

Reconhecimento de emoções: um estudo de caso utilizando biossensores

Juciano Cardoso, Carine G. Webber

¹Área de Conhecimento em Ciências Exatas e Engenharias – Universidade de Caxias do Sul (UCS)
Rua Francisco Getúlio Vargas, 1130 – 95.070-560 – Caxias do Sul – RS – Brasil

jcardoso3@ucs.br, cgwebber@ucs.br

Abstract. *The goal of this research is to develop a user independent emotion classifier to infer the emotions of happiness, anger, sadness and pleasure, by using biosensors. In order to achieve this goal, the electrodermal activity, peripheral temperature and blood volume pulse were collected from 15 subjects (10 males and 7 females). These signals were collected by a hardware built for this research. A software was developed to exhibit the emotional stimuli, manage the exposition time, transmit and store the data from the hardware. Also, the experiment allowed to apply tests to label data and capture pictures to be used to create labels when submitted to a facial emotion recognition system. As a preliminary result, the classifier achieved similar results compared to those available in literature, when concerning user independence data.*

Resumo. *Nesta pesquisa foi desenvolvido um classificador de emoções independente de usuário capaz de inferir as emoções de felicidade, raiva, tristeza e neutralidade, utilizando biossensores. Os sinais de atividade eletrodermal, temperatura periférica e volume de pulso sanguíneo foram coletados de 15 indivíduos, 10 homens e 7 mulheres. Estes sinais foram coletados através de um hardware construído para o propósito desta pesquisa. Um software foi desenvolvido para exibir os estímulos emocionais, gerenciar o tempo de exposição estímulos, transmitir e armazenar os dados do hardware de coleta, aplicar testes para fins de rotulação e capturar fotografias para a rotulação através de reconhecimento emocional por emoções faciais. Como resultado preliminar, o classificador obteve resultados similares aos obtidos por trabalhos relacionados na literatura, em casos independentes dos usuários.*

1. Introdução

As emoções tem influência em diversas funções essenciais, como o convívio social, o aprendizado e a percepção. Implicitamente os seres humanos conseguem reconhecer emoções em seus iguais. O reconhecimento de emoções é uma área que se propõe, através da detecção de padrões em respostas emocionais, reproduzir essa habilidade nata aos seres humanos [Picard 2000].

Algumas das áreas que se beneficiam do estudo do reconhecimento emocional são: a interface humano-computador [Picard et al. 2001], o acompanhamento médico e o cuidado de incapazes [Costa et al. 2019]. No âmbito da interface humano-computador, fazer com que computadores reconheçam e reajam a emoções que os usuários estão manifestando, pode fazer com que a interação entre humanos e máquinas seja mais natural,

simulando de maneira mais precisa uma interação humano-humano, assim fortalecendo o engajamento dos usuários [Picard 2000]. O reconhecimento de emoções pode detectar alterações no humor de pessoas que não conseguem ou tem dificuldade em se expressar, como idosos, crianças e portadores de doenças físicas e mentais. O reconhecimento de alterações emocionais neste público, pode por exemplo, prevenir e detectar surtos e crises [Costa et al. 2019].

Este trabalho teve como objetivo geral coletar dados fisiológicos e criar um classificador independente de usuário para a predição de emoções. A fim de desenvolver esse classificador, buscou-se criar um banco de dados de sinais fisiológicos rotulados com a emoção que o indivíduo manifestara durante a coleta daquele sinal. Este banco de dados foi posteriormente submetido a técnicas de aprendizado de máquina supervisionado.

Neste estudo os sinais foram coletados de um grupo de 15 voluntários. Foram desenvolvidos um hardware que agrega diferentes sensores e um software responsável pelo gerenciamento da sessão de coleta. As emoções que ao final deste estudo foram inferidas são: felicidade, tristeza, raiva e neutralidade.

2. Manifestações emocionais e emoções

Picard destaca diversas formas de manifestações emocionais ou modulações sensoriais (do inglês: *Sentic Motion*). Algumas delas são: expressões faciais, entonação vocal, reconhecimento de gestos e bio-sinais [Picard 2000]. O reconhecimento emocional através de expressões faciais e entonação vocal são largamente explorados, enquanto o reconhecimento através de bio-sinais, devido a sua grande variabilidade entre indivíduos, ainda é pouco explorado [Jerritta et al. 2011].

Os bio-sinais consistem em sinais fisiológicos originados no Sistema Nervoso Autônomo. Estes sinais não podem ser intencionalmente dissimulados (uma das vantagens dessa forma de modulação sensorial) [Picard 2000]. Bio-sinais são também independente de cultura, ao contrário de expressões faciais, por exemplo, onde pessoas de determinadas culturas podem ser menos expressivas do que outras.

Bio-sinais não são independentes de gênero e idade. Pessoas idosas podem manifestar sinais fisiológicos de maneira distinta de pessoas jovens e mulheres muitas vezes manifestam emoções de maneira mais intensa do que homens [Rani and Sarkar 2006] [Park et al. 2007].

Alguns bio-sinais utilizados no reconhecimento emocional são: o eletrocardiograma (ECG), a atividade eletrodermal (EDA ou GSR), a eletromiografia (EMG), a temperatura periférica (SKT), o pulso de volume sanguíneo (BVP) e a respiração (RESP). Sinais estes medidos através de biossensores [Haag et al. 2004].

Emoções podem ser mapeadas em três dimensões: excitação, valência e dominância. A excitação varia de pouco impactante a muito impactante e corresponde a intensidade com que uma emoção afetou um indivíduo. A valência varia de negativa a positiva, correspondendo a emoções que provocaram bem ou mal estar a um indivíduo [Lang and Bradley 2007].

A dominância por sua vez, corresponde a o quanto uma emoção abalou o controle do indivíduo, esta dimensão é pouco usada, tem baixa relevância e por isso não será utilizada neste estudo [Lang and Bradley 2007]. A Figura 1 mostra as emoções distribuídas

no plano de valência e excitação.



Figura 1. Distribuição das emoções em virtude de valência e excitação [Kim and André 2008].

Estímulos emocionais são utilizados em ambiente laboratorial para se provocar respostas emocionais. As principais formas destes estímulos são: imagens, música e filmes [Coan and Allen 2007]. Lang e sua equipe desenvolveram um banco de imagens para esse fim, chamado: IAPS (*International Affective Picture System*) [Lang et al. 2008].

3. Estado da arte

Picard, Vyzas, e Healey conduziram uma pesquisa com apenas um indivíduo, utilizaram como estímulo um banco próprio de imagens personalizadas e mensuraram EMG, BVP, EDA e RESP. Extraíram as seguintes características estatísticas: média e desvio padrão dos sinais brutos, média dos valores absolutos das primeiras e segundas diferenças do sinal bruto e do sinal normalizado. As características foram reduzidas utilizando-se o método de seleção sequencial para frente (do inglês: *Sequential Forward Selection* ou *SFS*) [Picard et al. 2001]. O estudo de Picard, Vyzas, e Healey tinha o objetivo de identificar as emoções de raiva, neutralidade, luto, amor platônico, amor romântico, alegria e admiração. Foi utilizado um classificador baseado em máquina de vetores de suporte (do inglês: *Support Vector Machine* ou *SVM*). Obteve-se 81% de acurácia, como apenas um indivíduo foi utilizado esse resultado é considerado dependente de usuário [Picard et al. 2001].

Haag, Gononzy, Schaich e Williams, não especificaram o número de indivíduos em seu estudo. Utilizaram como estímulo o banco de imagens IAPS e mensuraram EMG, ECG, EDA, SKT, BVP e RESP. A média móvel (do inglês: *Running mean*) e o desvio padrão móvel (do inglês: *Running standard deviation*) foram as características utilizadas. Buscou-se classificar os valores para valência e excitação utilizando para isso uma rede neural. A acurácia obtida foi de 89,7% para valência e 63,7% para excitação, ambos dependente de usuário [Haag et al. 2004].

Rigas, Katsis, Ganiatsas e Fotiadis conduziram uma pesquisa envolvendo 9 indivíduos. Utilizaram a base IAPS como estímulo e coletaram os sinais de EMG, ECG, EDA e RESP. Após o processo de redução as características utilizadas foram: frequência respiratória, frequência cardíaca, desvio padrão do masseter direito, média dos valores absolutos da primeira diferença do sinal EDA, desvio padrão do músculo esquerdo da testa, desvio padrão do músculo direito da testa, desvio padrão do músculo direito nasal e desvio padrão do músculo esquerdo nasal. Buscavam identificar as emoções de alegria, nojo e medo. Foram utilizados dois classificadores, um baseado em *k*-vizinhos próximos (*K-Nearest Neighbors* ou *KNN*) e outro baseado no algoritmo de *Random forest*. O algoritmo *KNN* obteve uma acurácia de 62,7% e o baseado em *Random forest* teve uma acurácia de 62,4%, ambos independente de usuário [Rigas et al. 2007].

Costa, Rincon, Carrascosa, Julian e Novais utilizando 20 indivíduos, e a base de imagens IAPS como estímulo, coletaram os sinais de BVP, EDA e SKT. Foram extraídas as mesmas características estatísticas que Picard, Vyzas e Healey [Picard et al. 2001] utilizaram. Buscou-se inferir as emoções de raiva, satisfação, nojo, medo, alegria, neutralidade, tristeza e surpresa. Uma rede neural foi usada como classificador e foi obtida uma acurácia de 80% dependente de usuário [Costa et al. 2019].

Kim e André utilizaram três indivíduos em sua coleta onde estes foram expostos a um conjunto de músicas por eles selecionado. O sinais de EMG, ECG, SKT e RESP foram capturados, estes sinais passaram por codificações de sub-banda (do inglês: *Sub-band*) e forem extraídas características estatísticas e baseadas em energia [Kim and André 2008]. Kim e André visavam inferir as emoções de alegria, raiva, tristeza e prazer. Um classificador baseado em análise discriminante linear foi utilizado e foi obtida uma acurácia de 95% dependente de usuário e 70% independente de usuário [Kim and André 2008].

Zhu utilizou o sinal de EMG do conjunto de dados do estudo de Kim e André [Kim and André 2008], do qual extraiu características a partir do processo de transformação de ondas de Daubechies. Zhu buscou identificar as emoções de alegria, raiva, tristeza e neutralidade utilizando um classificador baseado em rede neural de retropropagação (do inglês: *Backpropagation neural network*). Obteve uma acurácia de 82,2% dependente de usuário [Zhu 2010].

Hönig, Wagner, Batliner e Nöth utilizaram dois bancos de dados, um proveniente do estudo de Kim e André [Kim and André 2008] e outro do estudo de Hönig [Hoenig et al. 2007], que corresponde a um total de informações de 25 indivíduos. Os sinais de ECG, EMG, EDA e RESP foram escolhidos nesses bancos de dados. Características baseadas em movimentação e uma variação destas, chamadas de características "deslizantes" (do inglês: *Sliding Features*) foram extraídas. Buscou-se identificar as emoções de alegria, tristeza, raiva e neutralidade. Um classificador baseado em análise discriminante linear foi utilizado e obteve-se uma acurácia de 83,4% dependente de usuário [Hönig et al. 2009].

Kim, Bang e Kim construíram dois conjuntos de dados um com 125 indivíduos outro com 50, proveniente da coleta de dados em crianças de 5 a 8 anos de idade. Utilizaram um conjunto de 4 estímulos para elicitção emocional. Este conjunto era composto de dois estímulos visuais, um em forma de iluminação e outro em forma de imagem, um estímulo sonoro na forma de um som de fundo e um estímulo cognitivo que era uma

narração com uma voz apelativa. Os sinais de ECG, EDA e SKT foram coletados. Foram extraídas dos sinais características estatísticas e baseadas em altas e baixas frequências. O objetivo de reconhecer as emoções de tristeza, raiva, estresse e surpresa. Utilizou-se um classificador baseado em SVM e obteve-se uma acurácia de 78,4% para 3 emoções, dependente de usuário e 61,8% para 4 emoções, independente de usuário [Kim et al. 2004].

4. Metodologia

A metodologia para se desenvolver este tipo de classificador geralmente passa pelas seguintes etapas: coleta/rotulação, pré-processamento, extração de características, seleção de características (ou redução de dimensionalidade) e classificação [Jerritta et al. 2011].

4.1. Coleta e rotulação

O processo de coleta consiste em expor o indivíduo a um estímulo emocional e gravar seus sinais enquanto estimulado [Haag et al. 2004]. Neste estudo preferiu-se utilizar estímulos visuais na forma de imagens. Imagens são uma forma de estímulo conveniente para ser utilizada em laboratório, que permite uma exposição controlada em quesitos de tempo e intensidade [Lang and Bradley 2007].

O banco de imagens IAPS [Lang et al. 2008] foi utilizado. Este banco de imagens foi testado com indivíduos de diferentes idades, gêneros e culturas, mostrando-se independente em todos estes aspectos. Por sua fácil aplicação em ambiente laboratorial este banco de imagem foi utilizado como estímulo neste trabalho [Lang et al. 2008]. O acervo do IAPS corresponde a mais de 1000 imagens com diferentes atividades humanas. São disponibilizados juntos ao banco de imagens relatórios contendo a média e desvio padrão em termos de valência, excitação e dominância para cada imagem.

Essa classificação foi feita utilizando-se um teste de auto-avaliação, chamado SAM (*Self-Assessment Manikin*). Este teste consiste em um conjunto de pictogramas onde o indivíduo após ser exposto a um estímulo emocional, classifica sua reação emocional em uma escala hedônica de 9 pontos para cada dimensão [Bradley and Lang 1994]. No contexto deste estudo, o SAM foi aplicado aos voluntários para fins de rotulação dos dados obtidos.

Para se fazer a seleção das imagens utilizadas neste estudo utilizou-se o relatório do IAPS que trazia os valores para todos os indivíduos. As sessões de coletas seguem o modelo de Haag, que consiste em expor as imagens por 15 segundos ao indivíduo lendo os sinais do indivíduo a uma frequência de 50 Hz. A ordem das imagens também seguem o modelo de Haag, exibindo os estímulos em ordem crescente de excitação intercalando com imagens neutras [Haag et al. 2004].

Utilizando-se os valores das médias no relatório do IAPS, os estímulos foram separados em 7 grupos descritos na Tabela 1. Para cada sessão os estímulos são exibidos na seguinte ordem: 10 imagens neutras, 5 imagens de valência positiva e excitação baixa, 10 imagens neutras, 5 imagens de valência negativa e excitação baixa, 10 imagens neutras, 5 imagens de valência positiva e excitação neutra, 10 imagens neutras, 5 imagens de valência negativa e excitação neutra, 10 imagens neutras, 5 imagens de valência positiva e excitação alta, 10 imagens neutras, 5 imagens de valência negativa e excitação alta [Haag et al. 2004].

Tabela 1. Subconjuntos de estímulos.

	Valência média	Excitação média
Imagens neutras	[4, 6)	[1, 4)
Imagens com baixa excitação e valência positiva	[6, 10)	[1, 4)
Imagens com excitação neutra e valência positiva	[6, 10)	[5, 6)
Imagens com alta excitação e valência positiva	[6, 10)	[6, 10)
Imagens com baixa excitação e valência negativa	[1, 4)	[1, 4)
Imagens com excitação neutra e valência negativa	[1, 4)	[5, 6)
Imagens com alta excitação e valência negativa	[1, 4)	[6, 10)

As imagens de excitação alta devem ser exibidas no final da sessão para que seu impacto emocional não afete o indivíduo ao ponto dele não se recuperar a tempo do próximo estímulo, podendo gerar algum tipo de ruído nos sinais [Haag et al. 2004]. A intercalação de imagens neutras tem o objetivo de neutralizar os sinais [Haag et al. 2004].

Os sinais fisiológicos podem variar para o mesmo indivíduo em dias diferentes e por isso é necessário coletar de um mesmo indivíduo seus sinais em dias distintos [Picard et al. 2001]. Um total de 15 voluntários participaram deste estudo, 10 homens e 5 mulheres. Cada indivíduo teve seus sinais coletados durante 3 sessões, com intervalos médios de 7 dias entre as sessões.

As imagens separadas para exibição compreendem as que se adequam aos intervalos citados na Tabela 1, ordenadas em ordem decrescente de desvio padrão de excitação dentro de cada intervalo, com o intuito de selecionar as imagens com menos variação entre indivíduos.

Preferiu-se neste estudo utilizar sensores não intrusivos e de baixo custo. Foi utilizado um sensor de fotopletismografia (do inglês: *photoplethysmography* ou PPG) que mensura o pulso de volume sanguíneo (do inglês: *blood volume pulse* ou BVP), um sensor para atividade eletrodermal (do inglês: *electrodermal activity* ou EDA) e um sensor de temperatura periférica (do inglês: *peripheral temperature* ou SKT). Uma descrição mais detalhada sobre o coletor e seus sensores é fornecida na Seção 4.1.1.

A fim de facilitar o processo de classificação, no final da coleta espera-se obter um banco de dados de sinais fisiológicos já rotulados com a emoção que cada amostra corresponde. Apenas as quatro principais emoções, mapeadas nos quatro quadrantes do plano de valência e excitação foram consideradas: raiva, felicidade, tristeza e neutralidade. Este mapeamento é ilustrado na Figura 1.

O processo de coleta foi desenvolvido para que cada amostra de sinal pudesse ser rotulada com três rótulos, obtidos de maneiras distintas, um servindo de contra-prova para o outro:

1. **Rótulo extraído do IAPS:** consiste nos resultados de médias de valência e excitação do estímulo utilizado, este valor obtido do relatório proveniente do IAPS. Estes valores são mapeados no modelo de valência e excitação ilustrado pela Figura 2 [Lang et al. 2008].
2. **Rótulo extraído do SAM:** foi atribuído a cada amostra de sinais foi o resultado do SAM. Após cada exibição de um estímulo, os indivíduos se auto-

classificavam em termos de valência e excitação como foi sua reação a aquela imagem [Bradley and Lang 1994]. O resultado deste teste também foi mapeado no modelo ilustrado pela Figura 2 para se obter a emoção que correspondia.

3. **Rótulo extraído da *Microsoft Emotion API***: este rótulo foi desenvolvido baseado na metodologia de Costa [Costa et al. 2019]. Durante o processo de exposição uma fotografia do rosto do indivíduo é capturada a cada segundo. Estas fotografias foram posteriormente submetidas a *Microsoft Emotion API* ¹.

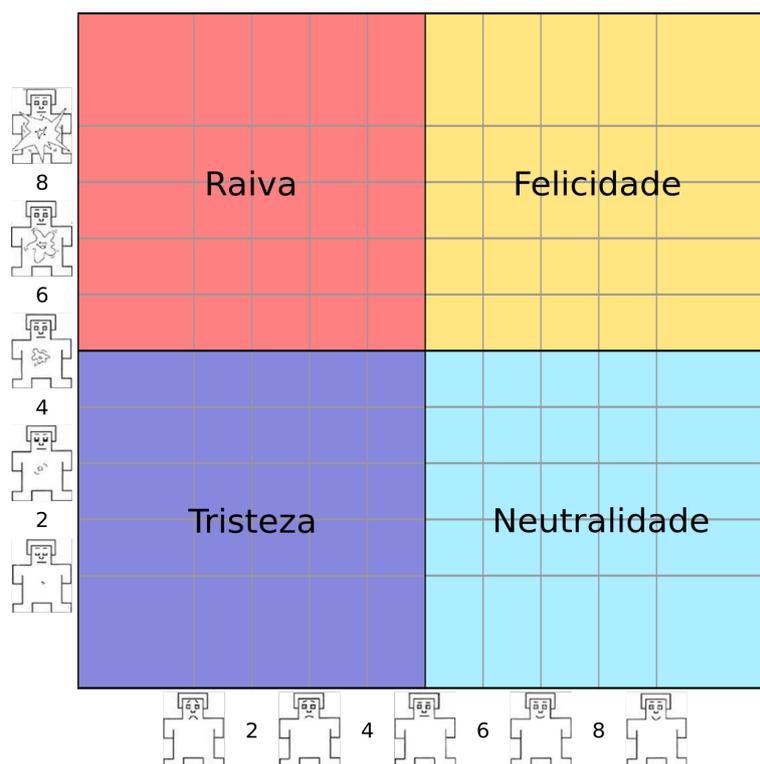


Figura 2. Mapeamento do SAM em valência e excitação. Fonte: adaptado de [Lang 1995].

A *Microsoft Emotion API* avalia cada rosto presente em uma imagem em oito emoções: raiva, neutralidade, nojo, medo, felicidade, contentamento, tristeza e surpresa. A API dá um valor para cada emoção, quanto maior esse valor, maior a probabilidade de ser aquela a emoção que está sendo manifestada. A emoção escolhida como representante deste rótulo corresponda a emoção que obteve a maior probabilidade em média para todas as fotografias capturadas durante a exposição da imagem de estímulo [Costa et al. 2019].

Como ilustra a Figura 3, neste estudo o processo de coleta consiste em: exibir um estímulo, coletar os sinais fisiológicos e fotografar o rosto do indivíduo enquanto isso, após a exibição aplicar o SAM. Repetir essas etapas até que o conjunto de imagens para a sessão chegue ao fim.

¹<https://azure.microsoft.com/en-us/services/cognitive-services/emotion/>

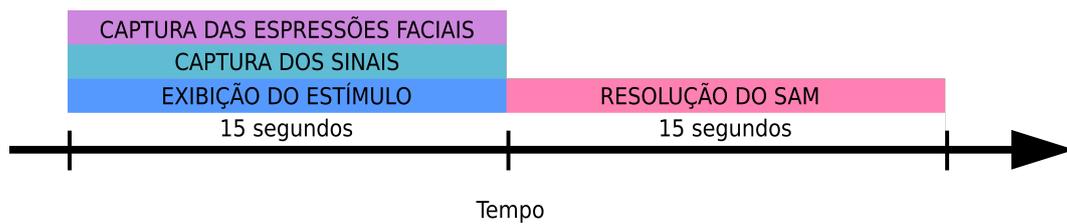


Figura 3. Processo de coleta.

Um software que integra o coletor, a exibição do estímulo, a captura do rosto do participante e a aplicação do SAM foi desenvolvido ². Este software recebe como parâmetros, o conjunto de imagens usadas como estímulo, o identificador do indivíduo, o identificador da sessão, o tempo de exibição do estímulo, o tempo de captura das fotografias e a porta serial onde o coletor esta conectado.

O programa opera com três processos: um processo principal e dois processos auxiliares responsáveis pela leitura do coletor e a captura das fotografias. Os processos auxiliares operam apenas durante a exibição de um estímulo e são suspensos no restante do tempo.

O processo principal é responsável pelo disparo dos eventos que suspendem e liberam os processos auxiliares além da exibição do estímulo, da aplicação do SAM e por serializar dos dados obtidos. Os dados obtidos são salvos em banco de dados relacional *SQLite* ³ em uma tabela cuja estrutura é apresentada pela Tabela 2.

Tabela 2. Estrutura da tabela de sinais.

Sinais
id: Integer (PK)
id_subject: Integer
id_session: Integer
id_collect: Integer
date_time: DateTime
ppg_signal: Float
eda_signal: Float
skt_signal: Float
sam_arousal: Integer
sam_valence: Integer
id_iaps: Integer

O banco de dados no fim do processo de coleta não possui os rótulos, mas todos os dados necessários para a rotulação já estão preparados. Os rótulos são adicionados na etapa de extração de características, descrita na Seção 4.2.

²<https://github.com/JucianoC/emotion-biosignal-collector>

³<https://www.sqlite.org/index.html>

4.1.1. Coletor

Diversas pesquisas relacionadas utilizam sensores intrusivos ao usuário, como eletrocardiograma e eletromiograma. Nesta pesquisa buscou-se utilizar sensores que fossem de fácil instalação, não intrusivos aos voluntários e possuísem baixo custo.

No estudo de Costa, Rincon, Carrascosa, Julian e Novais, eles construíram um coletor que batizaram de *Emotional Smart Wristband* (ESW). Este coletor utiliza os sensores de fotopletismografia, temperatura periférica e atividade eletrodermal [Costa et al. 2019]. O coletor utilizado neste estudo foi baseado neste modelo.

Utilizou-se um sensor de temperatura periférica desenvolvido pela fabricante *Melexis* modelo *MLX90615*⁴, que utiliza luz infra-vermelho para ler temperaturas de objetos e de ambientes. O *MLX90615* possui uma resolução de 0.02 °C e é capaz de ler temperaturas de -20 °C a +85 °C. O sensor foi posicionado a uma distância média de 5 centímetros da mão não dominante do voluntário. Em virtude de apenas a temperatura de objeto, ou seja da mão do indivíduo, ser relevante neste estudo, apenas o valor para temperatura de objeto foi armazenado. O valor lido por este sensor já é convertido internamente para graus Celsius.

O sensor utilizado para a leitura de atividade eletrodermal é desenvolvido pela fabricante *Grove* modelo *Grove-GSR v1.0*⁵. Este sensor é capaz de mensurar a voltagem de 0 a 1023 V. Sua instalação consiste em fixar o eletrodo positivo na parte inferior da falange proximal do dedo polegar e seu eletrodo negativo na parte inferior da falange média do dedo médio da mão não dominante, conforme indica o fabricante.

Para evitar ruídos neste sinal provenientes das mãos do indivíduo estarem em condições de umidade alteradas, todos os voluntários aplicaram uma solução de álcool gel em suas mãos antes da coleta. A utilização de soluções condutividade nos eletrodos não era especificada pelo fabricante e em seus exemplos em vídeo nada era aplicado, por essa razão soluções deste tipo não foram utilizadas.

Inicialmente buscou-se utilizar para fotopletismografia o mesmo sensor utilizado por Costa, Rincon, Carrascosa, Julian e Novais, fabricado pela *Pulse Sensor*⁶. Todavia, os exemplares adquiridos apresentaram mal funcionamento com leituras de valores demasiadamente elevados, além de, por não terem uma estrutura de fixação nos dedos apresentarem distorções e ruídos frequentes.

A fim de reduzir as chances das leituras possuírem ruídos provenientes do mal funcionamento deste sensor, este foi substituído por um outro sensor desenvolvido pela *Embedded Lab* modelo *Easy Pulse v1.1*⁷. Este sensor possui um encaixe em silicone para o dedo do usuário o que reduz a chance de ruídos por movimentação. Neste encaixe está disposto na parte superior um LED e na parte inferior um receptor infra-vermelho, responsáveis pela emissão da luz no tecido e leitura desta após transpassá-lo.

⁴<https://www.melexis.com/-/media/files/documents/datasheets/mlx90615-datasheet-melexis.pdf>

⁵https://github.com/SeeedDocument/Grove-GSR_Sensor/raw/master/res/Grove-GSR_Sensor_WiKi.pdf

⁶<https://pulsesensor.com/>

⁷<http://embedded-lab.com/blog/easy-pulse-version-1-1-sensor-overview-part-1/>

Testes foram feitos com todos os dedos da mão não dominante utilizando o software fornecido pelo fabricante. Cruzou-se a informação de frequência cardíaca fornecida por este software com a de um sensor de pulsos para atividades físicas. O polegar foi o dedo que apresentou os melhores resultados e conseqüentemente foi o dedo utilizado.

A Figura 4 apresenta a disposição e interligação dos componentes do coletor. Além dos sensores, foi utilizado uma pequena placa de ensaio para se interligar os componentes e um microcontrolador *Arduino UNO*⁸ para fazer a alimentação e leitura dos sensores.

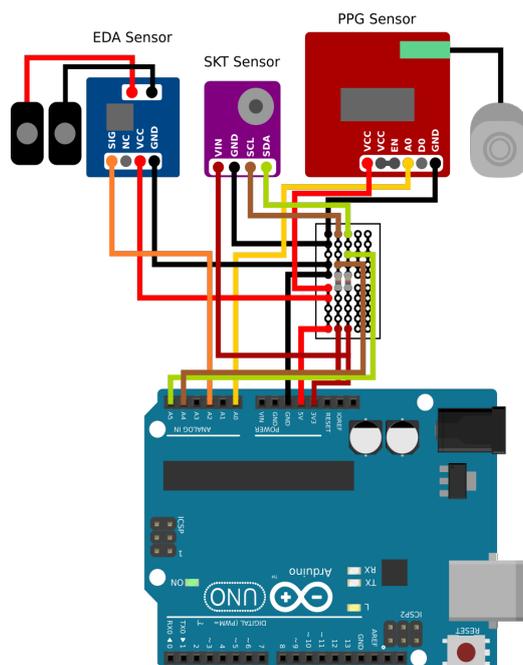


Figura 4. Disposição dos componentes do coletor.

Todos os sensores utilizados possuíam bibliotecas compatíveis com o *Arduino*. Para a integração entre sensores e *Arduino* foi desenvolvido um software⁹ responsável pela captura dos sinais e envio destas para a porta serial que é lida no software de exibição e coleta.

4.2. Pré-processamento e extração de características

Para cada estímulo exibido a um indivíduo, 6 características estatísticas foram extraídas de cada sinal, mesmas utilizadas por Picard, Vyzas e Harley [Picard et al. 2001].

Durante a exibição de uma estímulo pode-se haver diferenças no número total de amostras de sinais extraídas dos sensores. Picard, Vyzas e Harley indicam que as características devem ser extraídas de um conjunto de amostras de tamanho fixo [Picard et al. 2001]. A fim de corrigir eventuais diferenças no número de amostras de sinais obtidas durante cada exibição de estímulos, as amostras foram truncadas em um número que corresponde a aproximadamente 13 segundos de exibição de um estímulo.

⁸<https://store.arduino.cc/usa/arduino-uno-rev3>

⁹https://github.com/JucianoC/emotion-biosignal-collector/blob/master/arduino_collector/arduino_collector.ino

A junção de todas as amostras de um sinal para uma exibição de estímulo forma o vetor $\mathbf{x} = [x_1, x_2, \dots, x_N]$.

O sinal normalizado $\widetilde{\mathbf{x}}_n$ é um pré-processamento utilizado no cálculo de algumas características e é definido por

$$\widetilde{\mathbf{x}}_n = \frac{x_n - \mu_{\mathbf{x}}}{\sigma_{\mathbf{x}}} \quad (1)$$

, onde $\mu_{\mathbf{x}}$ e $\sigma_{\mathbf{x}}$ representam a média e o desvio padrão do sinal bruto, respectivamente:

$$\mu_{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N x_n \quad (2)$$

$$\sigma_{\mathbf{x}} = \left(\frac{1}{N-1} \sum_{n=1}^N (x_n - \mu_{\mathbf{x}})^2 \right)^{\frac{1}{2}} \quad (3)$$

A média dos valores absolutos das primeiras diferenças do sinal bruto é expressa por:

$$\delta_{\mathbf{x}} = \frac{1}{N-1} \sum_{n=1}^{N-1} |x_{n+1} - x_n| \quad (4)$$

A média dos valores absolutos das primeiras diferenças do sinal normalizado é dado por:

$$\widetilde{\delta}_{\mathbf{x}} = \frac{1}{N-1} \sum_{n=1}^{N-1} |\widetilde{x}_{n+1} - \widetilde{x}_n| = \frac{\delta_{\mathbf{x}}}{\sigma_{\mathbf{x}}} \quad (5)$$

A média dos valores absolutos das segundas diferenças do sinal bruto:

$$\gamma_{\mathbf{x}} = \frac{1}{N-2} \sum_{n=1}^{N-2} |x_{n+2} - x_n| \quad (6)$$

E por fim, a média dos valores absolutos das segundas diferenças do sinal normalizado:

$$\widetilde{\gamma}_{\mathbf{x}} = \frac{1}{N-2} \sum_{n=1}^{N-2} |\widetilde{x}_{n+2} - \widetilde{x}_n| = \frac{\gamma_{\mathbf{x}}}{\sigma_{\mathbf{x}}} \quad (7)$$

Um conjunto de dados contendo as características extraídas e seus respectivos rótulos é obtido a partir deste processo, a estrutura deste conjunto é representado na Tabela 3.

Tabela 3. Formato do conjunto de dados de características rotuladas.

Características + rótulos
ppg_mean: Float
ppg_stdev: Float
ppg_mean_first_diff_raw: Float
ppg_mean_first_diff_normalized: Float
ppg_mean_second_diff_raw: Float
ppg_mean_second_diff_normalized: Float
eda_mean: Float
eda_stdev: Float
eda_mean_first_diff_raw: Float
eda_mean_first_diff_normalized: Float
eda_mean_second_diff_raw: Float
eda_mean_second_diff_normalized: Float
skt_mean: Float
skt_stdev: Float
skt_mean_first_diff_raw: Float
skt_mean_first_diff_normalized: Float
skt_mean_second_diff_raw: Float
skt_mean_second_diff_normalized: Float
emotion_label_iaps: String
emotion_label_eapi: String
emotion_label_sam: String
label: String

Nos resultados provenientes da *Microsoft Emotion API* não tivemos nenhuma imagem facial mapeada para nojo e medo. Contentamento foi mapeado nas quatro emoções que buscamos identificar como neutralidade.

A emoção que aparece em ao menos dois dos três rótulos (*emotion_label_iaps*, *emotion_label_eapi* e *emotion_label_sam*) foi atribuída a coluna *label* e esta coluna foi usada como rótulo no aprendizado supervisionado do classificador. 1403 conjuntos de amostras de sinal tiveram ao menos duas emoções iguais em dois rótulos. Os números de amostras mapeadas para cada emoção é descrita pela Tabela 4.

Tabela 4. Número de amostras para cada emoção.

Emoção	Número de amostras
Felicidade	75
Raiva	82
Tristeza	196
Neutralidade	1050
Total	1403

4.3. Classificação e redução de dimensionalidade

A fim de possibilitar o processo de classificação a partir de um conjunto de dados restrito optou-se por reduzir a dimensionalidade utilizando PCA (*Principal Component Analysis*) [Tipping and Bishop 1999]. Essa prática é extensivamente aplicada em na redução de dimensionalidade, como mostra [Johnstone and Paul 2018] e [Xiaoxiao Wu et al. 2018].

O conjunto de dados de características foi redimensionado assim para 10 dimensões correspondentes aos componentes principais. Testes foram efetuados variando o número de componentes e observou-se que 10 componentes representavam 90% da variabilidade, abaixo deste valor a representatividade de variabilidade caía de forma significativa.

Uma rede neural perceptron de multicamadas foi usada para a classificação, uma vez que esse tipo de classificador apresentou resultados satisfatórios em outros estudos com o mesmo conjunto de sinais e características [Costa et al. 2019]. A implementação deste classificador¹⁰ feito pela biblioteca *scikit-learn* [Pedregosa et al. 2011] foi utilizada.

Dois métodos foram aplicados a fim corrigir o desbalanceamento no número de amostras que ocasiona impossibilidade de inferência de classes minoritárias. O primeiro chamado super amostragem aleatória (do inglês: *Random Oversampling*), consiste em, no conjunto de treinamento, duplicar amostras aleatórias de classes minoritárias até que o total de amostras nas classes minoritárias seja igual ao da classe majoritária [Tantithamthavorn et al. 2018].

O segundo método para balanceamento utilizado é chamado de aprendizado sensível a custo (do inglês: *Cost sensitive learning*). Este método tem como princípio aumentar o custo para classificações incorretas, o que faz com que quando o classificador prediz uma classe minoritária como majoritária ele sofra uma penalidade mais severa. Este método faz com que o classificador ao invés de minimizar a taxa de erro de classificação minimize os custos de classificações incorretas [Prati et al. 2009]. A implementação de perceptron multicamadas do *scikit-learn* não possui suporte para o aprendizado sensível a custo, o pacote *Keras*¹¹ foi utilizado nesta implementação.

Ambos os métodos foram submetidos a validações cruzadas utilizando o método *K-fold*. Testes foram realizados variando o número de partições e neurônios na camada intermediária. A função de ativação da camada oculta que melhor performou e foi utilizada em todos os casos foi a tangente hiperbólica ($f(x) = \tanh(x)$) [Karlik and Olgac 2011].

5. Resultados

Os resultados apresentados a seguir foram obtidos através de uma validação cruzada de 10 partições, ou seja, o conjunto de dados foi dividido em 10 partes, mantendo a proporção de classes em cada parte. O número de amostras para cada classe em cada partição é apresentado pela Tabela 5.

¹⁰https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

¹¹<https://keras.io/>

Tabela 5. Partições utilizadas na validação cruzada

Partição \ Classe	Felicidade	Raiva	Tristeza	Neutralidade
#1	8	9	20	105
#2	8	9	20	105
#3	8	8	20	105
#4	8	8	20	105
#5	8	8	20	105
#6	7	8	20	105
#7	7	8	19	105
#8	7	8	19	105
#9	7	8	19	105
#10	7	8	19	105

A Tabela 6 apresenta a matriz de confusão para o classificador que obteve os melhores resultados utilizando o método de balanceamento sensível a custo. O classificador foi testado com cada uma das partes e treinado com as outras nove.

O custo para a inferência de uma classe minoritária utilizado foi proporcional ao número de amostras em relação a classe majoritária.

O número de neurônios nas camadas ocultas foi de 21 na primeira camada e 18 na segunda. A acurácia apresentou resultado de 66% e a precisão média de 70,1%.

Tabela 6. Balanceamento sensível a custo - Matriz de confusão

Real \ Predito	Felicidade	Raiva	Tristeza	Neutralidade	Σ
Felicidade	30	7	11	27	75
Raiva	5	34	6	37	82
Tristeza	13	7	84	92	196
Neutralidade	50	57	162	781	1050
Σ	98	105	263	937	1403

A gráfico apresentado na Figura 5 ilustra a taxa de verdadeiros positivos e a precisão para cada classe. Em média a proporção de amostras corretamente classificadas foi de 40% para as classes minoritárias.

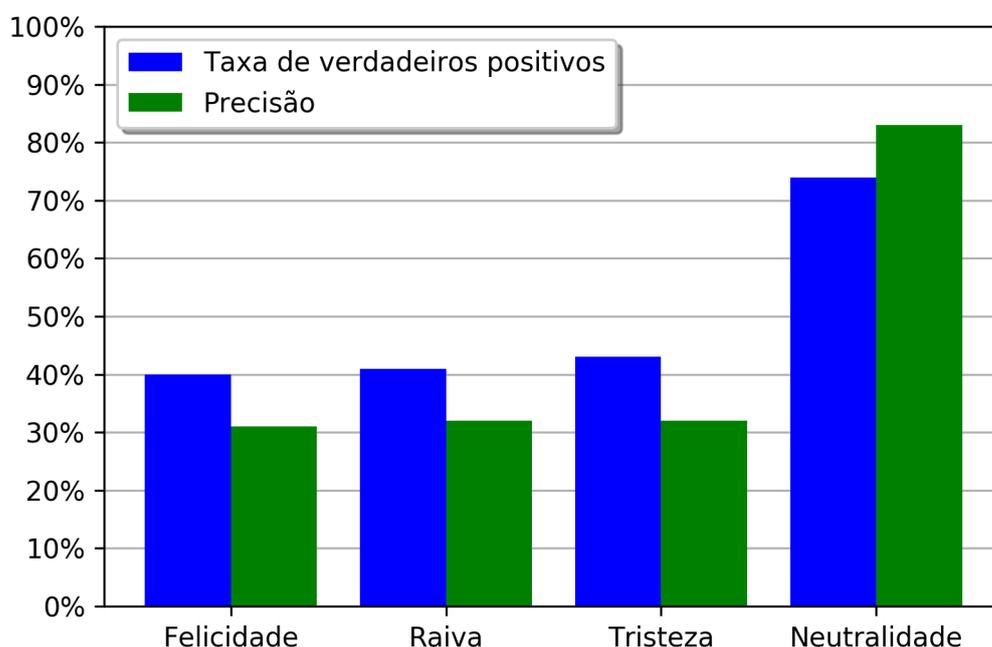


Figura 5. Balanceamento sensível a custo - Resultados para cada classe

Os valores duplicados através da super amostragem aleatória devem ser utilizados apenas para treinamento. Este método foi aplicado juntamente com uma validação cruzada, onde os conjuntos separados para treinamento foram rebalanceados enquanto o conjunto teste permaneceu com suas amostras inalteradas.

A Tabela 7 apresenta os melhores resultados utilizando o método de super amostragem aleatória. O número de neurônios nas camadas ocultas foi de 45 para a primeira camada e 49 para a segunda camada. A acurácia apresentou resultado de 58% e a precisão média de 62,8%.

Tabela 7. Balanceamento através de super amostragem aleatória - Matriz de confusão

Predito \ Real	Felicidade	Raiva	Tristeza	Neutralidade	Σ
Felicidade	11	8	10	46	75
Raiva	10	11	11	50	82
Tristeza	14	14	57	111	196
Neutralidade	69	82	165	734	1050
Σ	104	115	243	941	1403

A Figura 6 ilustra a taxa de verdadeiros positivos e a precisão para cada classe utilizando super amostragem aleatória. Em média a proporção de amostras corretamente classificadas foi de 19% para as classes minoritárias.

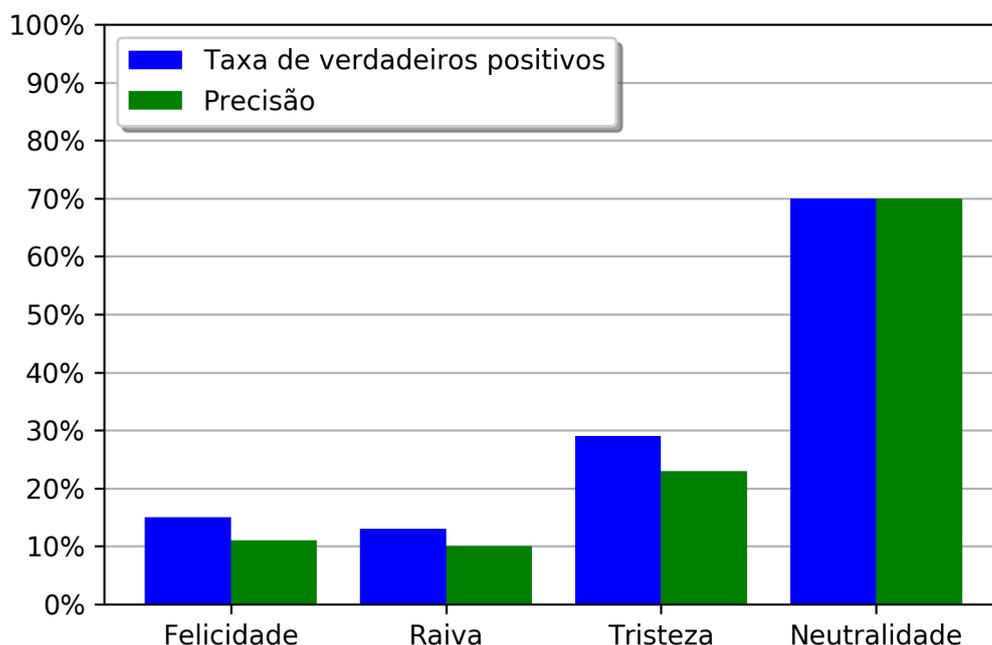


Figura 6. Balanceamento através de super amostragem aleatória - Resultados para cada classe

Os resultados obtidos neste trabalho se aproximam dos resultados obtidos em trabalhos anteriores que desenvolveram classificadores independentes de usuário. Rigas, Katsis, Ganiatsas e Fotiadis tiveram uma acurácia de 62,7% com KNN e 62,4% com *Random Forest* [Rigas et al. 2007]. Para Kim, Bang e Kim a acurácia foi de 78,4% para 3 emoções e 61,8% para 4 emoções [Kim et al. 2004]. Kim e André buscaram inferir o mesmo grupo de emoções que este estudo buscou, eles treinaram um classificador independente de usuário baseado em análise discriminante linear que obteve uma acurácia de 70% [Kim and André 2008].

6. Conclusão

O número de amostras utilizadas para treinamento e teste dos classificadores é pequena, seriam necessários mais dados para afirmar a validade do modelo. O coletor construído e o software de gerenciamento de coleta operaram com êxito, podendo ser utilizados para futuras coletas e conseqüentemente na expansão da base de dados fisiológicos.

Para os dados coletados neste estudo, o grande número de amostras para neutralidade é explicado pelo fato da maioria das imagens usadas como estímulo serem neutras com o objetivo de normalizar os sinais. O alto índice de falsos positivos nesta emoção pode se dar pelo motivo de que, estímulos mapeados para esta emoção serem exibidos logo após uma série de estímulos não neutros, o que pode ter contaminado algumas amostras.

O baixo número de amostras não neutras pode estar associada a escolha das imagens utilizadas como estímulo. Visou-se utilizar imagens com baixo desvio padrão médio

para excitação, uma abordagem diferente utilizando imagens com maiores valores para excitação poderia ter resultado em um número maior de amostras com emoções não neutras.

Outro fato a se observar é que dois terços dos indivíduos que participaram deste estudo são do sexo masculino, o que pode indicar que os resultados obtidos podem não serem independente de gênero. Os bio-sinais variam também conforme a faixa etária, neste estudo todos os indivíduos estavam na faixa de 20 a 30 anos, logo a independência de idade pode não ser abrangente.

Existe um grande número de metodologias que poderiam ter sido aplicadas ao banco de dados criado neste estudo que ficam para trabalhos futuros, como por exemplo, aplicação de filtros nos sinais, extração de outros tipos de características, análise e limpeza de *outliers* e a aplicação de outros algoritmos de classificação. Estas outras metodologias poderiam resultar em um processo mais simples e possivelmente em resultados melhores.

Referências

- Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59.
- Coan, J. A. and Allen, J. J. (2007). *Handbook of emotion elicitation and assessment*. Oxford university press.
- Costa, A., Rincon, J. A., Carrascosa, C., Julian, V., and Novais, P. (2019). Emotions detection on an ambient intelligent system using wearable devices. *Future Generation Computer Systems*, 92:479–489.
- Haag, A., Goronzy, S., Schaich, P., and Williams, J. (2004). Emotion recognition using bio-sensors: First steps towards an automatic system. *Affective Dialogue Systems*, pages 36–48.
- Hoenig, F., Batliner, A., and Noeth, E. (2007). Real-time recognition of the affective user state with physiological signals.
- Hönig, F., Wagner, J., Batliner, A., and Nöth, E. (2009). Classification of user states with physiological signals: On-line generic features vs. specialized feature sets. In *2009 17th European Signal Processing Conference*, pages 2357–2361. IEEE.
- Jerritta, S., Murugappan, M., Nagarajan, R., and Wan, K. (2011). Physiological signals based human emotion recognition: a review. In *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, pages 410–415. IEEE.
- Johnstone, I. M. and Paul, D. (2018). Pca in high dimensions: An orientation. *Proceedings of the IEEE*, 106(8):1277–1292.
- Karlik, B. and Olgac, A. V. (2011). Performance analysis of various activation functions in generalized mlp architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, 1(4):111–122.
- Kim, J. and André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12):2067–2083.

- Kim, K., Bang, S., and Kim, S. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42(3):419–427.
- Lang, P. and Bradley, M. M. (2007). The international affective picture system (iaps) in the study of emotion and attention. *Handbook of emotion elicitation and assessment*, 29.
- Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *American psychologist*, 50(5):372.
- Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (2008). International affective picture system (iaps): affective ratings of pictures and instruction manual. university of florida, gainesville. Technical report, Tech Rep A-8.
- Park, C., Ryu, J., Sohn, J., and Cho, H. (2007). An emotion expression system for the emotional robot. In *2007 IEEE International Symposium on Consumer Electronics*, pages 1–6. IEEE.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Picard, R. W. (2000). *Affective computing*. MIT press.
- Picard, R. W., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):1175–1191.
- Prati, R. C., Batista, G. E., and Monard, M. C. (2009). Data mining with imbalanced class distributions: concepts and methods. In *IICAI*, pages 359–376.
- Rani, P. and Sarkar, N. (2006). A new approach to implicit human-robot interaction using affective cues. In *Mobile Robots: towards New Applications*. IntechOpen.
- Rigas, G., Katsis, C. D., Ganiatsas, G., and Fotiadis, D. I. (2007). A user independent, biosignal based, emotion recognition method. In *International Conference on User Modeling*, pages 314–318. Springer.
- Tantithamthavorn, C., Hassan, A. E., and Matsumoto, K. (2018). The impact of class rebalancing techniques on the performance and interpretation of defect prediction models. *IEEE Transactions on Software Engineering*.
- Tipping, M. E. and Bishop, C. M. (1999). Mixtures of probabilistic principal component analyzers. *Neural computation*, 11(2):443–482.
- Xiaoxiao Wu, S., Wai, H.-T., Li, L., and Scaglione, A. (2018). A review of distributed algorithms for principal component analysis. *Proceedings of the IEEE*, 106:1321–1340.
- Zhu, X. (2010). Emotion recognition of emg based on bp neural network. In *Proc Int Symposium Network. Network Security*, pages 227–229. Jingtangshan.