FGSCare: A Feature-driven Grid Search-based Machine Learning Framework for Coronary Heart Disease Prediction

Zhenqi Li[a,*], Jiayi Zhang[b,*], Yuchen Wang[c,*], Jinjiang You[d], Yonghao Lin[a], Junhao Zhong[e], Soumyabrata Dev[f,g,**]

[a]*The Fu Foundation School of Engineering and Applied Science, Columbia University*
[b]*Faculty of Science and Engineering, University of Nottingham Ningbo China*
[c]*Department of Electrical and Computer Engineering, University of Washington*
[d]*School of Computer Science, Carnegie Mellon University*
[e]*Biomedical Engineering Department, Pratt School of Engineering, Duke University*
[f]*The ADAPT SFI Research Centre*
[g]*School of Computer Science, University College Dublin*

Abstract

Although machine learning has become a powerful tool for coronary heart disease (CHD) prediction, its effectiveness is often hindered by the complexity and nonlinear interactions among medical risk factors. A major challenge lies in feature selection, where the absence of systematic strategies may lead to information loss, overfitting, or the inclusion of irrelevant variables, ultimately degrading predictive performance. Additionally, different ML models exhibit varying predictive capabilities depending on the selected features. However, many studies fail to systematically evaluate how feature selection influences the performance of both traditional and deep learning approaches, limiting the understanding of optimal feature selection strategies and their impact on improving CHD prediction. To address these limitations, we propose a Feature-driven Grid Search-based Machine Learning Framework (FGSCare) for CHD prediction. FGSCare systematically filters and retains critical features through a data-driven selection process, enhancing model interpretability and generalization. We assess the impact of feature selection by comparing the performance of traditional machine learning classifiers. We assess the impact of feature selection by comparing the performance of traditional machine learning classifiers (e.g., k-Nearest Neighbors, ElasticNet, and Decision Tree) and deep learning models such as Transformers before and after feature selection. Experimental results demonstrate that feature selection significantly influences model performance across various evaluation metrics, including accuracy, precision, recall, F1-score, and AUC. Our findings provide valuable insights into the trade-offs between traditional ML and deep learning methods in CHD prediction, contributing to the development of more robust, data-driven healthcare applications. Our implementation is available at: `https://github.com/zl3508/Heart`.

Keywords: medical prediction, feature engineering, machine learning, deep learning

1. Introduction

Coronary heart disease (CHD) remains the leading cause of death worldwide, accounting for nearly 17.9 million deaths annually [1, 2]. Early detection and prevention of CHD can significantly reduce mortality rates and improve patient outcomes [3]. Traditional risk assessment models, such as the Framingham Risk Score, have been widely used for decades, relying on statistical methods to

*Authors contributed equally to this research.
**Corresponding author. Tel.: + 353 1896 1797.
Email addresses: zl3508@columbia.edu (Zhenqi Li), smyjz19@nottingham.edu.cn (Jiayi Zhang), yuchwang@uw.edu (Yuchen Wang), jinjiany@andrew.cmu.edu (Jinjiang You), yl5763@columbia.edu (Yonghao Lin), jz455@duke.edu (Junhao Zhong), soumyabrata.dev@ucd.ie (Soumyabrata Dev)

estimate an individual's probability of developing heart disease [4]. However, these models often suffer from limitations in capturing complex, non-linear interactions between risk factors [5].

In recent years, machine learning and deep learning techniques have demonstrated superior performance in medical diagnosis and prognosis [6, 7]. ML models, including decision trees, ensemble learning methods, and neural networks, can leverage large datasets to identify patterns that may be difficult to capture using traditional statistical methods [8, 9]. Additionally, ML models can be fine-tuned to improve accuracy while maintaining interpretability, a key factor in medical decision-making [10, 11].

Despite their potential, the application of ML techniques in CHD prediction still faces several challenges, including data quality, model interpretability, and fairness across demographic groups [12]. Feature selection also plays a crucial role in improving model performance by eliminating irrelevant or redundant information [13, 14]. This study aims to systematically compare various ML algorithms for CHD prediction using the Framingham Heart Study dataset [15]. Specifically, we evaluate models such as Decision Tree, Logistic Regression, and ElasticNet while also exploring deep learning models like KAN [16].

Our objectives are to:

- Leverage the well-established Framingham Heart Study dataset and apply preprocessing techniques, including handling missing values, outlier detection, and feature scaling, to ensure data quality.
- Evaluate and compare traditional ML and DL models to determine the most effective approach for CHD prediction.
- Investigate the effect of feature selection on model performance by comparing results using full and reduced feature sets. Additionally, model interpretability is enhanced using Shapley Additive Explanations, providing insights into feature importance and the decision-making process of different classifiers.

The structure of this paper is as follows: Section 2 reviews related work, discussing various machine learning techniques for CHD prediction, including supervised and deep learning models, feature selection methods, and interpretability approaches, highlighting their strengths and limitations. Section 3 describes the dataset and outlines the assumptions made for the experiments. Section 4 details the data preprocessing and feature engineering steps. Section 5 provides experimental results with quantitative analysis for traditional machine learning models and deep learning models. Section 6 interprets Dicision Tree model using Shapley Additive Explanations. Finally, Section 7 concludes the paper by discussing findings and potential directions for future research.

2. Related Work

Machine learning and deep learning have emerged as indispensable tools across a wide range of domains, transforming how we address complex challenges. In entertainment, they enable accurate popularity prediction for media content [17, 18, 19]; in medical diagnosis, they enhance the detection and prognosis of diseases [20, 21, 22]; in environmental analysis, they improve climate modeling and resource management [23, 24, 25, 26]; in finance, they power fraud detection and market forecasting [27]; and in transportation, they drive advancements in autonomous systems and traffic optimization [28]. Among these applications, ML and DL have shown significant promise in healthcare, particularly in cardiovascular disease risk assessment [29]. Coronary heart disease (CHD), a leading cause of mortality worldwide [30], has been a primary focus of ML-based predictive modeling. Traditional ML methods, including Decision Trees [31], Logistic Regression [32], K-Nearest Neighbors (KNN) [33], and Gradient Boosting [34], have been widely applied in CHD risk assessment [35, 36]. These models analyze clinical datasets—comprising demographics, lipid profiles, and blood pressure

measurements—to generate probabilistic risk scores [37]. Among them, Gradient Boosting demonstrates the effectiveness of ensemble learning, improving predictive accuracy and generalization [38, 39].

Although traditional ML models excel at structured data analysis, their performance is often limited when dealing with high-dimensional images and sequential data [40]. Deep learning overcomes these challenges by automatically extracting hierarchical feature representations from raw data, enabling more effective modeling of complex patterns in tasks [41, 42, 43]. For instance, CNNs represent a well-known type of learning-based vision model, commonly employed as dense feature extractor in various vision-related applications such as self-driving and perception systems [44, 45], camera calibration [46], medical vision [47, 48], salient object detection (SOD) vision algorithm [49, 50], and robotics perception system [51, 52, 53]. In general, CNNs have demonstrated solid performance in processing image data and learning some information from it [54, 55, 56] and cardiac imaging data, such as echocardiograms and coronary angiograms [57]. In addition, Recurrent Neural Networks (RNN) [58] and Long Short-Term Memory (LSTM) networks [59] effectively capture temporal dependencies in ECG signals and continuous blood pressure monitoring data [60]. These models have been widely applied to physiological trend analysis and disease progression prediction [61].

Beyond model architecture, feature selection and interpretability are crucial in CHD prediction due to the high dimensionality of medical data [62]. Redundant or irrelevant variables can degrade model performance, and techniques like recursive feature elimination (RFE) [63], mutual information-based selection [64], genetic algorithm-based optimization [65], Lasso regression, and SHAP-based selection [66] help identify the most predictive features. Ensuring interpretability is essential for clinical adoption, as black-box models can obscure decision-making processes. Shapley Additive Explanations (SHAP) [67] and Local Interpretable Model-agnostic Explanations (LIME) [68] are widely used to quantify feature importance and enhance transparency [69]. SHAP, in particular, allows for a direct comparison of feature importance across tree-based models and deep neural networks [70]. This study integrates SHAP-based analysis to evaluate interpretability differences between ML and DL models in CHD prediction.

This study evaluates traditional machine learning and deep learning models for CHD prediction, focusing on predictive performance, interpretability, and the role of feature selection. By comparing model effectiveness before and after feature processing, we provide insights into how different architectures respond to structured feature engineering. Furthermore, SHAP-based interpretability analysis improves transparency in decision-making, ensuring clinically meaningful predictions. This work aims to advance CHD risk assessment by developing more accurate and interpretable prediction frameworks.

3. Dataset

The dataset is derived from the Framingham Heart Study, containing 4,240 records with 16 variables describing demographic, lifestyle, and clinical risk factors associated with CHD. Key attributes include demographic information (gender, age, education), lifestyle factors (current smoking status, cigarettes per day), and medical history indicators (prevalent stroke, prevalent hypertension, diabetes). Clinical measurements such as total cholesterol (totChol), systolic and diastolic blood pressure (sysBP, diaBP), body mass index (BMI), heart rate, and glucose levels are also included. The target variable, TenYearCHD, is a binary indicator denoting whether a patient developed CHD within a ten-year period. The dataset contains some missing values, particularly in education, cigsPerDay, BPMeds, totChol, BMI, heart rate, and glucose attributes, which are handled through imputation techniques. The goal of this study is to develop machine learning models that accurately predict CHD risk based on these features.

4. Data Precessing

In our proposed FGSCare framework, we handled missing values in features such as education, cigsPerDay, BPMeds, totChol, BMI, heartRate, and glucose using statistical imputation. Outliers in numerical features were identified and capped to prevent extreme values from affecting model performance. Categorical variables were transformed using Label Encoding, while numerical variables were normalized with MinMax scaling to ensure consistency. These basic preprocessing steps helped clean the dataset and make it suitable for training machine learning models.

4.1. Handling missing data

Handling missing data is a critical step in data preprocessing, as it directly impacts the reliability and accuracy of subsequent analyses. In this study, we addressed missing values by identifying high-missingness features, categorizing remaining features, and applying imputation strategies. To further quantify the extent of missingness, Figure 1 was generated to visualize the proportion of missing values for each feature.

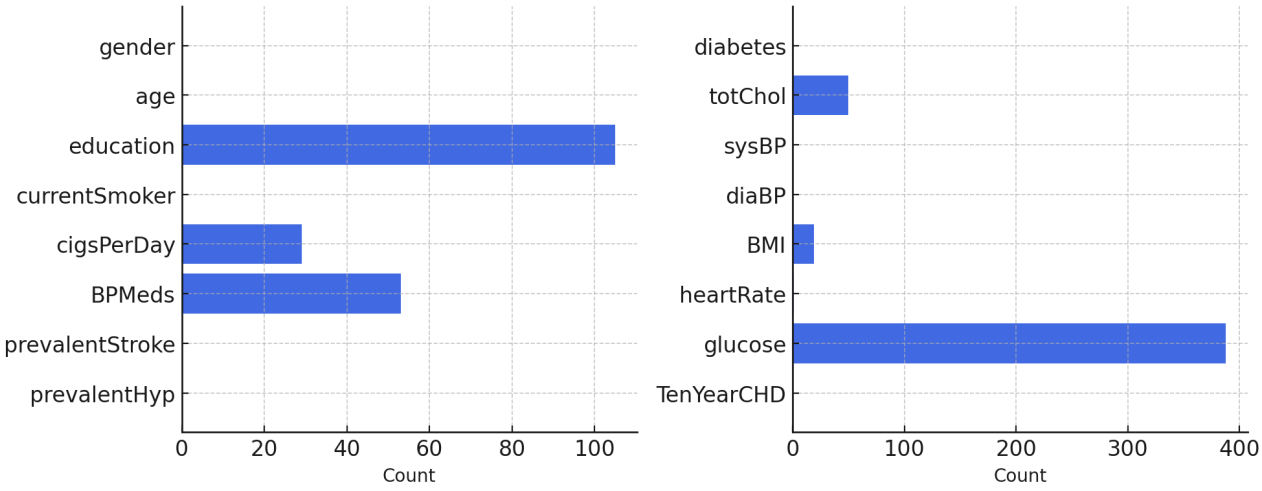


Figure 1: Bar chart of missing values. Features such as 'education' and 'glucose' show the highest missingness, while others like 'BPMeds' and 'BMI' have near-complete data.

Based on Figure 1, we applied different imputation strategies: for numerical features, mean imputation was used, while for categorical features, median imputation helped maintain class balance. These ensured that data integrity was preserved while minimizing potential biases introduced by missing values. After imputation, all features remained in the dataset. These features were categorized into two groups based on their data types and missingness levels.

The comparison between Figure 2 and Figure 3 demonstrates the effect of missing data imputation. Initially, some numerical features exhibited significant gaps, making their distributions uneven. However, after applying mean and median imputation, the distributions became smoother, ensuring better data quality for subsequent analyses. This step was crucial in maintaining statistical integrity and preventing potential biases in model training. As Figure 4 and Figure 5 show, the impact of our imputation methods on numerical features was assessed using histograms, which compared the distributions of selected features before and after imputation. These histograms confirmed that the central tendencies and overall distributions of features were preserved, validating the effectiveness of our imputation methods.

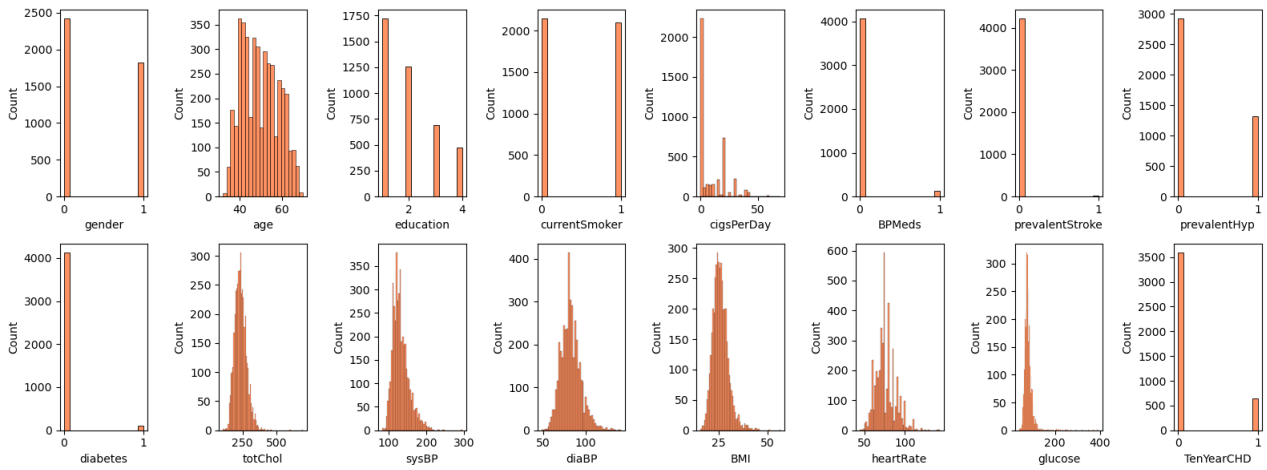


Figure 2: The distribution of features before imputation shows noticeable gaps due to missing values in numerical features, particularly in 'glucose' and 'totChol'.

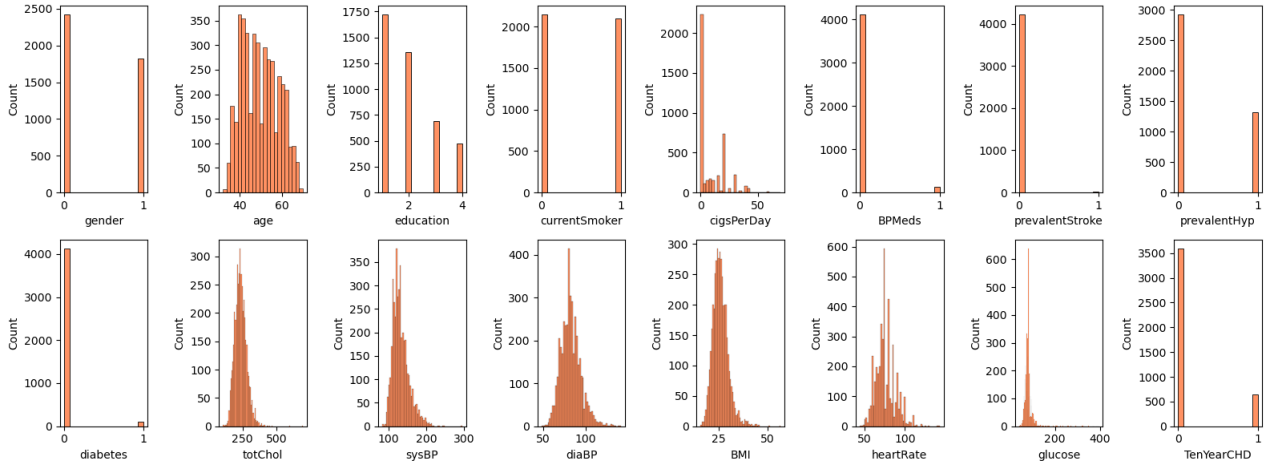


Figure 3: After imputation, the feature distributions appear smoother as the missing values have been filled, especially in numerical features.

4.2. Outlier Detection & Treatment

Outlier detection and handling are crucial preprocessing steps in healthcare datasets, as extreme values may result from measurement errors, data entry mistakes, or significant anomalies in patient conditions. To ensure robust analysis, we employed a combination of boxplot visualization and statistical techniques to identify and mitigate the impact of outliers.

As illustrated in Figure 6, we began by visualizing numerical features using boxplots to detect potential anomalies. Features such as dlaBP, heart rate, and glucose exhibited values well beyond their typical ranges, with dlaBP showing particularly high values that could impact the reliability of TenYearCHD predictions. To further examine the distribution of outliers, we separately analyzed boxplots for individuals with TenYearCHD. This subgroup analysis helped determine whether certain extreme values were inherently linked to disease presence or appeared arbitrarily across the dataset, allowing for more informed outlier-handling strategies.

To mitigate the impact of extreme values while preserving meaningful variations in the dataset, we applied a combination of Interquartile Range (IQR) filtering [71], Winsorization [72], and Clipping. Using the IQR method, we identified outliers as values beyond 1.5 times the interquartile range and capped them at the respective lower and upper bounds instead of removing them. Winsorization

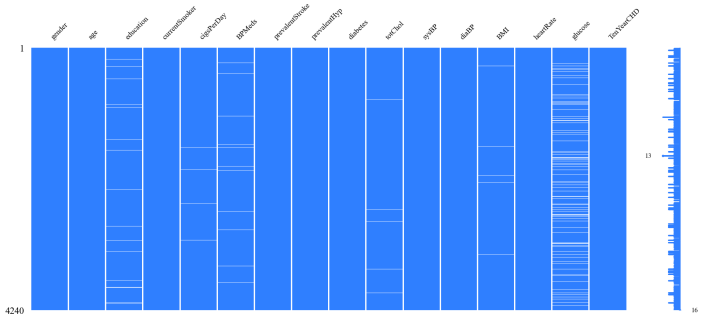


Figure 4: Missing value matrix. The missing value matrix displays the distribution of missing values across different features.

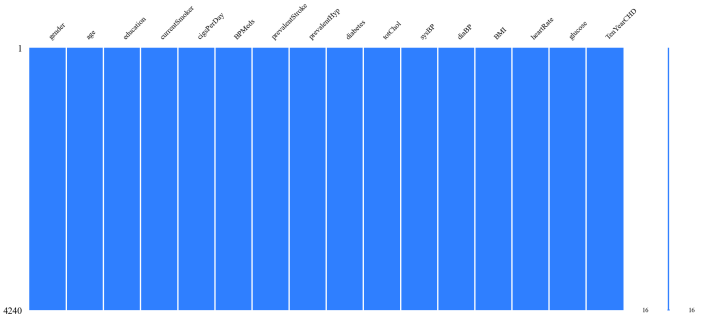


Figure 5: Post-imputation missing value matrix. The post-imputation missing value matrix confirms the resolution of all missing data.

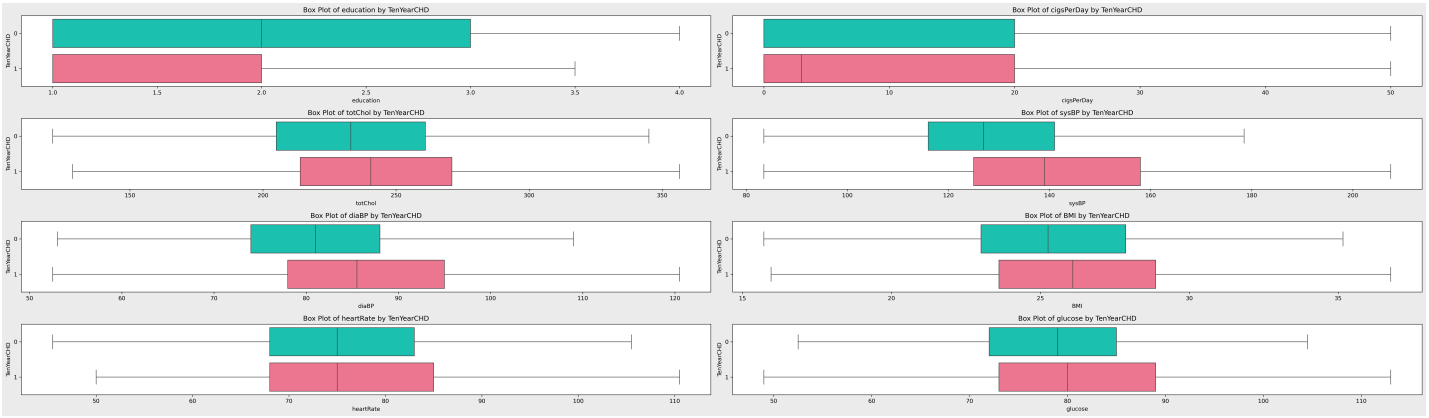


Figure 6: Boxplots of numerical features before outlier handling.

was performed by replacing extreme values beyond the 5th and 95th percentiles to reduce their influence while maintaining the overall distribution. Additionally, clipping was applied to constrain values within predefined statistical limits, ensuring that excessive deviations did not distort the dataset. These preprocessing steps effectively minimized the impact of extreme values. Post-treatment boxplots are shown in Figure 7. The adjusted distributions no longer contained extreme values while preserving their central tendencies, ensuring that the dataset remained robust and reliable for subsequent analysis and modeling.

4.3. Feature Selection & Reconstruction

After handling outliers, we performed a series of feature processing steps to optimize the dataset for predictive modeling. These steps included feature correlation analysis, feature selection, feature scaling, and data balancing using Synthetic Minority Oversampling Technique (SMOTE). The goal was to reduce redundancy and address class imbalance to improve model performance.

To identify redundant features and potential multicollinearity, we computed the Pearson correlation matrix for all numerical variables. The correlation matrix was visualized using a heatmap, where features with high correlation coefficients were easily identifiable. Beyond correlation analysis, recent studies have explored more sophisticated techniques for evaluating feature redundancy. Lin et al. [73]introduced the Integrated Transportation Distance (ITD), which provides a rigorous mathematical framework to quantify the similarity between stochastic kernels in Markov systems. ITD can be adapted to feature selection by identifying redundant variables based on their probabilistic distributions rather than just their linear correlation. In our study, we leverage Pearson correlation for initial selection, but future work may integrate ITD to further refine feature importance ranking. Additionally, we observed that diaBP and

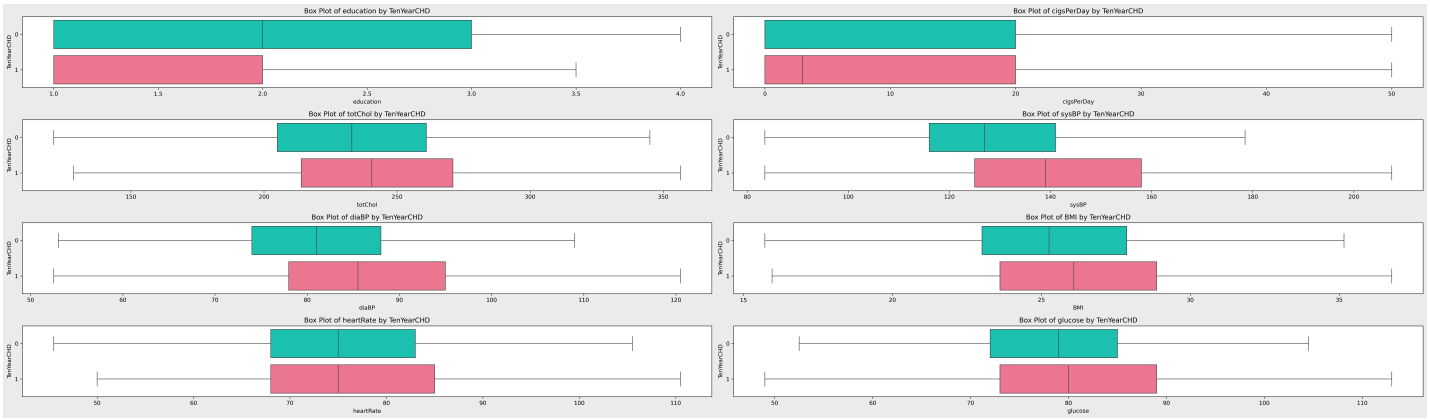


Figure 7: Boxplots of numerical features after outlier handling.

sysBP provide more detailed information related to blood pressure compared to the binary int64 feature prevalentHyp. In our study, we leverage Pearson correlation for initial selection, but future work may integrate ITD to further refine feature importance ranking. Additionally, we observed that diaBP and sysBP provide more detailed information related to blood pressure compared to the binary int64 feature prevalentHyp. Therefore, we decided to retain diaBP and sysBP as they offer a more comprehensive representation of blood pressure variations, while removing prevalentHyp to enhance the dataset's efficiency.

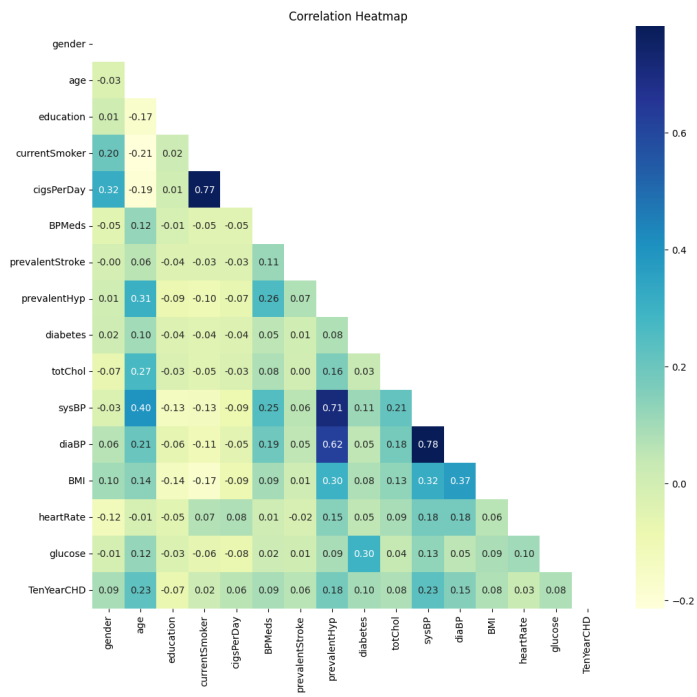


Figure 8: Correlation heatmap before feature selection.

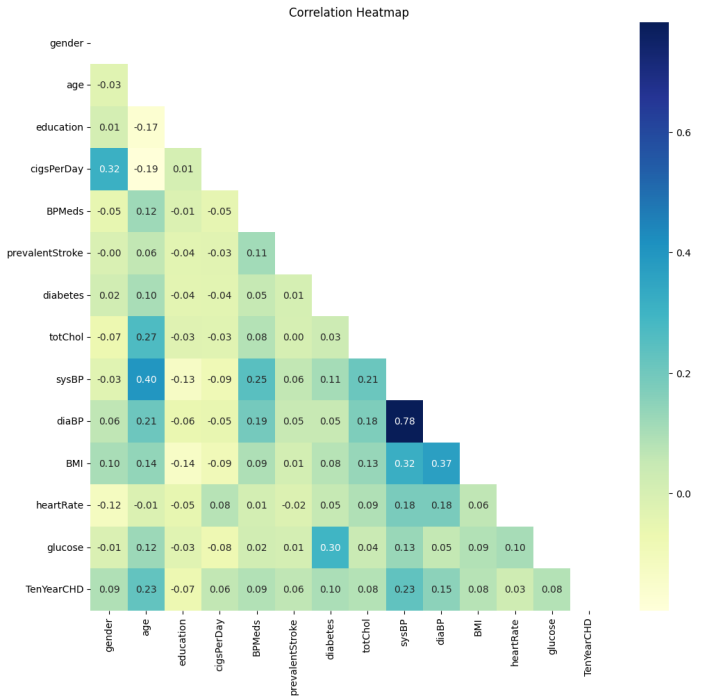


Figure 9: Correlation heatmap after feature selection.

Based on the correlation analysis, we rank the features according to their absolute correlation with the target variable, TenYearCHD. A pie chart in Figure 10 was created to visualize the top 15 most correlated features. Features such as currentSmoker and prevalentHyp were removed due to high multicollinearity predictive performance. After removing redundant features, we recalculated the correlation matrix to confirm that the remaining features exhibited lower multicollinearity. The updated correlation heatmap is shown in Figure 9.

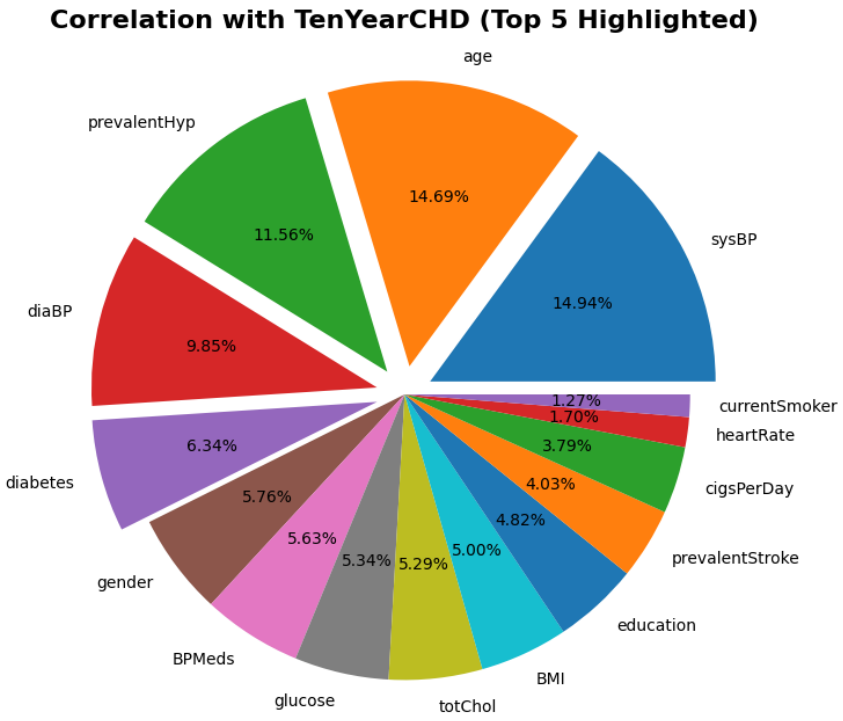


Figure 10: Pie chart showing top 5 features correlated with TenYearCHD.

To ensure that all features contributed equally to the model, we applied MinMax scaling, transforming all numerical variables to a standardized range between 0 and 1. This scaling technique is particularly effective for algorithms sensitive to feature magnitudes, such as neural networks, as it helps accelerate convergence during model training and ensures that no single feature dominates due to its scale. In addition to scaling, we addressed the class imbalance inherent in the dataset, where cases of TenYearCHD were significantly fewer than non-cases. To mitigate this issue, we employed the SMOTE, which generates synthetic samples for the minority class by creating new instances based on the feature space similarities of existing minority samples. This approach helps to prevent overfitting, which can occur with simple duplication of minority instances, and improves the model's ability to generalize to unseen data.

Through the combination of correlation analysis, feature selection, scaling, and data balancing, we refined the dataset to ensure it was robust, well-structured, and optimized for predictive modeling. These preprocessing steps played a crucial role in enhancing the model's accuracy, stability, and generalization capabilities.

4.4. Principal Component Analysis

To better understand the contribution of different features in the dataset, we performed t-Distributed Stochastic Neighbor Embedding (t-SNE) and Principal Component Analysis (PCA). PCA is a dimensionality reduction technique that transforms correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component captures the highest variance in the dataset, followed by subsequent components that explain decreasing amounts of variance. It is widely used for feature selection, noise reduction, and data visualization. On the other hand, t-SNE is a nonlinear technique that helps visualize high-dimensional data by preserving local similarities between data points. Unlike PCA, which captures global variance through linear projections, t-SNE focuses on maintaining the relative distances between points in lower-dimensional space, making it particularly

effective for cluster visualization. The t-SNE and PCA results are shown in Figure 11.

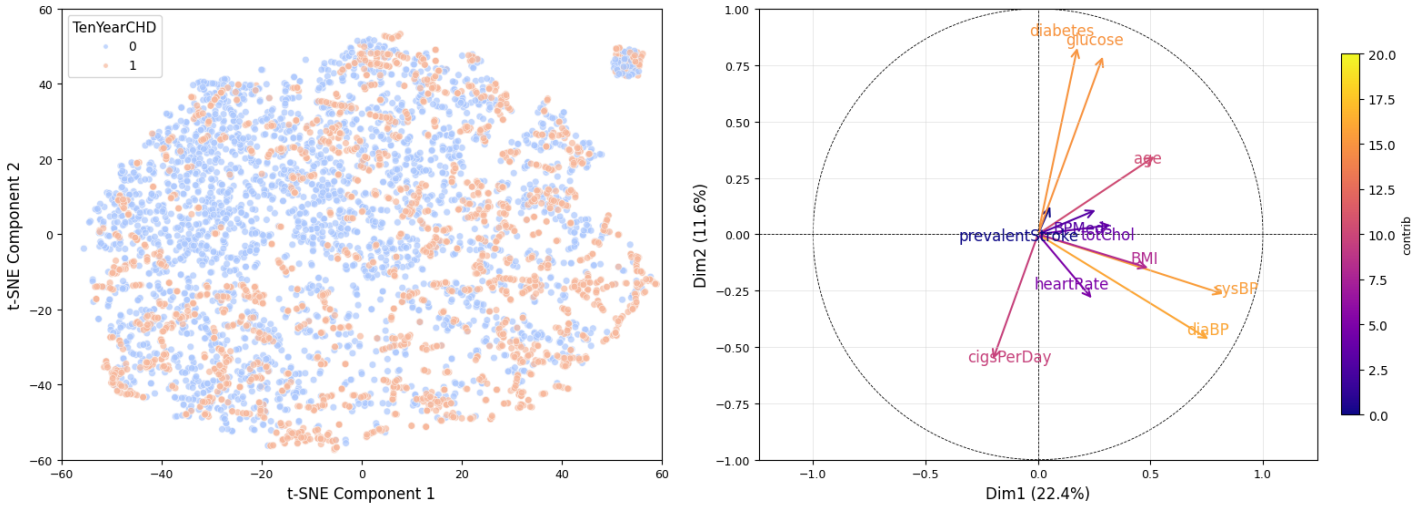


Figure 11: The left side shows the t-SNE visualization of CHD classes, while the right side presents the PCA biplot illustrating feature contributions to the first two principal components.

As Figure 11 shows, the t-SNE plot visualizes the distribution of people without CHD over a 10-year period. Given that the data set was balanced before analysis, both classes are equally represented at 50% each. Although some degree of overlap is evident, distinct clusters with a higher concentration of CHD cases suggest potential underlying patterns that may aid predictive modeling. The PCA biplot elucidates the relationships among features and their contributions to the first two principal components, which collectively explain 34.0% of the variance. In particular, age, glucose, and diabetes contribute strongly to the first principal component, while blood pressure (sysBP, diaBP), BMI, and total cholesterol contribute significantly to both components. Furthermore, CigsPerDay shows a negative correlation with the primary component. To refine the set of characteristics, gender and education were excluded based on heat map analysis, as their influence on CHD prediction was minimal while occupying a disproportionately large part in PCA.

These visualizations demonstrate the effectiveness of t-SNE and PCA in analyzing the dataset, providing additional evidence that our prior feature processing steps were correctly applied. In the next section, we outline the methodology for training and evaluating various machine learning and deep learning models for CHD prediction and analyze the preferences of different models.

5. Experiment

5.1. Evaluation Metrics

To comprehensively assess the performance of our selective model, multiple evaluation metrics are utilized, including Accuracy, Precision, Recall, Area Under the ROC Curve (AUC), and F-measure. Each of these metrics captures different aspects of model performance, ensuring a holistic evaluation across various dimensions.

Accuracy reflects the overall correctness of predictions, making it a general measure of model performance. However, it may be less informative when dealing with imbalanced datasets. **Precision** (positive predictive value) quantifies the proportion of correctly predicted positive cases, which is particularly crucial in scenarios where false positives are costly. **Recall** (sensitivity) evaluates the model's ability to identify all actual positive instances, making it essential in cases where missing positive samples are detrimental.

AUC measures the model's discriminative ability between classes by analyzing the trade-off between the true positive rate (TPR) and false positive rate (FPR) across classification thresholds. Finally, the **F-measure** (or F_β-score) provides a balanced representation of Precision and Recall through a weighted harmonic mean, with β determining the relative importance of Recall.

The mathematical definitions of these metrics are given as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$\text{AUC} = \int_0^1 \text{TPR}(t), d\text{FPR}(t) = P(\hat{y}^+ > \hat{y}^-), \tag{4}$$

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}}. \tag{5}$$

In the case of $\beta = 1$, the F_1-score provides an equally weighted measure of Precision and Recall, making it a widely used metric in classification tasks where both aspects are equally significant.

5.2. Interpretability Analysis

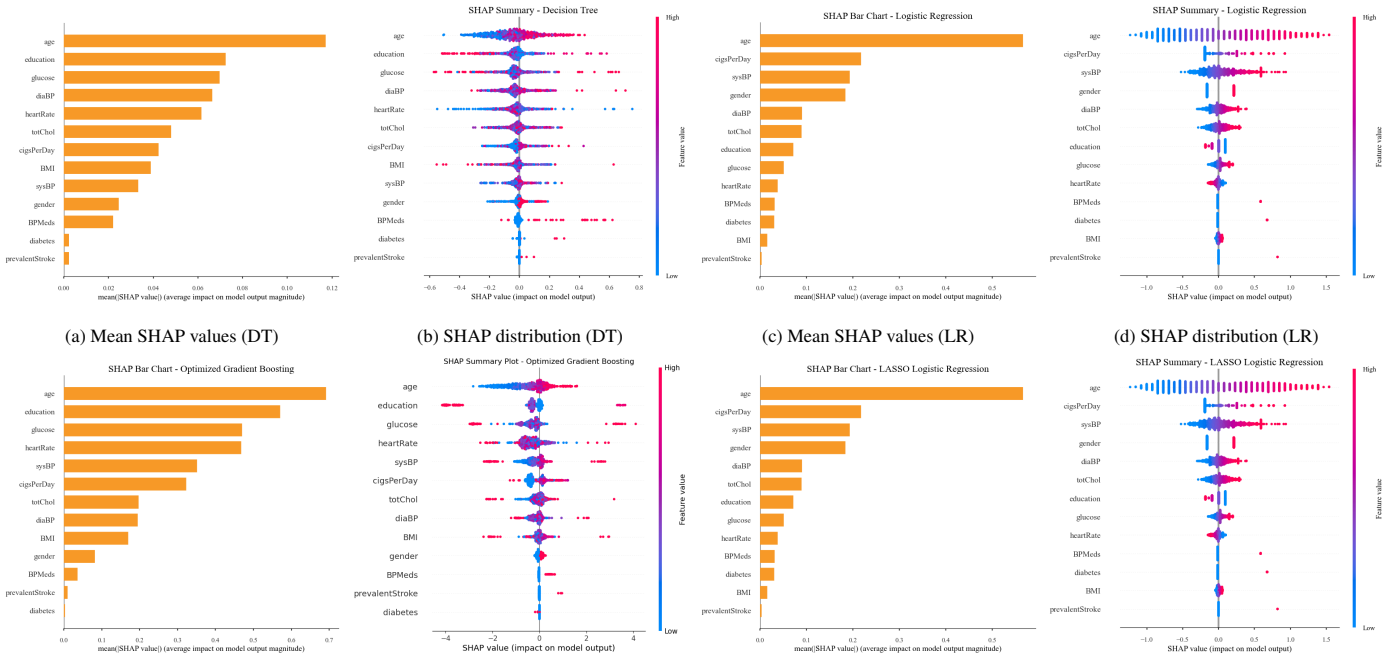
To improve the interpretability of our machine learning models, we leverage Shapley Additive Explanations (SHAP) [67], a game-theoretic method designed to quantify each feature's contribution to the model's predictions, thereby enhancing transparency and explainability.

SHAP values are derived from Shapley values in cooperative game theory, ensuring a fair allocation of each feature's impact across various feature subsets. The SHAP value for a given feature i is computed as follows:

$$\phi_i = \sum_{S \subseteq N \setminus i} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \Big[f(S \cup i) - f(S) \Big] \tag{6}$$

where ϕ_i denotes the SHAP value of feature i, N represents the full set of features, S is a subset of features excluding i, $f(S)$ refers to the model's prediction when considering only features in S, and $f(S \cup i) - f(S)$ captures the influence of adding feature i on the model's output.

In this work, SHAP is applied to interpret the feature importance of the most effective DTT classifier, with two summary visualizations produced. Figure 12a presents a global feature importance bar chart, illustrating the mean SHAP values for different features. The x-axis represents the mean absolute SHAP value (mean(|SHAP value|)), while the y-axis lists various feature names, including 'age', 'education', and 'glucose'. The bar lengths correspond to the magnitude of the mean SHAP values, where the feature 'age' exhibits the highest mean SHAP value, approaching 0.12, indicating its dominant influence on the model's output. Conversely, 'prevalent stroke' has the lowest mean SHAP value, around 0.005, implying a minimal effect on the model's predictions. This visualization enables a



(a) Mean SHAP values (DT) (b) SHAP distribution (DT) (c) Mean SHAP values (LR) (d) SHAP distribution (LR)

(e) Mean SHAP values (GB) (f) SHAP distribution (GB) (g) Mean SHAP values (LASSO) (h) SHAP distribution (LASSO)

Figure 12: SHAP summary and distribution plots for different models.

quick assessment of feature relevance, facilitating feature selection and model refinement by prioritizing the most impactful variables in predictive analysis.

As illustrated in Figure 12b 12d 12f 12h, the detailed feature importance distribution plots provide a SHAP-value dependence analysis for different models, including Decision Tree (DT), Logistic Regression (LR), Gradient Boosting (GB), and LASSO Regression. These plots reveal how each feature influences model predictions under varying conditions. The horizontal axis represents the SHAP values, ranging approximately from -0.6 to 0.8, while the vertical axis lists different feature names. To the right, a color bar labeled 'Feature value' transitions from blue to red, indicating that feature values span from 'Low' to 'High'. Each row of points corresponds to a feature, where the color of the points reflects the actual feature values, and their horizontal positions indicate the SHAP values. For instance, in the Decision Tree model, the 'age' feature exhibits a wide distribution of points with colors spanning from blue to red, signifying that the impact of 'age' on model predictions varies significantly in both direction and magnitude. In contrast, the 'education' feature in the Logistic Regression model predominantly shows red points with mostly positive SHAP values, suggesting that higher education levels have a consistently positive influence on model outputs. Similarly, the 'glucose' feature in the Gradient Boosting model reveals a strong positive correlation, indicating its crucial role in prediction outcomes. These SHAP dependence plots enhance the interpretability of machine learning models by revealing complex interactions between features and their contributions to predictions. The distinct distributions across different models highlight the varying mechanisms through which different algorithms leverage feature information, underscoring the necessity of model-specific interpretation when analyzing feature importance. Such insights are particularly valuable in feature selection, risk assessment, and decision-making processes, ensuring a transparent and data-driven approach to model evaluation.

5.3. Quantitative Experiment

In this experiment, the proposed framework, FGSCare, integrates both traditional machine learning models and deep learning architectures. Inspired by the work of stroke prediction [74], which demonstrated the effectiveness of machine learning and neural networks to enhance predictive accuracy, we adopt a similar approach to explore various models for CHD prediction. Their findings highlight the importance of leveraging both conventional and deep learning models to capture complex feature interactions and improve predictive accuracy. Classical approaches include Decision Tree (DT), Logistic Regression (LR), Naive Bayes (NB), K-Nearest Neighbors (KNN), Gradient Boosting (GB), and ElasticNet, while deep learning-based models encompass Multilayer Perceptron (MLP), CNN, ResNet, RNN, Transformers, and KAN. Empirical analysis demonstrates that the 8:1:1 dataset split consistently achieves superior performance compared to alternative ratios. Consequently, this configuration has been adopted as the standard for all subsequent experiments to ensure consistency and robustness. For hyperparameter optimization, we employed GridSearchCV to evaluate parameter configurations systematically through cross-validation. Instead of conducting an exhaustive search, parameter ranges were refined based on observed patterns, reducing the search space while maintaining the ability to converge to optimal solutions. However, grid search is computationally expensive, particularly when dealing with high-dimensional feature spaces. Lin et al. [75] proposed a Fast Dual Subgradient Optimization method to efficiently optimize stochastic kernels, which has potential applications in machine learning hyperparameter tuning. By leveraging kernel-based optimization, it may be possible to reduce the search complexity while maintaining or even improving model accuracy. Future work could explore integrating these optimization techniques into the CHD prediction pipeline. This strategy enhances computational efficiency and minimizes the likelihood of suboptimal selections.

For the traditional machine learning models, the optimal hyperparameter configurations are shown in Table 1, and the quantitative results are shown in Table 2. The results highlight variations in model performance, with some algorithms benefiting from feature processing while others showing marginal or negative changes.

Optimal Hyperparameters for Traditional Machine Learning Models	
Model	**Hyperparameters**
DT	criterion = ‘entropy’, max_depth = 40, max_features = ‘sqrt’, min_samples_leaf = 1, min_samples_split = 2
LR	estimator_C = 1, solver = ‘liblinear’, estimator_max_iter = 1000
NB	alpha = 10, binarize = 0.5
KNN	n_neighbors = 80, p = 1, weight = ‘distance’
GB	learning_rate = 0.1, n_estimators = 500
ElasticNet	C = 50, l1_ratio = 0.0001

Table 1: Optimal hyperparameter settings for traditional machine learning models derived via grid search optimization.

According to Table 2, the six traditional machine learning models exhibit varying predictive capabilities in CHD prediction. GB achieves the highest overall performance, demonstrating strong generalization ability due to its ensemble nature, which iteratively refines predictions by correcting previous errors. DT follows, benefiting from its hierarchical splitting strategy, though it is more prone to overfitting compared to GB. Logistic Regression LR and ElasticNet, which rely on linear decision boundaries and regularization,

Traditional Machine Learning Models (Original Data)					
Method	**Acc ↑**	**Prec ↑**	**Recall ↑**	**AUC ↑**	**F1 ↑**
DT	0.752	0.274	0.380	0.600	0.318
LR	0.672	0.259	0.620	0.703	0.365
NB	0.685	0.257	0.566	0.681	0.354
KNN	0.612	0.237	0.698	0.684	0.354
GB	0.860	0.556	0.388	0.771	0.457
ElasticNet	0.670	0.254	0.605	0.700	0.358

Traditional Machine Learning Models (Feature Processing)					
Method	**Acc ↑**	**Prec ↑**	**Recall ↑**	**AUC ↑**	**F1 ↑**
DT	0.758	0.317	0.512	0.657	0.392
LR	0.671	0.255	0.605	0.697	0.359
NB	0.672	0.246	0.558	0.677	0.341
KNN	0.624	0.225	0.605	0.676	0.328
GB	0.874	0.634	0.403	0.784	0.493
ElasticNet	0.671	0.256	0.612	0.697	0.362

Table 2: Performance comparison of traditional machine learning models before and after feature processing.

perform moderately well but are limited in capturing complex feature interactions. KNN and NB, on the other hand, generally perform less effectively. KNN's reliance on distance metrics makes it sensitive to feature scaling and distribution, while NB's assumption of feature independence may not hold for complex medical datasets.

Feature processing had varying effects on different models. Tree-based models like DT and GB improved as noise reduction and feature selection enhanced decision boundaries and generalization. GB further benefited from ensemble learning, refining predictions iteratively. In contrast, KNN and NB were negatively impacted; changes in feature distributions disrupted KNN's distance-based classification, while increased feature correlations weakened NB's independence assumption. Regularized models like LR and ElasticNet showed minimal change as they inherently mitigate redundant features.

For the deep learning models, the optimal hyperparameter configurations are shown in Table 3 and the quantitative results are shown in Table 4. According to Table 4, the deep learning models exhibit varying predictive performance on CHD prediction. KAN and MLP achieved relatively higher overall performance, benefiting from their ability to model complex feature interactions and nonlinear relationships effectively. KAN, in particular, demonstrated strong generalization ability, likely due to its structured representation learning, which enables more flexible function approximation. Transformer-based models showed moderate performance, indicating their capability to capture temporal or long-range dependencies in the dataset, though they may require larger data volumes for optimal effectiveness. CNN and RNN exhibited lower accuracy compared to other deep learning models, possibly due to their sensitivity to feature dependencies and the nature of tabular medical data, which may not align well with their architecture originally designed for sequential or spatial information. ResNet showed the lowest predictive performance, which suggests that its deep hierarchical structure may not be well-suited for CHD prediction tasks, likely due to the lack of spatially correlated features in the dataset.

Feature processing had mixed effects on the performance of deep learning models. KAN and MLP exhibited performance improvements, likely due to the removal of noisy or irrelevant features, allowing these models to focus on more informative patterns. Feature processing particularly benefited KAN, as its function approximation capabilities were enhanced with cleaner and more representative data. Transformer models also saw slight improvements, suggesting that refined feature representations helped stabilize their learning dynamics. Conversely, CNN and RNN experienced performance declines, which may be attributed to changes in feature distributions that disrupted their ability to capture meaningful patterns. CNNs, typically designed for spatial data, may have struggled with transformed tabular data, while RNNs could have been affected by altered feature relationships. ResNet's performance remained relatively poor even after feature processing, likely due to its deep structure not aligning well with the dataset's characteristics.

Optimal Hyperparameters for Deep Learning Models	
Model	**Hyperparameters**
MLP	learning_rate = 0.001, hidden_sizes = [128, 64, 32], activation = 'relu', alpha = 0.001
Transformer	learning_rate = 0.001, embed_dim = 16, nhead = 4, num_layers = 1, dim_feedforward = 64, activation = 'relu'
KAN	learning_rate = 0.001, Q = 2, hidden_dim = 64, activation = 'relu'
RNN	learning_rate = 0.001, hidden_size = 64, num_layers = 2, activation = 'relu'
CNN	learning_rate = 0.001, batch_size = 64, num_filters = 32, kernel_size = 3, activation = 'relu'
ResNet	learning_rate = 0.001, batch_size = 64, num_blocks = 2, hidden_dim = 32, kernel_size = 3, activation = 'relu'

Table 3: Optimal hyperparameter settings for deep learning models derived via grid search optimization.

Deep Learning Models (Original Data)					
Method	**Acc ↑**	**Prec ↑**	**Recall ↑**	**AUC ↑**	**F1 ↑**
MLP	0.754	0.273	0.372	0.647	0.315
Transformer	0.669	0.250	0.589	0.677	0.351
KAN	0.756	0.329	0.581	0.768	0.420
RNN	0.663	0.244	0.581	0.696	0.344
CNN	0.587	0.217	0.659	0.672	0.327
ResNet	0.153	0.151	0.992	0.568	0.263

Deep Learning Models (Feature Processing)					
Method	**Acc ↑**	**Prec ↑**	**Recall ↑**	**AUC ↑**	**F1 ↑**
MLP	0.767	0.298	0.395	0.649	0.340
Transformer	0.685	0.252	0.543	0.684	0.344
KAN	0.789	0.370	0.550	0.771	0.442
RNN	0.640	0.240	0.628	0.694	0.347
CNN	0.574	0.224	0.729	0.661	0.342
ResNet	0.206	0.161	1.000	0.668	0.277

Table 4: Performance comparison of deep learning models before and after feature processing.

In conclusion, traditional machine learning models, particularly tree-based methods, provided strong baseline performance with clear interpretability, while deep learning models like KAN and MLP showed greater flexibility in capturing complex feature interactions. Feature processing significantly improved tree-based models but had mixed effects on distance-based and probabilistic classifiers. Deep learning models exhibited varied responses, with some benefiting from refined feature representations while others, like CNN and RNN, struggled with tabular data. Overall, selecting the appropriate model and preprocessing strategy is crucial for optimizing CHD prediction performance.

6. Conclusion and Future Work

This study introduces FGSCare, integrating traditional machine learning classifiers and deep learning models for CHD prediction. A structured data preprocessing pipeline, including feature engineering and SMOTE for class imbalance mitigation, contributed to improved model performance. Additionally, hyperparameter tuning using GridSearchCV further enhanced predictive accuracy. Ex-

perimental results indicate that GB achieved the best performance among traditional models, while KAN outperformed other deep learning methods, demonstrating strong generalization ability. Model interpretability analysis using SHAP provided valuable insights into feature importance, reinforcing the necessity of informed feature selection strategies. These findings highlight the importance of ensemble learning, model interpretability, and the balance between precision and recall in real-world predictive tasks.

This study employs traditional feature engineering to enhance model performance, yet improvements remain. CNN and RNN struggled with structured tabular data, suggesting future work could integrate hybrid models, such as CNN with Vision Transformers (ViT), to improve feature extraction [76]. Extending this approach to structured medical data may enhance predictive performance. Automated feature selection, including self-supervised and reinforcement learning, can improve adaptability and mitigate biases. Recent reinforcement learning-based methods effectively balance feature diversity and relevance [77], offering potential for CHD prediction. Further exploration of causal inference and knowledge graphs could uncover hidden feature relationships, enhancing accuracy and interpretability. Refining model architectures and feature extraction strategies will strengthen predictive robustness across medical datasets.

References

[1] S. Nalluri, R. Saraswathi, S. Ramasubbareddy, K. Govinda, E. Swetha, Chronic Heart Disease Prediction Using Data Mining Techniques, in: Data Engineering and Communication Technology, Springer, 2020, pp. 903–912.

[2] M. Singh, A. Kumar, N. N. Khanna, J. R. Laird, A. Nicolaides, G. Faa, A. M. Johri, L. E. Mantella, J. F. E. Fernandes, J. S. Teji, et al., Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review, EClinicalMedicine 73 (2024).

[3] Smith, S.C. and Collins, A. and Ferrari, R. and Holmes, D.R. and Logstrup, S. and McGhie, D.V., Our time: A call to save preventable death from cardiovascular disease (heart disease and stroke), European Heart Journal 39 (4) (2018) 279–289.

[4] Wilson, P.W. and D'Agostino, R.B. and Levy, D. and Belanger, A.M. and Silbershatz, H. and Kannel, W.B., Prediction of coronary heart disease using risk factor categories, Circulation 97 (18) (1998) 1837–1847.

[5] Lloyd-Jones, D.M. and Nam, B.H. and D'Agostino, R.B. and Levy, D. and Murabito, J.M. and Wang, T.J., Implications of age and sex distributions for cardiovascular risk factor prediction: Insights from the framingham heart study, JAMA Cardiology 74 (2019) 674–681.

[6] A. Rajkomar, J. Dean, I. Kohane, Machine learning in medicine, New England Journal of Medicine 380 (14) (2019) 1347–1358.

[7] H. Guo, J. Li, H. Liu, J. He, Learning dynamic treatment strategies for coronary heart diseases by artificial intelligence: real-world data-driven study, BMC Medical Informatics and Decision Making 22 (1) (2022) 39.

[8] K. Kourou, T. Exarchos, K. Exarchos, M. Karamouzis, D. Fotiadis, Machine learning applications in cancer prognosis and prediction, Computational and Structural Biotechnology Journal 13 (2015) 8–17.

[9] M. Pal, S. Parija, Prediction of Heart Diseases using Random Forest, Journal of Physics: Conference Series 1817 (1) (2021) 012009. doi:10.1088/1742-6596/1817/1/012009.
URL https://iopscience.iop.org/article/10.1088/1742-6596/1817/1/012009

[10] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, N. Elhadad, Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission, in: Proc. 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2015, pp. 1721–1730.

[11] S. S. Kumar, S. S. Manogaran, R. Varatharajan, Intelligent Bi-LSTM with Architecture Optimization for Heart Disease Prediction in Wireless Body Area Networks, Computers, Materials & Continua 66 (3) (2021) 3311–3327. doi:10.32604/cmc.2021.014684.

[12] I. Chen, F. Johansson, D. Sontag, Why is my classifier discriminatory?, Advances in Neural Information Processing Systems 34 (2021) 20475–20488.

[13] W. Zhang, et al., Weighted recursive feature selection for chd prediction, Journal of Biomedical Informatics 139 (2023) 104276. doi:10.1016/j.jbi.2023.104276.

[14] A. Pathan, et al., Feature selection techniques in medical data science, Machine Learning in Healthcare 5 (2) (2022) 123–135. doi:10.1007/s41666-022-00123-4.

[15] M. S. Pathan, A. Nag, M. M. Pathan, S. Dev, Analyzing the impact of feature selection on the accuracy of heart disease prediction, Healthcare Analytics 2 (2022) 100060. doi:10.1016/j.health.2022.100060.

[16] Y. Dun, K. Tu, C. Chen, C. Hou, X. Yuan, Kan: Knowledge-aware attention network for fake news detection, Proceedings of the AAAI Conference on Artificial Intelligence 35 (1) (2021) 81–89. doi:10.1609/aaai.v35i1.16080.

[17] J. Xu, Z. Chen, J. Li, S. Yang, H. Wang, E. C. Ngai, AlignGroup: Learning and Aligning Group Consensus with Member Preferences for Group Recommendation, in: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, 2024, pp. 2682–2691. doi:https://doi.org/10.1145/3627673.3679697.

[18] J. Xu, Z. Chen, S. Yang, J. Li, H. Wang, E. C.-H. Ngai, MENTOR: Multi-level Self-supervised Learning for Multimodal Recommendation, arXiv preprint arXiv:2402.19407 (2024). doi:https://doi.org/10.48550/arXiv.2402.19407.

[19] H. Wang, Y. Li, K. Gong, M. S. Pathan, S. Xi, B. Zhu, Z. Wen, S. Dev, MFCSNet: A Musician–Follower Complex Social Network for Measuring Musical Influence, Entertainment Computing 48 (2024) 100601. doi:https://doi.org/10.1016/j.entcom.2023.100601.

[20] D. Dey, S. Ghosh, S. Biswas, et al., Machine learning based prediction models for cardiovascular diseases, European Heart Journal-Digital Health 6 (1) (2023) 7–17. doi:10.1093/ehjdh/ztad002.

[21] C. Krittanawong, K. W. Johnson, R. S. Rosenson, et al., Machine learning prediction in cardiovascular diseases: a meta-analysis, Scientific Reports 10 (1) (2020) 16057. doi:10.1038/s41598-020-72685-1.

[22] A. Alim, M. M. Rahman, S. Nasrin, et al., A systematic review on machine learning approaches for cardiovascular disease prediction, Health Information Science and Systems 10 (1) (2022) 1–16. doi:10.1007/s13755-022-00173-3.

[23] Y. Li, H. Wang, J. Xu, Z. Ma, P. Wu, S. Wang, S. Dev, CP2M: Clustered-Patch-Mixed Mosaic Augmentation for Aerial Image Segmentation, arXiv preprint arXiv:2501.15389 (2025).

[24] H. Wang, M. S. Pathan, Y. H. Lee, S. Dev, Day-ahead Forecasts of Air Temperature, in: 2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium), 2021, pp. 94–95. doi:10.23919/USNC-URSI51813.2021.9703606.

[25] Y. Li, H. Wang, S. Wang, Y. H. Lee, M. Salman Pathan, S. Dev, UCloudNet: A Residual U-Net with Deep Supervision for Cloud Image Segmentation, in: IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, 2024, pp. 5553–5557. doi:10.1109/IGARSS53475.2024.10640450.

[26] Y. Li, H. Wang, J. Xu, P. Wu, Y. Xiao, S. Wang, S. Dev, DDUNet: Dual Dynamic U-Net for Highly-Efficient Cloud Segmentation, arXiv preprint arXiv:2501.15385 (2025). doi:https://doi.org/10.48550/arXiv.2501.15385.

[27] S. Chen, H. He, Stock prediction using convolutional neural network, IOP Conference Series: Materials Science and Engineering 435 (2018) 012026. doi:10.1088/1757-899X/435/1/012026.

[28] M. Veres, M. Moussa, Deep learning for intelligent transportation systems: A survey of emerging trends, IEEE Transactions on Intelligent Transportation Systems PP (2019) 1–17. doi:10.1109/TITS.2019.2929020.

[29] C. Krittanawong, H. Zhang, Z. Wang, M. Aydar, T. Kitai, Artificial intelligence in precision cardiovascular medicine, Journal of the American College of Cardiology 69 (2017) 2657 – 6 4. doi:10.1016/j.jacc.2017.03.571.

[30] N. D. Wong, Epidemiological studies of CHD and the evolution of preventive cardiology, Nature Reviews Cardiology 11 (5) (2014) 276–289.

[31] C. Kingsford, S. L. Salzberg, What are decision trees?, Nature biotechnology 26 (9) (2008) 1011–1013.

[32] T. G. Nick, K. M. Campbell, Logistic regression, Topics in biostatistics (2007) 273–301.

[33] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, KNN model-based approach in classification, in: On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings, Springer, 2003, pp. 986–996.

[34] J. H. Friedman, Stochastic gradient boosting, Computational statistics & data analysis 38 (4) (2002) 367–378.

[35] M. Zreik, R. W. van Hamersvelt, J. M. Wolterink, T. Leiner, M. A. Viergever, I. Išgum, A recurrent cnn for automatic detection and classification of coronary artery plaque and stenosis in coronary ct angiography, IEEE Transactions on Medical Imaging 38 (7) (2019) 1588–1598. doi:10.1109/TMI.2018.2883807.
URL https://pubmed.ncbi.nlm.nih.gov/30507498/

[36] A. Dutta, T. Batabyal, M. Basu, S. T. Acton, An efficient convolutional neural network for coronary heart disease prediction, Expert Systems with Applications 159 (2020) 113408.

[37] R. D'Agostino, R. Vasan, M. Pencina, P. Wolf, M. Cobain, J. Massaro, W. Kannel, General cardiovascular risk profile for use in primary care: The framingham heart study, Circulation 117 (2008) 743–53. doi:10.1161/CIRCULATIONAHA.107.699579.

[38] T. N. Rincy, R. Gupta, Ensemble learning techniques and its efficiency in machine learning: A survey, in: 2nd international conference on data, engineering and applications (IDEA), IEEE, 2020, pp. 1–6.

[39] I. D. Mienye, Y. Sun, A survey of ensemble learning: Concepts, algorithms, applications, and prospects, Ieee Access 10 (2022) 99129–99149.

[40] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–44. doi:10.1038/nature14539.

[41] Y. Wang, Y. Fan, P. Bhatt, C. Davatzikos, High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables, Neuroimage 50 (4) (2010) 1519–1535.

[42] R. Miotto, F. Wang, S. Wang, X. Jiang, J. T. Dudley, Deep learning for healthcare: review, opportunities and challenges, Briefings in bioinformatics 19 (6) (2018) 1236–1246.

[43] K. Cui, R. Li, S. L. Polk, Y. Lin, H. Zhang, J. M. Murphy, R. J. Plemmons, R. H. Chan, Superpixel-based and spatially-regularized diffusion learning for unsupervised hyperspectral image clustering, IEEE Transactions on Geoscience and Remote Sensing (2024).

[44] Z. Li, H. Wang, Y. Li, S. Dev, G. Zuo, VGRISys: A Vision-Guided Robotic Intelligent System for Autonomous Instrument Calibration*, in: 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2023, pp. 1–6. doi:10.1109/ROBIO58561.2023.10354843.

[45] H. Wang, B. Zhu, Y. Li, K. Gong, Z. Wen, S. Wang, S. Dev, SYGNet: A SVD-YOLO based GhostNet for Real-time Driving Scene Parsing, in: 2022 IEEE International Conference on Image Processing (ICIP), 2022, pp. 2701–2705.

[46] J. You, H. Wang, Y. Li, M. Huo, L. V. T. Ha, M. Ma, J. Xu, P. Wu, S. Garg, W. Pu, Multi-cali anything: Dense feature multi-frame structure-from-motion for large-scale camera array calibration, arXiv preprint arXiv:2503.00737 (2025).

[47] W. Tang, K. Cui, R. H. Chan, Optimized hard exudate detection with supervised contrastive learning, in: 2024 IEEE International Symposium on Biomedical Imaging (ISBI), IEEE, 2024, pp. 1–5.

[48] F. Pan, Y. Wu, K. Cui, S. Chen, Y. Li, Y. Liu, A. Shakoor, H. Zhao, B. Lu, S. Zhi, et al., Accurate detection and instance segmentation of unstained living adherent cells in differential interference contrast images, Computers in Biology and Medicine 182 (2024) 109151.

[49] Y. Li, H. Wang, Z. Li, S. Wang, S. Dev, G. Zuo, DAANet: Dual Attention Aggregating Network for Salient Object Detection*, in: 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, 2023, pp. 1–7. doi:10.1109/ROBIO58561.2023.10354933.

[50] Y. Li, H. Wang, A. Katsaggelos, CPDR: Towards Highly-Efficient Salient Object Detection via Crossed Post-decoder Refinement, in: 35th British Machine Vision Conference 2024, BMVC 2024, Glasgow, UK, November 25-28, 2024, BMVA, 2024.

[51] X. Zhu, R. Tian, C. Xu, M. Huo, W. Zhan, M. Tomizuka, M. Ding, Fanuc manipulation: A dataset for learning-based manipulation with fanuc mate 200id robot (2023).

[52] M. Huo, M. Ding, C. Xu, T. Tian, X. Zhu, Y. Mu, L. Sun, M. Tomizuka, W. Zhan, Human-oriented Representation Learning for Robotic Manipulation, ArXiv abs/2310.03023 (2023).

[53] H. Lin, Y. Wang, M. Huo, C. Peng, Z. Liu, M. Tomizuka, Joint Pedestrian Trajectory Prediction through Posterior Sampling, 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2024) 5672–5679.

[54] H. Wang, M. S. Pathan, S. Dev, Stereo Matching Based on Visual Sensitive Information, in: 2021 6th International Conference on Image, Vision and Computing (ICIVC), 2021, pp. 312–316. doi:10.1109/ICIVC52351.2021.9527014.

[55] H. Wang, Y. Li, S. Xi, S. Wang, M. S. Pathan, S. Dev, AMDCNet: An attentional multi-directional convolutional network for stereo matching, Displays 74 (2022) 102243. doi:https://doi.org/10.1016/j.displa.2022.102243.

[56] S. Batra, H. Wang, A. Nag, P. Brodeur, M. Checkley, A. Klinkert, S. Dev, DMCNet: Diversified model combination network for understanding engagement from video screengrabs, Systems and Soft Computing 4 (2022) 200039.

[57] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE transactions on neural networks and learning systems 33 (12) (2021) 6999–7019.

[58] L. R. Medsker, L. Jain, et al., Recurrent neural networks, Design and Applications 5 (64-67) (2001) 2.

[59] A. Graves, A. Graves, Long short-term memory, Supervised sequence labelling with recurrent neural networks (2012) 37–45.

[60] E. Choi, A. Schuetz, W. F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure from longitudinal electronic health record data, Journal of the American Medical Informatics Association 24 (2) (2017) 361–370. doi:10.1093/jamia/ocw112.

[61] Z. Jiang, Machine learning models in cardiovascular disease prediction: A comprehensive review, Journal of Intelligent Healthcare Systems 1 (1) (2024) 26–38. doi:10.52810/JIR.2024.003.

[62] G. Saranya, A. Pravin, A novel feature selection approach with integrated feature sensitivity and feature correlation for improved prediction of heart disease, Journal of Ambient Intelligence and Humanized Computing 14 (9) (2023) 12005–12019.

[63] B. F. Darst, K. C. Malecki, C. D. Engelman, Using recursive feature elimination in random forest to account for correlated variables in high dimensional data, BMC genetics 19 (2018) 1–6.

[64] N. Hoque, D. K. Bhattacharyya, J. K. Kalita, MIFS-ND: A mutual information-based feature selection method, Expert systems with applications 41 (14) (2014) 6371–6385.

[65] U. Maulik, S. Bandyopadhyay, Genetic algorithm-based clustering technique, Pattern recognition 33 (9) (2000) 1455–1465.

[66] F. F. Firdaus, H. A. Nugroho, I. Soesanti, A review of feature selection and classification approaches for heart disease prediction, International Journal of Information Technology and Electrical Engineering 4 (3) (2020) 78–83.
URL https://www.researchgate.net/publication/354166578_A_Review_of_Feature_Selection_and_Classification_Approaches_for_Heart_Disease_Prediction

[67] S. Lundberg, A unified approach to interpreting model predictions, arXiv preprint arXiv:1705.07874 (2017).

[68] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.

[69] A. Chaddad, J. Peng, J. Xu, A. Bouridane, Survey of explainable AI techniques in healthcare, Sensors 23 (2) (2023) 634.

[70] S. Lundberg, G. Erion, H. Chen, A. DeGrave, J. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, Explainable ai for trees: From local explanations to global understanding (05 2019). doi:10.48550/arXiv.1905.04610.

[71] V. H P, B. Poornima, B. Sagar, Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset, 2018, pp. 511–518. doi:10.1007/978-981-10-7563-6_53.

[72] A. Abuzaid, E. Alkronz, A comparative study on univariate outlier winsorization methods in data science context, Italian Journal of Applied Statistics 36 (2024) 85–99. doi:10.26398/IJAS.0036-004.

[73] Z. Lin, A. Ruszczyński, An integrated transportation distance between kernels and approximate dynamic risk evaluation in markov systems, SIAM Journal on Control and Optimization 61 (6) (2023) 3559–3583. doi:10.1137/22m1530665.
URL http://dx.doi.org/10.1137/22M1530665

[74] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, D. John, A predictive analytics approach for stroke prediction using machine learning and neural networks, journal = Healthcare Analytics 2 (2022) 100032. doi:https://doi.org/10.1016/j.health.2022.100032.

[75] Z. Lin, A. Ruszczynski, Fast dual subgradient optimization of the integrated transportation distance between stochastic kernels (2023). arXiv:2312.01432.
URL https://arxiv.org/abs/2312.01432

[76] R. Yulvina, S. A. Putra, M. Rizkinia, A. Pujitresnani, E. D. Tenda, R. E. Yunus, D. H. Djumaryo, P. A. Yusuf, V. Valindria, Hybrid vision transformer and convolutional neural network for multi-class and multi-label classification of tuberculosis anomalies on chest x-ray, Computers 13 (12) (2024) 343. doi:10.3390/computers13120343.
URL https://doi.org/10.3390/computers13120343

[77] W. Fan, K. Liu, H. Liu, P. Wang, Y. Ge, Y. Fu, Autofs: Automated feature selection via diversity-aware interactive reinforcement learning, arXiv preprint arXiv:2008.12001 (2020).
URL https://arxiv.org/abs/2008.12001