Check for updates

A hybrid self-supervised model predicting life satisfaction in South Korea

Hung Viet Nguyen and Haewon Byeon*

Department of Digital Anti-Aging Healthcare (BK21), Inje University, Gimhae, Republic of Korea

Objective: Life satisfaction pertains to an individual's subjective evaluation of their life quality, grounded in their personal criteria. It stands as a crucial cognitive aspect of subjective wellbeing, offering a reliable gauge of a person's comprehensive wellbeing status. In this research, our objective is to develop a hybrid self-supervised model tailored for predicting individuals' life satisfaction in South Korea.

Methods: We employed the Busan Metropolitan City Social Survey Data in 2021, a comprehensive dataset compiled by the Big Data Statistics Division of Busan Metropolitan City. After preprocessing, our analysis focused on a total of 32,390 individuals with 51 variables. We developed the self-supervised pre-training TabNet model as a key component of this study. In addition, we integrated the proposed model with the Local Interpretable Model-agnostic Explanation (LIME) technique to enhance the ease and intuitiveness of interpreting local model behavior.

Results: The performance of our advanced model surpassed conventional tree-based ML models, registering an AUC of 0.7778 for the training set and 0.7757 for the test set. Furthermore, our integrated model simplifies and clarifies the interpretation of local model actions, effectively navigating past the intricate nuances of TabNet's standard explanatory mechanisms.

Conclusion: Our proposed model offers a transparent understanding of AI decisions, making it a valuable tool for professionals in the social sciences and psychology, even if they lack expertise in data analytics.

KEYWORDS

explainable AI, hybrid model, life satisfaction, self-supervised, TabNet

1 Introduction

Life satisfaction pertains to an individual's subjective evaluation of their life quality, grounded in their personal criteria. This comprehensive evaluation covers multiple areas such as family life, career, and social interactions (1). It stands as an essential cognitive component of subjective wellbeing, offering a reliable gauge of a person's comprehensive wellbeing status (2). Notably, it exerts a significant influence on various facets of an individual's mental wellbeing (2). A study conducted by Lewis et al. (3) has demonstrated an inverse relationship between life satisfaction and depressed symptoms, with enhancements in life satisfaction yielding a reduction in such symptoms. Similarly, Fergusson et al. (4) have documented the consequential role of life satisfaction in shaping an individual's mental wellbeing. Beyond its individual-level implications, life satisfaction also serves as a pivotal metric for assessing the quality of life within a given society (2). The research by Wong et al. (5) based on empirical data, has underscored the differential impacts of social policies on

individuals' life satisfaction. Consequently, the development of an effective predictive model for assessing life satisfaction among individuals carries significant relevance in the realms of mental health research and practice, as well as in formulating policies aimed at enhancing the wellbeing of the broader public.

In the past, numerous scholars have undertaken extensive investigations into the determinants of individuals' life satisfaction. However, most of these studies have relied upon conventional statistical techniques like regression or mediation analysis, often incorporating control variables in their analyses (6–11). These methods often rely on simplified relationships between variables and may lack the predictive power required to model the complexity of life satisfaction outcome. In contrast, contemporary advancements in computational methodologies, particularly machine learning (ML) and deep learning (DL), present new opportunities for enhancing our understanding of mental health outcomes on an individual basis (12). More recent studies (2, 13) have employed ML models to predict life satisfaction; nonetheless, the application of DL models, capable of revealing intricate patterns within large datasets, is yet inadequately investigated in this domain. Moreover, a notable constraint of deep learning models is their "black box" characteristic, which complicates the interpretation of results and the comprehension of the underlying causes influencing predictions.

To address those challenges, our research seeks to answer these three key research questions:

- Question 1: How effectively can a DL model, specifically the self-supervised pretraining TabNet (SSP-TabNet) model (14), predict life satisfaction from large-scale social survey data compared to traditional supervised ML models?
- Question 2: What are the most significant predictors of life satisfaction, and how can we interpret the model's decision-making process in a manner that is accessible to non-technical users?
- Question 3: How does the integration of the Local Interpretable Model-agnostic Explanation (LIME) (15) interpretability framework enhance the explanation of local model behaviors, particularly in the context of predicting life satisfaction?

In the context of the rapidly evolving field of artificial intelligence, recent advancements have highlighted the efficacy of self-supervised learning in acquiring valuable data representations (16). Notably, this success has primarily been evident in data modalities like images (17, 18), audio (19), and text (20, 21). This achievement hinges on the exploitation of inherent spatial, temporal, or semantic structures within the data (22, 23). However, when it comes to tabular datasets, frequently employed in domains such as healthcare, such structural characteristics are often limited or absent. In recent, Arik and Pfister (14) presented a groundbreaking DL architecture known as the TabNet model, specifically designed for handling tabular data. In addition to its supervised architecture, TabNet was the first use of self-supervised pre-training technique to tabular data, resulting in significant performance improvements. The TabNet model has laid the foundation for further exploration and advancements in leveraging self-supervised learning technique within tabular data applications, particularly in vital fields like healthcare.

In this research endeavor, our primary objective is to develop a self-supervised pre-training methodology tailored for the prediction of individuals' life satisfaction in the South Korean context. We employed the SSP-TabNet model as a key component of this approach. A notable advantage of SSP-TabNet is that it eliminates the need for feature selection, a critical step in traditional supervised machine learning models. Additionally, SSP-TabNet provides strong interpretability, enabling both localized and global insights into the decision-making process. Nonetheless, the localized interpretability of TabNet, reliant on access to its decision masks, is not easy to understand for professionals within fields such as social work and psychology, who may not possess advanced proficiency in data analysis techniques (24). To address this issue, we have integrated the SSP-TabNet model with the Local Interpretable Model-agnostic Explanation (LIME) (15) technique. This hybrid model enhances the ease and intuitiveness of interpreting local model behavior, surpassing the inherent complexities of SSP-TabNet's native interpretation features, thus making it more accessible to a broader audience.

2 Materials and methods

2.1 Dataset

In our research, we employed the Busan Metropolitan City Social Survey Data in 2021, a comprehensive dataset compiled by the Big Data Statistics Division of Busan Metropolitan City. This dataset provides a comprehensive understanding of the living conditions and civic engagement levels of the city's residents. It is specifically designed to assess the quality of civic life and overall welfare, forming the foundational data for policymaking and community development initiatives in Busan. The survey protocol received ethical approval (IRB approval no. 17339) from Statistics Korea, ensuring adherence to ethical research standards.

The survey encompassed all residents aged 15 years and above within the geographic boundaries of Busan Metropolitan City, with the sampling frame derived from the 2019 Population and Housing Census - a nationwide survey. Employing a probability proportional systematic sampling approach, the survey drew a final sample of 17,860 households (comprising 940 survey districts, with 19 households per district) residing in Busan Metropolitan City at the time of the survey administration. Our analysis, subsequently, focused on a total of 32,390 individuals who successfully completed the survey.

2.2 Data preprocessing

The original dataset consisted of 32,390 samples with 132 features. To prepare the dataset for model training, we applied several preprocessing steps, which are outlined below:

1. Removal of redundant columns: we first eliminated two columns containing superfluous serial number information, as they had no relevance to the analysis.
2. Handling missing data: given that the survey allowed for optional responses, some columns contained a high percentage of missing data. We adopted a threshold-based approach to exclude columns with a significant proportion of missing values. Specifically, any features where more than 50% of the values were missing were removed. This decision was based on

the assumption that such features would not provide reliable insights or sufficient information for model training. The excluded features were primarily non-critical or supplementary variables, which were not essential for achieving our research objectives. After removing columns with more than 50% missing values, the dataset included 32,390 samples and 51 variables, including the target feature. A detailed overview of these 51 variables is provided in Supplementary Table S1.

3   Target feature definition: the target feature for this study was defined as "life satisfaction during the COVID-19 pandemic," measured on a 10-point scale from 1 (least satisfied) to 10 (most satisfied). To simplify the classification task, we recategorized the target feature into two classes: scores of 5 or below were classified as "dissatisfied" (class 1), and scores of 6 or above were classified as "satisfied" (class 0). This recategorization yielded 16,906 samples in class 0 (satisfied) and 16,294 samples in class 1 (dissatisfied), as illustrated in Figure 1.

4   Data splitting: the dataset was randomly stratified into 80% for training (25,912 samples) and 20% for testing (6,478 samples), ensuring that both classes were equally represented in each set. The training set facilitated model development and hyperparameter optimization, while the test set assessed the model's efficacy on unseen data.

In our study, after removing columns with more than 50% missing values, we did not apply any additional feature selection methods to further diminish the dimensionality of the variables. The reason is because SSP-TabNet's architecture inherently performs automatic feature selection through its Mask layer, unlike conventional machine learning models, which often require a manual feature selection process. The Mask layer in SSP-TabNet identifies the most relevant features for each decision step during training, thereby eliminating the need for manual feature selection.

2.3 Development of self-supervised

TabNet, introduced by Arik et al. (14), builds upon the end-to-end retraining and representation learning feature characteristics inherent to deep neural networks (DNN). Notably, TabNet combines these traits with the interpretability and sparse feature selection capabilities often associated with tree-based models. In the realm of real-world dataset analysis, TabNet was demonstrated that outperformed traditional ML algorithms, highlighting its proficiency in delivering enhanced accuracy. Another notable advantage of TabNet is its elimination of the need for feature selection process by its Mask layer. Furthermore, TabNet provides interpretability by identifying the most critical attributes for each sample. This attribute selection process aids in understanding the model's decision-making rationale. In recent studies, TabNet has found applications across diverse domains, including healthcare (24), fraud detection (25), and energy management (26), showcasing its versatility and utility in a range of practical contexts.

The self-supervised pretraining TabNet (SSP-TabNet) architecture comprises two essential elements: a TabNet encoder and a TabNet decoder module. Central to this self-supervised learning approach is the acknowledgment of the intrinsic interconnections among various features within a single data sample (16). The approach initiates with the strategic masking of certain features. Subsequently, the encoder-decoder framework is utilized to predict these concealed features. This process effectively
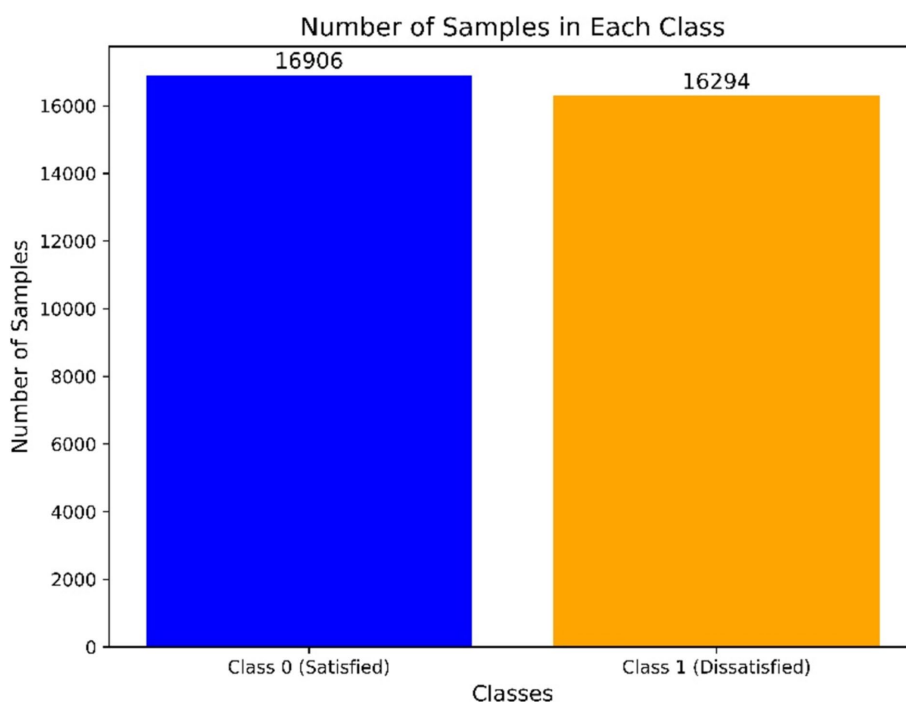

FIGURE 1
Target distribution.

equips the TabNet encoder module with the capacity to aptly characterize the distinctive features of each sample. This approach accelerates model convergence and augments overall model performance, enhancing its capacity to identify complex relationships within the data. In summary, we selected SSP-TabNet as the primary model due to its ability to handle tabular data effectively, its automatic feature selection capabilities, and its interpretability, which are particularly important in the context of social survey data.

2.3.1 TabNet encoder structure

The TabNet encoder structure, as illustrated in Figure 2, is composed of sequential multi-steps (N_{steps}). At each step i^{th}, it takes in refined data from the previous stage, denoted as the $(i-1)^{th}$, to determines the relevant features to employ. Subsequently, it generates processed feature representations, which are collectively integrated to inform the overall decision-making process. The model is designed to process datasets with a defined batch size (B) and features of D-dimensions, and it operates independently of global feature normalization. Prior to entering the Feature Transformer, the data is subject to batch normalization (BN).

The Feature Transformer, illustrated in Figure 3, encompasses a series of n distinct gated linear unit (GLU) blocks. Within each Feature Transformer block comprises three key layers: fully connected (FC), BN, and GLU. In configurations utilizing four GLU blocks, two are designed to function in tandem, while the remaining two operate independently. This design promotes efficient learning. Notably, a skip connection links consecutive blocks, and a normalization step with a factor of $\sqrt{0.5}$ follows each block to ensure stability, preventing significant variance fluctuations (27). Upon processing the features that have undergone batch normalization, the Feature Transformer conveys this refined information to the Attentive Transformer at the ith step via a split layer.

The Attentive Transformer (Figure 4) consists of four crucial layers: FC, BN, Prior Scales, and Sparsemax. The input sourced from the split layer is first processed via the FC and BN layers. Subsequently, it is directed to the Prior Scales layer, where the importance of features relevant to the present decision-making step is aggregated, as described in Equation 1:

$$P[i] = \prod_{j=1}^{i} (\gamma - M[j]) \qquad (1)$$

where γ is the relaxation parameter.

The primary role of an Attentive Transformer is to compute the Mask layer for the current step, building upon the results from the preceding step. This learnable Mask facilitates the sparse selection of the most critical features, thus improving the model's efficiency in terms of parameter utilization. By focusing the learning capabilities of each decision step on relevant features, this method minimizes the resources expended on non-essential elements. Importantly, the Attentive Transformer plays a crucial role in deriving masks from the processed features obtained in the preceding step, due to the multiplicative nature of the masking procedure, as demonstrated in Equation 2:

$$M[i] = sparsemax(P[i-1]) * h_i(a[i-1]) \qquad (2)$$

where h_i represents the trainable function used to represent the FC and BN layers, $P[i-1]$ denotes the preceding scales item, and the sparsemax layer is employed for coefficient normalizing leading to sparse feature selection.

The Mask layer is subsequently directed to the Feature Transformer, where an analysis of the filtered features occurs. Within this transformation process, the data is segregated into two distinct outputs, as presented in Equation 3:
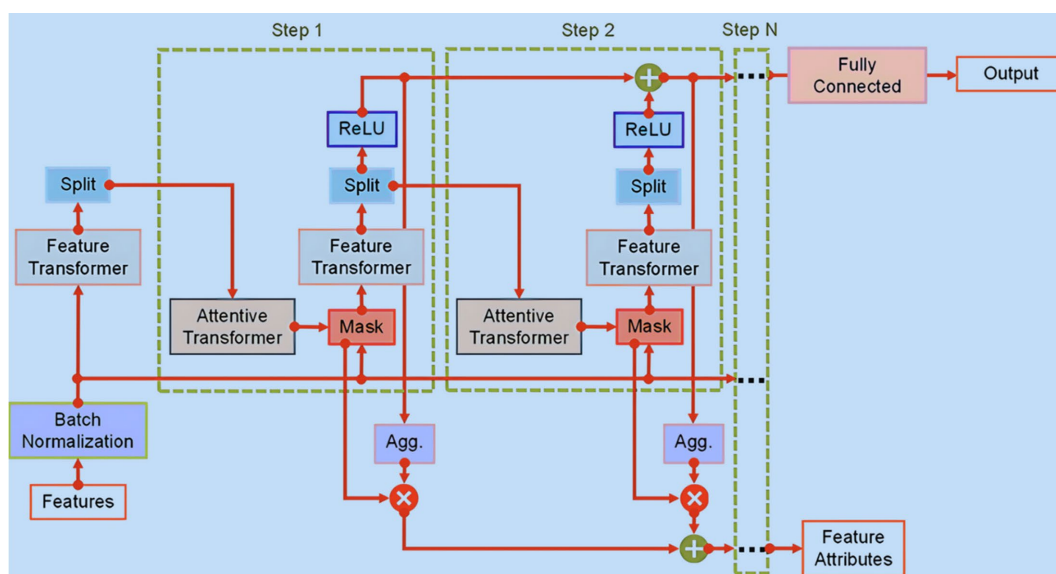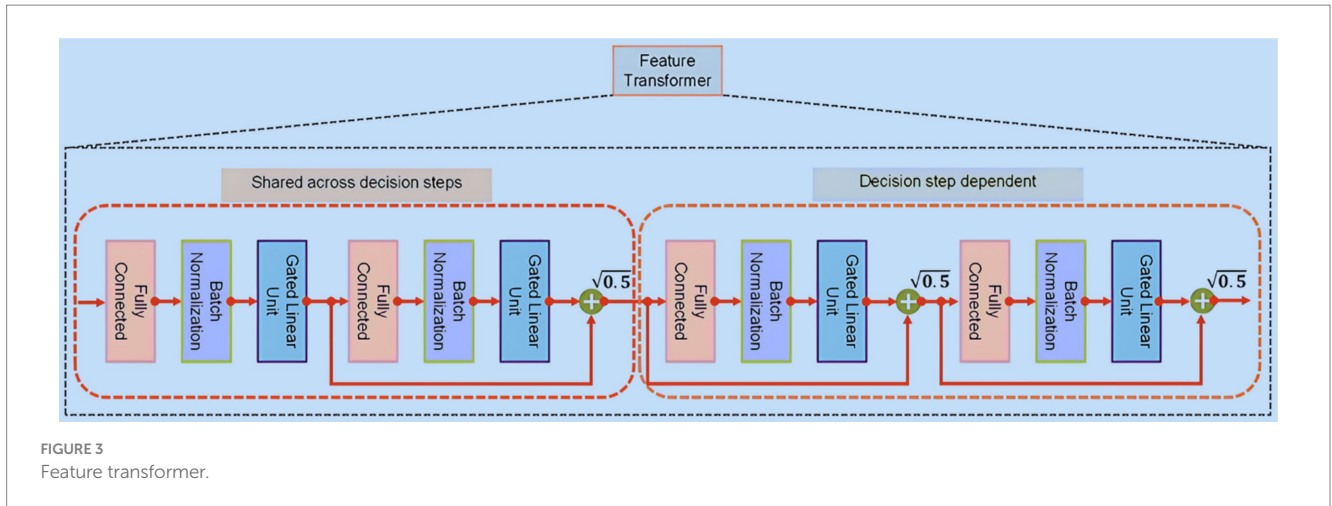

FIGURE 2
TabNet encoder structure.

FIGURE 3
Feature transformer.

$$\big[d[i],a[i]\big] = f_i\big(M[i]*f\big) \tag{3}$$

where $a[i] \in \mathbb{R}^{B*N_a}$ represents the data utilized in the subsequent stage of the Attentive Transformer, and $d[i] \in \mathbb{R}^{B*N_d}$ denotes the outcome generated by the decision step.

Drawing inspiration from the concept of aggregating tree models, the output vectors $d[i]$ generated from all the decision steps were consolidated into a unified vector, represented as d_{out}. Subsequently, a FC layer is employed to map this final aggregated output. This aggregation process is defined as shown in Equation 4:

$$d_{out} = \sum_{i=1}^{N_{steps}} ReLU\big(d[i]\big) \tag{4}$$

2.3.2 TabNet decoder structure

To fulfill the self-supervised learning objective of TabNet model, an accompanying decoder architecture (Figure 5) has been introduced to facilitate the reconstruction of masked features from the encoded phase. In this reconstruction process, a binary mask, labeled as S (with $S \in \{0,1\}^{B \times D}$), is used. Within this framework, both the FC layer and Feature Transformer layer work collaboratively to predict the masked feature at each step. This prediction is guided by the minimization of the reconstruction loss, as expressed in Equation 5:

$$\sum_{b=1}^{B}\sum_{j=1}^{D}\left| \frac{\left(\hat{f}_{b,j} - f_{b,j}\right).S_{b,j}}{\sqrt{\sum_{b=1}^{B}\left(f_{b,j} - 1/B\sum_{b=1}^{B}f_{b,j}\right)^2}} \right| \tag{5}$$

where $\hat{f}_{b,j}$ and $f_{b,j}$ respectively represent the expected feature importance score and the feature importance score attributed to the jth feature within the bth sample.

In our study, we harnessed the pytorch_tabnet package version 4.1 to craft the self-supervised pre-training TabNet (SSP-TabNet) model. The overarching objective of the self-supervised pre-training phase involves TabNet mapping the embedding space
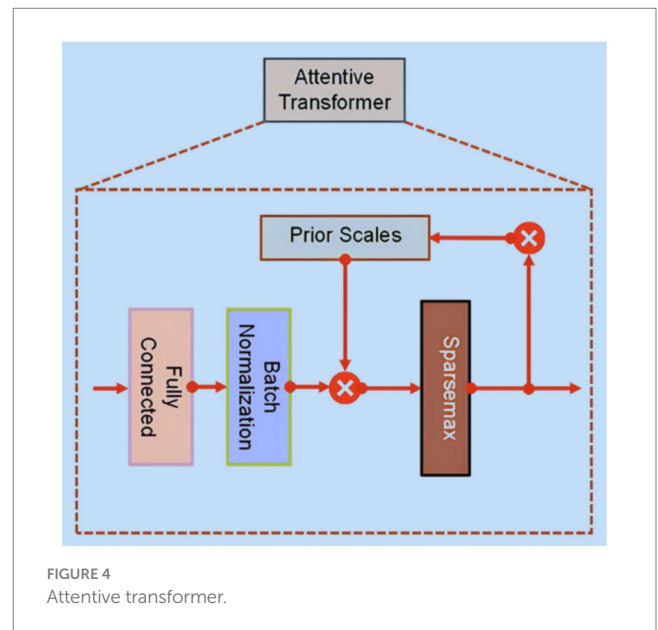

FIGURE 4
Attentive transformer.

of unlabeled data, from which we derived pretrained weights. This process entails the utilization of the TabNetPretrainer to acquire knowledge from the unlabeled dataset. Subsequently, we transitioned to a fine-tuning stage, where these pretrained weights were employed to train a TabNetClassifier. This classifier was further trained using labeled data to enhance its predictive capabilities. For the optimization of hyperparameters in both the pretraining and fine-tuning phases, we leveraged the Optuna (28) framework. The fine-tuned hyperparameters are meticulously detailed in Table 1. Ultimately, the optimized model is employed to evaluate performance scores through a 5-fold cross-validation procedure. To facilitate a performance comparison between the SSP-TabNet model and the standard TabNet model, we adopted a consistent approach by applying the same hyperparameters used for SSP-TabNet to the standard supervised TabNet model. However, it is worth noting that the "from_unsupervised" hyperparameter was not utilized when fitting the normal TabNet model.
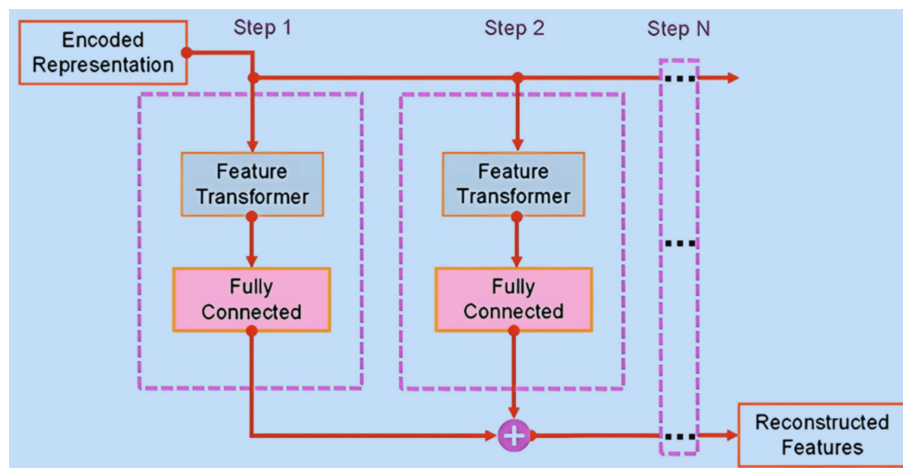
FIGURE 5
TabNet decoder structure.

TABLE 1 Hyperparameters when developing SSP-TabNet optimized by Optuna.

Model	Hyperparamters
pretrain	TabNetPretrainer: n_a = 40, n_d = 40, n_steps = 4, momentum = 0.12, lamda_sparse = 2.96e-05, optimizer_params = {'lr': 0.03}, optimizer_fn = torch.optim. adam.Adam, scheduler_params = {'patience': 5, 'mode': 'min', 'factor': 0.9, 'min_lr': 1e-05}, scheduler_fn = torch.optim.lr_scheduler.ReduceLROnPlateau, gamma = 1.9, mask_type = 'entmax', batch_size = 512, virtual_batch_size = 256, pretraining_ratio = 0.6, max_epochs = 500, patience = 30
SSP-TabNet	TabNetClassifier: n_a = 32, n_d = 32, n_steps = 9, momentum = 0.26, lamda_sparse = 6.39e-05, optimizer_params = {'lr': 0.009}, mask_type = 'sparsemax', optimizer_fn = torch.optim.adam.Adam, gamma = 1, virtual_batch_size = 128, batch_size = 1,024, eval_metric = ['auc'], max_epochs = 500, patience = 30, from_upsupervised = pretrain
TabNet	TabNetClassifier: n_a = 32, n_d = 32, n_steps = 9, momentum = 0.26, lamda_sparse = 6.39e-05, optimizer_params = {'lr': 0.009}, mask_type = 'sparsemax', optimizer_fn = torch.optim.adam.Adam, gamma = 1, virtual_batch_size = 128, batch_size = 1,024, eval_metric = ['auc'], max_epochs = 500, patience = 30

Unsupervised pretraining = pretrain, Self-supervised pretraining TabNet Classification model = SSP-TabNet, TabNet Classification model = TabNet.

2.4 Performance evaluation method

2.4.1 Comparing performance to ML models

While Arik et al. (14) demonstrated TabNet's superiority over traditional ML algorithms, recent studies (29–31) have continued to show instances where tree-based models still outperform DL when dealing with tabular data. This is due to tree-based models' ability to effectively handle high-dimensional data and perform well on structured datasets. Therefore, in our investigation, we opted to include a variety of tree-based models, specifically CatBoost (32), LightGBM (33), XGBoost (34), Gradient Boosting (GBC) (35), Extra Tree (ET) (36), Random Forest (RF) (37), and Decision Tree (DT) (38) to conduct a performance comparison with the SSP-TabNet model. Including these models ensures a fair and comprehensive assessment of SSP-TabNet's performance. Additionally, to guarantee a fair evaluation, we also fine-tuned the hyperparameters of these tree-based models using the Optuna framework. These models' hyperparameters were outlined in Table 2.

2.4.2 Performance metrics

In binary classification problems, the most commonly employed evaluation metrics encompass accuracy, recall, precision and the F1-score. These metrics collectively provide a statistical assessment of a classifier model's performance. The formulas for these metrics are presented in Equations 6–9:

$$Accuracy = \frac{True_{possitive} + True_{negative}}{Total\ predictions} \quad (6)$$

$$Recall = \frac{True_{possitive}}{True_{positive} + False_{negative}} \quad (7)$$

$$Precision = \frac{True_{possitive}}{True_{positive} + False_{positive}} \quad (8)$$

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (9)$$

where $True_{negative}$ and $True_{possitive}$ signify the accurate predictions made for the satisfied (class 0) and dissatisfied (class 1)

TABLE 2 Hyperparameters of tree-based models.

Model	Hyperparamters
CatBoost	CatBoostClassifier: depth = 10, l2_leaf_reg = 200, border_count = 254, n_estimators = 300, eta = 0.14
LightGBM	LGBMClassifier: max_depth = −1, learning_rate = 0.04, min_child_weight = 0.001, min_child_samples = 4, reg_alpha = 1.77e-09, reg_lambda = 0.02, bagging_freq = 2, feature_fraction = 0.4, bagging_fraction = 0.93, num_leaves = 256, n_estimators = 300
XGBoost	XGBClassifier: learning_rate = 0.07, max_depth = 11, colsample_bytree = 0.64, reg_alpha = 6.07e-06, reg_lambda = 10, subsample = 0.88, objective = 'binary:logistic', n_estimators = 300
GBM	GradientBoostingClassifier: max_features = 0.4, max_depth = 11, min_samples_leaf = 5, min_samples_split = 2, min_impurity_decrease = 1e-09, subsample = 0.86, n_estimators = 300, learning_rate = 0.05
ET	ExtraTreesClassifier: min_samples_split = 5, min_samples_leaf = 3, criterion = 'entropy', n_estimators = 150, max_depth = 20
RF	RandomForestClassifier: criterion = 'entropy', n_estimators = 120, max_depth = 25, min_samples_leaf = 4, min_samples_split = 6, min_impurity_decrease = 0.002, bootstrap = True
DT	DecisionTreeClassifier: min_samples_leaf = 2, min_impurity_decrease = 0.001, min_samples_split = 10, max_depth = 7

CatBoost Classification model = CatBoost, LightGBM Classification model = LightGBM, XGBoost Classification model = XGBoost, Gradient Boosting Classification model = GBM, Extra Tree Classification model = ET, Random Forest Classification mode = RF, Decision Trees Classification model = DT.

respectively; while $False_{negative}$ and $False_{positive}$ represent the erroneous predictions of the class 0 and the class 1, respectively.

Furthermore, the "Area under the Receiver Operating Characteristic Curve" (AUC) was also employed to assess the performance of models. The AUC score is calculated as shown in Equation 10:

$$AUC = \int_{1}^{0} TPR(t_i)\, d\left(FPR(t_i)\right) \qquad (10)$$

where $FPR(t_i)$ and $TPR(t_i)$ are "false positive rate" and "true positive rate" for a threshold t_i. In our study, we operated under the assumption that a model achieving the highest AUC exhibited the most robust predictive capability. In cases where multiple models displayed similar AUC values, we prioritized the model with the highest accuracy as the superior choice.

In this study, we conducted all our analyses using Python 3.10.13 (https://www.python.org, accessed on September 29, 2023). To rigorously evaluate the performance metrics across all models, we adopted the stratified 5-fold cross-validation technique. This method entails partitioning the dataset into k equally sized segments, with instances randomly distributed across these segments. Stratified cross-validation's distinctive aspect is its ability to preserve a distribution of class labels in each segment that mirrors the original dataset's distribution. In the k iterations of evaluating the model, each segment is utilized once as the validation set, while the others form the training data. This approach, by testing the model on various data subsets, provides a more comprehensive assessment of its generalization ability, thereby strengthening the credibility of our findings.

2.5 Local interpretable model-agnostic explanations (LIME)

Local Interpretable Model-Agnostic Explanations (LIME) is a widely adopted model-agnostic technique employed to elucidate the inner workings of black-box models. Its fundamental approach involves creating localized interpretable models centered around a specific data instance, thereby approximating how the black-box

model behaves in that particular context. LIME was originally developed by Ribeiro et al. (15) and has gained significant popularity for explaining the predictions made by complex ML models. Notably, researchers have harnessed LIME for various applications. For instance, Nguyen et al. (39) applied LIME to predict depression in Parkinson's Disease patients. Mardaoui et al. (40) employed LIME to interpret text data effectively. Mizanur et al. (41) used LIME to explain for loan approval prediction model. Additionally, Jain et al. (42) utilized LIME to offer insights into sentiment analysis results derived from social media texts. These studies collectively demonstrate LIME's utility in providing profound insights into the predictions of black-box models across diverse domains.

In order to provide an explanation for a particular observation, LIME operates by repeatedly perturbing an observation to create a set of replicated feature data. This perturbed data is then subjected to predictions using a model, such as TabNet. Each point in the perturbed dataset is compared with the original data point, and the Euclidean distance between them is calculated. This distance provides an indication of how much the perturbed data point diverges from the original observation. This metric is crucial in identifying which input features are considered significant by the model for its predictions. The ultimate objective of LIME is to create an explainer that is both dependable and interpretable. To achieve this goal, LIME focuses on minimizing the following objective function (Equation 11):

$$\xi(x) = argmin_{g \in G}\, \mathcal{L}(f, g, \pi_x) + \Omega(g) \qquad (11)$$

where f represents the original model, g denotes the interpretable model, x represents the original observation, π_x signifies the proximity measure computed across all permutations to the original observation. In addition, $\mathcal{L}(f, g, \pi_x)$ serves as a metric assessing the extent to which the interpretable model g faithfully approximates the behavior of the original model f within the locality defined by π. Lastly, $\Omega(g)$ represents a measure of model complexity. In the context of our study, we selected a specific instance for analysis to illustrate how the LIME model collaborates with the SSP-TabNet model to predict an individual's life dissatisfaction outcome. This focused examination allowed us to showcase the practical application of LIME in tandem with the SSP-TabNet model for predictive purposes.

TABLE 3 Performance comparison of optimized models on the training set.

Model	AUC	Accuracy	Precision	Recall	F1-score
SSP-TabNet	0.7778	0.7047	0.7066	0.7064	0.7065
CatBoost	0.7735	0.7012	0.6963	0.7057	0.701
XGBoost	0.7689	0.6982	0.6924	0.7031	0.6977
LightGBM	0.7683	0.6979	0.6994	0.7	0.6997
RF	0.7636	0.6945	0.6959	0.6965	0.6962
GBC	0.7598	0.6872	0.6914	0.6883	0.6898
ET	0.7521	0.6827	0.6928	0.6817	0.6872
DT	0.7321	0.6719	0.6709	0.6749	0.6729
TabNet	0.7321	0.6722	0.6703	0.6756	0.6729

Self-supervised pretraining TabNet Classification model = SSP-TabNet, TabNet Classification model = TabNet, LightGBM Classification model = LightGBM, XGBoost Classification model = XGBoost, Gradient Boosting Classification model = GBM, CatBoost Classification model = CatBoost, Extra Tree Classification model = ET, Decision Trees Classification model = DT, Random Forest Classification mode = RF.

3 Results

3.1 Performance comparisons results

Table 3 presents a summary of the predictive performance of 9 optimized models on the training set. The SSP-TabNet model showed the highest performance, achieving an AUC score of 0.7778, significantly higher than the other models. Models such as CatBoost, XGBoost, LightGBM, RF, and GBC followed, with AUC scores of 0.7735, 0.7689, 0.7683, 0.7636, and 0.7598, respectively. Conversely, the ET, DT, and TabNet models exhibited the lowest AUC scores among the nine models, each attaining an AUC value of 0.7521.

The SSP-Tabnet model not only excelled in AUC scores but also demonstrated outstanding performance in other evaluation metrics, such as accuracy, precision, recall, and F1-score. For accuracy, SSP-Tabnet obtained the best score at 0.7047, meaning that the model correctly predicted life satisfaction in 70.47% of cases. CatBoost (0.7012), XGBoost (0.6982), LightGBM (0.6979), and RF (0.6945) followed closely behind. Lower accuracies were observed for GBC (0.6872), ET (0.6827), TabNet (0.6722), and DT (0.6719). SSP-TabNet attained precision of 0.7066, recall of 0.7064, and F1-score of 0.7065, reflecting its balanced performance across different evaluation measures.

On the test set, the SSP-TabNet model continued to outperform the other models with an AUC score of 0.7757, as depicted in Figure 6. Following in the rankings were CatBoost (0.7682), XGBoost (0.7678), LightGBM (0.7641), RF (0.7628), GBC (0.7468), ET (0.7468), DT (0.7211), and TabNet (0.7209). According to these results, the SSP-TabNet model exhibited outstanding performance relative to all other models examined in this study, particularly in predicting individuals' life satisfaction, as observed in both the training and test datasets.

3.2 Results of global and local interpretability by TabNet

Figure 7 provides a visual representation of the global importance attributed to each feature following the training of SSP-TabNet. The analysis reveals that the top five crucial features, ranked by importance, for SSP-TabNet are as follows.

- code120 (Changes in Daily Life Due to COVID-19: Employment activities)
- code34 (Do you commute or go to school?)
- code110 (Age)
- code58 (Effects of School Education: Personality Development)
- code60 (Effects of School Education: Utilization in Daily Life and Employment)

In terms of local explanation by SSP-TabNet, Figure 8 illustrates how SSP-TabNet selects and weighs features at different decision steps in the model. The visualization utilizes color intensity to convey the feature weight assigned at each decision step. Brighter colors indicate features that were given more weight in that step, meaning they had a greater impact on the model's decision at that point. Each decision step assigns different weights to individual features, demonstrating that the model considers each case separately and gives different importance to features depending on the individual data point.

3.3 Local interpretation SSP-TabNet with LIME

Figure 9 provides a detailed description of an individual for whom the predicted outcome is "dissatisfied" with life. Given the individual's states and attributes, the SSP-TabNet model predicted a "dissatisfied" rate of 95%, as illustrated in Figure 9a. In Figure 9b visualization, the blue bars symbolize those variables that significantly contribute to negating the prediction, aligning with the "dissatisfied" outcome. Conversely, the orange bars depict variables that support the prediction, correlating with a "satisfied" outcome. According to the explanation provided in Figure 9b, at the time of the prediction, code26 (Residential Environment Satisfaction-Housing) emerged as the most influential factor contributing to the prediction, with a weight of 0.1.

In Figure 9c, we offer a summary of this individual's state and the contributing circumstances, which are derived from 10 out of the 51 features. This individual's attributes can be explained as:

- code26: 4 (Residential Environment Satisfaction-Housing: Not quite)
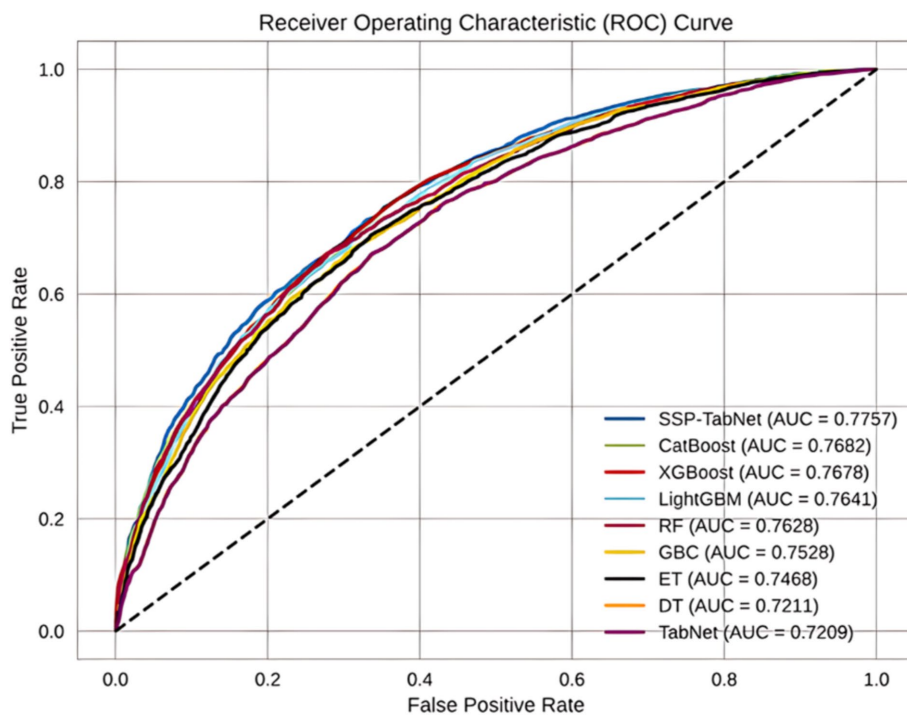- code41: 3 (Satisfaction with Leisure Activities: Average)

FIGURE 6
ROC curves of all models evaluated on the test set.

- code114: 4 (Marital Status: Widowed)
- code45: 1 (Weekend and Holiday Leisure Activities: Watching TV)
- code126: 6 (Changes in Daily Life Due to COVID-19 - Gatherings with Family, Friends, and Colleagues: Not applicable)
- code57: 2 (Effects of School Education - Acquisition of Knowledge and Skills: Somewhat)
- code34: 3 (Do you commute or go to school?: Do not commute or go to school)
- code109: 2 (Gender: Female)
- code128: 1 (Behavior Changes for COVID-19 Prevention - Canceling Gatherings and Not Attending Events: Always)
- code131: 1 (Behavior Changes for COVID-19 Prevention - Using Soap and Hand Sanitizer: Always)

4 Discussion

This research represents, to our knowledge, the first instance of developing a DL model to predict life satisfaction using tabular data extracted from social survey in general population. In this research, we evaluated the effectiveness of the novel SSP-TabNet model in predicting life satisfaction indices. Notably, our self-supervised pre-training approach showcased superior performance compared to conventional tree-based algorithms, including CatBoost, LightGBM, XGBoost, GBC, ET, RF, and DT, particularly when applied to social survey tabular datasets. Our comparisons revealed that while the standard supervised TabNet model yielded an AUC of 0.7321 on the training set and 0.7209 on the test set; while the SSP-TabNet model resulted in a commendable AUC of 0.7778 and 0.7757 on the respective sets. In addition, the integration of the SSP-TabNet with the LIME interpretability framework offers a transparent understanding of AI decisions, making it a valuable tool for professionals in the social sciences and psychology, even if they lack expertise in data analytics.
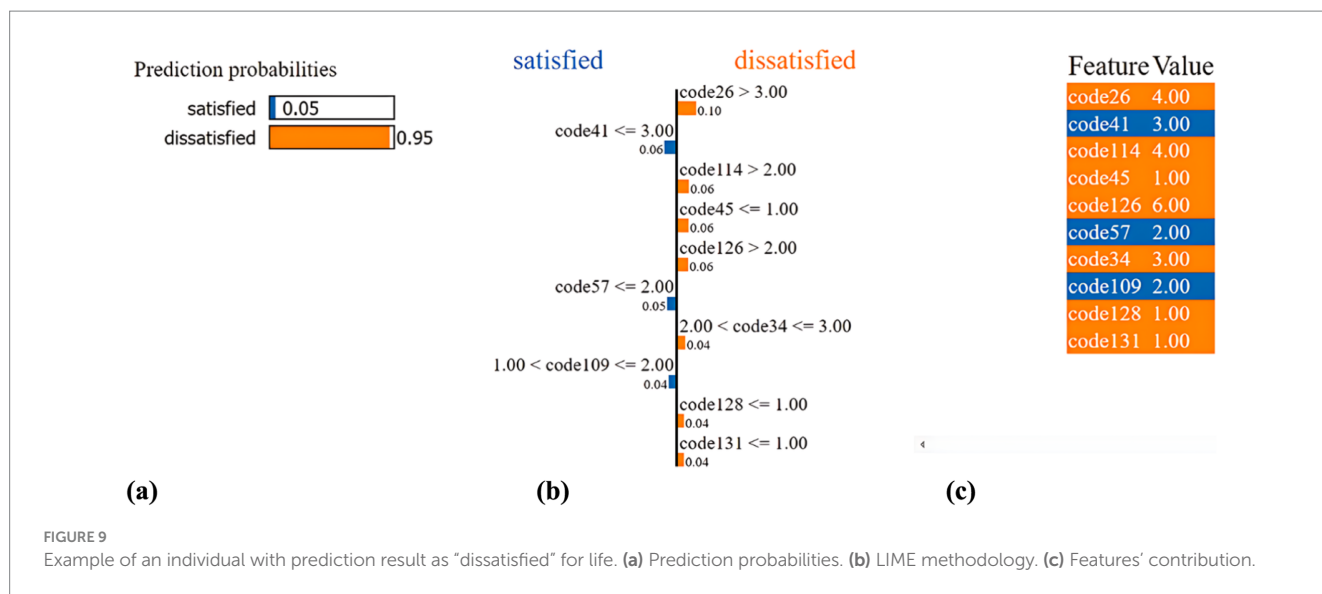
In the realm of data science, deciphering TabNet's mask as illustrated in Figure 8 is straightforward. Yet, for experts in social sciences and psychology, without an in-depth background in data analytics, grasping the intricacies of TabNet's mask can be challenging (24). While TabNet provides local explanations via feature masks, our proposed model offers a more intuitive insight. Social science professionals and psychologists can swiftly understand the model's rationale by merely contrasting code values. Consequently, our integrated model streamlines the interpretation of local model behaviors, bypassing the intrinsic nuances of TabNet's inherent explanatory features. This enhancement widens its applicability to a more diverse audience.

A significant benefit of TabNet's architecture is its ability to obviate the necessity for feature pre-processing. Through TabNet's Mask layer, there is a streamlined selection of the most pivotal features (14). Such a mechanism bolsters the model's efficiency, channeling its learning capability toward the most relevant features and circumventing unnecessary computations on superfluous ones. Hence, in our research, we trained the model directly without employing any feature selection techniques before model training. Yet, for conventional ML models to achieve optimal performance, the incorporation of an adept feature selection method is essential. Future studies might delve into the integration of tree-based models with several feature selection methods, subsequently contrasting their efficacy against our proposed SSP-TabNet model.

FIGURE 7
Global feature importance ranking.

FIGURE 8
TabNet's feature importance masks.
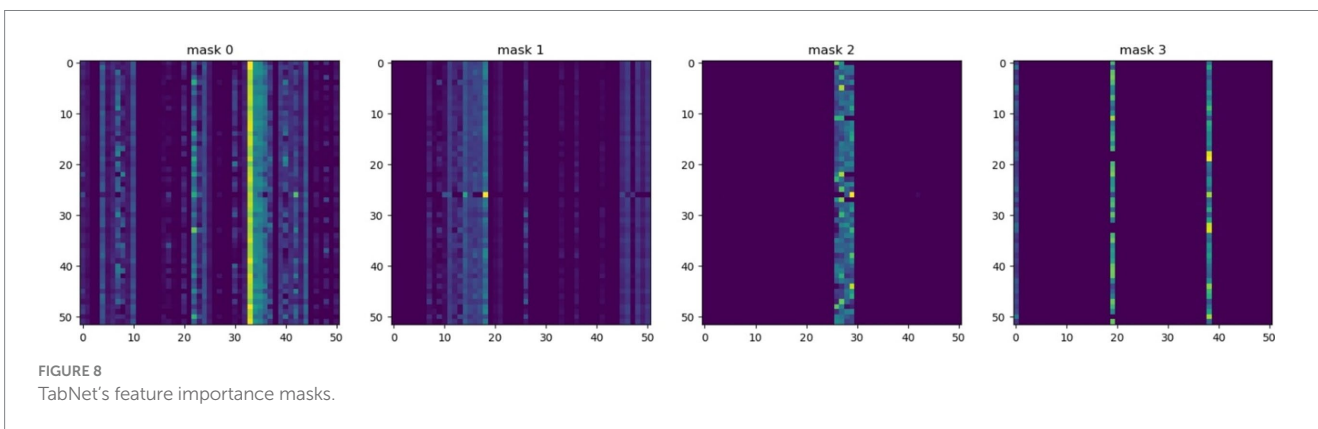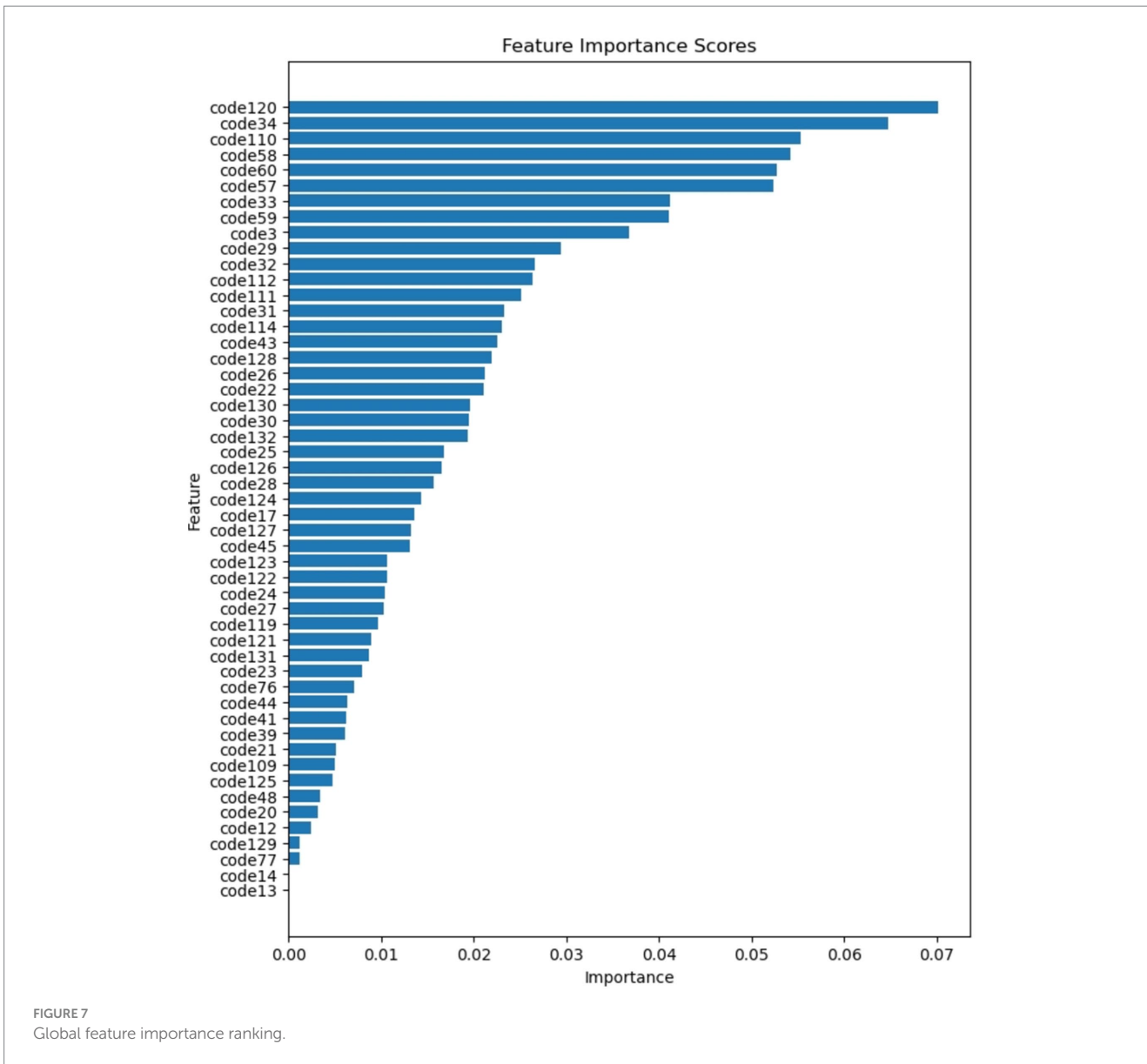
While some similar studies (2, 13, 43, 44), which applied ML models to predict life satisfaction, have reported slightly higher performance scores, it is crucial to acknowledge that these models often rely on extensive feature engineering or manual selection of relevant variables. In contrast, the SSP-TabNet model automatically identifies the most relevant features through its Mask layer. In our

FIGURE 9
Example of an individual with prediction result as "dissatisfied" for life. **(a)** Prediction probabilities. **(b)** LIME methodology. **(c)** Features' contribution.

work, we seeks to illustrate that the standard architecture of our proposed SSP-TabNet is more robust than the standard architectures of tree-based ML models, which often require experimentation with various manual feature selection methods to achieve optimal performance. Additionally, many higher-performing models in previous studies (2, 13, 43) tend to function as "black-box" models, offering limited interpretability. Our model, therefore, strikes a balance between predictive accuracy and interpretability, which is critical for real-world applications, particularly in social science and mental health research.

5 Limitations and future works

This study presents several limitations. Firstly, the process of training and fine-tuning the TabNet model was more time-intensive compared to traditional machine learning models. Secondly, the generalizability of our predictive models is limited due to the specific cultural and socioeconomic context of the dataset, which originates from South Korea and includes individuals aged 15 and above. As a result, applying these models to different populations and cultural contexts may not be straightforward. Third, by relying on static, cross-sectional data, the model may not capture the temporal and dynamic aspects of life satisfaction, potentially failing to account for inherent fluctuations and evolving circumstances over time. Lastly, the stability of LIME explanations can occasionally be affected by variations in sample selection or the criteria used for including local data points in the model.

In future research, we intend to test and potentially strengthen the robustness of our models by integrating data from a variety of nations, with the goal of developing a more globally applicable predictive model. We intend to expand our study to encompass various demographic and cultural situations, allowing for a comprehensive assessment of the generalizability of the findings across varied populations. Additionally, we intend to subject the LIME interpretations derived from our model to meticulous review and evaluation by experts in social science and psychology.

6 Conclusion

This study introduced the SSP-TabNet model, a synergy of the self-supervised pre-training TabNet model and the LIME interpretative methodology, tailored to predict life satisfaction among South Koreans. The performance of our advanced model surpassed conventional tree-based ML models, registering an AUC of 0.7778 for the training set and 0.7757 for the test set. Furthermore, our integrated model simplifies and clarifies the interpretation of local model actions, effectively navigating past the intricate nuances of TabNet's standard explanatory mechanisms. This refinement paves the way for wider accessibility and understanding across diverse audiences.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/participants or the patients'/participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

HN: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. HB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2024.1445864/full#supplementary-material

References

1. Pavot W, Diener E. Review of the satisfaction with life scale In: E Diener, editor. Assessing well-being: The collected works of Ed Diener. Dordrecht: Springer Netherlands (2009). 101–17.

2. Song M, Zhao N. Predicting life satisfaction based on the emotion words in self-statement texts. *Front Psych*. (2023) 14:1121915. doi: 10.3389/fpsyt.2023.1121915

3. Lewis CA, Dorahy MJ, Schumaker JF. Depression and life satisfaction among northern Irish adults. *J Soc Psychol*. (1999) 139:533–5. doi: 10.1080/00224549909598413

4. Fergusson DM, McLeod GFH, Horwood LJ, Swain NR, Chapple S, Poulton R. Life satisfaction and mental health problems (18 to 35 years). *Psychol Med*. (2015) 45:2427–36. doi: 10.1017/S0033291715000422

5. Wong CK, Wong KY, Mok BH. Subjective well-being, societal condition and social policy–the case study of a rich Chinese society. *Soc Indic Res*. (2006) 78:405–28. doi: 10.1007/s11205-005-1604-9

6. Choi Y-K, Joshanloo M, Lee J-H, Lee H-S, Lee H-P, Song J. Understanding key predictors of life satisfaction in a nationally representative sample of Koreans. *Int J Environ Res Public Health*. (2023) 20:6745. doi: 10.3390/ijerph20186745

7. Gutiérrez M, Tomás JM, Galiana L, Sancho P, Cebrià MA. Predicting life satisfaction of the Angolan elderly: a structural model. *Aging Ment Health*. (2013) 17:94–101. doi: 10.1080/13607863.2012.702731

8. Itahashi T, Kosibaty N, Hashimoto R-I, Aoki YY. Prediction of life satisfaction from resting-state functional connectome. *Brain Behav*. (2021) 11:e2331. doi: 10.1002/brb3.2331

9. Macků K, Caha J, Pászto V, Tuček P. Subjective or objective? How objective measures relate to subjective life satisfaction in Europe. *ISPRS Int J Geo Inf*. (2020) 9:50320. doi: 10.3390/ijgi9050320

10. Marion D, Laursen B, Zettergren P, Bergman LR. Predicting life satisfaction during middle adulthood from peer relationships during mid-adolescence. *J Youth Adolesc*. (2013) 42:1299–307. doi: 10.1007/s10964-013-9969-6

11. Verme P. Life satisfaction and income inequality. *Rev Income Wealth*. (2011) 57:111–27. doi: 10.1111/j.1475-4991.2010.00420.x

12. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatr Cognit Neurosci Neuroimag*. (2018) 3:223–30. doi: 10.1016/j.bpsc.2017.11.007

13. Shen X, Yin F, Jiao C. Predictive models of life satisfaction in older people: a machine learning approach. *Int J Environ Res Public Health*. (2023) 20:2445. doi: 10.3390/ijerph20032445

14. Arik SÖ, Pfister T. TabNet: attentive interpretable tabular learning. *AAAI*. (2021) 35:6679–87. doi: 10.1609/aaai.v35i8.16826

15. Ribeiro M, Singh S, Guestrin C. Why should I trust you?": explaining the predictions of any classifier In: J Denero, M Finlayson and S Reddy, editors. Proceedings of the 2016 conference of the north American chapter of the Association for Computational Linguistics: Demonstrations. San Diego, CA: Association for Computational Linguistics (2016). 97–101.

16. Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J, et al. Self-supervised learning: generative or contrastive. *IEEE Trans Knowl Data Eng*. (2023) 35:1–876. doi: 10.1109/TKDE.2021.3090866

17. Grill J-B, Strub F, Altché F, Tallec C, Richemond P, Buchatskaya E, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Adv Neural Inf Proces Syst*. (2020) 33:21271–84.

18. He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning In: IEEE, editor. 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Seattle, WA: IEEE (2020). 9726–35.

19. Liu Y, Ma J, Xie Y, Yang X, Tao X, Peng L, et al. Contrastive predictive coding with transformer for video representation learning. *Neurocomputing* (2022) 482:154–162. doi: 10.1016/j.neucom.2021.11.031

20. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *J Mach Learn Res*. (2020) 21:5485–551.

21. Song K, Tan X, Qin T, Lu J, Liu T-Y. MPNet: masked and permuted pre-training for language understanding. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS'20*. Red Hook, NY: Curran Associates Inc. (2020). p. 16857–67.

22. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: *Proceedings of the 37th International Conference on Machine Learning*. Vienna: ICML (2020). p. 1597–607.

23. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding In: J Burstein, C Doran and T Solorio, editors. Proceedings of the 2019 conference of the north American chapter of the Association for Computational Linguistics: Human language technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2–7, 2019, volume 1 (long and short papers). Stroudsburg, PA: Association for Computational Linguistics (2019). 4171–86.

24. Nguyen HV, Byeon H. Prediction of out-of-hospital cardiac arrest survival outcomes using a hybrid agnostic explanation TabNet model. *Mathematics*. (2023) 11:2030. doi: 10.3390/math11092030

25. Cai Q, He J. Credit payment fraud detection model based on TabNet and Xgboot. In 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE). Chester: ICCECE (2022). p. 823–6.

26. Borghini E, Giannetti C, Flynn J, Todeschini G. Data-driven energy storage scheduling to minimise peak demand on distribution systems with PV generation. *Energies*. (2021) 14:3453. doi: 10.3390/en14123453

27. Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN. Convolutional sequence to sequence learning. In Proceedings of the 34th International Conference on Machine Learning - Volume 70. Sydney, NSW: ICML'17. (2017). p. 1243–52.

28. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation Hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD international conference on Knowledge Discovery & Data Mining*. New York, NY, USA: Association for Computing Machinery (2019). p. 2623–2631.

29. Fayaz SA, Zaman M, Kaul S, Butt MA. Is deep learning on tabular data enough? An assessment. *IJACSA*. (2022) 13:130454. doi: 10.14569/IJACSA.2022.0130454

30. Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. *Inf Fusion*. (2022) 81:84–90. doi: 10.1016/j.inffus.2021.11.011

31. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? In: S Koyejo, S Mohamed, A Agarwal, D Belgrave, K Cho and A Oh, editors. Advances in neural information processing systems. Red Hook, NY: Curran Associates Inc (2022). 507–20.

32. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: *Proceedings of the 32nd International*

Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc. (2018). p. 6639–49.

33. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W. LightGBM: a highly efficient gradient boosting decision tree. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc. (2017). p. 3149–57.

34. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: Association for Computing Machinery (2016). p. 785–794.

35. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. (2001) 29:3451. doi: 10.1214/aos/1013203451

36. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. (2006) 63:3–42. doi: 10.1007/s10994-006-6226-1

37. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324

38. Ayyadevara VK. (ed.) Decision tree. In *Pro Machine Learning Algorithms: A Hands-on Approach to Implementing Algorithms in Python and R*. Berkeley, CA: Apress (2018). p. 71–103.

39. Nguyen HV, Byeon H. Prediction of Parkinson's disease depression using LIME-based stacking ensemble model. *Mathematics*. (2023) 11:30708. doi: 10.3390/math11030708

40. Mardaoui D, Garreau D. An analysis of lime for text data. *PMLR*. (2021) 18:3493–501.

41. Mizanur Rahman SM, Golam Rabiul Alam MD. Explainable loan approval prediction using extreme gradient boosting and local interpretable model agnostic explanations In: X-S Yang, RS Sherratt, N Dey and A Joshi, editors. Proceedings of Eighth International Congress on Information and Communication Technology. Singapore: Springer Nature Singapore (2023). 791–804.

42. Jain R, Kumar A, Nayyar A, Dewan K, Garg R, Raman S, et al. Explaining sentiment analysis results on social media texts through visualization. *Multimed Tools Appl*. (2023) 82:22613–29. doi: 10.1007/s11042-023-14432-y

43. Prati G. Correlates of quality of life, happiness and life satisfaction among European adults older than 50 years: a machine-learning approach. *Arch Gerontol Geriatr*. (2022) 103:104791. doi: 10.1016/j.archger.2022.104791

44. Khan AE, Hasan MJ, Anjum H, Mohammed N, Momen S. Predicting life satisfaction using machine learning and explainable AI. *Heliyon*. (2024) 10:e31158. doi: 10.1016/j.heliyon.2024.e31158