



## UMA REVISÃO DE LITERATURA SOBRE TÉCNICAS PARA CLASSIFICAÇÃO AUTOMÁTICA DE ELETROCARDIOGRAMA (ECG)

<sup>1</sup> Jandson Santos Nunes (SENAI-CIMATEC) – [ian.nunes@gmail.com](mailto:ian.nunes@gmail.com); <sup>2</sup> Valter de Senna (SENAI-CIMATEC) – [valter.senna@gmail.com](mailto:valter.senna@gmail.com);

**Resumo:** Este trabalho discorre sobre o estado da arte das técnicas de classificação automática de ECG. A produção científica internacional, publicada entre 2008 e 2018 em periódicos que compõem a plataforma *Science Direct*, caracteriza o escopo desta revisão de literatura. Ao todo, 130 artigos constituem as amostras do estudo. Elas evidenciam a realização sequencial do tratamento dos sinais, extração de informações relevantes e a classificação. Múltiplas abordagens em cada etapa sugerem uma análise comparativa das técnicas utilizadas.

**Palavras-Chaves:** eletrocardiograma; ECG; classificação automática de ECG;

## A REVIEW ON TECHNIQUES FOR AUTOMATIC CLASSIFICATION OF ELECTROCARDIOGRAM (ECG)

**Abstract:** This paper discusses the state-of-the-art of automatic ECG classification techniques. International scientific production, published between 2008 and 2018 in periodicals compiling the Science Direct platform, characterizes the scope of this literature review. In all, 130 articles constitute the study samples. They evidence sequential completion of signal processing, extraction of relevant information, and classification. Multiple approaches at each stage suggest a comparative analysis of the techniques used.

**Keywords:** eletrocardiogram; ECG; automatic ECG classification;



## 1. INTRODUÇÃO

Em condições hostis ou quando submetidos a estresse de qualquer natureza, humanos podem apresentar oscilações de raciocínio que afetam sua percepção e capacidade de tomar decisões. Neste cenário, os computadores são tolerantes a ambientes insalubres, imunes a questões emocionais e resistentes a fadiga. A realização de tarefas laboriosas pelo computador, quando bem definidas, tendem a ser mais precisas e isentas de distrações, tédio ou cansaço.

O estudo de fenômenos temporais, sejam eles naturais como o comportamento das marés, ou artificiais como a flutuação do valor de ações em uma bolsa, é considerado um dos 10 maiores desafios na mineração de dados, de acordo com Yang e Wu [1]. Por conta de suas peculiaridades, há décadas pesquisadores discorrem sobre modelos de análises de dados em séries temporais.[2,3] Dentre as diversas abordagens para a modelagem dos dados sequenciais, algumas incluem a estimação de parâmetros que podem ser utilizados como recursos em classificadores [4-6]. Entretanto, contextos reais podem ser dinamicamente complexos ou povoados de fatores intervenientes, inserindo informação de ruído a série e dificultando a definição de equações analíticas que as modelem [7].

A informação dos pulsos elétricos de batimentos cardíacos em relação ao tempo definem ECG. A utilização desta para diagnosticar problemas coronários é uma realidade comum na medicina. Daí a relevância em realizar uma classificação precisa, conforme anomalias identificadas. Sendo o ECG um caso de série temporal, conceitualmente apresenta as mesmas características inerentes ao tema. Além disso, Moavenian expande a problemática da classificação de ECG, afirmando que existem casos de pacientes com diagnósticos similares e diferentes formas de onda do ECG. A oposição disso também é verdade, ou seja, existem doenças distintas que afetam os sinais de ECG de forma semelhante [8]. A união destes aspectos dificulta a classificação das alterações, e conseqüentemente, estimula a busca por soluções viáveis do ponto de vista clínico.

O intuito desta revisão é visitar os trabalhos que se propõem a classificar automaticamente ECGs, e desta forma identificar as principais abordagens, normas e critérios de avaliação dos resultados, atualmente utilizados pela comunidade científica neste contexto. Esta é uma das etapas da minha pesquisa no Mestrado em Modelagem Computacional e Tecnologia Industrial (MCTI) do Centro Universitário SENAI CIMATEC, cujo tema está diretamente relacionado a esta proposta.

O documento está estruturado como segue: a seção 2 descreve o procedimento de definição e execução da sentença de busca na plataforma Science Direct. Posteriormente, na seção 3, apresenta os resultados e discussão dos conteúdos obtidos na busca. Por fim, são expostas as conclusões do trabalho na seção 4.



## 2. MÉTODO

Internacionalmente reconhecida como importante fomentadora de conteúdos científicos, desde 1997 a editora Elsevier reúne em sua plataforma *Science Direct* mais de 12 (doze) milhões de trabalhos, publicados em aproximadamente 3500 revistas e 34000 livros, de diversas áreas do conhecimento [9,10]. O acesso aos resumos dos artigos é livre e gratuito, porém o conteúdo completo das publicações somente é disponibilizado aos assinantes ou compradores (através de mecanismo pay-per-view). Estas características fazem dela, ambiente propício a realização desta revisão de literatura.

A elaboração da sentença lógica: "**(ECG) AND (automatic OR computer) AND (classification)**", objetivou a síntese da literatura relevante a este trabalho, considerando a especificação da sintaxe da plataforma [11]. O termo ECG indica o objeto de estudo. *Classification* diz respeito a ação que se pretende realizar sobre o conjunto de objetos. Este trabalho investiga os métodos computacionais de classificação de ECG, sendo assim o termo (*automatic OR computer*) reflete este requisito.

A sentença definida serviu como valor para o campo "*Title, abstract or Keywords*" da busca avançada disponibilizada na ferramenta. Pela proposta de investigação do estado da arte sobre o tema, restringiu-se o escopo da busca ao período compreendido entre 2008-2018 e os seguintes tipos de publicações: *Review articles, Research articles, Conference abstracts, Case reports, Data articles, Discussion, Mini reviews e Short communications*.

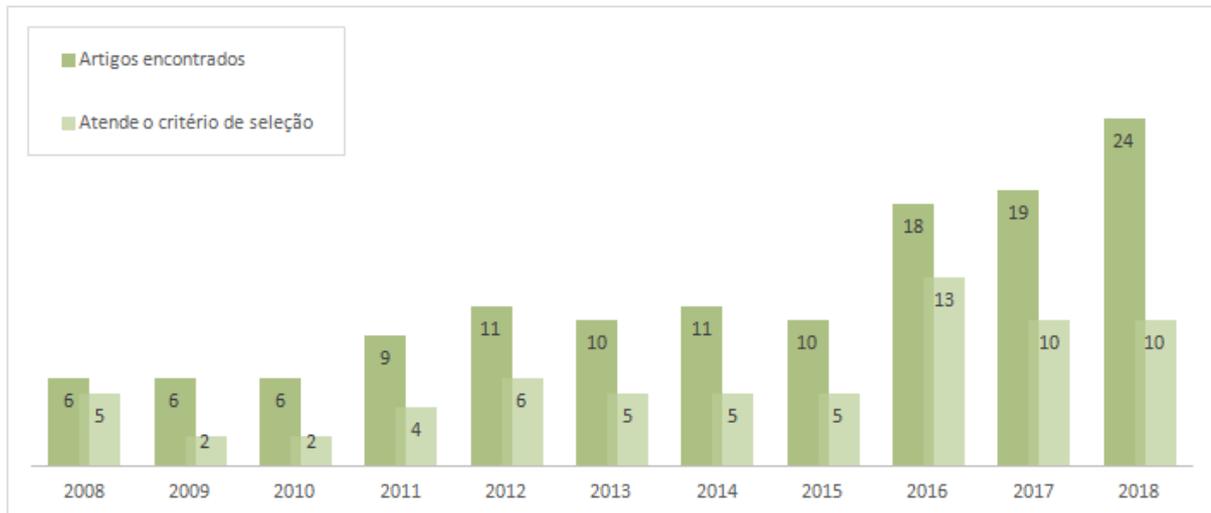
O critério de inclusão dos trabalhos relevantes considera as produções que apresentam ou discutem classificações automáticas de batimentos cardíacos, cujas categorias definidas estejam relacionadas apenas com os tipos de ondas apresentadas no ECG. Estas informações devem constar nos resumos das publicações. Embora outras classificações tenham relevância científica, não faz parte do escopo deste trabalho discutir técnicas que utilizem as características extraídas de ECGs, com a finalidade de identificar outros comportamentos que não sejam os de batimentos cardíacos, ainda que utilizem exclusivamente este mecanismo para obter informação.

## 3. RESULTADOS E DISCUSSÃO

A execução do método (definido na seção 2) em 12 de Julho de 2018, identificou 130 publicações, das quais 67 atenderam aos critérios de inclusão. A distribuição temporal dos trabalhos encontrados pode ser observada na Figura 1. O gráfico ilustra crescimento na quantidade de publicações, porém a diferença entre a quantidade de artigos encontrados e os que atendem ao critério de inclusão sugere que novos tipos de classificações estão sendo feitas utilizando ECG.

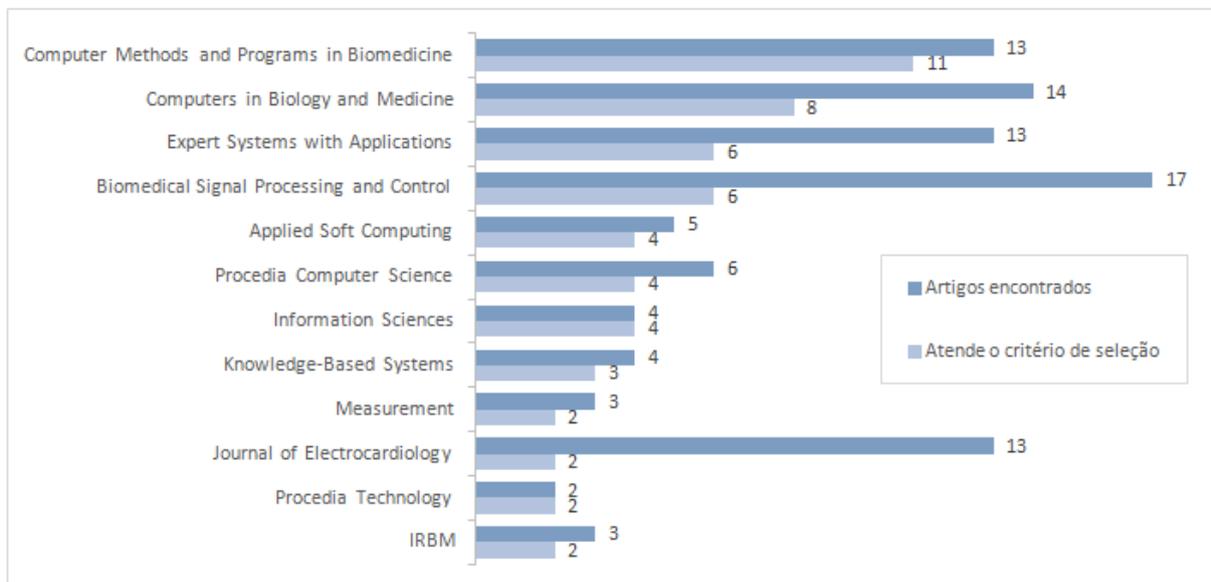


Figura 1. Distribuição da quantidade de publicações encontradas por ano. (Fonte: Autor)



A distribuição dos trabalhos nos diferentes periódicos que compõe a plataforma *Science Direct* é ilustrada na Figura 2. Este gráfico expõe um panorama sobre onde os pesquisadores costumam publicar seus trabalhos. A natureza de cada periódico pode indicar uma maior ou menor ocorrência de publicações de caráter técnico, e requer aprofundamento para que seja possível estabelecer uma relação entre o periódico e a probabilidade de encontrar publicações com critérios específicos.

Figura 2. 12 periódicos que publicaram mais conteúdos que atendem aos critérios de inclusão. (Fonte: Autor)





Alguns trabalhos descreveram o processo de classificação de ECG em 3 (três) etapas. A primeira delas é dedicada a remoção de ruídos nos sinais, provocados por perturbações que não correspondem à realidade, e que podem afetar negativamente o resultado final. O passo seguinte é a extração das características mais relevantes para a classificação, de forma que elas sejam as mínimas necessárias para distinguir as categorias propostas. Por fim, as soluções estabelecem o grau de pertinência do conjunto de valores característicos à cada classe definida, utilizando os valores limítrofes estabelecidos por uma inteligência artificial ou humana.

Dentre as diversas técnicas observadas nos trabalhos, algumas se mostraram recorrentes. A transformada Wavelet é a decomposição de uma série temporal, que pode ocorrer no domínio da frequência ou tempo. Ela é a principal técnica citada nas publicações para etapa de remoção de ruído. A extração de características é realizada, na maioria das publicações, através de análise de componentes principais. Este é um procedimento matemático cujo propósito é a obtenção de um subconjunto de variáveis linearmente não correlacionadas, originado de um conjunto de variáveis possivelmente correlacionadas. Quanto as técnicas aplicadas a classificação, os autores se dividem entre Máquina de Vetores de Suporte e Redes Neurais, e suas variações para ambas. O consenso está na utilização de inteligência artificial para interpretar os dados de ECG e classifica-los de acordo com suas características.

A quantidade de categorias estabelecidas nas publicações varia conforme as propostas. A *Association for the Advancement of Medical Instrumentation* (AAMI) é uma organização que apoia a adoção segura da tecnologia de saúde. O Comitê de ECG (EC) desta instituição estabelece na ANSI/AAMI EC57:1998, diretrizes que permitem avaliar o desempenho de algoritmos que medem o ritmo cardíaco [12]. Esta foi a referência mais utilizada pelos trabalhos para determinar as classes de batimentos cardíacos identificáveis em suas respectivas aplicações.

O desenvolvimento de novas abordagens com inteligência artificial estimulou a produção de trabalhos correlacionados a classificação de ECG. Na medida em que diferentes técnicas surgiam, observou-se sua aplicação com a demanda de melhorar o desempenho em comparação a outros trabalhos e/ou ampliar a quantidade de classes identificáveis. Ainda atendendo as diretrizes da ANSI/AAMI EC57:1998, [12] os principais critérios para avaliação de desempenho aplicados aos trabalhos são: precisão, sensibilidade e especificidade.

Foram poucas as publicações que não citaram a utilização da base de dados de arritmia do MIT-BIH. Esta base de dados foi desenvolvida pelo Massachusetts Institute of Technology e pelo Boston's Beth Israel Hospital [13]. Atualmente é mantida pela PhysioNet, [14] que disponibiliza acesso gratuito ao conteúdo.



## 4. CONCLUSÃO

A fragmentação do processo de classificação em etapas sugere o mesmo desdobramento para as pesquisas, de forma que a otimização das partes contribua para a melhoria da solução geral. Assim, o surgimento de uma nova técnica de classificação, pode estimular uma extração de características mais adequada, que pode demandar um filtro mais eficaz ao destaque destes aspectos. Este ciclo virtuoso proporciona a evolução científica deste contexto.

Utilizar o mesmo conjunto de ECGs em diferentes trabalhos facilita a comparação de resultados. Assim, a contribuição da base de dados de arritmia do MIT-BIH, associado ao cumprimento das recomendações descritas na ANSI/AAMI EC57:1998, [12] permite a construção de cenários e o estabelecimento de metas igualitárias as proposições dos trabalhos.

Classificação automática de ECG é uma atividade bastante explorada pela comunidade científica. A quantidade de publicações encontradas em apenas uma plataforma é a expressão disso. Contudo, características extraídas em ECG podem ser utilizadas como parâmetros para a classificação de comportamentos que não sejam os de batimentos cardíacos. Dos trabalhos encontrados nesta revisão que não atenderam aos critérios de inclusão, muitos estão relacionados a classificação de comportamento do sono, emoções ou fadiga. A sentença de busca precisa contemplar a exclusão destes tipos de classificação, para melhorar a qualidade dos resultados.

Embora a plataforma *Science Direct* seja representativa para a comunidade científica, a realização dos procedimentos descritos na seção 2, em outras plataformas com propostas semelhantes, ampliará a observação das produções científicas.

## Agradecimentos

Ao Prof<sup>o</sup> Dr. Valter de Senna, meu orientador no Mestrado Acadêmico em Modelagem Computacional e Tecnologia Industrial - MCTI, pelo direcionamento paciente e preciso dos estudos. A Prof<sup>a</sup> Dra. Ingrid Winkler, pelo incentivo e suporte a produção de conteúdos científicos. Ao Prof<sup>o</sup> Dr. Roberto Badaró pelos esclarecimentos e apoio pertinentes aos aspectos médicos da pesquisa. Ao Adhvan Furtado e demais colegas de trabalho na Área de Desenvolvimento de Software do SENAI – CIMATEC, pelo incentivo ao meu ingresso no Programa de Pós-Graduação.



## 5. REFERÊNCIAS

- <sup>1</sup> YANG, Qiang and WU, Xindong. **10 challenging problems in data mining research.** International Journal of Information Technology & Decision Making, 5(04):597–604, 2006.
- <sup>2</sup> DIETTERICH, Thomas G. **Machine learning for sequential data: A review.** In Terry Caelli, Adnan Amin, Robert P. W. Duin, Dick de Ridder, and Mohamed Kamel, editors, Structural, Syntactic, and Statistical Pattern Recognition, pages 15–30, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.
- <sup>3</sup> KEOGH, Eamonn and KASETTY, Shruti. **On the need for time series data mining benchmarks: a survey and empirical demonstration.** Data Mining and knowledge discovery, 7(4):349–371, 2003.
- <sup>4</sup> LUENBERGER, David G. **Introduction to dynamic systems: theory, models, and applications**, volume 1. Wiley New York, 1979.
- <sup>5</sup> RABINER, L. and JUANG, B. **An introduction to hidden markov models.** IEEE ASSP Magazine, 3(1):4–16, Jan 1986.
- <sup>6</sup> LÜTKEPOHL, Helmut. **New introduction to multiple time series analysis.** Springer Science & Business Media, 2005.
- <sup>7</sup> TAYLOR, Graham William. **Composable, distributed-state models for high-dimensional time series.** University of Toronto, 2009.
- <sup>8</sup> MOAVENIAN, Majid and KHORRAMI, Hamid. **A qualitative comparison of artificial neural networks and support vector machines in ecg arrhythmias classification.** Expert Systems with Applications, 37(4):3088–3093, 2010.
- <sup>9</sup> BENGTON, Jason. **Scimedirect through sciverse: A new way to approach elsevier.** Medical Reference Services Quarterly, 30(1):42–49, 2011. PMID: 21271451.
- <sup>10</sup> TENOPIR, Carol. et al. **Academic users' interactions with scimedirect in search tasks: Affective and cognitive behaviors.** Information Processing & Management, 44(1):105–121, 2008.
- <sup>11</sup> ELSEVIER, B.V. **How do i use the new advanced search?**, 2018.
- <sup>12</sup> ANSI-AAMI EC57. **Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms.** Association for the Advancement of Medical Instrumentation, Arlington, VA, 1998.



<sup>13</sup> MOODY, George B and Mark, Roger G. **The impact of the mit-bih arrhythmia database.** IEEE Engineering in Medicine and Biology Magazine, 20(3):45–50, 2001.

<sup>14</sup> PhysioBank PhysioToolkit. **Physionet: components of a new research resource for complex physiologic signals.** Circulation. v101 i23. e215-e220.