

Article

Prediction of the Health Status of Older Adults Using Oversampling and Neural Network

Yue Li , Qingyu Hu, Guilan Xie and Gong Chen *

Institute of Population Research, Peking University, Beijing 100871, China; 2301213053@pku.edu.cn (Q.H.); xieguilan@stu.pku.edu.cn (G.X.)

* Correspondence: chengong@pku.edu.cn

Abstract: Self-rated health (SRH) serves as an important indicator for measuring the physical and mental well-being of older adults, holding significance for their health management and disease prevention. In this paper, we introduce a novel classification method based on oversampling and neural network with the objective of enhancing the accuracy of predict the SRH of older adults. Utilizing data from the 2020 China Family Panel Studies (CFPS), we included a total of 6596 participants aged 60 years and above in our analysis. To mitigate the impact of imbalanced data, an improved oversampling was proposed, known as weighted Tomek-links adaptive semi-supervised weighted oversampling (WTASUWO). It firstly removes the features that are not relevant to the classification by ReliefF. Consequently, it combines undersampling and oversampling. To improve the prediction accuracy of the classifier, an improved multi-layer perception (IMLP) for predicting the SRH was constructed based on bagging and adjusted learning rate. Referring to the experimental results, WTASUWO can effectively improve the prediction performance of a classifier when being applied on an imbalanced dataset, and the IMLP using WTASUWO achieves a higher accuracy. This method can more objectively and accurately assess the health status and identify factors affecting the SRH of older adults. By mining relevant information related the health status of older adults and constructing the prediction model, we can provide policymakers and healthcare professionals with targeted intervention techniques to focus on the health needs of older adults. Meanwhile, this method provides a practical research basis for improving the health level of older adults in China.

Keywords: feature selection; class imbalance; neural network; self-rated health; older adults**MSC:** 68U99

Citation: Li, Y.; Hu, Q.; Xie, G.; Chen, G. Prediction of the Health Status of Older Adults Using Oversampling and Neural Network. *Mathematics* **2023**, *11*, 4985. <https://doi.org/10.3390/math11244985>

Academic Editor: Manuel Franco

Received: 17 November 2023

Revised: 10 December 2023

Accepted: 15 December 2023

Published: 17 December 2023



Copyright: © 2023 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

As China transitions into an aging society, the depth of ageing is on the rise. According to the seventh national census data, individuals aged 60 and above in China constitute now 18.7% of the total population, with those aged 65 and above accounting for 13.5%. Projections indicate that the peak will reach 34.9% around 2053. Notably, China's aging rate is 1.5 times that of the global population [1]. In the wave of silver hair, health issues have garnered widespread attention, particularly for the elderly population. With the rapid development of the social economy and the increasing trend of population aging, health problems in old age are anticipated to become increasingly prominent [2].

In this context, utilizing artificial intelligence (AI) information technology to enhance the quality of health and aging services is beneficial to meet their needs for a happy life and may address the challenge of social aging [3]. Furthermore, the health status of older adults is not only related to the basic survival state, but also depends on the long-term development of society. Gaining insights into the health status of older adults and predicting their health needs holds significant importance in formulating targeted health policies and services. The content of the study addresses a crucial topic in the pursuit of

achieving healthy and active aging in China. It not only broadens the research scope in this field but also carries substantial theoretical and practical significance for improving the health level of older adults and facilitating the realization of healthy aging.

At present, the assessment of the health status of older adults primarily relies on objective physiological indicators and medical data. However, these often fall short in reflecting the subjective health of older adults [4]. Self-rated health (SRH) emerges as a comprehensive health assessment index, allowing older adults to provide a self-assessment based on their physiological, psychological, and social suitability. It has increasingly been employed by researchers to measure the health status of older adults [5]. The prediction of SRH among older adults is beneficial to prevent potential health problems in advance and improve their overall quality of life. Therefore, the accurate prediction of the SRH of older adults, enabling the provision of more personalized and effective health services, has emerged as a research hotspot. Simultaneously, this method can offer valuable insights for health prediction and evaluation in various fields, thereby fostering the development of health science research.

In the research on predicting SRH in older adults, the imbalance in the number of older adults with various health states can lead to sample imbalance during the training process of commonly used classification methods. This imbalance significantly affects the practicability of the classifier. Effectively improving the accuracy of minority classes and the overall performance of classifiers has thus become a hot topic in the field of prediction [6,7]. Imbalanced datasets, characterized by a significant difference in the number of samples among different classes [8,9], poses a challenge. Currently, addressing the issue of data imbalance in classification involves existing research carried out in two aspects: data equalization processing and classification optimization [10].

At the data level, enhancing the accuracy of minority classes primarily involves altering the distribution of samples and selecting features in the training dataset. Resampling techniques are commonly employed for this purpose, with random undersampling and oversampling being the most utilized. In more detail, random undersampling refers to the random elimination of majority samples; however, this method may lose some useful information, resulting in affecting the classifier performance. As for the random oversampling, it mainly consists of randomly copying minority samples to make the class distribution of data balanced. However, this method increases the possibility of overfitting the model due to repeated replication of minority samples [11,12]. At the algorithm level, enhancing classification algorithms involves improving existing ones or designing new algorithms based on the inherent characteristics of imbalanced datasets. This is primarily achieved through ensemble and cost-sensitive learning to optimize the models [13]. As a powerful machine learning technique, neural networks (NNs) have demonstrated remarkable results in several fields in recent years [14]. Particularly in health prediction, NNs excel at accurately predicting individual health status by learning complex features and nonlinear relationships within large datasets [15].

To sum up, the primary objective of the current study is to develop a SRH prediction model for older adults with high accuracy, aiming to improve the refinement of health management. The main contributions include:

- A data processing method, based on oversampling technique, is proposed to address class imbalance in datasets related to the health status of older adults;
- An improved NN model is introduced to predict the health status of older adults, providing higher accuracy and stability in predicting the health status of older adults compared to traditional machine learning algorithms;
- Based on the empirical and publicly available datasets, the proposed method further demonstrates its effectiveness;
- By predicting the health status of older adults, potential health problems can be identified in advance and personalized health advice as well as interventions for the elderly, therefore improving their quality of life.

2. Literature Review

An imbalanced dataset refers to the number of samples in one class that is significantly larger than those in other classes. Due to sampling disparity, the traditional methods often fail to acquire the accurate prediction results. Nowadays, the methods applied to solve the classification problem for imbalanced dataset can be divided into two categories: the first one starting with the dataset, it transfers the imbalanced data into a balanced data by changing the data distribution or selecting the features that can better express the data information. As for the second category, it starts from the classification algorithm, where it enhances the accuracy according to the limitations resulting in the application of the algorithm on the imbalanced dataset.

2.1. Research on Oversampling Techniques

Regarding the oversampling techniques, Chawla et al. [16] proposed a novel technique, known as the synthetic minority oversampling technique (SMOTE). This technique achieves the class distribution balance by adding artificial minority oversampling samples to the data. Moreover, SMOTE has provided a new direction for addressing the imbalanced data and has become an effective oversampling technique that has been applied in several fields [17]. However, there are still some shortcomings regarding this technique: the fuzzy boundary problem, caused by SMOTE, ignores the imbalance in minority classes, and it is easily affected by noise. Therefore, it is difficult to get the best performance of a classifier by empirically selecting the different parameters.

Regarding these problems, some researchers have made improvements to SMOTE to adapt to different data characteristics. For instance, Han et al. [18] considered the minority class samples distributed near the class boundary and introduced the Borderline-SMOTE algorithm to improve the learning ability of the boundary samples. Moreover, He et al. [19] proposed the adaptive synthetic (ADASYN) sampling method based on sample distribution, effectively improving the prediction performance of the classifier [20]. In addition, Safe-level SMOTE [21] focused on the class overlap problem caused by SMOTE. It assigned a safety factor to each minority class sample before synthesizing the new samples to ensure that the new samples were distributed in the safe area. Furthermore, Douzas et al. [22] proposed a cluster-based oversampling method, called K-means SMOTE, which effectively addresses the problem of inter- and intra-class imbalance [23]. However, the oversampling results were contingent on the effectiveness of K-means, potentially leading to errors in minority class samples. In a different approach, Tao et al. [24] proposed a weighted oversampling method based on the support vector data description (SVDD) to tackle inter-class overlap and internal imbalanced problems.

In addition, Nekooimehr et al. [25] proposed the adaptive semi-supervised weighted oversampling (ASUWO) algorithm. This method adopted a semi-supervised learning and adaptively adjusted the weights of samples to achieve a balance in the number of samples between classes, thereby enhancing the generalization performance of the model. However, attention should be given to computational complexity and parameter sensitivity. In case of dealing with severely imbalanced datasets, combining it with other methods may be necessary. In practical applications, variations in base models may lead to significant differences, and appropriate adjustment and optimization can be applied based on specific problems [26].

In terms of hybrid sampling, Batista et al. [27] proposed SMOTE Tomek and SMOTE edited nearest neighbor (SMOTE ENN) by combining both common methods. SMOTE Tomek first oversamples the minority class samples to balance the data distribution. Consequently, it is combined with the Tomek-links technique to remove the blurred samples between the majority and minority classes. Moreover, Seiffert et al. [28] combined Adaboost with oversampling and undersampling, and proposed SMOTEBoost and RUSBoost algorithms to increase the robustness of the classifier and improve its accuracy.

2.2. Research on Neural Network Models

At the algorithm level, modification primarily involves adjusting the bias of the algorithm on the dataset level, directing decisions to be biased towards the minority class. Techniques, such as ensemble learning and feature selection, play a crucial role in achieving the bias. In the realm of machine learning applications, approaches include logistic regression (LR), K-nearest neighbor (KNN), decision tree (DT), random forest (RF), support vector machine (SVM), and other single algorithms. On the other hand, some researchers use the voting methods, adaptive boosting (AdaBoost), gradient boosting decision tree (GBDT), and extreme gradient boosting (XGBoost) for prediction and classification [29]. As a widely used method in machine learning, NNs have been widely used in the healthcare domain, addressing challenges such as classification, regression, and clustering.

Moreover, artificial neural network (ANN) is a nonlinear mapping and adaptive dynamic technique formed by several simple neurons widely interconnected [30]. Added to that, back propagation (BP) and radial basis function (RBF) networks have the disadvantages of slow learning speed, being easy to fall into local extremes, and having low accuracy of prediction results. As for the self-organizing map (SOM) network, it adopts unsupervised learning rules and lacks classification information. For instance, Rumelhart et al. proposed multiple-layers perceptron (MLP), a kind of neural network using forward propagation and error backward propagation; it is mainly applied to solve the multi-class prediction problem of nonlinear data [31,32]. The generalization ability and processing efficiency of MLP have several advantages compared to BP, RBF, and SOM networks. If MLP is applied to fit more complex mapping functions with higher accuracy, the main challenges become the structural design and parameter optimization of the model.

The model structure refers to the depth of the model and the width of each layer as well as the activation function. MLP can automatically extract the features and abstraction levels from the data. Generally, the deeper the depth is, the fewer nodes could be used in each layer. However, due to the gradient problems, it is difficult to train such a model. However, the accuracy can be enhanced by increasing the model width even if it is shallow. The wider the width of the model is, the better the fitting ability will be, but the easier the overfitted will appear. The advantages and limitations of various activation functions are not conclusive, and it is required to be selected according to the application scenarios and the datasets. Therefore, to get a suitable model structure for the specific application, it is necessary to adjust it after computing several experiments.

Concerning the model training, the optimization algorithms mainly improve the learning factors and the training methods. Among the first-order optimization algorithms, stochastic gradient descent (SGD) and its variants are among the most widely used in deep learning (DL). SGD is simple and has low computational complexity, but its convergence rate is slow. Given the learning rate significantly influences model performance, researchers have proposed several adaptive learning rate algorithms based on SGD. Among these, the adaptive moment (Adam) estimation algorithm has gained popularity. For example, Khan et al. [33] used the genetic algorithm (GA) to optimize the BP network, comparing the optimized prediction results obtained using some technical indicators and original BP. The results indicated that the prediction accuracy of the GA-based BP was higher. Moreover, Kar et al. [34] applied the evolutionary algorithm (EA) to train the ANN model. The synergy of NNs with optimization algorithms presents an avenue for further exploration. Studying how to apply improved algorithms to enhance the effectiveness and accuracy of NNs for prediction is a worthwhile pursuit.

As a crucial parameter in NNs, a fixed learning rate has no adaptability to the objective function. Setting the learning rate too high may result in training shock and divergence, while setting it too low may lead to slow convergence. In practical applications, adjusting the learning rate based on the specific task and dataset becomes necessary. A common approach is to employ dynamic methods that adjust this parameter, such as gradually reducing the learning rate during the learning process. This helps prevent the model from

learning too quickly at the beginning of training and potentially falling into a local optimal solution [35].

In addition, deep neural networks (DNNs) have a deeper network structure compared to shallow neural networks and can learn more complex features and higher level abstractions. For instance, Lin et al. [36] trained transformer protein language models with up to 15 billion parameters using experimental and high-quality predicted structures. They converted protein sequences into textual forms processable by the language model, leveraging the model's learning ability to predict protein structure. This approach significantly improves the accuracy and efficiency of predictions, particularly when dealing with large datasets, showcasing robust parallel processing capabilities. However, language models face limitations in certain fields with limited protein sequence data due to their dependency on large training datasets. To address this challenge, Lin et al. [37] proposes an approach called ghost convolution with BottleneckCSP and a tiny target prediction head incorporating YOLOv5 for PV panel defect detection. The approach enhances the network representation by introducing ghost convolution and BottleneckCSP. Furthermore, the prediction of small targets is improved by tiny target prediction head. These improvements are combined with YOLOv5, resulting in a new and enhanced detection model. While this method improves the prediction of small targets, enabling the detection of smaller defects, it is important to note that the YOLOv5 model has a large computational load and may require longer processing time for large-scale detection tasks. Overall, DNNs represent a very promising approach, but the structure of such models is more complex and they may pose a risk of overfitting for some simple detection tasks. Therefore, the specific applicability still requires evaluation based on the actual problem and data situation.

2.3. Major Findings

In recent years, classification methods based on oversampling and NNs have been widely applied in the field of health prediction. In practical applications, NN prediction methods also face some challenges. Assessing the health status of older adults involves multiple dimensions, including physiological, psychological, and social aspects. The effective extraction and integration of information becomes pivotal for the prediction model. Notably, there is a limited discussion on this issue in the context of health assessment for older adults in China.

(1) Data quality

For the SRH prediction of older adults, the selection of the data features is crucial to the prediction effect of the model. In more detail, undersampling may lead to lose information of majority classes, and oversampling poses a challenge in ensuring the generated samples conformity with the real data distribution. It is also robust to outliers, mitigating the impact of large deviations in predictions caused by outlier data points. Although hybrid sampling combines undersampling and oversampling methods, it also yields limitations. To improve the prediction rate of a classifier, it is necessary to optimize the oversampling technique.

(2) Model optimization

The traditional statistical methods are difficult to cope with the task regarding feature extraction and classification of complex data. Meanwhile, due to the large heterogeneity of elderly population, the health status of distinct individuals is obviously different. Therefore, NN has good nonlinear fitting ability and high fault tolerance. Even for small sample size data, it can generate a good prediction effect. Meanwhile, MLP has a strong learning ability. However, there are limited studies deploying MLP to predict the health status of older adults. Nevertheless, there is certain research significance of improving the application of MLP in the SRH prediction for older adults.

This paper focused on using oversampling and NNs to predict the health status of older adults. Although there are many research studies in this area, we believe that our work is of innovative value in the following ways: The oversampling technique is applied to balance the data and improve the generalization ability of the model. Moreover, the NN structure is applied to grasp the complex features of the health status of older adults, further

improving the prediction accuracy. This method is applied to train the model with existing datasets and can effectively identify health risk factors of older adults. Theoretically, the results can enrich the data processing methods as well as the theoretical framework of the health status prediction research. In practice, it serves as a basis for the implementation and improving the health status of older adults.

3. Methods

In this section, we mainly explore the way to use oversampling and NNs to improve the prediction accuracy of older adults' health status, as well as explore the influencing factors on the health status, in order to provide a scientific basis for health management and policy formulation. The basic principles of the study include the following:

- (1) Description and analysis of experimental data;
- (2) Proposition of a new oversampling technique and exploration of the main influencing factors;
- (3) Development of a new high-accuracy NN prediction model.

3.1. Data Sources and Analysis

To study the SRH prediction of older adults in China, the experimental data were gathered from the 2020 China Family Panel Studies (CFPS). The reasons include the following:

- (1) Representativeness: CFPS is a nationwide comprehensive social longitudinal survey, with high reliability, the changes in the society, economy, population, and health in China.
- (2) Richness: The data not only contain the basic information of older adults, but also cover their health status, lifestyle, and other aspects; this provides rich data support for studying the influencing factors on the SRH of older adults.
- (3) Timeliness: The data reflect the current health status and the factors affecting it and can provide policymakers and relevant departments with the latest reference basis [38].

In addition, it is relevant to the problem solved in this paper, summarized by the following two points:

- (1) It helps to reveal the current situation and characteristics of older adults' health status and can explore the influencing factors and provide a theoretical basis for enhancing their health level;
- (2) The data have high representativeness, reliability, and richness, helping in promoting and applying the results of the study and providing empirical support for solving the aging problem in China.

In this paper, the health status of older adults aged 60 and above was selected as the research object. First, the original data were preprocessed, including missing data handling and normalization. For each feature, samples containing missing values were removed, and a total of 6596 samples with 36 features were included for modeling. Normalization is a commonly used data preprocessing method to remove the effect of data magnitude and accelerate model converge. The feature attributes are normalized to between [0, 1], as follows:

$$x_{ij} = \frac{x_{ij} - x_{ij}^{\min}}{x_{ij}^{\max} - x_{ij}^{\min}} \quad (1)$$

where x_{ij} is the j th feature of the i th sample, and x_{ij}^{\max} and x_{ij}^{\min} are the maximum and minimum values of the j th feature, respectively. This dataset contains detailed information of older adults, such as basic information, working status, health status, lifestyle, network services, network importance, and importance of access to information. Table 1 describes the features of the data.

Table 1. Description of experimental data based on the 2020 CFPS.

Modules	Features	Descriptions
Basic information	x1: Sex	1 = male; 2 = female
	x2: Age	1 = 60–70; 2 = 70–80; 3 = 80 or above
	x3: Education	1 = no education; 2 = illiterate/semi-illiterate; 3 = primary school; 4 = junior high school; 5 = senior high school; 6 = junior college; 7 = bachelor’s degree or above
	x4: Marriage status	1 = unmarried; 2 = remarried; 3 = cohabiting; 4 = divorced; 5 = widowed
	x5: Living area	1 = urban; 2 = rural
	x6: Family numbers	1 = 1–3; 2 = 4–6; 3 = 7–9; 4 = 10 or above
Working status	x7: Household registration status	1 = agricultural; 2 = non-agricultural; 3 = no household registration; 4 = resident; 5 = unknown
	x8: Working condition	1 = employed; 2 = unemployed; 3 = no longer labor force
	x9: After-tax wage income in the past year	1 = 0–1000; 2 = 1000–2000; 3 = 2000–3000; 4 = 3000–4000; 5 = 4000–5000; 6 = 5000 or above (RMB)
	x10: Whether receiving pensions and retirement pay	1 = yes; 2 = no; 3 = unknown
Health status	x11: Whether receiving old-age insurance	1 = yes; 2 = no; 3 = unknown
	x12: Health changes	1 = worse; 2 = unchanged; 3 = better; 4 = unknown
	x13: Any chronic diseases within six months	1 = yes; 2 = no; 3 = unknown
Lifestyle	x14: Where do you usually go to see a doctor	1 = general hospitals; 2 = specialized hospitals; 3 = service centers; 4 = service stations; 5 = clinics; 6 = other
	x15: Frequency of physical exercise (times)	1 = less than 1 time per month; 2 = more than 1 time per month, but less than 1 time per week; 3 = 1–2 times per week; 4 = 3–4 times per week; 5 = 5 or more times per week; 6 = 1 time per day; 7 = 2 or more times per day; 8 = never attend; 9 = unknown
	x16: Have you smoked in the past month	1 = yes; 2 = no; 3 = unknown
	x17: Have you had a drink three times a week in the past month	1 = yes; 2 = no; 3 = unknown
	x18: Whether to read	1 = yes; 2 = no; 3 = unknown
	x19: Mobile Internet access	1 = yes; 2 = no
Network services	x20: Whether surfing online	1 = yes; 2 = no
	x21: Whether playing video games	1 = yes; 2 = no; 3 = unknown
	x22: Whether doing shopping online	1 = yes; 2 = no; 3 = unknown
	x23: Whether watching short videos	1 = yes; 2 = no; 3 = unknown
	x24: Whether learning online	1 = yes; 2 = no; 3 = unknown
	x25: Whether using WeChat	1 = yes; 2 = no; 3 = unknown
Network importance	x26: Importance of network to work	1~5, 1 = very unimportant, 5 = very important
	x27: Importance of network to entertainment	1~5, 1 = very unimportant, 5 = very important
	x28: Importance of network to keeping in touch with family and friends	1~5, 1 = very unimportant, 5 = very important
	x29: Importance of network to studying	1~5, 1 = very unimportant, 5 = very important
	x30: Importance of network to daily life	1~5, 1 = very unimportant, 5 = very important
Importance of access to information	x31: Television	1~5, 1 = very unimportant, 5 = very important
	x32: Internet	1~5, 1 = very unimportant, 5 = very important
	x33: Newspapers, periodicals, and magazines	1~5, 1 = very unimportant, 5 = very important
	x34: Broadcast	1~5, 1 = very unimportant, 5 = very important
	x35: Text messages	1~5, 1 = very unimportant, 5 = very important
	x36: Told by others	1~5, 1 = very unimportant, 5 = very important

According to the question “How healthy do you think you are?”, the answers are classified into unhealthy, average, relatively healthy, very healthy, and super healthy. The numbers of each class were 1607, 1081, 2448, 722, and 738, respectively. This paper combined the classes “relatively healthy”, “very healthy”, and “super healthy” into “healthy”; thus, the final categories were “unhealthy”, “average”, and “healthy”, whose data amounts were 1607, 1081, 3908, having the values of 1, 2, and 3, respectively. Currently, there are no unified criteria for defining the classification of various health levels, which are generally based on the purpose and problem of the study [39]. Therefore, the SRH of older adults tends to be good. Among them, the “healthy” older adults’ category accounts for the highest proportion, reaching 59.25% of the total number of participants, indicating that most older adults can maintain good health in daily life. This class was followed by “unhealthy” and “average” older adults, accounting for 24.36% and 16.39%, respectively, indicating that some older adults face health problems.

3.2. Oversampling Technique Based on WTASUWO

Oversampling is a technique yielding to data augmentation. It consists of increasing the number of minority class samples to balance the number of samples among the other classes, reducing, therefore, the influence of imbalanced class problem on model performance. In this paper, oversampling was selected due to several reasons:

- (1) Oversampling is an operation based on using the raw data and does not lose information;
- (2) Oversampling facilitates the model in more effective learning features from the samples belonging to minority classes during the training;
- (3) Oversampling is easier to implement compared to other data preprocessing techniques;
- (4) Oversampling has been widely applied to the prediction problems in several unbalanced data scenarios.

For example, some studies have used the oversampling techniques to solve the imbalanced problem in the diabetes risk prediction, and experiments have proved that it can effectively enhance the prediction performance of the model [40]. Moreover, the ASUWO algorithm was proposed in 2016 based on the SMOTE oversampling; this algorithm balances the number of samples by sampling with retraction in minority class samples as well as the application of weights to improve the performance of the model on the imbalanced dataset. However, some problems still reside in this algorithm.

(1) Calculate sample weights: Due to the noise and outliers in the samples, this can lead to uneven weight distribution of ASUWO, affecting the model’s prediction performance.

(2) Synthesize sample outliers: The synthesized samples by the boundary between the minority and majority classes may contain both data characteristics, resulting in outliers in the synthesized samples. Therefore, the imbalanced distribution of samples may affect the accuracy of the outlier detection based on clustering [41].

Thus, in this paper, we proposed a ReliefF-based of feature weighted Tomek-links ASUWO (WTASUWO) oversampling, which consists of combining Tomek-links undersampling and ASUWO oversampling. The improvements of this techniques can be summarized as follows:

- (1) A weighting strategy that was used to improve the generalization ability and accuracy of the model.

Problems like sparse data distribution, feature redundancy, and irrelevance in high-dimensional imbalanced data pose difficulties for traditional learning algorithms in identifying minority class samples. A prevalent approach is to address this by reducing the dimensionality of the data before classification, operating in a new dimension space. Feature selection aims to identify an “optimal” subset of relevant features. This not only reduce the complexity of the problem and enhances the processing efficiency but also effectively streamlines the design of the classifier [42].

The Relief algorithm, initially proposed by Kira, was limited to the classifying of two types of data. ReliefF, an improved feature selection method building upon Relief, introduces the concept of positive and negative samples. It calculates feature correlation coefficients, utilizing the square root to measure the distance between samples. ReliefF is versatile, applicable to both classification and regression problems, enhancing model performance by selecting features with higher weights. In practice applications, ReliefF can also be combined with various machine learning algorithms. Notably, it assigns substantial weights to features that significantly contribute to classification, offering a simple and fast approach for feature selection. This approach has no restriction regarding the type of data [43,44]. Therefore, ReliefF is employed to sort the importance of the features to reduce its dimension. The algorithm is described as follows:

Step 1: Initialize the weights of all features to 0.

Step 2: For $i = 1$ to m do (m is the number of sampling times)

- (1) Randomly select a sample R .
- (2) Finds k nearest neighbor samples H_j from the samples of the same class R , and finds k nearest neighbor samples M_j from the samples of different class R . ($j = 1, 2, \dots, k$).
- (3) For $i = 1$ to N do (N is the number of runs)

$$W(A) = W(A) - \sum_{j=1}^k diff(A, R, H_j) / (mk) + \sum_{C \neq class(R)} \left[\frac{p(C)}{1 - p(class(R))} \sum_{j=1}^k diff(A, R, M_j(C)) \right] / (mk) \quad (2)$$

where $diff(A, R_1, R_2)$ represents the distance between sample R_1 and sample R_2 on the feature A ; $M_j(C)$ represents the j th nearest neighbor sample of class C ; and $p(C)$ represents the probability of the class C .

- (4) Sort the weight W values.

Step 3: Calculate the weight value for each feature.

The feature weights obtained by ReliefF all belong to the interval $[-1, 1]$. The greater the weight, the more important the feature is. Figure 1 shows the importance and ranking of each factor. This study reduced the complexity of the model by keeping important variables and removing variables that contribute less or are not useful. The number of final features was 15, including x_{12} , x_{13} , x_9 , x_5 , x_1 , x_8 , x_{35} , x_{34} , x_{20} , x_{11} , x_{33} , x_{31} , x_6 , x_{14} , and x_{17} .

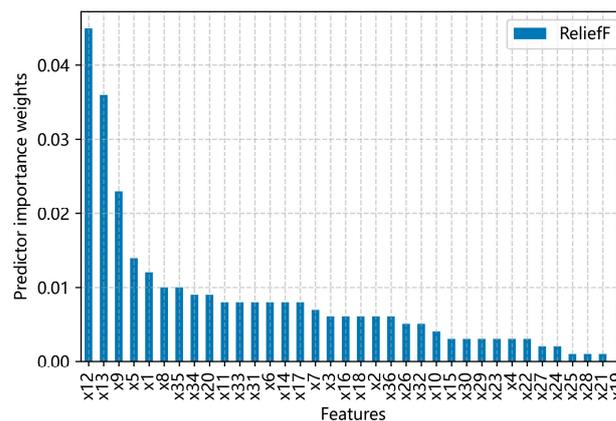


Figure 1. Importance of feature variables after ReliefF.

- (2) A resampling technique combining undersampling and oversampling was presented.

Hybrid sampling is a means to remove the noise samples in the imbalanced problems, but the combination of noise filtering technology can also eliminate the wrong samples [45]. For the problem of synthesizing sample errors, Tomek-links was used to eliminate outliers, and it calculated that the samples with the closest distance between classes form a Tomek chain. Then, according to the Tomek chain, the minority class samples located at the

boundary between classes were identified, and the overlapping samples were deleted. Next, the adaptive semi-supervised clustering in ASUWO was used for oversampling, so that the data distribution was basically balanced while reducing the number of noisy samples [46]. By combining labeled and unlabeled data for sample selection, ASUWO can learn more characteristics of the data, and it has better balance compared with SMOTE.

Tomek-links refers to the nearest neighbors of each other in two samples from different classes, providing a better decision boundary for the classifier. If there are two samples x and y belonging to different classes, the distance $d(x, y)$ is calculated. At this time, if another sample z cannot be found, so that the distance from any sample to z is smaller than $d(x, y)$, it is called a Tomek pair. For each Tomek pair, one of the samples is removed.

The WTASUWO algorithm (Algorithm 1) is a combination of ReliefF, Tomek-links, and ASUWO, aiming to improve the prediction performance through feature weighting, denoising undersampling, and oversampling. The ReliefF feature weights reflect the importance of each feature for classification. The Tomek-links algorithm removes samples that are more similar to the nearest neighbor samples of its own category by comparing the similarity between the samples, which reduces the effect of noisy samples on classification. The ASUWO algorithm removes those samples with low affiliation, which allows the classifier to be trained with more samples, thus improving the prediction performance.

Algorithm 1. WTASUWO

Inputs: training dataset, category labels, number of nearest neighbor samples.

Output: filtered and weighted balanced dataset.

1: Initialization: number of iterations, feature weights, undersampling rate, etc.

2: For each sample

a. Feature weight calculation: ReliefF is used to calculate the feature weights.

b. Undersampling processing: based on the weighted data, the Tomek-links algorithm is used for undersampling to remove noisy samples.

c. Oversampling processing: based on the weighted and noise denoised feature data, the ASUWO algorithm is used for oversampling to increase the number of samples in minority classes.

3: Output the filtered and weighted training dataset.

3.3. SRH Prediction Model of Older Adults Based on IMLP

In general, traditional statistical methods rely on specific data distribution assumptions and have poor prediction performance when considering high-dimensional and large-scale unknown data. As for MLP, it is a type of NN that describes a complex mapping between a set of input variables and an output variable. This technique is suitable for solving the multi-classification problem of nonlinear fractional data. This feature makes the MLP model highly valuable and promising in the prediction studies. Moreover, Figure 2 illustrates the main structure of the network model, divided into input, output, and hidden layers, where the layers of MLP are fully connected. The numbers below Figure 2 represent the number of nodes in the different layers. As for the number of selected features, it represents the number of nodes in the input layer. Moreover, the determination of the number of nodes in the hidden layer is described in Section 3.3, whereas the number of nodes in the output layer represents the classified health status category.

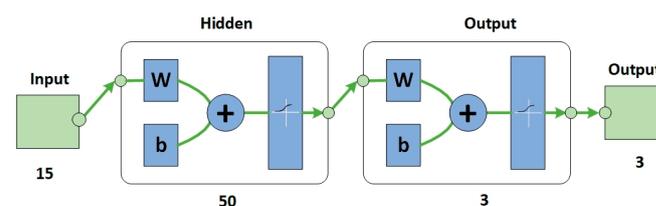


Figure 2. Basic structure of the MLP model.

The construction process of MLP can be divided into forward propagation and back propagation. Forward propagation calculates the outputs of the nodes through the weights and thresholds of the previous iteration. Back propagation was used for parameter training. It is a process of calculating the influence of each weight and threshold on the total error from the output layer, and then adjusting the weights and thresholds to minimize the error.

In the forward propagation, the feature vector of the input layer is set as x_{ij} . The weight and threshold values from the input layer to the hidden layer are represented by w_{ij} and b_j . The output value of the nodes in the hidden layer is S_j . M, N , and K are the number of nodes in the input, hidden, and output layers, respectively. L is the number of samples.

$$S_j = f\left(\sum_{i=1}^M w_{ij}x_{ij} + b_j\right), j = 1, 2, \dots, N \tag{3}$$

where $f(\cdot)$ represents the activation function, and the sigmoid function is generally used, as in Equation (4). After obtaining the output values of the hidden layer, they are calculated as the input values of the output layer. The network results are shown in Equation (5).

$$f(x) = \frac{1}{1 + e^{-x}} \tag{4}$$

$$\hat{y}_k = f\left(\sum_{j=1}^N w_{jk}x_{jk} + b_k\right), k = 1, 2, \dots, K \tag{5}$$

In back propagation learning, \hat{y}_k is the output signal of the network, \hat{y}_l is the predicted value of the network, and y_l is the true value. The purpose of NN learning is to reduce the error between \hat{y}_l and y_l to the set accuracy, and the commonly used error function E is the mean square error (MSE), expressed as follows:

$$E = \sqrt{\frac{1}{L} \sum_{l=1}^L (\hat{y}_l - y_l)^2 + \lambda \sum |w|} \tag{6}$$

where λ is a regularized coefficient that represents the degree of constraint on the function space. When λ is large, it limits the function space to be small; when λ is small, it allows the function space to become large. $\lambda \sum |w|$, also known as structural risk, is the sum of the absolute values of the weights w in the model. The commonly used $\sum |w|$ include L_1 and L_2 norms. With this item, the function space corresponding to the MLP can be controlled more flexibly, thus better avoiding the problem of overfitting.

The gradient descent method was used to update the weight parameters of the hidden and output layers. If the gradient of w is 0, that is, $= 0$, then the next weight w is adjusted, as in Equation (7).

$$\begin{cases} w \leftarrow w + \Delta w \\ \Delta w = -\eta \frac{\partial E}{\partial w} \end{cases} \tag{7}$$

where η is the learning rate, and Δw represents the magnitude of the gradient. If the output value \hat{y} is equal to the expected output y , w does not change. Otherwise, adjust w until the error is less than a certain value to end the training.

However, there still exists some problems with MLP, such as the possibility of disappearing gradients or gradient explosions when dealing with the complex tasks. Meanwhile, using NNs to solve classification problems requires to set values for the model hyperparameters. These hyperparameters include the training algorithm, the initial learning rate, the number of hidden layers, the number of hidden layer neurons, the activation function of each layer, and the training times. To improve the generalization ability of the network, some regularization methods are also deployed to avoid overtraining. The improved MLP (IMLP) consists of the following:

- (1) Design the network structure: number of nodes.

The input layer serves as a buffer, integrating data sources into the network. In this model, 15 feature dimensions are derived through feature selection. When a network with a hidden layer includes enough neurons, it can effectively approximate a complex system with arbitrary precision. However, increasing the number of hidden layers may extend the computing time and lead to potential overfitting problems. Hence, for simplicity and efficiency, the model with only one hidden layer is considered [47]. The utilization of an implicit layer enables a multilayer perceptron with nonlinear mapping capability, feature extraction, and abstract representation; however, this configuration will reduce the model complexity and improving training efficiency. Moreover, a NN with one hidden layer with a sufficient number of hidden neurons holds the capability to approximate any continuous function.

Few nodes in the hidden layer are not able to fully express the nonlinear relationship in the system, and many nodes may cause overfitting. Through the application of different number of nodes, Figure 3 demonstrates the accuracy of the model corresponding to the number of nodes in the hidden layers. Moreover, referring to Figure 3, the model has the optimal performance when the number of nodes in the hidden layer reaches 50. The number of neurons in the output layer represents the number of output variables (SRH), so the number of neurons in the output layer is three.

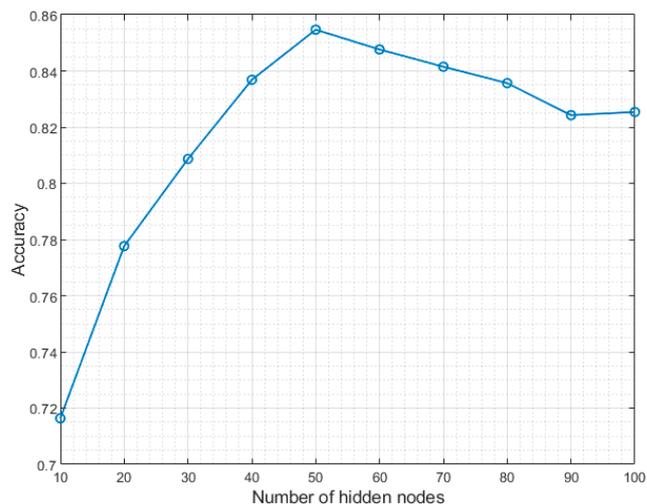


Figure 3. Accuracy of the model corresponding to the number of nodes in the hidden layers.

In this paper, defining the number of hidden nodes is an empirical process in designing the network structure [48]. Initially, the NN structure is evaluated based on the problem and data characteristics. When building a NN, a small number of hidden nodes can be set, ensuring that the network has a certain learning ability but may not be sufficient to solve the complex problem. Therefore, the range is set between 10 and 100, and the number of nodes gradually increases using a step of 10 to observe the performance variations of the network. As the number of nodes in the hidden layer increases, the accuracy of the network progressively improves. When it reaches a certain level, the performance of the network becomes invariant, and thus the optimal network structure will be suitable for the problem.

(2) Setup parameter optimization: ensemble learning and learning factors.

Bagging is a method applied to improve the accuracy of weak classifiers by constructing a series of prediction functions. They are then combined into a single function using a certain way. Bagging is suitably combined with unstable learning algorithms, such as NNs and DTs. By voting on these classifiers, weak classifiers are trained to generate stronger classifiers, as illustrated in Figure 4. Considering the instability of the model, bagging is applied to build a global model, generating multiple model sequences to improve the prediction accuracy of MLP [49]. At the same time, using Dropout to improve the training

process helps in solving the overfitting problems. As a result, the probability value of discarded neurons is set to 0.2 and the K-fold cross-verification is applied for training.

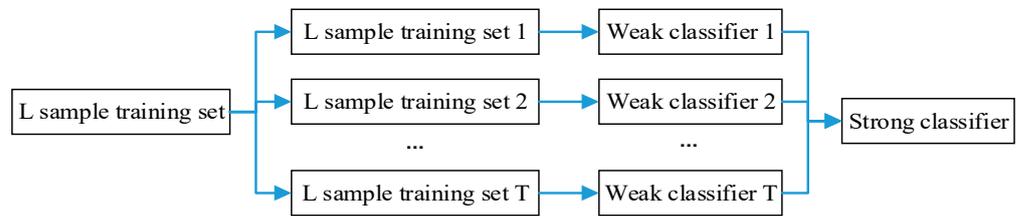


Figure 4. Bagging-based network model training architecture.

For this model, the learning factor is specified according to experience. If the learning factor is too small, the training speed may be slow. If it is too large, it may cause oscillations, and the error function will not converge during training. Therefore, the attenuation factor was introduced to dynamically adjust the learning rate in the training process of the model. When the error tends to the target in a reduced way, the learning rate can be increased by correcting the direction. When the error increases beyond a certain range, it indicates that the previous step correction is incorrect, and the step size should be reduced. The learning rate updating method is shown in Equation (8). The connection weights of different neurons are dynamically adapted to the learning rate to ensure that the optimization in each direction can find the optimal solution, in order to maximize the adjustment rate of weights and speed up the convergence of the model.

$$\eta = \begin{cases} a \cdot t, & t < T' \\ b \cdot e^{-r \cdot t} - c, & t > T' \end{cases} \tag{8}$$

where a , b , and c are constants; r is the attenuation rate; t is the number of iterations; and T' is the default number of iterations. The learning rate change curve of this method is shown in Figure 5, which shows the changes of fixed, linear, exponential decay, and adaptive learning rates. At the beginning of the iteration, the learning rate increases, while in the later stages of the iteration, it decreases exponentially.

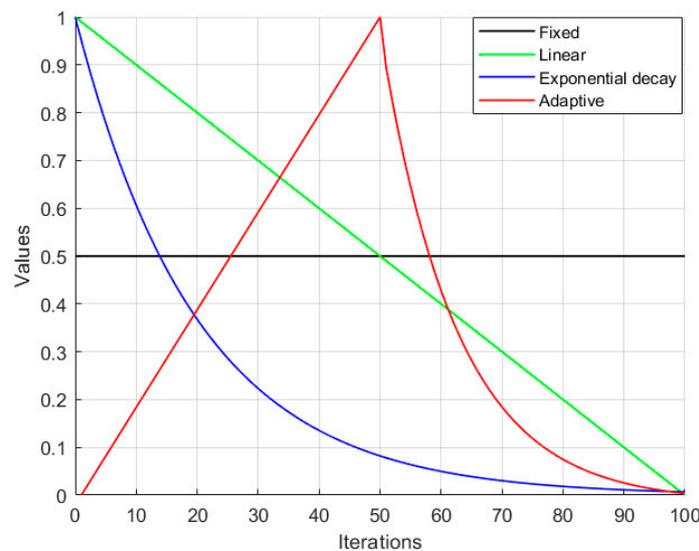


Figure 5. Change curves for different learning rate methods.

3.4. Overall Flowchart

The proposed method (e.g., WTASUWO-IMLP (Algorithm 2)) mainly includes feature extraction, undersampling, oversampling, and classification algorithms. ReliefF is first

applied to obtain the best feature subset to simplify the input values. Moreover, Tomek-links and ASUWO are used to construct the overall distribution of the data covering the learnable data. Strategies are provided to sample both minority and majority classes to reach data balance. Finally, the synthesized samples are applied to train the classifier. In the modeling of IMLP, it is necessary to clean and normalize the original dataset; consequently, these data are divided into the training and test datasets after being preprocessed. The training set is imported, and the network structure and optimized parameters are determined. Finally, the test dataset is input to verify the accuracy of the model. The modeling process of the SRH prediction of older adults is displayed in Figure 6.

Algorithm 2. WTASUWO-IMLP

Input: original dataset.

Output: Prediction results.

- 1: Preprocess the original dataset, including data cleaning and normalization.
 - 2: Using the ReliefF algorithm to obtain the optimal feature subset and weights.
 - 3: Using Tomek-links and ASUWO algorithms to undersample and oversample the training set, generating a new training set.
 - 4: Divide the new dataset into training and testing sets.
 - 5: Use the IMLP algorithm to train the model using the new training set.
 - a. Determine the network structure and optimization parameters.
 - b. Using the new training set to train models: forward propagation, computational loss, and backpropagation.
 - 6: Verify the accuracy of the model using the test set.
 - 7: Generate the prediction results.
-

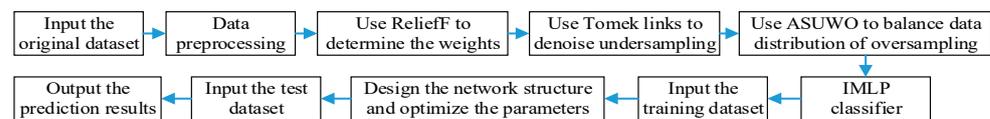


Figure 6. Flowchart of the SRH prediction model of older adults.

The steps of the WTASUWO-IMLP algorithm are as follows:

Step 1: Data preprocessing: data cleaning and normalization is performed on the original dataset to remove noise and outliers and to maintain the data at the same scale;

Step 2: Feature selection: the ReliefF algorithm is used to select the optimal feature subset from the original features based on the weights to simplify the input values;

Step 3: Data sampling: the Tomek-links and ASUWO algorithms are used to under-sample and oversample the training set for data balancing;

Step 4: Data partitioning: the preprocessed dataset is partitioned into training set and test set, where the former is used to train the model and the latter is applied to verify the accuracy of the model;

Step 5: Model training: on a new training set, the IMLP algorithm is applied to model training, including determining network structure and optimizing the parameters;

Step 6: Model validation: use the test set as input to verify the accuracy of the model;

Step 7: Output results: output the trained classifier to predict SRH for older adults.

4. Results

In this section, from the perspective of algorithm performance analysis, including the determination of the base classifier and feature selection, performance analysis of the oversampling and NNs, and significance analysis of the proposed method are verified based on the empirical 2020 CFPS dataset. In addition, we also conducted a robustness analysis of the proposed method on the public datasets to further validate its generalization ability and stability.

4.1. Evaluation Indicators

When dealing with the imbalanced data, it is necessary to select the evaluation indicators, which are usually based on the confusion matrix. In this paper, accuracy, precision, recall, specificity, area under the receiver operating characteristic (AUC), and running time (RT) were used to evaluate and compare the performances of different classifiers [50,51]. Due to the research on the health status prediction, it is a labeled classification problem based on the characteristics of the dataset. These metrics can evaluate the performances of classifiers from different perspectives and provide a reference basis for model optimization. For example, accuracy, precision, and recall are concerned with the ability of the classifier to recognize the positive class samples, specificity is concerned with the ability of the classifier to recognize the non-positive class samples, and AUC and RT are concerned with the overall performance and efficiency of the classifier.

(1) Accuracy: the ratio of the number of correctly classified samples to the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

(2) Precision: the proportion of samples classified as positive samples that are true positive samples.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

(3) Recall: the proportion of positive samples in the total sample that are predicted correctly.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

(4) Specificity: the proportion of negative samples in the total sample that are predicted correctly.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (12)$$

(5) The receiver operating characteristic (ROC) curve is usually presented in the case of cross, where the advantages and disadvantages of the classifiers cannot be directly distinguished [52]. AUC can be used as an extension of the judging basis of ROC. The larger the AUC value, the higher the accuracy and the better the classifier.

$$\text{AUC} = \frac{\sum_{i \in PC} \text{rank}_i - \frac{M(1+M)}{2}}{M \times N} \quad (13)$$

where M is the number of positive samples; N is the number of negative samples; rank_i is the sequence number of sample i ; PC indicates the positive samples; TP is the positive samples that are predicted to be positive; TN is the negative samples that are predicted to be negative; FP is the negative samples that are predicted to be positive; and FN is the positive samples that are predicted to be negative. The values of the above five indicators range from 0 to 1. The higher the value, the better the performance of the model.

Experimental environment: Windows 10 64-bit operating system, processor AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz, RAM 16.0 GB (15.9 GB available). The programming languages were Matlab 2022a and Python 3.8.8. To objectively compare various methods and improve the generalization ability of the model, the fivefold cross-validation was used to obtain the results. A total of 70% of samples of the original dataset were used as the training dataset, and these samples and the synthesized samples were used to train the classifier, and the remaining 30% samples were used as the test dataset for model testing.

The concept definitions of the key terms are shown in Table 2.

Table 2. Concept definitions of the key terms in this paper.

Terms	Concepts
Older adults	Refers to adults aged 60 and above.
SRH	Refers to an individual’s SRH status, including “unhealthy”, “average”, and “healthy”.
Feature selection	The process of selecting a feature subset from the original features that are relevant and valuable to the current learning task.
Undersampling	A technique used to handle the imbalanced data by reducing the number of samples in the majority classes.
Oversampling	A technique used to handle the imbalanced data by increasing the sample size of the minority classes.
Neural network	A computational model that simulates the neurons for processing and analyzing large amounts of complex data.
Class imbalance	Refers to the fact that some categories in a dataset have far more samples than others.
Prediction	An assessment of unknown health status based on existing data or information.
Machine learning	Refers to the algorithms by which machines learn from large amounts of historical data and thus generate empirical models.
CFPS	A long-term and large-scale survey project to collect socio-economic information on Chinese households and individuals.
ReliefF	It is a feature selection algorithm used to identify features related to a specific classification.
Tomek-links	It is an undersampling method used to remove noise or duplicate samples by calculating the distance between samples.
ASUWO	An oversampling method for handling the imbalanced dataset by adaptively generating new samples in minority classes according to their weights.
MLP	It is a feed-forward ANN model that can simulate complex nonlinear mapping relationships.
Bagging	An ensemble learning method that constructs multiple base learners by randomly extracting different subsets of training data.
Classifier	An algorithm or model used to classify data into two or more categories.

The abbreviations of terms are shown in Table 3.

Table 3. Abbreviations of the key terms in this paper.

Terms	Abbreviations	Terms	Abbreviations	Terms	Abbreviations	Terms	Abbreviations
Self-rated health	SRH	Logistic regression	LR	Radial basis function	RBF	Running time	RT
China Family Panel Studies	CFPS	K-nearest neighbor	KNN	Self-organizing map	SOM	Receiver operating characteristic	ROC
Adaptive semi-supervised weighted oversampling	ASUWO	Decision tree	DT	Multiple-layers perceptron	MLP	Area under ROC	AUC
Weighted Tomek-links ASUWO	WTASUWO	Random forest	RF	Stochastic gradient descent	SGD	Extreme learning machine	ELM
Improved multi-layer perception	IMLP	Support vector machine	SVM	Deep learning	DL	Cuckoo search	CS
Artificial intelligence	AI	Adaptive boosting	AdaBoost	Adaptive moment	Adam	Simulated annealing	SA
Neural networks	NNs	Gradient boosting decision tree	GBDT	Genetic algorithm	GA	Firefly algorithm	FA
Synthetic minority oversampling technique	SMOTE	Extreme gradient boosting	XGBoost	Evolutionary algorithm	EA	Particle swarm optimization	PSO
Adaptive synthetic	ADASYN	Artificial neural network	ANN	Deep neural networks	DNNs	False positive rate	FPR
Support vector data description	SVDD	Back propagation	BP	Mean square error	MSE	True positive rate	TPR

4.2. Determination of The Base Classifier and Feature Selection

In this paper, some base classifiers were used to choose the algorithm that was suitable for the studied problem, including LR, KNN, DT, RF, SVM, AdaBoost, GBDT, XGBoost, BP, extreme learning machine (ELM), and MLP. The reasons for selecting these classifiers are as follows: (1) Different machine learning algorithms have their own characteristics and show different performances when dealing with the problem in this paper. (2) In practice, there is no absolute algorithm. The only way to find a suitable algorithm as a base classifier is for a specific problem and dataset. Meanwhile, the prediction of the health status involves complex nonlinear relationships and multidimensional features, which requires strong generalization ability and the ability to process the complex data.

NNs have an advantage over other traditional machine learning algorithms in handling nonlinear problems. Comparison of multiple algorithms can help us to find a more suitable model for the current problem to meet the application requirements. For each classifier, we trained and tested the same datasets to ensure the objectivity of the results, and calculated the indicators of accuracy, precision, recall, specificity, AUC, and RT. The evaluation results of different classifiers are shown in Table 4, where the bold values are the relative optimal values. The visual effect is shown in Figure 7.

Table 4. Prediction results of different base classifiers.

Classifiers	Accuracy	Precision	Recall	Specificity	AUC	RT
LR	0.571	0.504	0.553	0.432	0.538	0.082
KNN	0.527	0.447	0.477	0.564	0.619	0.131
DT	0.549	0.491	0.456	0.523	0.641	0.042
RF	0.428	0.428	0.423	0.565	0.684	1.021
SVM	0.552	0.488	0.581	0.355	0.737	2.552
AdaBoost	0.578	0.510	0.757	0.440	0.742	0.504
GBDT	0.591	0.525	0.661	0.536	0.728	37.544
XGBoost	0.581	0.516	0.670	0.511	0.722	0.529
BP	0.575	0.514	0.564	0.583	0.690	128.548
ELM	0.588	0.524	0.599	0.515	0.669	0.051
MLP	0.602	0.538	0.660	0.606	0.701	10.178

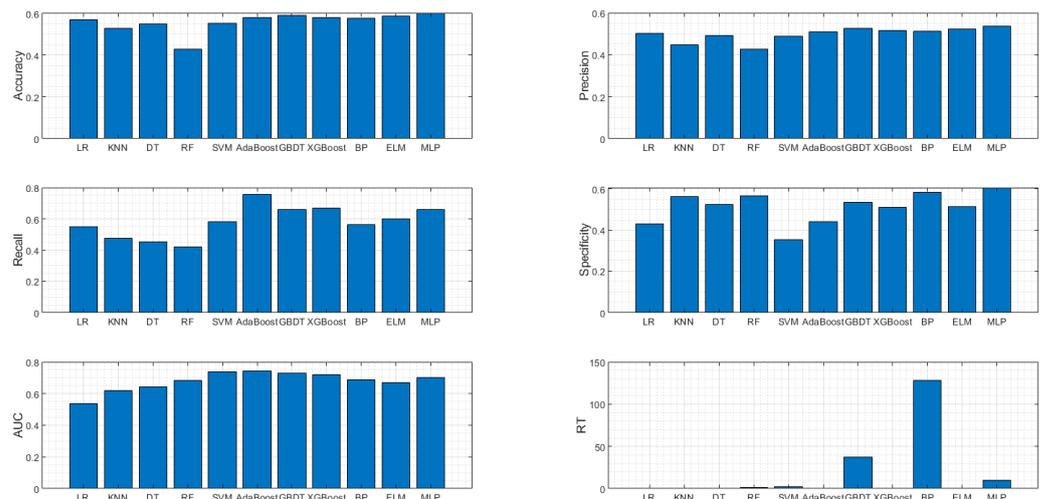


Figure 7. Comparison of the prediction results based on different classifiers.

From Table 4 and Figure 7, by comparing and evaluating these classifiers, the MLP model with the best performance to solve the studied problem was selected. LR is easy to implement and interpret, and it performs well for linear problems, but it is susceptible to overfitting. KNN does not need to transform the feature of the data, which is good for the nonlinear problems, but it is sensitive to the noise samples. DT is sensitive to feature

selection, it can process the missing values and class features. RF is robust to noise and missing values, but it can lead to feature selection bias. SVM performs well for linear problems and has certain advantages in solving samples with small sample size, as well as nonlinear and high-dimensional datasets. AdaBoost has a strong generalization ability, but it is easily affected by noise samples. GBDT can combine multiple weak learners, which has strong robustness to noise, but it has high computational complexity. XGBoost has strong robustness to missing values and noise and high computational efficiency, but it is easily affected by overfitting. BP can deal with nonlinear problems and has strong generalization ability, but it is sensitive to the initial weights. ELM has high computational efficiency and is not sensitive to the initial weights, but it is easily affected by noise and has poor performance for linear problems. MLP is a deep learning model with multiple hidden layers. By learning the nonlinear network structure, potential feature rules were obtained from the dataset, nonlinear activation functions were used to improve the expression ability of the model, and the model parameters were updated by the optimizer. It can learn the essential features of a dataset from the minority class samples.

However, too many features will increase the burden of the model. On the other hand, the invalid features will also decline the model performance, so some feature selection methods were used to select features. Feature selection methods can be divided into filter, wrapper, and embedded [53]. The difference lies in whether the feature selection process is combined with the specific classification learning methods. In this paper, different feature selection algorithms were compared, including variance selection and the chi-squared test in Filter, GA in Wrapper, tree-based model in Embedded, and Lasso and ReliefF machine learning algorithms. Based on the MLP classifier, the prediction performances of the feature selection algorithms were compared, which are shown in Table 5 and Figure 8, with the bold values being the relative optimal values.

Table 5. Prediction results of different feature selection algorithms based on MLP.

Feature Selection Methods	Accuracy	Precision	Recall	Specificity	AUC	RT
Variance	0.474	0.449	0.487	0.161	0.698	2.165
Chi-squared	0.439	0.439	0.435	0.321	0.696	1.020
GA	0.554	0.494	0.716	0.427	0.720	2.818
Tree-based	0.574	0.510	0.744	0.441	0.734	1.856
Lasso	0.585	0.522	0.659	0.528	0.723	9.937
ReliefF	0.596	0.527	0.731	0.567	0.735	5.509

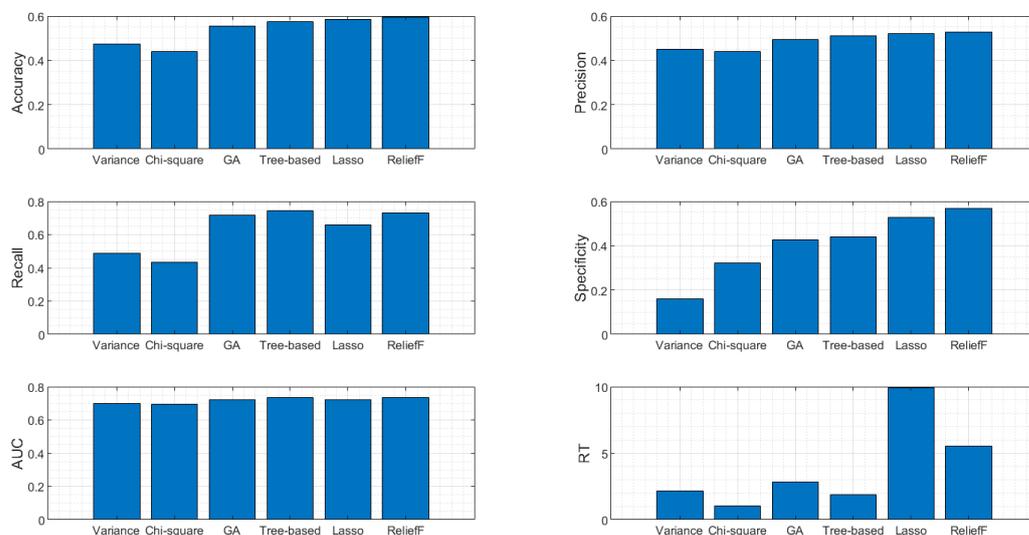


Figure 8. Prediction performances of different feature selection algorithms.

Filter is a preprocessing method that selects features based on their variance or correlation. The variance filtering method filters the features by the variance of the features. The chi-squared test is to determine whether there is a dependency between features by calculating chi-squared statistics. Wrapper selects features in the process of model training. It selects features according to the output of the model. GA is a common Wrapper method. It selects the optimal feature subset by simulating the process of natural selection and genetic evolution. Embedded is conducted by embedding feature selection into the process of model training. The tree-based model is conducted by constructing a decision tree for feature selection. Both Lasso and ReliefF methods conduct feature selection by limiting the complexity of the model. Lasso construes the complexity of the model by adding L1 regularization terms to select the relevant features. ReliefF selects features by calculating the relative importance of each feature, employing a new distance metric that is better able to process the high-dimensional data. As shown in Table 5 and Figure 8, ReliefF can select the most relevant features, thus improving the performance of the model.

4.3. Performance Analysis of The Oversampling Technique

Oversampling is a method to deal with the imbalance problem, which can reduce the imbalance among classes and improve the prediction performance of a classifier. In this paper, different oversampling techniques were compared, including random under-sampling, random oversampling, SMOTE, BorderLine SMOTE, SVM SMOTE, Safe-level SMOTE, ASUWO, SMOTE Tomek, WCNN ASUWO, and WTASUWO. Based on the base classifier and feature selection, the evaluation results of different oversampling algorithms are shown in Table 6 and Figure 9, with the bold values being the relative optimal values.

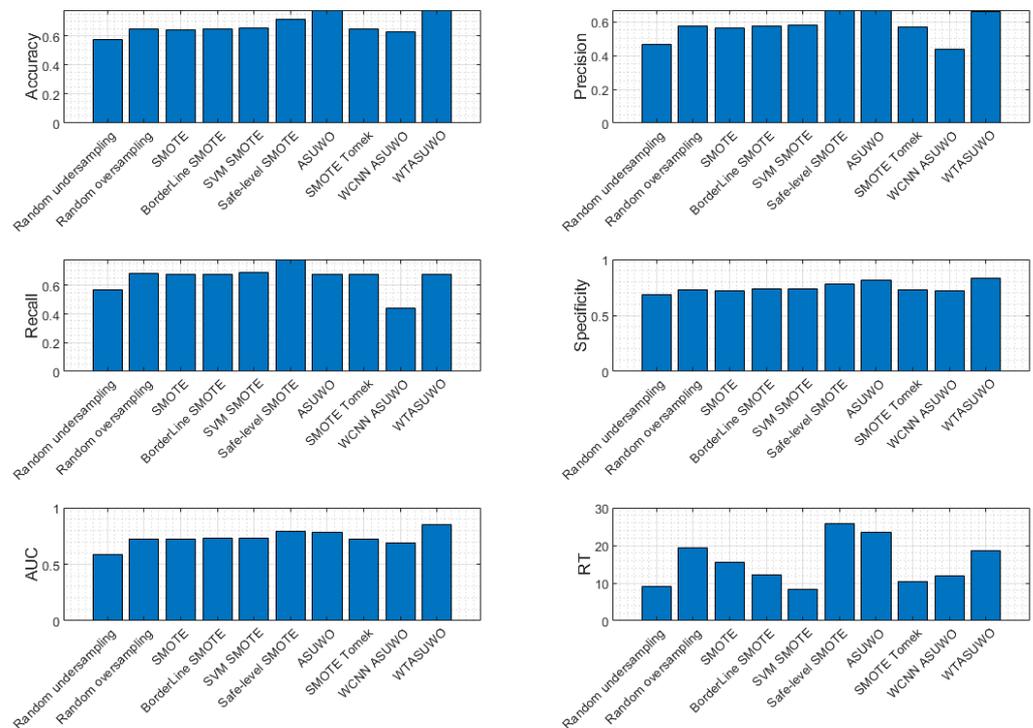


Figure 9. Comparison of the prediction performances of different oversampling algorithms based on MLP and ReliefF.

Table 6. Prediction results of different oversampling algorithms based on MLP and ReliefF.

Sampling Techniques	Accuracy	Precision	Recall	Specificity	AUC	RT
Random undersampling	0.576	0.465	0.565	0.682	0.588	9.108
Random oversampling	0.648	0.573	0.681	0.732	0.726	19.355
SMOTE	0.642	0.563	0.676	0.724	0.724	15.540
BorderLine SMOTE	0.650	0.575	0.676	0.737	0.729	12.246
SVM SMOTE	0.655	0.582	0.687	0.739	0.735	8.351
Safe-level SMOTE	0.714	0.670	0.780	0.782	0.794	25.781
ASUWO	0.773	0.667	0.674	0.813	0.785	23.337
SMOTE Tomek	0.646	0.570	0.679	0.730	0.726	10.434
WCNN ASUWO	0.626	0.439	0.439	0.719	0.690	11.996
WTASUWO	0.777	0.662	0.675	0.828	0.852	18.543

Random undersampling selects samples from the majority classes and deletes them. Random oversampling selects samples from the minority classes and copies them. These two methods are simple but can lead to data loss and redundancy. SMOTE is a common oversampling method, with good effect and high efficiency. BorderLine SMOTE finds the boundaries of minority class samples and then generates synthetic samples. SVM SMOTE uses the characteristics of SVM to oversample the minority class samples, which can improve the performance of the model. Safe-level SMOTE generates samples by introducing a safe distance. ASUWO adjusts the proportion of undersampling and oversampling according to the degree of imbalance to produce the synthetic samples. SMOTE Tomek is the combination of SMOTE and Tomek-links, which uses Tomek to oversample the minority class samples. WCNN ASUWO calculates the distance between samples by weighting and combines ASUWO to find the minority class samples. WTASUWO improves on ASUWO by first using weighted Tomek-links for undersampling, and then using ASUWO for oversampling to generate the most representative minority class samples. It should be noted that oversampling may increase the computational time, so it is necessary to choose the appropriate oversampling method according to the specific situation. After oversampling, the data distribution of each category will be more balanced, as shown in Figure 10.

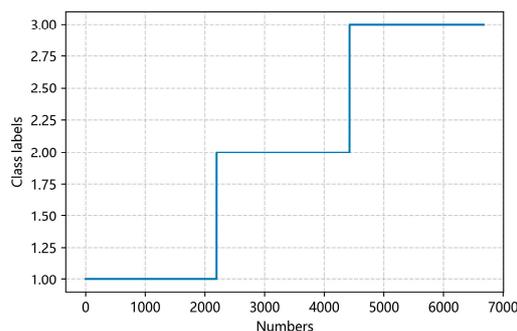


Figure 10. Distribution of data categories after oversampling based on WTASUWO.

4.4. Performance Analysis of The Neural Network Model

ELM and MLP have relatively strong representations that can fit complex nonlinear relationships, but they are also prone to overfitting and tuning parameters. The emergence of intelligent algorithms can be used to optimize the weights and bias parameters of NNs with better global search capabilities, which helps to improve the network performance. To verify the prediction performance of IMLP, it was compared with cuckoo search (CS)-based ELM, simulated annealing (SA)-based ELM, firefly algorithm (FA)-based ELM, particle swarm optimization (PSO)-based MLP, MLP using WTASUWO, and IMLP using WTASUWO. It is mainly based on the ELM and MLP classifiers, which are used for the

improved algorithm analysis. The prediction results are shown in Table 7 and Figure 11, with the bold values being the relative optimal values.

Table 7. Prediction results of the improved classification algorithms based on WTASUWO.

Improvement Methods	Accuracy	Precision	Recall	Specificity	AUC	RT
CS-based ELM	0.671	0.507	0.511	0.752	0.750	11.580
SA-based ELM	0.643	0.465	0.474	0.727	0.713	23.901
FA-based ELM	0.651	0.476	0.482	0.735	0.727	4.785
PSO-based MLP	0.665	0.498	0.551	0.723	0.752	25.654
MLP using WTASUWO	0.777	0.662	0.675	0.828	0.852	18.543
IMLP using WTASUWO	0.855	0.742	0.745	0.871	0.892	29.970

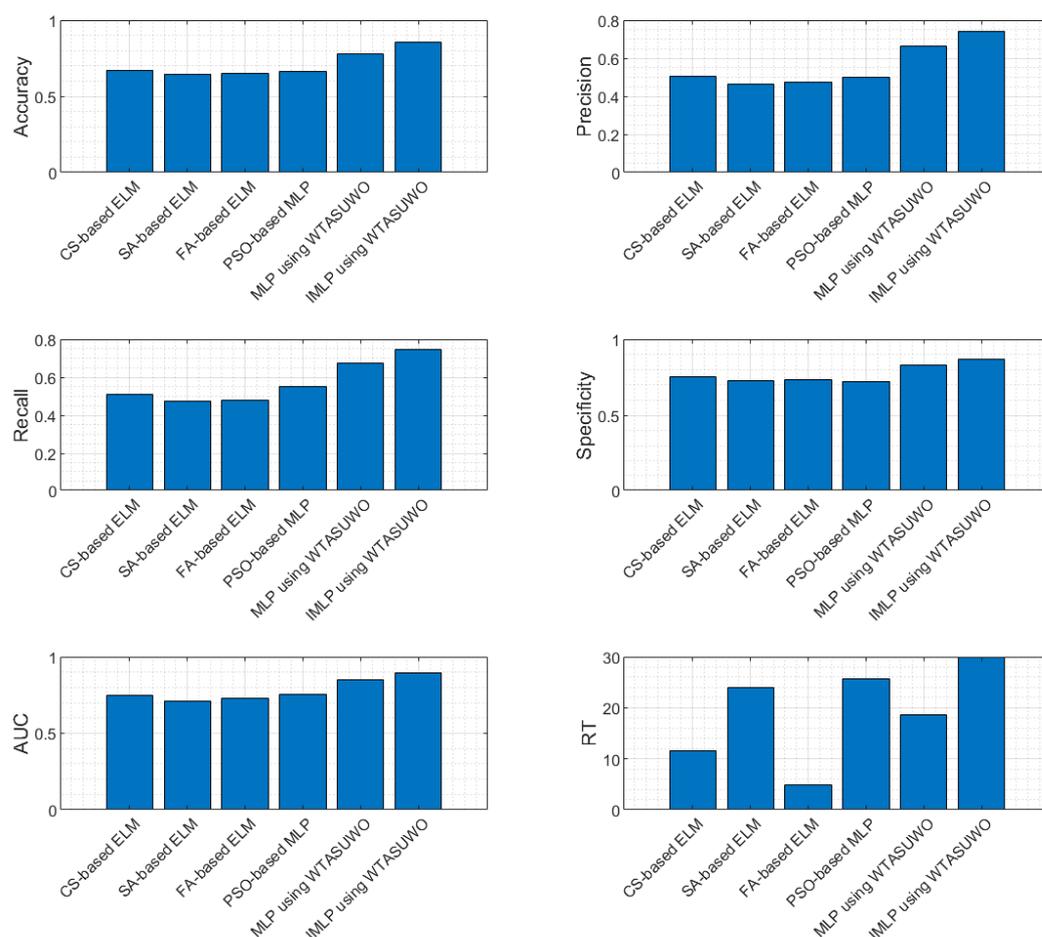


Figure 11. Comparison of the prediction performances of the improved classification algorithms based on WTASUWO.

MLP is a feedforward NN in which each neuron is connected to all neurons in the previous layer and nonlinearly transforms the input through an activation function. As shown in Table 7 and Figure 11, IMLP using WTASUWO had the best performance in various evaluation indicators, and the prediction accuracy was improved from 0.602 to 0.855. Followed by MLP using WTASUWO, it shows that the prediction performance of MLP was significantly improved. For the ELM and MLP algorithms optimized by intelligent algorithms, their prediction results were not as good as that of IMLP in this dataset. The reason may be because IMLP has a stronger generalization ability when dealing with such problems, or the WTASUWO oversampling strategy improves the performance of IMLP more significantly. At the same time, the RT is acceptable to the research problem in this paper.

CS, SA, FA, and PSO are optimization algorithms that find the optimal solution by training NNs such as ELM and MLP. These algorithms can optimize the model parameters. MLP is a supervised learning that improves the model performance by updating the parameters through back-propagation. In terms of performance, ELM is a fast-learning algorithm, while MLP requires many iterations to converge. In summary, the comparison between this method and other algorithms depends not only on the characteristics of the algorithm itself, but also on the specific application scenarios and problems.

Figure 12 shows the effect of the test process of IMLP using WTASUWO. ROC was used to evaluate the prediction performance of the model under different thresholds. The horizontal coordinate represents the false positive rate (FPR), and the vertical coordinate represents the true positive rate (TPR). The ideal ROC curve should be TPR increasing with FPR until it reached the maximum.

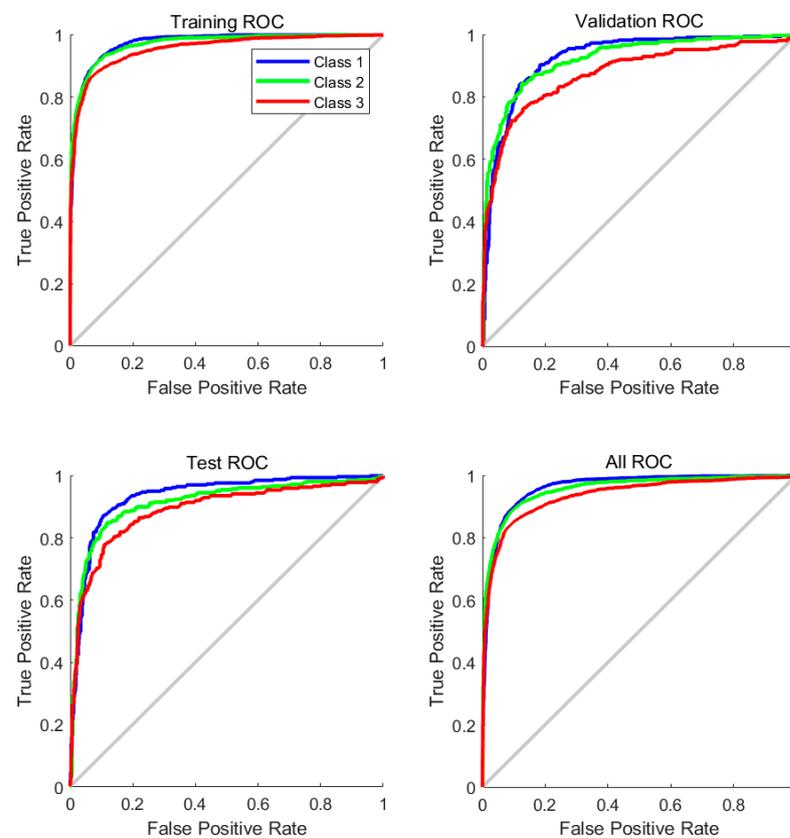


Figure 12. ROC curves with the adaptive learning rate based on IMLP using WTASUWO.

4.5. Significance and Robustness Analysis of WTASUWO-IMLP

(1) Statistical significance tests of IMLP using WTASUWO.

Due to the test errors between different algorithms, it is necessary to conduct the statistical significance tests on the accuracy of the six algorithms. In this paper, we used the Friedman rank test, which is a nonparametric post method that can be used to test whether there are significant differences between the algorithms. This method does not require the assumption that the data has a normal distribution or variance-aligned [54].

Null hypothesis: CS-based ELM, SA-based ELM, FA-based ELM, PSO-based MLP, MLP using WTASUWO, and IMLP using WTASUWO have the same distribution, meaning there is no significant difference between the different algorithms. These algorithms were tested in 10 experiments, with each training and test dataset randomly distributed. The Friedman test results are shown in Table 8.

Table 8. Friedman results between different algorithms and IMLP using WTASUWO.

	CS-Based ELM	SA-Based ELM	FA-Based ELM	PSO-Based MLP	MLP Using WTASUWO
Significance	0.012	<0.001	<0.001	0.028	0.252

The Friedman test results ($p = 0.000$) indicate that there were at least two of the different algorithms are significantly different from each other ($p < 0.001$), and the multiple comparison test (Nemenyi) can be further used to determine which algorithms were significantly different. Moreover, it can be seen from Table 8 that IMLP using WTASUWO was significant compared to CS-, SA-, FA-based ELM, and PSO-based MLP. The Nemenyi test was used to calculate the critical domain for the difference in the average order value. $CD = q_\alpha \sqrt{\frac{k(k+1)}{6 \times N}} = 2.38$, where k is the number of algorithms, N is the number of experiments, and q_α is a constant under the significance α . CD plots were used to represent the performance differences between the algorithms. Friedman average order values of different algorithms are shown in Figure 13.

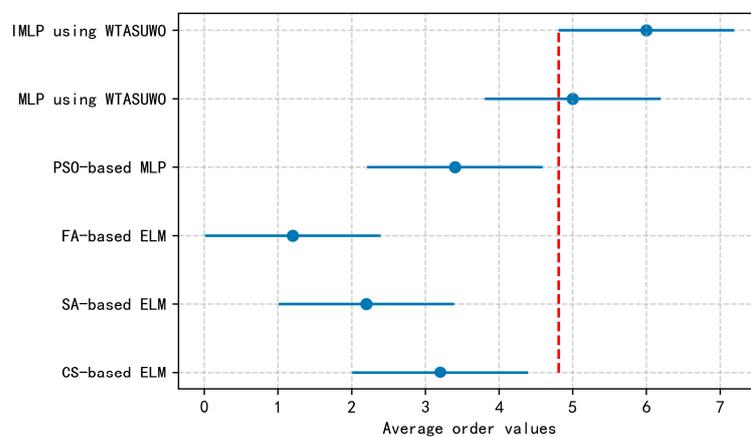


Figure 13. Friedman average order values of different algorithms.

In Figure 13, the vertical axis shows the comparison algorithms, and the horizontal axis shows the average order values. For each algorithm, a dot is used to show its average order value, and a horizontal line segment centered in the dot represents the size of the critical value domain. If the segments of the two algorithms overlap, it indicates that there is no significant difference in performance between the two algorithms within the confidence interval. Otherwise, it means that the performance is significantly different. From this figure, there is shown to be no significant difference between IMLP using WTASUWO and MLP using WTASUWO because their horizontal segments have overlapping areas. It also illustrates the significant improvement in the prediction performance of the classifier with the WTASUWO oversampling technique. In addition, IMLP using WTASUWO was found to be significantly superior to the CS-, SA-, FA-based ELM, and PSO-based MLP algorithms because their horizontal segments did not overlap. Therefore, from the results of the statistical significance tests, the method of this paper is significantly better than other algorithms.

(2) Robustness analysis of WTASUWO-IMLP.

To evaluate the robustness of the proposed method (WTASUWO-IMLP), and considering the limited number of samples, we used four imbalanced datasets with different ratios in the University of California Irvine (UCI) machine learning repository (<http://archive.ics.uci.edu/>, accessed on 12 October 2022). The characteristics of these datasets are summarized in Table 9. For each dataset, the data were divided into the training and test datasets, and they were preprocessed. Based on the data before and after oversampling, the proposed method was compared with the existing methods. The results are shown in Tables 10–13, and in Figures 14–17.

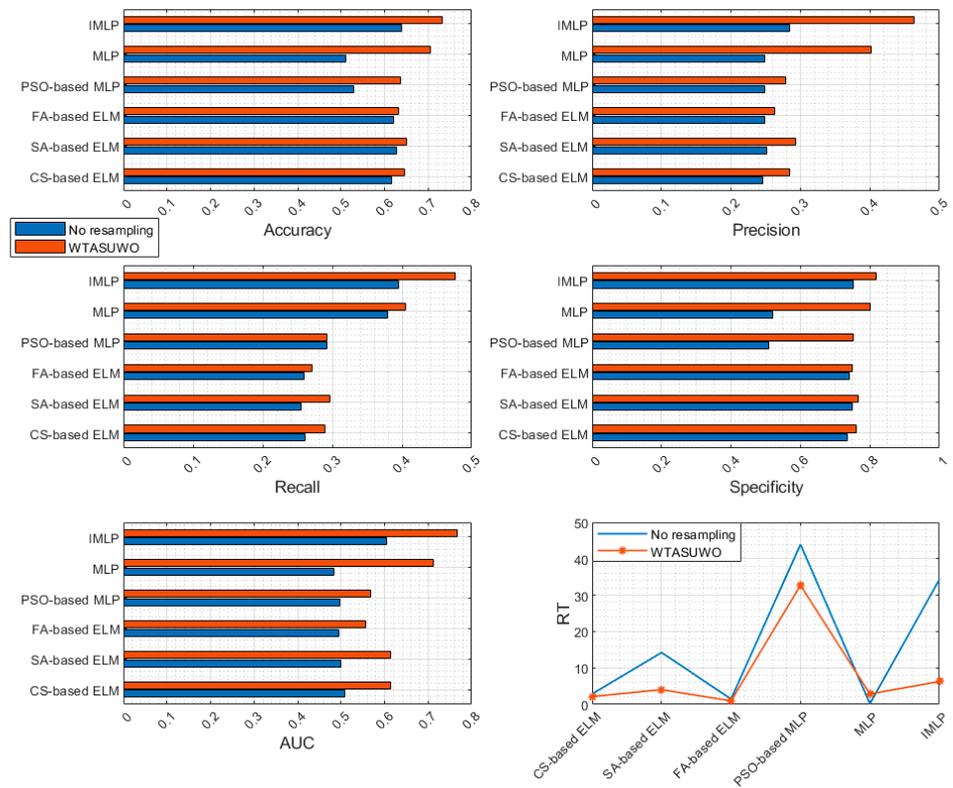


Figure 14. Comparison of the prediction performances of the proposed classification method based on the HCV dataset.

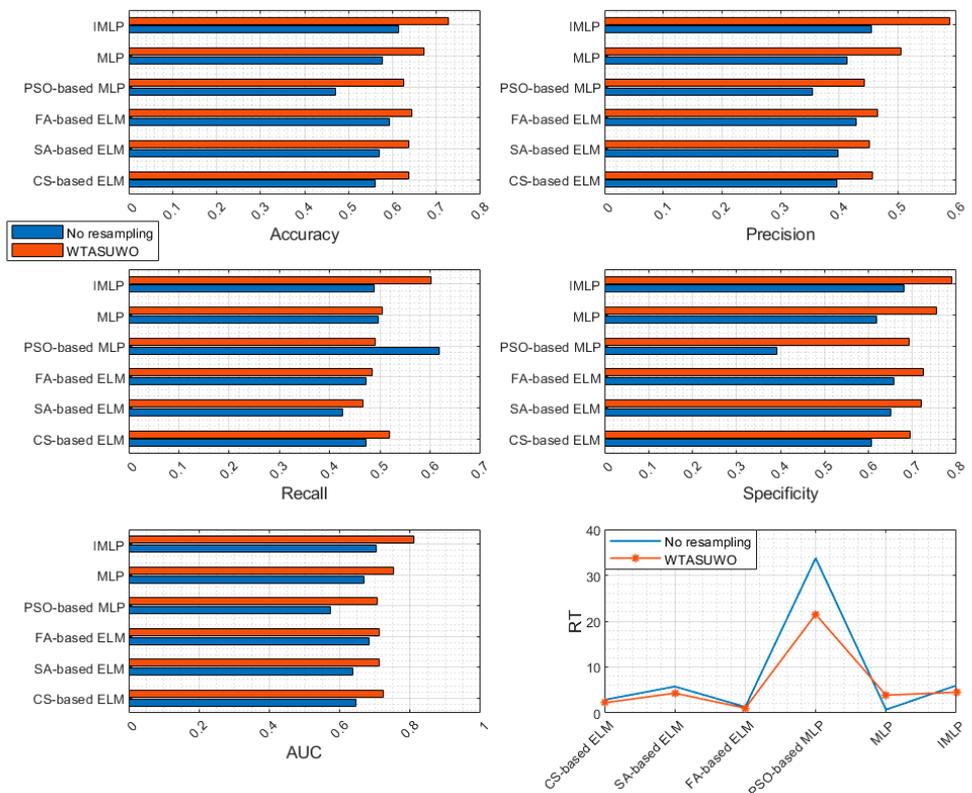


Figure 15. Comparison of the prediction performances of the proposed classification method based on the CM dataset.

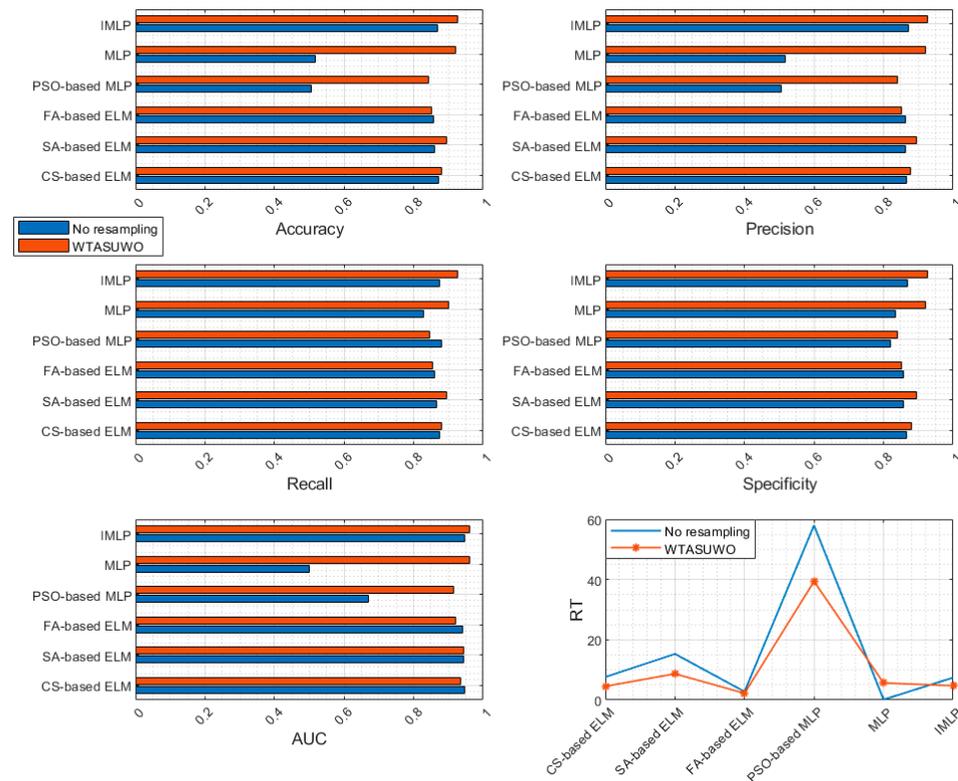


Figure 16. Comparison of the prediction performances of the proposed classification method based on the Rice dataset.

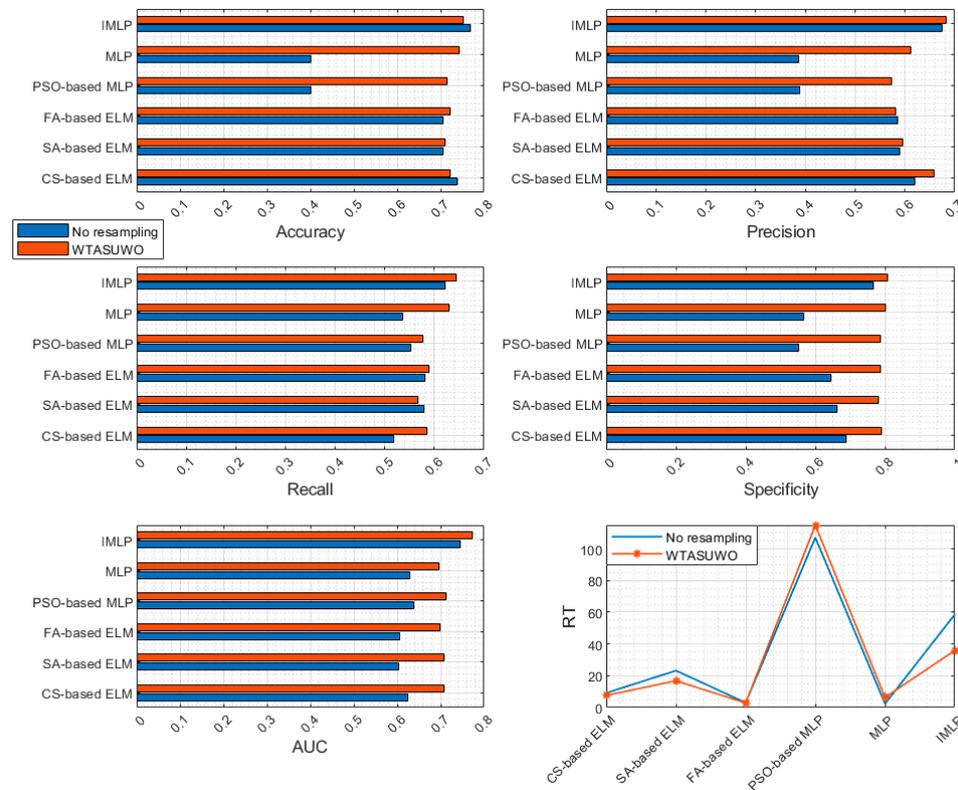


Figure 17. Comparison of the prediction performances of the proposed classification method based on the PS dataset.

Table 9. Information of the UCI datasets.

	Hepatitis C Virus (HCV) for Egyptian Patients	Contraceptive Method (CM) Choice	Rice (Cammeo and Osmancik)	Predict Students' (PS) Dropout and Academic Success
Instances	1385	1473	3810	4424
Attributes	28	9	7	36
Classes	4	3	2	3
Distribution	336/332/355/362	629/333/511	1630/2180	1421/794/2209
Imbalance ratio (IR)	1.01:1:1.07:1.09	1.89:1:1.53	1:1.34	1.79:1:2.78
Subject area	Life science	Life science	Computer science	Other
Distribution after WTASUWO	210/219/207/207	336/328/313	1331/1348	1253/1282/1286
Attributes after WTASUWO	10	8	7	24

Table 10. Prediction results of the oversampling and classification algorithms based on the HCV dataset.

Oversampling	Methods	Accuracy	Precision	Recall	Specificity	AUC	RT
No resampling	CS-based ELM	0.617	0.246	0.261	0.735	0.508	2.840
	SA-based ELM	0.627	0.252	0.255	0.750	0.499	14.218
	FA-based ELM	0.621	0.248	0.259	0.740	0.494	1.437
	PSO-based MLP	0.529	0.249	0.292	0.508	0.498	43.911
	MLP	0.510	0.249	0.379	0.521	0.484	0.162
	IMLP	0.638	0.284	0.395	0.752	0.604	34.324
WTASUWO	CS-based ELM	0.645	0.285	0.289	0.762	0.613	2.142
	SA-based ELM	0.650	0.293	0.296	0.766	0.613	4.026
	FA-based ELM	0.632	0.263	0.271	0.750	0.557	0.954
	PSO-based MLP	0.637	0.279	0.292	0.751	0.568	32.699
	MLP	0.704	0.402	0.405	0.802	0.713	2.848
	IMLP	0.732	0.464	0.477	0.817	0.766	6.267

Table 11. Prediction results of the oversampling and classification algorithms based on the CM dataset.

Oversampling	Methods	Accuracy	Precision	Recall	Specificity	AUC	RT
No resampling	CS-based ELM	0.560	0.396	0.473	0.608	0.647	2.854
	SA-based ELM	0.571	0.398	0.425	0.650	0.637	5.755
	FA-based ELM	0.592	0.429	0.473	0.657	0.683	1.312
	PSO-based MLP	0.470	0.354	0.618	0.391	0.574	33.743
	MLP	0.576	0.414	0.496	0.619	0.669	0.719
	IMLP	0.613	0.455	0.488	0.681	0.703	5.981
WTASUWO	CS-based ELM	0.636	0.457	0.518	0.695	0.724	2.218
	SA-based ELM	0.636	0.452	0.467	0.720	0.711	4.304
	FA-based ELM	0.645	0.466	0.485	0.725	0.713	1.047
	PSO-based MLP	0.625	0.443	0.490	0.693	0.707	21.489
	MLP	0.672	0.505	0.505	0.755	0.753	3.870
	IMLP	0.727	0.588	0.601	0.790	0.810	4.515

Table 12. Prediction results of the oversampling and classification algorithms based on the Rice dataset.

Oversampling	Methods	Accuracy	Precision	Recall	Specificity	AUC	RT
No resampling	CS-based ELM	0.871	0.867	0.875	0.867	0.948	7.630
	SA-based ELM	0.861	0.862	0.866	0.856	0.944	15.299
	FA-based ELM	0.858	0.862	0.859	0.856	0.942	2.761
	PSO-based MLP	0.507	0.507	0.879	0.820	0.670	57.994
	MLP	0.516	0.516	0.828	0.833	0.500	0.114
	IMLP	0.870	0.871	0.875	0.869	0.948	7.424
WTASUWO	CS-based ELM	0.879	0.878	0.879	0.879	0.935	4.539
	SA-based ELM	0.896	0.896	0.896	0.896	0.945	8.694
	FA-based ELM	0.853	0.852	0.854	0.852	0.920	2.220
	PSO-based MLP	0.843	0.841	0.846	0.840	0.914	39.383
	MLP	0.921	0.921	0.902	0.921	0.960	5.647
	IMLP	0.927	0.927	0.927	0.927	0.962	4.717

Table 13. Prediction results of the oversampling and classification algorithms based on the PS dataset.

Oversampling	Methods	Accuracy	Precision	Recall	Specificity	AUC	RT
No resampling	CS-based ELM	0.738	0.620	0.518	0.687	0.625	9.171
	SA-based ELM	0.706	0.589	0.578	0.661	0.604	23.256
	FA-based ELM	0.705	0.585	0.580	0.645	0.606	3.037
	PSO-based MLP	0.402	0.389	0.552	0.553	0.639	107.070
	MLP	0.402	0.387	0.537	0.566	0.628	2.373
	IMLP	0.769	0.674	0.622	0.767	0.746	58.603
WTASUWO	CS-based ELM	0.723	0.658	0.586	0.791	0.707	7.629
	SA-based ELM	0.710	0.596	0.567	0.781	0.708	16.845
	FA-based ELM	0.723	0.582	0.590	0.788	0.699	2.898
	PSO-based MLP	0.716	0.573	0.576	0.786	0.712	114.901
	MLP	0.744	0.612	0.630	0.800	0.697	6.406
	IMLP	0.752	0.682	0.644	0.806	0.773	35.923

Firstly, for the oversampling methods, it can be seen from the results that compared with that without oversampling, the prediction performances of the classifiers based on each dataset were significantly improved after WTASUWO. The application of WTASUWO made the RTs of the classifiers relatively reduced, which shows that the algorithm not only improved the performances of the classifiers, but also had high computational efficiency. Secondly, for the comparison of different classification algorithms, IMLP had high prediction accuracy and strong generalization ability for solving the classification problems, and it was able to effectively identify different categories of samples. Based on the above results, the WTASUWO oversampling technique and IMLP model had better effects in improving the performances and efficiencies of the classifiers. In practical applications, WTASUWO can be used for oversampling, and then IMLP can be used for classification and prediction.

5. Discussion

In this section, we focus on exploring the SRH of older adults in China based on the CFPS dataset from the perspective of the key factors affecting their health status, as well as suggesting useful policy recommendations.

5.1. Discussion of Experimental Results

This study indicated that most older people rate their health as good, accounting for 59.25%. The older adults with unhealthy and average health status accounted for 24.36% and 16.39%, respectively. Through feature selection, the final 15 features were used as the influencing factors, containing health changes, chronic diseases, after-tax wage

income, living area, and so on. Taking these four key factors as examples, we conducted the chi-squared analysis on the SRH of older adults. The results are shown in Table 14.

Table 14. Chi-squared analysis of key factors on the SRH of older adults (%).

Features	Groups	Unhealthy	Average	Healthy	χ^2	<i>p</i> Values
Health changes	worse	42.66	20.01	37.33	1060.718	<0.05
	unchanged	10.78	13.89	75.33		
	better	15.09	13.48	71.43		
	unknown	50.00	50.00	0.00		
Chronic diseases	yes	45.48	16.78	37.75	665.251	<0.05
	no	16.21	16.25	67.54		
	unknown	33.33	0.00	66.67		
After-tax wage income	0–1000	27.20	16.61	56.19	170.695	<0.05
	1000–2000	16.00	24.00	60.00		
	2000–3000	16.67	20.83	62.50		
	3000–4000	12.90	22.58	64.52		
	4000–5000	20.00	15.00	65.00		
5000 or above	9.47	14.70	75.83			
Living area	urban	20.37	18.52	61.11	62.794	<0.05
	rural	28.21	14.34	57.45		

The difference analysis employs the hypothesis testing method to assess whether influencing factors can account for changes in the dependent variable. Given that the SRH and influencing factors are categorical variables, the chi-squared test was employed to analyze differences in the impact of key factors on the SRH of older adults. Referring to Table 14, it is evident that these features significantly affect the SRH of older adults. Health changes directly influence the SRH of older adults, with those experiencing chronic diseases exhibiting poorer health status. Additionally, a higher after-tax wage income correlates with better health status of older adults. Moreover, older adults living in urban areas have a higher level of SRH than those living in rural areas. The findings of this study conform with previous studies.

The second data analysis of the Population Health Index Survey (PHIS) verified that chronic diseases are associated with lower SRH [55]. The lower SRH may not solely be attributed to the severity of chronic diseases but rather to the disparity between actual and expected health status. Additionally, the disruptive perception of pain, dysfunction, and inconvenience of daily life stemming from chronic diseases exacerbates the SRH of older adults.

A study conducted in the UK identified the positive effects of income increasing on SRH, emphasizing that income stability is strongly correlated with SRH and overall well-being [56]. Individuals with higher income levels typically enjoy better access to health knowledge and healthcare services, and they adopt healthier lifestyles compared to those with lower incomes. This, in turn, contributes to better SRH among individuals with high incomes.

Consistent with prior evidence, it has been established that the SRH of individuals in urban areas tends to be better than those in rural areas. The discrepancy is attributed to the varying availability of sanitation, preventive medicine, and healthcare services [57]. Urban areas typically boast sanitary conditions, abundant medical resources, and advanced facilities, fostering a conducive environment for the health status of older adults. However, these relatively poor sanitary conditions and limited resources in rural areas contribute to lower SRH among older adults residing in those regions.

The Canadian Longitudinal Study on aging, which enrolled community-dwelling adults aged 45 to 85, investigated the intersectional impact of sex and social factors on SRH. The findings revealed that women were less likely to report poor SRH [58]. The sex difference can be clarified by the different consideration between men and women. In

the health assessment, men often concentrate on physical health, whereas women adopt a more comprehensive approach, considering factors such as physical health, function, and mental health.

5.2. Policy Implications

There are several factors affecting the SRH of older adults in China. Special attention should be paid to older adults with lower income, with chronic diseases, in rural areas, and in worse health status. Comprehensive and consecutive efforts based on different aspects are still necessary to improve the health status of older adults.

(1) Reinforce publicity and education to guide older adults to develop good living habits.

Healthcare institutions should establish health archives for older adults, especially for those with hypertension, diabetes, and other chronic diseases, helping to manage the health status of sensitive population and slow the development of chronic diseases. Meanwhile, older adults should keep healthy lifestyles, including regular physical exercise, balanced diet patterns, and positive social connections. It is important to keep good attitudes and strengthen the prevention of serious diseases to improve the SRH of older adults [59]. Through health promotion, the occurrence of chronic diseases for older adults can be effectively prevented and controlled.

(2) Establish a sound old-age security system to improve the income level of older adults with financial difficulties.

The economic living standard has a significant positive effect on older adults' SRH. The health status of such a population, with more adequate financial sources, is better. Therefore, the social pension security system should be established and improved in line with the national conditions, and financial assistance to older adults with economic difficulties should be provided. The government can increase the transfer income of older adults with poor family status based on income redistribution or encourage social organizations to provide help for poor older adults.

(3) Improve the level of medical and health services through the government's investment and the joint efforts of the society.

Medical conditions can affect the health status of older adults, as the SRH of older adults in urban areas is significantly higher. The primary medical institutions in rural areas should make full use of the existing health resources by improving the community care and health network and helping in the enhancement of the health status of older adults in rural areas. Therefore, the investment in medical resources should increase and their distribution should expand. Moreover, a multi-level and rationally distributed medical and health service system should be developed to ensure that older adults with diseases can receive timely and effective treatment.

6. Limitations

The proposed method has advantages in the accuracy and stability of prediction. However, it still has some limitations that are described here below:

(1) Sample size and research perspective.

Research on the prediction of the health status of older adults necessitates a large amount of sample data to ensure the stability and accuracy of the model. Furthermore, the SRH of older adults is influenced by various factors, such as physical, psychological, and social aspects. We applied seven aspects and fail to cover the whole data information to explore the interrelationship between these factors. The diversity of sample data and the generalization ability of the model still requires to be further improved.

(2) Model optimization perspective.

Although NNs have certain advantages in predicting the health status of older adults, AI techniques still have limitations, such as overfitting and the inability to explain the model decision-making process. The difficulty in NN modeling lies in the determination of the number and threshold of hidden neuron layers.

(3) Discuss strategies

In the discussion section, the impact of certain factors on the health status of older adults is analyzed, but it is necessary to consider multiple factors to develop a comprehensive management program.

7. Conclusions

With the aging of the population in China, health issues for older adults have become an increasing concern. Therefore, this study aimed to develop a classification method based on oversampling and NNs to accurately predict the health status of older adults. The main findings are as follows:

(1) WTASUWO was proposed by combining the Weighted Tomek-links strategy and ASUWO oversampling. Compared to different data processing methods, the results show that WTASUWO is very important for improving the accuracy of a classifier.

(2) IMLP was proposed by setting the network structure and optimizing parameters. Based on the oversampled data, IMLP was compared to CS-, SA-, FA-based ELM; PSO-based MLP; and MLP using WTASUWO, and the accuracy was improved from 0.602 to 0.855, showing that this method achieved remarkable improvement in the accuracy of prediction.

To sum up, the experimental results show that the proposed method based on oversampling and NNs has high feasibility in practical applications. Finally, we discuss the experimental results and provide some policy implications.

8. Future Work

In future works, we can discuss the already listed limitations and the way to make targeted improvements in future research. Therefore, we will continue to optimize the model structure, expand the sample size, and explore influencing factors to improve the universality and practicability of the prediction problem. In more detail, here are the main future ideas:

(1) The effect of sample size on model performance can be analyzed using the CFPS dataset, and more large-scale and diverse data can be collected. Thus, we will explore the factors influencing the health status of older adults from other perspectives and develop personalized interventions.

(2) NNs have great potential for SRH prediction, and in the future, we can try to optimize the model structure, such as using more advanced network architectures and optimization algorithms to improve the prediction accuracy. Moreover, we will consider DNNs and deeper layers. In addition, it is possible to combine other machine learning and traditional statistical methods with NNs to achieve more efficient and accurate prediction.

(3) Promoting the attention of the government and all society sectors to the health problems of older adults and increasing policy support related to health benefits, achieving an effective health management for older adults in the future.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L.; software, Y.L.; formal analysis, Y.L.; resources, G.C.; writing—original draft preparation, Y.L., Q.H., and G.X.; writing—review and editing, Y.L.; visualization, Y.L.; supervision, G.C.; funding acquisition, G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Wuhan East Lake High-Tech Development Zone (also known as the Optics Valley of China, or OVC) National Comprehensive Experimental Base for Governance of Intelligent Society (grant #2023003437).

Data Availability Statement: Data employed in this study are public data with sources provided in the article.

Acknowledgments: We are grateful to the Institute for Artificial Intelligence, Peking University, and PKU-WUHAN Institute for Artificial Intelligence, Dingsheng Luo, for providing funding acquisition. The authors are grateful to the editor and the anonymous referees for their many helpful comments on earlier version of our paper.

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

References

1. Bai, C.; Lei, X. New trends in population aging and challenges for China's sustainable development. *China Econ. J.* **2020**, *13*, 3–23. [[CrossRef](#)]
2. Alanazi, H.; Daim, T. Health technology diffusion: Case of remote patient monitoring (RPM) for the care of senior population. *Technol. Soc.* **2021**, *66*, 101662. [[CrossRef](#)]
3. Elreedy, D.; Atiya, A.F. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Inf. Sci.* **2019**, *505*, 32–64. [[CrossRef](#)]
4. Liu, H.; Kong, L.; Sun, Q.; Ma, X. The effects of mindfulness-based interventions on nurses' anxiety and depression: A meta-analysis. *Nurs. Open* **2023**, *10*, 3622–3634. [[CrossRef](#)] [[PubMed](#)]
5. Chen, Y.; While, A.E.; Hicks, A. Self-rated health and associated factors among older people living alone in Shanghai: SRH of older people living alone. *Geriatr. Gerontol. Int.* **2015**, *15*, 457–464. [[CrossRef](#)] [[PubMed](#)]
6. Bahnsen, A.C.; Aouada, D.; Stojanovic, A.; Ottersten, B. Feature engineering strategies for credit card fraud detection. *Expert Syst. Appl.* **2016**, *51*, 134–142. [[CrossRef](#)]
7. Prati, R.C.; Batista, G.E.A.P.A.; Silva, D.F. Class imbalance revisited: A new experimental setup to access the performance of treatment methods. *Knowl. Inf. Syst.* **2015**, *45*, 247–270. [[CrossRef](#)]
8. Manokaran, J.; Vairavel, G. GIWRF-SMOTE: Gini impurity-based weighted random forest with SMOTE for effective malware attack and anomaly detection in IoT-Edge. *Smart Sci.* **2023**, *11*, 276–292. [[CrossRef](#)]
9. Shilaskar, S.; Ghatol, A.; Chatur, P. Medical decision support system for extremely imbalanced datasets. *Inf. Sci.* **2017**, *384*, 205–219. [[CrossRef](#)]
10. He, H.; Garcia, E.A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [[CrossRef](#)]
11. Dai, Q.; Liu, J.W.; Liu, Y. Multi-granularity relabeled under-sampling algorithm for imbalanced data. *Appl. Soft Comput.* **2022**, *124*, 109083. [[CrossRef](#)]
12. Zhao, J.; Liu, N. Semi-supervised classification based mixed sampling for imbalanced data. *Open Phys.* **2019**, *17*, 975–983. [[CrossRef](#)]
13. Li, Y.; Qi, J.; Jin, H.; Tian, D.; Mu, W.; Feng, J. An improved genetic-XGBoost classifier for customer consumption behavior prediction. *Comput. J.* **2023**, bxad041. [[CrossRef](#)]
14. Wang, W.; Wu, Y. Risk analysis of the Chinese financial market with the application of a novel hybrid volatility prediction model. *Mathematics* **2023**, *11*, 3937. [[CrossRef](#)]
15. He, B.; Zhang, Y.; Zhou, Z.; Wang, B.; Liang, Y.; Lang, J.; Lin, H.; Bing, P.; Yu, L.; Sun, D.; et al. A neural network framework for predicting the tissue-of-origin of 15 common cancer types based on RNA-seq data. *Front. Bioeng. Biotechnol.* **2020**, *8*, 737. [[CrossRef](#)]
16. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
17. Espinosa, E.; Figueira, A. On the quality of synthetic generated tabular data. *Mathematics* **2023**, *11*, 3278. [[CrossRef](#)]
18. Han, H.; Wang, W.Y.; Mao, B.H. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In *Advances in Intelligent Computing, Proceedings of the International Conference on Intelligent Computing (ICIC), Hefei, China, 23–26 August 2005*; Lecture Notes in Computer Science, Huang, D.S., Zhang, X.P., Huang, G.B., Eds.; Springer: Berlin/Heidelberg, Germany, 2005. [[CrossRef](#)]
19. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *Proceedings of the IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008*; IEEE: Piscataway, NJ, USA, 2008. [[CrossRef](#)]
20. Yang, L.; Zhang, J.; Wang, X.; Li, Z.; Li, Z.; He, Y. An improved ELM-based and data preprocessing integrated approach for phishing detection considering comprehensive features. *Expert Syst. Appl.* **2021**, *165*, 113863. [[CrossRef](#)]
21. Bunkhumpornpat, C.; Sinapiromsaran, K.; Lursinsap, C. Safe-level-SMOTE: Safe-level-synthetic minority over-sampling Technique for handling the class imbalanced problem. In *Advances in Knowledge Discovery and Data Mining, Proceedings of the Lecture Notes in Computer Science, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Bangkok, Thailand, 27–30 April 2009*; Theeramunkong, T., Kijssirikul, B., Cercone, N., Ho, T.B., Eds.; Springer: Berlin/Heidelberg, Germany, 2009. [[CrossRef](#)]
22. Douzas, G.; Bacao, F.; Last, F. Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Inf. Sci.* **2018**, *465*, 1–20. [[CrossRef](#)]
23. Zhu, T.; Lin, Y.; Liu, Y. Improving interpolation-based oversampling for imbalanced data learning. *Knowl.-Based Syst.* **2020**, *187*, 104826. [[CrossRef](#)]
24. Tao, X.; Zheng, Y.; Chen, W.; Zhang, X.; Qi, L.; Fan, Z.; Huang, S. SVDD-based weighted oversampling technique for imbalanced and overlapped dataset learning. *Inf. Sci.* **2022**, *588*, 13–51. [[CrossRef](#)]
25. Nekooimehr, I.; Lai-Yuen, S.K. Adaptive semi-supervised weighted oversampling (A-SUWO) for imbalanced datasets. *Expert Syst. Appl.* **2016**, *46*, 405–416. [[CrossRef](#)]
26. Li, Y.; Jia, X.; Wang, R.; Qi, J.; Jin, H.; Chu, X.; Mu, W. A new oversampling method and improved radial basis function classifier for customer consumption behavior prediction. *Expert Syst. Appl.* **2022**, *199*, 116982. [[CrossRef](#)]

27. Batista, G.E.A.P.A.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 20–29. [[CrossRef](#)]
28. Seiffert, C.; Khoshgoftaar, T.M.; Hulse, J.V.; Napolitano, A. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Trans. Syst. Man Cybern.* **2010**, *40*, 185–197. [[CrossRef](#)]
29. Xie, X.; Xie, B.; Xiong, D.; Hou, M.; Zuo, J.; Wei, G.; Chevallier, J. New theoretical ISM-K2 Bayesian network model for evaluating vaccination effectiveness. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 12789–12805. [[CrossRef](#)] [[PubMed](#)]
30. Hamidzadeh, J.; Moradi, M. Improved one-class classification using filled function. *Appl. Intell.* **2018**, *48*, 3263–3279. [[CrossRef](#)]
31. Aljarah, I.; Faris, H.; Mirjalili, S. Optimizing connection weights in neural networks using the whale optimization algorithm. *Soft Comput.* **2018**, *22*, 1–15. [[CrossRef](#)]
32. Bi, C.; Tian, Q.; Chen, H.; Meng, X.; Wang, H.; Liu, W.; Jiang, H. Optimizing a multi-layer perceptron based on an improved gray wolf algorithm to identify plant diseases. *Mathematics* **2023**, *11*, 3312. [[CrossRef](#)]
33. Khan, A.U.; Bandopadhyaya, T.K.; Sharma, S. Comparisons of stock rates prediction accuracy using different technical indicators with backpropagation neural network and genetic algorithm based backpropagation neural network. In Proceedings of the International Conference on Emerging Trends in Engineering and Technology, Nagpur, India, 16–18 July 2008; IEEE: Piscataway, NJ, USA, 2008. [[CrossRef](#)]
34. Kar, B.P.; Nayak, S.K.; Nayak, S.C. Opposition-based GA learning of artificial neural networks for financial time series forecasting. In *Advances in Intelligent Systems and Computing, Proceedings of the International Conference on Computational Intelligence in Data Mining (ICCIDM), Bhubaneswar, India, 5–6 December 2015*; Behera, H., Mohapatra, D., Eds.; Springer: New Delhi, India, 2015. [[CrossRef](#)]
35. Malalur, S.S.; Manry, M.T. Multiple optimal learning factors for feed forward networks. In Proceedings of the Conference on Independent Component Analyses, Wavelets, Neural Networks, Biosystems, and Nanoengineering VIII, Orlando, FL, USA, 7–9 April 2010. [[CrossRef](#)]
36. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379*, 1123–1130. [[CrossRef](#)]
37. Li, L.; Wang, Z.; Zhang, T. GBH-YOLOv5: Ghost convolution with BottleneckCSP and tiny target prediction head incorporating YOLOv5 for PV panel defect detection. *Electronics* **2023**, *12*, 561. [[CrossRef](#)]
38. Xin, Y.; Ren, X. Predicting depression among rural and urban disabled elderly in China using a random forest classifier. *BMC Psychiatry* **2022**, *22*, 118. [[CrossRef](#)] [[PubMed](#)]
39. Damaskinos, P.; Koletsi-Kounari, C.; Mamai-Homata, H.; Papaioannou, W. Social, clinical and psychometric factors affecting self-rated oral health, self-rated health and wellbeing in adults: A cross-sectional survey. *Health* **2022**, *14*, 104–124. [[CrossRef](#)]
40. Thamrin, S.A.; Sidik, D.; Kuswanto, H.; Lawi, A.; Ansariadi, A. Exploration of obesity status of Indonesia basic health research 2013 with synthetic minority over-sampling techniques. *Indones. J. Stat. Its Appl.* **2021**, *5*, 75–91. [[CrossRef](#)]
41. Xia, S.; Zheng, Y.; Wang, G.; He, P.; Li, H.; Chen, Z. Random space division sampling for label-noisy classification or imbalanced classification. *IEEE Trans. Cybern.* **2022**, *52*, 10444–10457. [[CrossRef](#)]
42. Liu, Y.; Zheng, Y.F. FS_SFS: A novel feature selection method for support vector machines. *Pattern Recognit.* **2006**, *39*, 1333–1345. [[CrossRef](#)]
43. Sun, L.; Kong, X.; Xu, J.; Xue, Z.; Zhai, R.; Zhang, S. A hybrid gene selection method based on ReliefF and ant colony optimization algorithm for tumor classification. *Sci. Rep.* **2019**, *9*, 8978. [[CrossRef](#)]
44. Vasilyeva, M.; Tyrylgina, A. Machine learning for accelerating macroscopic parameters prediction for poroelasticity problem in stochastic media. *Comput. Math. Appl.* **2021**, *84*, 185–202. [[CrossRef](#)]
45. Seiffert, C.; Khoshgoftaar, T.M.; Hulse, J.V. Hybrid sampling for imbalanced data. In Proceedings of the IEEE International Conference on Information Reuse and Integration, Las Vegas, NV, USA, 13–15 July 2008; IEEE: Piscataway, NJ, USA, 2008. [[CrossRef](#)]
46. Pereira, R.M.; Costa, Y.M.G.; Silla, C.N., Jr. MLTL: A multi-label approach for the Tomek Link undersampling algorithm. *Neurocomputing* **2020**, *383*, 95–105. [[CrossRef](#)]
47. Kaastra, I.; Boyd, M. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* **1996**, *10*, 215–236. [[CrossRef](#)]
48. Li, Y.; Chu, X.; Fu, Z.; Feng, J.; Mu, W. Shelf life prediction model of postharvest table grape using optimized radial basis function (RBF) neural network. *Br. Food J.* **2019**, *121*, 2919–2936. [[CrossRef](#)]
49. Schwenk, H.; Beng, Y. Boosting neural networks. *Neural Comput.* **2000**, *12*, 1869–1887. [[CrossRef](#)] [[PubMed](#)]
50. Fernandes, E.R.Q.; de Carvalho, A.C.P.L.F. Evolutionary inversion of class distribution in overlapping areas for multi-class imbalanced learning. *Inf. Sci.* **2019**, *494*, 141–154. [[CrossRef](#)]
51. Luque, A.; Carrasco, A.; Martín, A.; Heras, A.D.L. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit.* **2019**, *91*, 216–231. [[CrossRef](#)]
52. Krupinski, E.A. Receiver operating characteristic (ROC) analysis. *Frontline Learn. Res.* **2017**, *5*, 31–42. [[CrossRef](#)]
53. Apolloni, J.; Leguizamón, G.; Alba, E. Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments. *Appl. Soft Comput.* **2016**, *38*, 922–932. [[CrossRef](#)]
54. Li, Y.; Chu, X.; Tian, D.; Feng, J.; Mu, W. Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm. *Appl. Soft Comput.* **2021**, *113*, 107924. [[CrossRef](#)]

55. Ge, L.; Ong, R.; Yap, C.W.; Heng, B.H. Effects of chronic diseases on health-related quality of life and self-rated health among three adult age groups. *Nurs. Health Sci.* **2019**, *21*, 214–222. [[CrossRef](#)]
56. Akanni, L.; Lenhart, O.; Morton, A. Income trajectories and self-rated health status in the UK. *SSM Popul. Health* **2022**, *17*, 101035. [[CrossRef](#)]
57. Duboz, P.; Boëtsch, G.; Gueye, L.; Macia, E. Self-rated health in Senegal: A comparison between urban and rural areas. *PLoS ONE* **2017**, *12*, e0184416. [[CrossRef](#)]
58. Vafaei, A.; Yu, J.; Phillips, S.P. The intersectional impact of sex and social factors on subjective health: Analysis of the Canadian longitudinal study on aging (CLSA). *BMC Geriatr.* **2021**, *21*, 473. [[CrossRef](#)] [[PubMed](#)]
59. Yang, F.; Zhang, J. Traditional Chinese sports under China's health strategy. *J. Environ. Public Health* **2022**, *2022*, 1381464. [[CrossRef](#)] [[PubMed](#)]

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.