A Hybrid Strategy for Feature Selection in Classification Tasks

Combining Four Vector Intelligent Metaheuristic (FVIM) and Black-Winged Kite Algorithm (BKA)

Presenter: Thai Pham
Supervisor: Dr. Mohamed Tawhid

December 4, 2024

THOMPSON RIVERS UNIVERSITY | Faculty of Science

Outline

Introduction

Feature Selection (FS)

- Reduces dimensionality by selecting relevant features.
- Enhances interpretability and performance, especially in domains like bioinformatics and image recognition.

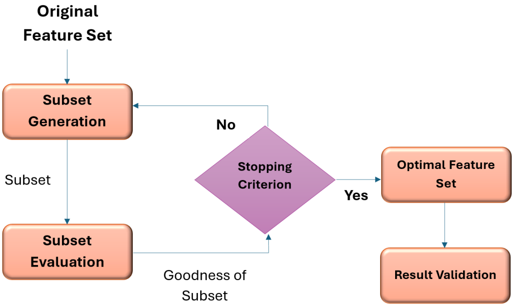


Figure: Overview of the Feature Selection Process

Introduction

Filter/Filter Approach

- Combines multiple filter methods to improve feature relevance and reduce redundancy.
- Enhances model accuracy and reduces overfitting in high-dimensional datasets.
- Uses a sequential filter process to select optimal features efficiently, without reliance on classifiers.

Introduction

Filter/Filter Approach

- Combines multiple filter methods to improve feature relevance and reduce redundancy.
- Enhances model accuracy and reduces overfitting in high-dimensional datasets.
- Uses a sequential filter process to select optimal features efficiently, without reliance on classifiers.

Hybrid Metaheuristic Algorithm

- Integrates strengths of different metaheuristics for better exploration and exploitation.
- Adapts to complex problem landscapes and avoids local optima.
- Recent studies show that hybrid metaheuristic strategies outperform single algorithms in feature selection..

Literature Review

Motivation

- High-dimensional data introduces noisy, redundant features.
- Traditional methods (GA, PSO) often struggle with high-dimensional data and complex feature interactions.
- Feature selection improves model accuracy and reduces overfitting.
- Hybrid methods can balance exploration (global search) and exploitation (local refinement).

Related Work

- **FVIM**: Strong in local search but limited exploration.
- **BKA**: Effective global search via migration but needs improved refinement.
- Combining FVIM and BKA leverages strengths of both for feature selection.

Four Vector Intelligence Metaheuristic (FVIM)

Key Features:

- Overcome premature convergence in traditional swarm algorithms
- Utilizes four guiding agents to improve solution search.
- Adaptive search strategy with dynamic parameters.

Mathematical Basis:

$$X_{t,i} = \begin{cases} P_{t,i} + (\alpha \times 2 \times r_1 - \alpha) \times \|r_2 \times P_{t,i} - P_i\|, & \text{if } r_3 < 0.5 \\ P_{t,i} - (\alpha \times 2 \times r_1 - \alpha) \times \|r_2 \times P_{t,i} - P_i\|, & \text{otherwise} \end{cases} \tag{3.1}$$

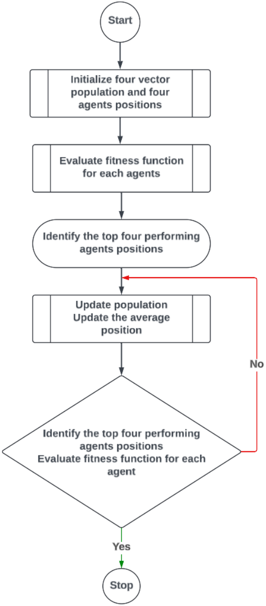
$$\bar{P}_i = \frac{X_{1,i} + X_{2,i} + X_{3,i} + X_{4,i}}{4} \tag{3.2}$$

Position Update Equation

$$\alpha = 1.5 - \left(\frac{4}{\text{Max\_iter}}\right)$$

Dynamic parameter

Flow chart of the FVIM algorithm



Methodology: **Standard Algorithms**

Black-Winged Kite Algorithm (BKA)

Figure: Hovering in the air, waiting for attack

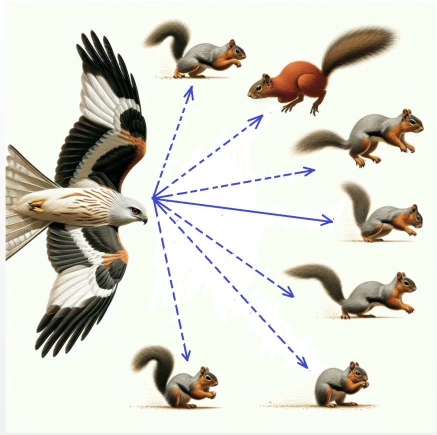


Figure: Hovering in the air, searching for prey

Black-Winged Kite Algorithm (BKA)

- **Biological Inspiration**: Models the adaptive hunting behavior of black-winged kites.
- **Cauchy Mutation Strategy**: Helps escape local optima, boosting performance in high-dimensional searches.
- **Leadership Strategy**: Guides search direction, enhances exploration.
- **Dynamic Search Capability**: Efficient in exploring new areas, adapting to changing targets.

Methodology: **Standard Algorithms**

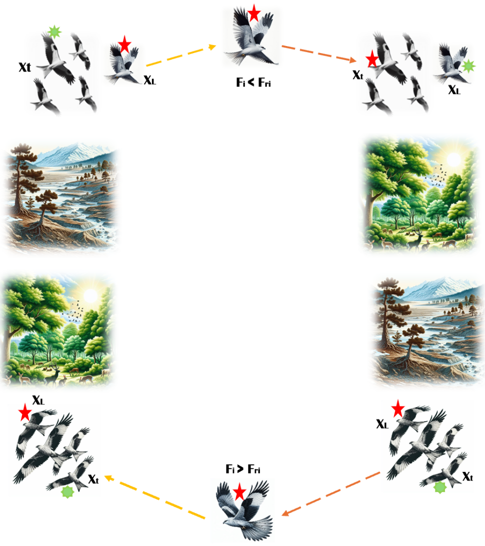


Figure: The strategic changes of Black-winged kites during migration

Black-Winged Kite Algorithm (BKA)

Mathematical Basis:

$$v_{i,j}^{t+1} = \begin{cases} v_{i,j}^{t} + n(1 + \sin(\text{Rand})) \times v_{i,j}^{t} & \text{if } p < r \\ v_{i,j}^{t} + n \times (2\text{Rand - } 1) \times v_{i,j}^{t} & \text{else} \end{cases}$$

$$v_{i,j}^{t+1} = \begin{cases} v_{i,j}^{t} + C(0,1) \times v_{i,j}^{t}, & \text{if } F_i < F_{ri} \\ v_{i,j}^{t} + C(0,1) \times (L_j^t - L_j^t - m \times v_{i,j}^{t}), & \text{else} \end{cases}$$

The parameter n and m are defined as:

$$n = 0.05 \times e^{-2\left(\frac{t}{T}\right)^2}$$

$$m = 2 \times \sin\left(\text{Rand} + \frac{\pi}{2}\right)$$

Methodology: **Proposed Hybrid Algorithm**

Feature Subset Evaluation Criteria

- **Mutual Information:** Measures the dependence between two variables, indicating how much information is shared.
- **Fisher Score:** Evaluates the discriminative power of features by comparing the ratio of between-class variance to within-class variance.
- **Chi-squared:** Assesses the independence of features by comparing observed and expected frequencies in categorical data.
- **ReliefF:** A filter-based algorithm that evaluates feature relevance by considering the nearest neighbors and their contributions to classification accuracy.

Methodology: **Proposed Hybrid Algorithm**

ReliefF

- Ranks features based on ability to distinguish between close instances of different classes.
- Updates feature weights based on nearest neighbor differences.
- MATLAB built-in function:
  `[weights, gains] = relieff(X, y, k)`
  The higher the weight, the more important the feature is for classification.

$$W[A] = W[A] - \frac{1}{m} \sum_{i=1}^{m} \left(\text{diff}(A, x_i, \text{nearestHit}(x_i)) - \sum_{c \neq y_i} \frac{P(c)}{1 - P(y_i)} \text{diff}(A, x_i, \text{nearestMiss}_c(x_i)) \right) \tag{3.9}$$

Fitness Function: Minimum Redundancy Maximum Relevance (mRMR)

- **mRMR function**: Identify a subset of features S that maximizes relevance to target class C while reducing redundancy.
- **mRMR formula**:

$$\text{mRMR} = \text{Rel} - \text{Red} \tag{1}$$

$$\text{Rel} = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; C) \tag{2}$$

$$\text{Red} = \frac{1}{|S|^2} \sum_{i \neq j} I(x_i; x_j) \tag{3}$$

where $I(X, Y)$ is the mutual information between variables X and Y.

Solution representation

- **Encoding Method:** Solutions are represented as vectors of integers, each corresponding to a feature.
- **Feature IDs:** For a dataset D with N features, each feature is assigned a unique ID from 1 to N.
- **Subset Selection:** Random vectors of feature IDs are generated, and the first P IDs in each vector form the selected feature subset.

Memory Buffer Integration

- **Purpose:** Enhances efficiency by serving as a temporary storage area for data processing.
- **Functionality:** Retains processed solutions and intermediate results to facilitate comparison.
- **Solution Validation:** Each new solution is validated against stored entries to assess novelty and eliminate duplicates.
- **Outcome:** Reduces redundant computations, allowing focus on exploring innovative and superior solutions.

Hybrid Algorithm Structure

- **Algorithm Name:** **Four Vector - Black winged Kite (FVBK)**
- **Purpose:** achieve a balanced trade-off between exploration and exploitation
- **Core Structure:**
  - Integrates the migration behavior of BKA in the FVIM framework
  - Embeds BKA's robust migration for exploration.
  - Utilizes FVIM's refined vector-based position updates.
- **Alternative Comparison:** **BKVP** (Black-winged Kite with Vector-based Position Update)

Hybrid Algorithm Structure

Pseudocode of FVBK Algorithm for Feature selection

Input: Dataset, MaxIter, PopSize, NFeat

Output: SelectedFeat (fstPos), BestFit (bstVal), ConvergenceCurve

Initialize

Initialize memory buffer with given size.

Initialize population *Pop* and positions (fstPos, sndPos, thrdPos, frthPos).

For each particle in the *Pop*:

Randomly select feature indices.

Calculate fitness of each particles using mRMR.

Store selected features in the memory buffer.

End For

Hybrid Algorithm Structure

Iterative Optimization Loop

While iter $\leq$ MaxIter:

For each particle in the *Pop*:

Apply boundary conditions.

Evaluate fitness using mRMR.

Update solution hierarchy (fstPos, sndPos, thrdPos, frthPos) using Algorithm 1

Store fitness and position history.

End For

Migration Behavior

For each particle in the *Pop*:

Generate a new position using equation 3.4.

Check new position's fitness using mRMR.

Update memory buffer if the new position is unique.

Store fitness and position history.

End For

Hybrid Algorithm Structure

***Update Positions*:**

For each particle in the *Pop*:

Adjust particle positions using equation 3.1

Update the average position using equation 3.2

End For

Identify the best-performing particles based on fstPos, sndPos, thrdPos, frthPos

Update memory buffer

Update the best value and convergence curve.

End While and return SelectedFeat, BestFit, ConvergenceCurve

Methodology: **Proposed Hybrid Algorithm**

Framework for Feature Selection

- **Stage 1: Filtering Phase**
  - Apply statistical metrics: **Chi-squared, ReliefF, Fisher score,** and **Mutual Information**.
  - Objective: Identify and retain **elite features** with high relevance to the target variable.
- **Stage 2: FVBK Optimization**
  - Refine feature selection using the **FVBK algorithm**.
  - Optimize with **mRMR criterion** to select features with high relevance and minimal redundancy.
- **Evaluation:** The performance is evaluated using five classifiers: SVM, LDA, Naive Bayes (NB), k-Nearest Neighbors (kNN), and CART.

Methodology: **Proposed Hybrid Algorithm**

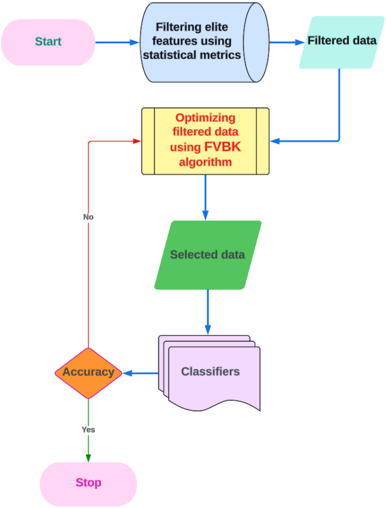


Figure: Flowchart illustrating the process for feature selection

Dataset Overview

Dataset	Type	# Features	# Instances	# Classes
GLI_85 (GLI)	Biological	22283	85	2
Colon Cancer (CLN)	Biological	2000	62	2
DBWorld e-mails (DBE)	Text	4702	64	2
Dexter (DEX)	Text	20000	300	2
Lymphoma (LYM)	Biological	4026	96	9
NCI9 (NCI)	Biological	9712	60	9
Orlraws10P (ORP)	Image	10304	100	10
SMK_CAN_187 (SMK)	Biological	19993	187	2
TOX_171 (TOX)	Biological	5748	171	4

Parameter Settings

Dataset	LDA Type	SVM Kernel	NB Distribution
GLI	Linear	Linear	Normal
CLN	Diaglinear	Linear	Multivariate Multinomial
DBE	Diaglinear	Linear	Multivariate Multinomial
DEX	Diaglinear	Gaussian	Kernel Density Estimate
LYM	Linear	Linear	Multivariate Multinomial
NCI	Diaglinear	Linear	Multivariate Multinomial
ORP	Linear	Linear	Normal
SMK	Linear	Linear	Kernel Density Estimate
TOX	Linear	Linear	Normal

Experiments and Discussion: **Statistical Analysis**

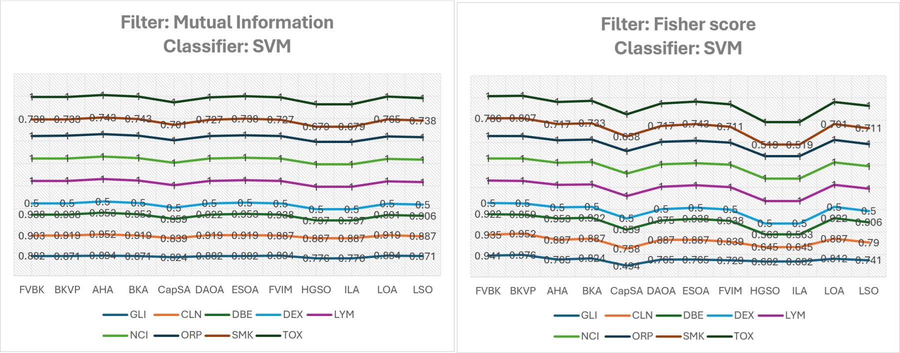
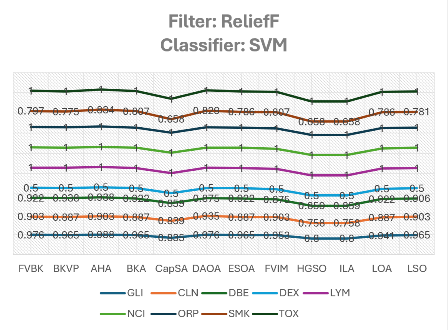


Figure: Accuracy by filters for the **SVM** classifier

Experiments and Discussion: **Statistical Analysis**



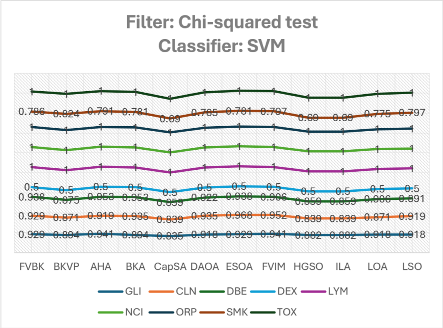
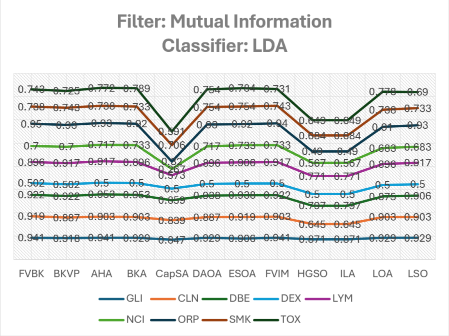


Figure: Accuracy by filters for the **SVM** classifier

Experiments and Discussion: **Statistical Analysis**



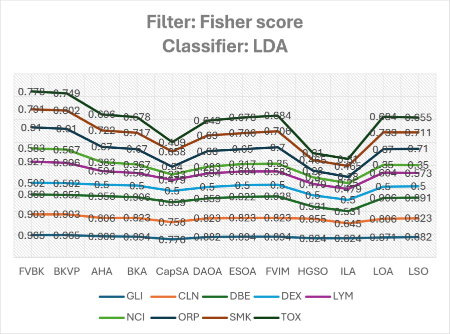


Figure: Accuracy by filters for the **LDA** classifier

Experiments and Discussion: **Statistical Analysis**

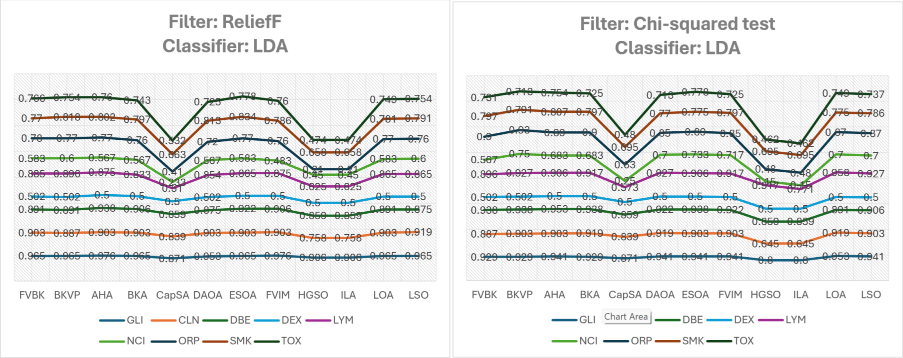
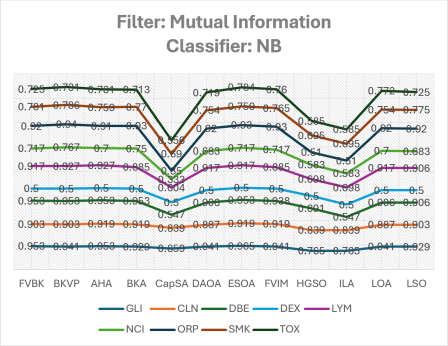


Figure: Accuracy by filters for the **LDA** classifier

Experiments and Discussion: **Statistical Analysis**



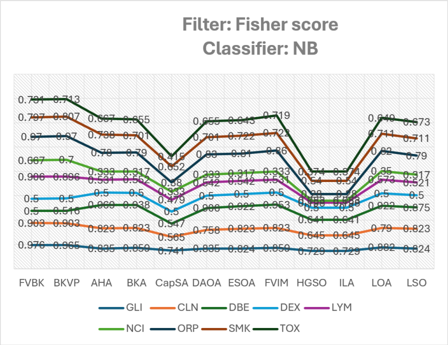
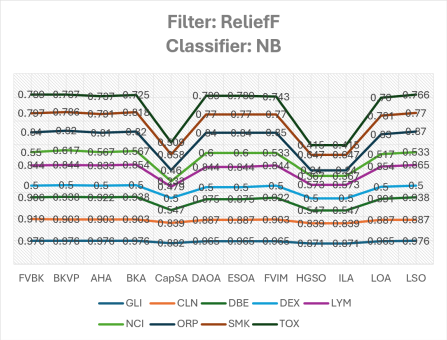


Figure: Accuracy by filters for the **NB** classifier

Experiments and Discussion: **Statistical Analysis**



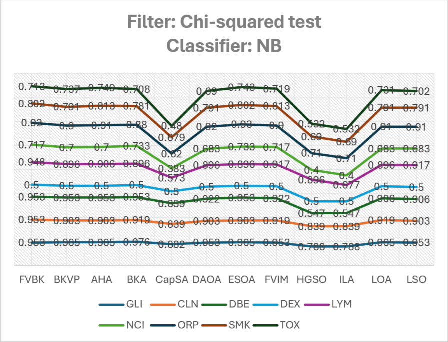


Figure: Accuracy by filters for the **NB** classifier

Experiments and Discussion: **Statistical Analysis**

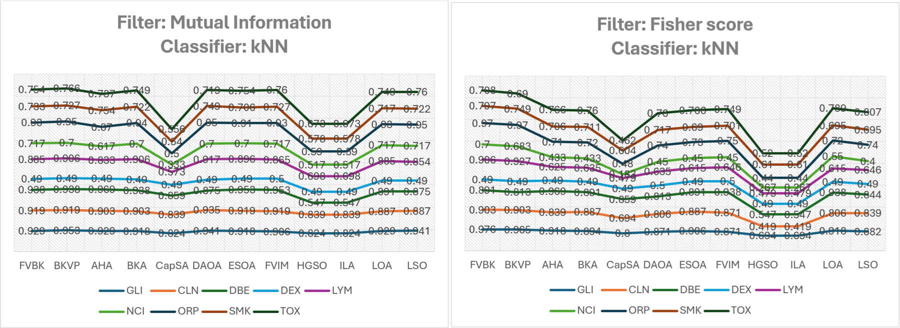


Figure: Accuracy by filters for the **kNN** classifier

Experiments and Discussion: **Statistical Analysis**

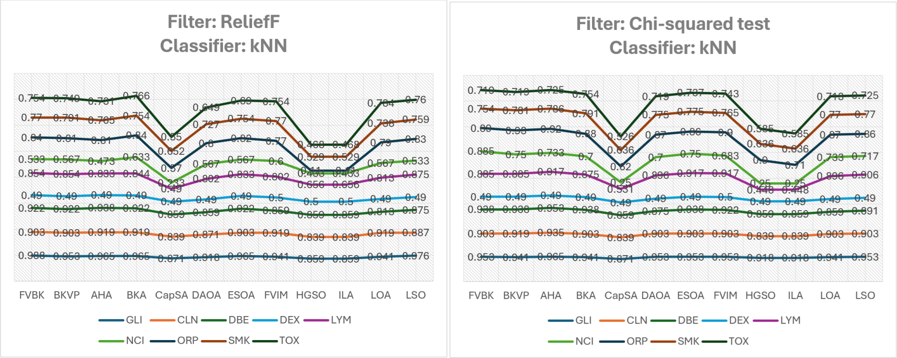
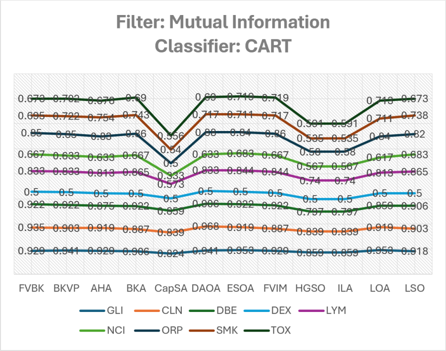


Figure: Accuracy by filters for the **kNN** classifier

Experiments and Discussion: **Statistical Analysis**



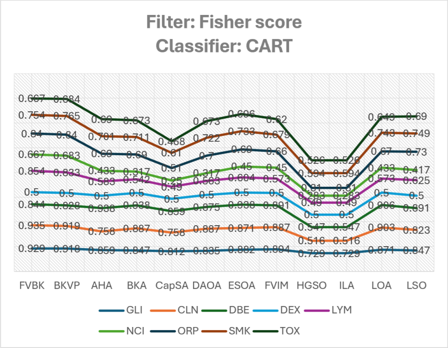


Figure: Accuracy by filters for the **CART** classifier

Experiments and Discussion: **Statistical Analysis**

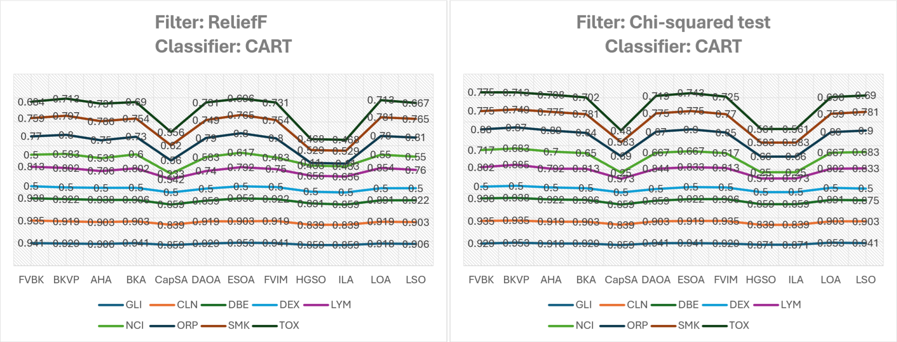


Figure: Accuracy by filters for the **CART** classifier

Experiments and Discussion: **Statistical Analysis**

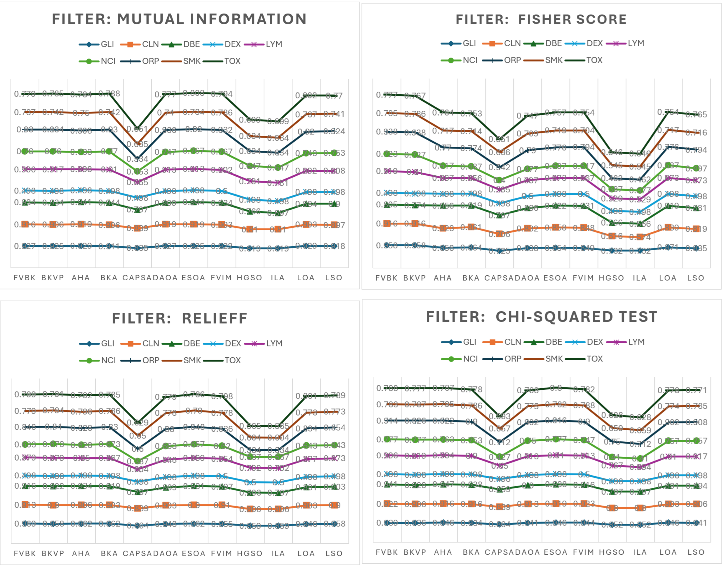


Figure: Overall accuracy by filters

Wilcoxon Signed-Rank Test

- Purpose: Non-parametric test to assess if there's a significant difference in paired classification accuracy between FVBK and other algorithms.
- Findings:
  - FVBK **outperforms FVIM** with the **ReliefF** filter.
  - FVBK shows **superior performance over BKA** with **ReliefF** and **Chi-squared filters**.

Friedman Test

- Purpose: Non-parametric test to rank algorithms across multiple datasets.
- Outcome: Consistent ranking indicating FVBK's superior performance in various feature selection tasks.

Table 4.11: Wilcoxon signed-rank test between FVBK and FVIM

Filter	**p_value**	**Signed Rank Statistics**	**Winner**	**Losser**
Mutual Information	0.9453125	19	4	4
Fisher score	0.07421875	7	7	2
ReliefF	**0.01953125**	**3**	**8**	**1**
Chi-squared test	0.49609375	16	4	5

Table 4.12: Wilcoxon signed-rank test between FVBK and BKA

Filter	**p_value**	**Signed Rank Statistics**	**Winner**	**Losser**
Mutual Information	0.7421875	21	3	5
Fisher score	0.0546875	6	8	1
ReliefF	**0.02578125**	**15**	**7**	**2**
Chi-squared test	**0.01953125**	**3**	**8**	**1**

Figure: Wilcoxon signed-rank test

Table 4.13: Ranked by Friedman test based on the Best Fitness Value

Rank	Engineering problem	CEC2013	CEC2017	CEC2006
1	AHA	CapSA	AHA	ESOA
2	CapSA	AHA	CapSA	CapSA
3	ILA	LSO	**FVIM**	AHA
4	**BKA**	LOA	ESOA	**FVIM**
5	**FVIM**	ESOA	**BKA**	**BKA**
6	ESOA	**FVIM**	LOA	LOA
7	HGSO	HGSO	HGSO	HGSO
8	LSO	**BKA**	LSO	LSO
9	DAOA	ILA	ILA	DAOA
10	LOA	DAOA	DAOA	ILA
p-value	1.69E-15	2.91E-07	3.02E-45	1.68E-30

Table 4.14: Ranked algorithms by Friedman test

Rank	**SVM**	**LDA**	**NB**	**kNN**	**CART**	**Overall accuracy**
1	AHA	**FVBK**	**FVBK**	**FVBK**	**FVBK**	**FVBK**
2	**FVBK**	**BKVP**	**BKVP**	**BKVP**	**BKVP**	**BKVP**
3	FVIM	AHA	FVIM	FVIM	ESOA	AHA
4	ESOA	FVIM	AHA	AHA	FVIM	ESOA
5	BKA	ESOA	ESOA	BKA	AHA	FVIM
6	DAOA	LOA	BKA	ESOA	LSO	BKA
7	**BKVP**	BKA	LOA	LSO	BKA	LOA
8	LOA	LSO	LSO	DAOA	DAOA	DAOA
9	LSO	DAOA	DAOA	LOA	LOA	LSO
10	CSA	HGSO	CSA	HGSO	CSA	CSA
11	HGSO	CSA	HGSO	ILA	HGSO	HGSO
12	ILA	ILA	ILA	CSA	ILA	ILA
p-value	**8.45E-02**	**3.81E-08**	**5.63E-08**	**1.29E-07**	**1.71E-07**	**4.21E-09**

Results achieved

- **Hybrid Design:** Combines FVIM's position updates with BKA's migration for effective feature selection.
- **Exploration-Exploitation Balance:** Enhances search depth and diversity, benefiting high-dimensional datasets.
- **Redundancy Reduction:** Prioritizes relevant features, minimizing redundancy for quality subsets.
- **Performance:** Achieves superior accuracy across diverse datasets, outperforming baselines.
- **Filter Effectiveness:** ReliefF filter optimally boosts FVBK results.

Future Research Directions

- **Enhanced Filtering Algorithms:** Develop more efficient filters to improve initial dimensionality reduction.
- **Optimized Subset Size:** Focus on methods for automatic subset size determination and adaptive parameter settings.
- **Advanced Memory Structures:** Explore multi-layered memory or reinforcement-based retention for dynamic problem spaces.
- **Alternative Objective Functions:** Test new fitness functions, like correlation-based measures, to improve feature selection quality.

Thank you!