

UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA
GRADUAÇÃO EM ENGENHARIA MECATRÔNICA
TRABALHO DE CONCLUSÃO DE CURSO

Guilherme Rodrigues Monteiro
Vinícius Del Manto Cavalcante

IA para medição de escala de esforço em fisioterapia

São Paulo
2022

GUILHERME RODRIGUES MONTEIRO
VINÍCIUS DEL MANTO CAVALCANTE

IA para medição de escala de esforço em fisioterapia

— Versão Original —

Monografia apresentada à Graduação em Engenharia Mecatrônica da Escola Politécnica da Universidade de São Paulo como parte dos requisitos para conclusão do curso.

Orientador: Profa. Dra. Larissa Driemeier ,
Prof. Dr. Rafael Traldi Moura

São Paulo
2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo-na-publicação

Monteiro, Guilherme Rodrigues; Cavalcante, Vinícius Del Manto
IA para medição de escala de esforço em fisioterapia/ G.R.Monteiro,
V.D.M.Cavalcante – São Paulo, 2022.
150p.

Monografia (Trabalho de Conclusão do Curso Graduação em Engenharia Mecatrônica) – Escola Politécnica da Universidade de São Paulo.
PMR – Departamento de Engenharia Mecatrônica.

1. Inteligência Artificial 2. Machine Learning 3. Rede Neural 4. Classificador SVM 5. Nível de esforço 6. Expressão facial.
I. Universidade de São Paulo. Escola Politécnica. PMR – Departamento de Engenharia Mecatrônica. II.t.

Agradecimentos

Agradecimento especial ao Professor Doutor Marco C. Uchida, por todo o auxílio para que o banco de dados ficasse o mais completo possível.

But it ain't about how hard ya hit. It's about
how hard you can get hit and keep moving
forward. How much you can take
and keep moving forward.

— *Rocky Balboa*

People's Dreams... Have No End!!

— *Marshall D. Teach*

Sumário

Sumário • *iv*

Resumo • *vi*

Abstract • *vii*

Lista de Figuras • *viii*

Lista de Tabelas • *x*

1 Introdução • *1*

1.1 Problema • *1*

1.2 Objetivo geral • *3*

1.3 Revisão da literatura • *3*

1.4 Hipótese • *5*

1.5 Requisitos • *6*

2 Materiais e Métodos • *7*

2.1 Elaboração da Base de Dados • *8*

2.1.1 Base de dados previamente existente • *8*

2.1.2 Coleta de dados • *9*

2.1.3 Pré-processamento dos Dados • *12*

2.2 Elaboração do classificador SVM com PCA • *13*

2.2.1 PCA • *13*

2.2.2 Classificador SVM • *18*

2.3 Elaboração da Rede Neural Convolucional • *22*

2.4 Validação dos Resultados • *27*

3 Resultados e Discussões • *29*

3.1 Resultados • *29*

3.1.1 Banco de dados • *29*

3.1.2 Classificador SVM com PCA • *29*

3.2 Discussões • *43*

4 Conclusão • *46*

4.1 Conclusão • *46*

4.2 Próximos passos • *47*

Referências • 48

Apêndices • 50

Resumo

MONTEIRO, G.R.; CAVALCANTE, V.D.M. *IA para medição de escala de esforço em fisioterapia*. 2022. Monografia (Trabalho de Conclusão do Curso Graduação em Engenharia Mecatrônica) – Escola Politécnica da Universidade de São Paulo. PMR – Departamento de Engenharia Mecatrônica. Universidade de São Paulo, São Paulo, 2022.

O treino de força se faz necessário em diversos contextos da área da saúde. Para atingir os resultados desejados, o profissional deve fazer seu paciente se exercitar em um nível de esforço específico. Atualmente, a medição do esforço do paciente se dá por meio de escalas de dor subjetivas e perguntadas entre séries de exercício.

Para auxiliar esse processo, o presente projeto tem o objetivo classificar de forma automática o nível de esforço de um indivíduo dada a sua expressão facial, sem depender da subjetividade do paciente e que possa ser feita de forma contínua durante a prática do exercício. Para isso, criou-se uma base de dados composta por 1499 imagens de rostos rotulados de forma subjetiva, que serviram como base de treino, validação e teste no desenvolvimento de um classificador SVM de Kernel Gaussiano com o auxílio da decomposição das faces por PCA e uma Rede Neural Convolucional.

O classificador SVM Gaussiano obteve uma pontuação F1 ponderada de 91% para pessoas presentes no treinamento e de 73% para pessoas não presentes no treinamento. Já a rede neural teve resultados mais robustos, e se mostrou menos suscetível a *overfitting*. Foi obtida uma exatidão das previsões de 96% para pessoas presentes no treinamento e de 94% para pessoas não presentes no treinamento. Esses resultados demonstram que é possível usar Inteligência Artificial para realizar a classificação desejada e colocam a rede neural como o método mais promissor para tanto.

Palavras-chave: Inteligência Artificial. Aprendizado de máquina. Rede Neural. Classificador SVM. Nível de Esforço. Expressão facial.

Abstract

MONTEIRO, G.R.; CAVALCANTE, V.D.M. *AI for effort rating in fisioterapia*. 2022. Monografia (Trabalho de Conclusão do Curso Graduação em Engenharia Mecatrônica) – Escola Politécnica da Universidade de São Paulo. PMR – Department of Mechatronics Engineering. University of São Paulo, São Paulo, Brazil. 2022.

Strength training is necessary in a variety of healthcare situations. In order to achieve his goals, the professional must make his patient exercise on a specific effort level. Currently, the measurement of effort is made by subjective pain scales and asked between exercise series.

To help this process, this project aims to automatically classify the effort level of a person given its facial expression, without depending on the patient subjectiveness and that can be done continuously during the exercise. To do so, it was created a data base of 1499 face images subjectively classified, that worked as a training, validation and test on the development of a SVM classifier with Gaussian kernel with aid of face decomposition using PCA and a Convolutional Neural Network.

The Gaussian SVM classifier obtained a 91% of weighted average f1-score for people within the classifier training and a 73% for people outside the classifier training. As for the CNN, it showed greater results and presented itself as a more resilient method against overfitting. It achieved a accuracy of 96% for people within its training dataset and 94% for people outside its training dataset. This results supports that it is possible to use AI to make the desired classification and present the CNN as the most promising method to do so.

Keywords: Artificial Intelligence. Machine Learning. Neural Network. SVM Classifier. Effort level. Facial expression.

Lista de Figuras

1.1	Exemplo de escala de dor.	2
1.2	Modelo de controle dos exercícios atualmente.	3
1.3	Modelo de controle dos exercícios desejado.	3
2.1	Diagrama generalizado da metodologia.	8
2.2	Esquema da realização do exercício.	10
2.3	Exemplo de <i>frame</i> da gravação inicial do participante em repouso.	11
2.4	Exemplo de <i>frame</i> da gravação durante o momento de esforço máximo.	11
2.5	Diagrama de pré processamento de dados.	12
2.6	Diagrama da estrutura de salvamento do banco de dados.	12
2.7	Exemplos das primeiras componenetes principais, as <i>eigenfaces</i>	15
2.8	Curva de porcentagem da variância explicada versus número de componentes principais utilizadas de imagens de 2914 dimensões.	16
2.9	Comparação imagem original e recuperada por 200 componentes principais de 2914 dimensões.	17
2.10	Curva de porcentagem da variância explicada versus número de componentes principais utilizadas de imagens de 11750 dimensões.	18
2.11	Exemplo de hiperplano separando regiões do espaço para duas dimensões.	19
2.12	Exemplo de hiperplanos na mesma situação feitos por kernels diferentes.	20
2.13	Arquitetura de uma Rede Neural Artificial simples.	23
2.14	Processo de convolução exemplificado.	23
2.15	Detecção de bordas verticais.	24
2.16	Exemplo de operação de Max Pooling.	24
2.17	Arquitetura da VGG16	25
2.18	Arquitetura da VGG16	26
2.19	Diagrama de validação dos resultados.	27
3.1	Matriz de confusão em validação do SVM com um terço da base e fotos de 2914 dimensões.	31
3.2	Matriz de confusão em validação do SVM com um terço da base e fotos de 2914 dimensões.	32
3.3	Matriz de confusão em validação do SVM com base completa e fotos de 2914 dimensões.	33
3.4	Matriz de confusão em validação do SVM com base completa e fotos de 2914 dimensões.	34
3.5	Matriz de confusão em validação do SVM com base completa e fotos de 11750 dimensões.	35

3.6	Matriz de confusão em validação do SVM com base completa e fotos de 11750 dimensões.	36
3.7	Valor da Função Custo durante o treinamento das camadas densas da rede. .	37
3.8	Valor da Exatidão durante o treinamento das camadas densas da rede.	38
3.9	Previsões realizadas para o grupo de teste após o treinamento das camadas densas da rede.	39
3.10	Matriz de confusão em validação da CNN, após o treinamento das camadas densas.	39
3.11	Valor da Função Custo após a sintonia fina da rede.	41
3.12	Valor da Exatidão após a sintonia fina da rede.	41
3.13	Matriz de confusão em validação da CNN, após a sintonia fina das camadas convolucionais.	42
3.14	Previsões realizadas para o grupo de teste após a sintonia fina das camadas convolucionais.	42

Lista de Tabelas

3.1	Tabela reporte de classificação em validação do SVM com um terço da base e fotos de 2914 dimensões.	30
3.2	Tabela reporte de classificação em teste do SVM com um terço da base e fotos de 2914 dimensões.	31
3.3	Tabela reporte de classificação em validação do SVM com base completa e fotos de 2914 dimensões.	32
3.4	Tabela reporte de classificação em teste do SVM com base completa e fotos de 2914 dimensões.	33
3.5	Tabela reporte de classificação em validação do SVM com base completa e fotos de 11750 dimensões.	35
3.6	Tabela reporte de classificação em teste do SVM com base completa e fotos de 11750 dimensões.	36
3.7	Valor da Função Custo e Acurácia da rede após o treinamento das camadas densas.	38
3.8	Tabela de reporte de classificação em teste da CNN, após o treinamento das camadas densas.	40
3.9	Valor da Função Custo e Acurácia da rede após a sintonia fina das camadas convolucionais.	43
3.10	Tabela de reporte de classificação em teste da CNN, após sintonia fina das camadas convolucionais.	43

Introdução

1.1 Problema

É comum a prática de exercícios guiados para atingir o limite motor do ser humano. Dentre as principais motivações constam: treinos de resistência física, desenvolvimento muscular e reabilitação muscular. O último é especialmente utilizado por fisioterapeutas e terapeutas ocupacionais para a recuperação de força e movimento, além do alívio da dor sentida por um paciente durante o pós-operatório.

Em sessões de fisioterapia e terapia ocupacional é comum o controle do esforço do paciente a partir de seu feedback a respeito do nível de esforço empreendido. A partir dela, o profissional responsável pelo tratamento pode aumentar, diminuir ou manter a dificuldade do exercício, garantindo que este tenha a maior eficiência possível na recuperação do paciente ao mantê-lo em seu limiar de esforço físico.

Por se tratar de uma situação delicada e crucial, é preciso um cuidado constante para garantir que o paciente não se encontre em uma situação pendendo para um sobrecarregamento e nem para um relaxamento. Com o feedback do esforço do paciente, os profissionais da saúde têm como controlar uma maior ou menor carga de exercícios para manter o esforço no limite desejado, o que torna o feedback uma das partes mais cruciais para o controle deste sistema humano.

Atualmente, o feedback é realizado através da ação humana por métodos como observação médica e perguntas diretas, inclusive com o auxílio de imagens de escalas de dor, como mostrada na figura 1.1, assim o paciente é capaz de indicar a intensidade de seu esforço. Tais métodos de medição do feedback são realizados em intervalos de tempos, em especial entre as sequências de exercícios para que estas não sejam interrompidas pelas perguntas de medição. O intervalo entre as medições caracteriza um tempo discreto, isto é, só é possível medir a resposta do paciente entre intervalos e não de modo contínuo.

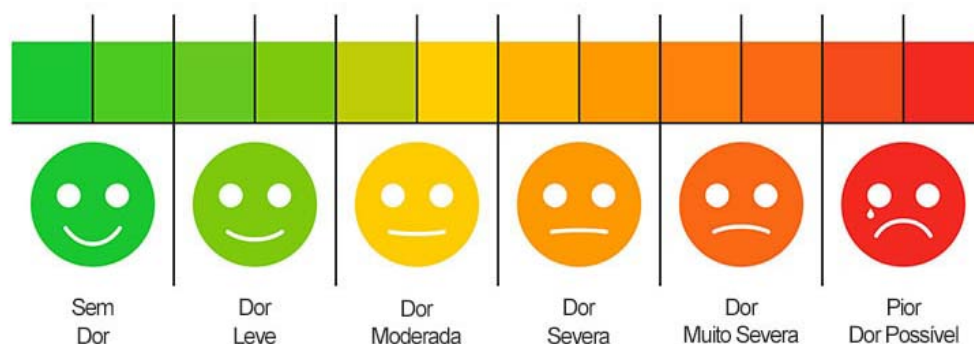


Figura 1.1: Exemplo de escala de dor.

FONTE: Extraído do site da [CAFISIO](#) - Centro de Atendimento Fisioterapêutico

Como a medição atual é feita através de ação humana de forma discreta, ela é sujeita a limitações que reduzem o desempenho do exercício e a eficiência das sessões de exercícios. É um processo suscetível a falhas de comunicação: pacientes podem ter dificuldade em enquadrar seu esforço na escala apresentada, ou omitir sua real sensação sobre o esforço auferido por não quererem aumentar a carga de exercício.

Também é um processo que tira tempo da sessão para realizar medições, o que por si só já é uma desvantagem, porém ainda reduz a escalabilidade do processo. Isto significa que em sessões fisioterápicas em grupo, os pacientes não conseguem ser monitorados individualmente sem atrapalhar o andamento da sessão, fazendo com que seu desempenho seja abaixo do que um processo monitorado automaticamente apresentaria.

Por fim, há também a questão do monitoramento discreto, que apresenta um desempenho de controle inferior ao apresentado em regimes contínuos. Na prática, quer dizer que se o paciente estiver tendo dificuldades com os exercícios ele deve esperar uma próxima medição para que sua demanda seja atendida, enquanto que em uma medição contínua o instrutor estaria constantemente monitorando a adequação da carga do exercício ao esforço realizado.

O modelo de controle do esforço realizado no exercício como é feito atualmente pode ser representado como mostrado na figura 1.2. Verifica-se, pela figura, que existe a atuação humana (HITL, do inglês, human-in-the-loop) no processo em pelo menos dois instantes que não necessariamente precisariam ter a atuação humana, evidenciados em vermelho.

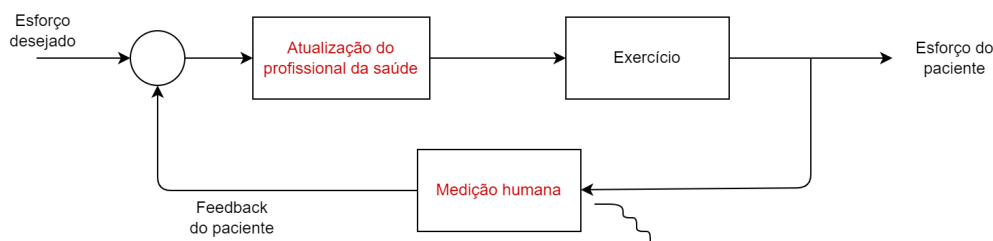


Figura 1.2: Modelo de controle dos exercícios atualmente.

FONTE: Autoria própria.

1.2 Objetivo geral

O presente trabalho tem por objetivo o desenvolvimento de um classificador do nível de esforço humano em três categorias: **baixo esforço**, **médio esforço** ou **alto esforço**. A classificação deve ser feita a partir da expressão facial do indivíduo, a fim de melhorar o processo de feedback nos exercícios controlados em sessões de fisioterapia e terapia ocupacional. O modelo de controle apresentado na figura 1.2 passa a ser melhor descrito pela figura 1.3.

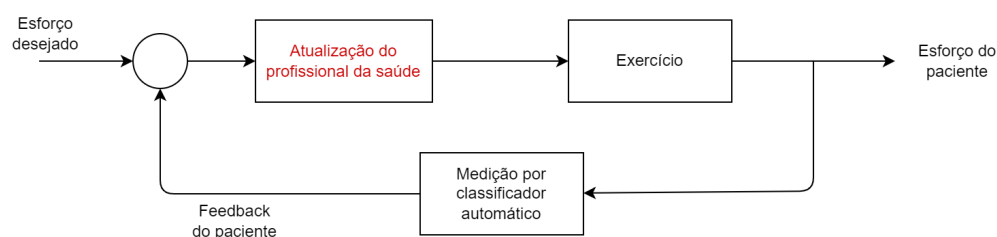


Figura 1.3: Modelo de controle dos exercícios desejado.

FONTE: Autoria própria.

O sistema proposto permite a futura implementação de exercícios físicos com um ajuste automático de carga a partir do sinal digital enviado pelo classificador implementado. Assim, abre portas para tornar o tratamento mais automatizado de forma a facilitar o trabalho do profissional da saúde e, mais importante, aumenta a eficiência do tratamento no paciente.

1.3 Revisão da literatura

As expressões faciais são uma das formas mais naturais e espontâneas de comunicação não-verbal de sensações – físicas ou emocionais. O esforço e sucesso na portabilidade dessa linguagem natural para sistemas computacionais, permitindo a classificação dessas

expressões através de métodos de Machine Learning (ML), cresceu consideravelmente na última década (Htay 2021).

O presente trabalho toma como base artigos recentes e relevantes que abordam um entre dois temas: aplicação de técnicas de ML na identificação de expressões faciais, não necessariamente indicativas de fadiga muscular, e técnicas tradicionais na identificação de expressões faciais direcionadas a classificação de dor ou desconforto em atividades físicas.

Do primeiro, os trabalhos realizados em anos consecutivos com a finalidade de classificar a intensidade de exercício físico através do reconhecimento facial, (Khanal, Barroso et al. 2018) e (Khanal, Sampaio et al. 2019), estão fortemente alinhados com os objetivos propostos e servem como referência para a elaboração da metodologia e, posteriormente, para a avaliação dos resultados obtidos.

No trabalho de Khanal, Barroso et al. 2018 o enfoque é no método de análise computacional de expressão facial. Para tanto, os pesquisadores usaram o *Luxand Code*, uma aplicação que detecta 70 pontos notáveis na face humana. Dos 70 pontos, foram selecionados 26, e em cima desses foram utilizados algoritmos de Inteligência Artificial (IA) para análise. Os algoritmos de IA usados foram *K-Nearest Neighbor* (KNN), *Discriminant Analysis* e *Support Vector Machine* (SVM). Em suma, são três algoritmos com abordagens diferentes que têm por objetivo fazer uma análise da expressão facial através da disposição dos 26 pontos notáveis selecionados na face, e assim classificar a intensidade do exercício físico. A menor acurácia apresentada do algoritmo SVM foi de 92,8%.

Ainda no enfoque de análise computacional, o trabalho de (Roy et al. 2016) traz grandes contribuições à discussão ao utilizar um método simular, porém com a ajuda de outro método para complementar. No estudo, também na área de detecção de nível de dor por expressão facial, utilizaram algoritmos de *Support Vector Machine* (SVM) e *Principal Component Analysis* (PCA). Este último método permite extrair as componentes principais de um dado espaço vetorial. No caso, o espaço vetorial é formado pelas imagens das faces, cada uma sendo um vetor de dimensão igual ao número de pixels; e os componentes principais, chamados de *eigenfaces* (analogia ao nome *eigenvectors* da teoria de álgebra linear por trás do método computacional), são autovetores da matriz de covariância das imagens de faces. A acurácia final para detecção de nível de dor com base em imagens foi de 95,5%.

Indo por outra abordagem, Khanal, Sampaio et al. 2019 apresenta forte argumentação em favor do uso de Redes Neurais Convolucionais (*Convolutional Neural Networks/CNN*) para o problema abordado. Além disso, (Li e Lam 2015) também defende o uso de classificadores SVM e Redes Neurais na identificação de expressões faciais, ainda que não exatamente indicativas de esforço físico, e a extração de componentes principais

para se atingir melhores resultados; (Kawashima, Nomiya e Hochin 2021) e (Lopes et al. 2017) lidam com e apresentam possíveis soluções para o problema de uma quantidade insatisfatória de dados, um exemplo sendo o uso de ajuste fino de camadas convolucionais.

Todavia, como a futura implementação do projeto em um contexto médico-hospitalar é um ponto chave da proposta, a abordagem adotada pelos artigos supracitados para a elaboração da base de dados pode não ser a mais apropriada ou aceita pela comunidade fisioterápica. Portanto, houve o cuidado em usar também como base uma referência mais ligada a área.

O trabalho de (Uchida et al. 2018) fornece uma metodologia consonante com as exigências no ramo da educação física e acompanhamento clínico de reabilitação. Ainda que não tenha envolvido diretamente o uso de Inteligência Artificial (IA), a metodologia para a identificação de fadiga utilizada como referência é robusta: é realizada a eletromiografia de superfície (EMG) em grupos musculares selecionados no intuito de medir a correlação entre o nível de atividade de tais músculos e a intensidade avaliada na unidade de ação AU12, de acordo com o sistema FACS (*Facial Acting Coding System*) proposto em (Ekman e Friesen 1978). Esse sistema é amplamente aceito como um dos mais coerentes e acurados para a identificação de expressões faciais e, ao demonstrar a correlação positiva entre essas duas grandezas (atuação dos grupos musculares medida através da EMG e a avaliação do grupo Zigomático Maior).

1.4 Hipótese

A principal hipótese do trabalho é que a expressão facial pode ser um indicador de intensidade de esforço físico possível de ser computado e de ser usada para monitoramento de atividade física. Isso implica que as expressões faciais seriam comportamentos padrão em seres humanos e com reprodutibilidade intra e inter pessoal.

Essa hipótese é o pilar de sustentação do trabalho, sem o qual os resultados não podem ser atingidos. É preciso que exista essa correlação direta entre expressão facial e nível de esforço físico e que ela possa ser generalizada para seres humanos. Deve ser padronizada para a mesma pessoa e para pessoas diferentes. Sem esse padrão, o trabalho de treinar IAs para serem as classificadoras propostas seria extremamente difícil, se não impossível.

Por mais que a correlação entre expressão facial e esforço físico pareça óbvia, ela não deve ser tomada como garantida. Felizmente, pesquisas sustentam a hipótese. Em (Uchida et al. 2018) é concluído que a expressão facial pode estar diretamente correlacionada com diferentes intensidades de exercícios de resistência e fadiga; e que, independentemente da intensidade, a fadiga leva a um aumento de indicadores de expressões faciais. Já em (Almeida 2015) é dito que, apesar de serem necessários mais estudos para validar

a expressão facial como um indicador de fadiga muscular, os resultados da pesquisa constataam que há uma relação direta e forte entre a cadência de movimento do exercício e as unidades de ação da face humana.

Assim, há embasamento teórico na literatura para que se possa adotar a hipótese proposta. Pode-se seguir adiante com a intenção de que a hipótese se sustente e assim possam ser atingidos os resultados desejados com o trabalho.

1.5 Requisitos

Espera-se que a solução elaborada constitua um classificador automático capaz de discernir três níveis de esforços distintos: alto, médio e baixo. Por se tratar de um problema com pouca quantidade de dados, espera-se que o treinamento do classificador seja razoavelmente rápido (tomando não mais do que uma hora) para que não se torne um gargalo futuramente quando houver maior disponibilidade de dados. A acurácia do classificador deve ser maior ou igual a 85%, tomando como base resultados na literatura. A classificação também deve ser rápida, sem ultrapassar um segundo, para que possa ser usado em tempo real e de forma contínua durante os exercícios caso seja implementado no acompanhamento de algum no futuro.

Materiais e Métodos

Definido o Estado da Arte da identificação de expressões faciais, com enfoque em reconhecimento e escalas de dor, é possível avaliar com mais precisão quais tecnologias e métodos estão disponíveis para que o projeto atinja seus objetivos e resultados. Alinhando os objetivos com uma análise crítica dos pontos fortes e fracos de cada linha de ação para solucionar o problema, será definida a metodologia adotada para cumprir os requisitos.

Para uma primeira triagem sobre a solução a ser utilizada, considera-se o objetivo de que a resposta seja rápida o suficiente para que humanos a percebam como contínua. Assim, é preciso abandonar os métodos convencionais de escalas analógicas, que envolvem pausas entre medições e dependem de comunicação entre humanos, o que demanda tempo. Portanto, o projeto deverá realizar a escalação da dor das faces analisadas através de métodos computacionais, que são mais rápidos e alinhados com tal objetivo.

Foram abordados dois métodos computacionais atuais e eficazes para se atingir este objetivo. O primeiro é baseado em análise computacional de expressão facial, com ênfase em algoritmo SVM as faces decompostas por PCA, e o segundo em Redes Neurais Convolucionais (CNN). A base para comparação dos métodos será os trabalhos mais recentes e relevantes a respeito de cada um, ambos da mesma equipe: (Khanal, Barroso et al. 2018) - que utiliza K-Nearest Neighbors(KNN) e SVM para classificação facial - e (Khanal, Sampaio et al. 2019) - que aborda o mesmo problema, utilizando redes neurais. Por mais que o primeiro método tenha apresentado resultados mais acurados para duas e três divisões de escala, os autores concluíram que há perda significativa de acurácia com o aumento do número de divisões, chegando inclusive a mencionar para trabalhos futuros um aumento no número de parâmetros considerados na análise. Já para o uso de CNN, (Khanal, Sampaio et al. 2019) apresentam resultados que indicam uma curva menos aguda para a perda de acurácia com o aumento das divisões de escala.

O projeto tem como objetivo uma classificação em apenas 3 divisões, porém os métodos analógicos amplamente usados atualmente costumam ter 10 divisões. Pensando

em trabalhos futuros e em uma evolução da classificação para se aproximar mais da resolução analógica, é fortuito utilizar um método que possibilite este novo passo com menos prejuízo à acurácia do método.

Posto isso, o método que apresenta a maior receptividade para avanço futuro do projeto é o que se baseia em CNN, sendo a utilização de SVM com PCA também um método viável, com boa acurácia, velocidade de treinamento e menor complexidade. Portanto, ambos os métodos serão adotados no projeto e será feita uma comparação de resultados ao final. Com isso em mente, os requisitos para se atingir os objetivos gerais do projeto tornam-se palpáveis e factíveis.

As IAs precisam ser treinadas com um banco de imagens de faces já classificadas. Portanto, o primeiro passo para a implementação da solução é o desenvolvimento do banco de imagens. Em seguida, deve-se desenvolver as IAs em si, e treiná-las com a base de dados para que elas sejam capazes de classificar sozinhas o esforço analisando expressões faciais. Com as IAs treinadas, é preciso testar e comparar seus resultados, em especial suas acurácias.

A figura 2.1 mostra um passo a passo generalizado da metodologia: primeiramente, coleta-se gravações de voluntários realizando o exercício; depois, há a extração e pré-processamento de *frames* desses vídeos, que servirão como base de dados para o treinamento da IA; finalmente, valida-se os resultados analisando a acurácia obtida com a realização de predições para uma base de teste.

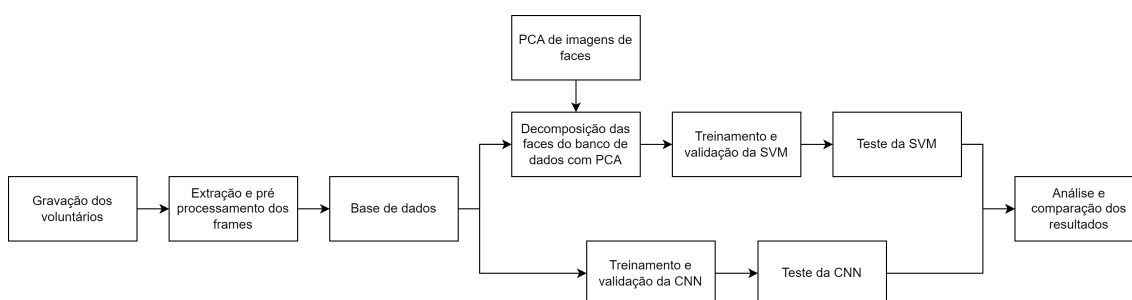


Figura 2.1: Diagrama generalizado da metodologia.

FONTE: Autoria própria.

2.1 Elaboração da Base de Dados

2.1.1 Base de dados previamente existente

A base para treinamento das IAs será composta parcialmente por voluntários recrutados pela dupla e majoritariamente pelos resultados do estudo (Uchida et al. 2018), em que sua equipe da Faculdade de Educação Física da UNICAMP fez experimentos análogos aos

descritos na próxima subseção (pois será replicado o método deles) a fim de se identificar fadiga muscular através de expressões faciais.

O estudo da equipe da UNICAMP respeita as normas éticas: todos os voluntários assinaram um TCLE à época do estudo e novamente outro atualizado para autorizar a utilização de sua imagem para o aprendizado das IAs a serem desenvolvidas. Tudo foi submetido para aprovação na [Plataforma Brasil](#).

2.1.2 Coleta de dados

Para o aumento da base de dados a ser usada no treinamento e avaliação das IAs foram coletadas amostras - vídeos de trinta segundos focalizados na face dos voluntários durante o treino de praticantes regulares de atividade física, habituados com treinamento de força. Cada voluntário informou e confirmou sua altura, peso e idade, bem como participou de uma breve entrevista para esclarecer quaisquer dúvidas a respeito do procedimento. Cada voluntário também leu e assinou um termo de consentimento de uso de imagem (direito de imagem) e um Termo de Consentimento Livre e Esclarecido (TCLE). O TCLE explicita os exercícios que os voluntários farão, seus objetivos e eventuais consequências, como fadiga muscular; e ao assiná-lo os voluntários declaram ciência do que leram e consentem o uso de suas filmagens para classificação de seu esforço e treinamento da IA.

Para a coleta de imagens é importante que o voluntário atinja a fadiga muscular e, portanto, foi desenhada uma série de repetições do exercício de rosca direta com halter, esquematizado na figura 2.2. Cada participante, inicialmente, foi gravado em repouso por 5 segundos - para que houvesse uma referência de rosto sem fadiga alguma no treinamento da rede - e, feito isso, realizou uma série de 12 repetições do exercício com uma carga de aproximadamente 50% da carga usual utilizada pelo voluntário ao realizar o exercício, em uma espécie de aquecimento. Após o aquecimento, foram feitas séries de 10 a 12, 8 a 10 e 3 a 5 repetições com cargas progressivamente maiores. Nos casos em que o voluntário não atingiu a fadiga muscular nas três primeiras séries de repetições, foram feitas séries de 1 a 3 repetições com cargas progressivas até que não fosse possível completar o movimento.

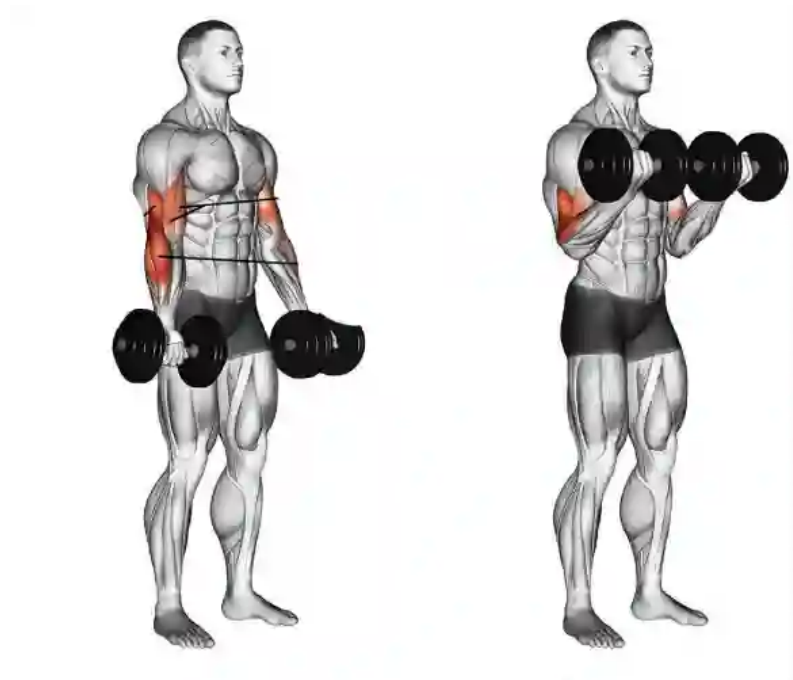


Figura 2.2: Esquema da realização do exercício.

FONTE: Extraído do site da [Base Fitness](#).

Todo o procedimento foi gravado com uma câmera digital montada sobre um tripé, posicionada a 1.30m de distância do voluntário e ajustada para a captura completa de seu rosto. As gravações eram interrompidas durante o período de descanso entre séries de repetições - que duravam dois minutos - e, a partir delas, extraiu-se os *frames* que foram pré-processados e utilizados no treinamento e validação da rede. As figuras 2.3 e 2.4 são exemplos de capturas retiradas de um dos voluntários durante o procedimento.



Figura 2.3: Exemplo de *frame* da gravação inicial do participante em repouso.

FONTE: Autoria própria. Voluntário: Autor Monteiro, G. R.



Figura 2.4: Exemplo de *frame* da gravação durante o momento de esforço máximo.

FONTE: Autoria própria. Voluntário: Autor Monteiro, G. R.

2.1.3 Pré-processamento dos Dados

O processo de transformação dos vídeos coletados em dados de entrada apropriados é ilustrado no fluxograma da figura 2.5. Ele consiste em, primeiramente, extrair um *frame* (foto) a cada segundo de vídeo; em seguida, é verificado se há uma face nesta imagem e, apenas em caso positivo, feito um recorte desta imagem de forma que ela contenha apenas o rosto do voluntário. Essa verificação e recorte inteligente são feitos usando as ferramentas da API (*Application Programming Interface*) *MediaPipe* de detecção de faces. Cada imagem passa por uma re-escala de forma a ter dimensões de 96x96 pixels e é salva em escala cinza como um arquivo de extensão .jpeg e nome padronizado. Extraídas as imagens de um vídeo, é feito o rotulamento de cada uma delas em um dos três níveis de esforço propostos, de acordo com a percepção subjetiva da dupla.

Para rotular uma imagem, basta movê-la para o diretório apropriado: se o voluntário fizer parte do grupo de treinamento, suas fotos pertencerão ao diretório-mãe "Treinamento"; caso contrário, suas fotos devem ser movidas para o diretório-mãe "Teste". A divisão entre grupo de Treinamento e grupo de Teste foi feita com uma proporção aproximada de 4:1. Dentro de ambos os diretórios há três pastas nomeadas "Alto", "Médio" e "Baixo" e cada imagem é salva necessariamente dentro de uma delas, de acordo com o nível de esforço percebido. A estrutura final de pastas do banco de dados onde as imagens serão salvas podem ser vistas na figura 2.6.

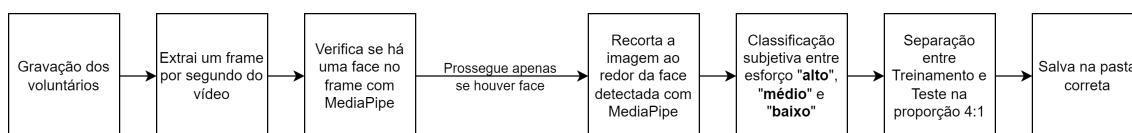


Figura 2.5: Diagrama de pré processamento de dados.

FONTE: Autoria própria.

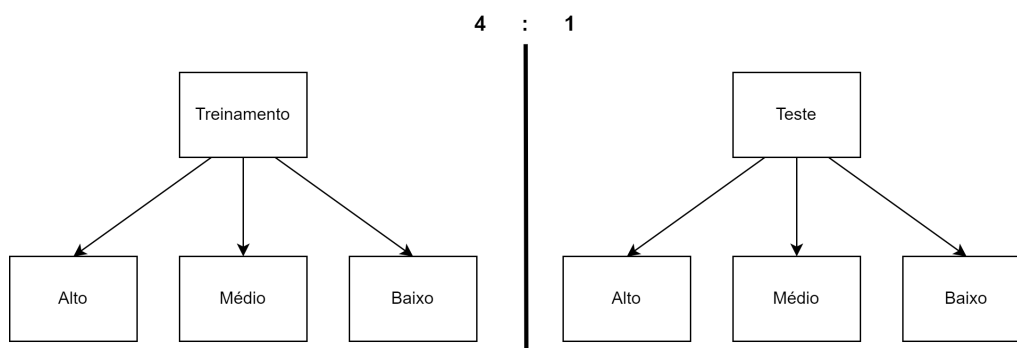


Figura 2.6: Diagrama da estrutura de salvamento do banco de dados.

FONTE: Autoria própria.

O banco de dados para treinamento é composto de 1242 imagens de faces com classificação de nível de esforço, sendo 270 de nível baixo, 698 de nível médio e 274 de nível alto. Já o banco de dados para testes é composto de 257 imagens, sendo 28 de nível baixo, 149 de nível médio e 80 de nível alto.

2.2 Elaboração do classificador SVM com PCA

O classificador SVM com PCA será implementado em linguagem Python, usando Google Colab, com as ferramentas fornecidas pela biblioteca *sklearn* para fazer o PCA e o treinamento do classificador SVM.

2.2.1 PCA

O PCA, ou *Principal Component Analysis*, é um método que retorna os autovetores da matriz de covariância de uma determinada matriz passada, chamados de Componentes Principais. Essas Componentes Principais descrevem um espaço vetorial e pode-se decompor os dados da matriz original entre os autovetores do PCA. A grande vantagem é que, por serem Componentes Principais, utilizando-se poucas delas pode-se explicar/reconstituir os dados originais com uma alta porcentagem de variância explicada. A variância explicada é a porcentagem da variância do dado original que pode ser reconstituída ao decompor o dado pelas Componentes Principais do PCA e depois realizar a transformada inversa de volta para as dimensões originais.

Ao realizar o PCA de uma matriz composta por imagens de faces - número de linhas igual ao número de imagens e número de colunas igual à dimensão de cada imagem (pixels de largura vezes pixels de altura) -, os autovetores (*eigenvectors*) passam a ser Componentes Principais para decompor e reconstruir uma face da matriz passada com alta variância explicada, e por isso são denominadas *eigenfaces*. Pode-se ver exemplos de *eigenfaces* na figura 2.7, por mais que sejam escuras e em tons de cinza, as sombras dão ênfase ao que é predominante na face para cada componente principal.

O PCA gera um número de Componentes Principais igual ao número de dimensões dos dados iniciais, porém, justifica-se o uso do PCA por poder realizar análises computacionais com menos dimensões perdendo pouca informação. Tomando como exemplo uma imagem pequena de face, de 62 por 47 pixels, as análises dela sozinha teria de levar em conta 2914 dimensões (pixels), porém com o PCA pode usar 200 dimensões (Componentes Principais) perdendo pouca variância explicada. É possível ver um exemplo de imagem decomposta por poucas Componentes Principais lado a lado com sua original na figura 2.9 e o quanto a variância explicada consegue expressar a face original. Esse uso de menos dimensões deixa as análises computacionais feitas a seguir com as faces mais rápidas,

utilizando menos poder computacional e com melhor desempenho, com baixa perda de precisão.

O desafio torna-se realizar um PCA que possa decompor, sumariamente, qualquer face humana com alta variância explicada pelas suas eigenfaces. Para tanto, é preciso que a matriz de faces passadas para o PCA contemplem toda a diversidade humana. A solução se apresenta com o banco de dados *Labeled Faces in the Wild* (LFW). É um banco de dados público de fotografias de faces designado para verificação facial (Vision 2018), feito com os mais variados fenótipos humanos, contendo imagens de pessoas de diversas etnias, idades e gêneros. Assim, seu uso possibilita resultados abrangentes e genéricos, que contemplam a grande diversidade de aparências humanas. A base LFW pode ser acessada em Python através da biblioteca *sklearn.datasets*. Será feita uma re-escala nas imagens do LFW para diminuir as dimensões das fotos, que originalmente são de 125 pixels de altura por 94 de largura.

Em um primeiro momento, o PCA foi feito com um *resizing* de 0.5 nas fotos presentes no LFW, dando medidas de 62 pixels de altura por 47 de largura, dando ao total 2914 dimensões para as imagens. O PCA pode ser feito em Python através da biblioteca *sklearn.decomposition*. O PCA feito com o LFW com esse *resize* gera *eigenfaces* que são ilustradas na figura 2.7.

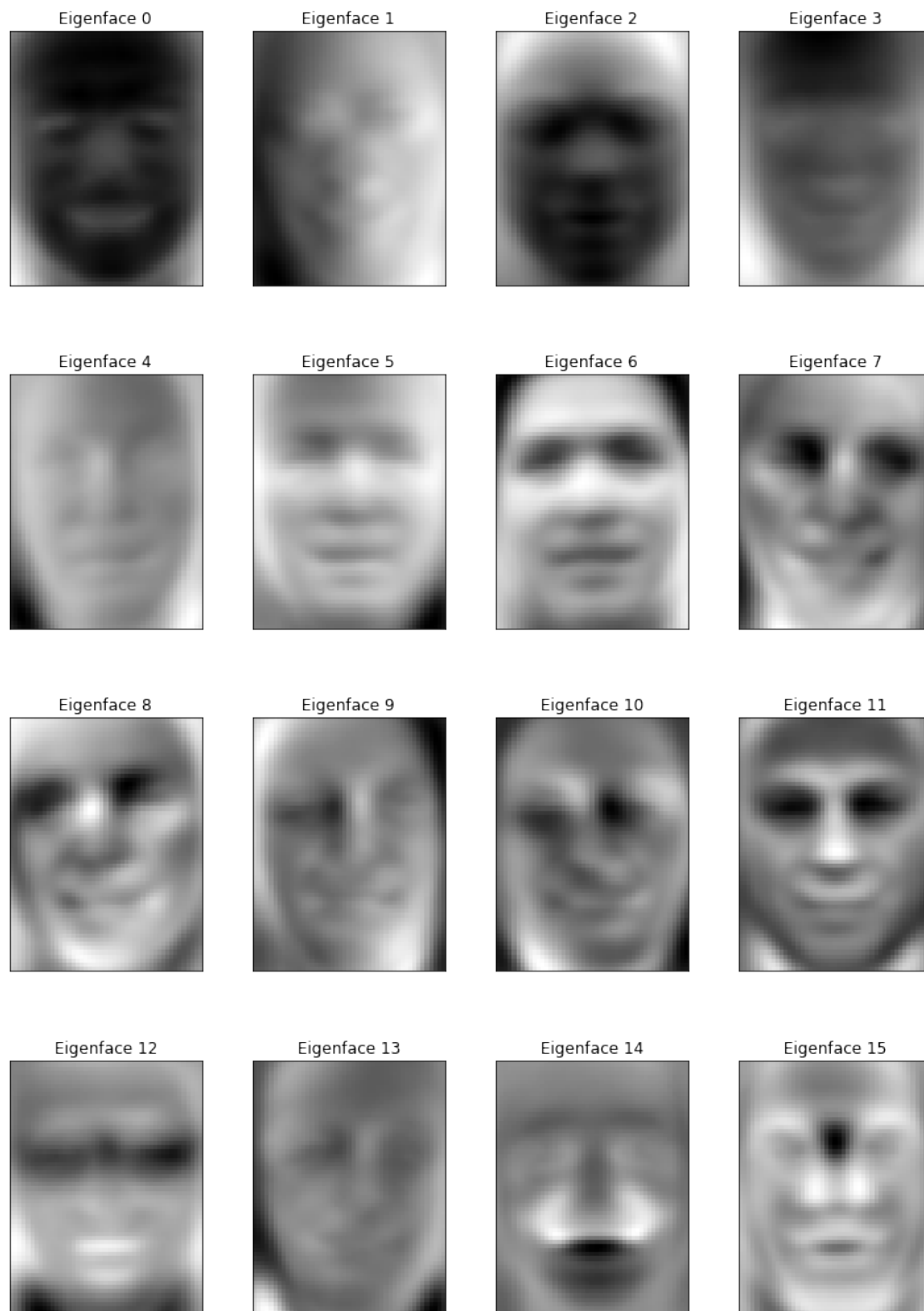


Figura 2.7: Exemplos das primeiras componentes principais, as *eigenfaces*.

FONTE: Autoria própria.

Com a finalidade de deixar o treinamento e a classificação do algoritmo SVM mais rápido, são usadas apenas algumas componentes principais. A figura 2.8 mostra a curva degrau da porcentagem da variância explicada versus o número de componentes utilizadas.

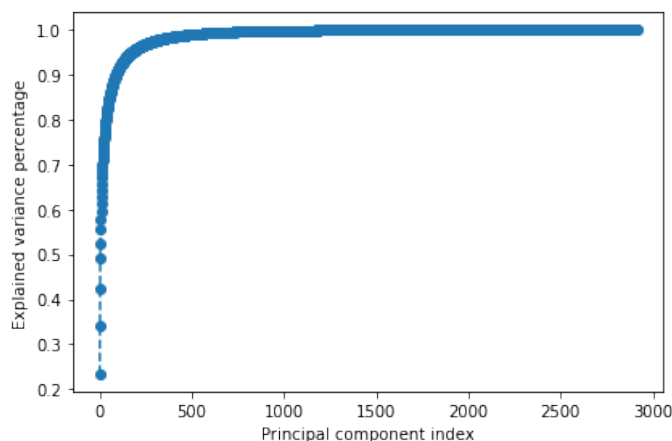


Figura 2.8: Curva de porcentagem da variância explicada versus número de componentes principais utilizadas de imagens de 2914 dimensões.

FONTE: Autoria própria.

Com apenas 200 componentes principais, a porcentagem de variância da face explicada pelo PCA é de 95,69%, sendo considerado o suficiente para expressar a face analisada pelo classificador. Como o gráfico ilustra, após o degrau é necessário usar muito mais componentes para aumentar apenas um pouco a porcentagem de variância explicada. Assim, considera-se 200 componentes um ponto ótimo de utilização.

Fazendo a decomposição da imagem 2.4 nas 200 componentes principais do PCA e depois a transformada inversa de volta para o formato de imagem padrão, pode-se comparar a versão original e a recuperada na figura 2.9.

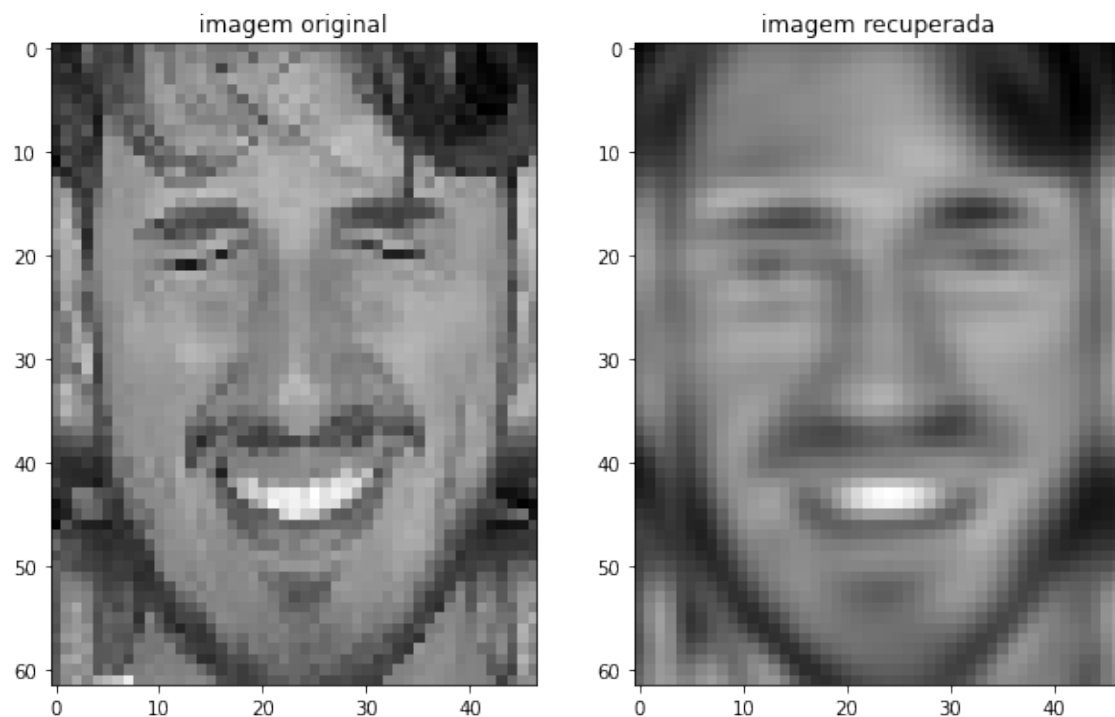


Figura 2.9: Comparação imagem original e recuperada por 200 componentes principais de 2914 dimensões.

FONTE: Autoria própria.

Como uma tentativa de aumentar a acurácia das previsões do classificador SVM, também foi feito o PCA das mesmas imagens da LFW sem o *resizing*, ou seja, usando um tamanho de imagem de 125 pixels de altura por 94 de largura, gerando uma imagem com 11750 dimensões. É mostrado abaixo o gráfico de variância explicada pelo número de componentes principais utilizada para as novas dimensões das fotos.

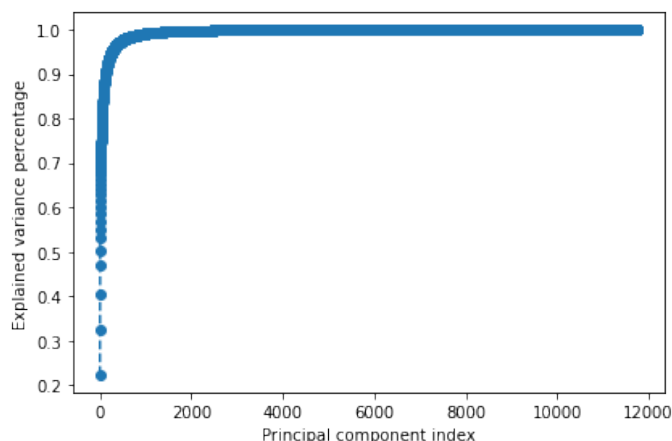


Figura 2.10: Curva de porcentagem da variância explicada versus número de componentes principais utilizadas de imagens de 11750 dimensões.

FONTE: Autoria própria.

Com 400 componentes principais, a porcentagem de variância da face explicada pelo PCA é de 96,87%, sendo considerado o suficiente para expressar a face analisada pelo classificador sem aumentar consideravelmente o esforço computacional. Além de ser o suficiente, vale ressaltar que essa segunda versão do PCA é feita com imagens de maior resolução e com uma variância explicada ainda maior ao decompor as faces (para o número de componentes determinado em cada caso).

2.2.2 Classificador SVM

Classificador SVM, ou *Support Vector Machine*, é um algoritmo de *machine learning* que procura dividir o espaço das variáveis explicadas em regiões separadas por hiperplanos (um plano em múltiplas dimensões). O SVM faz o cálculo dos hiperplanos para delimitar as regiões do espaço de forma que os pontos (dados) passados para treinamento sejam agrupados conforme sua classificação alvo. O conceito de um SVM calculando hiperplanos para separar pontos de mesma classificação pode ser visto simplificadaamente na figura 2.11, porém enquanto na figura a simplificação apresenta duas dimensões, na prática os dados podem ter múltiplas dimensões. No caso do problema abordado, a dimensão seria definida pelo número de pixels da imagem que se deseja classificar, porém graças à utilização do PCA a dimensão do problema é simplificada para o número de Componentes Principais adotadas ao realizar o PCA.

O cálculo dos hiperplanos para separar as classes que o SVM faz é feito otimizar as margens. Os pontos de treinamento mais próximos das margens são os exemplos mais difíceis de se classificar. Para sua otimização, o cálculo é feito de forma que estes influenciam a posição e a orientação do hiperplano. Por conta disso, os pontos próximos

das margens são chamados de vetores de suporte e são os responsáveis pelo nome da técnica. Na imagem 2.11, as margens podem ser vistas como as linhas pontilhadas, o hiperplano classificador é a linha contínua e os vetores de suporte são os pontos em destaque em cima da margem.

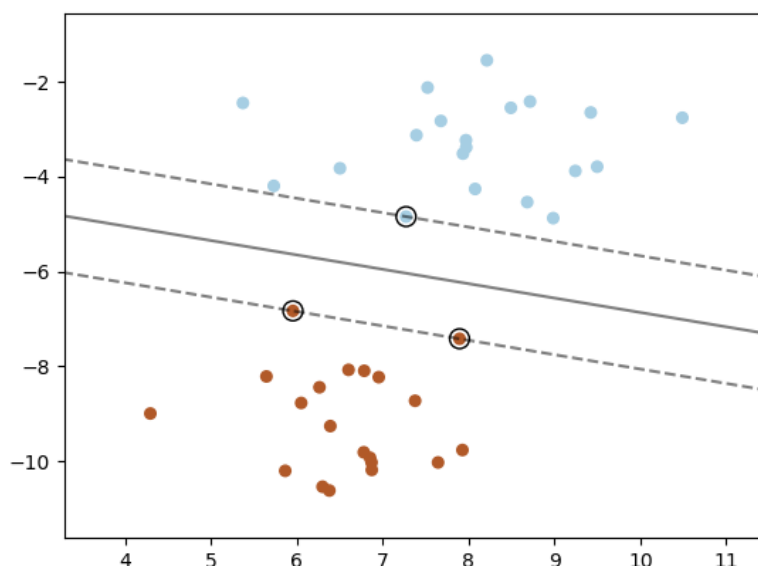


Figura 2.11: Exemplo de hiperplano separando regiões do espaço para duas dimensões.

FONTE: Extraído de [SciKit Learn](#), site oficial da biblioteca *sklearn* utilizada.

Quando o hiperplano é não linear, como acontece na maioria dos casos práticos, os classificadores SVM não podem ser usados. Então, utiliza-se uma mudança de dimensão dos dados, através de uma função denominada de kernel, onde o classificador torna-se linear. Existem várias opções padrões de kernel, além de ser possível utilizar uma nova específica. Pode-se ver alguns exemplos de hiperplanos feitos na mesma situação porém com kernels diferentes na figura 2.12.

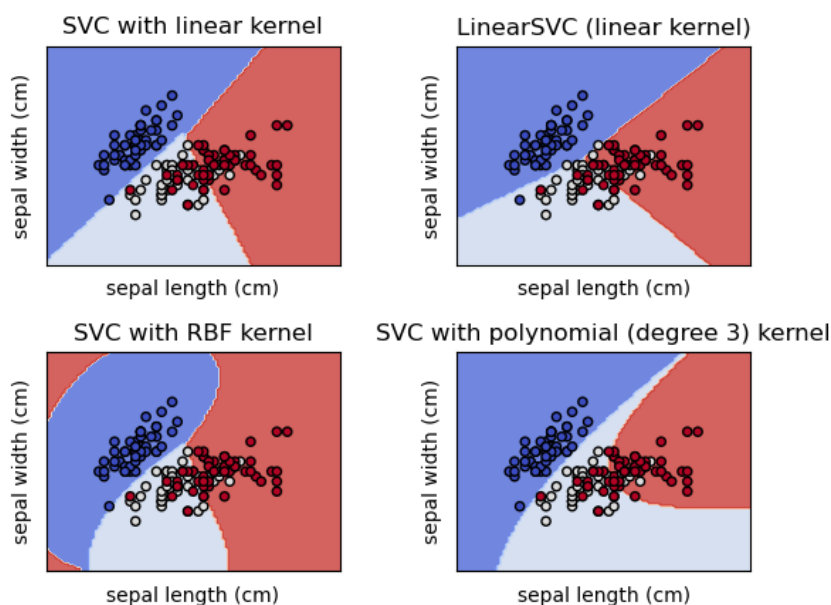


Figura 2.12: Exemplo de hiperplanos na mesma situação feitos por kernels diferentes.

FONTE: Extraído de [SciKit Learn](https://scikit-learn.org/), site oficial da biblioteca *sklearn* utilizada.

O classificador SVM para o problema proposto será feito com kernel Gaussiano, ou RBF (*Radial Basis Function*), um kernel padrão presente na biblioteca *sklearn.svm* definido pela equação 2.1. Será utilizado o kernel Gaussiano pois, como pode-se ver na figura 2.12, ele permite superfícies de hiperplanos mais complexas e com curvas, o que é uma vantagem ao fazer um classificador de faces humanas que possa englobar o máximo de diversidade possível.

$$\exp(-\gamma\|x - x'\|^2) \quad (2.1)$$

Os kernels possuem alguns hiperparâmetros que precisam ser declarados para o cálculo dos hiperplanos. No caso, será declarado o hiperparâmetro γ , específico do kernel Gaussiano escolhido, e define quanta influência um dado de teste tem individualmente. Quanto mais alto ele for, mais os dados devem estar próximos para que afetem o classificador (Learn 2022).

Será declarado também o parâmetro de regularização C , comum a todos os SVMs. Ele balanceia erros de classificação dos dados de treinamento contra a simplicidade do hiperplano; um baixo C torna a superfície do hiperplano mais suave enquanto um alto C mira em classificar todos os dados de teste corretamente; já o

Para escolher os melhores parâmetros, será feita de modo iterativo testando possíveis valores com a função *GridSearchCV* presente na biblioteca *sklearn.model_selection*, que retorna a melhor combinação dentre as possíveis com as opções passadas.

Em um primeiro momento, usando um terço do banco de dados e com as dimensões de imagem próprias do LFW com resizing de 0,5, ou seja, 62 por 47 pixels, o primeiro teste para a definição dos parâmetros foi com os valores:

$$'C' : [1e3, 2.5e3, 5e3, 1e4, 5e4, 1e5], \quad (2.2)$$

$$' \gamma ' : [0.0001, 0.00025, 0.0005, 0.001, 0.005, 0.01, 0.1] \quad (2.3)$$

O retorno foi de que o melhor par seria ($C=1e3$, $\gamma=0.005$). Por ter escolhido o C mais baixo, foi necessário testar mais valores a fim de se encontrar o ponto ótimo de verdade, e não apenas a beirada dos valores testados. Foi-se aumentado a lista de parâmetros passados para o teste até que chegou-se à seguinte configuração:

$$'C' : [1, 5, 10, 15, 25, 50, 0.1e3, 0.25e3, 0.5e3, 1e3, 2.5e3, 5e3, 1e4, 5e4, 1e5], \quad (2.4)$$

$$' \gamma ' : [0.0001, 0.00025, 0.005, 0.001, 0.005, 0.01, 0.1] \quad (2.5)$$

Conforme o limite inferior de C s possíveis foi diminuindo, o C escolhido também foi diminuindo e em compensação o γ foi aumentando. Ao fim, a dupla de parâmetros escolhidos foi ($C=10$, $\gamma=0.01$), e têm-se a certeza que são realmente os melhores, e não apenas a borda dos limites testados.

Em um segundo momento, com o banco de dados completo e ainda com as dimensões de imagem próprias do LFW com resizing de 0,5, ou seja, 62 por 47 pixels, o teste para a definição dos parâmetros seguiu com os mesmos intervalos definidos anteriormente pelas listas 2.4 e 2.5. O retorno foi de que o melhor par seria ($C=5$, $\gamma=0.005$). Por nenhum dos valores estarem no limite dos valores testados não foi preciso mudar as listas de possíveis parâmetros.

E por fim, em um terceiro e último momento, foi treinada a rede com o banco de dados completo e com as dimensões de imagem próprias do LFW sem resizing, ou seja, 125 por 94 pixels. O aumento na resolução dos dados foi feito com o intuito de melhorar a acurácia do classificador SVM, e vale ressaltar que com o aumento da resolução também foi mudada a decomposição PCA que agora decompõem em 400 componentes que conseguem explicar 96,87% da variância da imagem (ante os 95,69% de variância explicada com o PCA que decompunha as imagens de menor resolução).

O teste para a definição dos parâmetros seguiu com os mesmos intervalos definidos anteriormente pelas listas 2.4 e 2.5. O retorno foi de que o melhor par seria ($C=10$,

$\gamma=0.005$). Por nenhum dos valores estarem no limite dos valores testados não foi preciso mudar as listas de possíveis parâmetros.

O classificador é treinado com o banco de dados sendo decomposto pelas Componentes Principais do PCA. O treinamento ocorre apenas com os dados presentes na pasta raiz "Treinamento", como indicado pelo esquema do banco de dados presente na figura 2.6. Os dados dessa pasta são separados em grupo de treino e grupo de validação do treinamento que apresentam as mesmas pessoas (apesar de a mesma imagem não participar tanto do treinamento como da validação do treinamento). Essa separação entre grupo de treino e grupo de validação do treino é feita com a função *train_test_split* da biblioteca *sklearn.model_selection* em uma proporção de 4 para 1, respectivamente.

Além disso, após o treinamento, há o teste para saber se o classificador é realmente bem sucedido em sua tarefa, feito com os dados presentes na grande pasta "Teste", como indicado pelo esquema do banco de dados presente na figura 2.6, apenas com imagens de pessoas que não foram usadas no treinamento do classificador. Esse teste com separação é feito justamente para avaliar o grau de acerto do classificador quando é generalizado para situações em que ele não treinou; e também para verificar se os resultados apresentam *overfitting*, caso as classificações sejam satisfatórias para o grupo de pessoas do treinamento mas não para o grupo de pessoas de teste.

2.3 Elaboração da Rede Neural Convolutacional

Rede Neural Artificial

Uma Rede Neural Artificial (RNA) é uma técnica computacional inspirada no funcionamento do sistema nervoso humano, onde unidades de processamento (células) são organizadas em camadas, divididas em três tipos: camadas de entrada, que não realiza nenhum tipo de cálculo e apenas define a quantidade de informação a ser processada, camadas intermediárias ou escondidas extraem características dessa informação de entrada e o resultado de maior probabilidade é apresentado na camada de saída. A figura 2.13 mostra uma arquitetura básica de uma RNA com duas camadas escondidas.

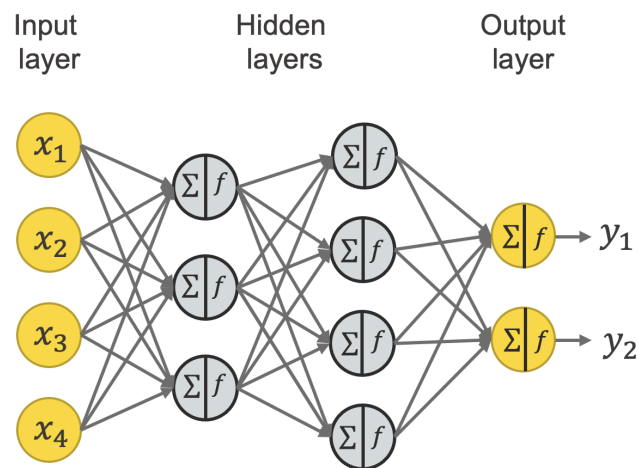


Figura 2.13: Arquitetura de uma Rede Neural Artificial simples.

FONTE: Extraído do site [KNIME](#)

Rede Neural Convolucional

Uma Rede Neural Convolucional (do inglês, *Convolutional Neural Network* ou CNN) é um caso especial de RNA onde a principal operação realizada pelas camadas escondidas é a de convolução. Imagens podem ser interpretadas como tensores de dimensão $m \times n$, onde m é a largura da imagem e n sua altura. Ao realizar o processo de convolução com um filtro (Kernel) de dimensão $f \times f$, gera-se como resultado uma matriz de dimensão $(m - f + 1) \times (n - f + 1)$, onde cada valor é representativo do tipo de característica que o Kernel é capaz de extrair. A figura 2.14 ilustra um passo do processo de convolução.

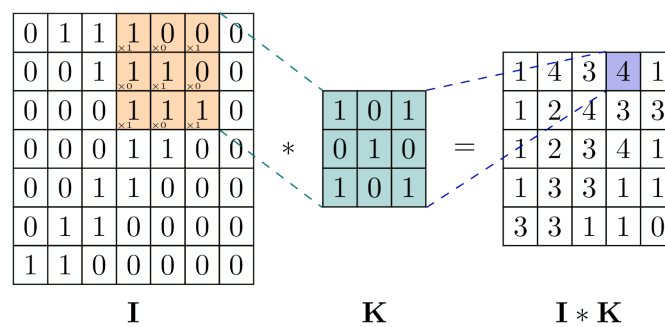


Figura 2.14: Processo de convolução exemplificado.

FONTE: Extraído do site [TIKZ](#)

Os valores dos elementos presentes no filtro são o que determinam qual a característica que será extraída da camada anterior. A título de exemplo, a figura 2.15 apresenta um Kernel básico utilizado na detecção de bordas verticais.

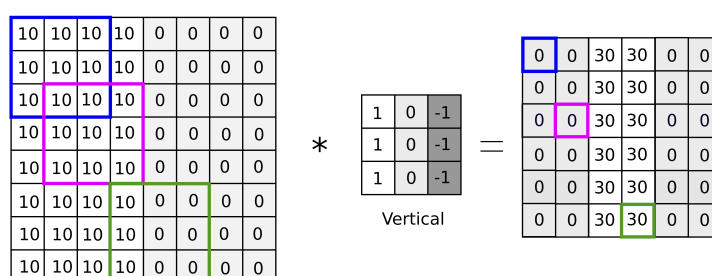


Figura 2.15: Detecção de bordas verticais.

FONTE: Extraído do site [anhreynolds](#)

Outro tipo de camada presente em CNNs é a camada de *pooling*, que não possui parâmetros treináveis e é utilizada para diminuir *overfitting* e reter características importantes. A figura 2.16 ilustra um exemplo de saída de uma camada de Max Pooling com um Kernel 2x2 e *Stride* (deslocamento do Kernel) igual a 2. Essas camadas não acrescentam parâmetros treináveis à rede.

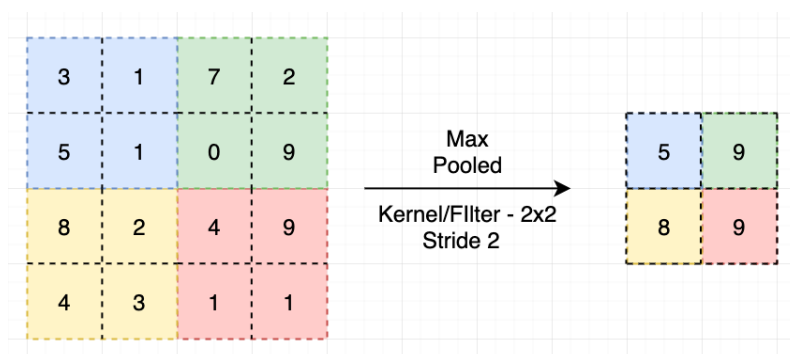


Figura 2.16: Exemplo de operação de Max Pooling.

FONTE: Extraído do site [Papers With Code](#)

Arquitetura Utilizada

A CNN será implementada em linguagem Python, com as ferramentas fornecidas pela biblioteca Keras e, os experimentos, no ambiente colaborativo Google Colab. A estrutura inicial da rede foi baseada no trabalho de (Simonyan e Zisserman 2014) e possui uma arquitetura simples, composta pela repetição de camadas de convolução, filtros 3x3 e *stride* 1, com max-pooling de janela 2x2 e *stride* igual 2.

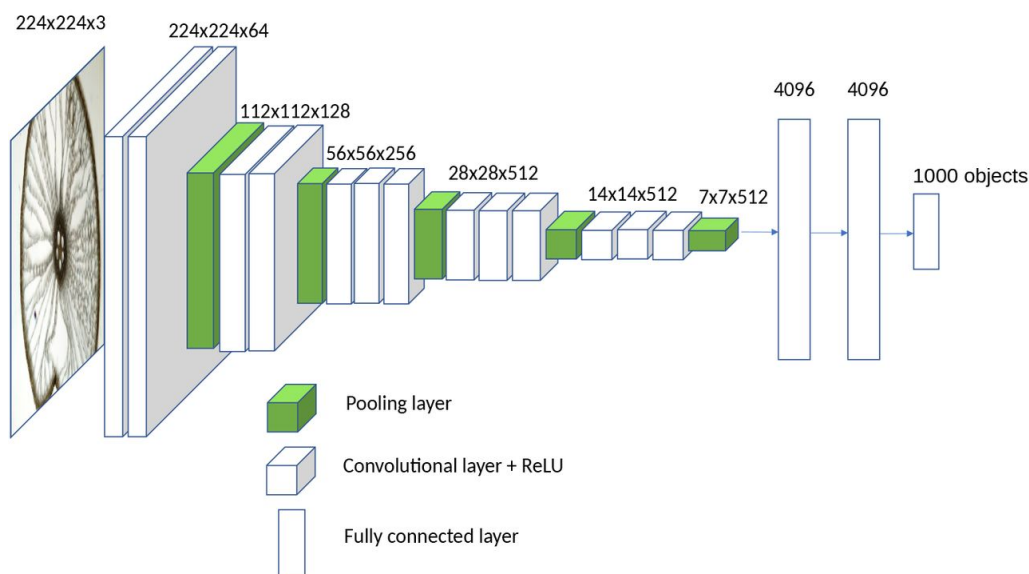


Figura 2.17: Arquitetura da VGG16

FONTE: extraído de [TowardsDataScience](#)

A figura 2.17 esquematiza a arquitetura da rede. As dimensões da camada de entrada podem ser alteradas de forma a serem compatíveis com os dados de entrada (no caso deste trabalho, imagens de dimensão 96x96x3) enquanto as dimensões dos filtros mantém o padrão apresentado - com as primeiras duas dimensões iguais as da camada de entrada e caindo pela metade conforme a quantidade de filtros dobra. Há um total de 16 camadas treináveis, somando aproximadamente 14.7 milhões de parâmetros.

Como a VGG16 não foi treinada especificamente para a classificação proposta no trabalho, há a necessidade de realizar uma transferência de aprendizado na rede, a fim de se obter resultados desejáveis. Esse processo consiste em descartar as camadas densas da rede, que, por estarem parametrizadas de forma muito específica ao treinamento generalista, não contribuem à correta classificação das classes. É então adicionada e treinada uma camada densa completamente nova, dimensionada para a tarefa específica. Durante esse treino inicial, as camadas convolucionais da rede não são alteradas, isto é, permanecem congeladas para que não haja a propagação de erros das camadas densas.

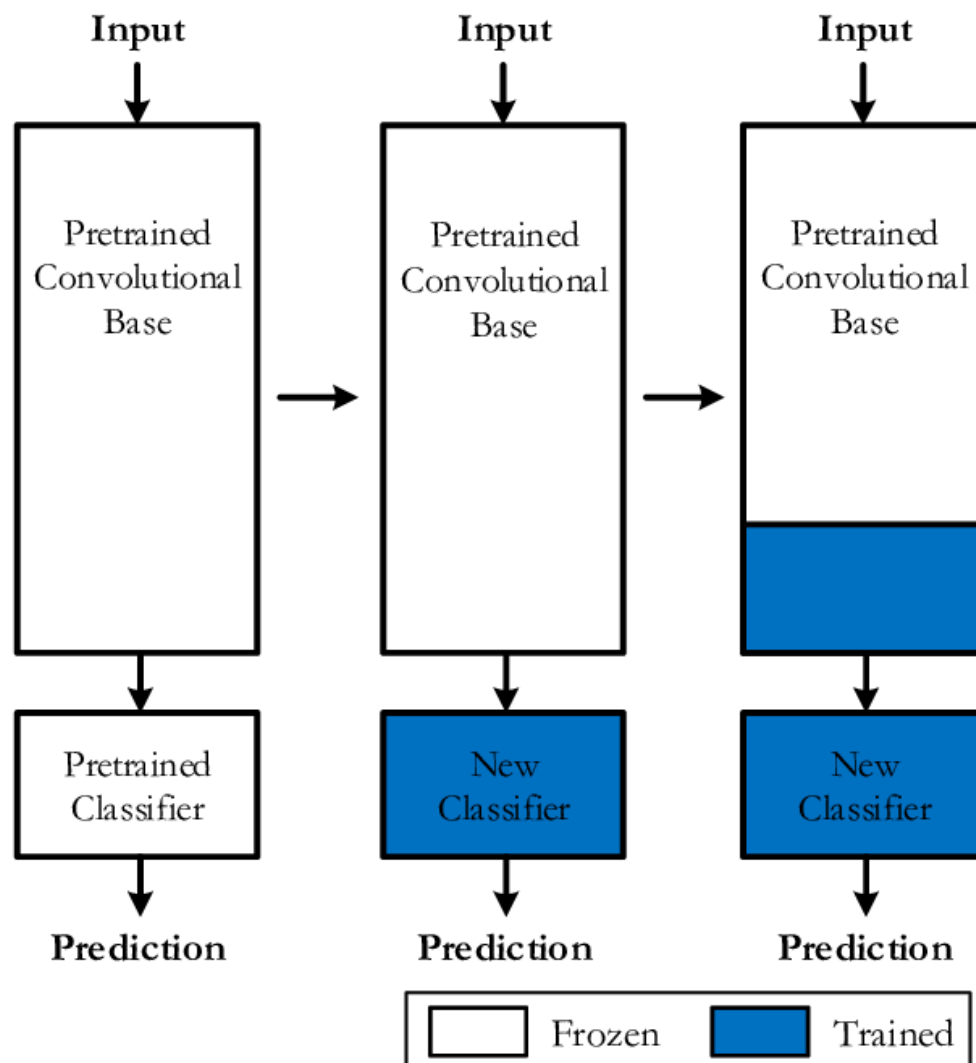


Figura 2.18: Arquitetura da VGG16

FONTE: extraído de [ResearchGate](#)

Após o treinamento inicial, algumas das últimas camadas convolucionais são descongeladas para treinamento - o que é chamado de sintonia fina da rede - e um novo treinamento é realizado. A figura 2.18 esquematiza todo esse processo. No presente trabalho, apenas o último bloco convolucional (composto por três camadas convolucionais) foi descongelado para treinamento.

2.4 Validação dos Resultados

O fluxo geral da validação dos resultados pode ser visto no diagrama da figura 2.19. Após o classificador SVM e a CNN serem desenvolvidos e treinados com a base de imagens coletada, estes deverão ser validados e testados. A validação é feita com dados separados do grupo de treino e o teste é feito com dados do grupo de teste. Nenhum voluntário pertence a ambos os grupos, para não macular os resultados, como explicitado na imagem 2.6.

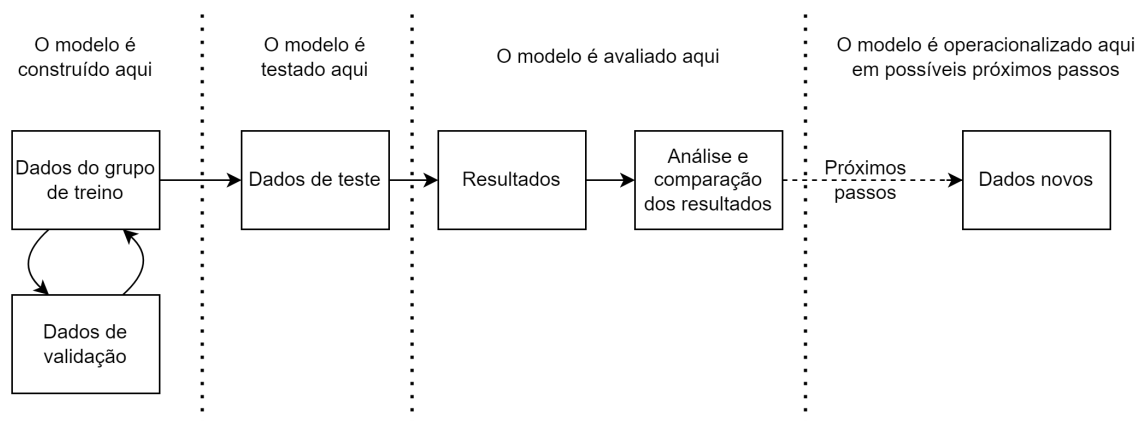


Figura 2.19: Diagrama de validação dos resultados.

FONTE: Autoria própria.

Sua análise deve ser crítica para avaliar a eficiência dos classificadores e buscar possíveis problemas na classificação. Dentre eles, destacam-se dificuldades em classificar algum nível de esforço específico, *overfitting* e hiper especialização, problemas comuns em *Machine Learning*.

Os dados principais para análise dos resultados dos classificadores são a tabela reporte de classificação e a matriz de confusão, ambos obtidos com funções prontas da biblioteca *sklearn.metrics*. A tabela mostra, para cada nível da classificação, a precisão (quantidade de previsões certas para uma classe, dada todas as previsões daquela classe), o *recall* (a capacidade do classificador em pegar todos os verdadeiros positivos da classe) e o *f1-score* (a média harmônica entre precisão e *recall*). A tabela também mostra essas métricas não só para cada classe, mas também para a acurácia geral do classificador, a média geral do classificador e a média ponderada geral do classificador (ponderada pela quantidade de dados em cada classe).

Já a matriz de confusão mostra a distribuição das classificações em uma matriz em que as linhas são a classificação real do dado e as colunas são a classificação prevista pelo classificador. Portanto, na diagonal principal da matriz estão as classificações corretas,

e fora da diagonal estão os casos de erro na classificação. A matriz é bastante útil para entender como e porquê um classificador está errando.

Resultados e Discussões

3.1 Resultados

3.1.1 Banco de dados

Tanto os vídeos coletados pela equipe da Unicamp quanto os vídeos coletados pelos autores serão tratados pelo mesmo arquivo Python *make_db* que pode ser conferido na íntegra no [GitHub do projeto](#).

O banco de dados foi coletado e pré processado de acordo com o que foi descrito em 2.1 e salvo em pastas organizadas em uma estrutura ilustrada na imagem 2.6. O banco de dados para treinamento é composto de 1242 imagens de faces com classificação de nível de esforço, sendo 270 de nível baixo, 698 de nível médio e 274 de nível alto. Já o banco de dados para testes é composto de 257 imagens, sendo 28 de nível baixo, 149 de nível médio e 80 de nível alto.

3.1.2 Classificador SVM com PCA

PCA

O Jupyter Notebook *PCA.ipynb* onde o PCA foi feito pode ser conferido na íntegra no [GitHub do projeto](#).

Graças ao uso da base LFW, o PCA possui as componentes principais abrangentes para as mais variadas faces humanas.

Como descrito na metodologia em 2.2, não é preciso usar todas as componentes principais das faces. No primeiro momento, o PCA foi feito com um *resizing* de 0.5 nas fotos presentes no LFW, dando medidas de 62 pixels de altura por 47 de largura, dando ao total 2914 dimensões para as imagens. Com a finalidade de deixar o treinamento e a classificação do algoritmo SVM mais rápido, são usadas apenas as 200 componentes

principais no PCA, o que garante 95,69% de porcentagem de variância da face explicada, como visto no gráfico 2.8. Pode-se comparar uma foto original com a reconstrução pelo PCA na figura 2.9 e podem ser vistos alguns exemplos de *eigenfaces* do PCA na figura 2.7.

O compilador demorou cerca de 36 segundos para fazer o gráfico de variância explicada pelo número de componentes e apenas 6 segundos para gerar o PCA das 200 primeiras componentes, tudo no ambiente *default* do Google Colab. A parte mais demorada do processo, na verdade, foi a importação da base LFW, que levou cerca de 1 minuto.

O compilador demorou cerca de 14 minutos para fazer o gráfico de variância explicada pelo número de componentes sem o resizing e apenas 27 segundos para gerar o PCA das 400 primeiras componentes, tudo no ambiente *default* do Google Colab.

Classificador SVM

Com o PCA feito, o classificador SVM pode ser treinado, de acordo com o que foi explicitado em 2.2.

No primeiro momento, usando um terço do banco de dados e com as dimensões de imagem próprias do LFW com resizing de 0,5, ou seja, 62 por 47 pixels, a melhor dupla de hiper parâmetros testados foi ($C=10$, $\gamma=0.01$).

O classificador SVM com kernel Gaussiano levou 5 segundos para descobrir os melhores hiper parâmetros com as listas passadas e treinar com os dados do banco de faces.

Os resultados obtidos dentro das validações no treinamento do classificador SVM de kernel Gaussiano podem ser conferidos na tabela 3.1 e na imagem 3.1.

	Precision	Recall	f1-score	Support
Alto	0.85	1.00	0.92	17
Médio	0.98	0.93	0.95	43
Baixo	0.96	0.92	0.94	25
Accuracy			0.94	85
Macro avg	0.93	0.95	0.94	85
Weighted avg	0.95	0.94	0.94	85

Tabela 3.1: Tabela reporte de classificação em validação do SVM com um terço da base e fotos de 2914 dimensões.

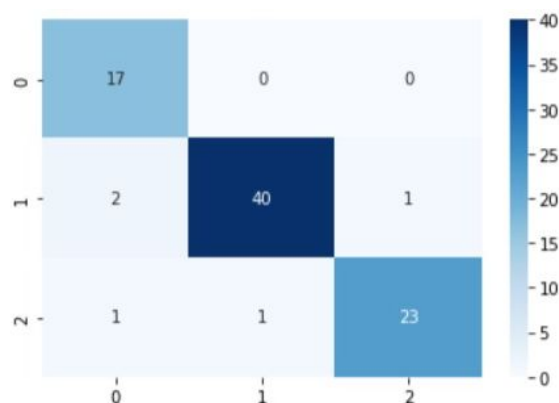


Figura 3.1: Matriz de confusão em validação do SVM com um terço da base e fotos de 2914 dimensões.

FONTE: Autoria própria.

Já usando esse classificador treinado com um terço de base e com fotos de 2914 dimensões, ao fazer os testes definitivos com dados que não participaram do treinamento e validação, os resultados mostraram uma queda significativa de desempenho, como explicitado na tabela 3.2 e na imagem 3.2.

	Precision	Recall	f1-score	Support
Alto	0.00	0.00	0.00	18
Médio	0.35	0.43	0.39	28
Baixo	0.00	0.00	0.00	4
Accuracy			0.24	50
Macro avg	0.12	0.14	0.13	50
Weighted avg	0.20	0.24	0.22	50

Tabela 3.2: Tabela reporte de classificação em teste do SVM com um terço da base e fotos de 2914 dimensões.

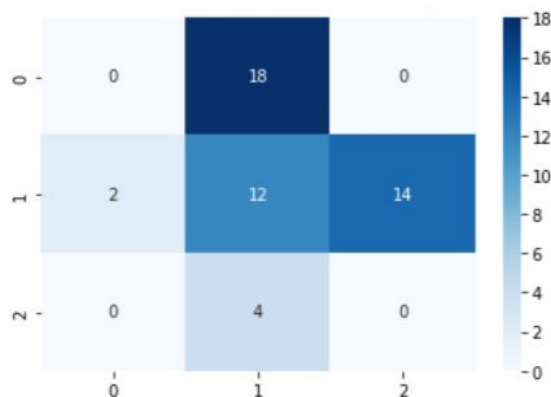


Figura 3.2: Matriz de confusão em validação do SVM com um terço da base e fotos de 2914 dimensões.

FONTE: Autoria própria.

No segundo momento, com o banco de dados completo e ainda com as dimensões de imagem próprias do LFW com resizing de 0,5, ou seja, 62 por 47 pixels, o teste para a definição dos hiperparâmetros retornou que o melhor par é ($C=5$, $\gamma=0.005$).

O classificador SVM com Kernel Gaussiano levou 50 segundos para descobrir os melhores hiperparâmetros com as listas passadas e treinar com os dados do banco de faces.

Os resultados obtidos dentro das validações no treinamento do classificador SVM de Kernel Gaussiano podem ser vistos na tabela 3.3 e na figura 3.3.

	Precision	Recall	f1-score	Support
Alto	0.94	0.85	0.89	54
Médio	0.89	0.96	0.92	140
Baixo	0.94	0.85	0.90	55
Accuracy			0.91	249
Macro avg	0.92	0.89	0.90	249
Weighted avg	0.91	0.91	0.91	249

Tabela 3.3: Tabela reporte de classificação em validação do SVM com base completa e fotos de 2914 dimensões.

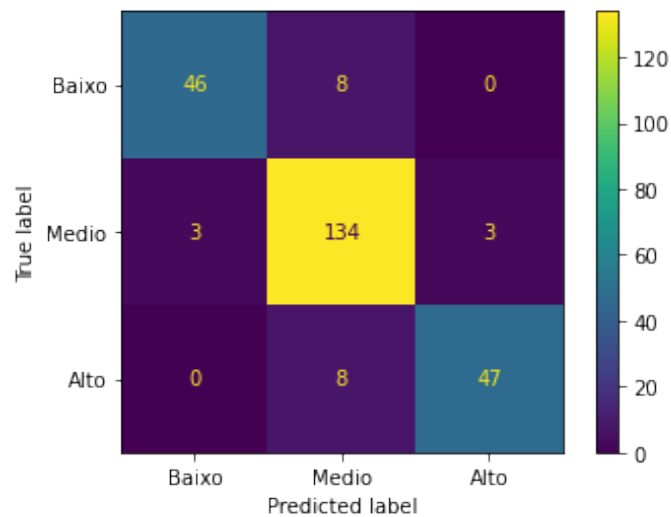


Figura 3.3: Matriz de confusão em validação do SVM com base completa e fotos de 2914 dimensões.

FONTE: Autoria própria.

Usando esse classificador treinado com a base completa e com fotos de 2914 dimensões, ao fazer os testes definitivos com dados que não participaram do treinamento e validação, os resultados mostraram uma leve queda na precisão, mas ainda mantendo resultados elevados, como explicitado na tabela 3.4 e na figura 3.4.

	Precision	Recall	f1-score	Support
Alto	0.86	0.43	0.57	28
Médio	0.71	0.97	0.82	149
Baixo	0.93	0.47	0.63	80
Accuracy			0.75	257
Macro avg	0.83	0.62	0.67	257
Weighted avg	0.80	0.75	0.73	257

Tabela 3.4: Tabela reporte de classificação em teste do SVM com base completa e fotos de 2914 dimensões.

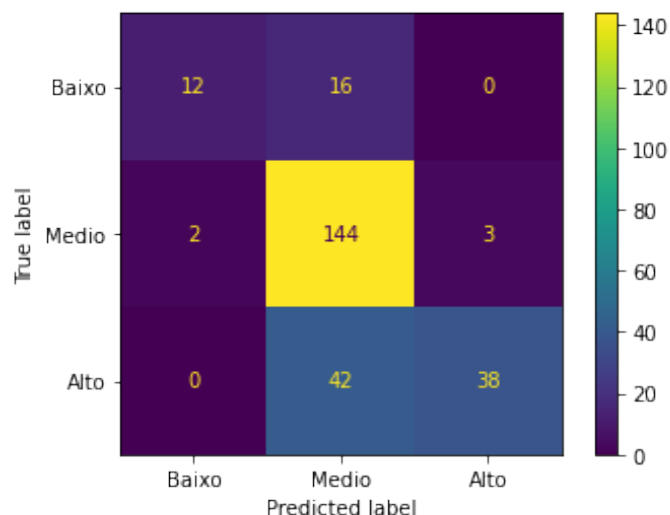


Figura 3.4: Matriz de confusão em validação do SVM com base completa e fotos de 2914 dimensões.

FONTE: Autoria própria.

E por fim, no terceiro e último momento, foi treinada a rede com o banco de dados completo e com as dimensões de imagem próprias do LFW sem resizing, ou seja, 125 por 94 pixels. O aumento na resolução dos dados foi feito com o intuito de melhorar a acurácia do classificador SVM, e vale ressaltar que com o aumento da resolução também foi mudado a decomposição PCA que agora decompõem em 400 componentes que conseguem explicar 96,87% da variância da imagem (ante os 95,69% de variância explicada com o PCA que decompunha as imagens de menor resolução). O teste para a definição dos hiper parâmetros retornou que o melhor par é ($C=10$, $\gamma=0.005$).

O classificador SVM com kernel Gaussiano novamente levou 50 segundos para descobrir os melhores hiper parâmetros com as listas passadas e treinar com os dados do banco de faces, mesmo trabalhando com 400 componentes principais, o dobro do que anteriormente.

Os resultados obtidos dentro das validações no treinamento do classificador SVM de kernel Gaussiano podem ser vistos na tabela 3.5 e na figura 3.5.

	Precision	Recall	f1-score	Support
Alto	0.96	0.85	0.90	54
Médio	0.89	0.96	0.93	140
Baixo	0.94	0.85	0.90	55
Accuracy			0.92	249
Macro avg	0.93	0.89	0.91	249
Weighted avg	0.92	0.92	0.92	249

Tabela 3.5: Tabela reporte de classificação em validação do SVM com base completa e fotos de 11750 dimensões.

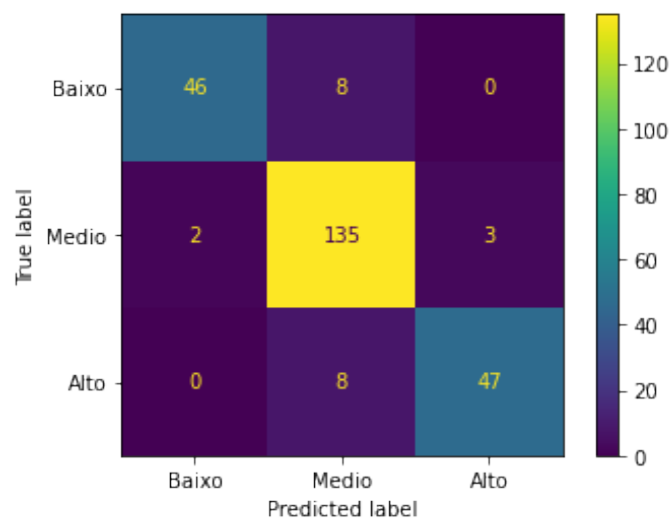


Figura 3.5: Matriz de confusão em validação do SVM com base completa e fotos de 11750 dimensões.

FONTE: Autoria própria.

Usando esse classificador treinado com a base completa e com fotos de 11750 dimensões, ao fazer os testes definitivos com dados que não participaram do treinamento e validação, os resultados mostraram novamente uma leve queda na precisão, praticamente a mesma, apesar de ainda mantendo resultados elevados, como explicitado na tabela 3.6 e na figura 3.6.

	Precision	Recall	f1-score	Support
Alto	0.86	0.43	0.57	28
Médio	0.72	0.97	0.83	149
Baixo	0.93	0.50	0.65	80
Accuracy			0.76	257
Macro avg	0.84	0.63	0.68	257
Weighted avg	0.80	0.76	0.74	257

Tabela 3.6: Tabela reporte de classificação em teste do SVM com base completa e fotos de 11750 dimensões.

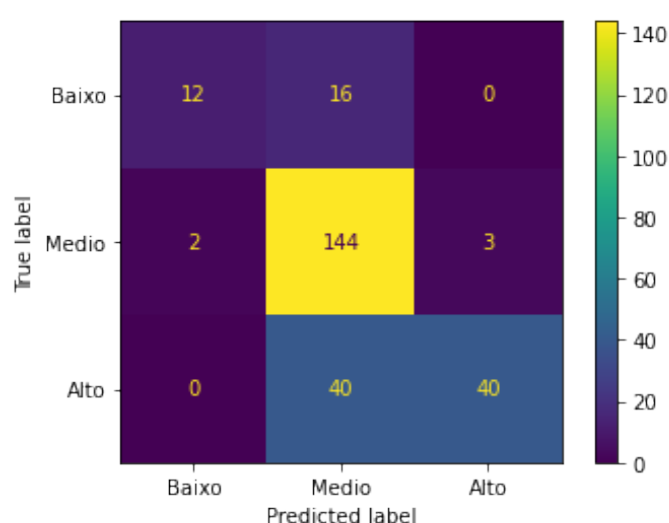


Figura 3.6: Matriz de confusão em validação do SVM com base completa e fotos de 11750 dimensões.

FONTE: Autoria própria.

Nos dois casos da SVM treinada com a base completa, as 257 faces que não participaram do treinamento (apenas do teste com o classificador já treinado) foram decompostas pelo PCA e classificadas pela SVM em menos de 1 segundo. Assim, estima-se que o classificador leva menos de 3 milissegundos para realizar o processo de classificação de uma face a partir de sua imagem.

Para comparação geral entre os classificadores SVM será usada a medida *weighted average f1-score* presente nas Tabelas reporte de Classificação. É a medida que considera a precisão e o recall de cada classe individual ponderadas por quantos dados cada classe tem.

Classificador com CNN

Ao utilizar uma arquitetura já treinada, como a VGG16, espera-se remediar as dificuldades advindas de uma base de dados escassa, como a do presente trabalho. Com a maioria de seus hiperparâmetros já sintonizada, as chances da rede sofrer um *overfitting* diminuem e a expectativa por resultados melhores aumenta.

Resultados Após Treinamento das Camadas Densas da Rede

As figuras 3.7 e 3.8 ilustram a evolução na função custo e na exatidão conforme o treinamento das camadas densas, respectivamente.

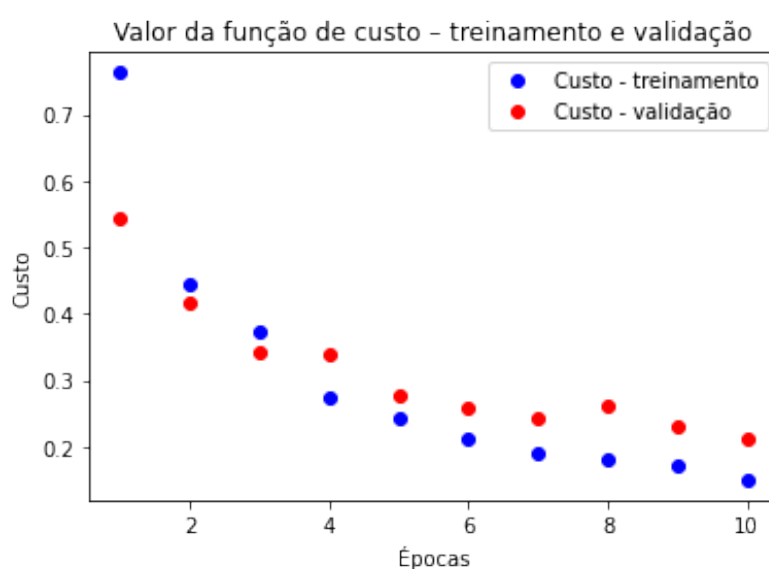


Figura 3.7: Valor da Função Custo durante o treinamento das camadas densas da rede.

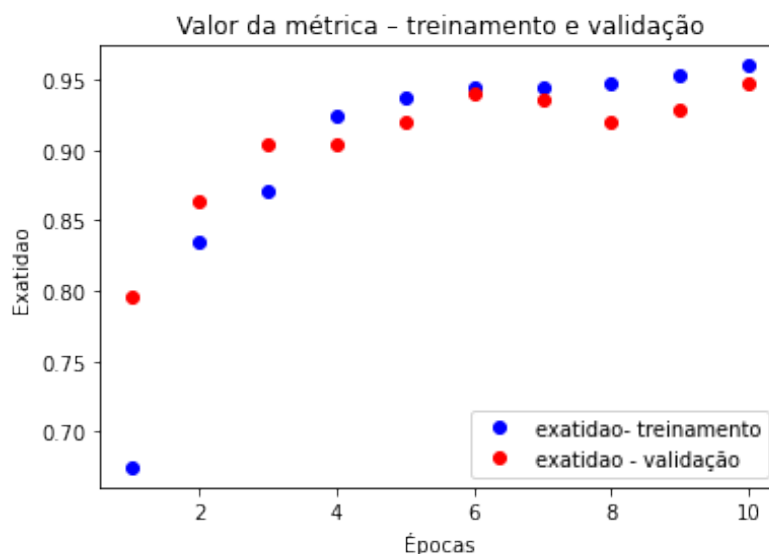


Figura 3.8: Valor da Exatidão durante o treinamento das camadas densas da rede.

Observa-se uma clara evolução na qualidade da métrica conforme a quantidade de Épocas (uma Época é definida como um uso completo de todos os dados de treinamento). Um bom indicativo de pouco ou nenhum *overfitting* é o comportamento similar das curvas de treinamento e validação - não há uma melhora na métrica para a base de treino sem simultânea melhora na acurácia para os dados de validação. Além disso, a diferença entre as métricas também é satisfatoriamente baixa, na ordem de menos de 1%.

No total, o treinamento das camadas densas durou 1185 segundos ou 19 minutos e 40 segundos. A tabela 3.1.2 consolida o valor da métrica de exatidão e o valor da função de custo obtida ao fim do treinamento, para as bases de treino, validação e teste.

	Valor Função Custo	Acurácia
Treino	0.1335	0.9718
Validação	0.2107	0.9478
Teste	0.1862	0.9494

Tabela 3.7: Valor da Função Custo e Acurácia da rede após o treinamento das camadas densas.

Observa-se que, de fato, os resultados obtidos para os diferentes conjuntos de dados foram muito próximos. Contudo, como há uma distribuição desigual na quantidade de imagens por classe, pode haver um desempenho insatisfatório da rede na classificação de uma delas, que esteja sendo mascarado por essa desproporção de dados por classe. Para ter uma melhor noção de seu desempenho para cada classe, a figura 3.9 apresenta o resultado de todas as previsões realizadas no grupo de teste.

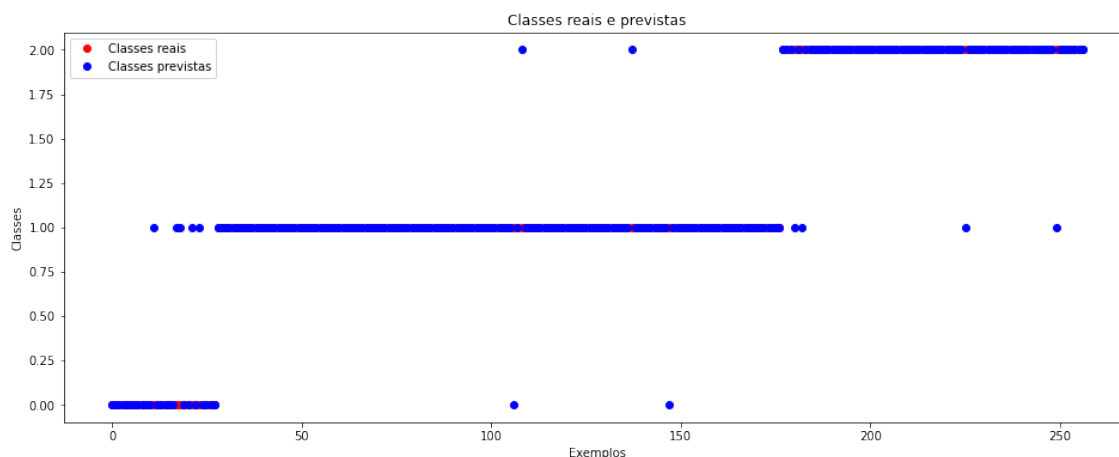


Figura 3.9: Previsões realizadas para o grupo de teste após o treinamento das camadas densas da rede.

Como mencionado, a classificação é feita em três classes distintas, sendo a classe 0 equivalente ao esforço baixo, a classe 1 equivalente ao esforço médio e a classe 2 equivalente ao esforço alto. Pontos em vermelho indicam que a foto pertencente a classe em questão não foi corretamente identificada (caso contrário, um ponto em azul se sobreporia). Observa-se que as identificações incorretas concentram-se na classe 0, com a rede classificando-as como classe 1. Para uma visão mais clara dessas avaliações, na figura 3.10 tem-se a matriz de confusão das previsões e a tabela 3.8 consolida os valores de Precisão, *Recall* e *f1-score* para os dados de teste.

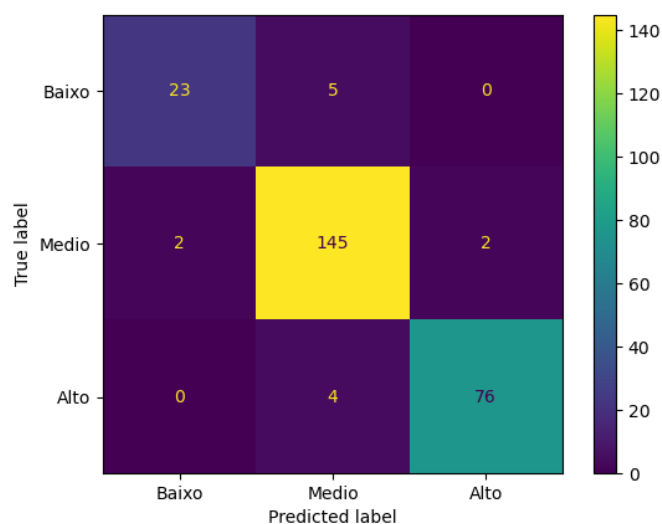


Figura 3.10: Matriz de confusão em validação da CNN, após o treinamento das camadas densas.

	Precision	Recall	f1-score	Support
Alto	0.97	0.95	0.96	28
Médio	0.94	0.97	0.96	149
Baixo	0.92	0.82	0.90	87
Accuracy			0.94	257
Macro avg	0.95	0.91	0.93	257
Weighted avg	0.95	0.95	0.95	257

Tabela 3.8: Tabela de reporte de classificação em teste da CNN, após o treinamento das camadas densas.

Percebe-se que, de fato, há uma variação nas métricas de acordo com a classe analisada, principalmente na métrica de *recall* na classificação de imagens de nível baixo de esforço. Contudo, a média ponderada das previsões resulta uma precisão, *recall* e *f1-score* de 95%, que é um valor satisfatório. Dentre as três classes, a que mais pode ser avaliada com leniência é justamente a de baixo esforço - tanto pelo baixo tempo em que o voluntário permanece neste nível durante o exercício (daí a baixa proporção de imagens com essa classificação frente ao total) quanto pela necessidade de acompanhamento e possíveis riscos a saúde do voluntário.

Resultados Após Sintonia Fina da Rede

Após o treinamento das camadas densas, ainda resta descongelar algumas das últimas camadas convolucionais da rede - que costumam extrair características mais específicas ao problema de classificação imposto - e treiná-las novamente, reconfigurando seus hiperparâmetros.

As figuras 3.11 e 3.12 ilustram os resultados obtidos após a sintonia fina da rede, descongeladas as três últimas camadas convolucionais (quinto bloco convolucional).

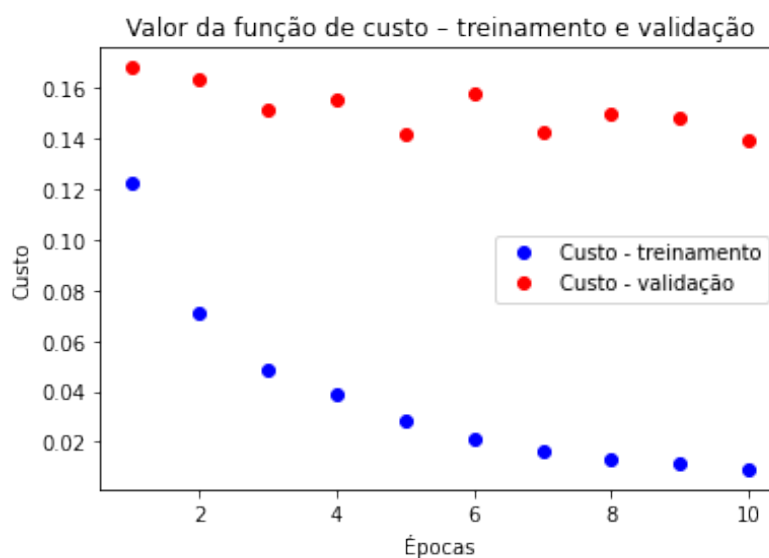


Figura 3.11: Valor da Função Custo após a sintonia fina da rede.

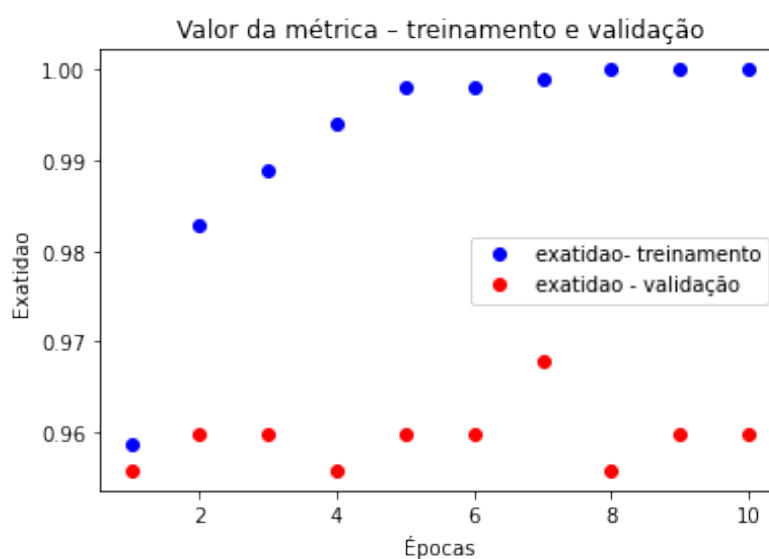


Figura 3.12: Valor da Exatidão após a sintonia fina da rede.

Percebe-se que a métrica de exatidão para o conjunto de validação não evolui tão bem quanto a mesma no conjunto de treinamento, mas ambas superaram, após a décima época, as métricas pré-sintonia. A margem para melhora na exatidão da rede também não é grande o suficiente para que se espere uma melhora brusca nos resultados.

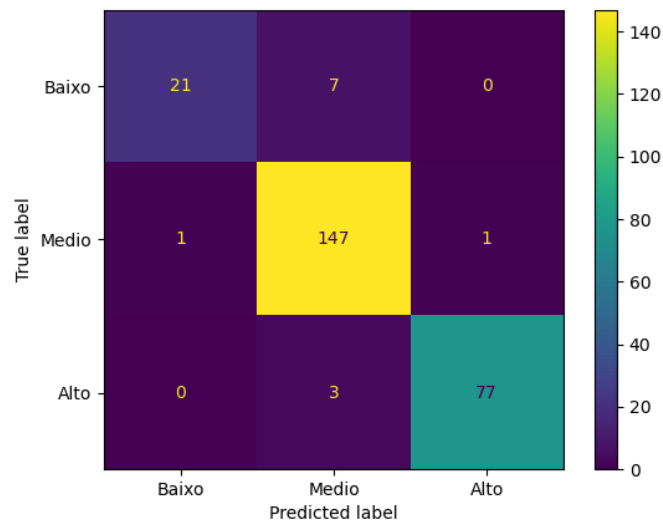


Figura 3.13: Matriz de confusão em validação da CNN, após a sintonia fina das camadas convolucionais.

Importante notar que a escala do gráfico da figura 3.12 é substancialmente menor do que a do gráfico em 3.8, o que dá a falsa impressão da margem entre as curvas ser muito maior. De fato, houve uma estagnação no avanço da curva referente aos dados de validação, o que indica que este treinamento não foi tão eficaz quanto o anterior. De todo modo, as tabelas 3.1.2 e 3.10 trazem as mesmas métricas consolidadas no passo anterior, enquanto a figura 3.13 mostra a matriz de confusão obtida e na figura 3.14 observa-se todas as previsões feitas para o conjunto de teste.

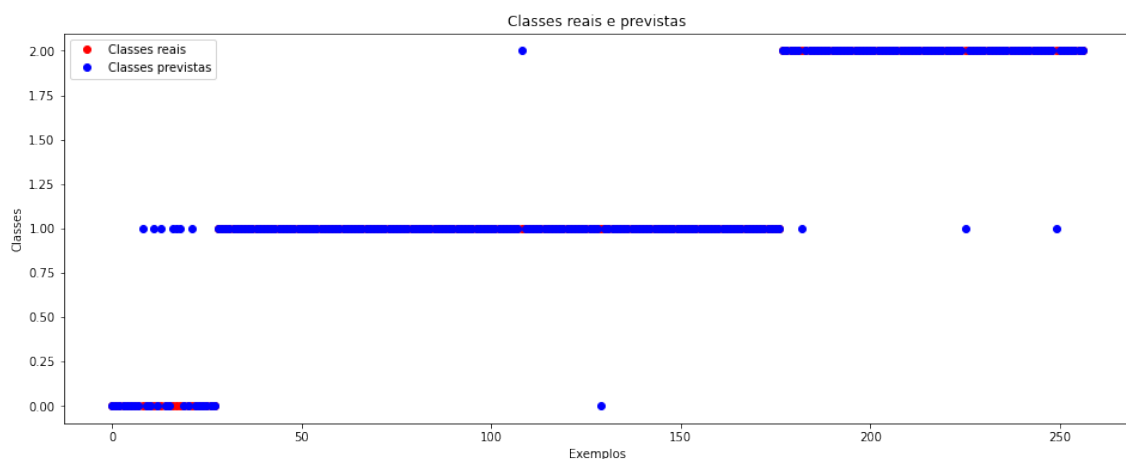


Figura 3.14: Previsões realizadas para o grupo de teste após a sintonia fina das camadas convolucionais.

	Valor Função Custo	Acurácia
Treino	0.0066	1.0000
Validação	0.1393	0.9598
Teste	0.1515	0.9533

Tabela 3.9: Valor da Função Custo e Acurácia da rede após a sintonia fina das camadas convolucionais.

	Precision	Recall	f1-score	Support
Alto	0.99	0.96	0.97	28
Médio	0.94	0.99	0.96	149
Baixo	0.95	0.75	0.84	87
Accuracy			0.95	257
Macro avg	0.96	0.90	0.93	257
Weighted avg	0.95	0.95	0.95	257

Tabela 3.10: Tabela de reporte de classificação em teste da CNN, após sintonia fina das camadas convolucionais.

3.2 Discussões

Ao analisar os resultados do primeiro treinamento parcial do classificador SVM - onde havia apenas um terço da base de dados completa em sua versão final, imagens com 2914 dimensões e PCA com 95,69% de variância explicada - nota-se grande *overfitting* do classificador, dada a discrepância entre os resultados obtidos para os dados de treinamento e teste.

As previsões de faces durante o treinamento, apresentaram um *weighted average f1-score* igual 94%. Já para os dados de teste, a mesma medida ficou em torno de 22%, chegando a errar todas as previsões de faces que demonstravam baixo e alto esforço, as classificações com menos dados disponíveis para treinamento.

Por mais que os resultados iniciais do classificador não sejam os definitivos e nem promissores, eles ajudam a ilustrar dois pontos. O primeiro é o cuidado que se deve ter ao lidar com o classificador SVM, em especial em treinamentos complexos, por ser um método suscetível a hiper especialização *overfitting*. O segundo é a importância de se trabalhar com uma base de treinamento robusta.

Isso se comprova ao analisar os resultados do segundo treinamento do classificador - agora com a base de dados completa - que traz resultados mais promissores. O indicador principal, *weighted average f1-score*, das previsões com faces que participaram do treinamento continuou elevado em 91% (da tabela 3.3), apesar de uma pequena queda

comparado ao anterior com um banco de dados menor. Essa diferença é explicada pelo desempenho com os dados de teste, vistos na tabela (3.4): com 73% de *weighted average f1-score*, uma melhora substancial com relação ao seu último treinamento que se deve ao fato do classificador ser mais generalista (muito melhor prevendo dados externos e pouco pior prevendo dados internos à base). De todo modo, a queda ainda é presente e indica traços de hiperespecialização e *overfitting*.

De todo modo, não é necessário descartar esse classificador justamente por observar grande melhora em seus resultados conforme se aumenta a quantidade de dados disponíveis; contanto que se continue adquirindo dados como parte dos próximos passos, há boas perspectivas para esse classificador, principalmente por conta de sua velocidade de treinamento, muito superior se comparada a CNN.

A importância da quantidade de dados é observada comparando o *recall* entre as tabelas 3.3 e 3.4: as categorias que tiveram maior queda foram as de alto e baixo esforço, categorias com menos dados de suporte para treinamento do classificador. Inclusive, observa-se que, quanto menor a quantidade de dados, menor o *recall*.

Por fim, foi feito um terceiro e último treinamento do classificador SVM - treinado com a base de dados completa, imagens com 11750 dimensões e PCA com 96,87% de variância explicada - com o objetivo de aumentar a precisão geral do classificador aumentando a resolução das imagens e da variância explicada pelo PCA; porém essas variações não melhoraram em muito os resultados.

Com as alterações feitas, o *weighted average f1-score* subiu apenas um ponto percentual, chegando a 92% quando classifica o conjunto de treinamento (tabela 3.5). Esse mesmo aumento é observado no conjunto de teste, com o *weighted average f1-score* para 74% (tabela 3.6).

Se por um lado os resultados não atingiram o objetivo esperado de melhorar significativamente a precisão do modelo, eles ao menos serviram para mostrar o poder das simplificações adotadas ao realizar o treinamento do classificador SVM. Com o *resize* de 0.5% a resolução e as dimensões das imagens analisadas foram divididas por 4 e com o PCA com 200 componentes a menos foram economizadas centenas de dimensões que não precisaram ser consideradas pelo classificador ao treinar, economizando muito tempo e poder computacional. E mesmo com as economias e simplificações, os resultados caíram apenas 1 ponto percentual entre os dois treinamentos, mostrando-se boas simplificações com relação ao seu custo benefício e consolidando o PCA como uma ótima ferramenta para o classificador SVM.

Já os resultados obtidos com a VGG16 foram consideravelmente melhores; apenas com o treinamento das camadas densas, obteve-se *f1-score*, precisão e *recall* ponderados de 94.89%, 94.94% e 94.94%, respectivamente, corroboram com a hipótese de que a rede atende

melhor ao problema de classificação apresentado. A métrica de pior performance, nessa etapa do treinamento, foi o *recall* da classe de baixo esforço, de 82.14%, mas que pode ser explicada pela baixa quantidade de fotos disponíveis para teste. Essa desproporcionalidade, por sua vez, é consequência da parcela de tempo em que a dupla julgou o esforço dos voluntários como deste nível, isto é, durante a maior parte do exercício os voluntários apresentam expressões faciais rotuláveis como média ou alta; portanto, esse valor é considerado aceitável frente aos valores médios.

Com a sintonia fina da rede houve uma melhora marginal nas médias ponderadas de precisão, *recall* e *f1-score*, iguais a 95.41%, 95.33% e 95.20%, respectivamente. Essa melhora equivale, em termos práticos, a um acerto a mais nas previsões da rede. Houve também uma redistribuição neste total de 12 erros de previsão da rede, como se pode observar comparando as matrizes de confusão 3.10 e 3.13.

Uma boa justificativa para a diferença entre as performances das metodologias está no fato da VGG16 ser treinada com uma base robusta, a Imagenet (Deng et al. 2009), o que previne, de certo modo, o *overfitting* da rede com os dados de treinamento. No mais, o desempenho da CNN a torna a metodologia mais faorável para uma continuidade do projeto.

Com relação aos erros de classificação, pode-se notar um padrão nas matrizes de confusão, tanto para a SVM como para a CNN, em que não há erros de classificação extremos (um alto classificado como baixo e vice-versa). Os erros se concentram, em sua grande maioria, em níveis de esforço baixo ou alto sendo classificados como médio. Muito disso se dá pelo fato de que as classificações baixa e alta eram as que menos tinham dados disponíveis, e isso pode ser resolvido com mais dados em próximos passos. No caso de um esforço baixo ser classificado como médio não haveriam implicações perigosas ao se utilizar na prática. Já no caso em que o alto for classificado como médio isso pode gerar problemas na medida em que o paciente pode ser levado a cargas ainda maiores mesmo já realizando um esforço elevado. É, portanto, o erro mais perigoso e deve ser eliminado antes da implementação, ou restringir o uso do classificador para apenas diminuir o esforço (e não aumentá-lo) enquanto o problema não for sanado.

Conclusão

4.1 Conclusão

O treino de força e resistência é utilizado em diversas áreas da saúde, em especial para reabilitação e terapia. Os profissionais que monitoram tais treinos devem estar em constante vigilância do nível de esforço e carga que estão aplicando em seus pacientes. Hoje em dia, esse monitoramento se dá de forma analógica, demanda tempo e é suscetível a subjetividade humana.

O presente projeto se propôs a desenvolver um classificador automático do nível de esforço a partir da análise de expressão facial, a fim de facilitar o trabalho desses profissionais de saúde. Para tanto, utilizou-se duas abordagens - ambas com respaldo na literatura - a fim de encontrar a mais adequada ao problema e propô-la como solução. Uma delas foi desenvolver um classificador SVM com Kernel Gaussiano, com uso de PCA a fim de diminuir o custo computacional. O segundo método foi retreinar uma Rede Neural Convolucional - a VGG16 - para a tarefa específica do trabalho.

O classificador SVM teve um *f1-score* ponderado de 91% para a validação interna da base de treinamento e, para o grupo de teste, que conta apenas com voluntários que não possuem nenhuma foto na base de treinamento (ao invés de apenas uma extração aleatória de imagens), o valor obtido foi de 73%.

A CNN baseada na arquitetura da VGG16 trouxe resultados melhores, com um *f1-score* ponderado de 95% para o conjunto de teste; a principal atribuição a diferença de desempenho entre as metodologias é a quantidade de dados disponíveis para treinamento: como a rede pré-treinada está menos propensa a *overfitting* de seus hiperparâmetros, sua performance na base de testes foi menos afetada se comparada ao classificador SVM. Desta forma, enquanto a disponibilidade de dados for baixa, trabalhar com os resultados do classificador CNN é a forma mais confiável de dar prosseguimento ao trabalho.

4.2 Próximos passos

Com o intuito de melhorar a eficiência do classificador, pode-se focar em aprimorar a base de dados construída, buscando mais voluntários que, com suas contribuições, aumentarão e diversificarão a quantidade de imagens disponíveis. A metodologia para a rotulação das imagens também pode ser revista e aderir mais ao trabalho de (Uchida et al. 2018). Também pode-se considerar o uso de ambos os métodos abordados em um sistema de votação ponderada como abordagem de classificação, caso se observe uma melhora nos resultados.

O classificador pode ser embarcado em um *single board computer* que opere em tempo real, provendo *feedback* contínuo do usuário. Na medida em que a confiabilidade se tornar satisfatória para os profissionais de saúde envolvidos, pode ser possível desenvolver um sistema de controle totalmente automatizado, que ajuste a carga da atividade de acordo com a resposta do classificador.

Referências

- Almeida, Alice Squillante Pereira de (2015). *Expressão facial: uma nova perspectiva para monitoramento do exercício de força*. TCC. Faculdade de Educação Física - Universidade Estadual de Campinas.
- Deng, Jia et al. (2009). “Imagenet: A large-scale hierarchical image database”. Em: *2009 IEEE conference on computer vision and pattern recognition*. Ieee, pp. 248–255.
- Ekman, Paul e Wallace V Friesen (1978). *Facial action coding system: Investigator’s guide*. Consulting Psychologists Press.
- Htay, Moe Moe (2021). “Feature extraction and classification methods of facial expression: A survey”. Em: *Computer Science and Information Technologies 2.1*, pp. 26–32.
- Kawashima, Tomoki, Hiroki Nomiya e Teruhisa Hochin (2021). “Facial Expression Intensity Estimation using Deep Convolutional Neural Network”. Em: *Proceedings of the The 8th International Virtual Conference on Applied Computing & Information Technology*, pp. 7–12.
- Khanal, Salik Ram, João Barroso et al. (2018). “Classification of physical exercise intensity by using facial expression analysis”. Em: *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, pp. 765–770.
- Khanal, Salik Ram, Jaime Sampaio et al. (2019). “Classification of physical exercise intensity based on facial expression using deep neural network”. Em: *International Conference on Human-Computer Interaction*. Springer, pp. 455–467.
- Learn, SciKit (2022). *SciKit Learn - 1.4. Support Vector Machines*. URL: <https://scikit-learn.org/stable/modules/svm.html> (acesso em 30/11/2022).
- Li, Junnan e Edmund Y Lam (2015). “Facial expression recognition using deep neural networks”. Em: *2015 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, pp. 1–6.

- Lopes, André Teixeira et al. (2017). “Facial expression recognition with convolutional neural networks: coping with few data and the training sample order”. Em: *Pattern recognition* 61, pp. 610–628.
- Roy, Sourav D et al. (2016). “An Approach for Automatic Pain Detection through Facial Expression”. Em: vol. 84. Elsevier B.V., pp. 99–106. DOI: [10.1016/j.procs.2016.04.072](https://doi.org/10.1016/j.procs.2016.04.072).
- Simonyan, Karen e Andrew Zisserman (2014). “Very deep convolutional networks for large-scale image recognition”. Em: *arXiv preprint arXiv:1409.1556*.
- Uchida, Marco C et al. (2018). “Identification of muscle fatigue by tracking facial expressions”. Em: *PLoS One* 13.12, eo208834.
- Vision, UMass (2018). *Labeled Faces in the Wild*. URL: <http://vis-www.cs.umass.edu/lfw/> (acesso em 28/11/2022).

Apêndices

Apêndice I - Projeto Plataforma Brasil

Arquivo do projeto que foi submetido na plataforma Brasil para aprovação da coleta de imagens de pessoas se exercitando.

IA para medição de escala de esforço em fisioterapia.

Pesquisador principal:

Prof. Dr. Rafael Traldi Moura

Assistente da pesquisa:

Guilherme Rodrigues Monteiro,
Vinícius Del Manto Cavalcante,
Prof. Dr. Marco Carlos Uchida,
Prof. Dr. Arturo Forner-Cordero,
Profa. Dra. Larissa Driemeier

Finalidade:

Trabalho de Conclusão de Curso

Local:

Escola Politécnica – USP

São Paulo
Outubro - 2022

Orientadores:

Profa. Dra. Larissa Driemeier,

Tel.: (011) 97979-4999

Currículo Plataforma Lattes: <http://lattes.cnpq.br/0367058546602662>

Prof. Dr. Rafael Traldi Moura,

Tel.: (011) 98123-8476

Currículo Plataforma Lattes: <http://lattes.cnpq.br/5846131329464125>

Pesquisadores:

Guilherme Rodrigues Monteiro

Tel.: (013) 99622-0053

Vinícius Del Manto Cavalcante

Tel.: (011) 94295-4098

Local de trabalho:

Departamento de Engenharia Mecatrônica e Sistemas Mecânicos –
Escola Politécnica da Universidade de São Paulo (USP)

Aspectos éticos da pesquisa

O Termo de Consentimento Livre e Esclarecido (TCLE) será entregue aos voluntários antes da coleta de dados. Os voluntários serão recrutados através de informativos divulgados via internet (redes sociais) pelos responsáveis da pesquisa. Nos informativos serão disponibilizadas informações de contato dos responsáveis (e-mail, facebook e telefone) para que os interessados em participar do estudo manifestem-se. Assim serão enviados o TCLE e os possíveis horários para realização da coleta de dados.

Os possíveis riscos incluem dor muscular de início tardio, perda de força muscular e de flexibilidade e um leve edema na musculatura alvo (músculos peitoral maior, tríceps braquial e deltoide) até 96 horas após a coleta. Possíveis desconfortos podem existir devido a filmagem direta do rosto e movimento dos membros superiores e utilização de maquiagem (lápis de olho preto) para marcação de alguns pontos na face. Durante a execução do exercício, auxiliares estarão acompanhando o movimento da barra guiada para, caso o voluntário entre em fadiga muscular, evitem a queda da barra sobre o mesmo.

Os benefícios incluem uma sessão de treino gratuita, disponibilização dos dados dos pesquisadores e esclarecimento dos resultados das análises e de eventuais dúvidas dos voluntários. Para os profissionais da saúde, que trabalham com fisioterapia, o estudo pretende treinar e validar uma nova forma de monitorar o exercício de força, para que isso os auxilie na prescrição das cargas dos exercícios, de acordo com a necessidade individual de seus pacientes, beneficiando indiretamente a população em geral.

Todas as imagens serão digitais e armazenadas em computador protegido com senha ou cloud do Google Drive disponibilizado pela instituição de ensino e protegido com acesso e senha. O sujeito terá a opção de consentir ou não o armazenamento dessas amostras (imagens e filmagens). As senhas serão de acesso dos pesquisadores e conhecidas apenas por estes e ninguém mais. Os dados serão mantidos armazenados dessa maneira protegida com senha apenas caso os voluntários permitam, e, em caso de pedido de exclusão, serão prontamente apagados. Somente os pesquisadores envolvidos com o projeto terão acesso aos dados, com fins estritamente de pesquisa; ou pesquisadores autorizados pela plataforma Brasil a realizar novas pesquisas na mesma área com os dados já obtidos e com o devido consentimento dos voluntários.

Resumo Informativo

Introdução: o exercício de força tem sido estudado nos mais diversos aspectos relacionados aos benefícios para a saúde. Para que este atue efetivamente nestes parâmetros é necessário que se adeque a carga de treinamento para cada indivíduo. A classificação do nível de esforço pela expressão facial pode ser um novo método que irá colaborar para a adequação da carga de treinamento. **Objetivo:** Coletar, com imagens e filmagens, as mudanças da expressão facial no decorrer da execução do exercício supino reto até a falha concêntrica, treinar e validar uma inteligência artificial com os dados coletados. **Metodologia:** Indivíduos de 18 a 50 anos, praticantes de musculação realizarão três séries com aproximadamente 10RM, e três minutos de intervalo entre essas, no exercício supino reto guiado (aparelho Smith Press). Durante a realização do exercício, uma câmera filmadora registrará as imagens do rosto para que, posteriormente, possam ser analisadas, classificadas e alimentem o aprendizado de máquina desejado. **Análise dos dados:** Os dados serão analisados no treinamento do machine learning por códigos classificadores SVM de kernel Gaussiano e também de rede neural convolucional, ambos métodos de Inteligência Artificial. **Resultados esperados:** O estudo pretende conseguir treinar e validar uma Inteligência Artificial classificadora do nível de esforço pela expressão facial. O classificador poderá agir como um indicador de fadiga muscular e um possível auxílio para monitoramento de carga de treino, esperando, portanto, um reprodutibilidade inter e intra pessoal e um padrão de comportamento da expressão facial durante o exercício de força.

Sumário

Aspectos éticos da pesquisa.....	3
Resumo Informativo.....	4
Introdução.....	6
Justificativa.....	7
Objetivos.....	7
Geral.....	7
Específicos:.....	7
Hipótese.....	7
Sujeitos e métodos.....	8
Estudo.....	8
Amostra.....	8
Avaliação da expressão facial.....	8
Exercício supino horizontal <i>smith press</i> (barra guiada).....	10
Protocolo da sessão de exercício.....	10
Coleta de dados.....	11
Filmagem da expressão facial durante a execução do exercício.....	11
Filmagem dos membros superiores durante a execução do exercício.....	11
Percepção subjetiva de esforço (PSE).....	12
Velocidade do movimento.....	12
Forma de análise dos resultados.....	12
Cronograma de execução.....	13
Referências bibliográficas.....	14

Introdução

Os benefícios da atividade física são amplamente estudados no meio acadêmico e uma das principais alterações positivas sobre o sistema músculo-esquelético é através do exercício de força. Este contribui para a manutenção ou melhora das capacidades funcionais através do aumento da força e potência muscular, sendo importante para o desempenho esportivo, melhora da estética e manutenção da saúde geral. Além dos benefícios, o treino da força é usado em sessões focadas em reabilitação, comandados por fisioterapeutas e terapeutas ocupacionais, em que o objetivo é o fortalecimento ou melhora do desempenho muscular.

Por este motivo, o presente estudo se propõe a analisar uma variável muito aparente e pouco destacada no treino de força: a expressão facial. Para que o treinamento físico promova adaptações fisiológicas significativas, como a hipertrofia e ganho de força, é necessário que se realize os exercícios até a falha concêntrica (ou próximo), esta pode ser acompanhada de uma sensação de dor muscular durante, algo que se relaciona a proximidade do esforço máximo (COOK et al. 1997). Ao realizar o esforço elevado, é natural que o indivíduo apresente uma alteração na sua expressão facial, que pode ser identificada e fragmentada em unidades de ação (UA). UA são bases anatômicas de unidades mínimas de ação muscular, das quais Ekman e Friesen definiram 46, cada uma correspondendo à atividade em um músculo ou grupos musculares faciais diferentes e produzem alterações características que podem ser identificadas por imagens (TOURETZKY, MOZER, HASSELMO, 1996). Conjuntos específicos dessas UA formam padrões de expressões faciais conhecidas, tais como medo, tristeza, raiva, felicidade e dor.

A percepção subjetiva de esforço possui diversas funções importantes como fornecer informações sobre dificuldade da tarefa, limitar o comportamento que afeta no consumo de energia e contribuir na percepção da capacidade de agir (PRESTON, WEGNER, 2009). A expressão facial de esforço, por sua vez, aumenta a importância dessa sensação pelo fato de revelá-la aos outros (MORREE, MARCORA, 2010). A ideia do presente estudo é aproveitar esta revelação espontânea do indivíduo, enquanto realiza exercício de força até a fadiga muscular, explorando as mudanças de expressão facial durante a execução de um exercício de força até a falha concêntrica para treinar Inteligência Artificial a reconhecer e classificar o nível de esforço do indivíduo pela expressão facial.

Justificativa

Os estudos que relacionam expressão facial e esforço são poucos (MORREE, MARCORA, 2010), mas a relevância desta relação pode ser muito importante, principalmente para os profissionais que estão envolvidos com a prática da musculação ou treinamento de força e para os profissionais da saúde que usam o exercício de força a fim de ajudar seus pacientes em tratamento. Este estudo buscará treinar e validar uma Inteligência Artificial a usar a expressão facial como uma variável para monitoramento do exercício, podendo oferecer o resultado a estes profissionais para que, ao acompanhar a mudança facial do paciente, seja possível afirmar se a intensidade e o volume estão adequados ao objetivo proposto. Deverá ser uma forma rápida e automática de se analisar a percepção de esforço do indivíduo em atividade.

Objetivos

Geral:

Verificar e documentar as mudanças da expressão facial no decorrer da execução do exercício supino reto até a falha concêntrica para realizar aprendizado de máquina com os dados coletados.

Específicos:

Coletar dados em imagens e vídeos da relação entre as mudanças da expressão facial e a velocidade e/ou o tempo de execução de cada repetição até o momento da fadiga;

Treinar algoritmos de Inteligência Artificial com os dados coletados para que consigam classificar automaticamente o nível de esforço de um indivíduo com base em sua expressão facial;

Verificar a reprodutibilidade inter e intraindividual da expressão facial dos voluntários enquanto realizam exercício de força.

Hipótese

A principal hipótese do trabalho é que a expressão facial pode ser um indicador de fadiga muscular e uma nova variável de monitoramento de carga. Espera-se, portanto, um comportamento padrão e uma reprodutibilidade inter e intraindividual das mudanças da expressão facial dos voluntários durante o exercício proposto.

Sujeitos e métodos

Estudo

O estudo é um estudo de caso que utilizará a metodologia quantitativa, com desenho experimental, testando o impacto da intervenção sobre as variáveis a serem avaliadas, controlando da melhor forma os fatores que podem ter algum tipo de intervenção sobre essas (Creswell, 2014).

A pesquisa será realizada em academias frequentadas pelos pesquisadores, cuja autorização e reserva para a utilização do espaço serão previamente providenciadas.

Amostra

A amostra será composta por praticantes de treinamento de força de qualquer gênero, entre 18 a 50 anos (não haverá, portanto, participação de menores de idade no estudo), que estão inseridos na prática há pelo menos um semestre e familiarizados com o exercício supino horizontal no aparelho *Smith Press*. Antes de iniciar a coleta, será esclarecido e entregue o Termo de Consentimento Livre e Esclarecido (TCLE) aos voluntários, o qual apresentará todos os possíveis riscos, desconfortos e benefícios da pesquisa. O voluntário só poderá participar do experimento depois de assinar este documento. Também estará presente no TCLE os direitos de imagem dos voluntários e sua opção de consentir ou não o armazenamento das amostras de imagens.

Os fatores de exclusão são quaisquer tipos de problemas musculares, articulares ou neurológicos nos membros superiores e possibilidade de alergia do produto utilizado para a marcação dos pontos na pele dos indivíduos.

Avaliação da expressão facial

O rosto do sujeito deverá estar limpo e sem cabelo sobre a superfície da face. Serão feitas marcações com lápis de olho da cor preta no rosto de cada voluntário em três pontos baseados nas UA, figura 1, características de dor aguda (WILLIAMS, 2002), pois esta expressão facial possui uma semelhança com a expressão de esforço (CRILE, 1916; FRIJDA, 2002; SALZEN, 2002 apud MORREE, MARCORA, 2010). Serão escolhidas, para este projeto, as seguintes U.A.:

Número da UA	Nome da UA	Músculos agonistas
4	abaixador de sobrancelha	depressor do supercílio, corrugador do Supercílio
6	levantador de bochechas	orbicular do olho
12	puxador de canto do lábio	zigomático maior

Logo após feitas as marcações na face do sujeito, será tirada uma foto com a câmera fotográfica, da marca Casio, modelo EX-Z29, enquanto o voluntário realiza a expressão facial de maior intensidade, apenas uma representação do esforço, orientado pelo responsável pelo projeto. Será feito isso a fim de tornar a avaliação o mais fiel possível, já que a aparência exata varia de indivíduo para indivíduo, dependendo da estrutura óssea, variações na musculatura facial, depósito de gordura, rugas permanentes, forma das feições, etc (EKMAN, FRIESEN, HAGER, 2002). Depois haverá a coleta de dados através da filmagem, para aprendizado de máquina a posteriori.

Exercício supino horizontal *smith press* (barra guiada)

O exercício supino horizontal no aparelho *smith press* consiste no afastamento e na aproximação de uma barra em relação ao tórax, estando o indivíduo em decúbito dorsal. A escolha deste exercício foi em função da sua grande popularidade entre os praticantes e atletas de força, fácil execução e posicionamento, pois nesse a cabeça fica relativamente estável durante a execução do movimento, uma vez que o presente grupo já realizou este tipo de coleta de dados com o exercício rosca direta, flexão do cotovelo, e conforme o sujeito se aproxima da fadiga, muitos tendem a movimentar-se com o intuito de terminar a tarefa, o que popularmente chamam de “roubar”, isso leva a dificuldade da análise de forma mais clara e detalhada.

O indivíduo deverá segurar a barra em uma posição confortável, em geral com as mãos posicionadas 10-20cm além da largura do ombro (KIM et al., 2002), será solicitado ao sujeito que mantenha a cabeça e o tronco sobre o banco, sem movimentá-los. Durante o exercício supino, o indivíduo realiza

movimentos nas articulações do ombro, escápulo-torácica e cotovelo. No momento do afastamento da barra, ação muscular concêntrica: adução horizontal dos ombros (músculos responsáveis: peitoral maior, deltóide parte clavicular e acromial - e coracobraquial), abdução das escápulas (músculos responsáveis: peitoral menor e serrátil anterior) e extensão dos cotovelos (músculos responsáveis: tríceps braquial e ancôneo). Durante a aproximação da barra, por sua vez, a ação muscular é excêntrica e portanto, utilizando os mesmos grupos musculares, são realizados os movimentos contrários aos da fase concêntrica: abdução horizontal dos ombros, adução das escápulas e flexão dos cotovelos. (MARCHETTI et al., 2010).

Protocolo da sessão de exercício

O protocolo do presente estudo consistirá em um dia de coleta para cada sujeito, no qual os voluntários realizarão três séries de aproximadamente dez repetições máximas (RM) com intervalo de três minutos entre as séries (i.e. 3 séries x 10RM, com intervalo de 3 minutos) do exercício supino horizontal, todas até a fadiga muscular. O peso para a realização das 10RM será determinada previamente, através de uma avaliação prévia, uma semana antes do dia da coleta de dados. Um dos pesquisadores sempre ficará próximo ao voluntário durante o exercício, a fim de garantir segurança no momento da fadiga, auxiliando a travar a barra e evitando que essa não caia ou permaceça sobre o tórax do voluntário por um longo período.

Será orientado aos voluntários que realizem o exercício com velocidade usual e costumeira no exercício supino, tanto na ação concêntrica e excêntrica, porém que nesta última não seja inferior a um segundo (1seg), ou seja mantendo a tensão e não relaxando a musculatura envolvida. É importante lembrar que os indivíduos não deverão realizar atividade de alta intensidade para os grupamentos musculares dos membros superiores 72 horas antes do dia de coleta de dados.

Coleta de dados

Filmagem da expressão facial durante a execução do exercício

Será utilizada uma filmadora digital (câmera 1) da marca Casio modelo EX-Z29, acoplada a um tripé, posicionada de modo que ela tenha visão de topo, captando imagens da face do voluntário, ver figura 2, de forma que o rosto deles fique centralizado e nítido.

Filmagem dos membros superiores durante a execução do exercício

Para realizar as filmagens dos movimentos no exercício supino, será utilizada uma filmadora digital (câmera 2) da marca *Nikon* modelo *AW100*, acoplada a um tripé, posicionada de modo que tenha uma visão frontal, captando o movimento dos membros superiores (mmss) dos voluntários, afim de analisar as mudanças da angulação nas articulações do cotovelo (extensão e flexão) e ombro (adução horizontal e abdução horizontal) durante a execução do exercício, ver figura 2.

Logo após o início da filmagem, antes que o voluntário comece a realizar as repetições, será dado um sinal luminoso com flash da marca *Nikon* e modelo *SpeedLight SB-27* e sonoro a fim de, posteriormente sincronizar as imagens com as da segunda câmera. Assim, será possível acompanhar em qual ângulo os indivíduos possuem mais dificuldade para a execução do movimento e a consequente ativação das UA do rosto nestes momentos críticos.

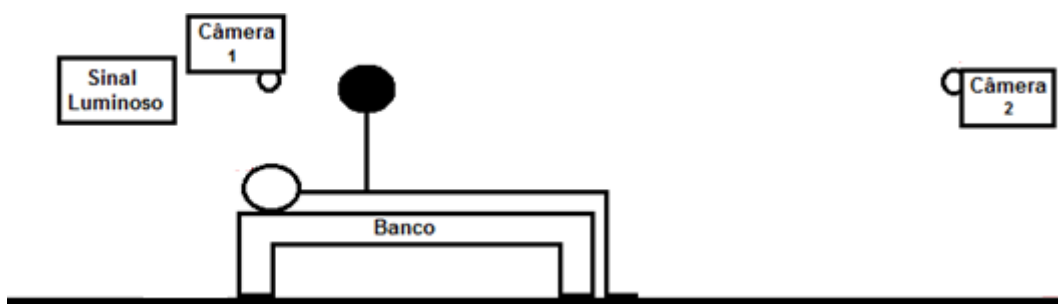


Figura 2. Apresentação da disposição das câmeras.

Percepção subjetiva de esforço (PSE)

Ao final de cada série e de ao fim da sessão de treino será questionado ao voluntário sobre a sensação de esforço utilizando a escala de percepção subjetiva de esforço (CR-10, adaptada de Borg por Foster et al. 2001). A escala varia de zero (nenhum esforço) a dez (esforço máximo) e inclui padrões verbais de percepção de esforço como leve, moderada e difícil, para valores intermediários.

Velocidade do movimento

A velocidade dos movimentos da barra do aparelho *smith press* serão medidos através do equipamento *Peak Power 4.0* da marca *Cefise*. O equipamento é composto por um sistema eletrônico e um software que mede a velocidade de deslocamento de um corpo, medida por um cronômetro de precisão ligado a um fio que, no caso, será acoplado à barra a ser utilizada. O sistema mede o tempo e monitora o deslocamento do ponto de fixação e terá

grande importância para o projeto, pois a partir dos dados gerados, será possível relacionar a cadência dos movimentos com as mudanças da expressão facial.

(<http://www.cefise.com.br/produtosdetalhes.php?Peak%20Power%204.0&id=39&produto=30>)

Forma de análise dos resultados

Os valores das intensidades avaliadas serão tabulados por repetição e por série de cada indivíduo. Com a classificação, as séries documentadas em vídeos serão separadas em imagens dos frames e agrupadas em pastas para cada nível de intensidade, onde não terão nenhuma informação de caráter pessoal do voluntário, como nome ou identificação. Os algoritmos consumirão os dados diretamente das pastas, onde poderão treinar para serem classificadores de nível de esforço e validar a classificação com base no nível da pasta de origem da imagem.

Os algoritmos utilizados serão de classificador SVM com kernel Gaussiano e de rede neural convolucional. O treinamento se dá de forma automática, onde apenas o código acessa os dados e, após treinados, eles não mantêm nenhum dado do indivíduo. Isto significa que mesmo se a Inteligência Artificial for implementada em dispositivos ou se for compartilhada, não é possível ter acesso a nenhuma imagem que as treinou, nem nada que remeta aos voluntários, nada disso permanece no algoritmo treinado.

Após o aprendizado de máquina, as classificações serão validadas considerando eficácia, ou seja, número de classificações corretas pelo número total de classificações realizadas pelo código. O cálculo será feito para cada nível de esforço, para considerar também se as expressões faciais de níveis específicos se tornam mais evidentes ou fáceis de serem classificadas corretamente.

Cronograma de execução

Tabela 3. . Atividades referentes ao ano de 2022.

	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
Revisão Bibliográfica	X	X	X	X	X	X	X		
Algoritmo SVM				X	X	X			
Algoritmo Rede Neural Convolucional				X	X	X			
Recrutamento dos voluntários para o experimento							X	X	
Aplicação do experimento								X	X
Análise dos resultados do experimento									X
Confecção da Monografia									X

* Início dos experimentos pode variar e depender de quando eles serão aprovados pelos devidos comitês de ética, e não iniciarão antes disso

Referências bibliográficas

COOK, DANE B. et al. Naturally occurring muscle pain during exercise: assessment and experimental evidence. **Medicine and science in sports and exercise**, v. 29, n. 8, p. 999-1012, 1997.

CRESWELL, John W. **Research design: Qualitative, quantitative, and mixed methods approaches**. Sage, 2013.

DE MORREE, Helma M.; MARCORA, Samuele M. The face of effort: frowning muscle activity reflects effort during a physical task. **Biological psychology**, v. 85, n. 3, p. 377-382, 2010.

DE OLIVEIRA RAMALHO, Gustavo Henrique Rodrigues et al. O teste de 1RM para predição da carga no treino de hipertrofia e sua relação com número máximo de repetições executadas. **CEP**, v. 25, p. 160, 2011.

Disponível em: <<http://unmaskingtheface.blogspot.com.br/2012/05/sistema-de-expressao-facial-facs-parte.html>>. Acesso em 02 jul. 2014.

Disponível em:

<<http://www.cefise.com.br/produtosdetalhes.php?Peak%20Power%204.0&id=39&produto=30>>. Acesso em: 03 jul. 2014.

EKMAN, P.; FRIESEN, W.; HAGER, J. Facial Action Coding Systems–The Manual. 2002.

FOSTER, Carl et al. A new approach to monitoring exercise training. **The Journal of Strength & Conditioning Research**, v. 15, n. 1, p. 109-115, 2001.

HOEGER, Werner WK et al. Relationship between repetitions and selected percentages of one repetition maximum: a comparison between untrained and trained males and females. **The Journal of Strength & Conditioning Research**, v. 4, n. 2, p. 47-54, 1990.

KIM, Paul S.; MAYHEW, JERRY L.; PETERSON, D. FRED. A modified YMCA bench press test as a predictor of 1 repetition maximum bench press strength. **The Journal of Strength & Conditioning Research**, v. 16, n. 3, p. 440-445, 2002.

MARCHETTI, Paulo Henrique et al. EXERCÍCIO SUPINO: uma breve revisão sobre os aspectos biomecânicos BENCH PRESS EXERCISE: a brief review in the biomechanical aspects. **Brazilian Journal of Sports and Exercise Research**, v. 1, n. 2, p. 135-142, 2010.

PRENDERGAST, Peter M. Anatomy of the Face and Neck. In: **Cosmetic Surgery**. Springer Berlin Heidelberg, 2013. p. 29-45.

PRESTON, Jesse; WEGNER, Daniel M. Elbow grease: when action feels like work. **Oxford handbook of human action**, p. 569-586, 2009.

TAN, Benedict. Manipulating resistance training program variables to optimize maximum strength in men: a review. **The Journal of Strength & Conditioning Research**, v. 13, n. 3, p. 289-304, 1999.

TOURETZKY, David S.; MOZER, Michael C.; HASSELMO, Michael E. (Ed.). **Advances in Neural Information Processing Systems 8: Proceedings of the 1995 Conference**. Mit Press, 1996.

WILLIAMS, Amanda C. de C. Facial expression of pain, empathy, evolution, and social learning. **Behavioral and brain sciences**, v. 25, n. 04, p. 475-480, 2002.

WINETT, Richard A.; CARPINELLI, Ralph N. Potential health-related benefits of resistance training. **Preventive medicine**, v. 33, n. 5, p. 503-513, 2001.

Apêndice II - TCLE

Arquivo do Termo de Consentimento Livre e Esclarecido submetido na Plataforma Brasil e dado aos voluntários do estudo.

Universidade de São Paulo

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

TÍTULO DA PESQUISA: IA PARA MEDIÇÃO DE ESCALA DE ESFORÇO EM FISIOTERAPIA

ORIENTADOR/PESQUISADOR RESPONSÁVEL: PROF. DR. RAFAEL TRALDI MOURA. PROFESSOR DOUTOR DO DEPARTAMENTO DE ENGENHARIA MECATRÔNICA E DE SISTEMAS MECÂNICOS, POLI/USP

ENDEREÇO PROFISSIONAL: ESCOLA POLITÉCNICA (POLI) – USP AV. PROFESSOR MELLO MORAES, 2231 - CIDADE UNIVERSITÁRIA, CEP: 05508-010 SÃO PAULO/SP. PRÉDIO DE ENGENHARIA MECÂNICA E NAVAL.

CONVITE: Convidamos você a participar desta pesquisa que visa criar uma Inteligência Artificial para medição de escala de dor através das mudanças da expressão facial no decorrer da execução do exercício. Você terá garantia de esclarecimentos antes, durante e após o término da pesquisa sobre todos os procedimentos além dos resultados que serão obtidos nesta pesquisa.

Justificativa: Propor a Inteligência Artificial como uma ferramenta adicional para monitorar o exercício de força, a fim de auxiliar, de forma mais prática e rápida, o controle da carga das sessões de fisioterapia. Não há base de dados disponíveis para treinamento de máquina com essa finalidade, mas a relevância deste classificador automático pode ser muito importante, já que se trata de uma reação espontânea do indivíduo enquanto realiza exercício de força até a fadiga muscular. Classificar automaticamente a alteração da expressão facial durante a execução do exercício poderia ajudar o fisioterapeuta ou terapeuta ocupacional a adequar a intensidade e volume do exercício, ou seja, uma melhor prescrição da sessão.

Objetivos: Coletar imagens das mudanças da expressão facial no decorrer da execução do exercício supino reto até a falha concêntrica, sua reprodutibilidade inter e intra pessoal, identificar uma possível relação entre as mudanças da expressão facial e a velocidade e/ou tempo de execução das repetições e, por fim, apresentar uma escala de percepção de expressão facial relacionada com esforço para treinar a Inteligência Artificial.

Procedimentos da Pesquisa: Você realizará três séries de aproximadamente 10 repetições máximas do exercício supino horizontal com barra guiada, três minutos de intervalo entre essas. Enquanto você realiza as repetições, serão filmados o seu rosto e o movimento de seus braços com duas câmeras. Além disso, seu rosto será marcado, com lápis de olho (cor preta), em alguns pontos para facilitar a avaliação das filmagens posteriormente.

É importante, portanto, que seu rosto esteja limpo e sem barba no dia da coleta de dados e que você não tenha realizado exercícios de alta intensidade há, no mínimo, 72h antes da coleta, para que não tenha risco de você estar com dores musculares e isto interferir no seu desempenho e/ou expressão facial.

Desconforto e Possíveis Riscos Associados à Pesquisa: Ao participar desta pesquisa, você poderá sentir dores ou desconforto musculares durante e após a sessão, resultando em um leve edema e diminuição da força na musculatura exercitada que poderá perdurar até 96h após a sessão de exercício. Talvez você sinta algum desconforto ao ser filmado, pois a filmagem é focada diretamente sobre o seu rosto e dos movimentos dos membros superiores durante a execução do exercício; ou ainda por conta da utilização de maquiagem (lápis de olho preto) para a marcação de alguns pontos no seu rosto.

Benefícios da Pesquisa: Os benefícios incluem uma sessão de treino gratuita, disponibilização dos dados dos pesquisadores e esclarecimento dos resultados das análises e de eventuais dúvidas suas. Para os profissionais da saúde, que trabalham com fisioterapia, o estudo pretende treinar e validar uma nova forma de monitorar o exercício de força, para que isso os auxilie na prescrição das cargas dos exercícios, de acordo com a necessidade individual de seus pacientes. Além disso, o estudo poderá facilitar uma nova linha de classificadores relacionando expressão facial com as mais diversas atividades físicas e populações (e.g. idosos, atletas, mulheres).

Esclarecimentos e Direitos: A qualquer momento, você poderá obter esclarecimentos sobre todos os procedimentos utilizados na pesquisa e nas formas de divulgação dos resultados. Terá também a liberdade e o direito de recusar a sua participação ou retirar seu consentimento em qualquer fase da pesquisa, bastando entrar em contato com o pesquisador. É importante ressaltar que tanto sujeitos como empresas envolvidas não receberão qualquer ressarcimento financeiro pela participação no estudo, tampouco poderão utilizá-lo como meio de propaganda de seus serviços. Você receberá também uma via deste TCLE. ***Caso você tenha alguma reclamação ou queira denunciar qualquer abuso ou improbidade desta pesquisa, entre em contato com o comitê de ética e pesquisa com seres humanos da USP, unidade EEFÉ:*** Rua da Reitoria, 374 – CEP 05508-220 – São Paulo – SP. Telefone: (11) 3091-3097 – e-mail: cep39@usp.br

Confidencialidade e Avaliação dos Registros: a sua identidade e de todos os voluntários serão mantidas em total sigilo por tempo indeterminado tanto pelo executor como pela instituição onde será realizada a pesquisa. O mesmo vale para qualquer dado sensível ou de caráter pessoal: nome, sobrenome, e-mail, telefone ou qualquer dado revelador de identidade ou possibilitador de contato. Os resultados dos procedimentos executados na pesquisa serão analisados e alocados em vídeos ou imagens e divulgados em palestras, conferências, periódico científico ou outra forma de divulgação que propicie o repasse dos conhecimentos para a sociedade e para autoridades normativas em saúde nacionais ou

internacionais, de acordo com as normas/leis legais regulatórias de proteção nacional ou internacional. Todas as imagens serão digitais e armazenadas em computador protegido com senha ou cloud do Google Drive disponibilizado pela instituição de ensino e protegido com acesso e senha. As senhas serão de acesso dos pesquisadores e conhecidas apenas por estes e ninguém mais. Os dados serão mantidos armazenados dessa maneira protegida com senha apenas se você permitir, e, em caso de pedido de exclusão, serão prontamente apagados. Somente os pesquisadores envolvidos com o projeto terão acesso a eles, com fins estritamente de pesquisa. Caso uma nova pesquisa deseje revisitar os dados classificados com o nível de esforço, o mesmo será feito apenas com seu consentimento e os dados sensíveis não serão abertos.

É importante lembrar que você tem a opção de consentir ou não com o armazenamento dessas amostras (imagens e filmagens).

Consentimento Pós-Informação

- ☐ Eu consinto com meu registro através das gravações em vídeo
- ☐ Não consinto com o meu registro através de gravações em vídeo

Eu, _____, portador (a) da Carteira de identidade nº _____ expedida pelo Órgão _____, por me considerar devidamente informado (a) e esclarecido (a) sobre o conteúdo deste termo e da pesquisa a ser desenvolvida, expresse livremente meu consentimento para inclusão como sujeito da pesquisa e armazenagem de meus registros para aprendizado de máquina que visa classificar automaticamente o nível de esforço pela expressão facial.

Fui informado que meu número de registro na pesquisa é _____ e recebi via desse documento por mim assinado.

Responsabilidade do Pesquisador:

Asseguro ter cumprido as exigências da resolução 466/2012 CNS/MS e complementares na elaboração do protocolo e na obtenção deste Termo de Consentimento Livre e Esclarecido. Asseguro, também, ter explicado e fornecido uma via deste documento ao participante. Informo que o estudo foi aprovado pelo CEP perante o qual o projeto foi apresentado. Comprometo-me a utilizar o material e os dados obtidos nesta pesquisa exclusivamente para as finalidades previstas neste documento ou conforme o consentimento dado pelo participante.

_____/_____/2022

Assinatura do Participante Voluntário

Data

_____/_____/2022

Assinatura do Pesquisador Responsável