

Article

Feature Selection and Machine Learning Approaches for Detecting Sarcopenia Through Predictive Modeling

Akhrorbek Tukhtaev ¹, Dilmurod Turimov ¹ , Jiyoun Kim ² and Wooseong Kim ^{1,*} 

¹ Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 13120, Republic of Korea; akhrorbek1599@gachon.ac.kr (A.T.); dilmurod@gachon.ac.kr (D.T.)

² Department of Exercise Rehabilitation & Welfare, Gachon University, Incheon 21936, Republic of Korea; eve14jiyoun@gachon.ac.kr

* Correspondence: wooseong@gachon.ac.kr

Abstract: Sarcopenia is an age-associated condition characterized by a muscle mass and function decline. This condition poses significant health risks for the elderly. This study developed a machine-learning model to predict sarcopenia using data from 664 participants. Key features were identified using the Local Interpretable Model-Agnostic Explanations (LIME) method. This enhanced model interpretability. Additionally, the CatBoost algorithm was used for training, and SMOTE-Tomek addressed dataset imbalance. Notably, the reduced-feature model outperformed the full-feature model, achieving an accuracy of 0.89 and an AUC of 0.94. The results highlight the importance of feature selection for improving model efficiency and interpretability in clinical applications. This approach provides valuable insights into the early identification and management of sarcopenia, contributing to better patient outcomes.

Keywords: sarcopenia; machine learning; feature selection; LIME; CatBoost; random forest; support vector machine; XGBoost; SMOTE-Tomek

MSC: 68T07



Academic Editor: Guoxian Yu

Received: 6 November 2024

Revised: 23 December 2024

Accepted: 26 December 2024

Published: 29 December 2024

Citation: Tukhtaev, A.; Turimov, D.; Kim, J.; Kim, W. Feature Selection and Machine Learning Approaches for Detecting Sarcopenia Through Predictive Modeling. *Mathematics* **2025**, *13*, 98. <https://doi.org/10.3390/math13010098>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sarcopenia, defined by the gradual decline in skeletal muscle mass, strength, and function, poses a major health issue for aging populations. Moreover, it contributes to increased morbidity, reduced quality of life, and higher mortality rates in older adults [1]. Sarcopenia has multiple causes. Central contributors include physiological aging, hormonal changes, inflammatory processes, and decreased physical activity [2].

The causes of sarcopenia are complex and diverse. Aging plays a central role, as the gradual decline in anabolic hormones such as growth hormone and testosterone, coupled with an increase in catabolic cytokines, results in muscle atrophy [3]. Furthermore, chronic diseases, malnutrition, and sedentary lifestyles exacerbate this process. Recent studies have emphasized the importance of identifying at-risk individuals in the early stages to prevent or mitigate the progression of sarcopenia.

With advancements in computational modeling, machine learning (ML) has become an exciting method for the early identification and forecasting of sarcopenia. ML algorithms can analyze complex and large datasets, including clinical, genetic, and imaging data, to identify patterns and markers predictive of sarcopenia before clinical symptoms become apparent [4]. These predictive models offer significant potential in personalized healthcare, enabling interventions tailored to individuals before muscle loss reaches critical thresholds.

This study aims to develop a machine-learning model utilizing raw data related to sarcopenia. The LIME technique identifies the most significant features contributing to sarcopenia [5–7]. Both the complete and selected features are incorporated into the model to enhance model performance and interpretation, and their respective outcomes are analyzed. This approach facilitates the identification of the most influential factors from the sarcopenia dataset, thereby contributing to a deeper understanding of the condition and potential clinical applications.

The motivation for this study stems from the increasing prevalence of sarcopenia, which poses considerable health hazards, especially among older individuals. Despite advancements in medical diagnostics, the complexity of sarcopenia's multifactorial nature makes it challenging to predict and manage effectively. Current clinical tools often lack the precision necessary for early detection, which is critical for intervention and treatment. This study seeks to address this gap by utilizing machine learning methods to improve the precision of sarcopenia prediction by identifying key predictive features [4,8]. By doing so, the research aims to provide a more robust and interpretable model that can be utilized in clinical settings to better identify individuals at risk and improve patient outcomes.

The key contributions of this research include the preprocessing of the sarcopenia dataset to ensure data quality and consistency. The main contributions are as follows:

- The study addressed the issue of class imbalance in the dataset by employing the SMOTE-Tomek technique, which helped to balance the dataset, improving the performance of the predictive model [9,10].
- The study developed a machine-learning model specifically for predicting sarcopenia, leveraging the CatBoost algorithm [11].
- The use of the LIME method helped identify the most important features influencing sarcopenia, improving model interpretability and providing valuable clinical insights [12].
- Both full-feature and reduced-feature models were evaluated, and the reduced-feature model demonstrated superior performance with an accuracy of 0.89 and an AUC of 0.94. This shows that feature selection enhanced both model efficiency and interpretability.
- The study highlights the importance of feature selection for improving the predictive accuracy and clinical applicability of machine-learning models in early sarcopenia identification, which can lead to better patient outcomes.

2. Literature Review

Recent developments in machine learning have greatly improved the early identification and prediction of sarcopenia. This literature review examines the application of various ML techniques in sarcopenia research, focusing on model performance, feature selection, and clinical implications.

Supervised learning models, including random forest (RF), support vector machines (SVM), and extreme gradient boosting (XGB), have demonstrated high accuracy in sarcopenia prediction. The Wild & Deep model achieved an impressive AUC of 0.970, outperforming other classifiers in primary healthcare applications for aging populations [13]. In a comparative study, LightGBM exhibited a prediction accuracy of 93%, making it one of the top-performing models for early sarcopenia detection [14]. These findings underscore the potential of ML algorithms in enhancing diagnostic precision.

The incorporation of feature selection techniques has proven crucial in improving model performance and interpretability. Genetic algorithm-based feature selection significantly enhances diagnostic accuracy by prioritizing key predictors [15]. SHAP (Shapley Additive Explanations) and LIME were utilized to identify waist circumference and grip strength as the most influential predictors of sarcopenia in older adults [16]. Similarly,

handgrip strength (HS), BMI, and sex were highlighted as critical predictive features, emphasizing the models' practical application in personalized medicine [14].

Deep learning approaches and multi-layer perceptron (MLP) networks have shown promise in automatically extracting features from imaging data. A deep learning algorithm achieved high Dice Similarity Coefficient (DSC) scores when segmenting muscle tissue from CT scans, facilitating automatic sarcopenia diagnosis [17]. Integrating body composition variables such as BMI and fat percentage with physical activity metrics significantly boosted model performance in deep learning-based sarcopenia prediction [18,19].

The integration of ML models into clinical workflows represents a promising frontier in sarcopenia diagnosis and management. A deep learning model capable of automatically assessing skeletal muscle index (SMI) from cervical CT scans and predicting survival outcomes based on sarcopenia status in head and neck cancer patients has been developed [20]. This automated approach addresses the limitations of manual assessments and paves the way for integrating artificial intelligence into routine clinical practice.

To tackle the challenge of class imbalance in sarcopenia datasets, researchers have employed advanced techniques such as SMOTE-Tomek. This hybrid method combines the Synthetic Minority Over-sampling Technique (SMOTE) with Tomek links to balance class distributions effectively [21]. Its efficacy in fault classification for electrical machines has been demonstrated, achieving high accuracy across various classifiers [22]. The technique's ability to generate synthetic samples while removing ambiguous instances makes it particularly useful in scenarios with noisy datasets or boundary overlap between classes [23].

Synthetic data generation might introduce samples not truly representative of the minority class, potentially leading to model overfitting [24]. Furthermore, concerns about unintended biases propagating through synthetic data generation have been highlighted, which could impact fairness in sensitive applications like healthcare or credit scoring [25].

The CatBoost algorithm has emerged as a powerful tool in sarcopenia prediction, offering several advantages over traditional methods. Its ordered boosting approach effectively reduces overfitting, particularly in small datasets, as demonstrated in corporate failure prediction models [26]. CatBoost's superior performance in handling categorical features without preprocessing makes it especially valuable in domains like loan risk assessment and credit scoring [27]. CatBoost's versatility is further enhanced when combined with feature selection methods and optimization techniques. Integrating CatBoost with the Sparrow Search Algorithm (SSA) improved classification and prediction accuracy in credit rating tasks [28]. Additionally, CatBoost's effectiveness in handling imbalanced data when paired with SMOTE-Tomek has been demonstrated [29]. Despite its strengths, CatBoost faces limitations in training complexity and hyperparameter tuning. CatBoost can be slower in training compared with other gradient-boosting methods, particularly with large, high-dimensional datasets [13]. The critical need for proper hyperparameter tuning to achieve optimal performance has been emphasized, which can be challenging in resource-limited settings [30].

LIME has emerged as a valuable tool for interpreting complex machine-learning models used in sarcopenia prediction. LIME's model agnosticism enables it to provide explanations for various algorithms, including deep learning and ensemble methods [31]. This flexibility is crucial in healthcare and finance applications, where understanding decision-making processes is essential for building trust and ensuring fairness. LIME's local interpretability allows for the generation of locally faithful explanations by approximating model behavior near specific predictions [32]. This feature is particularly useful for understanding nonlinear decision boundaries. LIME's ability to provide explanations in human-understandable terms makes it accessible to non-experts and helps meet regulatory requirements for explainable AI [33]. Concerns about the stability of LIME explanations

have been noted, as repeated runs can yield different results due to random perturbations [34]. Additionally, LIME's focus on local fidelity might not capture a model's global behavior accurately [31]. Potential biases in LIME explanations due to sampling methods have also been highlighted, which can be problematic in datasets with complex feature interactions [32].

The application of machine learning in sarcopenia prediction and diagnosis has shown significant promise, with various algorithms demonstrating high accuracy and interpretability. Techniques such as SMOTE-Tomek for addressing class imbalance, CatBoost for handling categorical data, and LIME for enhancing interpretability have further advanced the field. However, researchers must remain mindful of the limitations and potential biases associated with these methods. Future work should focus on refining these techniques, addressing their limitations, and further integrating ML models into clinical practice to improve sarcopenia management and patient outcomes.

3. Materials and Methods

The materials and methods section outlines the comprehensive steps to preprocess the dataset, balance the class distribution, and optimize model performance for sarcopenia classification. Initially, the data were standardized and normalized to ensure comparability of features. A comprehensive exploratory data analysis (EDA) was performed to evaluate the relationships between features and their distributions, after which the dataset was divided into training and testing sets. A strategy combining SMOTE and Tomek methods was employed to tackle the class imbalance within the training data, utilizing both oversampling and undersampling approaches. A CatBoost 1.2.7 classifier was employed, with hyperparameters fine-tuned using GridSearchCV. The model was evaluated based on metrics such as accuracy, F1-score, and ROC AUC score, with further optimization of the decision threshold to enhance recall. Feature importance was determined using LIME and permutation importance, leading to selecting the top features for model retraining. Cross-validation was implemented to assess the model's robustness, and final evaluations were conducted to evaluate the effectiveness of models with all features and with the selected subset.

3.1. Data Collection and Sample Size

The dataset used in this study was obtained from the Institute of Human Convergence Health Science at Gachon University. Data collection was conducted across various community settings in Incheon, including social welfare centers, daycare centers, and senior welfare centers, over nine months, from 1 September 2022 to 31 May 2023. This extended period allowed for the acquisition of a comprehensive and diverse sample from different locations. The prolonged data collection period contributed to a representative sample of the target population.

Informed consent was obtained from all 664 participants before data collection, ensuring ethical compliance. Standardized protocols were employed throughout the process to maintain consistency and accuracy in the data. The dataset consists of 76 variables categorized into demographic information, physical measurements, functional assessments, medical evaluations, and health conditions, offering a detailed and multifaceted view of the participant's health and lifestyle characteristics, refer to Table A1.

3.2. Data Preprocessing

The target variable, *sarcopenia_2*, is a binary variable that indicates the presence or absence of sarcopenia—a condition characterized by a loss of skeletal muscle mass and function. A value of 0 represents the absence of sarcopenia, while a value of 1 indicates its

presence. Predicting sarcopenia is the primary objective of our study, making this variable central to our analysis.

To ensure the comparability of features, the data were standardized and normalized. Standardization was performed using the Z-score normalization technique:

$$X_{\text{standardized}} = \frac{X - \mu}{\sigma}$$

where X represents the feature values, μ is the mean, and σ is the standard deviation.

Normalization was applied to rescale the standardized data into the range $[0, 1]$, using the min–max scaler:

$$X_{\text{normalized}} = \frac{X_{\text{standardized}} - X_{\min}}{X_{\max} - X_{\min}}$$

where $X_{\min}$ and $X_{\max}$ represent the lowest and highest values of the feature, respectively.

Z-score normalization is a widely used preprocessing technique that standardizes variables to have a mean of 0 and a standard deviation of 1. Although it assumes variables follow a normal distribution, it remains effective for machine learning models even with non-normally distributed data. This is particularly applicable in models like CatBoost, which are non-parametric and rely on decision-tree-based ensemble learning [11]. These models do not require input variables to follow a specific distribution, as they are inherently robust to variations in feature scaling and distribution.

Applying z-score normalization ensures consistent scaling of variables, preventing features with larger numerical ranges from dominating the learning process [35]. Despite the potential mismatch with non-normal distributions, z-score normalization retains the relative relationships among values within a feature, which is critical for machine learning workflows. Additionally, the model's robustness to feature scaling was validated empirically through cross-validation, confirming that z-score normalization did not degrade performance but instead enhanced consistency and comparability across features.

3.3. Data Splitting and Resampling

The dataset was divided into training (70%) and testing (30%) subsets. This was performed using the `train_test_split` function from scikit-learn. A random seed of 42 ensured consistent results.

In this study, SMOTE+Tomek was selected to mitigate the class imbalance within the sarcopenia dataset. In particular, this approach addresses the disparity in the representation of sarcopenia-positive cases. By generating minority class samples, SMOTE guarantees that the predictive model encounters an adequate number of sarcopenia instances. At the same time, Tomek links enhance class boundaries by eliminating near-boundary instances that create uncertainty. This hybrid technique enhances class separation and reduces overfitting, leading to more accurate boundary definition between classes and improving the model's sensitivity to sarcopenia cases. Consequently, SMOTE+Tomek contributes to a more balanced dataset that optimizes recall, precision, and generalization, aligning with the study's goal of reliable sarcopenia prediction [22]. This resulted in a balanced dataset for training, as visualized in the post-resampling target distribution (Figure 1).

The figure illustrates the class distribution for binary classification in different stages of data processing. In the full dataset (Figure 1a), the majority class (class 0) accounts for 78.92% (524 samples), while the minority class (class 1) represents 21.08% (140 samples), highlighting an imbalanced dataset. In the test dataset before applying SMOTE+Tomek (Figure 1b), the majority class maintains a similar imbalance with 76.50% (153 samples), and the minority class constitutes 23.50% (47 samples). For the training dataset before resampling (Figure 1c), the majority class is 79.96% (371 samples), and the minority class

is 20.04% (93 samples), continuing the imbalance. After applying SMOTE+Tomek to the training dataset (Figure 1d), the class distribution is perfectly balanced, with both classes contributing equally at 50% each (371 samples per class), ensuring that the training data are suitable for addressing the class imbalance issue.

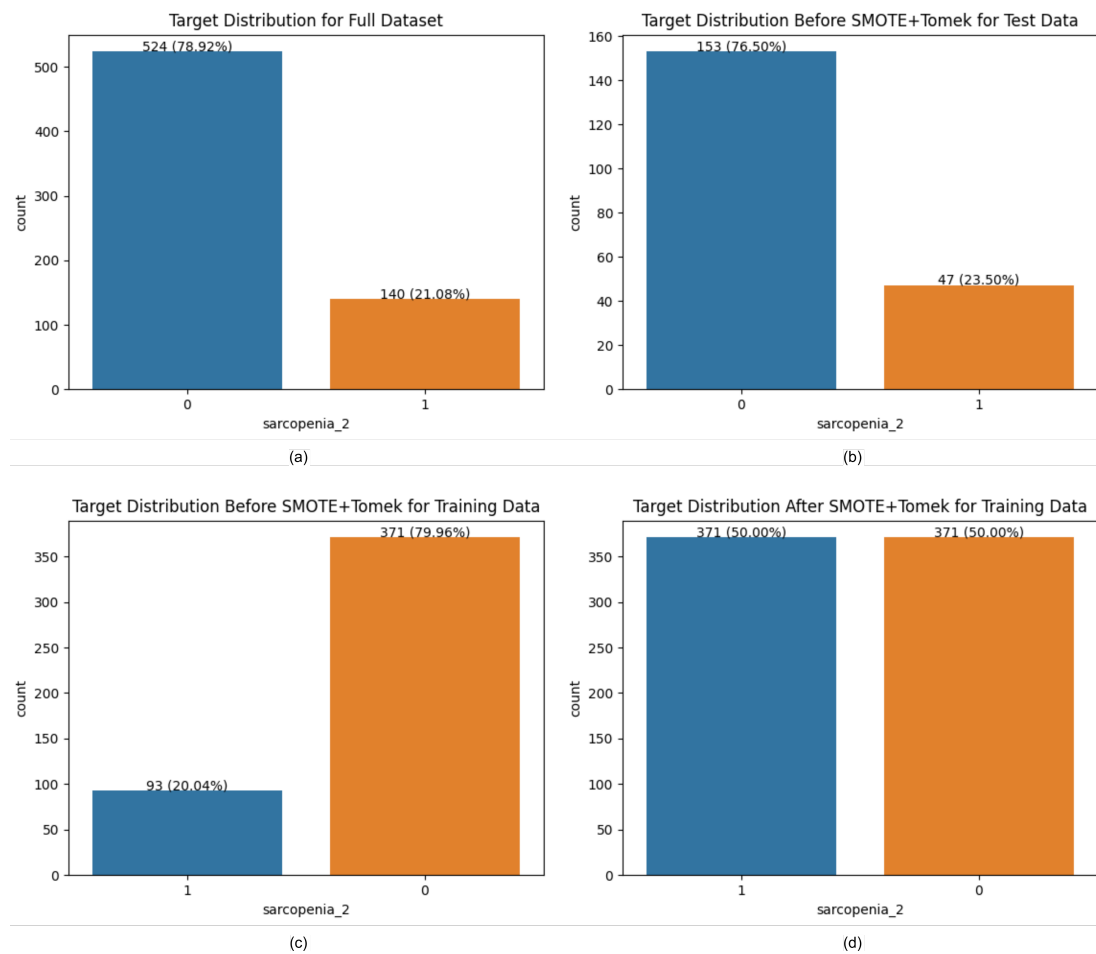


Figure 1. Target distribution across the dataset and after applying SMOTE+Tomek. (a) Target distribution for the full dataset before SMOTE+Tomek. (b) Target distribution for the test data before SMOTE+Tomek. (c) Target distribution for the training data before SMOTE+Tomek. (d) Target distribution for the training data after SMOTE+Tomek, showing balanced classes.

3.4. Training Model

CatBoost is a gradient-boosting technique that creates an ensemble of decision trees to iteratively reduce a loss function. The model prediction is represented as

$$G(x) = \sum_{j=1}^M \beta_j g_j(x),$$

where $G(x)$ is the overall model prediction for the input x , $g_j(x)$ are weak learners, β_j are their respective weights, and M is the total number of weak learners in the ensemble. Each new learner is fitted to the negative gradient of the loss function at each iteration.

$$\gamma_j(x) = -\frac{\partial \mathcal{L}(y, G_{j-1}(x))}{\partial G_{j-1}(x)},$$

where $\gamma_j(x)$ is the negative gradient for the loss function, $\mathcal{L}(y, G_{j-1}(x))$ is the loss function at iteration $j - 1$, y represents the true target values, and $G_{j-1}(x)$ is the model prediction at iteration $j - 1$.

For classification, CatBoost typically uses the logarithmic loss function:

$$\mathcal{L}(y, G(x)) = - \sum_{k=1}^N [y_k \log(\hat{y}_k) + (1 - y_k) \log(1 - \hat{y}_k)],$$

where $\mathcal{L}(y, G(x))$ is the loss function for classification, N is the total number of data points, y_k represents the true label for the k -th data point ($y_k \in \{0, 1\}$), and $\hat{y}_k$ represents the predicted probability of the positive class for the k -th data point.

For regression, the squared error loss is

$$\mathcal{L}(y, G(x)) = \frac{1}{N} \sum_{k=1}^N (y_k - G(x_k))^2,$$

where $\mathcal{L}(y, G(x))$ is the loss function for regression, y_k is the true target value for the k -th data point, and $G(x_k)$ is the predicted value for the k -th data point.

To prevent overfitting, CatBoost uses L2 regularization and learning rate shrinkage, represented as

$$G_j(x) = G_{j-1}(x) + \eta \cdot \beta_j g_j(x),$$

where $G_j(x)$ is the updated model prediction after iteration j , $G_{j-1}(x)$ is the model prediction from the previous iteration ($j - 1$), η is the learning rate, which controls the contribution of each new weak learner to the model, and β_j and $g_j(x)$ are the weight and prediction of the j -th weak learner, respectively.

Additionally, CatBoost incorporates ordered boosting to reduce prediction shifts and improve model stability.

3.5. Feature Importance

We used the LIME method to pinpoint the key features that influence the predictions made by the model. LIME clarifies specific predictions by locally approximating the model and determining the most significant features for each case. This approach proved especially beneficial in gaining a clear insight into the model's reasoning process.

LIME was chosen in this study because of its ability to improve model interpretability, a critical factor for clinical applications of machine learning in the prediction of sarcopenia. By providing local explanations, LIME facilitates insight into the contribution of individual features, such as basal metabolic rate (BMR) and body fat percentage, to each specific prediction, thereby aligning with personalized healthcare approaches [36]. The model-agnostic nature of LIME allows it to be applied flexibly to the CatBoost algorithm without requiring access to its internal structure, supporting the study's focus on accuracy and interpretability. This feature is particularly valuable for identifying influential predictors within a complex, multifactorial condition like sarcopenia, enabling clinicians to understand better and trust the model's decisions. Furthermore, LIME's local interpretability assists in refining feature selection and optimizing the model by highlighting the most impactful variables, which enhances both computational efficiency and predictive accuracy. By increasing transparency in decision-making, LIME also helps meet ethical and regulatory standards for explainable AI, making it a suitable choice for healthcare applications that prioritize clarity, fairness, and reliability [37].

LIME is a technique designed to interpret complex machine learning models by generating local approximations of the model's behavior around individual predictions. The method assumes that while the global model may be complex and nonlinear, the model can be approximated locally by a simpler, interpretable model, such as a linear regression or decision tree.

Given an instance x , LIME perturbs the input by generating several samples in its local vicinity and observes how the model's predictions change. It then fits a simpler interpretable model $g(x)$ to approximate the original model $f(x)$ within this local neighbourhood. The objective is to minimize the following weighted loss function:

$$\mathcal{L}(f, g, \mu_x) = \sum_{z \in Z} \mu_x(z) (f(z) - g(z))^2,$$

where $\mu_x(z)$ is a proximity measure that assigns higher weights to samples closer to the original instance x . The simplified model $g(x)$ provides insight into which features were most important in the original model's prediction.

By focusing on local rather than global behavior, LIME makes model predictions interpretable without requiring access to the model's internal structure, ensuring it remains model-agnostic.

3.6. Comparative Feature Selection Approaches

In this study, we compare LIME with RFE and MRMR as feature selection techniques to identify the most predictive features for sarcopenia. LIME offers instance-level interpretability, which enhances understanding of individual predictions, an important aspect of clinical decision-making. RFE and MRMR, on the other hand, provide globally optimized feature sets by focusing on overall model performance and minimizing redundancy, respectively. By applying and comparing these methods, we aim to identify a feature selection approach that optimally balances accuracy, interpretability, and clinical applicability, offering insights into the most effective feature selection strategies for predictive modeling in sarcopenia (Table 1).

The LIME algorithm is a method for interpretability that aims to clarify the predictions made by intricate machine-learning models by locally approximating these predictions with a simpler, more interpretable surrogate model. LIME perturbs an instance within its local neighborhood to create synthetic samples; subsequently, a more straightforward and interpretable model (like linear regression) is applied to these samples, with weights reflecting their closeness to the initial instance. This approach offers insights into which features most influence specific predictions, making it particularly valuable for model interpretability in fields requiring decision transparency, such as healthcare [31,38].

The key steps in LIME are as follows:

1. Data perturbation: perturbs the input instance around a local neighborhood, generating synthetic samples by altering feature values randomly.
2. Weighting samples: applies a proximity weighting kernel to emphasize samples closer to the instance being explained.
3. Local model fitting: applies a straightforward, easily interpretable model (such as linear regression) to the weighted synthetic samples to locally estimate the decision boundary of the original model.
4. Interpretability: the local model's coefficients provide feature weights that indicate their importance for the specific prediction.

Recursive Feature Elimination (RFE) is a technique for selecting features that progressively eliminate the least significant ones according to importance scores related to the model, like coefficients in linear models or feature importances in models based on trees. RFE trains a model repeatedly, ranking features and discarding those with the lowest relevance until the desired number of features remains. This method effectively optimizes model accuracy by retaining only the most predictive features, though it depends heavily on the chosen model's inherent importance metrics [39].

Minimum Redundancy Maximum Relevance (MRMR) is a feature selection method based on information theory that seeks to find features that are highly pertinent to the target variable while being as non-redundant as possible among themselves. MRMR ranks features based on mutual information to maximize relevance to the target and minimize redundancy among selected features, making it ideal for high-dimensional datasets with potential multicollinearity. This approach has been widely used in complex predictive tasks for its ability to enhance model performance by focusing on unique, non-overlapping features [40,41].

Table 1. Comparative analysis of LIME, RFE, and MRMR based on recent studies.

Criterion	LIME	RFE	MRMR
Type of Interpretation	Local interpretability: explains individual predictions for instance-specific insights [33]	Global interpretability: identifies feature importance across the entire dataset [42]	Global interpretability with focus on minimizing feature redundancy [40]
Feature Selection Mechanism	Model-agnostic surrogate model; approximate model predictions locally with weighted samples [38]	Iterative feature elimination based on feature importance derived from model coefficients or feature scores [42]	Mutual information-based, ranking features by relevance to target variable and minimizing redundancy among selected features [40]
Computational Complexity	Moderate to high; depends on model complexity and number of samples used in perturbation [34]	Generally moderate; grows with dataset size and feature count [42]	Low to moderate; efficient for high-dimensional data due to the direct computation of mutual information [41]
Model Dependence	Model-agnostic; applicable to any model, including deep learning and ensemble methods [36]	Model-dependent; relies on feature importance from specific models, typically tree-based or linear models [39]	Model-independent; does not require a model, making it versatile across feature selection tasks [41]
Sensitivity to Data Distribution	Sensitive to instance-specific data; explanations may vary significantly between instances [31]	Sensitive to the selected model's inherent feature importance metrics; susceptible to overfitting if model complexity is high [13]	Less sensitive due to use of information-theoretic metrics; generally robust across data distributions [40]
Primary Use Cases	Model interpretability, personalized explanations, understanding complex models in healthcare and finance [37]	Feature ranking and dimensionality reduction in high-dimensional predictive modeling tasks [39]	High-dimensional data analysis with feature relevance and low redundancy, such as in genomics and text processing [10,40]
Limitations	Instability in explanations due to random perturbations; lacks global feature importance [31]	Relies on model quality for feature ranking; iterative process can be computationally intensive [39]	Potential loss of relevant features if they exhibit redundancy; lacks instance-specific insights [10]

3.7. Model Training with Selected Features

After feature reduction, standardization and normalization were reapplied to ensure consistency across the reduced feature set. This step is necessary because feature reduction methods can alter the statistical properties of the remaining variables, including their mean

and variance. Without re-standardization, differences in feature scales could negatively impact the performance and stability of machine learning models [43]. Standardizing the reduced feature set ensures that the selected features remain comparable, thereby optimizing the training process and improving model convergence. The CatBoost classifier was retrained using these reduced features, and its robustness was assessed through 10-fold cross-validation.

A streamlined version of the CatBoost model was developed, applying the same train-test split, feature preprocessing, and hyperparameter tuning as the original model. The performance of both the full and reduced models was then compared across several predictive metrics to evaluate their effectiveness.

3.8. Evaluation Metrics

Accuracy, serving as a foundational metric, measured the ratio of correctly identified cases to the total number of samples, while the confusion matrix offered a detailed breakdown of true positives, true negatives, false positives, and false negatives, thus highlighting the model's error patterns. To provide a more specific assessment of performance per class, a classification report was created, including precision, recall, and F1-score for each category, which illuminated the model's equilibrium between false positives and false negatives.

The ROC AUC Score (Area Under the Receiver Operating Characteristic Curve) was computed to evaluate the model's discrimination capacity across varying threshold settings, with values approaching 1.0 indicative of strong classification performance. The ROC curve, plotted from the calculated False Positive Rate (FPR) and True Positive Rate (TPR) values, visually represented the trade-off between these two metrics. The area under the curve (AUC) acted as a solid indicator of the model's capacity to differentiate between categories, with elevated AUC values indicating better sensitivity and specificity. This comprehensive evaluation provided an in-depth perspective on the model's accuracy, reliability, and generalizability within the clinical context.

4. Result

The results section presents the outcomes of the implemented methodologies, including model training and evaluation. The findings are structured to provide insights into model performance both before and after feature selection and the impact of data balancing techniques. Specifically, metrics such as accuracy, F1-score, ROC AUC score, and confusion matrices are used to assess the predictive capabilities of the CatBoost model. Additionally, the comparison between the full-featured model and the reduced-feature model is highlighted to demonstrate the effect of feature selection on model efficiency and performance.

4.1. Full-Featured Model Results

The CatBoost model employed in this research was fine-tuned through grid search with cross-validation to identify the most effective parameters. The final selection of hyperparameters for the model is shown in Table 2. These parameters were adjusted to achieve a balance between accuracy and generalization, taking into account the unique features of the dataset:

- **Depth:** a tree depth of 7 was selected to allow the model to capture sufficient complexity in the data while avoiding overfitting.
- **Iterations:** the number of boosting iterations was set to 500, ensuring adequate training cycles without excessive computation time.

- **Learning rate:** a learning rate of 0.2 was chosen to allow the model to converge faster, balancing the need for a lower error rate and shorter training time.
- **Scale positive weight:** this parameter was set to 1, indicating that the positive and negative classes are treated equally in the model.
- **Subsample:** a subsampling rate of 0.7 was selected, meaning that the model randomly selects 70% of the data for each iteration, reducing overfitting while maintaining performance.

These hyperparameters were critical in achieving the optimal performance of the CatBoost model for this specific dataset.

Table 2. Optimal parameters for CatBoost model.

Parameter	Value
Depth	7
Iterations	500
Learning Rate	0.2
Scale Positive Weight	1
Subsample	0.7

The classification report offers comprehensive metrics regarding the effectiveness of the binary classification model, highlighting precision, recall, and F1-score for every class. The model exhibits an overall accuracy of 0.87, indicating that it accurately classified 87% of the instances.

The classification report gives a summary of how well the model is performing in terms of precision, recall, and F1-score. A precision of 0.82 signifies that 82% of the instances that the model predicted as positive are correct. The recall of 0.81 shows that the model successfully detects 81% of the positive cases found in the dataset. The F1-score, reflecting a balance between precision and recall, is measured at 0.82, suggesting that the model strikes a solid balance between these two metrics. These values illustrate that the model operates effectively, with only slight differences between precision and recall, suggesting it can reliably detect positive cases while maintaining a manageable rate of false positives and false negatives. This level of performance suggests that the model is well balanced and appropriate for applications where both precision and recall are crucial.

The confusion matrix presented offers a detailed assessment of the model's classification effectiveness for both classes (0 and 1). It is structured to display the number of accurate and inaccurate predictions made by the model:

- The cell in the top-left corner (141) indicates the count of true negatives, signifying that the model accurately identified class 0. This indicates that the majority of instances belonging to class 0 were accurately identified.
- The top-right cell (12) denotes false positives, where the model incorrectly predicted class 1 for instances that belong to class 0. These are cases of misclassification that contribute to the error rate for class 0.
- The bottom-left cell (14) contains false negatives, where the model failed to identify instances of class 1, mistakenly classifying them as class 0. This contributes to the recall error for class 1.
- The cell in the bottom-right corner (33) indicates the true positives, reflecting the instances where the model accurately predicted class 1. Although fewer in number than class 0 predictions, this indicates that a substantial portion of class 1 instances were correctly identified (Figure 2).

Figure 2 depicts the confusion matrix for the test dataset, which retains the original class imbalance. The test dataset contains 30% of the full dataset (199 cases), with 153 samples (76.5%) in the majority class and 47 samples (23.5%) in the minority class. The total number of cases in Figure 2 aligns with the test dataset size and reflects the model's performance in an unbiased evaluation scenario.

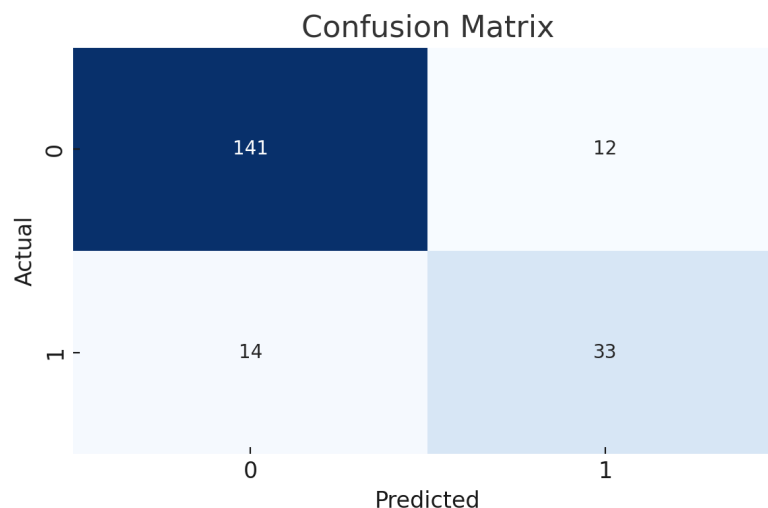


Figure 2. Confusion matrix for the classification model.

The ROC curve shown in Figure 3 offers an extensive assessment of the classifier's performance by illustrating the balance between the true positive rate (TPR) and false positive rate (FPR) at various decision thresholds. With an AUC (Area Under the Curve) of 0.93, the model exhibits strong discriminative capability, signifying its effectiveness in differentiating the positive from the negative classes. The curve ascends sharply towards the top-left corner, indicating a high true positive rate with few false positives. This steep rise is indicative of the model's high sensitivity and specificity, minimizing both type I and type II errors. The AUC value approaching 1.0 underscores the robustness of the model, making it well suited for binary classification tasks where precise prediction is critical. Overall, the ROC curve suggests that the classifier performs exceptionally well, providing reliable predictive capabilities across a wide range of thresholds.

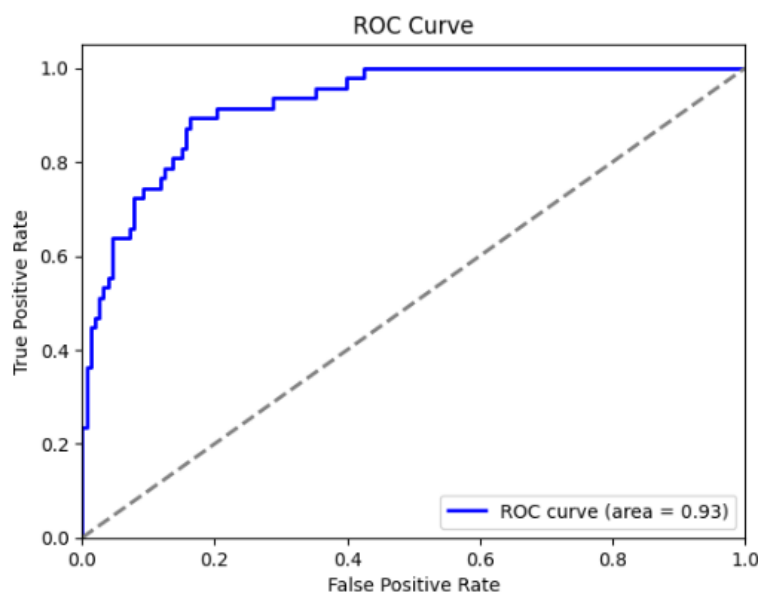


Figure 3. ROC curve for the classification model.

4.2. Selected Features

The study selected a precise subset of 10 features based on their significant contributions to sarcopenia prediction accuracy and interpretability, as identified by LIME. This focused selection enhanced the model's computational efficiency while retaining predictive power, allowing clinicians to interpret key physiological and functional indicators linked to sarcopenia. These features were chosen to ensure that the model captures relevant information without unnecessary complexity, contributing to an optimal balance between model simplicity and performance.

Table 3 provides a brief explanation of the key features used in the predictive model. Each feature is listed alongside its corresponding description, offering insight into the physiological or assessment metrics involved in the model.

Table 3. Description of top features in the model.

Feature	Description
BMR	Basal Metabolic Rate
Sarc_calf	Sarcopenia self-assessment questionnaire + calf circumference
G_TUG	Grade of Timed Up and Go test
G_SS	Grade of Stand and Sit
ND_SLS	Nondominant Single leg stance test
FVC	Forced Vital Capacity
SS	Stand and Sit, the number of stand up in 30 s
MWT2	2-min walking test
D2_s	Locomotion
Percent_BF	Percent Body Fat

The LIME method was applied to assess the contribution of individual features to the prediction process. The plot in Figure 4 illustrates the top features ranked by their contributions to the model's output. Notably, $BMR_{kcal} > 0.38$ had the highest negative contribution (-0.33), suggesting that elevated basal metabolic rate reduces the likelihood of a positive prediction. Several other features, including $SARC_Calf \leq 0.15$, $G_TUG \leq 0.25$, and $G_SS \leq 0.25$, also negatively impacted the prediction, with contributions around -0.11 each. These features primarily relate to physical assessment metrics, implying that specific physiological thresholds are critical in driving negative outcomes in this model.

Conversely, the feature $Percent_BF \leq 0.44$ exhibited the strongest positive contribution (0.38), indicating that lower body fat percentage significantly increases the likelihood of a positive prediction. Other features, such as $0.51 < D2_s \leq 0.69$ and $MWT2 > 0.54$, contributed positively, albeit to a lesser extent, with contributions of 0.05 each. These features reflect the subjects' physical mobility and locomotion capabilities, reinforcing their importance in determining positive health outcomes within this context.

Table 3 and Figure 4 illustrate the importance of the selected features as identified by the LIME framework. The "feature contribution" values in Figure 4 represent the magnitude and direction of each feature's influence on the model's predictions. Positive values (orange bars) indicate features that contribute positively to the prediction of sarcopenia (presence), while negative values (blue bars) indicate features that oppose the prediction of sarcopenia (absence). The "top contributions" refer to the subset of features with the largest absolute contributions to the prediction, as identified by LIME for specific instances. These top features are quantitatively assessed by their cumulative contributions compared with all 76 features, demonstrating that the selected features capture the majority of predictive power while reducing model complexity.

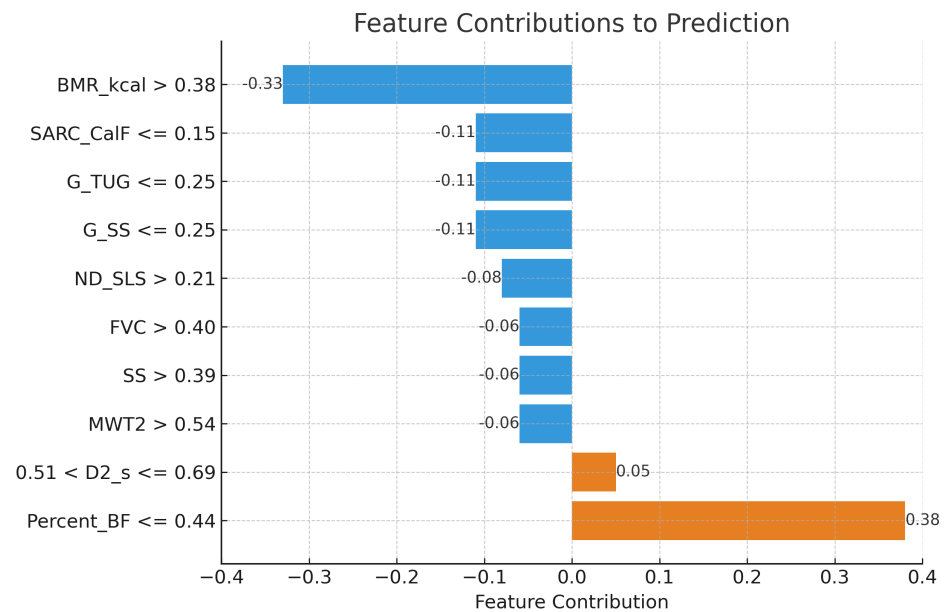


Figure 4. The bar chart shows the top features contributing to the model's prediction as interpreted by the LIME method.

Figure 5 shows the ranking of features based on their contribution to the classification model, with BMR_kcal having the highest importance, followed by FVC and D2_s, highlighting their significant role in sarcopenia prediction. This visualization demonstrates how these key physiological and functional indicators contribute disproportionately to the prediction outcomes, reinforcing their relevance in the context of sarcopenia assessment.

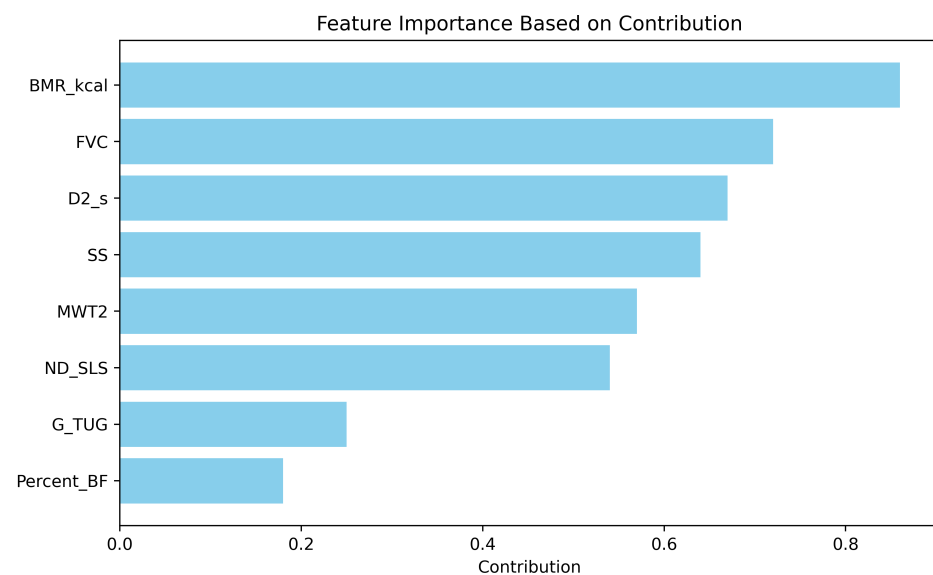


Figure 5. Feature importance based on contribution scores for predicting sarcopenia.

4.3. Comparison of Feature Selection Methods

To compare the feature selection results from LIME, RFE, and MRMR, we examine each method's contribution to predictive accuracy, interpretability, and feature diversity, highlighting the distinct attributes each selection yields for the sarcopenia prediction model.

LIME offers instance-level interpretability by identifying the most influential features for specific predictions rather than providing a global ranking across the dataset. The LIME-based selection includes features that contribute directly to local predictions, such as basal metabolic rate (BMR), Timed Up and Go (TUG) scores, and percent body fat (Percent_BF).

These features are individually relevant for clinical assessments, aligning with interpretability goals in healthcare applications where understanding each prediction's driving factors is crucial. LIME's selection emphasizes features directly affecting individual outcomes, a notable advantage in patient-specific evaluations.

RFE optimizes feature selection by iteratively removing the least significant features according to the model's performance across the entire dataset. The RFE selection incorporates features like age, percent body fat (Percent_BF), and forced expiratory volume (FEV1). These features offer a global relevance that enhances model performance by consistently emphasizing statistically significant predictors. While RFE lacks LIME's local interpretability, it is robust in identifying features that optimize model accuracy across all instances, essential in generalizing predictions within larger datasets (Table 4).

MRMR prioritizes features with maximum relevance to the target variable and minimal redundancy among selected features. The MRMR feature set includes Sarcopenia Calf (SARC_CalF), BMR, and Dorsal Muscle Strength (Dorsal_L_1), which provide diverse and unique information on sarcopenia indicators. MRMR reduces multicollinearity, allowing for a compact and informative feature subset that avoids redundant predictors. This method's information-theoretic foundation provides an optimal balance between feature relevance and diversity, reducing complexity without sacrificing predictive capacity (Table 5).

Table 4. RFE-based feature selection.

Feature	Description
Age	Age of Participant
Percent_BF	Percent Body Fat
BMR_kcal	Basal Metabolic Rate
Plantar_L_1	Left Plantar Muscle Strength
Dorsal_L_1	Left Dorsal Muscle Strength
D_Dorsal	Dominant Dorsal Muscle Strength
SLS_R	Right Single Leg Stance Test
SARC_CalF	Sarcopenia self-assessment questionnaire + calf circumference
FEV1	Forced Expiratory Volume in 1 Second
Hyperlipidemia	Presence of elevated levels of lipids

Table 5. MRMR-based feature selection.

Feature	Description
SARC_CalF	Sarcopenia self-assessment questionnaire + calf circumference
BMR_kcal	Basal Metabolic Rate
G_SS	Grade of Stand and Sit
PEF	Peak Expiratory Flow
Dorsal_L_1	Left Dorsal Muscle Strength
ND_Dorsal	Nondominant Dorsal Muscle Strength
Area_Dong	Geographic Area (Dong)
ND_Plantar	Nondominant Plantar Muscle Strength
D_Dorsal	Dominant Dorsal Muscle Strength
Place	Location of Testing Facility

Across the three methods, there is partial overlap, particularly with features like BMR and SARC_CalF, selected by both LIME and MRMR. This convergence suggests these features are universally significant for sarcopenia prediction. However, there are notable differences in the selection sets: age and percent body fat are prioritized in RFE,

while MRMR emphasizes muscle strength indicators such as dorsal muscle strength and nondominant plantar strength. These distinctions reflect each method’s underlying priorities—LIME and MRMR align with clinical relevance and diversity, while RFE emphasizes features that maximize global model performance.

4.4. Reduced-Feature Model Results

The features used in the “reduced-feature model” are the features listed in Table 3 that were selected for this model. These features were chosen based on their importance as identified through the feature selection process. Table 3 provides a comprehensive description of these selected features, which represent the most significant contributors to the model’s predictive performance while effectively reducing dimensionality. The reduced-feature model achieved an accuracy of 0.89, indicating that the model using a chosen group of attributes performs comparably to the full-feature model. The classification report reveals a precision of 0.85, meaning that 85% of the instances identified as positive were accurately classified. The recall was recorded at 0.83, which signifies that the model correctly identified 83% of the actual positive instances. The F1-score, a harmonic mean of precision and recall, was 0.84, reflecting a balanced trade-off between the two metrics. These results suggest that the reduced set of features is sufficient to maintain robust predictive performance while simplifying the model, which could improve interpretability and computational efficiency. The high performance of the reduced-feature model demonstrates that the selected features are both informative and predictive, supporting the model’s ability to generalize well to unseen data.

The confusion matrix plot in Figure 6 illustrates the performance of the model using the selected features. It shows the following:

- True Negatives (144): correct predictions for class 0.
- False Positives (9): instances where the model incorrectly predicted class 1 instead of class 0.
- False Negatives (13): instances where the model incorrectly predicted class 0 instead of class 1.
- True Positives (34): correct predictions for class 1.

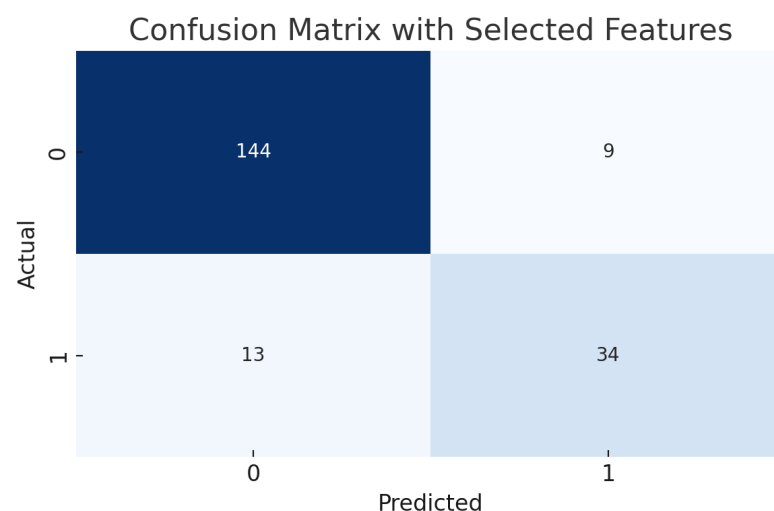


Figure 6. The confusion matrix illustrates the performance of the model using a selected subset of features.

Feature interpolation was applied as a data augmentation technique to enhance the training data. Prior to augmentation, the model achieved an accuracy of 0.89, precision of 0.85, recall of 0.83, and an F1-score of 0.84. After augmentation, the accuracy slightly

decreased to 0.88, while the precision remained stable at 0.84, recall improved to 0.85, and the F1-score remained consistent at 0.84. These results indicate that data augmentation improved the model's recall, reflecting better detection of the positive class while maintaining overall predictive performance. Model changes can be shown in Figure 7 after data augmentation.

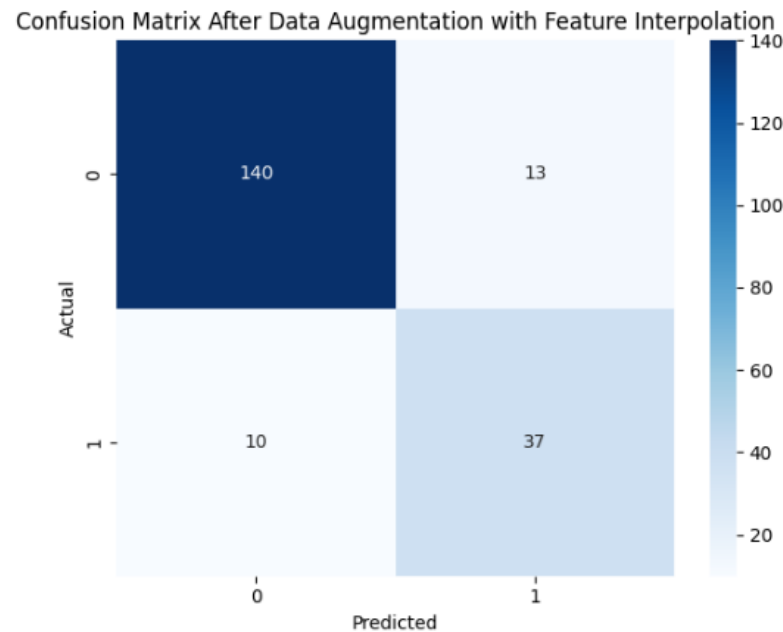


Figure 7. The confusion matrix illustrates the performance of the model after data augmentation.

The ROC curve, as illustrated in Figure 8, provides a comprehensive evaluation of the reduced-feature model's ability to differentiate between the two classes across varying decision thresholds. The graph illustrates the relationship between the true positive rate (sensitivity) and the false positive rate ($1 - \text{specificity}$), highlighting how well the model performs in achieving a balance between these important metrics. The area under the curve (AUC) is calculated to be 0.94, signifying the model's strong discriminatory capability. The sharp rise toward the top-left corner of the graph emphasizes the model's capability to attain high sensitivity while keeping the false positive rate, underscoring its robustness and efficacy when utilizing the selected subset of features.

Figure 9 illustrates the convergence of training loss and validation loss over iterations for the CatBoost model. Both curves decrease steadily and stabilize after approximately 200 iterations, indicating that the model successfully minimizes the loss without overfitting. The close alignment between the training and validation loss curves demonstrates good generalization performance, ensuring that the model is learning meaningful patterns from the data rather than fitting noise.

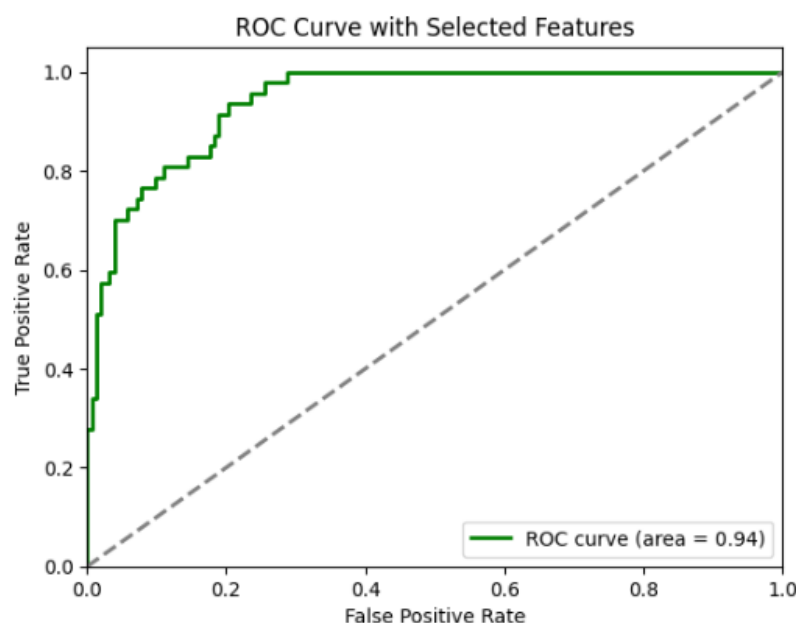


Figure 8. The ROC curve illustrates how well the model performs with the chosen features, achieving an area under the curve (AUC) of 0.94.

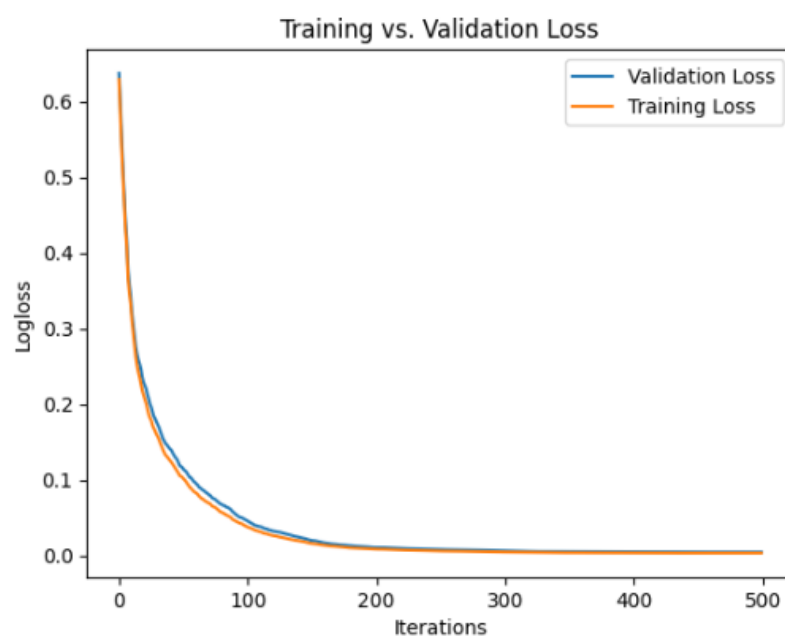


Figure 9. Training and validation loss curves for the CatBoost model during training.

4.5. Model Accuracy Comparison Across Feature Selection Techniques

Table 6 provides a comparative analysis of model accuracy across four machine learning algorithms: CatBoost, RF, SVM, and XGBoost, using different feature selection techniques: raw data (no feature selection), LIME-based selection, RFE-based selection, and MRMR-based selection. This comparison highlights the impact of each feature selection method on the predictive performance of the models. Table 6 represents the performance metrics on the test dataset.

Table 6. Accuracy of models with different feature selection techniques.

Feature Selection Technique	CatBoost	RF	SVM	XGBoost
Raw Data	0.87	0.80	0.84	0.86
LIME	0.89	0.86	0.87	0.88
RFE	0.88	0.87	0.85	0.87
MRMR	0.83	0.84	0.81	0.85

We made the following observations:

1. CatBoost demonstrates the highest accuracy across all feature selection methods, with LIME-based selection achieving the best accuracy (0.89). The raw data accuracy is 0.87, indicating that feature selection with LIME and RFE provides a slight improvement, while MRMR reduces accuracy to 0.83.
2. RF shows a notable increase in accuracy when feature selection techniques are applied. Starting with a base accuracy of 0.80 using raw data, RF achieves 0.86 with LIME and 0.87 with RFE, highlighting these methods' positive impact. MRMR results in a moderate accuracy of 0.84.
3. SVM's performance is improved with LIME-based feature selection, reaching an accuracy of 0.87, the highest for this algorithm. The raw data accuracy is 0.84, while RFE and MRMR achieve 0.85 and 0.81, respectively.
4. XGBoost performs consistently well across feature selection techniques, with the highest accuracy of 0.88 observed for LIME-based selection. RFE maintains a high accuracy of 0.87, while the raw data and MRMR selections yield slightly lower accuracies of 0.86 and 0.85, respectively.

Overall, LIME-based feature selection consistently yields the highest accuracy across the models, particularly enhancing the performance of CatBoost, RF, and SVM. The RFE-based selection also improves accuracy, especially for RF and SVM, suggesting its robustness in identifying globally relevant features. MRMR-based selection, while providing a diverse feature set, demonstrates slightly lower accuracy across models, likely due to its emphasis on minimizing redundancy over global relevance.

This comparison underscores that the choice of feature selection method impacts model performance. LIME and RFE show significant improvements, making them suitable choices for optimizing predictive accuracy in sarcopenia detection models.

5. Discussion

The manuscript focuses on developing a machine learning model for predicting sarcopenia, leveraging advanced techniques like the CatBoost algorithm and LIME for interpretability. A balanced dataset was created using SMOTE-Tomek, and feature selection enhanced model efficiency and clinical relevance. The results highlight the model's ability to achieve high accuracy and interpretability by prioritizing critical features such as basal metabolic rate and physical performance metrics. Future work aims to validate findings across diverse populations. Additionally, it seeks to incorporate broader clinical variables to improve generalizability.

To address the limitations of using surface electromyography (sEMG) in clinical settings, we acknowledge the challenges posed by the need for specialized equipment and variability in measurement protocols. Recent advancements, however, have made sEMG more feasible for clinical use. Portable and miniaturized systems now reduce the equipment burden, enabling their application in diverse environments [44]. Improvements in signal preprocessing and denoising techniques further enhance the reliability of sEMG measurements, even under variable conditions [45]. Ongoing efforts to standardize protocols and

include larger cohorts aim to mitigate sample size limitations and improve generalizability [46]. Additionally, the integration of physical activity and obesity features, such as moderate-to-vigorous physical activity (MVPA) and waist circumference, plays a critical role in sarcopenia prediction. Objective methods, such as accelerometer-based activity tracking, have been shown to correlate with reduced risks of severe sarcopenia and sarcopenic obesity, while stratified analyses address variability across populations [18,47,48]. These strategies enhance model reliability and generalizability, ensuring applicability in real-world scenarios.

Despite its effectiveness, the reliance on LIME for feature selection and interpretation presents challenges. LIME's instability, caused by random perturbations, and its focus on local explanations limit its ability to capture global feature interactions. Recent advancements, such as OptiLIME and DLIME, address these issues by improving stability through optimized perturbation strategies and deterministic sampling [49,50]. Furthermore, hybrid approaches integrating LIME with SHAP provide both local and global interpretability, offering a more comprehensive understanding of feature interactions [51]. Similarly, while the SMOTE-Tomek technique effectively addresses class imbalance, synthetic samples may not accurately represent the minority class, leading to potential overfitting and reduced generalizability [9,52]. Exploring advanced techniques, such as metaheuristic-based feature selection, could further enhance predictive accuracy while mitigating these biases [53]. Finally, the inclusion of additional factors, such as genetic predispositions, hormonal changes, and nutritional status, as highlighted by recent studies, could provide a more comprehensive assessment of sarcopenia risk and improve the model's clinical applicability [54,55].

The comparison between the full-feature model and the reduced-feature model reveals notable differences in performance metrics, indicating the impact of feature selection on model efficiency and accuracy; refer to Table 7.

Table 7. Comparison of full-feature and reduced-feature models.

Metric	Full-Feature Model	Reduced-Feature Model
Accuracy	0.87	0.89
Precision	0.82	0.85
Recall	0.81	0.83
F1-Score	0.82	0.84
True Positives (Class 1)	33	34
True Negatives (Class 0)	141	144
False Positives (Class 0 misclassified as 1)	12	9
False Negatives (Class 1 misclassified as 0)	14	13
ROC AUC Score	0.93	0.9609

With a sample size of 664 participants, our study provides a robust methodological advantage over the 210-participant study by [14]. A larger dataset enhances model stability and reduces overfitting risks, allowing for better generalization across diverse clinical populations. This greater representativeness enables our model to capture a broader spectrum of sarcopenia-related features, increasing statistical power and yielding reliable performance metrics like accuracy and recall. These strengths make our model a more robust and clinically applicable tool for early sarcopenia detection, supporting its practical utility in real-world healthcare settings.

The comprehensive model reached an accuracy of 0.87, demonstrating a precision of 0.82, a recall of 0.81, and an F1-score of 0.82. The confusion matrix indicates that the model accurately identified 141 instances of class 0 and 33 instances of class 1 but incorrectly

labeled 12 instances as false positives and 14 as false negatives. The ROC curve for the full-feature model yielded an AUC of 0.93, reflecting strong but not optimal discriminative power in distinguishing between the two classes. These results demonstrate that the full set of features allows the model to perform with reasonable accuracy and balance between precision and recall.

In contrast, the reduced-feature model exhibited an improvement in accuracy, reaching 0.89, with precision of 0.85, recall of 0.83, and an F1-score of 0.84. The confusion matrix for this model indicates a slight improvement, with fewer false positives (9) and false negatives (13), suggesting a better classification performance overall. Additionally, the ROC curve for the reduced-feature model displayed a higher AUC of 0.9609, indicating superior discriminatory capability compared with the full-feature model.

The increased accuracy and AUC in the reduced feature model suggest that the selected subset of features not only retains critical predictive information but also enhances the model's performance by removing redundant or irrelevant features. This result implies that feature selection contributes to both computational efficiency and improved classification performance. Additionally, the higher precision and recall in the reduced model indicate a better balance in correctly identifying both positive and negative classes, reducing the number of misclassified instances.

The contributions in Table 8 were calculated using the LIME framework. For each prediction instance, LIME assigns contribution values to features based on their influence on the local prediction. These instance-level contributions were aggregated and normalized across all instances to compute the overall contribution values, which are used to rank the features.

The ranking of features, as shown in Table 8, provides insight into the relative importance of each variable in contributing to the predictive performance of the model. BMR_kcal ranks the highest, with a contribution of 0.86, indicating that basal metabolic rate plays a crucial role in influencing the model's predictions. This suggests that metabolic factors are key determinants in the classification process, likely reflecting their strong association with the target outcome.

Following this, FVC with a contribution of 0.72 and D2_s with 0.67 are also among the top contributors. These features, which measure respiratory function and locomotion, respectively, highlight the importance of physical capacity and mobility in the model's predictive power. The features SS (Stand and Sit Test) and MWT2 (2-minute walking test) also rank highly, further supporting the importance of physical performance metrics in the model's decision-making.

Table 8. Ranking of selected features based on contribution.

Rank	Feature	Contribution
1	BMR_kcal	0.86
2	FVC	0.72
3	D2_s	0.67
4	SS	0.64
5	MWT2	0.57
6	ND_SLS	0.54
7	G_TUG	0.25
8	Percent_BF	0.18

6. Conclusions

Specifically, the study successfully developed a machine learning model using the CatBoost algorithm to predict sarcopenia, employing both full and reduced feature sets. Using the LIME method, the model identified key features that significantly contribute to the prediction of sarcopenia, such as basal metabolic rate (BMR), Forced Vital Capacity (FVC), and various physical performance tests (e.g., 2 min walking test, Stand and Sit Test). These features provide a clearer understanding of the condition's risk factors.

In summary, the reduced-feature model outperformed the full-feature model, with an accuracy of 0.89 versus 0.87. This indicates that the selected features retained critical information while reducing complexity. The reduced model also achieved a higher AUC score (0.94), indicating superior discriminative power.

Furthermore, the application of SMOTE-Tomek balancing improved the handling of class imbalance, enhancing the model's generalization. The research highlights that feature selection not only optimizes model performance but also facilitates easier interpretation, making it potentially valuable for clinical applications in identifying high-risk patients with sarcopenia.

The study contributes to the existing literature by demonstrating the effectiveness of machine learning and feature selection techniques in medical data analysis, emphasizing the significance of physiological and functional assessments in sarcopenia prediction. Future research should explore the generalizability of these findings across different populations and the integration of additional clinical variables.

Author Contributions: Conceptualization, A.T. and D.T.; methodology, D.T.; software, A.T.; validation, A.T., D.T. and W.K.; formal analysis, D.T.; investigation, A.T.; resources, D.T.; data curation, J.K. and W.K.; writing—original draft preparation, A.T.; writing—review and editing, J.K. and W.K.; visualization, A.T.; supervision, W.K.; project administration, J.K. and W.K.; funding acquisition, W.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Gachon University Research Fund (GCU-202300740001), the Ministry of Education of the Republic of Korea, and the National Research Foundation of Korea (NRF-2022S1A5C2A07090938).

Informed Consent Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Gachon University Institutional Bioethics Committee (approval no. 1044396-202301-HR-020-01). All participants provided informed consent.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We would like to thank Minje Seok for his assistance in this research.

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

Appendix A

Table A1. List of features with descriptions for sarcopenia dataset.

ID	Feature	Description
1	sarcopenia_2	Sarcopenia indicator
2	Place	Location where data were collected
3	sex	Gender of the subject
4	Age	Age of the subject in years
5	Percent_BF	Percent Body Fat
6	BFM	Body Fat Mass

Table A1. *Cont.*

ID	Feature	Description
7	SMM	Skeletal Muscle Mass
8	BMI	Body Mass Index
9	BMR	Basal Metabolic Rate
10	SBP	Systolic Blood Pressure
11	DBP	Diastolic Blood Pressure
12	BP_stage	Blood Pressure Stage
13	ASM	Appendicular Skeletal Muscle
14	CC	Calf Circumference
15	HG_R_1	Hand grip strength right (1st trial)
16	HG_R_2	Hand grip strength right (2nd trial)
17	HG_L_1	Hand grip strength left (1st trial)
18	HG_L_2	Hand grip strength left (2nd trial)
19	HG_R_M	Maximum hand grip strength, right
20	HG_L_M	Maximum hand grip strength, left
21	D_HG	Dominant hand grip strength
22	ND_HG	Non-dominant hand grip strength
23	Plantar_R_1	Plantar flexion strength, right foot
24	Plantar_L_1	Plantar flexion strength, left foot
25	Dorsal_R_1	Dorsal flexion strength, right foot
26	Dorsal_L_1	Dorsal flexion strength, left foot
27	D_Plantar	Dominant plantar flexion strength
28	D_Dorsal	Dominant dorsal flexion strength
29	ND_Plantar	Non-dominant plantar flexion strength
30	ND_Dorsal	Non-dominant dorsal flexion strength
31	Plantar_MAX	Plantar flexion maximum strength
32	Dorsal_MAX	Dorsal flexion maximum strength
33	SLS_R	Single leg stance test, right leg
34	SLS_L	Single leg stance test, left leg
35	D_SLS	Dominant leg single-leg stance test
36	ND_SLS	Non-dominant leg single-leg stance test
37	SLS_MAX	Maximum single-leg stance test
38	SS	Stand and Sit count (30-second repetitions)
39	SS_SPPB	Stand and Sit time for five repetitions (SPPB)
40	CSR	Chair sit and reach test
41	2MWT	2-minute walking test
42	TUG	Timed up and go test
43	Gait Speed_SPPB	Gait speed for 4-meter walk
44	SPPB	Short Physical Performance Battery score
45	SarQoL_Total_a	Total SarQoL score (absolute value)
46	SarQoL_Total_s	Total SarQoL score (adjusted)
47	D1	Physical and Mental Health domain
48	D2	Locomotion domain
49	D3	Body Composition domain
50	D4	Functionality domain
51	D5	Activities of daily living domain
52	D6	Leisure activities domain
53	D7	Fears domain
54	FES	Fall efficacy scale
55	SARC_F	Sarcopenia self-diagnosis questionnaire
56	SARC_Calf	Sarcopenia questionnaire + calf circumference
57	DM	Diabetes Mellitus
58	Area_city	Residential area (city/province)
59	Area_Dong	Residential area (neighborhood/township)
60	Education level	Participant's education level

Table A1. Cont.

ID	Feature	Description
61	Religion	Religion
62	Smoking	Smoking status
63	Smoking_d	Smoking duration
64	Smoking_a	Smoking amount
65	Drinking_f	Drinking frequency
66	Drinking_d	Drinking duration
67	Drinking_a	Drinking amount
68	Family	Family type
69	House	Housing type
70	Income	Monthly income
71	Education level_p	Education level of parents/guardian
72	Regular PA	Regular physical activity participation
73	Type of PA	Type of physical activity
74	SAF	Skin autofluorescence level (e.g., measure of tissue glycation)
75	HbA1c	Hemoglobin A1c level (average blood sugar)
76	Hypertension	Hypertension status

References

1. Cruz-Jentoft, A.J.; Sayer, A.A. Sarcopenia. *Lancet* **2019**, *393*, 2636–2646. [\[CrossRef\]](#)
2. Bilski, J.; Pierzchalski, P.; Szczepanik, M.; Bonior, J.; Zoladz, J.A. Multifactorial mechanism of sarcopenia and sarcopenic obesity. Role of physical exercise, microbiota and myokines. *Cells* **2022**, *11*, 160. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Tezze, C.; Sandri, M.; Tessari, P. Anabolic resistance in the pathogenesis of sarcopenia in the elderly: Role of nutrition and exercise in young and old people. *Nutrients* **2023**, *15*, 4073. [\[CrossRef\]](#)
4. Turimov Mustapoevich, D.; Kim, W. Machine learning applications in sarcopenia detection and management: A comprehensive survey. *Healthcare* **2023**, *11*, 2483. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Shin, J. Feasibility of local interpretable model-agnostic explanations (LIME) algorithm as an effective and interpretable feature selection method: Comparative fNIRS study. *Biomed. Eng. Lett.* **2023**, *13*, 689–703. [\[CrossRef\]](#)
6. Zafar, M.R.; Khan, N. Deterministic local interpretable model-agnostic explanations for stable explainability. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 525–541. [\[CrossRef\]](#)
7. Kumarakulasingham, N.B.; Blomberg, T.; Liu, J.; Leao, A.S.; Papapetrou, P. Evaluating local interpretable model-agnostic explanations on clinical machine learning classification models. In Proceedings of the 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Rochester, MN, USA, 28–30 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 7–12.
8. Olshvang, D.; Harris, C.; Chellappa, R.; Santhanam, P. Predictive modeling of lean body mass, appendicular lean mass, and appendicular skeletal muscle mass using machine learning techniques: A comprehensive analysis utilizing NHANES data and the Look AHEAD study. *PLoS ONE* **2024**, *19*, e0309830. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Wang, J.H.; Liu, C.Y.; Min, Y.R.; Wu, Z.H.; Hou, P.L. Cancer diagnosis by gene-environment interactions via combination of SMOTE-Tomek and overlapped group screening approaches with application to imbalanced TCGA clinical and genomic data. *Mathematics* **2024**, *12*, 2209. [\[CrossRef\]](#)
10. Zeng, M.; Zou, B.; Wei, F.; Liu, X.; Wang, L. Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data. In Proceedings of the 2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS), Chongqing, China, 28–29 May 2016; pp. 225–228. [\[CrossRef\]](#)
11. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*.
12. Aljadani, A.; Alharthi, B.; Farsi, M.A.; Balaha, H.M.; Badawy, M.; Elhosseini, M.A. Mathematical modeling and analysis of credit scoring using the lime explainer: A comprehensive approach. *Mathematics* **2023**, *11*, 4055. [\[CrossRef\]](#)
13. Zhang, H.; Yin, M.; Liu, Q.; Ding, F.; Hou, L.; Deng, Y.; Cui, T.; Han, Y.; Pang, W.; Ye, W.; et al. Machine and deep learning-based clinical characteristics and laboratory markers for the prediction of sarcopenia. *Chin. Med. J.* **2023**, *136*, 967–973. [\[CrossRef\]](#)
14. Ozgur, S.; Altinok, Y.A.; Bozkurt, D.; Saraç, Z.; Akçiçek, S. Performance Evaluation of Machine Learning Algorithms for Sarcopenia Diagnosis in Older Adults. *Healthcare* **2023**, *11*, 2699. [\[CrossRef\]](#)
15. Lee, J.; Choi, Y.; Yoon, Y. Genetic Algorithm-based Feature Selection for Machine Learning System Diagnosing Sarcopenia with Survey Data. In Proceedings of the Companion Conference on Genetic and Evolutionary Computation, Lisbon, Portugal, 15–19 July 2023. [\[CrossRef\]](#)

16. Ko, J.; Kim, K.B.; Shin, Y.; Han, H.; Han, S.K.; Jung, D.Y.; Hong, J. Predicting Sarcopenia of Female Elderly from Physical Activity Performance Measurement Using Machine Learning Classifiers. *Clin. Interv. Aging* **2021**, *16*, 1723–1733. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Blanc-Durand, P.; Schiratti, J.B.; Schutte, K.; Jehanno, P.; Hérent, P.; Pigneur, F.; Lucidarme, O.; Benaceur, Y.; Sadate, A.; Luciani, A.; et al. Abdominal musculature segmentation and surface prediction from CT using deep learning for sarcopenia assessment. *Diagn. Interv. Imaging* **2020**, *101*, 789–794. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Seok, M.; Kim, W. Sarcopenia Prediction for Elderly People Using Machine Learning: A Case Study on Physical Activity. *Healthcare* **2023**, *11*, 1334. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Seok, M.; Kim, W.; Kim, J. Machine learning for sarcopenia prediction in the elderly using socioeconomic, infrastructure, and quality-of-life data. *Healthcare* **2023**, *11*, 2881. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Naser, M.; Wahid, K.; Grossberg, A.; Olson, B.; Jain, R.; El-Habashy, D.; Dede, C.; Salama, V.; Abobakr, M.; Mohamed, A.; et al. Deep learning auto-segmentation of cervical skeletal muscle for sarcopenia analysis in patients with head and neck cancer. *Front. Oncol.* **2022**, *12*, 930432. [\[CrossRef\]](#)
21. Ahmed, Z.; Das, S. A Novel Hybrid Resampling Approach to Address Class-Imbalanced Issues. *SN Comput. Sci.* **2024**, *5*, 865. [\[CrossRef\]](#)
22. Swana, E.F.; Doorsamy, W.; Bokoro, P. Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset. *Sensors* **2022**, *22*, 3246. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Abdullahi, D.S.; Aliyu, D.M.S.; Abdullahi, U.M. Comparative analysis of resampling algorithms in the prediction of stroke diseases. *UMYU Sci.* **2023**, *2*, 88–94. [\[CrossRef\]](#)
24. Mukherjee, M.; Khushi, M. SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features. *arXiv* **2021**, arXiv:2103.07612. [\[CrossRef\]](#)
25. Baumann, J.; Castelnovo, A.; Crupi, R.; Inverardi, N.; Regoli, D. Bias on Demand: A Modelling Framework That Generates Synthetic Data With Bias. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023. [\[CrossRef\]](#)
26. ben Jabeur, S.; Gharib, C.; Mefteh-Wali, S.; Arfi, W.B. CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technol. Forecast. Soc. Change* **2021**, *166*, 120658. [\[CrossRef\]](#)
27. Wang, H.; Cheng, L. CatBoost model with synthetic features in application to loan risk assessment of small businesses. *arXiv* **2021**, arXiv:2106.07954.
28. Yang, R.; Wang, P.; Qi, J. A novel SSA-CatBoost machine learning model for credit rating. *J. Intell. Fuzzy Syst.* **2022**, *44*, 2269–2284. [\[CrossRef\]](#)
29. Hussain, S.; Mustafa, M.; Jumani, T.; Baloch, S.K.; Alotaibi, H.; Khan, I.; Khan, A. A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection. *Energy Rep.* **2021**, *7*, 4425–4436. [\[CrossRef\]](#)
30. Sysoev, A. On Factors Selection in CatBoost Models Construction. In Proceedings of the 2023 5th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russia, 8–10 November 2023; pp. 338–340. [\[CrossRef\]](#)
31. Garreau, D.; Luxburg, U.V. Explaining the Explainer: A First Theoretical Analysis of LIME. *arXiv* **2020**, arXiv:2001.03447.
32. Shi, S.; Du, Y.; Fan, W. An Extension of LIME with Improvement of Interpretability and Fidelity. *arXiv* **2020**, arXiv:2004.12277.
33. Shawi, R.E.; Mohamed, Y.; Al-Mallah, M.; Sakr, S. Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Comput. Intell.* **2020**, *37*, 1633–1650. [\[CrossRef\]](#)
34. Visani, G.; Bagli, E.; Chesani, F.; Poluzzi, A.; Capuzzo, D. Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models. *J. Oper. Res. Soc.* **2022**, *73*, 91–101. [\[CrossRef\]](#)
35. Cabello-Solorzano, K.; Ortigosa de Araujo, I.; Peña, M.; Correia, L.; Tallón-Ballesteros, A.J. The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis. In Proceedings of the International Conference on Soft Computing Models in Industrial and Environmental Applications, Salamanca, Spain, 5–7 September 2023; Springer: Cham, Switzerland, 2023; pp. 344–353.
36. Dieber, J.; Kirrane, S. Why model why? Assessing the strengths and limitations of LIME. *arXiv* **2020**, arXiv:2012.00093.
37. Wu, Y.; Zhang, L.; Bhatti, U.A.; Huang, M. Interpretable machine learning for personalized medical recommendations: A LIME-based approach. *Diagnostics* **2023**, *13*, 2681. [\[CrossRef\]](#)
38. Ribeiro, M.T.; Singh, S.; Guestrin, C. “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, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
39. You, W.; Yang, Z.; Ji, G. Feature selection for high-dimensional multi-category data using PLS-based local recursive feature elimination. *Expert Syst. Appl.* **2014**, *41*, 1463–1475. [\[CrossRef\]](#)
40. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [\[CrossRef\]](#) [\[PubMed\]](#)

41. Wang, L.; Wang, Y.; Chang, Q. Feature selection methods for big data bioinformatics: A survey from the search perspective. *Methods* **2016**, *111*, 21–31. [[CrossRef](#)] [[PubMed](#)]
42. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [[CrossRef](#)]
43. 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](#)]
44. Cerone, G.; Botter, A.; Gazzoni, M. A Modular, Smart, and Wearable System for High Density sEMG Detection. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 3371–3380. [[CrossRef](#)]
45. Merletti, R.; Cerone, G. Tutorial. Surface EMG detection, conditioning and pre-processing: Best practices. *J. Electromyogr. Kinesiol.* **2020**, *54*, 102440. [[CrossRef](#)] [[PubMed](#)]
46. Junquera-Godoy, I.; Gomis-Tena, J.; Saiz, J.; de Juan, J.L.M.; Carot-Sierra, J.; Prats-Boluda, G. Evaluation of Surface Electromyography Parameters as Possible Biomarkers for the Assessment of Sarcopenia. In Proceedings of the 2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE), Virtual, 4–6 December 2023; pp. 285–291. [[CrossRef](#)]
47. Aggio, D.A.; Sartini, C.; Papacosta, O.; Lennon, L.; Ash, S.; Whincup, P.; Wannamethee, S.; Jefferis, B. Cross-sectional associations of objectively measured physical activity and sedentary time with sarcopenia and sarcopenic obesity in older men. *Prev. Med.* **2016**, *91*, 264–272. [[CrossRef](#)] [[PubMed](#)]
48. Bouchard, D.; Dionne, I.; Brochu, M. Sarcopenic/Obesity and Physical Capacity in Older Men and Women: Data From the Nutrition as a Determinant of Successful Aging (NuAge)—the Quebec Longitudinal Study. *Obesity* **2009**, *17*, 2082–2088. [[CrossRef](#)]
49. Visani, G.; Bagli, E.; Chesani, F. OptiLIME: Optimized LIME Explanations for Diagnostic Computer Algorithms. *arXiv* **2020**, arXiv:2006.05714.
50. Zafar, M.R.; Khan, N. DLIME: A Deterministic Local Interpretable Model-Agnostic Explanations Approach for Computer-Aided Diagnosis Systems. *arXiv* **2019**, arXiv:1906.10263.
51. An, J.; Zhang, Y.; Joe, I. Specific-Input LIME Explanations for Tabular Data Based on Deep Learning Models. *Appl. Sci.* **2023**, *13*, 8782. [[CrossRef](#)]
52. Huang, M.W.; Chiu, C.H.; Tsai, C.F.; Lin, W.C. On Combining Feature Selection and Over-Sampling Techniques for Breast Cancer Prediction. *Appl. Sci.* **2021**, *11*, 6574. [[CrossRef](#)]
53. Lee, J.; Yoon, Y.; Kim, J.; Kim, Y.H. Metaheuristic-Based Feature Selection Methods for Diagnosing Sarcopenia with Machine Learning Algorithms. *Biomimetics* **2024**, *9*, 179. [[CrossRef](#)]
54. Cui, M.; Gang, X.; Gao, F.; Wang, G.; Xiao, X.; Li, Z.; Li, X.; Ning, G.; Wang, G. Risk Assessment of Sarcopenia in Patients with Type 2 Diabetes Mellitus Using Data Mining Methods. *Front. Endocrinol.* **2020**, *11*, 123. [[CrossRef](#)] [[PubMed](#)]
55. Zupo, R.; Moroni, A.; Castellana, F.; Gasparri, C.; Catino, F.; Lampignano, L.; Perna, S.; Clodoveo, M.L.; Sardone, R.; Rondanelli, M. A Machine-Learning Approach to Target Clinical and Biological Features Associated with Sarcopenia: Findings from Northern and Southern Italian Aging Populations. *Metabolites* **2023**, *13*, 565. [[CrossRef](#)]

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.