbook.org

[43] Margeloiu, A., Simidjievski, N., Lio', P., Jamnik, M.: Weight predictor network with feature selection for small sample tabular biomedical data. In: AAAI Conference on Artificial Intelligence (2022). https://doi.org/10.1609/aaai.v37i8.26090

[44] Ali, F.M., Kay, R., Finlay, A.Y., Piguet, V., Kupfer, J., Dalgard, F., Salek, M.S.: Mapping of the dlqi scores to eq-5d utility values using ordinal logistic regression. Quality of Life Research **26**(11), 3025–3034 (2017) https://doi.org/10.1007/s11136-017-1607-4

[45] Ameri, H., Yousefi, M., Yaseri, M., Nahvijou, A., Arab, M., Sari, A.A.: Mapping the cancer-specific qlq-c30 onto the generic eq-5d-5l and sf-6d in colorectal cancer patients. Expert Review of Pharmacoeconomics & Outcomes Research **19**, 89–96 (2018) https://doi.org/10.1080/14737167.2018.1517046

[46] Cheung, Y., Luo, N., Ng, R., Lee, C.: Mapping the functional assessment of cancer therapy - breast (fact-b) to the 5-level euroqol group's 5-dimension questionnaire (eq-5d-5l) utility index in a multi-ethnic asian population. Health and quality of life outcomes **12**, 180 (2014) https://doi.org/10.1186/s12955-014-0180-6