Article

Dementia and Heart Failure Classification Using Optimized Weighted Objective Distance and Blood Biomarker-Based Features

Veerasak Noonpan [1], Supansa Chaising [2], Georgi Hristov [3] and Punnarumol Temdee [1,*]

[1] Computer and Communication Engineering for Capacity Building Research Center, School of Applied Digital Technology, Mae Fah Luang University, Chiang Rai 57100, Thailand; 6471501501@lamduan.mfu.ac.th

[2] School of Management, Mae Fah Luang University, Chiang Rai 57100, Thailand; supansa.cha@mfu.ac.th

[3] Telecommunications Department, University of Ruse, 7017 Ruse, Bulgaria; ghristov@uni-ruse.bg

[*] Correspondence: punnarumol@mfu.ac.th

Abstract

Dementia and heart failure are growing global health issues, exacerbated by aging populations and disparities in care access. Diagnosing these conditions often requires advanced equipment or tests with limited availability. A reliable tool distinguishing between the two conditions is essential, enabling more accurate diagnoses and reducing misclassifications and inappropriate referrals. This study proposes a novel measurement, the optimized weighted objective distance (OWOD), a modified version of the weighted objective distance, for the classification of dementia and heart failure. The OWOD is designed to enhance model generalization through a data-driven approach. By enhancing objective class generalization, applying multi-feature distance normalization, and identifying the most significant features for classification—together with newly integrated blood biomarker features—the OWOD could strengthen the classification of dementia and heart failure. A combination of risk factors and proposed blood biomarkers (derived from 10,000 electronic health records at Chiang Rai Prachanukroh Hospital, Chiang Rai, Thailand), comprising 20 features, demonstrated the best OWOD classification performance. For model evaluation, the proposed OWOD-based classification method attained an accuracy of 95.45%, a precision of 96.14%, a recall of 94.70%, an F1-score of 95.42%, and an area under the receiver operating characteristic curve of 97.10%, surpassing the results obtained using other machine learning-based classification models (gradient boosting, decision tree, neural network, and support vector machine).

Keywords: dementia; heart failure; risk factors; blood biomarkers; objective distance; weighting features

1. Introduction

The proportion of older adults is increasing in every country, possibly due to improvements in education, nutrition, and healthcare. As a result, a new life-course model has been introduced, encompassing early life (younger than 45 years), midlife (45–65 years), and later life (older than 65 years) [1]. Dementia and heart failure are two of the most diagnosed chronic conditions in older adults, frequently co-occurring due to shared risk factors and interconnected biological mechanisms [2–4]. With increasing age, individuals are more likely to develop conditions such as high blood pressure, diabetes, high cholesterol, and obesity, all of which are known to increase the risk of heart disease and neurodegenerative

diseases [5]. There exists a clear physiological link between these two disorders. For example, heart failure can cause reduced blood flow to the brain, which may accelerate cognitive deterioration. Dementia is also associated with systemic inflammation and metabolic disturbances, which adversely affect cardiovascular function [4]. This two-way interaction highlights the importance of maintaining both cognitive and cardiovascular health in older adults. By managing these comorbidities simultaneously, more effective diagnoses can be developed, allowing for improved treatment outcomes and a higher quality of life for older adults.

Dementia is a growing public concern, impacting the social and economic sectors in terms of medical and social care costs. The global societal cost of dementia was estimated at USD 1.3 trillion in 2019, with costs expected to surpass USD 2.8 trillion by 2030. Additionally, in 2019, the impact on families and caregivers was found to be significant, with an average of 5 h per day spent caring for individuals with dementia. Women were especially impacted; the majority of dementia patients who passed away were female, accounting for up to 65% of cases. Currently, there is no cure for dementia; the primary goal of care is early diagnosis, which is key to optimal management [6]. The increase in the older adult population has given rise to a greater number of dementia prevalence studies [7] that utilize digital healthcare with appropriate technology [8–10] and various features [11].

Heart failure is defined as a condition in which the heart has a reduced ability to pump or fill with blood or exhibits structural abnormalities that result in inadequate cardiac output. Heart failure has been described as a global pandemic, affecting approximately 64.3 million people worldwide as of 2017. The prevalence is expected to rise, primarily because individuals live longer after diagnosis due to lifesaving interventions and general improvements in life expectancy. However, the financial burden is significant—in the U.S. alone, the cost of heart failure was estimated at USD 30.7 billion in 2012, and projections suggest that costs could rise by 127% to USD 69.8 billion by 2030 [12]. Given the high prevalence and severe health consequences of heart failure, numerous studies have focused on identifying its risk factors and developing effective prevention strategies.

Biomarkers are generally used for dementia detection. These include cerebrospinal fluid biomarkers, blood-based biomarkers, neuroimaging biomarkers, and genetic biomarkers. Given the high cost of positron emission tomography/magnetic resonance imaging (PET/MRI), blood-based biomarkers offer a practical alternative for detecting dementia. They are also noninvasive and widely accessible, making them suitable for large-scale and repeated assessments. Dementia and heart failure share several risk factors, including body weight, blood cholesterol levels, hypertension, serum lipids, and diabetes [5,6]. Leveraging noninvasive and affordable methods enhances accessibility and equity, particularly in resource-limited settings. However, using such biomarkers from electronic health records (EHRs) remains challenging and requires comprehensive data integration and validation.

Digital healthcare technology is being rapidly implemented among older adults, as it improves monitoring, communication, and health data collection. Currently, healthcare knowledge-based systems, which support personalized care for older adults and their healthcare providers, are widely accessible through smart devices, such as smartphones and smart watches. In addition, innovations in healthcare and medical support are required to enhance personalized healthcare and increase life expectancy. Machine learning (ML) methods have emerged as powerful tools in medical research. In dementia studies, ML methods have been applied in both preclinical and clinical investigations. Examples include studies using various cognitive tests for dementia [13], a deep learning framework to identify dementia from DNA datasets [14], gradient boosting (GB) for dementia classification based on human genetics [15], analysis of the dementia-related gene APOE4 [16], ensemble classifiers for multi-class classification using the Alzheimer's Disease Neuroimaging Initia-

tive (ADNI) dataset [17], and dementia classification based on PET/MRI [18–20], with most datasets derived from the ADNI. ML methods have also been used in the classification of heart failure, such as support vector machines (SVMs) [21–23], decision trees (DTs) [24], GB [25], extreme gradient boosting (XGB) [26], and deep learning, across datasets of varying types and sizes. However, ML approaches often require large datasets to perform effectively, and performance can degrade when data are limited. Furthermore, ML methods may struggle with interpretability, which is a key consideration in medical diagnostics. To address these limitations, it is essential to explore an alternative approach that could enhance classification performance while maintaining interpretability.

Distance measurement methods, such as the weighted objective distance (WOD), have been successfully applied in the classification of various diseases, including hypertension [27] and type 2 diabetes [28], which are risk factors for dementia and heart failure [1,29–31]. However, the WOD is limited by its reliance on predefined thresholds and fixed values tied to specific diseases, leading to model generalization issues. To address these limitations, this study introduces the optimized weighted objective distance (OWOD). The OWOD enhances the original WOD approach by incorporating an objective function that derives data representations directly from the dataset. It identifies the target values of both dementia and heart failure risk factors and calculates the WOD accordingly. Thus, this study proposes a binary method based on distance measurement for the classification of dementia and heart failure using blood biomarkers and established risk factors from EHRs. The OWOD is expected to strengthen the binary classification of dementia and heart failure by enhancing objective class generalization, applying multi-feature distance normalization, and identifying the most significant features for classification.

2. Literature Review

Numerous studies have investigated various aspects of dementia and heart failure. These studies can be categorized into three groups, which are discussed in this section.

2.1. Feature Studies of Dementia and Heart Failure

Several studies have focused on the risk factors for both dementia and heart failure. Dementia risk factor studies can be divided into three groups. The first group encompasses the relationship between dementia and body weight [32], blood cholesterol [33], and hypertension [29–31]. Such studies have demonstrated significant associations between dementia development and various clinical parameters that can be obtained from routine physical check-ups and blood tests. These include hypertension [29–31], body weight [32], blood cholesterol levels [33], serum lipids [34], and diabetes [1]. The second group comprises preclinical studies, such as speech assessment [35,36], lifestyle activity [37], handwriting assessment [38,39], sleep disturbances [40,41], sex differences [42], physical assessment, risk score models [43–45], and hearing loss [1]. The third group includes brain imaging clinical studies using PET/MRI [19].

Similarly, heart failure studies can be categorized into three groups. The first group identifies risk factors such as age [46,47], sex [46,47], hypertension [46], hypercholesterolemia [47], low-density lipoprotein cholesterol (LDL-C) [46], diabetes mellitus [46], overweight/obese [47], elevated body mass index (BMI) [46], smoking [46], low physical activity [46], family history of premature cardiovascular disease (CVD) [47], chronic kidney disease [47], and lipid biomarkers [47]. Key risk factors for heart failure—such as obesity, hypercholesterolemia, hypertension, lipid biomarkers, diabetes, sex, and age—are also strongly associated with dementia [12,29–31,46,47]. The second group includes studies on preclinical heart failure, such as the incidence of preclinical heart failure in healthy community individuals using health status records during visits. These include age, BMI, systolic

blood pressure (SBP), diastolic blood pressure (DBP), heart rate, diabetes, hypertension, obesity, hyperlipidemia, and echocardiogram [48]. Some studies assess the risk of heart failure based on preclinical information [47]. The third group comprises studies on the clinical identification and prevention of heart failure, such as heart disease identification using ML methods [21–23], such as DT [24], GB [25], and XGB [26].

While the reviewed studies provide valuable insights across the dementia and heart failure risk continuum—from risk factor identification to preclinical and clinical assessments—several limitations and potential biases warrant consideration. Risk factor studies often rely on observational or retrospective data, which may be subject to confounding variables and lack causal inference. Preclinical studies may introduce cultural or educational biases, limiting generalizability across diverse populations. Furthermore, risk score models in the case of dementia, although practical, can oversimplify complex disease pathways and may not fully capture interactions between multiple features. Clinical studies using advanced equipment and laboratory tests are typically conducted in well-resourced settings, potentially excluding underrepresented populations and limiting real-world applicability. These limitations underscore the need for integrative, accessible, and representative approaches in the classification of dementia and heart failure, balancing precision with inclusivity.

Although dementia and heart failure often co-occur and share multiple risk factors, their distinct clinical manifestations and biomarker profiles justify the use of a binary classification model. Developing such a model is challenging, as it must differentiate between two interrelated conditions with overlapping etiologies yet diverging diagnostic and therapeutic pathways—particularly in cases where early symptoms are ambiguous. The dementia and heart failure risk factors identified in previous studies are shown in Table 1.

Table 1. Summary of dementia and heart failure risk factors identified in previous studies.

Dementia	Heart Failure
Body weight [32]	Being overweight/obese [46]
Blood cholesterol [33]	Hypercholesterolemia [46]
Hypertension [29–31]	Hypertension [12,46]
Serum lipids [34]	Lipid biomarkers [47]
Diabetes [1]	Diabetes mellitus [46]
Sex [42]	Sex [12,46,47]
ApoE4 gene [16]	Age [12,46,47]
Air pollution (NO_2, PM2.5) [1]	
Sleep disturbances [40]	

This study proposes a new feature and method for classifying dementia and heart failure, with blood biomarkers mainly used to construct the binary classification model. Two groups of blood-oriented features were used in this study. The first group consisted of features associated with dementia and heart failure development from existing works, termed related risk factors (R). The second group consisted of blood biomarker features also obtainable from blood testing, but with no established associations with the two conditions; these were newly proposed for this study and termed potential features (P). Therefore, the features used to develop the classification model were derived from both existing risk factors and a new set of blood biomarker-oriented features. The integration of the new set of features was expected to sharpen the boundary of classes within the dataset.

2.2. ML-Based Classifications for Dementia and Heart Failure

Several ML models have been proposed for constructing classification models from different types of data. For dementia studies, the models can be categorized into two groups. First are the models that use imaging datasets for classification [18–20,49–52]. The images

are usually obtained from advanced laboratories and tend to have high resolutions, necessitating sophisticated models, such as deep learning-based models. The limitations of imaging-based studies include the need for advanced instrumentation and a sufficiently large sample size to ensure reliable results—factors that are often difficult to achieve in practical healthcare settings. Second are the models that use nonimaging datasets for classification [21–23,27,28,43,53]. Such studies utilize a range of biomarkers, such as demographic and psychiatric information [43], DNA methylation profiles [14], genetic data from ADNI-1 [15], and medical attributes [27,28,54].

For heart failure studies, SVMs have been widely adopted across different datasets, achieving prediction accuracies between 92.37% and 98.47% depending on preprocessing techniques, such as principal component analysis or hybrid feature selection [21–23]. DT, while simpler, has still proven effective, with an accuracy of 93.00% on Kaggle heart failure data [24]. More recent approaches have leveraged physiological signals—such as electrocardiograms—processed through recurrent neural networks and deep learning models—producing high accuracies of 99.86% [53] and 97.93% [51], respectively. These results revealed the effectiveness of the incorporation of time-series and image-based representations of cardiac signals, which appear to significantly enhance the accuracy and robustness of heart failure classification models.

Despite the high comorbidity rate of heart failure and dementia, few studies have addressed their joint classification. Previous work by the team of the corresponding author explored both binary and multi-class classification models. For binary classification, an extra trees model with data balancing achieved 89.11% accuracy on a small dataset (4297 records) [55], increasing to 96.47% on a larger dataset (14,763 records) using a data enrichment framework [54]. For multi-class classification, GB with 59 features across more than 16,000 records classified heart failure, aortic stenosis, and dementia with 83.81% accuracy [27]. XGB with 108 augmented features on 26,474 records achieved 91.42% accuracy [26]. These studies highlight the importance of large-scale, feature-rich data and robust ML algorithms for classification problems. A persistent challenge in applying ML to medical/healthcare data is the insufficiency of clinical data, as many datasets are incomplete or imbalanced, limiting the model's ability to generalize and increasing the risk of overfitting, particularly in complex classification tasks. In addition, the complexity and limited interpretability of such models necessitate the development of more transparent methods, such as distance measurements.

2.3. Classification with Distance Measurements

Classification using distance measurements is expected to outperform ML-based models in these specific scenarios as it relies on direct comparisons between feature vectors, making it more robust, less sensitive to overfitting, and more interpretable. This is especially the case when the feature space is well defined and clinically meaningful. Classification methods using distance measurements have been introduced and widely applied in various domains over the decades. These include Euclidean distance [56], Manhattan distance [57], distance measures on fuzzy c-means algorithms [58], and objective distance [59–61]. Recently, a WOD [27] was proposed to solve personalized care for older adults with hypertension. This method measures the distance between the current health status of an individual and their defined level of hypertension and then generates personalized feedback based on the distances obtained. Additionally, the average weighted objective distance (AWOD) [28] has been applied for predicting type 2 diabetes. Both the WOD and AWOD leverage the concepts of objective distance, demonstrating their potential for measuring the distance between an individual's status and expected health goals. The WOD utilizes weighting factors derived from information gain to prioritize these factors in distance calculations. The AWOD builds

upon this framework by focusing on the average distances between features and is particularly effective for binary classification problems. However, both methods have limitations: the WOD often struggles with comparability in distances due to its reliance on static weighting, while the AWOD, despite improving upon the WOD, may overlook the complexities in the interactions between risk factors. These limitations highlight a critical gap in the effective modeling of the influence of diverse health conditions. To address these challenges, this study proposes the OWOD, which modifies the distance metric for each feature, enhances comparability, and employs a refined objective function derived directly from the dataset of dementia and heart failure. This new distance aims to improve the performance of the classification model and promote model generalization.

2.4. The Proposed Study

The proposed OWOD introduces an optimized target value, defined as the point of maximum divergence between class histograms. This value represents the most discriminative threshold reflecting the class boundaries and is used to determine the optimal decision point for classification. Compared with the WOD and AWOD, the OWOD enhances the modification of each feature's distance values to facilitate comparability. It also improves on the previous distances' predefined and standard values, in terms of finding the objective function that represents values from the dataset, thereby enhancing generalization.

Similar to the WOD and AWOD, the weighting factors of the OWOD are derived from information gain. Information gain, derived from entropy, measures uncertainty or impurity in a dataset. Entropy quantifies the level of uncertainty or disorder within data. In the classification problem, entropy is primarily utilized to determine the irrelevant attributes of a dataset [62]; accordingly, information gain is one of the most popular methods for feature selection problems [63,64]. Information gain is used as a novel feature selection for text classification problems [65]. It reduces the dimensionality of features available in the document for improving classification performance. In addition, it is applied as a weight coefficient to enhance the effectiveness of the classification algorithm [66]. Weighting attributes can improve classification accuracy [67] by prioritizing attributes from the least to the most important.

For the current study, the classification model using the OWOD was constructed with an equal number of dementia and heart failure records, ensuring both conditions were treated as equally important. This approach addressed the limitations of previous measurements, including the WOD and AWOD, by refining distance calculations and employing a more effective objective function derived from the dataset. The risk factors and additional blood biomarker features were employed to sharpen the boundary between the two dataset classes: patients with dementia and patients with heart failure. It is hypothesized that new blood biomarker features and the OWOD can promote generalization and provide better classification performance compared with other ML-based binary classification models.

3. Research Methodology

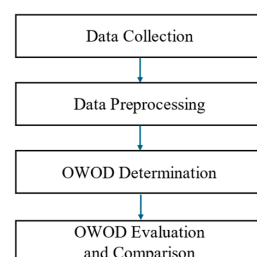This study consisted of four main procedures, as shown in Figure 1.

Figure 1. Research methodology consisting of four main processes: data collection, data preprocessing, optimized weighted objective distance (OWOD) determination, and OWOD evaluation and comparison.

The details of each procedure are discussed in the following sections.

3.1. Data Collection

The dataset used in this study is the EHR from the Chiang Rai Prachanukroh Hospital, Chiang Rai, Thailand, spanning the years 2016 to 2022. The dataset includes blood tests and clinical records of individuals aged 60 and over, comprising a total of 12,222 records (dementia = 5563; heart failure = 6659). No individual was diagnosed with both diseases simultaneously. Table 2 shows the features used.

Table 2. Description of the features used in the current study.

Features	Acronym	Data Range	Group
Body weight (kg)	W	40.1–116.3	R
Height (cm)	H	150.1–185.0	P
Body mass index (kg/m^2)	BMI	11.76–39.57	R
Systolic blood pressure (mmHg)	SBP	76–199	R
Diastolic blood pressure (mmHg)	DBP	61–126	R
Fasting blood sugar (mg/dL)	FBS	62–495	R
Triglycerides (mg/dL)	TGS	51–199	R
Total cholesterol (mg/dL)	TC	101–429	R
High-density lipoprotein cholesterol (mg/dL)	HDL	31–93	R
Low-density lipoprotein cholesterol (mg/dL)	LDL	51–196	R
Hemoglobin (g/dL)	HB	10.10–20.10	P
White blood cell (count/μL)	WBC	3100–19,900	P
Polymorphonuclear neutrophils (percentage)	NEUT	30.20–89.90	P
Thrombocytes (count/μL)	PLAT	101,000–585,000	P
Lymphocyte cells (percentage)	LYMP	10.10–59.00	P
Creatinine (mg/dL)	CREA	0.35–2.99	P
Blood urea nitrogen (mg/dL)	BUN	4–49	P
Thyroid stimulating hormone (mIU/L)	TSH	0.01–5.97	P
Potassium (mEq/L)	K	1.40–7.80	P
Sodium (mEq/L)	NA	109–167	P
Carbon dioxide (mEq/L)	CO_2	11–45	P

As previously discussed (and shown in Table 2), the features used in this study were divided into two groups: R and P. The R group (including body weight, blood cholesterol, and blood pressure) was employed as it was associated with dementia and heart failure in the existing research. The P group was obtained through blood tests. Physical activity was not considered a risk factor in this study due to insufficient data; this includes smoking and alcohol history.

3.2. Data Preprocessing

The dataset originally consisted of 12,222 records, including 5563 dementia records and 6659 heart failure records. The raw dataset subsequently went through the preprocessing procedure shown in Figure 2. Following this, the final dataset was balanced to include 10,000 records, with 5000 dementia and 5000 heart failure cases.
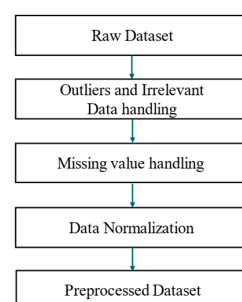

Figure 2. Data preprocessing workflow comprising outlier and irrelevant data handling, missing value imputation, and data normalization.

Prior to normalization, erroneous or invalid entries were addressed by handling outliers and irrelevant data. This included filtering or correcting values deviating significantly from the expected distribution or deemed implausible due to data entry errors (e.g., extremely low or high body weights). Missing values were addressed using mode imputation, based on the most frequent values observed in the data distribution. Data normalization was applied as part of the preprocessing workflow to ensure consistency and reliability before proceeding to OWOD determination. This step involved adjusting the raw data to a standard scale, typically by standardizing values with a mean of 0 and a standard deviation of 1, or by scaling them to a defined range. Such preprocessing ensured that the final dataset was statistically valid and suitable for subsequent analysis.

An example histogram of the body weight feature is presented in Figure 3, demonstrating its applicability in the data preprocessing step.
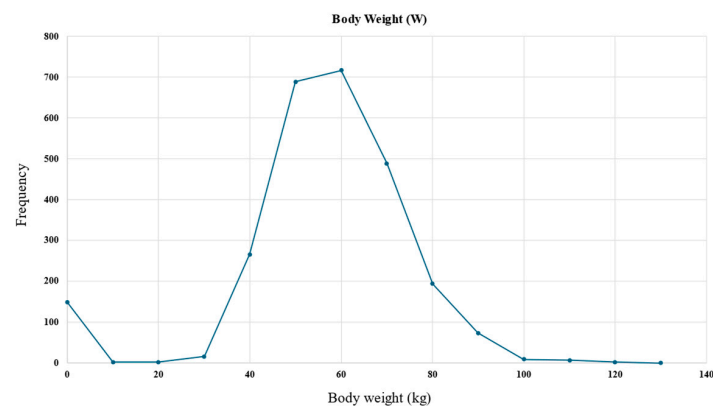

Figure 3. Body weight feature histogram used in the data preprocessing step to handle outliers and missing values.

It can be observed in Figure 3 that weight values within the 0–40 kg range were erroneous entries. To address this, the data were filtered to exclude values falling outside the expected distribution range. The missing body weight entries were attributed to the most commonly occurring value, which was 60 kg.

3.3. The OWOD Concept

The underlying principle of the OWOD method is based on the distance between current and objective states, with the weights derived from information gain. The weights reflect the actual impact on models and represent the diverse health conditions of individuals, as well as the general diagnostic procedures used by healthcare professionals. This concept is illustrated in Figure 4.
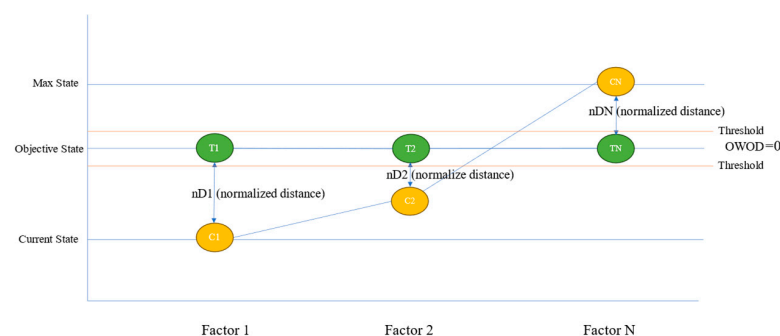

Figure 4. The concept of the optimized weighted objective distance (OWOD), a weighted distance-based measure between current and objective states, reflecting the influence of each feature at the individual level.

Figure 4 shows that three main steps are required for OWOD determination. The first step involves determining the objective value for distance calculation. The objective value (T_i) is the state level of each feature (i), which represents the group of dementia patients in the dataset. The current value (C_i) is the current state for each feature. The acceptable value (A_i) is the health status level for each feature acceptable for an individual. All three level values are used to determine the OWOD.

3.4. OWOD Determination

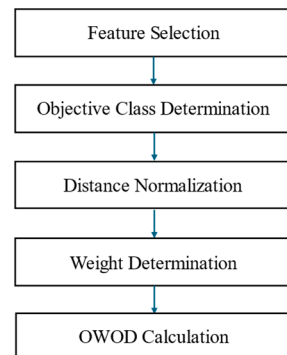The OWOD determination process is illustrated in Figure 5.

Figure 5. The optimized weighted objective distance (OWOD) determination process comprising feature selection, objective class determination, distance normalization, weight determination, and OWOD calculation.

The OWOD algorithm consists of five processes, beginning with feature selection, where the number of features used in the calculation is determined. The objective value represents the target level of each feature derived from the dataset. The next step involves calculating the normalized distance between the current state and the objective state. Entropy and information gain are then used to compute the weight for each feature. Factor weighting is determined based on information gain, and the final step involves calculating the OWOD.

3.4.1. Feature Selection

This study utilized a mixed selection of R and P features, testing various sets of 8, 12, 16, and 20 features to determine the feature combination of the best model. These selections were intended to investigate how varying levels of feature dimensionality influence OWOD model performance. These numbers represented incremental increases that helped examine the trade-off between feature richness and overall classification performance.

3.4.2. Objective Class Determination

To calculate the OWOD, the objective class was determined. A histogram was used to find the data distribution of the objective class. The histogram was calculated using Equation (1).

$$n = \sum_{i=1}^{k} m_i \tag{1}$$

where n refers to the total number of observations, k refers to the total number of bins, and m_i refers to the histogram data. The target value was defined as the value of observation k corresponding to the maximum positive difference between the histograms of the positive and negative classes. This objective value was determined from the class-wise histogram distribution presented in Figure 6.
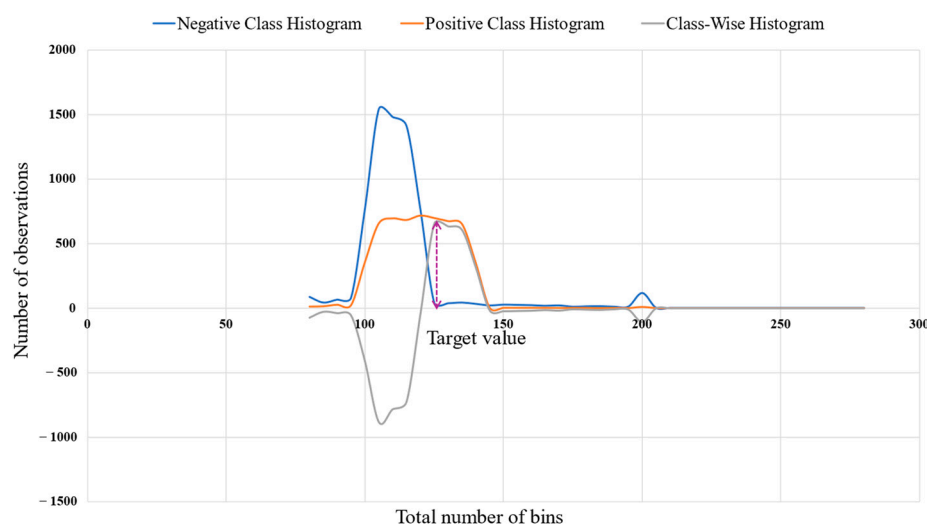
Figure 6. Determination of the target value (dotted arrow) for each feature based on class-wise histogram distributions, illustrating the distributional differences between positive and negative classes.

The target value of this feature was calculated using Equation (2).

$$T_{mi} = Mode\ (Max\ (\ Histogram\ (PositiveClass) - Histogram\ (NegativeClass)))\ \quad (2)$$

where T_{mi} denotes the target value of the observation feature and i corresponds to the point of maximum difference between the positive and negative class distributions.

3.4.3. Distance Normalization

The distance calculation method was used to determine the distance between the current and objective state features. The distance was calculated using Equation (3).

$$dTC = \sqrt{(T - C)^2} \quad (3)$$

where T refers to the target value of attributes, C refers to the current state value of attributes, and dTC refers to the distance between the target and current states.

$$dTA = \sqrt{(T - A)^2} \quad (4)$$

where T refers to the target value of attributes, A refers to the acceptable state value of attributes, and dTA refers to the distance between the target and acceptable states.

Data from multiple features were combined in the calculation. As distance was a key measurement, any bias could lead to issues in the final OWOD calculation. The normalization method was therefore applied. Normalization was calculated using Equation (5).

$$ndTC = \frac{dTC}{A_{max}} \quad (5)$$

where dTC is the distance from Equation (3), A_{max} is the acceptable value and the maximum value, and $ndTC$ is the normalized distance between the target and current states. This is shown in Equation (6).

$$ndTA = \frac{dTA}{A_{max}} \quad (6)$$

where dTA is the distance from Equation (4), A_{max} is the acceptable value and the maximum value, and $ndTA$ is the normalized distance between the target and acceptable states.

The normalized distance ratio used for the weight calculation rTC is the normalized distance ratio of the normalized distance between the target and current states compared to the total normalized distance. This is shown in Equation (7).

$$rTC = \frac{ndTC}{ndTA + ndTC} \tag{7}$$

rTA is the normalized distance ratio of the normalized distance between the target and acceptable states compared to the total normalized distance. This is shown in Equation (8).

$$rTA = \frac{ndTA}{ndTA + ndTC} \tag{8}$$

3.4.4. Weight Determination

To determine the entropy of the target class with respect to all attributes, an equal proportion of the target class was initially determined using Equation (9). This resulted in an equal proportion of positive and negative classes.

$$Oc_+ = Nc_- = \frac{Na}{N} \tag{9}$$

where Oc_+ is the value of the equal proportion for the positive class (+) with respect to all attributes, Nc_- is the value of the equal proportion for the negative class (−) with respect to all attributes, Na is the total number of attributes, and N is the total number of target classes.

Next, the fraction of the target class was determined using Equation (10). This value represents the fraction of the chance of being in a positive or negative class with respect to all attributes.

$$fOc_+ = \frac{Pe_+}{Na} = fNc_- = \frac{Pe_-}{Na} \tag{10}$$

where fOc_+ represents the fraction of the positive class (+) and fNc_- represents the fraction of the negative class (−), in which both classes are equal.

The weight determination relies on entropy and information gain.

- Entropy.

Shannon Entropy, an information theory method, quantifies the average level of information, surprise, uncertainty, or complexity for a given random variable based on its historical occurrences. Entropy $(EP(X))$ was calculated using Equation (11).

$$EP(X) = -\sum_{i=1}^{n} P(x_i) \, log_2 P(x_i) \tag{11}$$

where X is the class of attributes, $P(x_i)$ is the proportion of the samples belonging to class X, and log_2 is log based 2.

- Information Gain.

Information gain is the reduction in entropy produced by partitioning a set with attributes a and finding the optimal candidate that produces the highest value. The information gain $(IG\ (T, a))$ was calculated using Equation (12).

$$IG\ (T, a) = H(T) - H(T|a) \tag{12}$$

where T is a random variable and $H(T|a)$ is the entropy of T given the value of an attribute a.

The split information value is a positive integer that describes the potential worth of splitting a branch from a node. This, in turn, is the intrinsic value the random variable

possesses and is used to remove bias from the information gain ratio calculation. The split information value ($S(X)$) was calculated using Equation (13).

$$S(X) = -\sum_{i=1}^{n} \frac{N(t_i)}{N(t)} * log_2 \frac{N(t_i)}{N(t)} \tag{13}$$

where X is a discrete random variable with possible value X_i and $N(t_i)$ is the number of times that t_i occurs, divided by the total event count $N(t)$, where t is the set of events.

The information gain ratio ($IGR\ (T,a)$), the ratio of the information gain and the split information value for the T variable $S(T)$. $IGR\ (T,a)$, was calculated using Equation (14).

$$IGR\ (T,a) = \frac{IG(T,a)}{S(T)} \tag{14}$$

The weight information (IW_i) was calculated using Equation (15).

$$IW_i = \frac{R_{eg}}{\sum_{i=1}^{N} R_{eg}} \tag{15}$$

where R_{eg} is the ratio of entropy from Equation (11) and the information gain from Equation (14), and W_i is the calculated weight information.

3.4.5. OWOD Calculation

To determine the OWOD of all attributes, the WOD of each attribute ($owoD_i$) was first determined using Equation (16). Let $T_{mi} = t_{mi}$, $A_i = a_i$, $C_i = c_i$.

$$owoD_i = W_i * n \left| \sqrt{(T_{mi} - A_i)^2} - \sqrt{(T_{mi} - C_i)^2} \right| \tag{16}$$

After obtaining the $owoD_i$ of each person, the normalized WOD of each attribute ($nowoD_i$) was determined using Equation (17).

$$nowoD_i = \frac{owoD_i - owoD_{min}}{owoD_{max} - owoD_{min}} \tag{17}$$

where $owoD_{max}$ is the maximum OWOD among all attributes and $owoD_{min}$ is the minimum OWOD among all attributes.

Therefore, the OWOD of all attributes for each person ($OWOD_i$) was determined using Equation (18).

$$OWOD_i = \frac{\sum_{i=1}^{Na} nowoD_i}{Na} \tag{18}$$

where $OWOD_i$ denotes the OWOD of all attributes for the ith individual.

To identify the class based on the obtained $OWOD_i$, the following condition was applied, as shown in Equation (19):

$$OWOD_i = \begin{cases} Negative\ class,\ if\ OWOD_i > OWOD_c \\ Positive\ class,\ if\ OWOD_i \leq OWOD_c \end{cases} \tag{19}$$

where $OWOD_c$ is the cut-off OWOD calculated from threshold Ts_{mi} of all features. Ts_{mi} refers to the threshold of the target values for each feature and is defined as the deviation of 5% above and below the T_{mi} of all features. For this study, the negative class was heart failure, and the positive class was dementia.

3.4.6. OWOD Algorithm

The algorithm used to perform the OWOD calculation is presented in the pseudocode shown in Algorithm 1 as follows:

Algorithm 1: OWOD calculation

Input: List of selected features (**F**)
Output: Table of computed OWOD values
1: **Initialize** storage results
2: **For** each sample **S** in the dataset **do**
3: **For** each feature **F do**
4: **Retrieve C** (current level), **T** (target level), **A** (acceptable level)
5: **Compute** Euclidean distances:
$$dTC \leftarrow \sqrt{((T - C)^2)}$$
$$dTA \leftarrow \sqrt{((T - A)^2)}$$
6: **Compute** Normalized distances:
$$ndTC \leftarrow dTC/A$$
$$ndTA \leftarrow dTA/A$$
7: **Compute** Normalized distance ratios:
$$rTC \leftarrow ndTC/(ndTC + ndTA)$$
$$rTA \leftarrow ndTA/(ndTC + ndTA)$$
8: **Compute** Entropy, information gain, and gain score:
$$EP \leftarrow \textbf{calculate entropy } (rTC, rTA)$$
$$IG \leftarrow \textbf{calculate information gain } (EP)$$
$$G \leftarrow \textbf{calculate gain } (IG)$$
9: **Compute** Entropy–gain ratio and weight:
$$rEG \leftarrow EP/G$$
$$SrEG \leftarrow \textbf{sum } (rEG)$$
$$W \leftarrow rEG/SrEG$$
10: **Compute** Distance difference, weight*distance and normalize:
$$DF \leftarrow |ndTC - ndTA|$$
$$WD \leftarrow W \times DF$$
$$MxWD \leftarrow \textbf{max } (WD), MnWD \leftarrow \textbf{min } (WD)$$
$$nWD \leftarrow (WD - MnWD)/(MxWD - MnWD)$$
11: **Compute** OWOD:
$$OWOD \leftarrow \textbf{average } (nWD)$$
12: **Store** S, F, and OWOD values
13: **End for**
14: **End for**
15: **Return** OWOD result table

3.4.7. Sample Calculation

This section demonstrates a sample calculation of the OWOD for classifying a potential group of people with dementia and a group with heart failure. To compute the OWOD, both the target and acceptable levels for each feature are applied. The target level represents the desired or optimal value for each feature; the acceptable level indicates the range of values considered tolerable. The acceptable level for each feature is defined as the standard deviation from the average value of the feature.

The objective function relies on obtaining the boundary of two classes by plotting the histogram of each feature, including the histogram of the objective and the other class; the

target level is calculated by histogram subtraction of the two classes (grey line in Figure 7). The peak of the grey line indicates the target value.
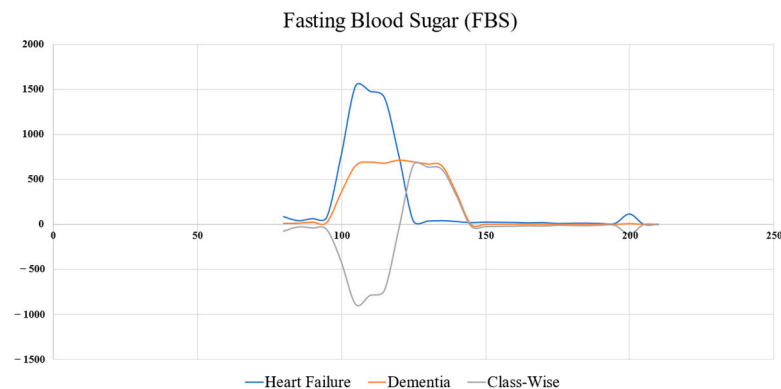

Figure 7. Target value determination of FBS, which is derived from a class-wise histogram distribution between the positive and negative classes.

As illustrated in Figure 7, the target level of the fasting blood sugar (FBS) (T_{FBS}) can be determined using Equation (2), resulting in a value of 130. Accordingly, when T_{FBS} = 130, the acceptable level (A_{FBS}) = 280. More samples of target and acceptable levels are shown in Table 3. Table 4 shows the samples of records used to calculate the OWOD.

Table 3. Sample target and acceptable values.

Feature (*i*)	Target Level (T_{mi})	Acceptable Level (A_i)
W	55 kg	90 kg
SBP	150 mg/dL	190 mg/dL
DBP	85 mg/dL	110 mg/dL
FBS	130 mg/dL	280 mg/dL
TGS	140 mg/dL	160 mg/dL
TC	200 mg/dL	260 mg/dL
HDL	55 mg/dL	70 mg/dL
LDL	135 mg/dL	165 mg/dL

Table 4. Examples of the dataset.

No.	W	SBP	DBP	FBS	TGS	TC	HDL	LDL
1	65	130	75	120	115	175	35	105
2	70	140	80	105	95	165	50	120
3	55	155	90	110	110	195	60	140
4	65	135	75	100	100	185	40	120
5	50	150	75	115	155	205	50	135
6	60	135	85	115	100	185	60	145
7	92	155	85	120	125	220	50	140
8	77	160	80	115	100	210	50	135
9	47	121	82	130	85	215	55	130
10	45	162	77	115	120	205	60	125
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
8000	65	135	75	100	100	185	40	120

The classification of dementia and heart failure using the proposed measurement can be illustrated by applying the information of Person No. 1. The datasets related to the current health status of Person No. 1 are shown in Table 4. The target and acceptable levels of all features are shown in Table 3. The sample attributes comprise eight features. The target group is an objective class (OC) group and a non-objective (NC) class group. OC refers to the positive class (with dementia) and NC to the negative class (with heart failure).

Following Algorithm 1, the sample calculation of the OWOD for identifying Person No. 1 is demonstrated below.

To determine the entropy of the target group with respect to all features, an equal proportion of the target group was initially determined using Equation (9). The value of the OC group (Oc_+) and NC group (Nc_-) is equal to 4, as follows:

$$Oc_+ = Nc_- = \frac{8}{2} = 4$$

Next, the fraction of the target group with respect to all features was determined using Equation (10). The fraction of the OC group (fOc_+) and the NC group (fNc_-) with respect to all features is as follows:

$$fOc_+ = \frac{4}{8} \quad fNc_- = \frac{4}{8}$$

Thereby, the entropy of the target group with respect to all features ($EP(C)$) was determined using Equation (11). The value of $EP(C)$ is equal to 1 as follows:

$$EP(C) = -\frac{4}{8} \times log_2\left(\frac{4}{8}\right) - \frac{4}{8} \times log_2\left(\frac{4}{8}\right) = 1$$

To determine the entropy of each feature, an example of the entropy calculation for the weight factor is presented.

Firstly, the current distance (dTC_w) and acceptable distance (dTA_w) were determined using Equations (3) and (4), respectively. According to the condition (*i*), $T_w = t_w = 55$, $C_w = c_w = 65$, and $A_w = a_w = 90$, the dTC_w and the dTA_w are equal to 10 and 35 as follows:

$$dTC_w = \sqrt{(55 - 65)^2} = 10$$

$$dTA_w = \sqrt{(55 - 90)^2} = 35$$

The distances of dTC_w and dTA_w must be normalized and converted to percentages using Equations (5) and (6), respectively, due to the different range of each feature.

$$ndTC_w = \left(\frac{10}{90}\right) * 100 = 11.11$$

$$ndTA_w = \left(\frac{35}{90}\right) * 100 = 38.89$$

Next, the proportion of the acceptable distance (rTA_w) and the current distance (rTC_w) were calculated using Equations (7) and (8), respectively. The rTA_w and rTC_w values are equal to 0.22 and 0.78, as follows:

$$rTC_w = \frac{11.11}{38.89 + 11.11} = 0.22$$

$$rTA_w = \frac{38.89}{38.89 + 11.11} = 0.78$$

The fractions of the OC group (fOc_{W+}) and the NC group (fNc_{W-}) are as follows:

$$fOc_{w+} = \frac{0.22}{0.78 + 0.22} = \frac{0.22}{1}$$

$$fNc_{w-} = \frac{0.78}{0.78 + 0.22} = \frac{0.78}{1}$$

Thereby, the entropy of W feature $(EP(C_w))$ was determined using Equation (11). The $EP(C_w)$ is equal to 0.76, as follows:

$$EP(C_w) = -\frac{0.22}{1} * log_2\left(\frac{0.22}{1}\right) - \frac{0.78}{1} * log_2\left(\frac{0.78}{1}\right) = 0.76 \tag{20}$$

Thus, the entropies of SBP, DBP, FBS, TGS, TC, HDL, and LDL are equal to 0.92, 0.86, 0.34, 0.99, 0.87, 0.99, and 1.00, respectively.

To determine the information gain of the target group with respect to all features, the entropy of all features was calculated using Equation (12). The $IG(C_a)$ is equal to 0.84, as follows:

$$
\begin{aligned}
IG(C_a) \quad &= 0.76 * \left(\frac{0.22+0.78}{8}\right) + 0.92 * \left(\frac{0.33+0.67}{8}\right) \\
&+0.86 * \left(\frac{0.29+0.71}{8}\right) + 0.34 * \left(\frac{0.06+0.94}{8}\right) \\
&+0.99 * \left(\frac{0.46+0.54}{8}\right) + 0.87 * \left(\frac{0.29+0.71}{8}\right) \\
&+0.99 * \left(\frac{0.57+0.43}{8}\right) + 1.0 * \left(\frac{0.5+0.5}{8}\right) \\
&= 0.84
\end{aligned}
$$

Thereby, the information gain of the target group with respect to all features $(G(C,\ T))$ is equal to 0.16 as follows:

$$G(C,T) = 1 - 0.84 = 0.16$$

To determine the weight of each feature, the weight calculation for the weight factor is presented as an example. The significant ratio value of the weight factor (R_w) was calculated, which is equal to 4.75, as follows:

$$R_W = \frac{0.76}{0.16} = 4.75$$

Thus, $R_{sbp} = 5.80$, $R_{dbp} = 5.45$, $R_{FBS} = 2.13$, $R_{tgs} = 6.26$, $R_{tc} = 5.52$, $R_{hdl} = 6.22$, and $R_{ldl} = 6.32$.

Thereby, the weight of the weight factor (W_w) was determined using Equation (15), which is equal to 0.11, as follows:

$$W_w = \frac{4.75}{4.75 + 5.80 + 5.45 + 2.13 + 6.26 + 5.52 + 6.22 + 6.32} = 0.11$$

Thus, $W_{sbp} = 0.14$, $W_{dbp} = 0.13$, $W_{FBS} = 0.05$, $W_{tgs} = 0.15$, $W_{tc} = 0.13$, $W_{hdl} = 0.15$, and $W_{ldl} = 0.15$.

To determine the OWOD of all features, the WOD of the weight factor $(owoD_w)$ was calculated using Equation (16), which is equal to 3.15, as follows:

$$owoD_w = 0.11 * |\,38.89 - 11.11\,| = 3.15$$

Thus, $owoD_{sbp} = 1.44$, $owoD_{dbp} = 1.75$, $owoD_{FBS} = 2.50$, $owoD_{tgs} = 0.46$, $owoD_{tc} = 1.75$, $owoD_{hdl} = 1.05$, and $owoD_{ldl} = 0$.

From the values of $owoD$ among the eight features, $owoD_{max}$ is 3.15 and $owoD_{min}$ is 0. The normalized WOD of weight (nwD_w) was determined using Equation (17), which is equal to 1.00, as follows:

$$nowoD_w = \frac{3.15 - 0}{3.15 - 0} = 1.00$$

Thus, $nowoD_{sbp} = 0.46$, $nowoD_{dbp} = 0.55$, $nowoD_{FBS} = 0.79$, $nowoD_{tgs} = 0.15$, $nowoD_{chol} = 0.55$, $nowoD_{hdl} = 0.33$, and $nowoD_{ldl} = 0$.

Therefore, the OWOD of all features classifying the group of Person No. 1 ($OWOD_1$) is equal to 0.48, as follows:

$$OWOD_1 = \frac{1.00 + 0.46 + 0.55 + 0.79 + 0.15 + 0.55 + 0.33 + 0}{8} = 0.48$$

According to the conditions, $OWOD_1 = 0.48$; before identifying the class of $OWOD_1$, $OWOD_c$ was calculated with a similar algorithm by changing the current status value (C_i) to the threshold value (Ts_{mi}) of all features, as illustrated in this calculation of an eight-feature dataset. The list of Ts_{mi} values is as follows: $Ts_{mi-w} = 52.5$, $Ts_{mi-sbp} = 142.5$, $Ts_{mi-dbp} = 80.75$, $Ts_{mi-FBS} = 123.5$, $Ts_{mi-tgs} = 133$, $Ts_{mi-tc} = 190$, $Ts_{mi-hdl} = 52.25$, and $Ts_{mi-ldl} = 128.25$. After determining that the $OWOD_c$ value = 0.56, it could be identified that Person No. 1 was in the positive class ($0.48 \leq 0.56$) and, therefore, in the dementia group.

3.4.8. OWOD Evaluation and Comparison

This section presents the OWOD evaluation process, as illustrated in Figure 8. The OWOD method is compared with the OWOD with differences in feature selection and other ML classification models.
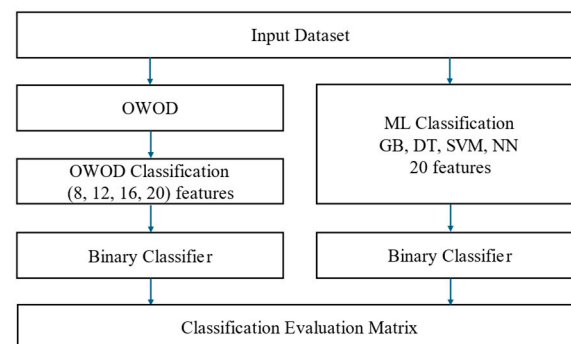

Figure 8. The optimized weighted objective distance (OWOD) evaluation framework for determining the optimal number of features used in constructing the OWOD and comparing classification performance with other machine learning (ML) models.

The impact of increasing the number of features (8, 12, 16, 20) was investigated to assess how varying levels of feature dimensionality influence the performance of the OWOD model. The best-performing OWOD model, identified based on the optimal number of features, was compared with several ML models, including GB, DT, neural network (NN), and SVM. For training and testing processes, model performance was assessed using standard evaluation metrics: accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). A 5-fold cross-validation was applied to the training dataset (80%) to ensure robust validation, and the results were compared with the testing performance on the remaining 20% of the data.

The OWOD algorithm builds upon the previously established WOD method requiring predefined feature ranges and target values, which must be manually specified based on expert knowledge. The OWOD eliminates this requirement by directly deriving the feature ranges and target values from the dataset through an objective function. This data-driven approach makes the OWOD particularly well suited for classifying dementia and heart failure, where biomarkers and their optimal ranges are often poorly established. Consequently, the WOD is not applicable for direct comparison with the OWOD in this context.

4. Results and Discussion

This section presents the results of the proposed measurement for classifying groups of patients with dementia and heart failure. The input data used for training and validating consisted of 8000 records containing risk factors and blood biomarker features for both groups. These records were used to calculate the OWOD values and train the ML models. The evaluation was performed using 2000 unseen records to assess the classification performance of both the OWOD approach and the ML models.

4.1. Optimal Feature Dimension Results

Table 5 presents the results of the OWOD classification model using varying numbers of selected features, based on 8000 sample records representing 80% of the dataset.

Table 5. Comparison of optimized weighted objective distance classifications by feature selection number.

No. of Features	Accuracy %	Precision %	Recall %	F1-Score %	AUC-ROC %
8	56.38 ± 1.94	85.05 ± 3.04	15.68 ± 5.13	26.44 ± 0.97	56.40 ± 0.02
12	61.26 ± 1.52	89.26 ± 1.84	25.62 ± 3.48	39.81 ± 0.76	61.40 ± 0.016
16	70.43 ± 0.28	86.61 ± 0.86	48.34 ± 0.86	62.05 ± 0.14	71.30 ± 0.003
20	94.95 ± 0.96	95.64 ± 0.95	94.20 ± 1.11	94.91 ± 0.48	96.60 ± 0.013

As seen in Table 5, the model's performance improved with an increasing number of features. At eight features, its performance was limited by low recall (15.68%) and F1-score (26.44%), indicating poor sensitivity. As the number increased to 12 and 16, recall and F1-score improved notably, reflecting enhanced balance. With 20 features, all metrics—accuracy (94.95%), precision (95.64%), recall (94.20%), F1-score (94.91%), and AUC (96.60%)—reached optimal levels. This demonstrates that higher feature dimensionality significantly boosts the model's effectiveness, particularly in detecting minority class instances. The comparison of the ROC curves is illustrated in Figure 9.
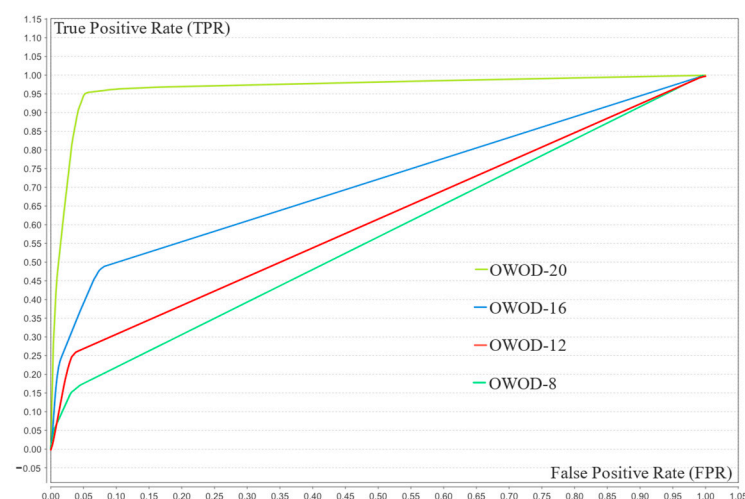

Figure 9. Receiver operating characteristic (ROC) chart of optimized weighted objective distance (OWOD) 8–20 features comparing the OWOD with 8 features (green), 12 features (red), 16 features (blue), and 20 features (lime).

The OWOD configuration with 20 features achieved the highest and most stable TPR across all FPR values, suggesting that 20 features represent the optimal balance between model complexity and performance.

4.2. OWOD Classification Results

Table 6 shows the results of the OWOD classification using 20 features. These are labeled as "dementia" for positive class groups and "heart failure" for negative groups.

Table 6. Optimized weighted objective distance (OWOD) classification results.

No.	Weight								OWOD Result	Classification Result	Actual Result	Matching Result
	SBP	DBP	W	FBS	TGS	TC	HDL	LDL				Correct /Incorrect
1	0.0003	0.0003	0.1404	0.0531	0.2095	0.1010	0.2582	0.2371	0.4051	Dementia	Dementia	Correct
2	0.1034	0.0003	0.1117	0.1216	0.1484	0.1484	0.1995	0.1667	0.4258	Dementia	Dementia	Correct
3	0.2727	0.0004	0.0004	0.0611	0.2484	0.0004	0.2409	0.1757	0.4057	Dementia	Dementia	Correct
4	0.1920	0.2168	0.0003	0.0467	0.2251	0.0003	0.1843	0.1344	0.5858	Dementia	Heart failure	Incorrect
5	0.2555	0.0005	0.2705	0.1849	0.0005	0.2872	0.0005	0.0005	0.3583	Dementia	Dementia	Correct
6	0.1356	0.0960	0.1301	0.0815	0.1149	0.1476	0.1475	0.1468	0.4822	Heart failure	Heart failure	Correct
7	0.1364	0.1225	0.0772	0.1281	0.1417	0.1378	0.1152	0.1411	0.4411	Heart failure	Heart failure	Correct
8	0.1465	0.1284	0.1137	0.0654	0.1207	0.1366	0.1444	0.1444	0.8029	Heart failure	Heart failure	Correct
9	0.1231	0.1257	0.1283	0.0862	0.1181	0.1337	0.1435	0.1414	0.6858	Heart failure	Heart failure	Correct
10	0.1456	0.1446	0.1158	0.0511	0.1192	0.1229	0.1515	0.1493	0.5363	Heart failure	Heart failure	Correct
...	...	...	...	...	...	...	...	...	...	...	...	...
8000	0.1265	0.1292	0.1144	0.0973	0.1374	0.1081	0.1497	0.1374	0.4680	Heart failure	Dementia	Incorrect

These results demonstrate the $OWOD_i$ value and weight information for each person. The weight information indicates the degree to which a feature contributes to developing target conditions by representing the actual effects of each feature. Weight information with a value of 0 refers to an insignificant feature indicating control at the accepted level. Accordingly, the insignificant feature was eliminated to enhance classification performance. Weight information with a value of more than 0 refers to a significant feature affecting dementia development.

In addition, information gain computed from a reduction in entropy was employed to assign weights to each feature. The weighted features represent a degree of feature priority affecting classification. The summation of the OWOD of all features can be utilized to classify groups. Older adults with $OWOD \leq OWOD_c$ can be considered as belonging to the positive group (dementia), and those with $OWOD > OWOD_c$ to the negative group (heart failure). Thus, Person No. 1, with $OWOD = 0.4051$ and $OWOD_c = 0.4324$, was classified as belonging to the positive group.

Confusion Matrix Evaluation

The classification performance of the OWOD was also evaluated using a confusion matrix, as shown in Table 7. The classified class was derived from the proposed measurement, and the actual class was obtained from the doctor's decision record. TP denotes true positive, representing older adults correctly classified as belonging to the OC group. TN denotes true negative, representing those correctly classified as belonging to the NC group. FP denotes a false positive, representing those incorrectly predicted as being in the OC group. FN denotes a false negative, representing those incorrectly predicted as being in the NC group.

Table 7. Confusion matrix of the proposed optimized weighted objective distance.

	Dementia	Heart Failure	Class Precision
Predict—Dementia	3768 (TP)	172 (FP)	95.63%
Predict—Heart failure	232 (FN)	3828 (TN)	94.29%
Class recall	94.20%	95.70%	

The confusion matrix of the proposed OWOD investigating 8000 records reveals that 7596 people were correctly classified. This suggests that prioritizing features by applying different weights obtained from information gain to eliminate the features that do not influence the classification is feasible. In total, 404 individuals were incorrectly classified. Variability in a particular feature measurement may account for these inaccurate results. For instance, blood pressure was recorded manually; thus, the circumstances during measurement could affect exactness. Recent caffeine use or smoking can raise SBP readings above the baseline. These findings demonstrate the possibility of misclassifications and emphasize the difficulty in classifying cases with shared overlapping risk factors, especially when the input features lack adequate discriminatory power.

4.3. Model Validation Results

This study used five-fold cross-validation to verify the OWOD model. A comparison of the OWOD and ML classification models for the validation process is shown in Table 8.

Table 8. Comparison of the validation results of the optimized weighted objective distance (OWOD) and other machine learning models.

Classification Method	No. of Features	Accuracy %	Precision %	Recall %	F1-Score %	AUC-ROC %
OWOD	20	94.95 ± 0.96	95.64 ± 0.95	94.20 ± 1.11	94.91 ± 0.48	96.60 ± 0.013
Gradient boosting (GB)	20	88.58 ± 0.77	91.34 ± 1.16	85.26 ± 1.38	88.19 ± 0.39	92.90 ± 0.006
Decision tree (DT)	20	86.75 ± 0.72	91.83 ± 1.51	80.72 ± 1.59	85.90 ± 0.36	88.10 ± 0.005
Support vector machine (SVM)	20	84.96 ± 1.03	85.81 ± 1.03	83.78 ± 1.91	84.78 ± 0.52	90.70 ± 0.007
Neural network (NN)	20	83.34 ± 0.87	88.19 ± 2.41	77.08 ± 1.59	82.23 ± 0.44	89.30 ± 0.006

The validation results demonstrate the superior performance of the OWOD compared with traditional and ensemble ML methods using 20 features. The OWOD achieved the highest overall performance across all key metrics, with an accuracy of 94.95% ± 0.96, a precision of 95.64% ± 0.95, a recall of 94.20% ± 1.11, an F1-score of 94.91% ± 0.48, and an AUC-ROC of 96.60% ± 0.013. GB, while relatively precise (91.34% ± 1.16), underperformed in recall and F1-score, indicating limitations in sensitivity. The DT, SVM, and NN models exhibited even lower recall and AUC-ROC values, suggesting a weaker ability to detect positive dementia cases accurately. These findings underscore the robustness and effectiveness of the OWOD, particularly its capacity to maintain high discrimination power, which is critical for reliable medical classification.

Figure 10 presents the ROC curves comparing the classification performance of five models—OWOD, GB, NN, DT, and SVM—using 20 features. The OWOD (green curve) outperformed all other methods with a consistently higher true positive rate (TPR) across nearly all false positive rates (FPRs), approaching a near-perfect classification boundary. The traditional ML models, particularly SVM and DT, demonstrated lower TPRs at equivalent FPRs, suggesting a reduced ability to distinguish between classes. This indicates that the OWOD offers a more effective solution for accurate classification in this context.

To ensure fair and consistent model comparisons, each method was implemented with carefully selected hyperparameters and evaluated using five-fold cross-validation. The hyperparameter details of both the ML and OWOD classifications are presented in Table 9. These configurations, outlined in Table 10, were selected based on grid search to ensure reliable and reproducible performance evaluation across models.
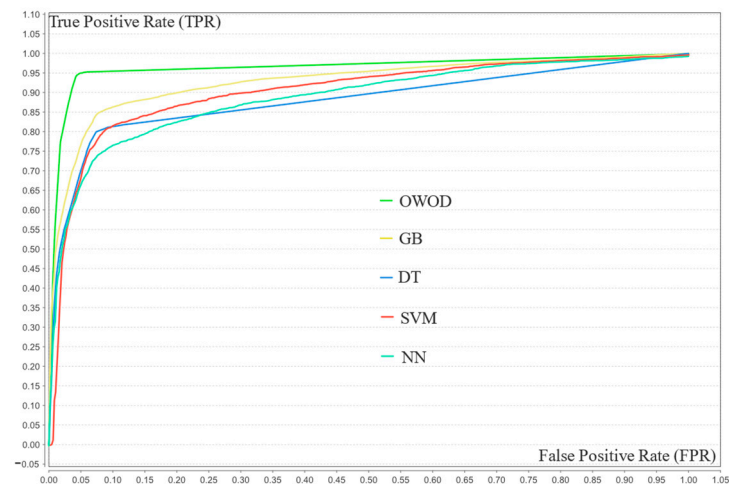
Figure 10. Receiver operating characteristic (ROC) curve comparison of the validation results of the optimized weighted objective distance (OWOD) and other machine learning models.

Table 9. Summary of the hyperparameters of the machine learning classifications.

Method	Hyperparameters
Gradient boosting	Number of trees = 50; maximal depth = 5; min rows = 10.0; number of bins = 20; learning rate = 0.01; sample rate = 1.0; cross-validation folds = 5
Decision tree	Criterion = gain ratio; maximal depth = 10; confidence = 0.1; minimal gain = 0.01; minimal leaf size = 2; cross-validation folds = 5
Support vector machine	Kernel type = dot; kernel cache = 200; C = 0.0; convergence epsilon = 0.001; max iterations = 100,000; cross-validation folds = 5
Neural network	Hidden layer = 2; training cycle = 200; learning rate = 0.01; momentum = 0.9; cross-validation folds = 5
OWOD	Cross-validation folds = 5

4.4. Statistical Significance Comparison

Statistical validation of the model's performance was carried out by pairing classifiers. A McNemar's test was applied to the paired predictions of all models. The significance matrix is illustrated in Figure 11.
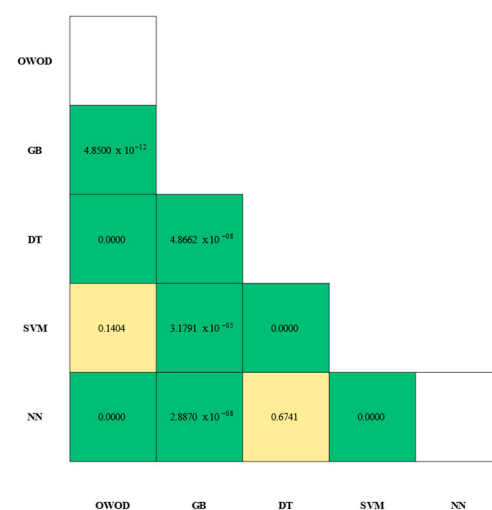
Figure 11. Statistical significance matrix from McNemar's tests comparing model classifier pairs. Each cell represents the *p*-value of the comparison between two models. Green cells indicate statistically significant differences ($p < 0.01$), whereas yellow cells indicate non-significant differences ($p \geq 0.01$).

The results showed a statistically significant value for OWOD–GB ($p = 4.85 \times 10^{-12}$) in which the OWOD had the first-place performance (OWOD AUC = 96.60%; GB AUC = 92.90%), with GB's performance showing significant differences against all models ($p < 0.01$). In contrast, there were some overlapping efficiencies in the classification of OWOD–SVM ($p = 0.1404$) and DT–NN ($p = 0.6741$).

4.5. Model Comparison and Discussion

A performance comparison of the OWOD and other ML-based methods is presented in Table 10. These results were obtained from the evaluation using 20% unseen data.

Table 10. Comparison of the testing results of the optimized weighted objective distance (OWOD) and other machine learning models.

Classification Method	No. of Features	Accuracy %	Precision %	Recall %	F1-Score %	AUC-ROC %
OWOD	20	95.45	96.14	94.70	95.42	97.10
Gradient boosting (GB)	20	88.20	90.90	84.90	87.80	92.40
Decision tree (DT)	20	86.20	91.04	80.30	85.33	87.30
Support vector machine (SVM)	20	84.40	84.89	83.70	84.29	90.10
Neural network (NN)	20	83.75	88.14	78.00	82.76	88.80

The testing results confirm the superior classification performance of the OWOD compared with conventional ML algorithms. The OWOD achieved the highest scores across all evaluation metrics, with an accuracy of 95.45%, a precision of 96.14%, a recall of 94.70%, an F1-score of 95.42%, and an AUC-ROC of 97.10%. This reflects the ability of the OWOD to accurately identify positive and negative classes, maintaining a high balance between sensitivity and specificity. GB demonstrated reasonably strong precision (90.90%) but a poorer recall (84.90%) and F1-score (87.80%), indicating a greater likelihood of missing TP cases. Traditional models such as DT, SVM, and NN exhibited even lower recall and AUC-ROC values, indicating limited effectiveness in differentiating between classes. These findings demonstrate that the OWOD delivers high classification performance and offers enhanced generalization capability when applied to previously unseen data.

Evaluation of Model Performances

The results from both the validation and testing phases demonstrated the superior performance and robustness of the OWOD classification model over conventional ensemble and traditional ML models. During validation, the OWOD achieved the highest metrics across all performance indicators (accuracy 94.95% $\pm$ 0.96; precision 95.64% $\pm$ 0.95; recall 94.20% $\pm$ 1.11; F1-score 94.91% $\pm$ 0.48; AUC-ROC 96.60% $\pm$ 0.013). The OWOD maintained top-tier performance in the testing phase with an accuracy of 95.45%, a precision of 96.14%, a recall of 94.70%, an F1-score of 95.42%, and an AUC-ROC of 97.10%, demonstrating excellent generalization to unseen data.

While ensemble methods such as GB are known for improving accuracy by combining multiple weak learners, the OWOD surpassed GB in all metrics. This suggests that the OWOD's class-wise feature interpretation and target-based thresholding mechanism enable a more refined and interpretable classification boundary. Despite leveraging ensemble principles, GB showed relatively lower recall (85.26% $\pm$ 1.38 in validation; 84.90% in testing), indicating potential underperformance in identifying TP cases—an important consideration in health-related applications where sensitivity is crucial.

Traditional classifiers such as DT, SVM, and NN exhibited further limitations, with lower recall, F1-score, and AUC-ROC values in both validation and testing. While computationally efficient, these models often struggle with complex decision boundaries and noisy/overlapping feature distributions. The OWOD incorporates a distribution-based

entropy framework and utilizes optimal thresholds derived from class-wise histograms, allowing for the capture of fine-grained patterns that rigid or margin-based decision mechanisms in DT, SVM, and NN may overlook. In this context, the OWOD outperformed ensemble and traditional classifiers in predictive accuracy and robustness.

It can therefore be concluded that the proposed OWOD has a higher potential for classifying groups of individuals with dementia and heart failure compared with other ML classification methods. The assumption of this study is consequently verified, in that the proposed measurement can improve classification performance. In addition, the OWOD can handle complex attributes with multiple conditions during the attribute weighting process, which further enhances performance.

4.6. Suggestions and Future Study

This study demonstrated the effectiveness of the proposed OWOD classification method in classifying individuals with dementia and heart failure using blood biomarker features. The model achieved strong performance across all evaluation metrics, underscoring its potential as a practical, noninvasive, and widely accessible decision support tool. To support clinical adoption, future research should focus on validating the model across diverse populations and healthcare settings, as well as exploring additional blood-based biomarkers to further enhance performance. While these findings are promising, the exclusion of individuals diagnosed with both dementia and heart failure in order to maintain clear class boundaries is a limitation and may reduce applicability in real-world settings. Incorporating more granular clinical information and multi-label classification approaches would allow for more sophisticated modeling and improved classification performance among complex patient populations.

The OWOD offers clear interpretability, as its decisions are based on the distance from target feature values, allowing clinicians to understand the contribution of each factor. Additionally, the OWOD can offer personalized recommendations according to each individual's significant features after classification as efficiently as healthcare providers. These features could be used to guide lifestyle changes on a broad scale. As a result, older adults could better monitor their health status, which may lead to dementia/heart failure prevention. Potential barriers to real-world implementation include validation across diverse populations, seamless integration with existing EHR systems, and acceptance by clinicians and patients.

5. Conclusions

This study proposes the OWOD as a novel measurement to distinguish groups of patients with dementia and heart failure. The OWOD is a modification of the original WOD using entropy-based weighting features and an objective function from class-wise histogram differences, promoting a data-driven approach. The weighted features indicate a degree of feature priority affecting classification and eliminate features that do not influence classification. This study evaluated the proposed OWOD using 20 features and 10,000 EHRs and compared it with other ML classification models, including GB, DT, SVM, and NN. The OWOD outperformed the other ML models during validation and testing processes for overall performance comparison in terms of accuracy, precision, recall, F1-scores, and AUC-ROC. Future studies should include more multi-site datasets and refine features to ensure generalization and practical applicability.

References

1. Livingston, G.; Huntley, J.; Liu, K.Y.; Costafreda, S.G.; Selbæk, G.; Alladi, S.; Ames, D.; Banerjee, S.; Burns, A.; Brayne, C.; et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing commission. *Lancet* **2024**, *404*, 572–628. [CrossRef] [PubMed]

2. Zhang, Y.; Gao, D.; Liang, J.; Ji, M.; Zhang, W.; Pan, Y.; Zheng, F.; Xie, W. The role of life's crucial 9 in cardiovascular disease incidence and dynamic transitions to dementia. *Commun. Med.* **2025**, *5*, 223. [CrossRef] [PubMed]

3. Ye, S.; Huynh, Q.; Potter, E.L. Cognitive dysfunction in heart failure: Pathophysiology and implications for patient management. *Comorbidities* **2022**, *19*, 303–315. [CrossRef]

4. Liu, J.; Xiao, G.; Liang, Y.; He, S.; Lyu, M.; Zhu, Y. Heart–brain interaction in cardiogenic dementia: Pathophysiology and therapeutic potential. *Front. Cardiovasc. Med.* **2024**, *11*, 1304864. [CrossRef]

5. Khan, S.S.; Matsushita, K.; Sang, Y.; Ballew, S.H.; Grams, M.E.; Surapaneni, A.; Blaha, M.J. Development and validation of the American Heart Association's PREVENT equations. *Circulation* **2024**, *149*, 430–449. [CrossRef] [PubMed]

6. World Health Organization. Dementia. WHO Fact Sheets. 31 March 2025. Available online: https://www.who.int/news-room/fact-sheets/detail/dementia (accessed on 8 September 2025).

7. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **2022**, *7*, e105–e125. [CrossRef]

8. Sohn, M.; Yang, J.; Sohn, J.; Lee, J. Digital healthcare for dementia and cognitive impairment: A scoping review. *Int. J. Nurs. Stud.* **2023**, *140*, 104413. [CrossRef]

9. Stoppa, E.; Di Donato, G.W.; Parde, N.; Santambrogio, M.D. Computer-aided dementia detection: How informative are your features? In Proceedings of the IEEE 7th Forum on Research and Technologies for Society and Industry Innovation (RTSI), Paris, France, 24–26 August 2022. [CrossRef]

10. Chaiyo, Y.; Temdee, P. A comparison of machine learning methods with feature extraction for classification of patients with dementia risk. In Proceedings of the Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Chiang Rai, Thailand, 26–28 January 2022. [CrossRef]

11. Hu, S.; Yang, C.; Luo, H. Current trends in blood biomarker detection and imaging for Alzheimer's disease. *Biosens. Bioelectron.* **2022**, *210*, 114278. [CrossRef]

12. Savarese, G.; Becher, P.M.; Lund, L.H.; Seferovic, P.; Rosano, G.M.C.; Coats, A.J.S. Global burden of heart failure: A comprehensive and updated review of epidemiology. *Cardiovasc. Res.* **2022**, *118*, 3272–3287. [CrossRef]

13. Li, R.; Wang, X.; Lawler, K.; Garg, S.; Bai, Q.; Alty, J. Applications of artificial intelligence to aid early detection of dementia: A scoping review on current capabilities and future directions. *J. Biomed. Inform.* **2022**, *127*, 104030. [CrossRef] [PubMed]

14. Mahendran, N.; Durai Raj Vincent, P.M. A deep learning framework with an embedded-based feature selection approach for the early detection of Alzheimer's disease. *Comput. Biol. Med.* **2022**, *141*, 105056. [CrossRef]

15. Ahmed, H.; Soliman, H.; Elmogy, M. Early detection of Alzheimer's disease using single nucleotide polymorphisms analysis based on gradient boosting tree. *Comput. Biol. Med.* **2022**, *146*, 105622. [CrossRef]

16. Park, J.E.; Kim, H.J.; Kim, Y.E.; Jang, H.; Cho, S.H.; Kim, S.J.; Na, D.L.; Won, H.; Ki, C.; Seo, S.W. Analysis of dementia-related gene variants in APOE ε4 noncarrying Korean patients with early-onset Alzheimer's disease. *Neurobiol. Aging* **2020**, *85*, 155.e5–155.e8. [CrossRef]

17. El-Sappagh, S.; Ali, F.; Abuhmed, T.; Singh, J.; Alonso, J.M. Automatic detection of Alzheimer's disease progression: An efficient information fusion approach with heterogeneous ensemble classifiers. *Neurocomputing* **2022**, *512*, 203–224. [CrossRef]

18. Shanmugam, J.V.; Duraisamy, B.; Simon, B.C.; Bhaskaran, P. Alzheimer's disease classification using pre-trained deep networks. *Biomed. Signal Process. Control* **2022**, *71*, 103217. [CrossRef]

19. Di Benedetto, M.; Carrara, F.; Tafuri, B.; Nigro, S.; De Blasi, R.; Falchi, F.; Gennaro, C.; Gigli, G.; Logroscino, G.; Amato, G. Deep networks for behavioral variant frontotemporal dementia identification from multiple acquisition sources. *Comput. Biol. Med.* **2022**, *148*, 105937. [CrossRef] [PubMed]

20. Rallabandi, V.P.S.; Seetharaman, K. Deep learning-based classification of healthy aging controls, mild cognitive impairment and Alzheimer's disease using fusion of MRI-PET imaging. *Biomed. Signal Process. Control* **2023**, *80*, 104312. [CrossRef]

21. Li, J.P.; Haq, A.U.; Din, S.U.; Khan, J.; Khan, A.; Saboor, A. heart disease identification method using machine learning classification in e-healthcare. *IEEE Access* **2020**, *8*, 107562–107582. [CrossRef]

22. Ahmad, A.A.; Polat, H. Prediction of heart disease based on machine learning using jellyfish optimization algorithm. *Diagnostics* **2023**, *13*, 2392. [CrossRef] [PubMed]

23. Aslam, M.U.; Xu, S.; Hussain, S.; Waqas, M.; Abiodun, N.L. Machine learning based classification of valvular heart disease using cardiovascular risk factors. *Sci. Rep.* **2024**, *14*, 24396. [CrossRef] [PubMed]

24. Chulde-Fernández, B.; Enríquez-Ortega, D.; Tirado-Espín, A.; Vizcaíno-Imacaña, P.; Guevara, C.; Navas, P.; Villalba-Meneses, F.; Cadena-Morejon, C.; Almeida-Galarraga, D.; Acosta-Vargas, P. Classification of heart failure using machine learning: A comparative study. *Life* **2025**, *15*, 496. [CrossRef]

25. Yongcharoenchaiyasit, K.; Arwatchananukul, S.; Temdee, P.; Prasad, R. Gradient boosting-based model for elderly heart failure, aortic stenosis, and dementia classification. *IEEE Access* **2023**, *11*, 48677–48696. [CrossRef]

26. Yongcharoenchaiyasit, K.; Arwatchananukul, S.; Hristov, G.; Temdee, P. Enhanced multi-model machine learning-based dementia detection using a data enrichment framework: Leveraging the blessing of dimensionality. *Bioengineering* **2025**, *12*, 592. [CrossRef]

27. Chaising, S.; Temdee, P.; Prasad, R. Weighted objective distance for the classification of elderly people with hypertension. *Knowl.-Based Syst.* **2020**, *210*, 106441. [CrossRef]

28. Nuankaew, P.; Chaising, S.; Temdee, P. Average weighted objective distance-based method for type 2 diabetes prediction. *IEEE Access* **2021**, *9*, 49018–49031. [CrossRef]

29. Daugherty, A.M. Hypertension-related risk for dementia: A summary review with future directions. *Semin. Cell Dev. Biol.* **2021**, *116*, 82–89. [CrossRef]

30. Brown, J.; Xiong, X.; Wu, J.; Li, M.; Lu, Z.K. CO76 Hypertension and the risk of dementia among medicare beneficiaries in the U.S. *Value Health* **2022**, *25*, S318. [CrossRef]

31. Campbell, G.; Jha, A. Dementia risk and hypertension: A literature review. *J. Neurol. Sci.* **2019**, *405*, 146. [CrossRef]

32. Park, K.; Nam, G.; Han, K.; Hwang, H. Body weight variability and the risk of dementia in patients with type 2 diabetes mellitus: A nationwide cohort study in Korea. *Diabetes Res. Clin. Pract.* **2022**, *190*, 110015. [CrossRef] [PubMed]

33. Iwagami, M.; Qizilbash, N.; Gregson, J.; Douglas, I.; Johnson, M.; Pearce, N.; Evans, S.; Pocock, S. Blood cholesterol and risk of dementia in more than 1.8 million people over two decades: A retrospective cohort study. *Lancet Healthy Longev.* **2021**, *2*, e498–e506. [CrossRef] [PubMed]

34. Gong, J.; Harris, K.; Peters, S.A.E.; Woodward, M. Serum lipid traits and the risk of dementia: A cohort study of 254,575 women and 214,891 men in the UK Biobank. *Lancet* **2022**, *54*, 101695. [CrossRef]

35. Bertini, F.; Allevi, D.; Lutero, G.; Calzà, L.; Montesi, D. An automatic Alzheimer's disease classifier based on spontaneous spoken English. *Comput. Speech Lang.* **2022**, *72*, 101298. [CrossRef]

36. Milana, S. Dementia classification using attention mechanism on audio data. In Proceedings of the 2023 IEEE 21st World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia, 19–21 January 2023. [CrossRef]

37. Meng, X.; Fang, S.; Zhang, S.; Li, H.; Ma, D.; Ye, Y.; Su, J.; Sun, J. Multidomain lifestyle interventions for cognition and the risk of dementia: A systematic review and meta-analysis. *Int. J. Nurs. Stud.* **2022**, *130*, 104236. [CrossRef] [PubMed]

38. Cilia, N.D.; De Gregorio, G.; De Stefano, C.; Fontanella, F.; Marcelli, A.; Parziale, A. Diagnosing Alzheimer's disease from on-line handwriting: A novel dataset and performance benchmarking. *Eng. Appl. Artif. Intell.* **2022**, *111*, 104822. [CrossRef]

39. Masuo, A.; Ito, Y.; Kanaiwa, T.; Naito, K.; Sakuma, T.; Kato, S. Dementia screening based on SVM using qualitative drawing error of clock drawing test. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, UK, 11–15 July 2022. [CrossRef]

40. Wong, R.; Lovier, M.A. Sleep disturbances and dementia risk in older adults: Findings from 10 years of national U.S. prospective data. *Am. J. Prev. Med.* **2023**, *64*, 781–787. [CrossRef] [PubMed]

41. Khosroazad, S.; Abedi, A.; Hayes, M.J. Sleep signal analysis for early detection of Alzheimer's disease and related dementia (ADRD). *IEEE J. Biomed. Health Inform.* **2023**, *27*, 3235391. [CrossRef]

42. Chiu, S.Y.; Wyman-Chick, K.A.; Ferman, T.J.; Bayram, E.; Holden, S.K.; Choudhury, P.; Armstrong, M.J. Sex differences in dementia with Lewy bodies: Focused review of available evidence and future directions. *Park. Relat. Disord.* **2023**, *107*, 105285. [CrossRef]

43. Alvi, A.M.; Siuly, S.; Wang, H.; Wang, K.; Whittaker, F. A deep learning based framework for diagnosis of mild cognitive impairment. *Knowl.-Based Syst.* **2022**, *248*, 108815. [CrossRef]

44. Lin, H.; Tsuji, T.; Kondo, K.; Imanaka, Y. Development of a risk score for the prediction of incident dementia in older adults using a frailty index and health checkup data: The JAGES longitudinal study. *Prev. Med.* **2018**, *112*, 88–96. [CrossRef]

45. Zhang, Y.; Xu, W.; Zhang, W.; Wang, H.; Ou, Y.; Qu, Y.; Shen, X.; Chen, S.; Wu, K.; Zhao, Q.; et al. Modifiable risk factors for incident dementia and cognitive impairment: An umbrella review of evidence. *J. Affect. Disord.* **2022**, *314*, 160–167. [CrossRef]

46. Sigala, E.G.; Vaina, S.; Chrysohoou, C.; Dri, E.; Damigou, E.; Tatakis, F.P.; Sakalidis, A.; Barkas, F.; Liberopoulos, E.; Sfikakis, P.P.; et al. Sex-related differences in the 20-year incidence of CVD and its risk factors: The ATTICA study (2002–2022). *Am. J. Prev. Cardiol.* **2024**, *19*, 100709. [CrossRef] [PubMed]

47. Goldsborough III, E.; Osuji, N.; Blaha, M.J. Assessment of cardiovascular disease risk: A 2022 update. *Endocrinol. Metab. Clin. N. Am.* **2022**, *51*, 483–509. [CrossRef]

48. Young, K.A.; Scott, C.G.; Rodeheffer, R.J.; Chen, H.H. Incidence of preclinical heart failure in a community population. *J. Am. Heart Assoc.* **2022**, *11*, e025519. [CrossRef] [PubMed]

49. Sevilla-Salcedo, C.; Imani, V.; Olmos, P.M.; Gómez-Verdejo, V.; Tohka, J. Multi-task longitudinal forecasting with missing values on Alzheimer's disease. *Comput. Methods Programs Biomed.* **2022**, *226*, 107056. [CrossRef]

50. Kherchouche, A.; Ben-Ahmed, O.; Guillevin, C.; Tremblais, B.; Julian, A.; Fernandez-Maloigne, C.; Guillevin, R. Attention-guided neural network for early dementia detection using MRS data. *Comput. Med. Imag. Graph.* **2022**, *99*, 102074. [CrossRef]

51. Ahmad, M.; Ahmed, A.; Hashim, H.; Farsi, M.; Mahmoud, N. Enhancing heart disease diagnosis using ECG signal reconstruction and deep transfer learning classification with optional SVM integration. *Diagnostics* **2025**, *15*, 1501. [CrossRef]

52. Menagadevi, M.; Mangai, S.; Madian, N.; Thiyagarajan, D. Automated prediction system for Alzheimer detection based on deep residual autoencoder and support vector machine. *Optik* **2023**, *272*, 170212. [CrossRef]

53. Darmawahyuni, A.; Firdaus, S.N.; Tutuko, B.; Yuwandini, M.; Rachmatullah, M.N. Congestive heart failure waveform classification based on short time-step analysis with recurrent network. *Informat. Med. Unlocked* **2020**, *21*, 100441. [CrossRef]

54. Chaiyo, Y.; Rueangsirarak, W.; Hristov, G.; Temdee, P. Improving early detection of dementia: Extra trees-based classification model using inter-relation-based features and K-means synthetic minority oversampling technique. *Big Data Cogn. Comput.* **2025**, *9*, 148. [CrossRef]

55. Phanbua, P.; Arwatchananukul, S.; Temdee, P. Classification model of dementia and heart failure in older adults using extra trees and oversampling-based technique. In Proceedings of the 2025 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON), Nan, Thailand, 29 January–1 February 2025. [CrossRef]

56. Dhar, A.; Dash, N.; Roy, K. Classification of text documents through distance measurement: An experiment with multi-domain Bangla text documents. In Proceedings of the 2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA) (Fall), Dehradun, India, 15–16 September 2017. [CrossRef]

57. Greche, L.; Jazouli, M.; Es-Sbai, N.; Majda, A.; Zarghili, A. Comparison between Euclidean and Manhattan distance measure for facial expressions classification. In Proceedings of the 2017 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), Fez, Morocco, 19–20 April 2017. [CrossRef]

58. Han, J.; Park, D.C.; Woo, D.M.; Min, S.Y. Comparison of distance measures on Fuzzyc-means algorithm for image classification problem. *AASRI Procedia* **2013**, *4*, 50–56. [CrossRef]

59. Chaichumpa, S.; Temdee, P. Assessment of student competency for personalised online learning using objective distance. *Int. J. Innov. Technol. Explor. Eng.* **2018**, *23*, 19–36. [CrossRef]

60. Chaising, S.; Temdee, P. Determining recommendations for preventing elderly people from cardiovascular disease complication using objective distance. In Proceedings of the 2018 Global Wireless Summit (GWS), Chiang Rai, Thailand, 25–28 November 2018. [CrossRef]

61. Chaising, S.; Prasad, R.; Temdee, P. Personalized recommendation method for preventing elderly people from cardiovascular disease complication using integrated objective distance. *Wirel. Pers. Commun.* **2019**, *117*, 215–233. [CrossRef]

62. Alhaj, T.A.; Siraj, M.M.; Zainal, A.; Elshoush, H.T.; Elhaj, F. Feature selection using information gain for improved structural-based alert correlation. *PLoS ONE* **2016**, *11*, e0166017. [CrossRef] [PubMed]

63. Pereira, R.B.; Carvalho, A.P.D.; Zadrozny, B.; Merschmann, L.H.D.C. Information gain feature selection for multi-label classification. *J. Inf. Data Manag.* **2015**, *6*, 48–58.

64. Azhagusundari, B.; Thanamani, A.S. Feature selection based on information gain. *Int. J. Innov. Technol. Explor. Eng.* **2013**, *2*, 18–21.

65. Patil, L.H.; Atique, M. A novel feature selection based on information gain using WordNet. In Proceedings of the 2013 International Conference on Information and Network Security (ICINS 2013), London, UK, 7–9 October 2013.

66. Baobao, B.; Jinsheng, M.; Minru, S. An enhancement of K-Nearest Neighbor algorithm using information gain and extension relativity. In Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis, Beijing, China, 21–24 April 2008. [CrossRef]

67. Gupta, M. Dynamic k-NN with attribute weighting for automatic web page classification (Dk-NNwAW). *Int. J. Comput. Appl.* **2012**, *58*, 34–40. [CrossRef]