

# Personalized prediction of diabetic foot ulcer recurrence in elderly individuals using machine learning paradigms

Shichai Hong<sup>a,1</sup>, Yihui Chen<sup>a,1</sup>, Yue Lin<sup>a</sup>, Xinsheng Xie<sup>a</sup>, Gang Chen<sup>a</sup>, Hefu Xie<sup>b</sup> and Weifeng Lu<sup>a,\*</sup>

<sup>a</sup>Department of Vascular Surgery, Zhongshan Hospital (Xiamen), Fudan University, Huli District, Xiamen, Fujian, China

<sup>b</sup>Xiamen University School of Information Science and Technology, Xiamen University Xiang'an Campus, Xiamen, Fujian, China

## Abstract.

**BACKGROUND:** This study utilizes machine learning to analyze the recurrence risk of diabetic foot ulcers (DFUs) in elderly diabetic patients, aiming to enhance prevention and intervention efforts.

**OBJECTIVE:** The goal is to construct accurate predictive models for assessing the recurrence risk of DFUs based on high-risk factors, such as age, blood sugar control, alcohol consumption, and smoking, in elderly diabetic patients.

**METHODS:** Data from 138 elderly diabetic patients were collected, and after data cleaning, outlier screening, and feature integration, machine learning models were constructed. Support Vector Machine (SVM) was employed, achieving an accuracy rate of 93%.

**RESULTS:** Experimental results demonstrate the effectiveness of SVM in predicting the recurrence risk of DFUs in elderly diabetic patients, providing clinicians with a more accurate tool for assessment.

**CONCLUSIONS:** The study highlights the significance of machine learning in managing foot ulcers in elderly diabetic patients, particularly in predicting recurrence risk. This approach facilitates timely intervention, reducing the likelihood of patient recurrence, and introduces computer-assisted medical strategies in elderly diabetes management.

Keywords: Elderly diabetic patients, machine learning, diabetic foot ulcers, recurrence risk

## 1. Introduction

Diabetic foot ulcer (DFU) [1], a severe complication of diabetes characterized by ulceration in the lower extremities or feet, significantly diminishes patients' quality of life, burdens healthcare professionals, and escalates healthcare costs. Timely diagnosis and treatment are crucial. Particularly prevalent in elderly diabetic patients, DFU contributes to an alarming 80% of diabetes-related amputations in this demographic, resulting in a mortality rate more severe than many common cancers. Current management emphasizes patient education, regular self-foot examinations, and annual diabetes foot assessments to

<sup>1</sup>Shichai Hong and Yihui Chen contributed equally to this work.

\*Corresponding author: Weifeng Lu, Department of Vascular Surgery, Zhongshan Hospital (Xiamen), Fudan University, No.668, Jinhu Road, Huli District, Xiamen, Fujian, China. Tel.: +86 15859292112; E-mail: lu.weifeng@zsmhospital.com.

identify DFU early; however, these approaches have limited efficacy in elderly patients. Additionally, DFU carries a notable risk of recurrence, and early identification and treatment do not guarantee freedom from recurrence risk.

To timely identify the recurrence risk of DFU in elderly diabetic patients and significantly reduce or prevent neglect-related DFU recurrence and severe complications, comprehensive analysis of medical data by healthcare professionals is essential. Traditional methods for DFU recurrence diagnosis involve substantial manpower and are susceptible to human errors [2]. The use of computer-aided diagnostic programs can reduce costs while improving performance. Artificial intelligence (AI) has made substantial contributions to healthcare and is gaining widespread attention. In the healthcare sector, machine learning is intelligently applied, with datasets being a primary requirement for disease prediction analysis. It finds broad applications in various fields such as healthcare, robotic vision, and facial recognition.

Clinical systems have been developed that include risk factors for DFU to categorize diabetic foot diseases [3]. However, there remains a lack of sufficient research exploring the severity of recurrence risk in elderly diabetic patients concerning factors such as alcohol consumption, smoking, and blood glucose control. This underscores the need for further investigation in this area to enhance our ability to reliably predict disease recurrence in elderly diabetic foot ulcer patients.

Our study proposes a machine learning (ML) model to predict the long-term recurrence probability of foot ulcers in elderly diabetic patients [4]. Strengthening the prediction of recurrence rates in elderly diabetic foot ulcer patients will ultimately lead to improved control of modifiable risk factors and enhance long-term prognosis [5].

## **2. Related work**

In the process of investigating the recurrence risk of diabetic foot ulcers (DFU) in elderly patients with diabetes, we referenced prior works and studies in related fields. Here are some works relevant to our research:

### *2.1. Previous studies on recurrence risk of diabetic foot ulcers*

There have been several studies focusing on the recurrence risk of diabetic foot ulcers, but most of them have emphasized patients across all age groups. Studies often employed traditional statistical methods such as regression analysis to explore the influence of patients' biological and clinical characteristics on recurrence risk [6].

### *2.2. Application of machine learning in diabetes management*

In recent years, significant progress has been made in the application of machine learning in the medical field. Some studies have attempted to use machine learning algorithms to predict the occurrence and recurrence of diabetic foot ulcers. These methods typically leverage clinical and biological data to identify patients' risk factors [7]. However, research on elderly patients with diabetes is relatively scarce and requires more attention.

### *2.3. Factors and diabetic foot ulcer recurrence in elderly patients*

Our study specifically focuses on elderly patients with diabetes and considers a range of specific factors

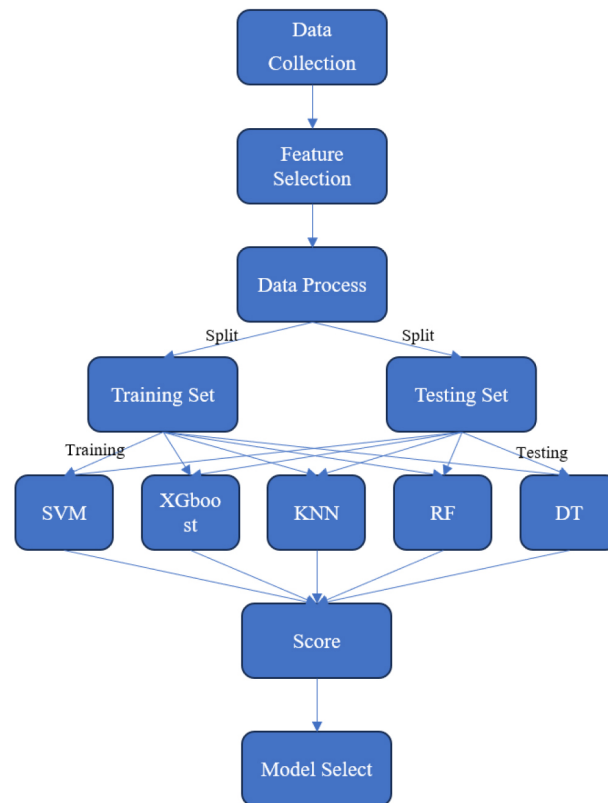


Fig. 1. The overall framework for predicting the recurrence of foot ulcers.

such as age, blood sugar control, alcohol consumption, smoking, etc., which may impact foot ulcer recurrence. Further research should delve into the complex relationships between these factors to gain a more comprehensive understanding of the recurrence risk in elderly patients with diabetes [8].

### 3. Method

In order to establish a more sophisticated predictive model for the recurrence risk of diabetic foot ulcers (DFUs) in elderly diabetic patients, we have designed a comprehensive framework, as illustrated in Fig. 1. The aim is to harness the combined strengths of multiple machine learning classifiers. This framework not only contributes to enhancing predictive performance but also provides healthcare professionals with a more comprehensive tool for risk assessment.

#### 3.1. Model selection

In this study, we opted for the Support Vector Machine (SVM) machine learning classifier to predict the recurrence risk of foot ulcers [9]. We then compared it with several common machine learning classifiers, including Support Vector Machine (SVM), XGBoost, KNN, Random Forest, and Decision Tree.

**Support Vector Machine (SVM):** SVM is a powerful classifier suitable for high-dimensional data and nonlinear relationships [10]. We selected SVM as the primary classifier due to its significant achievements

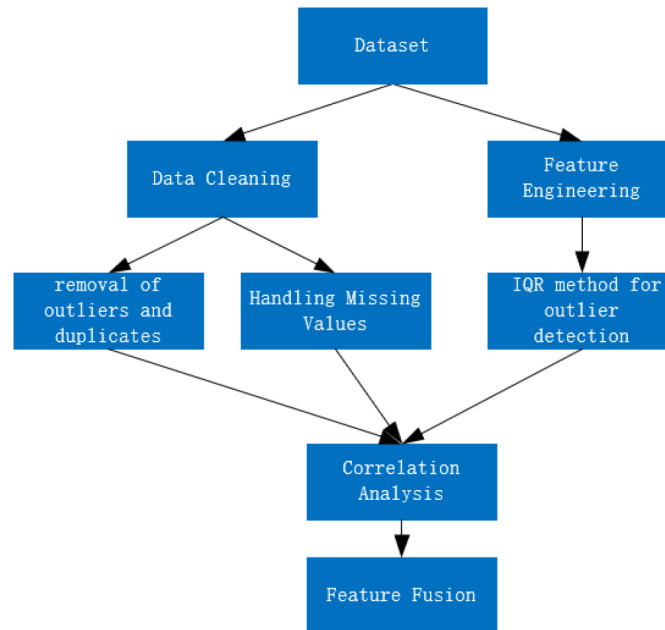


Fig. 2. Data preprocessing workflow.

in the medical field. We fine-tuned SVM hyperparameters, including C value, kernel type, and degree, to achieve optimal performance.

**XGBoost:** XGBoost is a gradient boosting algorithm suitable for large-scale data and efficient feature engineering [11]. We used XGBoost as a comparative model, optimizing its performance by adjusting parameters such as learning rate, tree depth, and the number of trees.

**KNN (K-Nearest Neighbors):** KNN is an instance-based algorithm, with predictions relying on the labels of previous instances most similar to the new instance [12]. We set K value and distance weights to explore its effectiveness in predicting foot ulcer recurrence risk.

**Random Forest:** Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve predictive performance [13]. We set parameters such as the number of forests, tree depth, and others to ensure the model's robustness and generalization ability.

**Decision Tree:** Decision trees are simple and intuitive classifiers helpful for understanding relationships between features. We set parameters like maximum depth and minimum leaf node number to balance model complexity and performance.

### 3.2. Data preprocessing

During data preprocessing, we initiated by cleansing the data, removing outliers, duplicates, and missing values. For missing values, we employed appropriate imputation methods such as mean or median filling to ensure data integrity.

Next, we conducted feature engineering, binary encoding discrete data for model understanding. For continuous data, we utilized the IQR method for outlier screening to enhance the model's robustness.

Lastly, to better comprehend the relationship between data features and foot ulcer recurrence, we performed correlation analysis, plotting correlation matrices to assist in selecting highly correlated features with the target variable for integration. The overall workflow is depicted in Fig. 2.

### 3.3. Genetic factors

In order to enhance the comprehensiveness of our predictive model, we recognize the importance of incorporating genetic factors into the analysis. Genetic markers can potentially play a crucial role in predicting the recurrence of diabetic foot ulcers (DFUs) in elderly diabetic patients, contributing to a more holistic understanding of the risk factors involved.

As part of our data collection process, we propose the inclusion of genetic data from study participants. This may involve obtaining information on specific genetic markers associated with diabetes and its complications. Collaborating with geneticists or utilizing existing genetic databases could provide valuable insights into the genetic predisposition to DFU recurrence.

Once genetic data is collected, it should be integrated into our machine learning models alongside other high-risk factors, such as age, blood sugar control, alcohol consumption, and smoking. The combined analysis of clinical and genetic data aims to improve the accuracy and comprehensiveness of our predictive model.

This addition not only strengthens the predictive power of our model but also contributes to advancing our understanding of the complex interplay between genetics and diabetic foot ulcer recurrence in elderly individuals.

### 3.4. Model evaluation

To assess the performance of the models, we employed common classification metrics, including Recall, Precision, and Accuracy [14]. These metrics provide a comprehensive understanding of the model's performance in various aspects.

During the experimental process, we utilized the K-fold cross-validation algorithm, dividing the dataset into multiple subsets to train and validate the model more comprehensively. This helps prevent overfitting and enhances the model's generalization ability.

Through the mentioned methods and experimental design, our goal is to establish an accurate, robust, and interpretable predictive model for the recurrence risk of foot ulcers.

## 4. Experiment and results

### 4.1. Data collection and processing

We collected high-risk factors for ulcer recurrence from 138 elderly diabetic patients, including gender, age over 70, thrombosis, alcohol consumption, smoking, strenuous physical labor, surgical intervention, blood sugar control, use of hormones and immunosuppressants, wound occurrence time, wound size (in CM), wound location, compression stockings, and BMI.

Raw data often contains missing values (NA or blanks), making it unsuitable for direct training. Additionally, the presence of outliers (values beyond the normal range) can negatively impact model performance [15]. Effective data processing is fundamental to improving predictive accuracy.

In data processing, we initially conducted data cleansing, removing outliers, duplicates, and handling missing values through deletion or imputation. Subsequently, we performed a correlation analysis on the cleaned data, as shown in Fig. 3. Highly correlated data were fused to reduce fitting complexity.

Afterwards, the data was categorized into continuous and discrete variables based on their value ranges. For discrete variables, we transformed them into binary encoding. For continuous variables, we initially

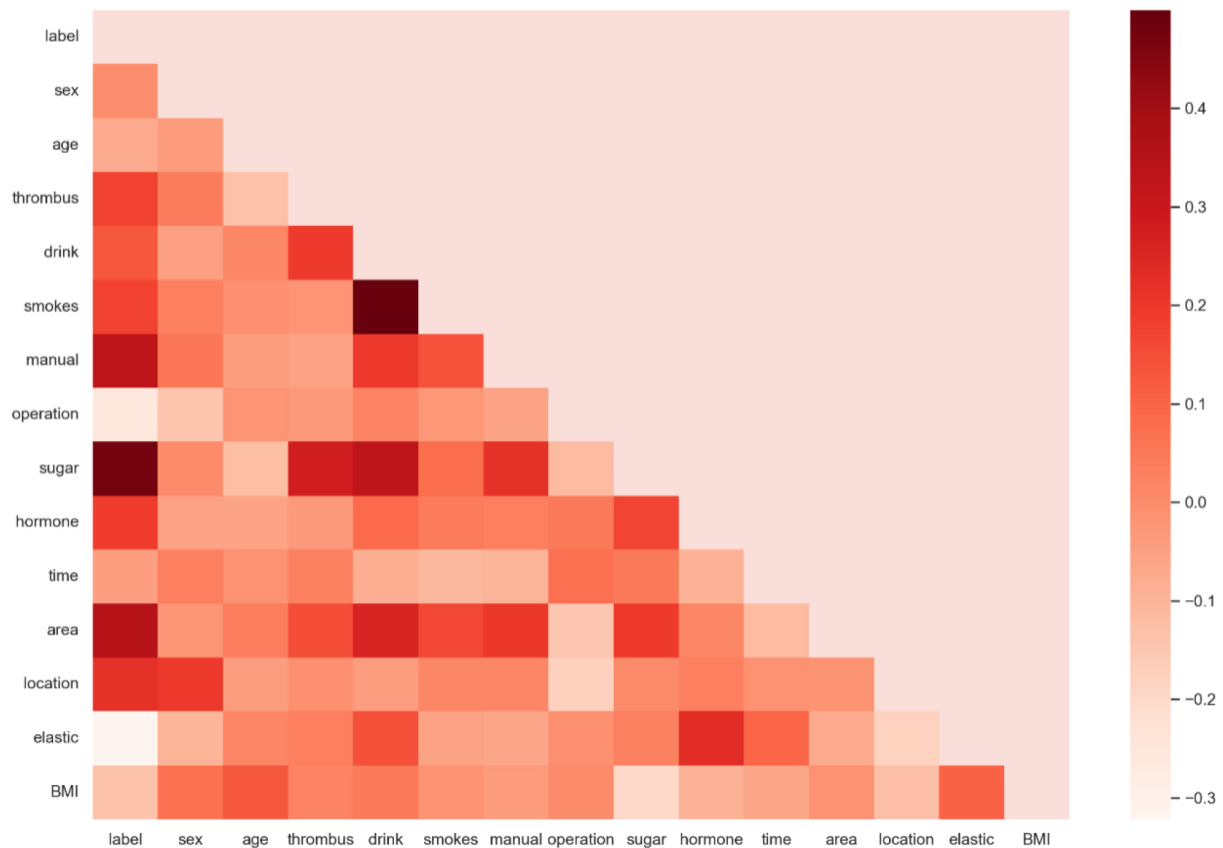


Fig. 3. Correlation analysis of high-risk factors for ulcer recurrence data.

applied the IQR method for outlier detection. The distribution of continuous data is illustrated in Fig. 4. Upon the identification of potential outliers, a cautious and iterative process was undertaken. Outliers were carefully examined to determine the validity of their values. We opted for a pragmatic approach, considering domain knowledge and potential impact on the overall dataset. Outliers were then either corrected through appropriate imputation methods or, in cases where their values were deemed erroneous, they were removed from the dataset.

Following the removal of outliers, we subjected the continuous data to scaling procedures. Feature scaling aims to mitigate the impact of feature scales on the model, thereby enhancing accuracy [16]. Additionally, for a more comprehensive examination of the relationship between features and ulcer occurrences, we integrated feature data with ulcer incidents for visualization, as depicted in Figs 5 and 6.

#### 4.2. Experimental procedure

In this study, Support Vector Machine (SVM) is employed as the predictive classifier, and it is compared with XGBoost, KNN, Random Forest, and Decision Tree. To ensure a fair and accurate comparison, each algorithm undergoes rigorous parameter tuning, aiming to optimize its performance. Table 1 provides a detailed overview of the model parameters for each classifier, showcasing the specific configurations chosen during the tuning process. The careful adjustment of parameters for each algorithm aims to

Table 1  
Model parameters

| Model         | Parameters  |
|---------------|---|
| SVM           | c = 20 degree = 2<br>gamma: scale, kernel type: linear  |
| XGBoost       | learning_rate = 0.015, max_depth = 10<br>n_estimators = 100   |
| KNN           | neighbors = 6, weights: distance<br>p: Manhattan distance   |
| Random forest | bootstrap: True, criterion = entropy<br>max_depth = 3, min_samples_leaf = 3<br>min_samples_split = 2, n_estimators = 10 |
| Decision tree | Criterion = gini, max_depth = 3<br>min_samples_leaf = 1<br>min_samples_split = 2  |

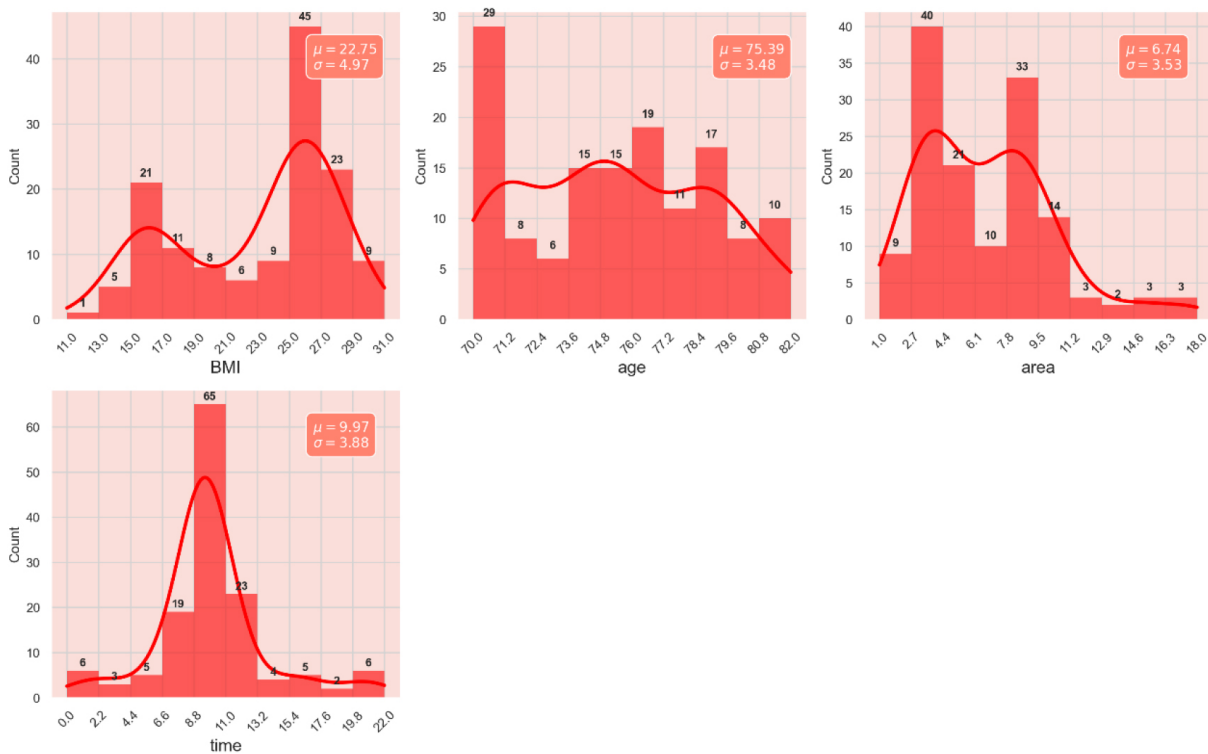


Fig. 4. Distribution of continuous high-risk factors for ulcer recurrence data.

capture the nuances of their individual strengths and characteristics, contributing to a robust and unbiased comparative analysis. This approach ensures that each algorithm is equipped with optimal settings, fostering a comprehensive evaluation of their effectiveness in predicting the recurrence risk of diabetic foot ulcers (DFUs) in elderly diabetic patients.

**Accuracy:** Summarizing the Performance of the Classification Task and assessing the count of accurately predicted instances among all data instances.

**Recall:** The proportion of participants with ulcer attacks and correctly classified among all participants.

**Precision:** How many people identified as having an ulcer attack belong to this category.

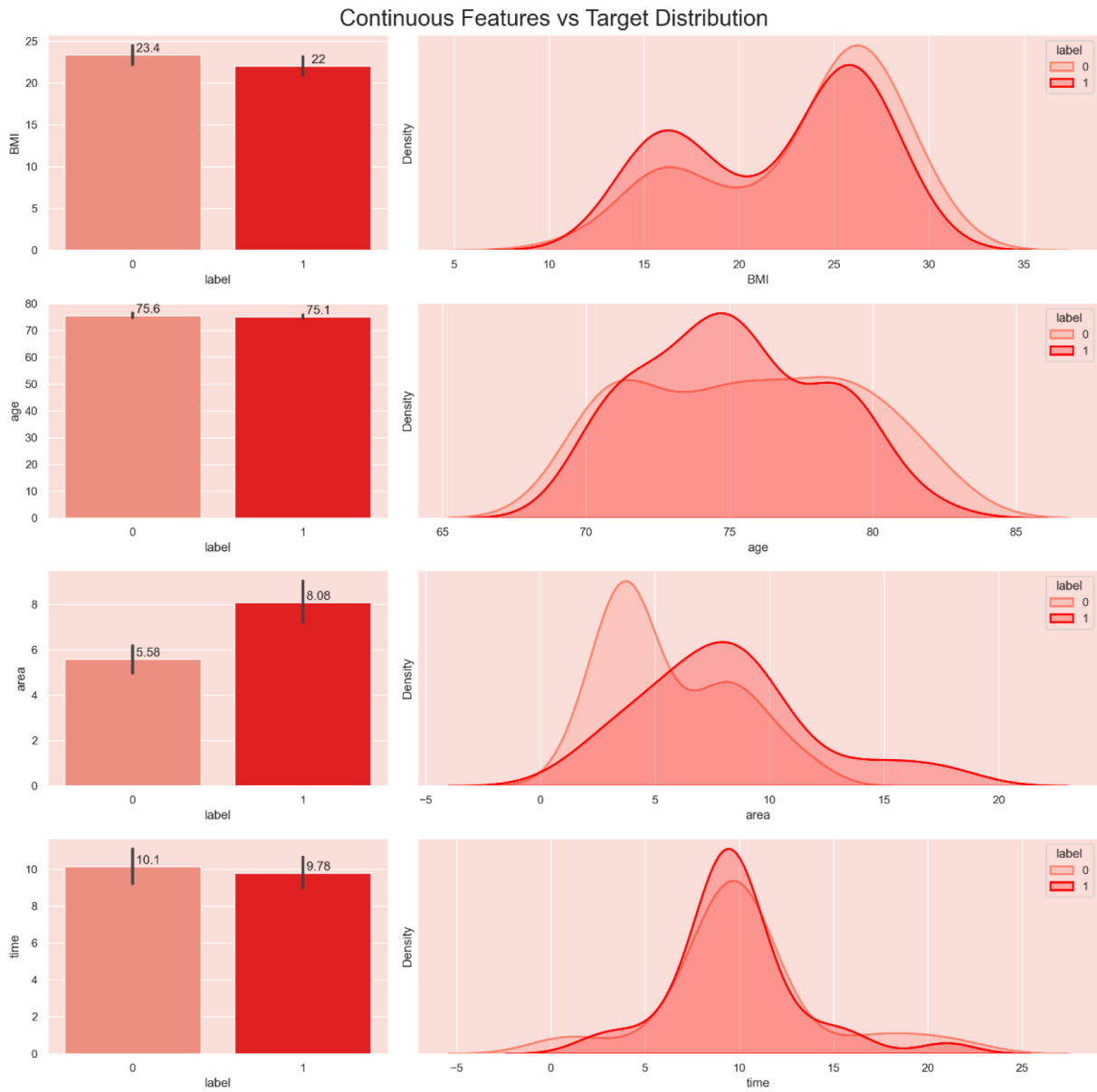


Fig. 5. Integration of feature data and ulcer: Continuous feature fusion and target distribution.

The formulas for calculating Recall, Precision, and Accuracy are as follows:

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$



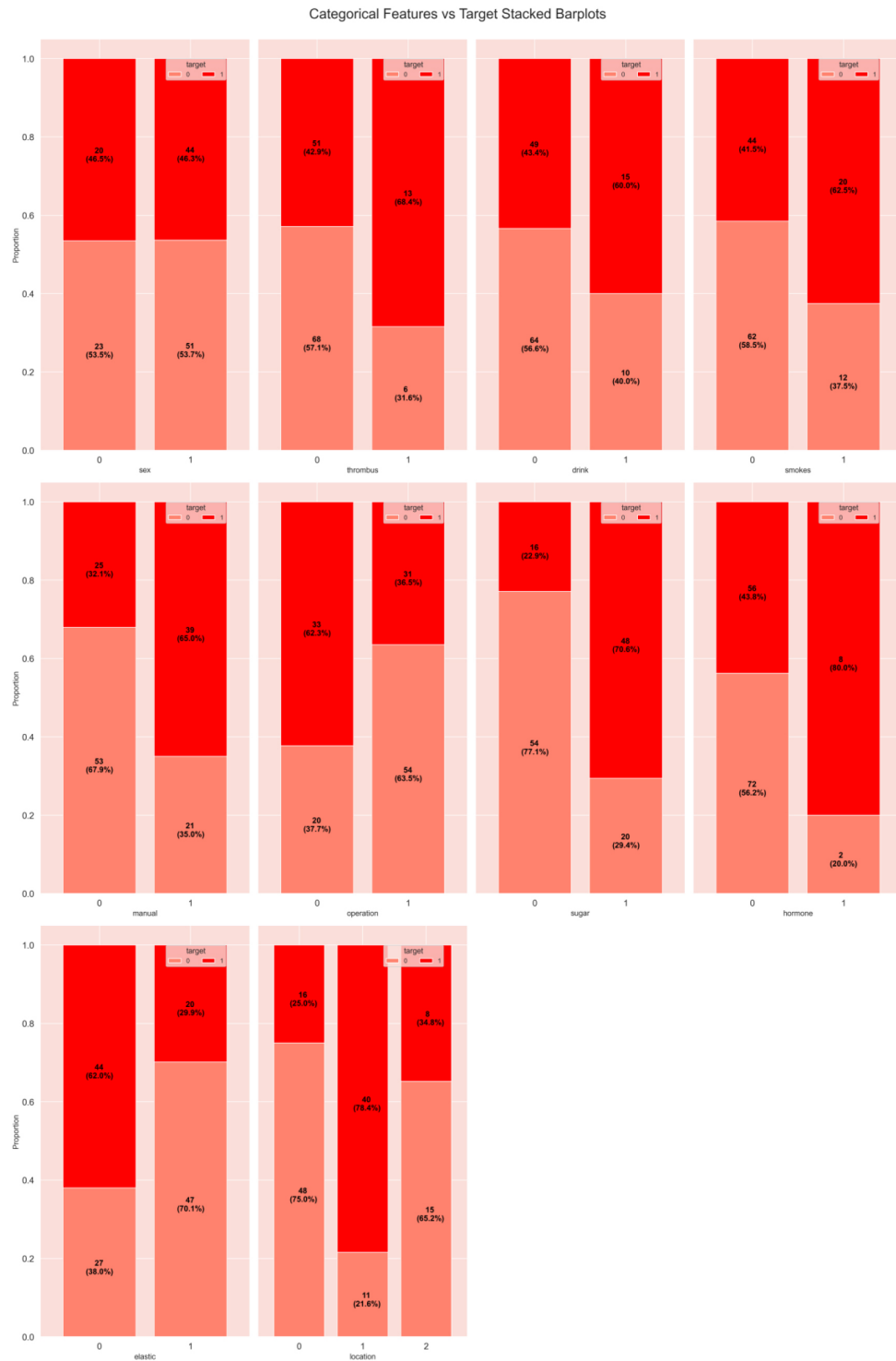


Fig. 6. Integration of feature data and ulcer: Stacked bar chart of categorical features and target.

Table 2  
Model results

| Model         | Recall | Precision | Accuracy |
|---------------|--------|-----------|----------|
| SVM           | 92%    | 92%       | 93%      |
| XGBoost       | 69%    | 90%       | 82%      |
| KNN           | 77%    | 83%       | 82%      |
| Random forest | 69%    | 90%       | 82%      |
| Decision tree | 62%    | 89%       | 79%      |

Here, TP stands for true positive, TN for true negative, FP for false positive, and FN for false negative. After preprocessing and standardization, we split the dataset into training and testing sets with a ratio of 7:3. Subsequently, the K-fold cross-validation algorithm is applied to train the model. The training dataset is evenly divided into k subsets, treating each subset as a validation fold once. Cross-validation is repeated k times, each time training the model with the remaining folds and validating the model's performance on the respective validation fold for training optimization. In our experiment, we set k to the empirical value of 5, striking a balance between computation and accuracy. Finally, the accuracy of the model on the test set is used as the ultimate performance metric.

#### 4.3. Experimental results

We divided the data into training and validation sets at a ratio of 7:3. The validation set was excluded from data processing and model training, serving only for the final validation of model performance. The performance of the model on the validation set reflects its generalization ability. Model evaluation is based on the model's performance on the validation set. The final results are presented in Table 2.

From Table 2, it can be observed that SVM achieved the highest accuracy at 93%, while Decision Tree exhibited the lowest performance with only 79%.

#### 4.4. Interpretability of machine learning models

Our analysis reveals that specific features play a pivotal role in the predictive outcomes of our model. Notably, surgical intervention demonstrate significant influence, shedding light on the critical decision-making factors behind our predictions.

We outline the decision process of our machine learning models, emphasizing the sequential steps taken in making predictions. Understanding the pathway to each prediction enhances transparency and aids healthcare professionals in grasping the rationale behind the model's decisions.

This section provides a user-friendly interpretation of the model's predictions, making it accessible and useful for healthcare professionals and patients alike. In addition to our current approach, we consider the exploration of alternative interpretable machine learning methods or post-hoc interpretability techniques.

## 5. Discussion

In our study, the application of machine learning models demonstrated potential advantages in analyzing the recurrence risk of foot ulcers in elderly diabetic patients. However, several aspects warrant further discussion and consideration.

Firstly, One notable limitation is the potential impact of data collection constraints. The availability of comprehensive datasets, especially concerning certain variables such as genetic factors or specific

lifestyle details, may have been limited. As a result, our model may not fully capture the entire spectrum of factors influencing diabetic foot ulcer (DFU) recurrence in elderly diabetic patients. Future research could consider incorporating additional factors that might influence foot ulcer recurrence, such as genetic factors, inflammation markers, etc., to enhance the comprehensiveness and predictive performance of the model [17].

Secondly, the interpretability of machine learning models remains a challenge. While our model exhibited excellent predictive performance, a deeper understanding of how the model makes predictions is crucial for healthcare professionals and patients. In future work, exploring more interpretable machine learning methods could enhance the model's understandability [18].

Third, machine learning models, including the Support Vector Machine (SVM) employed in our study, are susceptible to biases present in the training data. We recognize the importance of mitigating biases to ensure fair and equitable predictions. Efforts were made to curate a diverse and representative dataset, and ongoing monitoring for biases during model training and application is emphasized.

Lastly, While our study demonstrates promising results in predicting DFU recurrence in the specific population of elderly diabetic patients, generalizability to broader demographics should be approached with caution. External validation using datasets from diverse populations is essential to confirm the model's robustness and applicability beyond the specific cohort studied. Collaboration with multi-center study datasets would allow for a more comprehensive assessment of the model's performance and applicability [19].

## **6. Conclusion**

Our research introduces a machine learning approach to predict the long-term disease recurrence rate in elderly diabetic patients with foot ulcers. The results of our study indicate that the use of non-invasive and low-cost clinical and biological predictive factors can accurately forecast whether patients will experience a recurrence of foot ulcers in the future.

Our research provides empirical support for the machine learning analysis of the recurrence risk of foot ulcers in elderly diabetic patients. The application of machine learning models can assist healthcare professionals in more accurately identifying high-risk patients and taking appropriate preventive and intervention measures. Future work should further validate the reliability of the model and explore additional factors influencing foot ulcer recurrence to enhance the predictive performance of the model. This study offers a new perspective and possibilities for improving the management of foot ulcers in elderly diabetic patients.

## **Acknowledgments**

This work was supported by the Fujian Provincial Health and Middle-aged Youth Backbone Talent Training Project (2022GGB017) and Xiamen Natural Science Foundation Joint Project (3502Z202374023).

## **Conflict of interest**

None to report.

## References

- [1] Armstrong DG, Boulton AJ, Bus SA. Diabetic foot ulcers and their recurrence. *New England Journal of Medicine*. 2017; 376(24): 2367-2375.
- [2] Wu SC, Driver VR, Wrobel JS, Armstrong DG. Foot ulcers in the diabetic patient, prevention and treatment. *Vascular Health and Risk Management*. 2007; 3(1): 65-76.
- [3] Lavery LA, Armstrong DG, Wunderlich RP, Tredwell J, Boulton AJ. Diabetic foot syndrome: Evaluating the prevalence and incidence of foot pathology in Mexican Americans and non-Hispanic whites from a diabetes disease management cohort. *Diabetes Care*. 2003; 26(5): 1435-1438.
- [4] Jeffcoate WJ, Price PE, Phillips CJ, Game FL, Mudge E. Randomised controlled trial of the use of three dressing preparations in the management of chronic ulceration of the foot in diabetes. *Health Technology Assessment*. 2004; 8(29): iii-iv, 1-114.
- [5] Snyder RJ, Frykberg RG, Rogers LC, Applewhite AJ. The management of diabetic foot ulcers through optimal off-loading: Building consensus guidelines and practical recommendations to improve outcomes. *Journal of the American Podiatric Medical Association*. 2010; 100(5): 369-378.
- [6] Huang ZH, Li SQ, Kou Y, et al. Risk factors for the recurrence of diabetic foot ulcers among diabetic patients: A meta-analysis. *International Wound Journal*. 2019; 16(6): 1373-1382.
- [7] Pereira MG, Vilaça M, Braga D, et al. Healing profiles in patients with a chronic diabetic foot ulcer: an exploratory study with machine learning. *Wound Repair and Regeneration*. 2023.
- [8] Afsaneh E, Sharifdini A, Ghazzaghi H, et al. Recent applications of machine learning and deep learning models in the prediction, diagnosis, and management of diabetes: A comprehensive review. *Diabetology & Metabolic Syndrome*. 2022; 14(1): 1-39.
- [9] Nanda R, Nath A, Patel S, et al. Machine learning algorithm to evaluate risk factors of diabetic foot ulcers and its severity. *Medical & Biological Engineering & Computing*. 2022; 60(8): 2349-2357.
- [10] Farooqui ME, Ahmad DJ. Disease prediction system using support vector machine and multilinear regression. *International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN*. 2020; 2347-5552.
- [11] Deng L, Xie P, Chen Y, et al. Impact of acute hyperglycemic crisis episode on survival in individuals with diabetic foot ulcer using a machine learning approach. *Frontiers in Endocrinology*. 2022; 13: 974063.
- [12] Siregar SD, Ginting YUR, Sintami N, et al. Implementation of KNN algorithm in classifying diabetic ulcers in patients with diabetes mellitus. *Jurnal Mantik*. 2023; 7(2): 691-701.
- [13] Tulloch J, Zamani R, Akrami M. Machine learning in the prevention, diagnosis and management of diabetic foot ulcers: A systematic review. *IEEE Access*. 2020; 8: 198977-199000.
- [14] Mortazavi H, Safi Y, Baharvand M, et al. Diagnostic features of common oral ulcerative lesions: an updated decision tree. *International journal of dentistry*. 2016, 2016.
- [15] Wouter B, Abu-Hanna A, Bus SA. Development of a multivariable prediction model for plantar foot ulcer recurrence in high-risk people with diabetes. *BMJ Open Diabetes Research & Care*. 2020; 8(1).
- [16] Abaker AA, Saeed FA. A comparative analysis of machine learning algorithms to build a predictive model for detecting diabetes complications. *Informatica*. 2021; 45(1).
- [17] Hicks CW, Canner JK, Mathioudakis N, et al. Incidence and risk factors associated with ulcer recurrence among patients with diabetic foot ulcers treated in a multidisciplinary setting. *Journal of Surgical Research*. 2020; 246: 243-250.
- [18] D'Angelo G, Della-Morte D, Pastore D, et al. Identifying patterns in multiple biomarkers to diagnose diabetic foot using an explainable genetic programming-based approach. *Future Generation Computer Systems*. 2023; 140: 138-150.
- [19] Wei Z, Wang S, Wang Z, et al. Development and multi-center validation of machine learning model for early detection of fungal keratitis. *EBioMedicine*. 2023; 88.