

Reconhecimento de expressão facial assistido por EMG e aprendizado de máquina

1st Paulo Cirillo S. Barboza

Departamento de Engenharia de Teleinformática (DETI)
Universidade Federal do Ceará (UFC)
Centro de Tecnologia, Fortaleza, Ceará
paulocir@alu.ufc.br

2st Raul F. Santana

Departamento de Engenharia de Teleinformática (DETI)
Universidade Federal do Ceará (UFC)
Centro de Tecnologia, Fortaleza, Ceará
raulfontenele@alu.ufc.br

3st George A. P. Thé

Departamento de Engenharia de Teleinformática (DETI)
Universidade Federal do Ceará (UFC)
Centro de Tecnologia, Fortaleza, Ceará
george.the@ufc.br

4th Guilherme de Alencar Barreto

Departamento de Engenharia de Teleinformática (DETI)
Universidade Federal do Ceará (UFC)
Centro de Tecnologia, Fortaleza
gbarreto@ufc.br

Resumo—O presente trabalho apresenta um estudo sobre análise das atividades de sinais de Eletromiografia (EMG) a fim de realizar o reconhecimento de cinco gestos faciais. Em uma etapa inicial, é realizada a aquisição desses sinais estudando a escolha das expressões faciais e um melhor posicionamento dos sensores. Na etapa de classificação, utilizaram-se cinco diferentes classificadores e fez-se uma análise exploratória de seus hiperparâmetros, a fim de buscar um melhor desempenho para resolver o problema do reconhecimento de padrões. Além disto, essas análises que são baseadas na capacidade de acerto do classificador e custo computacional, são utilizadas em um estudo comparativo a fim de identificar quais dos métodos propostos, tem a capacidade de lidar com um conjunto de dados generalizado. Como resultados, identificou-se que para as expressões e métodos escolhidos, foi possível obter taxas de acerto superiores a 95% destacando como melhores resultados obtidos pelos os algoritmos QC, kNN e MLP.

Index Terms—Instrumentação, EMG, Reconhecimento de Padrões.

I. INTRODUÇÃO

As reações emocionais humanas são um assunto de grande relevância, e não raro se encontram estudos relatando investigações de percepção sensorial e/ou detecção de emoções a partir de informações faciais em diferentes campos do conhecimento. De fato, o reconhecimento, a interpretação, o processamento e a simulação de emoções assistidos por dispositivos eletrônicos têm sido objeto de interesse em aplicações que vão da segurança à saúde, passando pelo entretenimento. [1] [2]. No contexto de reconhecimento de emoções pela face, os trabalhos de Paul Ekman são uma referência e um bom ponto de partida, pois consistem em uma descrição de expressões universais a partir de movimentos de músculos da face, que permitem identificar um espectro de emoções derivado de seis emoções básicas [3].

Com frequência, a percepção do estado emocional a partir de informações faciais é realizada a partir de imagens RGB provenientes de câmera CCD ou similares, vista a natureza não-invasiva de aquisições dessa natureza. Sabe-se, contudo,

que limitações como a sensibilidade às condições de luminosidade ambiente, à postura na cena e o uso de artefatos estranhos afetam sobremaneira a percepção de informações da face. A utilização de alternativas à aquisição é, portanto, recomendável não em substituição mas, preferencialmente, como complemento àquela realizada por sistemas de visão, numa perspectiva de integração multimodal.

Este trabalho encontra lugar justamente no contexto de percepção sensorial de expressões faciais a partir de sinais eletromiográficos da face. Eletromiografia (EMG), em essência, consiste em fazer uso da atividade elétrica de membranas excitáveis, as quais são componentes do tecido nervoso, muscular ou glandular, de que seja possível monitorar a atividade muscular em resposta a movimentos. O sinal chamado de eletromiograma pode ter amplitude típica na faixa entre 0.1 mV e 90 mV e com resposta espectral entre 80Hz e 5KHz [4], [5].

Sendo assim, no presente trabalho, apresenta-se em detalhes o desenvolvimento de um sistema de aquisição de biosinais da face e o projeto de um sistema de classificação de expressões faciais, a partir de algoritmos de aprendizado de máquina de amplo conhecimento da comunidade científica. Em vista dos resultados obtidos, o trabalho pode ser visto como um bom caso de uso de inteligência computacional em processamento de sinais de EMG.

A. Trabalhos Relacionados

O processamento de biosinais associado a algoritmos de aprendizagem de máquina constituem importante tópico de pesquisa para a comunidade de reconhecimento de padrões. De fato, destacam-se trabalhos com temas similares à pesquisa deste, ao se utilizar de sinais biomédicos com algoritmos de aprendizagem de máquina, para resolver problemas de reconhecimento de padrões.

Neste sentido, destacam-se os seguintes trabalhos: Utilização de técnicas de extração de atributos de sinais de

EMG, associados à algoritmos de aprendizagem de máquina para realizar reconhecimento de gestos faciais [6]; Sistema de monitoramento em pacientes, que utiliza aquisição em tempo real de sinais EMG faciais, associados ao algoritmo kNN, a fim de realizar o reconhecimento de padrão quando o paciente está com dores [7]; Reconhecimento de emoções faciais utilizando LabView para aquisição de sinais EMG, processamento de sinal e classificação [8]; Utilização de biosinais EMG e Eletro-oculografia (EOG) para detectar estados de emoções em pacientes [9]

Além desses trabalhos, pode-se descrever sobre outros que realizaram análises de desempenho similares ao presente trabalho, como por exemplo: Reconhecimento de expressões faciais através de sistema robusto [10], que utiliza em sistema embarcado, aquisição e filtragem de sinais EMG faciais. Neste trabalho, o autor faz um processo investigativo ao utilizar diversos métodos de extração de atributos e algoritmos de aprendizagem de máquina para classificação dos gestos propostos. Para avaliar os métodos, utilizou-se a medida de informação mútua normalizada e verificou-se que os métodos de RMS e MAV para extração de atributos, associados ao algoritmo de Máxima Verossimilhança (MV), obtendo as maiores taxas de acerto com valores $99,25\% \pm 0,31$ e $99,31\% \pm 0,55$ respectivamente [10]; Reconhecimento de sorrisos espontâneos e forçados utilizando instrumentação e processamento de biosinal EMG facial com comparativo em 16 usuários. No experimento proposto pelos autores, utiliza-se processo de estímulos para que o sistema identifique os sorrisos. Neste trabalho utilizou-se Independent Component Analysis (ICA) para extração de atributos e na etapa de classificação, fez-se o uso do classificador *Support Vector Machine* (SVM) com análises temporais e espaciais. As taxas de acerto dos pacientes nessas duas análises ficaram respectivamente entre 56,10% até 97,80% e 85,23% até 96,43%. Os autores justificam a causa dos erros em classificar pelo fato que um sorriso forçado tende a ser mais duradouro, ao passo que um espontâneo pode ser realizado em um quarto de segundo. [11]

II. FUNDAMENTAÇÃO TEÓRICA

A presente seção contém uma breve descrição dos classificadores utilizados no presente estudo. A escolha destes está relacionada com a tentativa de avaliar a etapa de classificação através de algoritmos lineares e não lineares.

A. *k*-Nearest Neighbor - (kNN) e Distância Mínima ao Centróide - (DMC)

A partir de uma métrica de dissimilaridade, os classificadores kNN e DMC realizam associação entre as amostras e as classes originalmente definidas, as quais podem ser vetores presentes no conjunto de dados, no caso do kNN, ou vetores médios representativos das classes presentes nos dados, no caso do DMC. Durante a tomada de decisão, o kNN realiza uma contagem de frequência das classes associadas as k menores distâncias, e associa a amostra à classe com maior frequência [12]. Já no DMC, a associação é com a classe cujo centróide esteja mais próximo da amostra. [13].

B. Linear de Mínimos Quadrados - (LMQ)

Este classificador adiciona uma etapa de determinação da matriz de estimação do modelo, a qual relaciona os dados do conjunto de treino, \mathbf{X} , às medidas, \mathbf{Y} , relativas às classes conhecidas [14], conforme Eq. 1, (aqui já apresentada na sua versão com regularização por Tikhonov no parâmetro λ , para mitigação de eventual mau condicionamento na inversão):

$$\mathbf{W} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T + \lambda\mathbf{I}_n)^{-1} \quad (1)$$

Determina-se, portanto, uma associação a cada classe do problema sempre que uma nova amostra é apresentada, conforme Eq. 2, e decide-se pela classe correspondente à posição do vetor $\mathbf{y}_{predito}$ que contém o maior elemento.

$$\mathbf{y}_{predito} = \mathbf{W} \cdot \mathbf{X}_{amostra} \quad (2)$$

C. Quadratic Classifier - (QC)

É um classificador que utiliza métrica de dissimilaridade a partir de distância quadrática (Mahalanobis) em que, para uma nova amostra não classificada, atribui-se o rótulo a essa amostra em modo a atender a Eq. 3. Na equação, \mathbf{m}_i é o vetor médio associado a i -ésima classe, e \mathbf{C}_i é a matriz de covariância associada à i -ésima classe [15].

$$i^* = \arg \min_{i=1, \dots, K} (\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i) \quad (3)$$

No QC a invertibilidade também é uma questão; uma forma de contornar o eventual mau condicionamento na inversão de \mathbf{C}_i é usar o método da matriz de covariância agregada, realizando um somatório ponderado das matrizes de covariâncias das classes, conforme Eq. 4, em que $p(w_i)$ é a probabilidade a priori de uma amostra pertencer à classe i . Pode-se usar a regularização por Friedman, que faz a combinação linear da matriz agregada, com as matrizes de covariância de cada classe, de acordo com a Eq. 5 [15].

$$\Sigma_{pool} = \sum_{i=1}^K p(w_i) \Sigma_i = \frac{n_1}{N} \Sigma_1 + \frac{n_2}{N} \Sigma_2 + \dots + \frac{n_K}{N} \Sigma_K \quad (4)$$

$$\Sigma_i^\lambda = \frac{(1 - \lambda) \mathbf{S}_i + \lambda \mathbf{S}_{pool}}{(1 - \lambda) n_i + \lambda N} \quad (5)$$

D. Multilayer Perceptron - (MLP)

O perceptron de múltiplas camadas (MLP), ilustrado na Fig. 1, é uma rede neural artificial (RNA) que consegue aprender a relação de entrada e saída dos dados mediante ajuste dos pesos sinápticos associados aos neurônios que compõem sua estrutura multi-camadas, o qual é realizado através de um procedimento iterativo de retro-propagação do erro entre a saída produzida e a desejada.

III. METODOLOGIA

Nesta seção, descreve-se o experimento de reconhecimento facial realizado a partir de resposta de músculos da face adquiridas por sensores de EMG. Detalhes a respeito do arcabouço prático relacionado à realização de expressões faciais e sua aquisição por meios eletrônicos são trazidos.

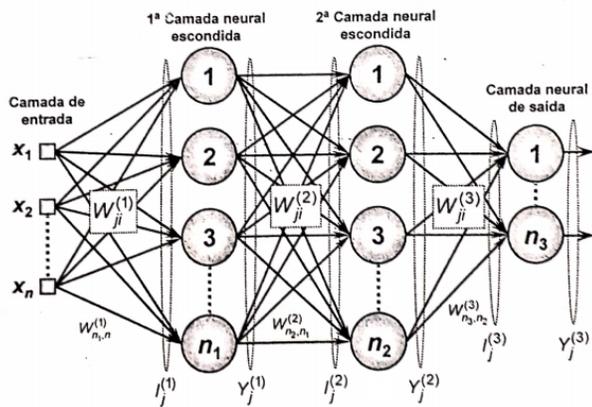


Figura 1. Representação de rede MLP [16].

A. Arcabouço prático

Do ponto de vista prático, a questão da realização de expressões pode ser bem compreendida a partir do conceito de unidades de ação, que representam grupos musculares presentes na face. De acordo com [17], os gestos e expressões faciais são resultado do movimento combinado de tais unidades. Dada a pluralidade de movimentos possíveis, a compatibilidade entre quantidade de sensores disponíveis e os grupos musculares associados às diversas expressões e para fins de delimitação de escopo, no presente trabalho optou-se por selecionar o conjunto de expressões mostrado na Fig. 2.

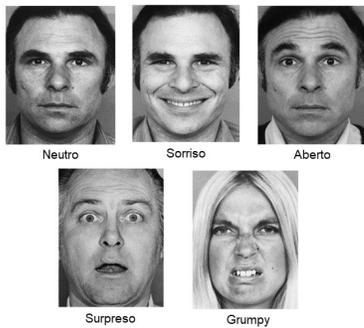


Figura 2. Expressões faciais escolhidas [17].

A solução eletrônica adotada para a percepção sensorial foi o conjunto formado por um par de sensores de EMG chamados *Myoware Muscle Sensor*, associados a eletrodos de Ag/AgCl e conectados a um microcontrolador NodeMCU-ESP32, que permite comunicação sem fio, assim dispensando excesso de fiação que poderia piorar a relação sinal-ruído dos sinais musculares. Como um dispositivo de instrumentação, o sensor realiza aquisição diferencial, e possui uma eletrônica integrada voltada a filtragem de passagem de banda, para mitigar efeitos de ruídos inerentes à aquisição (vide esquema da Fig. 3) [18]. Portanto, como essa saída já tem esse processamento do sinal bruto, no presente trabalho não utilizou-se métodos de extração de atributos.

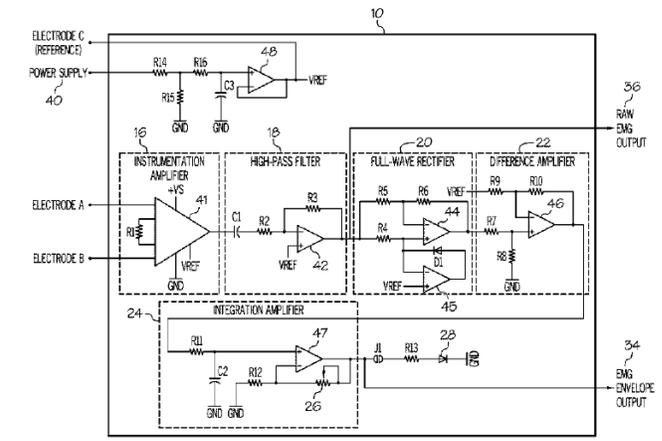


Figura 3. Esquemático do sensor de EMG [18].

B. Ensaio Preliminar

O posicionamento dos sensores na face é uma questão importante. Inicialmente, a escolha de posicioná-los sobre o zigomático maior e o corrugador do supercílio não se mostrou oportuna (vide Fig. 4b), embora tenham sido consideradas as unidades de ação conforme [17]. Com os sensores nesta que convencionamos chamar posição B, foram realizadas aquisições de: *i. conservação* de expressão simulada e de *ii. transição* de expressão simulada. Na primeira, o participante do ensaio era orientado a manter a expressão fixa durante toda a medição, ao passo que na segunda, era permitida uma transição para o estado neutro ao final da medição.

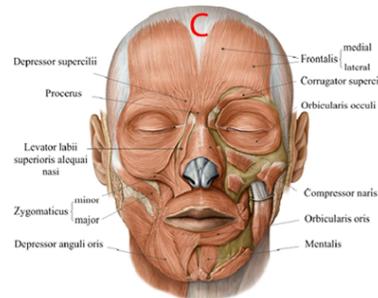
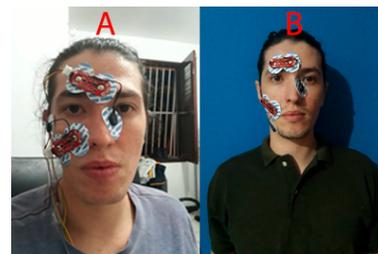


Figura 4. Posicionamentos inicial (B) e final (A) dos eletrodos, e localização muscular em face (C).

Na Figura 5, são apresentados os resultados da aquisição de *transição* para as expressões que compõem o universo definido, para transição ocorrendo no instante $t = 7s$. As

séries temporais são trazidas em escala de unidades arbitrárias de intensidade que indicam a leitura correspondente a 12 bits do conversor AD do microcontrolador. Chama a atenção na Figura o fato de que apenas a medição relativa à expressão Sorriso apresentou o comportamento esperado, qual seja, o de percepção de transição elétrica apenas ao final da janela de aquisição.

Especulou-se, em vista disso, um hipotético mau posicionamento dos sensores, que levaria ao afastamento da unidade muscular associada à expressão, portanto adotou-se a posição A indicada na Fig. 4a para as medições seguintes. Esta escolha mantinha ao menos um dos eletrodos posicionado sobre o músculo mais relevante para a expressão facial. Os experimentos preliminares de aquisição de *i. conservação* e *ii. transição* foram repetidos e seus resultados são trazidos na Fig. 6. Nota-se que, para expressões sendo realizadas entre os instantes de transição $t = 1s$ e $t = 6s$, os sinais dos sensores parecem representar melhor a existência de a atividade muscular.

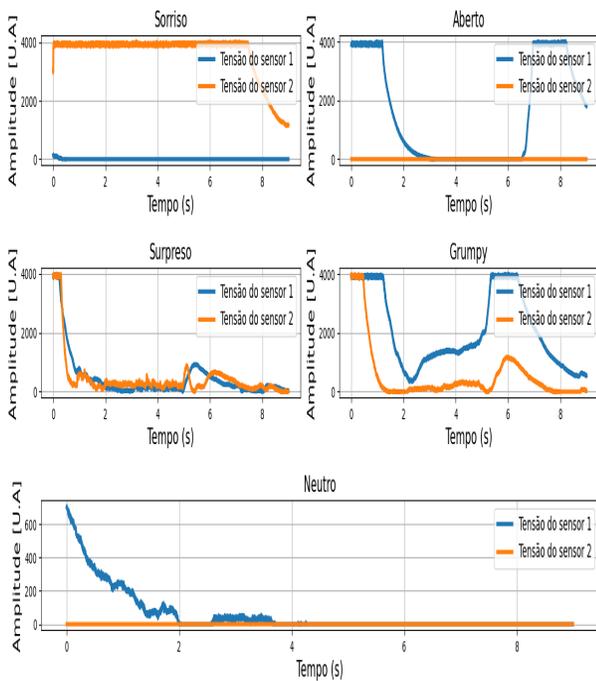


Figura 5. Aquisição de transição com posicionamento B dos sensores

Em complemento, a Fig. 7a apresenta o gráfico de dispersão das séries temporais dos dois sensores, para o inteiro conjunto de classes, e considerando o cenário de aquisição de *conservação* da expressão facial. Percebe-se uma sobreposição entre algumas das classes, o que pode estar associada à limitada repetibilidade das aquisições em virtude da dificuldade em se realizar repetições ideais dos movimentos faciais. De modo a mitigar esta questão, o conjunto de dados foi aumentado através da inclusão de diferentes rodadas de aquisição, e os gráficos de dispersão das rodadas 6 e 9 são trazidos na Fig. 8.

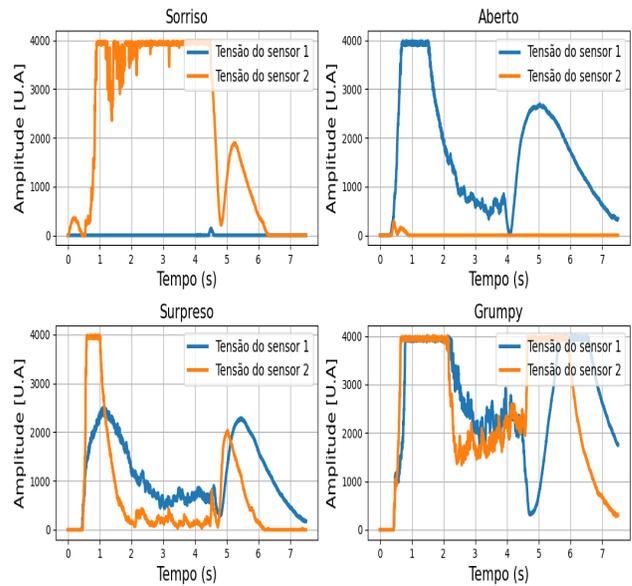


Figura 6. Aquisição de transição com posicionamento A de sensores

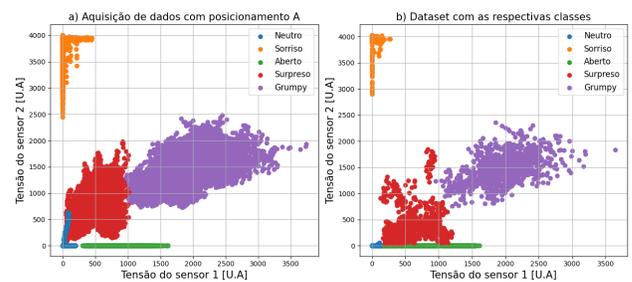


Figura 7. Aquisição sem variação do gesto facial com posicionamento A de sensores e dataset

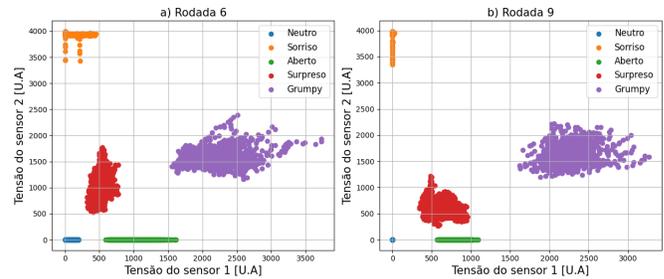


Figura 8. Gráfico de dispersão das rodadas 6 e 9 com posicionamento A de sensores

C. Criação do dataset

Com vistas a definir um *dataset* suficientemente representativo dos dados e da incerteza a eles associada (sobretudo aquela ligada à incapacidade de o participante do ensaio em repetir os movimentos faciais), foi realizada a extração de uma amostra com 100 dados, escolhida de forma aleatória, de cada um dos sensores e para cada uma das cinco expressões faciais

por rodada, resultando em um conjunto com 5 mil dados por sensor, ou 10% do superconjunto de aquisição. Um gráfico de dispersão do conjunto de dados está mostrado na Fig. 7b.

A Tabela I apresenta algumas informações estatísticas das classes, como seu ponto médio (P.M), desvio padrão (D.P) e valores mínimos e máximos ao longos dos eixos X e Y, que representam os sensores. Como pode ser visualizado na Tabela I, as expressões Surpreso e Grumpy, apresentam maior desvio padrão em ambas as direções, o que é um indicativo de expressões difíceis de serem repetidas. Também pode ser visto que as expressões Sorriso e Aberto, apresentam suas maiores variações sobre os eixos Y e X respectivamente, uma vez que esses eixos representam os sensores associados aos músculos com maior ativação para cada expressão.

Merece destaque que um *dataset* mais adequado para o problema de classificação de expressões seria construído a partir de amostras de movimentos faciais de um conjunto mais plural e diverso de participantes, de diferentes gêneros, idades e tamanhos, pois todas estas são variáveis que afetam a resposta muscular. Contudo, as restrições impostas pelo cenário de emergência sanitária atual limitaram sobremaneira a capacidade de investigação da generalização dos sistemas de classificação implementados. Portanto, os dados foram coletados apenas para um indivíduo do sexo masculino e com idade de 31 anos.

Tabela I
INFORMAÇÕES ESTATÍSTICAS DAS CLASSES

Classe	P. M.		D. P.		Mín		Máx	
	X	Y	X	Y	X	Y	X	Y
Neutro	16,86	0,21	45,05	2,91	0	0	231	53
Sorriso	9,74	3921,55	33,25	123,97	0	2896	270	4032
Aberto	904,67	0	371,53	0	176	0	1611	0
Surpreso	1966,79	1468,83	384,24	272,47	951	757	3646	2357
Grumpy	624,57	409,89	279,03	324,72	129	25	1203	1845

IV. RESULTADOS E DISCUSSÃO

Esta seção discute a validação dos experimentos de classificação e os seus resultados. Para os experimentos, foram escolhidos e implementados 5 algoritmos, dentre os quais estão o kNN e o DMC, que se baseiam na distância euclidiana entre os pontos, o algoritmo LMQ, que é baseado em regressão linear e modificado para realizar classificações, o algoritmo QC, que se baseia na distância de Mahalanobis, e a rede MLP. As experimentações foram divididas em duas etapas; inicialmente, foi realizada uma investigação paramétrica para definição da configuração dos algoritmos. Então, a partir de uma quantidade maior de dados, procedeu-se à validação da configuração definida, etapa em que se selecionou a melhor configuração de cada classificador em que o desempenho médio de acerto superou 95%.

A. Investigação paramétrica dos classificadores

Como já mencionado, o primeiro experimento visava aferir o desempenho dos classificadores com variação dos seus hiperparâmetros. Para isso, cada um dos algoritmos baseados em distância foi executado 100 vezes enquanto o algoritmo baseado em redes neurais foi executado duas vezes. Essa distinção foi feita devido à escala de tempo para a execução de cada um dos algoritmos, que será apresentada e discutida posteriormente. Durante as execuções, foi utilizado o *dataset* citado na subseção anterior, com 80% dos dados sendo utilizado para controle ou treinamento e 20% para validação ou classificação. As amostras de treinamento e classificação variavam entre cada uma das execuções, mantendo sempre a proporção entre as classes. Para cada um dos classificadores será apresentada a média (TA), a menor (Min) e a maior (Max) das taxas de acerto, seu desvio padrão (DP) e o tempo de execução.

O primeiro algoritmo a ser apresentado é o DMC, o qual não conta com etapa de ajuste de hiperparâmetros; resultados vêm na Tabela II. O segundo algoritmo a ser apresentado é o kNN, em que a dimensão da vizinhança é estudada em tamanhos ímpares (com saltos de 10 unidades) desde 1 a 141, de modo a minimizar a chance por empate entre classes candidatas. Como apresentado na Tabela III, é possível perceber uma tendência de queda do desempenho com o aumento de k.

Tabela II
RESULTADOS DO ALGORITMO DMC

T. A.(%)	D. P.(%)	Mín(%)	Máx(%)	Tempo(s)
90,57	0,83	88,30	92,30	0,05

Tabela III
RESULTADOS DO KNN COM K VARIANDO DE 1 A 141

k	T. A. (%)	D. P. (%)	Mín (%)	Máx (%)	Tempo (s)
1	99,65	0,15	99,3%	100%	16
11	99,65	0,18	99,1%	100%	16
21	99,54	0,21	98,9%	99,9%	16
31	99,43	0,23	98,9%	99,9%	16
41	99,23	0,27	98,5%	99,8%	17
61	98,52	0,37	97,7%	99,4%	16
81	98,16	0,39	97,1%	99,0%	16
101	97,90	0,46	96,90%	98,90%	16
121	97,57	0,46	96,50%	98,60%	16
141	97,36	0,47	96,10%	98,40%	16

Na sequência, o algoritmo LMQ, o qual não conta com hiperparâmetros a serem configurados na presente versão. Resultados vêm na Tabela IV e revelam que o acerto médio foi significativamente inferior ao dos algoritmos anteriores. Especula-se que isto esteja associado à baixa capacidade do algoritmo LMQ de rotular corretamente este conjunto de dados, que possui dois preditores e cinco classes. Outro fator pode estar associado a um mau condicionamento da matriz inversa, discutido na seção II. Procedeu-se à regularização por Tikhonov, em que o parâmetro λ foi variado de 0 até 0,95

com passo de 0,05, mas os resultados revelaram pouco efeito na taxa de acerto média.

Tabela IV
RESULTADOS DO ALGORITMO LMQ PADRÃO.

T. A. (%)	D. P. (%)	Max (%)	Min (%)	Tempo (s)
57,25	1,39	60,9	53,3	0,00036

O quarto classificador utilizado neste trabalho é o QC, que é dividido em três algoritmos. Os resultados dos dois primeiros são exibidos através da Tabela V, em que o QC Simples utiliza a matriz de covariância associada à cada classe conhecida do conjunto de dados, e o QC Agregada utiliza o cálculo através da Equação 4 exibida na seção II, para contornar eventual mau condicionamento da matriz de covariância.

Tabela V
RESULTADOS OBTIDOS QC SIMPLES E QC AGREGADA

	T.A. (%)	D.P. (%)	Mín (%)	Máx (%)	Tempo (s)
QC Simples	83,05	1,19	80,20	85,60	0,193371
QC Agregada	90,66	0,84	88,00	92,90	0,195089

O terceiro algoritmo associado ao classificador em questão é a versão que utiliza a regularização pelo método de Friedman, representado pela Equação 5 na seção II. Como este método possui um hiperparâmetro, avaliou-se seu efeito para λ assumindo valores de 0 a 1 (vide Tabela VI).

Tabela VI
VARIÇÕES DA REGULARIZAÇÃO POR FRIEDMAN UTILIZADA NO QC

λ	T.A. (%)	D.P. (%)	Mín (%)	Máx (%)	Tempo (s)
0	83,12	1,32	78,60	86,10	0,1872345
0,1	97,23	0,85	94,60	98,70	0,1898092
0,2	97,30	0,65	95,60	98,70	0,2224838
0,3	97,13	0,89	94,00	98,70	0,3421870
0,4	96,82	0,91	94,50	99,10	0,1686336
0,5	96,57	0,96	93,80	98,60	0,1532719
0,6	96,54	0,83	93,50	98,30	0,1595738
0,7	96,01	0,82	94,20	97,60	0,1538600
0,8	95,57	0,77	92,80	97,30	0,1547057
0,9	94,69	0,76	92,90	96,60	0,1767676
1	94,30	1,32	78,60	86,10	0,1657677

Por fim, foram realizados experimentos utilizando a rede MLP. Fixou-se a taxa de aprendizado em 0,3 nos testes iniciais e se definiu uma topologia com duas camadas ocultas, com quantidade de neurônios como parâmetro de investigação. Para o primeiro teste, cada camada escondida possuía a mesma quantidade de neurônios, que variaram entre 1 e 17, com passo de 2. Na Tabela VII é possível visualizar o resultados dessa variação de parâmetros, mostrando algumas taxas médias de acerto e desvio padrão com variações menor que 0,5%, porém com uma diferença na quantidade de épocas necessárias para a convergência, o que pode indicar uma taxa de aprendizado

elevada. Também se percebe a diferença na ordem de grandeza do tempo de execução com respeito aos demais algoritmos.

Também pode ser observado na Tabela VII que para o caso em que foram utilizadas duas camadas com apenas um neurônio em cada, a taxa de acerto é inferior aos resultados obtidos pelo QC e kNN. Tal fato pode ter ocorrido possivelmente por conta de um *underfitting* e, por isso, resolveu-se aumentar a quantidade de neurônio por camada afim de aumentar o desempenho. Para tal, também foram utilizadas duas camadas escondidas com a mesma quantidade de neurônios em cada, porém desta vez com o número variando entre 20 e 120 com passo de 10. A partir das execuções pôde-se perceber que a partir da configuração com 90 neurônios, há uma queda progressiva nas taxas de acerto, tempo e quantidade de épocas necessários para a convergência. Isto pode ter acontecido em função de um possível *overfitting*, e portanto, escolheu-se a topologia com 70 neurônios em cada camada.

Tabela VII
RESULTADOS MLP COM DUAS CAMADAS ESCONDIDAS E VARIANDO-SE A QUANTIDADE ENTRE 1 E 120

Top.	T. A. (%)	D. P. (%)	Mín (%)	Max (%)	Tempo	Épocas
1	75,65	2,85	73,40	77,90	00:41:03	7203
3	99,75	0,76	99,70	99,80	01:04:59	10739,5
7	99,85	0,07	99,80	99,90	00:45:38	7260
11	99,70	0,00	99,70	99,70	01:17:01	12000
15	99,70	0,14	99,60	99,80	01:16:25	12000
20	99,80	0,10	99,70	99,90	01:12:45	10926,5
40	99,50	0,40	99,10	99,90	00:54:46	7060,5
60	99,80	0,20	99,60	100,00	00:50:20	12000
70	99,90	0,00	99,90	99,90	00:58:50	12000
80	67,00	32,50	34,50	99,50	00:43:55	6000,5
90	30,35	12,15	18,20	42,50	00:31:15	2661
100	20,00	0,20	19,80	20,20	00:04:01	185,5
110	10,05	8,95	1,10	19,00	00:00:47	33,5
120	21,00	0,40	20,60	21,40	00:00:11	6,5

Após o final deste conjunto de experimentos, conclui-se que os algoritmos DMC e LMQ obtiveram os piores resultados de taxa de acerto, em relação aos outros algoritmos, com índices inferiores a 90%, com uma tendência de queda nas taxas de acerto para os casos onde os dados estiverem mais sobrepostos.

B. Validação dos classificadores selecionados

O segundo experimento, como já mencionado, visava validar a melhor configuração dos classificadores que obtiveram médias de taxas de acerto superiores a 95%, utilizando um conjunto de dados com uma maior quantidade de dados em relação ao *dataset* mencionado na subseção III-C. Os algoritmos utilizados nesse experimentos foram o kNN, o QC e MLP. O primeiro *dataset* criado, com 5000 amostras, foi utilizado como definidor dos hiperparâmetros para o kNN e para o QC, e como treinamento para o MLP. Cada algoritmo foi executado 50 vezes afim de classificar um novo *dataset* criado a cada execução. O processo para criação do novo conjunto de dados foi o mesmo utilizado anteriormente, porém com uma amostragem de 500 dados por sensor, por gesto e por rodada, totalizando 25000 para cada sensor.

Neste segundo experimento, também foram realizadas análises com relação à capacidade da tomada de decisão dos melhores classificadores através da composição de matrizes de confusão. Portanto, para cada uma das 50 execuções de cada algoritmo, compôs-se uma matriz de confusão e desta, extraiu-se informações de sensibilidade e especificidade, que são informativos relacionados à predição de uma das classes. Para realizar o cálculo destas taxas foi utilizado o método um contra todos para gerar uma análise individual de cada um dos gestos, e que transforma a matriz de confusão original em cinco matrizes de ordem 2×2 . Com estas matrizes foi realizado o cálculo das taxas de acordo com as Equações 6, 7. Como a execução dos classificadores repete-se 50 vezes, calculam-se as médias das taxas que são exibidas nos resultados individuais para cada método.

$$Sens = \frac{VP}{VP + FN} \quad (6) \quad Espe = \frac{VN}{VN + FP} \quad (7)$$

Para o classificador kNN foram escolhidas as configurações com valores de k iguais a 1 e a 11, pois seus resultados foram semelhantes e, ainda, para $k = 1$ o algoritmo é *Nearest Neighbor* e pode ter perda de desempenho. A Tabela VIII sintetiza os resultados.

Tabela VIII
RESULTADOS DO KNN NO SEGUNDO EXPERIMENTO

k=1	Taxa de Acerto Média		DP	
	99,727%		0,02%	
	Sens. Média	Sens. DP	Especif. Média	Especif. DP
Neutro	99,78%	0,02%	99,88%	0,02%
Sorriso	100,00%	0,00%	99,66%	0,03%
Aberto	100,00%	0,00%	99,66%	0,03%
Surpreso	99,93%	0,01%	99,73%	0,02%
Grumpy	99,95%	0,01%	99,70%	0,03%
k=11	Taxa de Acerto Média		DP	
	99,731%		0,02%	
	Sens. Média	Sens. DP	Especif. Média	Especif. DP
Neutro	99,34%	0,08%	99,83%	0,02%
Sorriso	100,00%	0,00%	99,66%	0,03%
Aberto	100,00%	0,00%	99,66%	0,03%
Surpreso	99,74%	0,06%	99,73%	0,03%
Grumpy	99,58%	0,06%	99,77%	0,02%

Com a execução dos três algoritmos do QC, foi possível identificar que, ao se utilizar a regularização pelo método de Friedman e para um $\lambda = 0,2$, obteve-se uma maior taxa de acerto média, atingindo 97,30%, com desvio padrão de 0,64%. Logo, para este segundo experimento e fixando λ em 0,2 pôde-se compor a Tabela IX que traz o desempenho para as diversas classes. É possível verificar que o desempenho para cada expressão é inferior ao do classificador anterior, destacando o gesto Sorriso que não teve todas as suas amostras classificadas corretamente nas 50 execuções do algoritmo.

Como dito anteriormente, para o algoritmo do MLP foi utilizada a configuração com duas camadas escondidas e 70

Tabela IX
RESULTADOS DO QC COM $\lambda = 0,2$ EXPERIMENTO

$\lambda = 0,2$	Taxa de Acerto Média		DP	
	99,08%		0,04%	
	Sens. Média	Sens. DP	Especif. Média	Especif. DP
Neutro	97,63%	0,15%	99,49%	0,03%
Sorriso	99,25%	0,07%	99,09%	0,05%
Aberto	100,00%	0,00%	98,90%	0,04%
Surpreso	99,66%	0,05%	98,99%	0,04%
Grumpy	99,07%	0,09%	99,13%	0,04%

neurônios em cada uma dessas camadas. Além disto, definiu-se como taxa de aprendizado o valor de 0,3. Esta configuração, assim como as demais, foi escolhida por ter apresentado a maior taxa média de acerto com o menor desvio padrão. Como, em ensaios anteriores esta configuração atingiu o limite de 12000 épocas, procedeu-se a alterar a condição de parada para 14000, de modo a tentar reduzir adicionalmente o erro quadrático médio. O presente classificador, que difere dos dois anteriores devido à necessidade de treinamento, foi treinado apenas uma vez e, posteriormente utilizado nas 50 execuções de classificação. A Tabela X traz os resultados de desempenho por cada classe e revela que a MLP se mostrou mais eficiente que as demais soluções avaliadas.

Tabela X
RESULTADOS DO MLP COM DUAS CAMADAS ESCONDIDAS NO SEGUNDO EXPERIMENTO

Top. 70 Max. Epo: 14000	Taxa de Acerto Média		DP	
	99,79%		0,02%	
	Sens. Média	Sens. DP	Especif. Média	Especif. DP
Neutro	99,19%	0,10%	99,94%	0,01%
Sorriso	100,00%	0,00%	99,74%	0,03%
Aberto	100,00%	0,00%	99,74%	0,03%
Surpreso	99,92%	0,03%	99,76%	0,03%
Grumpy	99,86%	0,03%	99,78%	0,03%

A fim de expressar graficamente as informações obtidas, escolheu-se aleatoriamente uma das 50 execuções para produzir matrizes de confusão para cada método. Ilustradas na Figura 9, exibem informações condizentes com os trazidos nas tabelas anteriores. Em uma análise inicial, pode-se verificar, pela diagonal principal de todas as matrizes, que a maioria dos gestos foi predito corretamente, porém nesta rodada em específico, os quatro classificadores fizeram predições erradas para a expressão Neutro, que foi predito como Surpreso algumas vezes. Destaca-se, ainda, a maior ocorrência de confusões pelo classificador QC, como esperado pelas taxas relatadas.

V. CONCLUSÃO

O presente trabalho apresentou um estudo de caso sobre a aquisição de sinais de EMG para reconhecimento de expressões faciais. Foram realizados experimentos com diversos classificadores a fim de averiguar quais os mais adequados para o problema proposto. Os classificadores DMC e LMQ não se mostraram eficientes, se comparados aos demais, assim como o algoritmo do QC simples, pois apresentaram taxas de

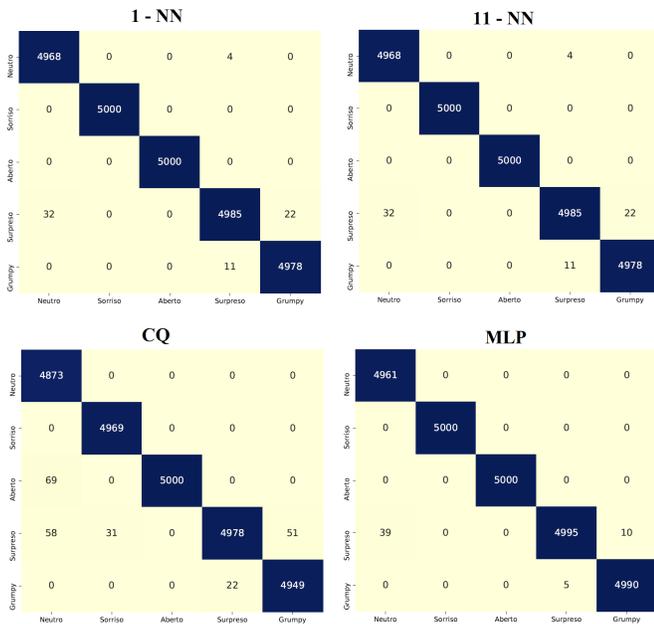


Figura 9. Matrizes de Confusão para uma das 50 execuções

classificação inferiores a 90%, enquanto os demais superaram 95%. Os classificadores que apresentaram os melhores resultados foram o QC, com a regularização de Friedman, o kNN e o MLP, que apresentaram resultados próximos uns aos outros, com aproximadamente 98% de acerto médio, e com baixíssimo desvio padrão.

Para uma avaliação justa na escolha do classificador a ser utilizado, deve-se levar em conta a aplicação dos mesmos, sua suscetibilidade a erros e a capacidade de processamento. Neste sentido, o algoritmo do QC possui o menor custo computacional, porém com menor taxa de acerto, enquanto o MLP possui maiores taxas de acerto, mas demanda mais tempo na etapa de treinamento. Para dados que apresentem maior separabilidade entre as classes, espera-se que os algoritmos baseados em distância euclidiana apresentem taxas de acerto cada vez mais altas e mais próximas daquelas atingidas pelo classificador baseado em redes neurais, tornando uma solução mais interessante, enquanto para classes mais sobrepostas, a diferença entre as taxas de acerto tendem a aumentar em favor do MLP.

Analisando o desempenho dos classificadores escolhidos em relação a outros trabalhos publicados, como o kNN utilizado em [10], que obtém uma taxa média de acerto próxima a 98%, ou quando comparado com [6] que utilizou o algoritmo de SVM para a classificação e obteve taxas de acerto de 99,9%, mesmo que não seja possível uma comparação direta devido a diferença metodológicas os resultados indicam que o presente trabalho é promissor. Isto, por fim, motiva a realização de trabalhos futuros voltados a adquirir amostras pertencentes a diferentes indivíduos, assim contribuindo para uma base balanceada e com boa representatividade das classes,

permitindo, portanto, avaliar a generalização dos resultados e a eficiência dos classificadores com mais assertividade.

Por fim, os autores do presente trabalho disponibilizam os conjuntos de dados gerados, bem como os algoritmos implementados para a etapa de classificação, através do endereço https://github.com/raulfontenele/Database_Facial_Expression.

VI. AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Os autores agradecem também à FUNCAP (processo no. BMD-0008-01413.01.17/20) e ao CNPq (processo no. 309379/2019-9) pelo apoio financeiro para realização desta pesquisa.

REFERÊNCIAS

- [1] Schuller, B.; Wimmer, D.; Arsic, D.; Moosmayr, T.; Rigoll, G. Detection of security related affect and behaviour in passenger transport, in: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, ISCA-INT SPEECH COMMUNICATION ASSOC, Brisbane, AUSTRALIA, 2008, pp. 265–268.
- [2] Luneski, A.; Konstantinidis, E.; Bamidis, P. D. Affective medicine a review of affective computing efforts in medical informatics. *Methods of Information in Medicine* 49 (3) (2010) 207–218. doi:10.3414/me0617.
- [3] Ekman, P. Universals and cultural differences in facial expressions of emotions. W: J. Cole (red.). In: *Nebraska symposium on motivation*. 1971. p. 207-283.
- [4] Webster, J. G. (Ed.). *Medical instrumentation: application and design*. John Wiley & Sons, 2009.
- [5] Marchetti, P. H.; DUARTE, Marcos. *Instrumentação em eletromiografia*. Laboratório de Biofísica, Escola de Educação Física e Esporte. São Paulo: Universidade de São Paulo, 2006.
- [6] Kelati, A.; Plosila, J.; Tenhunen, H. Machine Learning for sEMG Facial Feature Characterization. In: *2019 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. IEEE, 2019. p. 169-174.
- [7] Yang, G. et al. IoT-based remote pain monitoring system: From device to cloud platform. *IEEE journal of biomedical and health informatics*, v. 22, n. 6, p. 1711-1719, 2017.
- [8] Thulkar, D.; Bhaskarwar, T.; Hamde, S. T. Facial electromyography for characterization of emotions using LabVIEW. In: *2015 International Conference on Industrial Instrumentation and Control (ICIC)*. IEEE, 2015. p. 683-686.
- [9] Perdiz, J.; Pires, G.; Nunes, U. J. Emotional state detection based on EMG and EOG biosignals: A short survey. In: *2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)*. IEEE, 2017. p. 1-4.
- [10] Hamed, M. et al. Robust facial expression recognition for MuCI: a comprehensive neuromuscular signal analysis. *IEEE Transactions on Affective Computing*, v. 9, n. 1, p. 102-115, 2016.
- [11] Perusquía-Hernández, M.; Hirokawa, M.; Suzuki, K. Spontaneous and posed smile recognition based on spatial and temporal patterns of facial EMG. In: *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2017. p. 537-541.
- [12] James, G. et al. *An introduction to statistical learning*. New York: springer, 2013.
- [13] Andrade Filho, L. S. de. *Projeto de classificadores de padrões baseados em protótipos usando evolução diferencial*. 2014.
- [14] Bishop, C. M. *Pattern recognition and machine learning*. springer, 2006.
- [15] Webb, A. R. *Statistical pattern recognition*. John Wiley & Sons, 2003.
- [16] Silva, IN da; Spatti, D. H.; Flauzino, R. A. *Redes neurais artificiais para engenharia e ciências aplicadas*. São Paulo: Artliber, v. 23, n. 5, p. 33-111, 2010.
- [17] Ekman, P.; Hager, J. C.; Friesen, W. V. *Facial action coding system (FACS). A human face*, 2002.
- [18] Kaminski, B. E. EMG circuit. Google Patents. US Patent 10,010,259. July 3 2018.