1828
AKADÉMIAI KIADÓ

Low-burden data-driven work addiction screening: Comparing demographic and psychosocial predictors using machine learning

NATALIA A. WOROPAY-HORDZIEJEWICZ[1]*, EDYTA CHARZYŃSKA[2]**, ALEKSANDRA BUŹNIAK[3], STANISŁAW K. CZERWIŃSKI[3], ZUZANNA SCHNEIDER[4] and PAWEŁ A. ATROSZKO[3,5]***

Journal of Behavioral Addictions

15 (2026) 2, 900–914

DOI:
10.1556/2006.2026.00060

FULL-LENGTH REPORT

[1] Faculty of Psychology, University of Warsaw, Warsaw, Poland

[2] Institute of Psychology, Institute of Pedagogy, Faculty of Social Sciences, University of Silesia in Katowice, Katowice, Poland

[3] Institute of Psychology, Faculty of Social Sciences, University of Gdańsk, Gdańsk, Poland

[4] Institute of Psychology, Faculty of Social Sciences, University of Silesia in Katowice, Katowice, Poland

[5] Department of Psychological Sciences, FLAME University, Pune, India

Received: February 8, 2026 • Revised manuscript received: April 25, 2026 • Accepted: June 8, 2026
Published online: June 30, 2026

ABSTRACT

Background and Aims: Identifying employees at risk for work addiction is critical for early intervention, yet systematic screening approaches face practical challenges for organizations. This study compared two screening strategies: a demographic-only model that requires no additional data collection versus a psychosocial model that incorporates job satisfaction and job stress assessments. *Methods:* We analyzed data from employees across 85 countries or territories. Dataset 1 (psychosocial model, $n = 22{,}136$) included demographic, occupational, and psychosocial variables. Dataset 2 (demographic model, $n = 23{,}219$) used only readily available information. Work addiction risk was assessed using the International Work Addiction Scale. Five machine learning algorithms (Logistic Regression, Random Forest, XGBoost, Neural Networks, Ensemble Voting) were evaluated on both full and Elastic Net-selected feature sets using cross-validation. *Results:* The psychosocial model demonstrated superior performance. The best-balanced model (Logistic Regression with Elastic Net selection) achieved balanced accuracy of 0.72 ± 0.01, AUC-ROC of 0.80 ± 0.01, and Youden's J of 0.45 ± 0.02, correctly identifying 72.7% of at-risk individuals while maintaining 72.2% specificity. The demographic model demonstrated moderate performance (Ensemble Voting: balanced accuracy 0.61 ± 0.01, AUC-ROC 0.65 ± 0.01), yet retained practical utility, detecting 57.0% of at-risk cases. High-sensitivity configurations achieved detection rates of 71%, but with elevated false-positive rates, presenting important trade-offs for screening design. *Discussion and Conclusions:* While psychosocial variables provide superior prediction, demographic-only models offer a practical screening alternative that can identify over half of at-risk individuals without an additional data-collection burden. This trade-off between predictive performance and implementation feasibility has important implications for organizational screening programs.

KEYWORDS

work addiction, workaholism, machine learning, screening, job stress, occupational health

*Corresponding author.
E-mail: n.woropay-hordziejewicz@uw.edu.pl

**Corresponding author.
E-mail: edyta.charzynska@us.edu.pl

***Corresponding author.
E-mail: pawel.atroszko@flame.edu.in

INTRODUCTION

Work addiction, commonly referred to as workaholism (Atroszko et al., 2025), has increasingly been recognized as a significant yet underdetected occupational health risk and

problematic workplace behavior, associated with impaired well-being, mental health problems, emotional exhaustion and burnout, potential long-term somatic consequences, less prosocial behavior at work, more counterproductive work behavior, bullying behaviors towards colleagues and subordinates, and decreased productivity (Atroszko, 2022; Charzyńska et al., 2025; Clark, Michel, Zhdanova, Pui, & Baltes, 2016; Taris & de Jonge, 2024). Workaholism may directly or indirectly negatively affect the work environment and social and family functioning (Atroszko et al., 2025; Kenyhercz, Mervó, Lehel, Demetrovics, & Kun, 2024). Meta-analysis of 53 studies found that approximately 14% of workers meet criteria for work addiction (Andersen, Djugum, Sjåstad, & Pallesen, 2023), with higher rates consistently reported among managers and self-employed individuals (Atroszko & Atroszko, 2020; Moskalewicz et al., 2019). Moreover, a recent global study across 85 cultures (individual samples) showed that the average prevalence of work addiction based on the high-risk profile in latent profile analysis was about 19% (Charzyńska et al., 2025). Work addiction tends to be among the most prevalent problematic behaviors when compared to problematic gambling, shopping, internet and social media use in nationally representative samples (Moskalewicz et al., 2019), underscoring the public health and organizational relevance of systematic screening.

Despite growing empirical evidence and advances in theory, the systematic identification of individuals at risk remains limited in applied settings, largely due to the reliance on self-report instruments that require additional assessment effort and are rarely implemented at scale. From a prevention and public health perspective, there is a need for screening approaches that are feasible, scalable, and compatible with existing organizational or administrative data infrastructures. Demographic characteristics constitute one such underutilized resource. They are routinely collected, non-intrusive, and stable over time. However, their utility for identifying elevated risk of work addiction has rarely been evaluated in direct comparison with psychosocial predictors, nor using analytical approaches optimized for prediction rather than explanation. The present study addresses this gap by applying machine-learning methods to compare the screening value of demographic and psychosocial predictors of work addiction, with the aim of informing low-burden, administratively feasible approaches to early risk identification.

Work addiction is conceptualized as a behavioral addiction characterized by a persistent and compulsive pattern of excessive work engagement, accompanied by impaired control over working behavior, preoccupation with work-related activities, and continuation despite negative consequences (Atroszko, Demetrovics, & Griffiths, 2019; Charzyńska et al., 2025; Kun, Takacs, Richman, Griffiths, & Demetrovics, 2020). Unlike normative hard work, work addiction is not defined by the number of hours worked per se but by the underlying motivational and self-regulatory dysfunction that drives the behavior and distinguishes it clearly from healthy work engagement (Bereznowski, Konarski, Pallesen, & Atroszko, 2025). Empirical research has consistently linked work addiction to individual harms and impairments, social pathology, and economic costs (Atroszko, 2022). Importantly, work addiction shows substantial stability over time, cross-cultural generalizability, and meaningful associations with personality traits, job demands, and broader sociocultural contexts, supporting its status as a distinct and clinically relevant construct (Charzyńska et al., 2025; Clark et al., 2016; Kun et al., 2020). At the same time, accumulating evidence suggests that work addiction often develops and persists in environments where excessive work is socially rewarded or structurally encouraged, complicating detection and intervention (Atroszko et al., 2025). This underscores the importance of screening approaches that do not rely solely on self-disclosure or symptom awareness but can be integrated into broader occupational health monitoring frameworks.

Based on prior research and the availability of routinely collected administrative information, the present study focused on a set of sociodemographic and occupational variables commonly examined in the work addiction literature. These included age and gender, as well as indicators of human capital and career positioning such as highest educational attainment, occupational sector, managerial status, company size, length of employment at the current organization, and approximate monthly income level (for an overview, see Atroszko, 2022). To capture structural exposure to job demands, average weekly working hours, including overtime, were included as a core occupational characteristic. In addition, several variables reflecting nonwork context and potential constraints on time and energy investment were considered, including household size, number of children, and partnership status. Together, these variables represent relatively stable characteristics that are typically available in organizational or survey-based records. Only some of them have been linked in previous research, sometimes inconsistently, to work addiction (Clark et al., 2016). Managerial status and working time are highly consistently associated with work addiction risk (Atroszko & Atroszko, 2020). Nevertheless, there is more consistent evidence for their role in shaping general work intensity, role expectations, and patterns of work investment (Eurofound, 2017). Moreover, inconsistent associations, such as those involving gender, age, education, or sector, suggest moderating effects that can be effectively explored with machine learning. Examining their combined and relative predictive value allows for an assessment of whether low-burden demographic and occupational data can meaningfully contribute to screening for elevated risk of work addiction. The psychosocial model additionally incorporated two work-related well-being variables, routinely included in occupational health surveys: perceived job stress and job satisfaction. These constructs were selected because they represent among the most consistently studied psychosocial correlates of work addiction (Atroszko, 2022; Clark et al., 2016), they are proximal work environment indicators within the job demands-resources framework, and they are widely available in organizational wellness monitoring

contexts. Including them allows a direct comparison between a purely administrative screening approach and one that requires minimal additional self-report data.

The present study compared demographic-only and psychosocial screening approaches across multiple machine learning algorithms, with primary focus on detection rates of work addiction at-risk individuals. Specifically, we addressed four key research questions. First, can machine learning models achieve meaningful discrimination between at-risk and no-risk employees using demographic versus psychosocial predictors? Second, what proportion of at-risk individuals can be successfully identified using easily available demographic data compared to psychosocial assessment? Third, do models optimized for maximum detection accept prohibitively high false positive rates, or can balanced approaches maintain adequate sensitivity while controlling misclassification? Fourth, which predictors contribute most to risk detection, and how does the importance hierarchy differ between demographic-only and psychosocial models? Addressing these questions informs both theoretical understanding of work addiction determinants and practical decisions regarding the implementation of screening in organizational settings.

METHODS

Participants and data collection

Data were collected as part of a larger international collaborative project examining work addiction and its correlates across diverse cultural contexts. The initial sample consisted of 34,899 full-time employees from 91 countries or territories spanning six continents.

Participants were required to meet specific inclusion criteria: citizenship and current residence in the country or territory of data collection, legal adult status according to national legislation, full-time employment, employment in an organization with at least 10 employees, and tenure with the current employer of at least one year. These restrictions ensured that participants had sufficient organizational experience and work exposure to develop patterns relevant to work addiction assessment. In addition, two attention check items were included to ensure data quality.

Data were collected through online surveys beginning in autumn 2022, using convenience and snowball sampling methods coordinated by local research teams in each participating country or territory. Participation was voluntary and anonymous, with no compensation provided. Data collection was completed in all participating countries or territories by winter 2023, with the exception of Cambodia, where data were collected in autumn 2024 due to delays in obtaining ethical approval.

Final sample preparation

From the initial sample of 34,899 participants, we applied additional filtering criteria to ensure data quality and analytical appropriateness. Cases were excluded if they had missing values on any predictor variables or the work addiction outcome measure. We also excluded participants who worked fewer than 30 h per week to ensure sufficient work exposure, and participants from countries or territories with fewer than 150 observations to ensure adequate representation for cross-cultural analyses. The final sample consisted of 85 countries or territories.

Two analytical datasets were constructed from the filtered sample. Dataset 1 (psychosocial model) comprised 22,136 employees with complete data on all demographic, occupational, and psychosocial variables (job satisfaction and job stress). Dataset 2 (demographic model) comprised 23,219 employees with complete demographic and occupational data, excluding the psychosocial measures, to simulate a screening scenario using only readily available human resources information.

Model development and validation employed stratified k-fold cross-validation ($k = 5$; Kohavi, 1995) to ensure robust performance estimates. This approach divides the data into k equal-sized folds, while maintaining class proportions, trains models on k-1 folds, and evaluates on the remaining fold, repeating until all folds serve as test sets. The stratification maintained consistent class distributions across folds (~88% no-risk, ~12% at-risk in Dataset 1; ~87% no-risk, ~13% at-risk in Dataset 2). Performance metrics represent averages across the five test folds, with standard deviations indicating stability. Random state was fixed to 42 for reproducibility.

Measures

Work Addiction. Work addiction was assessed using a five-item version of the International Work Addiction Scale (IWAS; Charzyńska et al., 2025). The IWAS-5 was developed to assess core dimensions of work addiction based on established behavioral addiction criteria, including salience, problems, conflict, mood modification, and relapse (WHO, 2019). An example item is "How often during the last year have you tried to reduce the amount of your work but failed?" Items were rated on a five-point Likert scale ranging from 1 ("never") to 5 ("always"), referring to work-related behaviors and experiences during the past year. A total IWAS score was computed by summing responses across the five items, yielding a possible range from 5 to 25, with higher scores indicating greater work addiction risk. In the current sample, internal consistency was good, with Cronbach's alpha of .83 for both datasets.

For the present analysis, work addiction risk was operationalized using a validated cutoff score approach. Following a recent psychometric validation study (Charzyńska et al., 2025), participants scoring 18 or above on the IWAS-5 total score were classified as "at-risk" for work addiction, while those scoring below 18 were classified as "no risk". We acknowledge that transforming a continuous score into a binary outcome reduces available variance and may limit predictive performance relative to regression-based continuous outcome modeling. However, this dichotomization is consistent with the study's primary aim

of developing a practical screening tool. Screening requires a decision boundary to guide organizational action, and the chosen cutoff reflects an empirically established risk threshold (Charzyńska et al., 2025).

Job Satisfaction. Job satisfaction was measured with a single item developed by Dolbier, Webster, McCalister, Mallon, and Steinhardt (2005): "Taking everything into consideration, how do you feel about your job as a whole?" Responses were recorded on a seven-point Likert scale ranging from 1 ("extremely dissatisfied") to 7 ("extremely satisfied"). This short measure has been frequently used in previous studies and has demonstrated adequate reliability and validity (see Dolbier et al., 2005).

Job Stress. Perceived job stress was assessed with a single item developed by Houdmont et al. (2021): "In general, how do you find your job?" rated on a seven-point Likert scale from 1 ("not at all stressful") to 7 ("extremely stressful"). The reliability and validity of the item have been supported in previous studies (Houdmont et al., 2021).

Demography and Work-Related Variables. Participants also reported a set of demographic and work-related characteristics. These included age, gender, highest educational level attained (categorized as primary, secondary, bachelor's degree, master's degree, or doctorate), household size measured as the number of individuals living in the participant's household, number of children, average weekly work hours including overtime, length of employment at the current organization in years, occupational sector (private, public), company size measured by number of employees on an ordinal scale, managerial status (non-managerial/junior/mid-level/senior), approximate monthly income level on an ordinal scale, and partnership status including categories for single, partnered living together/apart, married, divorced, or widowed.

Machine learning analyses

Data preprocessing and model development. We evaluated ten model configurations representing five algorithmic approaches applied to both full and Elastic Net-selected feature sets. Elastic Net regularization (Zou & Hastie, 2005) with alpha = 0.005, l1_ratio = 0.7, and threshold = 1×10^{-5} was performed to reduce dimensionality while maintaining predictive performance. This approach selected 10 of 24 features (41.7%) in Dataset 1 and 8 of 22 features (36.4%) in Dataset 2. Five algorithm families were selected to span a range of model complexity and balance interpretability with predictive performance: Logistic Regression as a transparent linear baseline, providing a clear reference for linear relationships between predictors and the outcome; Random Forest as an ensemble of decision trees that reduces overfitting through aggregation and improves generalization; XGBoost as a gradient-boosting approach that captures complex non-linear relationships between variables; Neural Networks for flexible learning of more intricate patterns in the data; and an Ensemble Voting classifier that combines the strengths of selected base models to improve overall predictive performance. Throughout this paper, "higher sensitivity configurations" refers to models predicting more at-risk cases due to optimized thresholds or stronger class weighting, increasing sensitivity at the cost of more false positives. For the Neural Network and Ensemble models, this threshold was optimized within each training fold to maximize Youden's *J*. The remaining classifiers (Logistic Regression, Random Forest, XGBoost) were evaluated using the standard 0.5 decision threshold, with Youden's *J* reported post-hoc from the resulting predictions. Elastic Net combines L1 (Lasso) and L2 (Ridge) penalties: L1 encourages variable selection by driving some coefficients to zero, while L2 stabilizes estimates when predictors are correlated, making this combination appropriate given moderate intercorrelations among predictors. The l1_ratio of 0.7 favors variable selection, while alpha controls overall regularization strength. Elastic Net feature selection was performed on the full dataset prior to cross-validation; while this approach is common in applied research, it may introduce a small degree of optimistic bias in performance estimates, which is acknowledged as a limitation.

Algorithms

Logistic Regression. Logistic regression models employed elastic net regularization (Zou & Hastie, 2005) with balanced class weighting, l1_ratio = 0.5, $C = 0.5$, and saga solver (Defazio, Bach, & Lacoste-Julien, 2014) with a maximum of 2,000 iterations.

Random Forest. Random forest classifiers (Breiman, 2001) used 300 trees with balanced subsample weighting, maximum depth of 12, minimum samples per split of 15, minimum samples per leaf of 5, and sqrt maximum features following established recommendations for imbalanced data (Chen, Liaw, & Breiman, 2004).

XGBoost. Gradient boosting models employed the XGBoost algorithm (Chen & Guestrin, 2016) with 300 estimators, learning rate 0.05, maximum depth 5, minimum child weight 3, and subsample ratios of 0.8 for both samples and features. Scale_pos_weight parameters provided a 1.5 × boost for the minority class to address class imbalance.

Neural Networks. Deep neural networks utilized a four-layer architecture (128→64→32→16 neurons) implemented in TensorFlow and Keras (Abadi et al., 2016). Models incorporated batch normalization (Ioffe & Szegedy, 2015), dropout regularization (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) with rates of 0.3, 0.25, 0.2, and 0.15 per layer, L2 penalties of 0.001, and focal loss (Lin, Goyal, Girshick, He, & Dollár, 2017) with gamma = 2.0 and alpha = 0.25 to address class imbalance. Enhanced class weighting provided 1.5 × boost for the minority class. Decision thresholds were optimized using Youden's *J* statistic (Youden, 1950) rather than default 0.5 probability cutoffs. Training employed early stopping with patience = 15 epochs across 150 maximum epochs, batch size 32, and Adam optimizer (Kingma & Ba, 2014) with learning rate 0.001.

Ensemble Voting. Ensemble models combined predictions from Logistic Regression, Random Forest, and XGBoost using soft voting with Youden's *J*-optimized thresholds (Dietterich, 2000). This approach leverages complementary strengths of different algorithms to improve overall prediction.

Training was implemented in Python 3.13 using TensorFlow 2.20 with the Keras API (Abadi et al., 2016) and scikit-learn 1.7 (Pedregosa et al., 2011). Data manipulation was performed using Pandas 2.3 and NumPy 2.3, and visualizations were created using Matplotlib 3.10. All experiments were run in Visual Studio Code.

Model evaluation. Primary evaluation focused on detection rates of at-risk individuals (sensitivity/recall for the minority class) and balanced performance metrics. We calculated overall accuracy, balanced accuracy (arithmetic mean of sensitivity and specificity; Brodersen, Ong, Stephan, & Buhmann, 2010), and AUC-ROC (Hand & Till, 2001). For each class, we computed precision, recall, and F1-score. Utility was assessed through sensitivity (true positive rate), specificity (true negative rate), positive predictive value, negative predictive value, and Youden's *J* statistic (Youden, 1950), which maximizes balanced discrimination. Additionally, the Matthews Correlation Coefficient (MCC; Matthews, 1975) was computed as a summary measure of classification quality under class imbalance. MCC ranges from −1 to +1, where +1 indicates perfect prediction, 0 indicates performance no better than chance, and −1 indicates systematic inverse prediction. Unlike accuracy or F1-score, MCC accounts for all four cells of the confusion matrix simultaneously, making it particularly informative when class distributions are unequal.

Feature importance was quantified using model-specific methods: absolute coefficients for logistic regression, mean decrease in Gini impurity for Random Forest (Breiman, 2001), and gain for XGBoost (Chen & Guestrin, 2016). For ensemble models, importance values represent unweighted averages of Random Forest and XGBoost importance scores.

Model stability was assessed using the coefficient of variation (CV% = standard deviation/mean × 100) across validation folds (Kohavi, 1995). All models demonstrated low variability with CV% consistently below 5%.

Ethics

The study was conducted in accordance with the Declaration of Helsinki. The Ethics Committee at the University of Silesia in Katowice, Poland, approved the study (KEUS266/06.2022). Where applicable, ethical approval was also obtained from local ethics committees. The responses to the survey were anonymous, and participation in the study was voluntary. All subjects were informed about the purpose and content of the study and all provided online informed consent.

RESULTS

Detection rates

The central question for practical screening is how many at-risk individuals each approach successfully identifies. Figures 1a and 1b display sensitivity, specificity, F1-score, and accuracy across all models for both datasets.

In the psychosocial dataset, Neural Network with full features achieved one of the highest sensitivity of 75.3 ± 1.6%, correctly identifying approximately 396 of 526 at-risk cases. However, this came at substantial cost to specificity (68.6 ± 1.8%), incorrectly flagging approximately 1,226 of 3,900 no-risk cases (31.4% false positive rate). For balanced performance, Logistic Regression with Elastic Net selection achieved optimal results (Youden's $J = 0.45 \pm 0.02$), correctly identifying 72.7 ± 1.2% of at-risk cases while maintaining 72.2 ± 0.7% specificity.

In the demographic dataset, the Ensemble Voting approach with full features achieved best balance (Youden's $J = 0.23 \pm 0.004$), detecting 57.0 ± 1.2% of at-risk cases with 65.8 ± 1.0% specificity. While lower than psychosocial performance, this represents meaningful screening capability using only readily available across human resources in organizational settings. Demographic models with higher sensitivity configurations (XGBoost) detected up to 71.1% of at-risk cases but with elevated false positive rates (54.6%). Descriptive demographic and work-related characteristics of participants by dataset are provided in Table A1.

Overall model performance comparison

Table 1 presents performance metrics for all model configurations. The psychosocial model demonstrated superior performance across all algorithms. The best psychosocial model (Logistic Regression with Elastic Net selection) achieved balanced accuracy of 0.72 ± 0.01, AUC-ROC of 0.80 ± 0.01, and Youden's *J* of 0.45 ± 0.02. In the demographic dataset, three models tied for optimal balanced performance (Youden's $J = 0.23 \pm 0.004$): Logistic Regression with Elastic Net selection, Ensemble Voting with Elastic Net selection, and Ensemble Voting with full features. These top demographic models achieved balanced accuracy of 0.61 ± 0.01 and AUC-ROC of 0.65 ± 0.01, detecting 54–57% of at-risk cases with 66–68% specificity.

Random Forest models achieved the highest overall accuracy in both datasets (79.8 ± 0.8% psychosocial, 77.5 ± 1.6% demographic) but at substantial cost to sensitivity, correctly identifying only 57.0 ± 1.4% and 37.5 ± 2.5% of at-risk cases respectively. This illustrates the limitation of overall accuracy as a primary metric for imbalanced classification problems, where high accuracy can be achieved by predominantly predicting the majority class.

Feature selection through Elastic Net generally maintained model performance while reducing dimensionality. For Logistic Regression in the psychosocial dataset, Elastic Net selection improved Youden's *J* from 0.447 to 0.448 while reducing features from 24 to 10 (58.3% reduction).

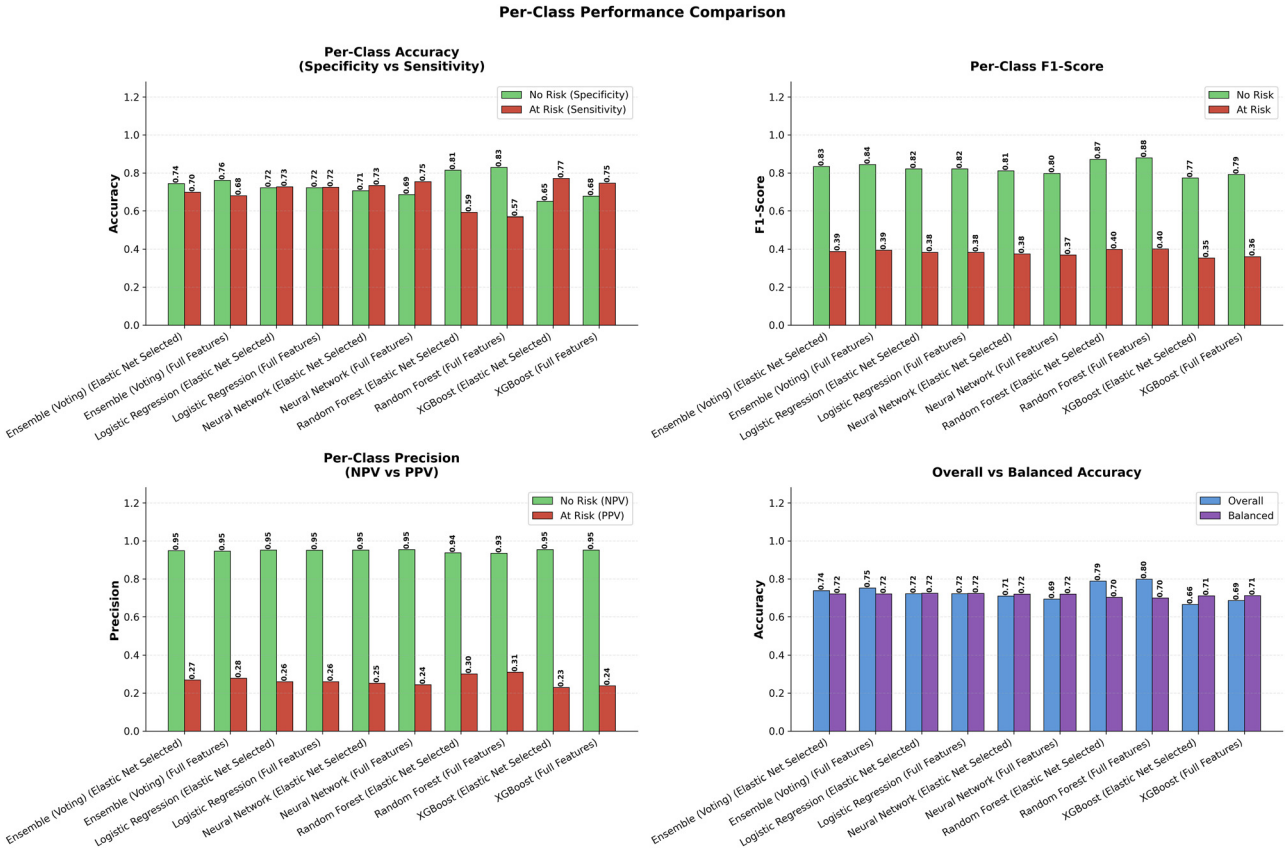

Fig. 1a. Detection rates of at-risk and no-risk individuals across models for psychosocial dataset

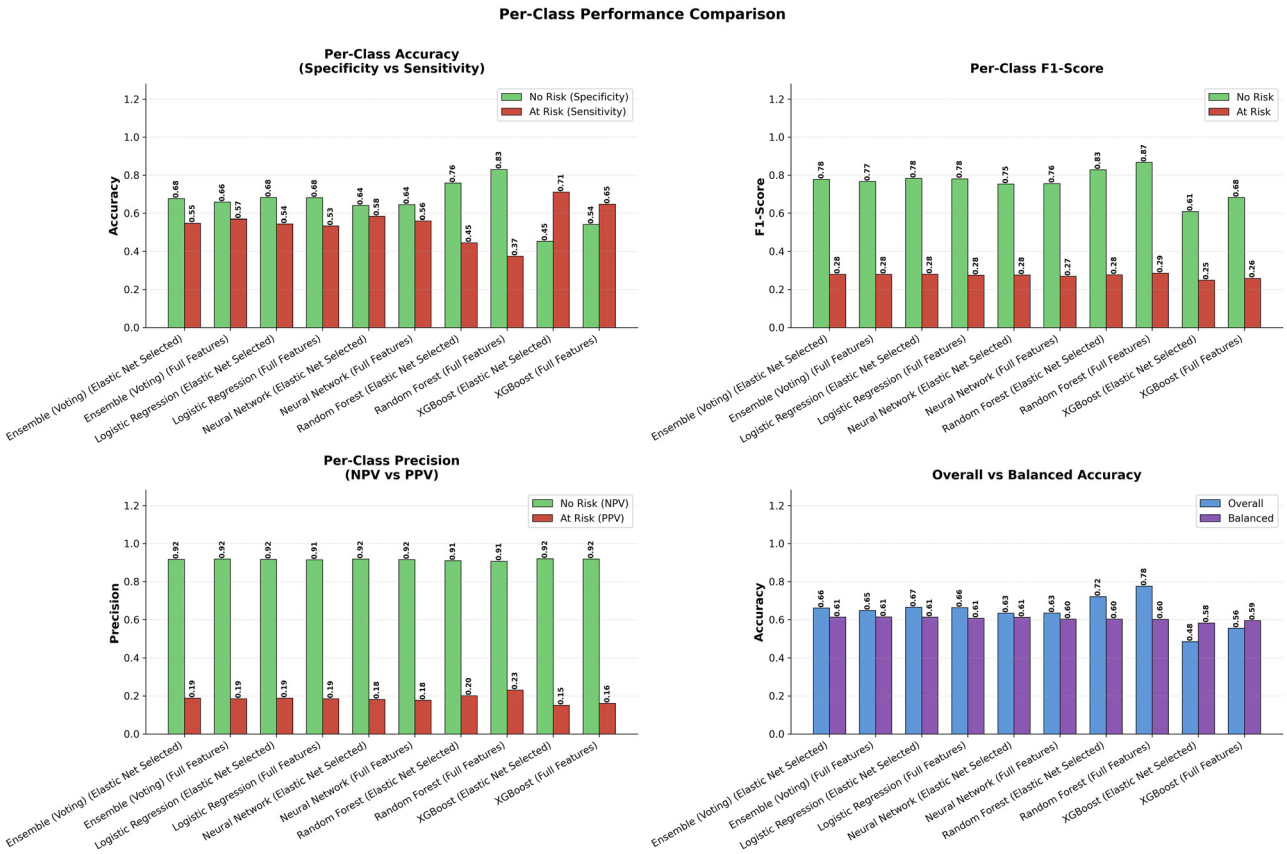

Fig. 1b. Detection rates of at-risk and no-risk individuals across models for demographic dataset

Table 1. Overall performance metrics across models and datasets

Model	Dataset	Accuracy	Balanced accuracy	AUC-ROC	Youden's *J*	Sensitivity	Specificity
Logistic Regression (EN)	Psychosocial	0.72	0.72	0.80	0.45	0.73	0.72
Logistic Regression (Full)	Psychosocial	0.72	0.72	0.79	0.45	0.73	0.72
Neural Network (EN)	Psychosocial	0.71	0.72	0.79	0.44	0.73	0.71
Neural Network (Full)	Psychosocial	0.69	0.72	0.79	0.44	0.75	0.69
Random Forest (EN)	Psychosocial	0.79	0.70	0.79	0.41	0.59	0.81
Random Forest (Full)	Psychosocial	0.78	0.70	0.80	0.40	0.57	0.84
XGBoost (EN)	Psychosocial	0.67	0.71	0.79	0.42	0.77	0.65
XGBoost (Full)	Psychosocial	0.69	0.71	0.79	0.42	0.75	0.68
Ensemble (EN)	Psychosocial	0.74	0.72	0.80	0.44	0.70	0.74
Ensemble (Full)	Psychosocial	0.75	0.72	0.80	0.44	0.68	0.76
Logistic Regression (EN)	Demographic	0.67	0.61	0.65	0.23	0.54	0.68
Logistic Regression (Full)	Demographic	0.66	0.61	0.65	0.22	0.53	0.68
Neural Network (EN)	Demographic	0.63	0.61	0.64	0.22	0.58	0.64
Neural Network (Full)	Demographic	0.64	0.60	0.65	0.21	0.56	0.65
Random Forest (EN)	Demographic	0.72	0.60	0.62	0.21	0.45	0.76
Random Forest (Full)	Demographic	0.78	0.60	0.65	0.20	0.38	0.83
XGBoost (EN)	Demographic	0.49	0.58	0.64	0.16	0.71	0.45
XGBoost (Full)	Demographic	0.56	0.60	0.64	0.19	0.65	0.54
Ensemble (EN)	Demographic	0.66	0.61	0.64	0.23	0.55	0.68
Ensemble (Full)	Demographic	0.65	0.61	0.65	0.23	0.57	0.66

Note. Values represent means across five validation folds. All models showed excellent stability with coefficients of variation <5%. EN = Elastic Net feature selection. Psychosocial dataset: n = 22,136, 24 features (10 after Elastic Net selection). Demographic dataset: n = 23,216, 22 features (8 after Elastic Net selection). Balanced Accuracy = (Sensitivity + Specificity)/2.

All models demonstrated excellent stability across validation folds, with coefficients of variation consistently below 5% for the primary metrics, indicating reliable and reproducible performance. Figures 2a and 2b present AUC-ROC comparisons, F1-scores for at-risk, balanced accuracy and the Matthews Correlation Coefficient (MCC) for all models.

Classification performance and error patterns

Table 2 presents per-class metrics revealing important trade-offs. Table 3 presents outputs from confusion matrices. The psychosocial Logistic Regression model (Elastic Net) correctly classified on average 383 ± 11 true positives and 2,815 ± 26 true negatives per fold, with 143 ± 10 false negatives (missed at-risk cases) and 1,086 ± 30 false positives (incorrect at-risk flags). This yielded an at-risk F1-score of 0.38 ± 0.01 with a precision of 26.1 ± 0.8% and a recall of 72.7 ± 1.2%.

The demographic Ensemble model correctly classified 317 ± 6 true positives and 2,690 ± 24 true negatives, with 239 ± 6 false negatives and 1,397 ± 22 false positives. This yielded an at-risk F1-score of 0.28 ± 0.005 with a precision of 18.5 ± 0.4% and a recall of 57.0 ± 1.2%. The lower precision reflects the challenge of identifying at-risk individuals from demographic data alone, while still detecting over half of the actual cases. Table 3 displays confusion matrix components for selected models.

Predictor importance and model interpretability

Table 4 presents feature importance rankings for ensemble models. In the psychosocial dataset with full feature selection, job stress dominated the importance rankings, with values of .29, substantially exceeding those of all other predictors. Job satisfaction ranked second (.10), while work hours per week emerged as the third most important feature (.09), but remained far below the psychosocial variables.

In the demographic dataset without psychosocial measures, work hours per week became the most important predictor (.19), followed by age (.09) and job tenure (.08). Income, education, gender, and company size showed moderate importance (.03–.06). The employment sector and partnership status indicators contributed minimally.

ROC analysis and discriminative ability

Figures 3a and 3b display ROC curves for all models. The psychosocial models clustered with AUC-ROC values between 0.79 and 0.80, indicating good discrimination. Demographic models showed lower but still meaningful discrimination with AUC-ROC values between 0.62 and 0.65. The ~0.14-point difference in AUC-ROC represents the predictive cost of using readily available data rather than administering additional questionnaires.

DISCUSSION

This study compared demographic-only and psychosocial approaches to work addiction screening, with primary focus on detection rates of at-risk individuals and practical implementation feasibility. While psychosocial assessment achieved superior overall performance, demographic screening offers a meaningful alternative that can identify substantial proportions of at-risk cases without an additional data-collection burden.

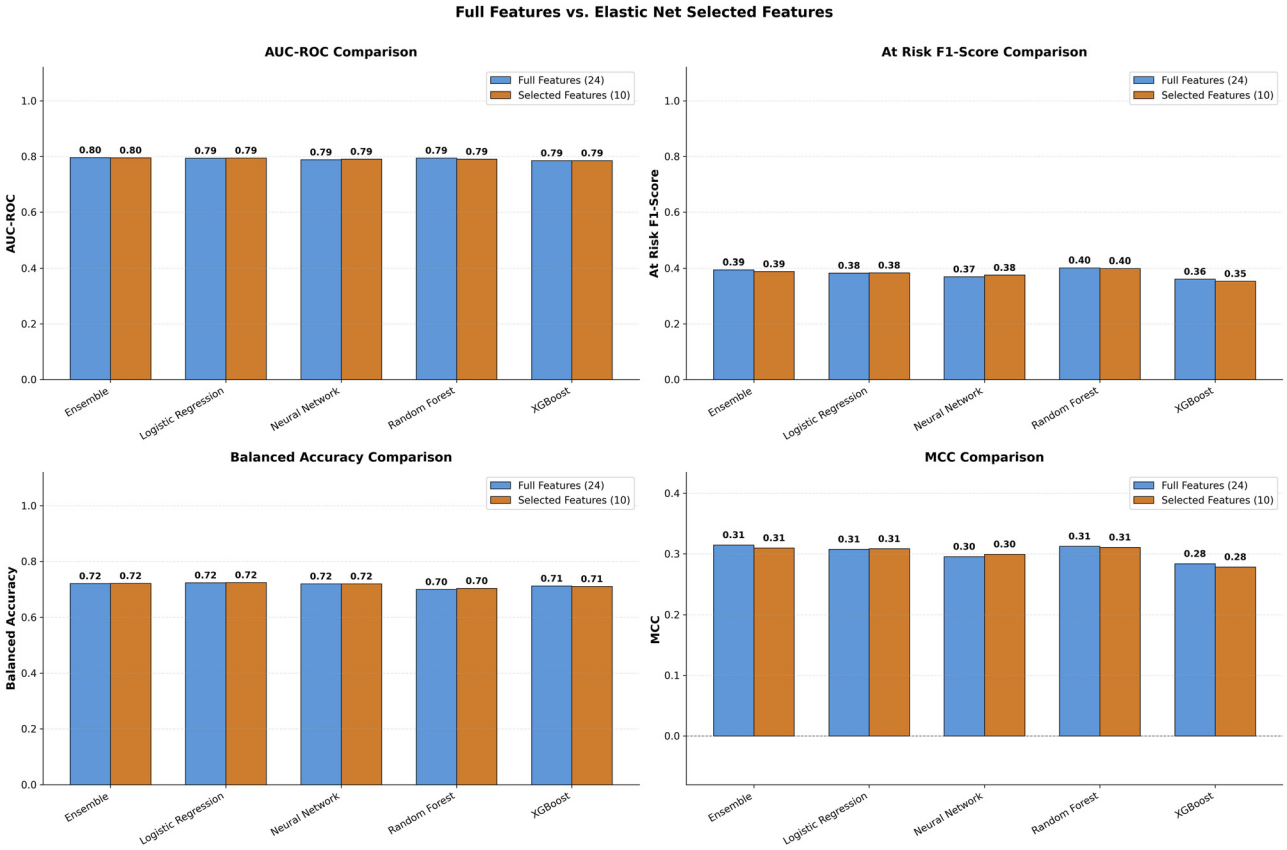

Fig. 2a. Performance comparison across key metrics. Multi-panel visualization showing AUC-ROC, F1-score, balanced accuracy, and Matthews correlation coefficient (MCC) for all models for the psychosocial dataset

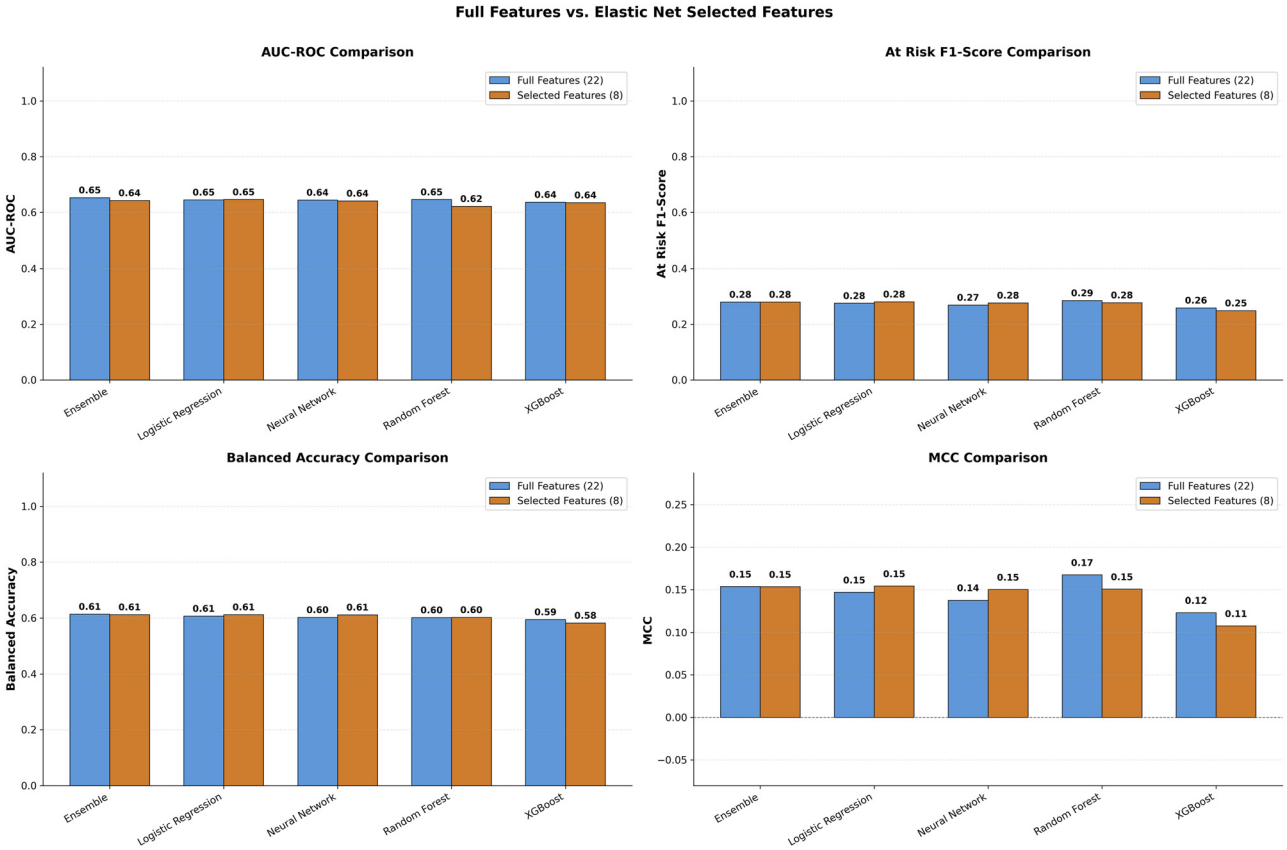

Fig. 2b. Performance comparison across key metrics. Multi-panel visualization showing AUC-ROC, F1-score, balanced accuracy, and Matthews Correlation Coefficient (MCC) for all models for the demographic dataset

Table 2. Per-class performance metrics for all models

Model	Dataset	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	No-Risk F1-Score	At-Risk F1-Score
Logistic Regression (EN)	Psychosocial	72.7	72.2	26.1	95.1	0.82	0.38
Logistic Regression (Full)	Psychosocial	72.5	72.2	26.0	95.1	0.82	0.38
Neural Network (EN)	Psychosocial	73.4	70.6	25.1	95.2	0.81	0.37
Neural Network (Full)	Psychosocial	75.3	68.6	24.5	95.4	0.80	0.37
Random Forest (EN)	Psychosocial	59.2	81.4	30.1	93.5	0.87	0.41
Random Forest (Full)	Psychosocial	57.0	82.9	31.0	93.5	0.88	0.40
XGBoost (Elastic Net)	Psychosocial	77.1	65.0	23.3	95.5	0.78	0.36
XGBoost (Full)	Psychosocial	74.6	67.8	24.2	95.3	0.79	0.37
Ensemble (EN)	Psychosocial	69.8	74.4	27.5	94.8	0.84	0.39
Ensemble (Full)	Psychosocial	68.0	76.1	27.8	94.6	0.84	0.39
Logistic Regression (EN)	Demographic	54.4	68.2	18.9	91.7	0.78	0.28
Logistic Regression (Full)	Demographic	53.4	68.1	18.7	91.6	0.78	0.27
Neural Network (EN)	Demographic	58.4	64.0	18.2	91.9	0.75	0.28
Neural Network (Full)	Demographic	56.0	64.5	18.0	91.7	0.76	0.27
Random Forest (EN)	Demographic	44.6	75.9	21.1	90.9	0.85	0.28
Random Forest (Full)	Demographic	37.5	83.0	23.0	90.7	0.87	0.29
XGBoost (EN)	Demographic	71.1	45.4	14.6	92.7	0.60	0.24
XGBoost (Full)	Demographic	64.7	54.3	16.1	91.9	0.67	0.26
Ensemble (EN)	Demographic	54.9	67.6	18.8	91.7	0.78	0.28
Ensemble (Full)	Demographic	57.0	65.8	18.5	91.8	0.77	0.28

Note. Values represent means across five folds. Sensitivity = recall for at-risk class (true positive rate). Specificity = recall for no-risk class (true negative rate). PPV = Positive Predictive Value (precision for at-risk). NPV = Negative Predictive Value (precision for no-risk). F1-Score = harmonic mean of precision and recall. EN = Elastic Net feature selection.

Table 3. Confusion matrix components for all models

Model	Dataset	TN	FP	FN	TP
Logistic Regression (EN)	Psychosocial	2,815	1,086	143	383
Logistic Regression (Full)	Psychosocial	2,817	1,083	145	381
Neural Network (EN)	Psychosocial	2,754	1,147	140	386
Neural Network (Full)	Psychosocial	2,674	1,226	129	396
Random Forest (EN)	Psychosocial	3,175	725	214	312
Random Forest (Full)	Psychosocial	3,233	667	226	299
XGBoost (EN)	Psychosocial	2,535	1,366	120	406
XGBoost (Full)	Psychosocial	2,644	1,257	133	392
Ensemble (EN)	Psychosocial	2,900	1,000	159	367
Ensemble (Full)	Psychosocial	2,969	931	168	357
Logistic Regression (EN)	Demographic	2,786	1,300	253	302
Logistic Regression (Full)	Demographic	2,785	1,302	259	297
Neural Network (EN)	Demographic	2,615	1,472	231	324
Neural Network (Full)	Demographic	2,635	1,452	245	311
Random Forest (EN)	Demographic	3,101	987	308	248
Random Forest (Full)	Demographic	3,391	696	348	208
XGBoost (EN)	Demographic	1,857	2,230	161	395
XGBoost (Full)	Demographic	2,219	1,868	196	360
Ensemble (EN)	Demographic	2,762	1,325	251	305
Ensemble (Full)	Demographic	2,690	1,397	239	317

Note. Values represent averages per fold. TN = True Negatives (correctly identified no-risk), FP = False Positives (no-risk incorrectly flagged as at-risk), FN = False Negatives (at-risk cases missed), TP = True Positives (correctly identified at-risk). EN = Elastic Net feature selection.

The central finding concerns the trade-off between predictive performance and practical feasibility. The best demographic model correctly identified 57.0% of at-risk individuals using only readily available human resources data. Lower classification thresholds detected up to 71.1% of at-risk cases, though at the cost of higher false-positive rates. For organizations unable or unwilling to administer additional questionnaires, these detection rates represent meaningful screening capability.

The practical implications depend critically on organizational context and screening objectives. In settings where

Table 4. Feature importance rankings from ensemble models: comparison of full and elastic net selected feature sets

Feature	Psychosocial (Full features)	Psychosocial (Elastic net)	Demographic (Full features)	Demographic (Elastic net)
Work Hours per Week	.09	.13	.19	.42
Job Stress	.29	.39	–	–
Education Level	.03	–	.05	.14
Job Satisfaction	.10	.13	–	–
Employment at Current Job	.05	.09	.08	–
Age	.05	–	.09	–
Company Size	.02	–	.04	.09
Gender (Female)	.02	.04	.04	.08
Household Size	.04	.07	.06	–
Managerial Position (Junior)	.02	.04	.03	.07
Managerial Position (Middle)	.02	.04	–	.07
Managerial Position (Senior)	–	.03	–	.07
Marital Status (Single)	.02	.04	–	.07
Income	.04	–	.06	–
Number of Children	.03	–	.05	–

Note. Feature importance values derived from tree-based ensemble components (Random Forest and XGBoost), averaging Gini impurity and gain importance scores. Full features include all available variables; Elastic Net selection retained 10 of 24 features for psychosocial dataset and 8 of 22 for demographic dataset. Missing values (–) indicate features not included in that specific model configuration. Values reflect relative contribution to model predictions and do not necessarily sum to 1.0 when considering only displayed features (additional features with lower importance not shown).

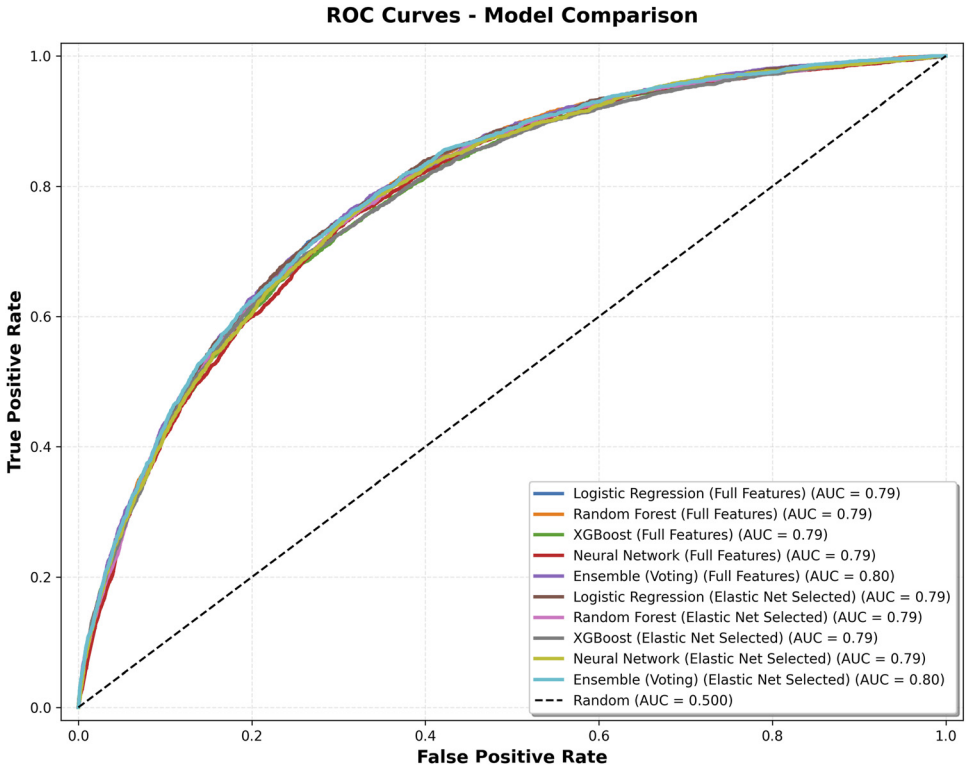

Fig. 3a. ROC curves comparing discriminative ability for psychosocial dataset

work addiction prevalence is moderate to high, with meta-analytic estimates suggesting approximately 14% prevalence in working populations (Andersen et al., 2023) and higher rates among self-employed individuals and managers (Atroszko & Atroszko, 2020; Moskalewicz et al., 2019), detecting 55–75% of cases may justify implementation despite imperfect accuracy. The demographic approach is particularly well-suited for large-scale preliminary screening, where identified individuals can undergo a more comprehensive assessment. Organizations could implement demographic screening immediately using existing data systems, identifying individuals for targeted follow-up rather than administering universal questionnaires.

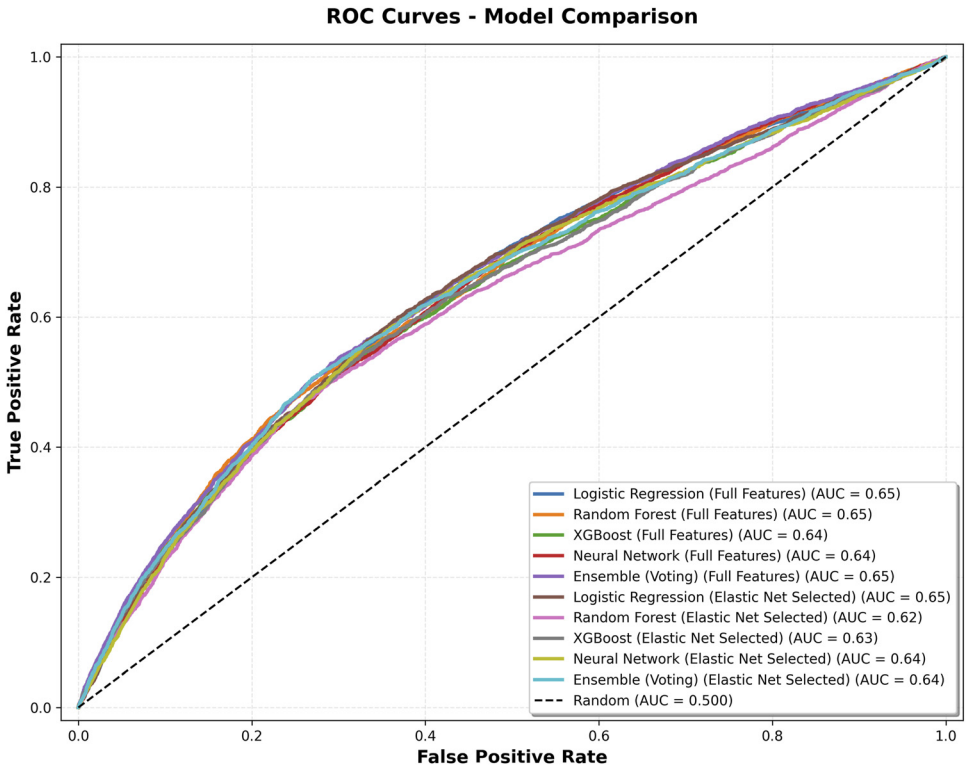

Fig. 3b. ROC curves comparing discriminative ability for demographic dataset

The psychosocial approach's superior performance came primarily from job stress and job satisfaction measures. These variables dominated the importance rankings, with job stress showing 2–3 times the importance of any demographic predictor. This indicates that work environment-related characteristics are more proximally associated with work addiction than demographic attributes. However, this theoretical insight does not negate the practical value of demographic screening in resource-constrained settings. Job stress and job satisfaction are also independent indicators of occupational well-being. For organizations already conducting wellness surveys, the psychosocial model can be applied at no additional cost. However, when resources permit, directly administering the IWAS-5 remains the most accurate identification approach. We propose a tiered framework: (1) demographic or psychosocial screening as a zero-burden first stage using existing HR records; (2) targeted direct IWAS-5 administration as the second-stage validation assessment for flagged individuals; (3) potential in-depth interviews and more comprehensive psychosocial assessment as a final confirmatory stage. This multi-stage approach progressively increases identification accuracy while managing data collection burden.

Two important factors and corresponding implementation approaches should be considered. First, work addiction is a sensitive issue, as it is not only an organizational phenomenon but also closely associated with mental health (Andreassen, 2014; Atroszko, 2022; Charzyńska et al., 2025). Consequently, screening, prevention, and intervention efforts must comply with legal and ethical regulations, ensuring employee safety, confidentiality, and voluntary participation. Second, due to these ethical and confidentiality considerations, organizations may prefer to conduct aggregate-level screening to estimate overall risk across the organization or within specific departments. Such an approach can inform the implementation of universal preventive measures, such as structural changes to reduce excessive demands and increase support, training programs (particularly for managers), awareness initiatives, and access to psychological support services, rather than targeted individual-level interventions (Cossin, Thaon, & Lalanne, 2021).

Among demographic variables, work hours per week emerged as the strongest predictor, consistent with work addiction's defining characteristic of excessive time investment. Age and job tenure showed moderate predictive value, possibly reflecting accumulated exposure to work-addictive patterns or cohort effects in work attitudes. The relatively weak contributions of income, education, and the employment sector suggest that work addiction transcends socioeconomic boundaries.

The sensitivity-specificity trade-off presents important decision points for screening program design. High-sensitivity models (75% detection) yield substantial false-positive rates (30–60%), meaning many individuals flagged as at risk are not actually at risk. Organizations must weigh the costs of over-identification against the costs of missed cases. In contexts where intervention is low-cost and non-stigmatizing, higher false positive rates may be acceptable. Where the consequences of positive identification are significant, more conservative thresholds become appropriate.

It is worth making explicit why organizations should invest in work addiction screening, as this may not be obvious to practitioners who associate long working hours with productivity. Work addiction is linked to a range of adverse outcomes, including burnout, turnover intention, counterproductive work behavior, impaired team dynamics, and deteriorating physical and mental health (Atroszko, 2022; Clark et al., 2016; Kenyhercz et al., 2024). These carry direct economic costs through productivity losses, recruitment, and occupational health. Importantly, work addiction's resemblance to high engagement means it is often overlooked or even rewarded by managers, making structured screening particularly valuable. The costs of employee turnover are particularly substantial for highly skilled professionals and managers, whose development requires significant organizational investment. When such employees leave due to burnout or health-related issues associated with excessive work, organizations incur not only financial losses but also disruptions to continuity, expertise, and team functioning. This underscores the importance of fostering sustainable work environments that support long-term employee well-being and performance, rather than relying on implicit assumptions of employee replaceability, which may yield short-term gains but ultimately undermine organizational stability and resilience, knowledge retention and expertise, and overall effectiveness (Martinez et al., 2025). This is particularly important in sectors such as health care, education, information technology, and professional services, where work relies heavily on specialized expertise, continuity of knowledge, and sustained cognitive and emotional engagement (Shanafelt, Goh, & Sinsky, 2017).

Limitations

Several limitations warrant consideration. First, the cross-sectional design precludes causal inference about the relationships between demographic characteristics and the development of work addiction. Second, reliance on self-report measures introduces potential common method variance, though the substantial performance difference between psychosocial and demographic datasets argues against this fully explaining our results. Third, the brief measures used for job stress and job satisfaction lack the nuance of multi-dimensional scales. Single-item assessments may miss important facets, such as work-life conflict, job control, social support, and intrinsic motivation, that could enhance predictive validity.

Fourth, even with psychosocial variables, our models achieved moderate but not excellent discrimination (AUC-ROC ~0.80), suggesting important unmeasured predictors. Personality characteristics (e.g., perfectionism, conscientiousness), work values, organizational culture, and supervisor behaviors are likely to contribute additional predictive value beyond the variables examined here.

Fifth, while five-fold cross-validation provides robust internal validation, external validation on independent organizational samples from different industries and cultural contexts would strengthen generalizability conclusions. The international nature of our sample (85 countries or territories) enhances breadth but does not replace true external validation.

Sixth, the Elastic Net feature selection was performed on the full dataset prior to cross-validation, introducing a degree of optimistic bias in the reported performance estimates. Future studies should employ nested cross-validation for fully unbiased evaluation. Seventh, model hyperparameters were fixed rather than systematically tuned, which represents a limitation of the current study. Systematic hyperparameter optimization (e.g., via grid search within nested cross-validation) represents an important avenue for improving performance in applied deployments.

It should be noted that both models show relatively low positive predictive values (psychosocial PPV ~26%; demographic PPV ~19%), which reflects the inherent statistical challenge of screening for a condition with ~12–14% prevalence rather than a limitation of the models themselves. Positive screening results should therefore be interpreted as indicators for priority follow-up assessment. A tiered approach, combining demographic screening with targeted psychosocial assessment and, where indicated, direct administration of the IWAS-5, may help improve overall screening accuracy while maintaining the practical advantages of low-burden initial identification.

Future directions

Future research should address several key questions. Longitudinal studies could examine whether demographic screening identifies individuals who subsequently develop work addiction, validating predictive rather than concurrent validity. Implementation research could evaluate demographic screening in organizational settings, assessing feasibility, acceptability, and impact on identification rates. Cost-effectiveness analyses could compare demographic versus psychosocial screening strategies, accounting for administration costs, intervention costs, and outcomes.

Research combining demographic screening with targeted psychosocial assessment could optimize resource allocation. For example, demographic models might identify high-risk subgroups for comprehensive evaluation while bypassing low-risk groups. Multi-stage screening approaches deserve systematic evaluation. Studies incorporating additional predictor domains (personality, organizational factors, work-life balance) could improve prediction beyond current models. Future work should also test administrative HR variables not available in the present dataset, including leave utilization patterns (e.g., unused annual leave), after-hours system access logs (e.g., email/IT activity outside working hours), and type of employment contract (flexible vs. fixed hours). These could improve demographic model performance without any self-report burden. Additionally, systematic hyperparameter optimization (nested cross-validation with grid search) should be evaluated in future applied deployments.

Conclusions

In conclusion, demographic-only screening represents a practical alternative to comprehensive psychosocial

assessment for work addiction risk identification. Although psychosocial models incorporating job stress and job satisfaction measures substantially outperformed demographic approaches (AUC-ROC 0.79–0.80 vs 0.62–0.65, sensitivity 57–77% vs 38–71%), demographic-only models can still identify the majority of at-risk individuals using readily available human resources data. This trade-off between performance and feasibility merits serious consideration for organizational screening programs. The choice between approaches should consider organizational resources, screening objectives, intervention capacities, and acceptable error rates. For many organizations, demographic screening may provide sufficient detection capability to justify its implementation, particularly as a preliminary screening preceding a targeted comprehensive assessment.

Funding sources: This work was supported by the National Science Centre, Poland [grant number 2020/39/D/HS6/00198]. Data collection in Armenia was supported by the Science Committee of the Republic of Armenia, in the frames of the research project № 25RG-5A025. Data collection in the Czech Republic was supported by financing from NPO "Systemic Risk Institute" no. LX22NPO5101, funded by European Union - Next Generation EU (Ministry of Education, Youth and Sports, NPO: EXCELES). Data collection in Estonia was supported by the Estonian Research Council grant (PRG2190). Data collection in Hungary was supported by the Hungarian National Research, Development, and Innovation Office (Grant number: FK134807) and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, granted to Bernadette Kun.

Authors' contributions: NAW-H: conceptualization; data curation; formal analysis; investigation; methodology; resources; visualisation; writing - original draft; writing – review & editing; EC: conceptualization; data curation; funding acquisition; investigation; methodology; project administration; resources; supervision; writing - original draft; writing - review & editing; AB: conceptualization; investigation; methodology; resources; writing – review & editing; SKC: conceptualization; data curation; investigation; methodology; resources; writing – review & editing; ZS: conceptualization; investigation; methodology; resources; writing – review & editing; PAA: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; writing - original draft; writing - review & editing.

Conflict of interest: The authors declare no conflict of interest.

Data availability: The dataset and related materials for the study, including the Python code used for analyses, have been deposited in the Zenodo repository at https://doi.org/10.5281/zenodo.20714352. The study was part of a preregistered project, details of which are available on OSF Registries [https://osf.io/8asnm].

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process: During the preparation of this manuscript, the authors used Claude.ai to support grammar adjustments, to assist with literature search and synthesis, and to provide technical support during coding of machine learning models. After using this tool, the authors carefully reviewed and edited all content and take full responsibility for the accuracy, originality, and integrity of the published work.

Acknowledgment: The authors are deeply grateful to the members of the team involved in the Global Research on Work Addiction for their contributions to data collection and their overall involvement in the project. A list of international collaborators and their bios is available on the project's website at https://workaddiction.org/team.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. *arXiv*. 10.48550/arXiv.1603.04467 https://doi.org/10.48550/arXiv.1603.04467

Andersen, F. B., Djugum, M. E. T., Sjåstad, V. S., & Pallesen, S. (2023). The prevalence of workaholism: A systematic review and meta-analysis. *Frontiers in Psychology*, *14*, 1252373. https://doi.org/10.3389/fpsyg.2023.1252373

Andreassen, C. S. (2014). Workaholism: An overview and current status of the research. *Journal of Behavioral Addictions*, *3*(1), 1–11. https://doi.org/10.1556/jba.2.2013.017

Atroszko, P. A. (2022). Non-drug addiction: Addiction to work. In *Handbook of substance misuse and addictions: From biology to public health* (pp. 2981–3012). Springer International Publishing. https://doi.org/10.1007/978-3-030-67928-6_3-1

Atroszko, P. A., & Atroszko, B. (2020). The costs of work-addicted managers in organizations: Towards integrating clinical and organizational frameworks. *Amfiteatru Economic*, *22*(14), 1265–1282. https://doi.org/10.24818/EA/2020/S14/1265

Atroszko, P. A., Demetrovics, Z., & Griffiths, M. D. (2019). Beyond the myths about work addiction: Toward a consensus on definition and trajectories for future studies on problematic overworking. *Journal of Behavioral Addictions*, *8*, 7–15. https://doi.org/10.1556/2006.8.2019.11

Atroszko, P. A., Kun, B., Buźniak, A., Czerwiński, S. K., Schneider, Z., Woropay-Hordziejewicz, N., … Charzyńska, E. (2025). Perceived coworkers' work addiction: Scale development and associations with one's own workaholism, job stress, and job satisfaction in 85 cultures. *Journal of Behavioral Addictions*, *14*(1), 246–262. https://doi.org/10.1556/2006.2025.00011

Bereznowski, P., Konarski, R., Pallesen, S., & Atroszko, P. A. (2025). Similarities and differences between study addiction and study engagement and work addiction and work engagement: A network analysis. *International Journal of Mental Health and Addiction*, *23*(3), 2380–2401. https://doi.org/10.1007/s11469-023-01234-4

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

Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution.

In *Proceedings of the 20th international conference on pattern recognition* (pp. 3121–3124). IEEE. https://doi.org/10.1109/ICPR.2010.764

Charzyńska, E., Buźniak, A., Czerwiński, S. K., Woropay-Hordziejewicz, N., Schneider, Z., Aavik, T., … Atroszko, P. A. (2025). The International Work Addiction Scale (IWAS): A screening tool for clinical and organizational applications validated in 85 cultures from six continents. *Journal of Behavioral Addictions*, *14*(1), 220–245. https://doi.org/10.1556/2006.2025.00005

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Chen, C., Liaw, A., & Breiman, L. (2004). *Using random forest to learn imbalanced data*. Vol. *110*. (pp. 1–12). Berkeley: University of California. https://statistics.berkeley.edu/tech-reports/666.

Clark, M. A., Michel, J. S., Zhdanova, L., Pui, S. Y., & Baltes, B. B. (2016). All work and no play? A meta-analytic examination of the correlates and outcomes of workaholism. *Journal of Management*, *42*(7), 1836–1873. https://doi.org/10.1177/0149206314522301

Cossin, T., Thaon, I., & Lalanne, L. (2021). Workaholism prevention in occupational medicine: A systematic review. *International Journal of Environmental Research and Public Health*, *18*(13), 7109. https://doi.org/10.3390/ijerph18137109

Defazio, A., Bach, F., & Lacoste-Julien, S. (2014). SAGA: A fast incremental gradient method with support for nonstrongly convex composite objectives. In *Advances in neural information processing systems* (pp. 1646–1654).

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems. MCS 2000. Lecture notes in computer science*. Vol. *1857*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-45014-9_1

Dolbier, C. L., Webster, J. A., McCalister, K. T., Mallon, M. W., & Steinhardt, M. A. (2005). Reliability and validity of a single-item measure of job satisfaction. *American Journal of Health Promotion*, *19*(3), 194–198. https://doi.org/10.4278/0890-1171-19.3.194

Eurofound. (2017). *Working time patterns for sustainable work*. Publications Office of the European Union. https://www.eurofound.europa.eu/en/publications/all/working-time-patterns-sustainable-work.

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, *45*(2), 171–186. https://doi.org/10.1023/A:1010920819831

Houdmont, J., Randall, R., Kinman, G., Colwell, J., Kerr, R., & Addley, K. (2021). Can a single-item measure of job stressfulness identify common mental disorder? *International Journal of Stress Management*, *28*(4), 305–313. https://doi.org/10.1037/str0000231

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd international conference on machine learning, in proceedings of machine learning research*. Vol. *37*. (pp. 448–456). Available from: https://proceedings.mlr.press/v37/ioffe15.html.

Kenyhercz, V., Mervó, B., Lehel, N., Demetrovics, Z., & Kun, B. (2024). Work addiction and social functioning: A systematic review and five meta-analyses. *PLoS One*, *19*(6), e0303563. https://doi.org/10.1371/journal.pone.0303563

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*. https://doi.org/10.48550/arXiv.1412.6980

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on artificial intelligence* (pp. 1137–1143).

Kun, B., Takacs, Z. K., Richman, M. J., Griffiths, M. D., & Demetrovics, Z. (2020). Work addiction and personality: A meta-analytic study. *Journal of Behavioral Addictions*, *9*(4), 945–966. https://doi.org/10.1556/2006.2020.00097

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2999–3007). https://doi.org/10.1109/ICCV.2017.324

Martinez, M. F., O'Shea, K. J., Kern, M. C., Chin, K. L., Dinh, J. V., Bartsch, S. M., ... Lee, B. Y. (2025). The health and economic burden of employee burnout to US employers. *American Journal of Preventive Medicine*, *68*(4), 645–655. https://doi.org/10.1016/j.amepre.2025.01.011

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) — Protein Structure*, *405*(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Moskalewicz, J., Badora, B., Feliksiak, M., Głowacki, A., Gwiazda, M., Herrmann, M., … Roguska, B. (2019). *Oszacowanie rozpowszechnienia oraz identyfikacja czynników ryzyka i czynników chroniących hazardu i innych uzależnień behawioralnych – edycja 2018/2019: Raport z badań [Estimating the prevalence and identifying risk and protective factors of gambling and other behavioural addictions – 2018/2019 edition: Research report]*. Fundacja Centrum Badania Opinii Społecznej.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Shanafelt, T., Goh, J., & Sinsky, C. (2017). The business case for investing in physician well-being. *JAMA Internal Medicine*, *177*(12), 1826–1832. https://doi.org/10.1001/jamainternmed.2017.4340

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(1), 1929–1958.

Taris, T. W., & de Jonge, J. (2024). Workaholism: Taking stock and looking forward. *Annual Review of Organizational Psychology and Organizational Behavior*, *11*, 113–138. https://doi.org/10.1146/annurev-orgpsych-111821-035514

World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Appendix

Table A1. Demographic and work-related characteristics of participants by dataset

Characteristic	Psychosocial Dataset (n = 22,136)	Demographic Dataset (n = 23,219)
Work Addiction Risk, *n* (%)		
No risk	19,504 (88.1%)	20,437 (88.0%)
At risk	2,632 (11.9%)	2,782 (12.0%)
Gender, *n* (%)		
Female	14,046 (63.5%)	14,739 (63.5%)
Male	7,987 (36.1%)	8,369 (36.0%)
Other	103 (0.005%)	111 (0.005%)
Marital Status, *n* (%)		
Married	10,347 (46.7%)	10,855 (46.8%)
Single	5,985 (27.0%)	6,294 (27.1%)
Informal living together	3,026 (13.7%)	3,140 (13.5%)
Informal living apart	1,356 (6.1%)	1,434 (6.2%)
Divorced	947 (4.3%)	1,003 (4.3%)
Separated	253 (1.1%)	265 (1.1%)
Widowed	206 (0.9%)	211 (0.9%)
Managerial Position, *n* (%)		
Non-managerial	13,291 (60.0%)	13,930 (60.0%)
Lower management	3,224 (14.6%)	3,364 (14.5%)
Middle management	3,745 (16.9%)	3,928 (16.9%)
Senior management	1,876 (8.5%)	1,997 (8.6%)
Employment Sector, *n* (%)		
Private	11,705 (52.9%)	12,297 (53.0%)
Public	10,431 (47.1%)	10,922 (47.0%)
Organization Size, *n* (%)		
Large (≥250 employees)	10,500 (47.4%)	10,997 (47.4%)
Medium (50–249 employees)	5,826 (26.3%)	6,098 (26.3%)
Small (10–49 employees)	5,810 (26.2%)	6,124 (26.4%)
Age, years, M ± SD	39.85 ± 11.21	39.78 ± 11.18
Weekly Work Hours, M ± SD	43.60 ± 8.84	43.64 ± 8.89
Organizational Tenure, years, M ± SD	8.77 ± 8.40	8.72 ± 8.38
Education Level, M ± SD	4.76 ± 1.18	4.76 ± 1.18
Household Size, M ± SD	3.21 ± 1.80	3.22 ± 1.81
Number of Children, M ± SD	1.12 ± 1.29	1.12 ± 1.29
Job Stress (1–7 scale), M ± SD	3.48 ± 1.44	–
Job Satisfaction (1–7 scale), M ± SD	3.60 ± 1.53	–

Note. Values are presented as *n* (%) for categorical variables and M ± SD for continuous variables. Education Level measured on ordinal scale with higher values indicating greater educational attainment; specific labels varied by country due to diverse educational systems. Income measured on a country-adjusted 8-point scale with the middle category representing median national earnings. Minor discrepancies in sample sizes (psychosocial n = 22,136; demographic n = 23,219) reflect different data availability after preprocessing and handling missing values.