



# Analysis of health recommendations using longitudinal quality of life data: QoL@TbA – A transformer-based approach

Health Informatics Journal  
1–14

© The Author(s) 2024

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/14604582241291789

[journals.sagepub.com/home/jhi](https://journals.sagepub.com/home/jhi)



**Clairton Siebra** 

Quality of Life Technologies Lab, University of Geneva, Geneva, Switzerland

Informatics Center, Federal University of Paraíba, João Pessoa, Brazil

**Mascha Kurpicz-Briki**

Bern University of Applied Science, Bern, Switzerland

**Katarzyna Wac**

Quality of Life Technologies Lab, University of Geneva, Geneva, Switzerland

## Abstract

**Objective:** Health recommendation systems suggest behavioral modifications to improve quality of life. However, current approaches do not facilitate the generation or examination of such recommendations considering the multifeature longitudinal evolution of behaviors. This paper proposes the use of a deep learning transformer-based model that allows the analysis of recommendations for behavior changes. **Methods:** We adapted a prediction approach, namely Behavior Sequence Transformer (BST), which analyzes temporal human routines and patterns, generating inductive outcomes. The evaluation relied on a case study that employed the behavioral history and profile of the English Longitudinal Study of Ageing (ELSA) participants ( $n = 2682$ ), predicting their psychological mood (normal, pre-depressed, depressed) according to input recommendations for behavioral changes. Root mean squared error (RMSE) and learning curves were used to track the recommendation accuracy evolution and possible overfitting problems. **Results:** Experiments demonstrated lower RMSE values for the multifeature model (0.28/0.03) when compared to its single-feature versions (marital status, 0.59/0.001), (high pressure, 0.357/0.04), (diabetes, 0.36/0.01), (sleep quality, 0.57/0.02), (level of physical

## Corresponding author:

Clairton Siebra, Quality of Life Technologies Lab, University of Geneva, 7, route de Drize, Battelle Building A, Office 234, Geneva 1211, Switzerland.

Email: [clairton.dealbuquerque@unige.ch](mailto:clairton.dealbuquerque@unige.ch)



Creative Commons CC BY: This article is distributed under the terms of the Creative

Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which

permits any use, reproduction and distribution of the work without further permission provided the

original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

activity, 0.57/0.01). **Conclusions:** The results demonstrate the architecture's capability to analyze multifeatured longitudinal data, supporting the generation of suggestions for concurrent modifications across multiple input features. Moreover, these suggestions align with findings in specialized literature.

## Keywords

Healthy behavior, recommendations, inductive reasoning, deep learning

## Introduction

According to the World Health Organization (WHO), quality of life (QoL) is “*the individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns.*”<sup>1</sup>. The QoL assessment is important because a direct association exists between QoL and the health of individuals.<sup>2</sup> Moreover, according to Manea and Wac,<sup>3</sup> (un)healthy lifestyle over long periods significantly affects QoL. Thus, the evolution of behavioral patterns in multiple dimensions presents intrinsic information that leverages the QoL interpretation. Indeed, as several health issues are derived from multifeatured and evolving daily behavior, the specification of health care support systems must consider such behavior to properly anticipate health issues,<sup>4</sup> and possibly recommend behavioral modifications to avoid or attenuate such issues.

The current state of the art for capturing the long-term dependencies among events is the transformer architecture.<sup>5</sup> Given its success in handling natural language processing problems, several works were encouraged to adapt such an architecture to other areas<sup>6</sup> including health care.<sup>7,8</sup> However, the most recent health recommendation systems reviews<sup>9,10</sup> show that the main employed approaches are still based on text mining, KNN, clustering, matrix factorization, and neural network techniques. Moreover, such approaches do not consider longitudinal data.

Differently from these previous approaches and following the trend of using machine learning as a prediction and prevention-support tool for public health,<sup>11</sup> our study relied on a specific type of transformer, called *Behavior Sequence Transformer (BST)*,<sup>12</sup> proposed to support recommendation systems. Our main idea was to adapt this transformer, so it considers the multifeatured behavioral history and profile of individuals to predict health outcomes (e.g., mental state) according to input recommendations for behavioral changes. Therefore, the requirements considered for leveraging BST for QoL were: (1) The model must capture the intrinsic sequential information underlying the individuals' behaviors; (2) Behaviors must be represented by a multifeatured set of QoL data; (3) Predictions must involve generated recommendations for behavioral changes, and; (4) The models must consider long sequences and dependencies since this aspect is common in QoL data. The demonstration of this approach employed data from a subgroup of 2682 participants of the ELSA dataset.<sup>13</sup> We used this data to train and validate our BST-based model and to support the analysis of behavior changes and their relationship with depression issues. [Table 1](#) summarizes the more recent studies in the health recommendation domain, comparing them with our approach (Transformer-based approach – QoL@TbA).

## Methods

### *Theoretical framework*

Behavior Sequence Transformers (BSTs) are neural network models used to capture the dependency among items in individuals' behavior sequences.<sup>12</sup> They are currently used in domains such as

**Table I.** Summary of previous recent recommendation studies and comparison with our approach.

Year/Ref	What is recommended?	Inputs	Modelling technique	Perform longitudinal analysis?
2022 <sup>14</sup>	Diet, exercise, and medications	Glucose, blood pressure, BMI, skin thickness, age, pregnancies	Fuzzy inference	No. Data is analyzed in real-time (moment)
2023 <sup>15</sup>	Safe distance of potential risks	Spatial data	Consistency-focused strategy	No. Only consider current positions (moment)
2023 <sup>16</sup>	Diet and exercise	Height, weight, age	K-means and random forest classification	No. Consider the current scenario to generate a plan of action
QoL@TbA (our)	Modifications for 5 behavioral features	Marital status, quality of sleep, level of physical activity, high-pressure diagnostic, and diabetes diagnostic	DL transformers	Yes. The analyzed time span is a configurable parameter

products or movies recommendation systems. We can specify a BST domain using the tuple  $\langle I, P \rangle$ , where  $I = (i_1, i_2, \dots, i_n)$  is a set of  $n$  individuals, while  $P = (p_1, p_2, \dots, p_m)$  is a set of  $m$  products. At different temporal moments  $t$ , each individual  $i_j$  can buy and evaluate (e.g., attributing a score  $\alpha$ ) a product  $p_k$ . This means,  $i_j: (p_k, \alpha)_t$ . Thus, the following descriptions are possible, considering the  $\alpha$  range from 1 to 5:

- Leo: (shoes, 3)<sub>1</sub>, (cellphone, 4)<sub>2</sub>, (earphone, 4)<sub>3</sub>, (glass, 5)<sub>4</sub>.
- Eva: (shoes, 3)<sub>1</sub>, (cellphone, 3)<sub>2</sub>, (earphone, 4)<sub>3</sub>, (SiliconeCover, 4)<sub>4</sub>.
- Marie: (shoes, 3)<sub>1</sub>, (cellphone, 3)<sub>2</sub>, (earphone, 4)<sub>3</sub>, (ps,  $\alpha$ )<sub>4</sub>.

The idea behind BST, in this context, is to predict the evaluation of *Marie* for different products  $p_s$ , so a BST-based recommender can offer the products with higher evaluations to her. Therefore, the shopping behavior sequence of all individuals is analyzed to suggest candidate products to *Marie*. BST optionally includes the use of individuals' profiles. For example, the system could reduce the glass score since Leo is a male while *Marie* is a female, or they have very different ages. As this information is not temporal, it brings implications for the architecture specification.

While BSTs were created for and are generally used in recommendation systems, this same transformer can be adapted to analyze the impact of behavior changes on health attributes. The tuple  $\langle X, A \rangle$  represents this scenario, where  $X = (x_1, x_2, \dots, x_n)$  is a set of  $n$  individuals, while  $A = (a_1, a_2, \dots, a_m)$  is a set of  $m$  behavioral assessments. At different moments  $t$ , each individual  $x_j$  has a physical, psychological, or social attribute value  $\beta$  derived from an assessment  $a_k$ . For example, depression status can be derived from other quality of life attributes, which characterize the behavior of individuals.<sup>28</sup> Like the previous example, this scenario can be summarized as  $x_j: (a_k, \beta)_t$ . When a physical, psychological, or social attribute value  $\beta$  is derived from  $a_k$  at the moment  $t$ , this derivation must use the sequence  $(a_k, \beta)_{t-1}, (a_k, \beta)_{t-2} \dots (a_k, \beta)_{t-z}$ , rather than only  $a_k$ . This strategy considers the importance of the sequence of behaviors to predict health outcomes, as in the previous example that considers the shopping sequence to score future options for shopping.

### Conceptualization

The following schema (Figure 1) summarizes the method used in this work. The BST training stage generates a model using longitudinal data from several individuals. The prediction variable is defined as any physiological or psychological feature. For example, sleep quality or level of stress. Assessments are multifeature data rather than unique features. Thus,  $a_1 = [a_{11}, a_{12}, a_{13}, \dots]$  represents common subset of quality of life features such as level of physical activity ( $a_{11}$ ), sleep quality ( $a_{12}$ ), indicators of health problems ( $a_{13}$ ), and marital status ( $a_{14}$ ), as defined in WHOQOL.<sup>1</sup> After being compiled, the trained model returns values to the prediction variable based on an individual's profile and his/her sequence of behaviors. In this case, the last behavioral sequence ( $S_t$ ) assessment is one or more generated behavior patterns. This means the model will evaluate these patterns, returning their prediction values. As discussed in the following sections, we needed to adapt the BST architecture for dealing with static and behavioral data given the multifeature aspect of the assessments, which are also changing and unfolding over time.<sup>24</sup>

### Architecture

The following schema (Figure 2) illustrates the high-level perspective of the QoL@TbA architecture proposed in this paper, which is an adaptation of the BST.<sup>12</sup>

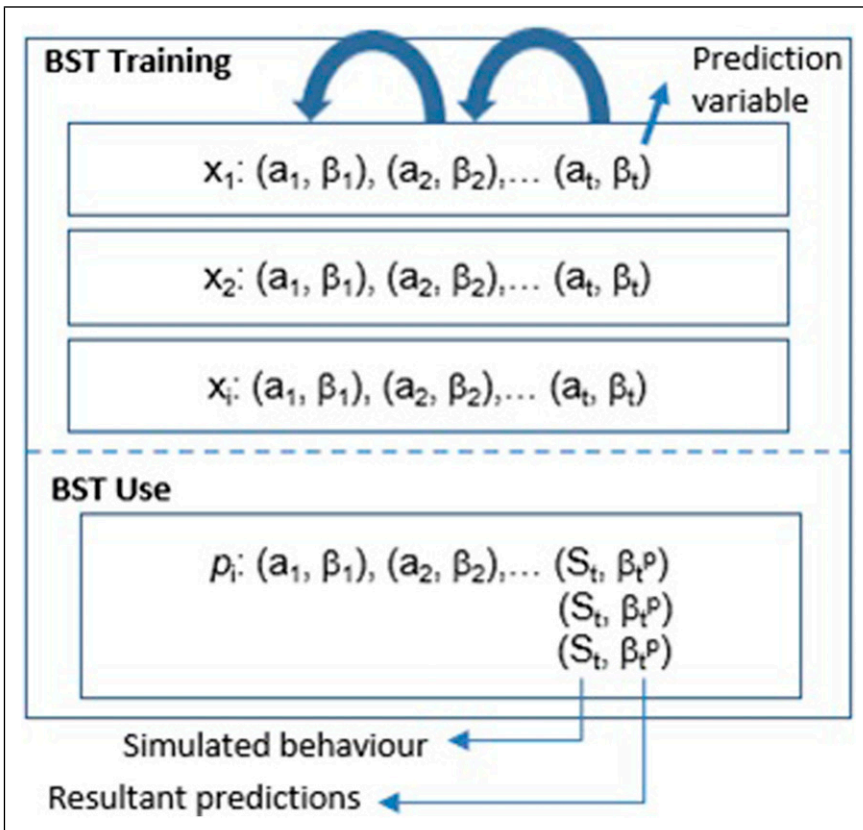
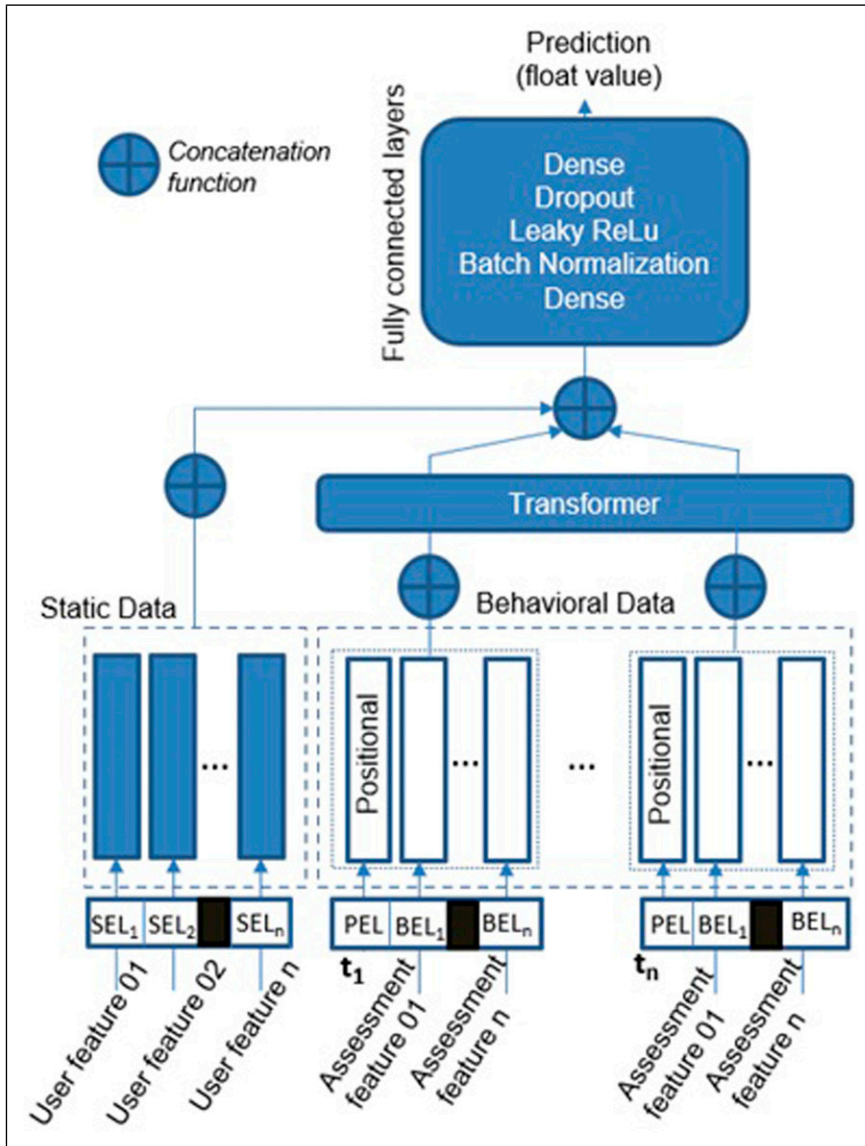


Figure 1. QoL@TbA: BST application schema for QoL analysis.



**Figure 2.** QoL@TbA architecture (adapted from Chen et al. proposal<sup>12</sup>).

*Static embedding layer (SEL).* Static data are represented by categorical features that do not evolve over time (e.g., biological sex). The module that processes such data has an embedding layer for each feature, and these layers encode categorical features to a dense vector of fixed size. However, as embedding layers only accept integer values as input, categorical values (strings) are first mapped into integer indexes. Moreover, each of the embedding layers is configured with an embedding dimension equal to the square root of the vocabulary size of the feature. For example, if the *age group* feature has four words in its vocabulary (child, young, adult, and elderly), the embedding dimension is 2. This is one of the machine learning design patterns commonly used in the area.<sup>17</sup>

Finally, the outputs of all static features are concatenated to form a single input tensor for the next module.

Our embedding layer for static data is simpler than the original layer of the BST version. For example, we do not combine different features to create new features. According to the BST authors,<sup>12</sup> such a combination has been demonstrated useful in their scenarios. However, we allow the model to automatically learn this cross-feature information if it is important since this is already an intrinsic feature of transformers architectures.

*Behavioral embedding layer (BEL).* Unlike static data, longitudinal behavioral data, such as sleep quality and level of physical activities, are assessed at different moments and change over time. Each assessment feature is encoded as a one-hot vector (Behavioral Embedding Layer - BEL) since they can only assume one value at each assessment. For example, the sleep quality can be “good” or “poor”. We defined the one-hot vector strategy because it can easily be adapted for a multi-hot vector if the assessment includes features that assume more than one value simultaneously. For example, an individual can have several diseases from a list. Moreover, we used a non-linear dense layer to process the outcomes of the embedding vector since it improves the results compared to directly sending the embedding vector to the next module. BEL uses the positional embedding layer (PEL) to add a positional vector to each set of BELs assessed at the same time, giving the notion of sequence for the input data.

*Transformer layer.* As in the original BST architecture, we implemented the only-encode transformer.<sup>5</sup> The main issue in this layer is to optimize its hyperparameters. The work of Michel et al.<sup>25</sup> discusses some greedy strategies to conduct such optimization. Similarly, automatic tuning algorithms, such as Random search and Bayesian Optimization, are usually applied to find optimal values for these hyper-parameters.<sup>26</sup>

*Fully connected layers.* The final module contains a set of fully connected layers (Dense, Batch Normalization, Leaky ReLu, and Dropout) (Figure 2). The role of these layers is to further learn the interactions among dense features. Our experiments used a linear layer with one output unit (float value) as last component since we intended to return a value that could represent the quantitative difference between predictions and reference values. In this case, dense layers do not present an activation function and only implement the dot product of matrices.

### *Case example: Motivation*

According to Razavi et al., “*depression is currently the second most significant contributor to non-fatal disease burdens globally.*”<sup>18</sup> This same study shows the possibility of screening for depressive symptoms using pervasive mobile technology.<sup>18</sup> This means, conducting a continuous capture of behavioral data. Thus, mobile technology enables the assessment of longitudinal data, which can be used to analyze possible pre-patterns for depression. This fact encourages investigations focused on inductive strategies that can efficiently analyze such data and provide the basis for interventions.

### *Case example: Dataset*

The English Longitudinal Study of Ageing (ELSA)<sup>13</sup> is a large-scale longitudinal study that involves participants aged 50 and over. This study is divided into waves, which occur every 2 years. During each wave, participants are requested to answer questions to figure out changes in their health, social,

and economic situations.<sup>13</sup> Thus, the ELSA study aims to “*complete the picture of what it means to grow older in the 21st century, and help us understand what accounts for the variety of patterns that are seen.*”<sup>13</sup> The resultant information provides, for example, data regarding demographics, physical and psycho-social health, cognitive function, social participation, and others. This study started with 7168 samples (in 2002). However, only 2959 continuously participated in the nine waves. Our study focused on these samples since our future aim is to use all the nine waves available in this dataset. This case example, in particular, uses data from the first four waves (2002-2008).

### Case example: Objectives

This case example aims to train a BST-based model using multifeature longitudinal (4 waves) data, comprising five QoL features. Then, the BST-based model uses behavioral history (3 waves) and profile of individuals to predict their psychological mood (normal, pre-depressed, depressed), according to input recommendations for behavioral changes. From the model perspective, each recommendation works as the fourth assessment. Thus, the model can evaluate several different recommendations. The following points detail the model input: **(1)** A set of static data representing the user profile (gender, age group, and number of children); **(2)** A fixed-length (3 waves) sequence of assessments, which represent five longitudinal features about marital status (single, married, remarried, divorced, separated, widowed), quality of sleep (good, poor), level of physical activity (under-active, active, very active), high-pressure diagnostic (yes, no), and diabetes diagnostic (yes, no); **(3)** A fixed-length (3 waves) sequence of mood values (normal, pre-depressed, depressed) for each assessment performed; and **(4)** A generated recommendation (behavior pattern - vector comprised of five features) for which we predict the mood. This recommendation has the same format as the input assessments (five QoL features), and it is automatically generated so the model can evaluate the effect of diverse behavioral patterns on the individuals’ moods.

### Case example: evaluation process

We have first created a baseline for our case example. As expected for datasets in the health domain, ELSA is imbalanced regarding the mood classes. From the 2682 samples, 2137 are classified as normal (value 1), 276 as pre-depressed (value 2), and 269 as depressed (value 3). We defined the baseline using the zero-rule strategy for classification, which predicts the class value that is the most common in the dataset. In our case, this class is the normal mood (value 1). Then, we obtain a baseline of 0.71 using the root-mean-square error (RMSE) measure. After that, we configured our experiments using the hyperparameters detailed in [Table 2](#), which were based on other works such as [19]. These experiments considered data from waves 1 to 4.

**Table 2.** Hyperparameters used in the experiments.

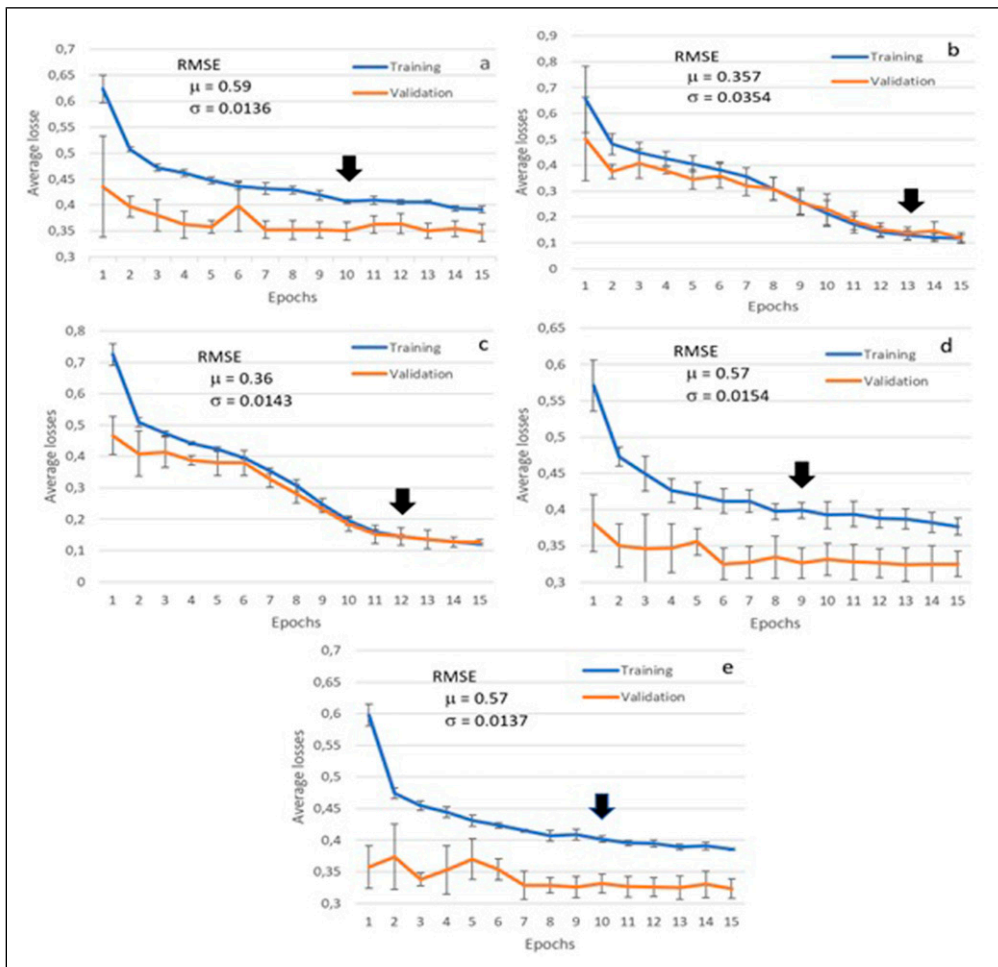
Hyperparameters	Value
Training and validation split rate	85/15%
Batch size	8
Number of epochs	15 (or until the learning process saturates)
Model compiler	Adagrad optimizer, learning rate = 0.01%
Loss function	RMSE (weighted)
Validation metric	RMSE

We observed the accuracy evolution and overfitting behavior using the loss for training and validation curves. Our network implemented the L2 Regularization (R2) to mitigate overfitting situations. The final parameters of the model, including the regularization parameter lambda, were obtained using a tuning process (Bayesian Optimization). We consider the output value a continuous float value to have a numeric idea about the errors. However, the Softmax function could also be used as the final layer.

## Results

### Single features experiments

The first experiments examined our BST model performance in relating single-longitudinal quality of life features to depression mood status (Figure 3). The curves represent the mean values



**Figure 3.** QoL@TbA training and validation loss curves for models using a unique feature: (a) marital status, (b) high-pressure diagnosis, (c) diabetes diagnosis, (d) sleep quality, and (e) level of physical activity.

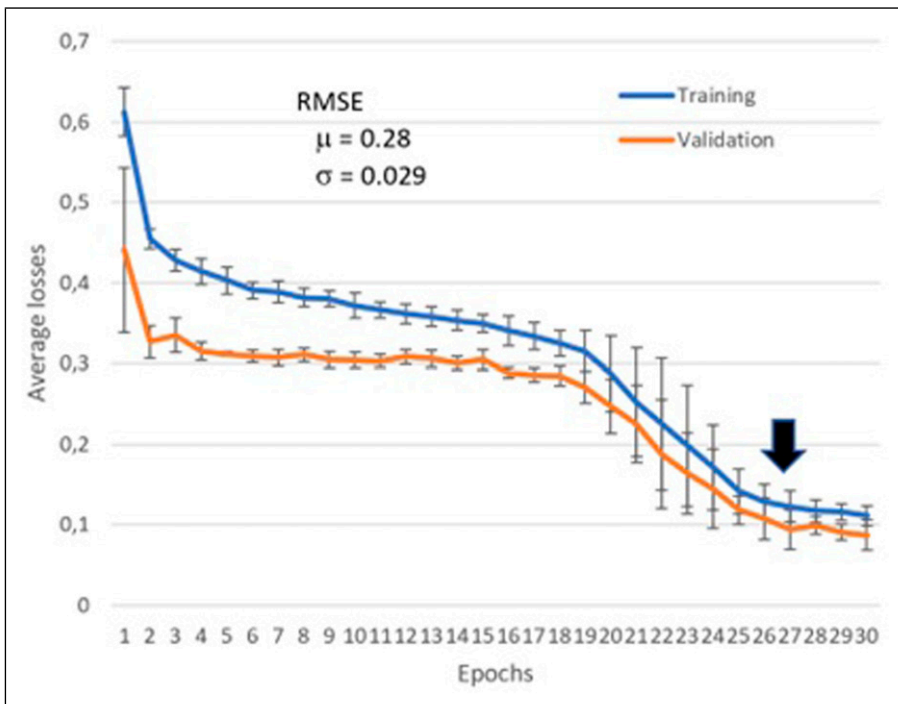
(10 rounds), and their standard deviations. The graphs show that the model can reduce the loss (difference between predicted and real mood representation values  $<1, 2 \text{ or } 3>$ ) during the training phase for all features. However, the validation curves are very smooth in most cases (Figure 3(a), 3(d), 3(e)), and their loss reductions are not very evident. The curves also show that the learning process saturates between epochs 6 and 13 for such experiments (black arrows in the graph), given the simplicity of the models (single feature embedding). Figure 3 also shows the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) for RMSE values. Such RMSE values are lower than the baseline (0.71), mainly when the model uses high-pressure diagnosis (0.357, Figure 3(c)) as the training feature.

### Multifeature experiment

This experiment used all features together (Figure 4). In this case, the training and validation loss curves present a sharper decreasing behavior. Moreover, the RMSE was improved to 0.28, with the learning saturation point at epoch 27. This high number of epochs was expected since the model presents a higher complexity. After that point, the model could not return any meaningful gains, and the curves became almost flat.

### Influence analysis of behavior changes

This experiment aimed at verifying if and which behaviors could mitigate a state of depression at wave 4. Therefore, we randomly selected 10 individuals who presented depression in their fourth wave. After that, we generated behavior patterns for this fourth wave and applied the BST model to



**Figure 4.** QoL@TbA training and validation loss curves for the multifeature model.

verify the influence of these generated patterns on the individuals' moods (e.g., what could have happened if the individual had slept better?).

Equation (1) shows the number of possible generated behavior patterns  $\gamma$ , considering the number of classes of each feature  $a$ , where  $n$  represents the number of features. For example, our five categorical attributes have the following possible classes: marital status = 6, high pressure = 2, diabetes = 2, sleep quality = 2, and level of physical activity = 3. Thus, the model should evaluate 144 patterns.

$$\gamma = \prod_{i=1}^n \text{Classes}(a_i) \quad (1)$$

The model identified behavioral patterns that could modify the mood from depressed to normal in only two cases (Table 3). The first case indicates, as a unique recommendation, that the individual should marry again and increase his/her level of physical activity to avoid depression.

The second case returns two recommendations. Part of the first recommendation (Recommendation 1) may indicate that the individual must still be married. However, the first and second recommendations for this individual again emphasize the importance of physical activity as an option to avoid depression.

## Discussion

### *Multifeature longitudinal data*

Multifeatured models obtained better results as they captured more complex relationships and interactions among various variables influencing the outcome. Theoretically, an increased number of features can enhance results, provided the correlations among these inputs are low and the dataset has sufficient data to avoid a sparse search space. Given the limitations of the ELSA dataset, we chose to maintain only five input features in our model. Additionally, we selected features that influence depression based on the specialized literature, including marital status,<sup>19</sup> high-pressure diagnosis,<sup>20</sup> diabetes diagnosis,<sup>21</sup> sleep quality,<sup>22</sup> and level of physical activities.<sup>23</sup> These features do not exhibit direct bidirectional correlations, thereby contributing valuable information to the problem. Furthermore, we observed that multifeatured models can effectively distinguish signal from noise, reducing the impact of random fluctuations associated with any single feature.

**Table 3.** Examples of recommendations for behavior changes.

Case	Assessments (features)	Mood
Case I	Widowed; noHP; nodiabetes; goodsleep; active	Current = 3
	Widowed; noHP; nodiabetes; badsleep; active Widowed; noHP; nodiabetes; goodsleep; underactive Recommendation	Resultant = 1.280
Case II	Married; noHP; nodiabetes; goodsleep; veryactive	Current = 3
	Married; noHP; nodiabetes; badsleep; active	
	Married; noHP; nodiabetes; badsleep; active Recommendation 1	Resultant = 1.286
	ReMarried; noHP; nodiabetes; goodsleep; veryactive Recommendation 2 Married; noHP; nodiabetes; goodsleep; veryactive	Resultant = 1.284

The use of multi-feature longitudinal data was also important to improve the accuracy of predictions and support the analysis of results (explainability). For example, the results in Table 3 allow a comparative analysis between the previous and recommended behavioral patterns. Moreover, we can confront new hypotheses with results from previous studies. For example, according to the literature, “*married people have comparatively low depression rates because they are, for several reasons, emotionally less damaged by stressful experiences that are non-married people*”.<sup>19</sup> This statement corroborates our results. However, this statement may only be valid if associated with other QoL features, such as a high level of physical activity. A multifeature approach allows this type of analysis, which differs from previous studies focusing only on one dimension.<sup>19–23</sup>

### Scalability

Our study presented and validated the concept of BST for longitudinal multifeature QoL data. However, this case example is simple when compared with real scenarios. For example, we only used five features as available in the ELSA dataset, configuring a scenario with low dimensionality. For example, these features allow 144 behavioral patterns (see equation (1)). Complex QoL domains may present more than 24 features (WHOQOL, 1995). If a domain has 30 features, each with five classes, the dimensionality exponentially increases to  $5^{30}$ , or approximately  $931 \times 10^{18}$  combinations. In this case, the use of meta-heuristics (e.g., genetic algorithms) is essential to reduce the search space of behavioral patterns.

### Practical implications

Health recommendations in the real world must consider three main aspects. First, recommendations for behavior change should closely align with current behaviors; for instance, we restrict the choice of new values to those in proximity to existing values. Second, the assessment of inputs should minimize user interactions and be transparent, utilizing passive data collection methods such as accelerometers. Finally, explainability is crucial for clarifying the reasons behind recommendations, which in turn supports user engagement.

### Limitations and future works

This study had three main limitations. We trained the model using only four waves. Thus, we could not evaluate the transformer’s ability to analyze long sequences when applied to multifeature data. As far as we know, the literature does not present this type of analysis yet. The number of samples (2682) is also low, considering the search space of the problem. Thus, overfitting was a challenge during the training stage. However, the main limitation is the reliability of the data. Some of its values are collected using questionnaires that summarize long periods. For example, ELSA summarizes the sleep quality of 2 years in a single binary class (“good”, “poor”). Thus, the use of a longer and larger (more samples) as well as a more granular dataset based on passive data is part of our ongoing work.

It is important to note that our architecture has a dense layer as the last network component. Thus, it returns a continuous value, characterizing a regression rather than a classification model. This strategy is interesting when the aim is to return probabilistic outputs and have insights into the likelihood of belonging to a particular class. Thus, RMSE is used since the resultant error is in the same units as the target variable. However, this strategy to use a regression rather than a

classification model also has disadvantages. Regression assumes a continuous output space, which might not align with the discrete nature of class labels. Thus, it can lead to misinterpretation of predictions and loss of class information, which could be a limitation depending on the task.

In future works, we intend to use the same strategy with data generated by wearable devices, which produce passive, continuous, and diverse behavioral data streams. For example, data about sleep quality (e.g., sleep duration, sleep efficiency), physical activity (e.g., level of intensity and distribution of activities along the day), physiological signals (e.g., heart rate variability, respiration), and others. Some of these data bring additional challenges since our approach relies on categorical input data. Thus, their use requires a process of data categorization that summarizes the data without losing quality. Besides own efforts in data collection, available datasets (e.g., UK Biobank and *All of Us* Research Program) provide long streams of health-related data and can be very useful for multifeature longitudinal recommendation research. Moreover, such long-term datasets are also helpful in supporting the validation of approaches regarding scalability and the use of long multifeature sequences.

The integration of explainable/interpretable strategies is another possible future work derived from our investigation. Our current approach returns a combination of features (i.e., recommendations) that, if implemented, could improve the long-term health aspects of individuals since such features work as a guide for possible behavioral changes toward healthy routines. Therefore, in this version, experts' opinions account for the interpretability of the recommendations' outcomes. However, considering the new trends regarding regulations for critical AI domains<sup>27</sup> (e.g., health), self-explainable models will be compulsory and part of our short-term research directions.

## Conclusion

Unlike traditional recommendation systems that aim to predict potentially interesting outputs for users, the QoL@TbA model supports a multifeatured longitudinal data analysis to provide recommendations for concurrent modifications across multiple input features. Additionally, static data (e.g., user profiles) is incorporated into the model to create more personalized recommendations for behavior change. To the best of our knowledge, this use of transformers represents a novel direction (multifeatured longitudinal data analysis for recommendation support and recommendations for concurrent modifications in multiple input features) with no prior efforts in this area. Supplementary material discusses the training algorithm, pre-processing strategies for missing and imbalanced data, and the computational complexity of this approach.

## Author contribution

Siebra: Methodology, conceptualization, experiments, writing. Kurpicz-Briki: Conceptualization and review. Wac: Conceptualization and review.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the European Commission Horizon 2020 research - Marie Skłodowska-Curie Action (H2020-MSCA-IF-2020-101024693).

## Ethical statement

Given the nature of the utilized data, which are publicly available, there is no cause for concern regarding ethical considerations in this study.

## ORCID iD

Clairton Siebra  <https://orcid.org/0000-0002-9742-2175>

## Data availability statement

The data were made available through the UK Data Archive (<https://www.elsa-project.ac.uk/>).

## References

1. Kim S. World Health Organization quality of life (WHOQOL) assessment. In: *Encyclopedia of quality of life and well-being research*, 2020, pp. 1–2.
2. Haraldstad K, Wahl A, Andenæs R, et al. A systematic review of quality of life research in medicine and health sciences. *Qual Life Res* 2019; 28(10): 2641–2650. DOI: [10.1007/s11136-019-02214-9](https://doi.org/10.1007/s11136-019-02214-9).
3. Manea V and Wac K. mQoL: mobile quality of life lab: from behavior change to QoL. In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 2018, pp. 642–647. DOI: [10.1145/3267305.3267549](https://doi.org/10.1145/3267305.3267549).
4. Virginia Anikwe C, Friday Nweke H, Chukwu Ikegwu A, et al. Mobile and wearable sensors for data-driven health monitoring system: state-of-the-art and future prospect. *Expert Syst Appl* 2022; 202: 117362. DOI: [10.1016/j.eswa.2022.117362](https://doi.org/10.1016/j.eswa.2022.117362).
5. Vaswani A, Shazeer M, Parmar N, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017; 30: 6000–6010. DOI:[10.5555/3295222.3295349](https://doi.org/10.5555/3295222.3295349).
6. Savcicens G, Eliassi-Rad T, Hansen LK, et al. Using sequences of life-events to predict human lives. *Nat Comput Sci* 2023; 4: 43–56.
7. Li Y, Rao S, Solares JRA, et al. BEHRT: transformer for electronic health records. *Sci Rep* 2020; 10(1): 7155–7162. DOI: [10.1038/s41598-020-62922-y](https://doi.org/10.1038/s41598-020-62922-y).
8. Rasmy L, Xiang Y, Xie Z, et al. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ Digit Med* 2021; 4(1): 13–86.
9. Tran TNT, Felfernig A, Trattner C, et al. Recommender systems in the healthcare domain: state-of-the-art and research issues. *J Intell Inf Syst* 2021; 57: 171–201.
10. Ko H, Lee S, Park Y, et al. A survey of recommendation systems: recommendation models, techniques, and application fields. *Electronics* 2022; 11(1): 141.
11. Delpino FM, Costa ÂK, Farias SR, et al. Machine learning for predicting chronic diseases: a systematic review. *Publ Health* 2022; 205: 14–25. DOI: [10.1016/j.puhe.2022.01.007](https://doi.org/10.1016/j.puhe.2022.01.007).
12. Chen Q, Zhao H, Li W, et al. Behavior sequence transformer for e-commerce recommendation in alibaba. In: *Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data*, 2019, vol 12, pp. 1–4. DOI: [10.1145/3326937.3341261](https://doi.org/10.1145/3326937.3341261).
13. Banks J. *English longitudinal study of ageing: waves 0-9, 1998-2019 [data collection]*. 36th edn. UK Data Service, 2021.
14. Nagaraj P and Deepalakshmi P. An intelligent fuzzy inference rule-based expert recommendation system for predictive diabetes diagnosis. *Int J Imag Syst Technol* 2022; 32(4): 1373–1396.
15. Amoon M, Altameem T and Hashem M. Consistent healthcare safety recommendation system for preventing contagious disease infections in human crowds. *Sensors* 2023; 23(23): 9394.

16. Sadhasivam S, Sarvesvaran MS, Prasanth P, et al. Diet and workout recommendation using ML. In: *2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, 2023, pp. 1–4.
17. Lakshmanan V, Robinson S and Munn M. *Machine learning design patterns*. O'Reilly Media, 2020.
18. Razavi R, Gharipour A and Gharipour M. Depression screening using mobile phone usage metadata: a machine learning approach. *J Am Med Inf Assoc* 2020; 27(4): 522–530. DOI: [10.1093/jamia/ocz221](https://doi.org/10.1093/jamia/ocz221).
19. Kessler RC and Essex M. Marital status and depression: the importance of coping resources. *Soc Forces* 1982; 61(2): 484–507. DOI: [10.2307/2578238](https://doi.org/10.2307/2578238).
20. Lee C and Kim H. Machine learning-based predictive modeling of depression in hypertensive populations. *PLoS One* 2022; 17(7): e0272330.
21. Nouwen A, Adriaanse MC, van Dam K, et al. Longitudinal associations between depression and diabetes complications: a systematic review and meta-analysis. *Diabet Med* 2019; 36(12): 1562–1572. DOI: [10.1111/dmc.14054](https://doi.org/10.1111/dmc.14054).
22. Tsuno N, Besset A and Ritchie K. Sleep and depression. *J Clin Psychiatry* 2005; 66(10): 1254–1269. DOI: [10.4088/jcp.v66n1008](https://doi.org/10.4088/jcp.v66n1008).
23. Matias TS, Lopes MVV, da Costa BGG, et al. Relationship between types of physical activity and depression among 88,522 adults. *J Affect Disord* 2022; 297: 415–420.
24. Siebra C, Amorim L, Quintino JP, et al. Behaviour recommendations with a deep learning model and genetic algorithm for health debt characterisation. *J Biomed Inf* 2023; 137: 104277.
25. Michel P, Levy O and Neubig G. Are sixteen heads really better than one? *Adv Neural Inf Process Syst* 2019: 32.
26. Morales-Hernández A, Van Nieuwenhuysse I and Rojas Gonzalez S. A survey on multi-objective hyperparameter optimization algorithms for machine learning. *Artif Intell Rev* 2023; 56(8): 8043–8093.
27. Hupont I, Micheli M, Delipetrev B, et al. Documenting high-risk AI: a European regulatory perspective. *Computer* 2023; 56(5): 18–27.
28. Tan X, He Y, Ning N, et al. Shared decision-making in the treatment of adolescents diagnosed with depression: a cross-sectional survey of mental health professionals in China. *J Psychiatr Ment Health Nurs* 2024; 31(3): 340–351.