**Construction of a Machine Learning-Based Model for Predicting Specialist Nurse Transfer Tendencies**

**Gan Wang1,2✝, Zheng Luo3✝, Huiying Wu2, Jiaojiao Wang2, Wen Ma1, Li Luo1,2**

*1Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai, 200032, China*

*2School of Public Health, Fudan University, Shanghai, 200032, China*

*3UM-SJTU Joint Institute, Shanghai Jiao Tong University, Shanghai, 200240, China*

Correspondence should be addressed to Li Luo; [liluo@fudan.edu.cn](mailto:liluo@fudan.edu.cn)

✝These authors have contributed equally to this work

**Abstract**

**Background:** 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, as well as determining the major causes of the transfer tendencies, aiming to provide hospitals with actionable insights for retention strategies.

**Methods:** We applied multiple classification algorithms, such as Multilayer Perceptron (MLP) Classifier, Random Forest Classifier, XGBoost and TabNet Classifier. The final model achieved an area under the receiver operating characteristic curve (AUC) score of // to do after cross-validation and hyperparameter optimisation using the Tree-structured Parzen Estimator (TPE). To better understand feature contributions, we utilised Shapley additive explanations (SHAP) values to perform feature importance analyses.

**Results:** Our findings identified that key predictors of nurse transfer include // to do. However, certain results, such as the unexpected positive correlation between satisfaction with performance distribution and job transfer, revealed anomalies that are likely attributable to class imbalances or data biases. We also discuss the tradeoffs between different data preprocessing techniques, particularly regarding oversampling within cross-validation to mitigate class imbalances. **Conclusions:** While the performance of the model was moderate, the insights from the SHAP analysis provided valuable directions for future research. Hence, we recommend collecting more representative data and refining feature selection to improve model accuracy. Therefore, this research contributes to the development of predictive tools for hospital management to address the specialist nurse shortage by identifying nurses who are at risk of job transfer.

**Strengths and Weaknesses**

### Strengths:

* **Multiple model comparison**: This study explores a wide range of algorithmic approaches to optimise predictive performance by implementing various machine learning models (TabNet Classifier, Random Forest, XGBoost, MLP Classifier, AdaBoost, LGBM Classifier, and CatBoost Classifier).
* **Cross-validation**: The use of cross-validation ensures robust model selection and reduces the risk of overfitting, thus enhancing the generalisability of the results.
* **Hyperparameter tuning**: The application of Optuna’s TPE-based optimisation allows for efficient hyperparameter selection, which leads to improved model performance.
* **Handling imbalanced data**: Our dataset is highly imbalanced. Among the 471 samples collected, 77.9% (367) of the participants intended to transfer while only 22.1% (104) of the participants intended to stay. This study addresses class imbalance by incorporating the synthetic minority over-sampling technique (SMOTE). SMOTE will oversample the minority class (nurses who do not intend to transfer) until it has the same number as the majority (nurses who intend to transfer), thus ensuring the better prediction of minority cases.
* **Feature importance analysis**: The use of SHAP to interpret feature importance adds transparency and interpretability, which facilitates understanding the decision-making processes of the models.

### Weaknesses:

* **Limited interpretability in some models**: While powerful, models such as XGBoost and CatBoost can be complex, which makes it harder to interpret the underlying relationships compared with simpler models, such as logistic regression.
* **One-hot encoding scalability**: Using one-hot encoding in preprocessing can lead to high-dimensional data, especially when categorical features have many unique values, which potentially impacts computational efficiency.
* **Potential data leakage risk**: The extensive preprocessing steps, including feature selection and scaling, require careful management to avoid leakage between the training and testing data.
* **Hyperparameter tuning computational cost**: Although Optuna optimises 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 do not perfectly represent real-world nurse resignation patterns, which could introduce noise into the training process.

**I. BACKGROUND**

Specialist nurses are essential in maintaining the quality of care in hospitals. However, the supply of specialist nurses is limited, and healthcare systems can be significantly impacted when specialist nurses transfer or leave their positions. 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 Organisation (WHO) (1).

In China, the situation is particularly critical. By the end of 2023, the country had 5.63 million registered nurses, translating to only 4 nurses per 1,000 people. This is far below the Organisation for Economic Co-operation and Development (OECD) average of 9.3 nurses per 1,000 people (2) and 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 ageing population (4), an increase in chronic diseases (5), and job transfers within the nursing profession (6). Job transfers 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, the intention to transfer reaches 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 organisational 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, both of which will ultimately help mitigate the effects of staff shortages.

**A. 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 and thereby ensure continuity in patient care as well as reduce 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.

**B. Research Objectives**

This study 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 analyse the key factors that most strongly correlate with transfer tendencies. By identifying these critical factors, hospitals can use the revealed insights to create a more supportive work environment, reduce nurses’ intentions to leave, and ultimately help to mitigate the ongoing shortage of specialist nurses.

**C. 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 utilised the synthetic minority oversampling technique (SMOTE) to address class imbalance through data oversampling.

To identify the final model, we evaluated multiple classification algorithms, including the 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 visualised 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 utilising data collected from Chinese nurses, our research provides hospitals in China with a predictive model that addresses the pressing issue of nurse turnover, which is 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. Therefore, the insights gained from our analysis of the key factors influencing transfer tendencies can serve as valuable references for hospitals globally. By understanding these factors, healthcare institutions can implement targeted strategies to create a more supportive work environment and ultimately lead 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**

Recent advances in machine learning have demonstrated remarkable potential in predictive modeling across various domains, including healthcare and agriculture. For instance, Alkhammash et al. (2023) and Elshewey et al. (2023, 2025) applied hybrid models—such as optimized gradient boosting, CNN-LSTM, and MLP integrated with metaheuristic algorithms—for tasks ranging from disease diagnosis to pandemic forecasting. Although these studies differ in application, they consistently highlight the importance of feature selection, handling data imbalance, and model interpretability. These insights guide our methodological approach to predicting specialist nurses’ transfer intentions using explainable machine learning techniques(10–12).

In the domain of employee retention, Kumar et al. (2023) conducted a comprehensive analysis using various machine learning models to predict turnover(13). Their findings suggest that predictive analytics can significantly help organizations reduce the costs associated with employee attrition and replacement. Alkhammash et al. (2022) further demonstrated that hybrid ML-optimization approaches, such as the optimized LR-MARS model using a social spider optimization algorithm, can outperform conventional models in forecasting complex behavioral patterns(14).

Building on these techniques, Zhang et al. (2023) focused specifically on newly graduated nurses, using longitudinal data and regression analysis to reveal how career self-efficacy and adaptability significantly influence turnover intentions. This underscores the importance of understanding the unique career challenges faced by early-career nurses and provides practical implications for healthcare workforce management(15).

**IV. METHODS**

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

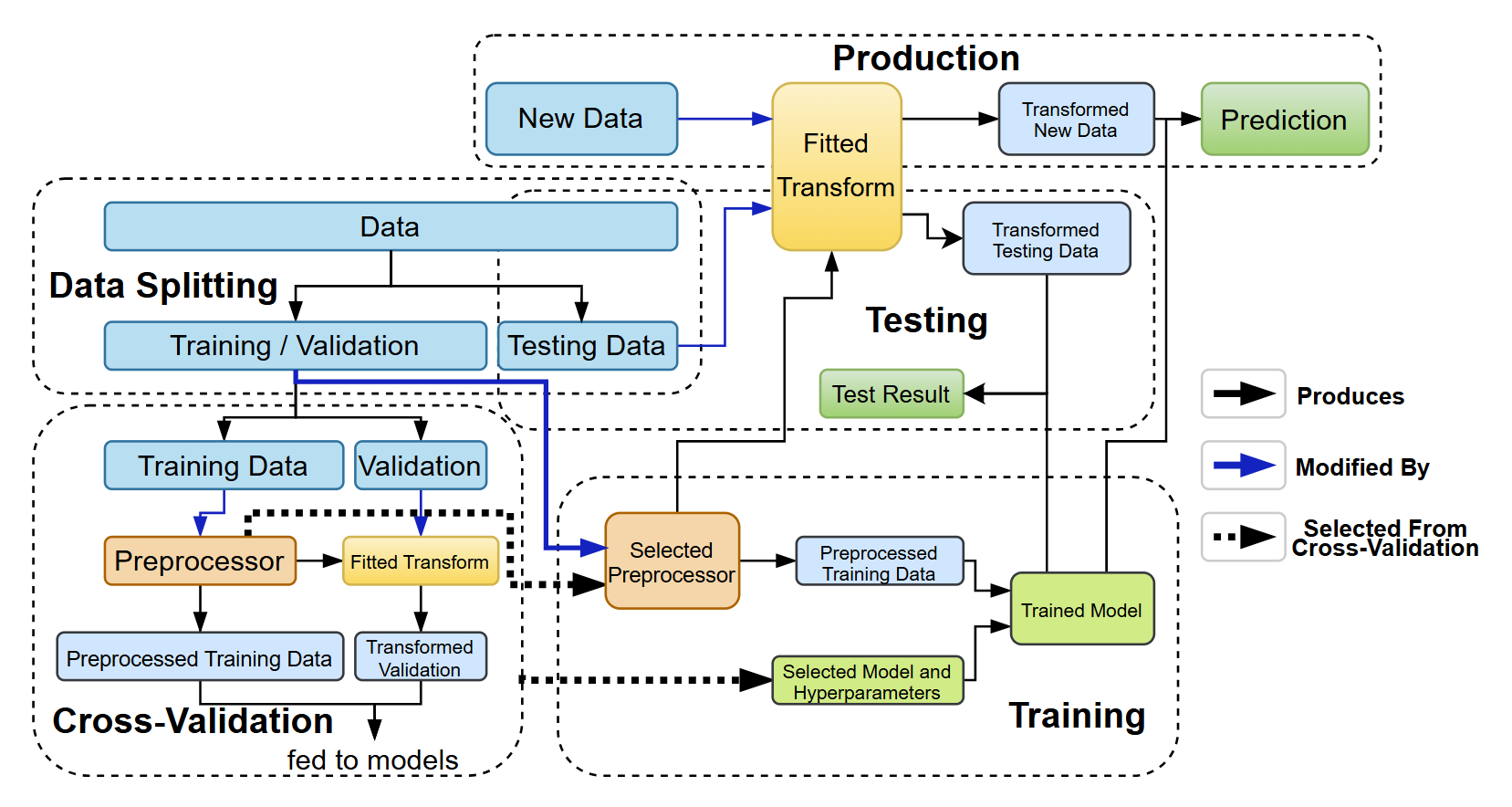
**A. Data Collection**

Data were collected from nurses in Shanghai. Anonymous questionnaires were distributed via the Wenjuanxing online platform to all district medical institutions, excluding one tertiary hospital. Inclusion criteria: 1. Have passed the nursing qualification examination and obtained certification as a Certified Nurse; 2. Obtaining the specialty competency certificate issued by the Shanghai Nursing Association; 3. Possess a secondary vocational education or higher; 4. Currently employed and actively engaged in clinical work; 5. Have at least six months of specialized work experience; 6. Nurses aged 55 or younger. Exclusion criteria: 1. Nurses who do not have a vocational secondary education or higher; 2. Individuals who have not passed the nursing qualification examination and do not hold a certified nursing qualification; 3. Nurses who are not currently employed, such as retired or unemployed nurses; 4. Nurses with less than six months of work experience or those who have not worked in a specialized field; 5. Nurses over the age of 55; 6. Nurses who are unable to engage in clinical work due to health reasons; 7. Nurses with significant mental health issues that affect their work ability; 8. Nurses who are restricted from transferring within specific medical institutions or who are subject to other professional ethics or legal constraints.

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 discretised 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. Figure 1 illustrates the overall workflow.



**Fig. 1. Machine learning workflow adopted in this study**

**B. Data Preprocessing**

The following preprocessing techniques were applied to the dataset. Since Wenjuanxing allows each question item to be designed as mandatory, no censoring value processing is required for the collected data. All questionnaires have been reviewed by three master's degree candidates majoring in public health to ensure that the filled results fall within a reasonable range. To maintain the integrity of the predictive model development process, blinding procedures were implemented to prevent any potential bias. Specifically, the assessment of the outcome to be predicted was concealed from the data analysts involved in model development. This was achieved through the anonymization of outcome labels and the assignment of a random code to each case. Additionally, the assessment of predictors for the outcome and other relevant variables was also blinded to ensure that the selection and weighting of variables were not influenced by prior knowledge of the outcome. This was accomplished by segmenting the dataset in such a way that the analysts responsible for feature selection and model training did not have access to the outcome information until the final model evaluation phase.

**1) Standard Scaling:** Standard scaling was used to normalise numeric features, ensuring that 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 such as job roles or departments.

**3) Feature Selection Via Logistic Regression and Random Forest: Feature selection is an important step in data preprocessing because it reduces model complexity and enhances interpretability.** We used two representative machine learning models as a guide to feature selection. Logistic Regression is a simple baseline model that captures individual feature contributions to the prediction result. However, Logistic Regression is vulnerable to multicollinearity of features. This means that if there are highly correlated features in the dataset, Logistic Regression may falsely determine their respective contributions. In light of this, we used Random Forest as a candidate model for feature selection since it is less sensitive to multicollinearity while maintaining simplicity. We used cross-validation to rank these feature selection models and determine which one to use.

**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 data. This technique improves the generalisability of the model by reducing its bias towards the majority class, which is an essential consideration when predicting rare outcomes, such as nurse transfer intentions.

**5) Preventing Data Leakage with the Correct Preprocessing Order: The** preprocessing steps mentioned above were applied after the train/test data split and within cross-validation to prevent data leakage. Data leakage means that a model is somehow able to peek the test set during training and overfit to it, causing the test statistics to look much better than what they should be. Data leakage can occur if the transformation uses global information from both the training and testing sets (16,17). For instance, applying standard scaling before splitting leaks the overall data distribution and leads to overestimated performance. This concern is particularly important with SMOTE, which synthesises new data from existing data and thereby amplifies the risk of overestimation compared with other transformations, such as one-hot encoding.

6) Making All Preprocessing Steps Optional for Broader Search Space: Although the preprocessing steps mentioned above are expected to benefit the model’s performance, we cannot guarantee it is indeed the case. Thus, we made all preprocessing steps optional and allow the cross-validation procedure to eliminate preprocessing steps that backfire and only keep the beneficial ones.

**C. Cross-Validation**

Cross-validation was used to select the model, hyperparameters, and preprocessing steps. It splits the training data into several segments and makes one of the segments as the validation set each time. Models with different hyperparameters are trained on the rest of the training set and metric scores like accuracy are collected on the validation set. Then the averaged metric scores are compared and the hyperparameter combination corresponding to the highest score is chosen.

**1) Model Selection:** We considered several machine learning models, including TabNet Classifier, Random Forest, XGBoost, MLP Classifier, AdaBoost, LGBM Classifier, and CatBoost Classifier. These models are chosen for their strong performance on medium sized datasets.

**2) Hyperparameter Selection: We used the Tree-structured Parzen Estimator from the Optuna Python library to automatically determine the best hyperparameters from specified ranges based on the cross-validation results. The ranges are mainly selected based on the recommendations from the API documents in the scikit-learn Python library and other model libraries. We do not emphasize too much on the range selection since the hyperparameters are automatically determined by the cross-validation procedure and will drift to the most optimal values as long as we provide a large enough range.** For each model, we optimised at most five critical hyperparameters. Other hyperparameters were left at their default values. The detailed hyperparameter range choices are as shown in table (1), (2), (3), (4), (5), (6), (7).

Table 1: Hyperparameter range choices for TabNet Classifier

|  |  |
| --- | --- |
| Parameter | Value Range |
| Width of the decision prediction layer | {4, 6, 8} |
| Width of the attention embedding | {4, 6, 8} |
| Number of steps in the architecture | {3, 4, 5} |
| Sparsity loss coefficient (lambda sparse) | {1e-4, 1e-3} |
| coefficient for feature reusage (gamma) | {1.0, 1.5} |

Table 2: Hyperparameter range choices for Random Forest Classifier

|  |  |
| --- | --- |
| Parameter | Value Range |
| Number of estimators | [10, 1000] |
| Maximum depth | [1, 30] |
| Minimum samples split | [2, 20] |
| Minimum samples leaf | [1, 10] |
| The number of features to consider | log2(n\_features) or sqrt(n\_features) |

Table 3: Hyperparameter range choices for XGBoost

|  |  |
| --- | --- |
| Parameter | Value Range |
| Number of estimators | [50, 1000] |
| Maximum depth | [3, 20] |
| Learning rate | [0.005, 0.2] |
| Subsample | [0.5, 1.0] |
| Colsample bytree | [0.3, 1.0] |

|  |  |
| --- | --- |
| Table 4: Hyperparameter range choices for MLP ClassifierParameter | Value Range |
| Hidden layer size (one hidden layer) | [5, 200] |
| Activation function | {identity, logistic, tanh, relu} |
| Solver | {lbfgs, sgd, adam} |
| Alpha (Strength of L2 regularization) | [0.0001, 0.01] |
| Learning rate | {constant, invscaling, adaptive} |

Table 5: Hyperparameter range choices for LGBM Classifier

|  |  |
| --- | --- |
| Parameter | Value Range |
| Number of estimators | [20, 1000] |
| Learning rate | [0.01, 0.2] |
| Number of leaves | [31, 1000] |
| Max depth | [2, 20] |
| Minimum data in leaf | [1, 50] |

Table 6: Hyperparameter range choices for AdaBoost Classifier

|  |  |
| --- | --- |
| Parameter | Value Range |
| Number of estimators | [50, 500] |
| Learning rate | [0.01, 1.0] |

Table 7: Hyperparameter range choices for CatBoost Classifier

|  |  |
| --- | --- |
| Parameter | Value Range |
| Iterations | [50, 200] |
| Learning rate | [0.01, 0.2] |
| Depth | [4, 8] |
| L2 leaf regularizer | [1, 10] |
| Early stopping rounds | [5, 20] |

**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 Optimisation Algorithm Selection:** We used the tree-structured Parzen estimator (TPE), which is the default algorithm in Optuna, for hyperparameter optimization. TPE is a sequential model-based optimisation method that models the objective function by using probability distributions to focus the search on promising regions of the hyperparameter space more efficiently than grid or random searches.

**5) Cross-Validation Performance Metric Selection:** The performances of the models were evaluated using the area under the receiver operating characteristic curve (AUC). The AUC was chosen because it provides a more nuanced understanding of model performance compared with metrics such as accuracy or F1-score. The AUC is particularly useful in predicting nurse transfer intentions because it allows for the adjustment of the decision threshold to balance false positives and false negatives based on their relative costs, thus making it the most suitable metric for this task.

**D. Training and Testing**

Based on the 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 that enables researchers and practitioners to understand the impacts of individual features on model predictions. Among the various methods available for feature importance analysis, the Shapley additive explanation (SHAP) has gained significant attention owing 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 of which has its own strengths and weaknesses. Traditional methods, such as 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 of feature importance analysis is derived from tree-based models, such as the 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 that it can be applied to any machine learning model, which enhances its versatility. Second, SHAP values ensure a fair distribution of importance among correlated features and thus provide a more accurate representation of the feature contributions. This is particularly advantageous in complex datasets in which 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 owing to its ability to provide meaningful insights into model predictions while ensuring interpretability. As our study aims to predict the transfer tendencies of specialist 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 explore the contribution of each feature to specific predictions, making our findings more actionable for hospital administrators and policymakers.

1. **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, whereas 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, then this feature pushes the predicted outcome higher by 0.3 units compared with the average prediction. This level of granularity enables stakeholders to identify not only which features are important but also how they influence individual decisions, thereby fostering a deeper understanding of the underlying dynamics in the data.

To ensure the rigor and transparency of the predictive model development and validation process in this study, we used the TRIPOD Checklist for verification, with details provided in ‘Appendix B-TRIPOD checklist’.

**V. RESULTS**

**A. Dataset**

A total of 523 questionnaires were distributed, and 471 valid questionnaires were collected. The demographic information was processed based on different dependent variables, and detailed results are provided in Appendix C for reference.

The dataset was highly imbalanced. Among the 471 samples, 77.9% (367) are positive (intend to transfer) and 22.1% (104) are negative (intend to stay).

We used a total of 471 samples for the training and testing purposes. Among the 471 samples, 80% (377) of the samples are used for training and validation, and the rest 20% (94) of the samples are used for testing. We used a 10-fold cross validation on the training and validation set in the model selection phase, and combined the training and validation set to train the final model.

**B. Model and Hyperparameter Selection Results**

Through the cross-validation of all candidate machine learning models, the MLPClassifier emerged as the best-performing model. The AUC results for each model, using their optimal hyperparameters, are summarised in Table 1. The optimal hyperparameter values for the MLPClassifier are presented in Table 2. The best hyperparameter configurations for the other models can be found in Appendix D.

**TABLE 1. AUC SCORES FOR CANDIDATE MODELS**

|  |  |
| --- | --- |
| **Model** | **AUC** |
| TabNet Classifier  Random Forest Classifier | 0.667  0.712 |
| XGBoost | 0.675 |
| MLP Classifier | 0.699 |
| AdaBoost | 0.668 |
| LGBM Classifier | 0.680 |
| CatBoost Classifier | 0.697 |

**TABLE 2. BEST HYPERPARAMETER VALUES FOR MLPCLASSIFIER**

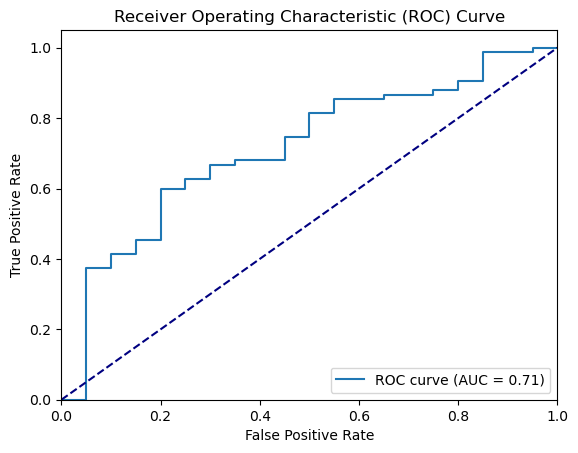
|  |  |
| --- | --- |
| **Hyperparameter** | **Value** |
| Standard scalar enabled | True |
| One-hot encoder enabled | True |
| Feature selection | None |
| SMOTE enabled | False |
| Number of estimators | 631 |
| Maxinum depth | 27 |
| Mininum sample split | 20 |
| Minimum sample leaf | 8 |
| Max features | sqrt(n\_features) |

**C. Test Results**

After selecting the Random Forest Classifier 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. Table 3 summarises the results. The predictive performance of the final model was evaluated using the Receiver Operating Characteristic (ROC) curve, as shown in Figure 2.

**TABLE 3. PERFORMANCE METRIC VALUES OF THE TRAINED MODEL**

|  |  |
| --- | --- |
| **Metric** | **Value** |
| AUC | 0.782 |
| F1-Score | 0.869 |
| Sensitivity | 0.987 |
| Accuracy | 0.768 |
| Specificity | 0.000 |
| Youden’s J Score | 0.280 |



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

The AUC curve is not considered satisfactory in a general sense since the curve does not cover most of the rectangular area in the chart, and the false positive rate increases proportionally to the TPR, which indicates that the model does not perform so well in classifying the positive class.

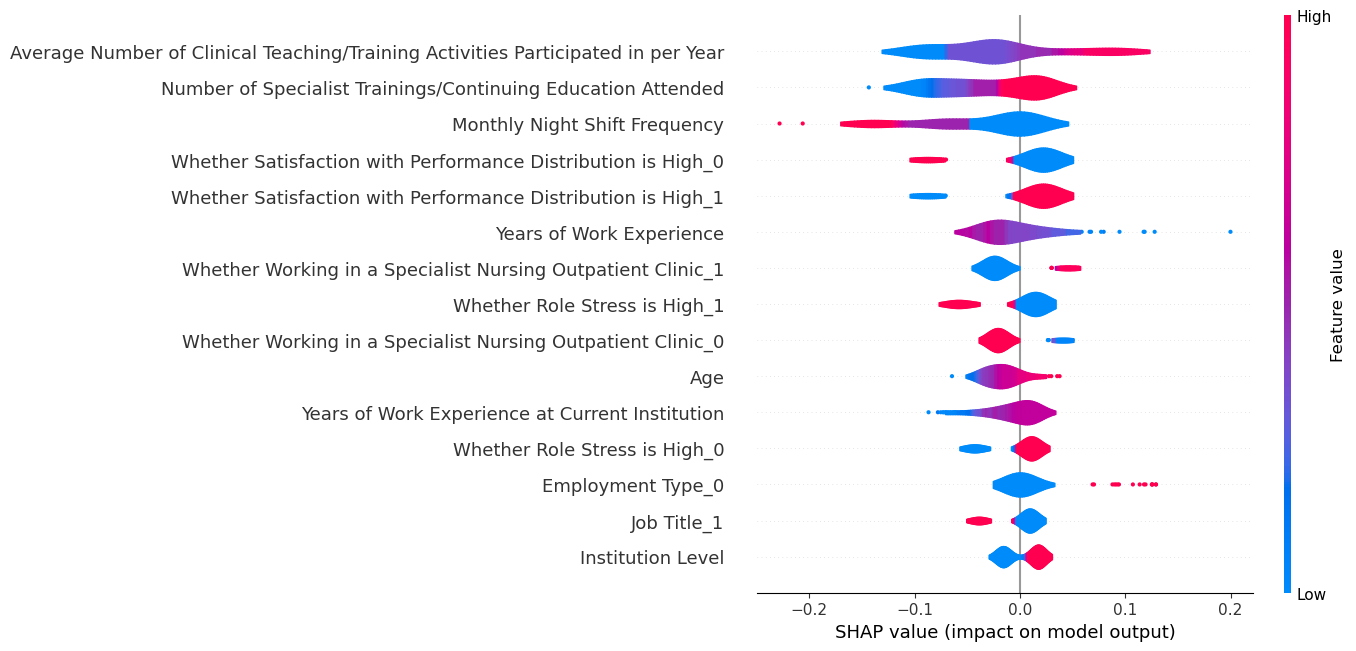
The result may be caused by several factors. First, we have a small dataset of 471 samples, which may not be sufficient for fitting a complex model. This is also shown by the fact that the simple MLP model prevailed in the cross-validation model selection phase. Another reason is that the features we collected might have a small correlation with the outcome, given that we focused mainly on nurses’ background information instead of their psychology characteristics, which may correlate with their transfer intention more. However, the goal of the project is to predict nurses’ transfer rate based solely on their background information, since we want to make predictions on a large scale without conducting interviews with the nurses one by one.

The biased dataset may also be a reason why the outcome is suboptimal. Although we used techniques like SMOTE, the lack of information about the minor class (those who do not intend to transfer) may still pose a challenge on discriminating between the two classes.

Although the performance of our final model is not considered ideal in a common sense, it is still benefitial to be put into use. The problem we want to address is to spot nurses that have the tendency to transfer, and allocate resources efficiently to prevent them from transferring using strategies like reducing their workload or giving them a raise. Therefore, with a guideline that is better than random guessing, we can move resources to places in need according to our model, potentially saving resources and reducing the transfer rate in the macroscopic level.

**D. SHAP Feature Importance Analysis Result**

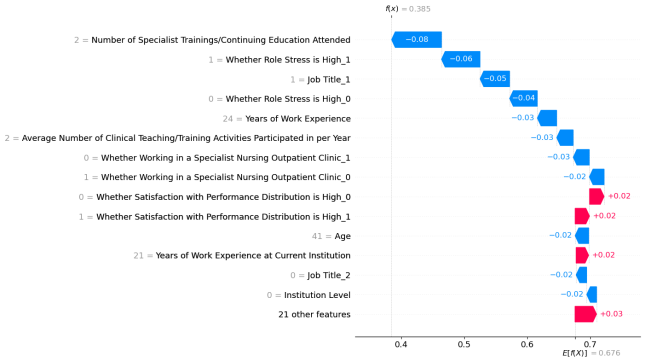
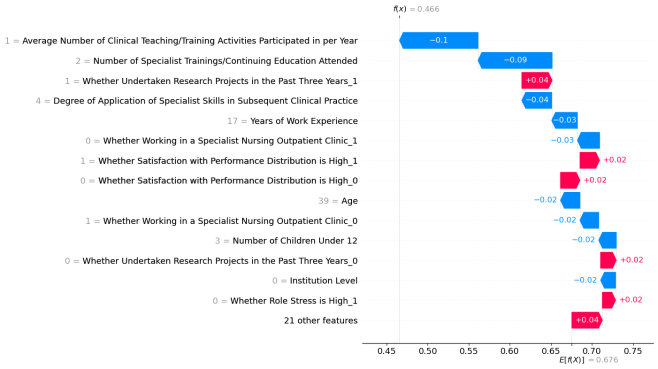
We calculated the SHAP values of the test data using the trained MLPClassifier model. The resulting visualizations include a summary violin plot (Figure 3), bar plot (Figure 4), and four waterfall plots (Figure 5) corresponding to the first four test samples.

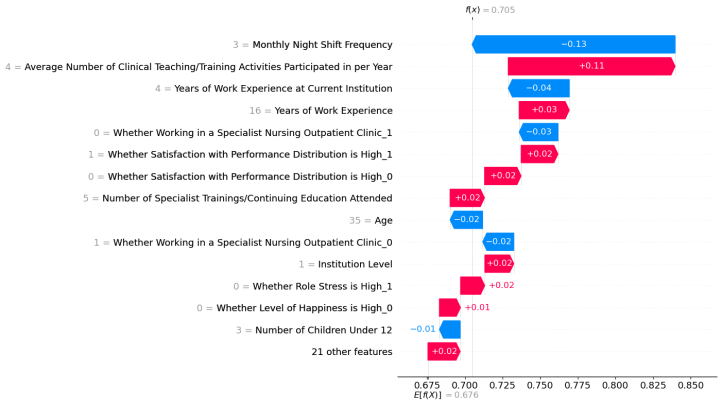


**Fig. 3. Summary Violin Plot of SHAP Values of the Trained Model**



**Fig. 4. Bar Plot of SHAP Values of the Trained Model**





**Fig. 5. SHAP Waterfall Plots for the First Four Test Samples**

**1) Key Features:** The violin and bar plots show 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 and impact on the model output of each feature. The y-axis labels display the name of each feature, with the features marked by ‘\_\*’ indicating one-hot-encoded categorical variables. For binary categorical features, those ending in ‘\_1’ indicate a ‘true’ evaluation, whereas those ending in ‘\_0’ indicate a ‘false’ evaluation.

The violins are colour-coded: red for high feature values and blue for low feature values. Red colours to the right of the central grey line show features with a positive correlation to the output, whereas blue colours 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, whereas low values reduce it.

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

**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’ correlate positively with job transfer, whereas ‘Monthly Night Shift Frequency’ and ‘Years of Work Experience’ correlate negatively.

Most of the correlations make sense. For example, nurses who undergo frequent clinical teaching and training are more willing to transfer because clinical teaching and training impose high burdens on them.

However, it is interesting to observe that one particular correlation is abnormal, which is between “whether satisfied with the performance distribution” and “whether willing to transfer”. Commonly, nurses who are satisfied with the performance distribution are likely to stay, but the model statistics indicate the opposite. The reasons might be as follows:

1. Nurses who are satisfied with their current salary are likely to perform well in the current hospital and may have greater ambitions to transfer to better hospitals.

2. Nurses who answered “not satisfied” might be honest and frank, and they are not upset enough to seek for a transfer, while those who answered “satisfied” might be easily annoyed by the verbose survey and gave dishonest and sarcastic answers, and the latter group are more likely to be upset about their current situation and seek for a transfer.

3. The feature may have loose correlation with our target (although ranked 4 in the most important features), and the noise in the dataset might easily lead the correlation to the opposite direction.

**VI. DISCUSSION**

**A. Controversy in Data Preprocessing**

We initially did data preprocessing including SMOTE oversampling technique before the cross-validation stage, which turned out to yield a significantly higher AUC score (above 0.9) than our current one, but then we realized that the synthesized data would appear in both the training and the validation set and can cause severe data leakage. Therefore, we modified our training pipeline to conduct oversampling within the cross-validation stage. That is to say, we first split the training and validation set, and then oversample the training set, leaving the validation and test set untouched. This approach prevents data leakage and avoids overestimating the performance of the model (16,17).

It is noteworthy that some studies do approve oversampling before the train–test split when dealing with biased data (18) and yielded better results with higher AUC scores (above 0.9). However, no concrete justification or guidelines exist for this approach. We opted for the reliability of our model’s performance metric and adopted the safer data preprocessing approach that strictly prevent data leakage, which yielded a performance that is significantly lower than one that could have been achieved using a bolder data preprocessing workflow like the studies mentioned above.

**B. Benchmarking Against Existing Turnover Prediction Models**

Our model's performance (AUC=0.713) aligns with established turnover prediction literature while revealing important contextual differences. Kim et al. (2023) developed a high-accuracy (98.9%) random forest model for predicting nurse turnover, identifying salary as the most critical factor, demonstrating machine learning's potential for cost-effective workforce management in healthcare systems(19). Gao et al. (2019) developed an improved random forest algorithm to predict employee turnover at a Chinese telecommunications company, achieving a prediction accuracy of 65.3% with the key factors influencing employee turnover were identified as monthly income, overtime, age, distance from home, years at the company, and percent of salary increase(20). This spectrum reflects fundamental differences in predictor stability across industries, with healthcare models typically benefiting from more structured institutional factors compared to the volatile market influences affecting corporate turnover.

**C. Performance Metric Evaluation**

A substantial body of work has been dedicated to the study, validation, and implementation of predictive models for job transfer, using machine learning models such as decision tree, logistic regression, and random forest to evaluate the predictive models(6,21,22). The final model achieved an AUC of 0.713. While this may appear modest, it reflects the model's capacity to capture complex non-linear relationships and feature interactions that traditional methods like logistic regression often overlook. By learning hierarchical representations, machine learning models can identify subtle, multifactorial patterns underlying nurse well-being and turnover risk. This highlights their advantage in modeling complex phenomena, even when conventional metrics seem comparable. This could be attributable to class imbalance or weak correlations between selected features and the target variable. All candidate models reached an AUC ceiling of approximately 0.7, suggesting that data quality may be limiting performance. Future improvements could involve collecting higher-quality data and refining feature selection. Furthermore, compared to turnover prediction models in existing literature report AUCs between 0.65 and 0.70, our approach provides a more nuanced understanding of complex, nonlinear feature interactions. This further supports the value of applying advanced machine learning methods to model turnover intentions in complex healthcare environments.

**C. SHAP Analysis Anomalies**

The SHAP analysis revealed unexpected correlations, such as nurses who were satisfied with performance distribution being more likely to transfer. Such results could stem from class imbalance, where the model learns incorrect patterns owing to insufficient information regarding minority classes. Alternatively, the correlations may be accurate and may reflect deeper, unexplored factors. Further research is needed to investigate these anomalies.

**D. Factors Influencing Job Transfer**

Similar to the findings of this study, most previous research identified specialist training as a major factor in job transfer (23,24). Nurses with insufficient education and training are unfamiliar with the job; hence, they can experience excessive workload and work stress, requiring that hospitals provide sufficient training to help staff become capable of handling their jobs. Monthly night shift frequency was considered the third most important factor in nurses’ job transfer, with a mean SHAP value of 0.170. In a study that predicted turnover among tertiary hospital nurses, nurses frequently worked night shifts, making sleep disorders the most fundamental problem in occupational characteristics. Good sleep can guarantee that the human body has abundant energy, enabling individuals to maintain interest and enthusiasm for their work. However, most nurses are in a long-term high-pressure state, leading to their willingness to leave their jobs. Hospitals can develop flexible shift scheduling to reduce the frequency of night shifts while ensuring sufficient manpower. Some studies concentrate on the elements that affect nurses’ job transfer, including individual variables (generation, education, certification, magnet status, educational background, and self-rated health) (24–26), job-related variables (working years, working hours, income, work pressure, social recognition, and hospital administrators), and job satisfaction (27,28). Although anticipating individual job transfer outcomes is challenging, with these factors, turnover control efforts are likely to be cohesive and well-directed.

**E. Limitations**

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 the performance distribution and transfer likelihood. These anomalies could be attributable to class imbalance, potential biases in the data, or underlying factors that were not fully captured in our study. In particular, SHAP analysis revealed these surprising associations, warranting deeper investigation. For instance, The SHAP values for "Years of Work Experience" are surprisingly low or even positive across multiple test samples (from -0.03 to +0.03). This contradicts conventional assumptions that senior nurses contribute more positively due to accumulated expertise. A possible explanation is that senior nurses, while more experienced, may face heavier administrative burdens or experience greater role fatigue, thereby diminishing their apparent contribution to the outcome variable. Additionally, tenure may not correlate directly with engagement in specialist roles, diluting its predictive value. To address this, we propose (1) conducting subgroup analyses to uncover hidden patterns; (2) incorporating expert interviews or qualitative methods to validate interpretations; and (3) integrating mixed-method data to identify latent factors that may not be captured by structured survey data.

Our investigation into the data preprocessing workflow emphasised 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 prioritised a more conservative approach to maintain the integrity of the evaluation process.

The main constraint is the reliance on data from a single study location (Xuhui District), which may limit the applicability of the findings to diverse environments(29). Hence, the representativeness of the findings may be limited.

In the future, we plan to improve our model and prediction accuracy by doing the following:

1. We will collect data from multiple districts in Shanghai, especially those with high population like Huangpu District and Yangpu District. This will significantly increase the number of samples we can obtain and the dataset will describe nurses’ behaviors across different regions in a more comprehensive way.

2. We will collect and incorporate additional features like organizational culture of the hospital, cultural background, lifestyle habits, regional development level and work-life balance. This can help models to learn complex patterns and make predictions more accurately.

3. We will try different imbalance mitigating algorithms like ADASYN, which may potentially outperform the current SMOTE oversampling method.

By collecting more and better training data and adopting better algorithm, we aim to improve model’s performance and provide better insight into understanding nurses’ internal transfer tendencies. Ultimately, we aim to reduce the specialist nurse shortage through proactive workforce management strategies.

**VII. CONCLUSIONS**

This study presented a machine learning algorithm designed to predict the likelihood of specialty nurses changing positions based on features such as personal information, professional background, working conditions, education, professional development, psychological state, and job satisfaction. The feature importance analysis identified key risk factors associated with turnover intentions, such as night-shift frequency and specialist training. Based on these insights, we recommend hospital administrators to implement flexible shift scheduling to reduce night-shift frequency. Prioritize continuous specialist training. Other interventions—sleep quality programs, clearer role definitions, and enhanced supervisory support—may further reduce turnover intentions. This study contributes to data-driven decision-making, efficiency of specialty nurse human resource management, and optimization of the work environment.

**Supplement material**

This content has been supplied by the authors.

Appendix A

Appendix B

Appendix C

Appendix D

**Acknowledgements**

We would like to express our sincere gratitude to the Xuhui District Health Commission of Shanghai for their assistance in data collection for this project.

**Authors’ contributors**

WG and LZ was the primary author of the manuscript and conducted the baseline data analysis as well as the construction, prediction, and evaluation of the machine learning model. WJ and WH contributed to writing parts of the background and discussion sections. MW was involved in the baseline data analysis. LL provided the overall framework and conceptual guidance for the study.

**Funding**

This research is funded by the “Research on Reconstructing the Family Doctor Service Model in the Context of Regional Medical Care” project, which is a major research project in philosophy and social sciences launched by the Ministry of Education (Grant No. 20JZD027).

**Data availability**

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

**Declarations**

**Ethics approval and consent to participate** The Medical Research Ethics Committee of the School of Public Health at Fudan University approved Professor Luo Li's ethics application for the study titled "Investigation on the Development of Nursing Talent in Xuhui District" on September 19, 2024 (Approval No. IRB#2024-09-1161). This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Informed consent was obtained from all individual participants included in the study.

**Consent for publication**

Not applicable

**Competing interests**

The authors declare no competing interests.

**Clinical Trial Statement**

Not applicable.

**Patient and public involvement**

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Provenance and peer view**

Not commissioned; externally peer reviewed.

**REFERENCES**

1. WHO. Global strategy on human resources for health: Workforce 2030 [Internet]. 2020 [cited 2024 Nov 4]. Available from: https://www.who.int/publications/i/item/9789241511131

2. OECD. Health at a Glance 2023 [Internet]. 2023. 234 p. Available from: https://www.oecd-ilibrary.org/content/publication/7a7afb35-en

3. Toyin-Thomas P, Ikhurionan P, Omoyibo EE, Iwegim C, Ukueku AO, Okpere J, et al. Drivers of health workers’ migration, intention to migrate and non-migration from low/middle-income countries, 1970-2022: a systematic review. BMJ Glob Health. 2023 May;8(5).

4. Shuai T, Xuan Y, Jiménez-Herrera MF, Yi L, Tian X. Moral distress and compassion fatigue among nursing interns: a cross-sectional study on the mediating roles of moral resilience and professional identity. BMC Nursing. 2024 Sep 11;23(1):638.

5. Fact sheet on Sustainable Development Goals (SDGs): health targets Climate change and health.pdf [Internet]. [cited 2024 Nov 4]. Available from: https://iris.who.int/bitstream/handle/10665/340820/WHO-EURO-2018-2374-42129-58024-eng.pdf?sequence=1

6. Kim SK, Kim EJ, Kim HK, Song SS, Park BN, Jo KW. Development of a Nurse Turnover Prediction Model in Korea Using Machine Learning. Healthcare (Basel). 2023 May 28;11(11).

7. Ayalew E, Workineh Y, Semachew A, Woldgiorgies T, Kerie S, Gedamu H, et al. Nurses’ intention to leave their job in sub-Saharan Africa: A systematic review and meta-analysis. Heliyon. 2021 Jun;7(6):e07382.

8. Wang Y, Yuan H. What is behind high turnover: a questionnaire survey of hospital nursing care workers in Shanghai, China. BMC Health Serv Res. 2018 Jun 22;18(1):485.

9. Dewanto A, Wardhani V. Nurse turnover and perceived causes and consequences: a preliminary study at private hospitals in Indonesia. BMC Nursing. 2018 Dec 19;17(2):52.

10. Alkhammash EH, Assiri SA, Nemenqani DM, Althaqafi RMM, Hadjouni M, Saeed F, et al. Application of Machine Learning to Predict COVID-19 Spread via an Optimized BPSO Model. Biomimetics. 2023 Oct;8(6):457.

11. Elshewey AM, Shams MY, Tawfeek SM, Alharbi AH, Ibrahim A, Abdelhamid AA, et al. Optimizing HCV Disease Prediction in Egypt: The hyOPTGB Framework. Diagnostics. 2023 Jan;13(22):3439.

12. Alzakari SA, Alhussan AA, Qenawy AST, Elshewey AM. Early Detection of Potato Disease Using an Enhanced Convolutional Neural Network-Long Short-Term Memory Deep Learning Model. Potato Res. 2025 Mar 1;68(1):695–713.

13. Kumar P, Gaikwad SB, Ramya ST, Tiwari T, Tiwari M, Kumar B. Predicting Employee Turnover: A Systematic Machine Learning Approach for Resource Conservation and Workforce Stability. Engineering Proceedings. 2023;59(1):117.

14. Alkhammash EH, Kamel AF, Al-Fattah SM, Elshewey AM. Optimized Multivariate Adaptive Regression Splines for Predicting Crude Oil Demand in Saudi Arabia. Discrete Dynamics in Nature and Society. 2022;2022(1):8412895.

15. Zhang J, Xia L, Wang Y, Yi T, Wang Y, Zhou E, et al. Predictive factors of turnover intention of newly graduated nurses in their first year of employment: a longitudinal study. BMC Nurs. 2024 Jul 30;23(1):522.

16. Brownlee J. Data Leakage in Machine Learning - MachineLearningMastery.com [Internet]. 2020 [cited 2024 Nov 4]. Available from: https://machinelearningmastery.com/data-leakage-machine-learning/

17. Santos MS, Soares JP, Abreu PH, Araujo H, Santos J. Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches [Research Frontier]. IEEE Computational Intelligence Magazine. 2018;13(4):59–76.

18. Mujahid M, Kına E, Rustam F, Villar MG, Alvarado ES, De La Torre Diez I, et al. Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering. Journal of Big Data. 2024 Jun 17;11(1):87.

19. Kim SK, Kim EJ, Kim HK, Song SS, Park BN, Jo KW. Development of a Nurse Turnover Prediction Model in Korea Using Machine Learning. Healthcare (Basel). 2023 May 28;11(11):1583.

20. Gao X, Wen J, Zhang C. An Improved Random Forest Algorithm for Predicting Employee Turnover. Mathematical Problems in Engineering. 2019;2019(1):4140707.

21. Yuan J. Research on Employee Turnover Prediction Based on Machine Learning Algorithms. In 2021. p. 114–20.

22. Suliman M, Aljezawi M, Almansi S, Musa A, Alazam M, Ta’an WF. Effect of nurse managers’ leadership styles on predicted nurse turnover. Nurs Manag (Harrow). 2020 Sep 24;27(5):20–5.

23. Guo N, Wu X, Jiao J, Cao J, Li J. Retrospective analysis of factors associated with nurse turnover. Chinese Nursing Research. 2018;32(03):492–4.

24. Koehler T, Olds D. Generational Differences in Nurses’ Intention to Leave. West J Nurs Res. 2022 May;44(5):446–55.

25. Jiang F, Zhou H, Rakofsky J, Hu L, Liu T, Wu S, et al. Intention to leave and associated factors among psychiatric nurses in China: A nationwide cross-sectional study. Int J Nurs Stud. 2019 Jun;94:159–65.

26. Wang XX, Wang LP, Wang QQ, Fang YY, Lv WJ, Huang HL, et al. Related factors influencing Chinese psychiatric nurses’ turnover: A cross-sectional study. J Psychiatr Ment Health Nurs. 2022 Oct;29(5):698–708.

27. Labrague LJ, Gloe DS, McEnroe-Petitte DM, Tsaras K, Colet PC. Factors influencing turnover intention among registered nurses in Samar Philippines. Appl Nurs Res. 2018 Feb;39:200–6.

28. Jung HS, Yoon HH. Error management culture and turnover intent among food and beverage employees in deluxe hotels: the mediating effect of job satisfaction. Service Business. 2017 Dec 1;11(4):785–802.

29. Al-Radaideh Q, Alnagi E. Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance. International Journal of Advanced Computer Science and Applications. 2012 Feb;3.