information

Article

Benchmarking Variants of Recursive Feature Elimination: Insights from Predictive Tasks in Education and Healthcare

Okan Bulut [1,*], Bin Tan [2], Elisabetta Mazzullo [2] and Ali Syed [3]

[1] Centre for Research in Applied Measurement and Evaluation, Faculty of Education, University of Alberta, Edmonton, AB T6G 2G5, Canada
[2] Measurement, Evaluation, and Data Science, Faculty of Education, University of Alberta, Edmonton, AB T6G 2G5, Canada; btan4@ualberta.ca (B.T.); mazzullo@ualberta.ca (E.M.)
[3] Pharmacology, Faculty of Science, University of Alberta, Edmonton, AB T6G 2G5, Canada; alisyed1@ualberta.ca
* Correspondence: bulut@ualberta.ca

Abstract: Originally developed as an effective feature selection method in healthcare predictive analytics, Recursive Feature Elimination (RFE) has gained increasing popularity in Educational Data Mining (EDM) due to its ability to handle high-dimensional data and support interpretable modeling. Over time, various RFE variants have emerged, each introducing methodological enhancements. To help researchers better understand and apply RFE more effectively, this study organizes existing variants into four methodological categories: (1) integration with different machine learning models, (2) combinations of multiple feature importance metrics, (3) modifications to the original RFE process, and (4) hybridization with other feature selection or dimensionality reduction techniques. Rather than conducting a systematic review, we present a narrative synthesis supported by illustrative studies from EDM to demonstrate how different variants have been applied in practice. We also conduct an empirical evaluation of five representative RFE variants across two domains: a regression task using a large-scale educational dataset and a classification task using a clinical dataset on chronic heart failure. Our evaluation benchmarks predictive accuracy, feature selection stability, and runtime efficiency. Results show that the evaluation metrics vary significantly across RFE variants. For example, while RFE wrapped with tree-based models such as Random Forest and Extreme Gradient Boosting (XGBoost) yields strong predictive performance, these methods tend to retain large feature sets and incur high computational costs. In contrast, a variant known as Enhanced RFE achieves substantial feature reduction with only marginal accuracy loss, offering a favorable balance between efficiency and performance. These findings underscore the trade-offs among accuracy, interpretability, and computational cost across RFE variants, providing practical guidance for selecting the most appropriate algorithm based on domain-specific needs and constraints.

Keywords: feature selection; educational data mining; dimensionality; recursive feature elimination; healthcare

check for updates

Academic Editor: Francesco Fontanella

Received: 17 April 2025
Revised: 31 May 2025
Accepted: 4 June 2025
Published: 6 June 2025

Citation: Bulut, O.; Tan, B.; Mazzullo, E.; Syed, A. Benchmarking Variants of Recursive Feature Elimination: Insights from Predictive Tasks in Education and Healthcare. *Information* **2025**, *16*, 476. https://doi.org/10.3390/info16060476

1. Introduction

The widespread adoption of educational technologies and online learning systems has significantly increased the volume and diversity of data available to educational institutions and researchers, creating what has been described as a "goldmine of educational data" [1]. These datasets often encompass a wide array of features, including administrative records, demographic information, digital traces of student interactions, and affective indicators collected through self-report instruments [1]. This rich and multifaceted data landscape offers promising opportunities to enhance learning experiences, improve educational outcomes, and guide more effective administrative decision-making [2]. In response to the growing need to extract meaningful insights from such complex data, the interdisciplinary field of Educational Data Mining (EDM) has emerged. EDM applies statistical analysis, data mining, and machine learning (ML) techniques to address pressing educational research challenges [3]. Key areas of focus include predicting student performance, generating personalized learning recommendations, refining course design, and informing strategic institutional planning [4–6].

Although rich and multifaceted educational datasets offer valuable opportunities for research, leveraging their full potential presents several challenges related to data management, processing, and analysis. A key issue is the high dimensionality often found in educational data, which complicates the use of standard machine learning (ML) algorithms in EDM. This problem is especially acute in cases where researchers work with relatively small sample sizes, such as data collected from individual classrooms or small-scale studies, yet have to contend with many features. This imbalance, commonly referred to as data sparsity, results in too few observations per feature, which weakens the reliability of ML models and can reduce predictive accuracy [7]. Moreover, high dimensionality introduces noise and increases the risk of overfitting, ultimately making models more complex, computationally demanding, harder to interpret, and less generalizable [8]. To address these challenges, researchers must adopt effective feature selection and dimensionality reduction techniques, along with modeling approaches tailored to the unique structure of educational data.

Statisticians and data mining researchers have developed diverse methods to effectively manage the challenges associated with high-dimensional datasets. Broadly, these approaches can be categorized into two main groups: (1) dimensionality reduction techniques, which generate p new features through linear or nonlinear transformations of the original n features, with $p < n$; and (2) feature selection methods, which identify and retain a subset of the most relevant original features. Dimensionality reduction methods, such as Principal Component Analysis (PCA), are powerful tools for simplifying the feature space by transforming original attributes into fewer composite features. However, a significant limitation of these methods is the potential loss of interpretability, as the transformed features often lack a clear, intuitive relationship with the original variables [9]. Conversely, in educational contexts, the interpretability and explainability of ML models are highly valued because they enhance transparency, build trust, promote fairness, and support informed decision-making among stakeholders who may not have technical expertise [10]. For this reason, feature selection methods, which build models using carefully chosen subsets of the original features, are often preferred in EDM.

Feature selection techniques can be broadly categorized into four groups: filter methods, wrapper methods, embedded methods, and hybrid methods that combine elements of the first three [11–14]. Among them, wrapper-based methods, such as Recursive Feature Elimination (RFE), directly leverage machine learning algorithms to evaluate subsets of features based on predictive performance. Originally developed in the healthcare domain to identify relevant gene expressions for cancer classification [13], RFE operates by iteratively removing the least important features and retaining those that best predict the target variable. Although wrapper methods are generally more computationally intensive than filter methods, they are widely recognized for their effectiveness in identifying influential features, leading to improved predictive accuracy [14]. RFE in particular has gained popularity due to its transparent and interpretable process, and it has been extensively applied in healthcare analytics, especially in high-dimensional biomedical data settings.

In contrast, the adoption of RFE in EDM has been more gradual. While RFE's strengths, especially its adaptability and performance in complex data environments, make it well-suited for educational research, recent reviews suggest that it remains underutilized in EDM [10,15]. To support EDM researchers in applying RFE more effectively, this study introduces the original RFE algorithm and presents a narrative synthesis of its variants, categorized into four methodological types based on their design enhancements. We highlight how each variant differs from the original method and illustrate their practical applications using representative examples from the EDM literature. Additionally, we conduct an empirical evaluation of five RFE variants across two distinct datasets, offering comparative insights into their performance, runtime, and stability. These contributions serve as both a conceptual and practical resource for researchers navigating the expanding landscape of RFE techniques. Collectively, our analyses aim to assist researchers in selecting and applying the most appropriate RFE algorithm for their specific research contexts.

2. A Narrative Review of RFE Algorithms and Their Applications in EDM

In this section, we begin by outlining the original RFE algorithm as introduced by Guyon et al. [13]. Next, we survey the diverse array of RFE variants developed to enhance its efficacy, scalability, or adaptability to specific contexts. These variants are categorized based on their difference from the original algorithm in terms of (1) the ML models employed for assessing and ranking feature importance, (2) the combinations of ML models and feature importance metrics, (3) modifications to the RFE process, or (4) hybrids of RFE with other feature selection or dimensionality reduction methods. Since many RFE algorithms, including the original, were initially developed and widely adopted in healthcare domains such as genomics, this review focuses on their application within the field of EDM. We take this approach for two reasons. First, although RFE has a longer history in healthcare domains, its adoption in EDM is more recent and less systematically documented. Second, the increasing availability of large-scale educational datasets, enabled by advancements in educational technologies, digital learning systems, and large-scale assessments, has created a growing need for effective feature selection methods in EDM research. To support future EDM researchers in using RFE more effectively, this review aims to illustrate how RFE variant methods have been applied to address challenges in analyzing educational data.

2.1. The Original RFE Algorithm

The RFE algorithm was initially introduced by Guyon et al. [13] in the context of gene selection for disease classification. Figure 1 illustrates the original RFE process. It begins by building an ML predictive model that includes the complete set of features. In the subsequent step, the importance of each feature is identified and assessed. The nature and information of feature importance vary depending on the ML model used (e.g., regression coefficients for linear models and feature relative importance for tree-based models). Once the importance of each feature is determined, the features are ranked accordingly. The least important features are then removed from the dataset. The next step is to check whether the algorithm should be terminated based on stopping criteria, such as the predefined number of features remaining in the dataset or if removing features no longer improves the model's prediction performance. If the stopping criteria are not met, a new predictive model is built using the remaining features. Therefore, this process of training the model, determining feature importance, ranking features, and dropping the least important ones is repeated until the stopping criteria are met, leaving a short but the most important subset of features that contribute the most to the predictive model.

The recursive process employed by RFE exemplifies backward feature elimination [13]. Initially, RFE trains ML models using the complete set of features, subsequently iteratively retraining models with progressively fewer features. This iterative process enables a more thorough assessment of feature importance compared to single-pass approaches, as feature relevance is continuously reassessed after removing the influence of less critical attributes [16]. Consequently, RFE is recognized as a greedy search strategy, as it does not explore all possible feature combinations exhaustively but rather selects locally optimal features at each iteration, aiming toward a globally optimal feature subset [17]. This greedy methodology substantially enhances computational efficiency compared to exhaustive evaluations, which can quickly become computationally infeasible due to the exponential growth of potential feature subsets as dataset dimensionality increases [18].

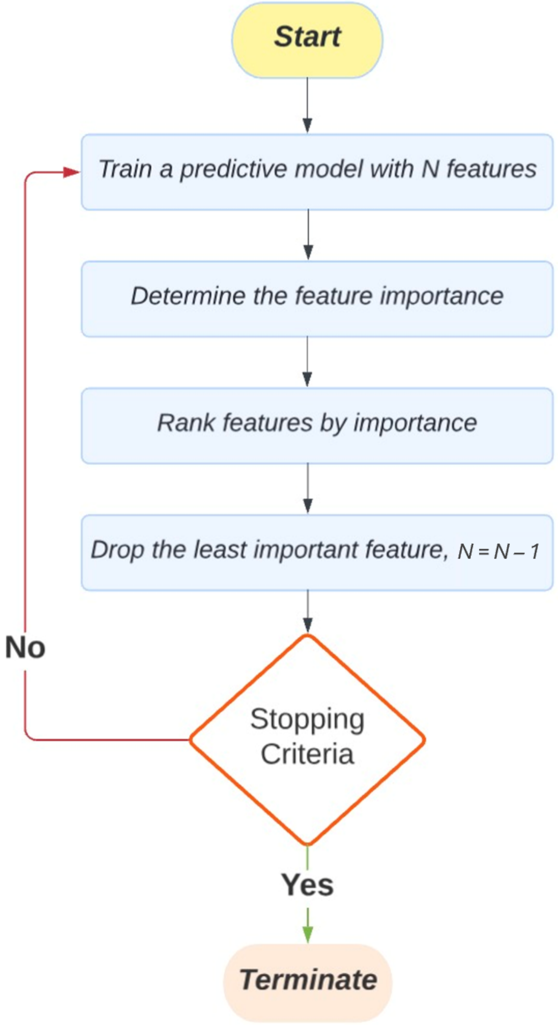
Start
Train a predictive model with N features
Determine the feature importance
Rank features by importance
Drop the least important feature, N = N – 1
No
Stopping Criteria
Yes
Terminate

Figure 1. The process of the original recursive feature elimination algorithm.

RFE has demonstrated its effectiveness as a feature selection approach, offering advantages such as dimensionality reduction, improved model accuracy and interpretability, and greater computational efficiency relative to exhaustive feature evaluations. These benefits have contributed significantly to RFE's popularity within EDM research. For instance, Yeung and Yeung [19] utilized a comprehensive educational dataset to predict whether students would pursue STEM or non-STEM careers post-graduation. After augmenting the original dataset with features extracted via deep knowledge tracing, several dimensionality management and overfitting prevention strategies—including PCA, RFE, and multiple oversampling methods—were compared. Among these, RFE emerged as the most effective strategy for mitigating overfitting and enhancing predictive model performance. Similarly, Pereira et al. [20] employed RFE for predicting student dropout rates. After balancing class labels through undersampling to address overfitting concerns, RFE retained only 5 of the original 20 features, yielding a high classification accuracy of 80%. This reduced feature set also facilitated a meaningful interpretation of student behaviors associated with adverse educational outcomes, enabling targeted recommendations and interventions.

2.2. Four Major Types of RFE Variants

We categorize the variants of RFE into four main types. Table 1 summarizes the defining characteristics and representative examples of each type. In the following sections, we provide a more detailed explanation of each type, accompanied by exemplar studies that illustrate their applications in EDM research.

Table 1. Summary of characteristics and examples of the four RFE variant types.

RFE Variant Type	Description	Examples
RFE wrapped with different ML models	Using the original RFE process to eliminate features based on feature importance metrics computed by an algorithm other than SVR/SVM.	DT-RFE [21] RF-RFE [22]
Combination of ML models or feature importance metrics	Using the original RFE process to eliminate features based on multiple feature importance metrics (i.e., feature importance values from distinct ML models or different feature importance metrics from the same model).	SVM, RF, and generalized boosted regression algorithms [23]; SVM, LR, and DT ensemble [24]; SVM-NB hybrid classifier [25]

Modification to the RFE process	Changing or adding one or more steps in the original RFE process.	RFE + CV [26,27];
		RFE + 4 resampling [28,29];
		Enhanced-RFE [30];
		Local search RFE [17];
		Marginal improvement-based RFE [31];
		Dynamic RFE [32]

RFE hybridized with other feature selection or dimension reduction methods	Using other feature selection methods or dimensionality reduction techniques together with RFE to select features.	TF-IDF + RFE [33];
		Chi-square + RFE [34];
		PCA + RFE [35];
		K-means + RFE [36];
		GI + RFE [37];
		Other hybrids [38,39]

Note. Abbreviations of RFE algorithms: RFE = Recursive Feature Elimination; SVM = Support Vector Machine; SVR = Support Vector Regression; DT = Decision Tree; RF = Random Forest; LR = Logistic Regression; NB = Naive Bayes; CV = Cross-Validation; TF-IDF = Term Frequency–Inverse Document Frequency; GI = Gini Index; PCA = Principal Component Analysis.

2.2.1. RFE Wrapped with Different ML Models

As mentioned earlier, a key step in the RFE process is to determine feature importance information, which is usually derived from the ML model used for prediction. This can lead to a variety of choices. For instance, the ML model used in the original RFE algorithm was the Support Vector Machine (SVM) [13]. This combination, referred to as SVM-RFE, has become one of the most commonly used feature selection methods. The SVM-RFE was originally applied to binary classification tasks; to generalize the use of SVM-RFE for multi-class classification prediction, Duan et al. [40] developed multiple SVM-RFE. Also, considering that the original SVM-RFE performed feature selection in a linear way, Mao et al. [41] extended the algorithm for use with nonlinear, complex data. Several additional extensions of RFE have been proposed in this line of research (e.g., [41–44]).

There are numerous examples of applying SVM-RFE in educational contexts. For instance, Chen et al. [45] classified Grade 4 students' digital reading performance as either high- or non-high-performing. The authors used SVM-RFE to identify 20 key contextual factors and then used these factors for prediction. Their model achieved an Area Under the ROC Curve (AUC-ROC) score of 89% and accuracy, sensitivity, and specificity of over 80%, indicating good prediction accuracy. In another study, Hu et al. [46] used SVM-RFE to identify the top 30 predictors of students' science achievement from 127 candidate variables about school, classroom, and student characteristics and background.

In addition to SVM models, RFE can be integrated with tree-based algorithms such as Decision Trees (DT) [21] and Random Forests (RF) [22], which evaluate feature importance based on metrics such as impurity reduction or permutation scores. RFE can also be applied with linear models, where feature importance is typically determined by the magnitude of regression coefficients [47]. For example, in an EDM study aimed at predicting students' teamwork styles (i.e., collaborative, cooperative, or solo-submit) on programming projects using GitHub logs, Gitinabard et al. [48] employed RFE in combination with both RF and logistic regression (LR) to identify the most important predictive features.

2.2.2. Combinations of ML Models or Feature Importance Metrics

Researchers have expanded the RFE algorithm by considering multiple feature importance metrics instead of relying on a single metric. These metrics can be generated by considering several distinct ML models or using different metrics from the same model. For example, Jeon and Oh [23] employed three ML models (SVM, RF, and generalized boosted regression algorithms) in the RFE process to determine the importance of features. They weighted the feature importance scores obtained from all three models to create their hybrid RFE algorithm, which yielded better performance in terms of feature selection and model performance improvement.

A similar variant of RFE was employed in an EDM study to investigate factors influencing students' online learning final results (passing or failing) [24]. Before performing RFE, the authors used each feature in the complete set to predict the target variable based on a DT classifier. This process led to the identification of the top 40 candidate features that alone were better able to predict the target variable. These 40 features were then used in RFE with three different modes (SVM, LR, and DT) until all but one predictor were eliminated. Lastly, the feature importance values obtained from the three models were averaged to obtain the final ranking of the candidate features.

In another EDM study, Alarape et al. [25] proposed a hybrid model consisting of an SVM classifier and a Naive Bayes (NB) classifier for predicting student performance. The two models were selected because they complement each other, with NB being resistant to noise and missing values while SVM is resistant to overfitting issues. The authors first compared the model performance between the NB classifier and the SVM classifier in predicting the target variable. The better-performing model was then used to wrap RFE and select key predictors, while the other algorithm was trained on the selected features to generate predictions. In their study, the SVM-RFE was the winner for model prediction on the first dataset, so it was used to select 18 out of the 50 original features; this subset was later used to train an NB classifier. Similarly, the NB classifier was the winner on the second dataset, with SVM being used to make predictions using the selected features determined by the NB classifier. This way, their hybrid methods can take advantage of both NB and SVM algorithms and reduce the bias introduced due to the algorithms' nature.

2.2.3. Modifications to the RFE Process

The original RFE process can be modified to achieve more robust, flexible, and effective feature selection performance. For instance, Han et al. [32] proposed a dynamic RFE that allows the elimination of more than one feature per iteration. Their method offers more flexible feature elimination operations. Another example is that the original RFE algorithm can benefit from the cross-validation (CV) framework to obtain more stable and reliable estimates of feature importance scores [26,27]. In this approach, CV divides the complete data into several subsets, and an ML model is trained with each subset to obtain multiple sets of feature importance scores. These scores are then averaged and used to iteratively eliminate the least important features until a stopping criterion is met. This allows the RFE algorithm to provide more robust results, further prevents overfitting, and increases the generalizability of the algorithm's performance to unseen data [26]. In a similar vein to the CV framework, researchers have also employed resampling and subsampling strategies to enhance the robustness of the feature elimination results [28,29].

In the education context, Chen et al. [49] used SVM-RFE with a 10-fold CV to predict students' skill mastery level in a game-based assessment. This allows the SVM-RFE algorithm to estimate the feature importance ten times because each data fold generates a feature importance score. Averaging the ten feature importance scores thus gives a more robust feature elimination decision. To further reduce the randomness of model performance, the authors repeated the 10-fold CV five times. Eventually, they chose to keep the top five features because the model's prediction performance peaked at five features. In another example, Sánchez-Pozo et al. [50] predicted students' mathematics performance based on their socioeconomic backgrounds and personal characteristics. They compared the performance of basic RFE and RFE with CV, showing that the ML classifier with the original RFE algorithm achieved only 60% on all three evaluation metrics (i.e., Recall, Accuracy, and F1 score) while the classifier with RFE with CV achieved at least 89% on these metrics. Moreover, Sivaneasharajah et al. [51] used user posts in a forum to predict students' learning behaviors and roles (e.g., information seeker, information giver) in MOOCs. They first extracted linguistic features using a text-mining tool and then applied RFE with a 10-fold CV to select 16 optimal features, which yielded good prediction performance.

Modifying the feature elimination process can also offer promising solutions to the limitations of the original RFE algorithm. For example, one commonly cited critique of the original RFE algorithm is that it did not consider the case where weak features, which are useless by themselves, may become very useful in predicting the target variable when combined with other important features [52]. Addressing this critique, Chen and Jeong [30] modified the original RFE process. Instead of directly removing the least important features, the authors considered whether the model's performance improves or worsens after removing those features in the subsequent prediction model. If the model's performance drops beyond a particular criterion, the least important features are retained instead of being removed outright, as in the original RFE process. In essence, the enhanced RFE removes features based on changes in model performance rather than solely relying on the importance of the features themselves.

Another limitation of the original RFE algorithm is its greedy search strategy, which determines feature importance based on the current feature set within each iteration (i.e., the local optimal choice). This approach does not guarantee the selection of the best possible feature subset because it does not consider all possible combinations of features. To mitigate this limitation, Samb et al. [17] proposed employing local search tools as an additional step following regular RFE to refine the suggested feature subset. These methods iterate through neighboring solutions, providing previously eliminated features an opportunity for reconsideration. If a neighboring solution is found to be superior to the current one, the algorithm updates to the best solution. This additional step in the feature selection process brought slight improvements in model performance in their empirical study. Another strategy to address the limitation of the greedy search strategy is to consider the marginal improvement in model performance. The original and general RFE algorithms calculate feature importance by building a model with the current complete set of features, but Ding et al. [31] proposed determining feature importance by building models with the current complete set of features, negating one feature at a time. This approach allows them to measure how the elimination of each feature can influence the model performance: if one feature provides the least marginal improvement in model performance, then that feature is considered the least important feature. This method is more computationally demanding than the original RFE process, but still less demanding than the exhaustive search and evaluation of all possible feature combinations.

2.2.4. RFE Hybridized with Other Feature Selection or Dimension Reduction Methods

RFE can function independently or be flexibly integrated with other feature selection methods. Most previously surveyed RFE variants are more complex and computationally demanding than the original RFE algorithm. However, combining RFE with other feature selection or dimension reduction techniques can help accelerate the feature elimination process. For instance, Lei and Govindaraju [35] employed PCA for dimensionality reduction and then used RFE to eliminate the extracted components, resulting in significantly faster processing. However, a major drawback of this approach is that using PCA prior to RFE reduces the interpretability of the resulting model. Similarly, Huang et al. [36] incorporated the K-means clustering technique to identify clusters of related features. They then used SVM-RFE to rank and select representative features from these clusters. In their approach, the representatives of feature clusters were used instead of individual features, thereby reducing computational complexity while preserving some interpretability of the original data structure.

Hybridizing RFE with other feature selection or dimensionality reduction methods not only accelerates the process but also provides a more effective strategy for identifying the most important features for predicting the target variable. For example, the SVM-RFE algorithm can be combined with the Gini index (GI) to form a hybrid feature selection approach [37]. Many other studies have proposed similar hybrid algorithms, such as those discussed in [38,39]. In the context of EDM research, Paddalwar et al. [34] compared the performance of machine learning models in predicting students' academic grades under three different feature selection settings: a filter-based method (chi-square test), basic RFE, and a hybrid approach combining the chi-square test with RFE. They found that the hybrid method—using the chi-square test followed by RFE—achieved the highest classification accuracy.

3. Methods

In addition to providing a narrative review of RFE variants, this paper also presents an empirical evaluation of five variants of RFE. These variants are structurally evaluated across two datasets, each corresponding to a different predictive task type (regression vs. classification). While this study primarily focuses on RFE within the context of EDM, our secondary analysis involving a health-related dataset offers important methodological insights. Specifically, we aim to evaluate whether the comparative performance of RFE variants observed in educational contexts holds true across diverse problem domains characterized by different feature structures, types of predictive outcomes, and data complexities. By benchmarking their performance in these heterogeneous scenarios, this analysis aims to inform future researchers in selecting and employing RFE algorithms. Figure 2 provides an overview of our analytical steps. A detailed description of the analyses is presented in the following sections.

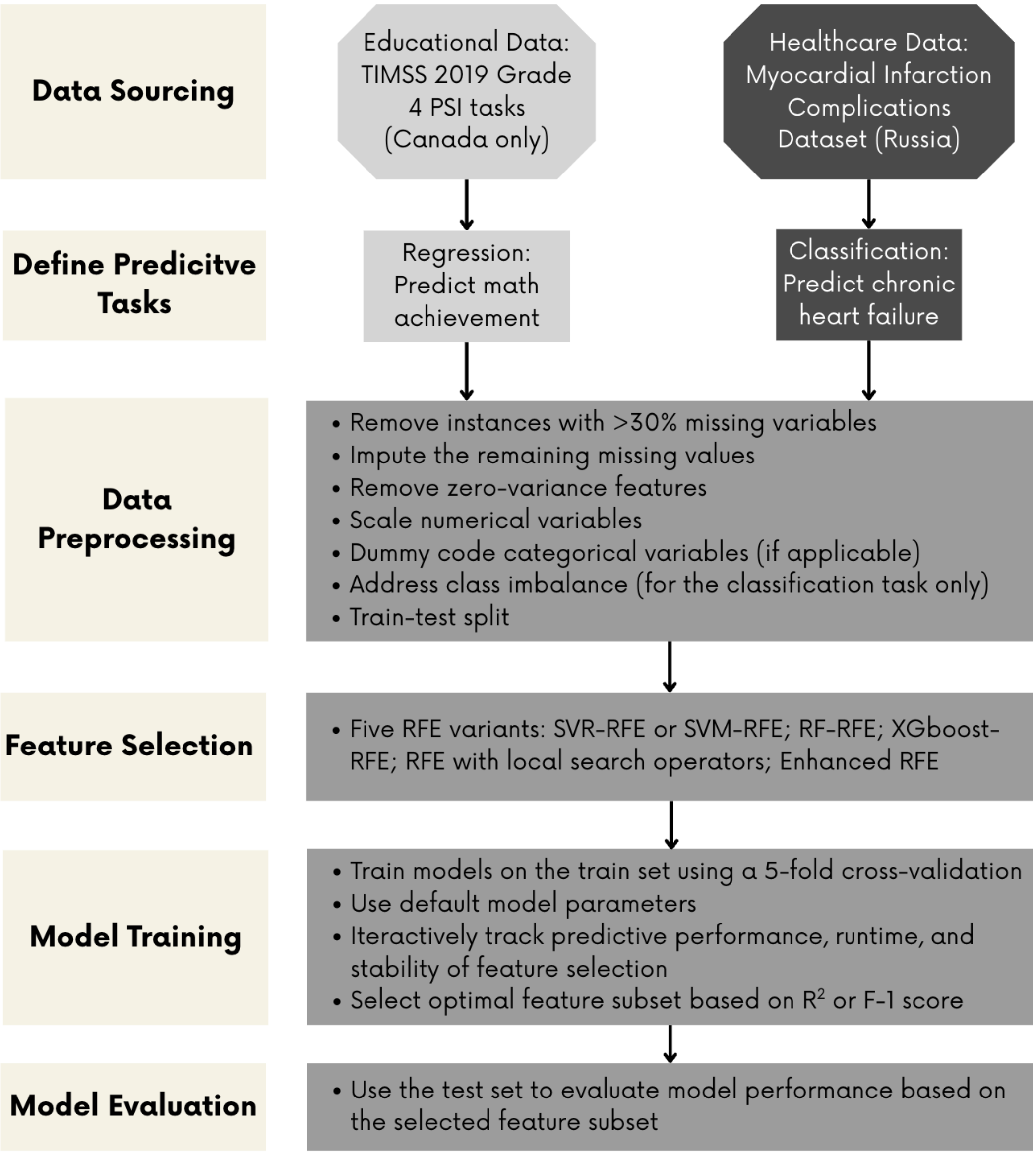
Data Sourcing
Educational Data: TIMSS 2019 Grade 4 PSI tasks (Canada only)
Healthcare Data: Myocardial Infarction Complications Dataset (Russia)
Define Predicitve Tasks
Regression: Predict math achievement
Classification: Predict chronic heart failure
Data Preprocessing
• Remove instances with >30% missing variables
• Impute the remaining missing values
• Remove zero-variance features
• Scale numerical variables
• Dummy code categorical variables (if applicable)
• Address class imbalance (for the classification task only)
• Train-test split
Feature Selection
• Five RFE variants: SVR-RFE or SVM-RFE; RF-RFE; XGboost-RFE; RFE with local search operators; Enhanced RFE
Model Training
• Train models on the train set using a 5-fold cross-validation
• Use default model parameters
• Iteractively track predictive performance, runtime, and stability of feature selection
• Select optimal feature subset based on R² or F-1 score
Model Evaluation
• Use the test set to evaluate model performance based on the selected feature subset

Figure 2. Overview of the analytical steps.

3.1. Datasets and Data Preprocessing

The educational dataset was sourced from Problem Solving and Inquiry (PSI) tasks, a component of the Trends in International Mathematics and Science Study (TIMSS) 2019 study. The PSI tasks are innovative computer-based assessments designed to evaluate students' higher-order mathematics and science skills (e.g., dynamic features, problem-solving processes) through digital-based interactive items in various formats [53]. TIMSS 2019 administered these tasks to approximately 22,000 fourth-grade students from

36 educational systems and 20,000 eighth-grade students from 27 educational systems worldwide. In addition to the PSI assessment data, TIMSS collected contextual questionnaire data from students, parents, teachers, and school principals. These questionnaires aim to explore the relationship between student achievement and contextual factors such as home environment, school climate, and attitudes toward learning. Detailed data collection procedures are documented in the TIMSS 2019 Technical Report [54]. The assessment and questionnaire datasets as well as the codebook for the surveyed variables are publicly available on the IEA's website (https://timss2019.org/international-database/index.html, accessed on 4 June 2025, [55]).

For this study, the focus was on Grade 4 mathematics PSI tasks. The target variable was students' mathematics achievement, a continuous variable representing their performance. Features included variables from the student questionnaire, capturing information such as home environment, attitudes toward learning, and demographic details. To ensure contextual consistency and reduce inter-country variations, data from six Canadian provinces (Alberta, British Columbia, Newfoundland and Labrador, Nova Scotia, Ontario, and Quebec; coded 9132 to 9137) were analyzed. To ensure data reliability, students with more than 30% missing responses were excluded, as high rates of missingness may indicate carelessness or other difficulties in completing the questionnaire [56,57]. Missing values in the remaining records were imputed using the RF method implemented in the *mice* R package [58]. Features with zero variance were removed, as they provide no predictive value for machine learning models. Numerical variables were scaled to improve the accuracy and efficiency of machine learning algorithms [59]. After preprocessing, the final dataset consisted of 16,137 students and 116 features. The data were split into training (80%) and test (20%) sets, resulting in 12,910 instances in the training set and 3227 instances in the test set.

The healthcare dataset was obtained from the Myocardial Infarction Complications Database, publicly available at https://doi.org/10.25392/leicester.data.12045261.v3 [60] (accessed on 4 June 2025). This dataset focuses on myocardial infarction (MI), a leading cause of global mortality. It was collected at the Krasnoyarsk Interdistrict Clinical Hospital (Russia) between 1992 and 1995 and contains 1700 patient records. The dataset comprises 111 features that describe clinical phenotypes and 12 binary variables representing potential myocardial infarction (MI) complications. A detailed description of the variables is available on its website.

For this study, the target variable was chronic heart failure, which is one of the key complications of MI. Chronic heart failure was selected because of its relatively high incidence rate ($n = 394$; 23.18%), which preserves the dataset's clinical relevance while alleviating data imbalance challenges. Similar data-preprocessing steps were applied to this dataset. Missing values were imputed using the RF algorithm in the *mice* R package. Numerical variables were scaled, and categorical variables were dummy-coded. This led to 144 features in total, including the generated dummy variables. The dataset was then split into training (80%) and test (20%) sets, with 1360 rows in the training set and 340 rows in the test set. To address the class imbalance in the binary target variable, the Synthetic Minority Oversampling Technique (SMOTE; [61]) was applied using the imbalanced-learn package (version 0.13.0) in Python. SMOTE ensures an equal distribution of classes in the training data, mitigating the risk of poor model generalizability and high classification error rates for rare categories [62]. This approach is empirically supported as superior to alternative resampling methods in various imbalanced scenarios [5,63,64]. After applying SMOTE, the training set contained 945 instances of chronic heart failure and 1044 instances of no chronic heart failure, resulting in a balanced total of 1989 rows.

3.2. Model Training, Validation, and Evaluation

To evaluate the effectiveness of RFE variants across two predictive tasks, the first task involved regression for the educational dataset, where the goal was to predict students' mathematics achievement. The second task focused on classification for the healthcare dataset, predicting chronic heart failure. As benchmarks without feature selection, the baseline models employed Support Vector Regression (SVR) for regression tasks and SVM for classification tasks. These models were selected because SVM was used in the original RFE algorithm [13], with SVR being the natural extension and comparable counterpart of SVM for regression tasks [65].

In this study, five variants of RFE were examined: the original RFE algorithm, Enhanced RFE [30], RF-RFE [22], Extreme Gradient Boosting (XGBoost)-RFE [66], and RFE with a local search operator [17]. These variants differ in how they determine feature importance and perform elimination, as previously detailed in the literature review. The original SVR/SVM-RFE algorithm ranks features based on the absolute magnitudes of their weights in the SVR or SVM model, iteratively removing the least important features. In contrast, both RF-RFE and XGBoost-RFE are extensions of the original RFE algorithm that integrate tree-based ensemble models to evaluate feature importance. RF-RFE uses RF models to rank features based on the decrease in predictive performance when individual feature values are permuted, making it robust to noise and well-suited for capturing non-linear interactions. Similarly, XGBoost-RFE leverages the XGBoost algorithm to assess feature importance based on each feature's contribution to improving prediction accuracy during the boosting process. Enhanced RFE extends the original algorithm by addressing the potential exclusion of weak but complementary features, iteratively reinstating features that significantly enhance performance. Finally, RFE with local search operators introduces an optimization layer that explores neighboring feature subsets at each iteration, dynamically reconsidering features that were previously eliminated.

All models were implemented using the *scikit-learn* library (version 1.6) in Python. Training and evaluation followed a 5-fold CV scheme to ensure robust performance assessment while mitigating overfitting risks. For RFE algorithms, performance was tracked iteratively, and the optimal number of features was identified based on the iteration yielding the highest average performance metric across folds. Default parameters were used for the ML models to reduce computational costs, as this study prioritized comparisons of RFE variants over hyperparameter optimization. Performance metrics were chosen to align with the goals of the respective predictive tasks. For regression, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 were utilized to assess error magnitude, variance explanation, and overall accuracy. For classification, Precision, Recall, and F1 score were used, reflecting the importance of addressing imbalanced class distributions, particularly in chronic heart failure prediction, where false negatives carry significant consequences. Among these metrics, R^2 (for regression) and F1 score (for classification) served as the primary criteria for evaluating model performance and determining the optimal feature subset. Once the models with the optimal number of features were selected, they were applied to holdout test sets to validate their performance. In addition, the total runtime across all iterations was recorded using system timestamps for each RFE variant. Feature selection stability was computed to evaluate how consistently features were retained across CV folds. Specifically, feature selection stability refers to the consistency of features being either consistently retained or excluded across cross-validation folds.

4. Results

4.1. Results for the Educational Dataset

Table 2 summarizes the number of features selected by each RFE variant and the corresponding performance metrics in predicting students' mathematics achievement. The dataset used for this analysis comprised a preprocessed set of 116 candidate features derived from the TIMSS student questionnaire.

Table 2. Number of features selected and performance metrics achieved by each RFE variant for the educational prediction task (regression).

Algorithm	Features	RMSE	MAE	R^2	Time (s)	Stability
SVR-RFE	82	57.357	45.957	0.359	430	0.633
RF-RFE	108	56.474	44.377	0.379	223,279	0.966
XGBoost-RFE	114	58.514	45.662	0.333	7321	0.949
Enhanced RFE	62	58.234	46.577	0.340	821	0.872
RFE with local search operator	85	57.442	46.091	0.357	473	0.581

4.1.1. Baseline: SVR-RFE

When applying the original SVR-RFE algorithm, 82 of the 116 initial features were retained for prediction. This reduced feature set accounted for approximately 35.9% of the variance in students' mathematics achievement while maintaining computational efficiency, completing the feature selection process in just 430 s. Many of the retained features pertained to students' personal and home backgrounds, their self-perceptions and attitudes toward mathematics and science, as well as aspects of their home environment. However, the moderate stability of SVR-RFE (0.633) indicates some variability across folds, suggesting that the importance of certain predictors may not be consistently emphasized.

4.1.2. RF-RFE

By integrating the RF model into the RFE process, RF-RFE selected 108 features and achieved the highest coefficient of determination ($R^2 = 0.379$) among all methods, alongside the lowest RMSE (56.474) and MAE (44.377). It also produced a high feature selection stability score of 0.966. However, this performance came at the cost of an extremely long runtime (over 223,000 s), making it less practical for time-sensitive applications. While RF's non-linear feature importance mechanism appears to capture complex interactions effectively, the relatively modest gain in predictive accuracy over simpler models such as SVR-RFE ($R^2 = 0.359$) raises concerns about efficiency, especially given the substantially larger computational burden and feature subset retained.

4.1.3. XGBoost-RFE

The XGBoost-RFE algorithm selected 114 features and eliminated 2. However, this feature selection led to a reduction in the model's predictive performance ($R^2 = 0.333$) while also requiring a longer runtime of 7321 s. Compared to the baseline algorithm, the stability improved to 0.949.

4.1.4. RFE with Local Search Operators

When local search operators were introduced, the method retained 85 features, 3 more than the baseline. However, this marginally increased subset did not yield better performance, as $R^2 = 0.357$, and the error metrics (RMSE and MAE) did not surpass those of the baseline. Additionally, this method required slightly more computational time (473 vs. 430 s) and exhibited reduced stability (0.581 vs. 0.633). While local search can reevaluate and reinstate previously removed features, these results suggest that, in this specific context, the added complexity did not lead to a meaningful improvement in predictive accuracy.

Of the 85 predictors, 81 matched the baseline SVR-RFE, indicating that many core features were similarly influential across both methods.

4.1.5. Enhanced RFE

Enhanced RFE produced the largest feature reduction, identifying a subset of only 62 features while maintaining a reduced R^2 of 0.340. The partial overlap (51 of the 62 features) with SVR-RFE underscores that Enhanced RFE identifies many of the same key variables while discarding others that appear redundant or weakly predictive. Although Enhanced RFE required more computation time than SVR-RFE (821 vs. 430 s), it was still significantly faster than other non-linear variants. The algorithm also achieved a higher stability score of 0.872. These results suggest that this RFE variant, which reintroduces potentially complementary predictors in its iterative steps, demonstrates that parsimony is achievable with minimal sacrifice in explanatory power.

4.1.6. Summary of Regression Findings

In the regression task, RF-RFE produced the highest R^2 value, leveraging its nonlinear ensemble nature to capture complex interactions. However, this performance required retaining relatively more features (108), resulting in reduced model interpretability and increased computational complexity. In contrast, Enhanced RFE—based on a linear SVR model but augmented by iterative feature reintroduction—achieved substantial dimensionality reduction (62 features) with only a modest drop in predictive accuracy. This result is particularly relevant in educational contexts such as TIMSS, where reducing survey length without sacrificing insight is a key goal. RFE with local search operators performed comparably to the baseline, offering slight improvements in predictive metrics but with decreased feature selection stability.

4.2. Results for the Healthcare Dataset

The performance of the five RFE variants for predicting chronic heart failure in a healthcare dataset containing 144 original features is summarized in Table 3. The main metrics used to evaluate the classification performance include Precision, Recall, F1 score, and AUC-ROC value, as well as total runtime and stability metrics for feature selection.

Table 3. Number of features selected and performance metrics achieved by each RFE variant for the healthcare prediction task (classification).

Algorithm	**Features**	**F1**	**Precision**	**Recall**	**AUC-ROC**	**Time (s)**	**Stability**
SVM-RFE	118	0.438	0.284	0.962	0.744	1011	0.632
RF-RFE	110	0.566	0.706	0.567	0.770	3358	0.938
XGBoost-RFE	56	0.665	0.725	0.645	0.799	5796	0.639
Enhanced RFE	120	0.604	0.600	0.624	0.692	12,940	0.778
RFE with local search operator	106	0.618	0.613	0.638	0.701	3334	0.563

4.2.1. Baseline: SVM-RFE

Using the baseline SVM-RFE, the model retained 118 of 144 features, resulting in the largest subset among all variants compared. This feature subset yielded a very high Recall of 96.2%, indicating that the model effectively captured most true positive cases. However, the low Precision (28.4%) reveals a high rate of false positives, resulting in a moderate overall F1 score of 0.438. While the relatively short runtime (1011 s) demonstrates computational efficiency, the moderate stability (0.632) suggests some inconsistency in feature selection across folds. Clinically, this method may be appropriate for screening purposes where the cost of missing a true case is high, such as early detection of chronic heart failure. However, the high false-positive rate could lead to over-testing or unnecessary interventions. The moderate AUC-ROC score of 0.744 further suggests that the model is better suited for broad identification rather than fine-grained risk stratification.

4.2.2. RF-RFE

Unlike the baseline model, the RF-RFE algorithm retained 110 features, with 89 overlapping those selected by the baseline approach. This subset achieved a significantly higher Precision of 70.6% but the lowest Recall rate (56.7%) among all methods. While RF-RFE more effectively filters out false positives, it also risks missing true-positive cases. The overall F1 score of 0.566 reflects this trade-off, with improved Precision coming at the cost of Recall. In a clinical diagnosis context, such a model may reduce unnecessary follow-up interventions but could fail to identify a substantial number of actual chronic heart failure patients—an issue when early detection is paramount. The algorithm also showed improved discrimination ability with an AUC-ROC of 0.770 and demonstrated high consistency in feature selection (stability = 0.938). Although it retained slightly fewer features than the baseline, the increased runtime (3358 s) reflects the added computational demand of RF-RFE.

4.2.3. XGBoost-RFE

Through the XGBoost-RFE algorithm, the largest feature reduction was observed, retaining only 56 features, with 44 overlapping those selected by SVM-RFE. Notably, XGBoost-RFE achieved the highest Precision (72.5%) while also maintaining a relatively high Recall (64.5%). This balance produced an F1 score of 0.665—the highest among all tested variants—indicating a more effective trade-off between capturing true positives and avoiding false positives. XGBoost-RFE also demonstrated strong discriminative capability with the highest AUC-ROC score (0.799). In clinical applications where both accurate detection and minimizing false alarms are important, this method presents a compelling option. However, these benefits come at a cost: a longer runtime of 5796 s and moderate feature selection stability (0.639).

4.2.4. Enhanced RFE

Enhanced RFE selected 120 features, with 110 overlapping those identified by SVM-RFE. This variant demonstrated a balanced classification performance, achieving a Recall of 62.4% and a Precision of 60.0%, leading to an overall F1 score of 0.604. These results reflect a moderate ability to both detect true positive cases and limit false positives, surpassing the baseline SVM-RFE in overall predictive quality. One notable strength of Enhanced RFE is its improved feature selection consistency, as indicated by a stability score of 0.778. This suggests that its iterative mechanism—reintroducing previously excluded features—helps retain relevant predictors while filtering out less informative ones. However, its AUC-ROC score of 0.692 indicates relatively weaker discrimination capability across risk thresholds, and its high computational cost (12,940 s) may limit practicality in time-sensitive applications. In a clinical context, Enhanced RFE may be well suited for scenarios that require a balanced detection of cases without overemphasizing either extreme, minimizing false alarms while still identifying a meaningful proportion of positive cases. Its strength lies in supporting general decision-making pipelines rather than fine-grained risk stratification.

4.2.5. RFE with Local Search Operators

RFE with local search operators retained 106 features, with 97 overlapping with SVM-RFE's selection. Its Recall (63.8%) and Precision (61.3%) were similar yet slightly higher than Enhanced RFE, resulting in a similar F1 score of 0.618 compared to Enhanced RFE's 0.604. The AUC-ROC score of 0.701 was also slightly higher than that of Enhanced RFE, along with requiring a much shorter operational runtime, but at the cost of stability. The local search mechanism appears to have helped the model reconsider critical features that might otherwise be excluded, allowing it to slightly surpass Enhanced RFE. For settings that require robust detection but also value reducing false positives, RFE with local search operators can offer a well-rounded strategy that adapts to complex interactions among features.

4.2.6. Summary of Classification Findings

In the classification task, XGBoost-RFE delivered the strongest performance overall, achieving the highest F1 score (0.665) and AUC-ROC (0.799) while using the smallest feature subset (56). Its tree-based boosting structure allowed for fine-grained prioritization of predictive variables but required much longer computational time. Enhanced RFE again struck a favorable balance, achieving a high F1 score (0.604) with improved stability over SVM-RFE and XGBoost-RFE, though with substantial runtime. RFE with local search operators provided similar classification accuracy (F1: 0.618) but showed the lowest stability.

5. Discussion

In predictive modeling contexts across multiple domains ranging from education to healthcare, the volume and complexity of available data have grown exponentially [1,2]. Effective feature selection is crucial in such high-dimensional contexts, as it serves several key purposes: reducing overfitting risk, improving model interpretability, and lowering computational costs [11,14]. By filtering out irrelevant or redundant variables, researchers can focus on the most meaningful predictors, thereby enhancing both the predictive accuracy and the practical utility of their findings [8,52]. Among the various feature selection techniques, RFE has emerged as a particularly powerful and interpretable method. Although the RFE algorithm was originally developed in bioinformatics and healthcare analytics fields [13,32], it has gained growing traction in EDM research. For instance, in the analysis of data from large-scale educational assessments, RFE enables researchers and policymakers to identify and prioritize the most influential factors affecting student achievement [67,68]. Similarly, when analyzing student data collected through digital learning management systems, RFE can facilitate accurate predictions of key outcomes, such as learning engagement, by focusing on the most relevant indicators [69,70].

While the original SVM-RFE or SVR-RFE algorithm is widely used and has demonstrated strong performance in feature selection tasks [13], the literature has introduced several enhanced or alternative variants aimed at addressing its known limitations. In this study, we synthesized these developments by categorizing RFE algorithms into the original method and four major variant types. This framework is intended to help future users, such as EDM researchers, better understand the design, characteristics, advantages, and trade-offs associated with each type of RFE algorithm. For example, Enhanced RFE [30] reintroduces previously eliminated features if they demonstrate added value in combination with others, capturing potential synergistic effects among variables. RF-RFE [22] and XGBoost-RFE [66] incorporate tree-based models along with ensemble methods capable of capturing nonlinear interactions, which is particularly important in complex domains such as healthcare and education. RFE with local search operators [17] further extends the method by dynamically reevaluating feature subsets to prevent suboptimal eliminations.

Following the narrative review of RFE algorithms and their applications in EDM research, this study conducted a structured evaluation of five RFE variants across two distinct tasks: a regression task predicting students' mathematics achievement using TIMSS data and a classification task identifying chronic heart failure using clinical data. We not only assessed the predictive performance of the features selected by each RFE algorithm but also recorded their total runtime and evaluated feature selection stability, defined as the consistency of feature selection across training folds. These metrics allowed us to benchmark each algorithm's effectiveness and practical utility across different domains.

Across both tasks, the performance of RFE variants varied considerably in terms of predictive accuracy, feature reduction, runtime, and stability. These differences largely reflect the underlying algorithmic structures of each method. For example, nonlinear ensemble-based variants such as RF-RFE and XGBoost-RFE achieved strong predictive performance and excelled at identifying complex feature interactions, but they incurred high computational costs and tended to retain larger feature subsets. Enhanced RFE demonstrated the most efficient trade-off of achieving substantial feature reduction while maintaining competitive predictive power, which makes it particularly suitable for domains that prioritize interpretability and efficiency.

In the EDM context, where interpretability often takes precedence over prediction accuracy, RFE variants that achieve substantial feature reduction with minimal loss in performance are generally preferred. This is especially beneficial in large-scale assessments like TIMSS, where reducing the number of predictors can streamline data collection, lower analytical costs, and enhance the usability of results for educational stakeholders. In contrast, healthcare applications—such as clinical classification—typically prioritize predictive performance, particularly metrics like Recall and Precision, depending on the diagnostic objective. In such cases, RFE methods integrated with tree-based models like Random Forest and XGBoost can be advantageous, as they effectively capture complex, nonlinear relationships among features. Although these approaches come with higher computational costs, this trade-off is often acceptable when classification accuracy is paramount. Overall, the findings of this study underscore the importance of selecting RFE variants based on the specific goals of the modeling task, domain requirements, and tolerance for error. Rather than seeking a one-size-fits-all solution, researchers should consider the unique strengths and trade-offs of each method in light of their application context.

Limitations and Directions for Future Research

This study has several limitations. First, the narrative review of RFE algorithms may be subject to selection bias. While we aimed to comprehensively survey RFE variants and their applications in EDM research, the narrative review inherently involves subjective interpretation in categorizing approaches into four major types. This classification framework holds the risk of omitting niche or emerging RFE variants and adaptations.

Second, our empirical evaluation focused on comparing five RFE variants. However, many other RFE algorithms exist and may yield different or even superior results under alternative conditions. Additionally, we used default parameters for SVR and SVM to manage computational demands, which inevitably limited our ability to explore the optimal configuration for each RFE variant. SVM-based models, in particular, are highly sensitive to hyperparameter settings (e.g., kernel type) and may perform quite differently when tuned using systematic approaches such as grid search, random search, or Bayesian optimization. Nevertheless, our goal was not to identify the single "best" RFE variant in terms of prediction accuracy. Rather, we aimed to demonstrate how different RFE techniques emphasize distinct aspects of model performance, including trade-offs in feature reduction, runtime, and stability. Future researchers could expand on this work by incorporating hyperparameter tuning and more robust training procedures to identify the algorithm configurations best suited to their specific research needs.

Third, although we reported the number of retained features for each RFE variant, we did not examine in depth why specific features emerged as important. We acknowledge that interpretability is critical in both healthcare and educational research. In contexts such as patient screening or educational policy-making, understanding why a feature is selected can be just as important as its contribution to predictive accuracy. Future research should therefore incorporate interpretability-focused analyses, such as evaluating feature importance in relation to domain-specific theories, to ensure that selected features are both data-driven and contextually meaningful. Additionally, visualization techniques can be employed to clearly demonstrate how individual features influence the prediction of target variables.

6. Conclusions

This study contributes to the field in two ways. First, it presents a narrative review that categorizes RFE algorithm variants into four major methodological types based on their design. These include (1) RFE combined with different machine learning models, (2) combinations of multiple models or feature importance metrics, (3) modifications to the RFE process, and (4) hybrid approaches that integrate RFE with other feature selection or dimensionality reduction techniques. We also present illustrative examples from the EDM literature to help researchers understand the advantages and trade-offs associated with different RFE variants. Second, this study includes an empirical evaluation of five representative RFE variants using datasets from both educational and healthcare domains. This evaluation benchmarked their performance in terms of prediction accuracy, runtime efficiency, and feature selection stability. These metrics are particularly important in applied research, where model interpretability, computational cost, and robustness are often critical considerations. Our results show that each RFE variant emphasizes different priorities and is better suited to different applications. For instance, RF-RFE achieved satisfactory predictive performance and excellent selection stability in both regression and classification tasks, but retained the largest feature subsets and required higher computational resources. In contrast, Enhanced RFE delivered strong predictive performance while significantly reducing the number of selected features and maintaining reasonable runtime and stability. Therefore, we emphasize that no single RFE variant is ideal for every situation. Future researchers should select the most appropriate algorithm based on the specific objectives, constraints, and data characteristics of their study. We recommend considering trade-offs among predictive accuracy, model complexity, interpretability, runtime, and feature selection stability when making methodological choices.

Author Contributions: Conceptualization, O.B.; methodology, E.M. and B.T.; software, E.M. and B.T.; validation, E.M., B.T. and A.S.; formal analysis, E.M. and B.T.; investigation, O.B., E.M., B.T. and A.S.; writing—original draft preparation, O.B., E.M., B.T. and A.S.; writing—review and editing, O.B., E.M., B.T. and A.S.; supervision, O.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this study are publicly available. The TIMSS 2019 International Database can be accessed via https://timss2019.org/international-database/ (accessed on 4 June 2025). Similarly, the health dataset can be downloaded from the Myocardial Infarction Complications Database, https://doi.org/10.25392/leicester.data.12045261.v3 (accessed on 4 June 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AUC	Area Under the Curve
CV	Cross-Validation
DT	Decision Trees
EDM	Educational Data Mining
GI	Gini Index
LASSO	Least Absolute Shrinkage and Selection Operator
LR	Logistic Regression
LLM	Large Language Models
MAE	Mean Absolute Error
ML	Machine Learning
MOOC	Massive Open Online Courses
PCA	Principal Component Analysis
PSI	Problem Solving and Inquiry
RF	Random Forests
RFE	Recursive Feature Elimination
RMSE	Root Mean Square Error
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
SVR	Support Vector Regression
TIMSS	Trends in International Mathematics and Science Study
XGBoost	Extreme Gradient Boosting

References

1. Romero, C.; Ventura, S. Data mining in education. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2013**, *3*, 12–27. [CrossRef]
2. Algarni, A. Data mining in education. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 12–27. [CrossRef]
3. Bulut, O.; Yavuz, H.C. Educational Data Mining: A Tutorial for the Rattle Package in R. *Int. J. Assess. Tools Educ.* **2019**, *6*, 20–36. [CrossRef]
4. Romero, C.; Ventura, S. Educational data mining: A review of the state of the art. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2010**, *40*, 601–618. [CrossRef]
5. Wongvorachan, T.; He, S.; Bulut, O. A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. *Information* **2023**, *14*, 54. [CrossRef]
6. Bulut, O.; Wongvorachan, T.; He, S.; Lee, S. Enhancing high-school dropout identification: A collaborative approach integrating human and machine insights. *Discov. Educ.* **2024**, *3*, 109. [CrossRef]
7. Cui, Z.; Gong, G. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage* **2018**, *178*, 622–637. [CrossRef]
8. James, T.P.G.; Karthikeyan, B.Y.; Ashok, P.; Dhaasarathy; Suganya, R.; Maharaja, K. Strategic Integration of CNN, SVM, and XGBoost for Early-stage Tumor Detection using Hybrid Deep Learning Method. In Proceedings of the 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 14–15 December 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
9. Palo, H.K.; Sahoo, S.; Subudhi, A.K. Dimensionality reduction techniques: Principles, benefits, and limitations. In *Data Analytics in Bioinformatics: A Machine Learning Perspective*; Wiley: New York, NY, USA, 2021; pp. 77–107.
10. Alalawi, K.; Athauda, R.; Chiong, R. Contextualizing the current state of research on the use of machine learning for student performance prediction: A systematic literature review. *Eng. Rep.* **2023**, *5*, e12699. [CrossRef]
11. Venkatesh, B.; Anuradha, J. A review of feature selection and its methods. *Cybern. Inf. Technol.* **2019**, *19*, 3–26. [CrossRef]
12. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [CrossRef]
13. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]
14. Liu, W.; Wang, J. Recursive elimination–election algorithms for wrapper feature selection. *Appl. Soft Comput.* **2021**, *113*, 107956. [CrossRef]
15. Albreiki, B.; Zaki, N.; Alashwal, H. A systematic literature review of student'performance prediction using machine learning techniques. *Educ. Sci.* **2021**, *11*, 552. [CrossRef]

16. Chen, R.; Manongga, W.; Dewi, C. Recursive Feature Elimination for Improving Learning Points on Hand-Sign Recognition. *Future Internet* **2022**, *14*, 352. [CrossRef]
17. Samb, M.L.; Camara, F.; Ndiaye, S.; Slimani, Y.; Esseghir, M.A. A novel RFE-SVM-based feature selection approach for classification. *Int. J. Adv. Sci. Technol.* **2012**, *43*, 27–36.
18. Reunanen, J. Overfitting in making comparisons between variable selection methods. *J. Mach. Learn. Res.* **2003**, *3*, 1371–1382.
19. Yeung, C.K.; Yeung, D.Y. Incorporating features learned by an enhanced deep knowledge tracing model for stem/non-stem job prediction. *Int. J. Artif. Intell. Educ.* **2019**, *29*, 317–341. [CrossRef]
20. Pereira, F.D.; Oliveira, E.; Cristea, A.; Fernandes, D.; Silva, L.; Aguiar, G.; Alamri, A.; Alshehri, M. Early dropout prediction for programming courses supported by online judges. In Proceedings of the Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, 25–29 June 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 67–72.
21. Lian, W.; Nie, G.; Jia, B.; Shi, D.; Fan, Q.; Liang, Y. An intrusion detection method based on decision tree-recursive feature elimination in ensemble learning. *Math. Probl. Eng.* **2020**, *2020*, 1–15. [CrossRef]
22. Granitto, P.M.; Furlanello, C.; Biasioli, F.; Gasperi, F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemom. Intell. Lab. Syst.* **2006**, *83*, 83–90. [CrossRef]
23. Jeon, H.; Oh, S. Hybrid-recursive feature elimination for efficient feature selection. *Appl. Sci.* **2020**, *10*, 3211. [CrossRef]
24. Chai, Y.; Lei, C.; Yin, C. Study on the influencing factors of online learning effect based on decision tree and recursive feature elimination. In Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning, Tokyo, Japan, 10–13 January 2019; pp. 52–57. [CrossRef]
25. Alarape, M.A.; Ameen, A.O.; Adewole, K.S. Hybrid students' academic performance and dropout prediction models using recursive feature elimination technique. In *Advances on Smart and Soft Computing: Proceedings of ICACIn 2021*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 93–106.
26. Nguyen, H.N.; Ohn, S.Y. Drfe: Dynamic recursive feature elimination for gene identification based on random forest. In Proceedings of the International Conference on Neural Information Processing, Hong Kong, China, 3–6 October 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1–10. [CrossRef]
27. Artur, M. Review the performance of the Bernoulli Naïve Bayes Classifier in Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of the best number of features. *Procedia Comput. Sci.* **2021**, *190*, 564–570. [CrossRef]
28. Wottschel, V.; Chard, D.T.; Enzinger, C.; Filippi, M.; Frederiksen, J.L.; Gasperini, C.; Giorgio, A.; Rocca, M.A.; Rovira, A.; De Stefano, N.; et al. SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of multiple sclerosis. *Neuroimage Clin.* **2019**, *24*, 102011. [CrossRef] [PubMed]
29. van der Ploeg, T.; Steyerberg, E.W. Feature selection and validated predictive performance in the domain of Legionella pneumophila: A comparative study. *BMC Res. Notes* **2016**, *9*, 147. [CrossRef]
30. Chen, X.W.; Jeong, J.C. Enhanced recursive feature elimination. In Proceedings of the Sixth International Conference on Machine Learning and Applications (ICMLA 2007), Cincinnati, OH, USA, 13–15 December 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 429–435. [CrossRef]
31. Ding, X.; Li, Y.; Chen, S. Maximum margin and global criterion based-recursive feature selection. *Neural Netw.* **2024**, *169*, 597–606. [CrossRef]
32. Han, Y.; Huang, L.; Zhou, F. A dynamic recursive feature elimination framework (dRFE) to further refine a set of OMIC biomarkers. *Bioinformatics* **2021**, *37*, 2183–2189. [CrossRef]
33. Nafis, N.S.M.; Awang, S. An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. *IEEE Access* **2021**, *9*, 52177–52192. [CrossRef]
34. Paddalwar, S.; Mane, V.; Ragha, L. Predicting students' academic grade using machine learning algorithms with hybrid feature selection approach. *ITM Web Conf.* **2022**, *44*, 03036. [CrossRef]
35. Lei, H.; Govindaraju, V. *Speeding Up Multi-Class SVM Evaluation by PCA and Feature Selection*; Technical Report; Center for Unified Biometrics and Sensors, State University of New York at Buffalo: Amherst, NY, USA, 2005.
36. Huang, X.; Zhang, L.; Wang, B.; Li, F.; Zhang, Z. Feature clustering based support vector machine recursive feature elimination for gene selection. *Appl. Intell.* **2018**, *48*, 594–607. [CrossRef]
37. Almutiri, T.; Saeed, F. A hybrid feature selection method combining Gini index and support vector machine with recursive feature elimination for gene expression classification. *Int. J. Data Min. Model. Manag.* **2022**, *14*, 41–62. [CrossRef]
38. Lin, X.; Wang, Q.; Yin, P.; Tang, L.; Tan, Y.; Li, H.; Yan, K.; Xu, G. A method for handling metabonomics data from liquid chromatography/mass spectrometry: Combinational use of support vector machine recursive feature elimination, genetic algorithm and random forest for feature selection. *Metabolomics* **2011**, *7*, 549–558. [CrossRef]
39. Louw, N.; Steel, S. Variable selection in kernel Fisher discriminant analysis by means of recursive feature elimination. *Comput. Stat. Data Anal.* **2006**, *51*, 2043–2055. [CrossRef]
40. Duan, K.B.; Rajapakse, J.C.; Wang, H.; Azuaje, F. Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Trans. Nanobiosci.* **2005**, *4*, 228–234. [CrossRef]

41. Mao, Y.; Zhou, X.; Yin, Z.; Pi, D.; Sun, Y.; Wong, S.T. Gene selection using Gaussian kernel support vector machine based recursive feature elimination with adaptive kernel width strategy. In Proceedings of the Rough Sets and Knowledge Technology: First International Conference, RSKT 2006, Chongqing, China, 24–26 July 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 799–806.
42. Zhou, X.; Tuck, D.P. MSVM-RFE: Extensions of SVM-RFE for multiclass gene selection on DNA microarray data. *Bioinformatics* **2007**, *23*, 1106–1114. [CrossRef]
43. Zhang, L.; Zheng, X.; Pang, Q.; Zhou, W. Fast Gaussian kernel support vector machine recursive feature elimination algorithm. *Appl. Intell.* **2021**, *51*, 9001–9014. [CrossRef]
44. Cao, J.; Zhang, L.; Wang, B.; Li, F.; Yang, J. A fast gene selection method for multi-cancer classification using multiple support vector data description. *J. Biomed. Inform.* **2015**, *53*, 381–389. [CrossRef] [PubMed]
45. Chen, F.; Sakyi, A.; Cui, Y. Identifying key contextual factors of digital reading literacy through a machine learning approach. *J. Educ. Comput. Res.* **2022**, *60*, 1763–1795. [CrossRef]
46. Hu, J.; Peng, Y.; Ma, H. Examining the contextual factors of science effectiveness: A machine learning-based approach. *Sch. Eff. Sch. Improv.* **2022**, *33*, 21–50. [CrossRef]
47. Zheng, S.; Liu, W. Lasso based gene selection for linear classifiers. In Proceedings of the 2009 IEEE International Conference on Bioinformatics and Biomedicine Workshop, Washington, DC, USA, 1–4 November 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 203–208. [CrossRef]
48. Gitinabard, N.; Okoilu, R.; Xu, Y.; Heckman, S.; Barnes, T.; Lynch, C. Student Teamwork on Programming Projects: What can GitHub logs show us? *arXiv* **2020**, arXiv:2008.11262. [CrossRef]
49. Chen, F.; Cui, Y.; Chu, M.W. Utilizing game analytics to inform and validate digital game-based assessment with evidence-centered game design: A case study. *Int. J. Artif. Intell. Educ.* **2020**, *30*, 481–503. [CrossRef]
50. Sánchez-Pozo, N.; Chamorro-Hernández, L.; Mina, J.; Márquez, J. Comparative analysis of feature selection techniques in predictive modeling of mathematics performance: An Ecuadorian case study. *Educ. Sci. Manag.* **2023**, *1*, 111–121. [CrossRef]
51. Sivaneasharajah, L.; Falkner, K.; Atapattu, T. Investigating Students' Learning in Online Learning Environment. In Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020), Virtual Conference, 10–13 July 2020.
52. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
53. Mullis, I.V.; Martin, M.O.; Fishbein, B.; Foy, P.; Moncaleano, S. Findings from the TIMSS 2019 Problem Solving and Inquiry Tasks. 2021. Available online: https://timssandpirls.bc.edu/timss2019/psi (accessed on 5 June 2025).
54. Martin, M.O.; von Davier, M.; Mullis, I.V.S. Methods and Procedures: TIMSS 2019 Technical Report. 2020. Available online: https://timssandpirls.bc.edu/timss2019/methods (accessed on 5 June 2025).
55. Fishbein, B.; Foy, P.; Yin, L. TIMSS 2019 User Guide for the International Database. Hentet Fra. 2021. Available online: https://timssandpirls.bc.edu/timss2019/international-database (accessed on 5 June 2025).
56. Ulitzsch, E.; Yildirim-Erbasli, S.N.; Gorgun, G.; Bulut, O. An explanatory mixture IRT model for careless and insufficient effort responding in self-report measures. *Br. J. Math. Stat. Psychol.* **2022**, *75*, 668–698. [CrossRef] [PubMed]
57. Wongvorachan, T.; Bulut, O.; Liu, J.X.; Mazzullo, E. A Comparison of Bias Mitigation Techniques for Educational Classification Tasks Using Supervised Machine Learning. *Information* **2024**, *15*, 326. [CrossRef]
58. Zhang, Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann. Transl. Med.* **2016**, *4*, 30.
59. Ahsan, M.M.; Mahmud, M.P.; Saha, P.K.; Gupta, K.D.; Siddique, Z. Effect of data scaling methods on machine learning algorithms and model performance. *Technologies* **2021**, *9*, 52. [CrossRef]
60. Golovenkin, S.; Gorban, A.; Mirkes, E.; Shulman, V.; Rossiev, D.; Shesternya, P.; Nikulina, S.Y.; Orlova, Y.V.; Dorrer, M. Complications of Myocardial Infarction: A Database for Testing Recognition and Prediction Systems. 2020. Available online: https://archive.ics.uci.edu/dataset/579/myocardial+infarction+complications (accessed on 5 June 2025).
61. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
62. Mao, A.; Huang, E.; Wang, X.; Liu, K. Deep learning-based animal activity recognition with wearable sensors: Overview, challenges, and future directions. *Comput. Electron. Agric.* **2023**, *211*, 108043. [CrossRef]
63. Blagus, R.; Lusa, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinform.* **2013**, *14*, 106. [CrossRef]
64. Van Hulse, J.; Khoshgoftaar, T.M.; Napolitano, A.; Wald, R. Feature selection with high-dimensional imbalanced data. In Proceedings of the 2009 IEEE International Conference on Data Mining Workshops, Miami, FL, USA, 6 December 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 507–514.
65. Gunn, S.R. *Support Vector Machines for Classification and Regression*; Technical Report; School of Electronics and Computer Science, University of Southampton: Southampton, UK, 1998.
66. Li, Y.F.; Xu, Z.H.; Hao, Z.B.; Yao, X.; Zhang, Q.; Huang, X.Y.; Li, B.; He, A.Q.; Li, Z.L.; Guo, X.Y. A comparative study of the performances of joint RFE with machine learning algorithms for extracting Moso bamboo (*Phyllostachys pubescens*) forest based on UAV hyperspectral images. *Geocarto Int.* **2023**, *38*, 2207550. [CrossRef]

67. Chen, J.; Zhang, Y.; Wei, Y.; Hu, J. Discrimination of the contextual features of top performers in scientific literacy using a machine learning approach. *Res. Sci. Educ.* **2021**, *51*, 129–158. [CrossRef]
68. Han, Z.; He, Q.; Von Davier, M. Predictive feature generation and selection using process data from PISA interactive problem-solving items: An application of random forests. *Front. Psychol.* **2019**, *10*, 2461. [CrossRef] [PubMed]
69. Syed Mustapha, S. Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods. *Appl. Syst. Innov.* **2023**, *6*, 86. [CrossRef]
70. Kilinc, M.; Teke, O.; Ozan, O.; Ozarslan, Y. Factors Influencing the Learner's Cognitive Engagement in a Language MOOC: Feature Selection Approach. In Proceedings of the 2023 Innovations in Intelligent Systems and Applications Conference (ASYU), Sivas, Turkiye, 11–13 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.