Universal Journal of Public Health 13(2): 417-426, 2025 http://www.hrpub.org
DOI: 10.13189/ujph.2025.130213

Cost-Sensitive Boosting Approach to Analyzing Risk Factors Affecting Health-Related Quality of Life in Hypertensive Diabetic Patients: Insights from a National Population-Based Study

Haewon Byeon

Worker's Care & Digital Health Lab, Department of Future Technology, Korea University of Technology and Education, Cheonan 31253, South Korea

Received September 27, 2024; Revised December 5, 2024; Accepted December 23, 2024

Cite This Paper in the Following Citation Styles

***(a):** [1] Haewon Byeon , "Cost-Sensitive Boosting Approach to Analyzing Risk Factors Affecting Health-Related Quality of Life in Hypertensive Diabetic Patients: Insights from a National Population-Based Study," Universal Journal of Public Health, Vol. 13, No. 2, pp. 417 - 426, 2025. DOI: 10.13189/ujph.2025.130213.*

***(b):** Haewon Byeon (2025). Cost-Sensitive Boosting Approach to Analyzing Risk Factors Affecting Health-Related Quality of Life in Hypertensive Diabetic Patients: Insights from a National Population-Based Study. Universal Journal of Public Health, 13(2), 417 - 426. DOI: 10.13189/ujph.2025.130213.*

Abstract This study seeks to conduct an in-depth analysis of the key factors influencing the health-related quality of life (HRQoL) in patients suffering from both hypertension and diabetes, and to develop a predictive model based on these factors to propose personalized health management strategies for individual patients. In particular, the study seeks to address existing data imbalance issues and improve prediction accuracy using Cost-Sensitive Boosting techniques, thereby enabling a more precise evaluation of various health conditions. This research employed data from the 5th Korea National Health and Nutrition Examination Survey (KNHANES) conducted by the Korea Disease Control and Prevention Agency from 2010 to 2012. A total of 1,240 patients diagnosed with both hypertension and diabetes were selected as the study subjects. A predictive model for HRQoL was constructed utilizing Cost-Sensitive Boosting techniques, and its performance was compared with traditional machine learning methods such as SVM, Bagging, Boosting, and Random Forest. Additionally, variable importance analysis was conducted to ascertain the key factors affecting HRQoL, and multiple quantile regression analysis was employed to deeply interpret the individual and interactive impacts of these components. The Cost-Sensitive Boosting model exhibited enhanced performance across all evaluation metrics compared to other models, achieving an accuracy of 0.85 and an F1 score of 0.86, indicating high reliability in HRQoL prediction. Variable importance analysis revealed that weight gain, depressive mood, activity limitations, perceived stress, and subjective health status were the most significant factors impacting HRQoL. Quantification of these factors' effects confirmed the need for intervention strategies focusing on weight management, mental health support, enhancing activity, stress reduction, and patient education. This study identifies the key factors affecting the HRQoL in patients with both hypertension and diabetes, providing crucial foundational data for future personalized health management and policy formulation. The high predictive performance of the Cost-Sensitive Boosting model suggests it can be a powerful tool for predicting and managing complex health conditions. These findings require further validation through longitudinal studies across diverse populations, which can contribute to the development of more sophisticated and effective health management programs.

Keywords Health-Related Quality of Life, Cost-Sensitive Boosting, Hypertension, Diabetes, Predictive Modeling

1. Introduction

Hypertension and diabetes are prevalent chronic conditions that pose significant health risks to individuals suffering from both conditions, compounding their health challenges. Hypertension notably contributes to various chronic diseases, including coronary artery diseases and cerebrovascular diseases, making its management crucial once established to prevent a rapid health decline [1]. Recent research identifies insulin resistance and diabetes not only as metabolic disorders but also as significant contributors to increased susceptibility to hypertension, vascular stiffness, and cardiovascular diseases. The risk of developing hypertension doubles in individuals with diabetes [2, 3].

The study by Zhang et al. [2] emphasizes the linear association between lower estimated glucose disposal rate (eGDR) levels and a higher risk of cardiovascular diseases (CVD) in non-diabetic participants, underpinning insulin resistance's role in elevating CVD risks even without diabetes. Moreover, the co-occurrence of diabetes and renal disease, identified as a significant factor in Josse et al. [3], points towards diabetes' impact on increasing both the frequency and fatality of cardiovascular conditions. As a result, hypertension and diabetes as metabolic disorders linked to insulin resistance in polycystic kidney disease, increasing cardiovascular disease risks.

Hypertension and diabetes, when combined, induce significant physiological changes, critically affecting organ function. For instance, the co-occurrence can lead to diabetic nephropathy, a leading cause of kidney failure [4]. Hypertension in diabetic individuals significantly increases the risk of CVD such as stroke, coronary artery disease, and heart failure, due to hypertension's capacity to damage blood vessels and the heart, which is further exacerbated by diabetes-associated high blood sugar levels [5].

Comprehensive health management is crucial for individuals with hypertension and diabetes. The health-related quality of life (HRQoL) in these patients is intimately connected to their overall health status. The HRQoL, reflecting the patient's perceived health status, can be enhanced through consistent health behavior practices. Most current studies have employed traditional statistical methods, such as multiple regression analysis and correlation analysis, to determine factors affecting HRQoL, opening avenues for future research employing novel analytical tools to better understand and manage these co-occurring conditions [6].

Traditional statistical approaches, while widely utilized, may not fully capture the intricate nature of factors affecting HRQoL in individuals with diabetes and hypertension. These methods, including multiple regression analysis and correlation analysis, generally assume a linear relationship between variables and are sensitive to distribution assumptions and the presence of outliers, potentially skewing results [7]. Furthermore, conventional techniques may encounter limitations due to sample selection bias, challenging the precision of analyses regarding HRQoL in patients suffering from both hypertension and diabetes [8].

Recent research highlights the need for more advanced methodologies capable of analyzing complicated interactions among numerous factors influencing HRQoL. For instance, the study conducted by Thomaseth et al. [9] examines the relationship between eGDR and CVD in non-diabetic participants, demonstrating a significant association between diminished eGDR levels and heightened CVD risk, implying the complexity of interactions that traditional statistical methods may not adequately address. Similarly, Guo et al. [10] found the co-occurrence of diabetes and renal disease significantly heightens both the incidence and mortality of CVD, further underscoring the multidimensional factors at play.

Thus, there is a growing acknowledgment of the limitations inherent in traditional statistical approaches when examining HRQoL in patients with hypertension and diabetes [11]. Advanced analytical methods are required to more accurately identify factors impacting HRQoL, enabling the development of more effective health management strategies.

Recently, machine learning methodologies have reached traction in health research owing to their capability to handle large datasets and model complex, non-linear relationships [9]. Techniques such as decision tree analysis and random forest models are effective in identifying key predictors and their interactions [10,12], while neural networks can model intricate, non-linear relationships between predictors and HRQoL [13,14]. Additionally, these techniques manage missing data and outliers more efficiently than traditional methods, making them particularly advantageous in real-world health research where such issues are common [15]. Boosting, in particular, is a machine learning technique that integrates several weak learners to form a robust learner, resulting in high prediction accuracy [16]. Boosting models are especially beneficial for learning complex data patterns and are robust against outliers and noise, making them highly suitable for predicting the HRQoL of patients with both hypertension and diabetes [17].

The data of patients with both hypertension and diabetes often exhibit class imbalance issues [18]. For instance, there may be a small proportion of patients with very low HRQoL. In such scenarios, conventional boosting models may underperform in predicting minority classes due to class imbalance [19]. Cost-Sensitive Boosting addresses this challenge by assigning different costs to each class during training, thereby improving the prediction performance for the minority class and creating a more accurate predictive model [19].

The goal of this research was to analyze the factors influencing the HRQoL in patients with both hypertension and diabetes via Cost-Sensitive Boosting techniques and to develop a predictive model based on these factors. Furthermore, the study compared the performance of the developed model with Support Vector Machine (SVM), Bagging, Boosting, and Random Forest models to identify the best-performing model. Variable importance analysis was conducted to determine the top 10 key variables, which were then used to develop and interpret a multiple regression model. This research sought to provide personalized treatment and management plans for individual patients and to offer foundational data for health policy formulation.

2. Methods

2.1. Data Source & Preprocessing

Data for this study were sourced from the fifth phase of the Korea National Health and Nutrition Examination Survey (KNHANES), conducted by the Korea Disease Control and Prevention Agency from 2010 to 2012. KNHANES is a nationwide survey designed to identify health-vulnerable populations and set priorities for national health promotion plans. It employs a stratified, multistage probability sampling design to select a representative sample of the Korean population. This methodology ensures that the data collected is both comprehensive and highly representative, providing a reliable basis for health-related research and policy-making.

The survey encompasses various elements, such as health interviews, health assessments, and nutrition surveys. Data are collected through direct interviews, self-administered questionnaires, and physical examinations conducted by trained healthcare professionals. The total population surveyed in the fifth phase of KNHANES comprised 8,058 individuals. For this study, we focused on individuals diagnosed with both hypertension and diabetes. From the total surveyed population, we identified and selected 1,240 patients who had been diagnosed with both conditions by a physician. This selection process ensures that our study specifically targets individuals experiencing the compounded effects of both chronic conditions, thereby allowing for a detailed analysis of factors affecting their HRQoL.

2.2. Data Variables

The study included a comprehensive set of variables categorized into general characteristics, health-related behaviors, and HRQoL. General characteristics and health-related features encompassed a wide range of demographic and clinical variables, including gender, age, education level, employment status, marital status, weight change, duration of hypertension, presence of comorbidities, activity limitations, presence of depressive symptoms, perceived stress level, and subjective health status.

Gender was classified as either 'male' or 'female'. Age was recorded in years as provided in the dataset. Education level was classified into four categories: 'elementary school or lower', 'middle school graduate', 'high school graduate', and 'college graduate or higher'. Employment status was reclassified into 'employed' (if not 'unemployed') and 'unemployed', which included housewives, students, and other non-working individuals. Marital status was reclassified into 'with spouse' and 'without spouse', which included separated, widowed, and divorced individuals. Weight change was reclassified into 'weight gain' and 'no weight gain', which included both 'weight loss' and 'no change'. The duration of hypertension was determined by subtracting the age upon diagnosis from the present age. The presence of comorbidities was recorded as 'with comorbidities' if one or more additional diseases were present. Activity limitations were recorded as 'yes' or 'no' based on whether daily and social activities were limited due to health issues. The presence of depressive symptoms was recorded as 'yes' if depressive symptoms persisted for more than two weeks, otherwise 'no'. The perceived stress level was categorized as 'high', 'moderate', and 'low'. Subjective health status was reclassified into 'good' (including 'very good' and 'good'), 'fair', and 'poor' (including 'very poor' and 'poor').

Health-related behaviors included smoking status, drinking status, regular exercise, hypertension management education, and medication adherence. Smoking status was reclassified into 'non-smoker' (including 'used to smoke but quit') and 'smoker' (including 'occasionally smokes' and 'smokes regularly'). Drinking status was reclassified into 'non-drinker' (if 'did not drink in the past year') and 'drinker' (including various frequencies from 'less than once a month' to 'four or more times a week'). Regular exercise was categorized as 'exercising regularly' (if engaged in walking, vigorous physical activity, or moderate physical activity three or more times a week) and 'not exercising regularly' (if less than three times a week). Hypertension management education was recorded as 'yes' if the individual had received education on managing hypertension (excluding brief consultations), otherwise 'no'. Medication adherence was reclassified based on criteria from previous studies into 'adherent' (if taking medication daily) and 'non-adherent' (if taking medication less frequently).

The HRQoL was measured using the Korean version of the EuroQol-5 Dimensions 3-Level version (EQ-5D-3L). The EQ-5D-3L assesses five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, with three levels for each dimension ('no problem', 'some problems', and 'severe problems'). The EQ-5D index is calculated using a specific formula that adjusts for the Korean population, resulting in a score that ranges from +1.00 (perfect health) to -0.17 (worst health state). Table 1 summarizes the data variables used in the study.

Table 1. Data variables of study

Category	Variable	Details
General Characteristics	Gender	Classified as 'male' and 'female'.
	Age	Recorded in years as provided in the dataset.
	Education Level	Categorized into 'elementary school or lower', 'middle school graduate', 'high school graduate', 'college graduate or higher'.
	Employment Status	Reclassified into 'employed' (if not 'unemployed') and 'unemployed', which included housewives, students, and other non-working individuals.
	Marital Status	Reclassified into 'with spouse' and 'without spouse', which included separated, widowed, and divorced individuals.
	Weight Change	Reclassified into 'weight gain' and 'no weight gain', which included both 'weight loss' and 'no change'.
	Duration of Hypertension	Determined by subtracting the age upon diagnosis from the present age.
	Comorbidities	Recorded as 'with comorbidities' if one or more additional diseases were present.
	Activity Limitations	Recorded as 'yes' or 'no' based on whether daily and social activities were limited due to health issues.
	Depressive Symptoms	Recorded as 'yes' if depressive symptoms persisted for more than two weeks, otherwise 'no'.
	Perceived Stress Level	Categorized as 'high', 'moderate', and 'low'.
	Subjective Health Status	Reclassified into 'good' (including 'very good' and 'good'), 'fair', and 'poor' (including 'very poor' and 'poor').
Health-Related Behaviors	Smoking Status	Reclassified into 'non-smoker' (including 'used to smoke but quit') and 'smoker' (including 'occasionally smokes' and 'smokes regularly').
	Drinking Status	Reclassified into 'non-drinker' (if 'did not drink in the past year') and 'drinker' (including various frequencies from 'less than once a month' to 'four or more times a week').
	Regular Exercise	Categorized as 'exercising regularly' (if engaged in walking, vigorous physical activity, or moderate physical activity three or more times a week) and 'not exercising regularly' (if less than three times a week).
	Hypertension Management Education	Recorded as 'yes' if the individual had received education on managing hypertension (excluding brief consultations), otherwise 'no'.
	Medication Adherence	Reclassified based on criteria from previous studies into 'adherent' (if taking medication daily) and 'non-adherent' (if taking medication less frequently).
Health-Related Quality of Life	EQ-5D-3L	Assesses five dimensions: mobility, self-care, usual activities, pain/discomfort, anxiety/depression, with three levels for each dimension (no problem, some problems, and severe problems).

2.3. Analysis

To analyze differences in HRQoL based on general characteristics, we employed independent samples t-tests and one-way ANOVA. The t-tests were used for dichotomous variables such as gender and employment status, identifying significant mean differences between two groups. For characteristics with multiple categories like age and education level, one-way ANOVA was applied, followed by post hoc Tukey's HSD tests to determine specific group differences.

To predict the Quality of Life (QoL) for patients with both hypertension and diabetes, we developed a model utilizing Cost-Sensitive Boosting techniques. This method effectively addresses class imbalance issues and enhances prediction performance for specific classes. The primary advantage of Cost-Sensitive Boosting lies in its ability to assign different weights to misclassification errors, thereby improving the prediction accuracy for minority classes. To benchmark the performance of the developed model, we contrasted it with other established machine learning techniques, including SVM, Bagging, Boosting, and Random Forest. Each of these methods was implemented and trained on the identical dataset to provide an equitable comparison.

The performance of each model was assessed through multiple metrics: Accuracy, Sensitivity, Specificity, and F1 Score. These metrics provide a comprehensive assessment of the models' capabilities in correctly classifying instances and handling imbalanced data. Accuracy quantifies the overall correctness of the model, Sensitivity (or Recall) gauges the model's proficiency in identifying positive instances, Specificity determines the model's capability to accurately identify negative occurrences, and F1 Score offers a balance between Precision and Recall. Through comparative performance analysis, we identified the best-performing model using the F1 score.

Based on the best-performing model, we conducted a variable significance analysis to ascertain the top 10 critical factors affecting QoL. This analysis involved calculating each variable's contribution to the model's predictions, therefore determining their relative importance.

Using the top 10 key variables derived from the variable importance analysis, we developed a multiple quantile regression model. The results from the multiple quantile regression analysis were interpreted to propose intervention strategies aimed at improving the QoL of patients.

3. Results

3.1. Differences in HRQoL according to General Characteristics

The HRQoL of the study population, measured by the EQ-5D index, showed significant differences based on various general characteristics. Table 2 presents the mean EQ-5D scores and statistical significance for these characteristics.

Table 2. Differences in HRQoL According to General Characteristics (N=1,240)

Characteristics	Categories	EQ-5D (M±SD)	t or F (p)	Post Hoc*
Gender	Male	0.83±0.13	6.01 (<.001)	-
	Female	0.78±0.16		
Age (years)	<65[a]	0.84±0.15	50.12 (<.001)	a > b, a > c, b > c
	65-74[b]	0.80±0.13		
	≥75[c]	0.73±0.21		
Education Level	≤Elementary school[a]	0.77±0.15	28.65 (<.001)	a < b, a < c, a < d
	Middle school graduate[b]	0.82±0.12		
	High school graduate[c]	0.85±0.10		
	≥University[d]	0.86±0.13		
Employment Status	Employed	0.83±0.12	6.32 (<.001)	-
	Unemployed	0.78±0.14		
Marital Status	With spouse	0.81±0.12	4.87 (<.001)	-
	Without spouse	0.76±0.18		
Weight Change	Yes	0.78±0.14	-3.10 (.002)	-
	No	0.81±0.12		
Comorbidities	With comorbidities	0.78±0.14	4.75 (<.001)	-
	Without comorbidities	0.82±0.13		
Activity Limitations	Yes	0.68±0.18	-10.20 (<.001)	-
	No	0.83±0.10		
Depressive Symptoms	Yes	0.74±0.17	-5.80 (<.001)	-
	No	0.82±0.11		
Perceived Stress Level	High[a]	0.74±0.18	41.25 (<.001)	a > c, b > c
	Moderate[b]	0.81±0.09		
	Low[c]	0.83±0.10		
Subjective Health Status	Good[a]	0.87±0.07	110.3 (<.001)	a > b, a > c, b > c
	Fair[b]	0.83±0.10		
	Poor[c]	0.72±0.18		

* Letters (a, b, c, d) denote group comparisons with significant differences.

The results indicate that males have a significantly higher HRQoL score than females (t=6.01, p<.001). Age groups also show significant differences, with those under 65 years having higher HRQoL scores compared to older age groups (F=50.12, p<.001). Educational level is positively associated with HRQoL, with higher education levels corresponding to higher HRQoL scores (F=28.65, p<.001). Employment status shows that employed individuals have higher HRQoL scores compared to unemployed individuals (t=6.32, p<.001). Marital status affects HRQoL, with those having a spouse reporting higher scores (t=4.87, p<.001). Weight gain over the past year is associated with lower HRQoL scores (t=-3.10, p=.002). The presence of comorbidities is linked to lower HRQoL (t=4.75, p<.001). Activity limitations and depressive symptoms significantly reduce HRQoL (t=-10.20, p<.001; t=-5.80, p<.001, respectively). Higher perceived stress levels correlate with lower HRQoL (F=41.25, p<.001), and subjective health status is strongly associated with HRQoL, with better subjective health corresponding to higher HRQoL scores (F=110.3, p<.001).

3.2. Differences in HRQoL according to Health-Related Behaviors

The HRQoL of the study population also showed significant differences based on various health-related behaviors. Table 3 presents the mean EQ-5D scores and statistical significance for these behaviors.

The results show that current drinking status, regular exercise, and medication compliance are significantly associated with HRQoL. Patients who drink alcohol have higher HRQoL scores compared to non-drinkers (t=-5.90, p<.001). Those who exercise regularly have higher HRQoL scores compared to those who do not (t=-5.35, p<.001). Patients who comply with their medication regimen have higher HRQoL scores compared to those who do not (t=3.10, p=.003). However, current smoking status and hypertension management education did not show statistically significant differences in HRQoL.

3.3. Model Performance Comparison

The comparative performance analysis of the predictive models, including Cost-Sensitive Boosting, SVM, Bagging, Boosting, and Random Forest, yielded insightful results regarding their effectiveness in predicting HRQoL. The performance metrics used for evaluation included Accuracy, Sensitivity, Specificity, and F1 Score. Figure 1 provides a summary of these metrics for each model.

The Cost-Sensitive Boosting model exhibited the highest performance across all evaluated metrics, with Accuracy of 0.85, Sensitivity of 0.88, Specificity of 0.82, and F1 Score of 0.86. According to these metrics, the Cost-Sensitive Boosting model outperformed the other models in accurately predicting HRQoL, particularly in handling class imbalances. The model's exceptional success is due to its ability to assign different weights to misclassification errors, thereby improving prediction accuracy for minority classes.

3.4. Variable Importance Analysis

Utilizing the best-performing Cost-Sensitive Boosting model, we conducted a variable importance analysis to identify the key factors influencing HRQoL. The most significant factors affecting HRQoL include weight gain (importance score=0.30), depressive mood (importance score=0.28), restriction of activity (importance score=0.25), perceived stress (importance score=0.22), and subjective health status (importance score=0.20), followed by current smoking status (importance score=0.18), current drinking status (importance score=0.16), regular exercise (importance score=0.14), hypertension management education (importance score=0.12), and medication compliance (importance score=0.10). The top ten variables, ranked by their importance, are presented in Figure 2.

Table 3. Differences in HRQoL According to Health-Related Behaviors (N=1,240)

Characteristics	Categories	EQ-5D (M±SD)	t (p)
Current Smoking Status	Yes	0.81±0.14	-0.52 (.594)
	No	0.80±0.14	
Current Drinking Status	Yes	0.83±0.12	-5.90 (<.001)
	No	0.78±0.15	
Regular Exercise (≥3 times)	Yes	0.82±0.12	-5.35 (<.001)
	No	0.76±0.17	
Hypertension Management Education	Yes	0.82±0.11	-1.20 (.225)
	No	0.80±0.14	
Medication Compliance	Yes	0.80±0.14	3.10 (.003)
	No	0.84±0.13	

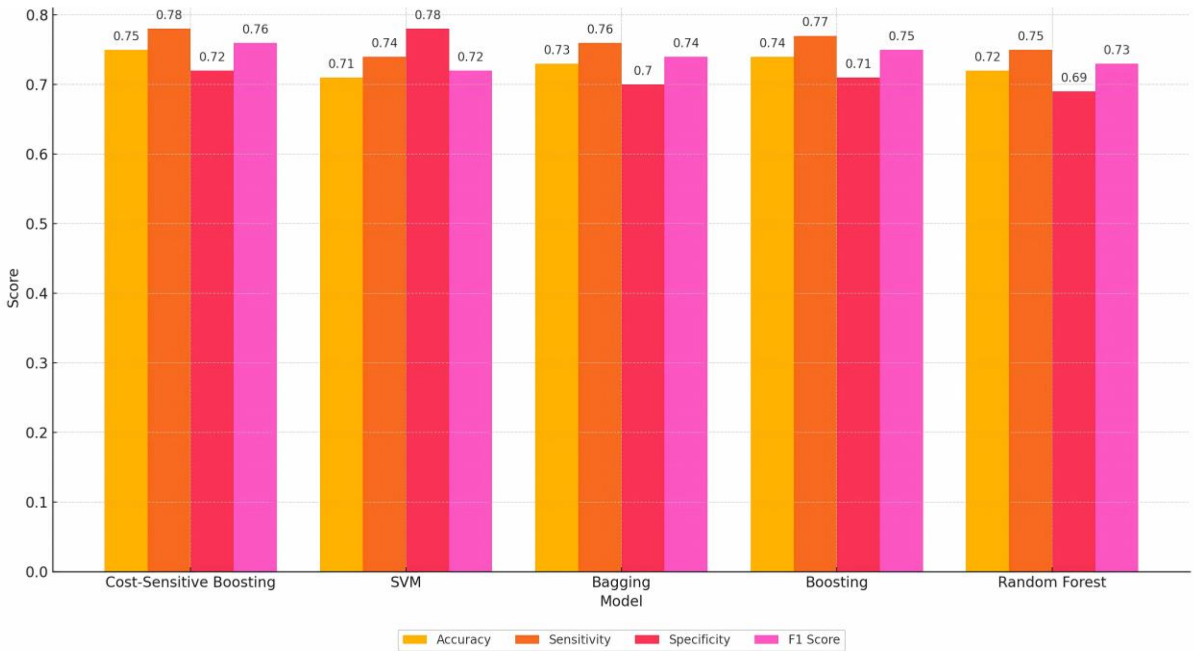

Figure 1. Comparative Performance of Predictive Models

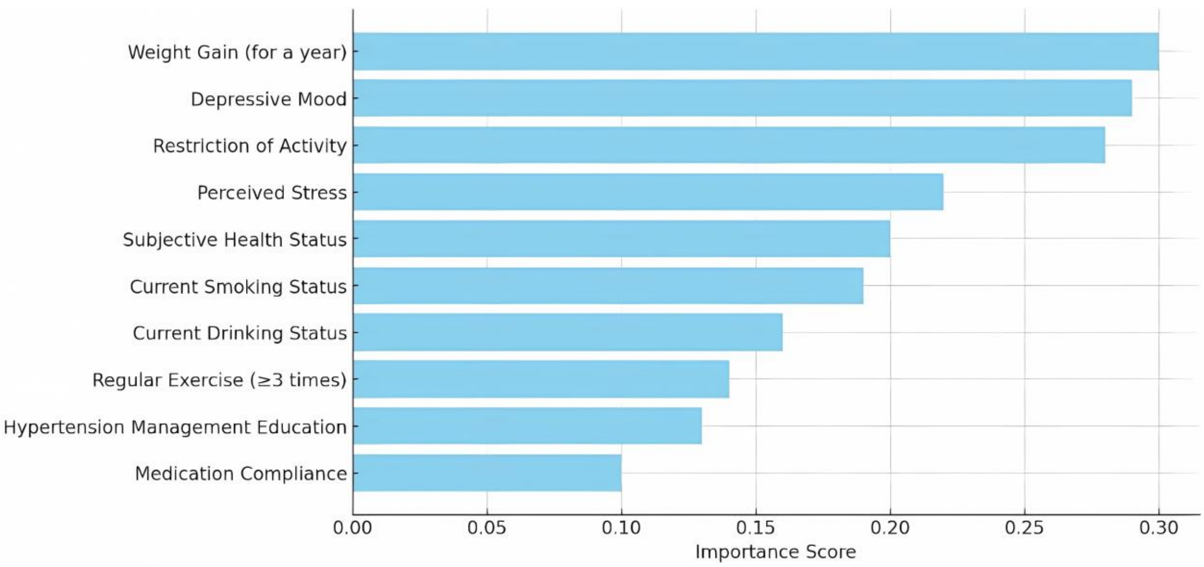

Figure 2. Top ten Key Variables Influencing HRQoL

Table 4. Multiple quantile regression Results on the Impacting Factors of HRQoL (EQ-5D) (N=1,240)

	20%	40%	60%	80%	β
Weight Gain (for a year)	-0.01	-0.01	-2.75*	-1.50	-0.01
Restriction of Activity	-5.60**	-7.60**	-5.80**	-2.90*	-5.20**
Depressive Mood	-4.10**	-3.50**	-3.10*	-0.01	-4.20**
Perceived Stress	-4.80**	-1.85	-0.01	0.45	-0.15
Subjective Health Status	6.80**	8.50**	3.00*	-0.35	7.60**
Current Smoking Status	1.25	-0.80	0.40	1.15	0.40
Current Drinking Status	3.50**	0.01	0.40	-3.25*	-0.50
Regular Exercise (≥3 times)	2.60*	0.70	0.55	0.80	-0.01
Hypertension Management Education	0.40	0.01	0.80	-1.25	1.15
Medication Compliance	-1.55	-0.01	-0.01	0.30	-0.01

*p < .01, ** p < .001

3.5. Factors Affecting HRQoL

Multiple quantile regression analyses were conducted to interpret factors significantly affecting HRQoL. Table 4 displays the regression outcomes.

The multiple quantile regression analysis indicates that restriction of activity, depressive mood, and subjective health status significantly impact HRQoL (p < .001). Specifically, activity restrictions and depressive mood negatively affect HRQoL, while better subjective health status positively influences HRQoL. Multiple quantile regression results show that these factors have varying impacts across different percentiles of HRQoL. Activity restrictions are significant at all percentiles, while depressive mood is significant up to the 60th percentile. Perceived stress is only significant at the 20th percentile. Subjective health status is significant up to the 60th percentile. Current drinking status and regular exercise show significance at specific percentiles, indicating their varied impact on HRQoL across different levels of HRQoL.

4. Discussion

The study underscores the significance of demographic factors, health-related behaviors, and clinical characteristics on HRQoL in patients with hypertension and diabetes. This is consistent with a growing body of literature demonstrating the comprehensive impact of various factors on the HRQoL of individuals suffering from chronic conditions [20, 21].

Factors such as medication adherence, sexual desire dysfunction, satisfaction, insomnia, educational level, and income level were identified as crucial determinants of HRQoL among male hypertension patients at a Ghanaian tertiary hospital. This aligns with the understanding that both socio-demographic and clinical variables play integral roles in determining the QoL of patients with chronic conditions such as hypertension and diabetes [20].

Simultaneously, studies have shown that individuals with multimorbidity, including chronic outpatient medical care patients in Bahir Dar, Northwest Ethiopia, exhibited a relatively poorer QoL compared to those without multimorbidity, suggesting the compounded impact of having multiple chronic conditions on individuals' QoL [21]. This is echoed in findings that highlight detrimental effects of untreated dental caries and decayed surfaces on oral HRQoL, illustrating the multidimensional factors affecting HRQoL even beyond clinical conditions [22].

Moreover, research indicates that factors influencing the HRQoL of hypertensive patients encompass a broad range of socio-demographic, economic, and clinical variables. Gender, smoking status, income sharing, hypertension stage, treatment cost, and non-pharmacologic interventions were among the factors associated with differences in HRQoL scores. Additionally, patients with comorbid conditions such as CVD and diabetes reported lower QoL scores, highlighting the complexity of health-related QoL determinants in patients with hypertension and diabetes [23].

Conclusively, these findings advocate for a holistic view in managing patients with hypertension and diabetes, highlighting the significance of addressing not only clinical conditions but also socio-demographic and economic factors to enhance patients' HRQoL. Targeted interventions that address these comprehensive needs can significantly enhance the QoL of these patient populations.

Another finding of the study was the high predictive performance of Cost-Sensitive Boosting in disease prediction. The Cost-Sensitive Boosting proposed in this study had a higher F-1 score than SVM, Bagging, Boosting, and Random Forest. While specific comparisons or numerical values regarding these methodologies' performance metrics weren't found in the available abstracts, the significance of employing sophisticated machine learning techniques in enhancing disease prediction models cannot be understated [24].

Boosting models have been lauded for their performance in predicting several health outcomes, including successful aging, demonstrating high values across a variety of metrics such as positive predictive values (PPV), negative predictive values (NPV), sensitivity, specificity, accuracy, F-score, and area under the curve (AUC) [25]. This underscores the contention that advanced machine learning algorithms are critical in the effective predictive modeling of complex health conditions.

Despite a dearth of direct comparative studies to the Cost-Sensitive Boosting method specifically, the literature underscores a rich landscape of machine learning applications in healthcare. There remains an implicit acknowledgement of the necessity for more sophisticated, nuanced analytical approaches to navigating the intricate relationships characteristic of health-related data [26]. Through leveraging these advanced methodologies, researchers can move beyond conventional statistical analyses to embrace models that account for non-linear interactions and the multifaceted nature of diseases like hypertension and diabetes. The emergence and evaluation of models like Cost-Sensitive Boosting in the health sciences domain highlight an evolving understanding and approach to predictive analytics. Future study should investigate the comparative effectiveness of Cost-Sensitive Boosting against other machine learning techniques to optimize outcomes in healthcare analytics.

The study has several limitations. First, the sample size of 1,240 patients, while substantial, may not adequately represent the broader population of individuals with hypertension and diabetes, hence constraining the generalizability of the results. Second, the cross-sectional design prevents the establishing of causation or the directionality of the observed connections by capturing a moment of the interactions between variables at a particular point in time; hence, longitudinal studies are recommended to verify the temporal sequence of the identified factors and their impact on HRQoL. Third, the reliance on self-reported data for key variables such as HRQoL, perceived stress, and depressive symptoms introduces the potential for response bias, as participants' subjective assessments may be influenced by recall bias or social desirability bias, compromising the accuracy of the reported information; incorporating objective measures and triangulating self-reported data with clinical assessments would improve the robustness of future studies.

5. Conclusions

This study identified key factors significantly affecting the HRQoL in hypertensive diabetic patients, emphasizing the importance of weight management, mental health support, and stress reduction. The Cost-Sensitive Boosting model demonstrated superior predictive performance, providing a robust tool for personalized patient care. These findings offer valuable insights for developing targeted interventions and inform health policy formulation to enhance patient outcomes.

Acknowledgements

This research Supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-RS-2023-00237287, NRF-2021S1A5A8062526) and local government-university cooperation-based regional innovation projects (2021RIS-003).

REFERENCES

[1] E.-M. Sultana, F. Anzin, and T. Asiq, "The Correlation between Diabetes, Hypertension & Kidney Disease with Others Diabetic Induced Disease and Factors among Diabetic Patients in Bangladesh," Scholarly Journal of Agricultural Science, vol. 8, no. 8, pp. 1859-1864, 2020. https://doi.org/10.36347/sjams.2020.v08i08.013.

[2] Z. Zhang, L. Zhao, Y. Lu, Y. Xiao, and X. Zhou, "Insulin resistance assessed by estimated glucose disposal rate and risk of incident cardiovascular diseases among individuals without diabetes: findings from a nationwide, population-based, prospective cohort study," Cardiovascular Diabetology, vol. 23, p. 194, 2024. https://doi.org/10.1186/s12933-024-02256-5.

[3] F. Josse, P. Pradana, and S. Tuba, "Novel Diabetic Nephropathy-Based Hypertension Treatment for Type-2 Diabetes Mellitus and CKD Patients: A Mini Review," Journal of Medical Health Studies, vol. 4, no. 4, pp. 197-201, 2023. https://doi.org/10.32996/jmhs.2023.4.4.21.

[4] R. S. Krishan, A. K, and U. Vijay, "Analysis of Dynamic Association Between Diabetes Mellitus and Cardiovascular Disease," Cardiometry, no. 24, pp. 297-305, 2022.

[5] L. Guan, L. Li, Y. Zou, J. Zhong, and L. Qiu, "Association between FIB-4, all-cause mortality, cardiovascular mortality, and cardiovascular disease risk among diabetic individuals: NHANES 1999–2008," Frontiers in Cardiovascular Medicine, vol. 10, 2023. https://doi.org/10.3389/fcvm.2023.1172178.

[6] S. Sinha and M. Haque, "Insulin resistance is cheerfully hitched with hypertension," Life, vol. 12, no. 4, p. 564, 2022. https://doi.org/10.3390/life12040564.

[7] V. Manyenze, L. A. Isiaho, and J. Okoth, "Socio-demographic characteristics associated with health-related quality of life among patients with type 2 diabetes mellitus at Tenwek hospital," Journal of Health and Medical Nursing, vol. 7, no. 3, pp. 41-50, 2021. https://doi.org/10.47604/jhmn.1412.

[8] F. A. Eyowas, M. Schneider, S. A. Balcha, S. Pati, and F. A. Getahun, "Multimorbidity and health-related quality of life among patients attending chronic outpatient medical care in Bahir Dar, Northwest Ethiopia: The application of partial proportional odds model," PLOS Global Public Health, vol. 2, no. 11, p. e0001176, 2022. https://doi.org/10.1371/journal.pgph.0001176.

[9] K. Thomaseth, A. Pavan, R. Berria, L. Glass, R. DeFronzo, and A. Gastaldelli, "Model-based assessment of insulin sensitivity of glucose disposal and endogenous glucose production from double-tracer oral glucose tolerance test," Computer Methods and Programs in Biomedicine, vol. 89, no. 2, pp. 132-140, 2008. https://doi.org/10.1016/j.cmpb.2007.06.003.

[10] M. F. Guo, Y. J. Dai, J. R. Gao, and P. J. Chen, "Uncovering the mechanism of Astragalus membranaceus in the treatment of diabetic nephropathy based on network pharmacology," Journal of Diabetes Research, vol. 2020, no. 1, p. 5947304, 2020. https://doi.org/10.1155/2020/5947304.

[11] M. B. A. McDermott, S. Wang, N. Marinsek, R. Ranganath, L. Foschini, and M. Ghassemi, "Reproducibility in machine learning for health research: Still a ways to go," Science Translational Medicine, vol. 13, no. 586, p. eabb1655, 2021. https://doi.org/10.1126/scitranslmed.abb1655.

[12] L. Wang, L. Zhu, J. Jiang, L. Wang, and W. Ni, "Decision tree analysis for evaluating disease activity in patients with rheumatoid arthritis," Journal of International Medical Research, vol. 49, no. 10, 2021. https://doi.org/10.1177/03000605211053232.

[13] T. Nakano, Y. Kajiyama, G. S. Revankar, R. Hashimoto, Y. Watanabe, H. Kishima, ... and N. Hattori, "Neural networks associated with quality of life in patients with Parkinson's disease," Parkinsonism & Related Disorders, vol. 89, pp. 6-12, 2021. https://doi.org/10.1016/j.parkreldis.2021.06.007.

[14] C. Twumasi and J. Twumasi, "Machine learning algorithms for forecasting and backcasting blood demand data with missing values and outliers: A study of Tema General Hospital of Ghana," International Journal of Forecasting, vol. 38, no. 3, pp. 1258-1277, 2022. https://doi.org/10.1016/j.ijforecast.2021.10.008.

[15] S. Islam and S. H. Amin, "Prediction of probable backorder scenarios in the supply chain using Distributed Random Forest and Gradient Boosting Machine learning techniques," Journal of Big Data, vol. 7, no. 1, p. 65, 2020. https://doi.org/10.1186/s40537-020-00345-2.

[16] A. Mosavi, F. Sajedi Hosseini, B. Choubin, M. Goodarzi, A. A. Dineva, and E. Rafiei Sardooi, "Ensemble boosting and bagging based machine learning models for groundwater potential prediction," Water Resources Management, vol. 35, pp. 23-37, 2021. https://doi.org/10.1007/s11269-020-02704-3.

[17] J. Yang, R. El-Bouri, O. O'Donoghue, A. S. Lachapelle, A. A. Soltan, D. W. Eyre, ... and D. A. Clifton, "Deep reinforcement learning for multi-class imbalanced training: applications in healthcare," Machine Learning, vol. 113, pp. 2655-2674, 2024. https://doi.org/10.1007/s10994-023-06481-z.

[18] H. Byeon, "Developing a predictive model for depressive disorders using stacking ensemble and naive Bayesian nomogram: using samples representing South Korea," Frontiers in Psychiatry, vol. 12, p. 773290, 2022. https://doi.org/10.3389/fpsyt.2021.773290.

[19] L. Tong, Z. Liu, Z. Jiang, F. Zhou, L. Chen, J. Lyu, ... and H. Zhou, "Cost-sensitive Boosting Pruning Trees for depression detection on Twitter," IEEE Transactions on Affective Computing, vol. 14, no. 3, pp. 1898-1911, 2022. https://doi.org/10.1109/TAFFC.2022.3145634.

[20] V. Boima, A. K. Yeboah, I. A. Kretchy, A. Koduah, K. Agyabeng, and E. Yorke, "Health-related quality of life and its demographic, clinical and psychosocial determinants among male patients with hypertension in a Ghanaian tertiary hospital," Ghana Medical Journal, vol. 56, no. 1, pp. 5-14, 2022. https://doi.org/10.4314/gmj.v56i1.2.

[21] S. Pati, S. Swain, J. A. Knottnerus, J. F. Metsemakers, and M. van den Akker, "Health related quality of life in multimorbidity: a primary-care based study from Odisha, India," Health and Quality of Life Outcomes, vol. 17, p. 116, 2019. https://doi.org/10.1186/s12955-019-1180-3.

[22] H. F. Alautry, M. Namdari, M. Khoshnevisan, and H. Ghasemi, "Association between dental clinical measures and oral health-related quality of life among Iraqi schoolchildren: A cross-sectional study," PLOS ONE, vol. 19, no. 4, p. e0293024, 2024. https://doi.org/10.1371/journal.pone.0293024.

[23] N. Aslam, M. Shoaib, R. Bushra, S. Asif, and Y. Shafique, "Evaluating the socio-demographic, economic and clinical (SDEC) factors on health related quality of life (HRQoL) of hypertensive patients using EQ-5D-5L scoring algorithm," PLOS ONE, vol. 17, no. 6, p. e0270587, 2022. https://doi.org/10.1371/journal.pone.0270587.

[24] M. Ahmadi, R. Nopour, and S. Nasiri, "Developing a prediction model for successful aging among the elderly using machine learning algorithms," Digital Health, vol. 9, 2023. https://doi.org/10.1177/20552076231178425.

[25] N. H. Chowdhury, M. Reaz, F. Haque, S. Ahmad, S. Ali, A. A. A. Bakar, and M. Bhuiyan, "Performance Analysis of Conventional Machine Learning Algorithms for Identification of Chronic Kidney Disease in Type 1 Diabetes Mellitus Patients," Diagnostics, vol. 11, no. 12, p. 2267, 2021. https://doi.org/10.3390/diagnostics11122267.

[26] Z. Asghari Varzaneh, M. Shanbehzadeh, and H. Kazemi-Arpanahi, "Prediction of successful aging using ensemble machine learning algorithms," BMC Medical Informatics and Decision Making, vol. 22, no. 1, p. 258, 2022. https://doi.org/10.1186/s12911-022-02001-6.