Predicting the Transfer Intentions of Specialist

Nurses through Machine Learning

Authors:

|  |  |
| --- | --- |
| Name | Gan Wang |
| Email |  |
| Department | Shanghai Institute of Infectious Disease and Biosecurity |
| University | Fudan University |

|  |  |
| --- | --- |
| Name | Zheng Luo |
| Email | luoz@umich.edu |
| Department | UM-SJTU Joint Institute |
| University | Shanghai Jiao Tong University |

|  |  |
| --- | --- |
| Name | Huiying Wu |
| Email |  |
| Department | School of Public Health |
| University | Fudan University |

|  |  |
| --- | --- |
| Name | Jiaojiao Wang |
| Email |  |
| Department | School of Public Health |
| University | Fudan University |

|  |  |
| --- | --- |
| Name | Wen Ma |
| Email |  |
| Department | Shanghai Institute of Infectious Disease and Biosecurity |
| University | Fudan University |

|  |  |
| --- | --- |
| Name | Li Luo |
| Email | liluo@fudan.edu.cn |
| Department | School of Public Health |
| University | Fudan University |

**Abstract**

The shortage of specialist nurses poses a significant challenge to healthcare systems worldwide, impacting both patient care and hospital operations. This study explores the application of machine learning to predict the job transfer tendencies of specialist nurses, aiming to provide hospitals with actionable insights for retention strategies. We applied multiple classification algorithms, including Random Forest, XGBoost, and MLPClassifier, with the final model achieving an AUC score of 0.713 after cross-validation and hyperparameter optimization using the Tree-structured Parzen Estimator (TPE). To better understand feature contributions, we utilized SHAP (SHapley Additive exPlanations) values to perform feature importance analysis. Our findings identified that key predictors of nurse transfer include clinical teaching participation, specialist training attendance, and job satisfaction. However, certain results, such as the unexpected positive correlation between satisfaction with performance distribution and job transfer, revealed anomalies likely due to class imbalance or data biases. We also discuss the trade-offs between different data preprocessing techniques, particularly regarding oversampling within cross-validation to mitigate class imbalances. While the model’s performance was moderate, the insights from SHAP analysis provide valuable directions for future research. We recommend collecting more representative data and refining feature selection to improve model accuracy. This research contributes to the development of predictive tools for hospital management to address the specialist nurse shortage by identifying nurses at risk of job transfer.

**Strengths and Weaknesses**

### Strengths:

* **Multiple model comparison**: By implementing various machine learning models (Random Forest, XGBoost, MLP Classifier, AdaBoost, LGBM Classifier, and CatBoost Classifier), the study explores a wide range of algorithmic approaches to optimize predictive performance.
* **Cross-validation**: The use of cross-validation ensures robust model selection and reduces the risk of overfitting, enhancing the generalizability of the results.
* **Hyperparameter tuning**: The application of Optuna’s TPE-based optimization allows for efficient hyperparameter selection, leading to improved model performance.
* **Handling imbalanced data**: By incorporating SMOTE (Synthetic Minority Over-sampling Technique), the study addresses class imbalance, ensuring better prediction of minority cases like resignations.
* **Feature importance analysis**: Using SHAP to interpret feature importance adds transparency and interpretability, making it easier to understand the models' decision-making processes.

### Weaknesses:

* **Limited interpretability in some models**: While powerful, models like XGBoost and CatBoost can be complex, making it harder to interpret the underlying relationships compared to simpler models like logistic regression.
* **One-hot encoding scalability**: One-hot encoding used in preprocessing can lead to high-dimensional data, especially when categorical features have many unique values, potentially impacting computational efficiency.
* **Potential data leakage risk**: The extensive preprocessing steps, including feature selection and scaling, require careful management to avoid leakage between training and testing data.
* **Hyperparameter tuning computational cost**: Although Optuna optimizes the process, the computational cost of hyperparameter tuning across multiple models can be high, particularly when combined with cross-validation.
* **SMOTE limitations**: While SMOTE helps balance the dataset, it may generate synthetic data that doesn't perfectly represent real-world nurse resignation patterns, potentially introducing noise into the training process.

**I. INTRODUCTION**

**A. Background**

Specialist nurses are essential in maintaining the quality of care in hospitals. However, the supply of these nurses is limited, and when they decide to transfer or leave their positions, it

can have a significant impact on healthcare systems. Globally, there were approximately 29.1 million nurses in 2020, but a shortage of 4.5 million nurses is projected by 2030, according to the World Health Organization (WHO) [1].

In China, the situation is particularly critical. By the end of 2023, the country had 5.63 million registered nurses, translating to just 4 nurses per 1,000 people. This is far below the Organization for Economic Co-operation and Development (OECD) average of 9.3 nurses per 1,000 people [2], and even the minimum threshold of 4.45 nurses per 1,000 people recommended by the WHO [3].

Several factors contribute to this global nursing shortage. These include the aging population [4], an increase in chronic diseases [5], and job transfers within the nursing profession [6]. Job transfers, in particular, are a major reason for staff shortages, as they directly affect the quality of clinical care and patient safety. For example, in sub-Saharan Africa, the intention to transfer among nurses is alarmingly high, reaching 50.74%, while in East Africa, it soars to 58.03% [7].

In China, the annual job transfer rate among hospital nurses ranges between 20% and 45% [8]. Even at the lower end of this range, a 20% transfer rate can lead to significant financial burdens for hospitals and reductions in medical quality. At the upper end, with a 45% transfer rate, hospitals face severe organizational and operational disruptions [9].

Given the critical role that nurse retention plays in healthcare, numerous studies have sought to predict job transfer tendencies among nurses. By identifying early indicators of transfer, hospitals can either prepare for the eventual vacancies or take preemptive steps to reduce turnover intentions, ultimately helping to mitigate the effects of staff shortages.

**B. Motivation**

Given the challenges posed by the limited supply of nurses and their high job transfer rates, it is essential for hospitals to anticipate potential transfers in advance. Accurate predictions of nurse transfer tendencies would enable hospitals to either prepare for staffing vacancies or implement targeted retention strategies. Such proactive measures can alleviate the negative impacts of nurse shortages, ensuring continuity in patient care and reducing the operational strain on hospital resources. Therefore, the development of reliable predictive models for nurse job transfers is crucial for maintaining a stable and efficient healthcare workforce.

**C. Research Objectives**

This paper aims to develop a machine learning model to predict the job transfer tendencies of specialist nurses using a variety of statistically collected data. In addition to building the prediction model, we analyze the key factors that most strongly correlate with transfer tendencies. By identifying these critical factors, hospitals can use the insights to create a more

supportive work environment and reduce nurses’ intentions to leave, ultimately helping to mitigate the ongoing shortage of specialist nurses.

**D. Methodological Approach**

The development of the predictive model followed a standard machine learning workflow, encompassing data preprocessing, cross-validation, model training, and testing. We implemented several preprocessing strategies, including standard scaling for numerical data, one-hot encoding for categorical variables, and feature selection using a logistic regression

estimator. Additionally, we utilized the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance through data oversampling.

To identify the final model, we evaluated multiple classification algorithms, such as Random Forest and XGBoost, based on their performance during cross-validation.

For feature importance analysis, we employed the Shapley Additive Explanations (SHAP) algorithm to assess the most significant features impacting the prediction outcomes. We visualized these relationships using SHAP value violin plots, which provided insights into feature importance and their contributions to the model’s predictions.

**II. SIGNIFICANCE AND CONTRIBUTION**

This study offers significant insights into the factors influencing job transfer tendencies among specialist nurses, with a particular focus on the Chinese healthcare context. By utilizing data collected from Chinese nurses, our research provides hospitals in China with a predictive model that addresses the pressing issue of nurse turnover, a challenge that has

substantial implications for healthcare delivery and patient care.

Moreover, the findings from this study have broader relevance beyond China, as many healthcare systems worldwide face similar challenges related to nurse retention and turnover. The insights gained from our analysis of the key factors influencing transfer tendencies can serve as a valuable reference for hospitals globally. By understanding these factors, healthcare institutions can implement targeted strategies to create a more supportive work environment, ultimately leading to improved job satisfaction and reduced turnover intentions among nursing staff.

The primary contribution of this research lies in its provision of a data-driven predictive model that equips hospitals with the tools to proactively address nurse retention. This model not only enhances the understanding of the complex dynamics influencing nurse job transfers but also empowers healthcare administrators to make informed decisions aimed at reducing turnover. In a time when healthcare systems are grappling with shortages of specialist nurses, our findings are crucial for developing effective workforce management strategies that ensure continuity of care and optimal patient outcomes.

**III. RELATED WORK**

The issue of nurse turnover has garnered significant attention in recent literature, particularly in the context of utilizing machine learning and statistical approaches to predict turnover intentions. Tiwari et al. (2023) conducted a comprehensive analysis employing various machine learning models to assess employee turnover. Their findings suggest that predictive analytics can substantially aid organizations in preserving workforce stability and reducing costs associated with rehiring and training new employees [10].

In a more focused study, Zhang et al. (2023) explored the turnover intentions of newly graduated nurses during their first year of employment. By utilizing longitudinal data and regression analysis, the study identified critical factors such as career self-efficacy and adaptability that significantly influence nurses’ decisions to leave their positions. This research highlights the importance of understanding the unique challenges faced by new nursing professionals and offers actionable insights for healthcare management [11].

**IV. METHODOLOGY**

This section describes how we constructed the predictive model and conducted the feature importance analysis.

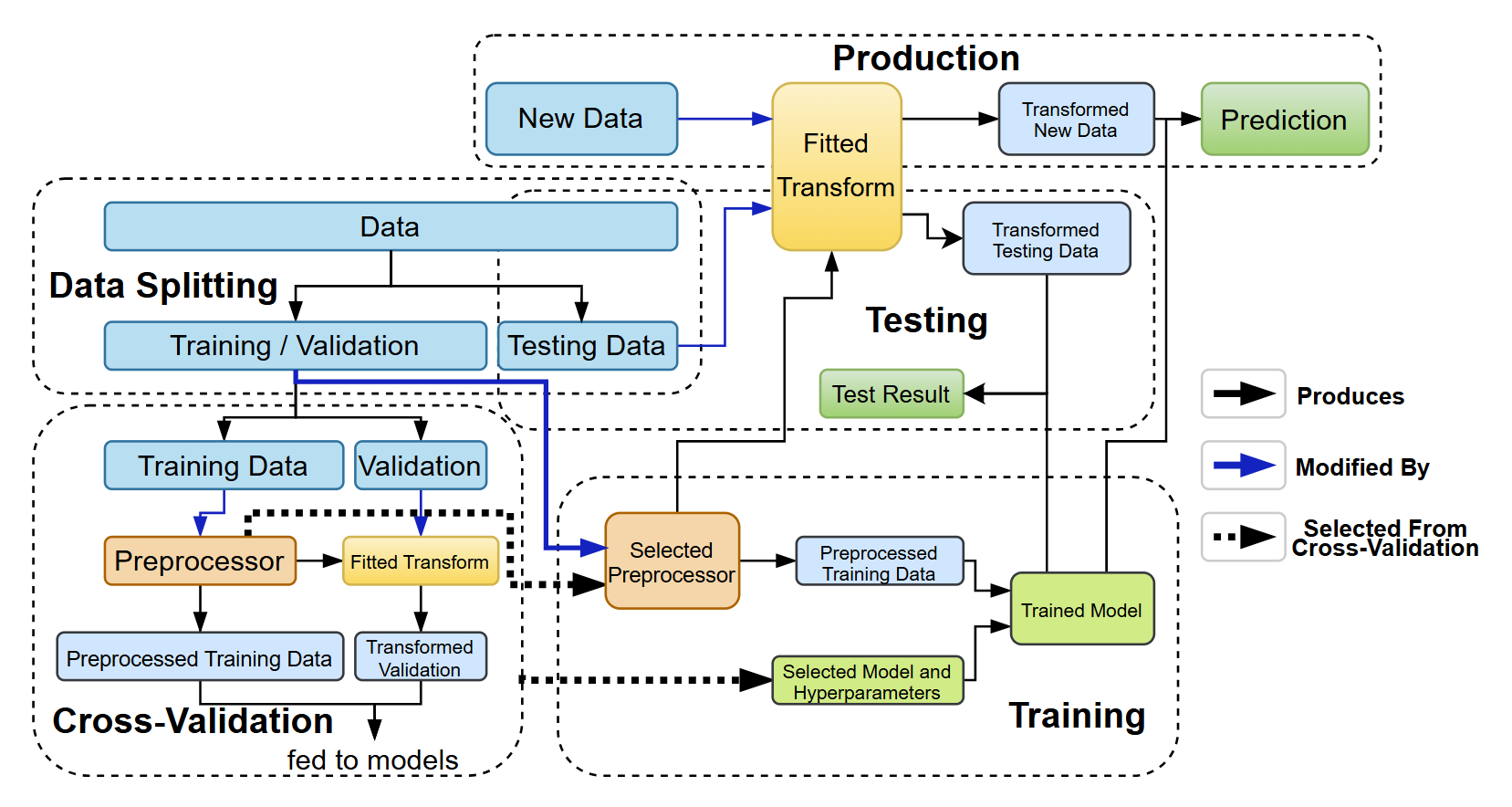
**A. Data Collection**

Data was collected from nurses in Shanghai, facilitated by the Shanghai Xuhui District Health Commission. Anonymous questionnaires were distributed via the online platform Wenjuanxing to all district medical institutions, excluding one tertiary hospital.

The questionnaire covered a range of features, including demographic characteristics, professional experience, education level, and psychosocial factors. The original questionnaire is provided in Appendix A for reference.

To reduce the burden on participants and encourage more accurate responses, we discretized numeric answers into categories. For instance, the question ”Night Shift Frequency Per Month” offered three options: “0-5”, “5-10”, and “>10”. This approach made the questions less intimidating and helped ensure truthful responses without overwhelming participants.

Once data collection was complete, we proceeded with model development, which involved four key stages: data splitting, cross-validation, model training, and testing. The overall workflow is illustrated in Figure 1.



**Fig. 1. The machine learning workflow adopted in this paper.**

**B. Data Preprocessing**

The following preprocessing techniques were applied to the dataset:

**1) Standard Scaling:** Standard scaling was used to normalize numeric features, ensuring they have a mean of zero and a standard deviation of one. This step prevents features with larger scales from disproportionately influencing the model.

**2) One-Hot Encoding:** To handle categorical data, we applied one-hot encoding, which converts categories into binary vectors. This allows models to process categorical variables without assuming an ordinal relationship between the categories, which is essential for features like job roles or departments.

**3) Feature Selection From Logistic Regression Model:** We leveraged the coefficients from a logistic regression model to select important features. This method reduces model complexity, enhances interpretability, and may improve performance by eliminating irrelevant or redundant variables.

**4) Synthetic Minority Oversampling Technique (SMOTE):** SMOTE was employed to address class imbalance by generating synthetic examples for the minority class, rather than duplicating existing ones. This technique improves the model’s ability to generalize, reducing bias toward the majority class—an essential consideration when predicting rare outcomes like nurse transfer intentions.

Preprocessing steps were applied after the training/testing data split and within cross-validation to prevent data leakage, which can occur if the transformation uses global information from both the training and testing sets [12] [13]. For instance, applying standard scaling before splitting would leak the overall data distribution, leading to overestimated performance. This concern is particularly important with SMOTE, as it synthesizes new data from existing data, amplifying the risk of overestimation compared to other transformations like one-hot encoding.

To maximize model performance, we made all preprocessing steps optional, allowing cross-validation to choose the best combination. This ensured that preprocessing steps would not negatively impact the model.

**C. Cross-Validation**

Cross-validation was used to select the model, hyperparameters, and preprocessing steps.

**1) Model Selection:** We considered several machine learning models, including Random Forest, XGBoost, MLP Classifier, AdaBoost, LGBM Classifier, and CatBoost Classifier. These models were chosen for their strong performance on medium-sized datasets, in contrast to deep learning approaches which are more computationally intensive.

**2) Hyperparameter Selection:** For each model, we optimized the five most critical hyperparameters, with ranges selected based on best practices and mainstream recommendations. For other hyperparameters, we left them at default values.

**3) Preprocessing Steps Selection:** All four preprocessing steps (standard scaling, one-hot encoding, feature selection, and SMOTE) were treated as optional, resulting in 16 possible combinations. This allowed for maximum flexibility in the preprocessing pipeline.

**4) Cross-Validation Optimization Algorithm Selection:** We used the Tree-structured Parzen Estimator (TPE), the default algorithm in Optuna, for hyperparameter optimization. TPE is a sequential model-based optimization (SMBO) method that models the objective function using probabilistic distributions, focusing the search on promising regions of the hyperparameter space more efficiently than grid or random search.

**5) Cross-Validation Performance Metric Selection:** The performance of the models was evaluated using the area under the receiver operating characteristic curve (AUC). AUC was chosen as it provides a more nuanced understanding of model performance compared to metrics like accuracy or F1-score. In predicting nurse transfer intentions, AUC is particularly

useful because it allows for adjusting the decision threshold to balance false positives and false negatives based on their relative costs, making it the most suitable metric for this task.

**D. Training and Testing**

Based on cross-validation results, we selected the best model, hyperparameters, and preprocessing steps. The final model was trained on the full training set and evaluated against the test set to generate performance metrics.

**E. SHAP feature Importance Analysis**

**1) Introduction:** Feature importance analysis is a crucial aspect of machine learning, enabling researchers and practitioners to understand the impact of individual features on model predictions. Among the various methods available for feature importance analysis, Shapley Additive Explanation (SHAP) has gained significant attention due to its unique approach grounded in cooperative game theory. SHAP values provide a consistent and interpretable measure of feature importance by attributing the prediction of an instance to its features based on their contributions.

**2) Comparison with Other Feature Importance Algorithms:**

Various algorithms exist for feature importance analysis, each with its strengths and weaknesses. Traditional methods like permutation importance evaluate feature importance by measuring the change in model performance when the values of a feature are randomly shuffled. While intuitive, this approach can be sensitive to correlated features and may not provide consistent results across different models.

Another popular method is feature importance derived from tree-based models, such as Gini importance or Mean Decrease Impurity. However, these methods often suffer from biases related to feature correlation and are model-specific, making them less applicable across different algorithms.

In contrast, SHAP values offer several advantages over these methods. First, SHAP is model-agnostic, meaning it can be applied to any machine learning model, enhancing its versatility. Second, SHAP values ensure a fair distribution of importance among correlated features, providing a more accurate representation of feature contributions. This is particularly advantageous in complex datasets where features may interact or be correlated, as it allows for a clearer understanding of their individual impacts.

**3) Rationale for Choosing SHAP:** In our analysis, we chose to use SHAP due to its ability to provide meaningful insights into model predictions while ensuring interpretability. As our study aims to predict the transfer tendency of specialty nurses, understanding the influence of each feature on individual predictions is essential for deriving actionable insights. The use of SHAP allows us to not only gauge overall feature importance but also to explore the contribution of each feature to specific predictions, making our findings more actionable for hospital administrators and policymakers.

**4) How to Interpret Individual SHAP Values:** An individual SHAP value represents the contribution of a specific feature to the prediction of a particular instance, relative to the average prediction across the dataset. Positive SHAP values indicate that the feature increases the prediction, while negative SHAP values suggest a decrease. The magnitude of the SHAP value quantifies the strength of this contribution. For instance, if a feature has a SHAP value of +0.3 for a given prediction, it means that this feature pushes the predicted outcome higher by 0.3 units compared to the average prediction. This level of granularity enables stakeholders to identify not just which features are important, but also how they influence individual decisions, fostering a deeper understanding of the underlying dynamics in the data.

**V. RESULT**

**A. Dataset**

A total of 471 questionnaires were collected from specialist nurses. The demographic information was processed based on different dependent variables, and detailed results are provided in Appendix B for reference.

**B. Model and Hyperparameter Selection Result**

Through cross-validation of all candidate machine learning models, the MLPClassifier emerged as the best-performing model. The Area Under the Curve (AUC) results for each model, using their optimal hyperparameters, are summarized in Table I. The optimal hyperparameter values for the MLPClassifier are presented in Table II. The best hyperparameter configurations for the other models can be found in Appendix C.

**TABLE I**

**AUC SCORES FOR CANDIDATE MODELS**

|  |  |
| --- | --- |
| Model Name | AUC |
| Random Forest Classifier | 0.708 |
| XGBoost | 0.691 |
| MLP Classifier | 0.720 |
| AdaBoost | 0.672 |
| LGBM Classifier | 0.713 |
| CatBoost Classifier | 0.701 |

**TABLE II**

**BEST HYPERPARAMETER VALUES FOR MLPCLASSIFIER**

|  |  |
| --- | --- |
| Hyperparameter Name | Value |
| Standard scalar enabled | False |
| One-hot encoder enabled | True |
| Feature selection enabled | False |
| SMOTE enabled | True |
| hidden\_layer\_sizes | 50 |
| activation | ‘logistic’ |
| solver | ‘adam’ |
| alpha | 0.003155 |
| learning\_rate | ‘constant’ |
| max\_iter | 102 |

**C. Test Result**

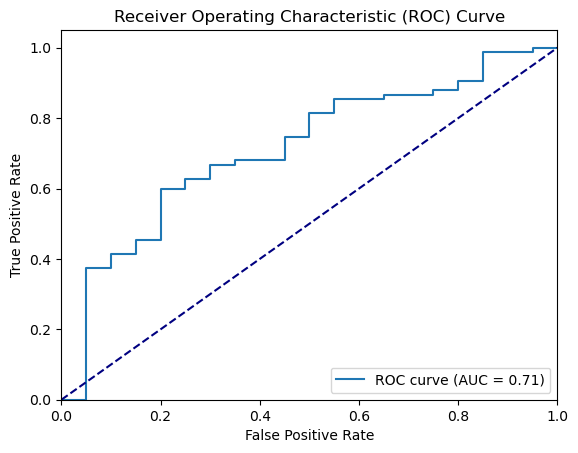
After selecting the MLPClassifier as the final model, we trained it using the hyperparameters listed above. The model was then tested against a reserved test set, and its performance was evaluated using various metrics. The results are summarized in Table III.

**TABLE III**

**PERFORMANCE METRIC VALUES OF THE TRAINED MODEL**

|  |  |
| --- | --- |
| Metric Name | Value |
| AUC | 0.713 |
| F1 Score | 0.761 |
| Sensitivity | 0.680 |
| Accuracy | 0.663 |
| Specificity | 0.600 |
| Youden’s J Score | 0.280 |

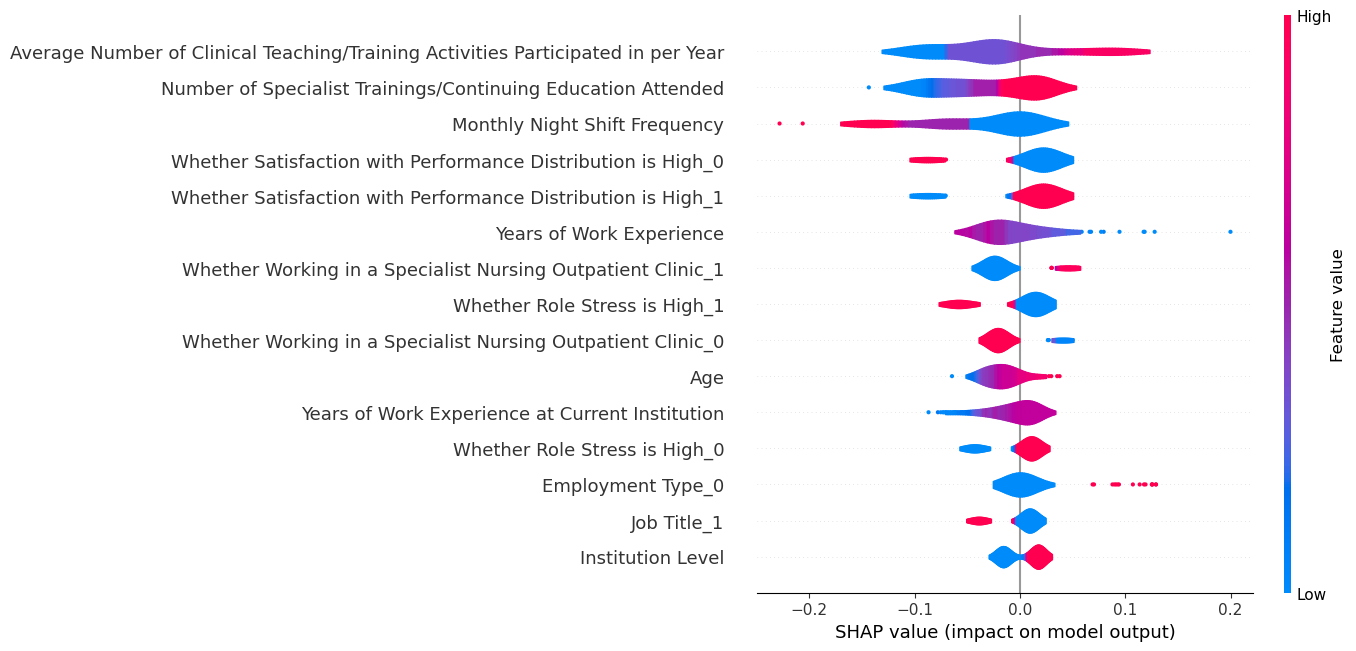
The ROC curve is shown in Figure 2.



**Fig. 2. The Receiver Operating Characteristic Curve of the Final Model**

**D. SHAP Feature Importance Analysis Result**

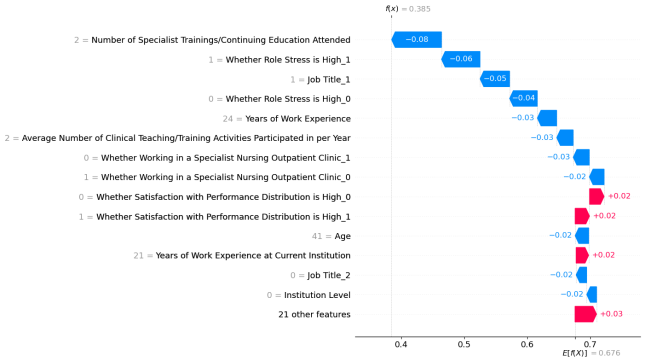
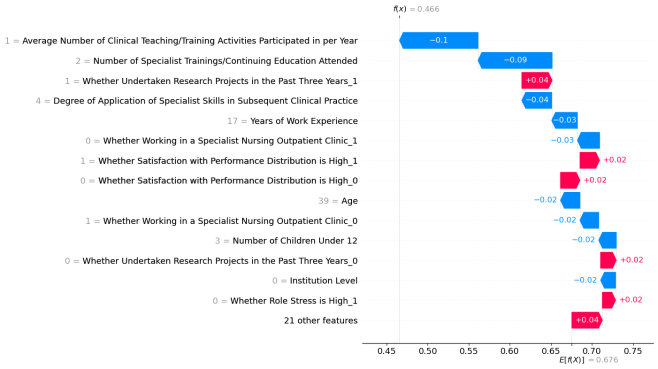
We calculated the SHAP values of the test data using the trained MLPClassifier model. The resulting visualizations include a summary violin plot (Figure 3), a bar plot (Figure 4), and four waterfall Plots (Figure 5) corresponding to the first four test samples.

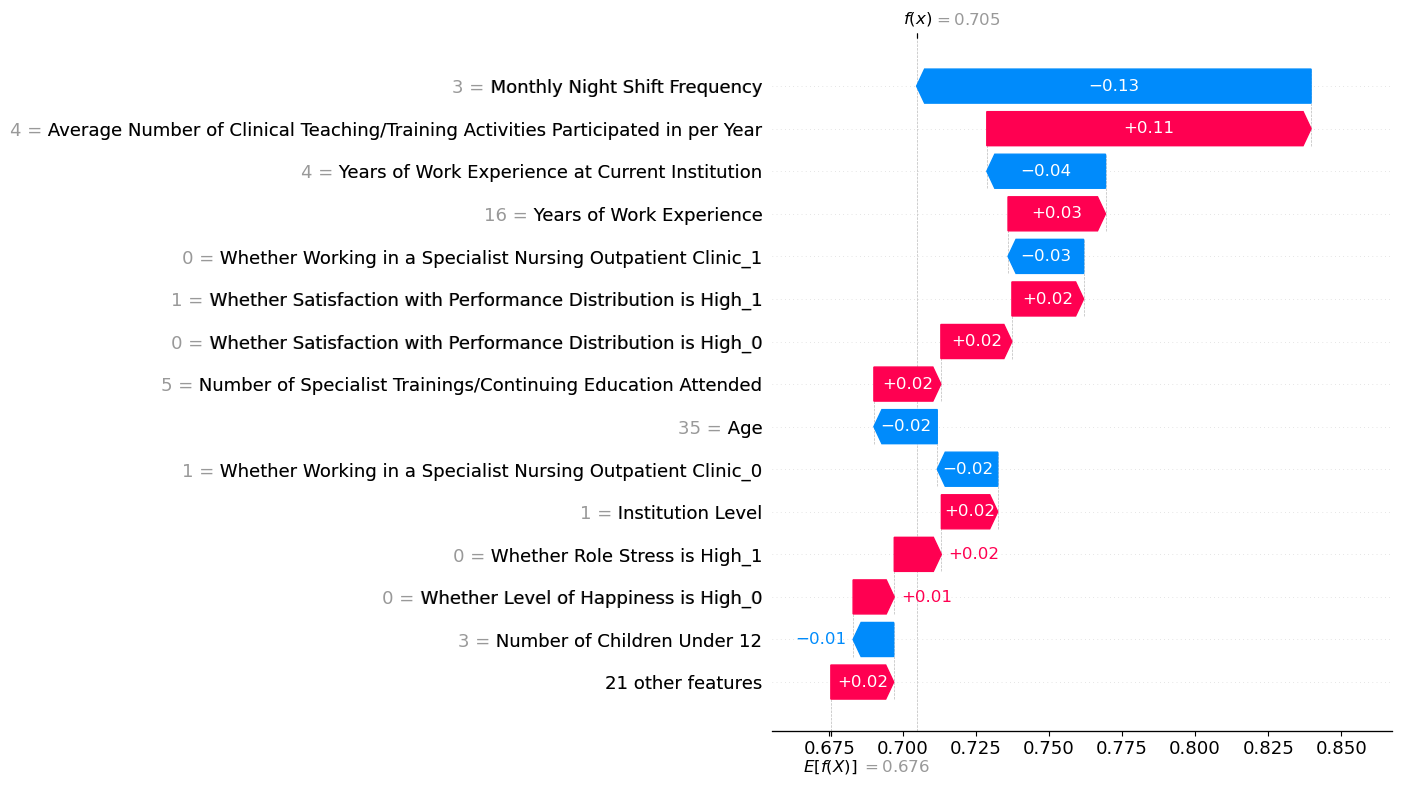
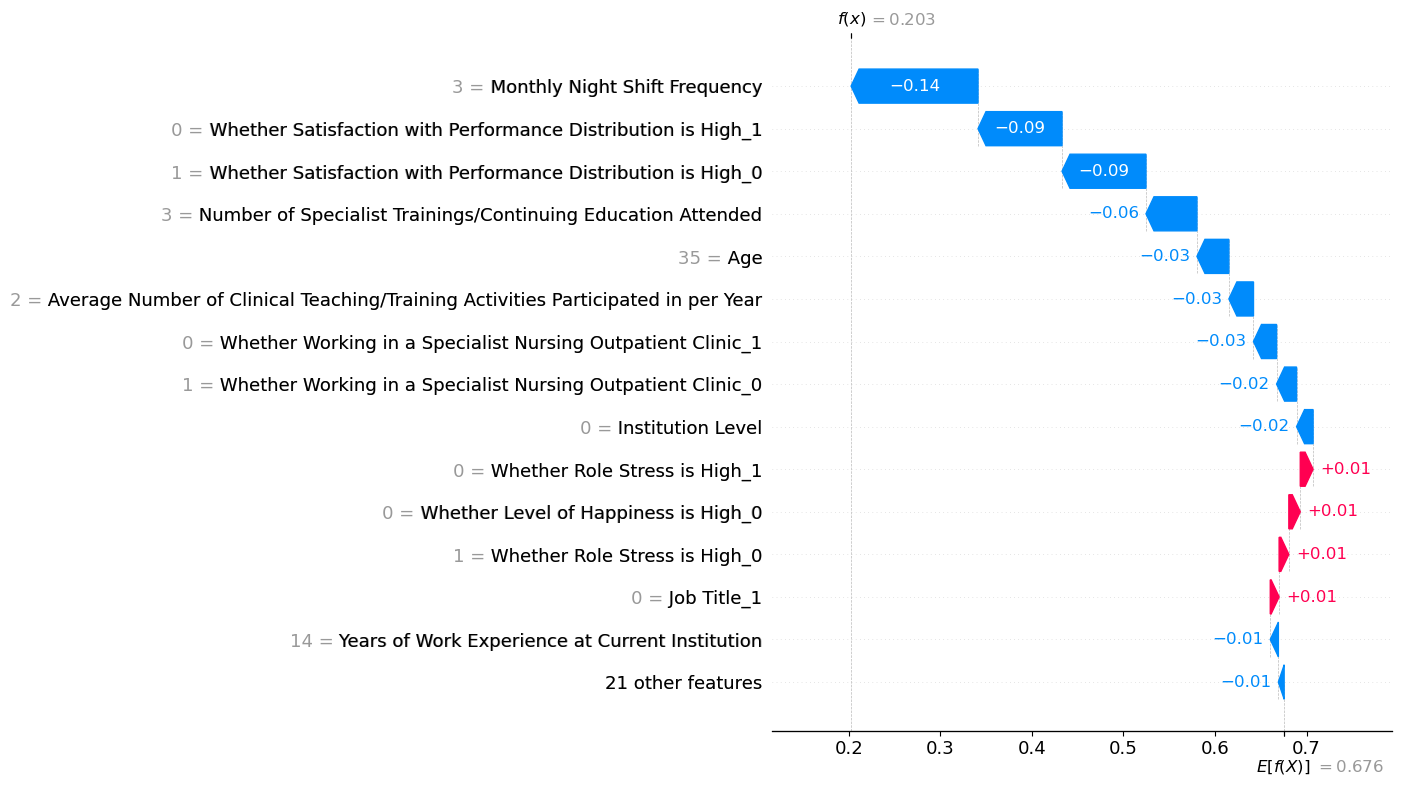


**Fig. 3. The summary violin plot of SHAP values of the trained model**



**Fig. 4. The bar plot of SHAP values of the trained model**





**Fig. 5. SHAP waterfall plots for the first 4 test samples**

**1) Key Features:** From the violin and bar plots, we observe that the five most important features influencing the model are:

• Average Number of Clinical Teaching/Training Activities Participated in per Year

• Number of Specialist Trainings/Continuing Education Attended

• Monthly Night Shift Frequency

• Satisfaction with Performance Distribution

• Years of Work Experience

**2) Interpreting the SHAP Violin Plot:** The SHAP violin plot highlights the relative importance of each feature and its impact on the model’s output. The y-axis labels display each feature’s name, with features marked by “\_\*” indicating one-hot-encoded categorical variables. For binary categorical features, those ending in “\_1” indicate a ”true” evaluation, while those ending in “\_0” indicate ”false.”

The violins are color-coded: red for high feature values, blue for low. Red colors to the right of the central gray line show features with a positive correlation to the output, while blue colors to the left indicate negative correlations. For example, the feature “Satisfaction with Performance Distribution\_1” has red values on the right and blue values on the left, indicating that high values increase the likelihood of job transfer, while low values reduce it.

**3) Interpreting the SHAP Waterfall Plot:** In the waterfall plot, arrows represent the impact of features on the model’s prediction, starting from the base value. Red arrows pointing to the right indicate a positive contribution to the prediction, while blue arrows pointing to the left indicate a negative contribution. The arrow length reflects the magnitude of each feature’s contribution.

**4) Correlations of Key Features:** The SHAP violin plot shows that “Average Number of Clinical Teaching/Training Activities,” “Number of Specialist Trainings,” and “Satisfaction with Performance Distribution” positively correlate with job transfer, while “Monthly Night Shift Frequency” and “Years of Work Experience” negatively correlate.

Some results are counterintuitive. For instance, nurses who are satisfied with performance distribution are expected to remain in their current positions. However, the model suggests the opposite. This anomaly could be due to factors such as sarcastic responses from participants or biases in the data, as most respondents reported high satisfaction. Further investigation is needed to verify these conclusions.

**VI. DISCUSSION**

**A. Controversy in Data Preprocessing**

We followed a standard data preprocessing workflow within the cross-validation stage, including oversampling to balance the class distribution. This approach prevents data leakage and avoids overestimating the model’s performance [12] [13]. However, given the strong class imbalance, this might underestimate the model’s performance. Some studies support oversampling before the train-test split when dealing with biased data [14], yielding better results with higher AUC scores (above 0.9). However, no concrete justification or guidelines exist for this approach. We opted for the standard method to avoid potential data leakage from the synthetic data created during SMOTE, acknowledging a possible slight underestimation of the model’s performance.

**B. Performance Metric Evaluation**

Our final model achieved an AUC score of 0.713, lower than expected. This could be attributed to class imbalance or weak correlations between selected features and the target variable. All candidate models reached an AUC ceiling around 0.7, suggesting data quality may be limiting performance. Future improvements could involve collecting higher-quality data and refining feature selection.

**C. SHAP Analysis Anomalies**

The SHAP analysis revealed some unexpected correlations, such as nurses satisfied with performance distribution being more likely to transfer. This could stem from the class imbalance, where the model learns incorrect patterns due to insufficient information about minority classes. Alternatively, the correlations may be accurate, reflecting deeper, unexplored factors. Further research is needed to investigate these anomalies.

**VII. CONCLUSION**

In this study, we developed a machine learning model to predict the transfer tendencies of specialist nurses using various classification algorithms. We employed feature importance analysis using SHAP values to better understand the impact of each feature on the model’s predictions. Through cross-validation and hyperparameter optimization, the MLPClassifier

demonstrated moderate performance, achieving an AUC score of 0.713.

Despite the performance limitations, the feature importance analysis highlighted key predictors of nurse transfer intentions, including clinical teaching participation, specialist training attendance, and job satisfaction. However, some findings were counterintuitive, such as the positive correlation between satisfaction with performance distribution and transfer likelihood. These anomalies could be due to class imbalances, potential biases in the data, or underlying factors not fully captured in our study.

Our investigation into the data preprocessing workflow emphasized the importance of adhering to standard practices to avoid data leakage, despite the risk of underestimating model performance in highly imbalanced datasets. While oversampling before splitting the train/test data could potentially yield higher AUC scores, we prioritized a more conservative approach to maintain the integrity of the evaluation process.

Moving forward, future work should focus on collecting more representative and balanced datasets, refining feature selection, and exploring the deeper reasons behind unexpected correlations in SHAP analysis. Such efforts would help improve the model’s accuracy and provide more actionable insights for hospitals to manage nurse retention more effectively. By addressing these challenges, we aim to contribute to reducing the specialist nurse shortage through proactive workforce management strategies.

**REFERENCES**

[1] W. H. Organization et al., “Global strategy on human resources for health: workforce 2030,” 2016.

[2] OECD, Health at a Glance 2023, 2023. [Online]. Available: <https://www.oecd-ilibrary.org/content/publication/7a7afb35-en>

[3] P. Toyin-Thomas, P. Ikhurionan, E. E. Omoyibo, C. Iwegim, A. O. Ukueku, J. Okpere, U. C. Nnawuihe, J. Atat, U. Otakhoigbogie, E. V. Orikpete, F. Erhiawarie, E. O. Gbejewoh, U. Odogu, I. C. G. Akhirevbulu, Y. K. Kwarshak, and O. Wariri, “Drivers of health workers’ migration, intention to migrate and non-migration from low/middle-income countries, 1970–2022: a systematic review,” BMJ Global Health, vol. 8, no. 5, 2023. [Online]. Available: <https://gh.bmj.com/content/8/5/e012338>

[4] T. Shuai, Y. Xuan, M. F. Jim´enez-Herrera, L. Yi, and X. Tian, “Moral distress and compassion fatigue among nursing interns: a cross-sectional study on the mediating roles of moral resilience and professional identity,” BMC nursing, vol. 23, no. 1, p. 638, 2024.

[5] W. H. O. R. O. for Europe, “Building resilience: a key pillar of health 2020 and the sustainable development goals: examples from the who european region,” World Health Organization, Tech. Rep. WHO-EURO-2018-2374-42129-58024, 2018, accessed: 2024-10-16. [Online]. Available: <https://iris.who.int/bitstream/handle/10665/340820/>WHO-EURO-2018-2374-42129-58024-eng.pdf?sequence=1

[6] S.-K. Kim, E.-J. Kim, H.-K. Kim, S.-S. Song, B.-N. Park, and K.-W. Jo, “Development of a nurse turnover prediction model in korea using machine learning,” Healthcare, vol. 11, no. 11, 2023. [Online]. Available: https://www.mdpi.com/2227-9032/11/11/1583

[7] E. Ayalew, Y. Workineh, A. Semachew, T. Woldgiorgies, S. Kerie, H. Gedamu, and B. Zeleke, “Nurses’ intention to leave their job in sub-saharan africa: A systematic review and meta-analysis,” Heliyon, vol. 7, no. 6, 2021.

[8] Y. Wang and H. Yuan, “What is behind high turnover: a questionnaire survey of hospital nursing care workers in shanghai, china,” BMC Health Services Research, vol. 18, pp. 1–7, 2018.

[9] A. Dewanto and V. Wardhani, “Nurse turnover and perceived causes and consequences: a preliminary study at private hospitals in indonesia,” BMC nursing, vol. 17, pp. 1–7, 2018.

[10] P. Kumar, S. B. Gaikwad, S. T. Ramya, T. Tiwari, M. Tiwari, and B. Kumar, “Predicting employee turnover: A systematic machine learning approach for resource conservation and workforce stability,” Engineering Proceedings, vol. 59, no. 1, 2023. [Online]. Available: https://www.mdpi.com/2673-4591/59/1/117

[11] J. Zhang, L. Xia, Y. Wang, and et al., “Predictive factors of turnover intention of newly graduated nurses in their first year of employment: a longitudinal study,” BMC Nursing, vol. 23, p. 522, 2024.

[12] J. Brownlee, “How to avoid data leakage when performing data preparation,” 2020, accessed: 2024-10-16. [Online]. Available: https://machinelearningmastery.com/data-leakage-machine-learning/

[13] M. Santos, J. Soares, P. Henriques Abreu, H. Araujo, and J. Santos, “Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches,” IEEE Computational Intelligence Magazine, vol. 13, pp. 59–76, 10 2018.

[14] M. Mujahid, E. Kına, F. Rustam et al., “Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering,” Journal of Big Data, vol. 11, p. 87, 2024. [Online]. Available: <https://doi.org/10.1186/s40537-024-00943-4>

**APPENDIX A**

**FILE RESOURCES**

The github repo https://github.com/LuoZheng2002/nurse\_transfer contains the raw questionnaire, the raw collected data and the complete python code for model development and SHAP analysis.

**APPENDIX B**

**DEMOGRAPHIC INFORMATION OF SPECIALIZED NURSES**

See Table D.

**TABLE D**

**DEMOGRAPHIC INFORMATION OF SPECIALIZED NURSES**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Dependent variable** | **Group** | **Job transfer** | **No job transfer** | **Chi2/t/Z** | **P-value** |
| Institution Level | Primary | 203(55.3) | 50(48.1) | 1.707 | 0.191 |
| Secondary | 164(44.7) | 54(51.9) |
| Age | — | 367(22-55) | 104(22-52) | 1.927 | 0.055 |
| Gender | Male | 7(1.9) | 6(5.8) | 4.503 | 0.034 |
| Female | 360(98.1) | 98(94.2) |
| Number of Children Under 12 | 1 | 170(46.3) | 45(43.3) | 0.892 | 0.849 |
| 2 | 159(43.3) | 47(45.2) |
| 3 | 37(10.1) | 12(11.5) |
| 4 | 1(0.3) | 0(0.0) |
| Years of Work Experience | — | 367(0-33) | 104(1-36) | -1.264 | 0.206 |
| Years of Work Experience at Current Institution | — | 367(0-36) | 104(0-33) | 1.653 | 0.099 |
| First Degree | Technical Secondary School | 201(54.8) | 53(51.0) | 1.970 | 0.582 |
| Associate Degree | 128(34.9) | 43(41.3) |
| Bachelor’s Degree | 37(10.1) | 8(7.7) |
| Master’s Degree | 1(0.3) | 0(0.0) |
| Highest Degree | Technical Secondary School | 2(0.5) | 0(0.0) | 0.533 | 1.000 |
| Associate Degree | 59(16.1) | 17(16.3) |
| Bachelor’s Degree | 303(82.6) | 86(82.7) |
| Master’s Degree | 3(0.8) | 1(1.0) |
| Job Title | Junior | 101(27.5) | 34(32.7) | 1.186 | 0.553 |
| Associate Senior | 257(70.0) | 67(64.4) |
| Senior | 9(2.5) | 3(2.9) |
| Position | Clinical Nursing Position (Direct Nursing Services) | 63(17.2) | 8(7.7) | 5.929 | 0.052 |
| Nursing Management Position (Hospital Nursing Management Work) | 284(77.4) | 91(87.5) |
| Other Nursing Positions (Non-Direct Nursing Services) | 20(5.4) | 5(4.8) |
| Monthly Night Shift Frequency | No Night Shifts | 235(64.0) | 52(50.0) | 7.119 | 0.057 |
| 0-5 Times | 87(23.7) | 34(32.7) |
| 5-10 Times | 41(11.2) | 17(16.3) |
| >10 Times | 4(1.1) | 1(1.0) |
| Employment Type | Staff (Formal Employment) | 313(85.3) | 97(93.3) | 4.581 | 0.032 |
| Labor Dispatch | 54(14.7) | 7(6.7) |
| Monthly Take-home Income | ≤5000 | 11(3.0) | 3(2.9) | 11.082 | 0.020 |
| 5001~8000 | 166(45.2) | 64(61.5) |
| 8001~12000 | 166(45.2) | 29(27.9) |
| 12001~15000 | 19(5.2) | 7(6.7) |
| >15000 | 5(1.4) | 1(1.0) |
| Number of Specialist Trainings/Continuing Education Attended | 0 | 17(4.6) | 6(5.8) | 8.019 | 0.090 |
| 1 | 98(26.7) | 40(38.5) |
| 2 | 80(21.8) | 24(23.1) |
| 3 | 34(9.3) | 8(7.7) |
| 4 Times or More | 138(37.6) | 26(25.0) |
| Whether Obtained Specialist Qualification Certificate and the Specialist Field | Yes | 367 | 104 | — | — |
| No | 0 | 0 |
| Whether Working in a Specialist Nursing Outpatient Clinic | Yes | 106(28.9) | 20(19.2) | 3.853 | 0.059 |
| No | 261(71.1) | 84(80.8) |
| Average Number of Clinical Teaching/Training Activities Participated in per Year | None | 99(27.0) | 36(34.6) | 6.508 | 0.089 |
| 1-5 Times | 174(47.4) | 53(51.0) |
| 5-10 Times | 36(9.8) | 7(6.7) |
| >10 Times | 58(15.8) | 8(7.7) |
| Degree of Application of Specialist Skills in Subsequent Clinical Practice | Always Used | 161(43.9) | 35(33.7) | 5.993 | 0.11 |
| Frequently Used | 137(37.3) | 40(38.5) |
| Occasionally Used | 57(15.5) | 22(21.2) |
| Not Used | 12(3.3) | 7(6.7) |
| Whether Undertaken Research Projects in the Past Three Years | Yes | 34(9.3) | 4(3.8) | 3.207 | 0.1 |
| No | 333(90.7) | 100(96.2) |
| Whether Role Stress is High | Yes | 93(25.3) | 44(42.3) | 11.310 | 0.001 |
| No | 274(74.7) | 60(57.7) |
| Whether Empathy Level is High | Yes | 358(97.5) | 101(97.1) | 0.000 | 1.000 |
| No | 9(2.5) | 3(2.9) |
| Whether Level of Happiness is High | Yes | 338(92.1) | 82(78.8) | 14.739 | 0.000 |
| No | 29(7.9) | 22(21.2) |
| Whether Satisfaction with Performance Distribution is High | Yes | 315(85.8) | 62(59.6) | 34.864 | 0.000 |
| No | 52(14.2) | 42(40.4) |

**APPENDIX C**

**BEST HYPERPARAMETER CONFIGURATIONS FOR ALL CANDIDATE MODELS**

See Table E.

**TABLE E**

**BEST HYPERPARAMETER CONFIGURATIONS FOR ALL CANDIDATE MODELS**

|  |  |  |
| --- | --- | --- |
| Model | Hyperparameter Name | Value |
| RandomForestClassifier | Standard Scalar Enabled | TRUE |
| One-Hot Encoder Enabled | FALSE |
| Feature Selection Enabled | FALSE |
| SMOTE Enabled | TRUE |
| n\_estimators | 255 |
| max\_depth | 10 |
| min\_samples\_split | 6 |
| min\_samples\_leaf | 1 |
| max\_features | 'sqrt' |
| XGBClassifier | Standard Scalar Enabled | TRUE |
| One-Hot Encoder Enabled | TRUE |
| Feature Selection Enabled | FALSE |
| SMOTE Enabled | TRUE |
| n\_estimators | 213 |
| max\_depth | 13 |
| learning\_rate | 0.00716 |
| subsample | 0.623 |
| colsample\_bytree | 0.412 |
| MLPClassifier | Standard Scalar Enabled | FALSE |
| One-Hot Encoder Enabled | TRUE |
| Feature Selection Enabled | FALSE |
| SMOTE Enabled | TRUE |
| hidden\_layer\_sizes | 50 |
| activation | 'logistic' |
| solver | 'adam' |
| alpha | 0.00316 |
| learning\_rate | 'constant' |
| max\_iter | 102 |
| LGBMClassifier | Standard Scalar Enabled | FALSE |
| One-Hot Encoder Enabled | FALSE |
| Feature Selection Enabled | TRUE |
| SMOTE Enabled | TRUE |
| n\_estimators | 716 |
| learning\_rate | 0.0294 |
| num\_leaves | 31 |
| max\_depth | 16 |
| min\_data\_in\_leaf | 33 |
| AdaBoostClassifier | Standard Scalar Enabled | FALSE |
| One-Hot Encoder Enabled | TRUE |
| Feature Selection Enabled | FALSE |
| SMOTE Enabled | TRUE |
| n\_estimators | 182 |
| learning\_rate | 0.633 |
| algorithm | 'SAMME' |
| CatBoostClassifier | Standard Scalar Enabled | TRUE |
| One-Hot Encoder Enabled | TRUE |
| Feature Selection Enabled | FALSE |
| SMOTE Enabled | FALSE |
| iterations | 61 |
| learning\_rate | 0.0245 |
| depth | 8 |
| l2\_leaf\_reg | 6 |
| early\_stopping\_rounds | 7 |