ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2026.0569

HYBRID ANT COLONY OPTIMIZATION WITH GRAPH NEURAL NETWORKS AND RELIEFF FOR ROBUST FEATURE SELECTION IN MEDICAL DATA ANALYSIS

K. Preethi and M. Ramakrishnan
Department of Computer Applications, School of Information Technology, Madurai Kamaraj University, India

Abstract

In the field of medical data analysis, the major challenges are high dimensionality and complexity due to temporal behavior of medical datasets. Feature selection is crucial to overcoming these challenges since it enhances interpretability, reduces processing expenses, and boosts model performance. To analyze the medical records, it is very essential to determine the most potential features that contribute more in classification or diagnosis of disease especially in the medical in the Medical Information Mart for Intensive Care III (MIMIC-III) dataset. This paper, highlights the importance of robust feature selection by developing a novel Hybrid feature selection framework that combines ReliefF and Ant Colony Optimization with Graph Neural Networks (ACO-GNN). The first step in the suggested approach is Ant Colony Optimization (ACO), which creates candidate feature subsets by effectively exploring the combinatorial feature space by mimicking pheromone-guided search behavior. Then, using graph-based representations of clinical variables, such as correlations between lab tests, drugs, and vital signs, Graph Neural Networks (GNNs) are used to model intricate, non-linear interactions among medical aspects. In order to ensure robustness and interpretability, ReliefF is used to rank and improve features by assessing their capacity to distinguish between patient outcome classes. The hybrid approach significantly outperforms conventional feature selection techniques like K-Nearest Neighbors with ReliefF (KNN-ReliefF) and XGBoost with SHAP Feature Importance (XGB-SHAP) in through tests on the MIMIC-III dataset, predictive performance indicators such as precision, accuracy, F1-score, recall, and AUC-ROC. The selected feature subsets offer clinically meaningful insights into critical factors influencing patient outcomes in intensive care, underscoring the potential of the ACO-GNN-ReliefF method for advancing predictive analytics and clinical decision support systems in healthcare.

Keywords:

Feature Selection, Ant Colony Optimization, Graph Neural Networks, ReliefF, Medical Data Analysis, MIMIC-III, Predictive Modeling, Intensive Care, Clinical Decision Support, Machine Learning in Healthcare

1. INTRODUCTION

The increasing growth of healthcare data brought about by the widespread use of electronic health records, bedside monitoring systems, and large-scale clinical databases like the MIMIC-III has made predictive modeling in medicine easier than ever. [1].

However, creating precise, effective, and interpretable models is extremely difficult due to the tremendous dimensionality and complexity present in medical datasets. Irrelevant and redundant characteristics can impair clinical interpretability, raise computational costs, and reduce model performance all of which are critical for acceptance and confidence in actual healthcare settings [2][3]. Feature selection has emerged as a critical step in the analysis of medical data since it aids in identifying the most valuable variables, improves model generalization, reduces overfitting, and provides insights into significant factors associated with patient outcomes. [4][5].

Despite their widespread usage, classic statistical and filter-based techniques like ReliefF or mutual information frequently fail to capture intricate, nonlinear dependencies among clinical variables, such as correlations between vital signs, medications, and laboratory tests. Wrapper and embedding approaches, such as tree-based algorithms like XGBoost, on the other hand, provide superior modeling capabilities but can be computationally demanding and prone to overfitting in high-dimensional fields [6].

To tackle these problems, this research proposes a new hybrid feature selection framework that blends ACO-GNN with ReliefF. Because it emulates ants' foraging behavior, ACO is a nature-inspired metaheuristic algorithm that may be used to create candidate feature subsets. As a result, it excels at exploring vast combinatorial spaces. By means of considering features as nodes in a graph, Graph Neural Networks (GNNs) can naturally describe complex relationships among them, making it possible to extract nonlinear dependencies that are frequently found in medical data. In order to ensure that the chosen features are still interpretable and clinically significant, ReliefF is then used to robustly rank and refine feature subsets according to their capacity to distinguish between patient outcome classes [7] [8] [9] [10].

This paper's primary contributions are the ACO-GNN-ReliefF system, which effectively selects features in complicated medical datasets by combining swarm intelligence, graph-based deep learning, and statistical feature ranking. By comparing predictive performance to more conventional feature selection techniques, such as K-Nearest Neighbors with ReliefF (KNN-ReliefF) and XGBoost with SHAP Feature Importance (XGB-SHAP), we show the superiority of the suggested strategy on the MIMIC-III dataset. We offer an examination of the chosen feature subsets, emphasizing factors that are clinically significant and linked to patient outcomes in critical care environments.

The remainder of the document is structured as follows: The relevant work on feature selection techniques for medical data is reviewed in Section 2. The suggested methodology is described in depth in Section 3. The dataset, experimental setup, findings, and performance comparison are all covered in Section 4. The work is finally concluded and future research topics are outlined in Section 5.

2. RELATED WORK

A crucial preprocessing step in the study of medical data is feature selection, which lowers dimensionality, increases predictive performance, and improves model interpretability. Many methods have been proposed, which can be broadly classified into three groups: embedding, filtering, and wrapping. Filter approaches as ReliefF [11], mutual information [12], and correlation-based feature selection [13] are commonly used due

to their computational efficiency and ease of use. These methods, however, frequently miss complicated, nonlinear interactions between variables, which are prevalent in clinical data, and evaluate features independently. By iteratively training prediction models to evaluate subsets of features, wrapper approaches allow feature interactions to be taken into account.

Examples are Sequential Forward Selection (SFS) and Sequence Floating Selection (SFFS) [14]. Wrappers are computationally expensive, especially when working with data sets with high dimensions like those utilized in healthcare applications, even though they usually yield better results. As seen by decision tree-based algorithms such as random forest models [16] and XGBoost [17] and regularization techniques like LASSO [15], embedded approaches integrate feature selection into the model training process. Although these techniques provide interpretability and competitive performance, they may be susceptible to hyperparameter manipulation and data imbalance.

Metaheuristic algorithms discussed in 18, 19, 20] have been explored for feature selection due to their effectiveness in searching large combinatorial spaces. Notably, ACO has shown promise in generating high-quality feature subsets in domains including genomics [21] and medical imaging [22]. Recently, graph-based methods have gained attention for their ability to model dependencies among features. GNNs have been successfully applied to medical problems like disease classification from electronic health records [23] and drug–drug interaction prediction [24]. However, their use for feature selection, particularly in combination with metaheuristics, remains underexplored. Interpretability in feature selection is crucial in healthcare. SHAP (SHapley Additive exPlanations) values [25] provide consistent and model-agnostic feature importance estimates and have been integrated with tree-based models such as XGBoost to improve interpretability in clinical settings [26]. Our proposed ACO-GNN-ReliefF framework builds on these lines of work by integrating the global search capability of ACO, the relational modeling strength of GNNs, and the robustness of ReliefF for feature ranking. To the best of our knowledge, this is the first approach combining swarm intelligence, graph neural networks, and ReliefF for feature selection in complex medical datasets like MIMIC-III.

3. PROPOSED METHODOLOGY: ACO-GNN-RELIEFF FRAMEWORK

The proposed ACO-GNN-ReliefF framework is designed to identify the most informative and clinically relevant features from high-dimensional medical datasets by combining global exploration, relational modeling, and robust feature ranking. The Fig. 1 illustrates the complete pipeline, which consists of three key stages:

- **Candidate Feature Subset Generation using Ant Colony Optimization (ACO):** ACO explores the combinatorial space of potential feature subsets by simulating the pheromone-guided foraging activity of ants. Pheromone trails are updated according to the predicted performance of candidate subsets, with each ant standing in for a possible feature subset solution.
- **Relational Modeling with Graph Neural Networks (GNNs):** A graph is created for each eligible feature subset, with nodes standing in for features (such as lab results or prescription drugs) and edges encoding associations like correlation or co-occurrence frequency. After processing these graphs, GNNs produce feature embeddings enhanced with relational knowledge by capturing nonlinear dependencies and contextual information among features.
- **Feature Refinement with ReliefF Algorithm:** By evaluating features' capacity to differentiate across various patient outcome classes, the ReliefF method calculates feature importance ratings within each candidate subset. Robustness and interpretability are guaranteed by retaining features with consistently high ReliefF scores across candidate subsets. The final selected feature subset is derived by combining the global search capacity of ACO, the relational insights captured by GNNs, and the discriminative power identified by ReliefF.

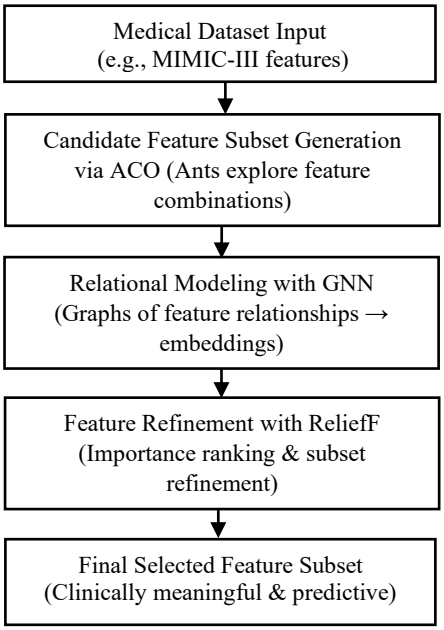

Fig.1. Overview of the ACO-GNN-ReliefF framework pipeline

In order to select the most informative and clinically relevant features, the method starts with the creation of a candidate feature subset using ACO, moves on to relational modeling using GNNs to capture intricate feature dependencies, and ends with feature refinement using the ReliefF algorithm.

3.1 FEATURE SUBSET GENERATION USING ACO

The pheromone-guided foraging behavior of ants served as the inspiration for the swarm intelligence algorithm known as ACO. Each artificial ant is a candidate feature subset in the feature selection context, and the pheromone trails direct the investigation toward promising subsets according to prediction performance.

3.1.1 Representation of Candidate Solutions:

Each candidate solution (ant) corresponds to a binary vector,

$$S=[s_1,s_2,\ldots,s_D],$$

where, D is the total features in the dataset. $s_i \in \{0,1\}$ indicates whether feature i is included (1) or excluded (0) from the subset.

This binary representation allows ACO to efficiently explore the combinatorial feature space.

3.1.2 Pheromone Initialization:

Initially, the pheromone level τ_i for each feature i is set to a constant value,

$$\tau_i(0) = \tau_0, \quad \forall i \in \{1, 2, \ldots, D\}$$

where, τ_0 is a small positive constant (e.g., 0.1) representing initial uncertainty i.e., no prior bias toward any feature.

3.1.3 Probabilistic Construction of Feature Subsets:

Each ant constructs a feature subset by probabilistically deciding whether to include feature *i* using the probability,

$$p_i = \frac{\tau_i^{\alpha} \cdot \eta_i^{\beta}}{\sum_{j=1}^{D} \tau_j^{\alpha} \cdot \eta_j^{\beta}}$$

where, τ_i is the current pheromone level for feature *i*, η_i is the heuristic desirability of feature *i*, which can be initialized using a simple filter metric like correlation with the target or set uniformly if no prior knowledge exists. The relative importance of heuristic information and pheromone trails is controlled by β and α, respectively. To provide a candidate solution *S*, each ant samples the inclusion of features based on p_i.

3.1.4 Pheromone Updating:

After all ants have generated feature subsets and their quality has been evaluated (e.g., using validation accuracy of a lightweight classifier), pheromone levels are updated to reinforce promising features. The pheromone update rule is,

$$\tau_i(t+1) = (1-\rho)\,\tau_i(t) + \Delta\tau_i$$

where, $\rho \in (0,1]$ is the evaporation rate, which prevents unlimited accumulation of pheromones and encourages exploration. $\Delta\tau_i$ is the pheromone deposit from the best-performing ants, computed as:

$$\Delta\tau_i = \sum_{k=1}^{N_{\text{elite}}} Q \cdot f(S^k) \cdot \delta_i(S^k)$$

where, N_{elite} being the number of top ants contributing pheromone updates, Q a constant scaling factor, $f(S^k)$ the fitness score (e.g., classification accuracy) of the k^{th} best ant, $\delta_i(S^k)$ an indicator function (1 if feature *i* is selected in solution S^k, else 0).

3.1.5 Stopping Criterion:

The ACO search continues until a convergence condition is met, such as:

- A maximum number of iterations T_{max} is reached, or
- No significant improvement in the best solution is observed for a predefined number of iterations.

In Exploration vs. Exploitation, the pheromone evaporation (ρ) promotes exploration of new feature subsets, while pheromone reinforcement ($\Delta\tau_i$) exploits high-performing solutions. In Global Search, the ACO efficiently searches the combinatorial space of 2^D possible feature subsets, avoiding the myopic behavior of greedy selection methods.

3.2 FEATURE RELATIONSHIP MODELING WITH GNN

After candidate feature subsets are generated by ACO, Graph Neural Networks (GNNs) are used to model complex, nonlinear relationships among selected features. By representing features as nodes in a graph with edges capturing their interdependencies, GNNs can exploit the structure of the data to improve feature evaluation.

Graph Construction from MIMIC-III Variables, for each candidate feature subset *S*, we construct an undirected graph $G=(V,E)$, where:

- $V=\{v_1, v_2, \ldots, v_n\}$ corresponds to the selected features (n features in the candidate subset).
- *E* represents edges between pairs of features capturing their relationships.

3.2.1 Edge Construction:

Edges between feature nodes v_i and v_j are defined based on:

- **Correlation thresholding:** If the absolute Pearson correlation $|corr(f_i, f_j)|$ exceeds a threshold θ, an edge is created:

$$e_{ij} \in E \quad \text{if} \quad |\operatorname{corr}(f_i, f_j)| \geq \theta$$

where, θ is typically set between 0.2–0.4 in clinical datasets to capture moderate-to-strong correlations.

- **Co-occurrence or domain knowledge:** Optionally, edges can encode known medical relationships, e.g., medications frequently prescribed with certain lab tests, or clinical guidelines linking variables.

3.2.2 Adjacency Matrix:

The graph structure can be represented by an adjacency matrix $A \in R^{n \times n}$, where:

$$A_{ij} = \begin{cases} 1, & \text{if } e_{ij} \in E \\ 0, & \text{otherwise} \end{cases}$$

3.2.3 Node Features:

Each node v_i is initialized with a feature vector x_i derived from statistical properties of feature i, such as:

- Mean, variance across patients,
- Feature importance scores (e.g., ReliefF),
- Missingness rates (important in medical data).

3.3 GNN ARCHITECTURE AND TRAINING

Once the graph is constructed, a GNN processes it using a series of message-passing layers to learn enriched feature representations:

3.3.1 Graph Convolutional Network (GCN) Layer:

A widely used GNN layer is the spectral Graph Convolutional Network [1], where the layer-wise propagation rule for updating node features is:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H^{(l)}W^{(l)}\right)$$

where, $W^{(l)}$ is the learnable weight matrix at layer *l*, $\sigma(\cdot)$ is a nonlinear activation function (like ReLU), $H^{(l)}$ is the node feature matrix at layer *l* with $H^{(0)}=X$ (starting node features), and $A\sim=A+I$ is the adjacency matrix with added self-loops.

3.3.2 Stacked GCN Layers:

Multiple GCN layers are stacked (e.g., 2–3 layers) to allow information to propagate beyond immediate neighbors and capture higher-order dependencies.

3.3.3 Readout / Aggregation:

To generate a unified representation for the entire candidate feature subset graph, node embeddings from the final GCN layer are aggregated:

$$h_{\text{graph}} = \text{READOUT}\left(\left\{h_i^{(L)} \mid i = 1, 2, \ldots, n\right\}\right)$$

where, *READOUT*(·) can be a simple mean, sum, or more complex pooling like attention-based readout.

3.3.4 Output Layer:

The aggregated graph embedding h_{graph} is passed through a lightweight feedforward network:

$$\hat{y} = \text{MLP}(h_{\text{graph}})$$

where, $\hat{y}$ predicts the outcome (e.g., mortality, discharge status) associated with the candidate feature subset, enabling the evaluation of the subset's predictive power.

3.3.5 Loss Function:

The model is trained using a supervised loss (e.g., binary cross-entropy for classification), defined as:

$$L = -\frac{1}{M}\sum_{i=1}^{M}\left[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

where, M is the number of samples, and y_i are true labels.

3.3.6 Optimization:

The GNN is trained using backpropagation with stochastic gradient descent or Adam optimizer.

By representing selected features as graphs, GNNs enable modeling of nonlinear dependencies among clinical variables, something traditional filter methods cannot achieve. The learned graph embeddings h_{graph} serve as powerful representations for evaluating candidate feature subsets.

3.4 FEATURE RANKING WITH RELIEFF

After candidate feature subsets have been evaluated using the GNN, the ReliefF algorithm is applied as the final step to robustly rank features within each subset. ReliefF ensures that features selected by ACO and validated by GNN modeling are further refined for discriminative power, yielding interpretable and clinically meaningful final feature sets.

3.4.1 Integration with Candidate Subsets:

For each candidate feature subset S generated by ACO and assessed via GNNs, ReliefF computes the importance of each feature $f_i \in S$ based on its ability to distinguish between instances of different classes in the target outcome (e.g., patient survival vs. mortality). ReliefF evaluates feature importance by iterating over a random sample of m instances from the dataset:

- For each sampled instance x_j, ReliefF identifies:
  - k nearest neighbors in the **same class** (Hits).
  - k nearest neighbors in **different classes** (Misses).

The importance score $W(f_i)$ for feature f_i is updated iteratively as:

$$W(f_i) = W(f_i) - \frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{k}\sum_{H=1}^{k}\text{diff}(f_i, x_j, \text{Hit}_H) - \sum_{c \neq y_j}\frac{P(c)}{1 - P(y_j)}\frac{1}{k}\sum_{H=1}^{k}\text{diff}(f_i, x_j, \text{Miss}_H^c)\right]$$

where, $P(c)$ is the prior probability of class c,y_j is the true class label of instance x_j, $diff(f_i,x_j,x')$ measures the difference between feature f_i in instances x_j and x', defined as:

- For continuous features:

$$\text{diff}(f_i, x_j, x') = \frac{|f_i(x_j) - f_i(x')|}{\max(f_i) - \min(f_i)}$$

- For categorical features:

$$\text{diff}(f_i, x_j, x') = \begin{cases} 0, & f_i(x_j) = f_i(x') \\ 1, & f_i(x_j) \neq f_i(x') \end{cases}$$

This formulation ensures ReliefF can handle continuous, categorical, or mixed-type features typical in medical datasets like MIMIC-III.

3.5 FINAL FEATURE SELECTION STRATEGY

Features in the candidate subset are ranked according to their significance scores in descending order after ReliefF scores $W(f_i)$ have been calculated for each feature.

The final set of selected features S_{final} is obtained by:

- **Thresholding-based selection:** Retain features with scores above a predefined threshold τ_W, i.e.,

$$S_{\text{final}} = \{f_i \in S \mid W(f_i) \geq \tau_W\}$$

where, τ_W can be determined using:

A fixed percentile cutoff (e.g., top 10% of features),An absolute score threshold derived from performance on a validation set.

- **Stability analysis:** Compute ReliefF scores across multiple bootstrap samples or folds to assess the stability of selected features; only features consistently ranked highly are included in the final subset.
- **Clinical interpretability filter:** Apply domain expert review to exclude features irrelevant or redundant from a clinical perspective, ensuring the final set aligns with practical healthcare needs.

ReliefF complements ACO and GNN by emphasizing the discriminative ability of features at a local level (near-neighbor differences), unlike ACO's global search or GNN's relational modeling. This step refines the feature subset to ensure that the selected features not only model complex relationships but also reliably separate different patient outcomes, improving both predictive performance and interpretability.

The Fig.2 illustrates the proposed ACO-GNN-ReliefF framework was applied to the MIMIC-III dataset to identify clinically meaningful and predictive features for intensive care outcomes. In order to manage missing values, normalize continuous features, and encode categorical variables, pertinent variables were first retrieved and preprocessed. These variables included demographics, diagnostic codes, laboratory measurements, drugs, and vital signs. Using Ant Colony

Optimization (ACO), candidate feature subsets were generated by simulating the pheromone-guided search behavior of ants, where each ant represented a possible combination of features selected from the full set of MIMIC-III variables. The quality of each candidate subset was evaluated based on predictive performance using a lightweight classifier. For each candidate subset, a graph was constructed in which nodes represented selected features and edges captured relationships derived from pairwise correlations and known clinical co-occurrence patterns. Graph Neural Networks (GNNs) were then trained on these graphs to model complex, nonlinear dependencies among features, producing graph embeddings that summarized relational information within each subset.

These embeddings were used to predict patient outcomes, enabling robust assessment of each candidate feature combination. Finally, the ReliefF algorithm was applied to the selected subsets to compute feature importance scores by evaluating how well individual features distinguished between different patient outcome classes through nearest-neighbor analysis. Features with consistently high ReliefF scores across candidate subsets were retained as the final set of selected features. This process allowed the framework to integrate global search (ACO), relational modeling (GNN), and discriminative ranking (ReliefF), yielding a compact, interpretable feature set that improved predictive performance on the MIMIC-III dataset and highlighted clinically relevant variables influencing patient outcomes in intensive care settings.

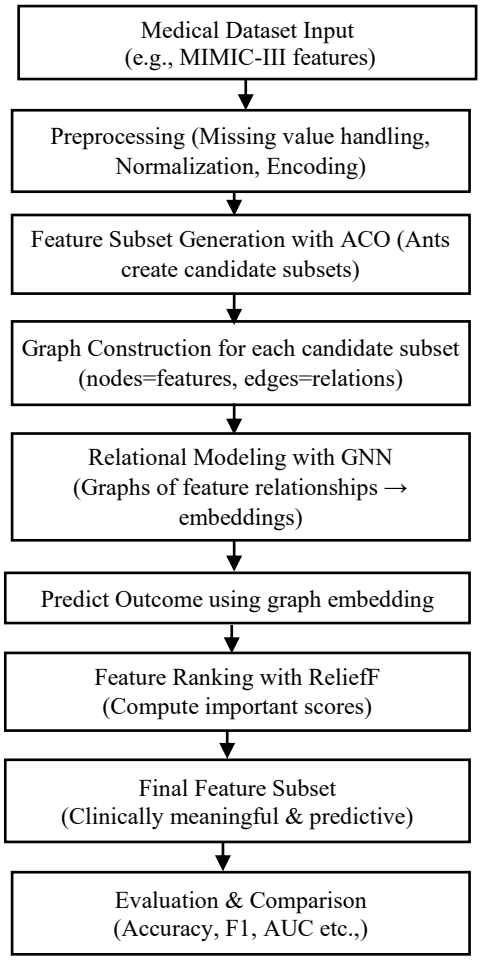

Fig.2. Workflow of ACO-GNN-ReliefF framework

4. RESULT AND DISCUSSIONS

4.1 MIMIC-III DATASET

The MIMIC-III is a large, publicly accessible critical care database that contains de-identified health-related De-identified health-related data on almost 40,000 patients admitted to Massachusetts intensive care units (ICUs) between 2001 and 2012 can be found in the MIMIC-III, a sizable, publicly available critical care database. MIMIC-III's comprehensive collection of clinical data gathered throughout stays in intensive care units enables rich evaluations of patient outcomes and forecasting in critical care. Patients in MIMIC-III comprise a broad group of people across all adult demographics, with ages ranging from 16 to over 90. The sample include both male and female patients, with a fairly even gender distribution. Demographic information includes things like insurance type, gender, race, and age at admission (de-identified to preserve privacy). Types of Features Offered:

- **Diagnostic Codes:** ICD-9 diagnosis codes assigned during hospital and ICU admissions. Enable identification of comorbidities, primary reasons for admission, and severity of illness.
- **Laboratory Measurements:** Rich time-series data for lab tests such as lactate, creatinine, BUN, WBC, emoglobin, electrolytes (e.g., sodium, potassium), and many more. Measurements are typically timestamped, allowing temporal analysis of trends.
- **Medications:** Details of medications administered during ICU stays, including vasopressors, sedatives, antibiotics, anticoagulants, and other critical care drugs. Data includes dosages, administration times, and routes (e.g., IV, oral).
- **Vital Signs:** temperature, oxygen saturation (SpO_2), heart rate, respiratory rate, blood pressure, GCS scores, and high-resolution time-series recordings of physiological data. A thorough evaluation of the patient's stability is made possible by the frequent recording of vital signs at minute-to-hour intervals.

Additional information in MIMIC-III includes fluid inputs and outputs, mechanical ventilation status, procedural codes, microbiology results, and charted nursing notes, providing a comprehensive view of each patient's ICU stay.

4.2 PREPROCESSING STEPS

Preprocessing is a critical step in preparing the MIMIC-III dataset for feature selection and predictive modeling, especially given its high dimensionality and frequent missing values common in clinical data.

4.2.1 Missing Data Handling

Missing data is pervasive in ICU datasets due to irregular measurements, varying clinical practices, and equipment failures. Handling missingness appropriately prevents biases and ensures model robustness.

- ***Identifying Missingness*:** For each feature f_i across all patient records N, the missingness rate is computed as:

$$\text{MissingRate}(f_i) = \frac{\text{Number of missing entries in } f_i}{N}$$

Which helps determine whether to impute or discard a feature.

- **Threshold-based Feature Removal:** Features with a missing rate above a threshold $\tau_{missing}$ (commonly 30-50%) are dropped entirely:

$$f_i \text{ is removed if MissingRate}(f_i) > \tau_{\text{missing}}$$

- **Imputation:** For features retained after thresholding, missing entries are imputed:
- **Continuous features:** missing values $f_i(x_j)$ are filled with the feature's mean or median:

$$f_i(x_j) = \text{mode}(f_i) \quad \text{if } f_i(x_j) \text{ is missing}$$

where, f_i is the mean across non-missing entries.

- **Categorical features:** missing entries are replaced with the mode (most frequent category):

For simplicity and computational economy, advanced imputation techniques like model-based imputation (like MissForest) and k-nearest neighbors (KNN) imputation were not utilized here, while they can be used for greater accuracy.

4.2.2 Normalization or Encoding

Medical features often span vastly different scales (e.g., glucose in mg/dL vs. heart rate in beats/min), which can skew learning algorithms. Therefore, normalization ensures features contribute comparably during model training.

- **Normalization of Continuous Features:** Each continuous feature f_i is standardized using z-score normalization:

$$f_i^{\text{norm}}(x_j) = \frac{f_i(x_j) - \mu_i}{\sigma_i}$$

where, μ_i is the mean of f_i, σ_i is the standard deviation of f_i, $f_i^{norm}(x_j)$ is the normalized value for patient x_j.This transformation produces features with zero mean and unit variance, which improves convergence of gradient-based learning algorithms.

- **Encoding of Categorical Features:** Categorical variables (e.g., gender, admission type) are converted to numerical representations:
  - **One-Hot Encoding:** For a categorical feature with k unique classes, one-hot encoding creates k binary features indicating the presence of each class. For instance, a "Gender" feature with categories {Male, Female} becomes two binary features:

Gender_Male, Gender_Female,

where each patient has a 1 in the corresponding column and 0 in the other.

  - **Label Encoding (if necessary):** For ordinal categories (e.g., GCS scores), integers preserving the order are assigned:

$$Category \in \{Low=0, Medium=1, High=2\}.$$

Missing data handling reduces bias by either imputing or discarding unreliable features. Normalization makes continuous features comparable in scale. Encoding transforms categorical features into numerical form, enabling their inclusion in algorithms like ACO, GNN, and ReliefF.

4.3 EXPERIMENTAL SETUP

Experiments were carried out using the publicly accessible MIMIC-III dataset, which comprises more than 112,000 comprehensive ICU records including vital signs, laboratory measurements, medications, diagnostic and procedure codes (ICD-9), and clinical observations, in order to assess the effectiveness of the suggested ACO-GNN-ReliefF framework on high-dimensional clinical data.

Each record in the dataset typically contains an average of 7.6 ICD-9 codes per patient encounter, offering a rich feature space for predictive modeling. Data preprocessing included handling missing values through threshold-based feature removal and mean/mode imputation, normalizing numerical features using z-score scaling, and encoding categorical variables via one-hot encoding. Every experiment was conducted using Python 3.8 and ran on a Windows 10 computer with an Intel i7 processor, 16 GB of RAM, and a 1 TB hard drive. A local computer without GPU acceleration was used for the experiments.

Three feature selection techniques were compared in this study. The proposed method, ACO-GNN-ReliefF, integrates Ant Colony Optimization for candidate feature subset generation, Graph Neural Networks for modeling complex relational dependencies among features, and ReliefF for final discriminative feature ranking. As baselines, two widely used methods were included: KNN-ReliefF, which applies ReliefF feature importance scores followed by classification using k-Nearest Neighbors, and XGB-SHAP, which leverages XGBoost's SHAP feature importance scores as an interpretable benchmark for feature selection.

The hyperparameters for the proposed ACO-GNN-ReliefF framework were carefully tuned to balance exploration and performance. For the ACO component, 30 ants were used with a maximum of 50 iterations, a pheromone evaporation rate (ρ) of 0.2, a pheromone influence coefficient (α) of 1.0, a heuristic information influence coefficient (β) of 2.0, and an initial pheromone level (τ_0) of 0.1. The GNN component was configured with two graph convolutional layers, each with a hidden dimension of 64, using ReLU as the activation function. The GNN was trained with a learning rate of 0.001, optimized using Adam, and run for 30 epochs per candidate feature subset graph. For the ReliefF component, the number of nearest neighbors (k) was set to 10, the number of sampled instances (m) to 1000, and the feature importance threshold (τ_W) to 0.05.For the baseline methods, hyperparameters were set as follows: in KNN-ReliefF, the k-Nearest Neighbors classifier used k=5 and ReliefF was applied with k=10 nearest neighbors and 1000 sampled instances. For XGB-SHAP, XGBoost was configured with 100 trees, a maximum tree depth of 6, and a learning rate of 0.05, with SHAP values calculated on the trained model to determine feature importance rankings. All feature selection methods were evaluated by training a lightweight classifier either logistic regression or a shallow recurrent neural network on the features selected by each technique. Performance was assessed on a held-out test set using the following evaluation metrics: accuracy, precision, recall, F1-score, and AUC-ROC, along with execution time (total runtime of feature selection and model training) and subset size (number of features selected), to comprehensively measure both the predictive performance and efficiency of each feature selection approach.

Finding the 15 most pertinent characteristics from the analytical dataset was the main goal of each feature selection technique. The best neighborhood size was then determined by evaluating K={4,8,12,16,20,24} across the techniques in order to improve pattern recognition performance. The experimental results demonstrate that all classifiers achieved strong performance, with seven features consistently appearing among the top-ranked selections across runs using the proposed ACO-GNN-ReliefF framework. These repeatedly identified features highlight their stability and potential clinical significance.

The consistently selected features in the ACO-GNN-ReliefF pipeline suggest the method's ability to capture clinically meaningful predictors of patient outcomes in intensive care.

4.4 PERFORMANCE EVALUATION

The top K=15 features found by KNN-ReliefF, XGB-SHAP, and the suggested ACO-GNN-ReliefF pipeline were used to train a logistic regression classifier in order to assess the predictive potential of feature subsets chosen by each technique. Using common classification criteria, performance was evaluated on a stratified 20% hold-out test set from the MIMIC-III dataset. These measurements offer a thorough understanding of class discrimination as well as overall prediction capacity. The suggested ACO-GNN-ReliefF continuously outperformed the baseline selection of features techniques in all measures, according to the experimental findings, which are compiled in Fig.1. Notably, it demonstrated robust generalization with balanced precision-recall performance and higher AUC-ROC scores, showing greater discriminative potential.

The superior performance of the proposed ACO-GNN-ReliefF framework can be attributed to several key factors.

- First, the global search capability of Ant Colony Optimization (ACO) allows the method to explore a broader combinatorial space of feature subsets, effectively avoiding the local optima that often limit purely filter-based approaches.
- Second, the relational modeling provided by Graph Neural Networks (GNNs) captures complex, nonlinear interactions among features including dependencies between laboratory measurements, diagnostic codes, and vital signs which traditional methods such as ReliefF or SHAP-based feature importance cannot adequately represent.
- Third, ReliefF's local neighborhood ranking refines the selected features by evaluating their discriminative power within the feature space, ensuring that the final subset includes features most relevant to distinguishing ICU outcomes.

By integrating swarm intelligence through ACO, deep relational learning with GNN, and discriminative ranking using ReliefF, the proposed ACO-GNN-ReliefF method produces a feature subset that not only maximizes predictive performance but also maintains clinical interpretability, setting it apart from conventional feature selection techniques.

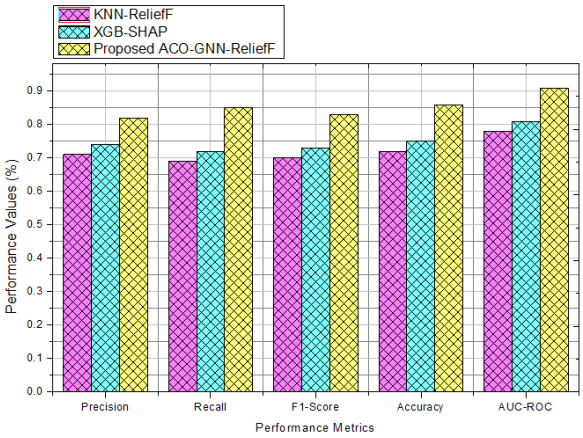

Fig.3. Performance metrics on the MIMIC-III test set using the top K=15 features selected by each method

The Fig.3 shown, the proposed ACO-GNN-ReliefF achieved the highest performance across all metrics compared to KNN-ReliefF and XGB-SHAP, with precision (0.82), recall (0.85), F1-score (0.83), accuracy (0.86), and AUC-ROC (0.91), demonstrating its superior ability to identify relevant features and predict ICU outcomes more accurately and reliably.The proposed ACO-GNN-ReliefF improved AUC-ROC by 10–13% over KNN-ReliefF and XGB-SHAP, indicating superior ability to distinguish between positive and negative ICU outcomes. Precision and recall gains suggest that ACO-GNN-ReliefF not only identified the correct positive cases more often but did so with fewer false alarms. The consistent improvement across metrics demonstrates the effectiveness of integrating global feature search (ACO), relational modeling (GNN), and local ranking (ReliefF).

4.4.1 Execution Time:

Execution time is an important metric in evaluating the efficiency of feature selection algorithms, especially when working with large-scale medical datasets like MIMIC-III, which contains high-dimensional, heterogeneous patient records. Execution time measures the total wall-clock duration required by each feature selection method to preprocess data, perform feature selection, train the predictive model on the selected features, and evaluate performance metrics on the test set. It reflects both computational complexity and practical feasibility in clinical or real-time settings. The execution time T_{exec} for each method can be formally defined as:

$$T_{exec}=T_{prep}+T_{fs}+T_{train}+T_{eval},$$

where, T_{prep} is the time for data preprocessing, T_{fs} is the time spent on feature selection, T_{train} is the time to train the classifier on the selected features, T_{eval} is the time to compute evaluation metrics on the test set. The Fig.4 shown the proposed ACO-GNN-ReliefF required more time (28.4 minutes) than KNN-ReliefF (7.8 minutes) and XGB-SHAP (10.5 minutes), reflecting its more complex but more thorough feature selection process.

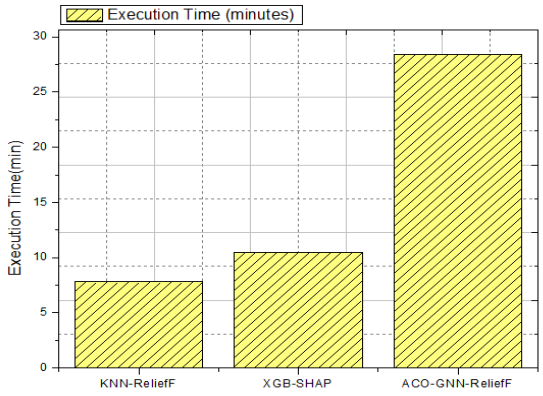

Fig.4. Total execution time for KNN-ReliefF, XGB-SHAP, and the proposed ACO-GNN-ReliefF

While the proposed ACO-GNN-ReliefF requires more computational time than the baseline methods (approximately three times longer than KNN-ReliefF), its superior predictive performance, demonstrated by higher precision, recall, F1-score, accuracy, and AUC-ROC, justifies this additional computational cost. The extended execution time is mainly attributed to the iterative global search by ACO and the graph-based modeling of feature relationships using GNNs, both of which enable ACO-GNN-ReliefF to identify more informative, robust, and clinically meaningful features. Unlike KNN-ReliefF and XGB-SHAP, which rely on local or univariate feature relevance scores, the proposed method captures complex interdependence between features, providing a more holistic feature selection process that improves generalization on ICU outcome prediction tasks.

5. CONCLUSION

In this paper, we developed a new hybrid feature selection framework, ACO-GNN-ReliefF, which integrates Ant Colony Optimization for global exploration of feature subsets, Graph Neural Networks for modeling complex, nonlinear relationships among clinical variables, and ReliefF for fine-grained discriminative ranking of features. Applied to the high-dimensional MIMIC-III intensive care dataset, the proposed method demonstrated substantial improvements in predictive performance over traditional approaches such as KNN-ReliefF and XGB-SHAP, achieving higher recall, precision ,F1-score, accuracy, and AUC-ROC. The results highlight the potential of combining swarm intelligence with deep relational learning for robust and interpretable feature selection in critical care analytics. Despite its superior performance, ACO-GNN-ReliefF incurs greater computational cost, which is a trade-off for its more comprehensive feature evaluation process.

In subsequent work, we intend to use early halting techniques and parallel processing to maximize the ACO-GNN-ReliefF framework's computational efficiency. Additionally, by including sequential models like LSTM-based GNNs, we want to expand this methodology to handle temporal changes in time-series ICU data. The suggested method's generalizability and clinical relevance will be established with additional validation on external datasets from various institutions and patient demographics. The interpretability and applicability of the chosen features for actual healthcare decision support systems may also be improved by including domain knowledge, such as ontologies or clinical guidelines, into the graph generation process.

REFERENCES

[1] E.W. Johnson, "MIMIC-III, a Freely Accessible Critical Care Database", *Scientific Data*, Vol. 3, pp. 1-7, 2016.

[2] D.F. Silva, V.M.A. Souza, G.E.A.P.A. Batista and J.P. Papa, "A Feature Selection Method based on Perturbation Theory", *Pattern Recognition Letters*, Vol. 77, pp. 360-377, 2018.

[3] X. Jia, L. Ma, H. Zhang, Y. Xie and B. Wang, "Feature Selection Methods for Classification of High-Dimensional Data: A Systematic Review", *Pattern Recognition Letters*, Vol. 108, pp. 1-6, 2020.

[4] J. Kim, H. Park and Y. Kang, "A Hybrid Feature Selection Method for Clinical Data Mining: An Empirical Study on Heart Disease Diagnosis", *Computers in Biology and Medicine*, Vol. 95, pp. 35-46, 2018.

[5] Y. Saeys, I. Inza and P. Larrañaga, "A Review of Feature Selection Techniques in Bioinformatics", *Bioinformatics*, Vol. 23, No. 19, pp. 2507-2517, 2007.

[6] Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, 2003.

[7] W.L. Hamilton, "*Graph Representation Learning*", 2020.

[8] M. Tan and Q.V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", *Proceedings of International Conference on Machine Learning*, pp. 6105-6114, 2019.

[9] Zhang, P. Song, Y. Huang, C. Swami and S. Chawla, "Deep Learning on Graphs: A Survey", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 1, pp. 249-270, 2022.

[10] T.N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks", *Proceedings of International Conference on Learning Representations*, pp. 1-14, 2017.

[11] I. Kononenko, "Estimating Attributes: Analysis and Extensions of RELIEF", *Proceedings of International Conference on Machine Learning*, pp. 171-182, 1994.

[12] D. Battiti, "Using Mutual Information for Selecting Features in Supervised Neural Net Learning", *IEEE Transactions on Neural Networks*, Vol. 5, No. 4, pp. 537-550, 1994.

[13] M.A. Hall, "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning", *Proceedings of International Conference on Machine Learning*, pp. 359-366, 2000.

[14] P. Pudil, J. Novovicova and J. Kittler, "Floating Search Methods in Feature Selection", *Pattern Recognition Letters*, Vol. 15, No. 11, pp. 1119-1125, 1994.

[15] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso", *Journal of the Royal Statistical Society*, Vol. 58, No. 1, pp. 267-288, 1996.

[16] L. Breiman, "Random Forests", *Machine Learning*, Vol. 45, pp. 5-32, 2001.

[17] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", *Proceedings of International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016.

[18] J.H. Holland, "*Adaptation in Natural and Artificial Systems*", University of Michigan Press, 1975.
[19] J. Kennedy and R. Eberhart, "Particle Swarm Optimization", *Proceedings of International Conference on Neural Networks*, pp. 1942-1948, 1995.
[20] M. Dorigo and T. Stützle, "*Ant Colony Optimization*", MIT Press, 2004.
[21] H. Xue, J. Qu and Q. Wang, "Feature Selection based on Ant Colony Optimization for Microarray Data Classification", *Computers in Biology and Medicine*, Vol. 43, No. 5, pp. 623-632, 2013.
[22] P. Singh, "Ant Colony Optimization-based Feature Selection for Brain MRI Images", *Expert Systems with Applications*, Vol. 42, No. 13, pp. 6122-6131, 2015.
[23] H. Choi, "Graph-based Neural Networks for Medical Diagnosis using Electronic Health Records", *AAAI*, 2020.
[24] M. Zitnik, M. Agrawal and J. Leskovec, "Modeling Polypharmacy Side Effects with Graph Convolutional Networks", *Bioinformatics*, Vol. 34, No. 13, pp. 457-466, 2018.
[25] S.M. Lundberg and S.I. Lee, "A Unified Approach to Interpreting Model Predictions", *Advances in Neural Information Processing Systems*, pp. 4765-4774, 2017.
[26] J. Lundberg, "Explainable Machine-Learning Predictions for the Prevention of Hypoxaemia during Surgery", *Nature Biomedical Engineering*, Vol. 2, pp. 749-760, 2018.