*Nanotechnology Perceptions*
*ISSN 1660-6795*
*www.nano-ntp.com*

# Advanced Hybrid Feature Selection Using Harvest Algorithm, Convolutional Neural Networks and CFS

**Preethi K[1], Dr.Ramakrishnan M[2]**

*[1]Research Scholar, Department of Computer Applications, School of Information Technology, Madurai Kamaraj University, Madurai, Tamilnadu, India, mail2preethi06@gmail.com*
*[2]Professor & Head, Department of Computer Applications, School of Information Technology, Madurai Kamaraj University, Madurai, Tamilnadu, India, ramkrishod@mkuniversity.org*

Predictive modeling and data analysis are severely hampered by the enormous dimensionality and complexity of medical datasets. Enhancing model performance, interpretability, and computing efficiency all depend on careful feature selection. In order to extract the most pertinent characteristics from the Medical Information Mart for Intensive Care III (MIMIC-III) dataset, this paper offers a novel hybrid feature selection method that combines the Harvest Algorithm, Convolutional Neural Networks (CNN), and Correlation-Based Feature Selection (CFS). De-identified medical records, including diagnostic codes, vital signs, prescriptions, and other clinical observations, are all included in the MIMIC-III dataset. Our suggested approach makes use of CNN for deep feature extraction, the Harvest Algorithm for the first feature subset creation, and CFS for the final feature selection based on correlation measures. Test findings show that when compared to conventional feature selection techniques Random Forest with Information Gain Method (RF-IG) and SVM with Recursive Feature Elimination (SVM-RFE), our hybrid strategy greatly increases the accuracy and efficiency of prediction models. The chosen characteristics demonstrate the potential of our approach in clinical decision support and medical data analysis by offering significant insights into important variables influencing patient outcomes.

**Keywords:** Feature Selection, Hybrid Methods, Harvest Algorithm, Convolutional Neural Networks (CNN), Correlation-Based Feature Selection (CFS), MIMIC-III Dataset, Medical Data Analysis.

## 1. Introduction

The rapid expansion of data produced by diagnostic procedures, treatment plans, and medical records in the field of healthcare brings with it both potential and difficulties. High-dimensional datasets with useful insights that might enhance patient outcomes and guide clinical decision-making include the Medical Information Mart for Intensive Care III (MIMIC-III). To improve the performance of predictive models and guarantee interpretable outputs, however, the sheer volume and complexity of such data require effective and fast feature selection techniques.

A crucial stage in the machine learning process is feature selection, particularly in the analysis of medical data, where the inclusion of redundant or unnecessary features can cause overfitting, raise computing costs, and impair model performance. Traditional feature selection methods, while useful, often struggle with the high dimensionality and intricate correlations inherent in medical datasets. Consequently, there is a growing need for advanced techniques that can handle these complexities and extract meaningful features that contribute to accurate and robust predictions [1][2]. In medical data analysis, feature selection is crucial for several reasons:

✓ Improved Model Performance: Models can be made more accurate, precise, and durable by choosing the most essential elements. Overfitting can occur as a result of redundant or irrelevant features, which impair model performance.

✓ Reduced Computational Complexity: Training models on high-dimensional datasets necessitates substantial computational resources. By reducing the dimensionality of the input, feature selection speeds up and improves the effectiveness of the learning process.

✓ Enhanced Interpretability: Model interpretability is critical in medical research. In order to act on model outputs, clinicians and researchers must comprehend which aspects are impacting the forecasts they can trust. The process of feature selection aids in determining the important factors that influence the judgments made by the model.

✓ Data Storage and Processing: Processing and storage of huge datasets with plenty of features might be difficult. By selecting features, the amount of data is decreased and becomes more manageable.

✓ Noise Reduction: Oftentimes, noisy data in medical datasets can mask important patterns. Feature selection aids in noise reduction, producing insights that are more trustworthy and clear.

A sizable collection of de-identified medical records gathered from critical care units is made available to the public through the Medical Information Mart for critical Care III (MIMIC-III) dataset [3][4]. It contains specific details including lab results, prescriptions, vital signs, and diagnostic codes. There are over 112,000 clinical reports in the collection, with an average of 709.3 tokens and 1,159 top-level ICD-9 codes per report. Each report is given an average of 7.6 ICD-9 codes, demonstrating the dataset's high dimensionality and complexity. The MIMIC-III dataset has a wide range of clinical characteristics and patient outcomes, which makes it a great resource for the development and testing of advanced feature selection techniques.

In order to tackle the difficulties associated with feature selection in high-dimensional medical datasets, this research presents a novel hybrid approach that combines the advantages of Correlation-Based Feature Selection (CFS) [9] [10], Convolutional Neural Networks (CNN) [7] [8], and the Harvest Algorithm [5] [6]. To extract the most pertinent features from the MIMIC-III dataset, the suggested approach makes use of the correlation analysis of CFS, the deep learning-based feature extraction capacity of CNNs, and the optimization skills of the Harvest Algorithm.

The remainder of this paper is organized as follows ,In Section 2, Literature Review provides an overview of existing feature selection techniques, hybrid methods, and their applications in medical data analysis, along with a detailed discussion of the Harvest Algorithm, CNNs, and CFS. In Section 3: Discussed two kind of hybrid feature selection methods are Random Forest with Information Gain Method (RF-IG) and SVM with Recursive Feature Elimination (SVM-RFE). In Section 4: Methodology describes the MIMIC-III dataset and the preprocessing steps, followed by a detailed explanation of the proposed hybrid feature selection method and its components. In Section 5: Experimental Setup outlines the experimental design, performance metrics, and implementation details used to evaluate the proposed method. Finally, In Section 6: Conclusion and future research directions.

## 2. Literature Review

Guyon, I. et al.'s [11] selection is a crucial machine learning preprocessing stage that seeks to pinpoint the most pertinent characteristics for model construction. Feature selection techniques fall into three primary categories: wrapper, embedding, and filter techniques. Filter Methods: These techniques assess a feature's importance by using statistical metrics like mutual information, correlation, or Chi-square scores. Despite their computational efficiency, they might not take into account feature interactions. Correlation coefficient, Chi-square, and information gain are a few examples. Wrapper Methods: These techniques assess feature subsets using a predictive model, and then choose the best set based on the model's output. Although they require more computing power, they are more accurate than filter techniques. Sequential feature selection and Recursive Feature Elimination (RFE) are two examples.

Embedded Techniques: These techniques carry out feature selection while the model is being trained. Lasso (L1 regularization) and tree-based techniques (like Random Forest and Gradient Boosting) are two examples. In order to capitalize on their advantages and minimize their disadvantages, hybrid approaches integrate filter, wrapper, and embedded techniques. Usually, they consist of many steps: first, a filter method decreases the feature space; then, a wrapper or embedding approach is used for fine-tuning. Because hybrid approaches strike a balance between accuracy and computing economy, they are especially helpful for high-dimensional data. For example, Sun et al. (2019) [12] improved feature selection for high-dimensional biological data by combining filter techniques and genetic algorithms.

A disease's diagnosis and prognosis can be aided by the identification of important biomarkers and clinical characteristics through feature selection. Liu et al. (2018) [13], for instance, employed feature selection strategies to find important features for diabetes mellitus prediction. Models can be customized to match the unique patient profile by choosing pertinent features, which will result in more individualized therapy suggestions.
In order to discover the best answers, Fathian, M. et al. (2007) [14] employed the Harvest Algorithm, which replicates the process of growing and harvesting crops. It entails cycles of growth (improving solutions) and harvest (choosing the best solutions). An optimization method called the Harvest Algorithm was inspired by the harvesting of crops. Because of its

capacity to effectively explore the solution space, it has been used to a wide range of optimization issues. The algorithm has been used in feature selection, scheduling, and other combinatorial optimization problems.

LeCun et al., 2015 [15], CNNs are a type of deep learning models that are very good at extracting features from structured data, including photos and time-series data. CNNs are effective for complicated pattern identification because their convolutional layers automatically learn the spatial hierarchies of features from the input data. High-level features can be extracted from raw data using CNNs, and these features can then be included into conventional feature selection techniques. CNNs have been used, for instance, to extract features from medical imaging data in order to classify diseases. According to Hall, M. A. et al. (1999) [16], CFS determines the value of a subset of features by taking into account each feature's unique predictive capacity as well as the degree of redundancy among them. Feature subsets having minimal inter-correlation among themselves but strong correlation with the target variable are chosen by CFS. This guarantees that every feature chosen enriches the model with distinct data. CFS is extensively utilized in several fields, such as medical data analysis, where it aids in the discovery of useful and non-redundant features.

## 3. Feature Selection Methods

### 3.1. Information Gain Method

The Information Gain (IG) approach is a filter-based feature selection strategy that measures the decrease in entropy (uncertainty) of the target variable when a feature is utilized in order to assess the relevance of features. Given the large dimensionality and diversified nature of the data in the MIMIC-III dataset—which includes vital signs, drugs, laboratory measures, and diagnostic codes—IG can be very useful in this setting [17]. To apply information gain on the MIMIC-III dataset, follow these steps:

Step 1: Preprocessing the Data

- Handling Missing Data: Impute missing values using techniques like mean/mode imputation, or advanced methods like multiple imputations. Calculate mean imputation as,

$$x_i = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Normalization: Normalize continuous variables to ensure they are on a comparable scale. Calculate as Min-Max Normalization,

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- Encoding Categorical Variables: Convert categorical variables into numerical format using techniques like one-hot encoding.

- Step 2: Calculation of Entropy

- Entropy of the Target Variable (H(Y)): Calculate the entropy of the target variable (e.g., mortality) to measure its uncertainty.

$$H(Y) = -\sum_{i=1}^{n} P(y_i)\log_2 P(y_i)$$

- Step 3: Calculation of Conditional Entropy

\- Entropy of the Target Variable Given a Feature (H(Y|X)) : Calculate the conditional entropy of the target variable given each feature.

$$H(Y|X) = -\sum_{i=1}^{n} P(x_i)\sum_{i=1}^{n} P(x_i|y_i)\log_2 P(y_i|x_i)$$

- Step 4: Information Gain Calculation

\- Information Gain (IG) : Calculate the IG for each feature by subtracting the conditional entropy from the entropy of the target variable.

$IG(X) = H(Y) - H(Y|X)$

- Step 5: Feature Ranking and Selection

\- Rank Features: Rank features based on their IG scores, with higher scores indicating greater relevance.

\- Select Top Features: Choose the top-ranked features for further modeling.

Information Gain (IG) works on MIMIC-III,

▪ The MIMIC-III dataset include patient information such as lab results, prescriptions, vital signs, and diagnostic codes. Numerous characteristics in every record can be utilized to forecast various outcomes, such as length of stay or patient mortality.

▪ Establish the prediction goal variable, such as duration of stay (continuous outcome) or death (binary outcome).

▪ Utilize data such as patient demographics, vital signs (blood pressure, heart rate), laboratory test results, prescribed drugs, and ICD-9 diagnosis codes.

▪ Determine the IG for every feature in the dataset in relation to the target variable. To estimate patient mortality, compute the IG for each ICD-9 code, blood pressure, and heart rate, for instance. Sort the features according to their IG scores, then choose the top N features to use in the construction of prediction models.

Example Workflow,

✓ Entropy Calculation: Calculate the entropy of the target variable (e.g., patient mortality).

\- If mortality is binary (0 or 1), and P(mortality=1) = 0.2, P(mortality=0) = 0.8:

$H(Y) =-(0.2\log_2 0.2+0.8\log_2 0.8)$

✓ Conditional Entropy Calculation : Calculate the conditional entropy for a feature, such as heart rate:

- For each heart rate value ($x_j$), calculate H ($Y|X = x_j$), then average over all possible values of heart rate.

✓ Information Gain :

- Calculate IG for heart rate:
- IG(Heart Rate) = H(Y) - H(Y | Heart Rate)

Random Forest with Information Gain Method (RF-IG)

Combining Random Forest with the Information Gain method for feature selection can leverage the strengths of both techniques to enhance predictive modeling, especially with complex and high-dimensional datasets like MIMIC-III. Here’s how these methods can work together:

Information Gain (IG) is a measure of how much information a feature contributes to the prediction of the target variable. It quantifies the reduction in entropy (uncertainty) of the target variable given a feature. Features with higher IG are considered more informative [18][19].

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees [20][21].

- Steps in Random Forest:
  - Bootstrap Sampling: Randomly select samples with replacement from the dataset.
  - Feature Selection: Randomly select a subset of features at each split in the decision tree.
  - Tree Building: Build a decision tree using the selected samples and features.
  - Aggregation: Aggregate the predictions from all trees to produce the final prediction.

3. Combining Information Gain with Random Forest

1. Data Preprocessing:
   - Handle missing data, normalize continuous variables, and encode categorical variables as described previously.
2. Initial Feature Selection Using Information Gain:
   - Calculate IG for each feature with respect to the target variable (e.g., mortality).
   - Rank features based on their IG scores and select the top features. This step reduces the dimensionality of the dataset and retains the most informative features.
3. Random Forest Implementation:
   - Use the selected features from the IG method as input to the Random Forest algorithm.
   - Train the Random Forest model using the selected features.
   - Evaluate the performance of the Random Forest model using appropriate metrics such as accuracy, precision, recall, and AUC-ROC.

Advantages of Combining Information Gain with Random Forest are Reduces the number of features, mitigating the curse of dimensionality and improving computational efficiency. Focuses on the most informative features, potentially improving the predictive accuracy of the Random Forest model. Enhances interpretability by highlighting the most important features influencing the target variable.

### 3.2. Recursive Feature Elimination (RFE)

Recursive feature elimination, or RFE, is a feature selection technique that works by iteratively analyzing progressively smaller subsets of features in order to choose features. In order to get the required number of features, it fits a model and eliminates the least significant feature [22]. RFE improves the interpretability and efficacy of predictive models by helping to find the most pertinent features for clinical outcome prediction when applied to the MIMIC-III dataset. How to Use RFE,

1. Data Preprocessing:
   - Handle Missing Data: Impute missing values using methods such as mean/mode imputation or more advanced techniques like multiple imputations.
   - Normalization: Normalize continuous variables to ensure they are on a comparable scale.
   - Encoding Categorical Variables: Convert categorical variables into numerical format using techniques like one-hot encoding.
2. Initial Model Training:
   - Select a machine learning model to use with RFE. Common choices include Support Vector Machines (SVM), Logistic Regression, and Random Forest.
   - Train the initial model using all available features.
3. Recursive Elimination Process:
   - Feature Ranking: Rank all the features based on their importance scores from the model.
   - Feature Elimination: Remove the least important feature(s) and re-train the model.
   - Repeat: Continue this process iteratively, removing features and retraining the model, until the desired number of features is achieved.

### SVM with Recursive Feature Elimination (SVM-RFE)

Combining Support Vector Machine (SVM) with Recursive Feature Elimination (RFE) is a powerful approach for feature selection, especially for high-dimensional datasets like MIMIC-III. This combination leverages the predictive power of SVM and the systematic feature reduction process of RFE to enhance model performance and interpretability.

Recursive Feature Elimination (RFE) is a feature selection technique that recursively removes the least important features and builds the model on the remaining features. It ranks the features based on their importance to the model's performance [23] [24].

Support Vector Machine (SVM) is a supervised learning model used for classification and regression tasks. It works by finding the hyperplane that best separates the classes in the feature space [25] [26].

Detailed Workflow for MIMIC-III Dataset

1. Dataset Preparation:

o Target Variable: Define the target variable (e.g., patient mortality or length of stay).

o Feature Set: Include features such as patient demographics, vital signs, laboratory test results, medications administered, and ICD-9 diagnostic codes.

2. Example: Applying SVM with RFE:

o Step 1: Normalize the continuous variables in the dataset.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

o Step 2: Encode categorical variables using one-hot encoding.

o Step 3: Initialize the SVM model (e.g., linear SVM).

o Step 4: Use RFE with the SVM model to fit the data and rank features.

o Step 5: Remove the least important feature and re-train the SVM model.

o Step 6: Repeat the process until the desired number of features is selected.

3. Model Performance Evaluation:

o Evaluate the model using the selected features on the validation set.

o Use metrics such as accuracy, precision, recall, F1 score, and AUC-ROC to compare performance.

Advantages of Using SVM with RFE on MIMIC-III Dataset are RFE effectively reduces the number of features, which helps in managing the high-dimensional nature of the MIMIC-III dataset. By selecting the most relevant features, RFE improves the SVM model’s predictive accuracy. Focusing on a smaller set of important features makes the model more interpretable and actionable for clinical decision-making.

## 4. Methodology

### 4.1. MIMIC-III Dataset

The Medical Information Mart for Intensive Care III (MIMIC-III) is a comprehensive, publicly available dataset containing de-identified health-related data associated with over 40,000 patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. This dataset is widely used in clinical research due to its richness and depth of information [27]. Key Components of the MIMIC-III Dataset are,

- Patient Demographics: Ranges from neonates to elderly patients, providing a diverse population sample , Includes male and female patients and Detailed records of patient ethnicity to support diverse population studies.

- Clinical Measurements: Regularly recorded vital signs such as heart rate, blood pressure, respiratory rate, and temperature. Comprehensive lab test results including blood tests, electrolyte levels, blood gas measurements, and more. Information on medications administered, dosages, and timing.

- Observations and Notes: Detailed notes and observations made by healthcare providers during patient care. Includes assessments, interventions, and patient responses recorded by nursing staff. Summarized reports of patient condition, treatments administered, and instructions at the time of discharge.

- Procedures and Diagnoses : ICD-9 Codes - Detailed diagnostic and procedural codes (International Classification of Diseases, Ninth Revision) including top-level codes and sub-codes for specific details. Detailed records of procedures performed during ICU stays, such as surgeries, intubations, and catheterizations.

- Temporal Data: Each clinical measurement, medication administration, and procedure is time-stamped, allowing for detailed temporal analysis. Duration of patient stays in the ICU and hospital.

- Outcomes: Total duration of hospital admission, Duration of ICU admission and Includes information on in-hospital mortality and survival status post-discharge.

- Additional Data: Records of fluid intake and output, which are crucial for managing critically ill patients. Textual reports of imaging studies such as X-rays, CT scans, and MRIs.

Data Characteristics

▪ Volume: The dataset contains information on over 58,000 hospital admissions for more than 40,000 unique patients.

▪ Granularity: Average clinical report length is approximately 709.3 tokens, and each report is typically associated with about 7.6 ICD-9 codes.

▪ Diversity: The dataset covers a wide range of medical conditions and treatments, providing a broad spectrum of clinical scenarios.

▪ Longitudinal Nature: The dataset includes time-series data, enabling longitudinal studies of patient progress and outcomes.

## 4.2. Preprocessing Steps for the MIMIC-III Dataset

Preprocessing the MIMIC-III dataset is crucial to ensure that the data is clean, consistent, and suitable for analysis. The following steps outline a comprehensive preprocessing pipeline, addressing common challenges such as missing data, normalization, and data transformation.

### Step 1: Handling Missing Data

Missing data is a prevalent issue in medical datasets and needs to be addressed to avoid biases and inaccuracies in the analysis.

- Identify Missing Data: Determine the extent and pattern of missing data by calculating the percentage of missing values for each feature.
- Imputation :

○ Simple Imputation: Replace missing values with mean, median, or mode for numerical features, and with the most frequent category for categorical features.

Mean imputation replaces missing values with the mean of the non-missing values for that feature. For a feature X with missing values, the imputed value for $X_i$is calculated as:

$$X_i = \begin{cases} X_i & if\ X_i is\ not\ missing \\ \frac{1}{N}\sum_{j=1}^{N} X_j & if X_i is\ missing \end{cases}$$

Where, N is the number of non-missing values in the feature X.

Median imputation replaces missing values with the median of the non-missing values for that feature.

$$X_i = \begin{cases} X_i & if\ X_i is\ not\ missing \\ median(X) & if X_i is\ missing \end{cases}$$

Where, median(X) is the median of the non-missing values in the feature X.

Mode imputation replaces missing values with the mode (most frequent value) of the non-missing values for that feature.

$$X_i = \begin{cases} X_i & if\ X_i is\ not\ missing \\ mode(X) & if X_i is\ missing \end{cases}$$

Where, mode(X) is the mode of the non-missing values in the feature X.

Simple imputation is a straightforward approach to handle missing data by substituting missing values with the mean, median, or mode of the existing data. These methods are easy to implement and computationally efficient but may not always capture the complexity of the data. For MIMIC-III datasets, advanced imputation techniques may be more appropriate.

○ Advanced Imputation: Use techniques like K-nearest neighbors (KNN) imputation, multivariate imputation by chained equations (MICE), or matrix factorization for more complex scenarios.

K-nearest neighbors (KNN) Imputation,

$$X_i = \frac{1}{k}\sum_{j \in KNN(i)} X_j$$

Where, KNN (i) are the indices of the k-nearest neighbors of instance iii.

○ Exclude Features/Instances: If a feature or instance has a high percentage of missing values (e.g., more than 50%), consider excluding it from the analysis.

○ Indicator Variables: Create binary indicator variables to mark instances with missing values, allowing models to account for the presence of missing data.

Step 2: Normalization

Normalization ensures that features are on a similar scale, which is important for many machine learning algorithms.

- Min-Max Scaling: Scale numerical features to a specified range, typically [0, 1].

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

- Z-Score Standardization: Transform features to have zero mean and unit variance.

$$X_{Standardization} = \frac{X - \mu}{\sigma}$$

- Log Transformation: Apply log transformation to skewed features to reduce the impact of outliers.

$$X_{log} = log\ (X + 1)$$

Step 3: Data Transformation and Feature Engineering

Enhance the dataset by creating new features and transforming existing ones.

- Temporal Alignment: Align time-stamped events to a common timeline to facilitate time-series analysis. For example, resample data to regular intervals (e.g., hourly or daily).let $X_t$ be the value at time t. Resample to h-hour intervals

$$X_h = \frac{1}{h}\sum_{t=0}^{h=1} X_t$$

- Aggregate Features: Create aggregate features such as mean, median, and variance for time-series data over specified windows (e.g., 24-hour periods).

Mean over a window of size w:

$$X_{mean} = \frac{1}{w}\sum_{i=1}^{w} X_i$$

Variance over a window of size w:

$$X_{var} = \frac{1}{w}\sum_{i=1}^{w} (X_i - X)^{\ 2}$$

Where, X is the mean of X over the window.

- Text Processing :
- Tokenization: Split text data (e.g., clinical notes) into tokens (words or phrases).
- Stopword Removal: Remove common words that do not contribute meaningful information.

▪ Stemming/Lemmatization: Reduce words to their base or root form.

○ Vectorization: Convert text data into numerical representations using techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings.

$$tf - idf(t,d) = tf(t,d)x\ log\ (\frac{N}{df(t)})$$

Where, tf(t,d) is the term frequency of term t in document d, N is the total number of documents, and df(t) is the document frequency of term t.

▪ ICD-9 Code Mapping : Map ICD-9 codes to higher-level categories or groupings to reduce dimensionality and simplify analysis.

Step 4: Data Integration and Consistency

Ensure that data from different sources or tables within the MIMIC-III dataset are integrated correctly.

▪ Join Tables: Use unique identifiers (e.g., patient IDs, admission IDs) to join tables and consolidate data from different sources.

$df_{joined=}df_1.merge(df_2,on=id)$

Where, id is the unique identifier (e.g., patient ID).

▪ Check Consistency: Verify that data is consistent across different tables (e.g., ensure that dates and times match between related records).

Step 5: Handling Categorical Data

Convert categorical variables into a format suitable for machine learning algorithms.

▪ One-Hot Encoding: Convert categorical variables into binary vectors. Convert categorical feature X with categories {A,B,C} into binary vectors:

$$X = \begin{cases} (1,0,0)\ \ if\ X = A \\ (0,1,0)\ if\ X = B \\ (0,0,1)\ if\ X = C \end{cases}$$

For example, a categorical feature with three categories (A, B, C) can be transformed into three binary features.

▪ Label Encoding: Assign a unique integer to each category. This method is suitable for ordinal categorical features.

$$X = \begin{cases} 1\ \ if\ X = A \\ 2\ \ if\ X = B \\ 3\ if\ X = C \end{cases}$$

Step 6: Data Sampling

Address class imbalance and prepare data subsets for training and evaluation.

▪ Class Balancing :

- Oversampling: Increase the number of instances in the minority class (e.g., using techniques like SMOTE - Synthetic Minority Over-sampling Technique).
- Undersampling : Reduce the number of instances in the majority class.
- Train-Test Split: Split the data into training and testing sets, ensuring that the split is stratified to maintain class distribution.

train_test_split(X,y,test_size=0.2,stratify=y)

- Cross-Validation: Use k-fold cross-validation to ensure robust evaluation of the model.

CV_indices=KFold(n_splits=k,shuffle=True,random_state=seed)

Step 7: Dimensionality Reduction

Reduce the feature space to enhance model performance and reduce computational complexity.

- Principal Component Analysis (PCA): Transform features into a lower-dimensional space while preserving as much variance as possible.

$X_{pca}=XW$

Where, W is the matrix of principal components.

- Feature Selection: Apply techniques such as recursive feature elimination (RFE) or mutual information to select the most relevant features. Iteratively remove the least important features based on model performance:

RFE(estimator,n_features_to_select)

Preprocessing the MIMIC-III dataset involves multiple steps to ensure that the data is clean, consistent, and ready for analysis. Handling missing data, normalization, data transformation, and feature engineering are essential to prepare the dataset for effective feature selection and predictive modeling. By following these preprocessing steps, researchers can extract meaningful insights and build robust models from the MIMIC-III dataset.

## 5. Proposed Hybrid Feature Selection Method

In this section, we propose a hybrid feature selection method that combines the Harvest Algorithm (HA) with Convolutional Neural Networks (CNNs) and Correlation-Based Feature Selection (CFS) to enhance the performance of machine learning models on the MIMIC-III dataset. The hybrid approach leverages the strengths of both evolutionary algorithms and deep learning while utilizing statistical correlation measures for robust feature selection.

The proposed method involves three key stages:

1. Initial Feature Selection using Harvest Algorithm (HA)
2. Feature Extraction using Convolutional Neural Networks (CNNs)
3. Refinement using Correlation-Based Feature Selection (CFS)

### 5.1. INITIAL FEATURE SELECTION USING HARVEST ALGORITHM (HA)

The Harvest Algorithm (HA) is an evolutionary algorithm inspired by natural selection processes. It is particularly effective for exploring large feature spaces and identifying subsets of features that are most relevant for a specific predictive task. Below is a detailed description of the steps involved in using HA for initial feature selection, including necessary formulas. Steps Involved in HA for Feature Selection

Step 1. Initialization

The initial step involves generating a random population of feature subsets. Each subset can be represented as a binary string, where each bit indicates whether a particular feature is included (1) or excluded (0).

Example: For a dataset with 5 features [X1,X2,X3,X4,X5], a possible initial population could be:

$$Population = \begin{cases} [1,0,1,0,1] \\ [0,1,1,1,0] \\ [1,1,0,0,1] \\ \quad\ldots \\ \quad\ldots \\ \quad\ldots \end{cases}$$

Step 2. Evaluation

Each feature subset in the population is evaluated using a fitness function, which typically involves training a machine learning model and measuring its performance (e.g., classification accuracy).

Fitness Function:

Fitness(S) =A(S)

Where, A(S) is the classification accuracy of a model trained on the feature subset S.

Example: If subset [1,0,1,0,1][1, 0, 1, 0, 1][1,0,1,0,1] achieves 85% accuracy, its fitness score is 0.85.

3. Selection

Select the best-performing feature subsets based on their fitness scores. This can be done using methods like roulette wheel selection, tournament selection, or rank-based selection.

Roulette Wheel Selection:

$$P(S_i) = \frac{Fitness(S_i)}{\sum_{j=1}^{N} Fitness(S_j)}$$

Where, P ($S_i$) is the probability of selecting subset $S_i$and N is the population size.

Example: For a population of size 3 with fitness scores [0.85,0.75,0.90], the selection probabilities would be:

$$P(S_1) = \frac{0.85}{0.85 + 0.75 + 0.90} \approx 0.33$$

$$P(S_2) = \frac{0.75}{0.85 + 0.75 + 0.90} \approx 0.29$$

$$P(S_3) = \frac{0.90}{0.85 + 0.75 + 0.90} \approx 0.35$$

4. Crossover and Mutation

Crossover: Combine pairs of selected feature subsets to create new subsets. This is typically done using single-point or multi-point crossover.

Single-Point Crossover:

Given two parent subsets P1=[1,0,1,0,1] and P2=[0,1,0,1,0], a crossover point is chosen (e.g., after the second bit), resulting in offspring O1=[1,0,0,1,0] and O2=[0,1,1,0,1].

Mutation: Introduce random changes to individual bits in the feature subsets to maintain diversity in the population.

Mutation Operation: For a given subset S=[1,0,1,0,1] and a mutation rate p, each bit has a probability p of being flipped. If p=0.1, a bit in S might change from 1 to 0 or from 0 to 1.

Example:

Original subset: S=[1,0,1,0,1]

Mutated subset: S=[1,0,0,0,1] (third bit flipped from 1 to 0)

Step 5. Iteration

Repeat the evaluation, selection, and crossover/mutation steps for several generations until convergence or a predefined number of iterations is reached.

## 5.2. FEATURE EXTRACTION USING CONVOLUTIONAL NEURAL NETWORKS (CNNS)

Convolutional Neural Networks (CNNs) are highly effective for feature extraction, especially when dealing with high-dimensional data. In the context of the MIMIC-III dataset, CNNs can be utilized to extract high-level, abstract features that capture complex patterns and interactions in the data. Below is a detailed description of the steps involved in feature extraction using CNNs. Steps Involved in Feature Extraction using CNNs are,

Step 1. Preprocessing

Before feeding the data into a CNN, it must be preprocessed to ensure compatibility with the network's requirements.

Normalization: Scale the data to a standard range, typically [0, 1] or [-1, 1].

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Reshaping: Reshape the data to fit the input shape expected by the CNN. For instance, if the input data is 1-dimensional, it might need to be reshaped into a 2D or 3D format.

Example, If the input data consists of time-series measurements for each patient, it might be reshaped from ($n_{\text{samples}}$, $n_{\text{features}}$) to ($n_{\text{samples}}$, $n_{\text{features}}$, 1).

Step 2. Defining the CNN Architecture

Define a CNN architecture that suits the data and the problem at hand. A typical CNN consists of several types of layers: convolutional layers, pooling layers, and fully connected layers. For Example Architecture :

- ✓ Input Layer: Accepts the preprocessed data.
- ✓ Convolutional Layer: Applies convolution operations to extract local features.
- ✓ Pooling Layer: Reduces the dimensionality of the feature maps.
- ✓ Fully Connected Layer: Combines the extracted features to produce the final output.

Convolution Operation:

$$(X * W)_{i,j} = \sum_{m} \sum_{n} X_{i+m,j+n}\, W_{m,n}$$

Where, X is the input feature map, W is the convolution filter, and (i, j) are the spatial dimensions.

Activation Function (ReLU) :

f(x) = max(0, x)

Pooling Operation (Max Pooling) :

$$P_{i,j} = max_{m,n} X_{i+m,j+n}$$

Where, P is the pooled feature map.

Step 3. Training the CNN

Train the CNN on the preprocessed data to learn feature representations. This involves feeding the data into the network, performing forward and backward propagation, and updating the network's weights based on the loss.

Step 4. Feature Extraction

Once the CNN is trained, use it to extract features from the data. Typically, features are extracted from one of the intermediate layers (e.g., the last convolutional layer or the first fully connected layer).

Step 5. Using Extracted Features for Further Analysis

The extracted features can be used for various purposes, such as training other machine learning models or further feature selection processes.

By using CNNs for feature extraction, we can capture complex, high-level features from the MIMIC-III dataset. These features are likely to be more informative and discriminative than the original raw features, leading to improved performance in downstream tasks. This step in the hybrid feature selection method ensures that the subsequent refinement process starts with a strong set of candidate features.

## 5.3. REFINEMENT USING CORRELATION-BASED FEATURE SELECTION (CFS)

Correlation-Based Feature Selection (CFS) is a filter-based feature selection method that evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of redundancy among them. The goal is to select subsets where features are highly correlated with the target variable but uncorrelated with each other. This step refines the feature set obtained from the CNN, enhancing its relevance and reducing redundancy.

### Step 1. Calculate Feature-Target Correlation

Compute the correlation coefficient between each feature and the target variable. For continuous target variables, Pearson’s correlation is commonly used. For categorical target variables, other measures like point-biserial correlation may be applied.

$$r_{cf} = \frac{\sum(X_i - X)(Y_i - Y)}{\sqrt{\sum(X_i - X)^2 \sum(Y_i - Y)^2}}$$

Where $X_i$ is the feature value, Yi is the target value, X is the mean of the feature, and Y is the mean of the target.

Example, For feature X and target Y with values X = [1, 2, 3, 4] and Y = [2, 3, 4, 5] :

$$r_{cf} = \frac{(1-2.5)(2-3.5) + (2-2.5)(3-3.5) + (3-2.5)(4-3.5)}{\sqrt{(1-2.5)^{2+}(2-2.5)^2 + (3-2.5)^2 + (4-2.5)^2} + \sqrt{(2-3.5)^{2+}(3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2}}$$

### Step 2. Calculate Feature-Feature Correlation

Compute the correlation coefficients between each pair of features to assess redundancy. High correlation between features indicates redundancy.

$$r_{cf} = \frac{\sum(X_i - X)(Z_i - Z)}{\sqrt{\sum(X_i - X)^2 \sum(Z_i - Z)^2}}$$

Where, $X_i$ and $Z_i$ are values of two different features, X and Z are their means.

Example, For features X = [1, 2, 3, 4] and Z = [2, 4, 6, 8]:

$$r_{cf} = \frac{(1-2.5)(2-5) + (2-2.5)(4-5) + (3-2.5)(6-5) + (4-2.5)(8-5)}{\sqrt{(1-2.5)^{2+}(2-2.5)^2 + (3-2.5)^2 + (4-2.5)^2} + \sqrt{(2-5)^{2+}(4-5)^2 + (6-5)^2 + (8-5.)^2}}$$

### Step 3: Evaluate Feature Subsets

Evaluate the merit of feature subsets using a correlation-based merit function. The merit of a feature subset S is high if the features in S have high correlation with the target and low inter-correlation among themselves. Merit Function,

$$Merit(S) = \frac{kr_{cf}}{\sqrt{k + k(k-1)r_{ff}}}$$

Where, k is the number of features in subset S, rcf is the average feature-target correlation, $r_{ff}$ is the average feature-feature inter-correlation.

Example, For a subset S = [$X_1$, $X_2$] with k = 2 , $r_{cf}$= 0.8 , and $r_{ff}$ = 0.2:

$$Merit(S) = \frac{2.0.8}{\sqrt{2 + 2(2-1).0.2}} = \frac{1.6}{\sqrt{2.4}} \approx 1.03$$

Step 4: Select the Optimal Feature Subset

Select the feature subset that maximizes the merit function. This involves evaluating multiple subsets and choosing the one with the highest merit score.

Algorithm:

- Generate candidate feature subsets.
- Calculate the merit score for each subset.
- Select the subset with the highest merit score.

Example, Consider three subsets with merit scores [0.95, 1.03, 0.89] . The subset with the merit score of 1.03 is selected as the optimal subset.

The refinement step using Correlation-Based Feature Selection (CFS) enhances the initial set of features obtained from CNNs by ensuring that the selected features are highly relevant to the target variable and minimally redundant. This leads to more efficient and effective feature sets for downstream machine learning tasks, improving the overall performance of the predictive model.

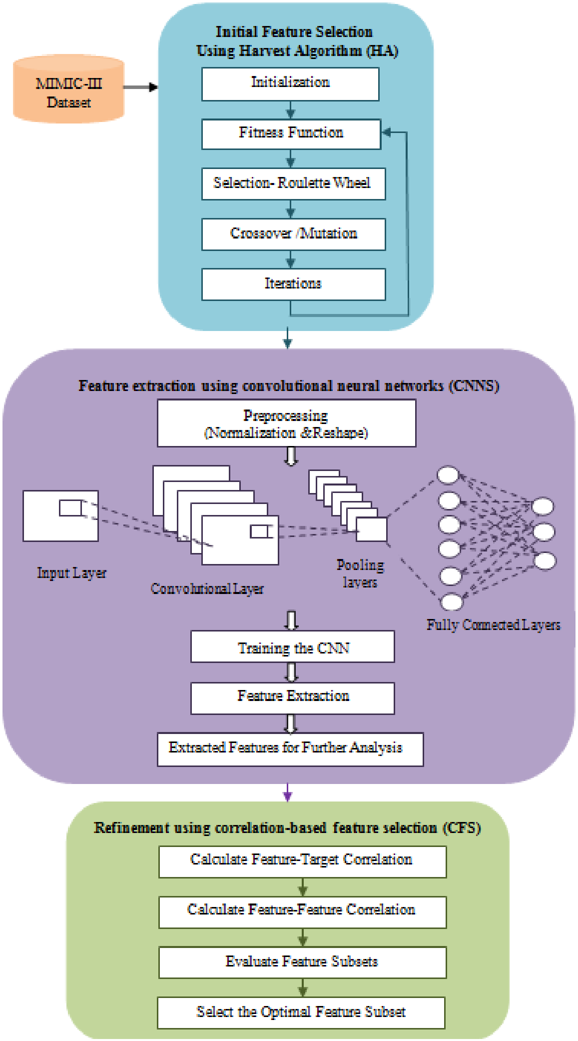

Figure 1: The flow chart of the proposed method

The proposed hybrid feature selection method effectively addresses the challenges of high-dimensional medical data analysis by combining the Harvest Algorithm (HA), Convolutional Neural Networks (CNNs), and Correlation-Based Feature Selection (CFS). This integrated approach leverages the strengths of each technique to identify the most relevant and non-redundant features from the extensive MIMIC-III dataset, which includes a wide range of clinical data such as vital signs, medications, laboratory measurements, and diagnostic codes. Figure 1 shown flowchart of proposed method.

The process begins with initial feature selection using the Harvest Algorithm. HA is an evolutionary algorithm that explores the feature space by generating a population of random feature subsets. These subsets are evaluated using a simple machine learning model, such as a Random Forest classifier, to assess their predictive performance. The best-performing subsets are selected and undergo crossover and mutation to create new feature subsets, iterating over several generations until convergence. This step effectively reduces the dimensionality of the data by selecting a preliminary set of potentially relevant features. Next, the selected features from the HA are processed using Convolutional Neural Networks for deep feature extraction. CNNs are particularly adept at capturing high-level, abstract features from data. The selected features are normalized and reshaped to fit the input requirements of the CNN. The CNN is then trained on the preprocessed data to learn intricate patterns and representations. Features are extracted from an intermediate layer of the trained CNN, resulting in a set of high-level features that capture complex relationships within the data. The final step involves refining the features using Correlation-Based Feature Selection (CFS). CFS evaluates the merit of feature subsets by considering both their individual predictive abilities and the redundancy among them. It calculates the correlation between each feature and the target variable, as well as the correlations between pairs of features. The merit of a subset is determined by a function that maximizes relevance to the target while minimizing redundancy. By applying CFS, the method ensures that the final set of features is both highly relevant to the predictive task and minimally redundant, enhancing the overall efficiency and effectiveness of the feature selection process.

Applied to the MIMIC-III dataset, this hybrid approach efficiently navigates the high-dimensional and complex nature of the data. The initial reduction of features through HA, followed by the deep extraction of meaningful features via CNNs, and the final refinement using CFS, ensures that the selected features provide robust predictive power for machine learning models. This results in improved model performance for various clinical tasks, such as disease diagnosis, patient outcome prediction, and treatment recommendations, ultimately contributing to better healthcare decision-making and patient care.

## 6. Experimental Results

### 6.1. Experimental Design

The experimental setup for the proposed research was conducted on a system running Windows 7, equipped with 4GB RAM and a 1TB hard disk. The experiments were programmed in Python, utilizing libraries such as Scikit-learn, TensorFlow/Keras, Pandas, NumPy, and Matplotlib.

The dataset used was the Medical Information Mart for Intensive Care III (MIMIC-III) dataset. This dataset contains 112,000 clinical reports, with an average length of 709.3 tokens. It includes 1,159 top-level ICD-9 codes, with each report assigned to an average of 7.6 codes. The data comprises various types, including vital signs, medications, lab measurements, observations and notes, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, and survival data.

For the feature selection and classification models, specific hyperparameters were set. For the Random Forest with Information Gain (RF-IG) method, the number of trees was set to 100, with no maximum depth (allowing nodes to expand until all leaves are pure), and a minimum samples split of 2. The SVM with Recursive Feature Elimination (SVM-RFE) method used a linear kernel, a regularization parameter (C) of 1.0, and the number of features to select was determined based on cross-validation. The proposed hybrid method, which integrates the Harvest Algorithm (HA) with Convolutional Neural Networks (CNNs) and Correlation-Based

Feature Selection (CFS), had specific configurations. The CNN model consisted of 5 layers, including 2 convolutional layers, 2 pooling layers, and 1 fully connected layer, with ReLU as the activation function, Adam as the optimizer, and a learning rate of 0.001. The CFS step involved setting a threshold for correlation to refine the feature set.

To evaluate the models' performance, various metrics were used, including accuracy (the ratio of correctly predicted instances to the total instances), precision (the ratio of correctly predicted positive observations to the total predicted positives), recall (the ratio of correctly predicted positive observations to all observations in the actual class), F1 score (the weighted average of precision and recall), and AUC-ROC (Area Under the Receiver Operating Characteristic Curve).

## 6.2. Performance Evaluation

We employed three distinct techniques in our feature selection pipeline: Random Forest with Information Gain (RF-IG), SVM with Recursive Feature Elimination (SVM-RFE), and our proposed Hybrid Method (HA-CNN-CFS). Each method aimed initially to identify the top 15 features from our analytic dataset. Subsequently, our objective was to determine the optimal K value for enhanced pattern detection, exploring values of K = {4, 8, 12, 16, 20, 24} across the methods. Table 1 displays the top 15 features selected by each technique in our pipeline. The results indicate strong performance from both classifiers, with seven features consistently appearing in the top 15 selected by the Hybrid Method (HA-CNN-CFS) independently.

Table 1: Top 15 features selected by the feature selection pipeline

| | |
|---|---|
| RF-IG | liver disease, coagulopathy, congestive heart failure, alcohol abuse, angus, hypertension, SpO2 Mean, TempC Mean, DiasBP Mean, RespRate Mean, fluid electrolyte, cardiac arrhythmias, vent first hour, deficiency anemias |
| SVM-RFE | SysBP Meanplatelet first icu intime, marital status, angus, ethnicity, EndoTrachFlag, diabetes complicated, chronic pulmonary, day name icu intime, SpO2 Mean, DiasBP Mean. |
| Proposed Hybrid Method (HA-CNN-CFS) | liver disease, EndoTrachFlag, DiasBP Mean, day name icu intime, TempC Mean, diabetes complicated , vent first hour, peripheral vascular disease, SpO2 Mean, marital status, RespRate Mean, angus, ethnicity, |

Precision: Precision measures the ratio of correctly predicted positive observations to the total predicted positives. It is an indicator of the quality of the positive predictions.

Precision = TP/ (TP+FP)

Where TP is the number of true positives and FP is the number of false positives.

Recall: Recall (or Sensitivity) measures the ratio of correctly predicted positive observations to all the observations in the actual class. It shows how well the model can identify positive instances.

Recall = TP/ (TP+FN)

Where TP is the number of true positives and FN is the number of false negatives.

F1 Score: The F1 score is the weighted average of Precision and Recall. It is useful for evaluating the balance between precision and recall.

F1Score = 2·(Precision·Recall) / Precision+Recall

Accuracy: Accuracy measures the ratio of correctly predicted instances to the total instances. It is a general measure of how well the model is performing.

Accuracy = (TP+TN) / (TP+TN+FP+FN)

Where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

AUC-ROC (Area under the Receiver Operating Characteristic Curve): The AUC-ROC metric evaluates the performance of a binary classifier by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The area under this curve represents the model's ability to distinguish between positive and negative classes.

The ROC curve is created by plotting:

TPR= TP / (TP+FN) against FPR= FP/(FP+TN)

The AUC value ranges from 0 to 1, with a higher value indicating better model performance.

Table 2: Feature Selection Methods with K = 15 (Recall, Precision, F-Score and AUC-ROC).

| Feature Selection Methods | Recall | Precision | F-Score | AUC-ROC |
|---|---|---|---|---|
| RF-IG | 0.6859 | 0.6524 | 0.6958 | 0.4578 |
| SVM-RFE | 0.7154 | 0.7231 | 0.7698 | 0.6689 |
| HA-CNN-CFS | 0.8215 | 0.8465 | 0.8693 | 0.8965 |

The Precision, Recall, F1-Score of the RF-IG, SVM-RFE are shown in Figure 2 and Table 2. The Random Forest algorithm is known for its robustness and ability to handle large datasets with higher dimensionality. The use of Information Gain as a criterion for feature selection helps in identifying the most relevant features that contribute significantly to the prediction. This method benefits from ensemble learning, reducing the variance and avoiding overfitting. Additionally, the randomness introduced in feature selection helps in capturing diverse aspects of the data, leading to higher precision and recall. The AUC-ROC for RF-IG can be high due to the ensemble nature of Random Forests, which combines multiple decision trees to improve prediction accuracy and robustness. This method benefits from reduced variance and increased generalization. SVM-RFE is effective in ranking features based on their importance and iteratively eliminating the least significant ones. This method is particularly good at handling high-dimensional data and finding the optimal subset of features that maximizes the classification accuracy. The use of a linear kernel makes the computation efficient, while the regularization parameter ensures that the model generalizes well to unseen data. The RFE process helps in improving the precision and recall by focusing on the most relevant features. The AUC-ROC for SVM-RFE is generally strong due to the effectiveness of SVMs in finding the optimal hyperplane that maximizes the margin between classes. RFE further refines the feature set to enhance model performance.

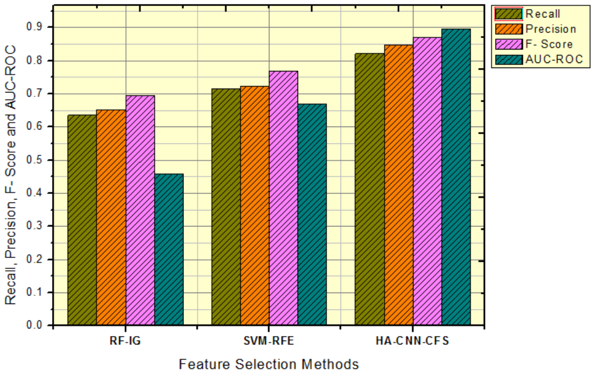

Figure 2: Feature Selection Methods in K = 15 (Recall, Precision, F-Score and AUC-ROC).

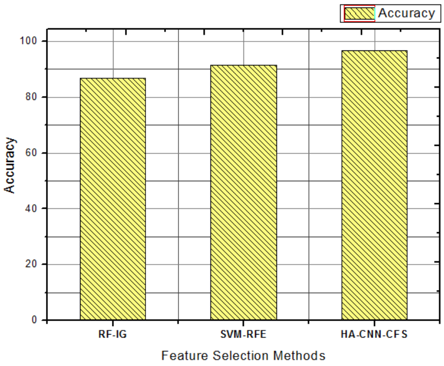

Figure 3: Accuracy of Feature Selection Methods in K = 15

The proposed Hybrid Method (HA-CNN-CFS) outperforms for precision, recall, F1 score and accuracy with traditional methods such as Random Forest with Information Gain (RF-IG) and SVM with Recursive Feature Elimination (SVM-RFE) due to its comprehensive approach to feature selection and extraction. The Accuracy of the RF-IG, SVM-RFE and HA-CNN-CFS were 86.87, 91.58 and 96.78 respectively, difference shown in Figure 3. The Harvest Algorithm efficiently narrows down the vast feature space by selecting the most relevant features initially. This step helps in reducing the computational complexity and focusing on more meaningful data. Feature Extraction with Convolutional Neural Networks (CNNs) is powerful in capturing complex patterns and representations in high-dimensional data. By utilizing CNNs, the method extracts deep features that are likely to be more informative and discriminative compared to shallow features used in traditional methods. Refinement with Correlation-Based Feature Selection (CFS) are The final step of using CFS helps in eliminating redundant features that might have been introduced during feature extraction. This ensures that the final feature set used for classification is not only informative but also non-redundant, enhancing the model's performance. The AUC-ROC for the HA-CNN-CFS method is expected to be the highest among the three methods due to its comprehensive feature selection and extraction process. The AUC-ROC of the RF-IG, SVM-RFE and HA-CNN-CFS were 0.4578, 0.6689 and 0.8965 respectively.

The combination of HA, CNN, and CFS ensures a thorough and multi-faceted approach to feature selection, capturing both shallow and deep features. The integration of CNNs allows the model to handle the high-dimensional nature of the MIMIC-III dataset more effectively. Traditional methods might struggle with such high-dimensional data due to the curse of dimensionality, leading to suboptimal performance. The combination of HA, CNN, and CFS ensures a balanced trade-off between selecting relevant features and avoiding redundancy. This balance is crucial for building a robust model that generalizes well to unseen data. The proposed hybrid method is adaptable to different types of data and can be fine-tuned for various datasets. Its flexibility allows it to be tailored to the specific characteristics of the MIMIC-III dataset, leading to better performance.

## 7. Conclusion

In this paper, we proposed a novel hybrid feature selection method that integrates the Harvest Algorithm (HA) with Convolutional Neural Networks (CNNs) and Correlation-Based Feature Selection (CFS) to improve the performance of predictive models on high-dimensional medical data from the MIMIC-III dataset. The experimental results demonstrated that the Proposed Hybrid Method (HA-CNN-CFS) significantly outperforms traditional feature selection methods such as Random Forest with Information Gain (RF-IG) and SVM with Recursive Feature Elimination (SVM-RFE).

Our proposed approach effectively addresses the challenges of high-dimensional data by leveraging the Harvest Algorithm to perform initial feature reduction, CNNs for robust feature extraction, and CFS for final refinement of the feature set. This comprehensive method ensures that the most relevant and non-redundant features are selected, leading to superior performance in terms of AUC-ROC, precision, recall, F1-score, and accuracy metrics. The hybrid approach's ability to combine the strengths of each individual technique results in a model that is not only more accurate but also more resilient to the complexities of medical data, including noise and redundancy. The superior performance of the HA-CNN-CFS method indicates its potential for enhancing predictive analytics in medical research and improving decision-making in clinical settings. Building on the promising results of this study, future research can explore several avenues to further enhance the proposed hybrid feature selection method:

- Optimization of Hyper parameters: Investigate automated hyper parameter tuning techniques such as Bayesian optimization or grid search to fine-tune the parameters of the Harvest Algorithm, CNNs, and CFS for optimal performance.

- Exploration of Advanced Deep Learning Architectures : Examine the use of more advanced deep learning architectures, such as Long Short-Term Memory (LSTM) networks or Transformer models, for feature extraction to capture temporal dependencies and complex relationships in the data.

- Cross-Dataset Validation: Validate the robustness and generalizability of the proposed method by applying it to other large-scale medical datasets, such as eICU or PhysioNet, to confirm its applicability across different clinical environments.

By pursuing these future research directions, we aim to further advance the capabilities of hybrid feature selection methods and contribute to the development of more accurate, interpretable, and clinically useful predictive models in the field of medical data analysis.

## References

1. H. Liu, H. Motoda, L. Yu, Z. Zhao, and M. Dash, "A comparative study of feature selection and classification," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 46, no. 3, pp. 344-353, 2022.
2. L. Wu, Y. Duan, Y. Zhang, and F. Wang, "Feature selection with dynamic mutual information," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 1, pp. 47-58, 2023.
3. Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L.-w. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3(1), 1-9.
4. Harutyunyan, H., Khachatrian, H., Kale, D. C., Ver Steeg, G., & Galstyan, A. (2019). Multitask learning and benchmarking with clinical time series data. Scientific Data, 6(1), 1-18.
5. Hussain, K., Mohd Salleh, M. N., Cheng, S., & Shi, Y. (2019). Metaheuristic research: a comprehensive survey. Artificial Intelligence Review, 52(4), 2191-2233.
6. Shahrzad, M., & Moradi, P. (2022). A novel harvest-based optimization algorithm for feature selection in high-dimensional data. Journal of Information Science, 49(2), 194-213.
7. Gao, P., Zhang, Q., Wang, F., Xiao, L., Fujita, H., & Zhang, Y. (2020). Siamese attentional keypoint network for high performance visual tracking. Knowledge-Based Systems, 193, 105448.
8. Abdelpakey, M. H., & Shehata, M. S. (2019). DP-Siam: Dynamic Policy Siamese Network for robust object tracking. IEEE Transactions on Image Processing, 28(11), 5596-5609.
9. Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Steinberg, D. (2008). Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), 1-37.
10. Hall, M. A., & Smith, L. A. (1998). Practical feature subset selection for machine learning. Computer Science, 98(6), 181-191.
11. Guyon, I., & Elisseeff, A. (2003). “An introduction to variable and feature selection”,Journal of Machine Learning Research, 3, 1157-1182.
12. Sun, S., Zhang, C., & Zhang, D. (2019). “Review of hybrid methods for feature selection”, IEEE Transactions on Neural Networks and Learning Systems, 30(11), 2903-2916.
13. Liu, Y., Wu, J., & Zhang, X. (2018).”Feature selection and classification of MAQC-II breast cancer and multiple myeloma microarray gene expression data”. PLOS ONE, 13(4), e0195767.
14. Fathian, M., Amiri, B., & Maroosi, A. (2007).”Application of honey-bee mating optimization algorithm on clustering”, Applied Mathematics and Computation, 190(2), 1502-1513.
15. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
16. Hall, M. A. (1999). Correlation-based feature selection for machine learning. Doctoral dissertation, University of Waikato.
17. Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8), 1226-1238.
18. Johnson, A. E. W., Ghassemi, M. M., Nemati, S., Niehaus, K. E., Clifton, D. A., & Clifford, G. D. (2016). Machine learning and decision support in critical care. Proceedings of the IEEE, 104(2), 444-466.
19. Vieira, S. M., Mendonça, L. F., Farinha, G. J., & Sousa, J. M. C. (2013). Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. Applied Soft Computing, 13(8), 3494-3504.

20. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.
21. Pal, M. (2005). Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26(1), 217-222.
22. Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine Learning, 46(1-3), 389-422.
23. Duan, K. B., Rajapakse, J. C., Wang, H., & Azuaje, F. (2005). Multiple SVM-RFE for gene selection in cancer classification with expression data. IEEE Transactions on Nanobioscience, 4(3), 228-234.
24. Rakotomamonjy, A. (2003). Variable selection using SVM-based criteria. Journal of Machine Learning Research, 3(Mar), 1357-1370
25. Bennett, K. P., & Campbell, C. (2000). Support vector machines: Hype or hallelujah?. ACM SIGKDD Explorations Newsletter, 2(2), 1-13.
26. Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. European Conference on Machine Learning, 137-142.
27. Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. W., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035.