mathematics

MDPI

Article

Enhancing Sarcopenia Prediction Through an Ensemble Learning Approach: Addressing Class Imbalance for Improved Clinical Diagnosis

Dilmurod Turimov and Wooseong Kim *

Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 461-701, Gyeonggi-do, Republic of Korea; dilmurod@gachon.ac.kr
* Correspondence: wooseong@gachon.ac.kr

Abstract: This study developed an advanced ensemble learning model aimed to improve the accuracy of predicting sarcopenia, a condition characterized by a gradual decline in muscle mass and strength, leading to increased disability and mortality. The study focused on enhancing model performance by combining various machine learning methods and addressing critical challenges, such as class imbalance and data complexity. Several foundational models were employed, including support vector machine, random forest, neural network, logistic regression, and decision tree. To address class imbalance, the adaptive synthetic sampling method was implemented, producing synthetic samples for the minority class to achieve a more balanced dataset. The data preprocessing stage included feature scaling and feature selection processes, utilizing recursive feature elimination to refine feature selection. Subsequently, a classifier selection algorithm was employed to select models that provided an optimal balance of diversity and performance. The effectiveness of the final ensemble model was evaluated using various metrics, such as accuracy, precision, recall, F1-score, and ROC AUC. The model achieved an accuracy of 88.5%, outperforming individual machine learning models and existing methods in the literature. These findings suggest that the classifier selection algorithm effectively addresses challenges in sarcopenia prediction, particularly in the case of imbalanced data. The model's strong performance indicates its potential for use in clinical environments, where it can facilitate early diagnosis and improve intervention strategies for sarcopenia patients. This study advances the field of medical machine learning by demonstrating the utility of ensemble learning in healthcare prediction.

Keywords: sarcopenia prediction; ensemble learning; adaptive synthetic sampling; recursive feature elimination; class imbalance; voting classifier

MSC: 68T99

check for updates

Academic Editors: Florin Leon, Mircea Hulea, Marius Gavrilescu, Jie Wen and Huawen Liu

Received: 17 October 2024
Revised: 2 December 2024
Accepted: 23 December 2024
Published: 25 December 2024

Citation: Turimov, D.; Kim, W. Enhancing Sarcopenia Prediction Through an Ensemble Learning Approach: Addressing Class Imbalance for Improved Clinical Diagnosis. *Mathematics* **2025**, *13*, 26. https://doi.org/10.3390/math13010026

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 is a progressive degenerative condition characterized by a reduction in skeletal muscle mass and strength [1]. It frequently causes physical impairment, a lower quality of life, and a higher death rate. Muscle mass deterioration leads to increased vulnerability to falls, frailty, and an overall decline in physical performance, substantially affecting quality of life [2]. The prevalence of sarcopenia differs across populations and is influenced by various diagnostic criteria; however, it tends to be significantly higher in certain groups. For example, in cancer patients, sarcopenia occurring before treatment

is observed in approximately 38% of cases and is highly correlated with poor treatment outcomes, including higher postoperative complication rates and lower survival rates [3].

Given the severe consequences of sarcopenia, early detection and timely intervention are critical. Advances in technology, particularly in machine learning, have demonstrated that algorithms such as ensemble models can enhance sarcopenia prediction and management [4]. These models, which integrate multiple machine learning techniques, have demonstrated improved accuracy and predictive performance over conventional methods [5,6]. For example, studies using machine learning approaches have revealed that sarcopenic obesity (a condition of concurrent sarcopenia and obesity) is a key predictive factor for poor outcomes in patients undergoing surgery for spinal metastases, underscoring the clinical relevance of these advanced methods [7].

Incorporating machine learning into sarcopenia management has great potential for personalized medicine, enabling earlier interventions and more precise therapeutic strategies to improve patient outcomes.

This study develops and evaluates an advanced ensemble learning model specifically designed for sarcopenia prediction. This model employs multiple machine learning algorithms and implements a dynamic selection process to enhance predictive accuracy, address class imbalance, and provide a comprehensive framework for sarcopenia detection, outperforming existing approaches. Its goal is to provide a reliable predictive tool that can be utilized in clinical settings to facilitate early diagnosis and intervention strategies for sarcopenia, ultimately improving patient outcomes.

This study is motivated by recent advancements in machine learning, in particular, the demonstrated efficacy of ensemble methods in improving disease prediction. Recent studies have reported that ensemble methods, such as stacking, consistently outperform individual models in predicting various diseases, such as diabetes and cardiovascular diseases, by addressing challenges such as class imbalance and improving predictive performance [8]. However, despite these advancements, sarcopenia has not been widely studied using ensemble methods, presenting an opportunity for innovation. This study aims to bridge this gap by applying ensemble learning to sarcopenia prediction, leveraging its strengths in handling complex datasets and improving predictive accuracy, thereby setting a new standard in clinical prediction for sarcopenia and contributing to improved health outcomes.

This study makes several important contributions to the field of medical machine learning, particularly in the context of sarcopenia prediction, which are outlined below:

- **Application of ensemble learning to sarcopenia prediction**: To the best of our knowledge, this study is among the first to apply ensemble learning specifically to sarcopenia prediction, illustrating how the integration of multiple machine learning models can significantly improve predictive accuracy for this condition.
- **Classifier selection algorithm**: This study introduces a process for selecting classifiers within the ensemble framework, enhancing both accuracy and model diversity and ensuring that only the most complementary and effective models are included in the final ensemble, thereby improving overall performance.
- **Improved predictive performance**: The effectiveness of the proposed approach is illustrated using a constructed ensemble model, which outperforms individual base models and previously reported ensemble models for various diseases, achieving an accuracy rate of 88.5%.
- **Addressing class imbalance**: This study successfully addresses the challenge of class imbalance in sarcopenia prediction using the ADASYN technique, thereby enhancing the model's ability to accurately detect sarcopenia cases, which is crucial for clinical applications.

- **Comprehensive comparison with existing models**: Through detailed comparative analysis with recent studies employing ensemble methods for other diseases, this study highlights the advantages of the proposed approach, setting a new benchmark for predictive accuracy in sarcopenia detection.

2. Related Work

Sarcopenia has been linked to increased frailty, disability, surgical complications, and mortality. Studies have confirmed its prevalence in cancer patients, where it serves as a critical predictor of adverse clinical outcomes. For example, in patients with pancreatic ductal adenocarcinoma, sarcopenia has been associated with postoperative complications, such as the development of pancreatic fistulas [9]. Similarly, in patients with non-small cell lung cancer, preoperative sarcopenia is a significant risk factor for severe complications [10]. In gastrointestinal surgeries, sarcopenia is correlated with lower survival rates and extended hospital stays [11,12].

Recent research suggests that wearable technology, such as activity trackers and smartwatches, can play a critical role in improving the detection and management of sarcopenia through real-time monitoring of disease progression. These innovations are particularly important for addressing the current challenges in data accessibility and the imbalance in sarcopenia datasets. However, the need for standardized diagnostic criteria remains. Emerging technologies such as blockchain and edge computing can help address these limitations by improving data security and privacy [13].

Traditional methods for detecting sarcopenia typically rely on physical performance assessments and imaging techniques, such as computed tomography and magnetic resonance imaging, to quantify muscle mass. Indices such as the skeletal muscle index and psoas muscle index are commonly used, although they primarily assess muscle quantity rather than function. These metrics are not always reliable in predicting clinical outcomes [14], and some studies have reported inconsistent relationships between imaging results and survival rates [15]. Moreover, the lack of standardized diagnostic criteria across studies complicates the direct comparison of results [16].

Machine learning has transformed healthcare by enabling earlier and more accurate disease detection, frequently outperforming traditional diagnostic methods. Machine learning algorithms have demonstrated remarkable success in predicting diseases such as diabetes and cardiovascular conditions [17,18]. For heart disease, machine learning models are increasingly applied to extract features from large datasets, resulting in earlier and more precise diagnoses [19].

Ensemble learning, which combines multiple machine learning models, is frequently employed to improve the accuracy of disease prediction. Techniques such as boosting, bagging, and stacking have been extensively utilized to predict cardiovascular diseases, diabetes, and sarcopenia. Studies have consistently demonstrated that ensemble methods outperform individual models. For example, stacking exhibits superior predictive performance for diseases such as diabetes and skin disorders [8]. Models such as random forest and gradient boosting are also widely used to improve the accuracy of heart disease and diabetes prediction [20,21].

A key challenge in applying machine learning to sarcopenia prediction is the lack of comprehensive studies in this area. Current research mainly focuses on traditional diagnostic approaches, not fully utilizing the potential of machine learning. Moreover, sarcopenia datasets are complex and imbalanced, reducing predictive accuracy. More advanced machine learning techniques, such as ensemble learning, can effectively address these challenges [22]. The limited research on machine learning applications specific to sarcopenia prediction highlights the need to fill this gap in the literature [23].

Dynamic model selection in ensemble learning provides a continuous evaluation of model performance, enhancing both diversity and robustness. To overcome class imbalance in illness prediction models, methods such as synthetic minority oversampling (SMOTE) and adaptive synthetic sampling (ADASYN) are frequently employed. For example, ADASYN has been used to generate synthetic samples to improve the detection of minority classes in cardiovascular disease and diabetes datasets [24]. These methods can be similarly effective for sarcopenia datasets, which are often imbalanced, to improve predictive outcomes.

Ensemble models such as random forest, AdaBoost, and gradient boosting have consistently outperformed traditional models in predicting heart disease and diabetes, providing higher accuracy and reliability [20,25]. Studies comparing these models indicate that stacking is particularly effective in managing complex datasets, offering notable improvements over individual models. This method can provide similar advantages in sarcopenia prediction, where conventional models often struggle with dataset complexity.

The taxonomy of ensemble learning methods illustrates their effectiveness in disease prediction. Bagging and boosting methods, such as those for heart disease and chronic kidney disease, enhance model stability and accuracy, while stacking improves predictive power for complex conditions such as cardiovascular disease. Voting ensembles enhance diagnostic robustness for various diseases. In this study, the ensemble model for sarcopenia aligns with these methods by employing a voting ensemble combined with feature selection and class balancing to optimize sarcopenia prediction (Table 1).

Table 1. Taxonomy of related works by ensemble learning techniques.

Ensemble Technique	Study	Disease Type	Ensemble Approach	Main Contribution
Bagging	[20]	Heart disease	Bagging with random forest	Improved accuracy of heart disease prediction by combining bagging with optimization techniques.
	[26]	Chronic kidney disease	Comparison of bagging and basic classifiers	Compared ensemble and basic classifiers, highlighting the advantages of bagging in chronic kidney disease prediction.
	[27]	Kidney stone type	Random forest and boosting techniques	Developed a method for predicting kidney stone type, using bagging and boosting for improved accuracy.
Boosting	[28]	Chronic kidney disease	Gradient boosting and feature ranking	Used boosting and feature ranking to predict chronic kidney disease, emphasizing cost-effective solutions.
	[25]	Heart disease	Boosting and voting ensemble	Developed an ensemble architecture combining boosting and voting for heart disease prediction.
Stacking	[21]	Cardiovascular disease	Stacking with XGBoost and random forest	Achieved improved cardiovascular disease prediction using a stacking model combining multiple algorithms.
	[29]	Skin disease	Hybrid feature selection with stacking	Proposed a hybrid feature selection technique with stacking ensemble for skin disease prediction.
Voting	[24]	Heart disease	Voting ensemble	Developed a voting ensemble for early heart disease prediction using multiple base classifiers.
	[30]	Erythemato-squamous disease	Multi-feature voting ensemble	Employed a voting ensemble with multi-feature selection to improve erythemato-squamous disease prediction.
	This study	Sarcopenia	Voting ensemble	Developed an optimized ensemble learning model that improves sarcopenia prediction by addressing class imbalance and enhancing feature selection.

3. Materials and Methods

The ensemble model for sarcopenia prediction was developed and evaluated following a structured and systematic methodology, as illustrated in Figure 1.

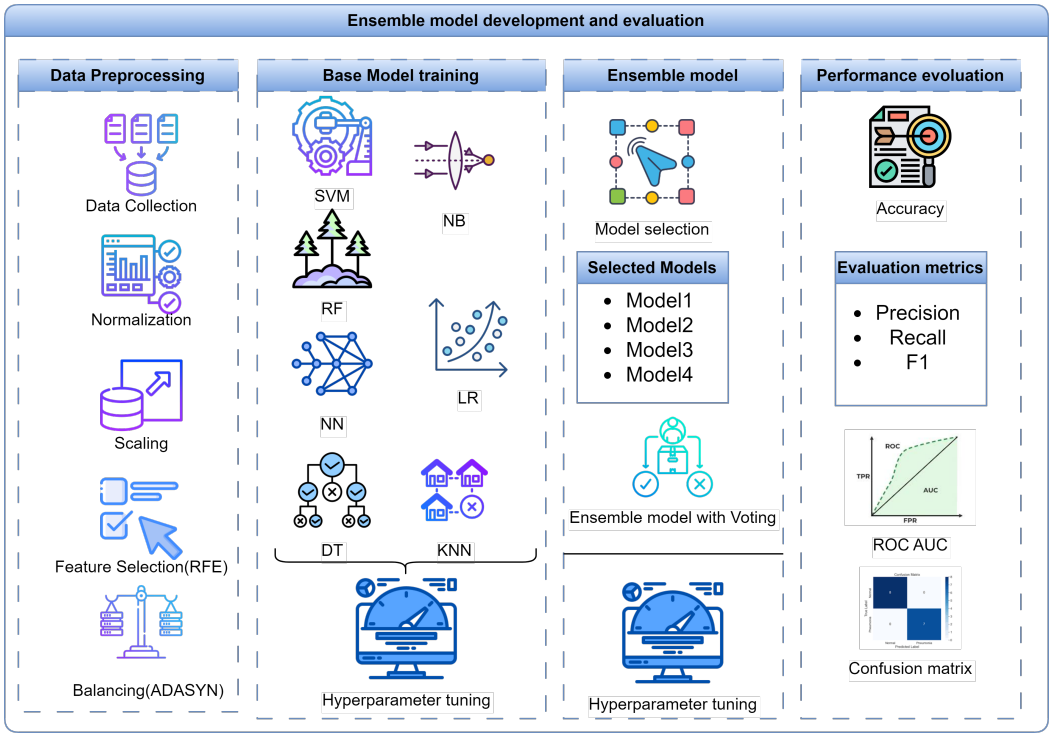


Figure 1. Overview of the ensemble model development and evaluation process. The diagram shows key stages, including data preprocessing, base model training, hyperparameter tuning, ensemble model selection, and performance evaluation using metrics like accuracy, ROC AUC, and confusion matrix.

The process is described in this section in the follows steps:

1. **Data collection**: Data were collected from multiple community sites over nine months, ensuring a diverse and representative sample across different locations.
2. **Rationale for variable collection**: The dataset included 76 variables selected based on their relevance to sarcopenia prediction. The variables spanned multiple categories, including demographic, body composition, vital signs, balance, mobility, lifestyle, and socioeconomic indicators.
3. **Data preprocessing**:
   - **Data splitting**: The dataset was divided into training, validation, and test sets using stratified sampling to maintain the distribution of the target variable across subsets.
   - **Feature scaling**: Features were scaled using normalization (MinMaxScaler) and standardization (StandardScaler) to ensure compatibility across different machine learning models.
   - **Feature selection**: Recursive feature elimination (RFE) was used to prioritize influential variables for prediction by iteratively removing less significant features.
   - **Handling class imbalance**: The ADASYN algorithm was employed to balance the dataset by synthesizing samples for the minority class, thereby reducing bias and enhancing model sensitivity.
4. **Base models**: Multiple machine learning algorithms were trained and optimized, each selected for its unique predictive capabilities. These algorithms included sup-

port vector machine (SVM), random forest, neural network, decision tree, logistic regression, k-nearest neighbors, and naive Bayes.

5. **Ensemble model development**:
    - **Diversity calculation**: Pairwise agreement between classifiers was calculated to ensure diversity within the ensemble.
    - **Classifier selection process**: A classifier selection algorithm was implemented to optimize ensemble performance by selecting complementary models.
6. **Voting classifier**: The ensemble was constructed using a voting classifier with soft voting, combining the probabilistic outputs of the selected base models.
7. **Model evaluation**: The final ensemble was evaluated on a separate test set using the following metrics: accuracy, precision, recall, F1-score, ROC AUC, and confusion matrix. These metrics assess the overall predictive performance of the model.
8. **Computational environment**: All experiments were conducted in the Kaggle cloud environment for consistency, with the code accessible on GitHub.

3.1. Data Collection

The dataset was acquired from the Institute of Human Convergence Health Science at Gachon University. Data collection was conducted in various community settings across Incheon, such as social welfare centers, daycare centers, and senior welfare centers. This process spanned nine months, from 1 September 2022 to 31 May 2023, resulting in a comprehensive and diverse sample across different locations. The extended duration of data collection ensured a representative dataset of the study population. All participants provided informed consent prior to their involvement in the study.

The dataset contained detailed health, physical, and lifestyle information for 664 participants, covering a wide range of variables. Standardized data collection protocols were employed to ensure precision and consistency of the collected information.

3.2. Rationale Behind Variable Collection

The dataset used in this study comprised 76 variables, each selected based on its relevance to the prediction and analysis of health outcomes, particularly for sarcopenia and related conditions. This section provides an overview of the key variables and the rationale behind their inclusion, grouped by functional and contextual significance:

3.2.1. Target Variable—Sarcopenia Classification (Sarcopenia_2)

This binary variable indicates the presence or absence of sarcopenia, a condition characterized by a loss of skeletal muscle mass and function. This variable is central to this study, whose focus is sarcopenia prediction.

3.2.2. Demographic Variables

- **Age**: Aging is a primary factor associated with sarcopenia, as muscle deterioration typically increases with age.
- **Sex**: There are well-documented sex-specific differences in muscle mass and function, making it a relevant factor.
- **Place**: Place or locality may capture environmental or lifestyle variations affecting sarcopenia risk.

3.2.3. Body Composition and Vital Signs

- **Percent body fat (Percent_BF)**: High body fat can impact muscle function and mobility, increasing sarcopenia risk.

- **Basal metabolic rate (BMR_kcal)**: The basal metabolic rate reflects overall metabolic health, correlating with muscle health and functionality.
- **Blood pressure (SBP, DBP, BP Stage)**: Blood pressure measurements are relevant due to the relationship between cardiovascular health and muscle maintenance, particularly in older populations.

3.2.4. Balance and Posture Metrics

- **Balance metrics (plantar, dorsal measurements)**: Collected from both dominant and non-dominant sides, these metrics assess postural stability, an important indicator of physical functioning that deteriorates in sarcopenia.
- **Single-leg stance (SLS_R, SLS_L, D_SLS, ND_SLS)**: The ability to balance on one leg is a critical measure of lower-body strength and balance, providing insights into muscle functionality and fall risk.
- **Composite stability ratio (CSR)**: CSR is a comprehensive measure of stability that reflects overall body balance and control, both of which are directly impacted by muscle weakness.

3.2.5. Mobility and Gait Metrics

- **Gait speed and mobility (MWT2, TUG, Gaitspeed_SPPB)**: These metrics assess walking speed and agility. Slow gait speed is a common predictor of sarcopenia and other mobility-related impairments.
- **Timed up and go (TUG)**: TUG is a practical assessment of mobility that correlates with muscle strength and coordination, which are essential for daily activities.

3.2.6. Self-Reported Quality of Life and Physical Assessment

- **SARC-F**: This questionnaire-based score assesses functional limitations due to sarcopenia, providing a patient-reported outcome measure.
- **SarQoL**: The sarcopenia quality of life (SarQoL) scale evaluates how sarcopenia affects patients' perceived quality of life, addressing factors such as physical discomfort, mobility, and daily activities.

3.2.7. Lifestyle and Socioeconomic Variables

- **Smoking and drinking habits**: These lifestyle factors directly impact muscle health, as excessive smoking and drinking can impair muscle function.
- **Socioeconomic indicators (education, income)**: These indicators may affect access to healthcare as well as lifestyle choices, indirectly affecting sarcopenia risk.
- **Physical activity types (TypeofPA_1 to TypeofPA_6)**: These variables represent different types of activity participation. Regular physical activity, particularly strength-based exercise, is a protective factor against sarcopenia.

3.2.8. Pulmonary and Biometric Indicators

- **Forced vital capacity (FVC)** and **forced expiratory volume (FEV1)**: These pulmonary function measures reflect respiratory health, which is correlated with physical endurance and muscle performance.
- **Mean inspiratory pressure (MIP_Ave)**: Mean inspiratory pressure assesses respiratory muscle strength, which may deteriorate with skeletal muscle in sarcopenia.
- **Glycemic control (HbA1c)**: Elevated HbA1c indicates poor blood sugar control, a risk factor for muscle degeneration over time.

3.2.9. Chronic Conditions and Sleep Quality

- **Diabetes (DM)**, **hypertension**, **hyperlipidemia**: These chronic conditions are associated with accelerated muscle loss and mobility impairments, impacting sarcopenia risk.
- **Sleep disorder**: Poor sleep quality can impact muscle recovery and overall physical health, making it a relevant factor in assessing sarcopenia risk.

Each variable was selected to provide a comprehensive view of the factors influencing sarcopenia. These variables span biological, functional, lifestyle, and socioeconomic dimensions, providing a holistic foundation for accurate sarcopenia prediction and analysis.

3.3. Data Preprocessing

3.3.1. Data Splitting

The dataset was divided into three subsets for training, validation, and testing to improve model robustness and reduce the risk of overfitting. Using a stratified sample technique, the dataset was first divided into a 70% training set and a 30% test set. This method ensured that the distribution of the target variable remained consistent across the training and test sets. Subsequently, the training data were further divided into training (80%) and validation (20%) subsets. This split allowed the model to be trained on one set, fine-tuned on another set, and evaluated on entirely unseen data for final testing.

3.3.2. Feature Scaling

Due to potential differences in feature scales, applying feature scaling was a critical preprocessing step. Two complementary techniques for scaling were employed:

1. **Normalization:** Each feature was normalized using MinMaxScaler, which scales each feature to a range between 0 and 1. This transformation is mathematically expressed as follows:
$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$
where x is the original feature, $\min(x)$ and $\max(x)$ represent its minimum and maximum values, respectively, and x' is the normalized feature.
2. **Standardization:** Following normalization, features were standardized using StandardScaler, which modifies features such that the standard deviation is 1 and the mean is 0. The standardization equation is expressed as follows:
$$x'' = \frac{x' - \mu}{\sigma}$$
where μ denotes the mean of the normalized feature, σ denotes the standard deviation, and x'' denotes the standardized feature.

3.3.3. Feature Selection

After data preprocessing, feature selection was performed using RFE to identify the most relevant subset of features with the largest contribution to the target prediction. RFE builds a model with features that remain after recursively eliminating the least significant ones. For a classifier $\mathcal{F}_{\text{rfe}}$ (e.g., a random forest classifier), the RFE process can be mathematically represented as follows:

$$\mathbf{x}^* = \arg\min_{|\mathbf{x}|=k} L(\mathcal{F}_{\text{rfe}}, \mathbf{x}, \mathbf{y})$$

where $\mathbf{x}^*$ is the subset of features selected, k is the desired number of features, and $L(\mathcal{F}_{\text{rfe}}, \mathbf{x}, \mathbf{y})$ is the loss function minimized during the feature elimination process.

The final set of selected features $\mathbf{x}^*$ is used to train the models in subsequent stages.

3.3.4. Handling Class Imbalance

The original dataset exhibited imbalance in the target classes, with the majority class significantly outnumbering the minority class. The ADASYN technique for imbalanced learning was utilized on the training data to address this problem. ADASYN generates synthetic examples for the minority class to balance the class distribution and improve the model's ability to learn from instances of the minority class. In formal terms, let N_{maj} and N_{min} represent the number of samples in the majority and minority classes, respectively. ADASYN synthesizes new samples for the minority class such that

$$N_{\text{new_samples}} = \gamma \times (N_{\text{maj}} - N_{\text{min}})$$

where γ is a parameter that controls the number of synthetic samples generated.

3.4. Base Models

Different supervised learning algorithms serve distinct purposes in classification and regression tasks, as illustrated in Figure 2. SVM utilizes hyperplane optimization in the feature space to delineate class boundaries, exhibiting particularly high performance in high-dimensional contexts through the use of kernel methods. Random forest, an ensemble method, aggregates the outputs of multiple decision trees, employing modal analysis for classification and mean calculation for regression. The robustness of this approach is attributed to its averaging mechanism, which reduces overfitting.

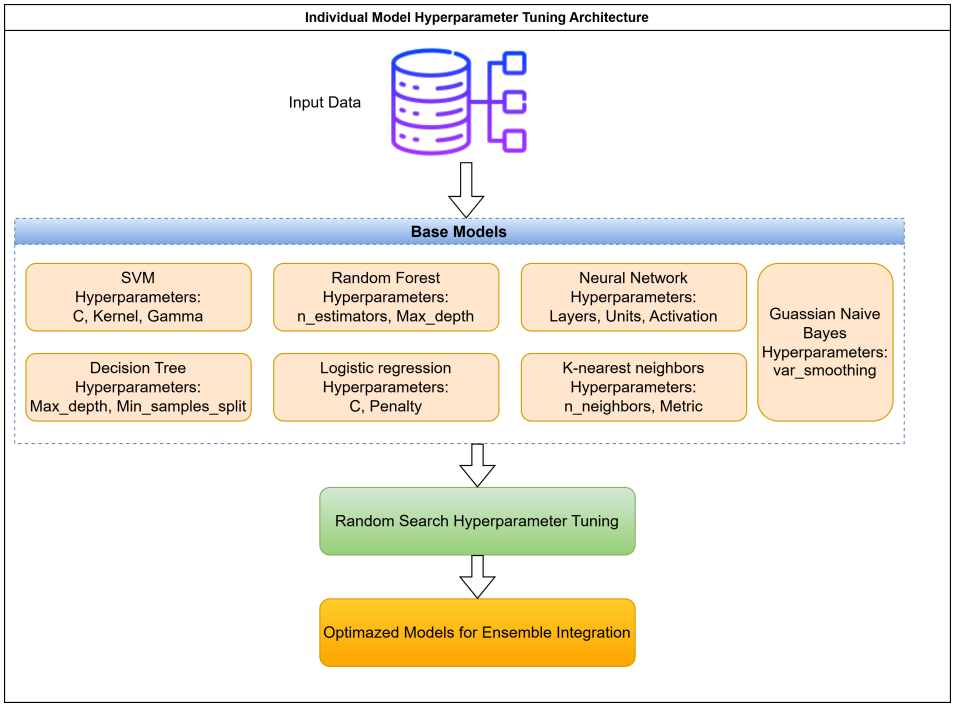


Figure 2. Individual model hyperparameter tuning architecture. The architecture shows base models with specific hyperparameters tuned via random search, resulting in optimized models for ensemble integration.

Multilayer perceptrons are artificial neural networks characterized by multiple interconnected layers of nodes, utilizing backpropagation for parameter optimization. These networks excel in modeling complex, nonlinear relationships between input and output variables. In contrast, decision trees employ a recursive partitioning method, creating

a hierarchical decision node structure that terminates in classificatory leaf nodes. This approach is inherently interpretable and accommodates diverse feature types.

The k-nearest neighbors algorithm is a non-parametric, instance-based learning method that bases classification decisions on proximity analysis within the feature space. Its optimal application domain typically involves lower-dimensional datasets. Naive Bayes classifiers leverage probabilistic principles, specifically Bayes' theorem, operating under the assumption of feature independence. Despite this simplifying assumption, they demonstrate high computational efficiency and effectiveness in specific domains, such as textual analysis.

Logistic regression, though fundamentally a linear classifier, applies a logistic transformation to feature combinations, enabling the estimation of probabilistic class membership. Its utility is attributed to its interpretability and computational simplicity, particularly in binary classification tasks.

By combining these models, the ensemble leverages various methodologies—linear, nonlinear, instance-based, and rule-based—thus ensuring a comprehensive and robust predictive framework. This diversity enhances the model's performance, particularly in the context of healthcare, where data variability and class imbalance are prevalent.

3.5. Ensemble Model

A model selection strategy was employed to further improve predictive performance. The ensemble approach combines multiple classifiers to create a model that is typically more accurate and robust than any individual classifier, as illustrated in Figure 3.

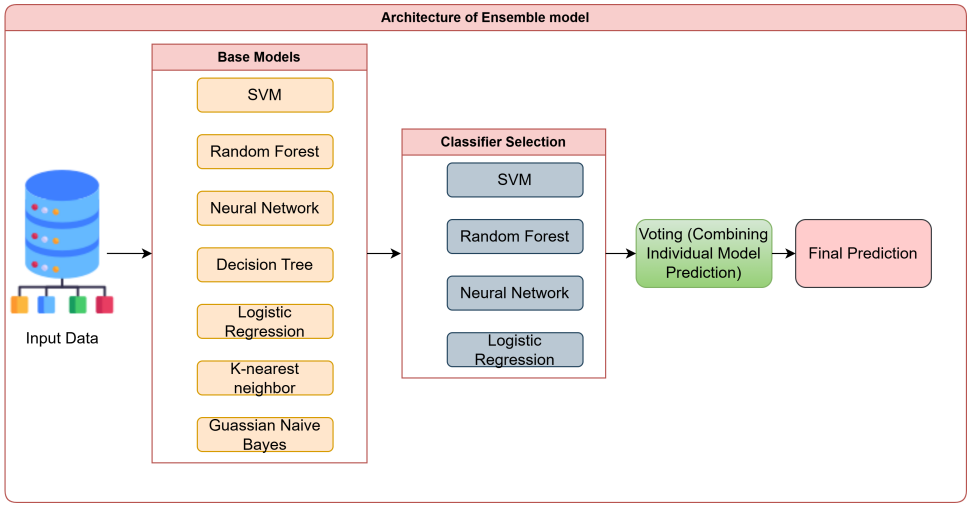


Figure 3. Architecture of ensemble model. The ensemble model integrates multiple base models (SVM, random forest, neural network, etc.) through classifier selection, followed by a voting mechanism to produce the final prediction.

3.5.1. Diversity Calculation

The diversity among classifiers in the ensemble was calculated based on pairwise agreement on predictions. Specifically, by letting **P** represent the set of predictions from all classifiers, the pairwise agreement between classifiers i and j can be computed as follows:

$$A_{ij} = \frac{1}{m} \sum_{k=1}^{m} \mathbb{I}(P_{ik} = P_{jk})$$

where $\mathbb{I}$ is an indicator function, P_{ik} and P_{jk} are the predictions of classifiers i and j for instance k, respectively, and m denotes the total number of instances. The overall diversity D is then quantified as follows:

$$D = 1 - \frac{1}{\binom{n}{2}} \sum_{i<j} A_{ij}$$

where $\binom{n}{2}$ represents the total number of unique pairs of classifiers in the ensemble, where n is the number of classifiers.

3.5.2. Classifier Selection Process

Only classifiers that achieved an accuracy greater than a certain threshold on the validation set were considered for inclusion in the ensemble. This thresholding ensured that only high-performing models contributed to the final prediction.

The selection of classifiers for the ensemble was guided by both their individual performance and the overall diversity of the ensemble. In cases where the diversity among the selected classifiers decreased below a specified threshold, the classifier with the lowest accuracy was excluded from the ensemble. This step was essential to ensure that the ensemble maintained sufficient diversity, which is crucial for improving model robustness.

The classifier selection process was employed in addition to the voting classifier approach for the following reasons:

1. Ensuring diversity and reducing redundancy: While the voting classifier aggregates predictions, it does not inherently account for model diversity or redundancy within the ensemble. Therefore, the classifier selection process is used to assess each model's contributions individually, only retaining those models that provide unique predictive value. This step is critical in preventing redundant models from influencing the ensemble, which can sometimes reduce the benefits of ensemble learning.
2. Balancing accuracy and diversity: Accuracy and diversity thresholds are applied during classifier selection to balance accuracy and diversity. By ensuring that the final ensemble includes models with high performance across different data perspectives, these thresholds help manage class imbalance, preventing dominance by a small number of similar models.
3. Enhanced flexibility for real-world applications: Through optimization of the ensemble composition, the model becomes more flexible and adaptable for clinical applications, as only high-performing and non-redundant models contribute to the final prediction.

While the voting classifier aggregates predictions, the classifier selection process adds a layer of refinement to the ensemble composition. This approach ensures an optimal balance between predictive accuracy, model diversity, and generalizability, enhancing the overall robustness of the ensemble for sarcopenia prediction.

The classifier selection algorithm incorporates diversity calculation to ensure the optimal combination of classifiers for the ensemble. The diversity calculation assesses each model's unique predictive contributions, selecting classifiers that not only achieve strong individual performance but also provide complementary perspectives, thereby improving the ensemble's robustness and predictive power.

The classifier selection algorithm evaluated a variety of models and, based on the performance metrics and diversity scores, selected the following five classifiers for inclusion in the ensemble: SVM, random forest, neural network, decision tree, and logistic regression.

The flow diagram in Figure 4 presents the algorithm for classifier selection.

This flow diagram illustrates the dynamic ensemble selection process, which involves evaluating classifier models based on both their accuracy and diversity to form an optimized ensemble. The process begins by initializing parameters, which include classifiers (a dictionary containing the classifier models to be evaluated), X_val (validation dataset features), and y_val (true labels for the validation dataset). Each classifier's performance

score is calculated in terms of accuracy, defined as the ratio of correct predictions to total predictions on the validation dataset.

For a classifier i, let $y_{\text{pred}}^{(i)}$ represent the predicted labels and y_{true} represent the true labels for the validation dataset X_{val} with n samples. The accuracy (performance score) of classifier i is then calculated as

$$\text{Performance Score}_i = \frac{1}{n} \sum_{j=1}^{n} 1\left(y_{\text{pred},j}^{(i)} = y_{\text{true},j}\right)$$

- $1(\cdot)$ is an indicator function that returns 1 if $y_{\text{pred},j}^{(i)} = y_{\text{true},j}$ and 0 otherwise.
- n denotes the total number of validation samples.

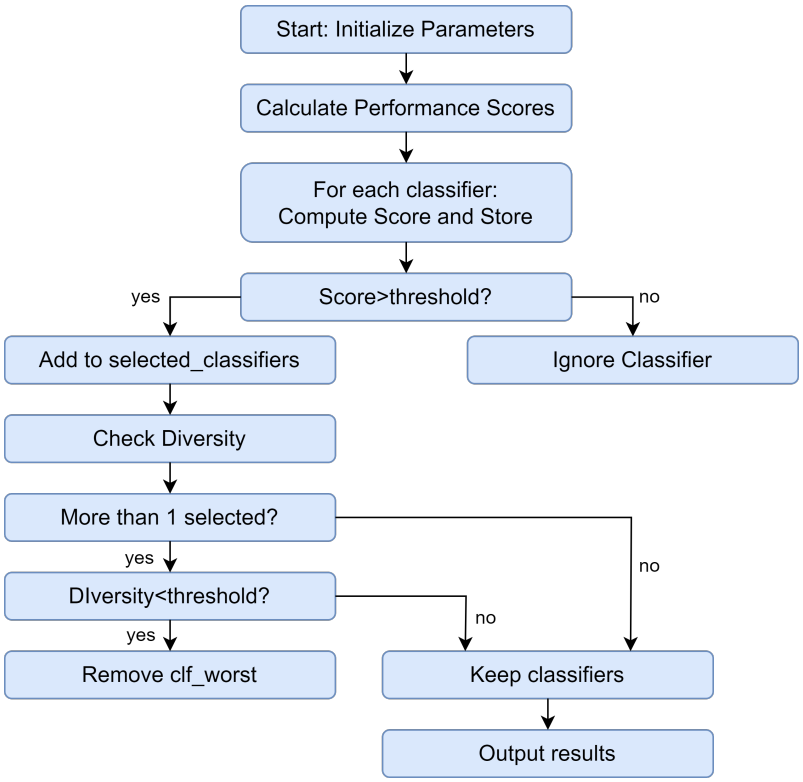


Figure 4. Classifier selection algorithm block scheme. The scheme shows parameter initialization, performance scoring, classifier selection based on thresholds, diversity checking, and the final output of selected classifiers with performance scores.

Only classifiers with scores exceeding a predefined threshold are selected for the ensemble. Next, diversity verification is performed to ensure that the selected models provide unique perspectives. If multiple classifiers meet the threshold, their diversity is assessed. If their diversity is below the diversity threshold, the algorithm removes the classifier with the lowest performance (Clf_worst) to increase the diversity of predictions. This structured approach balances individual model accuracy with ensemble diversity, enhancing the robustness of the final model (refer to Algorithm 1).

Algorithm 1: Classifier selection methodology.

Input: classifiers, X_{val}, y_{val}, score_threshold, diversity_threshold
Output: selected_classifiers
Initialize selected_classifiers as an empty list;
foreach *classifier in classifiers* **do**
  Calculate performance score using accuracy on (X_{val}, y_{val});
  if *performance score* $>$ *score_threshold* **then**
    Add classifier to selected_classifiers;
  end
end
if *length of selected_classifiers* > 1 **then**
  Calculate diversity among selected_classifiers;
  if *diversity* $<$ *diversity_threshold* **then**
    Remove classifier with lowest performance (Clf_worst);
  end
end
return selected_classifiers;

3.6. Voting Classifier

The final ensemble was constructed using VotingClassifier with soft voting. Soft voting is a flexible ensemble learning approach, where the predicted probabilities from each model are used instead of discrete class predictions. The final classification outcome $\hat{y}$ is determined by identifying the class that maximizes the weighted sum of probability estimates, formally expressed as follows:

$$\hat{y} = \arg\max_{y} \sum_{i=1}^{n} w_i \cdot p_i(y),$$

where w_i represents the weight assigned to the i-th classifier, $p_i(y)$ indicates the likelihood estimate for class y provided by the i-th classifier, and $\hat{y}$ denotes the predicted class, achieved by maximizing the weighted probabilistic sum. This methodology facilitates a more refined integration of model outputs by considering the confidence (probabilities) provided by each classifier, which can enhance performance, particularly when certain classifiers are more reliable than others.

3.7. Model Evaluation

The performance of the selected ensemble was assessed on a separate test set. Various metrics were used, including accuracy, precision, recall, and F1-score, utilizing the classification_report function in scikit-learn [31]. To evaluate the model's ability to discriminate, a receiver operating characteristic (ROC) curve was created, and the area under the curve (AUC) was calculated. The ROC curve was plotted using Matplotlib, with the AUC score included in the legend.

To illustrate the model's effectiveness across different classes, a confusion matrix was generated using ConfusionMatrixDisplay in scikit-learn. This confusion matrix provides insights into the model's true positive, true negative, false positive, and false negative predictions.

3.8. Computational Environment

All experiments were conducted using Kaggle notebooks to maintain consistency and reproducibility. Kaggle provides a cloud-based environment with standardized resources

and pre-configured libraries essential for data science and machine learning applications, ensuring a uniform computational setup. To promote transparency and enable replication of this study, the source code for this implementation has been made publicly accessible on GitHub at https://github.com/dilmurod86/Ensemble.git (accessed on 22 December 2024).

4. Results

This section presents the results of the dynamic ensemble selection process and the performance of the final ensemble model on the test set. Initially, individual classifiers were trained and evaluated on the validation set, with their hyperparameters optimized using RandomizedSearchCV, a method that performs random sampling of hyperparameter values to efficiently determine the optimal parameters within a specified range. Following this, a model selection strategy was employed to construct an ensemble that maximizes both accuracy and diversity. The performance of the final ensemble was assessed using several metrics, including accuracy, precision, recall, F1-score, and ROC AUC, which together provide a comprehensive evaluation of the model's predictive ability. The results reveal the effectiveness of the ensemble strategy in enhancing classification performance for detecting sarcopenia in contrast to standalone classifiers. Important results, such as the confusion matrix and ROC curve of the final model, are also presented to demonstrate the advantages and limitations of the proposed method.

4.1. Feature Selection

After applying RFE, a subset of the most relevant features was selected for use in the final model. RFE operates by systematically eliminating the least important features and re-fitting the model until the optimal set of features is identified. This method ensures that the model prioritizes the most significant and impactful variables, thereby enhancing overall predictive performance and minimizing the risk of overfitting.

Table 2 presents the features identified as the most important predictors for the classification task, highlighting their relevance in improving model accuracy and interpretability.

Table 2. Selected features after recursive feature elimination.

Feature	Description
Age	Age of the individual
Percent_BF	Percent Body Fat—Indicator of body composition
BMR_kcal	Basal Metabolic Rate—Calories required to maintain basic physiological functions
Plantar_L_1	Plantar Flexion—Measure of lower-limb flexibility and strength
Dorsal_L_1	Dorsal Flexion—Measure of lower-limb mobility
D_Dorsal	Dominant Dorsal Flexion—Flexibility measure of the dominant limb
SLS_R	Single-Leg Stance Test—Right—Balance test on the right leg
SARC_Calf	Sarcopenia self-assessment questionnaire + calf circumference—Composite measure of sarcopenia risk and muscle mass
FEV	Forced Expiratory Volume—Measure of lung function

4.2. Dataset Balancing

Before applying ADASYN, the dataset exhibited considerable class imbalance. The distribution of the target variable sarcopenia_2 was heavily skewed, with the majority class (class 0, indicating the absence of sarcopenia) comprising 524 instances and the minority class (class 1, indicating the presence of sarcopenia) comprising 140 instances. This imbalance, with the majority class outnumbering the minority class by a factor of

almost 4:1, introduced a risk of bias in the model, where the classifier may predominantly favor the majority class, leading to poor performance in identifying sarcopenia cases.

After applying ADASYN, the class distribution was successfully balanced. The number of instances in the minority class (class 1) increased to 301, while the number of instances in the majority class (class 0) was marginally modified to 297. This led to a nearly equal distribution of classes, with the minority class surpassing the majority class in some instances, refer to Figure 5.

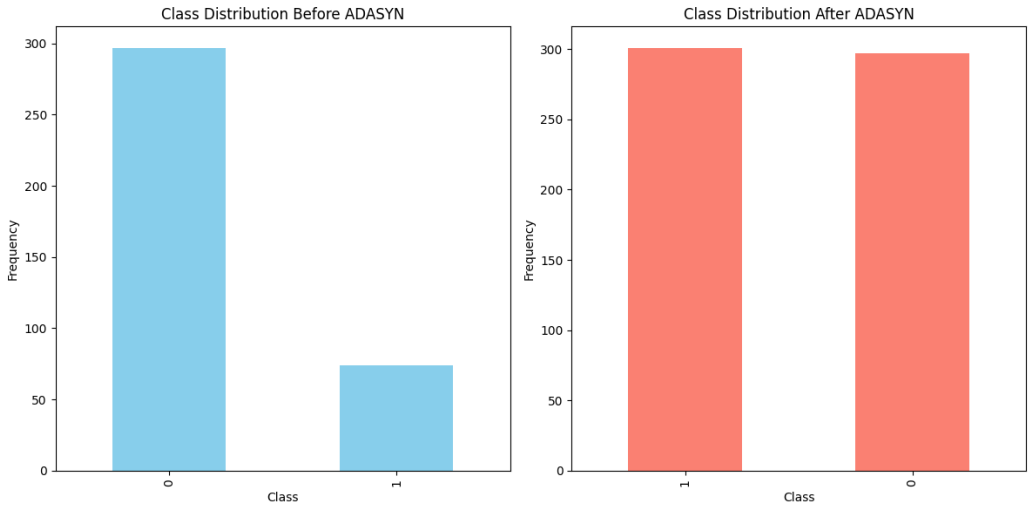


Figure 5. Class distribution before and after ADASYN application. The left panel shows the original class imbalance, while the right panel displays the balanced distribution achieved with ADASYN, enhancing the model's accuracy for the minority class.

ADASYN was selected for this study due to its unique approach to addressing class imbalance, which aligns with the characteristics and requirements of the dataset in this study. The rationale for selecting ADASYN over other methods is as follows:

1. Targeted synthesis for minority classes: Unlike traditional methods, such as SMOTE and random oversampling, ADASYN generates synthetic samples in regions of the feature space where the minority class is underrepresented. This approach improves the classifier's ability to learn challenging instances near decision boundaries, where minority class instances are sparse, thus enhancing sensitivity to minority class samples in complex datasets [32].
2. Adaptive nature of ADASYN: ADASYN adaptively assigns more weight to minority class instances that are harder to classify, based on their difficulty level. This adaptive sampling ensures a more balanced learning process by reducing the likelihood of introducing noise and redundancy, a common limitation of methods such as random oversampling and SMOTE [33].
3. Suitability for medical data imbalance: Medical datasets, such as the one used in this study, often exhibit significant imbalances such that the majority class overwhelms rare, critical cases. ADASYN's targeted approach is particularly useful in this context as it produces more informative synthetic samples without overwhelming the model with redundant samples that can arise from random oversampling or undersampling [34].
4. Preserving data distribution and class characteristics: ADASYN's adaptive mechanism balances classes and preserves the dataset's inherent distribution, making it an effective choice to avoid model overfitting or underfitting due to oversampling or undersampling [35].

4.3. Ensemble Model Performance

During cross-validation, the selected ensemble achieved an average accuracy of 95.48%, demonstrating its robustness and generalizability across different data subsets. However, when applied to the independent test set, the final ensemble achieved an accuracy of 88.50%. This minor reduction in accuracy is within the expected range and suggests that the model generalizes well to unseen data.

Overall, the ensemble achieved balanced performance across classes, with a macro-average F1-score of 0.84 and a weighted-average F1-score of 0.88. These results demonstrate the ensemble's ability to maintain high accuracy while ensuring fair representation of both classes in its predictions.

The performance evaluation matrix, illustrated in Figure 6, presents a comprehensive cross-tabulation of predicted classifications against ground-truth labels. The matrix indicates that out of 153 negative cases, the model correctly identified 143 as negative, resulting in a true negative rate of 93.46%. However, the model incorrectly classified 10 negative cases as positive (false positives).

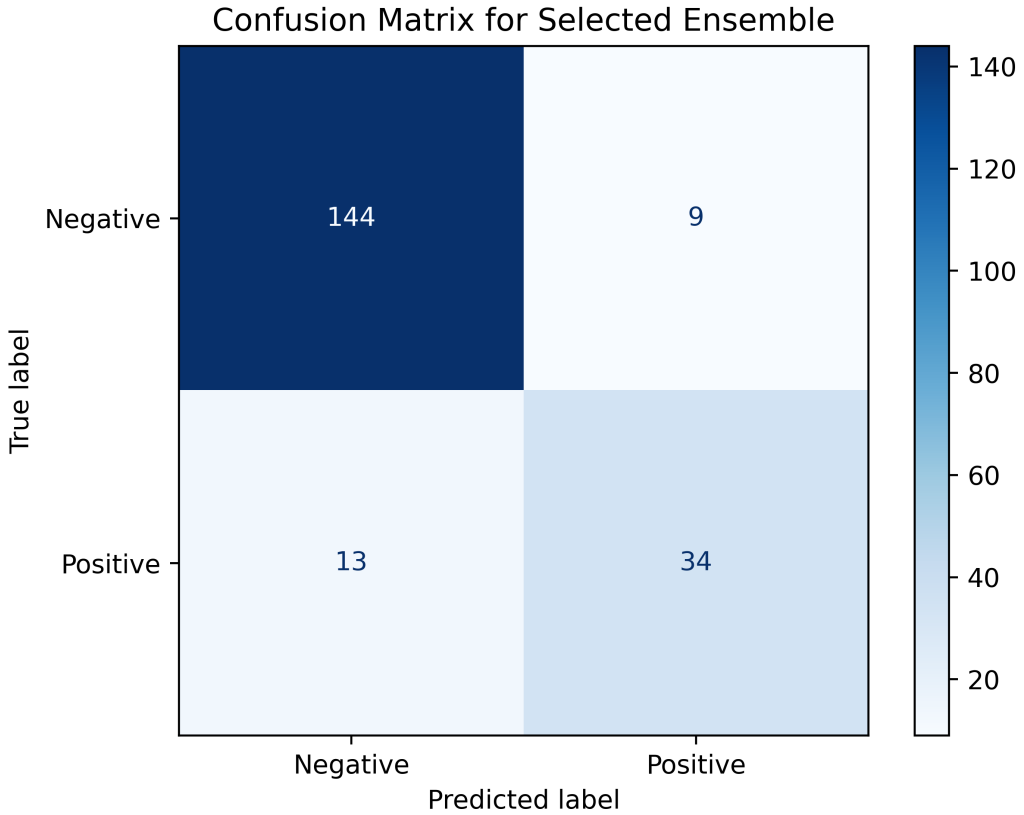


Figure 6. Contingency table of the optimized ensemble classifier. The matrix shows frequencies for true positives, true negatives, false positives, and false negatives, highlighting the classifier's strong specificity and balanced sensitivity across binary outcomes.

The model correctly identified 34 out of 47 positive cases, yielding a true positive rate (recall) of 72.34%. However, it misclassified 13 positive cases as negative (false negatives). These results demonstrate the model's strong performance in identifying negative cases and reasonable performance in identifying positive cases, which is particularly challenging due to class imbalance.

The ROC curve (Figure 7) illustrates the trade-off between the true positive rate (sensitivity) and false positive rate for various threshold settings. The AUC for the final ensemble model is 0.9327, indicating a high level of discrimination ability. This signifies

that the model is effective at distinguishing between positive and negative classes. An AUC close to 1.0 indicates excellent performance; therefore, the model demonstrates strong overall classification ability.

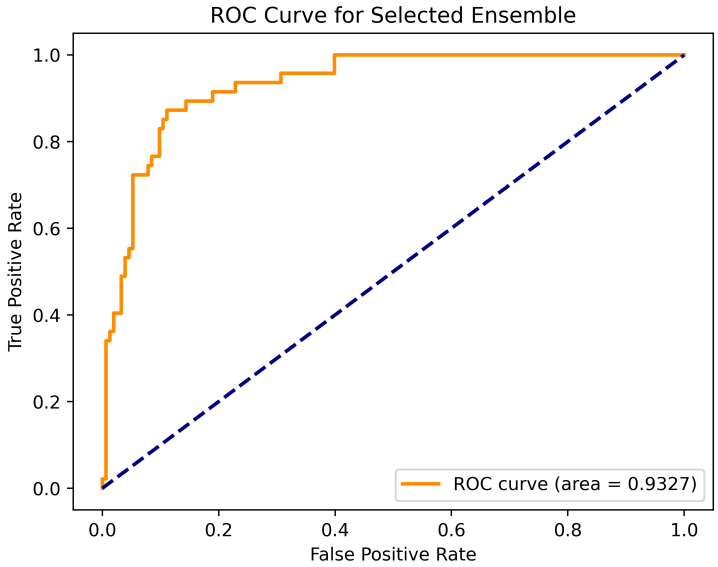


Figure 7. ROC curve for the tuned ensemble model. The area under the curve (AUC) of 0.9327 indicates that the model has a strong ability to discriminate between the positive and negative classes, highlighting its overall effectiveness in classification tasks.

To mitigate overfitting, the synthetic sample generation in ADASYN was controlled, focusing on minority class samples near the decision boundary. This selective approach prioritized the generation of only the most informative synthetic samples, avoiding overrepresentation of redundant patterns. Following this adjustment, the ROC curve analysis for the validation and test sets indicated a significant reduction in overfitting. The AUC values—0.9061 for validation and 0.9143 for test—are both high and closely aligned, indicating robust generalization. The ROC curves exhibit smooth and consistent trajectories across the datasets, demonstrating predictive stability. This alignment between the validation and test performance suggests that the model is not overly tuned to specific validation patterns, indicating its reliability for deployment and reflecting improved generalization, see Figure 8.

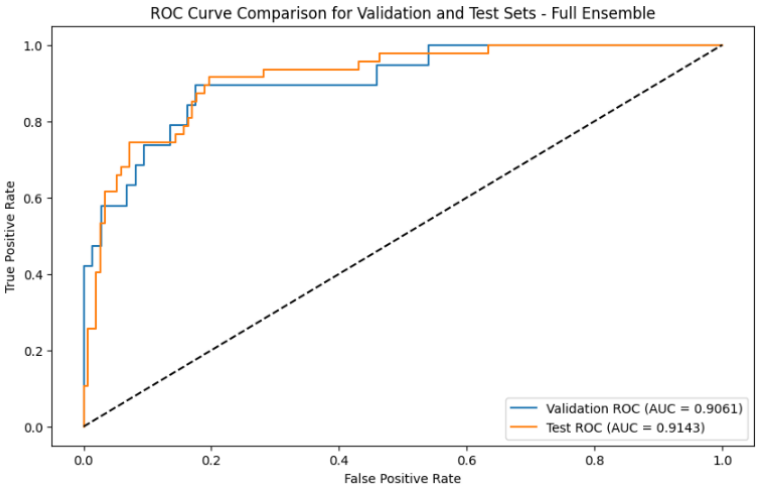


Figure 8. ROC curve comparison for validation and test sets: the ROC curves for the validation (AUC = 0.9061) and test sets (AUC = 0.9143) illustrate close alignment and high discriminative performance, indicating robust generalization and minimal overfitting in the ensemble model.

Cohen's kappa is a statistical metric used to evaluate the degree of agreement between predicted and actual classifications in a classification model. Unlike simple accuracy, Cohen's kappa accounts for random agreement, providing a more nuanced measure of model reliability, particularly for imbalanced datasets. Its values range from −1 to 1, where 1 indicates perfect agreement, 0 represents agreement equivalent to chance, and negative values suggest agreement that is worse than chance. In medical and clinical applications, where the practical implications of prediction accuracy are crucial, Cohen's kappa offers insights into a model's robustness and its effectiveness in distinguishing between classes. For the model in this study, a Cohen's kappa score of 0.6895 indicates substantial agreement beyond chance, supporting the model's reliability in identifying sarcopenia cases and highlighting its applicability in clinical decision-making.

The comparison of ensemble model metrics with and without the classifier selection algorithm reveals critical insights into the effectiveness of various feature selection methods on model performance (refer to Tables 3 and 4). When employing the classifier selection algorithm, RFE achieved the highest metrics, with an accuracy of 88.5%, precision of 84%, recall of 83%, and F1-score of 84%. This indicates that RFE, coupled with classifier selection, optimizes the model's predictive ability more effectively than other feature selection methods. In contrast, without the classifier selection algorithm, RFE achieved slightly lower metrics, with an accuracy of 87%, precision of 82%, recall of 83%, and F1-score of 83%. This pattern was also observed across other feature selection methods.

Mutual information, for instance, yielded an accuracy of 83.5% with classifier selection but decreased to 81% without it. Similarly, other methods such as minimum redundancy maximum relevance, Pearson, and Chi-square demonstrated reductions in all metrics when classifier selection was omitted. However, while Chi-square performed suboptimally with an F1-score of 73% with classifier selection, its performance slightly increased to 76% without it. These results indicate the classifier selection algorithm's role in enhancing model metrics, particularly for RFE and mutual information. In addition, they suggest that the combined application of feature selection and classifier selection algorithms can produce more robust and reliable predictive models.

Table 3. Comparison of ensemble model metrics using feature selection methods with classifier selection algorithm.

Feature Selection	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
RFE	88.5	84	83	84
MRMR	83	76	77	77
Mutual Information	83.5	77	79	78
Pearson	82	76	80	78
Chi-square	82	75	72	73

Table 4. Comparison of ensemble model metrics using feature selection methods without classifier selection algorithm.

Feature Selection	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
RFE	87	82	83	83
MRMR	80	74	73	74
Mutual Information	81	75	78	76
Pearson	80.5	74	79	76
Chi-square	83	76	76	76

5. Discussion

5.1. Performance Comparison of Base Models and Ensemble Model

The results of this study demonstrate the superior performance of the ensemble model compared to individual base models. As illustrated in Table 5, the final ensemble model achieved an accuracy of 88.5%, outperforming all the individual classifiers. This result underscores the effectiveness of the ensemble approach. Among the base models, the neural network achieved the highest accuracy of 87.4%, followed by the SVM with an accuracy of 86.91%. Nevertheless, the ensemble surpassed even the highest-performing base models by successfully combining the advantages of each classifier.

The random forest and logistic regression models also exhibited high performance, with accuracies of 86.01% and 85.2%, respectively. In contrast, models such as naive Bayes, with an accuracy of 77.12%, demonstrated lower performance, highlighting the limitation of relying solely on simpler models. The superior accuracy of the ensemble model suggests that it successfully leverages the complementary strengths of diverse classifiers, thereby enhancing its overall predictive performance.

Table 5. Accuracy comparison of base models and final ensemble model.

Model	Accuracy (%)
SVM	86.91
Random Forest	86.01
Neural Network	87.40
Decision Tree	83.50
KNN	83.00
Naive Bayes	77.12
Logistic Regression	85.20
Final Ensemble Model	**88.5**

5.2. Model Selection for Ensemble

The ensemble's model selection process involved evaluating the performance and diversity of the base classifiers. The selection was guided by two thresholds: an accuracy threshold of 83% and a diversity threshold of 0.1. Models that met these criteria were included in the final ensemble, ensuring that the ensemble was composed of high-performing and diverse classifiers.

The models selected for the ensemble included SVM, random forest, neural network, decision tree, and logistic regression. Each model offered unique advantages, contributing to the improved performance of the ensemble. The diversity threshold played a crucial role in maintaining the ensemble's robustness, ensuring that the selected models did not make redundant predictions, which could reduce the benefits of ensemble learning.

The success of the final ensemble is attributed to its balanced composition, where each model's strengths were harnessed to compensate for the limitations of other models. The ensemble's ability to surpass the performance of individual models demonstrates the power of combining different algorithms in a carefully constructed framework. This approach not only improves accuracy, but also enhances the model's generalizability, making it more reliable for real-world applications.

The ensemble model not only outperformed each individual base model, but also demonstrated the effectiveness of model selection. By setting accuracy and diversity thresholds, the ensemble was optimized to include only those models that positively contributed to the final prediction. This approach resulted in enhanced overall accuracy

and ensured that the model was well rounded and capable of addressing various aspects of the classification task. These results reinforce the value of ensemble learning in building robust and accurate predictive models, particularly in complex and imbalanced datasets such as the one used for sarcopenia classification.

5.3. Comparative Analysis of Ensemble Learning for Disease Prediction

To the best of our knowledge, this study represents one of the first applications of ensemble learning specifically for sarcopenia prediction. Comparing the results of this study with those of recent studies employing ensemble techniques to predict other diseases can highlight the proposed model's strengths and limitations and contribute to a broader understanding of ensemble methods in clinical settings. This comparative analysis can provide insights into how the proposed ensemble model aligns with or surpasses the current standards in medical predictive modeling, offering a detailed exploration of its effectiveness relative to recent advancements in the field.

The comparative analysis of ensemble models across various disease prediction tasks demonstrates the effectiveness of the proposed ensemble model, particularly in the context of sarcopenia prediction. By outperforming other ensemble approaches, the proposed model highlights the effectiveness of carefully designed model combinations in achieving superior predictive accuracy.

In contrast to previous ensemble approaches (Table 6), the proposed ensemble model, which combines SVM, random forest, neural network, decision tree, and logistic regression, achieved an accuracy of 88.5% for sarcopenia prediction. This considerable improvement in accuracy is attributed to the following key factors:

- **Diverse model selection**: The base models selected for the ensemble encompass a wide range of methodologies, including linear approaches, such as logistic regression, nonlinear models, such as neural networks and SVMs, and tree-based methods. This diversity allows the ensemble to capture different aspects of the data, resulting in more comprehensive and accurate predictions.
- **Model selection strategy**: The classifier selection process ensures that only the most effective and complementary models are included in the final ensemble. This approach not only enhances accuracy, but also maintains a balance between model complexity and generalizability.

Table 6. Comparison of ensemble models for disease prediction.

Author(s)	Disease Name	Accuracy	Combination of Base Models
[19]	Heart Disease	73.00%	Adaptive Boosting, SVM, DT, RF
[18]	Diabetes	73.50%	NB, RF, DT, XGBoost, LightGBM
[36]	Breast Cancer	85.38%	Deep learning based
Our model	Sarcopenia	88.5%	SVM, RF, NN, DT, LR

6. Conclusions

This study developed and validated an ensemble learning framework aimed at enhancing the accuracy of sarcopenia prediction, addressing the limitations of traditional diagnostic methods. By combining multiple machine learning models through a dynamic selection process, the proposed ensemble outperformed individual classifiers, achieving a prediction accuracy of 88.5%, demonstrating its effectiveness in sarcopenia detection.

A major strength of the proposed approach is its ability to manage class imbalance—a common challenge in medical datasets—by employing the ADASYN technique to generate synthetic samples for underrepresented classes. This improves the model's sensitivity

in detecting sarcopenia, which is crucial for early diagnosis and timely intervention in clinical practice.

The success of the ensemble model is attributed to its diverse combination of base classifiers, such as SVM, random forest, neural networks, and logistic regression, each contributing unique abilities. The classifier selection process ensures that only the most complementary and effective models are selected, enhancing both model accuracy and robustness.

This study contributes to the field of medical machine learning by presenting a novel application of ensemble learning for sarcopenia prediction. Future research can refine the model by incorporating additional patient data, such as real-time monitoring using wearable devices, and extend its use to other disease prediction tasks. The ensemble framework developed in this study provides a reliable tool for enhancing sarcopenia diagnosis, contributing to improved clinical outcomes and patient care.

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

Funding: This research was funded by the Gachon University Research Fund (GCU-202110230001) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2022R1F1A1074767).

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 the study are included in the article, further inquiries can be directed to the corresponding author.

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

References

1. Larsson, L.; Degens, H.; Li, M.; Salviati, L.; Lee, Y.I.; Thompson, W.; Kirkland, J.L.; Sandri, M. Sarcopenia: aging-related loss of muscle mass and function. *Physiol. Rev.* **2019**, *99*, 427–511. [CrossRef] [PubMed]
2. Rodrigues, F.; Domingos, C.; Monteiro, D.; Morouço, P. A review on aging, sarcopenia, falls, and resistance training in community-dwelling older adults. *Int. J. Environ. Res. Public Health* **2022**, *19*, 874. [CrossRef] [PubMed]
3. Couderc, A.; Liuu, E.; Boudou-Rouquette, P.; Poisson, J.; Frelaut, M.; Montégut, C.; Mebarki, S.; Geiss, R.; ap Thomas, Z.; Noret, A.; et al. Pre-Therapeutic Sarcopenia among Cancer Patients: An Up-to-Date Meta-Analysis of Prevalence and Predictive Value during Cancer Treatment. *Nutrients* **2023**, *15*, 1193. [CrossRef]
4. Kim, J.h. Machine-learning classifier models for predicting sarcopenia in the elderly based on physical factors. *Geriatr. Gerontol. Int.* **2024**, *24*, 595–602. [CrossRef] [PubMed]
5. Seok, M.; Kim, W. Sarcopenia prediction for elderly people using machine learning: A case study on physical activity. *Healthcare* **2023**, *11*, 1334. [CrossRef]
6. 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]
7. Khalid, S.I.; Massaad, E.; Kiapour, A.; Bridge, C.P.; Rigney, G.H.; Burrows, A.; Shim, J.; de la Garza Ramos, R.; Tobert, D.G.; Schoenfeld, A.J.; et al. Machine learning-based detection of sarcopenic obesity and association with adverse outcomes in patients undergoing surgical treatment for spinal metastases. *J. Neurosurg. Spine* **2023**, *40*, 291–300 . [CrossRef]
8. Mahajan, P.; Uddin, S.; Hajati, F.; Moni, M. Ensemble Learning for Disease Prediction: A Review. *Healthcare* **2023**, *11*, 1808. [CrossRef] [PubMed]
9. Vaccara, V.L.; Cascone, C.; Coppola, A.; Farolfi, T.; Cammarata, R.; Emerenziani, S.; Coppola, R.; Caputo, D. Role of preoperative sarcopenia in predicting postoperative complications and survival after pancreatoduodenectomy for pancreatic cancer. *Ann. Ital. Chir.* **2023**, *94*, 45–51.
10. Chang, B.S.; Peng, T.C.; Lue, K.; Wu, Y.F.; Huang, C.H. Comprehensive Assessment of the Clinical Risk Factors of Postoperative Adverse Events and Survival in Patients With Non-small-cell Lung Cancer. *In Vivo* **2023**, *37*, 1358–1364. [CrossRef]

11. Kurokawa, Y. How can sarcopenia be prevented after gastrointestinal surgery? *Ann. Gastroenterol. Surg.* **2022**, *6*, 604–605. [CrossRef] [PubMed]
12. Benedek, Z.; Coroș, M. The impact of sarcopenia on the postoperative outcome in colorectal cancer surgery. *Med. Pharm. Rep.* **2023**, *96*, 20–27. [CrossRef] [PubMed]
13. Mustapoevich, D.T.; Kim, W. Machine Learning Applications in Sarcopenia Detection and Management: A Comprehensive Survey. *Healthcare* **2023**, *11*, 2483. [CrossRef] [PubMed]
14. Martin, D.; Maeder, Y.; Kobayashi, K.; Schneider, M.; Koerfer, J.; Melloul, E.; Halkic, N.; Hübner, M.; Demartines, N.; Becce, F.; et al. Association between CT-Based Preoperative Sarcopenia and Outcomes in Patients That Underwent Liver Resections. *Cancers* **2022**, *14*, 261. [CrossRef] [PubMed]
15. Ozcan, S.G.G.; Kaçan, T.; Kacmaz, N.; Gundogan, B.; Vurgec, I.H.; Ozyildirim, S. Clinical significance of sarcopenia in patients undergoing treatment for gastric cancer. *Rev. Assoc. MéDica Bras.* **2023**, *69*, e20230568. [CrossRef]
16. Sato, R.; Oikawa, M.; Kakita, T.; Okada, T.; Abe, T.; Yazawa, T.; Tsuchiya, H.; Akazawa, N.; Yoshimachi, S.; Okano, H.; et al. Impact of Sarcopenia on Postoperative Complications in Obstructive Colorectal Cancer Patients Who Received Stenting as a Bridge to Curative Surgery. *J. Anus Rectum Colon* **2022**, *6*, 40–51. [CrossRef]
17. Dhande, B.; Bamble, K.; Chavan, S.; Maktum, T. Diabetes Heart Disease Prediction Using Machine Learning. In *ITM Web of Conferences*; EDP Sciences: Paris, France, 2022. [CrossRef]
18. Dutta, A.; Hasan, M.; Ahmad, M.; Awal, M.A.; Islam, M.; Masud, M.; Meshref, H. Early Prediction of Diabetes & Using an Ensemble of Machine Learning Models. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12378. [CrossRef] [PubMed]
19. Alshehri, G.A.; Alharbi, H.M. Prediction of Heart Disease using an Ensemble Learning Approach. *Int. J. Adv. Comput. Sci. Appl.* **2023**, *14*. [CrossRef]
20. Chandrasekhar, N.; Peddakrishna, S. Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. *Processes* **2023**, *11*, 1210. [CrossRef]
21. Rustamov, Z.; Rustamov, J.; Sultana, M.S.; Ywei, J.; Balakrishnan, V.; Zaki, N. Cardiovascular Disease Prediction using Ensemble Learning Techniques: A Stacking Approach. In Proceedings of the 2023 19th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) Kedah, Malaysia, 3–4 March 2023; pp. 93–98. [CrossRef]
22. Miao, L.; Wang, W. Cardiovascular Disease Prediction Based on Soft Voting Ensemble Model. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2023; Volume 2504. [CrossRef]
23. Kibria, H.B.; Nahiduzzaman, M.; Goni, M.O.F.; Ahsan, M.; Haider, J. An Ensemble Approach for the Prediction of Diabetes Mellitus Using a Soft Voting Classifier with an Explainable AI. *Sensors* **2022**, *22*, 7268. [CrossRef]
24. Chowdary, K.R.; Bhargav, P.; Nikhil, N.; Varun, K.; Jayanthi, D. Early heart disease prediction using ensemble learning techniques. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2022; Volume 2325. [CrossRef]
25. Jani, R.; Shanto, M.S.I.; Kabir, M.M.; Rahman, M.S.; Mridha, M.F. Heart Disease Prediction and Analysis Using Ensemble Architecture. In Proceedings of the 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 23–25 March 2022; pp. 1386–1390. [CrossRef]
26. Chaurasia, V.; Pandey, M.K.; Pal, S. Chronic kidney disease: A prediction and comparison of ensemble and basic classifiers performance. *Hum.-Intell. Syst. Integr.* **2022**, *4*, 1–10. [CrossRef]
27. Kazemi, Y.; Mirroshandel, S.A. A novel method for predicting kidney stone type using ensemble learning. *Artif. Intell. Med.* **2018**, *84*, 117–126. [CrossRef]
28. Ali, S.I.; Bilal, H.S.M.; Hussain, M.; Hussain, J.; Satti, F.A.; Hussain, M.; Park, G.H.; Chung, T.; Lee, S. Ensemble feature ranking for cost-based non-overlapping groups: A case study of chronic kidney disease diagnosis in developing countries. *IEEE Access* **2020**, *8*, 215623–215648. [CrossRef]
29. Verma, A.K.; Pal, S.; Tiwari, B.B. Skin disease prediction using ensemble methods and a new hybrid feature selection technique. *Iran J. Comput. Sci.* **2020**, *3*, 207–216. [CrossRef]
30. Igodan, E.C.; Thompson, A.F.B.; Obe, O.; Owolafe, O. Erythemato Squamous Disease prediction using ensemble multi-feature selection approach. *Int. J. Comput. Sci. Inf. Secur. IJCSIS* **2022**, *20*, 95–106.
31. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
32. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; IEEE: New York, NY, USA, 2008; pp. 1322–1328.
33. Susan, S.; Kumar, A. The balancing trick: Optimized sampling of imbalanced datasets—A brief survey of the recent State of the Art. *Eng. Rep.* **2021**, *3*, e12298. [CrossRef]
34. Dey, I.; Pratap, V. A Comparative Study of SMOTE, Borderline-SMOTE, and ADASYN Oversampling Techniques using Different Classifiers. In Proceedings of the 2023 3rd International Conference on Smart Data Intelligence (ICSMDI), Trichy, India, 30–31 March 2023; pp. 294–302. [CrossRef]

35. Eom, G.; Byeon, H. Searching for Optimal Oversampling to Process Imbalanced Data: Generative Adversarial Networks and Synthetic Minority Over-Sampling Technique. *Mathematics* **2023**, *11*, 3605. [CrossRef]
36. Ambrews, A.B.; Moung, E.G.; Farzamnia, A.; Yahya, F.; Omatu, S.; Angeline, L. Ensemble Based Machine Learning Model for Heart Disease Prediction. In Proceedings of the 2022 International Conference on Communications, Information, Electronic and Energy Systems (CIEES), Veliko Tarnovo, Bulgaria, 24–26 November 2022; pp. 1–6. [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.