

# Deep learning-based physiological risk stratification in night-shift hospital workers

Received: 21 November 2025

Accepted: 9 March 2026

Published online: 16 March 2026

Cite this article as: Lee I., Hong S., Lee J. *et al.* Deep learning-based physiological risk stratification in night-shift hospital workers. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-43982-y>

InHo Lee, SangHee Hong, JuneHee Lee, HwaYoung Lee, SoonChan Kwon, YoungSun Min, EunChul Jang & JeongBeom Lee

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

**Title Page****Deep Learning-Based Physiological Risk Stratification in Night-Shift Hospital Workers**

InHo Lee<sup>1,2</sup>, SangHee Hong<sup>3</sup>, JuneHee Lee<sup>4</sup>, HwaYoung Lee<sup>5</sup>, SoonChan Kwon<sup>1</sup>, YoungSun Min<sup>1</sup>, EunChul Jang<sup>1\*</sup>, JeongBeom Lee<sup>2\*</sup>

<sup>1</sup>Department of Occupational and Environmental Medicine, Soonchunhyang University Cheonan Hospital, Cheonan, Republic of Korea

<sup>2</sup>Department of Physiology, College of Medicine, Soonchunhyang University, Cheonan 31151, Republic of Korea

<sup>3</sup>Department of Medical Sciences, Graduate School, Soonchunhyang University, Asan, Republic of Korea

<sup>4</sup>Department of Occupational and Environmental Medicine, Soonchunhyang University Hospital, Seoul, Republic of Korea

<sup>5</sup>Department of Psychiatry, Soonchunhyang University Cheonan Hospital, Cheonan 31151, Republic of Korea.

**\*Corresponding authors:**

EunChul Jang, Department of Occupational and Environmental Medicine, Soonchunhyang University Cheonan Hospital, Cheonan, Republic of Korea

Email: oemdr10@gmail.com

JeongBeom Lee, Department of Physiology, College of Medicine, Soonchunhyang University, Cheonan 31151, Republic of Korea

Email: leejb@sch.ac.kr

Lead Contact: Dr. JeongBeom Lee(leejb@sch.ac.kr)

ARTICLE IN PRESS

**Abstract**

This cross-sectional study of 2,250 hospital workers attempted to identify potential physiological risk structures for night workers by utilizing the nonlinear manifold learning framework. Unlike conventional mean-centered linear analysis, this study applied the Potential of Heat-diffusion for Affinity-based Transition Embedding (PHATE) for topological visualization and Neural Additive Models (NAM) to predict interpretable risk. The 'cholesterol homeostasis profile' was defined as the main latent indicator.

Analysis showed that the systemic metabolic load (PHATE-1) was higher after age correction, despite the night-shift group being younger than the non-night shift group ( $p < 0.001$ ). This axis showed a positive correlation with body mass index ( $r = 0.71$ ) and triglyceride ( $r = 0.56$ ). The NAM model captured the risk slope with an average AUC of 0.864 (range 0.832-0.901) and confirmed a nonlinear risk transition when it exceeded the triglyceride 2 mmol/L threshold. The predicted risk of night workers in the highest risk cluster (cluster 1) was 0.71, and the metabolic load and cholesterol homeostasis profile risk were increased.

Ultimately, the PHATE-NAM framework moves beyond traditional group averages. It provides a purely data-driven way to pinpoint individual physiological vulnerabilities, giving hospitals the exact evidence needed to design safer, more sustainable shift-work environments.

Keywords: Night-shift, PHATE-NAM analysis, accumulated night exposure, circadian misalignment

## INTRODUCTION

Continuous care through 24-hour night shift work, maintaining a care system, requires a harsh biological price for hospital workers<sup>1,2</sup>. Night shift disrupts the natural connection between a person's biological clock causing sleep deprivation<sup>3</sup>. Night shift work is known to be associated with adverse metabolic and cardiovascular health effects<sup>4,5</sup>. The irregular blood pressure and metabolic changes observed in night shift workers indicate that the body is undergoing long-term physical load<sup>3,5</sup>. This evidence is also shown in prominent international research reports. Major international organizations, including the World Health Organization (WHO)<sup>6</sup>, the International Labor Organization (ILO)<sup>7</sup>, and the National Institute for Occupational Safety and Health (NIOSH)<sup>8</sup> have characterized night shift work as an occupational hazard. They point out the need for precise and individualized tracking systems to identify and monitor the physiological deterioration caused by night shifts.

Up to now, previous studies have focused on identifying differences in group averages between night and non-night workers<sup>9</sup>. Although these analytical methodologies have contributed to understanding the overall risk level of night work, it has been difficult to fully explain how consecutive working days or complex job factors are associated with the physiological variability of individual workers<sup>10-12</sup>. In addition, previous studies focused only on numerical changes in specific indicators due to night work, and there were limits to analyzing the heterogeneity of high-dimensional physiological responses in which biological indicators are intricately intertwined<sup>13</sup>. Night work affects metabolism, hemodynamics, and periodic biomarkers, and their complex and nonlinear interactions adversely affect workers' health. Considering the

characteristics of a cross-sectional study<sup>14</sup>, instead of analyzing the causal outcomes of certain diseases driven by night work<sup>2,15,16</sup>, this study aims to explore what correlation patterns and interactions the biochemical indicators of workers form within the health risk landscape associated with potential night work. This perspective is essential for capturing individual vulnerabilities that have been eclipsed by group averages and for identifying the diversity of physiological responses within the same working environment<sup>5</sup>.

To systematically analyze the nonlinearity and complexity of the patterns of indicators and physiological data, we utilized an analysis framework (Figure 1) that combines the Potential of Heat-diffusion for Affinity-based Trajectory Embedding (PHATE) technique and Neural Additive Models (NAM)<sup>17,18</sup>. PHATE shows workers' 'physiological maps' to compress diversified physiological data into a visualizable environment. This enables the visualization of natural interactions between physiological indicators and transfer patterns for the risk of working at night<sup>17</sup>. NAM is an 'interpretable glass box' model, and unlike 'black box' AI, which has an opaque decision-making process, it can identify risk probabilities nonlinearly. In fact, it is essential to secure the reliability and validity of the AI model in the decision-making stage for application in a clinical setting, so the nonlinear predictive performance of deep learning and the interpretability of the linear model can be secured at the same time<sup>18,19</sup>. This step-by-step analysis is significant in that it can help managers make data-based decisions in judging the risk of working at night in the actual clinical field, beyond simply enhancing predictive performance<sup>20-22</sup>.

The major latent axis derived from PHATE analysis, the Cholesterol Homeostasis Profile, is based on the evidence that the stable regulatory ability

of lipid metabolic indicators such as triglyceride, LDL, and HDL cholesterol correlates with night work<sup>23,24</sup>. This reflects the clinical significance of triglyceride to HDL ratio, a well-known indicator of metabolic homeostasis and insulin resistance in our preceding study<sup>24</sup>. Therefore, instead of looking at lipids as isolated levels, we identified them as indicators of the body's inherent ability to buffer PHATE-derived metabolic stress.

This study aims to verify that night workers form a unique potential physiological risk structure that distinguishes them from non-night workers through nonlinear representation learning. We hypothesized that night shift exposure is associated with the formation of certain high-risk clusters within the latent space beyond mere deterioration of indicator values, and that metabolic risk including cholesterol homeostasis profiles is a major factor characterizing risk structures. The results of this study provide an academic basis for preemptively identifying high-risk groups within the organization and establishing a sustainable working environment design and precise preventive health management system that takes into account the physiological limitations of night shift workers within the hospital's 24-hour operating system.

*Fig. 1. Conceptual overview of the PHATE-NAM analytical workflow*

## **METHODS**

### **Study Participant Dataset**

A total of 2,250 hospital employees were included in this study, comprising 1,151 night-shift workers and 1,099 non-night-shift workers who participated in the Soonchunhyang University Hospital Occupational Health Program (2025). Data were obtained from the Health Screening Program and integrated with the institution's Occupational Health Dataset, comprising real-world electronic medical records (EMR). Baseline demographic and occupational characteristics of night-shift and non-night-shift workers are summarized in Table 1.

To minimize selection bias and ensure the ecological validity of the study, we adopted a total census approach, enrolling the entire cohort of eligible employees who participated in the Soonchunhyang University Occupational Health Program (2025). Although the sample size was determined by this comprehensive enrollment strategy, a post-hoc power analysis confirmed the statistical robustness of the design. The final sample size ( $n = 2,250$ ) achieved a statistical power of  $>99.7\%$  ( $1-\beta > 0.99$ ) to detect even small effect sizes (Cohen's  $d = 0.2$ ) at a significance level of  $\alpha = 0.05$ , ensuring sufficient sensitivity to identify physiological differences between groups.

The inclusion criteria were as follows: (1) full-time employment for more than one year, (2) completion of fasting biochemical examinations, and (3) availability of verified shift-work records.

The exclusion criteria were as follows: (1) missing key biochemical indicators and (2) current active treatment for chronic metabolic or endocrine diseases.

### **Ethical Approval and Data Collection**

All participants provided written informed consent, and the study protocol was reviewed and approved by the Soonchunhyang University Cheonan Hospital

Institutional Review Board (IRB No. SCHCA-2025-10-010). All procedures involving human participants were conducted in accordance with the relevant guidelines and regulations, including institutional policies and the principles of the Declaration of Helsinki. The dataset included routine biochemical and hemodynamic parameters such as triglyceride (TG), total cholesterol (TC), fasting glucose, aspartate aminotransferase (AST), alanine aminotransferase (ALT), systolic blood pressure (SBP), and diastolic blood pressure (DBP). All laboratory tests were performed under fasting conditions, following standard hospital protocols.

### **Data preprocessing and statistical analysis schemes**

Cleaning and standardization were performed for data analysis. Continuous variables were normalized using Z-score transformation, and outliers were processed according to the interquartile range (IQR) standard ( $1.5 \times \text{IQR}$ ). As a result of checking the entire data, missing values accounted for less than 1.5%, for which a simple distribution-based imputation method was applied. In order to prevent overfitting of the model and increase the generalization performance, five-fold cross-validation (25 repetitions) was applied, and stability was verified based on the average and standard deviation of the calculated performance indicators.

For analysis, clusters were first identified based on the latent distribution inherent in the data using VaDE, and then PHATE was applied to project the data for those clusters into the latent space. Using this integrated technique, we were able to visualize the high-risk transition process.

All data preprocessing and integration processes were performed in the Python 3.11 environment using general-purpose analysis packages, including pandas, NumPy, and scikit-learn. The original dataset is restricted from external disclosure according to the agency's information protection policy, but it is accessible through the corresponding author (Lee Jeong-beom, leejb@sch.ac.kr) for reasonable research purposes.

## **METHOD DETAILS**

### **Nonlinear manifold embedding (PHATE)**

All continuous physiological variables underwent Z-score normalization prior to analysis to ensure scale consistency. To explore the potential structure formed by several physiological markers, we applied the PHATE algorithm (Python package `phate`, v3.0) to a standardized data matrix. PHATE constructs adaptive diffusion graphs based on local data density, allowing us to visualize continuous physiological gradients across the data instead of discrete classification. Similarities among participants were calculated through adaptive Gaussian kernels and converted to transition probabilities, while potential distances were calculated by applying the diffusion time ( $t$ ) parameter (see Supplementary Methods: Mathematical Formulation of PHATE). To ensure the structural performance of the embeddings, a sensitivity analysis to sweep was performed by setting diffusion times ( $t$ ) to 10, 30, 50, and attenuation to 20, 40, and 80 (Supplementary Table S1). To clarify the biological importance of the derived latent axes, quantitative correlation mapping was performed, and the final model parameters were selected based on the stability of the latent structure (see Supplementary Methods: Mathematical Formulation of PHATE).

Furthermore, a potential gradient distribution analysis was performed to quantify the slope of risk transfer in PHATE, and local gradient refers to the speed of physiological change and the sensitivity of specific conditions under night work and homeostasis. By associating this with occupational characteristics such as cumulative night work period and intensity of labor, we analyzed how a specific work type promotes the deterioration of health indicators.

### **Risk probability modeling (LightGBM)**

In order to quantitatively confirm how physiological data interact with job factors in the latent space, risk probability was modeled using LightGBM (see Supplementary Methods: LightGBM for the model formulation). A high-performance risk estimation was performed to build a continuous risk landscape and to define the risk density across the manifold. In addition, in order to prevent errors in circular reasoning, a practical interpretation of the physiological mechanism was confirmed through transparent Neural Additive Models (NAM). To prevent overfitting of the model and to ensure generalization performance, strict iterative stratified 5-fold cross-validation (a total of 25 runs) was applied instead of the traditional single-out. The performance evaluation of the model used the receiver manipulation characteristic curve area (AUC) as the main indicator, and the predictive stability of the model was secured through cross-validation. Through this, the calculated average predictive probability was projected onto the PHATE embedding to build a continuous risk landscape, which was used to visualize the metabolic risk changing along the PHATE-1 axis.

## **Deep interpretable risk learning (NAM)**

To transparently interpret nonlinear relationships between variables beyond the limits of tree-based models, a PyTorch-based (v2.3.1) NAM was implemented. Considering that the night shift group, which is a significant demographic difference, is younger than the day work group, we explicitly included age and sex as input covariates. In addition, we utilized NAM to mathematically separate these effects. Since NAM calculates the contribution by each feature through an independent function, metabolic risk associated with biochemical indicators can be assessed together with metabolic indicators, work intensity index, and PHATE coordinates regardless of aging effect. Each physiological and job variable is modeled through an independent subnetwork and then additionally combined to produce a final output (see Supplementary Methods: Structure of Nonlinear Additive Model (NAM)). As input variables, metabolic indicators, work intensity index, and PHATE coordinates were included. For reproducibility, all deep learning models were trained with a fixed random seed (seed=42).

For our high-risk cutoff, we isolated the top 30% of the probability distribution. We chose this specific metric as a slightly tighter variation of the classic top-tertile split (roughly 33%) widely used across epidemiological studies<sup>25</sup>. Furthermore, we confirmed through sensitivity checks that the core topological features and cluster compositions remained stable around nearby thresholds, demonstrating that our results are robust to minor variations in the cutoff definition. Zeroing in on this upper tail is simply a more efficient way to target clinical interventions<sup>26</sup>. We also had to ensure model stability. We ran repeated cross-validation for this, replicating our LightGBM protocol. This step

guarantees that the variable contributions we calculate are universal, rather than just noise from a single biased sample. To justify the nonlinear method, we put NAM up against standard Logistic Regression and Random Forest baselines. NAM enables visual tracking of nonlinear effects while maintaining predictive performance. When comparing model performance, the results support that NAM is optimized in that it provides a high level of accuracy and interpretability required in the clinical field (Supplementary Table S2). That is, NAM guarantees a level of accuracy similar to that of existing black box models and provides intuitive identification of nonlinear patterns in which each variable contributes to the risk probability.

### **Latent cluster identification (VaDE)**

After topological embedding, we applied a variational deep embedding (VaDE) framework combining a variational autoencoder (VAE) and a Gaussian mixture model (GMM) (see Supplementary Methods: Loss Function for Variational Deep Embedding (VaDE)). Using PHATE-induced latent coordinates input from the encoder, we compressed the data into a low-dimensional space and estimated stochastic cluster membership via GMM. Specifically, to verify the statistical stability of the identified cluster structure, iterative evaluations were performed using different initialization seeds, and the reproducibility of subtype induction was verified using the Adjusted Rand Index (ARI). For the final reported results, the model was initialized with a fixed random seed (seed=42) to ensure strict reproducibility. Detailed hyperparameters and architectural configurations for all implemented models (PHATE, LightGBM, NAM, and VaDE) are summarized in Supplementary Table S3. All deep learning analyses were performed using

PyTorch (v2.3.1), while visualization and clustering were performed using Scikit-learn (v1.5.0). The PHATE-NAM-VaDE code used in this study can be provided by the corresponding author upon reasonable request, and all key algorithms were implemented using the aforementioned open-source packages (PHATE, LightGBM, and PyTorch). Researchers wishing to access the data must go through a formal review process of their research plan to comply with privacy regulations before requesting an anonymized dataset.

## RESULTS

### Global PHATE topology of physiological variation in hospital staff

*Fig. 2. Physiological distribution of hospital workers mapped onto the PHATE manifold*

The PHATE-1 axis, representing 'Metabolic Load', showed positive correlations with body mass index (BMI,  $r = 0.71$ ), systolic blood pressure (SBP,  $r = 0.67$ ), and triglycerides (TG,  $r = 0.56$ ). The PHATE-2 axis, defined as the cholesterol homeostasis profile, showed significant correlations with LDL cholesterol ( $r = 0.67$ ) and HDL cholesterol ( $r = 0.51$ ).

The Gaussian kernel density estimation (KDE) contours in Figure 2 show a clear contrast. The night shift group had a significantly higher distribution on the PHATE-1 axis compared to the non-night shift group ( $p < 0.001$ ). This reveals a co-elevation pattern where metabolic burden markers, such as triglycerides, blood sugar, and liver enzymes, rise together in the night shift cohort. However, the wide distribution of PHATE-2 levels confirmed that significant physiological

heterogeneity, driven by individual cholesterol homeostasis control, exists even within the night shift group itself.

### **Spatial Distribution of Predicted Metabolic Risk**

*Fig. 3. Projection of predicted metabolic risk across the global PHATE landscape*

As a result of projecting the LightGBM model into the PHATE space (Fig. 3), it was confirmed that a high-risk region was formed in the area with a high PHATE-1 value. The average AUC calculated through repeated layered 5-layer cross-validation was 0.886 (SD = 0.018, range 0.852–0.915), demonstrating that the model stably captures the risk gradient in the latent space. Furthermore, we supported the predictability of the model by analyzing calibration curves that showed a high level of alignment between the predictive probability and the observed risk frequency. As we approach the 45-degree mark, we see that the model provides a reliable risk estimate. In particular, the high-risk group with a predictive risk of 0.7 or more was spatially concentrated and distributed in the right area of the manifold where the metabolic load was concentrated.

### **Local Occupational Topology of Night-shift Workers**

*Fig. 4. Occupational gradient patterns among night workers in latent space*

As a result of the latent gradient mapping analysis (Fig. 4) targeting only night workers, it was found that the high-density areas of (a) consecutive night workdays, (b) cumulative number of working hours, and (c) weekly work hours significantly overlapped with the high-risk areas of PHATE-1. These patterns suggest that accumulated night exposure and long working hours are closely

related to physiological stress weighting and show that the PHATE-1 axis effectively captures the changes in biomarkers overall in the night work environment.

### **NAM-based Interpretable Modeling: Prediction and Explanation**

*Fig. 5. Interpretable risk components derived from the NAM model*

Variable importance analysis of the NAM model (Figure 5a) identified triglycerides (TG), systolic and diastolic blood pressure (SBP/DBP), total cholesterol (TC), and working hours per week as major predictors. Compared with the baseline models, the AUC for logistic regression was lower than that of the NAM (Supplementary Table S2). Random Forest showed slightly higher accuracy (AUC = 0.914) than the NAM, but it has the fundamental problem of lacking the interpretability essential for setting clinical thresholds. The NAM is significant in that it achieves a stable mean AUC of 0.864 (SD = 0.018, range 0.832-0.901) while simultaneously ensuring this interpretability.

Partial dependence analysis (Figure 5b-d) revealed a nonlinear pattern where the predicted risk increased rapidly when TG levels exceeded approximately 2 mmol/L (Figure 5b) which is 'borderline high' category of NCEP ATP III (approx. 177 mg/dL). For weekly working hours, metabolic risk increased as hours deviated from the standardized mean ( $Z=0$ ) (Figure 5c). Furthermore, PHATE-2 levels showed a significant negative linear relationship with predicted risk reflecting the functional efficiency of reverse cholesterol transport (RCT) mediated by the antagonism between LDL and HDL (Figure 5d).

### **Model Discrimination and Calibration**

*Fig. 6. Model discrimination and calibration performance of the NAM framework*

Average AUC of 0.864 calculated through rigorous iterative cross-validation was recorded (ROC curve, Fig. 6a). The calibration curve (Fig. 6b) shows that the predicted probability closely matches the true event rate, indicating that the model is reliable and not overconfident. The calibration curve (Fig. 6b) shows a high degree of agreement between the risk probability predicted by the model and the actual observed risk distribution.

**High-risk Topography within the PHATE Manifold**

*Fig. 7. VaDE-based clustering with risk stratification on the PHATE manifold*

Three major physiological subtypes were identified because of the combination of VaDE clustering and risk probabilities calculated from NAM (Fig. 7). As a result of repeated evaluation with different initialization seeds, the average adjusted Rand Index (ARI) of 0.858 (range 0.689–0.985) was recorded, confirming the high reproducibility and statistical stability of the derived cluster structure.

Cluster 1, which was characterized by a negative PHATE-1 gradient, showed the highest average predicted risk (0.71) among night workers, and concomitant increases in triglyceride, blood pressure, blood sugar and liver enzyme levels were clearly observed. On the other hand, Cluster 2, which was characterized by a high PHATE-2 value, was mainly composed of low-risk groups and showed a pattern of maintaining a relatively stable lipid metabolism despite exposure to night work.

*Fig. 8. Density contours of high-risk regions within the PHATE space (night-shift subset)*

Nonlinear interaction structure was identified where risk is most pronounced, specifically manifesting as a high-density convergence in the region with negative PHATE-1 and low PHATE-2 levels (Fig. 8). The most concentrated risk in this context implies a high predicted risk probability. The two axes of PHATE represent that high-risk workers tend to be distributed at the critical point where physiological equilibrium collapses. By analyzing and visualizing the vulnerability of night workers as quantitative patterns within a multidimensional space, this approach explains the physiological mechanism of risk accumulation that is difficult to identify with only a single variable.

## **Discussion**

This study attempted to confirm distinct physiological adaptations and the pattern of changes in structural patterns under interaction between physiological variables beyond the numerical changes of individual indicators for night work. In Table 1, the night shift group was about 10 years younger than the non-night shift group (30.9 years vs. 40.7 years old), and the superficial metabolic index was also better. This is a typical 'healthy worker effect', and the physiological robustness of young workers showed 'statistical masking' to cover the cumulative metabolic load according to night shift and working conditions<sup>27,28</sup>. Due to these demographic effects, conventional linear analysis has limitations in detecting the health impact assessment of night shift. In this study, potential data structures were identified through PHATE analysis rather

than simple group comparisons to identify structural risks hidden behind the younger age group. Despite the young age of night shift workers, they were distributed along the PHATE-1 axis, which indicates metabolic load, and the exposure to night shift reflected the accumulation of 'allostatic load' due to incongruity in biological rhythms, indicating that individuals were located in specific high-risk areas in the potential physiological space<sup>23,24</sup>. The night workers in particular exhibited a wide distribution along the PHATE-2 axis, which, even in the same working environment, could significantly alter the physiological response depending on an individual's cholesterol homeostasis<sup>24</sup>. Each physiological indicator, blood sugar, blood pressure, and lipid levels, shows the process of systematic integration, indicating topologically that it is a loss of metabolic resilience due to chronic shift work. The spatial slope for the high-risk group identified in LightCBM<sup>29</sup> indicates that the influence of night shift exposure to key physiological domains is statistically associated<sup>28-30</sup>. This encompasses the multivariate risk correlation structures that univariate analysis often misses<sup>31</sup>. This suggests that beyond traditional single-indicator assessments, a precise industrial health approach that considers topographical flows is needed for various effects, including the working environment of night workers within the latent space.

Regarding the type of work, consecutive night shift days, cumulative working hours, and weekly working hours showed a distinct positive correlation with an increase in metabolic load (PHATE-1)<sup>1,32,33</sup>. This suggests that the physiological position within the potential risk area is not determined simply by the dichotomous exposure to night work, but is complexly related to the intensity of the job and the type of work. Previous studies also showed a significant

association between irregular shift work and cumulative exposure and metabolic syndrome indicators, which is in line with the metabolic fluctuation pattern shown in the latent space of this study<sup>2,11,34-36</sup>. The visualization of the latent space goes beyond simply checking the distribution of physiological markers in individuals, and it can be said that it is a tool to show the effect of the system of night work on physiological changes in the human body.

As a result of VaDE analysis, cluster 1, which is a high-risk group, was concentrated in the region where PHATE-1 (metabolic load) was negative and PHATE-2 (Cholesterol Homeostasis Profile) was low (Fig. 7). To ensure the reproducibility and stability of these risk clusters, a VaDE model was constructed using certain latent dimensions and Gaussian mixed model dictionaries. The optimal number of clusters  $K$  was determined based on the lower bound of evidence and the silhouette score. This suggests that risk gradients may appear differently depending on individual lipid regulation capabilities even under the same level of metabolic exposure<sup>36,37</sup>. Previous studies have reported that when the regulatory ability of the biological rhythm is deteriorated<sup>38</sup>, it may become more vulnerable to metabolic disease and inflammatory reactions<sup>4,39</sup>, and the instability of the lipid profile observed in this study can be interpreted as a biochemical result reflecting the functional deterioration of these regulatory systems<sup>20,40-42</sup>.

If the global map visualizes the spatial slope of the risk, the NAM is significant in that it identifies the contribution function of each variable and thus increases its applicability in industrial health sites beyond the limitations of the conventional black box model<sup>18,22</sup>. It should be noted that the threshold setting value of TG, 1.7-2.2 mmol/L (150-199 mg/dL), is consistent with the NCEP ATP

III guideline, which classifies it as 'borderline high'<sup>2,43</sup>. This suggests that the NAM model successfully captured the acceleration of the hidden physiological risk before reaching the high-risk criteria due to night work.<sup>48</sup> In addition, the U-shaped curve observed in working hours per week showed the lowest risk in the 40-50-hour interval, indicating that out-of-range work patterns are associated with a decrease in physiological stability<sup>1,14,36,37,45,46</sup>.

Beyond the dichotomous analysis of the presence or absence of night work, PHATE shows different physiological responses and structural distributions implemented in the latent space. Although we did not directly measure actigraphy data such as sleep quality or diet in this study, the resilience through cholesterol homeostasis projected on the PHATE-2 axis and the distribution of risk due to night work can be interpreted as the result of the convergence of various factors into physiological indicators<sup>47,48</sup>. Lipid regulatory abnormality acts as a strong biological proxy that integrates the downstream physiological effects of lifestyle factors that have not been measured, serving as a metabolic footprint of biological rhythm disturbance. This is consistent with existing literature suggesting that a decrease in biorhythm stability makes individuals more vulnerable to metabolic disease and inflammatory profile formation<sup>4,39,40</sup>. Therefore, different patterns may appear in this study even for two workers with identical clinical figures for the same metabolic burden. If workers have a strong cholesterol homeostasis profile (high PHATE-2), they may show resilience to metabolic stress. On the other hand, workers lacking this profile who are placed in high-risk belts require preemptive industrial health management even if they work the same night shifts. This suggests that latent coordinates enable preemptive and individualized monitoring of night shift

work and metabolic conditions. For night workers, health care requires a precise approach that takes into account inherent physiological capacities, including cholesterol homeostasis profiles, based on an individual's latent physiological coordinates rather than a uniform standard<sup>9,49,50</sup>.

Therefore, the step-by-step analysis pipeline, from PHATE identifying global topological structures to NAMs calculating the nonlinear contributions of individual variables, is designed to procedurally approach the complexity and interdependence of multidimensional physiological data. This is in line with probabilistic health modeling studies that incorporate predictive risk models and cluster analyses in a cross-sectional study environment<sup>51</sup>, and it can capture frequently overlooked potential risk structures. This analytical approach is useful for identifying the individual vulnerabilities of high-risk workers who are often obscured by group averages<sup>52</sup>, consistent with the findings of large-scale studies of night workers<sup>53</sup>. This provides foundational data for objectively monitoring workers' physiological loads in environments where night shifts are institutionalized<sup>21,44,54</sup>. While current guidelines from institutions such as EU-OSHA<sup>55</sup> mainly focus on general shift schedules, the results of this study suggest that individual and detailed approaches tailored to each workplace and worker can be provided to meet global standards.

Although this study provides a data-driven framework for identifying potential physiological risk structures, some limitations should be considered. First, since this is a cross-sectional study of a single university hospital, there are limits to establishing definitive causal relationships. Furthermore, because it is based on a single institution, bias may occur due to the shift work characteristics of a specific hospital or the unique demographic and sociological characteristics of

the region. The spatial distribution between the PHATE-NAM risk zones and night shift work provides evidence of spatial associations and physiological gradients rather than temporal causal effects. Health selection effects or issues of temporal precedence cannot be completely excluded. Future longitudinal studies should track individual pattern changes over time on the PHATE manifold, and follow-up verification should be conducted to determine whether the identified high-risk belts and potential slopes accurately predict the actual incidence of metabolic and cardiovascular diseases.

Second, there is a limitation that individual circadian behavior data, such as sleep quality or diet, have not been directly measured. However, the cholesterol Homeostasis Profile defined in this study functions as a physiological endpoint for which these unmeasured various biological background factors finally converge. In particular, the Mean ARI of 0.858 which demonstrates the stability of clustering structures, supports that the identified subtypes reflect a robust physiological structure rather than an accidental distribution of the data.

Third, although there is a possibility of health worker bias (which suggests our findings likely represent a conservative estimate of the true risk), the nonlinear manifold analysis in this study has differentiated value in that it presents specific and immediately viable clinical criteria, such as the triglyceride 2mmol/L threshold, even within the relatively healthy population. These results can be used as practical decision-making support tools for hospital health managers to proactively identify high-risk groups and establish sustainable work designs.

## **Conclusion**

The analysis system that combines PHATE and NAM clearly demonstrates how the health risks of night workers are organized and spread within the latent space. By identifying the phase structure of the entire data through PHATE and precisely analyzing the contribution of individual variables through NAM, it is possible to identify the number of night shifts, job intensity, and metabolic indicators such as triglycerides and blood sugar that form high-risk areas and connect them. This multidimensional approach deviates from the existing method of relying on the group average and quantitatively demonstrates the degree of risk and why it varies from individual to individual even in the same environment.

The results of this study provide a scientific basis for reducing the physiological load of night workers and designing sustainable work systems. This approach shifts global occupational health management toward data-driven, personalized protection for each worker. It moves beyond uniform corporate and social regulations on night work and provides the scientific evidence needed to update worker health guidelines.

This framework can contribute to establishing practical risk assessment and management strategies in various industrial sites where shift work is essential, including hospitals. Furthermore, precise intervention measures can be tailored by utilizing the latent physiological coordinates of individual workers. Future multicenter cohorts and longitudinal follow-up studies are required to verify the effectiveness of the latent physiological coordinate-based approach proposed in this study.

## **Acknowledgements**

We extend our thanks to the subjects whose participation made this study possible.

### **Author contributions**

Conceptualization: EC.J., JB.L., and IH.L.; Methodology: IH.L., JB.L., SH.H., and EC.J.; Software: IH.L.; Validation: JB.L. and EC.J.; Formal analysis: IH.L. and JB.L.; Investigation: IH.L., SH.H., JH.L., and HY.L.; Resources: SC.K., YS.M., and EC.J.; Data Curation: IH.L. and SH.H.; Visualization: IH.L.; Supervision: JB.L. and EC.J.; Writing - original draft: IH.L.; Writing - review & editing: IH.L., JB.L., and EC.J.

### **Data availability**

The datasets generated and/or analyzed during the current study are not publicly available due to the institution's Electronic Medical Record (EMR) data protection policy regarding participant privacy. However, the raw data supporting the conclusions of this article are available from the corresponding author (JeongBeom Lee, [leejb@sch.ac.kr](mailto:leejb@sch.ac.kr)) upon reasonable request and subject to Institutional Review Board (IRB) approval. To fully comply with the journal's Code Availability policy and support research reproducibility, the official Python implementation of the analytical pipeline (including PHATE, LightGBM, NAM, and VaDE models) has been made publicly available. The code repository can be accessed at: <https://github.com/Crescendo-OEM/DL-Physiological-Risk-Stratification>. To ensure permanent accessibility, the version of the code used in this study is also archived at: <https://doi.org/10.5281/zenodo.18872957>

### **Competing interests**

The authors declare that they have no competing interests.

### **Ethical approval and consent to participate**

All participants provided written informed consent, and the study protocol was reviewed and approved by the Soonchunhyang University Cheonan Hospital Institutional Review Board (IRB No. SCHCA-2025-10-010).

### **Consent for publication**

Not applicable.

### **Funding**

This research was supported by the Soonchunhyang University Research Fund. The funder had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

### **Reference**

- [1] Pan, A., Schernhammer, E. S., Sun, Q. & Hu, F. B. Rotating night shift work and risk of type 2 diabetes: two prospective cohort studies in women. *PLoS Med.* **8**, e1001141 (2011). <https://doi.org/10.1371/journal.pmed.1001141>
- [2] Wang, F. et al. Meta-analysis on night shift work and risk of metabolic syndrome. *Obes. Rev.* **15**, 709–720 (2014). <https://doi.org/10.1111/obr.12194>

- [3] Morris, C. J., Purvis, T. E., Hu, K. & Scheer, F. A. J. L. Circadian misalignment increases cardiovascular disease risk factors in humans. *Proc. Natl. Acad. Sci. USA* **113**, E1402–E1411 (2016). <https://doi.org/10.1073/pnas.1516953113>
- [4] Leproult, R., Holmbäck, U. & Van Cauter, E. Circadian misalignment augments markers of insulin resistance and inflammation, independently of sleep loss. *Diabetes* **63**, 1860–1869 (2014). <https://doi.org/10.2337/db13-1546>
- [5] Kecklund, G. & Axelsson, J. Health consequences of shift work and insufficient sleep. *BMJ* **355**, i5210 (2016). <https://doi.org/10.1136/bmj.i5210>
- [6] International Agency for Research on Cancer. *Night Shift Work* (IARC Monographs on the Identification of Carcinogenic Hazards to Humans, Vol. **124**) (IARC, 2020). <https://publications.iarc.fr/593>
- [7] International Labour Organization. Night Work Convention, 1990 (No. 171) (ILO, 1990). [https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100\\_ILO\\_CODE:C171](https://www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_ILO_CODE:C171)
- [8] Caruso, C. C. et al. *NIOSH Training for Nurses on Shift Work and Long Work Hours* (DHHS (NIOSH) Publication No. 2015-115, National Institute for Occupational Safety and Health, 2015). <https://www.cdc.gov/niosh/docs/2015-115/>
- [9] Vetter, C. et al. Night shift work, genetic risk, and type 2 diabetes in the UK Biobank. *Diabetes Care* **41**, 762–769 (2018). <https://doi.org/10.2337/dc17-1933>
- [10] Puttonen, S., Härmä, M. & Hublin, C. Shift work and cardiovascular disease: pathways from circadian stress to morbidity. *Scand. J. Work Environ. Health* **36**, 96–108 (2010). <https://doi.org/10.5271/sjweh.2894>

- [11] Sooriyaarachchi, P., Jayawardena, R., Pavey, T. & King, N. A. Shift work and the risk for metabolic syndrome among healthcare workers: a systematic review and meta-analysis. *Obes. Rev.* **23**, e13489 (2022). <https://doi.org/10.1111/obr.13489>
- [12] Biglari, H. et al. Relationship between occupational stress and cardiovascular diseases risk factors in drivers. *Int. J. Occup. Med. Environ. Health* **29**, 895–901 (2016). <https://doi.org/10.13075/ijomeh.1896.00125>
- [13] Khaleghi, S. et al. Association between blood pressure and oral temperature rate with sleepiness changes among clinical night workers. *Iran. J. Public Health* **49**, 2232–2234 (2020). <https://doi.org/10.18502/ijph.v49i11.4747>
- [14] Saberinia, A. et al. Investigation of relationship between occupational stress and cardiovascular risk factors among nurses. *Iran. J. Public Health* **49**, 1954–1958 (2020). <https://doi.org/10.18502/ijph.v49i10.4699>
- [15] Foster, R. G. Sleep, circadian rhythms and health. *Interface Focus* **10**, 20190098 (2020). <https://doi.org/10.1098/rsfs.2019.0098>
- [16] Shen, Y. et al. Personalized physician-assisted sleep advice for shift workers: algorithm development and validation study. *JMIR Form. Res.* **9**, e65000 (2025). <https://doi.org/10.2196/65000>
- [17] Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019). <https://doi.org/10.1038/s41587-019-0336-3>
- [18] Agarwal, R. et al. Neural additive models: interpretable machine learning with neural nets. *Adv. Neural Inf. Process. Syst.* **34**, 4699–4711 (2021). <https://doi.org/10.48550/arXiv.2004.13912>

- [19] Jiang, S. et al. Enhancing vitiligo stage diagnosis through a reliable multimodal model with uncertainty calibration. *Applied Intelligence* 55, 1002 (2025). <https://doi.org/10.1007/s10489-025-06839-x>
- [20] Wang, Y., Kung, L., Wang, W. Y. C. & Cegielski, C. G. An integrated big data analytics-enabled transformation model: application to healthcare. *Inf. Manag.* **55**, 64–79 (2018). <https://doi.org/10.1016/j.im.2017.04.001>
- [21] Rajkomar, A., Dean, J. & Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **380**, 1347–1358 (2019). <https://doi.org/10.1056/NEJMra1814259>
- [22] Sadeghi, Z. et al. A review of explainable artificial intelligence in healthcare. *Comput. Electr. Eng.* **118**, 109370 (2024). <https://doi.org/10.1016/j.compeleceng.2024.109370>
- [23] Morris, C. J. et al. Endogenous circadian system and circadian misalignment impact glucose tolerance via separate mechanisms in humans. *Proc. Natl. Acad. Sci. USA* **112**, E2225–E2234 (2015). <https://doi.org/10.1073/pnas.1418955112>
- [24] Chellappa, S. L., Vujovic, N., Williams, J. S. & Scheer, F. A. J. L. Impact of circadian disruption on cardiovascular function and disease. *Trends Endocrinol. Metab.* **30**, 767–779 (2019). <https://doi.org/10.1016/j.tem.2019.07.008>
- [25] Mavaddat, N. et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J. Natl. Cancer Inst.* **107**, djv036 (2015). <https://doi.org/10.1093/jnci/djv036>
- [26] Rose, G. Sick individuals and sick populations. *Int. J. Epidemiol.* **14**, 32–38 (1985). <https://doi.org/10.1093/ije/14.1.32>
- [27] Knutsson, A. Methodological issues in shift work research. *Occup. Med. (Lond)* **54**, 193–198 (2004). <https://doi.org/10.1093/occmed/kqh039>

- [28] McEwen, B. S. Protective and damaging effects of stress mediators. *N. Engl. J. Med.* **338**, 171–179 (1998). <https://doi.org/10.1056/NEJM199801153380307>
- [29] Vyas, M. V. et al. Shift work and vascular events: systematic review and meta-analysis. *BMJ* **345**, e4800 (2012). <https://doi.org/10.1136/bmj.e4800>
- [30] Reutrakul, S. & Knutson, K. L. Consequences of circadian disruption on cardiometabolic health. *Sleep Med. Clin.* **10**, 455–468 (2015). <https://doi.org/10.1016/j.jsmc.2015.07.005>
- [31] Czeisler, C. A. Perspective: casting light on sleep deficiency. *Nature* **497**, S13 (2013). <https://doi.org/10.1038/497S13a>
- [32] Buxton, O. M. et al. Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. *Sci. Transl. Med.* **4**, 129ra43 (2012). <https://doi.org/10.1126/scitranslmed.3003200>
- [33] Wong, I. S., McLeod, C. B. & Demers, P. A. Shift work trends and risk of work injury among Canadian workers. *Scand. J. Work Environ. Health* **37**, 54–61 (2011). <https://doi.org/10.5271/sjweh.3124>
- [34] Huang, H., Liu, Z., Xie, J. & Xu, C. Association between night shift work and NAFLD: a prospective analysis of 281,280 UK Biobank participants. *BMC Public Health* **23**, 1282 (2023). <https://doi.org/10.1186/s12889-023-16204-7>
- [35] De Bacquer, D. et al. Rotating shift work and the metabolic syndrome: a prospective study. *Int. J. Epidemiol.* **38**, 848–854 (2009). <https://doi.org/10.1093/ije/dyn360>
- [36] Partch, C. L., Green, C. B. & Takahashi, J. S. Molecular architecture of the mammalian circadian clock. *Trends Cell Biol.* **24**, 90–99 (2014). <https://doi.org/10.1016/j.tcb.2013.07.002>

- [37] Koruca, H. İ., Emek, M. S. & Gulmez, E. Development of a personalized staff-scheduling method with work-life balance perspective: case of a hospital. *Ann. Oper. Res.* (2023). <https://doi.org/10.1007/s10479-023-05244-2>
- [38] Whittaker, D. S. et al. Circadian modulation by time-restricted feeding rescues brain pathology and improves memory in mouse models of Alzheimer's disease. *Cell Metab.* **35**, 1704–1721.e6 (2023). <https://doi.org/10.1016/j.cmet.2023.07.014>
- [39] Adamovich, Y. et al. Oxygen and carbon dioxide rhythms are circadian clock controlled and differentially directed by behavioral signals. *Cell Metab.* **29**, 1092–1103.e3 (2019). <https://doi.org/10.1016/j.cmet.2019.01.007>
- [40] Tian, H., Zhao, X., Zhang, Y. & Xia, Z. Research progress of circadian rhythm in cardiovascular disease: a bibliometric study. *Heliyon* **10**, e28738 (2024). <https://doi.org/10.1016/j.heliyon.2024.e28738>
- [41] Gachon, F. et al. Potential bidirectional communication between the liver and the central circadian clock in MASLD. *NPJ Metab. Health Dis.* **3**, 15 (2025). <https://doi.org/10.1038/s44324-025-00058-1>
- [42] Shi, S. Q. et al. Circadian disruption leads to insulin resistance and obesity. *Curr. Biol.* **23**, 372–381 (2013). <https://doi.org/10.1016/j.cub.2013.01.048>
- [43] Chellappa, S. L., Morris, C. J. & Scheer, F. A. J. L. Daily circadian misalignment impairs human cognitive performance task-dependently. *Sci. Rep.* **8**, 3041 (2018). <https://doi.org/10.1038/s41598-018-20707-4>
- [44] Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *JAMA* **285**, 2486–2497 (2001). <https://doi.org/10.1001/jama.285.19.2486>

- [45] Zhang, L., Sun, D. M., Li, C. B. & Tao, M. F. Influencing factors for sleep quality among shift-working nurses. *Asian Nurs. Res.* **10**, 277–282 (2016). <https://doi.org/10.1016/j.anr.2016.09.002>
- [46] Eastman, C. I. How to reduce circadian misalignment in rotating shift workers. *ChronoPhysiol. Ther.* **6**, 41–46 (2016). <https://doi.org/10.2147/CPT.S111424>
- [47] Chellappa, S. L. et al. Daytime eating during simulated night work mitigates changes in cardiovascular risk factors. *Nat. Commun.* **16**, 57846 (2025). <https://doi.org/10.1038/s41467-025-57846-y>
- [48] Molzof, H. E. et al. Nightshift work and nighttime eating are associated with higher insulin and leptin levels in hospital nurses. *Front. Endocrinol.* **13**, 876752 (2022). <https://doi.org/10.3389/fendo.2022.876752>
- [49] Ezerins, M. E. et al. Advancing safety analytics: a diagnostic framework for assessing system readiness within occupational safety and health. *Saf. Sci.* **146**, 105569 (2022). <https://doi.org/10.1016/j.ssci.2021.105569>
- [50] Baron, K. G. & Reid, K. J. Circadian misalignment and health. *Int. Rev. Psychiatry* **26**, 139–154 (2014). <https://doi.org/10.3109/09540261.2014.911149>
- [51] Salehi, A. & Khedmati, M. Identifying at-risk patients for congenital heart disease using integrated predictive models and fuzzy clustering analysis: A cross-sectional study. *Heliyon* **10**, e39609 (2024). <https://doi.org/10.1016/j.heliyon.2024.e39609>
- [52] Epstein, M., Arakelian, E., Tucker, P. & Dahlgren, A. Managing sustainable working hours within participatory scheduling for nurses and assistant nurses: a qualitative interview study. *J. Nurs. Manag.* **2023**, 8096034 (2023). <https://doi.org/10.1155/2023/8096034>

- [53] Lee, I. et al. Metabolic risk stratification of night shift workers in a large retail workplace through clustering and SHAP interpretation. *Front. Public Health* **13**, 1704046 (2026). <https://doi.org/10.3389/fpubh.2025.1704046>
- [54] Chemkomnerd, N. et al. A scenario-driven simulation approach to sustainable hospital resource management: aging society, pandemic preparedness and referral enhancement. *BMC Health Serv. Res.* **25**, 982 (2025). <https://doi.org/10.1186/s12913-025-13221-7>
- [55] European Agency for Safety and Health at Work (EU-OSHA). *Current and emerging issues in the healthcare sector, including home and community care* (Publications Office of the European Union, 2014). <https://doi.org/10.2802/72773>

**Table****Table. 1. Baseline characteristics of night-shift and non-night workers**

<b>Variable</b>	<b>Night-shift (n = 1,151)</b>	<b>Non-night (n = 1,099)</b>	<b>P- value</b>
<b>Age (years)</b>	30.9 ± 8.8	40.7 ± 11.0	< 0.001
<b>BMI (kg/m<sup>2</sup>)</b>	22.4 ± 3.5	24.5 ± 4.0	< 0.001
<b>Waist circumference (cm)</b>	72.1 ± 9.6	84.2 ± 8.2	< 0.001
<b>SBP (mmHg)</b>	116.1 ± 12.6	121.4 ± 12.4	< 0.001
<b>DBP (mmHg)</b>	72.7 ± 9.7	76.6 ± 9.3	< 0.001
<b>Fasting glucose (mg/dL)</b>	94.8 ± 10.0	100.3 ± 19.8	< 0.001
<b>Triglyceride (mg/dL)</b>	88.3 ± 58.8	129.7 ± 104.7	< 0.001
<b>HDL-cholesterol (mg/dL)</b>	66.8 ± 14.9	59.4 ± 16.3	< 0.001
<b>LDL-cholesterol (mg/dL)</b>	113.1 ± 29.2	115.9 ± 34.8	0.084
<b>Job role, n (%)</b>			
↳ <b>Nursing staff</b>	483 (42.0)	444 (40.4)	—
↳ <b>Clinical staff</b>	299 (26.0)	293 (26.7)	—
↳ <b>Administrative staff</b>	219 (19.0)	215 (19.6)	—
↳ <b>Facility / support staff</b>	150 (13.0)	147 (13.3)	—

Values are presented as mean ± standard deviation or number (%). SBP,

systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index.

## **Legends for schematics and figures**

### **Fig. 1. Conceptual overview of the PHATE-NAM analytical workflow**

This scheme illustrates the analysis process that constitutes a potential risk landscape. The probability of metabolic risk was estimated based on LightGBM using physiological variables such as BMI, blood pressure, and geological indicators, as well as occupational variables related to night work. The right panel displays the results of projecting these multidimensional profiles into a low-dimensional space using the PHATE technique. In the formed latent space, PHATE-1 reflects the systemic metabolic burden, and PHATE-2 reflects the cholesterol homeostasis pattern. Each point represents the physiological position of an individual in that space.

### **Fig. 2. Physiological distribution of hospital workers mapped onto the PHATE manifold**

This figure is a visualization of the physiological distribution of hospital workers. Night workers (red) and day workers (blue) are distinct from each other in the phase space. The PHATE-1 axis reflects the systemic metabolic burden and shows positive correlations with body mass index ( $r=0.71$ ), systolic blood pressure ( $r=0.67$ ), and triglycerides ( $r=0.56$ ). The PHATE-2 axis represents the cholesterol homeostasis profile, which is defined as the conflicting contributions between LDL ( $r=0.67$ ) and HDL ( $r=0.51$ ). The density contour was calculated by Gaussian kernel density estimation. It means that higher densities indicate higher phase concentrations of individuals in the region. The

night shift distribution was significantly to the right in the PHATE-1 axis direction. This means that structural changes associated with increased metabolic burden have appeared ( $p < 0.001$ ).

### **Fig. 3. Projection of predicted metabolic risk across the global PHATE landscape**

This figure shows the result of projecting the predicted risk probability into the PHATE latent space using a lightweight gradient boosting machine (LightGBM) model. The axis consists of a dimensionless latent distance that preserves the global structure of the physiological state, PHATE-1 reflects systemic metabolic load, and PHATE-2 reflects cholesterol homeostasis-related variations. Regions with a predicted probability of 0.70 or more (red series) tend to be concentrated on the right side of the manifold, that is, in the section with high metabolic load. The average AUC of the model was 0.886 (standard deviation = 0.018), confirming that the risk distribution in the latent space was stably classified. This nonlinear risk distribution visually presents the change in risk level according to the physiological position in the PHATE space.

### **Fig. 4. Occupational gradient patterns among night workers in latent space**

As a result of the latent slope distribution analysis targeting only night workers, the high-density areas of (a) continuous night shift days, (b) cumulative working hours, and (c) weekly working hours significantly overlapped with the high-risk section of PHATE-1. This pattern was reflected as a nonlinear effect in the NAM model, and the risk probability tended to increase if it deviated from the section.

### **Fig. 5. Interpretable risk components derived from the NAM model**

(a) As a result of analyzing the variable contribution of the Natural Additive Model (NAM), triglyceride (TG), blood pressure (SBP/DBP), total cholesterol (TC), and working hours per week showed a direction associated with increased risk. On the other hand, PHATE-2 (cholesterol homeostasis profile) and continuous working days were associated with a risk reduction trend. (b-d) partial dependence curves representing nonlinear risk changes were presented. Biometric indicators were presented in standard clinical units, and triglycerides were presented in mmol/L, blood pressure was presented in mmHg, and working hours were presented in hours/week. (b) A nonlinear increase in risk was observed in the section where triglycerides exceeded about 150 mg/dL ( $\approx$  1.7 mmol/L). (c) Working hours per week showed a U-shaped curve with the lowest risk in the 40-50-hour interval. (d) As the PHATE-2 value increased, the risk tended to decrease overall.

Abbreviations: BMI, body mass index; LDL/HDL, low-/high-density lipoprotein; G-GTP, gamma-glutamyl transpeptidase; ALT/AST, alanine/aspartate aminotransferase.

### **Fig. 6. Model discrimination and calibration performance of the NAM framework**

(a) The ROC curve of the NAM model calculated through repeated cross-validation was presented. The average AUC was 0.864. (b) A calibration curve shows the degree of agreement between the predicted probability and the actual observed risk. Overall, good consistency was confirmed between the predicted and observed values.

Abbreviations: ROC, receiver operating characteristic; AUC, area under the curve.

**Fig. 7. VaDE-based clustering with risk stratification on the PHATE manifold**

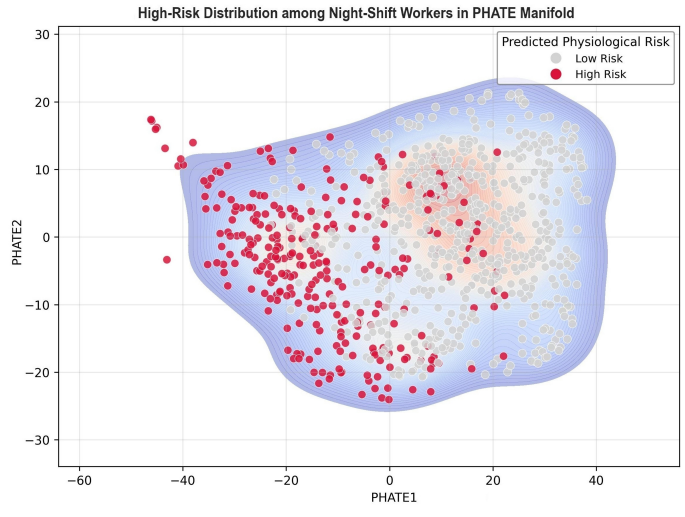
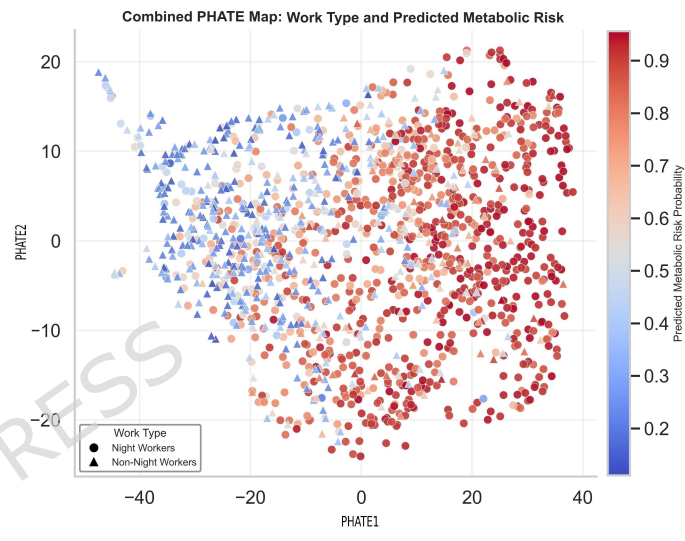
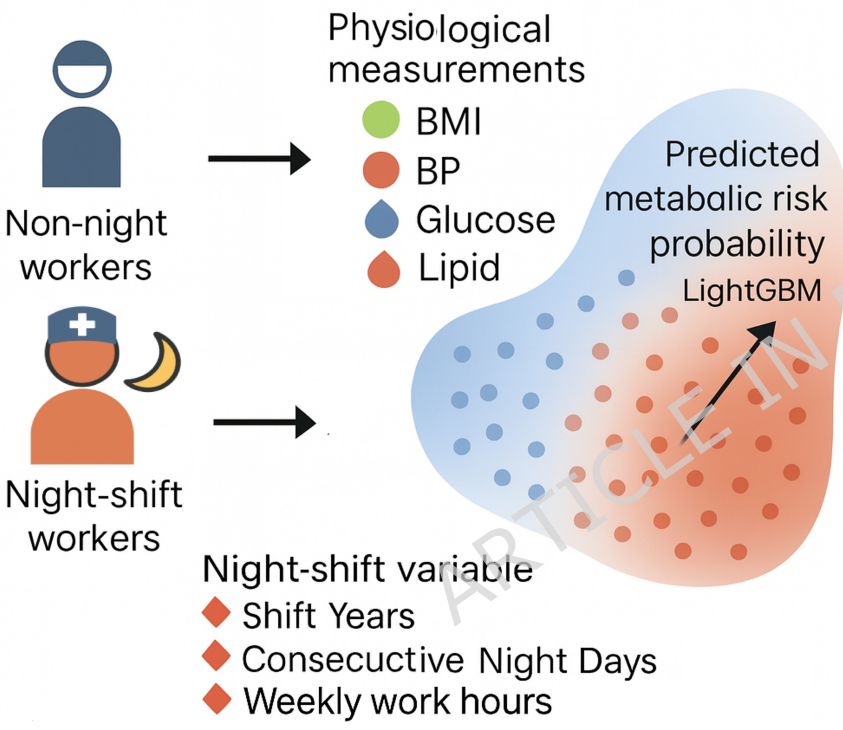
These are the three potential subtypes of night workers identified by Variational Deep Embedding (VaDE). The axes represent dimensionless potential distances mapping the global manifold geometry, where PHATE-1 captures the gradient of Metabolic Load and PHATE-2 reflects the Cholesterol Homeostasis Profile. Cluster 1 (orange) is the 'metabolic overload group' characterized by a negative PHATE-1 gradient, with the highest average predicted risk (0.71). Cluster 2 (green) is a low-risk group with a high PHATE-2 score, which maintains a relatively stable lipid metabolism. The size of the bubble refers to the individual risk probability predicted by NAM (average ARI = 0.858). From a clinical point of view, this stratification enables an occupational health monitoring system based on an individual's physiological risk level. Cluster 1 is identified as a group with relatively high cardiovascular risk characteristics, regardless of age, and can be a priority for more intensive risk management.

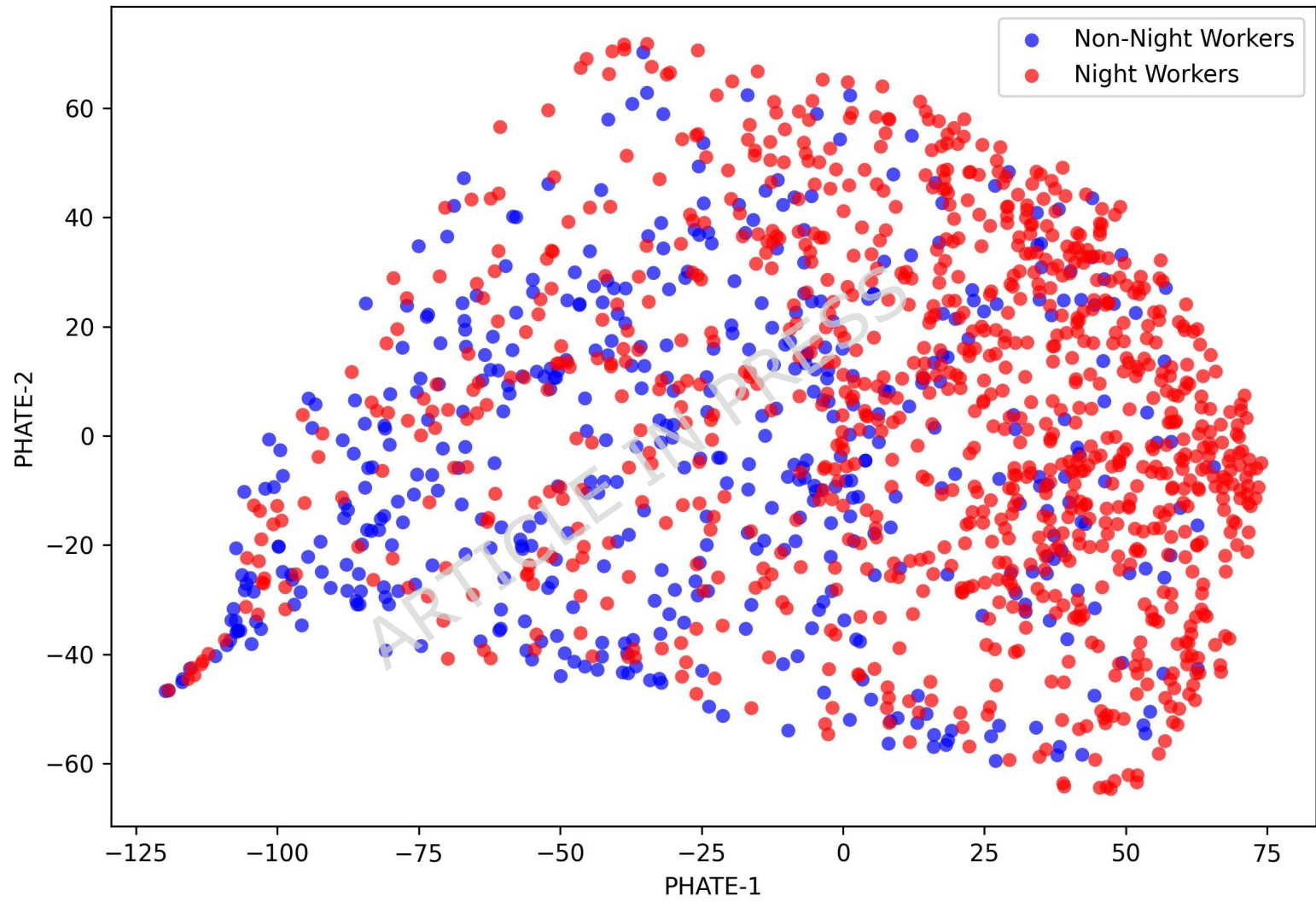
**Fig. 8. Density contours of high-risk regions within the PHATE space (night-shift subset)**

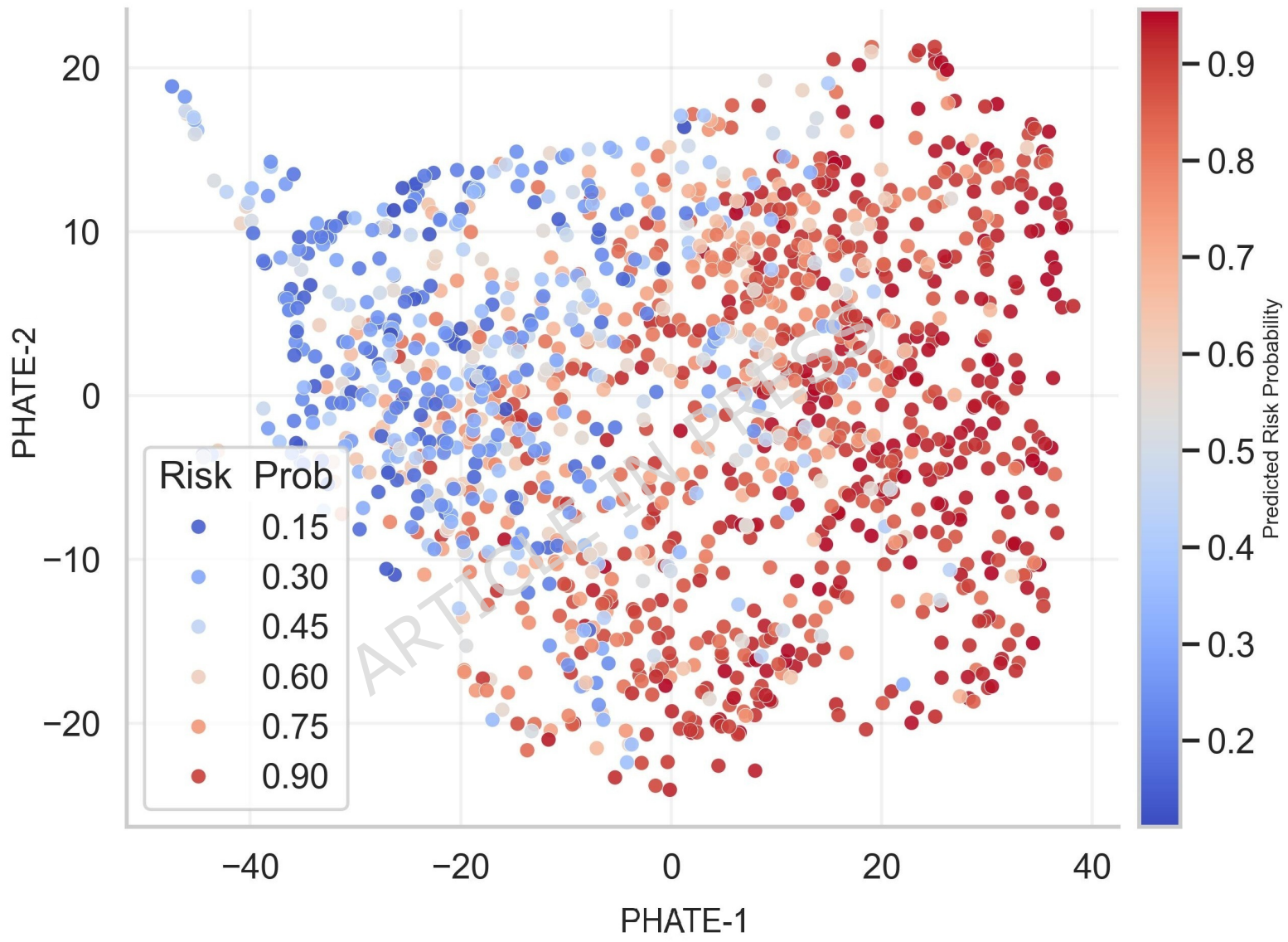
The distribution density of high-risk probabilities was visualized through Gaussian kernel density estimation (KDE). Density vertices were formed in the region where PHATE-1 (metabolic load) was negative and PHATE-2 (cholesterol homeostasis profile) was low. The contour lines spatially represent the nonlinear interaction structure between risk factors and present a boundary

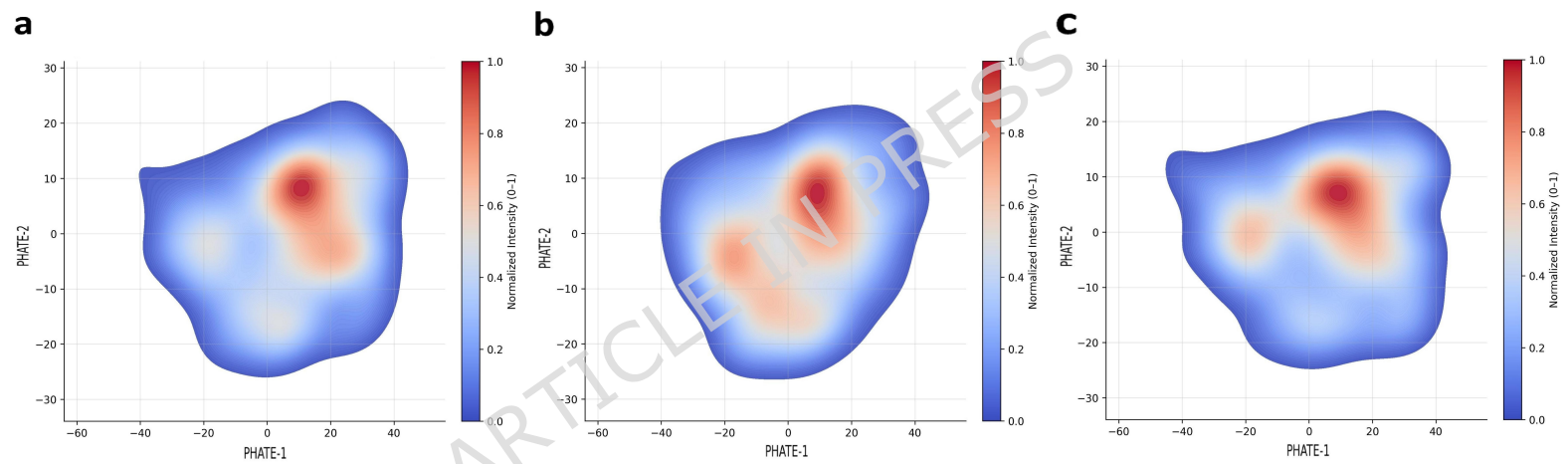
that can distinguish between sections where the cumulative workload exceeds the physiological buffer range. This distribution pattern suggests the necessity of designing a work structure that considers the transition to the high-density risk section.

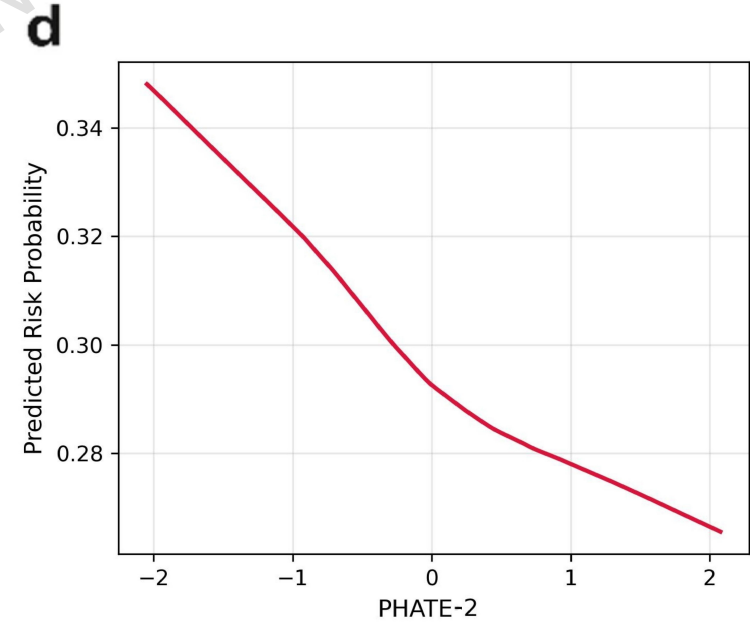
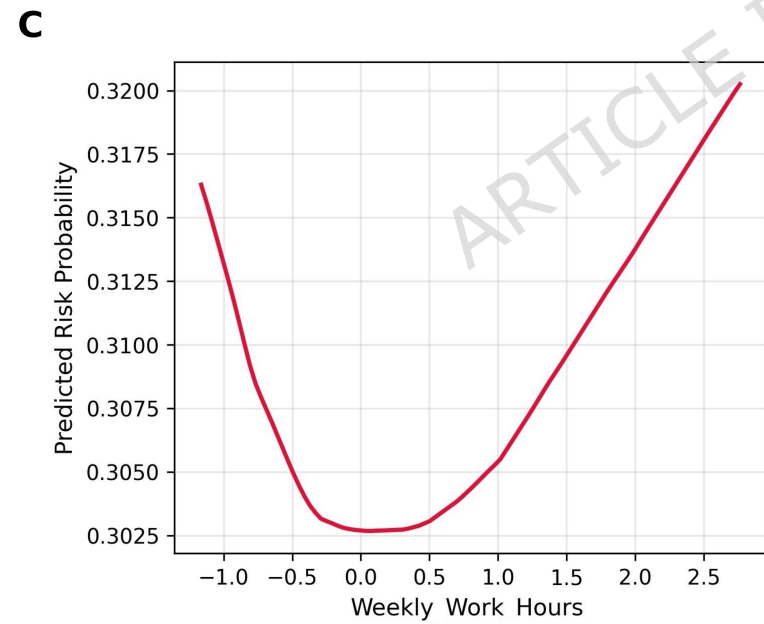
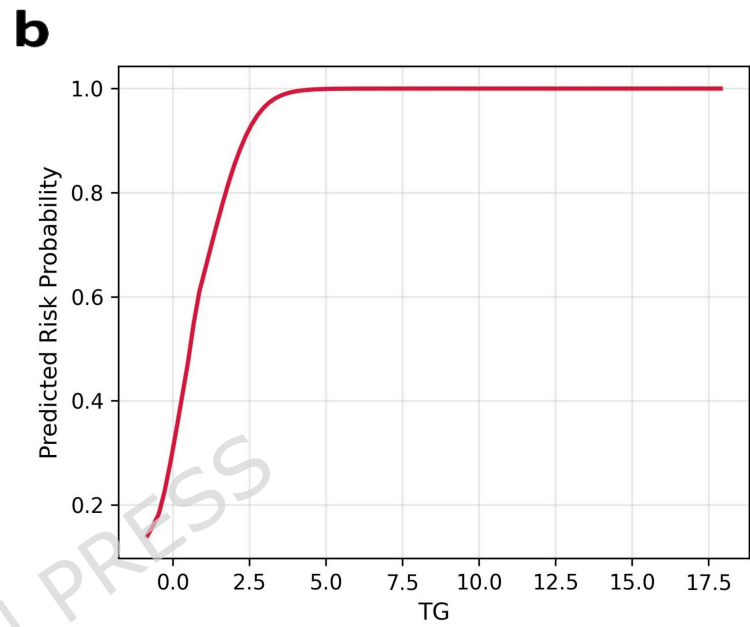
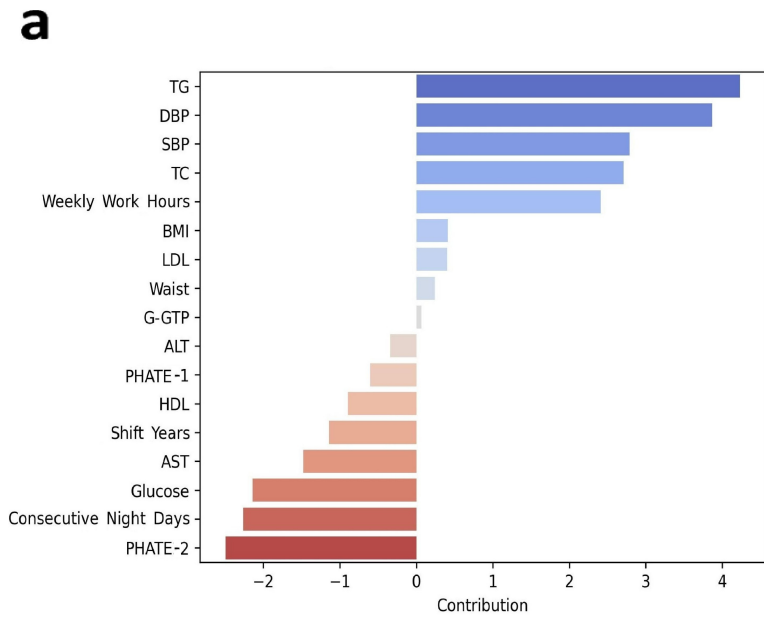
ARTICLE IN PRESS

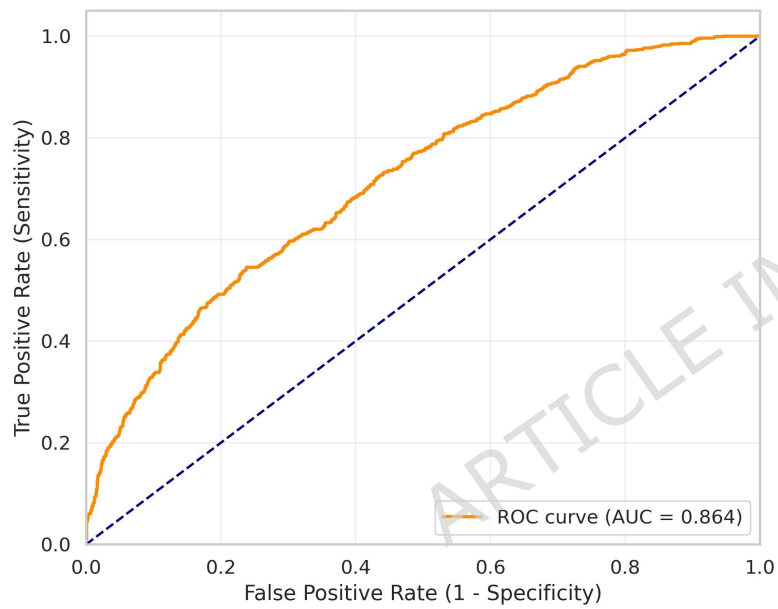










**a****b**