Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Development and validation of a machine learning-based predictive model for work readiness in Chinese vocational nursing students

**Xixi Du**
Shangqiu Medical college
**Lijuan Zhou**
Shangqiu Medical college
**Zhen Shi**
Inner Mongolia Medical University
**Dongyang Wang**
The Third People's Hospital of Henan Province
**Guozeng Zhang**
zhanggz@henu.edu.cn

Henan University

**Research Article**

**Keywords:** Work readiness, Machine learning, SHAP, predictive model, Vocational nursing students

**Posted Date:** September 2nd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10424750/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

## Abstract

# Background

Global nursing workforce shortages prevail worldwide, including in China. Vocational nursing students, accounting for over 60% of national nursing enrollments, constitute the main frontline care workforce. However, they commonly suffer from insufficient work readiness, leading to employment difficulties and poor post-employment adaptation. Early targeted interventions can improve employment quality and alleviate nursing workforce shortages. However, no validated tools are available for early screening of poor work readiness among these students.

# Methods

Cross-sectional survey study using multicenter data. A total of 854 vocational nursing students from four colleges in Henan Province completed standardized questionnaires including the Work Readiness Scale for Graduate Nurses (WRS-GN-C), Nurse Professional Identity Scale (NPIS), and Future Work Self-Clarity Scale (FWSS). Seven regression algorithms were compared for continuous WRS-GN score prediction, and seven classification algorithms for median-split binary classification. Feature importance was assessed using standardized coefficients for linear models and SHAP (TreeExplainer) for tree-based models.

# Results

The Elastic Net regression model achieved the highest predictive performance ($R^2$ = 0.705, RMSE = 28.14, MAE = 21.83), with excellent generalizability (CV/test = 0.952). Future Work Self-Clarity emerged as the dominant predictor (standardized |β| = 14.51), followed by Skills Confidence (|β| = 7.76) and Professional Identity's Social Support subscale (|β| = 6.91). In the sensitivity analysis, SVM achieved the highest AUC-ROC (0.894) and Random Forest demonstrated the highest recall (83.6%) for identifying at-risk students.

# Conclusions

Work readiness among vocational nursing students is highly predictable from a compact set of modifiable psychological factors. Future Work Self-Clarity represents the most actionable intervention target. The predictive models provide a methodological foundation for precision nursing education and early identification of at-risk students.

## 1. Introduction

Global healthcare systems are facing a severe workforce shortage of 5.8 million nurses, with population aging further exacerbating this crisis[1]. As the most populous country, China suffers from a prominent nursing workforce shortage. By the end of 2025, China's nurse density reached 4.32 per 1,000 population, which is markedly lower than that of developed countries such as Australia (12.6), Japan (11.2), and Canada (9.1) [2–3]. Beyond this quantitative shortage, the structural attributes of vocational nursing education represent a shared yet diversified challenge across global health systems. Internationally, established models, such as Australia's Technical and Further Education (TAFE) system, which prioritizes competency-based and industry-led training, and Germany's dual-system (Duales System), which integrates enterprise-based apprenticeships with part-time vocational schooling, have long served as primary pipelines for frontline nursing workforce preparation. Despite their pedagogical differences, these systems commonly confront a core tension: maximizing graduates' clinical adaptability while accommodating heterogeneous academic entry levels.

In China, vocational nursing students account for over 60% of the total nursing student population[4]. Compared with undergraduate nursing students, they undergo a three-year skills-oriented curriculum and consequently possess relatively weaker theoretical foundations. Moreover, unlike their Australian and German counterparts, who benefit from well-established industry-embedded career pathways, Chinese vocational nursing students face an additional structural bottleneck: the escalating credential inflation mandated by hospitals, which markedly intensifies their employment competition and post-entry psychological adaptation difficulties.

This prevailing workforce strain, combined with the structural constraints inherent in China's vocational nursing education, ultimately converges on a fundamental question: whether graduates possess sufficient work readiness to meet the demands of real-world clinical practice. Accumulating evidence demonstrates that work readiness exerts a direct impact on nursing students' employability, and sufficient pre-employment preparation enables graduates to better adapt to real-world clinical practice [5–6].

Work readiness is defined as the extent to which individuals master the knowledge and skills required for their chosen profession and their adaptability to future work [7]. Graduates with high work readiness can achieve a smooth transition from student to clinical nurse, cope with workplace challenges, and sustain professional development, thereby helping stabilize the nursing workforce and alleviate the global nursing shortage [8–10].

Given its practical value, vocational nursing students' work readiness has attracted growing research attention. Existing studies have identified a range of demographic, educational, psychological, environmental, and professional identity-related influencing factors, predominantly through conventional regression approaches [9, 11, 12]. These foundational studies have generated valuable insights into the correlates of work readiness. However, these studies typically operate under linearity assumptions and within single analytical frameworks. More importantly, there remains a lack of theoretically grounded, interpretable individualized prediction tools that can translate predictive factors into actionable risk

stratification strategies for clinical training and school education to facilitate the development and implementation of targeted intervention protocols.

To address these methodological and practical gaps, this study adopts Social Cognitive Career Theory (SCCT) as the theoretical framework and employs a comparative machine learning (ML) pipeline. Multiple linear and non-linear algorithms are systematically evaluated via rigorous cross-validation to construct optimal prediction models, aiming to inform precision educational practice.

The SCCT framework synthesizes occupational self-perception, clinical self-efficacy, environmental support, and adaptive coping resources into an integrated theoretical system, enabling scientific variable screening and credible interpretation of modifiable psychological predictors of work readiness. Machine learning has become a prevalent tool for medical predictive research, overcoming the limitations of conventional statistical analysis. Multi-algorithm comparison supports accurate individual-level prediction of work readiness[13, 14]. Furthermore, explainable artificial intelligence (XAI) represented by SHAP resolves the “black-box” defect of traditional ML models by quantifying variable contributions and enabling comprehensive global and local interpretation [15, 16]. Such interpretable analytical outputs can be further transformed into actionable decision-support tools for nursing education.

Accordingly, this study aims to develop theory-driven, interpretable ML models to predict work readiness among vocational nursing students. Both continuous and binary predictive models are established to compare the applicability of different modeling strategies in precision vocational nursing education. Integrating native model-based feature importance with the SHAP interpretability framework, this study quantifies and hierarchically ranks the contribution weights of all modifiable psychological predictors. The findings are expected to assist nursing educators in precisely identifying students at high risk of insufficient work readiness and implementing individualized, stratified educational interventions to improve graduate clinical competency and workforce sustainability.

## 2. Methods

## 2.1 Study design and participants

This study adopted a multicenter cross-sectional survey design. Data were collected from vocational nursing colleges across Henan Province, China, from May to August 2025. A stratified cluster sampling method was utilized to ensure representative geographic coverage and institutional diversity.

The inclusion criteria were as follows: (1) full-time students enrolled in a three-year vocational nursing diploma program; (2) students in their final semester of study; (3) completion of no less than 80% of the required clinical practicum hours; and (4) voluntary participation with signed informed consent. The exclusion criteria included: (1) students with interrupted studies or academic probation; (2) transfer students with incomplete academic training records; and (3) students engaged in off-campus employment during the study period. Prior to data collection, ethical approval was granted by the

Institutional Review Board. All participants provided written informed consent for inclusion before taking part in the study.

A total of 959 students participated in the survey. Strict data quality control was performed to screen for careless responses. Specifically, 105 questionnaires (10.9%) with zero standard deviation across all items (straight-line responding) were identified and excluded in accordance with established data cleaning protocols[17] (Curran, 2016), resulting in a final analytical sample of 854 participants.

The sample size estimation was based on the EPV rule for predictive models, with a minimum of 10 to 20 participants per predictor variable[18] (Riley et al., 2020). This study included 20 candidate predictors, including demographic indicators and scale subscale scores. Given that the 7:3 train-test split and 10-fold cross-validation would consume part of the sample, a stricter criterion of 40–50 participants per predictor was adopted, with a target sample size of 800–1000 students. The final sample size of 854 fully satisfied this standard, supporting stable model training and reducing overfitting.

# 2.2 Measurement Instruments

**General Information Questionnaire**. This self-design questionnaire was developed based on a systematic literature review and tailored to specific research objectives, consisting of 13 items. The collected variables included gender, age, family origin, parental medical background, family support for a nursing career, course satisfaction, examination confidence, preferred hospital tier for future employment, self-perceived clinical thinking ability, self-rated communication skills, perceived internship stress, attention to labor market information, and confidence in nursing practical skills.

**Work Readiness Scale for Graduate Nurses (WRS-GN-C)**. Originally developed by Walker et al.[7] and culturally adapted for the Chinese nursing population by Li et al. [19], this 37-item self-reported scale is employed to assess nursing students' work readiness across five dimensions: Job Competitiveness, Social Ability, Career Development, Organizational Acumen, and Personal Work Characteristics. All items are rated on a 10-point Likert scale (1 = strongly disagree to 10 = strongly agree), with total scores ranging from 37 to 370. Higher total scores indicate a higher level of work readiness. In the present study *Cronbach's α* was 0.929.

The continuous total score of WRS-GN-C was adopted as the primary outcome for machine learning modeling. For binary modeling and sensitivity analysis, participants were stratified into low and high work readiness groups based on the sample median (≤ 240.5 vs. > 240.5). Given the lack of universally recognized authoritative cut-off values for this scale, the median-split strategy was applied in accordance with standard data-driven analytical approaches for predictive modeling [20].

**Nurse Professional Identity Scale (NPIS)**. Developed by Liu et al. [21], this 30-item scale comprises five dimensions: professional cognition, social skills, social support, frustration response, and professional reflection. All items are scored on a 5-point Likert scale (1 = completely inconsistent to 5 = completely consistent), with total scores ranging from 30 to 150. The original scale reported a *Cronbach's α* of 0.938,

and in the current study was 0.910. Both subscale scores and the overall total score were included as candidate predictors in the prediction models.

**Future Work Self-Clarity Scale (FWSS)**. The Chinese version of the scale was adapted by Guan et al. [22] based on the original scale developed by Strauss et al. [23]. This 4-item scale is used to assess individuals' clarity regarding their future work self-representation (e.g., "I have a clear idea about what I want to do in my future work"). All items are rated on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree), with total scores ranging from 4 to 28. The *Cronbach's α* in this study was 0.881.

## 2.3 Statistical Methods and Model Development

All statistical and machine learning analyses were performed using Python 3.11 (scikit-learn 1.4), with a two-tailed significance level set at $\alpha = 0.05$. This study adopted two analytical pipelines: continuous regression modeling for overall work readiness prediction and dichotomous classification modeling for supplementary sensitivity analysis.

The continuous WRS-GN-C total score was used as the primary outcome. Seven regression algorithms were compared: Linear Regression, Ridge Regression, Elastic Net Regression, Random Forest Regression, Gradient Boosting Regression, Support Vector Regression, and K-nearest Neighbors Regression. Conservative hyperparameter settings were applied to prevent overfitting. Model performance was evaluated using $R^2$, RMSE, and MAE. For sensitivity analysis, participants were grouped into low (≤ 240.5) and high (> 240.5) work readiness groups via median split. Seven classification models were compared: Logistic Regression, Random Forest, Gradient Boosting, RBF-kernel SVM, KNN, Decision Tree, and Multilayer Perceptron. Classification performance was assessed using AUC-ROC, accuracy, precision, recall, F1-score, and AUC-PR.

The dataset was split into training and test sets at a 7:3 ratio. Stratified sampling was used for classification to preserve outcome distribution. Ten-fold stratified cross-validation was conducted on the training set for internal validation. All model training, hyperparameter tuning, and feature analyses were performed exclusively on the training set, while the isolated test set was used for final evaluation to avoid data leakage. Model interpretability was analyzed using model-matched approaches. Linear models were interpreted via standardized regression coefficients. For tree-based ensemble models, SHAP values calculated with TreeExplainer were applied to quantify individual feature contributions, enabling reliable global and local model interpretation[15]. This multi-method interpretive strategy ensures valid and robust feature importance ranking.

# 3. Results

## 3.1 Participant Characteristics and Descriptive Statistics

A total of 854 vocational nursing students from four colleges in Henan Province were included in the final analytical sample after excluding 105 straight-line responders. The mean age was 20.3 years (SD =

1.1), with the majority being female (n = 770, 90.2%). Approximately 61.8% (n = 528) originated from rural areas, and 28.7% (n = 245) reported having at least one parent with a healthcare occupation. Participants reported moderate-to-high levels of course satisfaction (mean = 3.72, SD = 0.81) and self-rated clinical thinking ability (mean = 2.74, SD = 0.79). The mean professional identity score was 105.4 (SD = 18.9, range 40–150) as measured by the NPIS, the mean future work self-clarity score was 19.0 (SD = 4.4, range 4–28) as measured by the FWSS, and the mean work readiness total score was 245.8 (SD = 50.7, range 52–357) as measured by the WRS-GN-C. Using the sample median split (≤ 240.5 = low readiness, > 240.5 = high readiness), 427 participants (50.0%) were classified as having low work readiness and 427 (50.0%) as having high work readiness.

## 3.2 Primary Analysis: Predicting Continuous WRS-GN Score

To identify the optimal modeling approach for predicting vocational nursing students' work readiness, seven regression algorithms were systematically compared. The core purpose was to screen the model with the best predictive accuracy while ensuring sufficient interpretability for educational practice. All seven regression models were trained and tested based on a 7:3 train-test split, including 598 training samples and 256 test samples, to predict continuous WRS-GN-C scores. The Elastic Net model yielded the best predictive performance with an $R^2$ of 0.705, followed by Linear Regression and Ridge Regression, both achieving an $R^2$ of 0.704. The highly consistent performance of the three linear models indicated a dominant linear additive relationship between the candidate predictors and work readiness, with L1/L2 regularization bringing only marginal predictive improvement.

The optimal Elastic Net model explained 70.5% of score variance, with an RMSE of 28.14 (7.8% relative error) and an MAE of 21.83. The consistent $R^2$ values between 10-fold cross-validation (0.672) and the test set (CV/test ratio = 0.95) demonstrated good generalizability and negligible overfitting. Tree-based models (Random Forest, $R^2$ = 0.689; Gradient Boosting, $R^2$ = 0.672) showed slightly lower performance, suggesting no predictive benefit from increased model complexity. SVR performed poorly ($R^2$ = 0.277), likely due to the inherent linear structure of the data and sensitivity to feature scaling. The narrow $R^2$ range across top-performing models confirmed stable and robust predictive signals. Figure 1 illustrates the actual versus predicted score distribution across all regression models, with the Elastic Net showing the tightest fit to the ideal prediction line. Detailed evaluation metrics for all models are presented in Table 1.

Table 1
Predictive performance of seven regression models for WRS-GN score prediction

| Algorithm | R-squared | RMSE | MAE | 10-CV $R^2$ |
|---|---|---|---|---|
| Elastic Net | 0.705 | 28.14 | 21.83 | 0.672 |
| Linear Regression | 0.704 | 28.19 | 21.80 | 0.667 |
| Ridge Regression | 0.704 | 28.19 | 21.80 | 0.668 |
| Random Forest | 0.689 | 28.92 | 22.30 | 0.656 |
| Gradient Boosting | 0.672 | 29.70 | 22.58 | 0.615 |
| KNN | 0.643 | 31.00 | 24.38 | 0.605 |
| SVR | 0.277 | 44.08 | 34.59 | 0.242 |

# 3.3 Feature Importance Analysis: Elastic Net Model

The most influential predictor was Future Work Self-Clarity (standardized |beta| = 14.51), followed by Skills Confidence (|β| = 7.76), Professional Identity—Social Support (|β| = 6.91), Professional Identity—Frustration Coping (|β|= 5.91), and Clinical Thinking (|β|= 3.86). These top five features collectively accounted for most of the predictive influence, with Future Work Self-Clarity alone contributing approximately 35% of the total model effect. More details are presented in Fig. 1 and Fig. 2.

# 3.4 Sensitivity Analysis: Binary Classification of Work Readiness

As a sensitivity analysis, WRS-GN total score was dichotomized at the sample median (≤ 240.5 = low readiness, > 240.5 = high readiness). Seven classification algorithms were trained and evaluated on the stratified 7:3 split. The results show that SVM achieved the highest AUC-ROC (0.894) and F1-score (0.830). Random Forest followed closely (AUC-ROC = 0.887) with the highest recall (83.6%), indicating superior ability to correctly identify at-risk students, a clinically important metric as false negatives carry more serious consequences than false positives. More details presented in Table 2 and Fig. 3. The Random Forest confusion matrix on the test set yielded: true negatives = 102, false positives = 25, false negatives = 20, true positives = 109, corresponding to an accuracy of 82.5%, sensitivity of 84.5%, and specificity of 80.3%.

Table 2
Predictive metrics of seven classification algorithms for work readiness stratification

| Algorithm | AUC-ROC | Accuracy | Precision | Recall | F1 | AUC-PR | 10-CV AUC |
|---|---|---|---|---|---|---|---|
| SVM (RBF) | 0.894 | 0.829 | 0.823 | 0.836 | 0.830 | 0.885 | 0.893 |
| Random Forest | 0.887 | 0.825 | 0.817 | 0.836 | 0.826 | 0.889 | 0.892 |
| Logistic Regression | 0.882 | 0.817 | 0.819 | 0.813 | 0.816 | 0.882 | 0.901 |
| Gradient Boosting | 0.877 | 0.809 | 0.797 | 0.828 | 0.812 | 0.886 | 0.884 |
| MLP | 0.869 | 0.790 | 0.776 | 0.813 | 0.7940 | 0.871 | 0.886 |
| KNN | 0.860 | 0.794 | 0.786 | 0.805 | 0.795 | 0.841 | 0.899 |
| Decision Tree | 0.813 | 0.735 | 0.763 | 0.680 | 0.719 | 0.766 | 0.824 |

## 3.5 SHAP Feature Importance: Random Forest (TreeExplainer)

Because Random Forest is a tree-based ensemble model lacking inherent coefficient estimates, SHAP values were computed exactly using Tree Explainer to provide theoretically grounded feature attribution. Professional Identity Total emerged as the top predictor (mean |SHAP| = 0.080), followed by Future Work Self-Clarity (0.062) and Social Support (0.057), more details presented in Fig. 4. The SHAP beeswarm plot (Fig. 5) reveals important directional patterns: for Future Work Self-Clarity, red dots (high feature values) consistently cluster on the right (positive SHAP values pushing toward high readiness), while blue dots (low values) cluster on the left, confirming the monotonic positive relationship identified in the regression analysis. The SHAP heatmap (Fig. 6) further illustrates that Professional Identity Total exhibits consistently strong positive contributions for high-probability samples (left panel) but weak or negative contributions for low-probability samples (right panel), suggesting context-dependent feature effects that coefficient-based methods cannot capture. Demographic variables (bottom rows) show minimal and inconsistent patterns across all samples.

# 4. Discussion

## 4.1 Principal Findings

This study developed and validated an interpretable machine learning prediction model for vocational nursing students' work readiness. The Elastic Net model's high explanatory power ($R^2$ = 0.705) and stable cross-validation performance demonstrate that selected psychological indicators adequately explain individual disparities in pre-employment preparation. Supplementary binary classification further highlights practical educational value: Random Forest achieved an 83.6% recall rate, which is critical for

identifying students at high risk of poor post-graduation clinical adaptation. Modifiable psychological constructs dominated predictive effects, while demographic variables contributed nearly no variance to work readiness scores. Convergent evidence from standardized regression coefficients and SHAP explainable artificial intelligence consistently ranked future work self-clarity as the top predictor, followed by skills confidence and two subdimensions of nurse professional identity. Distinct from traditional single linear regression used in prior literature, this dual-model interpretable framework provides quantified, theory-based evidence to advance precision nursing education and deliver individualized early intervention for vocational nursing trainees.

# 4.2 Interpretation of Key Predictors

Future work self-clarity emerged as the single most influential predictor across both analytical approaches, ranking first in Elastic Net regression (|β| = 14.51) and second in Random Forest SHAP analysis (mean |SHAP| = 0.062). This cross-model convergence provides robust empirical evidence for the core value of occupational self-concept in shaping vocational nursing students' pre-employment preparation. Based on Social Cognitive Career Theory (SCCT), future work self-clarity reflects an individual's clear cognitive representation of their future occupational self, serving as a core motivational driver that stimulates proactive career exploration, clinical practice engagement, and adaptive coping behaviors [22, 23]. Holding all other variables constant, a one-point increase on the 4–28 FWSS scale predicted a 14.5-point increase on the 37–370 WRS-GN scale. This practically significant effect size provides strong empirical support for structured career visualization interventions and mentorship programs that facilitate exploration of professional possible selves.

Skills confidence ranked as the second most influential predictor (standardized |β|= 7.76), which distinguishes this study from prior investigations focusing on baccalaureate nursing students. Li et al. [24] reported that professional identity total score outperformed clinical self-efficacy in predicting work readiness among four-year nursing undergraduates, whereas skills confidence occupied a superior predictive position in our vocational nursing cohort. This discrepancy can be attributed to the core training orientation of vocational nursing education: vocational colleges prioritize clinical operational competency cultivation, and students’ self-perceived mastery of nursing skills directly determines their willingness to engage in clinical internships and cope with workplace challenges. Consistent with Bandura’s social cognitive theory, clinical skill confidence acts as a critical self-efficacy signal: students with strong skill confidence actively participate in complex patient care scenarios to accumulate practical experience, while those lacking competency self-assurance adopt avoidant learning strategies that further widen gaps in work readiness [25]. Longitudinal follow-up evidence further corroborates this mechanism: nursing graduates with insufficient clinical self-efficacy suffer from severe transition shock and higher turnover intention within their first year of clinical practice [26, 27].

Two subscales of nurse professional identity—social support (|β| = 6.91) and frustration coping (|β| = 5.91)—ranked third and fourth in feature importance, exceeding the predictive weight of the overall professional identity total score. This nuanced layered association could not be captured by traditional single-variable correlation analysis, demonstrating the unique advantage of multi-factor machine

learning modeling in identifying targeted intervention dimensions. The prominent predictive effect of frustration coping aligns with Duchscher's transition shock theory, which proposes that emotional adjustment and stress-resilience resources constitute the decisive factor for new nurses' smooth role transformation[28]. Internship pressure, interpersonal conflicts and clinical work setbacks are pervasive for vocational nursing students, and poor frustration coping capacity will trigger compassion fatigue and weaken professional identity formation [29]. Meanwhile, perceived social support from teachers, clinical preceptors and peers provides external psychological buffering resources, mitigating internship stress and maintaining students' positive career cognition [30]. Differentiating the predictive value of professional identity subdimensions indicates that educators should design targeted resilience training and peer support activities rather than carrying out generalized professional identity enhancement programs with diluted intervention effects.

## 4.3 Practical Implications for Nursing Education

Based on the feature importance ranking and validated Elastic Net predictive model, this study proposes a Precision Vocational Nursing Education framework that translates statistical predictions into actionable, tiered interventions. An open web-based prediction tool[31]supports real-time risk screening for integration into routine teaching management.

Universal intervention for all students. Career visualization workshops centered on future work self-clarity are embedded as compulsory first-year modules. Guided by Possible Selves Theory, structured career exploration activities, mentor sharing, and clinical role visualization exercises help students construct clear, personalized nursing career blueprints at the formative stage of professional training. Quantifiably, a 2-point increase in future work self-clarity elevates the WRS-GN total score by 29 points, sufficient to shift students from low to high work readiness status. Concurrently, formative feedback-driven clinical simulation courses are scheduled throughout internships to generate mastery experiences, the core source of self-efficacy [25], thereby continuously building skills confidence.

Selective intervention for moderate-risk students. Following routine model screening each semester, students with low future work self-clarity or insufficient skills confidence are flagged for targeted support. One-on-one professional tutoring provides individualized career planning, clarifies developmental pathways, and supplements low-intensity skill training to elevate competency self-perception.

Indicated intensive intervention for high-risk students. Those with multiple concurrent risk factors (low future work self-clarity, poor skills confidence, insufficient social support, and weak frustration coping) receive combined intensive support: cognitive behavioral therapy–based resilience training to improve stress and setback coping[32]; structured peer support groups to enhance perceived social support; and high-fidelity simulation with individualized operational guidance to consolidate clinical skills.

At the curriculum level, three standardized modular courses are recommended for the first academic year, the critical formative window for occupational self-concept development[33]: career visualization

workshops, resilience and emotional regulation training, and competency-oriented clinical simulation practice.

At the policy level, demographic variables collectively explained less than 1% of model variance, indicating that work readiness is minimally constrained by socioeconomic background. This pattern aligns with evidence that supportive educational environments compensate for demographic disadvantage through self-efficacy mediation [30]. Therefore, resource allocation should prioritize universal investments in career counseling infrastructure, psychological intervention services, and clinical simulation training centers over background-based subsidies. Such reallocation can transform remedial teaching into precision talent cultivation, promoting equitable improvement of work readiness across all student groups and ultimately enhancing the quality and quantity of the vocational nursing workforce.

## 4.4 Limitations and Recommendations

Several limitations warrant acknowledgment. First, the cross-sectional design precludes causal inference; reciprocal effects may exist among psychological variables (e.g., future work self-clarity enhances clinical confidence, while successful internship experience further reinforces occupational clarity). Longitudinal cohort studies tracking students throughout the three-year training cycle are needed to identify critical intervention windows and establish temporal precedence. Second, all predictors were self-reported, introducing potential common method bias. Future research should incorporate objective measures such as standardized clinical skills assessments and clinical preceptor performance ratings. Third, the sample was confined to Henan Province; replication in coastal, western, and northeastern provinces is necessary to confirm national generalizability. Fourth, while conservative hyperparameters mitigated overfitting, external validation on independent cohorts from different academic years remains the gold standard for confirming model stability [34].

Four directions merit future pursuit. First, time-series longitudinal modeling can extract dynamic developmental trajectories of work readiness and test whether early-semester psychological indicators predict graduation outcomes. Second, natural language processing of unstructured data (internship journals, reflective diaries) could enrich the predictor space beyond questionnaire-based variables. Third, causal inference methods should be applied to quantify the independent effects of specific interventions, career counseling, resilience training, and simulation-based competency development, on readiness outcomes, providing causal evidence for policy formulation. Finally, randomized controlled trials comparing model-guided stratified personalized training against conventional uniform career education would generate high-level evidence to support large-scale implementation of precision nursing education.

# 5. Conclusion

This study established reliable machine learning predictive models, validated through convergent regression and classification analyses, to quantify hierarchical contributions of modifiable psychological

factors to vocational nursing students' work readiness. Future work self-clarity, skills confidence, professional identity social support, and frustration coping constitute the core intervention targets, while demographic background exerts negligible predictive impacts. The tiered precision intervention framework proposed herein offers a data-driven, interpretable solution enabling educators to screen at-risk students prospectively and implement differentiated training. These advances contribute to elevating the overall employment quality of vocational nursing graduates and alleviating the clinical nursing workforce shortage.

## Abbreviations

| | |
|---|---|
| FWSS | Future Work Self-Clarity Scale |
| NPIS | Nurse Professional Identity Scale |
| WRS-GN-C | Work Readiness Scale for Graduate Nurses (Chinese version) |
| SCCT | Social Cognitive Career Theory |
| SHAP | SHapley Additive exPlanations |
| ML | Machine Learning |
| XAI | eXplainable Artificial Intelligence |
| AUC-ROC | Area Under the Receiver Operating Characteristic Curve |
| RMSE | Root Mean Square Error |
| MAE | Mean Absolute Error |
| CV | Cross-Validation |

## Declarations

# Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

# Funding information

This study was funded by the Education Department of Henan Province (Grant Number: 2025YB0618). The funding body had no role in the study design, data collection, analysis, interpretation, or manuscript

preparation.

# Author Contribution

XXD, ZS, and LJZ. Conceptualized the study; XXD. Curated data; DYW and ZS. Performed formal analysis; XXD. Acquired funding; XXD and LJZ. Conducted investigation; XXD, DYW, ZS, and GZZ. Developed methodology; XXD. Administered the project; XXD. Provided resources; DYW. Developed software; GZZ. Supervised the research; GZZ. Validated the study; DYW. Created visualizations: XXD. Wrote the original draft of the manuscript; GZZ. Reviewed and edited the manuscript. All authors read and approved the final manuscript.

# Acknowledgement

We sincerely thank all nursing students who completed the questionnaires and provided valid research data. We also appreciate all participants who assisted in finishing this study. Finally, special thanks are given to the Henan Provincial Department of Education for funding support.

# Data Availability

All raw data and supplementary materials underlying this study are publicly available in the Science Data Bank at https://www.scidb.cn/en/s/fquUZr.

## References

1. World Health Organization. State of the World's Nursing 2025: investing in education, jobs, leadership and service delivery. Geneva: WHO; 2025.
2. National Health Commission of China. Statistical bulletin on national health development in 2025. Beijing: NHC; 2026.
3. Organisation for Economic Co-operation and Development. Health at a glance 2023: OECD indicators. Paris: OECD Publishing; 2024. 10.1787/health_glance-2023-en.
4. Zhang R, Li M, Chen Y, et al. Factors influencing vocational nursing students' career intentions in geriatric nursing in Central-South China. BMC Nurs. 2025;24:432. 10.1186/s12912-025-01573-7.
5. Kenny A, Dickson-Swift V, McKenna L, et al. Interventions to support graduate nurse transition to practice and associated outcomes: a systematic review. Nurse Educ Today. 2021;100:104860. 10.1016/j.nedt.2021.104860.
6. See ECW, Koh SSL, Baladram S, et al. Role transition of newly graduated nurses from nursing students to registered nurses: a qualitative systematic review. Nurse Educ Today. 2023;121:105702. 10.1016/j.nedt.2022.105702.

7. Walker A, Storey KM, Costa BM, et al. Refinement and validation of the work readiness scale for graduate nurses. Nurs Outlook. 2015;63(6):632–8. 10.1016/j.outlook.2015.07.003.
8. Hampton KB, Smeltzer SC, Ross JG. Evaluating the transition from nursing student to practicing nurse: an integrative review. J Prof Nurs. 2020;36(4):551–9. 10.1016/j.profnurs.2019.12.003.
9. Liu C, Li Q, Xu Y, et al. Nursing students' work readiness and its influencing factors before participating in a nurse residency program: a multicenter cross-sectional study. Front Med. 2024;11:1391907. 10.3389/fmed.2024.1391907.
10. Zheng R, Wang Y, Li Y, et al. The impact of workflow interruptions, work readiness and missed nursing care on patient safety competency among new nurses. J Clin Nurs. 2025;34(3):891–903. 10.1111/jocn.17305.
11. Tarhan N, Yildirim B. Investigation of factors affecting the career maturity of nursing students. Int J Caring Sci. 2021;14(1):423–31.
12. Patterson EEB, Boyd L, Mnatzaganian G, et al. The impact of undergraduate clinical teaching models on the perceptions of work-readiness among new graduate nurses: a cross sectional study. Nurse Educ Today. 2017;55:101–6. 10.1016/j.nedt.2017.05.007.
13. Naz M, Shams R, Parveen M. Statistical and machine learning models for predicting university dropout: a comparative study. PLoS ONE. 2024;19(9):e0311130. 10.1371/journal.pone.0311130.
14. Al Hashmi RAM, Ozturk I, Elmehdi HM. Early detection of at-risk health sciences students: a machine learning-based predictive study using midterm grades. BMC Med Educ. 2025;25(1):1651. 10.1186/s12909-025-08192-6.
15. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765–74.
16. Ghimire S. Explainable artificial intelligence-machine learning models to predict at-risk students in higher education. In: Artificial Intelligence in Education. Springer; 2024. pp. 1–15. 10.1007/978-3-031-64315-3_9
17. Curran PG. Methods for the detection of carelessly invalid responses in survey data. J Exp Soc Psychol. 2016;66:4–19. 10.1016/j.jesp.2015.07.006.
18. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ. 2020;368:m441. 10.1136/bmj.m441.
19. Li J, Gong Y, Chen J, et al. Psychometric properties assessment of Chinese version of work readiness scale for graduate nurses (WRS-GN) and associated factors of work readiness: two cross-sectional studies. Nurse Educ Today. 2020;89:104416. 10.1016/j.nedt.2020.104416.
20. DeCoster J, Gallucci M, Iselin AMR. Best practices for using median splits, artificial categorization, and their continuous alternatives. J Exp Psychopathol. 2011;2(2):197–209. 10.5127/jep.008109.
21. Liu M, Luo Y, Zhang Y. Development of the Nurse Professional Identity Scale. J Nurs Sci. 2007;22(9):1–3.

22. Strauss K, Griffin MA, Parker SK. Future work selves: how salient hoped-for identities motivate proactive career behaviors. J Appl Psychol. 2012;97(3):580–98. 10.1037/a0026423.
23. Guan Y, Wang Z, Chen Y, et al. The Chinese version of the Future Work Self-Clarity Scale: psychometric properties and validity. Chin J Clin Psychol. 2021;29(4):789–93.
24. Li J, Huang Y, Fong DYT, et al. Work readiness: its determinants and association with work-related outcomes among new graduate nurses. J Nurs Manag. 2022;30(7):2968–81. 10.1111/jonm.13691.
25. Bandura A. Social foundations of thought and action: a social cognitive theory. New Jersey: Prentice-Hall; 1986.
26. Ohue T, Ohue Y. Impact of COVID-19 pandemic-induced changes in clinical practicums on the mental health of newly graduated nurses: longitudinal study. JMIR Nurs. 2025;8:e79556. 10.2196/79556.
27. Kawashima T, Ota Y, Aikawa G, et al. Effectiveness of emotional intelligence training on nurses' and nursing students' emotional intelligence, resilience, stress, and communication skills: a systematic review and meta-analysis. Nurse Educ Today. 2025;141:106299. 10.1016/j.nedt.2025.106299.
28. Duchscher JEB. Transition shock: the initial stage of role adaptation for newly graduated registered nurses. J Adv Nurs. 2009;65(5):1103–13. 10.1111/j.1365-2648.2008.04898.x.
29. Zhong Q, Zhou Y, Li J, et al. How future work self-salience influences occupational sense of mission among medical students in the post-pandemic era: a dual-perspective analysis from variable-centered and person-centered approaches based on professional identity. Front Public Health. 2026;14:1753631. 10.3389/fpubh.2026.1753631.
30. Zhao ZH, Guo JY, Zhou J, et al. Perceived social support and professional identity in nursing students during the COVID-19 pandemic era: the mediating effects of self-efficacy and the moderating role of anxiety. BMC Med Educ. 2023;23:117. 10.1186/s12909-023-04164-0.
31. Prediction tool for. vocational nursing students' work readiness. https://vuq7g7fzwkryo.ok.kimi.link/. Accessed 16 July 2026.
32. Saikia M, George LS, Unnikrishnan B, et al. Thirty years of emotional intelligence: a scoping review of emotional intelligence training programme among nurses. Int J Ment Health Nurs. 2024;33(1):37–51. 10.1111/inm.13235.
33. Wang Y, Liu H, Chen X, et al. Effects of clinical learning environment on the professional identity of Chinese nursing students: multiple mediating effects of learning engagement and stress-related growth. BMC Nurs. 2024;23:425. 10.1186/s12912-024-01897-2.
34. Latham IM, Rush KL, Hatt L, et al. A landscape of evidence on RN transition to practice programs. J Nurs Regul. 2024;15(3):45–58. 10.1016/S2155-8256(24)00103-1.

# Figures

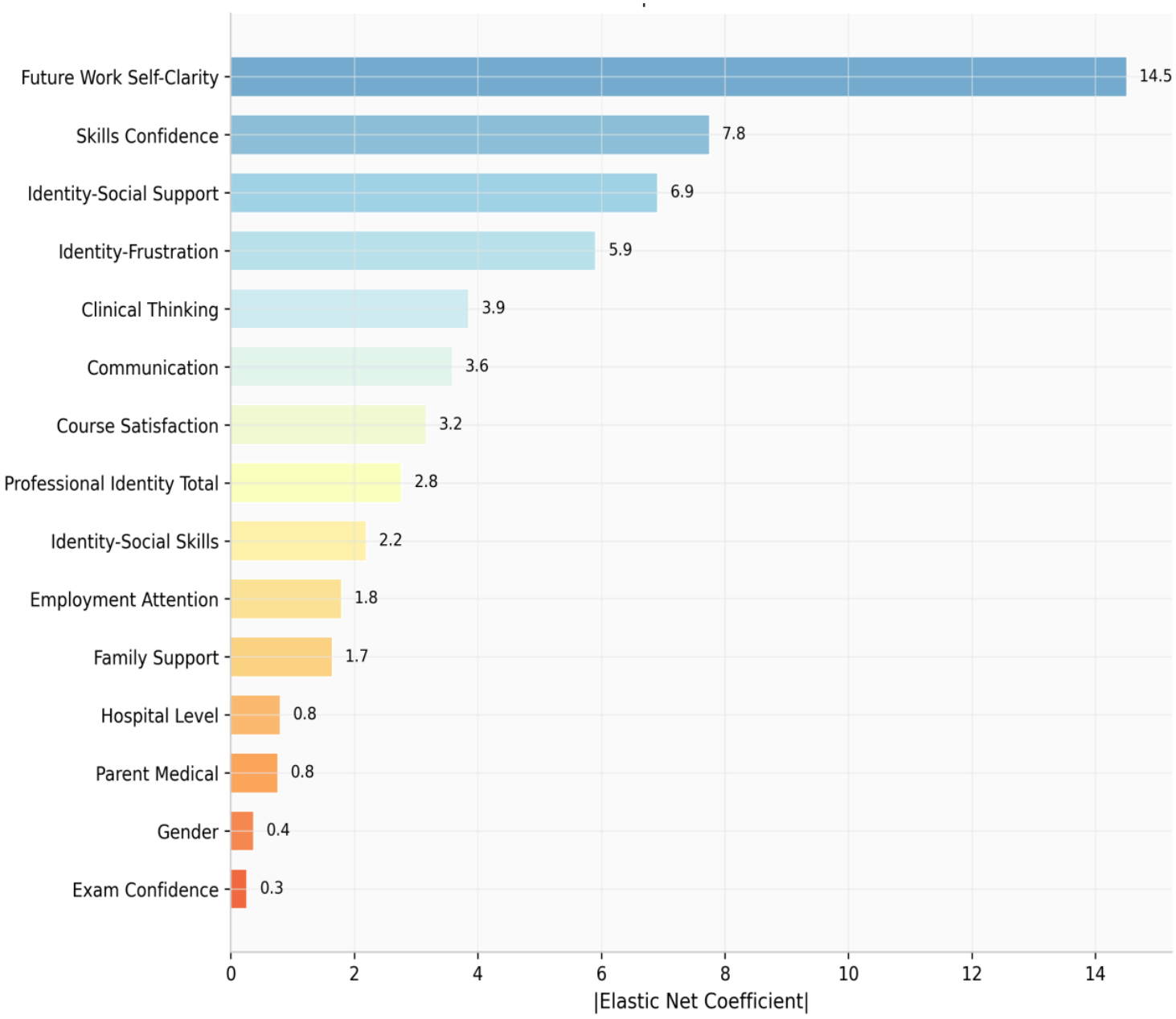
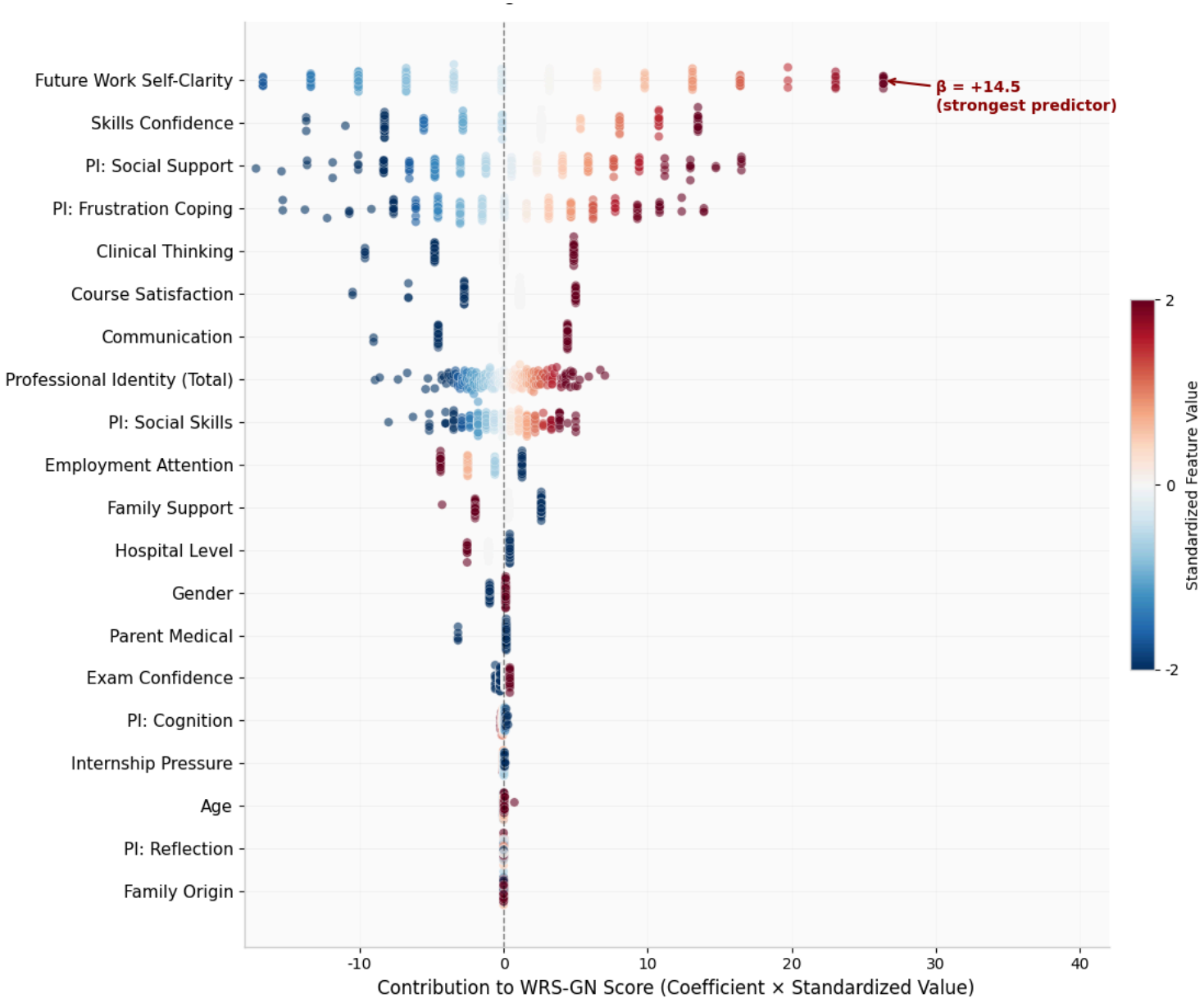



**Figure 1**

Feature importance of Elastic Net model (Top 15)




## Figure 2

Contribution Plot: Elastic Net Regression Model (Test Samples n = 256)

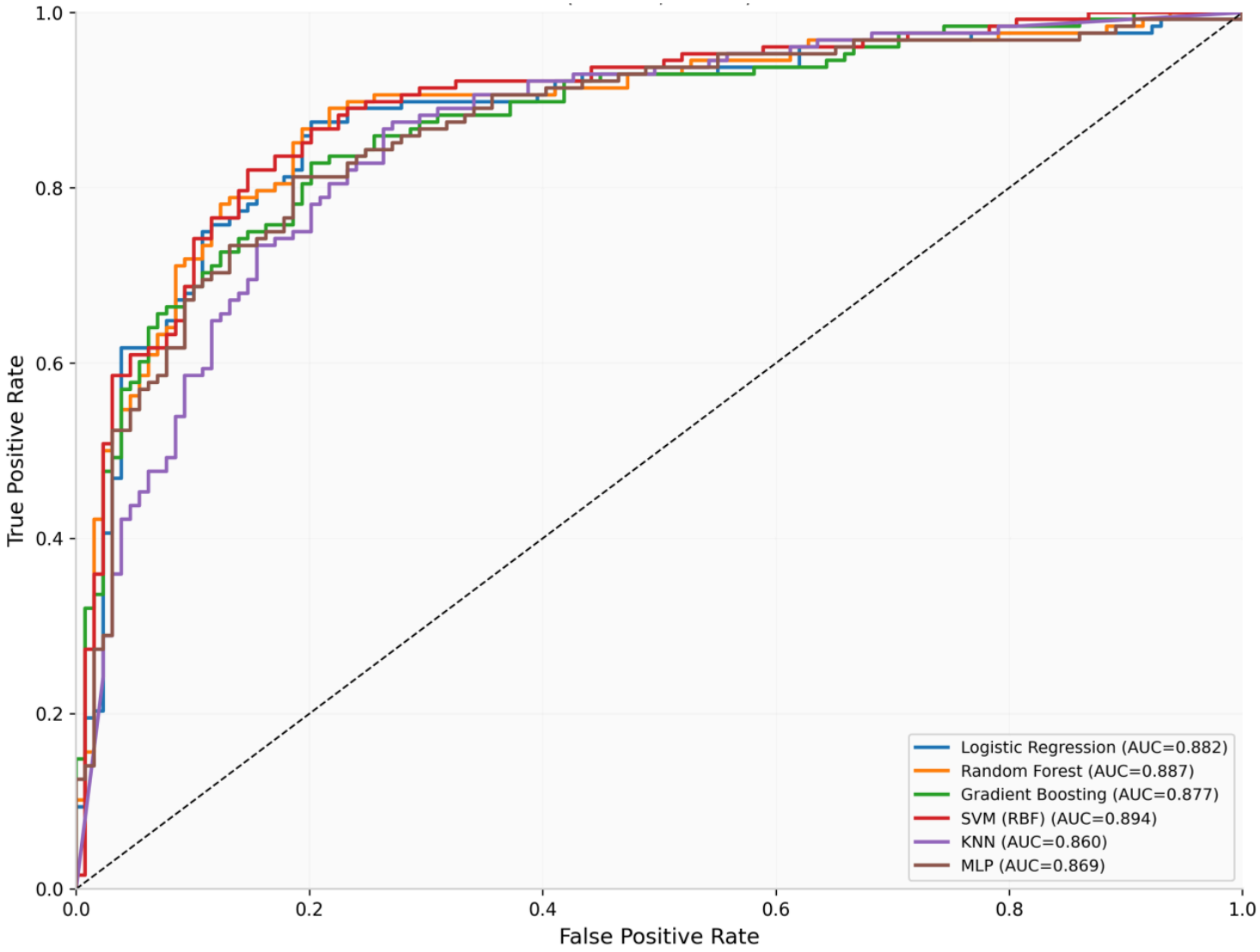



**Figure 3**

ROC Curves for Seven Classification Algorithms

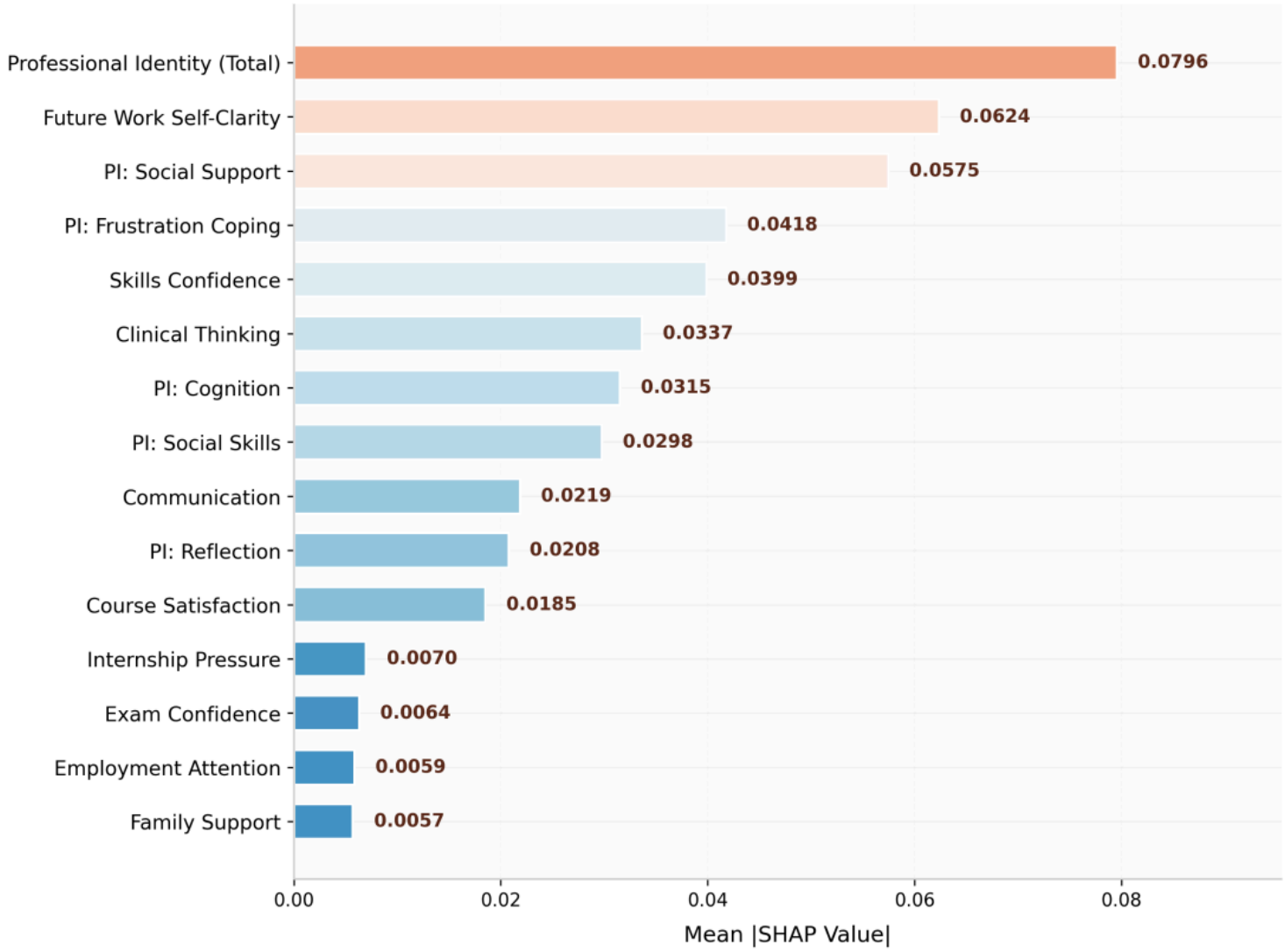
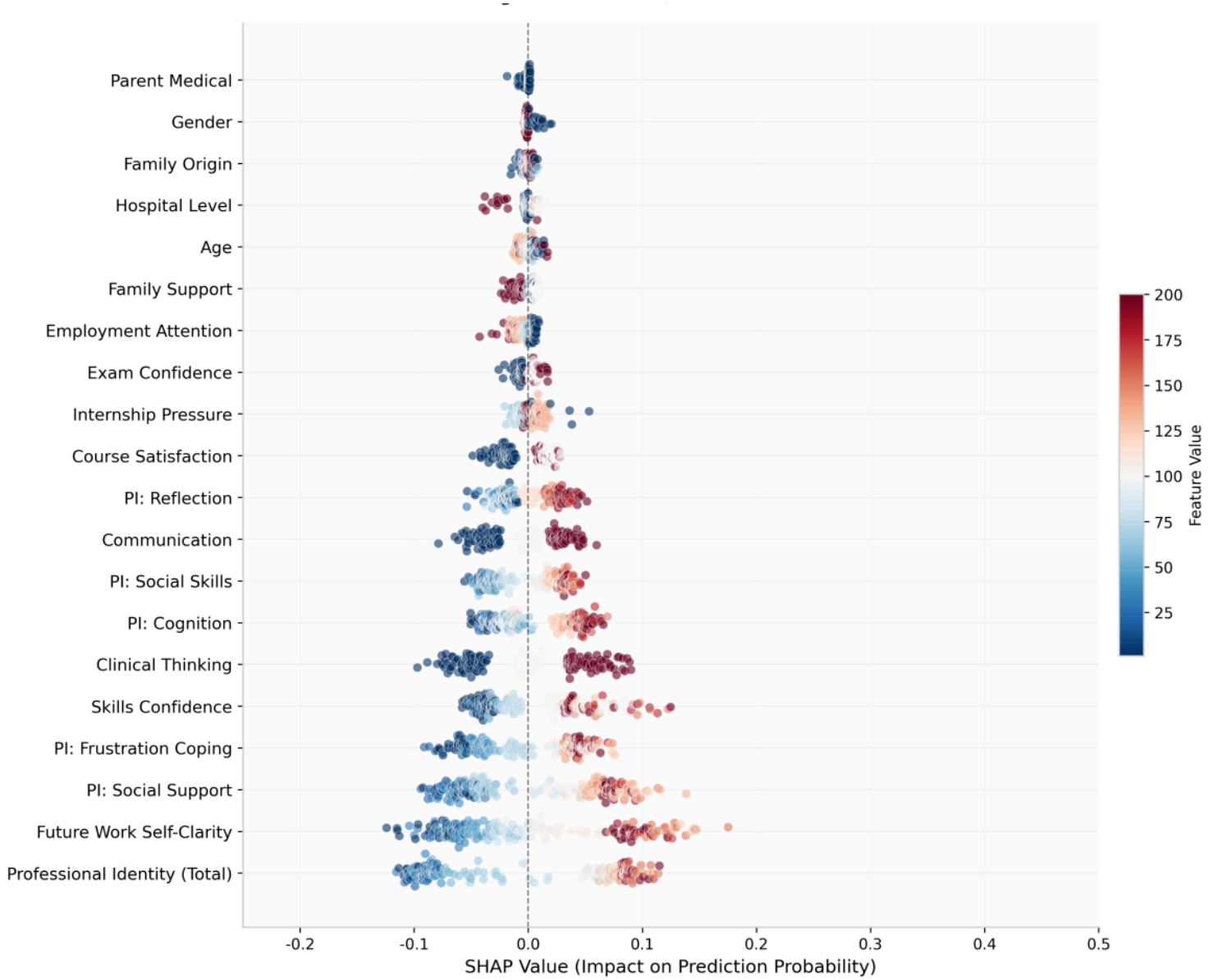


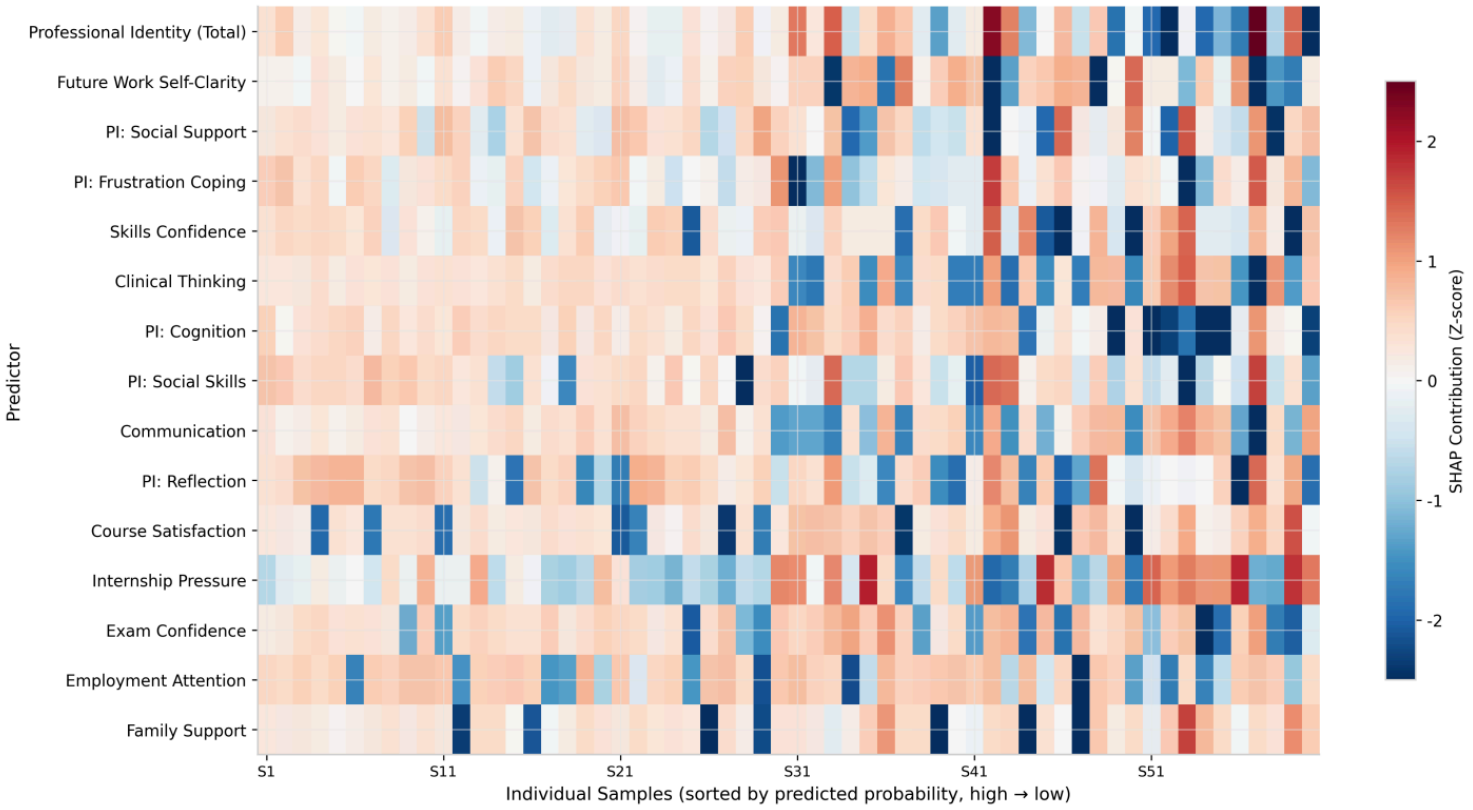

**Figure 4**

SHAP Feature Importance: Random Forest (Binary Classification, TOP 15)




## Figure 5

SHAP Beeswarm Plot: Random Forest (Binary Classification)




## Figure 6

SHAP Contribution Heatmap of Random Forest (Top 15 predictors)