



**KAROLINA CARDOSO FARIA**

**APRENDIZADO INCREMENTAL ONLINE A PARTIR DE  
FLUXOS DE DADOS MÉDICOS: FBEM E ECM**

**LAVRAS-MG**

**2020**

**KAROLINA CARDOSO FARIA**

**APRENDIZADO INCREMENTAL ONLINE A PARTIR DE FLUXOS DE DADOS  
MÉDICOS: FBEM E ECM**

Monografia apresentada à Universidade  
Federal de Lavras, como parte das exigências  
do Curso de Engenharia de Controle e  
Automação, para a obtenção do título de  
Bacharel.

Prof. Dr. Daniel Furtado Leite

Orientador

**LAVRAS-MG**

**2020**

**KAROLINA CARDOSO FARIA**

**APRENDIZADO INCREMENTAL ONLINE A PARTIR DE FLUXOS DE DADOS  
MÉDICOS: FBEM E ECM**

**INCREMENTAL LEARNING ONLINE FROM MEDICAL DATA STREAMS:  
FBEM AND ECM**

Monografia apresentada à Universidade  
Federal de Lavras, como parte das exigências  
do Curso de Engenharia de Controle e  
Automação, para a obtenção do título de  
Bacharel.

APROVADA em 17 de julho de 2020.

Dr. Daniel Furtado Leite	UFLA
Dr (a). Silvia Costa Ferreira	UFLA
Dr (a). Tatiane Carvalho Alvarenga	IF SUDESTE MG
Rafael Henrico de Souza	UFLA

Prof. Dr. Daniel Furtado Leite

Orientador

**LAVRAS-MG**

**2020**

## RESUMO

No contexto de computação e automação atual, os dados são gerados em grandes volumes e precisam ser processados em tempo real no sentido de que a essência da informação seja capturada e mantida em modelos. A produção de fluxos de dados ocorre por meio de sensores, aplicativos, GPS, câmeras, satélites, usuários de redes sociais e de vários ambientes virtuais. O principal desafio de se trabalhar com fluxo de dados inclui manter modelos classificadores, preditores ou controladores atualizados à medida que os dados são gerados. Há limites de tempo e memória a serem atendidos por algoritmos de aprendizado de máquina online. Além disso, fluxos de dados estão sujeitos a mudanças graduais e abruptas de conceito, ou seja, possuem comportamento imprevisível. Algoritmos de aprendizado voltados para esses fluxos devem ser capazes de perceber variações e modificar parâmetros e estrutura de modelos. O presente trabalho aplica dois algoritmos e modelos incrementais online, *evolving Clustering Method* (ECM) e *Fuzzy-Set-Based evolving Modeling* (FBeM) para a classificação de fluxos de dados no contexto médico. Consideramos bases de dados *benchmark* como Câncer de Mama, Doença do Coração, e Telemonitoramento da Doença de Parkinson. Apesar das bases de dados serem estáticas, e serem inteiramente disponibilizadas, consideramos que as amostras são disponibilizadas sequencialmente, e que suas classes são desconhecidas. Os algoritmos ECM e FBeM atuam, portanto, em modo não-supervisionado. Eles criam, e adaptam em tempo real, modelos que mapeiam variáveis mensuráveis em uma classe. Os modelos gerados são comparados em termos de compactação (número de *clusters* e somatório das áreas de cobertura), acurácia e através de índices de validação como: Xie-Beni e *Partition Coefficient*.

**Palavras-chave:** Fluxo de Dados, Modelos Incrementais Online, Clusterização Fuzzy.

## **ABSTRACT**

In the current computing and automation context, large volumes of data have been produced. The data should be processed in real time in the sense that the essence of the information must be captured and maintained in models. The production of data streams occurs through sensors, mobile apps, GPS, cameras, satellites, users of social networks and other virtual environments. The main challenge of dealing with data streams includes keeping classifiers, predictors and controllers updated as the data is being generated. There are time and memory constraints to be met by online machine learning algorithms. In addition, streams are subject to gradual and abrupt concept change, that is, they may behave unpredictably. Learning algorithms to handle data streams must be able to change the parameters and structure of models. In this work we apply two incremental methods, namely, evolving Clustering Method (ECM) and Fuzzy-Set-Based evolving Modeling (FBeM) for the classification of medical data. We considered benchmark databases such as Breast Cancer, Heart Disease, and Parkinson's Disease Telemonitoring. Although the datasets are static, and are entirely available, we consider that the samples are available sequentially, and that their classes are unknown. The ECM and FBeM algorithms therefore operate in an unsupervised mode. They create, and update models in real time to map measurable variables into a class. The generated models are compared in terms of compactness (number of clusters and covered area), accuracy and in terms of the validation indexes Xie-Beni and Partition Coefficient.

**Keywords:** Data Stream, Incremental Models Online, Fuzzy Clustering.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>6</b>
1.1	Contextualização.....	6
1.2	Motivação.....	9
1.3	Organização e Estrutura do Texto .....	14
<b>2</b>	<b>CONCEITOS IMPORTANTES E TRABALHOS RELACIONADOS.....</b>	<b>14</b>
2.1	Aprendizado de Máquina Incremental .....	14
2.2	Fluxos de Dados Não Estacionários.....	17
2.3	Formação Deriva de Conceito .....	18
2.4	Clusterização.....	20
2.5	Clusterização em Fluxo de Dados.....	23
<b>3</b>	<b>MODELOS EVOLUTIVOS A PARTIR DE FLUXOS DE DADOS .....</b>	<b>27</b>
3.1	Introdução.....	27
3.2	Método de Clusterização Evolutiva – ECM.....	28
3.3	Modelos evolutivos baseados em conjunto Fuzzy – FBeM.....	30
3.4	Correlação de Spearman para Seleção de Atributos.....	36
3.5	PCA – Redução de Dimensionalidade.....	37
<b>4</b>	<b>METODOLOGIA .....</b>	<b>38</b>
4.1	Suposições e Hipóteses .....	39
4.2	Fluxo de Trabalho.....	39
4.3	Base de Dados.....	40
4.4	Pré-processamento.....	44
4.5	Métricas de Avaliação dos Resultados .....	45
<b>5</b>	<b>RESULTADOS .....</b>	<b>47</b>
5.1	Algoritmo FBeM .....	48
5.2	Algoritmo ECM .....	54
<b>6</b>	<b>DISCUSSÃO .....</b>	<b>61</b>
<b>7</b>	<b>CONCLUSÃO.....</b>	<b>63</b>
	<b>REFERÊNCIAS.....</b>	<b>64</b>

# 1 INTRODUÇÃO

## 1.1 Contextualização

Toda revolução provoca transformações, seja no contexto político, econômico, cultural ou social. Historicamente, a sociedade passou por grandes revoluções, tais como: Revolução Francesa, Primeira Revolução Industrial, Revolução Cubana, Revolução de 1930, entre outras. Dentre estas, a única que não representava a luta de uma classe pelo poder ou pela igualdade dos direitos é a Primeira Revolução Industrial. As revoluções industriais são caracterizadas por provocarem uma mudança abrupta no contexto econômico, social e político com a incorporação de novas tecnologias.

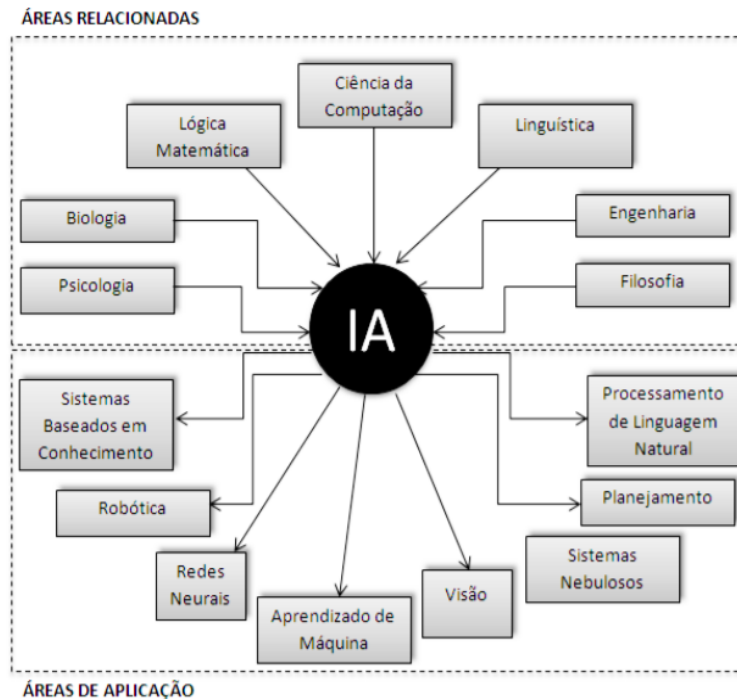
Para alguns especialistas, temos vivenciado a quarta revolução industrial, ou *Indústria 4.0*. Em seu livro *A Quarta Revolução Industrial*, publicado em 2016, o professor alemão, Klaus Schwab ressalta a importância que a quarta revolução industrial terá na forma como os seres humanos estabelecem suas relações comportamentais, sendo diferente de qualquer experiência que o ser humano já tenha vivido antes.

A Indústria 4.0 se define pela presença das seguintes tecnologias: Inteligência Artificial (IA), Robótica, Internet das Coisas, Veículos Autônomos, Impressão 3D, Nanotecnologia, Biotecnologia, Armazenamento de Energia e Computação Quântica (SANSON et al., 2007). A Fig. 1 evidencia as áreas relacionadas à IA.

Através da IA as máquinas realizam tarefas associadas aos seres humanos, tais como: aprender e raciocinar, com o objetivo de tomar decisões que solucionem problemas.

Como exemplos de aplicação da Inteligência artificial temos: sugestão de *tags* de foto, sistemas de recomendação de hotéis e restaurantes, agrupamento baseados em reconhecimento de imagem em redes sociais, recomendação de produtos pela Amazon, otimização de rotas em tempo real pelo Waze; criação de carros autônomos pelo Google, diagnóstico preciso de doenças pela plataforma Watson da IBM, entre outros.

Figura 1 - Áreas relacionadas a IA



Fonte: SETIAWAN et al., (2009).

Sistemas que utilizam inteligência artificial podem ser classificados de acordo com as técnicas que utilizam, como *Machine Learning* (Aprendizado de Máquina) e *Deep Learning* (Aprendizagem Profunda). *Machine Learning* é o uso de algoritmos para organizar dados (dados geralmente estruturados), reconhecer padrões e fazer com que computadores, *softwares*, robôs possam aprender com esses modelos e gerar *insights* inteligentes sem necessidade de pré-programação. *Deep Learning* é a parte do aprendizado de máquina que, por meio de algoritmos de alto nível, imita a rede neural do cérebro humano. Dados são submetidos a várias camadas de processamento, como camadas não-lineares e convolucionais, que simulam o processamento de neurônios biológicos. As aplicações incluem visão computacional, reconhecimento de fala, compreensão de linguagem natural e reconhecimento de imagens.

A aplicação de IA em medicina iniciou-se com o artigo de Shortliffe em 1963. Szilovits, em sua publicação “*The Coming Age of Artificial Intelligence in Medicine*”, admite que atualmente a IA está se tornando não apenas uma parte, mas um componente essencial da informática médica, além de ser um recurso importante na solução de problemas em atenção à saúde (PATEL et al., 2009).



A plataforma Watson da IBM e o Deep Mind do Google são consideradas importantes experiências do uso da IA na medicina. Pesquisadores da IBM, usando redes neurais, conseguiram obter uma acurácia de 86% no diagnóstico de retinopatia diabética feito em 35 mil imagens de retina acessadas através da tecnologia EyePacS da IBM de identificação de lesões e outros sinais observados em vasos sanguíneos. Registrando um grande número de casos com seu esquema de diagnóstico, tratamentos prescritos e resultados obtidos, esses sistemas permitem uma expansão do conhecimento médico e a sugestão de condutas a serem seguidas (LOBO, 2007).

Fabio Matoso, CEO da IBM Watson no setor de Oncologia, afirma que o Watson é uma plataforma de soluções na nuvem, onde através de uma leitura de artigos, livro texto, periódicos, além do histórico do paciente, ele aprende a pensar e devolver *insights* para médicos especialistas, estes dizem o quanto uma solução é adequada. Com tal retorno, Watson aprimora sua solução ao longo do tempo.

É evidente que o Watson não substitui o trabalho dos médicos, sendo apenas uma ferramenta de auxílio no diagnóstico do paciente, como um gestor de informação.

O principal desafio de trabalhar com dados médicos, consiste em transformar a vasta quantidade de dados brutos em informação e conhecimento, e acumular experiência ao longo do tempo para apoiar processos de tomada de decisão no futuro. Existe ainda o problema da infraestrutura, em relação ao tempo e a memória de processamento e armazenamento destes dados, uma vez que estes são gerados a um tempo infinito e devem ser tratados em um tempo finito. Se a velocidade de produção dos dados for maior que a velocidade de processamento, os dados se acumulam continuamente, e são descartados, tornando o aprendizado cada vez mais defasado (LEITE, 2002).

No presente trabalho iremos utilizar sistemas inteligentes evolutivos, para classificar dados médicos. Estes sistemas possuem características eficientes para lidar com as incertezas e grande quantidade dos dados, tais como: capacidade de adaptação contínua dos parâmetros e da estrutura de modelos sempre que novas informações são disponibilizadas. Os algoritmos utilizados são: *evolving Clustering Method* (ECM) (KASABOV et al., 2002) e *Fuzzy-Set-Based evolving Modeling* (FBeM) (LEITE et al., 2013).

As bases de dados são retiradas do repositório da Universidade da Califórnia em Irvine. São elas: Câncer de Mama, Doença do Coração, e Telemonitoramento da Doença de Parkinson. Os algoritmos ECM e FBeM atuam, em modo não-supervisionado, nas amostras que são disponibilizadas sequencialmente com suas classes desconhecidas.

De acordo com Leite et al., (2014) estes algoritmos, evolutivos, criam e adaptam em tempo real modelos que mapeiam variáveis mensuráveis em uma classe. Tais sistemas evolutivos podem ser inspirados no comportamento de redes neurais biológicas e na evolução de indivíduos durante seu ciclo de vida: aprendendo a partir da experiência, herança, mudança gradual, e informação incompleta e incerta.

Os modelos gerados serão avaliados em relação a acurácia, além dos índices de validação dos *clusters*: Xie-Beni e Partition Coefficient.

## **1.2 Motivação**

O diagnóstico médico realizado de forma precisa e precoce, atrasa o progresso da doença, proporciona o desenvolvimento de tratamentos individualizados, resultando em uma qualidade de vida melhor para os pacientes, e aumentando a probabilidade de sobrevivência (considerando doenças graves).

Através da IA, o médico tem a possibilidade de prover um diagnóstico mais preciso de determinadas doenças. Os algoritmos de *Machine Learning* e *Deep Learning* são um suporte para o médico, que poderá analisar laudos em menos tempo, priorizando os casos mais urgentes de maneira relativamente mais assertiva.

O presente trabalho aplica algoritmos evolutivos de *machine learning*, com o objetivo de auxiliar o médico no diagnóstico precoce e assertivo das doenças de Parkinson, do coração e câncer de mama.

### **1.2.1 Doença de Parkinson**

A doença de Parkinson tem caráter degenerativo e é caracterizada pela perda progressiva de neurônios da parte compacta da substância negra, situada no mesencéfalo. A degeneração nesses neurônios é irreversível e resulta na diminuição da produção de dopamina, que é um neurotransmissor essencial no controle dos movimentos (MOREIRA et al., 2007).

O diagnóstico baseia-se principalmente nos aspectos semiológicos, com destaque para a apresentação da síndrome extrapiramidal (característica da doença), manifestada por tremor de repouso, rigidez, perda do reflexo postural e hipocinesia. A marcha característica, consta de pequenos passos com velocidade crescente e também denuncia a presença do distúrbio. Alterações da mímica facial, do humor, da caligrafia e da voz também são marcos significativos da doença (LITTLE et al., 2009).

Os sintomas do Parkinson costumam ser sutis e surgem de forma gradual, podendo passar despercebidos por muito tempo e até serem considerados traços característicos do envelhecimento, o que dificulta o diagnóstico assertivo da doença.

Não existe cura para a doença de Parkinson, porém existe o tratamento da doença que pode ser através de medicamentos, psicoterápico e até cirúrgico em casos onde os medicamentos não apresentam mais efeito (RIBEIRO et al., 2018).

O método utilizado no presente trabalho para o diagnóstico da doença de Parkinson, analisa sinais de voz dos pacientes, visto que pesquisas recentes mostraram que os primeiros indicadores da DP, são exibidos pela alteração sutil na voz, hiposmia e distúrbio do sono (OLANOW et al., 2011; COOPER et al., 2008; LEITE et al., 2013). A análise dos sinais de voz é realizada usando ferramentas acústicas que registram as mudanças na pressão nos lábios ou dentro do trato vocal.

SPIELMAN et al., (2003), OLIVEIRA et al., (2004) e SILVEIRA et al., (2005) afirmam, com base em suas pesquisas, que todos os sujeitos com DP apresentam dificuldades de comunicação. Silveira e Brasolotto (2005) reportam queixa de dificuldade para pronunciar as palavras e, ocasionalmente, fala embaralhada e rápida. Essas alterações de fala, segundo os autores são geradas devido ao prejuízo da função do gânglio basal. Afirmam que se trata também de alterações vocais ocasionadas pelo fechamento incompleto das pregas vocais, pela redução da ativação da musculatura laríngea, atrofia ou fadiga muscular, assimetria de tensão ou movimento das pregas vocais, rigidez das pregas vocais ou dos músculos respiratórios.

Little (2009) do Media Labs do Massachusetts Institute of Technology (MIT), desenvolveu um software para detectar diferenças nos padrões de voz afim de classificar indivíduos saudáveis e indivíduos portadores da DP, obteve-se um desempenho de classificação correto de 91,4% usando o método de análise de regressão LASSO (Least Absolute Shrinkage and Selection Operator) (LITTLE et al., 2009).

A proposta de Gil e Johnson (2009) teve uma precisão de aproximadamente 90% no diagnóstico de DP, usando, uma Rede Neural Artificial do tipo feed-forward (RNA) e Máquina de Vetor Suporte (SVM). Classificadores Fuzzy C-Means e Gustafson-Kessel proposto em CAGLAR et al., (2010), para o grau de severidade da DP, apresentaram precisão de 83,9% e 85,4% respectivamente.

### 1.2.2 Câncer de Mama

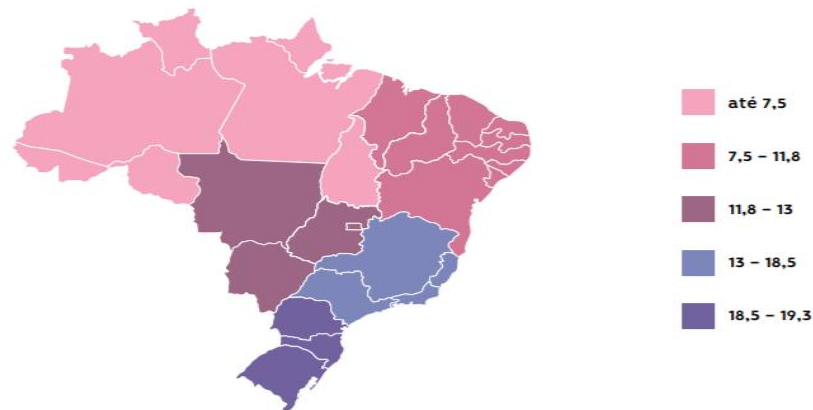
A definição de câncer dada pelo Instituto Nacional do Câncer (INCA) é: Câncer é o nome dado a um conjunto de mais de 100 doenças que têm em comum o crescimento desordenado de células, que invadem tecidos e órgãos.

As células ao se dividirem rapidamente, tendem a ser agressivas e incontroláveis, determinando a formação de tumores, que podem espalhar para outras regiões do corpo. Os tumores são divididos em dois grupos: benignos e malignos. Nos tumores benignos as células têm seu crescimento semelhante às do tecido normal. Na maioria dos casos o tumor é removido e o paciente curado por meio de cirurgia. Entretanto, nos tumores malignos, as células multiplicam-se rapidamente e possuem a capacidade de “invadir” estruturas próximas ao local de origem. A cura neste tipo de tumor depende do diagnóstico precoce e do tratamento adotado.

De acordo com o Ministério da Saúde, o câncer de mama é mais incidente em mulheres, representando 24,2% do total de casos de câncer no mundo, em 2018. É a quinta causa de morte por câncer em geral e a causa mais frequente de morte por câncer em mulheres (WORLD HEALTH ORGANIZATION, 2020). Em 2016, foram registrados, no Brasil, 16.069 óbitos de mulheres devido ao câncer de mama. A taxa bruta de mortalidade por esse câncer foi de 15,4 óbitos por 100 mil mulheres no país. A Fig. 2 representa esta taxa por regiões do Brasil, as regiões Sul e Sudeste tiveram as maiores taxas de mortalidade.

A principal causa deste número elevado, está atrelada a um diagnóstico tardio, onde o câncer encontra-se em estágio avançado, piorando o prognóstico. Neste cenário a detecção precoce desta doença é uma estratégia fundamental na redução do número de óbitos. Quanto mais cedo um tumor invasivo é detectado e o tratamento iniciado, maior a probabilidade de cura.

Figura 2 -Taxa bruta de mortalidade por câncer de mama nas Regiões geográficas do Brasil, por sexo feminino, em 2016



Fonte: INSTITUTO NACIONAL DE CÂNCER (2019).

No presente trabalho iremos utilizar algoritmos evolutivos para a classificação do câncer de mama em benigno ou maligno, através dos dados de biópsia realizada pelo método PAAF, que consiste na retirada de uma parte do tumor por agulha fina. Classificações mais precisas de tumores benignos podem evitar tratamentos desnecessários.

Existem vários estudos para o diagnóstico do câncer de mama com base nos dados *Wisconsin Diagnostic Breast Cancer* (WDBC). Em um estudo de (ALBRECHT et al.,2002) foi utilizado uma rede neural do tipo perceptron multicamadas (MLP) para obtenção de uma taxa de sucesso de 98,8%. Em um estudo que utilizou um método de agrupamento fuzzy, alcançou-se um sucesso de 95,6%. (NAUCK et al., 1999) e (PEÑA-REYES et al.,1999) alcançaram sucesso de 97,71% por meio da teoria fuzzy.

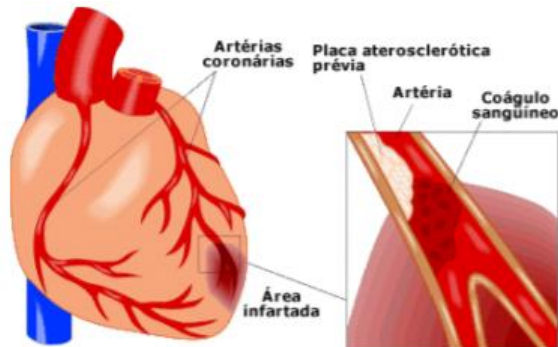
### 1.2.3 Doença do Coração

Segundo a Sociedade Brasileira de Cardiologia (SBC), as doenças cardiovasculares representam a principal causa de mortes no Brasil, responsáveis por mais de 30% dos óbitos registrados. São mais de mil mortes por dia, cerca de 43 por hora, isto é, uma morte a cada 90 segundos.

De acordo com o BRATS (2013), a doença arterial coronariana (DAC) é resultante do processo de estreitamento ou oclusão das artérias coronarianas por aterosclerose, uma doença que afeta o revestimento endotelial das grandes e médias artérias. A falta de

distribuição do oxigênio pelo corpo, provocado pelo estreitamento das artérias, provoca a formação de coágulo em uma região, o que pode ser observado na Fig. 3.

Figura 3 -Aterosclerose: formação de coágulo



Fonte: Adaptado de GOMES (2011).

Na área médica, o diagnóstico de doença cardíaca é a tarefa mais difícil. Depende da análise cuidadosa de diferentes dados clínicos e patológicos do paciente por médicos especialistas, o que é um processo complicado.

A DAC é uma enfermidade cujo diagnóstico pode oscilar de fácil à difícil, temos que considerar as queixas clínicas, quando sintomáticas, essas doenças sugerem exames bem específicos que podem confirmar ou excluir o diagnóstico da DAC. Contudo, quando assintomáticas, esses exames passam a ser questionados, devido a seu alto custo, por se tratar de um método invasivo, e, seus efeitos, são nocivos ao paciente. Em casos como este é importante pesquisar o histórico do paciente, como fatores de risco, como fumo, diabetes, sedentarismo entre outros. (PASSOS, 2014)

É necessário, portanto, métodos e ferramentas que auxiliem os médicos no diagnóstico precoce da DAC evitando o agravamento da doença.

Sistemas inteligentes de aprendizado de máquina podem ser usados para reduzir a taxa de mortalidade, melhorar precisão no diagnóstico da doença cardíaca e reduzir o tempo de diagnóstico (SHASHIKANT et al., 2011).

No presente trabalho utilizamos algoritmos evolutivos para a classificação da doença do coração (presença ou ausência da doença), os dados foram retirados do banco de dados de doenças cardíacas de Cleveland, do repositório da Universidade da Califórnia (UCI). Estudos foram realizados com a mesma base de dados, porém utilizando métodos de classificação diferentes: Almeida (2006) propõem em seu trabalho o desenvolvimento de uma aplicação

baseada em algoritmos genéticos e raciocínio baseado em casos que, através da análise de uma série de atributos do histórico clínico do paciente, determina qual a relação de cada um destes com o diagnóstico final, sendo ao o algoritmo capaz de definir a probabilidade de este paciente apresentar diagnóstico positivo ou não para a doença obstrutiva coronariana, obtendo taxas de até 90% de acertos.

### **1.3 Organização e Estrutura do Texto**

Esta dissertação está organizada da seguinte forma. No capítulo 2 são abordados os conceitos de aprendizado de máquina incremental, problemas encontrados ao se trabalhar com fluxo de dados não estacionários, além de conceito e aplicações do agrupamento evolutivo. No capítulo 3 são apresentados os métodos de Clusterização Evolutiva, os algoritmos: ECM e FBeM, além dos métodos para redução de dimensionalidade das variáveis. No capítulo 4 é mostrada a metodologia a ser utilizada no processo de classificação, bem como uma explicação sobre as bases de dados e as métricas que serão utilizados na avaliação dos resultados. No capítulo 5, os resultados são evidenciados, por meio de análise da acurácia e dos índices de validação dos clusters. O trabalho é concluído no capítulo 6, indicando possíveis melhorias por meio da IA na área de saúde.

## **2 CONCEITOS IMPORTANTES E TRABALHOS RELACIONADOS**

### **2.1 Aprendizado de Máquina Incremental**

O processo de aprendizagem é caracterizado pela aquisição do conhecimento, por meio da prática ou instrução, após aquisição, existe a etapa de organização deste conhecimento e por fim, a descoberta de novos fatos e teorias por meio da observação e experimentação. O aprendizado de máquina (AM) nada mais é, do que a implementação deste processo nos computadores, investigando métodos computacionais capazes de aprender e adquirir conhecimento de forma automática (MICHALSKI et al., 2013).

Um dos métodos mais utilizados em AM é o aprendizado indutivo. A inferência indutiva é um processo de raciocínio lógico para uma conclusão sobre todos os membros de uma classe por meio do exame de apenas uns poucos membros da classe. De uma maneira

mais formal, para um conjunto de objetos  $x = \{a,b,c,d\}$  se a propriedade  $P$  é verdadeira para  $a$  e  $b$ , então  $P$  é verdadeira para todo  $x$ . As duas principais técnicas de aprendizado indutivo são conhecidas como aprendizado supervisionado e não supervisionado.

No aprendizado supervisionado através de um conjunto de dados de entrada associados a um conjunto de resultados é necessário construir um modelo que aprende como prever  $Y_y$  baseado em  $X_x$ . O conjunto de resultados é denominado de classe ou rótulo. As técnicas de aprendizado supervisionado se diferenciam pelo tipo de rótulo presente nos exemplos. Para rótulos discretos, ou que possuem um valor nominal, são utilizados métodos de classificação, enquanto que para rótulos com valores contínuos são utilizados métodos de regressão. Exemplos desses métodos incluem árvores de decisão (QUINLAN, 1986), redes neurais (BISHOP, 1995) e abordagens evolutivas (GOLDBERG, 1989).

O aprendizado não supervisionado é utilizado em casos onde os dados não possuem informações de rótulos, ou seja, não existem conceitos prévios relacionados ao conjunto de exemplos. Um exemplo deste aprendizado é o retratado pela técnica de agrupamento, tal técnica realiza a divisão dos exemplos em grupos de forma que exemplos semelhantes pertençam a um mesmo grupo e exemplos distintos pertençam a grupos distintos (JAIN; MURTY; FLYNN, 1999). Para exemplos que possuem atributos inteiramente numéricos, a avaliação de similaridade mais comum baseia-se em métricas de distância (TAN; STEINBACH; KUMAR, 2005). De acordo com o modo de aprendizado existem duas classificações possíveis: modo não incremental (batch) e incremental. No modo não incremental todo o conjunto deve estar presente para o aprendizado, já no modo incremental a hipótese antiga é alterada sempre que novos exemplos são adicionados ao conjunto de treinamento.

Os algoritmos de AM podem ser executados no modo online ou offline. No modo online o algoritmo aprende enquanto está em operação, já no modo offline o algoritmo necessita de aprendizagem separada da fase de operação, a incorporação de novos dados requer a reconstrução do modelo de previsão a partir do processamento de toda a base de dados, o que torna o processo lento e ineficiente. É esperado que depois da fase de treinamento os algoritmos que possuem aprendizado offline tenham criado em suas estruturas internas uma função hipótese que seja capaz de generalizar corretamente em qualquer conjunto de dados daquele contexto.



No aprendizado online, não se espera que uma hipótese seja criada para aquele contexto de informações, o algoritmo trata todas as entradas de maneira única, uma por uma, sempre aglomerando os conhecimentos desta entrada (DEKEL, 2009).

A escolha de qual tipo de algoritmo de AM utilizar depende das características e do contexto do problema. No contexto, onde os dados são gerados continuamente, que são os denominados dados contínuos e com crescentes variações, trabalhar com os modelos de aprendizado de máquina offline, tornou-se obsoleto, uma vez que estes modelos não serão capazes de lidar com conceitos de aprendizagem que aparecem somente nas informações mais novas. O aprendizado de máquina incremental online surgiu como uma alternativa a este problema, sendo este capaz de capacitar modelos de Aprendizado de máquina a aprenderem continuamente, conforme novas entradas chegam ao sistema (SCHICK, 2019).

Existem várias aplicações que necessitam de aprendizado incremental, estas são: detecção de intrusos em rede de computadores (LANE e BRODLEY, 1998), detecção de fraudes em sistemas de cartão de crédito (WANG et al., 2003), problemas relacionados a finanças (LAST, 2002), dentre outros.

De acordo com a definição de (LANGLEY, 1995), para um sistema ser considerado de aprendizado incremental online, o mesmo deve ser submetido a três restrições:

- Deve processar os dados à medida que os mesmos chegam no sistema;
- Não deve processar experiências passadas;
- Deve reter apenas uma estrutura de conhecimento em memória.

A primeira restrição impõe que o sistema use o conhecimento adquirido em qualquer instante durante todo o processo de aprendizado, e não depois do processamento completo, a segunda restrição garante que o tempo de processamento do novo dado seja baixo e constante, e finalmente a terceira restrição assegura uso eficiente da memória.

A principal vantagem de um algoritmo incremental é o fato de não impor restrições à análise de novos dados, quer em tempo, quer em número, estando sempre o algoritmo disponível a qualquer momento. Além disso, é possível obter modelos que processam grandes quantidades de dados sem um gasto excessivo de recursos e de tempo, características que são desejáveis (LAN, 1995). Uma outra vantagem consiste em não desaproveitar o tempo gasto no processamento do algoritmo quando é necessário reavaliar a estrutura de conhecimento. Assim é possível resumir as vantagens em cinco itens fundamentais:

- Poupa memória do computador;

- Não desaproveitar o tempo de processamento que foi gasto para a construção do modelo;
- Maior rapidez de resposta;
- Importante em alguns domínios como o data mining;
- Aproxima-se da forma de aprendizagem dos humanos.

## 2.2 Fluxos de Dados Não Estacionários

Um fluxo de dados é uma sequência contínua e constante de itens, que são apresentados em tempo real segundo a ordem de chegada (GOLAB et al., 2003). Fluxos de dados não estacionários são caracterizados, por possuírem o fenômeno de mudança de conceito no processo de extração de conhecimento. Mudança de conceito, ou *concept drift*, é caracterizado pelas alterações nas distribuições de probabilidades dos dados, que podem introduzir conceitos novos, nunca identificados anteriormente, e também conceitos previamente identificados que desapareceram e voltaram a aparecer após algum tempo (GAMA et al., 2014).

Fluxos de dados não estacionários ocorrem desde aplicações que monitoram tráfego na internet até dados provenientes de sensores. A seguir podemos observar exemplos de aplicações de fluxos de dados não estacionários:

- Transporte: Os sistemas de gerenciamento de tráfego usam a mineração de dados para determinar os estados de tráfego, densidade do carro em uma área específica, acidentes. Os centros de controle de tráfego são os usuários finais desses sistemas. Os sistemas de transporte são dinâmicos (sempre em movimento). Os padrões de tráfego estão mudando sazonalmente e permanentemente, portanto, os sistemas precisam ser capazes de lidar com o *concept drift*.
- Perfil do cliente: O objetivo é segmentar os clientes de acordo com seus interesses. Como os interesses individuais estão mudando ao longo do tempo, os algoritmos de perfil do cliente devem levar em consideração essa não estacionaridade. A análise de redes sociais tem sido empregada na segmentação de clientes. Observa-se que os interesses do usuário não evoluem simultaneamente. Os usuários que costumavam ter interesses semelhantes no passado podem não compartilhar mais os interesses no futuro. As recomendações automáticas podem estar relacionadas

ao perfil do cliente e à assistência pessoal, como exemplo temos a recomendação de filmes Netflix.

- E-mails: A classificação dos e-mails em válido ou spam, depende da preferência dos usuários e estas podem ser alteradas com o tempo. Os e-mails que são considerados como spam também estão em constante alteração; estes além de dependerem das preferências dos usuários dependem também dos produtores de spam (spammers) cujo objetivo é passar o spam como verdadeiro.
- Sensores Inteligentes: Na classificação dos insetos sensores inteligentes captam informações sobre o vôo de insetos, utilizando uma fonte de luz infravermelha e foto transistores, e os classifica de acordo com suas espécies. Os sensores inteligentes monitoram o ambiente externo, desta forma estão sujeitos a alterações nas condições de temperatura e umidade, devido às alterações nas condições climáticas os metabolismos dos insetos são modificados, provocando variações nos dados utilizados para a classificação de suas espécies.

### 2.3 Formação Deriva de Conceito

Segundo Narasimhamurthy e Kuncheva (2007) e Zliobaite (2009), todo o problema de classificação, pode ser representado do seguinte modo:

- Seja  $X \in (R)^p$  uma instância do dado no espaço de característica p-dimensional.  $X \in c_i$ , onde  $c_1, c_2, \dots, c_k$  é o conjunto de rótulos de classe.

O Classificador ideal para classificar  $X \rightarrow c_i$  é completamente determinado pelas probabilidades anteriores das classes  $P(c_i)$  e pelas funções de densidade de probabilidade condicionadas às classes  $p(c_i)$ ,  $i = 1, \dots, k$ .

A teoria da decisão bayesiana, na decisão de classificação é feita com base na probabilidade máxima a posteriori, que para uma classe  $c_i$  é:

$$p(c_i | X) = \frac{P(c_i)p(X|c_i)}{P(X)} \quad (2.2.1.1)$$

onde  $p(X)$  é uma evidência de X, que é constante para todas as classes  $c_i$ .

A Fonte de dados S é definida como um conjunto das probabilidades anteriores e condicionais.

$$S = \{(P(c_1), p(X | c_1)), (P(c_2), p(X | c_2)), \dots, (P(c_k), p(X | c_k))\} \quad (2.2.1.2)$$

Conforme apresentado por Kelly et al., (1999) apesar de existirem diferentes conceitos para o termo *concept drift*, todas referem-se ao problema principal de alteração das probabilidades mencionadas na definição de fonte de dados, assim como mostrado a seguir:

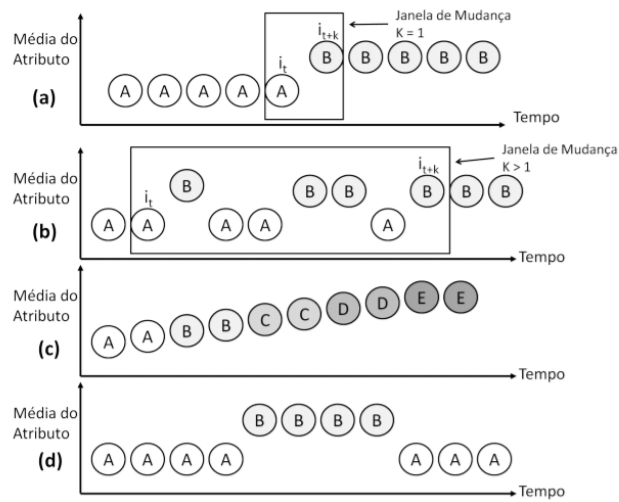
- $P(c)$ : número de exemplos das classes são alterados;
- $P(X/c)$ : mudança na definição das classes;
- $p(c | X)$ : alterações de atributos, possivelmente em alguns deles.

Segundo Gama et al., (2014), os desvios de conceito podem ocorrer de diversas formas:

- Abruptamente: O padrão mais simples de uma mudança é o desvio repentino [*sudden drift*], quando no momento  $t_0$  uma fonte  $S_I$  é subitamente substituída pela fonte  $S_{II}$ .
- Incrementalmente: A distribuição muda de forma lenta, podendo ser identificados conceitos intermediários.
- Gradualmente: O conceito muda repentinamente, mas continua voltando ao conceito anterior durante um tempo.
- Recorrentemente: Os conceitos aparecem e desaparecem da distribuição.

A Fig. 4 ilustra os principais tipos de mudança de conceito.

Figura 4- Tipos de mudanças de conceito: (a) Abrupta; (b) Gradual; (c) Incremental; (d) Recorrente



Fonte: ZLIOBAITE (2009).

Em situações onde há a presença de desvio de conceito os algoritmos incrementais devem ser capazes de fazer suposições sobre a fonte de dados futura [St+1]; identificar possíveis mudanças nos padrões dos dados; com base na mudança do padrão e na suposição futura, os algoritmos devem ajustar seus parâmetros, sendo o mais genérico possível no tratamento dos dados novos a serem rotulados (ZLIOBAITE, 2009).

## 2.4 Clusterização

A Clusterização é a classificação não-supervisionada dos dados dentro de grupos (clusters) com objetos similares. Existe uma homogeneidade entre os indivíduos que compõe o mesmo cluster e heterogeneidade entre clusters, pois os dados pertencentes a um mesmo cluster devem ser mais “similares” entre si do que em relação aos dados pertencentes a outros clusters.

De acordo com Hruschka e Ebecken (2001), a definição do processo de clusterização é: Dado um conjunto com  $n$  elementos  $X = \{X_1, X_2, \dots, X_n\}$ , o problema de clusterização consiste na obtenção de um conjunto de  $k$  clusters,  $C = \{C_1, C_2, \dots, C_k\}$ , tal que os elementos

contidos em um cluster  $C_i$  possuam uma maior similaridade entre si do que com os elementos de qualquer um dos demais clusters do conjunto  $C$ . O conjunto  $C$  é considerado uma clusterização com  $k$  clusters caso as seguintes condições sejam satisfeitas:

$$C_1 \cup C_2 \cup \dots \cup C_k = X \quad (2.3.1)$$

$$C_i \neq \emptyset, \text{ para } 1 \leq i \leq k \quad (2.3.2)$$

$$C_i \cap C_j = \emptyset, \text{ para } 1 \leq i, j \leq k \text{ e } i \neq j \quad (2.3.3)$$

Segundo Moscato e Von Zuben (2004), as principais etapas do processo de clusterização são as seguintes:

- Seleção dos atributos: Os padrões dos dados são descritos pelos atributos, estes devem ser selecionados de forma correta, de forma a codificar a maior quantidade de informações possíveis.
- Extração de características: Através de transformações nos atributos de entrada as características mais relevantes presentes nos dados são selecionadas.
- Medida de similaridade: É fornecida por uma função de distância definida entre dois pares de atributos. É possível incluir na medida de distância aspectos conceituais (qualitativos) ou então numéricos (quantitativos). De modo a evitar que um atributo seja dominante sobre o outro no cálculo da distância, os dados devem ser normalizados. A Seguir temos exemplos de técnicas de normalização.

$$\text{Min} - \text{Max} : n_i = \frac{(x_i - \min(x))}{(x) - \min(x)} \quad (2.3.4)$$

$$\text{Z} - \text{score} : n_i = \frac{(x_i - \text{mean}(x))}{\text{std}(x)} \quad (2.3.5)$$

$$\text{Soma} : n_i = \frac{(x_i)}{\sum_{i=0}^n(x)} \quad (2.3.6)$$

Sendo:  $X_i$ :  $i$ -ésima amostra dos dados de entrada;

$\text{Mean}(X)$ : Média das amostras dos dados de entrada;

$\text{Std}(X)$ : Desvio Padrão das amostras dos dados de entrada.

Como medidas de distância muito utilizadas podemos citar a distância Euclidiana e a distância *city-block*:

- Distância Euclidiana: considera a distância  $d$  entre dois elementos  $X_i$  e  $X_j$  no espaço  $p$ -dimensional:

$$d(X_i, X_j) = (\sum_{l=1}^p |x_i^l - x_j^l|^2)^{\frac{1}{2}} \quad (2.3.7)$$

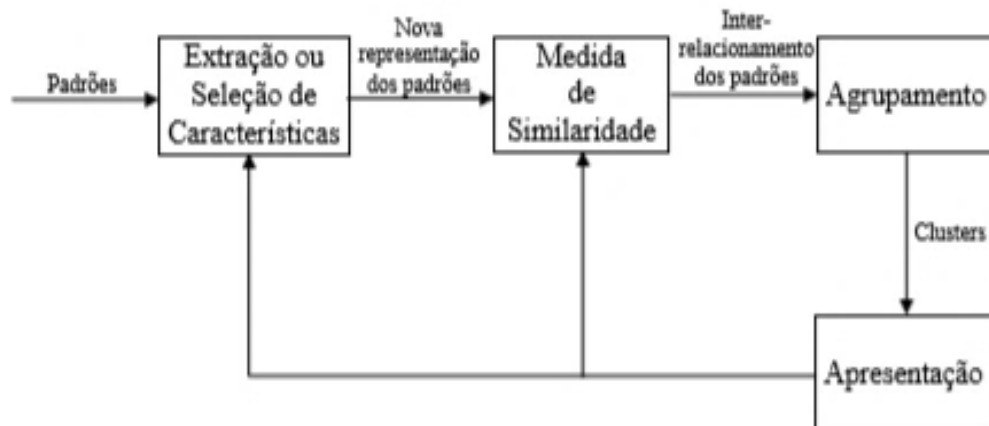
- Distância “city-block”: corresponde à soma das diferenças entre todos os  $p$  atributos de dois elementos  $X_i$  e  $X_j$ , não sendo indicada para os casos em que existe uma correlação entre tais atributos:

$$d(X_i, X_j) = \sum_{l=1}^p |x_i^l - x_j^l|^2 \quad (2.3.8)$$

- Agrupamento: Grupos podem ser definidos como conjuntos crisp (um padrão pertence ou não-pertence a um dado grupo) ou fuzzy (um padrão pode apresentar graus de pertinência aos grupos). O processo de agrupamento pode ser hierárquico, com um processo recursivo de junções ou separações de grupos, ou não-hierárquico, com o emprego direto de técnicas de discriminação de clusters.
- Algoritmo de Agrupamento: Adotado a medida de proximidade e o critério de agrupamento a ser utilizado deve-se escolher o algoritmo de agrupamento, para a formação dos clusters.
- Validação dos resultados: Deve-se após obter os cluster, verificar se a formação dos mesmos está correta. Em geral os resultados são analisados em conjuntos com outras evidências experimentais. Para fazer a validação recorre-se a critérios de otimalidade, definidos de forma subjetiva, alguns destes critérios são: Soma dos Erros Quadrados e Critérios de Dispersão.
- Realimentação: A realimentação do resultado do processo de clusterização pode levar à redefinição dos módulos de “extração ou seleção de características” e “medida de similaridade”.

A Fig.5 abaixo exhibe o processo de clusterização.

Figura 5- Processo de Clusterização



Fonte: Moscato e Von Zuben (2004).

Os algoritmos de clusterização são divididos em dez tipos principais, estes são:

- Métodos Hierárquicos;
- Métodos Particionais;
- Métodos Baseados em Densidade;
- Métodos Baseados em Grade;
- Métodos Baseados em Modelos;
- Métodos Baseados em Redes Neurais;
- Métodos Baseados em Lógica Fuzzy;
- Métodos Baseados em Kernel;
- Métodos Baseados em Grafos;
- Métodos Baseados em Computação Evolucionária.

## 2.5 Clusterização em Fluxo de Dados

A mineração de fluxo de dados é motivada por aplicações que envolvem conjuntos de dados massivos e gerados continuamente, tais como: fluxos de cliques do cliente, registros telefônicos, grandes conjuntos de páginas da Web, dados multimídia, transações financeiras e dados científicos observacionais. Para tais aplicações são necessários uma análise exploratória dos dados mais robusta, e é neste ponto que os clusters para fluxo de dados contínuos são utilizados.

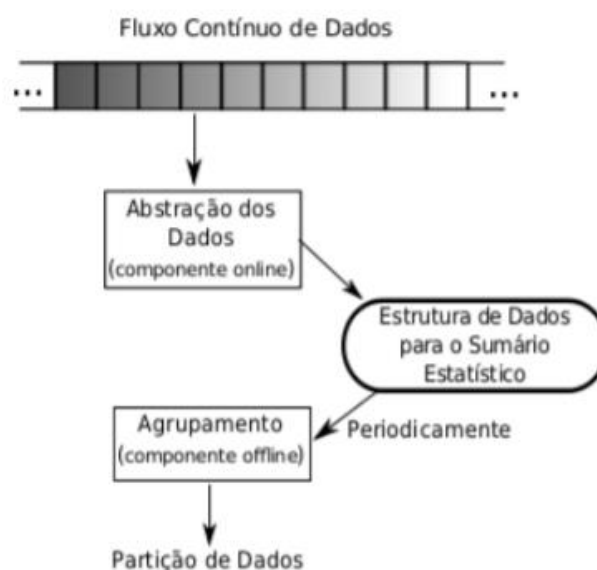


O cluster de fluxos de dados requer um processo capaz de agrupar objetos continuamente dentro de restrições de memória e tempo (GAMA, 2010). Tendo essas restrições em mente, os algoritmos para agrupar fluxos de dados devem idealmente cumprir os seguintes requisitos (BABCOCK et al., 2002; BIFET, 2010):

- Fornecer resultados oportunos executando um processamento rápido e incremental de objetos de dados;
- Adaptar-se rapidamente à dinâmica dos dados, o que significa que os algoritmos devem detectar quando novos clusters podem aparecer ou outros desaparecerem;
- Escalar o número de objetos que chegam continuamente;
- Fornecer uma representação de modelo que não seja apenas compacta, mas que também não cresça com o número;
- Detectar rapidamente a presença de outliers e agir em conformidade;
- Lidar com diferentes tipos de dados, por exemplo, árvores XML, sequências de DNA, informações temporais e espaciais do GPS.

Os algoritmos de agrupamento de fluxos de dados podem ser resumidos em duas etapas principais (CAO et al., 2006), a saber: etapa de abstração de dados (também conhecida como componente on-line) e etapa de agrupamento (também conhecida como componente off-line), como ilustrado na Fig. 6.

Figura 6- Etapas de algoritmos de clusterização baseado em objetos

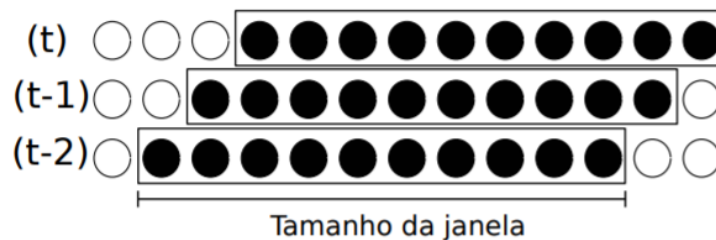


Fonte: ANDRADE (2016).

As janelas de tempo são utilizadas na etapa on-line com a finalidade de resumir o fluxo de dados e dar mais importância aos objetos atualizados. Existem três modelos comumente estudados em fluxos de dados (ZHU e SHASHA, 2002):

- **Modelo de Janela Deslizante:** Neste modelo apenas as informações mais recentes do fluxo de dados são armazenadas na estrutura de dados. Esta estrutura segue o sistema de organização do tipo FIFO (*First-In/First-Out*), onde o primeiro objeto adicionado à fila será o primeiro a ser removido, obedecendo a um intervalo específico de tempo. Na Fig. 7, apresentamos um exemplo do modelo de janela deslizante. O tamanho da janela depende do recurso computacional disponível.

Figura 7- Modelo de Janela Deslizante



Fonte: ANDRADE (2016).

- **Modelo de janela amortizada:** Diferentemente das janelas deslizantes, os objetos mais recentes recebem maior peso que objetos mais antigos, e o peso dos objetos diminui com o tempo. Um exemplo ilustrativo do modelo de janela amortecida é apresentado na Fig. 8, na qual o peso dos objetos decai exponencialmente de preto (mais recente) para branco (expirado).

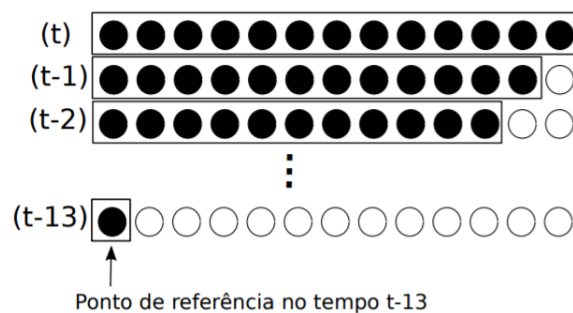
Figura 8- Modelo de Janela Amortecida



Fonte: ANDRADE (2016).

- Modelo de janela de referência: Os pontos de referência, que são objetos relevantes, podem ser definidos em termos de tempo (por exemplo, diariamente ou semanalmente) ou em termos do número de elementos observados desde o ponto de referência anterior. Quando um novo marco é atingido, todos os objetos mantidos na janela são removidos e os novos objetos do marco atual são mantidos na janela até que um novo marco seja alcançado. A Fig. 9 ilustra um exemplo de janela de referência.

Figura 9- Modelo de janela de referência



Fonte: ANDRADE (2016).

Na etapa de agrupamento *off-line*, os algoritmos de fluxo de dados obtêm uma partição dos dados, com base nos dados sumarizados pela etapa online. Um dos algoritmos mais utilizados na etapa *off-line* é o K-Médias (MACQUEEN,1967), devido à sua simplicidade e escalabilidade.

As aplicações de clusters em fluxo de dados são variadas, dentre elas podemos citar: análise de redes sociais, dados de sensores, preços de ações, entre outras. Na seção seguinte iremos entender o conceito de modelo evolutivo em fluxo de dados e usaremos, este modelo no problema de aprendizado incremental on-line não supervisionado para a classificação de dados médicos.

### 3 MODELOS EVOLUTIVOS A PARTIR DE FLUXOS DE DADOS

#### 3.1 Introdução

Os modelos evolutivos de sistemas estão sendo cada vez mais utilizados na classificação de dados stream, muitas áreas tais como: Sistemas IoT em Indústrias, Sensores, Aplicações Web (redes sociais), Sistemas anti-fraudes, biomedicina, e visão computacional já aplicam estes modelos. Os modelos evolutivos surgem como uma solução ao processamento de fluxo de dados no ambiente on-line, a fim de apoiar a análise de sistemas complexos que fazem uso destes dados, visto que estes são capazes de aprender de forma incremental, assimilam o conhecimento de forma rápida e requerem armazenamento mínimo de memória.

O principal desafio dos pesquisadores em reconhecimento de padrões consiste em classificar fluxo de dados usando modelos incrementais que adaptam ou evoluem suas estruturas e parâmetros baseado em possíveis novas informações contidas nos dados, o que denominados de desvio de conceito (LEITE et al., 2012).

Classificadores evolutivos buscam refinar o conhecimento existente sem recorrer à informação de um dado do passado, o principal fato consiste em como receber os novos dados de forma incremental mantendo o sistema em operação. Exemplos de classificadores evolutivos que têm sido discutidos na literatura incluem Growing Neural Gas (FRITZKE, 1995), Evolving Self Organizing Map (KASABOV, 2007), eClass (ANGELOV et al., 2007), Fuzzy Min-Max Neural Network FMM (SIMPSON, 1992).

Modelos evolutivos podem ser pensados em uma estrutura de três camadas, na primeira camada existem os modelos locais resultantes do agrupamento de dados. A partir da medida de similaridade, os dados são associados a um grupo existente ou formam um novo grupo. A segunda camada consiste de funções localmente válidas associadas aos grupos. Em problemas de classificação, a função é um valor constante (o rótulo de uma classe). A terceira camada consiste de um procedimento ou mecanismo para fusão de informação, i.e., as contribuições locais são agregadas para prover uma saída global (LEITE, 2014).

Nas seções a seguir iremos entrar em detalhes em dois modelos: Método de Clusterização Evolutiva (ECM) e Método evolutivo baseado em conjunto Fuzzy (FBEM).

### 3.2 Método de Clusterização Evolutiva – ECM

ECM é um algoritmo incremental usado para clusterização dinâmica e on-line de dados onde não há um número pré-definido de clusters (KASABOV et al., 2002).

No ECM os centros dos clusters são representados por nós evoluídos em um modo on-line. Inicia-se o processo de armazenamento em cluster com um conjunto vazio de clusters, após a chegada do primeiro dado, cria-se o primeiro cluster, localiza-se seu centro de cluster,  $C_c$  e seu raio de cluster,  $R_u$ , é inicialmente definido com um valor 0. Com a entrada de novos dados, alguns clusters já criados serão atualizados através da alteração de seus centros e aumentando seus raios de cluster.

A distância da amostra em relação ao centro do cluster, comparada com o valor limite  $D_{thr}$ , irá definir se o cluster deverá ser atualizado, deslocando a posição do centro do cluster ou se será necessário criar um novo cluster. Se a distância é maior que o valor limite, cria-se um novo cluster, caso contrário, o centro e o raio do cluster são ajustados no sentido de deslocar o cluster no domínio do dado de entrada, ou aumentar seu raio de alcance. O raio de um cluster não é mais atualizado quando ele atinge o limiar máximo de expansão definido pelo valor limite  $D_{thr}$ .

O algoritmo de aprendizado incremental ECM é representado a seguir em seis passos:

- Etapa 1: Crie o primeiro cluster  $C_1$  simplesmente tomando a posição do primeiro exemplo do fluxo de dados de entrada como o primeiro centro de cluster  $C_{c1}$  e definindo um valor 0 para seu raio de agrupamento  $R_{u1}$  (veja a Fig. 10a).
- Etapa 2: Se todos os exemplos do fluxo de dados tiverem sido processados, o processo de armazenamento em cluster será concluído. Caso contrário, o exemplo de entrada atual  $x_i$  é obtido e a distância Euclidiana normalizada  $d(i, j)$  entre este ponto da amostra e todos os  $n$  centros de clusters já criados  $C_{cj}$ , é calculada.

$$d(i, j) = \|x_i - C_{cj}\|_2 = \left( \sum_{i=1}^q |x_i - C_{cj}|^2 \right)^{\frac{1}{2}} \quad j = 1, 2, \dots, n \quad (3.2.1)$$

- Etapa 3: Se houver um cluster  $C_m$  com um centro  $C_{cm}$ , um raio do cluster  $R_u$ , o valor da distância  $d(i, m)$ , entre o centro do cluster  $C_{cm}$  e a amostra  $x_i$  é definida da seguinte forma:

$$d(i, m) = \min d(i, j) = \min(\|x_i - C_{cj}\|),$$

$$\text{para } j = 1, 2 \dots n \text{ e } d(i, m) \leq R_{um} \quad (3.2.2)$$

Para o valor de  $d(i, m) \leq R_{um}$ , considera-se que a amostra atual  $x_i$  pertence ao cluster  $C_m$ . Nesse caso, nem um novo cluster é criado, nem nenhum cluster existente é atualizado (por exemplo, vetores de dados  $x_4$  e  $x_6$  na Fig. 10). O algoritmo retorna à Etapa 2. Caso contrário, direciona-se para a etapa 4.

- Etapa 4: Encontre um cluster  $C_a$ , com seu centro  $C_{ca}$ ; raio do cluster  $R_{ua}$  e o valor da distância  $d(i, a)$  de todos os  $n$  clusters existentes através do cálculo dos valores da distância estendida:

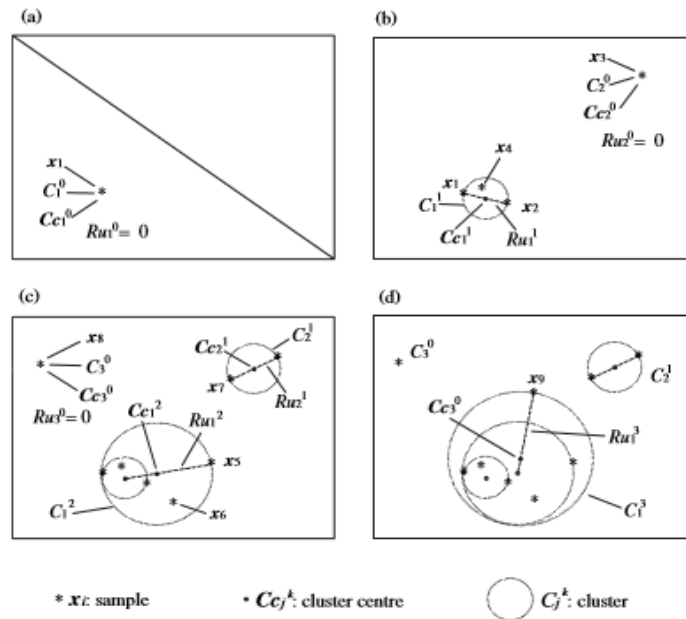
$$s(i, j) = d(i, j) + R_{uj}, j = 1, 2, \dots, n \quad (3.2.3)$$

e, em seguida, selecionando o cluster  $C_a$  com o valor mínimo  $s(i, a)$ :

$$s(i, a) = d(i, a) + R_{ua} = \min s(i, j), j = 1, 2, \dots, n. \quad (3.2.4)$$

- Etapa 5: Se  $s(i, a) > 2D_{thr}$ , a amostra  $x_i$  não pode pertencer a nenhum cluster existente. Um novo cluster deve ser criado da mesma maneira descrita na Etapa 1 (por exemplo, vetores de dados de entrada  $x_3$  e  $x_8$  na Fig. 10). O algoritmo retorna à Etapa 2. Caso contrário,
- Etapa 6: Se  $s(i, a) \leq 2D_{thr}$ , o cluster  $C_a$  é atualizado movendo seu centro,  $C_{ca}$ , e aumentando seu valor de raio,  $R_{ua}$ . O raio atualizado  $R_{uaneu}$  é definido como  $s(i, a) / 2$  e o novo centro  $C_{caneu}$  está localizado na linha que conecta o vetor de entrada  $x_i$  e o antigo centro de cluster  $C_{ca}$ , de modo que a distância do novo centro  $C_{caneu}$  ao ponto de amostra  $x_i$  é igual a  $R_{uaneu}$  (por exemplo, pontos de dados de entrada  $x_2$ ,  $x_5$ ,  $x_7$  e  $x_9$  na Fig. 10). O algoritmo retorna à Etapa 2. Dessa maneira, a distância máxima de qualquer centro de cluster até a amostra mais distante que pertence a esse cluster é mantida dentro do valor limite  $D_{thr}$ , embora o algoritmo não mantenha nenhuma informação das amostras passadas.

Figura 10- Um processo de clustering em evolução usando o ECM com exemplos consecutivos de  $x_1$  a  $x_9$  em um espaço 2D



Fonte: KASABOV et al., (2002).

A amostra  $x_1$  faz com que o ECM crie um novo cluster  $C_1^0$

$x_2$  : Atualiza o cluster  $C_1^0 \rightarrow C_1^1$ ;

$x_3$  : Cria um novo cluster  $C_2^0$ ;

$x_4$  : Não faz nada

$x_5$  : Atualiza o cluster  $C_1^1 \rightarrow C_1^2$ ;

$x_6$  Não faz nada;

$x_7$  : Atualiza o cluster  $C_2^0 \rightarrow C_2^1$ ;

$x_8$  : Cria um novo cluster  $C_3^0$

$x_9$  para atualizar o cluster  $C_1^2 \rightarrow C_1^3$ .

### 3.3 Modelos evolutivos baseados em conjunto Fuzzy – FBeM

FBeM é baseado em aprendizado incremental e fornece fronteiras não-lineares de separação de classes (LEITE et al., 2013). O método não requer dados *a priori* para aprender. Regras e grânulos de informação fuzzy são criados dinamicamente e adaptados ao longo do tempo (LEITE et al., 2011). Para cada grânulo existe uma regra correspondente. A parte

antecedente das regras FBeM consiste de hiper-retângulos fuzzy, e a parte consequente é uma classe basicamente.

Teorias e metodologias que utilizam grânulos para resolver problemas apresentados fornecendo uma enorme quantidade de dados, informação e conhecimento rotulam uma nova área de estudo multidisciplinar chamada computação granular (BARGIELA et al., 2002).

As unidades de processamento elementar nos sistemas granulares em evolução são conhecidas como grânulos de informação. Um grânulo define um subconjunto de um conjunto universal que contém uma representação. Uma estrutura granular é uma família de grânulos fechado por união (BARGIELA et al., 2002).

### 3.3.1 Estrutura e Processo

Modelos baseados em FBeM são representados através de uma combinação entre: função granular (grânulos de informação) e descrição linguística (regras *If-Then* extraídas do fluxo de dados). As regras criadas são responsáveis por gerenciar os grânulos de informação. Para cada grânulo de informação, existe uma regra correspondente.

As Regras  $R_i$  responsáveis por gerenciar os grânulos de informação  $\gamma_i$  são as seguintes:

$$R_i(x): \text{Se } (x_i \text{ é } (A_1)^i) E \dots E (x_n \text{ é } (A_n)^i). \text{ Então } (Y \text{ é classe } z_k) \quad (3.3.1.1)$$

Temos:

- $x_n$ : Variável de entrada;
- $Y$ : a classe estimada; A classe verdadeira é conhecida somente após a estimacão  $[y]$ ;
- $A_j^i$ : Funções de pertinências trapezoidais construídas à medida que os dados estão disponíveis;

As regras no FBeM são criadas e adaptadas sob demanda sempre que a estrutura dos dados exigir melhorias no modelo atual. Uma vez que estamos trabalhando com um problema de classificação. A equação que representa a regra consequente é só um valor, uma constante, representando a classe. Cada regra antecedente  $A_j^i$  está associada a uma classe, 1 ou 0, no respectivo trabalho. A classe relativa à regra mais ativa para uma amostra  $x[h]$  é a saída global do modelo. A regra mais ativa é dada por:

$$\alpha^* = S(\alpha^i, \dots, \alpha^i, \dots, \alpha^c) \quad (3.3.1.2)$$



Onde considera-se o operador máximo,  $\max x_i$ , para qualquer  $i$ , como co-norma triangular S. Além disso:

$$\alpha^i = T(A_1^i, \dots, A_j^i, \dots, A_n^i) \quad (3.3.1.3)$$

Onde adotou-se o operador de agregação mínimo,  $\min x_i$  para qualquer  $i$  como norma triangular T. Considerando a amostra de entrada  $x_j$  com  $n$  atributos dados de  $x_i = \{x_1, x_2, x_3, \dots, x_N\}$  e considerando que os trapézios  $A_1^i$  que montam o antecedente da  $i$ -ésima regra fuzzy são  $n$  um em cada dimensão de  $x$ .

- $X_1$  aciona o trapézio  $A_1^i$  com grau de pertinência  $M_{u1}$ ;
- $X_2$  aciona o trapézio  $A_2^i$  com grau de pertinência  $M_{u2}$ ;
- $X_3$  aciona o trapézio  $A_3^i$  com grau de pertinência  $M_{u3}$ .

O grau de ativação da  $i$ -ésima regra  $A_j^i$  é dado pela T norma, considerando o operador mínimo:

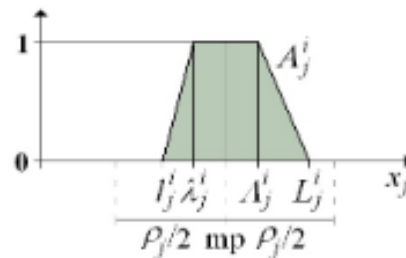
$$Tnorma(M_{u1}, M_{u2}, M_{u3} \dots M_{un}) = (M_{u1}, M_{u2}, M_{u3} \dots M_{un}) \quad (3.3.1.4)$$

Significando que a amostra  $n$  é  $Tnorma(M_{u1}M_{u2}, M_{u3} \dots M_{un})$  pertencente ao grânulo  $i$ .

Considere uma função de associação trapezoidal  $A_j^i = (l_j^i, \lambda_j^i, A_j^i, L_j^i)$  conforme ilustrado na Fig. 11.

$A_j^i$  é um subconjunto fuzzy, usado para modelar dados imprecisos sobre  $x_j$ . O subconjunto  $A_n^i = (A_1^i, \dots, A_j^i, \dots, A_n^i)$  montam um grânulo de informações  $\gamma_i$  governado por uma regra  $R_i$ . Para cada regra antecedente  $A_i$ , existe uma classe associada.

Figura 11- Função de associação trapezoidal



Fonte: LEITE et al., (2011).

Denotamos o apoio e o núcleo de uma função de associação trapezoidal  $A_j^i$  como:

$$\text{supp} (A_j^i) = [l_j^i, L_j^i]; \quad (3.3.1.5)$$

$$\text{núcleo} (A_j^i) = [\lambda_j^i, \Lambda_j^i] \quad (3.3.1.6)$$

A largura de uma função de associação  $A_j^i$  é o comprimento de seu suporte, a saber:

$$\text{wdt} (A_j^i) = L_j^i - l_j^i \quad (3.3.1.7)$$

Suponha que  $\rho_j$ , sejam a largura máxima que a função de associação  $A_j^i$  de um grânulo pode ter no espaço de entrada. Diferentes valores de  $\rho$  produzem diferentes representações do fluxo de dados em diferentes níveis de granularidade.

O ponto médio de um subconjunto  $A_j^i$  é:

$$\text{mp} (A_j^i) = \frac{\lambda_j^i + \Lambda_j^i}{2} \quad (3.3.1.8)$$

Sendo a região de expansão de  $A_j^i$  denotada por:

$$E_j^i = [\text{mp}(A_j^i) - \frac{\rho_j}{2}, \text{mp}(A_j^i) + \frac{\rho_j}{2}] \quad (3.3.1.9)$$

A região  $E_j^i$  auxilia na determinação de quando uma amostra  $x^h$  deve ser considerada ou não como pertencente a um grânulo já existente. Para dados normalizados, o valor de  $\rho$  varia em  $[0,1]$ .

Se  $\rho = 0$ , então o grânulo de FBEM não é capaz de se expandir. Portanto, o processo de aprendizado cria uma regra para cada amostra de dados. Se  $\rho = 1$ , então cria-se um único grânulo para representar todos os dados, o que não é interessante em ambiente não-estacionário.

O procedimento utilizado para evoluir  $\rho$  ao longo do tempo, depende da diferença entre o número de grânulos atual e de iterações passadas, comparado com uma taxa de crescimento específica. Seja  $r$  a diferença entre o número de grânulos atual e o número de grânulos a  $h_r$  iterações passadas:

$$r = c[h] - c[h - hr] \quad (3.3.1.10)$$

Se a quantidade de grânulos aumenta mais rápido do que uma taxa de crescimento  $\eta$ , ou seja,  $r > \eta$ , então  $\rho$  é aumentado conforme:

$$\rho^{[h]} = (1 + r/hr)\rho^{[h-hr]} \quad (3.3.1.11)$$

A granularidade  $\rho$  cresce a uma taxa menor do que  $\eta$ , ou seja,  $r < \eta$ , então  $\rho$  é reduzido:

$$\rho^{[h]} = (1 - (\eta - r)/hr)\rho^{[h-hr]} \quad (3.3.1.12)$$

Reduzindo a largura máxima permitida (reduzido o valor de  $\rho$ ) para grânulos requer a redução de grânulos maiores para ajustá-los ao novo valor. Nesse caso, o suporte de um conjunto  $A_j^i$  é estreitado como segue:

$$\text{SE: } mp(A_j^i) - \frac{p_{jnew}}{2} > l_j^i \text{ Então: } l_{jnew}^i = mp(A_j^i) - \frac{p_{jnew}}{2} \quad (3.3.1.13)$$

$$\text{SE: } mp(A_j^i) + \frac{p_{jnew}}{2} < L_j^i \text{ Então: } L_{jnew}^i = mp(A_j^i) + \frac{p_{jnew}}{2} \quad (3.3.1.14)$$

### 3.3.2 Criando Grânulos

Não é necessário a criação de regras antes do início do aprendizado. O processo incremental cria as regras quando um dado de entrada  $x = (x_1, \dots, x_n)$  não pertence à região de expansão  $(E_1^i, \dots, E_n^i), i = 1, \dots, c$ . Nesse caso,

$$A_j^{c+1} = (l_j^{c+1}, \lambda_j^{c+1}, A_j^{c+1}, L_j^{c+1}) = (x_j, x_j, x_j, x_j), \quad j = 1, \dots, n. \quad (3.3.2.1)$$

### 3.3.3 Adaptando Grânulos

A adaptação dos grânulos expande ou contrai o suporte e o núcleo dos conjuntos fuzzy  $A_j^i$ , um grânulo é escolhido para ser adaptado sempre que uma instância do fluxo de dados cai dentro de sua região de expansão. Em situações em que dois ou mais grânulos são qualificados para incluir os dados, a adaptação de apenas um dos grânulos é suficiente. Uma medida de similaridade útil para trapézios é:

$$S(x, A^i) = 1 - \frac{1}{4n} \sum |x_j - l_j^i| + |x_j - \lambda_j^i| + |x_j - A_j^i| + |x_j - L_j^i| \quad (3.3.3.1)$$

Essa medida quantifica o grau em que os dados de entrada correspondem o conhecimento atual. Retorna 1 para trapézios idênticos e diminui linearmente quando  $x$  e  $A^i$  se afastam. Naturalmente, entre todos os grânulos qualificados para acomodar uma amostra específica, aquela com maior similaridade deve ser escolhida. Este procedimento evita conflito e ajuda a manter a construção do FBeM simples.

A adaptação de parâmetros ocorre quando  $x \in E^i$  para algum  $i$ . Existem seis possibilidades:

- Se  $x_j^{[h]} \in [mp(A_j^i) - \frac{\rho_j}{2}, l_j^i]$ , então  $l_j^i(new) = x_j^{[h]}$  (expansão de suporte)

- Se  $x_j^{[h]} \in [l_j^i, \lambda_j^i]$  então  $\lambda_j^i$  (novo) =  $x_j^{[h]}$  (expansão do núcleo)
- Se  $x_j^{[h]} \in [\lambda_j^i, mp(A_j^i)]$  então  $\lambda_j^i$  (novo) =  $x_j^{[h]}$  (contração do núcleo)
- Se  $x_j^{[h]} \in [mp(A_j^i), \Lambda_j^i]$  então  $\Lambda_j^i$  (novo) =  $x_j^{[h]}$  (contração do núcleo)
- Se  $x_j^{[h]} \in [\Lambda_j^i, L_j^i]$  então  $\Lambda_j^i$  (novo) =  $x_j^{[h]}$  (expansão do núcleo)
- Se  $x_j^{[h]} \in [L_j^i, mp(A_j^i) + \frac{\rho_j}{2}]$  então  $L_j^i$  (novo) =  $x_j^{[h]}$  (expansão de suporte)

Operações nos parâmetros principais  $\lambda_j^i$  e  $\Lambda_j^i$  requer ajuste adicional do ponto médio do grânulo:

$$mp(A_j^i) \text{ (novo)} = \frac{(\lambda_j^i \text{ (novo)} + \Lambda_j^i \text{ (novo)})}{2} \quad (3.3.3.2)$$

Como resultado, a contração do suporte pode ocorrer em duas ocasiões:

- Se  $mp(A_j^i) \text{ novo} - \frac{\rho_j}{2} > l_j^i$ , então  $l_j^i$  (novo) =  $mp(A_j^i) \text{ novo} - \frac{\rho_j}{2}$
- Se  $mp(A_j^i) \text{ novo} + \frac{\rho_j}{2} > L_j^i$ , então  $L_j^i$  (novo) =  $mp(A_j^i) \text{ novo} + \frac{\rho_j}{2}$

### 3.3.4 Ampliando o modelo granular

As relações entre grânulos podem ser fortes o suficiente para justificar a montagem de um grânulo mais abstrato que herda a informação de grânulos de nível inferior. A medida de similaridade pode ser usada para quantificar a semelhança grânulo-grânulo se nós reafirmamos como:

$$S(A^{i1}, A^{i2}) = 1 - \frac{1}{4n} \sum_{j=1}^n |l_j^{i1} - l_j^{i2}| + |\lambda_j^{i1} - \lambda_j^{i2}| + |\Lambda_j^{i1} - \Lambda_j^{i2}| + |L_j^{i1} - L_j^{i2}| \quad (3.3.4.1)$$

FBeM combina grânulos no intervalo de  $hr$  passos considerando a menor entrada de  $S(A^{i1}, A^{i2})$  como um critério de decisão. A decisão pode ser com base se o novo grânulo obedece a largura máxima permitida  $\rho$ .

O novo grânulo  $\gamma_i$ , é formado por parâmetros derivados de  $A_j^{i1}$  e  $A_j^{i2}$  como segue:

$$l_j^i = \min(l_j^{i1}, l_j^{i2}) \quad (3.3.4.2)$$

$$\lambda_j^i = \min(\lambda_j^{i1}, \lambda_j^{i2}) \quad (3.3.4.3)$$

$$\Lambda_j^i = \max(\Lambda_j^{i1}, \Lambda_j^{i2}) \quad (3.3.4.4)$$

$$L_j^i = \max(L_j^{i1}, L_j^{i2}) \quad (3.3.4.5)$$

Procedimentos para remoção e mescla de grânulos e regras FBeM podem ser encontrados em (LEITE et al., 2012). O aprendizado FBeM é sumarizado a seguir:

- 1- Determinar  $\rho, hr, \eta$ ;
- 2- Para  $h = 1, \dots$ 
  - a. Receber amostra  $x^{[h]}$ ;
  - b. Prover classe estimada  $C^{[h]}$ ;
- 3- A classe real se torna disponível
- 4- Calcular erro de estimação;
- 5- Adaptação do modelo
- 6- Se  $x^{[h]} \notin E^i \forall i$ 
  - a. Criar novo grânulo  $\gamma^{[c+1]}$ ;
- 7- Senão
  - a. Adaptar grânulo mais ativo  $\gamma^i$ ;
- 8- Fim
- 9- Remover amostra  $x^{[h]}$ ;
- 10- Se  $h = \beta hr, \beta = 1, \dots$ 
  - a. Mesclar grânulos existentes;
  - b. Atualizar a granularidade  $\rho$ ;
  - c. Remover grânulos inativos
- 11- Fim
- 12- Fim

### 3.4 Correlação de Spearman para Seleção de Atributos

O coeficiente  $\rho$  de Spearman é usado para calcular o grau de relacionamento entre duas ou mais variáveis. Através do cálculo da correlação, apenas as variáveis relevantes são utilizadas, eliminando a chance de ocorrer overfitting e reduzindo a complexidade do problema.

Este coeficiente usa a ordem das observações, chamada de posto, se duas variáveis possuem a mesma medida, o valor da média delas é atribuído ao posto. Deste modo, este

coeficiente não é sensível a assimetrias na distribuição, nem à presença de outliers, não exigindo, portanto, que os dados provenham de duas populações normais, ao contrário da correlação de Pearson que assume que os dados são normalmente distribuídos (GUIMARÃES, 2013).

O coeficiente  $\rho$  de Spearman é mais apropriado, nos casos em que os dados não formam uma nuvem “bem-comportada”, com alguns pontos muito afastados dos restantes, ou em que parece existir uma relação crescente ou decrescente em formato de curva. Uma fórmula fácil para calcular o coeficiente  $\rho$  de Spearman é dada por:

$$\rho = 1 - 6 \sum_{i=1}^n (d_i^2) / (n^3 - n) \quad (3.4.1)$$

em que  $n$  é o número de pares  $(x_i, y_i)$  e  $d_i =$  (postos de  $x_i$  dentre os valores de  $x$ ) - (postos de  $y_i$  dentre os valores de  $y$ ).

O coeficiente  $\rho$  de Spearman varia entre -1 e 1. Quanto mais próximo estiver destes extremos, maior será a associação entre as variáveis. O sinal negativo da correlação significa que as variáveis variam em sentido contrário, isto é, as categorias mais elevadas de uma variável estão associadas a categorias mais baixas da outra variável. Quanto mais próximo de 0 é o valor do coeficiente de Spearman mais importante é o atributo. A ideia é que um atributo interessante possua baixa correlação com outros atributos, o que implica que este transmite informações únicas e diferentes. Como a velocidade de processamento é essencial em ambiente online, o método de seleção de atributos descrito nesta seção é útil para reduzir o problema de dimensionalidade (HUEI, 2005).

### 3.5 PCA – Redução de Dimensionalidade

A técnica de PCA, ou análise de componentes principais, é utilizada para redução de dimensionalidade dos dados numéricos.

A PCA consta de um procedimento algébrico que converte as variáveis originais (que são tipicamente correlacionadas) num conjunto de variáveis não correlacionadas (linearmente) que se designam por componentes principais (PCs) ou variáveis latentes. Assim, a PCA fornece um mapeamento de um espaço com  $N$  dimensões (em que  $N$  é o número de variáveis originais) para um espaço com  $M$  dimensões (onde  $M$  é tipicamente muito menor do que  $N$ ) (ROSSI, 2017).

Matematicamente, na PCA, a matriz  $\mathbf{X}$  (dados originais) é decomposta em um produto de duas matrizes, denominadas escores ( $\mathbf{T}$ ) e pesos ( $\mathbf{P}$ ), mais uma matriz de erros ( $\mathbf{E}$ ), como mostrado a seguir:

$$X = TP^T + E \quad (3.5.1)$$

Os escores representam as coordenadas das amostras no sistema de eixos formados pelos componentes principais. Cada componente principal é constituído pela combinação linear das variáveis originais e os coeficientes da combinação são denominados pesos. A primeira componente principal (PC1) é traçada no sentido da maior variação no conjunto de dados; a segunda (PC2) é traçada ortogonalmente à primeira, com o intuito de descrever a maior porcentagem da variação não explicada pela PC1 e assim por diante; enquanto os escores representam as relações de similaridade entre as amostras. A avaliação dos pesos permite entender quais variáveis mais contribuem para os agrupamentos observados no gráfico dos escores. Através da análise conjunta do gráfico de escores e pesos, é possível verificar quais variáveis são responsáveis pelas diferenças observadas entre as amostras (SOUZA et al., 2012).

O número de componentes principais a ser utilizado no modelo PCA é determinado pela porcentagem de variância explicada. Assim, seleciona-se um número de componentes de tal maneira que a maior porcentagem da variação presente no conjunto de dados originais seja capturada.

#### 4 METODOLOGIA

Analisando do ponto de vista do problema, este trabalho pode ser classificado como uma pesquisa quantitativa, isto pelo fato de permitir a descrição detalhada dos fenômenos observados, com a realização de um estudo das questões com mais profundidade e facilitar a comparação e a análise estatística dos dados utilizados. Sob a ótica da natureza, esta pesquisa classifica-se como aplicada, visto que possui o intuito de auxiliar os médicos no diagnóstico precoce das doenças, por meio da Inteligência Artificial com aplicação de algoritmos evolutivos, ou seja, é motivada pela necessidade de resolver problemas concretos. Quanto aos meios, esta pesquisa classifica-se como sendo de laboratório, pois a mesma trata simulações realizadas em computador, com auxílio de softwares específicos. E por fim,

quanto aos objetivos a pesquisa é exploratória e explicativa, pois foi necessário realizar uma pesquisa bibliográfica para analisar os exemplos e compreender o problema a ser resolvido, e a pesquisa explicativa foi realizada para identificar os atributos necessários para a aplicação dos algoritmos, explicando a causa e efeito da ocorrência dos fenômenos envolvidos na classificação dos dados médicos.

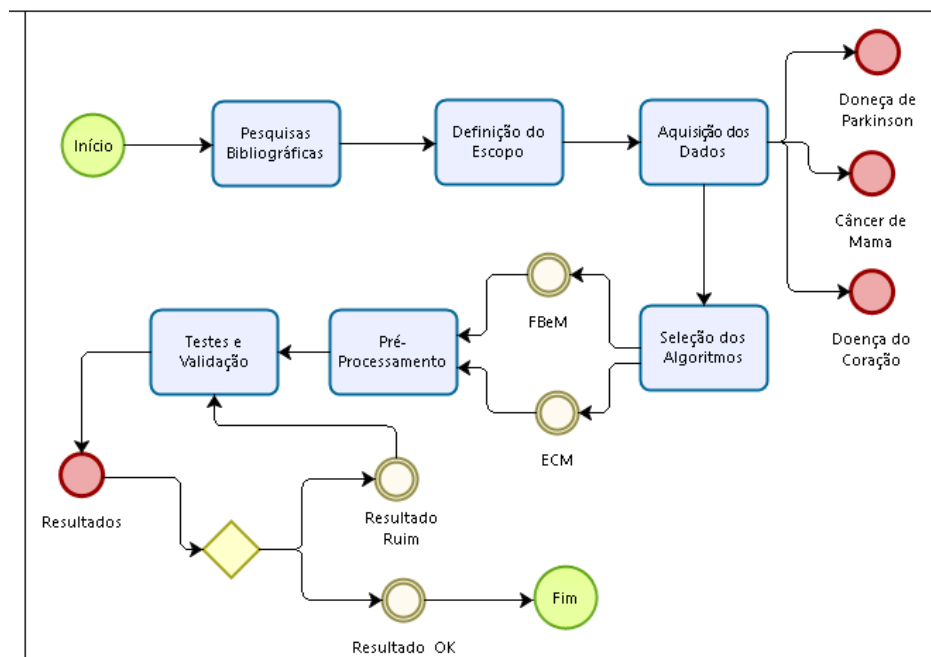
#### 4.1 Suposições e Hipóteses

Esta pesquisa pretende testar a seguinte hipótese: Através do uso de algoritmos evolutivos (ECM e FBeM), é possível aumentar a taxa de acertos (Acurácia) no diagnóstico das seguintes doenças: Parkinson, do Coração e Câncer de Mama, quando comparado com outros algoritmos de classificação encontrados na literatura.

#### 4.2 Fluxo de Trabalho

Para que o projeto fosse desenvolvido de forma sistemática e organizada foi criado e respeitado um fluxo simplificado de trabalho, como pode ser observado na Fig. 12.

Figura 12- Fluxograma: Elaboração do Projeto.



Fonte: O autor.



O fluxograma apresentado na imagem acima descreve de forma resumida o ciclo de vida da proposta em questão. Logo é possível entender cada etapa do processo, desde a concepção da ideia, até disponibilização da ferramenta para testes em campo.

### **4.3 Base de Dados**

As bases de dados benchmark Câncer de Mama, Doença do Coração e Telemonitoramento da Doença de Parkinson, utilizadas neste trabalho foram obtidas do repositório de aprendizado de máquina da UCI - Irvine, Universidade da Califórnia, Escola de Informação e Ciência da Computação.

#### **4.3.1 Doença do Coração**

Os criadores do conjunto de dados de doenças cardíacas foram: Instituto Húngaro de Cardiologia. Budapeste: Andras Janosi, MD; Hospital Universitário, Zurique, Suíça: William Steinbrunn, MD; Hospital Universitário, Basileia, Suíça: Matthias Pfisterer, MD; Centro Médico VA, Long Beach e Cleveland Clinic Foundation: Robert Detrano, MD, Ph.D. O banco de dados contém 76 atributos e 303 dados individuais, todos os trabalhos publicados utilizam apenas um subconjunto de 14 atributos que são:

- Idade: Idade em anos;
- Sexo: 1 = Homem; 0 = Mulher;
- Cp: Tipo de dor no peito; - Valor 1: angina típica ; - Valor 2: angina atípica; - Valor 3: dor não anginal; - Valor 4: assintótica;
- Trestbps assintomáticos: Pressão arterial em repouso (em mmHg na admissão no hospital);
- Chol: Exibe o colesterol sérico em mg / dl (unidade);
- Fbs: Açúcar no sangue em jejum > 120 mg / dl - (1 = verdadeiro; 0 = falso);
- Restecg: Resultados eletrocardiográficos em repouso: Valor 0: Normal; 1 = com anormalidade da onda ST-T; 2 = hipertrofia ventricular esquerda;
- Thalach: Frequência cardíaca máxima alcançada por um indivíduo;
- Exang: Angina induzida pelo exercício (1 = sim; 0 = não);

- Oldpeak: Depressão do ST induzida pelo exercício em relação ao repouso (ano da leitura do ECG do exercício);
- Inclinação: A inclinação do pico do exercício segmento ST: - Valor 1: subida; - Valor 2: plano; - Valor 3 descida;
- Ca: número de vasos principais (0-3) coloridos por flourosopy;
- Thal: Exibe a talassemia - 3 = normal; 6 = defeito fixo; 7 = defeito reversível;
- Num: Diagnóstico de doença cardíaca (status da doença angiográfica); exibe se o indivíduo está sofrendo de doença cardíaca ou não:0 = ausência, 1= presente.

#### 4.3.2 Câncer de Mama

Os Criadores deste banco de dados são: Dr. William H. Wolberg e os Hospitais da Universidade de Wisconsin em Madison, Wisconsin, EUA.

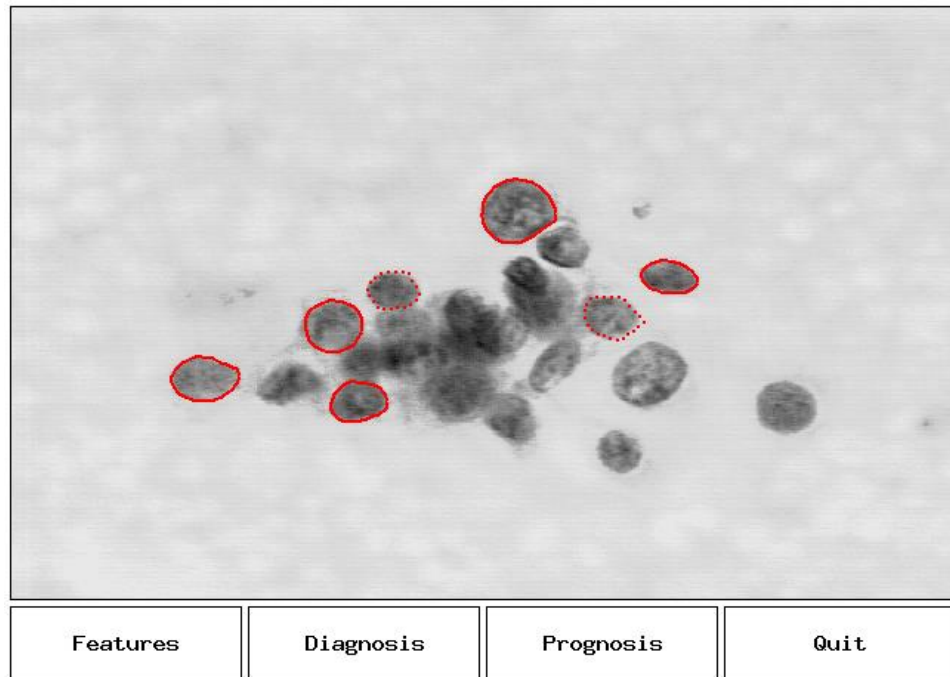
Os atributos utilizados do banco de dados câncer de mama são calculados a partir de uma imagem digitalizada de um aspirado por agulha fina (PAAF) de uma massa mamária. Os atributos são obtidos por meio do seguinte procedimento.

Um PAAF é retirado da massa mamária. Este material é então montado em uma lâmina de microscópio e corado para destacar os núcleos celulares. Uma parte do slide em que as células são bem diferenciadas é digitalizada usando uma câmera digital e uma placa de captura de quadro.

O usuário então isola os núcleos individuais usando o software **Xcyt** . Usando um ponteiro do mouse, o usuário desenha o limite aproximado de cada núcleo. Utilizando uma abordagem de visão computacional conhecida como "serpentes", essas aproximações convergem para os limites nucleares exatos. Esse processo interativo leva entre dois e cinco minutos por slide. A Fig. 13 mostra o **Xcyt** em uso.

Uma vez que todos (ou a maioria) dos núcleos foram isolados nesse fasion, o programa calcula valores para cada uma das dez características de cada núcleo, medindo tamanho, forma e textura. A média, erro padrão e valores extremos desses recursos são calculados, resultando em um total de 30 recursos nucleares para cada amostra.

Figura 13- Núcleos individuais das células mamárias capturadas pelo *software* Xcyt



Fonte: WILLIAM et al., (1990).

O resultado do estudo de Willian et al., (1990) através da análise estatística mostrou que as nove características, excluindo o número de identificação, a seguir diferiram significativamente entre amostras benignas e malignas do câncer de mama, representando os atributos a serem utilizado no presente estudo.

- Número do código da amostra: número de identificação
- Espessura do grupo (massa mamária): 1 – 10
- Uniformidade do tamanho da célula: 1 – 10
- Uniformidade do formato da célula: 1 – 10
- Adesão marginal: 1 – 10
- Tamanho da célula epitelial única: 1 – 10
- Núcleos desencapados: 1 – 10
- Cromatina branda: 1 - 10
- Núcleos normais: 1 – 10
- Mitoses: 1 – 10
- Classe: (0 para benignos, 1 para malignos)

### 4.3.3 Doença de Parkinson

O conjunto de dados foi criado por Max Little, da Universidade de Oxford, em colaboração com o Centro Nacional de Voz e Fala, Denver, Colorado, que registrou os sinais de fala. O Oxford Parkinson's Disease Center (OPDC) é um centro de pesquisa multidisciplinar exclusivo da Universidade de Oxford, o centro de pesquisa tem como objetivo: Entender a progressão da doença de Parkinson; Prever o início da doença de Parkinson; identificar possíveis alvos de medicamentos para o Parkinson e desenvolver novos tratamentos que impedirão o desenvolvimento de Parkinson em indivíduos em risco.

O conjunto de dados é composto por uma série de medidas biomédicas de voz de 31 pessoas, 23 com doença de Parkinson (DP). Cada coluna da tabela é uma medida de voz específica e cada linha corresponde a uma das 195 gravações de voz desses indivíduos. O principal objetivo dos dados é discriminar pessoas saudáveis daquelas com DP, de acordo com a coluna "status", que é definida como 0 para saudável e 1 para DP. A seguir temos informações dos atributos:

- Nome: nome do assunto ASCII e número de gravação;
- MDVP: Fo (Hz) - frequência fundamental vocal média;
- MDVP: Fhi (Hz) - frequência fundamental vocal máxima;
- MDVP: Flo (Hz) - frequência fundamental vocal mínima;
- MDVP: Tremulação (%), MDVP: Tremulação (Abs), MDVP: RAP, MDVP: PPQ e Tremulação: DDP: São 5 características de medidas de variação da frequência fundamental;
- MDVP: Shimmer, MDVP: Shimmer (dB), Shimmer: APQ3, Shimmer: APQ5, Shimmer: APQ11, Shimmer: DDA: São 6 características de medidas de variação da amplitude da fonação sustentada em certas frequências;
- NHR e HNR: Duas medidas de razão de ruído para componentes tonais no status da voz;
- Status de saúde do sujeito: 1 – Paciente com Parkinson, 0 – Paciente Saudável;
- RPDE- Duas medidas de complexidade dinâmica não linear
- DFA - Expoente de escalonamento fractal de sinais
- Spread1, Spread2, PPE - Três medidas não lineares de variação de frequência fundamental.

#### 4.4 Pré-processamento

O pré-processamento é essencial para a análise dos dados, sem esta fase o algoritmo não apresentará uma classificação adequada. As principais etapas do pré-processamento são:

- **Limpeza:** Retirar registros incompletos, valores errados e dados inconsistentes, de modo a não influenciar no resultado dos algoritmos utilizados;
- **Redução:** Devido ao alto volume de dados algumas estratégias são adotadas para reduzir a quantidade de dados, tais como, seleção de atributos e redução de dimensionalidade;
- **Transformação:** Alguns algoritmos utilizam valores numéricos e outros nominais, e nestes casos é necessário transformar os valores numéricos em nominais ou ao contrário, e em alguns algoritmos os dados precisam estar em um intervalo especificado. As técnicas empregadas nesta etapa são: suavização, agrupamento e normalização.

No presente trabalho utilizamos as etapas do pré-processamento como forma de melhorar a assertividade dos modelos evolutivos.

Na etapa de transformação os valores dos dados foram normalizados no intervalo de 0 a 1, por meio da seguinte equação:

$$Min - Max : n_i = \frac{(x_i - \min(x))}{(x) - \min(x)} \quad (4.4.1)$$

Na fase de redução dos dados foi utilizado a técnica de correlação de Spearman e a redução de dimensionalidade PCA, mencionadas nas seções 3.4 e 3.5 respectivamente.

## 4.5 Métricas de Avaliação dos Resultados

### 4.5.1 Medidas de Acurácia Diagnóstica

Os componentes da acurácia diante de um diagnóstico dicotômico (presença ou ausência da doença) são: sensibilidade e especificidade. Sensibilidade é a capacidade do método em reconhecer os doentes, é calculada pela razão entre: casos doentes com resultado positivo (VP- verdadeiro positivo), sobre o total de doente que inclui os casos falso-negativo (FN).

$$\text{Sensibilidade} = \frac{VP}{(VP+FN)} \quad (4.4.1.1)$$

Enquanto a especificidade é a capacidade do método em reconhecer os saudáveis, sendo calculada utilizando-se apenas os indivíduos sem doença, como a razão entre os casos não doentes com resultado negativo (VN – verdadeiro negativo) sobre o total de não doentes que inclui os casos falso positivos (FP).

$$\text{Especificidade} = \frac{VN}{(VN+FP)} \quad (4.4.1.2)$$

As razões de probabilidade, são os parâmetros que combinam sensibilidade e especificidade.

A razão de probabilidade positiva é um número que representa o quanto um método de resultado positivo aumenta a chance de um indivíduo ser doente é expressa por:

$$Rp = \text{sensibilidade} / (1 - \text{especificidade}) \quad (4.4.1.3)$$

Quando mais alto este número, melhor ou seja: RP positiva: > 10 (acurácia ótima); 5-10 (acurácia moderada); 2-5 (acurácia pequena); 1-2 (acurácia nula). Já a razão de probabilidade negativa é representada por:

$$Rn = \left(1 - \frac{\text{sensibilidade}}{\text{especificidade}}\right) \quad (4.4.1.4)$$

Sendo o seu valor expresso por quanto um método de resultado negativo influencia a chance de um indivíduo ser saudável. Quanto mais próximo de zero, melhor: RP negativa: < 0.1 (acurácia ótima); 0.1-0.2 (acurácia moderada); 0.2-0.5 (acurácia pequena); 0.5-1.0 (acurácia nula).

A Eficácia de Diagnóstico (ED) é uma medida global de acurácia diagnóstica, expressa em proporção de indivíduos classificados corretamente entre todos.

$$ED = (VP + VN) / (VP + VN + FP + FN) \quad (4.4.1.5)$$

É afetada pela prevalência da doença. Com uma mesma sensibilidade e especificidade, a precisão do diagnóstico de uma determinada doença aumenta quando a prevalência da doença diminui (BORGES, 2016).

No presente trabalho iremos utilizar a matriz de confusão, para avaliar os modelos de classificação que utilizaram o algoritmo FBeM. A matriz de confusão funciona da seguinte forma, dependendo da quantidade de classes é criada uma matriz (N x N), por exemplo, imagine um modelo de classificação binária (0 e 1) no presente trabalho teremos uma matriz de confusão da dimensão 2 x 2, indicando as classes de verdadeiras como P(positivo) e N(negativo), e as classes previstas pelo modelo com o S(sim) e N(não). A imagem a seguir mostra um exemplo da matriz de confusão.

Tabela 1- Matriz de Confusão

	Positivo	Negativo
SIM	Positivos verdadeiros	Positivos falsos
NÃO	Negativos Falsos	Negativos verdadeiros

Fonte: Autor

Através da matriz de confusão iremos obter os valores de sensibilidade, especificidade, das razões de probabilidade e da eficácia do diagnóstico para as seguintes bases de dados: Câncer de Mama, Doença do Coração e Mal de Parkinson.

#### 4.5.2 Medidas de validação de clusters

Medidas de validação são geralmente usadas para determinar a quantidade ideal de clusters existentes em uma base de dados (BABUSKA et al., 2002).

Há dois principais índices de validação de agrupamentos: índices internos e externos. Os externos comparam a estrutura dos clusters descobertos com uma estrutura já conhecida, já os índices internos analisam a estrutura de cluster descoberta, com base em algum critério: compacidade ou separabilidade.

- Compacidade: Distância entre grupos são maximizadas.
- Separabilidade: Distância dentro dos grupos são minimizadas.

Neste trabalho adotou-se duas medidas de validação com índices internos, conforme descrito a seguir.

- Critério de validação de Xie-Beni: o número ideal de clusters é determinado segundo o valor mínimo para o critério de Xie e Beni (XB) (XIE et al., 1991):

$$XB(c) = \sum_{i=1}^c \sum_{j=1}^N \frac{(\mu_{ij}^m \|x_j - v_i\|^2)}{(N \min_{ij} \|x_j - v_i\|^2)} \quad (4.4.2.1)$$

O numerador da Eq. (4.4.2.1) refere-se à minimização da distância Euclidiana entre as amostras e os centros de clusters,  $\|\cdot\|$  é a norma 2. O denominador refere-se à maximização da distância entre centros de clusters. Pequenos valores de  $XB$  são esperados para clusters compactos e bem separados. Os valores de  $XB$  dependem da constante de fuzzicidade;  $m = 2$  é o padrão adotado.

- Coeficiente de Partição (PC): É definido segundo:

$$PC(c) = 1/N (\sum_{i=1}^c \sum_{j=1}^N (\mu_{ij}))^2 \quad (4.4.2.2)$$

O valor de PC mede a quantidade de sobreposição entre clusters (BEZDEK, 1981). Os valores variam em  $[1/c, 1]$ . Quanto mais próximo do valor 1 for o coeficiente PC, mais “nítido” será o cluster. Um valor próximo de  $1/c$  indica que não há tendência da existência de qualquer agrupamento na base de dados ou que o algoritmo em questão não conseguiu revelá-los. O número ideal de clusters gera o maior valor do coeficiente PC.

## 5 RESULTADOS

A seção de resultados será exibida de acordo com os algoritmos evolutivos utilizados no presente trabalho, para classificação das bases de dados: Câncer de Mama, Doença de Parkison e Doença do Coração.

Para o algoritmo FBeM às métricas de validação dos resultados, serão as citadas na seção 4.1 de acurácia diagnóstica, tais são: Sensibilidade, Especificidade, Razões de Probabilidade, Eficácia Diagnóstica, todas obtidas com o auxílio da matriz de confusão.

Já para o algoritmo ECM as métricas de validação serão os índices internos de compacidade e separabilidade, que são o Coeficiente de Partição (PC) e Critério de validação de Xie-Ben (XB) respectivamente, além da alteração do meta-parâmetro,  $D_{thr}$ , distância limite para criação de um novo cluster, queremos analisar se o aumento do meta-parâmetro influencia na quantidade de clusters criados e no tempo de processamento do algoritmo.



## 5.1 Algoritmo FBeM

### 5.1.1 Doença do Coração

A base de dados possui 13 atributos, aplicou-se o algoritmo FBeM considerando 75%, 50% e 25% destes atributos, mantendo fixos os meta-parâmetros ( $\rho, hr, \eta$ ) do algoritmo.

O método de correlação de Spearman foi utilizado na seleção dos atributos, os que possuíam baixa correlação foram escolhidos, uma vez que estes transmitem informações únicas e diferentes.

Os resultados obtidos são mostrados na Tabela 2, levando em consideração a acurácia do modelo, tempo de processamento e número de regras médio obtido com seu respectivo desvio padrão.

Tabela 2- Resultados FBeM doença do coração

Nº_Atributos	Acc(%)	Tempo(s)	Nº_Regras
10	99,09	108	9,65 ± 3,19
<b>7</b>	98,01	95	9,14 ± 2,89
3	98,35	74	3,27 ± 0,79

Fonte: Autor

A matriz de confusão abaixo, Fig.14, retrata o resultado do modelo considerando a utilização de três parâmetros, visto que para tais parâmetros o modelo apresentou acurácia elevada de 98,35% e o menor número de regras de 3,27 ± 0,79, os parâmetros utilizados foram:

- Trestbps assintomáticos: Pressão arterial em repouso (em mmHg na admissão no hospital);
- Chol: Exibe o colesterol sérico em mg / dl (unidade);
- Fbs: Açúcar no sangue em jejum > 120 mg / dl - (1 = verdadeiro; 0 = falso).

Figura 14- Matriz de Confusão Doença do Coração com 3 atributos

**Confusion Matrix**

Output Class	0	133 43.9%	0 0.0%	100% 0.0%
	1	5 1.7%	165 54.5%	97.1% 2.9%
		96.4% 3.6%	100% 0.0%	98.3% 1.7%
		0	1	
		Target Class		

Fonte: Autor.

Por meio da matriz de confusão, Fig.14, podemos calcular a sensibilidade, especificidade, razões de probabilidade e eficácia do diagnóstico. Lembrando que o diagnóstico da doença cardíaca é feito considerando: 0- Ausência da Doença e 1- Presença da doença.

$$\text{Sensibilidade} = \frac{VP}{VP + FN} = \frac{165}{165 + 5} = 0,97 \quad (5.1.1.1)$$

$$\text{Especificidade} = \frac{VN}{(VN + FP)} = \frac{133}{133 + 0} = 1 \quad (5.1.1.2)$$

$$Rp = \frac{0,97}{(1 - 1)} > \infty \quad (5.1.1.3)$$

$$Rn = (1 - \text{sensibilidade} / \text{especificidade}) = 1 - 0,97/1 = 0,03 \quad (5.1.1.4)$$

$$ED = (VP + VN) / (VP + VN + FP + FN) = 0,98 \quad (5.1.1.5)$$

Percebe-se a eficácia do modelo, em classificar tanto os doentes (sensibilidade) como os saudáveis (especificidade), apenas 5 casos de falsos negativos foram encontrados, como mostrados na matriz de confusão Fig.14. As razões de probabilidade positiva ( $Rp > 10$ ) e

negativa ( $Rn < 0,1$ ), além da eficácia do diagnóstico evidenciaram que a acurácia do modelo foi ótima para a base de dados Doença do Coração.

### 5.1.2 Doença de Parkinson

Considerando 75%, 50% e 25% dos atributos da doença de Parkinson, o método de correlação de Spearman foi utilizado para a escolha dos atributos com os valores mais baixos de correlação. O algoritmo FBeM foi aplicado mantendo fixo os meta-parâmetros ( $\rho, hr, \eta$ ).

Os resultados obtidos são mostrados na Tabela 3 levando em consideração a acurácia do modelo, tempo de processamento e o número de regras médio obtido com seu respectivo desvio padrão.

Tabela 3- Resultados FBeM Doença de Parkinson

Nº_Atributos	Acc(%)	Tempo(s)	Nº_Regras
18	88,21	94	6,75 ± 1,94
11	88,72	103	6,61 ± 1,91
6	89,74	91	6,11 ± 2,26

Fonte: Autor

A matriz de confusão, Fig.15, abaixo retrata o resultado do modelo considerando a utilização dos seis parâmetros abaixo, visto que para tais parâmetros o modelo apresentou acurácia elevada (89,74%) com o menor tempo de processamento (91s) e com o menor número de regras (6,11 ± 2,26).

- MDVP: Fo (Hz) – Frequência fundamental vocal média;
- MDVP: Fhi (Hz) – Frequência fundamental vocal máxima;
- MDVP:APQ – Medidas de variação da amplitude da fonação sustentada em certas frequências;
- NHR – Duas medidas de proporção de ruído para componentes tonais de voz;
- D2 – Medida da complexidade dinâmica não linear;
- Spread2 – Medida não linear da variação da frequência fundamental.

Figura 15 - Matriz de Confusão Doença de Parkinson com 6 atributos

		Confusion Matrix		
		0	1	
Output Class	0	43 22.1%	16 8.2%	72.9% 27.1%
	1	5 2.6%	131 67.2%	96.3% 3.7%
		89.6% 10.4%	89.1% 10.9%	89.2% 10.8%
		0	1	
		Target Class		

Fonte: Autor

Por meio da matriz de confusão, Fig.15, e das fórmulas abaixo, podemos calcular a sensibilidade, especificidade, razões de probabilidade e eficácia do diagnóstico da doença de Parkinson. A classificação dos pacientes com doença de Parkinson é obtida através das classes 1 e 0, sendo a classe 1: pacientes doentes e a classe 0: pacientes saudáveis.

$$\text{Sensibilidade} = \frac{VP}{(VP+FN)} = \frac{131}{131+5} = 0,96 \quad (5.1.2.1)$$

$$\text{Especificidade} = \frac{VN}{(VN+FP)} = \frac{43}{43+16} = 0,72 \quad (5.1.2.2)$$

$$R_p = \text{sensibilidade} / (1 - \text{especificidade}) = 0,96 / (1 - 0,72) = 3,42 \quad (5.1.2.3)$$

$$R_n = (1 - \text{sensibilidade} / \text{especificidade}) = 1 - 0,96/0,72 = 0,09 \quad (5.1.2.4)$$

$$ED = (VP + VN) / (VP + VN + FP + FN) = 0,8923 \quad (5.1.2.5)$$

Percebe-se a eficácia do modelo, em classificar os doentes (sensibilidade), com apenas 5 casos de falsos negativos, como mostrados na matriz de confusão Fig.15. As razões

de probabilidade positiva ( $2 < Rp < 5$ ) e negativa ( $Rn < 0,1$ ), além da eficácia do diagnóstico evidenciou que a acurácia do modelo foi moderada para a base de dados Doença de Parkinson.

### 5.1.3 Câncer de Mama

Para a base de dados Câncer de Mama, foi utilizado o método de redução de dimensionalidade PCA, temos nesta base de dados 31 atributos aplicamos o modelo considerando 75%, 50% e 25% dos atributos com maior porcentagem da variação presente no conjunto de dados originais, mantendo fixos os meta-parâmetros do algoritmo ( $\rho, hr, \eta$ ).

O resultado obtido é mostrado na Tabela 4 levando em consideração a acurácia do modelo, tempo de processamento e número de regras médio obtido com seu respectivo desvio padrão.

Tabela 4- Resultados FBeM Câncer de Mama

<b>Nº_Atributos</b>	<b>Acc(%)</b>	<b>Tempo(s)</b>	<b>Nº_Regras</b>
8	87,30	115	9,2 ± 3,74
23	91,04	121	7,1 ± 2,16
15	89,20	114	7,3 ± 1,92

Fonte: Autor

A matriz de confusão abaixo, Fig.16, retrata o resultado do modelo considerando a utilização de vinte e três parâmetros, visto que para tais parâmetros o modelo apresentou acurácia elevada com o tempo de processamento razoável, levando em consideração a quantidade elevada de atributos.

Figura 16- Matriz de Confusão para Câncer de Mama com 23 atributos.

		Confusion Matrix		
		0	1	
Output Class	0	344 60.5%	38 6.7%	90.1% 9.9%
	1	13 2.3%	174 30.6%	93.0% 7.0%
		96.4% 3.6%	82.1% 17.9%	91.0% 9.0%
		0	1	
		Target Class		

Fonte: Autor

Por meio da matriz de confusão, Fig.16, e das fórmulas abaixo, podemos calcular a sensibilidade, especificidade, razões de probabilidade e eficácia do diagnóstico da doença de coração. Levando em consideração as classes temos: classe 0 - classificação do câncer em benigno e classe 1: classificação do câncer em maligno.

$$\text{Sensibilidade} = \frac{VP}{(VP+FN)} = \frac{174}{(174+13)} = 0,93 \quad (5.1.3.1)$$

$$\text{Especificidade} = \frac{VN}{(VN+FP)} = 344 / (344 + 38) = 0,90 \quad (5.1.3.2)$$

$$Rp = \text{sensibilidade} / (1 - \text{especificidade}) = 9,3 \quad (5.1.3.3)$$

$$Rn = \left(1 - \frac{\text{sensibilidade}}{\text{especificidade}}\right) = -0,032 \quad (5.1.3.4)$$

$$ED = (VP + VN) / (VP + VN + FP + FN) = 0,9104 \quad (5.1.3.5)$$

Percebe-se a eficácia do modelo, em classificar os doentes (sensibilidade), com apenas 13 casos de falsos negativos, como mostrados na matriz de confusão Fig.16. As

razões de probabilidade positiva ( $5 < R_p < 10$ ) e negativa ( $R_n < 0,1$ ), além da eficácia do diagnóstico evidenciou que a acurácia do modelo foi ótima para a base de dados *Breast\_Cancer*.

## 5.2 Algoritmo ECM

### 5.2.1 Doença do Coração

Por meio da interface ECMgui disponível nas versões inferiores a 7.0 do Matlab podemos aplicar o algoritmo ECM na base de dados Doença do Coração, por meio da correlação de Spearman, utilizamos como atributos os três com menor valor de correlação, visto que irão transmitir informações únicas e diferentes aos *clusters* criados. Foram estes os atributos utilizados:

- Trestbps assintomáticos: Pressão arterial em repouso (em mmHg na admissão no hospital);
- Chol: Exibe o colesterol sérico em mg / dl (unidade);
- Fbs: Açúcar no sangue em jejum > 120 mg / dl - (1 = verdadeiro; 0 = falso).

A acurácia do modelo foi alta, para estes atributos e o número de regras apresentado foi o valor mais baixo.

Queremos analisar o comportamento dos índices internos de validação Xien-Beni e Coeficiente de partição, quanto a separabilidade e compacidade dos *clusters*, para isso alteramos o valor do meta-parâmetro do algoritmo em questão, que é a distância limite para criação de um novo *cluster*, *Dthr*. A Tabela 5 mostra os resultados encontrados, considerando número de *clusters* e tempo de processamento, tendo como valor default de *Dthr* de 0.1 e 0.2.

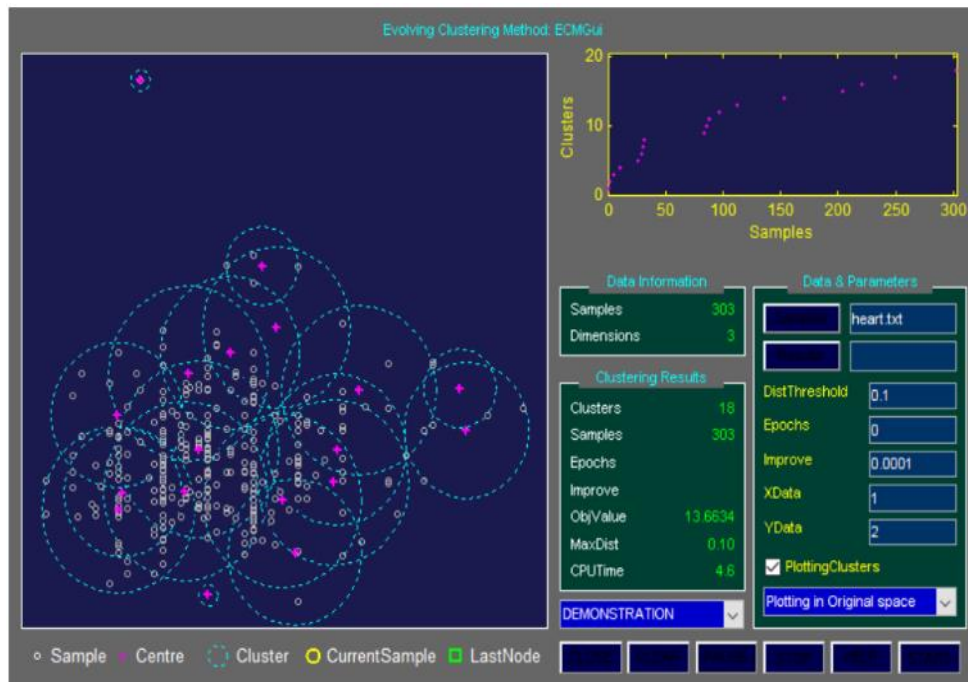
Tabela 5- Resultado ECM Doença do Coração

<i>Dthr</i>	Nº_Clusters	CPUtime	Nº_Atributos
0.1	18	4.6	3
0.2	7	4.3	3

Fonte: Autor

As imagens abaixo, Fig.17 e 18, representam o resultado do modelo considerando *Dthr* de 0.1 e de 0.2 respectivamente. Os índices de XB e PC encontrados para as duas situações acima são mostrados na Tabela 6.

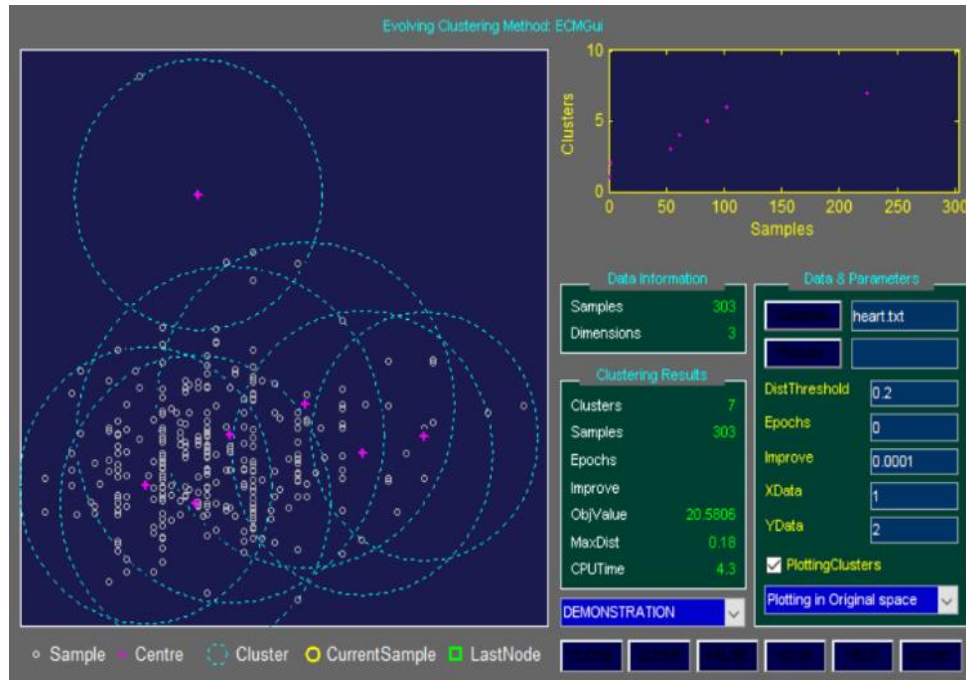
Figura 17 - Resultado do Modelo ECM com 3 atributos e Dthr de 0.1



Fonte: Autor



Figura 18 - Resultado do Modelo ECM com 3 atributos e Dthr de 0.2



Fonte: Autor

Tabela 6- Índices de Validação Doença do Coração

<i>Dthr</i>	Nº_Clusters	XB	PC
0.1	18	0.1806	0.0583
0.2	7	0.4097	0.1491

Fonte: Autor

As Figuras 17,18 e a Tabela 6, evidenciam que aumentar o valor do parâmetro *Dthr* significa alterar a sobreposição dos *clusters* (índice PC), e a distância entre as amostras e os centros de *clusters* (índice XB). Um valor maior de *Dthr*, provoca a criação de *clusters* mais nítidos, ou seja, menos sobrepostos, porém com o valor da distância entre os centros menores, resultando em *clusters* mais próximos entre si.

### 5.2.2 Doença de Parkinson

Através da correlação de Spearman, utilizamos os atributos com os menores valores de correlação. Foram estes os atributos utilizados:

- MDVP: Fo (Hz) - frequência fundamental vocal média;

- MDVP: Fhi (Hz) - frequência fundamental vocal máxima;
- MDVP: Flo (Hz) - frequência fundamental vocal mínima.

A Tabela 7 exhibe os resultados encontrados, considerando número de clusters e tempo de processamento, tendo como valor default de *Dthr* de 0.1 e o dobro deste valor 0.2.

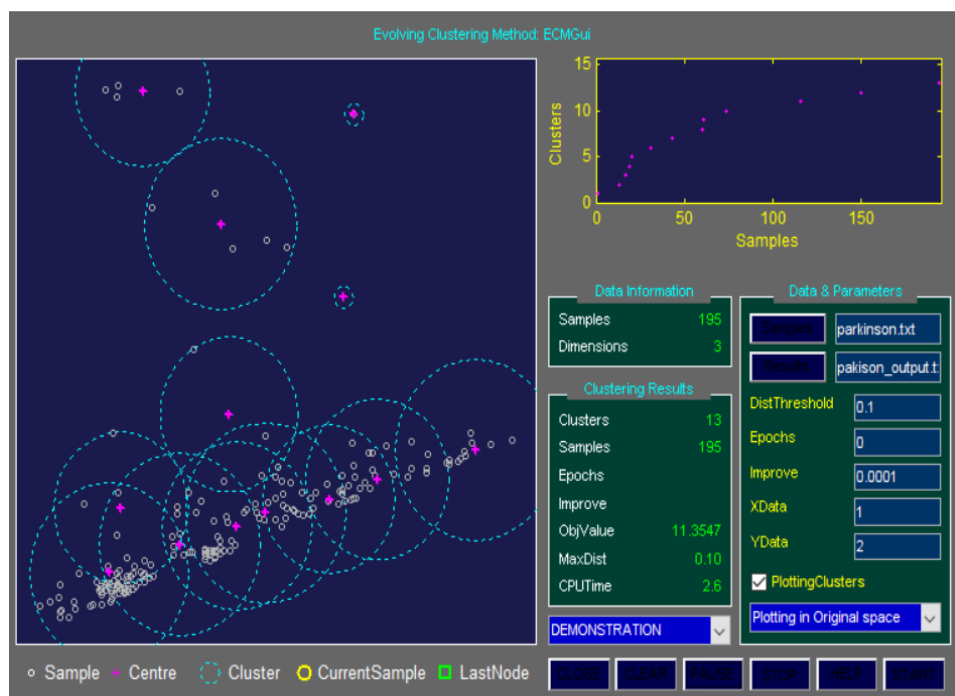
Tabela 7- Resultados ECM Doença de Parkison

<i>Dthr</i>	Nº_Clusters	CPUtime	Nº_Atributos
0.1	13	2.6	3
0.2	5	1.2	3

Fonte: Autor

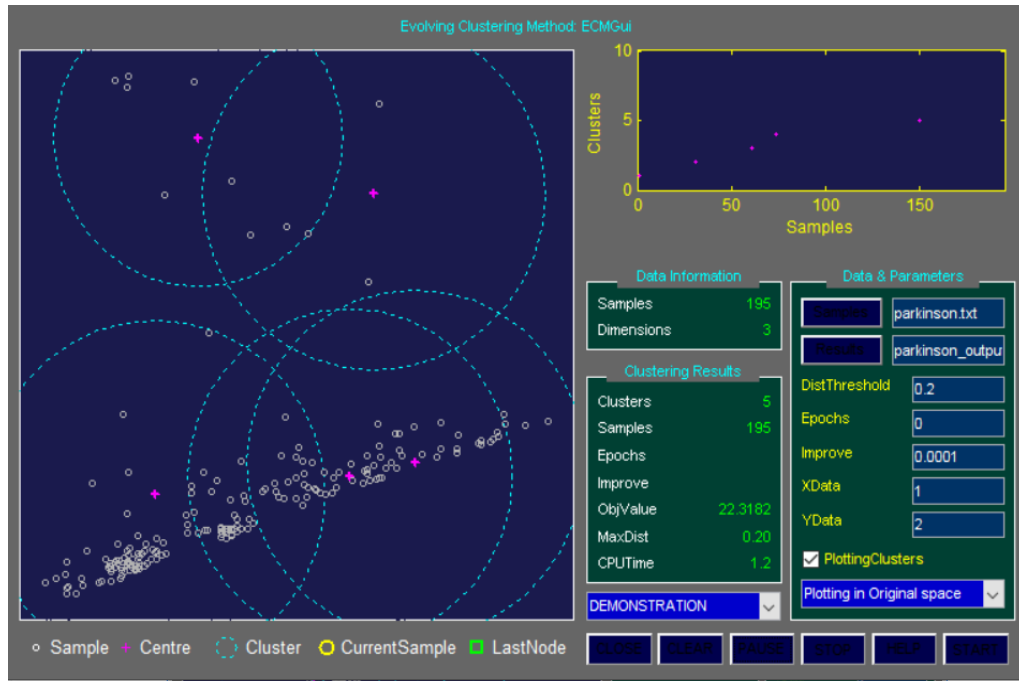
As imagens abaixo representam o resultado do modelo considerando *Dthr* de 0.1 e de 0.2 respectivamente.

Figura 19: Resultado do Modelo ECM com 3 atributos e Dthr de 0.1



Fonte: Autor

Figura 20: Resultado do Modelo ECM com 3 atributos e Dthr de 0.2



Fonte: Autor

Os índices de XB e PC encontrados para as duas situações acima são mostrados a seguir, na Tabela 8.

Tabela 8- Índices de Validação Doença de Parkison

<i>Dthr</i>	Nº_Clusters	<b>XB</b>	<b>PC</b>
0.1	13	0.2647	0.0825
0.2	5	0.5969	0.2128

Fonte: Autor

As imagens, Fig. 19 e 20, e a Tabela 8, evidenciam que aumentar o valor do parâmetro *Dthr*, provoca a criação de *clusters* mais nítidos, ou seja, menos sobrepostos, valor maior de PC, porém com *clusters* cuja distância entre os centros são menores, sendo mais próximos entre si, valor maior de XB.

### 5.2.3 Câncer de Mama

O método de redução de dimensionalidade PCA, foi aplicado na base de dados, os atributos com a maior porcentagem da variação, foram utilizados. Foram estes os atributos utilizados:

- Número do código da amostra: número de identificação;
- Espessura do grupo (massa mamária): 1 – 10;
- Uniformidade do tamanho da célula: 1 – 10.

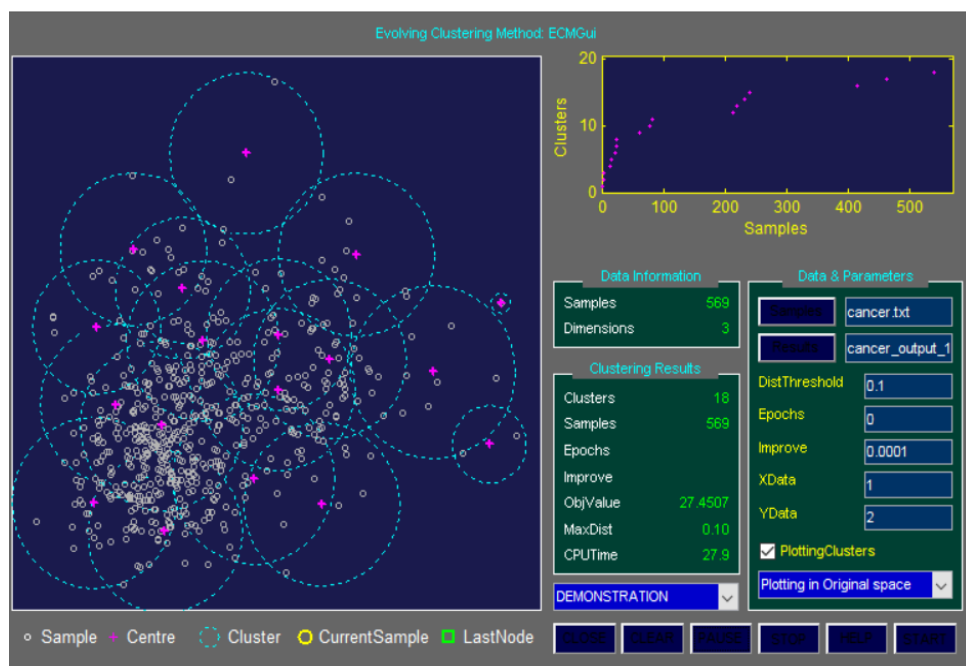
A Tabela 9 mostra os resultados encontrados, considerando número de clusters e tempo de processamento, tendo como valor default de *Dthr* de 0.1 e 0.2.

Tabela 9- Resultado ECM Breast Cancer

<i>Dthr</i>	Nº_Clusters	CPUtime	Nº_Atributos
0.1	18	27.9	3
0.2	7	8,5	3

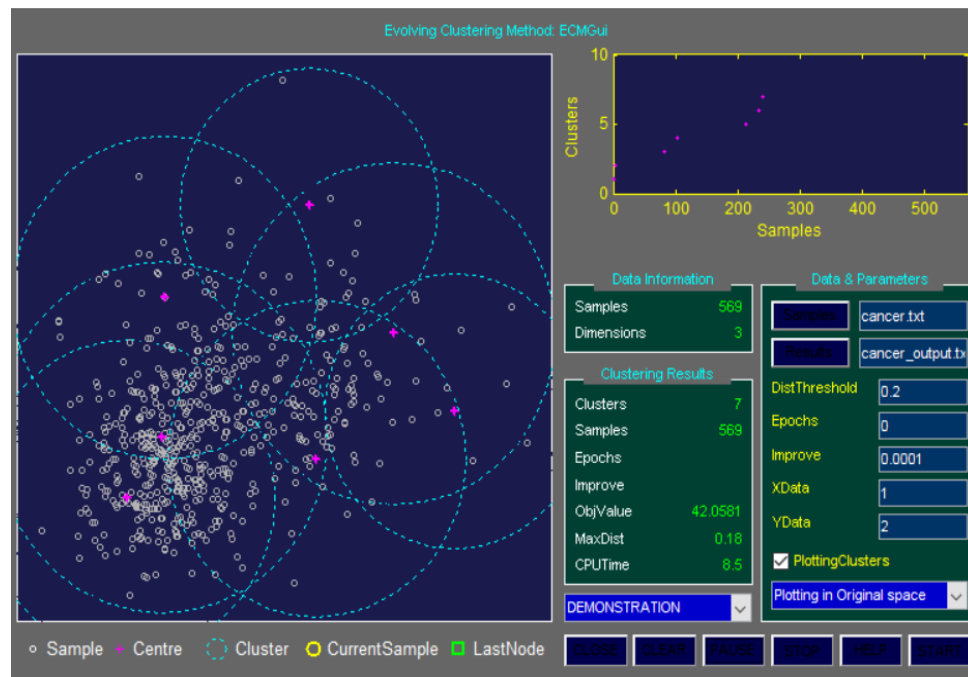
Fonte: Autor

Figura 21- Resultado do Modelo ECM com 3 atributos e Dthr de 0.1



Fonte: Autor

Figura 22- Resultado do Modelo ECM com 3 atributos e Dthr de 0.2



Fonte: Autor

Os índices de XB e PC encontrados para as duas situações acima são mostrados a seguir, na Tabela 10.

Tabela 10- Índices de Validação Câncer de Mama

<i>Dthr</i>	Nº Clusters	XB	PC
0.1	18	0.2197	0.0598
0.2	7	0.4971	0.1527

Fonte: Autor

As imagens, Fig. 21 e 22 e a Tabela 10, reafirmam o padrão observado nos outros resultados, de que aumentar o valor do meta-parâmetro *Dthr* reduz o número de *clusters*, porém os mesmos tornam-se menos sobrepostos e com o índice de distância entre os grupos maximizado, ou seja, aumenta a compacidade do cluster, distância entre os grupos, com menos sobreposição, porém aumenta a separabilidade dos *clusters*, com distância dentro dos grupo sendo minimizada.

## 6 DISCUSSÃO

O algoritmo FBeM utilizado no presente trabalho ao ser comparado com algoritmos da literatura, apresentou resultado próximo em relação ao valor da acurácia do diagnóstico das doenças de Parkinson, Câncer de Mama e Doença do Coração. Como podemos ver na Tabela 11 abaixo.

Tabela 11: Comparação do Algoritmo FBeM

Base de Dados	Algoritmo	Acurácia
<b>Parkinson</b>	Classificador Fuzzy C-Means (CAGLAR et al., 2010)	83,9%
	FBeM	89,74%
<b>Câncer de Mama</b>	Agrupamento Fuzzy (Nauck et al., 1999)	97,71%
	FBeM	91,04%
<b>Doença do Coração</b>	Algoritmo Genético (Almeida, S, 2006)	90%
	FBeM	98,35%

Fonte: Autor

Para a base de dados de Parkinson e Doença do Coração a acurácia foi maior do que o algoritmo evolutivo encontrado na literatura, porém para o câncer de mama a acurácia apresentou um valor inferior em relação ao encontrado na literatura.

Considerando o modelo ECM os clusters apresentaram os seguintes valores de acurácia, evidenciados na Tabela 12.

Tabela 12: Comparação do Algoritmo ECM

<b>Base de Dados</b>	<b>Algoritmo</b>	<b>Acurácia</b>
<b>Parkinson</b>	ECM (11 Clusters e 22 atributos) (LEITE et al., 2014).	83,58%
	ECM (5 Clusters e 3 atributos).	80%
<b>Câncer de Mama</b>	ECM (7 Clusters e 10 atributos) (LEITE et al., 2014).	95,75%
	ECM (7 clusters e 3 atributos).	85,72%
<b>Doença do Coração</b>	ECM (11 Clusters e 13 atributos) (LEITE et al., 2014).	69,25%
	ECM (7 Clusters e 3 atributos)	57,21%

Fonte: Autor

A análise do modelo ECM em relação ao algoritmo da literatura (LEITE et al., 2014), mostrou que o modelo da literatura apresentou valores de acurácia elevados em todas as bases de dados, porém não tão discrepantes com o valor encontrado no modelo do presente trabalho. A discrepância entre os resultados ocorreu, pois no trabalho de LEITE et al., 2014, foi considerado um número maior de clusters e de atributos, aumentando a acurácia do modelo.

## 7 CONCLUSÃO

O presente trabalho teve como objetivo a classificação de dados médicos utilizando algoritmos incrementais online.

Ao aplicarmos o algoritmos FBeM para as bases de dados Heart\_Disease, Parkinson\_Disease e Breast\_Cancer Disease, percebemos que as métricas de eficácia diagnóstica encontradas foram altas e que existe um relação diretamente proporcional entre o aumento do número de atributos com as seguintes variáveis: tempo de processamento e número de regras do algoritmo. Considerando o algoritmo ECM, ao alterarmos o metaparámetro *Dthr* percebemos que aumentar o mesmo significa reduzir o número de clusters e o tempo de processamento, porém aumenta os índices de validação Xien-Beni e Coeficiente de Partição PC.

De acordo com o índice Xie-Beni o classificador ECM é mais compacto e seus *clusters* mais separados entre si, para os casos onde o valor de *Dthr* é de 0.1, já para o índice PC, o número ideal de *clusters* para os dois modelos classificadores são os que apresentaram *Dthr* de 0.2.

Concluimos que os algoritmos de aprendizagem on-line e incremental usados no presente trabalho FBeM e ECM foram eficientes para a classificação de doenças consideradas relevantes.

No futuro pretende-se aplicar os métodos evolutivos, apresentados no presente trabalho, em problemas que envolvam a classificação de mais doenças, proporcionando uma qualidade de vida melhor para as pessoas, com um diagnóstico assertivo e antecipado.



## REFERÊNCIAS

- ALBRECHT, A.; LAPPAS, G.; VINTERBO, S.; WONG, C.; OHNO-MACHADO, L. Two applications of the LSA macrine. In: Proceedings of the 9th International Conference on Neural Information Processing; 2002 Nov 18-22; Singapore. p. 184-189.
- ALMEIDA, S. M. Algoritmo Genético para Determinação de Relevância de Sintomas no Diagnóstico de Cardiopatia Isquêmica, Brasília, 2006.
- ANDRADE, J. Agrupamento de dados em fluxo contínuos com estimativa automática do número de grupos. Maio de 2015.
- ANGELOV, P. (2007). “Architectures forevolving fuzzy rule-based classifiers”. IEEE Conf. on Systems, Man and Cybernetics, pp: 2050-2055.
- BABCOCK, B.; BABU, S.; DATAR, M.; MOTWANI, R.; WIDOM, J. Models and issues in data stream systems. In PODS’02: Proc. of the 21th ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, ACM Press, pp 1-16, 2002.
- BARGIELA, A.; PEDRYCZ, W. Granular computing: an introduction. [S.l.]: Springer Science & Business Media, 2012. v. 717.
- BIFET, A. Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams. [S.l.]: IOS Press, 2010. 1-212 p. (Frontiers in Artificial Intelligence and Applications, v. 207). ISBN 978-1-60750-090-2.
- BISHOP, C. M. (1995). Neural Networks for pattern recognition. Oxford: Oxford University Press.
- BORGES, L. Acurácia Diagnostico -- Medidas de Acurácia Diagnóstica na Pesquisa Cardiovascular. Universidade Federal de Uberlândia, Uberlândia, MG – Brasil, 2016.
- BRATS. Stents farmacológicos e stents metálicos no tratamento da doença arterial coronariana. Boletim Brasileiro de Avaliação de Tecnologias em Saúde (BRATS), número 22, Junho 2013. Disponível em: [www.anvisa.gov.br/divulga/newsletter/brats/2009/BRATS8.pdf](http://www.anvisa.gov.br/divulga/newsletter/brats/2009/BRATS8.pdf)
- CAGLAR, M. F.; CETISLI, B.; TOPRAK, I. B. (2010). “Automatic recognition of Parkinson’s disease from sustained phonation tests using ANN and adaptive neuro-fuzzy classifier”. Journal of Eng Sci and Design, vol.1, no.2, pp. 59-64.
- CAO, F.; ESTER, M.; QIAN, W.; ZHOU, A. Density-based clustering over an evolving data stream with noise. In: Proceeding of the Sixth SIAM International Conference on Data Mining, SIAM, pp. 328-339, 2006.
- COMMITTEE. The Global Parkinson’s Disease Survey Steering (2002). “Factors impacting on quality of life in Parkinson's disease: results from an international survey”. Mov Disord.,vol. 17, issue 1, 60-67.
- COOPER, G.; EICHHORN, G.; RODNITZKY, R.L. (2008). “Parkinson’s Disease”. In: Neuroscience in Medicine. Totowa, NJ: Humana Press. pp. 508-512.

- DEKEL, O. From online to batch learning with cutoff-averaging. 2009.
- FRITZKE, B. (1995). "A Growing Neural Gas Network Learns Topologies". *Advances in Neural Information Processing Systems*, pp: 625-632.
- GAMA, J. *Knowledge Discovery from Data Streams*. 1st. ed. [S.l.]: Chapman & Hall/CRC, 2010. ISBN 1439826110, 9781439826119.
- GAMA, J.; ŽLIobaite, I.; BIFET, A.; PECHENIZKIY, M.; BOUCHACHIA, A. A survey on concept drift adaptation. *ACM Computing Surveys*, v. 46, n. 4, p. 44:1–37, 2014.
- GIL, D.; JOHNSON, M. (2009). "Diagnosing Parkinson by using artificial neural networks and support vector machines". *Global Journal of Computer Science and Technology*, vol.9, issue 4, pp. 63-71.
- GOLAB, L.; OZSU, M.T. Issues in data stream management. *SIGMOD Record*, v.32, p.5-14, 2003.
- GOMES, R. C. Ministério da Saúde. Portal Brasil, p. Gomes, R. C. Doenças Cardiovasculares Causam quase 30% das Mortes no País. No Brasil, homens são os mais afetados pelas doenças cardiovasculares, como infarto e AVCs. Ministério da Saúde, 2011. Disponível em < <http://www.brasil.gov.br/saude/2011/09/doenca>, 06 set. 2011. Acesso em: 01 fev. de 2020.
- GUIMARÃES, P. *Análise de Correlação e medidas de associação*. DEST/UFPR, 2013.
- GUSTAVO, C. *Pesquisa e Desenvolvimento de um Método de Aprendizagem de Máquina: Agrupamento Incremental 2015*.
- HAN, J., KAMBER, M.; PEI, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition.
- HRUSCHKA, E. R., EBECKEN, N. F. A Genetic algorithm for cluster analysis. *IEEE Transactions on Evolutionary Computation*, 2001.
- INSTITUTO NACIONAL DE CÂNCER (Brasil). *Estimativa 2020. Incidência do Câncer no Brasil*. Rio de Janeiro: INCA, 2019.
- BEZDEK, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, 1981.
- MACQUEEN, J. B. "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1967, 1:281-297.
- JAIN, A. K.; DUBES, R. C. 1999. *Algorithms for Clustering Data*. Prentice-Hall advanced reference series. Prentice-Hall, Inc., Upper Saddle River, NJ
- JANKOVIC, J. (2008). "Parkinson's disease: clinical features and diagnosis". *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 79, issue 4, pp. 368-376.

KASABOV, N. (2007). *Evolving Connectionist Systems*. Springer, 2ª edição, 451p. Kasabov, N. (2001). “Evolving fuzzy neural networks for supervised/unsupervised online knowledge-based learning”. *IEEE Transactions on SMC - Parte B*, Vol. 31-6, pp: 902-918.

KASABOV, N.; SONG, Q. *ECM — A Novel On-line, Evolving Clustering Method and Its Applications*, 2002.

KEELY, M.G.; HAND, D.J.; ADAMS, N.M. The impact of changing populations on classifier performance. In: *Proceeding of the fifth ACM SIGKDD international conference on Knowledge Discovery and data mining*. New York, NY, USA: ACM, 1999. (KDD' 99), p. 367-371. ISBN 1- 581 13-143-7. Disponível em: <<https://doi.org/10.1145/312129.312285>>.

LANE, T.; BRODLEY, C. Approaches to online learning and concept drift for user identification in computer security. Em *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, páginas 259-263, 1998.

LANGLEY, P. *Learning in humans and machines: Towards an interdisciplinary learning Science – order effects in incremental learning*. Oxford, 1995.

LAN95.; P. LANGLEY. *Learning in humans and machines: Towards an interdisciplinary learning science*, chapter Order Effects in Incremental Learning. Oxford, 1995.

LAST, M. Online classification of nonstationary data streams. *Intelligent Data Analysis*, 6:129-147, 2002.

LEE, H. *Seleção de atributos importantes para a extração de conhecimento de base de dados*, 2005.

LEITE, D.; GOMIDE, F. *Evolving Linguistic Fuzzy Models from Data Streams*, 2013.

LEITE, D. et al. Fuzzy granular evolving modeling for time series prediction. In: *IEEE. Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*. [S.l.], 2011. p. 2794–2801.

LEITE, D et al. *Agrupamento evolutivo aplicado ao reconhecimento de padrões em dados médicos*, 2014.

LEITE, D.; COSTA P.; GOMIDE F. *Redes Neurais Granulares evolutivas*, 2002.

LEITE, D.; COSTA, P.; GOMIDE, F. (2013). “Evolving granular neural networks from fuzzy data streams”. *Neural Networks*, vol. 38, pp. 1-16.

LITTLE, M. A.; MCSHARRY, P. E.; HUNTER, E. J.; SPIELMAN, J.; RAMIG, L. (2009). “Suitability of dysphomia measurements for telemonitoring of Parkinson’s disease”. *IEEE Trans Biomed Eng*, vol. 56, issue 4, pp. 1015-1022.

LOBO L. *Inteligência artificial e medicina*, 2007.

MACQUEEN, J. B. Some Methods for classification and Analysis of MultiVariate Observations. In: CAM, L.; NEYMAN, J. (Eds). *5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281-297, 1967.

MEDIUM. Entendendo o que é Matriz de Confusão com Python. Disponível em: <https://medium.com/data-hackers/entendendo-o-que-%C3%A9-matriz-de-confus%C3%A3o-com-python-114e683ec509>. Acessado em: 01 fev de. 2020.

MICHALSKI, R. S.; CARBONELL, J. G.; MITCHELL, T. M. Machine learning: An artificial intelligence approach. [S.l.]: Springer Science and Business Media, 2013.

MINISTÉRIO DA SAÚDE. Cardiologia/Cirurgia Cardíaca. Disponível em: <https://www.saude.gov.br/atencao-especializada-e-hospitalar/especialidades/cardiologia-cirurgia-cardiaca>. Acesso em 01 fev. de 2020.

MOREIRA, C. S. et al. DOENÇA DE PARKINSON: COMO DIAGNOSTICAR E TRATAR: Parkinson's Disease: How to diagnose and to treat. Revista Científica da Faculdade de Medicina de Campo. Campos dos Goytacazes, RJ/Brasil., v. 2, n. 2, p.2, 11./2007.

MOSCATO.; VON, Z (2004). Uma Visão Geral de Clusterização de Dados, Análise de dados em Bioinformática.

NARASIMHAMURTHY, A.; KUNCHEVA, L.A framework for generation data to simulate changing environments. In: ACTA PRESS. Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications [S.l.], 2007.v.549, p.389.

NAUCK, D.; KRUSE, R. Obtaining interpretable fuzzy classification rules from medical data. Artificial Intelligence in Medicine 1999; 16:149-169

OLANOW, C. W.; STOCCHI, F.; LANG, A. E. (Eds.) (2011) Parkinson's Disease: Non-Motor and Non-Dopaminergic Features. Wiley-Blackwell.

OLIVEIRA, C. R.; ORTIZ, K. Z.; VIEIRA, M. M. Disartria: estudo da velocidade de fala. Pró-Fono Revista de Atualização Científica, v. 16, n. 1, p. 39-48. 2004

PAL, N. R.; BEZDEK, J. C. (1995). "On cluster validity for the Fuzzy C-Means model". IEEE Trans Fuzzy Syst, vol. 3, issue 3, pp.370-379.

PASSOS, C. U. Computação evolutiva e aprendizado de máquina aplicados ao apoio da cardiopatia isquêmica, 2014.

PATEL VL.; SHORTLIFE EH.; STEFANELLI M.; SZOLOVITS P.; BERTHOLD MR.; BELLAZI R.; ABU-HANNA A. The Coming Age of Artificial Intelligence in Medicine. Artif Intell Med 2009; 46(1):5-17.

PEÑA-REYES CA.; SIPPER M. A fuzzy-genetic approach to breast cancer diagnosis. Artificial Intelligence in Medicine 1999; 17:131-155.

QUINLAN, R. J. Induction of decision trees. Machine Learning, v. 1, n. 1, 1986. Citado na página 46.

BABUSKA, R.; VEEN, P.; KAYMAK, U. Improved covariance estimation for Gustafson-Kessel clustering. IEEE International Conference on Fuzzy Systems, pages 1081–1085, 2002.

RIBEIRO T.; BORGES M.; CARDOSO R.; COELHO R.; FERREIRA S.; LEITE D. Classificação Fuzzy de Padrões Não-Motores e Indicação da Severidade da Doença de Parkinson, 2018.

ROBERT, B.; GOLDBERGSUSAN, J.; BARKERLUISPEREZ-GRAU. (1989). Regulation of gene expression during plant embryogenesis.

ROSSI, R. Análise de Componentes Principais em Data Warehouses, 2017.

RUSTEMPASIC, I.; CAN, M. (2013). "Diagnosis of Parkinson's disease using fuzzyC-means clustering and pattern recognition". SouthEast Europe Journal of Soft Computing, Online available, pp. 42-49.

SANSON C.; CAMARGO J. Quarta Revolução Industrial 4.0. 2007. Disponível em: [http://www.ihu.unisinos.br/images/ihu/apresentacoes\\_palestrantes/30\\_05\\_17\\_cesar\\_sanson\\_revolucao\\_4.0.pdf](http://www.ihu.unisinos.br/images/ihu/apresentacoes_palestrantes/30_05_17_cesar_sanson_revolucao_4.0.pdf). Acesso em: 01 fev. de 2020.

SCHICK, L. Distibuted d-FuzzStream: Agrupamento Distribuído Não Supervisionado Fuzzy em Fluxo de Dados, 2019.

SETIAWAN, N. A.; VENKATACHALAM, P. A.; HANI, A. F. M. Diagnosis of Coronary Artery Disease Using Artificial Intelligence Based Decision Suport System. Proceedings of The Internacional Confernce of Man-Machine Systems (ICoMMS), 2009.

SHASHIKANT G.; CHETAN PATIL.; ASHOK GHATOL. Heart Disease Diagnosis using Support Vector Machine, 2011.

SIMPSON, P. (1992). "Fuzzy min-max neural net-works. Part I: Classification". IEEE Trans. on Neural Networks, Vol. 3-5, pp: 776-786.

SILVEIRA, D; BRASOLOTTO.; GHEDINI, A. Reabilitação vocal em pacientes com doença de Parkinson: fatores interferentes. Pró-Fono R. Atual. Cient., Barueri , v. 17, n. 2, p. 241-250, ago. 2005 . Disponível em <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0104-56872005000200013&lng=pt&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-56872005000200013&lng=pt&nrm=iso)>. acessos em 31 maio 2020. <https://doi.org/10.1590/S0104-56872005000200013>.

SIMPSON, P. (1992). "Fuzzy min-max neural net-works. Part I: Classification". IEEE Trans. on Neural Networks, Vol. 3-5, pp: 776-786.

SOUZA, A.; POPPI, R. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. Quím. Nova [online]. 2012, vol.35, n.1, pp. 223-229. ISSN 0100-4042. [100]

SPIELMAN, J. L.; BOROD, J. C.; RAMIG, L. O. The effects of intensive voice treatment on facial expressiveness in Parkinson disease: preliminary data. Cogn. Behav. Neurol., v. 16, n. 3, p. 177-188, sep. 2003

TAN, P.; STEINBACH, M.; KUMAR, V. (2005) Cluster Analysis: Basic Concepts and Algorithms. In Introduction to Data Mining, Addison-Wrsley, Boston.

TSANAS, A.; LITTLE, M. A.; MCSHARRY, P. E.; RAMIG, L. O. (2010). “Accurate telemonitoring of Parkinson’s disease progression by non-invasive speech tests”. IEEE Trans Biomed Eng, vol. 57, no. 4, pp. 884-893.

TSANAS, A.; LITTLE, M. A.; MCSHARRY, P. E.; SPIELMAN, J.; RAMIG, L. O. (2012). “Novel speech signal processing algorithms for high-accuracy classification of Parkinson’s disease”. IEEE Trans Biomed Eng. vol. 59, no. 5, pp. 1264-1271.

WANG, H.; W. FAN.; HAN, J. Mining concept drifting data streams using ensemble classifiers. Em Proceedings of the International Conference on Knowledge Discovery and Data Mining, páginas 226-235, 2003.

WILLIAM H. WOLBERG.; OLVI L. MANGASARIAN. Multisurface method of pattern separation for medical diagnosis applied to breast cytology (linear programming/pattern recognition/expert systems/cancer diagnosis, 1990.

WORLD HEALTH ORGANIZATION. International Agency for Research on Cancer. Globocan. Disponível em: <https://gco.iarc.fr/today/online-analysis-map?projection=globe> . Acesso em 05/02/2020.

XIE, X.; BENI, G. (1991). A Validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell., vol.13, no. 8, pp.841-847.

ZHU, Y.; SASHA, D. StatStream: statistical monitoring of thousands of data streams in real time. In: Proceedings of the 28th International Conference on Very Large Data Bases, VLDB Endowment, pp.358-469,2002.

ZLIOBAITE, I. Learning under concept drift: na overview. Lithuania, 2009.