

# Fuzzy C-Means com Método Wrapper Com Baixo Custo Computacional de Seleção de Atributos\*

Gabriel Marcondes dos Santos, Emmanuel Tavares F. Affonso, Alisson Marques da Silva e Gray Farias Moita  
Centro Federal de Educação Tecnológica de Minas Gerais - CEFET-MG

marcondescefet@gmail.com, emmanuelcomp@gmail.com, alisson@cefetmg.br, gray@cefetmg.br

*Abstract*—Nowadays the Computational Intelligence (IC) algorithms have shown a lot of efficiency in pattern classification and recognition processes. However, some databases may contain irrelevant attributes that may be detrimental to the learning of the classification model. In order to detect and exclude input attributes with little representativeness in the data sets presented to the classification algorithms, the Features Selection (FS) methods are commonly used. The goal of features selection methods is to minimize the number of input attributes processed by a classifier in order to improve its assertiveness. In this way, this work aims to analyze solutions to classification problems with three different classification algorithms. The first approach used for classification is the unsupervised Fuzzy C-Means (FCM) algorithm, the second approach is a supervised version of FCM and the third approach is a variation of supervised FCM with features selection. The method of features selection incorporated in FCM is called the Mean Ratio Feature Selection (MRFS), and was developed with the objective of being a method with low computational cost, without need for complex mathematical equations and can be easily incorporated into any classifier. For the experiments, the three versions of the unsupervised FCM, supervised FCM and FCM with attribute selection were performed with the aim of verifying whether there would be a significant improvement between the variations of the FCM. The results of the experiments showed that FCM with MRFS is promising, with results superior to the original algorithm and also to its supervised version.

*Index Terms*—Features Selection, Computational Intelligence, FCM, Wrapper Method, Low Cost Computational.

*Resumo*—Nos dias atuais os algoritmos de Inteligência Computacional (IC) tem apresentado bastante eficiência em processos de classificação e reconhecimento de padrões. Porém, algumas bases de dados podem conter atributos irrelevantes que podem proporcionar um detrimento no aprendizado do modelo de classificação. Para detectar e excluir atributos de entrada com pouca representatividade nos conjuntos de dados apresentados aos algoritmos de classificação, é comumente utilizado os métodos de seleção de atributos (FS, do inglês *Features Selection*). O objetivo dos métodos de seleção de atributos é minimizar a quantidade de atributos de entrada processados por um classificador de modo a melhorar sua assertividade. Desta maneira, este trabalho visa analisar soluções para problemas de classificação com três diferentes algoritmos de classificação. A primeira abordagem utilizada para classificação é o algoritmo não supervisionado *Fuzzy C-Means* (FCM), a segunda abordagem é uma versão do FCM supervisionada e a terceira abordagem é uma variação do FCM supervisionada com seleção de atributos. O método de seleção de atributos incorporado no FCM é chamado de razão das médias para seleção de atributos (MRFS, do inglês *Mean Ratio Feature Selection*), e foi desenvolvido com o objetivo de ser um método com baixo custo computacional, sem necessidade de equações matemáticas complexas e poder ser facilmente incorporado em qualquer classificador. Para os experimentos,

o FCM e suas duas novas variações propostas tiveram seus resultados de classificações comparadas em diferentes bases de dados, o objetivo foi avaliar se haveria uma melhora significativa entre as variações do FCM. Os resultados dos experimentos apontaram que o FCM com MRFS é o modelo mais promissor, com resultados superiores ao algoritmo original e também a sua versão supervisionada.

*Index Terms*—Seleção de Atributos, Inteligência Computacional, FCM, Método Wrapper, Baixo Custo Computacional.

## I. INTRODUÇÃO

Algoritmos de Inteligência Computacional (IC), são frequentemente utilizados para resolução de problemas de classificação e reconhecimento de padrões. Os algoritmos de classificação são desenvolvidos com o objetivo de rotular os dados processados em diferentes classes. Uma característica comum entre os algoritmos de classificação é utilizar uma base de dados para gerar seu conhecimento e possibilitar que novos dados sejam classificados entre as classes. Os modelos de classificação podem apresentar três diferentes maneiras de aprendizagem: aprendizagem supervisionada, aprendizagem não supervisionada e aprendizado por reforço [3].

O aprendizado supervisionado ocorre quando as amostras dos dados são apresentadas juntamente com suas respectivas classes, para o treinamento e formação do conhecimento no algoritmo de classificação. O aprendizado não supervisionado ocorre quando os conjuntos de amostras são apresentados sem as respectivas saídas e o algoritmo de classificação busca descobrir melhores representações e padrões dos dados apenas com os atributos de entrada. No aprendizado por reforço o algoritmo tem diferentes ações para decidir entre diferentes dados apresentados, seu aprendizado ocorre de forma que ao tomar uma decisão errada ou certa, o algoritmo armazena aquele aprendizado para repetir ou não essa ação quando um dado igual for apresentado. De modo geral, o algoritmo com característica de aprendizado por reforço tem seu aprendizado na tentativa e erro, até que desenvolva um conhecimento e comece a executar sua tarefa corretamente.

Os algoritmos de classificação podem encontrar alguns fatores no processamento dos conjuntos de dados que dificultam a classificação. Estes fatores podem estar relacionados a entradas com pouca representatividade, atributos altamente correlacionados, ruidosos e até conjuntos de amostra com classes de saídas desbalanceadas. Todos esses fatores podem ser prejudiciais para um algoritmo de classificação, que consequentemente, diminui a capacidade de gerar bons resultados e

pode até aumentar o tempo de resposta do seu processamento. Desta maneira, métodos de seleção de atributos (FS, do inglês *Features Selection*) são comumente utilizados com o propósito de reduzir o número de atributos não representativos a serem processados [10]. O objetivo dos métodos de seleção de atributos, é isolar um conjunto de atributos de maior representatividade e identificar de modo automatizado o relacionamento entre diferentes conjuntos de dados, proporcionando aos classificadores conjuntos de atributos mais representativos.

O *Fuzzy C-Means* (FCM) é um algoritmo de agrupamento não supervisionado que visa fazer uma análise dos dados com o objetivo de agrupar diferentes conjuntos de dados que apresentem alguma similaridade entre eles. O FCM funciona atribuindo um grau de pertinência de cada amostra de dados a um centro de grupo conhecido, quanto mais próximo a amostra estiver do centro, maior será seu grau de pertinência em relação a este centro. Consequentemente, maior será as chances da amostra pertencer a um grupo conhecido [1], [5].

Neste trabalho são introduzidas duas novas abordagens para o FCM. Inicialmente, a primeira abordagem, introduz uma nova versão do FCM com treinamento supervisionado para tarefas de classificação. Em seguida, na segunda abordagem, é incorporado ao FCM supervisionado um método *wrapper* de seleção de atributos de entrada. O método de seleção de atributos é baseado na razão das médias (MRFS, do inglês *Mean Ratio for Feature Selection*), um método de baixo custo computacional.

As seções seguintes, estão organizadas como se segue: na Seção II são apresentados os conceitos fundamentais dos métodos de seleção de atributos e do algoritmo de classificação FCM; a Seção III detalha as abordagens propostas; a Seção IV ilustra a metodologia e os experimentos, bem como, os resultados obtidos; por fim, a Seção V discorre com as considerações finais.

## II. FUNDAMENTAÇÃO TEÓRICA

Nesta seção é apresentada uma revisão sobre o modelo FCM e métodos de seleção de atributos que são bases utilizadas neste trabalho para problemas de classificação, procurando apresentar seus conceitos gerais e também as diferentes metodologias que podem ser encontradas na literatura. Na primeira subseção são apresentados diferentes conceitos e métodos para seleção de atributos (FS, do inglês *Feature Selection*). Na segunda subseção são apresentados os do algoritmo de agrupamento *Fuzzy C-Means* (FCM).

### A. Métodos de seleção de atributos

Os métodos de seleção de atributos visam reduzir o espaço de atributos de entrada em um menor e mais representativo conjunto de atributos a ser processado. O objetivo é melhorar o desempenho e a capacidade de classificação, tornando os algoritmos mais rápidos e com uma melhor assertividade no processamento dos dados [6].

Nos métodos de seleção de atributos, o processo para avaliar quais atributos apresentam maior representatividade a um classificador são baseados em equações estatísticas ou

heurísticos. Estes métodos, sejam heurísticos ou estatísticos, podem ser representados como *wrapper* (empacotamento), *filter* (filtro) e *embedded* (embarcado):

- **Wrapper:** nos métodos *wrapper* o algoritmo de classificação é executado com diferentes subconjuntos de entradas, onde a seleção é feita comparando diferentes subconjuntos de dados durante o treinamento, de maneira a encontrar um subconjunto que apresenta o melhor desempenho.
- **Filter:** nos métodos *filter*, a seleção é realizada em um pré-processamento ao algoritmo de IC, o método executa um processamento que faz um ranqueamento das entradas de acordo com sua representatividade, a seleção de quantos atributos serão excluídos é realizada pelo usuário.
- **Embedded:** métodos *embedded*, o processo de seleção de atributos está embutido no modelo de classificação, neste método o classificador e o processo de seleção de atributos não podem ser separados. A seleção de atributos no método *embedded* ocorre junto ao classificador.

Na literatura podemos encontrar diferentes aplicações da seleção de atributos utilizando os métodos citados anteriormente. Na proposta de [7], um método *wrapper* foi desenvolvido utilizando um modelo heurístico, o *Opposition-based Social Spider Optimization* (OBSSO), sendo implementado com os algoritmos de classificação *Random Forest* (RF) e no *K-Nearest Neighbors* (KNN).

Em [16] foi proposto um *software* que permite a utilização de onze métodos *filter* de seleção de atributos de entrada, que podem ser selecionados e aplicados nos classificadores KNN, *Decision Tree* (DT), *Naïve Bayes* (NB), RF e *Support Vector Machine* (SVM). Neste *software* é possível avaliar os conjuntos de atributos de entrada selecionados por cada método e seu impacto nos diferentes algoritmos de classificação.

Foi proposto em [15] um método *embedded* utilizando o *Multispectral Image Neighborhood* (MPN) para classificação multiespectral de sensoriamento remoto com um Sistema *Hardware/Software Codesign* (HSC), visando reduzir o tempo de processamento e a carga computacional do algoritmo.

### B. Fuzzy C-Means (FCM)

O FCM é um algoritmo de agrupamento não supervisionado proposto por [5] e estendido por [1]. Diferente dos modelos de agrupamento onde uma amostra pertence ou não a um grupo, no FCM uma amostra pode pertencer a vários grupos dado um grau de pertinência. Este grau de pertinência é representado por um intervalo entre  $[0,1]$ , de modo que quanto mais próximo de 1, maior a similaridade a um grupo. No cálculo de pertinência de uma amostra para todos os grupos conhecidos, a soma dos graus de pertinência é sempre igual a 1.

O FCM tem como valores de entrada um conjunto  $X$  de dados, uma quantidade pré-fixada de  $k$  grupos que serão representados pelos seus respectivos centroides e um parâmetro de fuzificação  $m$ . Em sua inicialização, o FCM gera uma matriz de graus de pertinências aleatórios, com o objetivo de criar a

posição inicial dos  $c$  centroides que serão atualizados posteriormente. A matriz de graus de pertinências é representada por  $\mu_{lj}$ , onde  $l$  é o índice de cada amostra do conjunto de dados  $X$  e  $j$  representa o índice do  $c$ -ésimo centroide. Posteriormente, os centroides são atualizados com base na matriz de graus de pertinência  $\mu_{lj}$ , representado por

$$c_j = \frac{\sum_{l=1}^l (\mu_{lj})^m x_l}{\sum_{l=1}^l (\mu_{lj})^m}, \quad (1)$$

em que para cada centroide  $c$  indexado por  $j$ , há um vetor  $\mu_l$  com o grau de pertinência que tem como somatório o valor de 1 para todos  $j$ . O  $l$  indexa cada amostra do conjunto de dados  $X$  e  $m$  é um parâmetro de fuzificação, que geralmente é utilizado um valor entre 1.25 e 2 [2].

Com os valores de centroide atualizados, um novo cálculo da similaridade das amostras do conjunto  $X$  é realizado. Este cálculo é responsável por gerar uma nova matriz de graus de pertinência, que possui valores diferentes da inicialização do FCM quando essa matriz foi gerada aleatoriamente. Para calcular a similaridade, é utilizado a distância Euclidiana (2) como um valor de dissimilaridade entre o ponto da amostra e os centroides  $c$  de cada grupo. A matriz de graus de pertinências pode ser calculada por Eq. (3).

$$d(x; c) = \sqrt{\sum_{i=1}^r |x_i - c_i|^2} \quad (2)$$

$$\mu_j = \frac{\left(\frac{1}{d(x_l; c_j)}\right)^{\frac{1}{1-m}}}{\sum_{j=1}^k \left(\frac{1}{d(x_l; c_j)}\right)^{\frac{1}{1-m}}} \quad (3)$$

em (2) a distância Euclidiana  $d$  entre a amostra  $x$  e o centroide  $c$  indexados pelos atributos de entrada  $i$  de cada amostra. O vetor  $\mu_j$  (3) é indexado por  $j$  que representa a similaridade da nova amostra  $x_l$  para o  $j$ -ésimo centroide,  $k$  é o total de centroides e  $l$  representa índice da amostra do conjunto total  $X$  processada. A matriz  $u_{lj}$  é formada após o processamento de todas amostras do conjunto  $X$ .

Em seguida, com os novos valores de similaridades calculados em  $\mu_{lj}$ , o FCM atualiza as novas posições dos centroides com a Eq. (1). Este ciclo com as repetições das equações (1) e (3) são executadas até que um número de iterações seja alcançado ou quando os centroides não sofrem mais mudanças com as novas iterações.

É possível encontrar diferentes aplicações do FCM na literatura, em [17] é proposto um modelo para segmentação de imagem utilizando *Distance Regularized Level Set Evolution* (DRLSE). O DRLSE realiza a segmentação de imagens com base em uma técnica de reconstrução morfológica dos dados baseada no FCM. O objetivo da reconstrução morfológica utilizando o FCM é reduzir ruídos de uma imagem a ser processada. Já em [11] foi proposto um sistema automatizado de detecção de tumor cerebral com base no FCM, neste trabalho o FCM é inicialmente utilizado como uma técnica de filtro médio para remover o ruído das imagens. Após a remoção do ruído das imagens são aplicados os algoritmos

CNN, KNN e *Backpropagation Neural Network* (BPNN), para classificação e avaliação dos resultados.

Em [4] o FCM foi utilizado para classificar empresas conforme as mudanças financeiras em comparação ao ano anterior. Os dados históricos de 166 empresas foram processados pelo FCM e, posteriormente, os resultados produzidos pelo FCM são processados por uma técnica chamada *Fuzzy Financial Health Scores* (F-FHSs). O F-FHSs tem o objetivo de ranquear a eficiência operacional e lucratividade de cada empresa. Em [18] o FCM foi implementado para avaliar o estado operacional de caixas de controle de visão. Para a avaliação desses equipamentos, foi criada uma tabela com valores para classificação de seu estado. Dessa forma, o FCM é utilizado para agrupar os dados e reduzir atributos irrelevantes, produzindo uma tabela com as saídas, possibilitando uma melhor avaliação dos estados dos equipamentos.

### III. MODELOS PROPOSTOS

Nesta seção, será apresentado os diferentes algoritmos de classificação propostos neste trabalho. Inicialmente, será apresentado uma nova variação do FCM, com um método supervisionado para o agrupamento dos dados, onde o número de grupos são gerados a partir das diferentes classes de saídas apresentadas na etapa de treinamento e com uma melhor escolha do posicionamento dos centroides. Por fim, será apresentado uma nova variação supervisionada do FCM com seleção de atributos de entrada (FS, do inglês *Feature Selection*), utilizando o método de razão das médias para seleção de atributos (MRFS, do inglês *Mean Ratio Feature Selection*). O MRFS é um método de fácil agregação como *wrapper* em algoritmos de classificação e com boa relevância no advento de seleção de atributos de entrada.

#### A. FCM Supervisionado

O método supervisionado proposto para utilização com FCM, é um modelo de classificação que necessita de um treinamento prévio para inicializar seus parâmetros e faz o armazenamento da melhor posição dos centroides durante as iterações do algoritmo. Nesta variação do FCM, também não é necessário que seja informado a quantidade de centroide.

Inicialmente, com os dados de treinamento o algoritmo avalia quantas classes distintas de saídas há no conjunto *Xtreino*, no qual serão utilizados para geração dos centroides associados a cada diferente classe de saída. Neste modelo, a posição inicial dos centroides não é gerada a partir de uma matriz de graus de pertinência criada com valores aleatórios. A geração inicial de cada centroide é realizada buscando os valores de uma amostra de cada classe de saída e tem seus valores iniciais centrados nessa amostra. Posteriormente uma cópia dos centroides é realizada no denominado *bcentroides*, que é responsável por armazenar as posições que tiveram melhor acurácia no treinamento.

Com a quantidade de grupos já definidas e centroides inicializados, o processo de treinamento e atualização a cada iteração do algoritmo é realizado, seguindo as equações (1) e (3). Porém, antes da conclusão de cada ciclo de iteração, o

modelo proposto utiliza a matriz de graus de pertinência  $\mu_{ij}$  para classificar quantas amostras são avaliadas corretamente naquela posição atual dos centroides.

A diferença mais significativa entre o modelo não supervisionado e supervisionado é que em cada iteração, quando se é calculado os graus de pertinência das amostras de treinamento, Eq. (3), o modelo supervisionado avalia a quantidade de amostras classificadas corretamente do conjunto de dados de treinamento na posição atual dos centroides. Se os centroides oferecerem uma maior quantidade de amostras classificadas corretamente que os  $bcentroids$  já conhecidos,  $bcentroids$  recebe os valores dos centroides, conforme apresentado em (4). Posteriormente, é atualizado os novos valores centroides, Eq. (1), e este ciclo é repetido até que as condições de parada do algoritmo sejam satisfeitas.

$$bcentroids = \begin{cases} centroide, & \text{se } centroide \text{ é melhor} \\ bcentroids, & \text{se } bcentroids \text{ é melhor} \end{cases} \quad (4)$$

No final, os  $bcentroids$  serão os valores que ofereceram uma melhor classificação entre todas as iterações do algoritmo para aquele conjunto de dados de treinamento. Esses valores são os mesmos que serão utilizados para predição do algoritmo para novas e desconhecidas amostras apresentadas. No pseudocódigo em (1) é apresentado as etapas de execução do modelo supervisionado proposto.

### B. FCM com Seleção de Atributos

Para uma melhor compreensão, esta seção foi dividida em duas partes, onde na primeira será apresentado os conceitos e fundamentação do método de seleção de atributos, o *Mean Ratio for Feature Selection* (MRFS); na segunda parte, será apresentado como o método MRFS foi incorporado no FCM, nomeado de FCM-MRFS.

1) *MRFS*: O MRFS é um método de seleção de atributos baseado na razão das médias associadas a cada classe de saída e utiliza apenas equações matemáticas básicas, tem bom desempenho e boa capacidade de seleção de atributos para os classificadores. O MRFS, foi inicialmente proposto como *filter* em [13], mas neste trabalho ele foi modificado e empregado como *wrapper*.

A execução do MRFS, além de possuir apenas equações matemáticas simples, pode ser concluída com somente 3 etapas, que são: (i) cálculo da média amostral de todos os atributos de acordo com suas respectivas saídas; (ii) cálculo da razão entre as médias de cada atributo para cada classe e; (iii) ranqueamento dos atributos do menor valor da razão das médias para o maior valor. No MRFS, sugere-se que os atributos com os menores valores da razão entre suas médias nas classes de saída são os mais relevantes para a classificação. Em outras palavras, os atributos que possuem valores mais dispersos entre as classes são os mais representativos para os algoritmos de classificação. A seguir, as 3 etapas do MRFS são detalhadas.

- (i) **Cálculo da Média Amostral**: inicialmente, para calcular a média amostral de cada atributo, o MRFS

---

### Algoritmo 1: Pseudocódigo para FCM Supervisionado.

---

```

Entradas: Xtreino, Ytreino, niteracao;
Saídas: bcentroids;
Define quantidade de classes tem em Ytreino;
Inicializa centroide com uma amostra de cada classe;
Copia centroide em bcentroids;
Define acurácia de bcentroids como zero;
iteracao = 0;
teste = Verdadeiro;
// Treinamento e atualização dos
  centroides
repita
  Classifica Xtreino usando os centroides - Eq. (3);
  Calcula a acurácia para  $u_{ij}$  de Xtreino;
  se Acurácia com bcentroids < Acurácia com
    centroide então
    | Copia centroide em bcentroids;
    | Armazena a acurácia dos dados com
    | bcentroids;
  fim
  Atualiza posição dos centroides - Eq. (1);
  Calcula o erro;
  se (error > 0) and (iteracao < niteracao) então
  | Teste = Falso;
  fim
  senão
  | iteracao = iteracao + 1;
  fim
até teste;
Define bcentroids como melhor;

```

---

identifica o número de classes e o número de amostras de cada classe no conjunto de dados. Em seguida, para cada amostra da classe  $j$ , é calculado o somatório do valor do atributo  $i$ . Por fim, calcula-se a média do valor de cada atributo  $i$  em cada classe  $j$  dividindo o valor do somatório pelo número de amostras da respectiva classe, conforme Eq. 5:

$$S_{ij} = \sum_{l=1}^{n_k} (\bar{X}_{ji} \in k_j) \quad (5)$$

na qual  $n_k$  é o número de amostras de cada classe,  $\bar{X}$  é a média do atributo  $i$  na classe  $j$  e  $k$  são as classes indexadas por  $j$ .

- (ii) **Cálculo da Razão das Médias**: a razão é uma medida que expressa uma fração ou porcentagem entre dois valores e é utilizada para comparar duas quantidades. Nesta etapa é calculada a razão  $R_i$  das médias do atributo  $i$  em cada classe  $j$ . A Eq. 6 ilustra o cálculo da razão das médias para duas ou mais classes.

$$R_i = \sum_{j=2}^k \left( \min \left( \frac{S_{ij}}{S_{i(j-1)}}, \frac{S_{i(j-1)}}{S_{ij}} \right) \right), \quad (6)$$

na qual  $k$  é o número de classes e  $R$  é a consequência direta do teorema do valor médio [9], [12]. Este teorema é dado pela densidade de uma função de probabilidade contínua para os intervalos médios de cada classe  $j$  de cada atributo  $i$ , com base nos intervalos entre o valor mínimo da razão da média  $j$  por  $j - 1$  ou  $j - 1$  por  $j$ .

- **(iii) Ranqueamento dos Atributos:** no contexto do MRFS, os atributos que possuem médias mais dispersas são mais relevantes para os classificadores. Portanto, os atributos são ranqueados em ordem crescente do valor da razão das médias  $R$ , considerando que os que possuem menores valores são os mais relevantes.

2) *FCM-MRFS*: O modelo *wrapper* proposto, incorporando o MRFS no algoritmo de classificação FCM. Foi desenvolvido utilizando a eliminação para trás, na qual o classificador é iniciado com todos os atributos e os atributos considerados irrelevantes são excluídos durante o processo de seleção.

O processamento do FCM-MRFS pode ser concluído em 7 etapas, onde as 3 primeiras etapas são idênticas às etapas apresentadas anteriormente e por isso, não serão detalhadas novamente. As etapas 4-7 são detalhadas a seguir.

- **(iv) Criação dos Modelos:** conforme a ordenação proveniente da razão das médias, Etapa (iii), são criados dois modelos: (M) modelo atual com o número atual de atributos, e que, na primeira iteração do treinamento do classificador, inicia o modelo atual com o número total de atributos; (M-1) modelo candidato que é construído excluindo o atributo menos relevante do conjunto atual.
- **(v) Execução dos Modelos e treinamento:** nesta etapa, os modelos (candidato e atual) são executados juntamente ao treinamento do FCM supervisionado, descrito na Seção III-A. A ideia é verificar se vale a pena substituir o conjunto atual pelo conjunto candidato, reduzindo assim a quantidade de atributos até um ponto que o modelo atual apresente melhor acurácia que o modelo candidato. Simultaneamente, ajustando as posições dos centroides do FCM, de modo que proporcionem uma melhor acurácias com os dados de treinamento.
- **(vi) Avaliação dos Modelos:** a avaliação é realizada para comparar o desempenho dos dois modelos. Caso o modelo candidato tenha um desempenho melhor, ele se torna o novo modelo atual e é criado um novo modelo candidato, excluindo o atributo menos relevante. Isto é, caso o modelo candidato apresente melhores resultados, volta-se a Etapa (iv), ou seja, o modelo atual assume o lugar do modelo candidato e o modelo candidato tem seu atributo menos relevante excluído. Esse processo se repete enquanto FCM não interromper seu treinamento, seja pelo número de iterações ou cálculo do erro.
- **(vii) Armazenamento dos atributos:** essa etapa é encerrada com o final do processo de treinamento do FCM.

Os atributos que apresentaram a melhor acurácia durante o treinamento são armazenados para o FCM utilizar somente os atributos selecionados no processo de classificação de novas amostras.

Em (2) é apresentado o pseudocódigo do algoritmo proposto do FCM com o modelo MRFS de seleção de atributos.

#### IV. RESULTADOS EXPERIMENTAIS

Nesta seção serão apresentados os resultados computacionais para avaliar as abordagens propostas. Os experimentos serão executados com três diferentes variações de um modelo de agrupamento *fuzzy*: o FCM não supervisionado, uma variação do FCM supervisionada e o FCM-MRFS. Na execução dos experimentos, todas as bases de dados utilizadas foram divididas aleatoriamente em 80% das amostras para treinamento e 20% para validação. Os resultados apresentados são da média da acurácia para 25 repetições de cada algoritmo em cada diferente base de dados utilizada.

O parâmetro do número de grupos no FCM supervisionado é igual à quantidade de classes de saída para cada diferente base de dados. Os modelos do FCM supervisionado e com seleção de atributos, nomeados de FCM-SUP e FCM-MRFS, geram a quantidade de grupos necessários no processo de treinamento. A medida utilizada para avaliar os resultados de todas as variações dos algoritmos e em todos os experimentos foi a acurácia (7):

$$Acc(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} 1(\hat{y}_i = y_i) \quad (7)$$

onde  $\hat{y}_i$  é o classe prevista pelo classificador da  $i$ -ésima amostra e  $y_i$  é a classe correta,  $n$  é o número total de amostras.

##### A. Bases de Dados

Para a execução dos experimentos, diferentes bases de dados de classificação foram executadas com versões do FCM. Na Tabela I é apresentado os detalhes das bases de dados, com suas diferentes quantidades de atributos de entrada de cada base, a quantidade de amostras, classes de saída e as plataformas que foram obtidas.

##### B. Resultados

Na Tabela II, os resultados apresentados são das acurácias médias de todos os experimentos nas diferentes bases de dados utilizadas. Em uma avaliação entre os algoritmos testados, o FCM-SUP apresentou melhor acurácia que a versão não supervisionada do FCM. Porém, o FCM-MRFS foi melhor que outras versões em 100% dos experimentos. Na Tabela II, também é apresentado a média geral de cada algoritmo para todas as bases, onde o FCM-MRFS foi aproximadamente 40% melhor que o FCM não supervisionado e 7% melhor que o FCM-SUP.

Na Figura IV-B, é apresentado as acurácias médias das três versões do FCM. É possível notar que a versão do FCM-MRFS mantém sempre melhores resultados que a versão FCM-SUP e FCM não supervisionado. Resultados que se

**Algoritmo 2:** Pseudocódigo para o FCM-MRFS.

```

Entradas: Xtreino, Ytreino, niteracoes;
Saídas: M, bcentroides;
// M = atributos de entradas
para  $i=1$  até  $M$  faça
  // n = amostras
  para  $j=1$  até  $n$  faça
    |  $S_{ij}$  = Calcula a Média Amostral - Eq. (5);
  fim
   $R_i$  = Calcula a Razão das Médias - Eq. (6);
fim
fim
Ordena Xtreino em  $R$ ;
Define quantidade de classes em Ytreino;
Inicializa centroides com uma amostra de cada classe;
Copia centroides em bcentroides;
Define acurácia de bcentroides como zero;
iteracao = 0;
Test = Verdadeiro;
// FS e Treinamento
repita
  Cria conjunto atual com  $M$  entradas;
  Cria conjunto candidato  $M - 1$  entradas;
  Classifica conjunto atual - Eq. (3);
  Calcula Acurácia do conjunto atual;
  Classifica conjunto candidato - Eq. (3);
  Calcula Acurácia do conjunto candidato;
  se Acurácia do conjunto atual < Acurácia do
  conjunto candidato então
    |  $M = M-1$  (remove a última entrada);
  fim
  // Utilizada a acurácia do conjunto
  atual
  se Acurácia com bcentroides < Acurácia com
  centroides então
    | Copia centroides em bcentroides;
    | Armazena o valor da acurácia;
  fim
  Atualiza posição dos centroides - Eq. (1);
  Calcula o erro;
  se (error > 0) and (iteracao < niteracao) então
    | Teste = Falso;
  fim
  senão
    | iteracao = iteracao + 1;
  fim
até Test;
Define bcentroides como melhor;
Define  $M$  atributos de entrada;

```

destacam principalmente nas bases de *Heart Disease* e *Wine Dataset*, onde a acurácia do FCM-MRFS foi bem melhor que ambas as outras versões apresentadas.

Table I  
BASES DE DADOS UTILIZADAS NOS EXPERIMENTOS.

Nome	Atrib. de Ent.	Amostras	C. Saída	Fonte
Iris Dataset	4	150	3	[14]
Heart Disease	13	303	2	[8]
Wine Dataset	13	178	3	[14]
Glass Ident.	10	214	6	[14]
QSAR Bio.	41	1054	2	[14]
Spambase	58	4601	2	[14]

Table II  
ACURÁCIA DOS ALGORITMOS(%).

Algoritmo	FCM	FCM-SUP	FCM-MRFS
Iris Dataset	41,46	91,46	<b>95,60</b>
Heart Disease	51,80	65,70	<b>80,39</b>
Wine Dataset	35,44	70,55	<b>82,77</b>
QSAR Biodegradation	52,35	65,66	<b>68,94</b>
Glass Identification	35,37	87,72	<b>88,00</b>
Spambase Dataset	46,14	67,88	<b>71,61</b>
Média	43,76	74,82	<b>81,21</b>

Figure 1. Acurácias do FCM



## V. CONCLUSÃO

Este artigo introduziu duas novas abordagens para o *Fuzzy C-Means*. Na primeira abordagem o FCM foi adaptado para executar tarefas de classificação de maneira supervisionada. Por outro lado, a segunda abordagem utilizou o FCM supervisionado na primeira abordagem e incluiu um método *wrapper* de seleção de atributos de entrada. O método de seleção de atributos incorporado ao FCM é baseado na razão das médias e possui baixo custo computacional e utiliza apenas operações matemáticas básicas.

O FCM-MRFS foi avaliado e comparado com duas versões do FCM em tarefas de classificação: (i) FCM original com treinamento não supervisionado; (ii) FCM com o mesmo treinamento supervisionado implementado no FCM-MRFS. Os resultados experimentais mostram que o FCM-MRFS obteve

um melhor desempenho na classificação, seguido pelo FCM Supervisionado.

Os bons resultados alcançados do método FCM-MRFS, indicam que o método é promissor, principalmente pelo baixo custo computacional e facilidade em incorporar o MRFS em algoritmos de classificação. Para trabalhos futuros o método MRFS pode ser explorado em novas variações como em algoritmos como *Support Vector Machine*, *Artificial Neural Networks*, *K-Nearest Neighbor*, entre outros.

#### ACKNOWLEDGMENT

Os autores agradecem ao apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

#### REFERENCES

- [1] Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Springer US (1981). <https://doi.org/10.1007/978-1-4757-0450-1>
- [2] Cox, E.: Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration. The Morgan Kaufmann Series in Data Management Systems, Elsevier Science (2005)
- [3] Dalal, K.R.: Analysing the Role of Supervised and Unsupervised Machine Learning in IoT. In: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). pp. 75–79. IEEE (2020). <https://doi.org/10.1109/ICESC48915.2020.9155761>
- [4] Duan, J., Sun, H.: Comprehensive financial analysis of a company relying on fuzzy c-mean (fcm) clustering algorithm. In: 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE). pp. 1904–1907 (2020). <https://doi.org/10.1109/ICMCCE51767.2020.00417>
- [5] Dunn, J.C.: A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* **3**(3), 32–57 (1973). <https://doi.org/10.1080/01969727308546046>
- [6] Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *The Journal of Machine Learning Research* **3**, 1157–1182 (2003)
- [7] Ibrahim, R.A., Elaziz, M.A., Oliva, D., Cuevas, E., Lu, S.: An opposition-based social spider optimization for feature selection. *Soft Computing* **23**(24), 13547–13567 (2019). <https://doi.org/10.1007/s00500-019-03891-x>
- [8] Kaggle: The Home of Data Science & Machine Learning. <https://www.kaggle.com/>
- [9] Lang, S.: The Mean Value Theorem, pp. 159–180. Springer New York, New York, NY (1986). <https://doi.org/10.1007/97814419853235>
- [10] Qin, H., Tang, S.: A Solution to Dimensionality Curse of BP Network in Pattern Recognition Based on RS Theory. In: 2009 International Joint Conference on Computational Sciences and Optimization. pp. 636–638. IEEE (2009). <https://doi.org/10.1109/CSO.2009.324>
- [11] Rao, L.J., Challa, R., Sudarsa, D., Naresh, C., Basha, C.Z.: Enhanced automatic classification of brain tumours with fcm and convolution neural network. In: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). pp. 1233–1237 (2020). <https://doi.org/10.1109/ICSSIT48917.2020.9214199>
- [12] Sahoo, P., Riedel, T.: Mean Value Theorems and Functional Equations. World Scientific (1998)
- [13] Tavares, E., da Silva, A.M., Moita, G.F., Cardoso, R.T.N.: A straightforward feature selection method based on mean ratio for classifiers. *Intelligent Decision Technologies: An International Journal* (2021)
- [14] UCI: UC Irvine Machine Learning Repository. <https://archive.ics.uci.edu/ml/index.php>
- [15] Villalon-Turrubiates, I.E.: Classification algorithm for embedded systems using high-resolution multispectral data. In: 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS. pp. 3582–3585 (2013). <https://doi.org/10.1109/IGARSS.2013.6723604>
- [16] Vora, S., Yang, H.: A comprehensive study of eleven feature selection algorithms and their impact on text classification. In: Proceedings of Computing Conference 2017. vol. 2018-January, pp. 440–449. Institute of Electrical and Electronics Engineers Inc. (2018). <https://doi.org/10.1109/SAI.2017.8252136>
- [17] Yan, X., Sun, N.: Improved fcm based distance regularization level set algorithm for image segmentation. In: 2020 IEEE 3rd International Conference of Safe Production and Informatization (ICSPI). pp. 278–281 (2020). <https://doi.org/10.1109/ICSPI51290.2020.9332456>
- [18] Yao, Z., Yang, Y., Li, Y., Liu, H.: State evaluation of sight control box based on rough set fusion improved fcm clustering algorithm. In: 2020 Prognostics and Health Management Conference (PHM-Besançon). pp. 207–213 (2020). <https://doi.org/10.1109/PHM-Besancon49106.2020.00041>