



**UNIVERSIDADE ESTADUAL DA PARAÍBA
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
MESTRADO PROFISSIONAL EM CIÊNCIA E TECNOLOGIA EM SAÚDE**

GENILSON MEDEIROS MARTINS

**CLASSIFICAÇÃO DE SINAIS ELETROMIOGRÁFICOS PARA
AUXÍLIO AO DIAGNÓSTICO MÉDICO**

**CAMPINA GRANDE
2018**

GENILSON MEDEIROS MARTINS

**CLASSIFICAÇÃO DE SINAIS ELETROMIOGRÁFICOS PARA
AUXÍLIO AO DIAGNÓSTICO MÉDICO**

Dissertação de Mestrado submetido à banca
de avaliação do Programa de Pós-
Graduação em Ciência e Tecnologia em
Saúde da Universidade Estadual da Paraíba

Orientador: Prof. Dr. Robson Pequeno de Sousa

**CAMPINA GRANDE
2018**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

M379c Martins, Genilson Medeiros.
Classificação de sinais eletromiográficos para auxílio ao diagnóstico médico [manuscrito] / Genilson Medeiros Martins. - 2018.
68 p.
Digitado.
Dissertação (Mestrado em Profissional em Ciência e Tecnologia em Saúde) - Universidade Estadual da Paraíba, Pró-Reitoria de Pós-Graduação e Pesquisa, 2018.
"Orientação : Prof. Dr. Robson Pequeno de Sousa, Departamento de Computação - CCT."
1. Eletromiografia. 2. Diagnóstico médico. 3. Classificação.
I. Título
21. ed. CDD 615.5

GENILSON MEDEIROS MARTINS

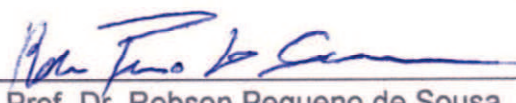
**CLASSIFICAÇÃO DE SINAIS ELETROMIOGRÁFICOS PARA
AUXÍLIO AO DIAGNÓSTICO MÉDICO**

Dissertação de Mestrado apresentada ao programa de Pós-Graduação em Ciência e Tecnologia em Saúde da Universidade Estadual da Paraíba, como requisito para obtenção título de Mestre.

Área de concentração: Tecnologia em Saúde

Aprovada em: 04/10/2018

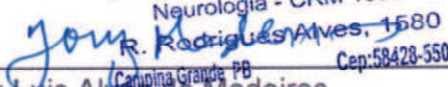
BANCA EXAMINADORA



Prof. Dr. Robson Pequeno de Sousa

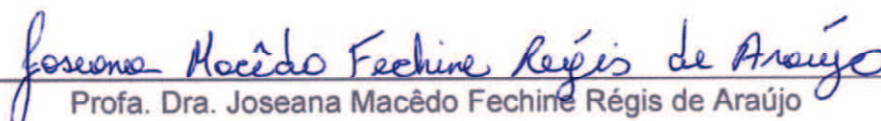
Orientador

Dr. Jovany Luis Alves de Medeiros
Neurologia - CRM 1682


R. Rodrigues Alves, 1580
Carolina Grande - PB Cep: 58428-550

Dr. Jovany Luis Alves de Medeiros

Membro Externo



Profa. Dra. Joseana Macêdo Fechine Régis de Araújo

Membro Externo

RESUMO

Estudo publicado em 2016, na revista *Improving Diagnosis in Health Care* pelo *Institute of Medicine* (IMO), afirma que um em dez diagnósticos estão incorretos e que pelo menos uma pessoa em cada três tem experiência relacionada a um erro de diagnóstico. Pesquisadores descobriram que erros em diagnósticos respondem pela maior fração de alegações de negligência e o maior quantitativo de pagamentos de penalidade como multas e indenizações. Na Eletromiografia, do ponto de vista clínico, o diagnóstico das doenças como a Esclerose Lateral Amiotrófica (ELA) e miopatias, nos estágios iniciais é de difícil diagnóstico, à medida que os sintomas dessas doenças podem ser facilmente confundidos com outras doenças. Contudo, neste trabalho é apresentado um sistema para auxílio ao diagnóstico médico (*Computer-Aided Diagnosis - CAD*), visando ao aprimoramento e ao aumento na eficiência, precisão e rapidez do diagnóstico clínico, capaz de classificar os sinais eletromiográficos automaticamente em saudável, miopatia e ELA, proporcionando uma segunda opinião ao especialista médico.

Palavras Chave: Eletromiografia, Classificação, Diagnóstico.

ABSTRACT

A study published in 2016 in the journal *Improving Diagnosis in Health Care* by the Institute of Medicine (IOM) states that one in ten diagnoses are incorrect and that at least one person in three has experience related to a diagnostic error. Researchers have found that errors in diagnoses account for the largest fraction of negligence claims and the highest amount of penalty payments such as fines and damages. From the clinical point of view, the diagnosis of diseases such as Amyotrophic Lateral Sclerosis (ALS) and myopathies in the early stages of diagnosis is difficult to diagnose, as the symptoms of these diseases can easily be confused with other diseases. However, this work presents a hybrid system to aid in the diagnosis (CAD), aiming at improving and increasing the efficiency, accuracy and speed of the clinical diagnosis, capable of automatically classifying electromyographic signals into healthy, myopathy and ELA, providing a second opinion to the medical specialist.

Keywords: Electromyography, Classification, Diagnosis.

LISTA DE FIGURAS

Figura 1 - Comparação entre um sistema computacional em 1993 com um sistema CAD recente.	11
Figura 2 - Sinal clássico de eletromiografia	13
Figura 3 - Unidade motora com um motoneurônio	14
Figura 4 - Potencial de ação que se propaga da esquerda para a direita. Mecanismos de condução no axônio	15
Figura 5 - Potencial de Unidade Motora - PUM.....	16
Figura 6 - Processo de aquisição do sinal.....	17
Figura 7 - EMG de um paciente com miopatia	24
Figura 8 - EMG de um paciente com ELA.....	24
Figura 9 - EMG de um paciente saudável	24
Figura 10 - Tendência de pesquisa sobre sistemas CAD	26
Figura 11 - Diagnóstico auxiliado por computador	28
Figura 12 - Conjunto de dados não linear	29
Figura 13 - Hiperplano ótimo para padrões linearmente separáveis.....	30
Figura 14 - Distâncias algébricas de pontos até o hiperplano ótimo para um caso bidimensional	32
Figura 15 - A arquitetura SVM empregando a função kernel.....	34
Figura 16 - Exemplo de classificação KNN	36
Figura 17 - Sistema Híbrido para Classificação de Sinais - SHCSEMG	37
Figura 18 - Sinal EMG com deslocamento de linha de base	40
Figura 19 - Sinal EMG após a filtragem.....	40
Figura 20 – Sinal EMG original e segmentado	42
Figura 21 - Etapas de classificação do sinal EMG	44
Figura 22 - Classificador SVM - ELA x Outro com Kernel de 0,1	46
Figura 23 - Classificador SVM - ELA x Outro com Kernel de 0,5.....	46
Figura 24 - Ilustração hipotética da classificação de um elemento desconhecido	48
Figura 25 - Avaliação de desempenho dos classificadores a partir da Matriz de Confusão	51
Figura 26 - Interface principal do protótipo	52
Figura 27 – Amostra de sinal EMG renderizado na tela do protótipo.....	52
Figura 28 - Selecionando um sinal EMG para análise	53

Figura 29 - Abrindo um arquivo EMG para análise.....	54
Figura 30 - Frequência de amostragem	54
Figura 31 - Exibição do gráfico EMG após seleção do arquivo de dados	55
Figura 32 - Exibição da classificação do Sinal EMG.....	55

LISTA DE TABELAS

Tabela 1 - Principais técnicas utilizadas para classificação de sinais	20
Tabela 2 - Áreas com mais pesquisas em sistemas CAD	27
Tabela 3 - Funções Kernel comuns.....	34
Tabela 4 - Classificador SVM para classificação de sinais com ELA, miopatia e saudável	56
Tabela 5 - Classificador KNN para classificação de sinais com ELA, miopatia e saudável	56
Tabela 6 - Matriz de Confusão (ELA x Outro).....	57
Tabela 7 - Matriz de Confusão (Miopatia x Saudável)	58
Tabela 8 - Desempenho do classificador SHCSEMG.....	58
Tabela 9 - Comparativo de desempenho entre os classificadores de sinais EMG utilizados pelos principais autores estudados e SHCSEMG	59

LISTA DE ABREVIATURAS

A/D - Analógico Digital
ACh - Acetilcolina
APSD - *Average Power Spectral Density*
CAD - *Computer-Aided Diagnosis*
ELA - Esclerose Lateral Amiotrófica
EMG - Eletromiografia
FFNN - *Feedforward Neural Network*
GBML - *Genetics Based Machine Learning*
KNN - *K-Nearest Neighbors*
LVQ - *Learning Vector Quantization*
MF - *Mean Frequency*
MLP - *Multi-Layer Perceptron*
MUAP - *Motor Unit Action Potential*
MUAPQ - *MUAP Quantity*
MeF - *Median Frequency*
P2P - *Peak to Peak Amplitude*
PAUM - Potenciais de Ação de Unidade Motora
PF - *Peak Frequency*
PM - Pico Máximo
PQ - *Phase Quantity*
RMS - *Root Mean Square*
RNA - Redes Neurais Artificiais
RP - Reconhecimento de Padrões
RS - *Rise Time*
SHCSEMG - Sistema Híbrido para Classificação de Sinais Eletromiográficos
SSC - *Slop Sign Changes*
SVM - *Support Vector Machine*
UM - Unidade Motora
WL - *Waveform Length*
ZCR - *Zero Crossing Rate*

SUMÁRIO

1 INTRODUÇÃO	10
1.1 RELEVÂNCIA.....	10
1.2 OBJETIVOS.....	12
2 FUNDAMENTAÇÃO TEÓRICA	13
2.1 ELETROMIOGRAFIA - EMG	13
2.2 PROCESSAMENTO DE SINAIS BIOLÓGICOS	17
2.3 ALGORITMOS DE EXTRAÇÃO DE ATRIBUTOS	18
2.4 CLASSIFICAÇÃO DE SINAIS EMG.....	19
2.5 PRINCIPAIS CARACTERÍSTICAS DAS DOENÇAS NEUROMUSCULARES	23
2.6 SISTEMA DE DIAGNÓSTICO AUXILIADO POR COMPUTADOR - CAD	25
3 CLASSIFICADOR DE SINAIS	29
3.1 MÁQUINAS DE VETORES DE SUPORTE – SVM.....	29
3.2 K-VIZINHOS MAIS PRÓXIMOS – KNN.....	34
4 MÉTODO PROPOSTO E RESULTADOS PRELIMINARES	37
4.1 SISTEMA HÍBRIDO PARA CLASSIFICAÇÃO DE SINAIS EMG - SHCSEMG.....	37
4.2 PRÉ-PROCESSAMENTO DO SINAL ELETROMIOGRÁFICO.....	38
4.2.1 Limiar Suave (<i>Soft-Thresholding</i>).....	39
4.2.2 Filtro Butterworth.....	40
4.3 SEGMENTAÇÃO	41
4.4 EXTRAÇÃO DE CARACTERÍSTICAS	42
4.5 CLASSIFICADOR DE SINAIS EMG	43
4.5.1 Classificador SVM	44
4.5.2 Classificador KNN.....	47
4.6 MÉTRICAS DE AVALIAÇÃO	48
5 RESULTADOS E DISCUSSÕES	50
5.1 DESENVOLVIMENTO DO PROTÓTIPO	50
5.1.1 Visão Geral.....	51
5.2 Funcionamento da ferramenta.....	53
5.2 CÁLCULO DE DESEMPENHO	55
6 CONSIDERAÇÕES FINAIS	60
7 PROPOSTAS FUTURAS	61
REFERÊNCIAS	62

1 INTRODUÇÃO

O diagnóstico a partir da análise computadorizada, conhecido como diagnóstico auxiliado por computador (*Computer-Aided Diagnosis - CAD*) é uma ferramenta de suporte importante para reduzir o tempo necessário para interpretação de imagens e sinais biológicos, bem como prover uma segunda opinião, tornando a decisão final do especialista mais rápida e com maior confiança (Astakhov, 2009).

Por outro lado, do ponto de vista clínico, o diagnóstico das doenças como a Esclerose Lateral Amiotrófica (ELA) e miopatias, nos estágios iniciais é difícil, à medida que os sintomas dessas doenças podem ser facilmente confundidos com outras doenças (Joshi *et al.*, 2017).

Consequentemente, o especialista médico tem que distinguir entre inúmeros tipos de patologias que muitas vezes não são facilmente interpretáveis e exigem certa experiência. Acrescenta-se prováveis situações em que os profissionais estão expostos a ambientes estressantes e longos períodos de trabalho, os quais, podem promover a fadiga visual, distração, estresse e cansaço mental, diminuindo, consequentemente, sua performance e a capacidade de observação (Fuglsang-Frederiksen, 2000).

1.1 RELEVÂNCIA

Os sistemas CAD têm sido desenvolvidos para ajudar especialistas a aumentar a acurácia no diagnóstico médico (Kong *et al.*, 2009). Uma grande diferença entre CAD e diagnóstico por computador é a forma como a saída do sistema computacional é utilizada. Com o CAD, os especialistas usam a informação obtida como uma "segunda opinião", e os profissionais tomam as decisões finais, dessa forma, tornando sua decisão mais rápida e mais precisa (Astakhov, 2009). Por outro lado, o diagnóstico automatizado de computadores é um conceito baseado em algoritmos de computador (Takahashi e Kajikawa, 2017).

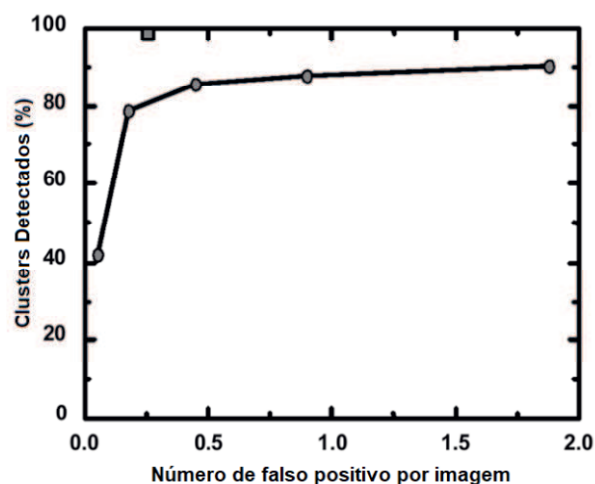
Com o diagnóstico automatizado de computadores, o desempenho dos equipamentos precisa ser comparável ou melhores que aqueles exercidos pelos

médicos. Com o CAD, no entanto, esse desempenho não precisa ser comparado, mas complementado (Dempsey, 2005).

Os Sistemas de Diagnóstico Auxiliado por Computador tornaram-se parte do trabalho clínico de rotina para detecção de câncer de mama em mamografias, doenças pulmonares, ultrassonografia, lesões na pele, doenças cerebrais, entre outras, em muitas clínicas e hospitais. Isto indica que as ferramentas CAD estão começando a ser aplicadas amplamente na detecção e diagnóstico de muitos tipos diferentes de anormalidades em imagens médicas e sinais biológicos, obtidos em vários exames (Takahashi e Kajikawa, 2017).

Na Figura 1 é exibido um comparativo do nível de desempenho anterior, marcado por círculos na detecção de microcalcificações agrupadas por um sistema computacional em 1993, em que se obteve 87% de sensibilidade em 1.0 falso positivo por imagem. Quando a tecnologia CAD foi licenciada para uma empresa em 2007, com o nível de desempenho atual estimado, marcado por um quadrado cinzento, proporcionou 98% de sensibilidade em 0,25 falso positivo por imagem. É possível perceber uma melhoria de 11%.

Figura 1 - Comparação entre um sistema computacional em 1993 com um sistema CAD recente.



Fonte: Doi (2007).

Existem inúmeras técnicas de classificação de sinais que podem ser empregadas em sistemas CAD. Adicionalmente, esta pesquisa propõe um sistema híbrido constituído de dois classificadores de sinais, capaz de classificar sinais eletromiográficos (EMG) em saudável, ELA e miopático.

1.2 OBJETIVOS

Os sinais biológicos, como os eletromiográficos, são de natureza aleatória, isto é, apresentam natureza não estacionária e possuem grande variação (Joshi *et al.*, 2017). Portanto, é uma tarefa difícil diagnosticar com precisão doenças a partir da análise desses sinais biológicos. O método tradicional do diagnóstico da doença é a partir da análise visual do sinal EMG.

O diagnóstico das doenças neuromusculares nos estágios iniciais é difícil à medida que os sintomas dessas doenças podem ser facilmente confundidos com outras doenças. Desta forma, o especialista médico tem que distinguir entre inúmeros tipos de patologias, que muitas vezes não são facilmente interpretáveis, exigindo certa experiência do especialista, acrescentando ainda, prováveis situações em que os profissionais estão expostos a ambientes estressantes e longos períodos de trabalho, podendo promover a fadiga visual, distração, estresse e cansaço mental diminuindo, conseqüentemente, sua performance e a capacidade de observação (Criswell, 2011). Neste sentido, a seguir, são apresentados os objetivos gerais e específicos referentes a esta pesquisa de mestrado.

Objetivo Geral

Estudos revelam que a dupla análise pode aumentar a sensibilidade do diagnóstico. Portanto, o objetivo geral desta pesquisa é desenvolver um protótipo para classificação de sinais eletromiográficos, capaz de identificar e classificar o sinal eletromiográfico em normal, miopático e Esclerose Lateral Amiotrófica, proporcionando e disponibilizando uma segunda opinião ao especialista médico.

Objetivos Específicos

- Melhorar a acurácia do diagnóstico;
- Auxiliar o especialista no diagnóstico de doenças neuromusculares como miopatia e Esclerose Lateral Amiotrófica;
- Diminuir o tempo gasto no processo de diagnóstico e
- Proporcionar uma segunda opinião durante o processo de diagnóstico.

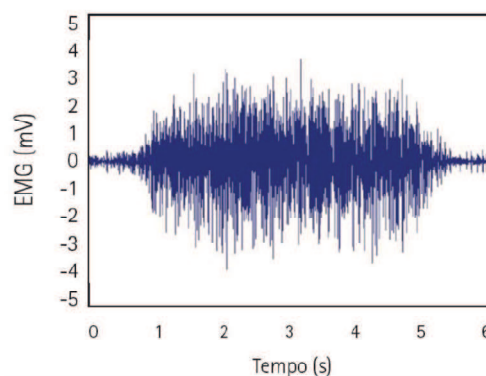
2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo fornecer embasamento teórico sobre conceitos necessários para melhor compreensão da pesquisa. Serão introduzidos os conceitos relacionados à eletromiografia, processamento de sinais biológicos, aos algoritmos de extração de atributos, a classificação de sinais EMG, as principais características das doenças neuromusculares e Sistema de Diagnóstico Auxiliado por Computador, assuntos relevantes para a pesquisa.

2.1 ELETROMIOGRAFIA - EMG

A eletromiografia é a representação gráfica e o estudo das características elétricas dos músculos que, em repouso são normalmente silenciosos. No entanto, quando o músculo está ativo, como durante a contração ou estimulação, conforme apresentado na Figura 2, são produzidas correntes elétricas e os sucessivos Potenciais de Ação de Unidade Motora (PAUM), que resultam da descarga repetitiva de grupos de fibras musculares, chamadas Unidade Motoras (UM). A gravação visual dos dados é chamada de eletromiograma (Hess *et al.*, 2016).

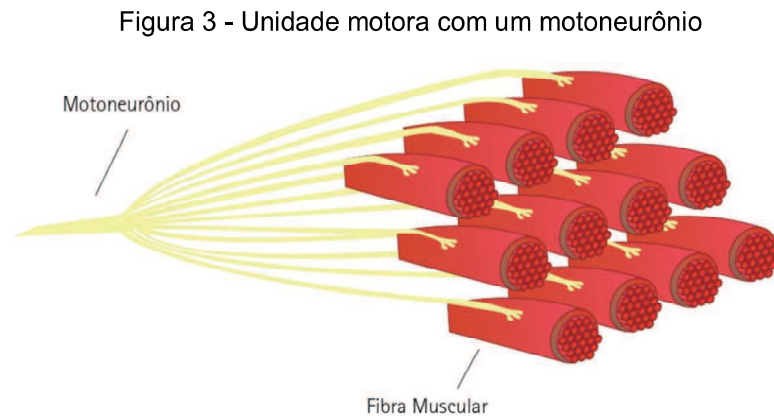
Figura 2 - Sinal clássico de eletromiografia



Fonte: SCHEEREN; KRUEGER; FREITAS, 2015, p.29

O termo UM refere-se coletivamente a um motoneurônio e a um grupo de fibras musculares por ele inervadas. A UM é a menor unidade de músculo esquelético que pode ser ativada por esforço físico. Elas são consideradas fibras de contração muscular, pois produzem uma resposta de contração mecânica para um

único estímulo e geram um potencial de ação que se propaga (Rangayyan, 2015). O modelo esquemático de uma unidade motora com um motoneurônio e 12 fibras por ele inervadas é representado na Figura 3.



Fonte: SCHEEREN; KRUEGER; FREITAS (2015)

Os PAUM de diferentes UM tendem a ter formas distintas, que permanecem semelhantes para cada descarga. Os PAUM podem, portanto, ser identificados e analisados utilizando técnicas de reconhecimento de padrões (Bhoi, Sherpa e Mallick, 2014), que serão abordados posteriormente. A informação resultante pode ser utilizada para determinar a origem da doença.

O nível básico de organização do sistema nervoso do músculo é a unidade motora juntamente com o sistema motor alfa associado, isto é, o neurônio motor inferior, seu axônio e as fibras musculares que inerva (Criswell, 2011).

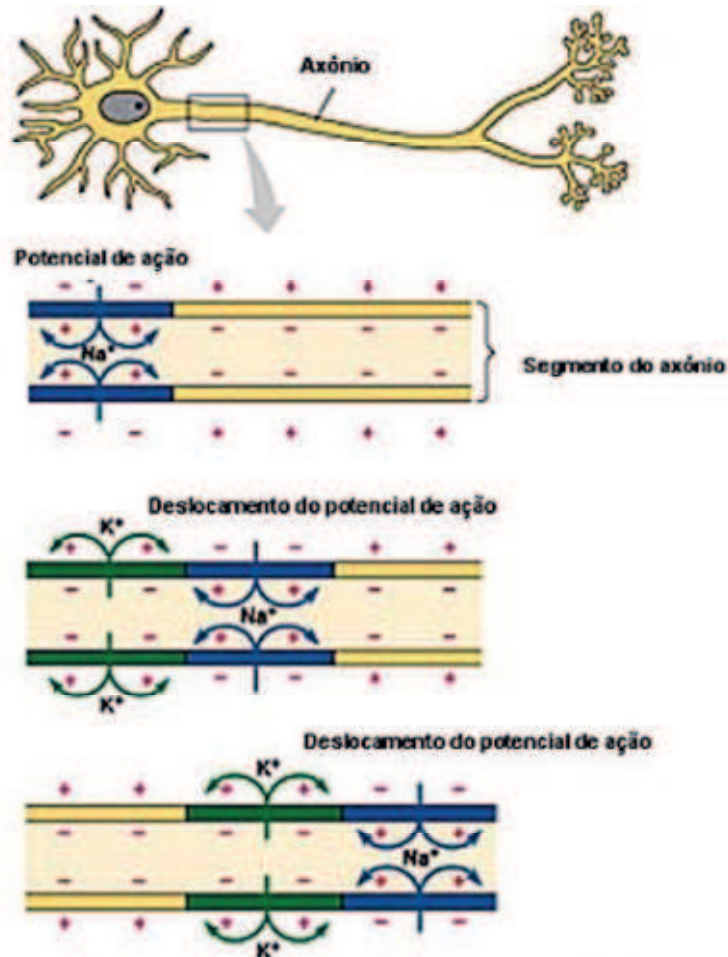
O axônio motor inferior ramifica-se de modo que possa se prender à fibra muscular na placa final do motor, criando sinapses neuromusculares (Figura 4). Quando um potencial de ação do nervo percorre o axônio, alcança a sinapse neuromuscular e libera acetilcolina (ACh), que causa a quebra da barreira iônica do tecido muscular e envia o sinal por todo o sistema a partir dos túbulos transversais. Isso cria o potencial de ação da unidade e a contração muscular (Scheeren, Krueger e Freitas, 2015).

O registro extracelular dessa troca de energia fornece a base para a eletromiografia. Os potenciais de ação estão associados a um aumento súbito na permeabilidade da membrana celular aos íons sódio (Na^+ , despolarização).

Isso faz com que um súbito influxo de Na^+ na fibra muscular que está associado a uma mudança mensurável no potencial de repouso da célula. Perto do

pico deste influxo, um rápido efluxo de íons de potássio (K^+) faz com que ocorra uma repolarização rápida da célula (Towe, 2004). Na Figura 4 é apresentado a direção da propagação dos potenciais de ação nas fibras musculares.

Figura 4 - Potencial de ação que se propaga da esquerda para a direita. Mecanismos de condução no axônio



Fonte: CRISWELL (2011)

O sinal biológico EMG é obtido da medição de tensão elétrica associada com correntes elétricas geradas a partir do músculo durante contrações, promovendo a aferição de atividades neuromusculares (Yousefi e Hamilton-Wright, 2014).

O potencial de unidade motora (PUM) é um somatório dos potenciais de ação das contrações das fibras musculares na UM, e são medidas inserindo um eletrodo de agulha no músculo ou, no caso da eletromiografia de superfície, utilizando-se eletrodos na superfície na pele. Segundo Katirji (2007), as tensões elétricas detectadas pelo eletrodo representam o somatório da atividade de todas as UM ativas, ou seja, a soma da contribuição de todos os PUM.

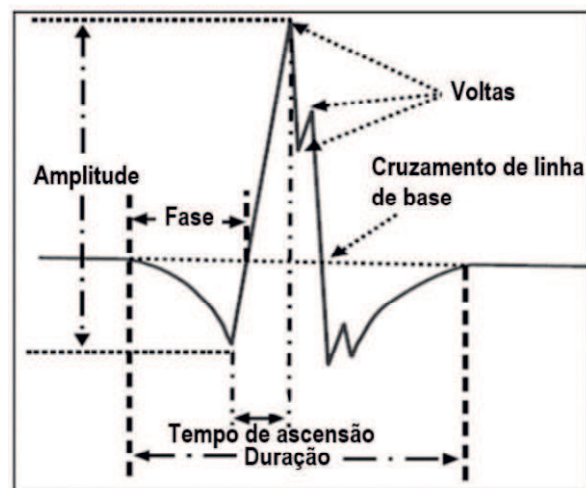
Matematicamente, o sinal EMG, usando um eletrodo para capturar uma contração muscular, pode ser representado conforme a Equação 1.

$$EMG_T = \sum_{m=1}^N MUPT_m t + n t \quad (1)$$

Em que, $MUPT_m$ refere-se à função que descreve a contribuição de tensão, medida durante o tempo da contração de um único PUM m de um total de N PUM. A variação do tempo é representada pela variável t e a função $n(t)$ representa todos os componentes do sinal que não estão associados ao PUM, entendidos como as contribuições de ruído biológico e ambiental no sinal gravado. Ambas as funções são parametrizadas pelo tempo (Raez *et al.*, 2006).

Um típico PAUM é exibido na Figura 5. Analisando a forma de onda no domínio do tempo, é possível apontar algumas características importantes como: amplitude, fase, duração, tempo de ascensão, voltas e cruzamento de linha de base (Yousefi e Hamilton-Wright, 2014).

Figura 5 - Potencial de Unidade Motora - PUM



Fonte: Yousefi, Hamilton-Wrihht (2014)

Conhecendo as características do sinal EMG, é possível extrair alguns atributos, utilizando técnicas matemáticas de extração de características. Também é possível extrair atributos no domínio da frequência, com por exemplo, frequência média, valor quadrado médio, dentre outros.

Ao extrair os atributos relevantes, os dados são armazenados em vetores de características, que posteriormente serão processados e classificados (Onsy *et al.*, 2012).

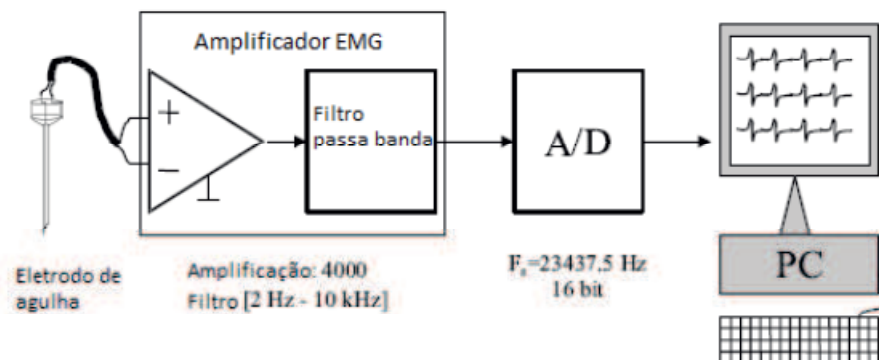
2.2 PROCESSAMENTO DE SINAIS BIOLÓGICOS

A bioeletricidade está presente em todos os seres vivos, na eletromiografia, a atividade elétrica é gerada após a contração muscular. As células nervosas e musculares (fibras musculares) são células excitáveis. Estas células, reagem a todos os eventos externos, também conhecidos como excitações, que são de natureza mecânica, química ou elétrica (Negi, 2016).

Potenciais elétricos constantes e variáveis no tempo podem ser medidos mediante o uso de um transdutor, também conhecidos como sensores, que podem ser invasivos ou de superfície (Towe, 2004).

Devido à baixa amplitude do sinal EMG durante sua aquisição, faz-se necessário amplificar o sinal, filtrar e quantizar para, posteriormente, ser processado (MARCHETTI; DUARTE, 2006). Na Figura 6 é apresentado o processo de aquisição do sinal EMG e suas etapas.

Figura 6 - Processo de aquisição do sinal



Fonte: JALI (2015)

Conforme exemplificado na Figura 6, um eletrodo de agulha concêntrica é inserido no músculo do paciente para captação dos pulsos elétricos gerados. Durante a captação do sinal EMG do tecido neuromuscular, outros sinais indesejados da rede elétrica, bem como de outros equipamentos elétricos, podem

interferir no registro do sinal. Como o sinal foi amplificado em 4000 vezes (no exemplo), os ruídos inerentes do processo também foram amplificados. Para tanto, é necessário o uso de filtros digitais para eliminar as frequências indesejadas.

Um filtro passa banda de 2 Hz à 10 KHz foi utilizado por Nikolic, Miki e Krarup (2011). Em seguida, o sinal foi amostrado e digitalizado em uma frequência de amostragem (F_s) de 23437,5Hz em resolução de 16 bits. Foi utilizado o dispositivo Motorola DSP56ADC16.

Com os dados digitalizados, a informação pode ser processada por um software e exibida de forma correta para o usuário especialista (MARCHETTI; DUARTE, 2006).

2.3 ALGORITMOS DE EXTRAÇÃO DE ATRIBUTOS

A extração de características é uma técnica na qual um subconjunto de recursos relevantes é selecionado, a partir de dados existentes, e aplicado em um algoritmo de aprendizado (Negi, 2016). Essa etapa é uma das mais importantes, pois os algoritmos de extração de características utilizados devem ser capazes de obter um bom desempenho, ao ponto de atribuir uma relação entre os atributos selecionados à classe em estudo.

Segundo Duda, Hart e Stork (2000) o limite conceitual entre a extração de características e a classificação propriamente dita é algo arbitrário: um extrator de características ideal produziria uma representação que tornasse trivial o trabalho do classificador, por outro lado, um classificador onipotente não precisaria da ajuda de um sofisticado extrator de características.

A distinção é imposta por razões práticas, e não teóricas. De um modo geral, a tarefa de extração de recursos é muito mais problemática e dependente do domínio do que a classificação adequada e, portanto, requer conhecimento do domínio (Lahmiri, 2017).

A seleção de características geralmente cria um modelo que facilita a generalização das dimensões invisíveis e pode melhorar substancialmente a compreensão do modelo classificador que é produzido (Valencia *et al.*, 2016).

Existem vários algoritmos utilizados na literatura para extração de características do sinal EMG, dentre os quais, os mais comumente utilizados são:

- Valor Médio Quadrado (*Root Mean Square - RMS*);
- Taxa de Cruzamento por Zero (*Zero Crossing Rate - ZCR*);
- Amplitude de Pico à Pico (*Peak to Peak Amplitude - P2P*);
- Mediana da frequência (*Median Frequency – MeF*);
- Média de Densidade Espectral de Potência (*Average Power Spectral Density – APSD*);
- Tamanho da forma de onda (*Waveform Length – WL*);
- Frequência de Pico (*Peak Frequency – PF*);
- Frequência Média (*Mean Frequency – MF*);
- Quantidade de MUAP (*MUAP Quantity – MUAPQ*);
- Pico Máximo (*Peak*);
- Tempo de Subida (*Rise Time – RS*);
- Quantidade de Fases (*Phase - PQ*);
- Mudança de Inclinação do Sinal (*Slop Sign Changes – SSC*).

Para cada classe à qual se deseja classificar o sinal EMG, bem como o classificador selecionado, um conjunto de algoritmos de extração de características deve ser selecionado com o objetivo de obter o melhor cenário entre eles.

2.4 CLASSIFICAÇÃO DE SINAIS EMG

Nos últimos vinte anos, a eletromiografia tem sido amplamente utilizada por pesquisadores e médicos como uma ferramenta valiosa para um diagnóstico preciso de distúrbios ou transtornos neuromusculares (Lahmiri, 2017). Transtorno neuromuscular é um termo geral que se refere a doenças que afetam qualquer parte do nervo ou músculo, incluindo neurônios motores, junções neuromusculares e tecido muscular.

Existem vários grupos críticos de doenças neuromusculares, e o discernimento entre suas características, bem como entre o estado de saudável e não saudável, são objetivos típicos de um sistema baseado em classificação de sinais.

Embora historicamente os dados EMG tenham sido abordados qualitativamente (Mishra *et al.*, 2016), nos últimos anos um grande interesse foi encontrado na análise EMG quantitativa chamada QEMG (Sengar *et al.*, 2017), a

partir do qual uma série de medidas quantitativas do sinal EMG são analisadas. Para caracterizar um músculo usando dados QEMG, os sinais adquiridos devem ser analisados, decompostos e classificados. Ao diagnosticar distúrbios neuromusculares, a classificação do sinal EMG, em diferentes grupos, é usada na detecção de anormalidades.

Na Tabela 1 é fornecido o resumo histórico das técnicas mais utilizadas na literatura, bem como resultados mais recentes e contexto histórico importantes, a fim de entender melhor as tendências recentes na classificação de sinais EMG. As técnicas são discriminadas pelo tipo de classificador utilizado e acurácia.

Tabela 1 - Principais técnicas utilizadas para classificação de sinais

Ano	Técnica	Referência	Acurácia (%)
1992	Self-organizing feature map (SOFM)	(Iordanova, Rialle e Vila, 1992)	76-83
1995	Integration of parametric pattern recognition algorithm (PPR) and artificial neural network (ANN) SOFM	(Pattichis, Schizas e Middleton, 1995)	80-90
1996	SOFM and learning vector quantization (LVQ)	(Abel <i>et al.</i> , 1996)	60-80
1998	Modular ANN	(Christodoulou e Pattichis, 1998)	79,6
1999	SOFM, LVQ and statistical methods based on Euclidean distance	(Christodoulou e Pattichis, 1999)	90
2004	Continuous wavelet transform (CWT) and a multi-channel ANN	(Gazzoni, Farina e Merletti, 2004)	-
2005	Multi-layer perceptron (MLP)	(Güler e Koçer, 2005)	91,6
2006	Wavelet-based neural network (WNN)	(Subasi, Yilmaz e Ozcalik, 2006)	90,7
2007	Radial basis networks (RBN) and decision trees	(Katsis <i>et al.</i> , 2007)	89
2012	SOFM and LVQ	(BHARDWAJ; MANISH; ARORA, 2012)	97,6
2012	Principal component analysis (PCA) and probabilistic neural network (PNN)	(SHAW; BAGHA, 2012)	91,72
2013	PCA and PNN	(Manuel e Acevedo, 2013)	68-94,3
Fuzzy			
2001	Fuzzy logic	(Chauvet <i>et al.</i> , 2001)	88,4
2006	Adaptive fuzzy k-NN classifier (AFNNC)	(Rasheed, Stashuk e Kamel, 2006)	96,6
2012	Fuzzy logic	(DISSELHORST-KLUG, 2012)	97
RNAs Híbridos			
1996	Combined ANN and genetics-based machine learning (GBML) models	(Pattichis e Schizas, 1996)	80
2004	Fuzzy integral of multiple ANN	(Xel, Huang e Wan, 2004)	88,58

2010	Neuro-fuzzy system (NFS)	(Koçer, 2010)	90
Máquinas de vetores de suporte			
2002	Support vector machine (SVM) with one against one training algorithm	(Katsis <i>et al.</i> , 2006)	89
2005	SVM	(Güler e Koçer, 2005)	92,3
2009	Multiclass SVM	(Kaur, Arora e Jain, 2009)	100
2010	Binary SVM	(Kaur, 2010)	100
2010	Fuzzy support vector machine (FSVM)	(Koçer, 2010)	99,6
2010	SVM	(Istenič <i>et al.</i> , 2010)	70,4
2012	SVM with wavelet technique for feature extraction	(Dobrowolski, Wierzbowski e Tomczykiewicz, 2012)	99,4
2012	SVM	(Tomczykiewicz, Dobrowolski e Wierzbowski, 2012)	99,3
2012	FSVM classifier combined with statistical features extracted from discrete wave transform (DWT)	(Subasi, 2012)	97,67
2013	Hybridization of the particle swarm optimization (PSO) and SVM	(Subasi, 2013)	96,75
Outros			
1995	Principle component analysis and multivariate discriminant algorithm	(Kaur, 2010)	70,4-76,5
2008	Bayesian aggregation	(Hamilton-Wright <i>et al.</i> , 2010)	88,7
2012	Decision tree	(KELEŞ; SUBAŞI, 2012)	96,33-96,50
Redes neurais artificiais			
2014	Machine Learning for Supporting Diagnosis of Amyotrophic Lateral Sclerosis Using Surface Electromyogram	(Xu Zhang <i>et al.</i> , 2014)	90
2017	Identification of Amyotrophic Lateral Sclerosis using EMG Signals	(Sengar <i>et al.</i> , 2017)	92
Abordagem Híbrida			
2017	A Hybrid Real-Time EMG Intelligent Rehabilitation Robot Motions Control based on Kalman Filter, Support Vector Machines and Particle Swarm Optimization	(ELBAGOURY; VLADAREANU, 2017)	84-96
2017	Hybrid Water Wave Optimization and Support Vector Machine to Improve EMG Signal Classification for Neurogenic Disorders	(FDWLRQ; LVRUGHUV, 2017)	91,44

Fonte: Adaptado de Yousefi; Hamilton-Wright (2014)

Vários estudos empregam métodos de classificação de sinais, incluindo técnicas bayesianas, perceptrons multicamadas, abordagens fuzzy, máquina de vetor de suporte e sistemas neuro-fuzzy. Os autores mostraram que as Redes Neurais Artificiais (RNA) é o método mais popular de classificação (medido pelo número de implementações), para a classificação de sinais EMG (Yousefi e Hamilton-Wright, 2014).

Vários modelos diferentes de RNA foram utilizados, incluindo o perceptron multicamadas (*multi-layer perceptron* - MLP), Feedforward (*Feedforward Neural*

Network - FFNN), *Backpropagation*, modular NN (Mod NN), a rede neural recorrente (*Recurrent Neural Network - RNN*), os mapas de características auto-organizados (*Self-Organizing Feature Maps - SOFM*) e a rede neural probabilística (*Probabilistic - PNN*).

Aplicar um algoritmo de aprendizado apropriado é fator muito importante para atingir maior precisão. Por exemplo, o *Backpropagation* necessita de longos períodos de aprendizagem, especialmente quando o conjunto de dados é grande, e também possui baixo desempenho na discriminação de aprendizagem na classificação do sinal devido à alta variabilidade nos dados EMG. Os SOFM podem superar esses inconvenientes.

Como consequência, pode-se perceber que a precisão dos modelos de RNA pode ser aprimorada usando o método de aprendizado SOFM combinado com LVQ, pode obter 97,6% de acurácia (Christodoulou e Pattichis, 1999). O uso adicional do LVQ melhora a precisão da classificação. Além disto, o modelo SOFM possui o menor custo computacional em comparação com o *Backpropagation* e os algoritmos GBML (PATTICHIS; SCHIZAS, 1996).

Bose, Samanta e Chatterjee (2016) apontam os classificadores SVM (*Support Vector Machine - SVM*) e KNN (*K-Nearest Neighbors - KNN*) como uns dos mais eficientes e utilizados por pesquisadores em diferentes áreas. Neste sentido, Goen (2014) propôs um classificador de sinais EMG para avaliar doenças neuromusculares, como neuropatia e miopatia, utilizando as SVM.

Para medir a eficiência do classificador foram calculados a sensibilidade e especificidade e obteve 87,5% e 88,8%, respectivamente. Entretanto, outros autores obtiveram melhores resultados como Subasi (2013), com 95% e Bose, Samanta, Chatterjee (2016) com 100% de acurácia para classificação de até duas classes.

Kucuk, Tepe e Eminoglu (2013) propuseram um classificador de sinais EMG utilizando KNN para identificação de duas classes de sinais: saudável e ELA. O estudo mostrou que o classificador KNN obteve acuraria de 96,1%, com desempenho um pouco melhor em relação aos classificadores que utilizam as SVM.

Em outro estudo, Bose, Samanta e Chatterjee (2016), demonstram que os sistemas híbridos vêm obtendo um melhor desempenho na classificação dos sinais EMG.

A abordagem híbrida é a junção de dois ou mais métodos e ainda pouco utilizada, porém percebe-se o crescimento de sua utilização nos trabalhos

relacionadas a classificação de sinais e imagens (Moraes, 2010). Na investigação de métodos para classificação de sinais, constata-se que cada método isolado tem uma deficiência específica. Ao aliar dois ou mais métodos em modo de compensação, os resultados podem ser potencializados, afirma Bose, Samanta e Chatterjee (2016).

2.5 PRINCIPAIS CARACTERÍSTICAS DAS DOENÇAS NEUROMUSCULARES

Os sinais EMG são amplamente usados no campo do diagnóstico clínico de diferentes distúrbios neuromusculares como a miopatia, Esclerose Lateral Amiotrófica, dentre outras (Zhao *et al.*, 2016).

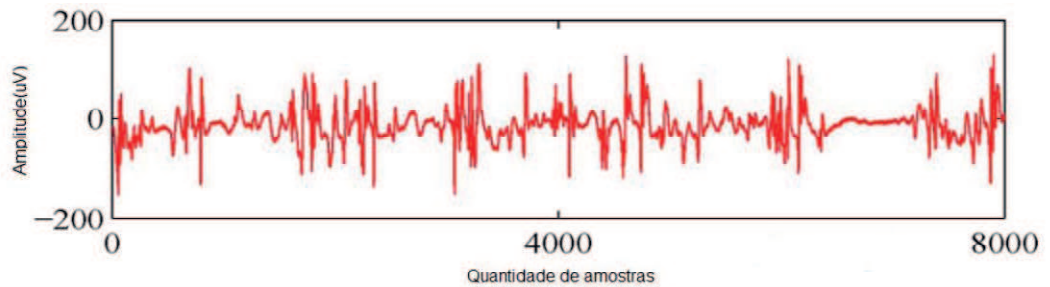
A miopatia descreve um grupo de doenças que afetam o tecido muscular esquelético. É uma das doenças hereditárias que causa fraqueza muscular devido ao mau funcionamento das fibras musculares ligadas aos ossos (Preston e Shapiro, 2013).

A Esclerose Lateral Amiotrófica (ELA) é outro tipo de doença neurológica fatal, rapidamente progressiva, que ataca os neurônios responsáveis pelo controle dos músculos voluntários. Esta doença se enquadra na categoria de doenças neuronais motoras, que são caracterizadas por degeneração gradual e, eventualmente, a morte de neurônios motores (Chawla, 2011).

Existem diferenças significativas nas características e tratamento das doenças das células musculares e lesão do nervo. Uma técnica de análise do sinal EMG pode ser realizada por meio do espectro de frequência, que fornece informações sobre a distribuição da potência em relação a um sinal EMG saudável. Também é possível realizar análise do sinal no domínio do tempo. Em geral, é possível visualizar algumas características como amplitude, duração e fase que se destacam visualmente.

No sinal miopático, percebe-se como principal característica, no domínio do tempo, os PAUM com curta duração e amplitude reduzida em comparação com o sinal saudável, conforme apresentado na Figura 7.

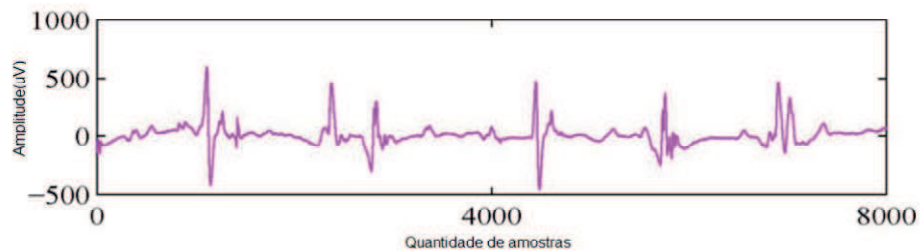
Figura 7 - EMG de um paciente com miopatia



Fonte: PHYSIONET.ORG (2017)

No sinal de um paciente com ELA (Figura 8), percebe-se como principal característica, no domínio do tempo, os PAUM com longa duração e amplitude maior em comparação com o sinal saudável (Figura 9).

Figura 8 - EMG de um paciente com ELA



Fonte: PHYSIONET.ORG (2017)

Ambos os sinais, dos pacientes com miopatia e ELA, são comparados ao sinal EMG de um homem saudável (Figura 9) de 44 anos, sem histórico de doença neuromuscular (PHYSIONET.ORG, 2017).

Figura 9 - EMG de um paciente saudável



Fonte: PHYSIONET.ORG (2017)

De acordo com Physionet.org (2017), os dados das Figuras 7, 8 e 9 foram obtidos a partir de um Sistema de Monitorização EMG, *Medelec Synergy N2 (Oxford Instruments Medical, Old Woking)*, Reino Unido. Um eletrodo de agulha concêntrico de 25 mm foi colocado no músculo tibial anterior de cada indivíduo. Pediu-se então ao paciente que flexibilizasse suavemente o pé contra uma resistência. O eletrodo de agulha foi reposicionado até que fossem identificados potenciais de unidade motora com um rápido tempo de resposta. Os dados foram, então, coletados durante vários segundos.

2.6 SISTEMA DE DIAGNÓSTICO AUXILIADO POR COMPUTADOR - CAD

Criswell (2011) afirma que a história da eletromiografia foi marcada pela descoberta da eletricidade e o desenvolvimento da capacidade de ver, por meio da ajuda de instrumentos, coisas que não podiam ser vistas, sentidas ou tocadas. Importantes avanços no campo da visão computacional em paralelo com técnicas de inteligência artificial e mineração de dados tem facilitado o desenvolvimento de ferramentas automáticas de análise médica e interpretação de dados (Meyer-Baese *et al.*, 2014).

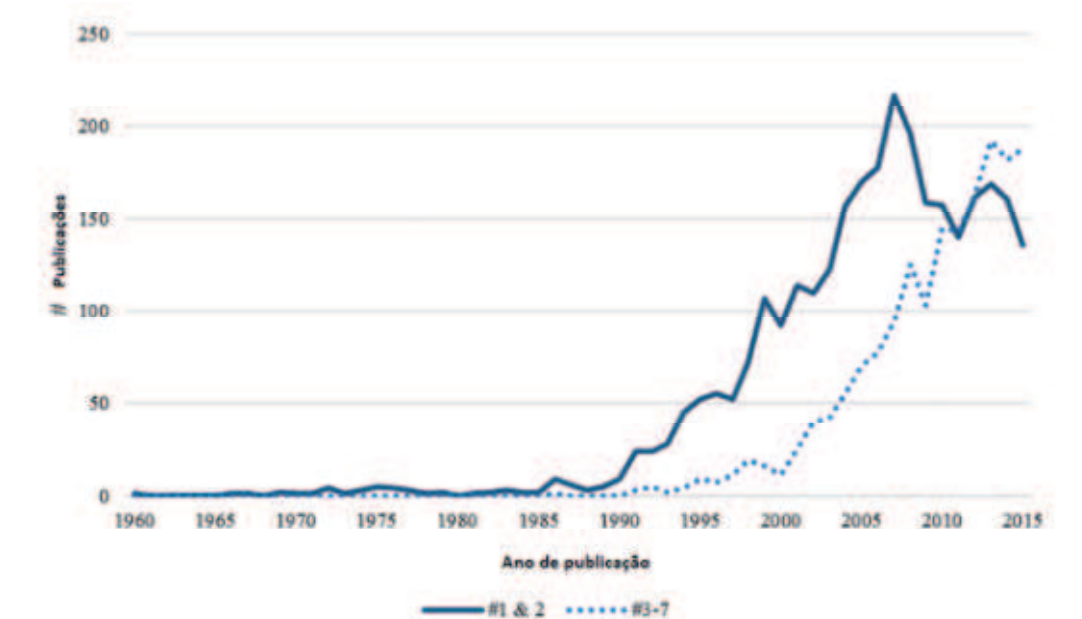
Os sistemas CAD são o resultado dessas pesquisas, uma vez que tem como objetivo aprimorar a eficiência e a precisão de diagnósticos clínicos, podendo detectar e classificar automaticamente anormalidades e doenças em exames médicos, proporcionando uma segunda opinião ao especialista médico (Astakhov, 2009).

Existem várias pesquisas em CAD visando otimizar o trabalho do especialista médico. Na Figura 10 é demonstrado o crescimento dessas pesquisas ao longo dos anos. A linha sólida mostra a soma da quantidade de publicações em cada ano nos agrupamentos # 1 (CAD de câncer de mama) e # 2 (CAD de doença pulmonar). Esses dois agrupamentos, que são os maiores e mais antigos, contém 2973 artigos (TAKAHASHI; KAJIKAWA, 2017).

Por outro lado, a linha pontilhada expressa a mesma informação para os agrupamentos # 3 (CAD de ultra-sonografia), # 4 (CAD de lesão de pele), # 5 (CAD de câncer colorretal), # 6 (CAD de doença cerebral) e # 7 (vários assuntos). Estes cinco agrupamentos continham 1732 publicações. Os resultados apresentados

(Figura 10) ajudam a concluir que a pesquisa relacionada a ferramentas CAD típica, ou seja, para câncer de mama e câncer pulmonar, caiu no início da primeira década do século XXI. Por outro lado, a pesquisa em outros domínios CAD vem progredindo de forma constante desde então.

Figura 10 - Tendência de pesquisa sobre sistemas CAD



Fonte: Takahashi e Kajikawa (2017)

Takahashi e Kajikawa (2017) realizaram uma pesquisa bibliométrica com mais de 4705 artigos sobre sistemas CAD e dividiu em grupos que dependiam de sua topologia de citações diretas. A Tabela 2 exibe a visão geral da pesquisa realizada. Ela está dividida em campo de pesquisa, ano de publicação, quantidade de artigos encontrados, número e taxa de citações, por fim, as palavras chaves utilizadas na pesquisa.

Finalmente, ainda baseando-se nos dados da Tabela 1, pode-se apontar que existem muitas pesquisas que tratam sobre os sistemas CAD e que demonstram o aumento do interesse dos pesquisadores por essa poderosa ferramenta, principalmente nas áreas os quais trabalham com artefatos visuais como imagem a exemplo da monografia, doenças pulmonares, dentre outras.

Todavia, é perceptível que, na revisão bibliométrica realizada por Takahashi e Kajikawa (2017), existem poucos trabalhos desenvolvidos em sistemas CAD com o foco em eletromiografia. Tal cenário é expresso no grupo #7, em que o campo de

estudo da eletromiografia está contido na menor e a mais variada área de pesquisa, em “CAD para vários assuntos”, contendo apenas 107 artigos publicados e mapeados na revisão bibliométrica.

Tabela 2 - Áreas com mais pesquisas em sistemas CAD

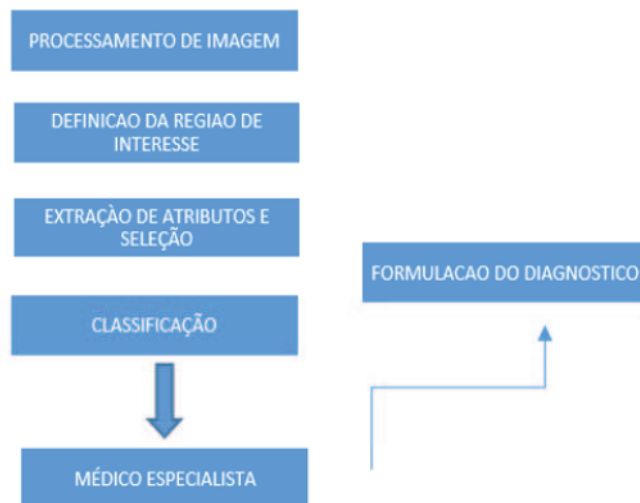
ID	Nome	Ano de publicação	Artigos	Citações	Taxa de citação por artigo	Palavra-chave
1	CAD para câncer de mama	2004	1628	10.085	6,2	mamografia peito massa mamografia câncer de mama
2	CAD para doenças pulmonares	2007	1345	7805	5,8	nódulo pulmão nódulo pulmonar pulmonar nódulo pulmonar
3	CAD com ultrassonografia	2010	652	2275	3,5	peito lesão tumor ultra-som próstata
4	CAD para lesão cutânea	2009	429	1062	2,5	dermoscopia melanoma retinal pele lesão na pele
5	CAD para câncer colorretal	2007	402	1498	3,7	pólipo colonografia colônico colonoscopia
6	CAD para doença cerebral	2011	142	400	2,8	Alzheimer espectro cérebro animal imagem do espectro
7	CAD para vários assuntos	2009	107	132	1,2	osteoporose fígado rim

Fonte: Takahashi e Kajikawa (2017)

Nas últimas duas décadas, progressos significantes foram alcançados no campo da tecnologia da informação. Diagnóstico por meio da análise computadorizada, foi originalmente desenvolvida para o câncer de mama usado na mamografia em 1960, e foi estendido para o diagnóstico de câncer de pulmão, câncer colorretal e assim por diante (Takahashi e Kajikawa, 2017).

Para desenvolver uma ferramenta CAD, segundo Theodoridis e Koutroumbas (2003), é preciso implementar sistemas de reconhecimento de padrões (RP), que seguem normalmente um conjunto de etapas: aquisição, segmentação, extração de características, classificação, pós-processamento e decisão, como sugerido na Figura 11.

Figura 11 - Diagnóstico auxiliado por computador



Fonte: MEYER-BAESE, A.,SCHMID (2014)

Porém, o reconhecimento de padrões no sinal EMG não é uma tarefa fácil, devido ao seu comportamento não linear. É necessário um bom classificador para distinguir cada padrão (Jali, M. H., Izzuddin, T. A., Bohari, Z. H., Jaafar, H. I. Nasir, 2015).

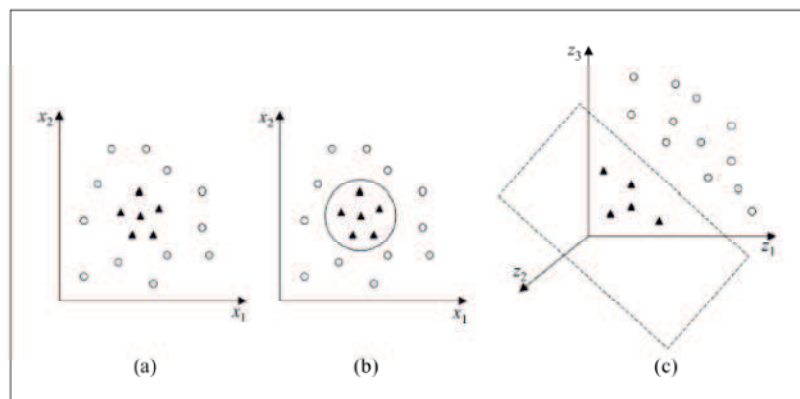
3 CLASSIFICADOR DE SINAIS

Neste capítulo, serão apresentados dois classificadores utilizados nesta pesquisa, Máquinas de Vetores de Suporte (*Support Vector Machines - SVM*) não lineares e K-Vizinhos mais próximos (*K-Nearest neighbors - KNN*). Ambos, serão apresentados a seguir.

3.1 MÁQUINAS DE VETORES DE SUPORTE – SVM

As SVMs lineares são eficazes na classificação de conjuntos de dados linearmente separáveis ou que possuam uma distribuição aproximadamente linear, sendo que a versão de margens suaves tolera a presença de alguns ruídos e outliers. Porém, há muitos casos em que não é possível dividir satisfatoriamente os dados de treinamento por um hiperplano. Um exemplo é apresentado na Figura 12, em que o uso de uma fronteira curva seria mais adequado na separação das classes.

Figura 12 - Conjunto de dados não linear



Fonte: THEODORIDIS; KOUTROUMBAS (2009)

O conjunto (a) representa o conjunto de dados não linear; (b) representa a fronteira não linear no espaço de entradas; (c) representa a fronteira linear no espaço de características.

As SVM lidam com problemas não lineares mapeando o conjunto de treinamento de seu espaço original, referenciado como entradas, para um novo espaço de maior dimensão, denominado espaço de características. Considere uma

amostra de treinamento $\{(x_f d_f)\}_{i=1}^N$, em que x_f é o padrão de entrada para o i -ésimo exemplo e d_f é a resposta desejada correspondente (saída-alvo). Para começar, assumimos que o padrão (classe) representado pelo subconjunto $d_f = +1$ e o padrão representado pelo subconjunto $d_f = -1$ são “linearmente separáveis”. A equação de uma superfície de decisão na forma de um hiperplano que realiza esta separação, conforme apresentado em Haykin (2009) é:

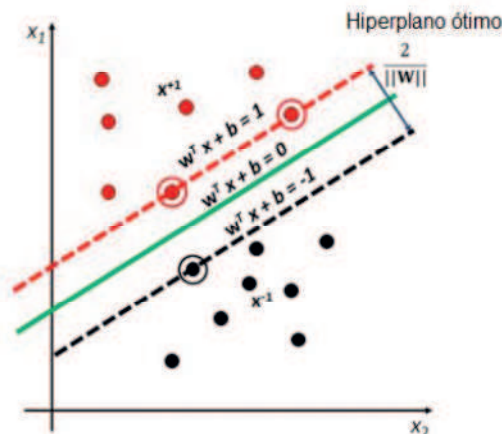
$$w^T x + b = 0 \quad (2)$$

Em que x é um vetor de entrada, w é um vetor peso ajustável e b é um bias. Foi descrita por Haykin (2009) conforme Equação 3.

$$\begin{aligned} w^T x_f + b &\geq 0 && \text{para } d_f = +1 \\ w^T x_f + b &< 0 && \text{para } d_f = -1 \end{aligned} \quad (3)$$

Para um dado vetor peso de w e bias b , a separação entre o hiperplano definido na Equação 2 e o ponto de dados mais próximo é denominada a *margem de separação*, representada por p . O objetivo de uma máquina de vetor de suporte é encontrar o hiperplano particular para o qual a margem de separação p é máxima. Sob esta condição, a superfície de decisão é referida como o *hiperplano ótimo*. A Figura 13 ilustra a construção geométrica de um hiperplano ótimo (representada pela Equação 2) para um espaço de entrada bidimensional.

Figura 13 - Hiperplano ótimo para padrões linearmente separáveis



Fonte: BONIFACIO (2017)

Considere que w_θ e b_θ representam os valores ótimos do vetor peso e do bias, respectivamente. Conseqüentemente, o *hiperplano ótimo*, representando uma superfície de decisão linear multidimensional no espaço de entrada é definido por

$$w_\theta^T x + b_\theta = 0 \quad (4)$$

A função discriminante

$$g(x) = w_\theta^T x + b_\theta \quad (5)$$

Fornece uma medida algébrica da distância de x até o hiperplano (Duda, Hart e Stork, 2000), também expresso como

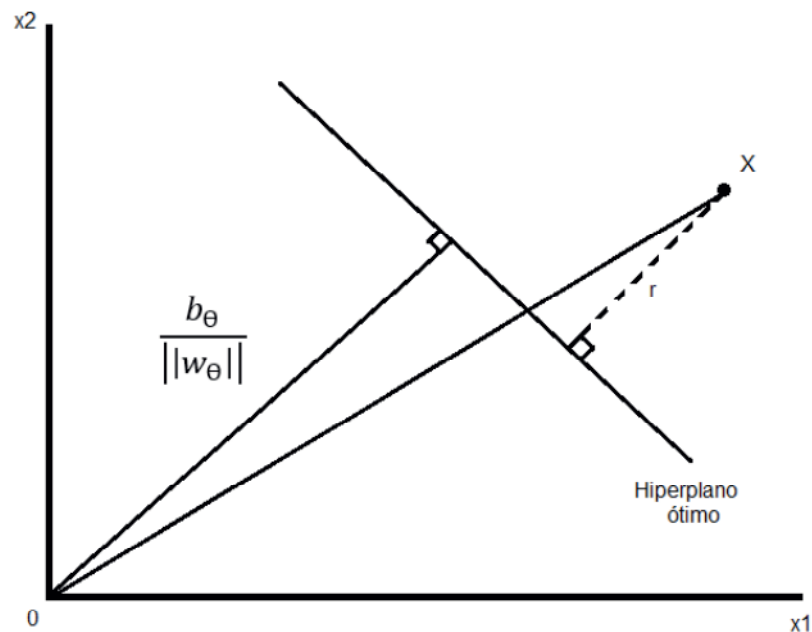
$$x = x_p + r \frac{w_\theta}{|w_\theta|} \quad (6)$$

Em que x_p é a projeção normal de x sobre o hiperplano ótimo, e r é a distância algébrica desejada; r é positivo se x estiver no lado positivo do hiperplano ótimo e negativo se x estiver no lado negativo. Como, por definição, $g(x_p) = 0$, resulta que

$$g(x) = w_\theta^T x + b_\theta = r |w_\theta| \quad (7)$$

A distância da origem até o hiperplano ótimo é dada por $b_\theta / |w_\theta|$. Se $b_\theta > 0$, a origem está no lado positivo do hiperplano ótimo; se $b_\theta < 0$, ela está no lado negativo. Se $b_\theta = 0$, o hiperplano ótimo passa pela origem. É possível visualizar a interpretação geométrica na Figura 14.

Figura 14 - Distâncias algébricas de pontos até o hiperplano ótimo para um caso bidimensional



Fonte: HAYKIN (2001)

Para encontrar os parâmetros w_θ e b_θ para o hiperplano ótimo, dado um conjunto de treinamento $T = \{(x_p, d_f)\}$, com base nos resultados da Figura 15, vemos que o par (w_θ, b_θ) deve satisfazer a restrição:

$$\begin{aligned} w_\theta^T x_f + b_\theta &\geq 1 && \text{para } d_f = +1 \\ w_\theta^T x_f + b_\theta &\leq -1 && \text{para } d_f = -1 \end{aligned} \quad (8)$$

Os pontos de dados particulares (x_p, d_f) para os quais a primeira ou segunda linha da Equação 6 é satisfeita com o sinal de igualdade são chamados de vetores de suporte, por isso o nome máquina de vetor de suporte. Estes vetores são importantes e desempenham um papel fundamental na operação desta classe de máquinas de aprendizagem. Em termos conceituais, os vetores de suporte são aqueles pontos de dados que se encontram mais próximos da superfície de decisão e são, portanto, os mais difíceis de classificar. Dessa forma, têm uma influência direta na localização ótima da superfície de decisão.

Um vetor de suporte $x^{(s)}$ para o qual $d^{(t)} = +1$. Então, por definição, temos

$$g x^{(s)} = w_{\theta}^T x^{(s)} \pm b_{\theta} \text{ para } d^{(s)} = \pm 1 \quad (9)$$

Da Equação 6, a distância algébrica do vetor de suporte $x^{(s)}$ até o hiperplano ótimo é

$$\begin{aligned} r &= \frac{g x^{(s)}}{w_{\theta}} \\ &= \begin{cases} \frac{1}{w_{\theta}} & \text{se } d^{(s)} = +1 \\ -\frac{1}{w_{\theta}} & \text{se } d^{(s)} = -1 \end{cases} \end{aligned} \quad (10)$$

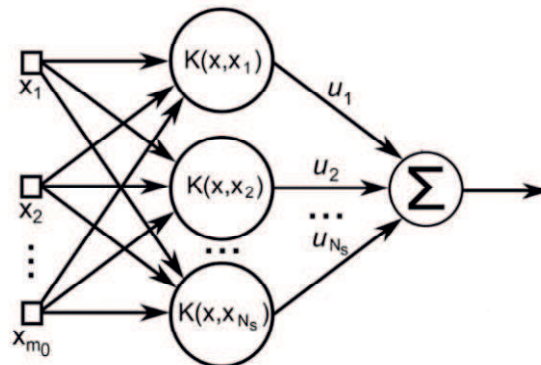
Onde o sinal positivo indica que $x^{(s)}$ se encontra no lado positivo do hiperplano ótimo e o sinal negativo indica de $x^{(s)}$ está no lado negativo do hiperplano ótimo. Considere que p represente o valor ótimo da margem de separação entre as duas classes que constituem o conjunto de treinamento T .

$$\begin{aligned} p &= 2r \\ &= \frac{2}{|w_{\theta}|} \end{aligned} \quad (11)$$

Em que a Equação 11 representa maximizar a margem de separação entre classes e que é equivalente a minimizar a norma euclidiana do vetor peso w . O hiperplano ótimo definido na Equação 4 é único no sentido de que o vetor peso w_{θ} fornece a máxima separação possível entre exemplos positivos e negativos. Esta condição ótima é alcançada minimizando-se a norma euclidiana do vetor peso w .

A Função de Base Radial Gaussiana (*Gaussian Radial Basis Function - RBF*) com função Kernel é uma função que recebe dois pontos x_i e x_j do espaço de entradas e computa o produto escalar desses dados no espaço de características. A Figura 15 apresenta a arquitetura SVM empregando a função kernel.

Figura 15 - A arquitetura SVM empregando a função kernel



Fonte: Theodoridis, Koutroumbas (2003)

As variáveis X_1 , X_2 e X_{m_0} correspondem a camada de entrada. As funções $K(x, x_1)$, $K(x, x_2)$ e $K(x, x_{N_s})$ representam a camada oculta de N_s vetores de suporte. Por fim, u_1 , u_2 e u_{N_s} correspondem aos pesos.

Alguns dos Kernels mais utilizados na prática são os Polinomiais, os Gaussianos ou RBF (*Radial-Basis Function*) e os Sigmoidais, listados na Tabela 3. Cada um deles apresenta parâmetros que devem ser determinados pelo usuário, indicados também na tabela. O Kernel Sigmoidal, em particular, satisfaz as condições de Mercer apenas para alguns valores de δ e K . Os Kernels Polinomiais com $d = 1$ também são denominados lineares.

Tabela 3 - Funções Kernel comuns

Tipo de Kernel	Função $K(K_i, K_j)$	Parâmetros
Polinomial	$(\delta(X_i, X_j) + k)^d$	δ, k e d
Gaussiano	$\exp(-\sigma \ X_i - X_j\ ^2)$	σ
Sigmoidal	$\tanh(\delta X_i, X_j + k)$	δ e k

Fonte: THEODORIDIS; KOUTROUMBAS (2009)

3.2 K-VIZINHOS MAIS PRÓXIMOS – KNN

KNN é um classificador onde o aprendizado é baseado na analogia. O conjunto de treinamento é formado por vetores n-dimensionais e cada elemento deste conjunto representa um ponto no espaço n-dimensional. Para determinar a

classe de um elemento que não pertença ao conjunto de treinamento, o classificador KNN procura K elementos do conjunto de treinamento que estejam mais próximos deste elemento desconhecido, ou seja, que tenham a menor distância.

Estes K elementos são chamados de K-vizinhos mais próximos. Verifica-se quais são as classes desses K vizinhos e a classe mais frequente será atribuída à classe do elemento desconhecido.

As métricas mais comuns no cálculo de distância entre dois pontos, são: distância Euclidiana, Manhattan e Minkowski, sendo que a mais utilizada, é a distância Euclidiana, conforme expresso na Equação 12.

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (12)$$

Seja $X = (x_1, x_2, \dots, x_n)$ e $Y = (y_1, y_2, \dots, y_n)$ dois pontos do \mathbb{R}^n . A distância Manhattan entre X e Y é dada, conforme Equação 13.

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n| \quad (13)$$

A distância Minkowski entre X e Y é dada por:

$$d(x, y) = \left(|x_1 - y_1|^q + |x_2 - y_2|^q + \dots + |x_n - y_n|^q \right)^{\frac{1}{q}} \quad (14)$$

Em que $q \in \mathbb{N}$. Esta distância é a generalização das duas distâncias anteriores. Quando $q = 1$, esta distância representa a distância de Manhattan e quando $q = 2$, a distância Euclidiana.

Se cada variável possuir um peso relativo à sua importância, a distância Euclidiana ponderada pode ser representada como:

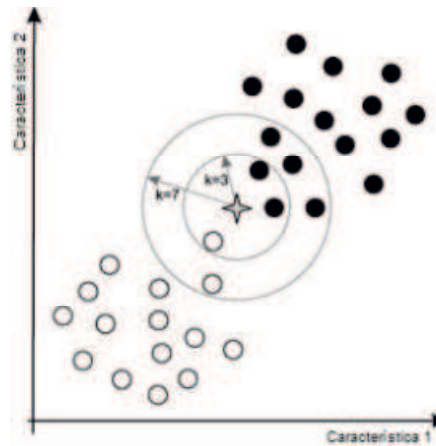
$$d(x, y) = \sqrt{w_1(x_1 - y_1)^2 + w_2(x_2 - y_2)^2 + \dots + w_n(x_n - y_n)^2} \quad (15)$$

Pesos também podem ser aplicados às distâncias Manhattan e Minkowski.

KNN é um classificador que possui apenas um parâmetro livre (o número de K-vizinhos) que é controlado pelo usuário com o objetivo de obter uma melhor classificação. A Figura 16 exemplifica o processo de classificação, em que a classe

predominante dos vizinhos mais próximos, o ponto desconhecido será classificado como um ponto pertencente a classe B (bolas pretas).

Figura 16 - Exemplo de classificação KNN



Fonte: BOSE; SAMANTA; CHATTERJEE (2016)

Este processo de classificação pode ser computacionalmente exaustivo, por este motivo, há uma variação mais rápida deste algoritmo. Esse outro processo seleciona pontos que estão dentro de uma hiper-esfera de raio R (decidido pelo usuário) e a classe predominante dentro desta hiper-esfera será a classe do ponto desconhecido. A desvantagem deste processo é que pode existir hiper-esfera sem qualquer ponto.

4 MÉTODO PROPOSTO E RESULTADOS PRELIMINARES

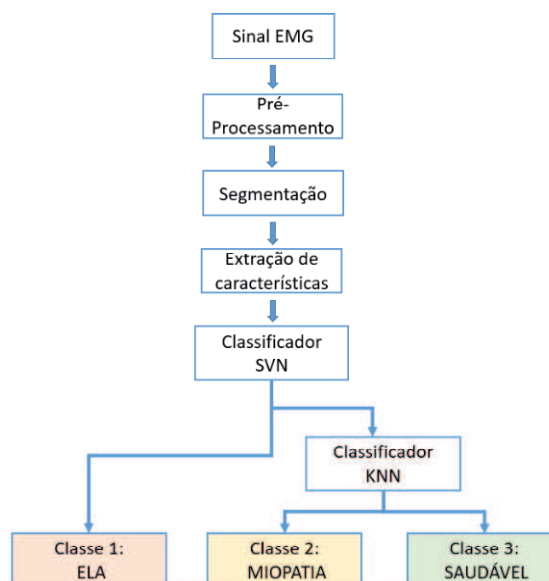
Do ponto de vista clínico, o diagnóstico das doenças neuromusculares nos estágios iniciais é difícil, à medida que os sintomas dessas doenças podem ser facilmente confundidos com outras doenças (Joshi *et al.*, 2017). Adicionalmente, estudos revelam que a dupla análise pode aumentar a sensibilidade do diagnóstico (Cram e Criswell, 2011). As soluções CAD vêm se mostrando, ao longo dos anos, uma opção viável e presente em vários segmentos da medicina.

Neste sentido, este capítulo descreve, portanto, a solução proposta para a problemática abordada na pesquisa, apresentando as etapas relativas ao desenvolvimento da ferramenta CAD, aqui chamado de Sistema Híbrido para Classificação de Sinais EMG (SHCSEMG), capaz de classificar sinais eletromiográficos em três classes: saudável, miopatia e ELA.

4.1 SISTEMA HIBRIDO PARA CLASSIFICAÇÃO DE SINAIS EMG - SHCSEMG

A literatura relacionada à classificação de sinais EMG tem apresentado os benefícios da combinação entre classificadores de sinais, pois ao aliar dois ou mais métodos em modo de compensação, os resultados podem ser potencializados.

Figura 17 - Sistema Híbrido para Classificação de Sinais - SHCSEMG



Fonte: O autor

Na Figura 17 é apresentado as principais etapas do processo de classificação de sinais EMG proposto. O sinal EMG são valores digitalizados e salvos em um dispositivo de armazenamento.

A etapa de pré-processamento consiste na remoção de sinais indesejados como oscilações na linha de base e outros ruídos provenientes de outros equipamentos elétricos ligados a rede elétrica.

A segmentação busca localizar e armazenar os PAUM representativos. A extração de características consiste em obter atributos discriminativos que caracterize cada classe de estudo.

Por fim, a etapa de classificação consiste de dois algoritmos de classificação, Máquina de Vetores de Suporte (*Support Vector Machine - SVM*) e K-ésimo Vizinho Mais Próximo (*K-Nearest Neighbors - KNN*).

Os sinais eletromiográficos digitalizados utilizados, foram obtidos na base de dados online aberta do Departamento de Neurofisiologia Clínica (*Department of Clinical Neurophysiology*) em Rigshospitalet (Hospital Nacional da Dinamarca) e consiste de grupos de pacientes saudáveis com miopatia e ELA.

O grupo de pacientes saudáveis consiste de 10 indivíduos com idades entre 21 e 37 anos, sendo 4 do sexo feminino e 6 do sexo masculino. Sessenta por cento (60%) do grupo de controle estavam em boa forma física, nenhum apresentava sinais ou histórico de distúrbios neuromusculares.

O grupo de pacientes com miopatia consiste de 7 indivíduos; 2 mulheres e 5 homens com idades entre 19 e 63 anos. Todos apresentavam sinais clínicos e eletrofisiológicos de miopatia.

O grupo de pacientes diagnosticado com ELA consiste de 8 indivíduos; 4 mulheres e 4 homens com idades entre 35 e 67 anos. Além dos sinais clínicos e eletrofisiológicos compatíveis com ELA, 5 deles morreram em poucos anos após o início do distúrbio.

4.2 PRÉ-PROCESSAMENTO DO SINAL ELETROMIOGRÁFICO

Na etapa de processamento, o sinal EMG é primeiramente filtrado para reduzir interferência de ruídos de alta e baixa frequência. Desta forma, a etapa de Processamento do Sinal Eletromiográfico, consistiu na implementação de filtros digitais, conforme descrito nas seções seguintes.

4.2.1 Limiar Suave (*Soft-Thresholding*)

A técnica Wavelet chamada Limiar Suave (*Soft-Thresholding*) foi utilizada com o objetivo de remover pequenas oscilações ao longo do sinal EMG. A técnica consiste em: os coeficientes wavelet, cujos valores absolutos forem maiores que o limiar, são substituídos por T , e os demais coeficientes serão iguais a zero (Valencia *et al.*, 2016).

A Equação 16, definida por Suresh (2017), expressa o modelo *Soft-Thresholding* em que, o tempo n é igualmente espaçado.

$$P_s \omega = \begin{cases} \text{sgn } \omega \cdot (\omega - T), & |\omega| \geq T \\ 0, & \omega < T \end{cases} \quad (16)$$

em que, ω é o coeficiente Wavelet, T é o valor do limiar, P_s é o *Soft-Thresholding*. Foi utilizada a técnica universal de limiarização para se obter o valor de T . Este trabalho escolheu um método relativamente simples de implementar e que mostrou bons resultados. A equação para estimar o Threshold é apresentada na Equação 17.

$$T = \sigma_n \sqrt{2 \log(N)}, \quad (17)$$

em que, N é o tamanho do sinal e σ_n é o desvio padrão do ruído da sub-banda que pode ser obtida a partir da Equação 18.

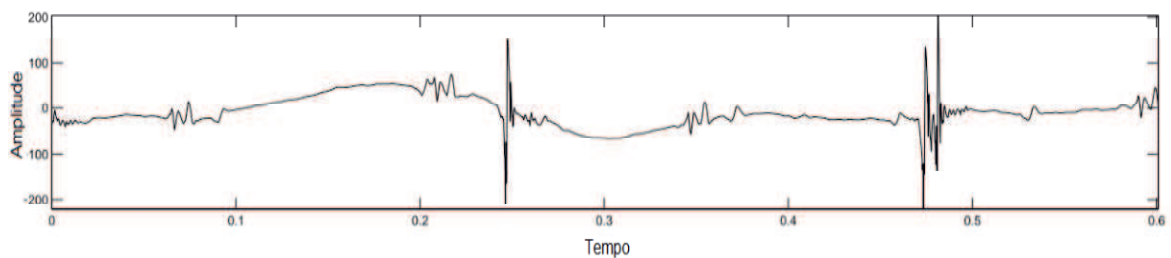
$$\sigma_n = \text{Mediana}(|Y_{x,y}|) / 0,6745, \quad (18)$$

O desafio em estimar T é determinar se um valor ω da transformada é significativo ou não perante os demais. Num primeiro pensamento, a média dos módulos seria um bom indicador: se $|\omega|$ estiver muito abaixo da média, ele seria considerado pequeno. Entretanto, o cálculo da média é prejudicado se um dos valores for muito grande comparado com o resto. Por isto, a mediana foi escolhida, a qual consegue calcular um valor intermediário sem ser tão afetada por valores extremos.

4.2.2 Filtro Butterworth

O filtro Butterworth é um filtro linear e em sua versão digital, quando aplicado adequadamente, não provoca distorção na fase do sinal filtrado (Khanna e Upadhyay, 2016). O filtro Butterworth foi utilizado para eliminar ruídos de baixa frequência, causado por uma componente contínua, que pode acarretar deslocamento na linha de base do sinal, conforme expresso na Figura 18.

Figura 18 - Sinal EMG com deslocamento de linha de base

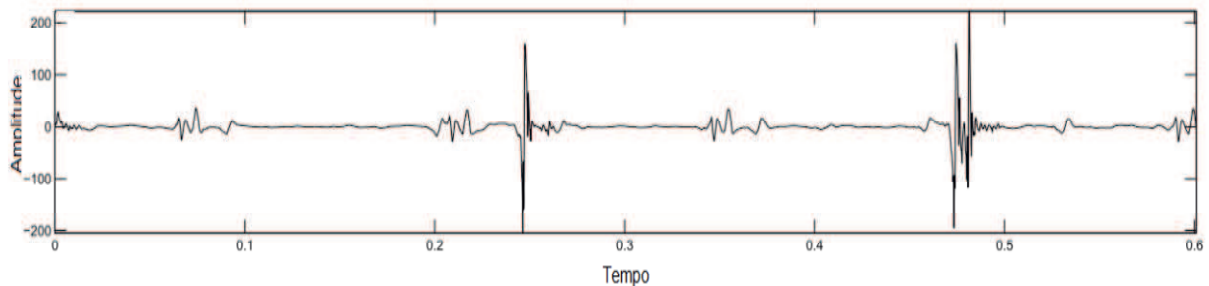


Fonte: NIKOLIC; KRARUP, 2011

Essa componente não tem relação com a atividade mioelétrica, sendo causada por fenômenos eletroquímicos que ocorrem entre a interface eletrodo/pele (Hosen *et al.*, 2015).

Para tanto, foi implementado o filtro Butterworth passa-alta com frequência de amostragem de 23437,5 Hz de primeira ordem e frequência de corte de 50Hz. Após a filtragem do sinal EMG, as oscilações devem ser eliminadas, conforme Figura 19.

Figura 19 - Sinal EMG após a filtragem



Fonte: NIKOLIC; KRARUP, 2011

4.3 SEGMENTAÇÃO

De acordo com Florestal, Mathieu e McGill (2017), a segmentação de sinais é uma técnica utilizada para a análise da estrutura, organização e função de uma unidade motora. O objetivo da segmentação do sinal é identificar as contribuições de cada unidade motora, durante a contração muscular, dentro de um conjunto de dados adquiridos. O processo de extração dos potenciais de ação equivale a visualização de suas formas de ondas e a detecção do tempo em que eles ocorrem. Neste sentido, a primeira tarefa a ser realizada na etapa de segmentação é localizar e estimar o início e o fim de todos os segmentos de intervalo de tempo contendo as atividades dos PAUM.

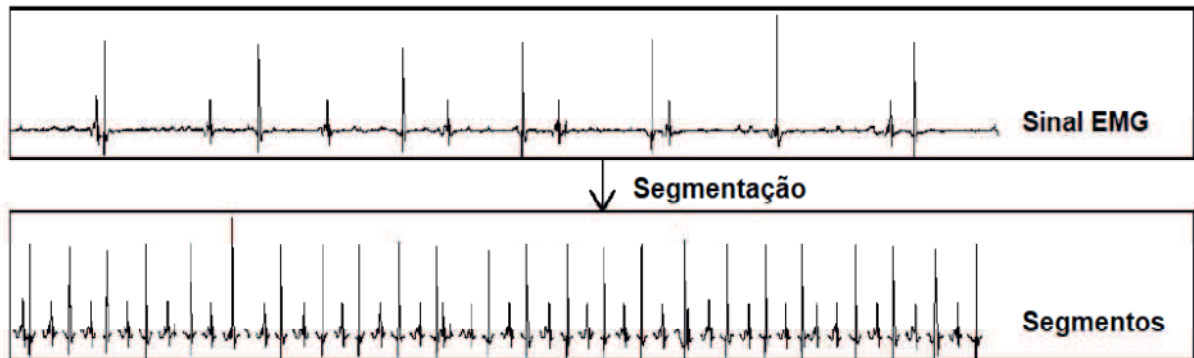
O método do único limiar foi usado por Goen (2014) para decompor os sinais EMG em PAUM candidatos. Usando o valor limite T (*threshold*), as áreas de baixa atividade foram eliminadas e os picos sobre o limite calculado T foram considerados como PAUM candidatos, conforme descrito na Equação 19. O método usado por Goen mostrou-se eficiente para o propósito desta pesquisa.

$$\begin{aligned}
 & \text{Se } \max_{i=1}^L x_i > \frac{30}{L} \quad x_i \\
 & \text{então } T = \frac{5}{L} \sum_{i=1}^L |x_i| \\
 & \text{caso contrário } T = \max(\{x_i/5\})
 \end{aligned} \tag{19}$$

Uma janela deslizante de comprimento de 200 pontos de amostragem foi utilizada para identificar os picos. A janela foi centrada no pico mais alto encontrado, então 200 pontos foram salvos como PAUM candidatas. O limiar T varia entre 30 e 100 μ V (Kim *et al.*, 2003). A variável x_i corresponde ao segmento de 200 pontos do sinal EMG.

Na Figura 20 é ilustrado o exemplo do processo de segmentação, exibindo o sinal EMG original e, em seguida, os segmentos correspondentes após a aplicação da segmentação.

Figura 20 – Sinal EMG original e segmentado



Fonte: Nikolic (2001)

4.4 EXTRAÇÃO DE CARACTERÍSTICAS

Inúmeros algoritmos de extração de características foram testados, conforme listado na Seção 2.3. Entretanto, com o objetivo de obter a combinação ideal entre o classificador de sinais EMG, que atenda aos requisitos do escopo da pesquisa e que se obtenha atributos discriminantes, foram selecionados os três modelos que proporcionaram o melhor resultado, sendo estes: RMS, PF e Número de Fase. A tabela contendo o resultado da extração de características, de todos os algoritmos estudados nesta pesquisa, pode ser conferida na íntegra no ANEXO A.

Para realizar as operações matemáticas necessárias, foram utilizadas funções implementadas na plataforma MATLAB, bem com seu toolbox, contendo recursos consolidados.

O Valor Quadrático Médio (*Root Mean Square - RMS*) de um vetor X , é calculado, conforme descrito na Equação 20.

$$X_{RMS} = \sqrt{\frac{1}{N} \sum_{n=1}^N |X_n|^2}, \quad (20)$$

Com o somatório realizado ao longo da dimensão especificada. Em que N corresponde a dimensão do vetor X .

A segunda característica extraída do sinal EMG é a Frequência de Pico (*Peak Frequency - PF*) de um vetor X , que é obtida a partir da frequência máxima

absoluta da Transformada Rápida de Fourier (*Fast Fourier Transform - FFT*), conforme descrito na Equação 21.

$$F_n = \max\left(\left| \sum_{i=0}^{N-1} x_i e^{\frac{2\pi j}{N} ni} \right|\right) \quad (21)$$

Finalmente, o número de fases de um sinal EMG é calculado a partir da Equação 22. A fase de um sinal EMG é uma seção do PAUM, localizada entre dois cruzamentos de linha de base, e que atinge um valor absoluto de amplitude maior que 0,02 milivolts.

$$\begin{aligned} z &= r e^{i\theta} \\ \log z &= \log r + i\theta \end{aligned} \quad (22)$$

A função de ângulo toma um número complexo $z = x + iy$ e calcula a tangente inversa de quatro quadrantes (*Four-quadrant inverse tangente*) a partir da função em MATLAB $\text{atan2}(y, x)$, para encontrar o ângulo formado no plano xy entre o eixo x positivo e um raio da origem ao ponto x, y . Este ângulo de fase é também a parte imaginária do logaritmo complexo, uma vez que:

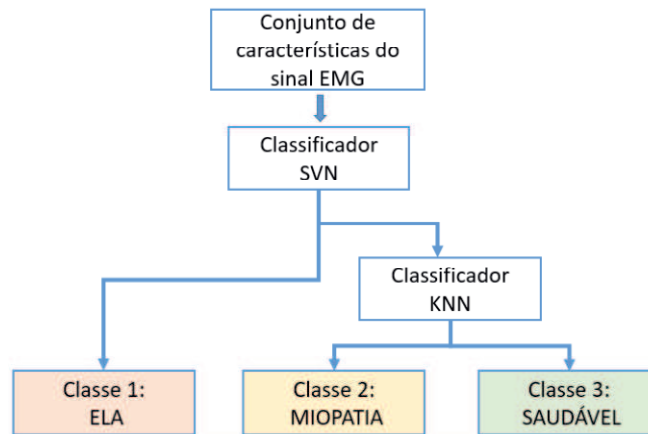
4.5 CLASSIFICADOR DE SINAIS EMG

Os sinais EMG têm propriedade não estacionária e possuem muitas características em cada classe de estudo. A variabilidade do sinal é, em parte, devido à existência de redundância no sistema neuromuscular, fazendo com que muitas combinações de atividades musculares produzam praticamente os mesmos potenciais de ação. É difícil para um único modelo de classificação supervisionado aprender as características dos dados não estacionários, como é o caso dos sinais EMG (Yousefi e Hamilton-Wright, 2014).

Neste sentido, foram utilizados dois algoritmos de classificação de sinais EMG neste trabalho, classificador baseado no vizinho mais próximo (*K-Nearest Neighbors - KNN*) e máquina de vetores de suporte (*Support Vector Machine - SVM*).

Em que, cada classificador pertence a uma etapa no fluxo de classificação do sinal EMG, conforme exemplificado no diagrama da Figura 21.

Figura 21 - Etapas de classificação do sinal EMG



Fonte: O autor

O conjunto de características do sinal EMG é constituído de um vetor de atributos, extraídos dos segmentos do sinal EMG em estudo, e que foram submetidos ao modelo do classificador SVM previamente treinado. Nesta etapa, os dados em análise serão classificados em duas classes: “ELA” ou “Outro”. A classe “Outro” corresponde aos sinais saudáveis e miopáticos.

Se os dados analisados pelo classificador SVM possuírem atributos da doença ELA, estes, serão classificados como Classe 1, caso contrário, a segunda etapa do sistema de classificação será executada. Na segunda etapa, o classificador KNN irá classificar os dados em miopatia (Classe 2) ou saudável (Classe 3).

4.5.1 Classificador SVM

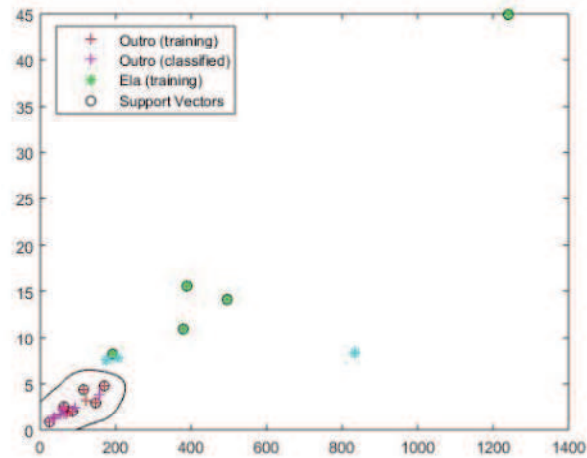
As Máquinas de Vetores de Suporte constituem uma técnica de aprendizado que vem recebendo atenção crescente da comunidade de Aprendizado de Máquina (AM). Os resultados da aplicação dessa técnica são comparáveis e muitas vezes superiores aos obtidos por outras técnicas de aprendizagem, como as Redes Neurais Artificiais (RNA) (Ara e May, 2016). Exemplos de aplicações de sucesso podem ser encontrados em diversos domínios, como na categorização de textos, na análise de imagens e em Bioinformática (Kim *et al.*, 2003).

Yousefi e Hamilton (2014) sugeriram um classificador SVM para a classificação de PAUM individuais a partir de sinais EMG intramusculares. Um experimento em 231 PAUM levou a uma precisão de 86,14%. Os resultados mostram que o SVM melhora a precisão da classificação, segundo Yousefi e Hamilton (2014). Duas vantagens desses trabalhos foram que os parâmetros ajustados e o cálculo orientado por dados foram minimizados. As vantagens de usar SVM incluem o baixo tempo de treinamento necessário e também seu desempenho satisfatório.

A maioria dos classificadores separa classes usando hiperplanos, mas o SVM amplia a ideia de separação de hiperplano para dados que não podem ser separados linearmente, mapeando os preditores em um novo espaço de maior dimensão no qual eles podem ser separados linearmente (Subasi, 2013). Porém, há muitos casos em que não é possível dividir satisfatoriamente os dados de treinamento por um hiperplano.

Neste sentido, com o objetivo de obter melhores resultados na classificação das classes em estudo, optou-se pela utilização de funções kernels, conforme listado no Capítulo 3. Verificou-se também, a partir dos testes realizados, que o tipo de Kernel Gaussiano variando entre 1 e 0.5, obtiveram resultados satisfatórios para o escopo desta pesquisa, conforme apresentado nas Figuras 22 e 23.

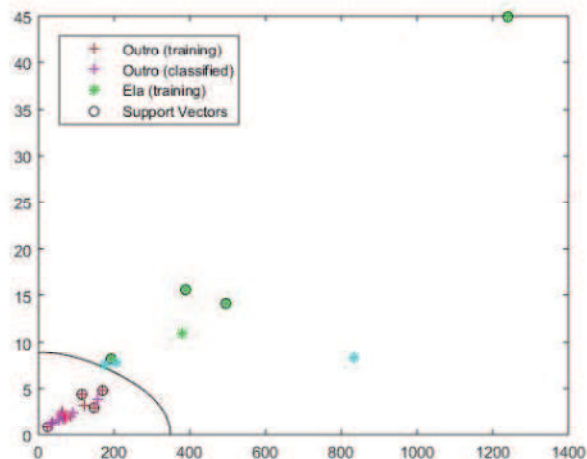
Figura 22 - Classificador SVM - ELA x Outro com Kernel de 0,1



Fonte: O autor

Para Kernel igual a 0,1, percebe-se que todos os dados de teste (três pontos em azul), foram classificados corretamente, ou seja, Classe 1 (ELA). Os dados de treino foram previamente selecionados, como ELA e “Outro” (saudável e miopia).

Figura 23 - Classificador SVM - ELA x Outro com Kernel de 0,5



Fonte: O autor

Por outro lado, atribuindo-se o valor de 0,5 para o kernel, percebe-se que o classificador SVM atua em seu limiar, conseqüentemente, podendo conduzir a uma diminuição na acurácia do classificador. Neste sentido, verificou-se, heurísticamente, que o melhor valor de K para o desempenho do classificador SVM utilizado é de $K = 0,1$ devido ao melhor resultado medido.

4.5.2 Classificador KNN

O algoritmo de classificação baseado no vizinho mais próximo (KNN) é uma técnica amplamente empregada para reconhecer padrões. O princípio do seu funcionamento está em descobrir o vizinho mais próximo para uma dada instância.

O algoritmo KNN pertence a um grupo de técnicas denominada de *Instance-based Learning*. Neste caso, são encontrados os k vizinhos mais próximos do padrão de consulta, ao invés de apenas o vizinho mais próximo (Bose, Samanta e Chatterjee, 2016).

O KNN classifica um dado elemento de acordo com as respectivas classes dos k ($k \geq 1$) vizinhos mais próximos pertencentes a uma base de treinamento. O algoritmo calcula a distância do elemento dado para cada elemento da base de treinamento e então ordena os elementos da base de treinamento do mais próximo ao de maior distância. Dos elementos ordenados, selecionam-se apenas os k primeiros, que servem de parâmetro para a regra de classificação.

Dois pontos importantes no KNN são: a regra de classificação e função que calcula a distância entre duas instâncias. A regra de classificação diz como o algoritmo vai tratar a importância de cada um dos k elementos selecionados mais próximos. À função de distância cabe a tarefa de mensurar a disparidade entre dois elementos de forma a poder identificar quais são os KNN.

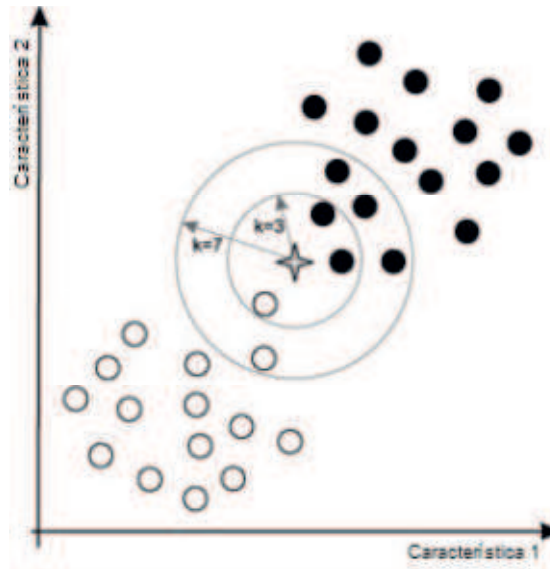
Para $k = 1$ a regra de classificação é irrelevante, pois só existe um elemento referente a uma classe. Porém, com $k > 1$, é preciso haver uma regra para discernir a qual classe se atribuirá o elemento. A literatura apresenta as duas regras de classificação clássicas: maioria na votação e peso pela distância. Na primeira, cada elemento tem uma influência igual, a classe escolhida é aquela que possui mais representantes entre o KNN, este, utilizado na pesquisa. No peso pela distância, cada k -vizinho tem um peso inversamente proporcional à sua distância.

Na literatura são as funções de distância que possuem o maior viés de estudo no KNN, pois possibilitam as maiores variações nos resultados obtidos por esse algoritmo, podendo inclusive proporcionar resultados em nível de estado da arte nessa área.

A Figura 24 ilustra a classificação hipotética de um elemento desconhecido que foi submetido ao classificador KNN com K igual a 3 e 7. Percebe-se que, para $k = 3$ o elemento seria classificado com classe B (bolas pretas), pois existem dois

elementos similares (de cor preta) próximos a ele e apenas um elemento da classe A (bolas brancas). Para $k = 7$, o elemento também seria classificado como classe B, entretanto o número de elementos similares é maior (sete elementos), ao contrário da classe A de apenas dois elementos.

Figura 24 - Ilustração hipotética da classificação de um elemento desconhecido



Fonte: BOSE; SAMANTA; CHATTERJEE (2016)

Para encontrar o melhor número de k vizinhos para comparação (valor de k) do classificador KNN, a base de dados foi dividida aleatoriamente em dois grupos, um para treinamento e outro para teste do sistema, designados por grupo A (saudável) e grupo B (miopático). O grupo A foi tomado para treino e o grupo B, para teste. Verificou-se que o valor ideal de k para o desempenho satisfatório, foi de $k = 3$.

4.6 MÉTRICAS DE AVALIAÇÃO

Uma vez que o sinal EMG foi classificado, é preciso medir o seu desempenho. Neste sentido, tem-se as seguintes definições para validação:

- Verdadeiro Positivo (VP): Instâncias corretamente classificadas como a classe de interesse.

- Verdadeiro Negativo (VN): Instâncias corretamente classificadas como não sendo a classe de interesse.
- Falso Positivo (FP): Instâncias incorretamente classificadas como a classe de interesse.
- Falso Negativo (FN): Instâncias incorretamente classificadas como não sendo a classe de interesse.

Assumindo estas definições, com o objetivo de comparar o desempenho dos classificadores atuando de forma independente e em conjunto, a partir de um sistema híbrido, foram utilizados três métricas distintas: Precisão, Acurácia e Sensibilidade.

A precisão é definida na Equação 22.

$$\text{Precisão \%} = \frac{VP}{VP + FP} \times 100, \quad (22)$$

Em que, VP representa a taxa de verdadeiros positivos e FP a taxa de falsos positivos, fornecendo a proporção de todos os atributos identificados a uma determinada classe de fato pertencerem àquela classe. Esta análise se mostra relevante ao problema porque a falsa identificação de atributos poderia levar a um falso alerta de patologia no sinal EMG.

A sensibilidade é definida na Equação 23.

$$\text{Sensibilidade \%} = \frac{VP}{VP + FN} \times 100, \quad (23)$$

Em que, FN representa a taxa de falsos negativos, fornecendo a capacidade do algoritmo identificar cada classe nos segmentos. Baixa sensibilidade na detecção das classes correspondentes aos atributos poderia levar a um retardamento na identificação da doença.

Por fim, a acurácia é definida na Equação 24.

$$\text{Acurácia \%} = \frac{VP + VN}{VP + VN + FP + FN} \times 100, \quad (24)$$

5 RESULTADOS E DISCUSSÕES

Como produto da pesquisa, obteve-se o protótipo de uma solução de auxílio ao diagnóstico médico, em eletromiografia, capaz de classificar sinais eletromiográficos em saudável, miopático ou ELA, a partir de uma interface gráfica de fácil utilização.

5.1 DESENVOLVIMENTO DO PROTÓTIPO

Para o desenvolvimento do protótipo, optou-se pela utilização do MATLAB da MathWorks, Inc., que Gonzalez et al (2004) descrevem como uma linguagem de alta performance para cálculos técnicos e integra recursos de: computação, visualização e programação em um ambiente fácil, no qual problemas e soluções são expressos com notações matemáticas familiares.

Gonzalez et al (2004) destacam ainda, que uma importante característica relativa à concepção de sistemas de processamento de sinais, é o nível significativo de testes e experimentos que normalmente é requerido antes de se chegar a uma solução aceitável. Desta maneira, o MATLAB juntamente com o seu *Toobox* são fortes aliados na resolução de problemas de processamento de sinais digitais.

É possível realizar teste de classificação para um determinado sinal, obtidos a partir de um diretório no disco rígido ou mídia externa. Também é possível realizar testes no protótipo com dados previamente armazenados durante o carregamento em memória ou executar a classificação de todos os dados de teste, nesse último, é possível obter o desempenho dos classificadores implementados na ferramenta a partir da análise da Matriz de confusão.

No campo de aprendizagem de máquina e especificamente o problema da classificação da classificação estatística, uma matriz de confusão, também conhecida como matriz de erro, é um layout de tabela específico que permite a visualização do desempenho de um algoritmo, tipicamente um aprendizado supervisionado.

A Figura 25 exibe a Matriz de Confusão para o classificador SVM com duas classes (Ela e Outro), para 10 amostras, e o classificador KNN, também para duas classes (Miopatia e Normal).

Figura 25 - Avaliação de desempenho dos classificadores a partir da Matriz de Confusão



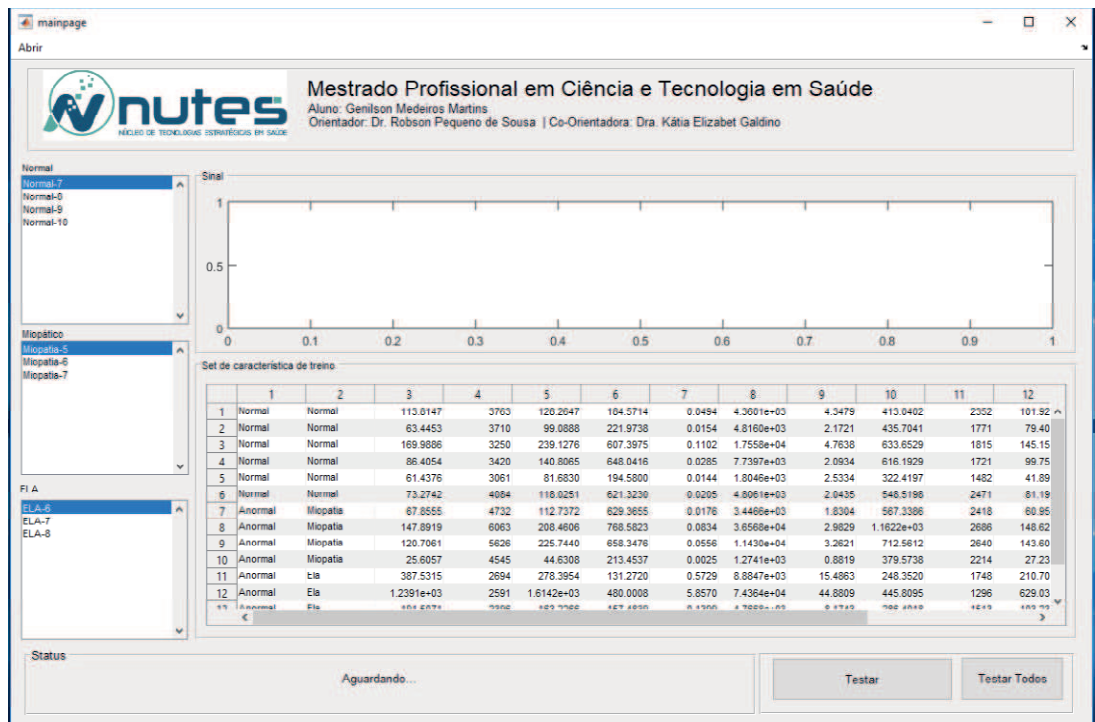
Fonte: O autor

Observa-se que para o classificador SVM, 3 amostras (30,0%), foram classificadas como Verdadeiro Positivo, ou seja, com ELA e 7 amostras (70,0%), como Verdadeiro Negativo, como classe "Outro". As seta amostras (70%), consideradas Verdadeiro Negativo, foram submetidas no classificador KNN. Percebe-se que 3 amostras (42,9%) foram consideradas como Verdadeiro Positivo, Miopatia e 57,1% como Verdadeiro Negativo, ou seja, Normal.

5.1.1 Visão Geral

O protótipo foi concebido visando uma interface prática e intuitiva, permitindo a visualização imediata de seus resultados. Na Figura 26 é apresentada a interface desenvolvida que consiste em um software de tela única, desenvolvido em linguagem MATLAB, no qual fez-se uso de suas funções e *toolbox* para construção do classificador de sinais.

Figura 26 - Interface principal do protótipo

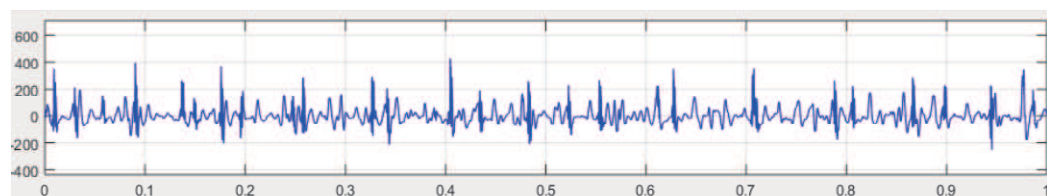


Fonte: O autor

O protótipo possui *feedback* visual (visualização gráfica) e textual (exibição do estado corrente software). Para fins acadêmicos, é possível visualizar o conjunto de característica extraídas anteriormente de alguns dados de testes e que são utilizados para treinamento dos classificadores utilizados nesta pesquisa. Também é possível realizar testes com algumas amostras de sinais pré-processados bem como visualizar o desempenho geral dos algoritmos de classificação.

Na Figura 27 é exibido uma amostra pré-processada do sinal EMG renderizado na tela do protótipo após sua seleção.

Figura 27 – Amostra de sinal EMG renderizado na tela do protótipo

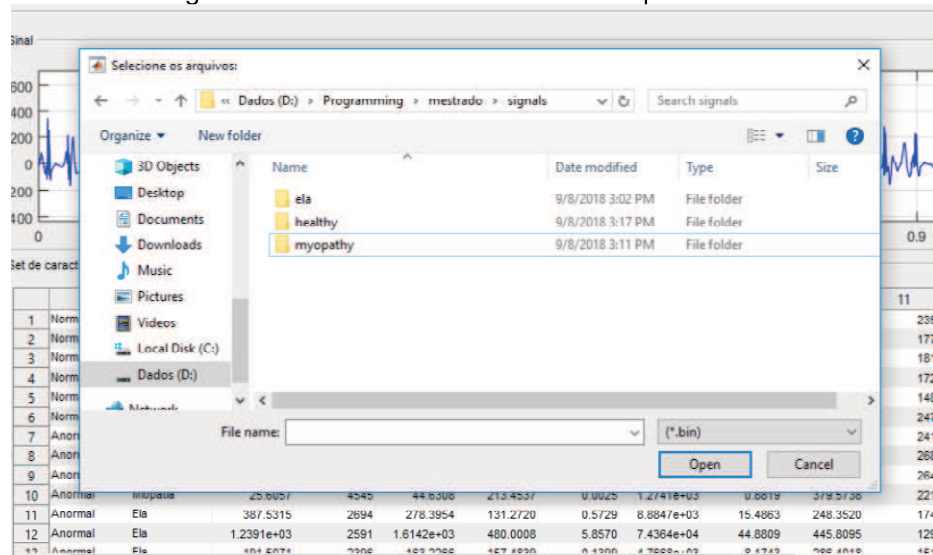


Fonte: O autor

5.2.2 Funcionamento da ferramenta

A interface do protótipo foi desenvolvida com o intuito de minimizar as ações do operador, bem como proporcionar uma interface gráfica o mais simples possível, e que proporcione uma boa experiência do usuário. Pensando nisto, com poucos cliques é possível executar a ferramenta e obter auxílio ao diagnóstico a partir da classificação da classe em que os dados em análise pertence. Para tanto, é necessário que o operador clique no menu “Abrir” para a visualização da tela padrão de seleção de arquivo do Windows, conforme apresentado na Figura 28.

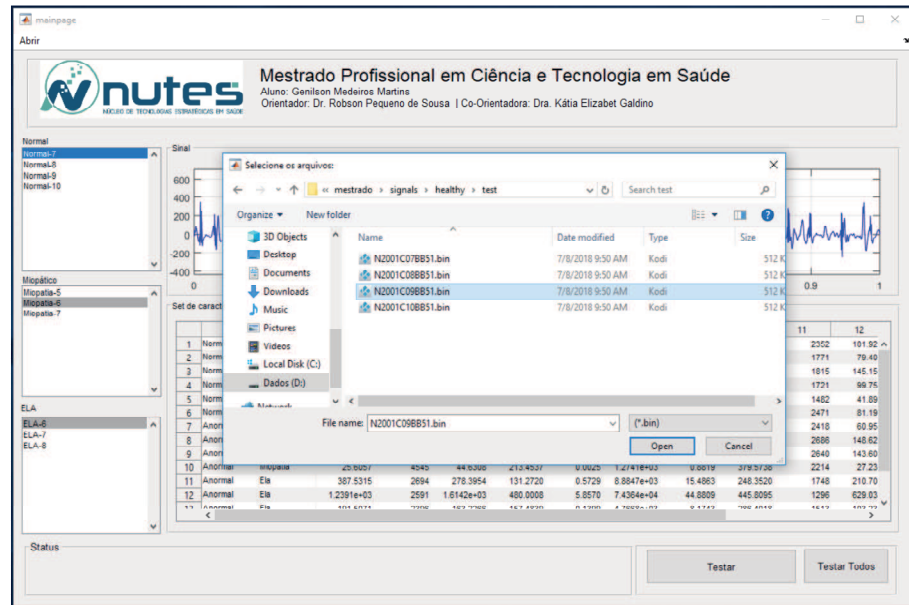
Figura 28 - Selecionando um sinal EMG para análise



Fonte: O autor

O operador deverá localizar o diretório do arquivo contendo os dados EMG salvos posteriormente. Após sua localização, basta selecioná-lo e abri-lo, conforme ilustrado na Figura 29.

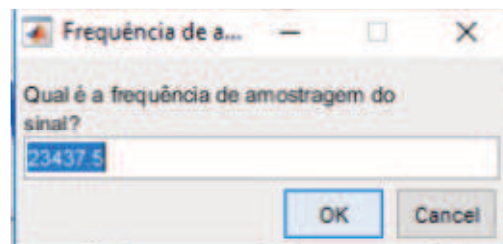
Figura 29 - Abrindo um arquivo EMG para análise



Fonte: O autor

Após a seleção do arquivo contendo os dados EMG, uma “caixa de diálogo” será exibida solicitando a informação referente a frequência de amostragem do sinal em análise (Figura 30). Por padrão, o campo de entrada de dados já é carregado com o valor de 23437,5 Hz referente aos dados de teste utilizados nesta pesquisa, sendo possível alterá-los para o valor desejado.

Figura 30 - Frequência de amostragem

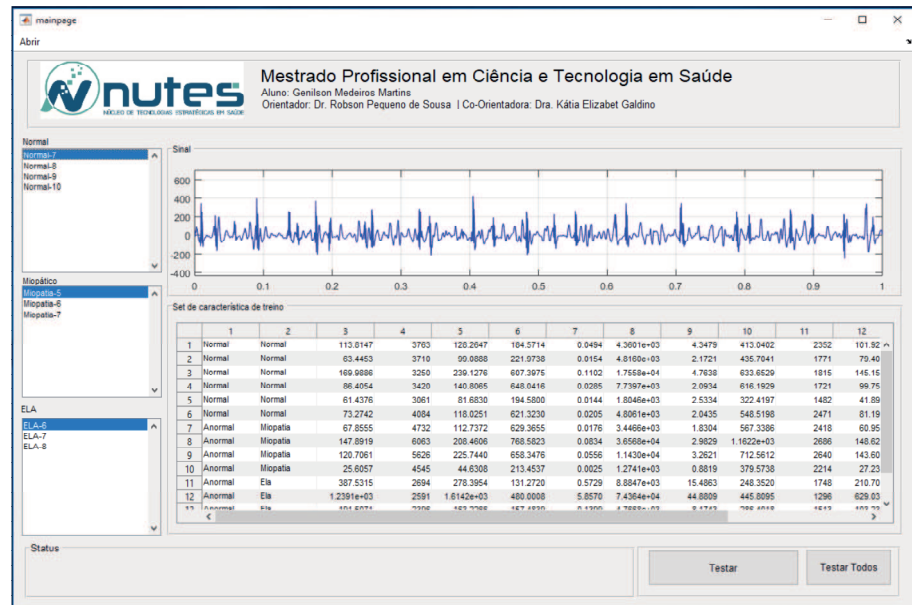


Fonte: O autor

O gráfico, referente ao arquivo de dados, será exibido na tela do aplicativo, conforme expresso na Figura 31. Nota-se, que o gráfico exibido foi submetido ao pré-processamento, em que, interferências e ruídos foram removidos. Após a etapa de filtragem, que foi executada logo após o carregamento dos dados do sinal em

memória, os algoritmos para extração de características são executados resultando no conjunto de atributos que caracterizam tal sinal.

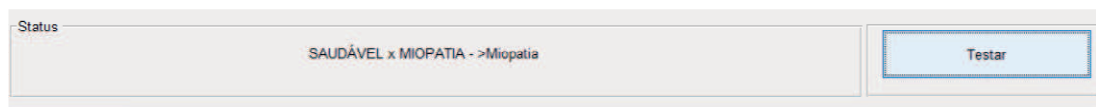
Figura 31 - Exibição do gráfico EMG após seleção do arquivo de dados



Fonte: O autor

A ação final que o operador precisa executar é clicar no botão “Testar” localizado no canto inferior direito da tela do aplicativo. Na Figura 32 é exibido o resultado obtido após a classificação do sinal no campo “Status” da ferramenta.

Figura 32 - Exibição da classificação do Sinal EMG



Fonte: O autor

5.2 CÁLCULO DE DESEMPENHO

O cálculo de precisão, sensibilidade e acurácia foram utilizados em dois momentos na pesquisa. No primeiro momento, para avaliar o desempenho dos classificadores SVM e KNN, ambos, operando isoladamente e no segundo momento foi verificado o desempenho do Sistema de Híbrido de Classificação de Sinais Eletromiográficos (SHCSEMG), que utiliza a abordagem híbrida.

A adoção desta metodologia é importante para a pesquisa, pois, desta forma, foi possível comparar o desempenho do classificador híbrido utilizado com a implementação dos classificadores de forma individual.

Na Tabela 4 é apresentado o desempenho do classificador SVM para classificação dos sinais EMG em três classes: saudável, miopatia e ELA. Os atributos utilizados no classificador foram RMS e PF.

Para classificação dos sinais EMG com ELA, o modelo SVM proporcionou o valor de 100% nas três métricas avaliadas, por outro lado, miopatia e saudável, proporcionaram baixa avaliação, correspondendo a 0% e 25%, respectivamente, nas três métricas de desempenho adotadas.

Tabela 4 - Classificador SVM para classificação de sinais com ELA, miopatia e saudável

	Precisão	Sensibilidade	Acurácia
Ela	100%	100%	100%
Miopatia	0%	0%	0%
Saudável	25%	25%	25%

Fonte: O autor

Para o classificador KNN com $k = 3$, foram utilizados os atributos FP e Número de fases. Na Tabela 5 é apresentado o desempenho do classificador para classificação dos sinais EMG para as três classes: saudável, miopatia e ELA.

Tabela 5 - Classificador KNN para classificação de sinais com ELA, miopatia e saudável

	Precisão	Sensibilidade	Acurácia
Ela	100%	66,6%	66,6%
Miopatia	100%	100%	100%
Saudável	100%	75%	75%

Fonte: O autor

Observa-se, que o classificador KNN testado, proporcionou um bom resultado para classificação de sinais EMG em miopatia, obtendo 100% nas avaliações das métricas de precisão, sensibilidade e acurácia, por outro lado, para sinais com ELA e saudável o desempenho foi inferior ao desejado.

Na segunda abordagem, o cálculo de desempenho do SHCSEMG foi medido utilizando as métricas, anteriormente descritas, em conjunto com a Matriz de Confusão para uma melhor visualização dos resultados.

Assim, uma saída positiva na primeira etapa de classificação, ou seja, no classificador SVM, significa que o sinal foi classificado como ELA. Na segunda etapa (classificador KNN), caso necessário, uma saída positiva significa que o sinal foi classificado como miopatia ou saudável.

Tabela 6 - Matriz de Confusão (ELA x Outro)

		Classe Prevista	
		ELA	Outro
Classe Real	ELA	3 Verdadeiro positivo	0 Falso Positivo
	Outro	0 Falso Negativo	7 Verdadeiro Negativo

Fonte: O autor

A Tabela 6 representa a Matriz de Confusão para a primeira etapa de classificação do sistema híbrido. Esta etapa corresponde ao classificador SVM, em que cada coluna corresponde as instâncias em uma classe prevista, enquanto cada linha representa as instâncias em uma classe real. Ao submeter 10 sinais EMG para análise do classificador, em que 3 foram diagnosticados previamente com ELA, percebe-se que o modelo obteve resultado satisfatório, ou seja, 3 sinais foram considerados como VP (classe de interesse) e 7 VN (não sendo a classe de interesse).

Os sete sinais considerados como VN, foram previamente diagnosticados, por especialistas em neurologia, e concluiu-se que três deles são miopáticos e quatro saudáveis. Ao submeter as sete amostras a segunda etapa de classificação do sistema híbrido, o classificador KNN (com $k = 3$), é possível perceber, na Tabela 7,

que o modelo obteve resultado satisfatório, pois 3 sinais foram classificados como miopático, ou seja, como VP (classe de interesse) e 4 como saudável VN.

Tabela 7 - Matriz de Confusão (Miopia x Saudável)

		Classe Prevista	
		Miopia	Normal
Classe Real	Miopia	3 Verdadeiro positivo	0 Falso Positivo
	Normal	0 Falso Negativo	4 Verdadeiro Negativo

Fonte: O autor

A Tabela 8 apresenta os valores obtidos no cálculo das métricas listadas nas Equações 22, 23 e 24 para o SHCSEMG.

Tabela 8 - Desempenho do classificador SHCSEMG

	Precisão	Sensibilidade	Acurácia
SVM (Ela x Outro)	100%	100%	100%
KNN (Miopático x Saudável)	100%	100%	100%

Fonte: O autor

Neste sentido, conforme apresentado na Tabela 8, foi observado que os classificadores SVM e KNN, utilizados nesta pesquisa, constituindo um sistema híbrido, foi possível potencializar seus benefícios e consequentemente obter melhor resultado.

Por fim, a Tabela 9 apresenta um comparativo entre o desempenho dos classificadores utilizados pelos principais autores estudados nesta pesquisa e o SHCSEMG.

Tabela 9 - Comparativo de desempenho entre os classificadores de sinais EMG utilizados pelos principais autores estudados e SHCSEMG

Referência	Classificador	Acurácia
(Goen, 2014)	SVM	88,5
(Bose, Samanta e Chatterjee, 2016)	SVM	95%
(Küçük e Emino, 2015)	KNN	96,1%
(Bose, Samanta e Chatterjee, 2016)	KNN	99,75%
SHCSEMG	Híbrido (SVM e KNN)	100%

Fonte: O autor

6 CONSIDERAÇÕES FINAIS

Diagnóstico a partir da análise computadorizada vem se mostrando bastante eficiente em vários segmentos da medicina. O Sistema de auxílio ao diagnóstico médico é uma ferramenta de suporte importante para reduzir o tempo necessário para interpretação de imagens e sinais biológicos, e principalmente em prover uma segunda opinião, tornando a decisão final do especialista mais rápida e mais confiável.

A necessidade de auxílio ao diagnóstico em eletromiografia se dá devido ao grande número de doenças neuromusculares, variações existentes entre as doenças, tempo gasto em testes e exames, bem como a similaridade das características visuais do sinal que, nos estágios iniciais, são acentuados e acarretando dificuldade da análise e possíveis erros.

Na investigação de métodos para classificação de sinais, constatou-se que cada método isolado tem uma deficiência específica. Ao aliar dois ou mais métodos em modo de compensação, os resultados podem ser potencializados.

Neste sentido, o protótipo de uma ferramenta para Auxílio ao Diagnóstico Médico constituído por um sistema híbrido de classificação de sinais desenvolvido, buscou iniciar o estudo no campo da eletromiografia, visando à diminuição da falta de ferramentas para os profissionais da área.

Concluiu-se que a implementação de um sistema híbrido constituído pelos classificadores SVM e KNN, para o escopo delimitado, é válido devido ao desempenho satisfatório.

Vale salientar, que não é possível substituir os processos atuais executados pelos especialistas, pois o diagnóstico de doenças neurológicas se dá após várias análises, exames e entrevistas com os pacientes. Entretanto, a ferramenta CAD foi considerada válida devido a sua característica de suporte como “segunda opinião”, o que permite proporcionar ao especialista maior segurança durante o processo de análise e diagnóstico.

7 PROPOSTAS FUTURAS

Com a conclusão da pesquisa, foi realizada uma análise e concluiu-se que há propostas para dar continuidade, com o objetivo de aperfeiçoar o que foi iniciado, como descrito a seguir.

- Implementação da norma internacional ISO/IEC 62304 que trata da análise de requisitos de software no processo de desenvolvimento para dispositivos médicos;
- Obtenção de mais amostras de sinais eletromiográficos para aumentar a acurácia no treinamento dos modelos matemáticos responsáveis pela classificação do sinal EMG;
- Realização de novos estudos experimentais é necessário em conjunto com a indústria, referente a outros sistemas existentes, para que sejam possíveis observações empíricas, que conduzam a resultados experimentais mais concretos e
- Acréscimo de métricas para mensurar a solução desenvolvida em relação a usabilidade e satisfação dos usuários.

REFERÊNCIAS

- ABEL, E.; ZACHARIA, P.; FORSTER, A.; FARROW, T. Neural network analysis of the EMG interference pattern. **Medical engineering & ...**, n. 1, p. 12–17, 1996.
- ARA, B.; MAY, O. Nöromüsküler Hastalıkların Teşhisinde SVM, K- EYK ve DA Algoritma Tabanlı Sınıflandırma Neuromuscular Disease Diagnosis of SVM, K- NN and DA Algorithm Based Classification Part-II. p. 8–11, 2016.
- ASTAKHOV, V. **Biomedical Informatics**. 2010. v. 569
- BHOI, A. K.; SHERPA, K. S.; MALLICK, P. K. A comparative analysis of neuropathic and healthy EMG signal using PSD. **International Conference on Communication and Signal Processing, ICCSP 2014 - Proceedings**, p. 1375–1379, 2014.
- BONIFÁCIO, A. **Estratégia computacional para avaliação de propriedades mecânicas de concreto de agregado leve**. 2017.
- BOSE, R.; SAMANTA, K.; CHATTERJEE, S. Cross-correlation based feature extraction from EMG signals for classification of neuro-muscular diseases. **2016 International Conference on Intelligent Control Power and Instrumentation (ICICPI)**, p. 241–245, 2016.
- C. GONZALEZ, R.; E. WOODS, R.; L. EDDINS, S. **Digital Image Processing Using Matlab**. 2005. v. 1
- CHAUVET, E.; FOKAPU, O.; HOGREL, J.-Y.; GAMET, D.; DUCHÊNE, J. A method of EMG decomposition based on fuzzy logic. **IEEE Engineering in Medicine and Biology (EMBS)**, v. 2, p. 1948–1950, 2001.
- CHAWLA, J. Stepwise approach to myopathy in systemic disease. **Frontiers in Neurology**, v. AUG, n. August, p. 1–10, 2011.
- CHRISTODOULOU, C. I.; PATTICHIS, C. S. Unsupervised pattern recognition for the classification of EMG signals. **IEEE transactions on bio-medical engineering**, v. 46, n. 2, p. 169–78, 1999.
- CHRISTODOULOU, C.; PATTICHIS, C. S. Combining neural classifiers in EMG diagnosis. **6th European Congress on Intelligent Techniques and Soft Computing. EUFIT '98**, v. 3, n. September 1998, p. 1837–1841, 1998.
- CRAM, J. R.; CRISWELL, E. **Cram's introduction to surface electromyography**. 2006.
- CRISWELL, E. **Cram's Introduction to Surface Electromyography**. 2009.
- DEMPSEY, P. J. **Can computer-aided detection with double reading of**

screening mammograms help decrease the false-negative rate? Initial experience**Breast Diseases**, 2005.

DOBROWOLSKI, A. P.; WIERZBOWSKI, M.; TOMCZYKIEWICZ, K. Multiresolution MUAPs decomposition and SVM-based analysis in the classification of neuromuscular disorders. **Computer Methods and Programs in Biomedicine**, v. 107, n. 3, p. 393–403, 2012.

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**New York: **John Wiley, Section**, 2000.

ELBAGOURY, B. M.; VLADAREANU, L. A hybrid real-time EMG intelligent rehabilitation robot motions control based on Kalman Filter, support vector machines and particle swarm optimization. **SKIMA 2016 - 2016 10th International Conference on Software, Knowledge, Information Management and Applications**, p. 439–444, 2017.

FDWLRQ, O.; LVRUGHUV, I. R. U. Hybrid Water Wave Optimization and Support Vector Machine to Improve EMG Signal.pdf. [s.d.].

FLORESTAL, J. R.; MATHIEU, P. A.; MCGILL, K. C. Automatic decomposition of multichannel intramuscular EMG signals. **Journal of Electromyography and Kinesiology**, v. 19, n. 1, p. 1–9, 23 abr. 2017.

FUGLSANG-FREDERIKSEN, A. The utility of interference pattern analysis. **Muscle & nerve**, v. 23, n. 1, p. 18–36, 2000.

GAZZONI, M.; FARINA, D.; MERLETTI, R. A new method for the extraction and classification of single motor unit action potentials from surface EMG signals. **Journal of Neuroscience Methods**, v. 136, n. 2, p. 165–177, 2004.

GOEN, A. Classification of EMG Signals for Assessment of Neuromuscular Disorders. **International Journal of Electronics and Electrical Engineering**, v. 2, n. 3, p. 242–248, 2014.

GÜLER, N. F.; KOÇER, S. Classification of EMG signals using PCA and FFT. **Journal of Medical Systems**, v. 29, n. 3, p. 241–250, 2005.

HAMILTON-WRIGHT, A.; MCLEAN, L.; STASHUK, D. W.; CALDER, K. M. Bayesian aggregation versus majority vote in the characterization of non-specific arm pain based on quantitative needle electromyography. **Journal of NeuroEngineering and Rehabilitation**, v. 7, n. 1, p. 8, 2010.

HAYKIN, S. **Redes Neurais - Princípios e Prática 2ª Ed**, 200, 2011.

HAYKIN, S. S. **Neural networks and learning machines, 3rd Edition**, 2005.

HESS, U.; ARSLAN, R.; MAUERSBERGER, H.; BLAISON, C.; DUFNER, M.; DENISSEN, J. J. A.; ZIEGLER, M. Reliability of surface facial electromyography. **Psychophysiology**, v. 54, p. 12–23, 2016.

HOSEN, R.; HASAN, S.; HASAN, M.; DAS, R. K. Age Classification Based on EMG Signal Using Artificial Neural Network. n. May, p. 21–23, 2015.

IORDANOVA, I.; RIALLE, V.; VILA, A. Use of unsupervised neural networks for classification tasks in electromyography. **Proc Anna Int Conf IEEE Eng Med Biol Soc**, p. 1014–1015, 1992.

ISTENIČ, R.; KAPLANIS, P. A.; PATTICHIS, C. S.; ZAZULA, D. Multiscale entropy-based approach to automated surface EMG classification of neuromuscular disorders. **Medical and Biological Engineering and Computing**, v. 48, n. 8, p. 773–781, 2010.

JALI, M. H., IZZUDDIN, T. A., BOHARI, Z. H., JAAFAR, H. I. NASIR, M. N. M. Pattern Recognition of EMG Signal During Load Lifting Using Artificial Neural Network. **IEEE International Conference on Control System, Computing and Engineering**, n. November, p. 27–29, 2015.

JOSHI, D.; TRIPATHI, A.; SHARMA, R.; PACHORI, R. B. Computer Aided Detection of Abnormal EMG Signals based on Tunable- Q Wavelet Transform. 2017.

KATIRJI, B. **Electromyography in Clinical Practice**, 2005.

KATSIKIS, C. D.; EXARCHOS, T. P.; PAPALOUKAS, C.; GOLETIS, Y.; FOTIADIS, D. I.; SARMAS, I. A two-stage method for MUAP classification based on EMG decomposition. **Computers in Biology and Medicine**, v. 37, n. 9, p. 1232–1240, 2007.

KATSIKIS, C. D.; GOLETIS, Y.; LIKAS, A.; FOTIADIS, D. I.; SARMAS, I.; TROJABORG, W. A novel method for automated EMG decomposition and MUAP classification. **Artificial intelligence in medicine**, v. 37, n. 1, p. 55–64, maio 2006.

KAUR, G. EMG Diagnosis via AR Modeling and Binary Support Vector Machine Classification. v. 2, n. 6, p. 1767–1772, 2010.

KAUR, G.; ARORA, A.; JAIN, V. Multi-class support vector machine classifier in EMG diagnosis. **WSEAS Transactions on Signal Processing**, v. 5, n. 12, p. 379–389, 2009.

KELEŞ, S.; SUBAŞI, A. Classification Of Emg Signals Using Decision Tree Methods. n. Ci, p. 354–366, 2003.

KHANNA, T.; UPADHYAY, D. K. Design and analysis of higher order fractional step Butterworth filters. **International Conference on Soft Computing Techniques and Implementations, ICSCIT 2015**, n. 2, p. 77–82, 2016.

KIM, H.-C.; PANG, S.; JE, H.-M.; KIM, D.; YANG BANG, S. Constructing support vector machine ensemble. **Pattern Recognition**, v. 36, n. 12, p. 2757–2767, dez. 2003.

KOÇER, S. Classification of Emg Signals Using Neuro-Fuzzy System and Diagnosis of Neuromuscular Diseases. **Journal of Medical Systems**, v. 34, n. 3, p. 321–329, 2010.

KONG, J.; SERTEL, O.; SHIMADA, H.; BOYER, K. L.; SALTZ, J. H.; GURCAN, M. N. Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation. **Pattern Recognition**, v. 42, n. 6, p. 1080–1092, 2009.

KÜÇÜK, H.; EMİNO, İ. Destek Vektör Makinesi Kullanarak ALS Hastalığının Sınıflandırılması Classification of ALS Disease Using Support Vector Machines. p. 0–3, 2015.

KUCUK, H.; TEPE, C.; EMİNOĞLU, İ. Classification of EMG signals by k-Nearest Neighbor algorithm and Support vector machine methods. **2013 21st Signal Processing and Communications Applications Conference (SIU)**, p. 1–4, 2013.

LAHMIRI, S. An Accurate Automatic System for Distinguishing Neuropathy and Healthy Electromyography Signals. p. 1–4, 2017.

MANUEL, C.; ACEVEDO, D. Optimization and classification of EMG signals through Pattern recognition methods. p. 1–2, 2013.

MARCHETTI, P. H.; DUARTE, M. Instrumentação em Eletromiografia, 2011.

MEYER-BAESE, A.; SCHMID, V.; MEYER-BAESE, A.; SCHMID, V. Chapter 1 – Introduction. *In: Pattern Recognition and Signal Analysis in Medical Imaging*, 2004. p. 1–20.

MISHRA, V. K.; BAJAJ, V.; KUMAR, A.; SINGH, G. K. Analysis of ALS and normal EMG signals based on empirical mode decomposition. **IET Science, Measurement & Technology**, v. 10, n. 8, p. 963–971, 2016.

MORAES, N. N. Desenvolvimento e implementação de um sensor híbrido não invasivo de eletromiografia e mecanomiografia para registros de potenciais de ação de uma unidade motora. 2010.

NEGI, S. Feature Extraction and Classification for EMG Signals Using Linear Discriminant Analysis. 2016.

NIKOLIC, M. Detailed analysis of clinical electromyography signals. 2001.

NIKOLIC, M.; KRARUP, C. EMGTools, an adaptive and versatile tool for detailed EMG analysis. **IEEE Transactions on Biomedical Engineering**, v. 58, n. 10 PART 1, p. 2707–2718, 2011.

ONSY, P.; ALIM, A.; MOSELHY, M.; MROUEH, E. F. EMG Signal Processing and Diagnostic of Muscle Diseases. p. 1–6, 2012.

PATTICHIS, C. S.; SCHIZAS, C. N. Genetics-based machine learning for the

assessment of certain neuromuscular disorders. **IEEE Transactions on Neural Networks**, v. 7, n. 2, p. 427–439, 1996.

PATTICHIS, C. S.; SCHIZAS, C. N.; MIDDLETON, L. T. Neural network models in EMG diagnosis. **IEEE Transactions on Biomedical Engineering**, v. 42, n. 5, p. 486–496, 1995.

PHYSIONET.ORG. **Examples of Electromyograms**. Disponível em:
<<https://physionet.org/physiobank/database/emgdb/>>.

PRESTON, D. C.; SHAPIRO, B. E. **Electromyography and neuromuscular disorders : clinical-electrophysiological correlations**. Elsevier Saunders, 2013.

RAEZ, M. B. I.; HUSSAIN, M. S.; MOHD-YASIN, F.; REAZ, M.; HUSSAIN, M. S.; MOHD-YASIN, F. Techniques of EMG signal analysis: detection, processing, classification and applications. **Biological procedures online**, v. 8, n. 1, p. 11–35, 2006.

RANGAYYAN, R. M. **Biomedical Signal Analysis**. 2. ed. [s.l.] Wiley-IEEE Press , 2015.

RASHEED, S.; STASHUK, D.; KAMEL, M. Adaptive fuzzy k-NN classifier for EMG signal decomposition. **Medical Engineering and Physics**, v. 28, n. 7, p. 694–709, 2006.

SCHEEREN, E. M.; KRUEGER, E.; FREITAS, C. DE L. R. **Eletromiografia: O que é e para que serve**. 1. ed. Curitiba: PUCPRes, 2015.

SENGAR, N.; DUTTA, M. K.; TRAVIESO, C. M.; IEEE. Identification of Amyotrophic Lateral Sclerosis using EMG Signals. **2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics**, p. 468–471, 2017.

SUBASI, A. Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines. **Computers in Biology and Medicine**, v. 42, n. 8, p. 806–815, 2012.

SUBASI, A.; YILMAZ, M.; OZCALIK, H. R. Classification of EMG signals using wavelet neural network. **Journal of Neuroscience Methods**, v. 156, n. 1–2, p. 360–367, 2006.

TAKAHASHI, R.; KAJIKAWA, Y. Computer-aided diagnosis: A survey with bibliometric analysis. **International Journal of Medical Informatics**, v. 101, p. 58–67, 2017.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 2010.

TOMCZYKIEWICZ, K.; DOBROWOLSKI, A. P.; WIERZBOWSKI, M. Evaluation of motor unit potential wavelet analysis in the electrodiagnosis of neuromuscular disorders. **Muscle and Nerve**, v. 46, n. 1, p. 63–69, 2012.

TOWE, B. C. Bioelectricity and its measurement. **Standard Handbook of Biomedical Engineering and Design**, p. 17.1-17.50, 2004.

VALENCIA, D.; OREJUELA, D.; SALAZAR, J.; VALENCIA, J. Comparison analysis between rigsure, sqtwolog, heursure and minimaxi techniques using hard and soft thresholding methods. **2016 21st Symposium on Signal Processing, Images and Artificial Vision, STSIVA 2016**, p. 1–5, 2016.

XEL, H.; HUANG, H.; WAN, Z. NETTRO-FUZZY SYSTEM FOR NURALMUSCULAR DISORDERS DIAGNOSIS. p. 5–8, 2004.

XU ZHANG; BARKHAUS, P. E.; RYMER, W. Z.; PING ZHOU. Machine Learning for Supporting Diagnosis of Amyotrophic Lateral Sclerosis Using Surface Electromyogram. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 22, n. 1, p. 96–103, 2014.

YOUSEFI, J.; HAMILTON-WRIGHT, A. Characterizing EMG data using machine-learning tools. **Computers in Biology and Medicine**, v. 51, p. 1–13, 2014.

ZHAO, Q.; LI, Y.; YANG, N.; YANG, Y.; ZHU, M. A Convolutional Neural Network Approach for Semaphore Flag Signaling Recognition. 2016.

ANEXO A

TABELA DE VALORES CORRESPONDENTES AOS ATRIBUTOS DAS CLASSES EM ESTUDO

Classes	Valor Médio Quadrado - (Root Mean Square - RMS)	Taxa de Cruzamento por Zero - (Zero Crossing Rate - ZCR)	Amplitude de Pico à Pico - (Peak to Peak Amplitude - P2P)	Mediana da Frequência - (Median Frequency - Mef)	Média de Densidade Espectral de Potência - (Average Power Spectral Density - APD)	Tamanho da forma de onda - (Waveform Length - WL)	Frequência de Pico - (Peak Frequency - PF)	Frequência Média - (Mean Frequency - MF)
Normal	113.8146742	3763	128.2647216	184.5713941	0.049414757	4360.056624	4.347912724	413.0402246
Normal	63.44532256	3210	99.08629448	221.8738144	0.015355336	4815.968611	2.112145186	435.2041347
Normal	169.988555	3230	239.1275584	607.3974912	0.110229907	17557.82869	4.769178314	633.6529352
Normal	86.40542209	3420	140.8064585	648.0416161	0.028480139	7739.66034	2.09344529	616.1929066
Normal	61.43761647	3061	81.68299618	194.5800324	0.014398891	1804.561749	2.534441624	322.4196503
Normal	73.2742153	4084	118.0250727	621.8229723	0.020481332	4806.070962	2.043500604	548.5198211
Normal	67.85547242	4732	112.7971717	629.3655433	0.01756426	3446.579062	1.830872647	567.3885708
Anormal	147.8919488	6053	208.4606236	768.5823454	0.083435198	36567.70803	2.98292548	1162.224901
Anormal	120.7080706	5626	225.7439947	658.3475974	0.05557997	11430.12919	3.262119131	712.5611827
Anormal	25.60565338	4545	44.6307398	213.4537121	0.002501105	1274.075621	0.881868238	379.5737531
Anormal	387.5315255	2694	278.5953675	131.770298	0.572893843	8884.686726	15.48627501	248.3719697
Anormal	1239.1001229	2591	1614.244217	480.007922	5.856968462	74364.15815	44.8808573	445.8095111
Anormal	191.5071405	2396	163.2266087	157.4839155	0.139903965	4766.798855	8.174303546	286.40181
Anormal	379.3837746	5367	718.2012548	652.970243	0.548057193	100830.5448	11.01811011	1006.109361
Anormal	496.7098988	3846	813.944894	507.0936319	0.94116488	21259.07322	14.09865081	439.254428

Quantidade de MUAP - (MUAP Quantity - MUAPQ)	Pico Máximo - (Peak)	Tempo de Subida - (Rise Time - RS)	Quantidade de Fases - (Phase - PQ)	Mudança de Inclinação do Sinal - (Slop Sign Changes - SSC)
2352	101.9278978	3.580696567	1881	8244
1771	79.40138808	4.599515602	1855	7888
1815	145.1578675	1.201282341	1625	6728
1721	99.75679	4.941114735	1710	7340
1482	41.89662103	0.06588378	1531	4290
2471	81.19510907	0.578762529	2042	9509
2418	60.95635091	0.74384578	2366	7944
2686	148.6265946	0.07410548	3031	11366
2640	143.6011727	0.573971814	2813	10719
2214	27.21817212	0.694155527	2272	8873
1748	210.7069695	4.53876789	1347	7030
1296	629.0336537	0.158834223	1295	6080
1513	103.230219	0.295414363	1198	6562
2000	500.2754566	0.634072628	2683	11845
1685	444.6824237	0.00395067	1923	7994