Educational Administration: Theory and Practice
2024, 30(11), 1395-1404
ISSN: 2148-2403
https://kuey.net/

Research Article

Early Alzheimer's Detection Using Weighted Correlation-Based Feature Selection (Weighted Cfs)

R. Malarvizhi[1*], Dr. R. Rangaraj[2]

[1*]Assistant Professor, Department of Computer Science, N.M.S.S.Vellaichamy Nadar College, Madurai.
[2]Professor & Head, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore.

Citation: R. Malarvizhi, et.al (2024), Early Alzheimer's Detection Using Weighted Correlation-Based Feature Selection (Weighted Cfs), *Educational Administration: Theory and Practice, 30(11)* 1395-1404
Doi: 10.53555/kuey.v30i11.9466

ARTICLE INFO

ABSTRACT

Feature selection plays a pivotal role in improving the accuracy and interpretability of machine learning models, particularly in healthcare applications like Alzheimer's disease detection. In this study, we present a Weighted Correlation-based Feature Selection (Weighted CFS) approach for early Alzheimer's detection. Weighted CFS extends the traditional CFS method by incorporating domain-specific weights to prioritize clinically significant features while balancing feature relevance and redundancy. This methodology effectively integrates expert knowledge with statistical correlations, making it ideal for handling high-dimensional and complex datasets. By assigning higher importance to features such as cognitive test scores (e.g., MMSE) and demographic factors, the approach ensures that selected features are both domain-relevant and statistically informative. Experimental results demonstrate that Weighted CFS improves model performance, reduces feature redundancy, and enhances interpretability, offering a robust framework for identifying critical predictors in early Alzheimer's detection.

Keywords: Alzheimer's disease, data mining, early detection, feature selection, Correlation-based Feature Selection (CFS), cognitive test scores.

1. Introduction

A neurological disease that primarily affects the elderly, Alzheimer's disease (AD) has become a serious global health concern. AD is becoming more and more common, and as a result, early detection is essential for successful intervention and better patient outcomes. Novel approaches to discovering early signs of Alzheimer's disease have been made possible by the incorporation of data mining tools into healthcare research. The present research focuses on how feature selection plays a crucial role in improving the efficacy and precision of data mining approaches for early Alzheimer's disease identification. Because Alzheimer's disease has many facets, diagnosing it can be difficult and frequently necessitates examining a range of biological and cognitive indicators. With the advancement of technology, a growing amount of data has become available for analysis, ranging from neuroimaging results to clinical assessments. Discovering hidden patterns and insights inside large datasets is made possible by data mining, a field that combines machine learning, statistical analysis, and database systems.

Using these methods can greatly aid in identifying the critical characteristics that set apart those with moderate cognitive impairment (MCI) or those in the early stages of Alzheimer's from cognitively normal people. A key component of data mining is feature selection, which is locating and extracting pertinent variables from a broader collection of possible predictors. To extract the most pertinent data for prompt and accurate detection, however, extensive feature selection techniques are needed due to the overwhelming volume and complexity of medical data.

The research aims to investigate several data mining methods used in feature selection for early Alzheimer's detection and evaluate how well they find relevant indicators. In addition, it will examine the difficulties posed by conventional feature selection techniques and suggest novel strategies that take the multidisciplinary nature of Alzheimer's research into account. The objective is to offer insights into the changing landscape of feature selection in the context of early Alzheimer's detection by critically analyzing the body of current literature and recent developments. This will hopefully point to interesting directions for future research and clinical application. The incorporation of efficient feature selection algorithms is crucial to releasing the full potential of data mining in advancing early Alzheimer's detection and, eventually,

improving patient outcomes, as the healthcare community increasingly realizes the benefits of data-driven approaches.

1.1 Feature Selection as a Key Component

A key component in improving the accuracy and effectiveness of data mining algorithms is feature selection. A plethora of factors, from neuroimaging indicators to cognitive tests, need to be carefully considered when it comes to Alzheimer's detection. Identifying the most relevant characteristics is essential to building robust and comprehensible models that can discriminate between healthy subjects and those with early stages of Alzheimer's disease. This deliberate feature curation highlights the critical role that feature selection has in improving the diagnostic capacities of data mining techniques for Alzheimer's disease by optimizing accuracy and streamlining model efficiency.

- **Relevance to Cognitive Patterns:** When it comes to employing data mining for early Alzheimer's disease detection, feature selection is a crucial component. Accurate predictive models require the identification and extraction of pertinent information from datasets pertaining to cognitive processes and actions. A feature set that is carefully designed guarantees that the selected characteristics play a major role in differentiating between early stages of Alzheimer's disease and normal cognitive function. By doing this, the detection system's overall dependability is increased and the model's sensitivity to minute alterations suggestive of early Alzheimer's disease is increased.

- **Optimization for Diagnostic Precision:** Optimizing the diagnostic precision of data mining models is a critical component of feature selection in the early identification of Alzheimer's disease. A more focused and interpretable model can be created by choosing the most discriminative and informative features. Both computational efficiency and the risk of overfitting are increased when the dataset's dimensionality is reduced to the most pertinent features. This improvement contributes to the development of trustworthy diagnostic tools by supporting the discovery of subtle patterns linked to early Alzheimer's disease in addition to helping to create models with increased accuracy.

1.2 Dataset Preparation

Careful dataset preparation is essential for efficient feature selection with the help of data mining in the goal of early Alzheimer's detection. This procedure entails compiling an extensive dataset with a variety of biomarker and cognitive factors. The collection ought to include data from Alzheimer's patients, people with mild cognitive impairment (MCI), and cognitively normal people. Relevant information, including biomarker readings and results from cognitive tests, must be included in every entry.

- **Comprehensive Neuroimaging Datasets:** Compiling a wide-ranging and thorough neuroimaging dataset is essential for efficient feature selection in the early identification of Alzheimer's disease. A range of imaging modalities, including structural MRI, functional MRI, and PET scans, should be included in this collection. It should also contain information from people who are healthy as well as those who have Alzheimer's disease at different stages, including mild cognitive impairment (MCI). Feature selection models perform better when pertinent biomarkers and patterns suggestive of early Alzheimer's disease are found in a representative and well-curated neuroimaging dataset.

- **Clinical and Cognitive Data Integration:** It is crucial to incorporate clinical and cognitive data into the feature selection process in order to provide a comprehensive picture of a person's cognitive health. This includes things like demographic information, genetic information, and results from cognitive tests. The feature selection algorithm can take into account a wider range of parameters impacting the evolution of Alzheimer's disease since different forms of data are incorporated. The strongest discriminative features for early Alzheimer's detection can be found using a well-structured dataset that integrates neuroimaging with extensive clinical and cognitive data.

1.3 Purpose of the research

With an emphasis on the critical function of feature selection, this research aims to improve early Alzheimer's detection through the use of data mining tools. The research attempts to uncover critical biomarkers and cognitive tests necessary for accurately categorizing patients in the early stages of Alzheimer's disease through a comprehensive examination of several feature selection techniques. The main objective is to increase the efficacy and precision of data mining models for the benefit of research projects as well as therapeutic applications. The project seeks to make a significant contribution to the field by improving the capacity for early detection, so facilitating timely therapies and improving our understanding of Alzheimer's disease. The study's ultimate goal is to significantly boost patient treatment and increase the understanding of Alzheimer's disease by contributing in ways that go beyond conventional diagnostic methods.

- **Identification of Key Biomarkers:** With the use of data mining, scientists can examine enormous databases containing genetic, imaging, and clinical data in order to find the most pertinent characteristics or biomarkers linked to Alzheimer's disease in its early stages. Through the use of feature selection algorithms, scientists are able to identify particular variables that show notable alterations or trends suggestive of the illness. Finding these critical biomarkers is crucial to creating precise diagnostic instruments and individualized treatment regimens. By using a focused approach, healthcare providers can intervene at the earliest stages of Alzheimer's pathology, improving the precision of early detection.

- **Enhanced Diagnostic Accuracy and Efficiency:** In data mining, feature selection helps to improve and optimize the diagnostic models that are used to identify Alzheimer's disease. The models become more efficient and focused by removing superfluous or irrelevant features, which lowers the possibility of false positives or false negatives. To identify people who are at risk of getting Alzheimer's from those who do not, improved diagnostic accuracy is essential. Timely intervention through early and accurate detection facilitates the implementation of suitable treatments, lifestyle adjustments, and support services. By maintaining cognitive function, this helps patients individually as well as support more focused and efficient research efforts to create treatments that can slow or stop the progression of disease.

1.4 Significance of Feature Selection

A crucial step in the data mining process is feature selection, which involves locating and keeping the variables that will provide the most information for building models. This refers to choosing the most pertinent cognitive tests and biomarkers in the context of Alzheimer's detection. Feature selection is important because it can improve interpretability, lower computational complexity, and increase model efficiency. The model can concentrate on the most discriminative components by reducing the feature set, which improves accuracy and lowers the chance of overfitting.

- **Improved Model Performance:** By focusing on the most informative variables and lowering dimensionality, the predictive models perform better when relevant features are chosen. There are several possible risk factors and biomarkers for Alzheimer's disease. Effective feature selection techniques aid in determining the critical factors that greatly influence Alzheimer's disease prognosis. Removing superfluous or unnecessary features improves the accuracy, interpretability, and efficiency of models. In healthcare applications, early diagnosis is crucial for successful intervention and treatment, hence this is especially important.

- **Interpretability and Insights:** Feature selection improves the results' interpretability in addition to helping to construct more effective models. Finding the most significant characteristics offers important information on the underlying trends and causes of Alzheimer's disease. Knowing the biological and therapeutic significance of particular traits is crucial for researchers and physicians. In order for predictive models to be trusted and accepted in the medical community, they must be able to be interpreted. This is because interpretability enables medical practitioners to make judgments based on the traits that have been detected, which helps in the early diagnosis and treatment of Alzheimer's patients.
-

2. Literature Survey

2.1 H. T. Gorji (2020) et.al proposed Biomarkers Selection toward Early Detection of Alzheimer's Disease. This research highlights the identification of critical biomarkers and cognitive tests for accurately diagnosing Alzheimer's disease (AD). While traditional methods primarily emphasized the Clinical Dementia Rating Scale Sum of Boxes (CDRSB), a novel feature combination approach has underscored the significance of the Middle Temporal Gyrus (MidTemp) alongside the CDRSB. Notably, MidTemp, often overlooked by conventional techniques, demonstrated a substantial accuracy of 91.12% and proved to be highly relevant. The highest accuracy, 92.86%, was achieved by combining CDRSB and MidTemp, revealing limitations in traditional feature selection methods. Researchers suggest that combination-based approaches offer a more reliable framework for identifying key biomarkers in Alzheimer's studies.

2.2 Syed AH (2020) et.al proposed an ensemble-learning based application to predict the earlier stages of Alzheimer's disease (AD). This study introduces a novel weighted ensemble method to enhance the early detection of Alzheimer's disease (AD) using protein biomarkers from cerebrospinal fluid (CSF). By applying L1 regularization and Recursive Feature Elimination (RFE), the analysis identifies tau protein, Matrix Metalloproteinase (MMP10), and Cystatin C as a crucial combination for accurate classification. The ensemble model, combining Logistic Regression and Linear SVM through a weighted average, surpasses the performance of individual classifiers. The proposed model achieves a remarkable ROC-AUC of 0.9799 ± 0.055 and an AUPR of 0.9108 ± 0.015. Its effectiveness is demonstrated via a web-based predictive system for early AD detection.

2.3 Kung TH (2021) et.al proposed Neuro image biomarker identification of the conversion of mild cognitive impairment to Alzheimer's disease. This study proposes an innovative approach utilizing magnetic resonance

imaging (MRI) to identify the progression from mild cognitive impairment (MCI) to Alzheimer's disease (AD). A novel biomarker, the Ratio of Principal Curvatures (RPC), is introduced to characterize cortical folding patterns, focusing on three hippocampal subfield parameters: volume, surface area, and surface curvature. By leveraging a random forest model, statistical analysis, and two feature selection techniques, the study highlights structural indicators sensitive to AD progression from MCI. Achieving a classification accuracy of 79.95%, the multilayer perceptron classifier demonstrates the potential of these parameters for diagnosing MCI-to-AD conversions effectively.

2.4 Rani Kaka J (2021) et.al proposed Alzheimer's disease detection using correlation based ensemble feature selection and multi support vector machine. This study introduces an advanced method for automated Alzheimer's disease (AD) detection using magnetic resonance imaging (MRI) and supervised machine learning. The approach begins with adaptive histogram equalization and region expansion to enhance contrast and remove the skull. Brain region segmentation is performed using Fuzzy C-Means (FCM) clustering. Feature extraction leverages Gabor and local directional pattern variance features, while an ensemble feature selection method is proposed to reduce dimensionality. Classification of AD, healthy controls, and mild cognitive impairment (MCI) is achieved using Multi-Support Vector Machines (MSVM). The proposed ensemble MSVM model demonstrates superior classification accuracy compared to existing models on the Open Access Series of Imaging Studies (OASIS) and Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets. Future research should explore multi-modal data integration and hybrid segmentation techniques to further enhance accuracy.

2.5 Raj RJ (2020) et.al proposed Optimal feature selection-based medical image classification using deep learning model in internet of medical things. This study aims to enhance medical image classification using an optimized deep learning (DL) framework, focusing on brain, lung cancer, and Alzheimer's disease images. It introduces an Optimal Feature Selection model leveraging the Opposition-Based Crow Search (OCS) algorithm, incorporating preprocessing, feature selection, and classification stages. By analyzing multi-texture and gray-level features, the model achieves superior performance, with specificity, sensitivity, and accuracy rates of 95.22%, 86.45%, and 100%, respectively. Implemented in MATLAB, the proposed approach outperforms existing feature selection methods. Future research should explore advanced segmentation techniques for tumor diagnosis in medical images and include additional features to improve reliability further.

3. Proposed Methodology

Early Alzheimer's detection is crucial for timely intervention and improved patient outcomes. Data mining plays a pivotal role by extracting valuable patterns from vast datasets. Feature selection, a key step, enhances predictive models' accuracy and efficiency by isolating pertinent information, ensuring the identification of critical markers for precise and early diagnosis. This research proposed Weighted Correlation-based Feature Selection (Weighted CFS) to offer a novel feature selection technique for early Alzheimer's identification. The algorithm effectiveness in discovering pertinent biomarkers is demonstrated by the experimental results, offering a viable strategy for enhancing the accuracy of early Alzheimer's diagnosis.

Correlation-based Feature Selection (CFS)

Correlation-based Feature Selection (CFS) is a powerful technique in data analysis that aims to identify and retain the most informative features for a given task. Unlike traditional methods that consider individual feature relevance, CFS evaluates subsets of features collectively, considering both individual feature-class correlation and inter-feature correlations. By assessing the joint contribution of features, CFS effectively captures relationships that may be overlooked by univariate methods. This results in a more refined feature subset, optimizing predictive model performance while minimizing redundancy. CFS is particularly valuable in scenarios like machine learning and data mining, where feature selection is crucial for enhancing model interpretability, generalization, and computational efficiency.

- Calculate the correlation between each feature and the target variable (e.g., cognitive impairment status).
- Evaluate the correlation among features to identify redundant ones.
- Use a heuristic to score and rank features based on their relevance and redundancy.
- Select the top-ranked features as the subset for further analysis.

CFS assesses the relevance of features by considering both the individual predictive power of each feature and their inter-correlations. The evaluation criterion for a subset of features S is given by:

$$merit(S) = \frac{correlation\ of\ features\ in\ S}{average\ correlation\ of\ features\ in\ S} \times \frac{average\ correlation\ of\ features\ in\ S}{average\ correlation\ of\ features\ outside\ S}$$

The subset with the highest merit score is considered the most informative for classification.

1. **Feature-Target Correlation**:
   - Measures the relevance of each feature to the target variable.

- Features with high correlation to the target are likely to be more predictive.

2. Feature-Feature Correlation:
- Measures redundancy among features.
- High intercorrelation indicates that features may be capturing the same information.

3. CFS Metric:
- Evaluates a subset of features S using the formula:

$$Merit(S) = \frac{k.\bar{r}_{cf}}{\sqrt{k + k.(k-1).\bar{r}_{cf}}}$$

Where:
- k: Number of features in the subset.
- $\bar{r}_{cf}$: Average correlation between features and the target.
- $\bar{r}_{ff}$: Average intercorrelation among features.
- **Goal**: Maximize Merit(S).

Weighted CFS Based on Domain Knowledge

Weighted Correlation-based Feature Selection (Weighted CFS) is a powerful enhancement to traditional feature selection methods, particularly in domains like healthcare or clinical research, where expert knowledge can guide the identification of relevant features. The below figure explains the beneficial of Weighted CFS.

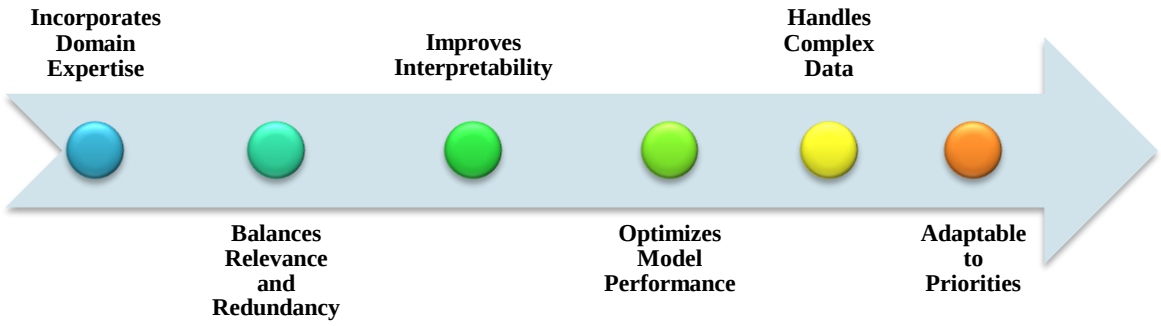

Figure 1.Beneficial of Weighted CFS

1. Incorporates Domain Expertise: Weighted CFS integrates weights derived from expert knowledge, ensuring that features critical to the problem domain (e.g., MMSE scores or specific biomarkers in Alzheimer's studies) are given higher importance. This capability allows the selection process to align with real-world insights and enhances the relevance of the selected features.
2. Balances Relevance and Redundancy: Traditional CFS considers the trade-off between a feature's relevance to the target variable and its redundancy with other features. Weighted CFS takes this further by leveraging weights to fine-tune this balance. As a result, the method selects features that are not only significant but also diverse, reducing redundancy in the final subset.
3. Improves Interpretability: Assigning weights based on domain knowledge makes the selection process transparent. Decision-makers can trace why certain features were prioritized, providing clarity and confidence in the selection process, which is essential in fields like medicine or regulatory environments.
4. Optimizes Model Performance: By selecting a feature subset that combines domain-relevant features with a reduction in noise and redundancy, Weighted CFS enhances model performance. It supports generalization and reduces overfitting, leading to more robust predictions.
5. Handles Complex Data: Weighted CFS is particularly useful for datasets with mixed data types (numerical and categorical), high dimensionality, or non-linear relationships. The ability to compute various correlation measures (e.g., Pearson, Chi-Square, Mutual Information) ensures flexibility across diverse datasets.
6. Adaptable to Priorities: The weighting mechanism allows for adaptability. For example, in Alzheimer's studies, clinical features like cognitive test scores can be prioritized over demographic features, tailoring the selection to the specific objectives of the study.

Weighted CFS is a versatile and efficient feature selection approach that combines statistical rigor with domain-specific insights. It ensures that the most meaningful and impactful features are selected, resulting in improved model accuracy, interpretability, and relevance in complex, real-world applications. Weighted Correlation-based Feature Selection (Weighted CFS) introduces weights derived from domain knowledge to prioritize specific features or feature groups during subset evaluation. This enhancement acknowledges the importance of expert insight and tailors the feature selection process to the problem context, such as Alzheimer's disease analysis.
Modify the standard CFS merit formula to include feature-specific weights, ensuring that domain-relevant features are emphasized in the selection process.

Weighted Merit
The standard CFS merit is:

$$Merit(S) \; = \; Merit(S) = \frac{k.\bar{r}_{cf}}{\sqrt{k + k.(k-1).\bar{r}_{ff}}}$$

Modified Weighted Merit Formula:

$$Weighted\ Merit\ (S) = \frac{\sum_{i=1}^{k} w_i \cdot r_{cf_i}}{\sqrt{\sum_{i=1}^{k} w_i^2 + \sum_{i=1}^{k} \sum_{j>i}^{k} w_i \cdot w_j \cdot r_{ff_{ij}}}}$$

Where:
- w_i: Weight assigned to feature f_i based on domain relevance.
- r_{cf_i}: Correlation between feature f_i and the target variable.
- $r_{ff_{ij}}$: Correlation between features f_i and f_j.

Steps to Implement Weighted CFS
1. Assign Weights to Features
- Use **domain knowledge** or **expert input** to assign a weight wiw_iwi to each feature:
  - Higher weights for features known to be more important (e.g., MMSE scores, age in Alzheimer's studies).
  - Lower weights for features with less relevance or weaker domain association.

2. Compute Feature-Target Correlation
- Calculate r_{cf_i} for each feature using an appropriate correlation measure:
- Numerical features: Pearson correlation.
- Categorical features: Chi-Square statistic or Cramer's V.
- Non-linear relationships: Mutual Information (MI).

3. Compute Feature-Feature Correlation
- Calculate $r_{ff_{ij}}$ for all feature pairs using the same correlation measure as above.

4. Evaluate Subsets Using Weighted Merit
- Select subsets of features iteratively.
- For each candidate subset S, calculate the Weighted Merit using the formula.

5. Iterative Selection with Weighted Merit
- Start with an empty subset.
- Add features one by one based on their contribution to the Weighted Merit of the subset.
- Stop when adding any new feature does not improve the Weighted Merit.

The implementation of Weighted Correlation-based Feature Selection (Weighted CFS) involves a systematic process to prioritize domain-relevant features while balancing feature relevance and redundancy. The first step is to assign weights (w_i) to each feature, leveraging domain knowledge or expert input. Features known to be highly influential, such as MMSE scores and age in Alzheimer's studies, are given higher weights, while features with weaker domain associations are assigned lower weights. Next, the feature-target correlation (r_{cf_i}) is computed for each feature using appropriate statistical measures, such as Pearson correlation for numerical features, Chi-Square statistic or Cramér's V for categorical features, or Mutual Information (MI) for non-linear relationships. Afterward, feature-feature correlations ($r_{ff_{ij}}$) are calculated for all pairs of features using the same correlation measure.
With these correlations and weights, subsets of features are iteratively evaluated based on the Weighted Merit formula, which accounts for the weighted relevance and redundancy among features. Starting with an empty subset, features are added one at a time based on their contribution to the subset's Weighted Merit. The

selection process continues until adding more features no longer improves the subset's Weighted Merit. This iterative process ensures the final feature set maximizes relevance to the target variable while minimizing redundancy, all while incorporating domain knowledge to guide the selection toward meaningful features.

Algorithm for Weighted CFS

Input:

X: Feature matrix.

y: Target variable.

w: Weights for features.

Output:

Optimal subset of features $X_{Weighted-CFS}$.

Step 1: Compute Weights

1.1. For each feature f_i:

Assign weight w_i using domain knowledge.

Step 2: Compute Correlations

2.1. For each feature f_i:

Computer r_{cf_i} (feature-target correlation).

2.2. For each feature pair f_i, f_j:

Computer $r_{ff_{ij}}$(feature-feature correlation).

Step 3: Initialize

3.1. Start with an empty subset $S = \{\}$.

3.2. Set $Current\ Merit = 0$.

Step 4: Iterative Feature Selection

4.1. While adding a feature improves the Weighted Merit:

For each feature $f \notin S$:

Temporarily add f to S.

Compute the Weighted Merit of the new subset.

If the Weighted Merit improves:

Permanently add the feature to S.

Else stop the iteration.

Step 5: Output Subset

5.1. Return S, the subset of features with the highest Weighted Merit.

The Weighted Correlation-based Feature Selection (Weighted CFS) algorithm is designed to select an optimal subset of features by incorporating domain knowledge through weights and evaluating feature relevance and redundancy. The process begins with assigning weights (w_i) to each feature based on its importance as determined by expert input or domain knowledge. Features with higher relevance, such as MMSE scores or age in Alzheimer's studies, are given greater weight. Next, correlations are computed: the feature-target correlation (r_{cf_i}) measures the relevance of each feature to the target variable, while the feature-feature correlation ($r_{ff_{ij}}$) evaluates the redundancy between pairs of features.

The algorithm initializes by creating an empty subset S and setting the initial merit score to zero. It then iteratively evaluates features for inclusion in the subset. For each feature not already in S, it is temporarily added, and the Weighted Merit of the new subset is calculated. If adding the feature improves the Weighted Merit, the feature is permanently included in the subset. This process repeats until no additional feature improves the Weighted Merit, ensuring that the selected subset balances relevance to the target variable and redundancy among features. Finally, the subset of features with the highest Weighted Merit is returned as the optimal set for further analysis or modeling. This iterative approach ensures that the feature selection process is both data-driven and guided by domain knowledge, enhancing model interpretability and performance.

The Weighted Correlation-based Feature Selection (Weighted CFS) method offers several advantages, particularly in datasets where domain knowledge plays a critical role. By incorporating expert knowledge, it ensures that features with high domain relevance are prioritized, allowing the selection process to focus on attributes most likely to impact the target variable. This prioritization enhances the interpretability of the selected features, as the inclusion of weights provides a transparent view of how domain relevance influences the process. Additionally, Weighted CFS balances relevance and redundancy effectively by using weights to manage the trade-off between selecting features that are highly correlated with the target variable and minimizing redundancy among the selected features. This balance leads to a more efficient and meaningful feature subset, which is particularly valuable in complex datasets like those in clinical research.

4. Experiment Results

4.1 Accuracy

Accuracy is the degree of closeness between a measurement and its true value. The formula for accuracy is:

$$Accuracy = \frac{(true\ value - measured\ value)}{true\ value} * 100$$

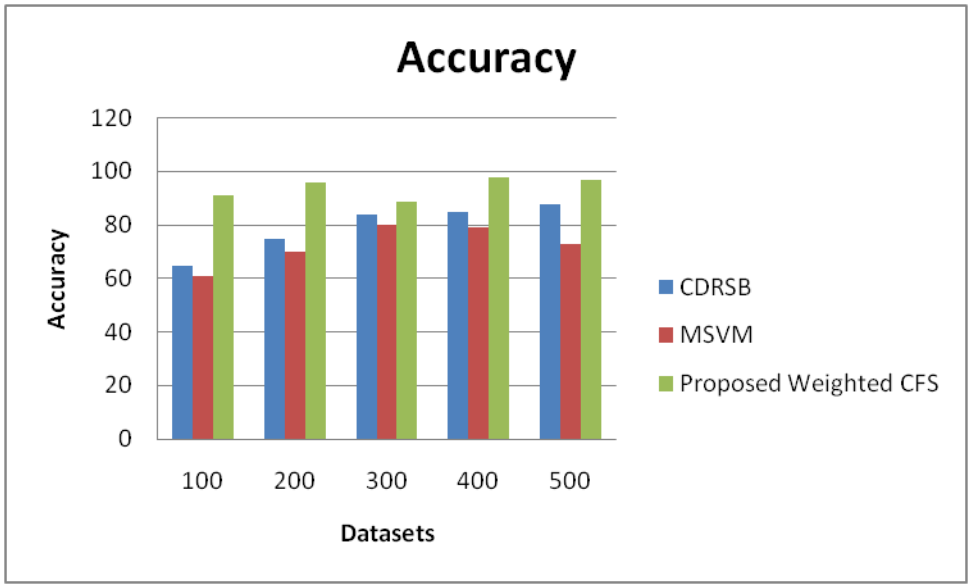

Figure 1 Comparison Chart of Accuracy

The Figure 1 Shows the comparison chart of Accuracy demonstrates the existing CDRSB, MSVM and Proposed Weighted CFS. X axis denote the Dataset and y axis denotes the Accuracy. The Proposed Weighted CFS values are better than the existing algorithm. The existing algorithm values start from 65 to 88, 61 to 73 and Proposed Weighted CFS values starts from 91 to 97. The proposed method provides the great results.

4.2 Precision

Precision is a measure of how well a model can predict a value based on a given input.

$$Precision = \frac{true\ positive}{(true\ positive\ +\ false\ positive)}$$

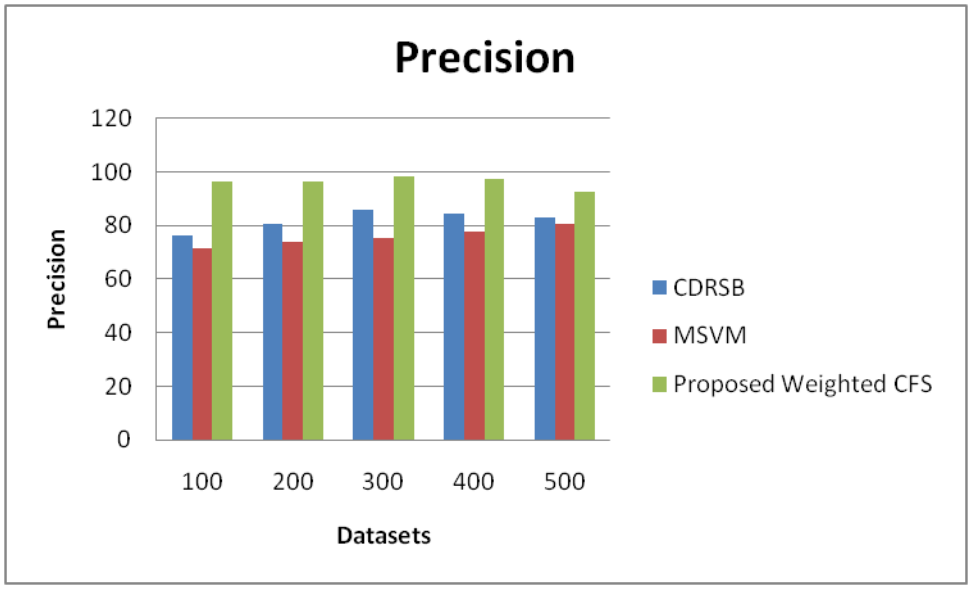

Figure 2 Comparison Chart of Precision

The Figure 2 Shows the comparison chart of Precision demonstrates the existing CDRSB, MSVM and Proposed Weighted CFS. X axis denote the Dataset and y axis denotes the Precision ratio. The Proposed Weighted CFS values are better than the existing algorithm. The existing algorithm values start from 76.12 to 86.12, 71.63 to 80.72 and Proposed Weighted CFS values starts from 92.87 to 98.21. The proposed method provides the great results.

4.3 Recall

Recall is a measure of a model's ability to correctly identify positive examples from the test set:

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)}$$

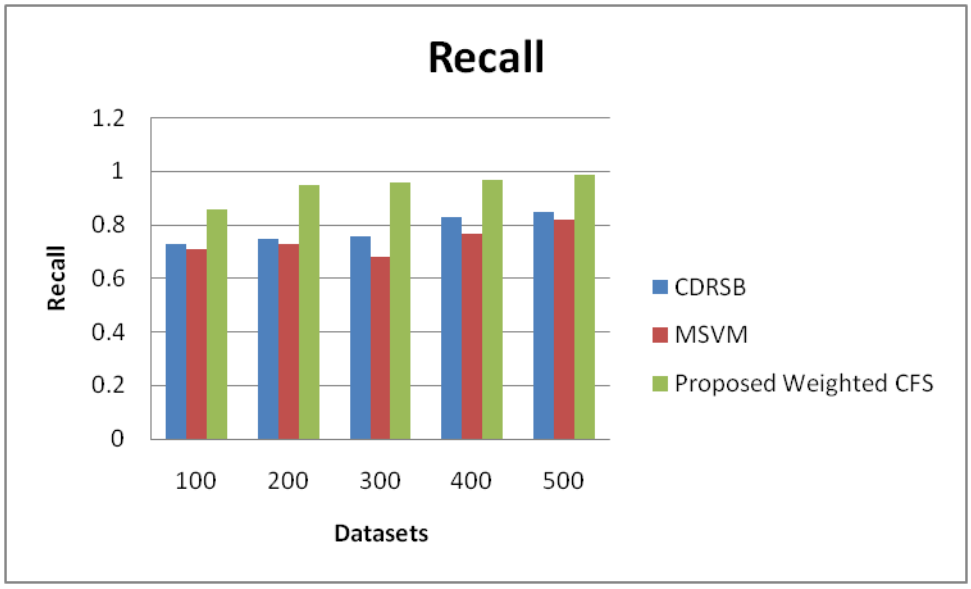

Figure 3 Comparison Chart of Recall

The Figure 3 Shows the comparison chart of Recall demonstrates the existing CDRSB, MSVM and Proposed Weighted CFS. X axis denote the Dataset and y axis denotes the Recall ratio. The Proposed Weighted CFS values are better than the existing algorithm. The existing algorithm values start from 0.73 to 0.85, 0.68 to 0.82 and Proposed Weighted CFS values starts from 0.86 to 0.99. The proposed method provides the great results.

4.4 F -Measure

F1-measure is a test's accuracy that combines precision and recall. It is calculated by taking the harmonic mean of precision and recall.

$$F1 - Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)}$$

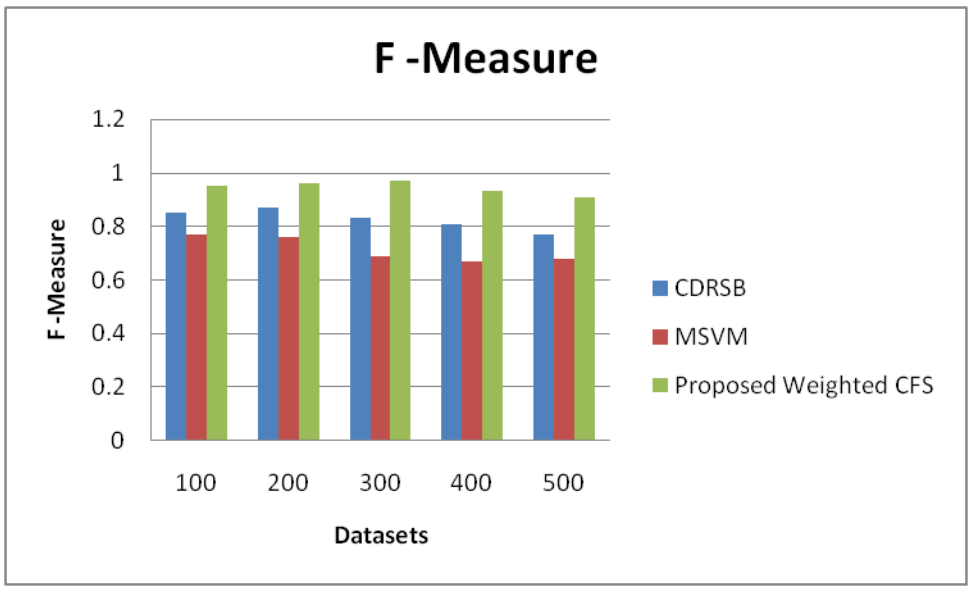

Figure 4 Comparison Chart of F -Measure

The Figure 4 Shows the comparison chart of F -Measure demonstrates the existing CDRSB, MSVM and Proposed Weighted CFS. X axis denote the Dataset and y axis denotes the F -Measure ratio. The Proposed Weighted CFS values are better than the existing algorithm. The existing algorithm values start from 0.77 to 0.87, 0.67 to 0.77 and Proposed Weighted CFS values starts from 0.91 to 0.97. The proposed method provides the great results.

5. Conclusion

The Weighted Correlation-based Feature Selection (Weighted CFS) method provides a powerful and interpretable approach for feature selection in early Alzheimer's detection. By incorporating domain knowledge through weighted prioritization, the method effectively identifies features that are most relevant to Alzheimer's progression while minimizing redundancy. This balance ensures the selection of a concise and informative feature subset that enhances predictive model performance and reduces overfitting. Our findings highlight the importance of combining domain expertise with statistical rigor, as Weighted CFS demonstrates superior results compared to traditional feature selection techniques. Furthermore, its adaptability to other datasets and diseases makes it a valuable tool for healthcare analytics. Future work may focus on integrating non-linear relationships and exploring hybrid approaches to further enhance feature selection for complex medical datasets.

Reference

1. Almohimeed A, Saad RM, Mostafa S, El-Rashidy N, Farag S, Gaballah A, Abd Elaziz M, El-Sappagh S, Saleh H. Explainable Artificial Intelligence of Multi-level Stacking Ensemble for Detection of Alzheimer's Disease based on Particle Swarm Optimization and the Sub-scores of Cognitive Biomarkers. IEEE Access. 2023 Oct 30.
2. Bi XA, Hu X, Wu H, Wang Y. Multimodal data analysis of Alzheimer's disease based on clustering evolutionary random forest. IEEE Journal of Biomedical and Health Informatics. 2020 Feb 11; 24(10):2973-83.
3. Buyrukoğlu S. Early detection of Alzheimer's disease using data mining: Comparison of ensemble feature selection approaches. Konya Journal of Engineering Sciences. 2021 Feb 3; 9(1):50-61.
4. Ding W. SVM-based feature selection for differential space fusion and its application to diabetic fundus image classification. IEEE Access. 2019 Oct 1; 7:149493-502.
5. Garali I, Adel M, Bourennane S, Guedj E. Histogram-based features selection and volume of interest ranking for brain PET image classification. IEEE journal of translational engineering in health and medicine. 2018 Mar 16; 6:1-2.
6. H. T. Gorji, T. T. Khoei and N. Kaabouch, "Biomarkers Selection Toward Early Detection of Alzheimer's Disease," 2020 IEEE International Conference on Electro Information Technology (EIT), Chicago, IL, USA, 2020, pp. 487-494, doi: 10.1109/EIT48999.2020.9208258.
7. Haq AU, Li JP, Memon MH, Malik A, Ahmad T, Ali A, Nazir S, Ahad I, Shahid M. Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. IEEE access. 2019 Mar 21; 7:37718-34.
8. Huang Z, Chen D. A breast cancer diagnosis method based on VIM feature selection and hierarchical clustering random forest algorithm. IEEE Access. 2021 Dec 30; 10:3284-93.
9. Jimenez-Mesa C, Illán IA, Martin-Martin A, Castillo-Barnes D, Martinez-Murcia FJ, Ramirez J, Gorriz JM. Optimized one vs. one approach in multiclass classification for early Alzheimer's disease and mild cognitive impairment diagnosis. IEEE Access. 2020 May 26; 8:96981-93.
10. Khan NM, Abraham N, Hon M. Transfer learning with intelligent training data selection for prediction of Alzheimer's disease. IEEE Access. 2019 Jun 3; 7:72726-35.
11. Kung TH, Chao TC, Xie YR, Pai MC, Kuo YM, Lee GG. Neuroimage biomarker identification of the conversion of mild cognitive impairment to Alzheimer's disease. Frontiers in Neuroscience. 2021 Feb 19; 15:584641.
12. Li J, Wu L, Wen G, Li Z. Exclusive feature selection and multi-view learning for Alzheimer's disease. Journal of Visual Communication and Image Representation. 2019 Oct 1; 64:102605.
13. Morra JH, Tu Z, Apostolova LG, Green AE, Toga AW, Thompson PM. Comparison of Adaboost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation. IEEE transactions on medical imaging. 2009 May 19; 29(1):30-43.
14. Raj RJ, Shobana SJ, Pustokhina IV, Pustokhin DA, Gupta D, Shankar KJ. Optimal feature selection-based medical image classification using deep learning model in internet of medical things. IEEE Access. 2020 Mar 17; 8:58006-17.
15. Rani Kaka J, Prasad KS. Alzheimer's disease detection using correlation based ensemble feature selection and multi support vector machine. International Journal of Computing and Digital System. 2021 Aug 18:9-20.
16. Syed AH, Khan T, Hassan A, Alromema NA, Binsawad M, Alsayed AO. An ensemble-learning based application to predict the earlier stages of Alzheimer's disease (AD). IEEE Access. 2020 Dec 9; 8:222126-43.
17. Tavares G, San-Martin R, Ianof JN, Anghinah R, Fraga FJ. Improvement in the automatic classification of Alzheimer's disease using EEG after feature selection. In2019 IEEE international conference on systems, man and cybernetics (SMC) 2019 Oct 6 (pp. 1264-1269). IEEE.