UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

MARIA EDUARDA DE ARAÚJO CARDOSO

Segmentação automática de Expressões Faciais Gramaticais com *Multilayer
Perceptrons* e Misturas de Especialistas

São Paulo

2018

MARIA EDUARDA DE ARAÚJO CARDOSO

Segmentação automática de Expressões Faciais Gramaticais com *Multilayer Perceptrons* e Misturas de Especialistas

Versão corrigida

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação.

Área de concentração: Metodologia e Técnicas da Computação

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 02 de outubro de 2018. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Profa. Dra. Sarajane Marques Peres

São Paulo

2018

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.

CATALOGAÇÃO-NA-PUBLICAÇÃO

(Universidade de São Paulo. Escola de Artes, Ciências e Humanidades. Biblioteca)
CRB-8 4936

Cardoso, Maria Eduarda de Araújo

Segmentação automática de expressões faciais gramaticais com
Multilayer Perceptrons e Misturas de Especialistas / Maria Eduarda
de Araújo Cardoso ; orientadora, Sarajane Marques Peres. – 2018.
153 f. : il.

Dissertação (Mestrado em Ciências) - Programa de Pós-
Graduação em Sistemas de Informação, Escola de Artes,
Ciências e Humanidades, Universidade de São Paulo.
Versão corrigida

1. Reconhecimento de padrões. 2. Aprendizagem
computacional. 3. Língua brasileira de sinais. 4. Gestos -
Análise. 5. Análise do movimento humano. I. Peres, sarajane
Marques, orient. II. Título.

CDD 22.ed.— 006.4

Dissertação de autoria de Maria Eduarda de Araújo Cardoso, sob o título “**Segmentação automática de Expressões Faciais Gramaticais com *Multilayer Perceptrons* e Misturas de Especialistas**”, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Sistemas de Informação, na área de concentração Metodologia e Técnicas da Computação, aprovada em 02 de outubro de 2018 pela comissão julgadora constituída pelos doutores:

Prof. Dr. João Luiz Bernardes

Universidade de São Paulo - EACH

Presidente

Profa. Dra. Arianne Machado Lima

Universidade de São Paulo - EACH

Prof. Dr. Clodis Boscarioli

Universidade Estadual do Oeste do Paraná

Prof. Dr. Leandro Augusto da Silva

Universidade Presbiteriana Mackenzie

Dedico esta dissertação de Mestrado aos meus pais e meu marido, pelo amor, apoio e dedicação constantes, durante toda a minha vida.

Agradecimentos

Primeiramente, agradeço a Deus por sempre ter me abençoado e guardado até o presente momento.

Agradeço à minha orientadora, Prof^a Dr^a Sarajane Marques Peres, por toda dedicação a mim e a este trabalho, pela amizade, pelo apoio nos momentos difíceis, pela paciência ao longo desses 2 anos e alguns meses, e pela confiança em mim depositada quando aceitou ser minha orientadora. Levarei cada conselho e cada “puxão de orelha” para o resto da minha vida. Obrigada Sara pela dedicação e competência, que a tornam um modelo a ser seguido. Muitas vezes em nossas reuniões você escutou meus problemas e sempre teve um conselho para me acalmar. Palavras não expressam o meu sentimento de gratidão, muito obrigada por tudo o que fez por mim.

Agradeço a meus pais que sempre estiveram ao meu lado e em todos momentos da minha vida nunca deixaram me faltar nada. Dedicaram suas vidas para que hoje eu estivesse aqui realizando meu sonhado Mestrado. Sempre estiveram em todas as plateias para me aplaudir de pé e também estender as mãos quando necessário. Agradeço a Deus pelos pais que tenho, que por muitas vezes tiraram deles para poder fazer minhas vontades. Posso afirmar que tenho os melhores pais do mundo. Agradeço à minha irmã Maria Clara que também sempre esteve ao meu lado e me ajudou em tudo que precisei. Agradeço pela família maravilhosa!

Agradeço ao meu esposo Henrique por toda paciência comigo ao longo desse Mestrado. Por todas as vezes que eu estava triste e ele sempre estava ali para me tirar um sorriso e me dizer o quão capaz eu sou. Obrigada meu amor pelo companheirismo e apoio.

Agradeço aos amigos que fiz nessa jornada Dani, Fer, Paulo e Williane que invariavelmente estavam ao meu lado para ouvir na hora das minhas aflições. Sempre ofereceram ombro amigo nas horas ruins. Irmãos que fiz no PPgSI. Agradeço minha mãe postiça Luciana e minha vovó postiça Sandra por toda a força que sempre me passaram, pelo amor e carinho que sempre me trataram mesmo nos meus momentos de “stress”.

“A ciência nunca resolve um problema sem criar pelo menos outros dez.”
(George Bernard Shaw)

Resumo

CARDOSO, Maria Eduarda de Araújo. **Segmentação Automática de Expressões Faciais Gramaticais com *Multilayer Perceptrons* e Misturas de Especialistas**, 2018. 153 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2018.

O reconhecimento de expressões faciais é uma área de interesse da ciência da computação e tem sido um atrativo para pesquisadores de diferentes áreas, pois tem potencial para promover o desenvolvimento de diferentes tipos de aplicações. Reconhecer automaticamente essas expressões tem se tornado um objetivo, principalmente na área de análise do comportamento humano. Especialmente para estudo das línguas de sinais, a análise das expressões faciais é importante para a interpretação do discurso, pois é o elemento que permite expressar informação prosódica, suporta o desenvolvimento da estrutura gramatical e semântica da língua, e ajuda na formação de sinais com outros elementos básicos da língua. Nesse contexto, as expressões faciais são chamadas de expressões faciais gramaticais e colaboram na composição no sentido semântico das sentenças. Entre as linhas de estudo que exploram essa temática, está aquela que pretende implementar a análise automática da língua de sinais. Para aplicações com objetivo de interpretar línguas de sinais de forma automatizada, é preciso que tais expressões sejam identificadas no curso de uma sinalização, e essa tarefa dá-se é definida como segmentação de expressões faciais gramaticais. Para essa área, faz-se útil o desenvolvimento de uma arquitetura capaz de realizar a identificação de tais expressões em uma sentença, segmentando-a de acordo com cada tipo diferente de expressão usada em sua construção. Dada a necessidade do desenvolvimento dessa arquitetura, esta pesquisa apresenta: uma análise dos estudos na área para levantar o estado da arte; a implementação de algoritmos de reconhecimento de padrões usando *Multilayer Perceptron* e misturas de especialistas para a resolução do problema de reconhecimento da expressão facial; a comparação desses algoritmos reconhecedores das expressões faciais gramaticais usadas na concepção de sentenças na Língua Brasileira de Sinais (Libras). A implementação e teste dos algoritmos mostraram que a segmentação automática de expressões faciais gramaticais é viável em contextos dependentes do usuários. Para contextos independentes de usuários, o problema de segmentação de expressões faciais representa um desafio que requer, principalmente, a organização de um ambiente de aprendizado estruturado sobre um conjunto de dados com volume e diversidade maior do que os atualmente disponíveis.

Palavras-chaves: Mistura de Especialistas. *Multilayer Perceptron*. Reconhecimento de padrões. Expressões faciais gramaticais. Língua de sinais. Língua Brasileira de Sinais.

Abstract

CARDOSO, Maria Eduarda de Araújo. **Automatic Segmentation of Grammatical Facial Expressions with Multilayer Perceptrons and Mixtures of Experts:** work subtitle. 2018. 153 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2018.

The recognition of facial expressions is an area of interest in computer science and has been an attraction for researchers in different fields since it has potential for development of different types of applications. Automatically recognizing these expressions has become a goal primarily in the area of human behavior analysis. Especially for the study of sign languages, the analysis of facial expressions represents an important factor for the interpretation of discourse, since it is the element that allows expressing prosodic information, supports the development of the grammatical and semantic structure of the language, and eliminates ambiguities between similar signs. In this context, facial expressions are called grammatical facial expressions. These expressions collaborate in the semantic composition of the sentences. Among the lines of study that explore this theme is the one that intends to implement the automatic analysis of sign language. For applications aiming to interpret signal languages in an automated way, it is necessary that such expressions be identified in the course of a signaling, and that task is called “segmentation of grammatical facial expressions”. For this area, it is useful to develop an architecture capable of performing the identification of such expressions in a sentence, segmenting it according to each different type of expression used in its construction. Given the need to develop this architecture, this research presents: a review of studies already carried out in the area; the implementation of pattern recognition algorithms using *Multilayer Perceptron* and mixtures of experts to solve the facial expression recognition problem; the comparison of these algorithms as recognizers of grammatical facial expressions used in the conception of sentences in the Brazilian Language of Signs (Libras). The implementation and tests carried out with such algorithms showed that the automatic segmentation of grammatical facial expressions is practicable in user-dependent contexts. Regarding user-independent contexts, this is a challenge which demands the organization of a learning environment structured on datasets bigger and more diversified than those current available.

Keywords: Mixtures of experts. Multilayer Perceptron. Pattern recognition. Grammatical facial expressions. Sign language. Brazilian Sign Language.

Lista de figuras

Figura 1 – Exemplo de expressões faciais afetivas	23
Figura 2 – Exemplo de expressões faciais gramaticais	24
Figura 3 – Arquitetura para processamento da língua de sinais. O quadro tracejado destaca o foco de estudo deste trabalho	28
Figura 4 – Espaço de realização dos sinais e suas quatro áreas principais de articulação	34
Figura 5 – Modelagem de uma sentença em LS usando EFGs	35
Figura 6 – Quadro de vídeo contendo EFGs usadas na execução de uma sentença afirmativa (esquerda) e uma sentença negativa (direita)	37
Figura 7 – Exemplo de EFG da palavra gordão na fala em LS (bochechas infladas)	38
Figura 8 – Esquema de unidade McCulloch - Pitts	57
Figura 9 – Exemplo de uma rede MLP de múltiplas camadas (4 entradas, 1 camada escondida e 3 saídas)	60
Figura 10 – Problema de discriminação de vogais. A nuvem de dados a esquerda representam as vogais [i] minúsculo e [I] maiúsculo e a nuvem de dados a direita representam as vogais [a] minúsculo e [A] maiúsculo	63
Figura 11 – Estrutura típica de uma arquitetura de mistura de especialistas	64
Figura 12 – Arquitetura de mistura de especialistas modelada ao problema	66
Figura 13 – Exemplo de uma sequência de quadros de vídeo capturadas pelo <i>Kinect</i>	69
Figura 14 – Exemplo de face neutra e da execução de uma EFG, e os respectivos pontos (x,y) extraídos da face pelo <i>Kinect</i>	69
Figura 15 – Intérprete <i>a</i> e intérprete <i>b</i> sinalizando as frases do conjunto de dados (ambos intérpretes realizando a frase “Eu vou”(afirmativa)	72
Figura 16 – Representação do pré-processamento dos dados	76
Figura 17 – Efeitos dos procedimentos de normalização e translação: quadros originais à esquerda; quadros normalizados e transladados à direita	77
Figura 18 – Indicação dos pontos usados nos subconjuntos com nove pontos	77
Figura 19 – Distância entre pontos da face: acima - um quadro com expressão neutra; (abaixo) um quadro com uma EFG	78
Figura 20 – Quadros reais e distâncias entre pontos da face para a expressão neutra e EFGs estudadas nos experimentos deste estudo	78

Figura 21 – Exemplo de problema atemporal identificado em um quadro de vídeo	79
Figura 22 – Exemplo de janela com 8 pontos centrado no quadro Q_{int}	80
Figura 23 – Exemplo de janelamento de tamanho 3 com rótulo no meio	81
Figura 24 – Estrutura de como foi realizada as experimentações	83
Figura 25 – Estratégia adotada para construção de subconjuntos de dados para treino e teste dos classificadores	84
Figura 26 – F -score dos testes dependentes do intérprete	89
Figura 27 – F -score dos testes independentes do intérprete	89
Figura 28 – F -score dos piores e melhores casos das redes MLP e ME, casos depen- dentes e independentes do intérprete	93
Figura 29 – Análise de erros na borda no problema de três classes: duas EFGs mais a expressão neutra com oito pontos da face. Análise em testes dependentes e independentes do intérprete	97
Figura 30 – Análise de erros na borda no problema de três classes: duas EFGs mais a expressão neutra com oito pontos da face. Análise em testes dependentes e independentes do intérprete	98
Figura 31 – Análise de erros gerais e de borda no problema de três classes com oito pontos da face. Análise em testes dependentes do intérprete	99
Figura 32 – Análise de erros gerais e de borda no problema de três classes com oito pontos da face. Análise em testes independentes do intérprete	100
Figura 33 – Análise de erros na borda no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete	103
Figura 34 – Análise de erros na borda no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete	104
Figura 35 – Análise de erros de borda e erros gerais no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete	105
Figura 36 – Análise de erros de borda e erros gerais no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete	105

Figura 37 – Análise de erros na borda no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete	109
Figura 38 – Análise de erros na borda no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete	110
Figura 39 – Análise de erros de borda e erros gerais no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes	111
Figura 40 – Análise de erros na borda e erros gerais no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise em testes independentes	112
Figura 41 – Análise de erros na borda no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete	115
Figura 42 – Análise de erros na borda no problema de três classes com a distância dos pontos : duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete	116
Figura 43 – Análise de erros na borda e erros gerais no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes	117
Figura 44 – Análise de erros na borda e erros gerais no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes independentes	117
Figura 45 – <i>A média das acurácias</i> dos piores e melhores casos das redes MLP e ME, casos dependentes e independentes do intérprete	119
Figura 46 – <i>A média das acurácias</i> dos piores e melhores casos das redes MLP e ME, casos dependentes e independentes do intérprete	123

Lista de quadros

Quadro 1 – Comparativo entre este trabalho e o trabalho de Freitas (2015)	27
Quadro 2 – Possibilidades de movimentos importantes para a construção da EFG na fala em LS.	37
Quadro 3 – Informações sobre os dados utilizados nas experimentações e aplicações dos estudos referentes as EFGs	45
Quadro 4 – Técnicas e métodos utilizados nos estudos levantados	46
Quadro 5 – Finalidade dos estudos levantados	54
Quadro 6 – Exemplos de frases usadas no conjunto de dados	72
Quadro 7 – Exemplo de adaptação do conjunto de dados para adequação à mode- lagem de problemas multiclasse	74
Quadro 8 – Exemplo de janelas de tamanhos distintos	80

Lista de tabelas

Tabela 1 – Quantidade total de quadros de vídeo de cada expressão; quadros positivos (contém expressão) e quadros negativos (não contém expressão)	73
Tabela 2 – Total de quadros com três EFGs combinadas	75
Tabela 3 – Total de quadros com todas EFGs combinadas	75
Tabela 4 – Número de neurônios na cada oculta da MLP para cada representação de dados e uso ou não de janelas: janelas de tamanho 3, 5, e 9. Rep: representação; tam: tamanho	84
Tabela 5 – <i>F-score</i> , precisão, revocação e acurácia de classificadores no problema de duas classes com oito pontos da face: uma EFG mais a expressão neutra. Análise dependente do intérprete	87
Tabela 6 – <i>F-score</i> , precisão, revocação e acurácia de classificadores no problema de duas classes com oito pontos da face: uma EFG mais a expressão neutra. Análise independente do intérprete	88
Tabela 7 – <i>F-score</i> , precisão, revocação e acurácia de classificadores no problema de duas classes com oito pontos da face: uma EFG mais a expressão neutra. Análise dependente do intérprete	91
Tabela 8 – <i>F-score</i> , precisão, revocação e acurácia de classificadores no problema de duas classes com oito pontos da face: uma EFG mais a expressão neutra. Análise independente do intérprete	91
Tabela 9 – <i>F-score</i> , precisão, revocação e acurácia de classificadores no problema de duas classes com oito pontos da face: uma EFG mais a expressão neutra. Análise dependente e independente do intérprete	92
Tabela 10 – Acurácia de classificadores no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete.	95
Tabela 11 – Acurácia de classificadores no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise independente do intérprete	96

Tabela 12 – Acurácia de classificadores no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete	101
Tabela 13 – Acurácia de classificadores no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise independente do intérprete	102
Tabela 14 – Acurácia de classificadores no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete	107
Tabela 15 – Acurácia de classificadores no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise independente do intérprete	108
Tabela 16 – Acurácia de classificadores no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete	113
Tabela 17 – Acurácia de classificadores no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise independente do intérprete	114
Tabela 18 – <i>F-score</i> , precisão, revocação e acurácia de classificadores no problema de duas classes com oito pontos da face: uma EFG mais a expressão neutra. Análise dependente do intérprete	118
Tabela 19 – Acurácia de classificadores no problema de sete classes com oito pontos da face e erros de borda e erros gerais: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete	120
Tabela 20 – Acurácia de classificadores no problema de sete classes e erros de borda e erros gerais com a distância dos pontos da face: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete	120
Tabela 21 – Acurácia de classificadores no problema de sete classes com oito pontos da face e erros de borda e erros gerais: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete	121

Tabela 22 – Acurácia de classificadores no problema de sete classes com a distância dos pontos da face e erros de borda e erros gerais: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete	122
Tabela 23 – Acurácia de classificadores no problema de sete classes com oito pontos da face e erros de borda e erros gerais: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete	122
Tabela 24 – Acurácia de classificadores no problema de três classes com nove pontos da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete	136
Tabela 25 – Acurácia de classificadores no problema de três classes com nove pontos da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete	137
Tabela 26 – Análise de erros de borda no problema de três classes nove pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete	138
Tabela 27 – Acurácia de classificadores no problema de três classes nove pontos da face: duas EFGs mais a expressão neutra. Análise independente do intérprete	139
Tabela 28 – Análise de erros de borda no problema de três classes nove pontos da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete	140
Tabela 29 – Análise de erros de borda no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete	141
Tabela 30 – Análise de erros de borda no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete	142
Tabela 31 – Análise de erros de borda no problema de três classes com distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete	143

Tabela 32 – Análise de erros de borda no problema de três classes com a distância dos pontos : duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete	144
Tabela 33 – Análise de erros de borda no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete	145
Tabela 34 – Análise de erros de borda no problema de três classes com oito pontos da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete	146
Tabela 35 – Análise de erros de borda no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete	147
Tabela 36 – Análise de erros de borda no problema de três classes com a distância dos pontos da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete	148
Tabela 37 – Acurácia de classificadores no problema de três classes com nove pontos da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete	149
Tabela 38 – Acurácia de classificadores no problema de três classes nove pontos da face: duas EFGs mais a expressão neutra. Análise independente do intérprete	150
Tabela 39 – Análise de erros de borda no problema de três classes com nove pontos da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete	151
Tabela 40 – Análise de erros de borda no problema de três classes com nove pontos da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete	152
Tabela 41 – Acurácia de classificadores no problema de sete classes e erros de borda com nove pontos da face: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete	153

Lista de abreviaturas e siglas

AM	Aprendizado de Máquina
AU	Áreas utilizadas
EF	Expressão Facial
EFA	Expressão Facial Afetiva
EFG	Expressão Facial Gramatical
LS	Língua de Sinais
ME	Mistura de Especialistas
MLP	<i>Multilayer Perceptron</i>
RNA	Redes Neurais Artificiais
SVM	<i>Support Vector Machines</i>

Lista de símbolos

W	Conjunto de pesos
x, y, z	Coordenada espacial
f'	Derivada da função de ativação
D	Dimensão
x_1, x_2, x_n	Entrada da rede
g_i	Saída da rede <i>gating</i> , interpretada como a probabilidade a <i>priori</i> , ou seja, a probabilidade de escolher o i -ésimo especialista, dada somente a entrada
$P(y x, \Theta_i^0)$	É a probabilidade do especialista i gerar a saída y baseado na entrada x e no vetor de parâmetros Θ_i
$P(i x, v^0)$	É a probabilidade de se escolher o especialista i , dados a entrada x e o vetor de parâmetros v^0
Θ_i	É o vetor de parâmetros do especialista i
μ	Saída do especialista i
f	Função de ativação
m	Número de especialistas
h	Número de neurônios na camada oculta
k	Número de neurônios na camada de saída
W_1, W_2, W_n	Pesos
z	Sinal produzido em um neurônio da camada oculta
y	Sinal produzido em um neurônio da camada de saída
$\mathbf{y}$	Saída desejada
Θ_{k+1}^j	Representa os parâmetros do j -ésimo especialista para a iteração $k+1$

η

Taxa de aprendizado

Sumário

1	Introdução	22
1.1	<i>Contextualização da pesquisa</i>	25
1.2	<i>Motivação</i>	27
1.3	<i>Hipótese</i>	29
1.4	<i>Objetivos</i>	29
1.5	<i>Método</i>	30
1.6	<i>Organização do documento</i>	31
2	Expressões faciais gramaticais na língua brasileira de sinais . .	33
2.1	<i>Estrutura básica da Língua de Sinais</i>	33
2.2	<i>Segmentação de expressões faciais gramaticais: levantamento bibliográfico</i>	38
2.2.1	Escopo dos estudos	39
2.2.2	Natureza dos dados	43
2.2.3	Bases de dados	44
2.2.4	Técnicas utilizadas e métodos de avaliação de desempenho	45
2.2.5	Considerações finais	54
3	<i>Rede Neural Artificial Multilayer Perceptron</i> e Mistura de Especialistas	56
3.1	<i>Multilayer Perceptron</i>	56
3.2	<i>Mistura de Especialistas</i>	61
3.2.1	Considerações finais	67
4	Reconhecimento automático das expressões faciais gramaticais	68
4.1	<i>Definição do problema</i>	68
4.2	<i>Conjuntos de dados</i>	71
4.2.1	<i>Grammatical Facial Expressions Data Set</i>	71
4.2.2	Extensão do conjunto de dados <i>Grammatical Facial Expressions Data Set</i>	73
4.3	<i>Pré-processamento</i>	76
4.4	<i>Configuração dos experimentos</i>	81

4.5	<i>Avaliação dos resultados</i>	85
4.6	<i>Reconhecimento de expressões faciais gramaticais: resultados e análises</i>	85
4.6.1	Resultados com o problema de classificação binária: Experimento #0	85
4.6.2	Resultados com o problema de classificação multiclasse: Experimento #1	93
5	Conclusão	125
5.1	<i>Contribuições</i>	126
5.2	<i>Limitações</i>	127
5.3	<i>Trabalhos Futuros</i>	127
	Referências ¹	129
	Apêndice A – Resultados	136

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

1 Introdução

Uma das maneiras mais representativas pela qual o ser humano demonstra seus sentimentos é por meio de expressões faciais (EFs). Recentemente, a identificação e a análise das expressões faciais têm recebido atenção especial de grupos de pesquisas de diferentes áreas, como a Linguística, o Reconhecimento de Padrões, áreas da Psicologia e algumas áreas comumente envolvidas com a acessibilidade. As emoções têm sido tema de muitas pesquisas recentes, e o seu estudo faz-se importante devido à necessidade de maior compreensão e controle das atuais patologias associadas ao aspecto emocional. Para a psicologia, o ser humano nasce com emoções básicas como o medo, a tristeza, a raiva e a alegria (PRETTE; PRETTE, 2017), e o rosto humano tem sido objeto de estudo para a compreensão de aspectos fisiológicos e psicológicos do comportamento das pessoas durante a sua interação com seu entorno, pois parte dessa interação é baseada nas expressões faciais (EKMAN; FRIESEN, 2003).

Pesquisas na área do reconhecimento automático de expressões faciais têm sido úteis para o desenvolvimento de muitas aplicações, como o controle de acesso a prédios e bancos, até sistemas de reconhecimento de faces acoplados em robôs que, inclusive, possuem a capacidade de perceber o estado emocional de seus operadores. Expressões faciais manifestam emoções e assim, por meio da análise das expressões é possível reconhecer emoções e, por exemplo, dotar agentes de software da capacidade de usar essa informação na melhoria da interação humano-computador. Emoções básicas são respostas afetivas pré-programadas evolutivamente, expressas e reconhecidas de modo semelhante por toda uma espécie, constituídas de respostas fisiológicas distintas para situações de sobrevivência e fundamentais à vida (TRACY; RANGLES, 2011). Além disso, as expressões faciais são usadas como um tipo de gesto não manual de informação referente ao discurso e à comunicação interpessoal.

No decorrer dos anos, as expressões faciais e prosódia emocional tiveram mais atenção da pesquisa científica (COSTA-VIEIRA; SOUZA, 2014), em detrimento de outros sinais não-verbais de expressão da emoção, como gestos, posturas corporais e movimentos. Estudos desenvolvidos por Chang e Huang (2010), atestam que a expressão facial compõe 55% da comunicação estabelecida entre os seres humanos, comprovando sua importância nas relações interpessoais. Estudos têm mostrado melhorias no uso de uma língua (neste

caso, língua de sinais) e no uso da linguagem gestual quando o estudo das expressões faciais está incluído em uma abordagem multimodal de análise. Como exemplos, estão os estudos de Kacorri e Huenerfauth (2016), que trabalharam na interpretação da Língua Americana de Sinais, e os estudos dos autores Uchibori e Matsuoka (2016), que estudaram Língua Japonesa de Sinais. As expressões faciais são essenciais na língua de sinais para transmissão de emoção, mas principalmente, para suportar variações semânticas. Pesquisas verificaram que as animações na língua de sinais americana com ausência de expressões faciais ou com pouca presença delas são significativamente menos compreensíveis para usuários surdos ((KACORRI, 2013), (KACORRI; HUENERFAUTH, 2014), (KACORRI; HUENERFAUTH, 2016)). A comunicação da língua de sinais inclui sinais manuais, como configuração de mão, movimento, direção da mão, ponto de articulação e número de mãos (XAVIER; BARBOSA, 2017), mas também inclui sinais não manuais como as expressões faciais.

Para a Psicologia, toda manifestação de uma expressão facial resulta da ocorrência de uma emoção, mesmo no caso de uma expressão neutra, a qual geralmente é utilizada como referência inicial para a análise das demais expressões. Há conjuntos de emoções possíveis, gerados a partir das relações e reações emocionais, e ele é suficiente para que se possa compreender as relações entre os seres humanos (WHISSELL et al., 1986). Esse conjunto é composto por seis emoções, sendo elas: felicidade, surpresa, raiva, nojo, medo e tristeza. Estudos recentes defendem que essas seis emoções podem ser resumidas em quatro: felicidade, tristeza, medo/surpresa e nojo/raiva (JACK; GARROD; SCHYNS, 2014). A figura 1 exemplifica as emoções de felicidade e tristeza (LYONS et al., 1998).

Figura 1 – Exemplo de expressões faciais afetivas

Fonte: Lyons et al. (1998)

No contexto das línguas de sinais (LS), as expressões faciais assumem um papel fundamental, pois além de possibilitar a expressividade da emoção de um indivíduo,

também é usada na formação da estrutura sintática e semântica da língua, assumindo um papel de expressão da prosódia¹. Nas LSs, a informação gramatical em uma sentença é transmitida por meio de expressões faciais. Stokoe (1960) foi um dos pioneiros nos estudos da descrição da estrutura de línguas de sinais, posicionando as expressões faciais como elementos constituintes da língua. Segundo Ferreira-Brito (1995), as EFs assumem um papel de grande destaque nas LSs, pois são essenciais para dar sentido ao que é dito. Ao assumir esse papel na estrutura da língua, as EFs são chamadas de Expressões Faciais Gramaticais (EFGs).

As EFGs se encontram presentes nos níveis morfológicos da língua, no qual podem assumir o papel de atribuição de adjetivação, e no nível sintático da língua, atuando como uma das estruturas responsáveis por construir sentenças negativas, interrogativas, afirmativas, condicionais, relativas, com tópicos e com foco. Exemplos de estudos de EFG na língua de sinais são os estudos de Uddin (2015), que usa *AdaBoost* para reconhecer expressões faciais gramaticais, e de Benitez-Quiroz, Wilbur e Martinez (2016) que versa sobre o uso das áreas utilizadas da face para a EFG de negação em LS. Um estudo recente dos autores Kumar, Roy e Dogra (2018) exemplifica o uso das EFGs como ilustrado pela figura 2. Os gestos das mãos “quem” e “o que” (duas mãos) parecem semelhantes, no entanto, esses a configuração das mãos são combinadas a diferentes expressões faciais para que os dois significados possam ser representados. Para a palavra-sinal “quem”, a boca do sinalizador tem forma arredondada e o movimento da cabeça vai para cima enquanto, já para a palavra-sinal “o que”, a boca do sinalizador está aberta e as sobrancelhas estão levantadas. Essa informação facial gramatical ajuda a discriminar os gestos e tem grande influência nos resultados de reconhecimento automático implementado para as LSs.

Figura 2 – Exemplo de expressões faciais gramaticais

Fonte: Kumar, Roy e Dogra (2018)

¹ A prosódia é expressada por meio da voz nas línguas orais. Ela é a parte da gramática dedicada às características da emissão dos sons da fala, como o acento e a entonação. Em línguas visual-espaciais, i.e. com ausência de informação sonora, a prosódia é expressada por meio de outros recursos.

Diante da importância das EFGs para a construção do discurso em uma língua de sinais, torna-se claro que aplicações que objetivam processar a informação proveniente deste discurso precisam considerar a identificação automática das EFGs. À tarefa de encontrar os trechos, dentro do discurso, nos quais uma EFG é realizada dá-se o nome de **segmentação de EFGs**. Pela literatura, observam-se esforços da comunidade científica a fim de compreender características referentes a tal tarefa. Além disso, na literatura são encontrados estudos que apresentam alternativas computacionais que lidam com a complexidade da análise automatizada das EFGs (ONG; RANGANATH, 2005; KAUSAR; JAVED, 2011; NEIVA; ZANCHETTIN, 2018a). Na prática, há algumas aplicações que traduzem um idioma falado para a língua de sinais, como o ProDeaf² e o Hand Talk Translator³, disponíveis em diferentes plataformas. Outra abordagem presente em aplicações é a substituição dos clássicos dispositivos de entrada em computador por dispositivos que aceitam gestos estáticos como forma de expressar entrada de dados. Alguns exemplos são: o Signily⁴, que substitui o teclado normal por um teclado composto de gestos estáticos de sinal (NEIVA; ZANCHETTIN, 2018b); e o jogo da força em Libras (MADEO et al., 2010; SOUZA; PERES; LIMA, 2015) que permite que a entrada das letras seja feita por meio da soletração manual usada na Libras.

É no contexto da língua de sinais gramaticais que este trabalho está inserido, sendo que o foco principal é a segmentação das expressões faciais gramaticais, via resolução de um problema de reconhecimento de padrões modelado como classificação multiclasse. Para melhor apresentar o estudo desenvolvido, as próximas seções apresentam a contextualização da pesquisa (seção 1.1), motivação (seção 1.2), a hipótese de pesquisa (seção 1.3), os objetivos (seção 1.4), o método adotado (seção 1.5) e a organização do presente documento (seção 1.6).

1.1 Contextualização da pesquisa

Em estudos realizados pelo grupo de pesquisa correlato a este projeto, foi explorado o problema do reconhecimento automático de expressões faciais gramaticais na Língua Brasileira de Sinais (Libras) a partir da modelagem de um problema de classificação

² <https://play.google.com/store/apps/details?id=com.Proativa.ProDeafMovel&hl=en>

³ <https://play.google.com/store/apps/details?id=br.com.handtalk&hl=en>

⁴ <https://itunes.apple.com/us/app/signily-keyboard-sign-language/id995022521?mt=8&ignmpt=uo%3D8>

binário. Nesse estudo, classificadores baseados na rede neural *Multilayer Perceptron* (MLP) foram aplicados para localizar uma determinada EFG dentro de uma sentença. As MLPs são redes neurais que surgiram a partir da criação de um modelo de neurônio artificial chamado *Perceptron* (ROSENBLATT, 1958; HAYKIN, 2009).

O trabalho seminal nesta área foi desenvolvido por Freitas (2015), dentro do mesmo grupo de pesquisa no qual o presente trabalho está inserido. O trabalho de Freitas et al. (2014) foi o primeiro que teve como objetivo analisar automaticamente as EFGs em Libras, estudando o problema sob uma modelagem de classificação binária com a rede neural *Multilayer Perceptron*. O presente trabalho representa uma evolução, uma vez que traz um estudo com base em um problema de classificação multiclasse, e aplica mistura de especialistas além da rede *Multilayer Perceptron*.

Este trabalho visou o desenvolvimento de uma arquitetura que seja capaz de solucionar o problema de classificação multiclasse, de forma que a partir de uma sentença expressa em língua de sinais, a arquitetura seja capaz de localizar cada uma das EFGs usadas na execução da sentença em Libras. Essa arquitetura tem como base a modelagem de um problema de classificação multiclasse a ser resolvido por classificador. Primeiramente, o problema é resolvido com uma MLP e, então, mistura de especialistas (ME) é aplicada ao problema. Mistura de especialistas foi apresentada há cerca de 20 anos, como uma arquitetura modular para aprendizado supervisionado, devidamente formalizada por Jacobs et al. (1991) e Jordan e Jacobs (1994), e vem sendo aplicada com sucesso junto a uma variedade de problemas em aprendizado de máquina (AM). Para a construção dos especialistas e do decisor (*gate*) de uma mistura de especialistas foi utilizada uma função linear, comum nesse tipo de abordagem, e dois tipos diferentes de inicialização: a inicialização aleatória e inicialização com base em grupos criados pelo algoritmo *K-means++*.

O problema de classificação binária modelado anteriormente e também usado neste trabalho tem como objetivo diferenciar uma EFG da expressão neutral. Esse tipo de análise permite o estudo da complexidade envolvida no reconhecimento de cada uma das EFGs. Contudo, apresenta uma limitação em termos práticos, visto que cada classificador é preparado para diferenciar uma EFGs da expressão neutra, mas não há garantias de que este classificador consegue aprender as características da EFG a ponto de ser usado para diferenciá-la de outras EFGs. O presente trabalho trata essa limitação apresentando um estudo em dois níveis: no primeiro, um problema mais simples com três classes (duas

EFGs e a expressão neutra) é estudado; no segundo, um problema com sete classes (seis EFGs e a expressão neutra) é explorado.

O conjunto de dados que suporta o presente estudo foi criado por Freitas et al. (2014) e Freitas (2015). Porém, o conjunto foi elaborado para suportar o estudo da classificação binárias. No presente trabalho, esse conjunto de dados foi reestruturado a fim de suportar o estudo da classificação multiclasse. O quadro 1 apresenta um resumo comparativo do presente trabalho com o desenvolvido por Freitas (2015).

Quadro 1 – Comparativo entre este trabalho e o trabalho de Freitas (2015)

Característica	Freitas et al. (2014)	Este trabalho
Modelo de classificação	Binário	Binário e multiclasse
Técnicas	MLP	MLP e mistura de especialistas
Conjunto de dados	<i>Grammatical Facial Expressions Data Set</i>	<i>Grammatical Facial Expressions Data Set</i> adaptado
Avaliação	F-score e tipos de erros	F-score, precisão, revocação, acurácia e tipos de erros

Fonte: Maria Eduarda de Araújo Cardoso, 2018

1.2 Motivação

Os dispositivos que usam técnicas de inteligência artificial têm crescido e se tornado essencial na vida das pessoas, como por exemplo, os assistente Siri da *Apple* e Cortana da *Microsoft*. Infelizmente, esses aplicativos e vários que usam a língua oral como entrada de dados não alcançam as pessoas surdas que não conseguem se expressar pela fala oral. Segundo a *World Federation of the Deaf* (DEAF, 2015), esses sistemas não alcançam cerca de 70 milhões de pessoas. A maioria dessas pessoas usa a língua de sinais como o modo de comunicação. Assim, há uma necessidade de implementar softwares capazes de detectar e interpretar a língua de sinais (o que inclui a necessidade de segmentação e reconhecimento das expressões faciais gramaticais) para que seja possível implementar interfaces que sejam mais adaptadas às pessoas que usam essa modalidade de comunicação. Para contribuir com esse segmento de comunicação, este trabalho segue a arquitetura de processamento da língua de sinais mostrada na Figura 3, a qual pode suportar o desenvolvimento de aplicações no mundo real.

Figura 3 – Arquitetura para processamento da língua de sinais. O quadro tracejado destaca o foco de estudo deste trabalho

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Na arquitetura mostrada na figura, o *Kinect* representa um sensor de captura da gesticulação. Então, a partir de uma sequência de quadros de vídeo, um sistema deve extrair os elementos constituintes dos sinais, que são: (a) configuração das mãos, (b) movimentos, (c) orientação, (d) ponto de articulação e (e) expressão facial gramatical. Este último elemento constitui o foco de atenção para este trabalho⁵.

Alguns estudos que já estão engajados na análise das expressões faciais gramaticais tem como objetivo auxiliar na tradução de uma LS (UDDIN, 2015), (BHUVAN et al., 2016), (FREITAS et al., 2017), na contribuição específica para análise de expressões faciais gramaticais (BENITEZ-QUIROZ; WILBUR; MARTINEZ, 2016), no suporte a diagnóstico de distúrbios de fala (MARSHALL et al., 2015), no desenvolvimento de sistema de reconhecimento automático para marcadores gramaticais (LIU et al., 2014), (FREITAS et al., 2017).

Estudos indicam que a automação do reconhecimento das expressões faciais melhoraram a taxa de reconhecimento da língua de sinais quando expressões faciais são incluídas em uma abordagem de análise multimodal. Por exemplo, Nguyen e Ranganath (2012) rastrearam os traços e expressões faciais utilizadas na comunicação da língua de sinais americana para impulsionar o desempenho do reconhecimento da língua de sinais. Outro estudo sobre a linguagem gestual alemã demonstrou que as características faciais podem melhorar as taxas de reconhecimento de isolamento e assinatura contínua 1,5% e 6%, respectivamente, quando comparados ao uso apenas de características manuais (AGRIS; KNORR; KRAISS,

⁵ Entretanto, por razões que serão explicadas no decorrer deste texto, apenas sete expressões faciais gramaticais, das nove existentes na Libras, são estudadas neste trabalho.

2008). Em um estudo com crianças surdas na língua de sinais britânica os autores afirmam que a falta de expressão facial modificaria o significado da sentença (MARSHALL et al., 2015). Portanto, a análise automatizada das expressões faciais gramaticais é benéfica para muitos estudos em diferentes campos, incluindo linguística descritiva.

1.3 Hipótese

A hipótese verificada neste trabalho de pesquisa é: “Expressões faciais gramaticais podem ser automaticamente localizadas e identificadas dentro de uma sentença da Libras. Ainda, defende-se que a identificação das EFGs poderá ser eficientemente realizada com o apoio da *Multilayer Perceptron* e da mistura de especialistas.”

As expressões faciais gramaticais podem ocorrer individualmente ou concomitantemente em uma fala na Libras. A hipótese adotada nesse trabalho defende que se um classificador baseado em *Multilayer Perceptron* e mistura de especialistas são capazes de especializar seus componentes em aspectos diferentes de um problema, então tais classificadores são capazes de analisar um trecho de vídeo e identificar a expressão facial gramatical que ocorreu nesse trecho; ou a não ocorrência de uma expressão facial gramatical.

Portanto o trabalho está baseado na capacidade do modelo classificador MLP e ME em gerar componentes especializados em analisar aspectos que discriminam expressões faciais gramaticais. A hipótese deve ser refutada se: (a) o classificador não for capaz de encontrar tais sequências; (b) o classificador não for capaz de identificar corretamente a expressão (ou expressões) facial gramatical que ocorre em tais sequências, considerando uma margem de erro empírico admissível seguindo avaliação de especialistas.

1.4 Objetivos

O objetivo deste estudo foi apresentar soluções para o problema de identificação das expressões faciais gramaticais que ocorrem na execução de uma sentença em Libras, segmentando-a de acordo com a ocorrência de cada expressão. A localização e consequente segmentação foi resolvida como um problema de reconhecimento de padrões modelado como classificação multiclasse.

Os objetivos específicos são:

- exploração do estado da arte na área do reconhecimento das EFGs;
- adaptação do *corpus dataset* “*Grammatical Facial Expressions*”⁶ (FREITAS et al., 2014) de forma que um problema de classificação multiclasse possa ser tratado sobre ele: trata-se de um *corpus* construído pelo grupo de pesquisa correlato a esse trabalho para suportar o primeiro estudo brasileiro referente à automação do reconhecimento de EFGs no contexto da Libras (FREITAS, 2015). Esse *corpus* foi preparado para suportar a resolução do problema modelado como classificação binária, e necessitou ser alterado e rotulado para que pudesse suportar a resolução e avaliação do problema modelado como classificação multiclasse;
- criação/treinamento de modelos classificadores para resolver o problema de reconhecimento das EFGs, avaliando o seu desempenho isoladamente a fim de verificar se a técnica MLP é adequada para o tratamento do problema de reconhecimento de padrões em questão;
- criação/treinamento de modelos classificadores baseado em mistura de especialistas;
- organização da avaliação dos desempenhos quantitativo e qualitativo dos modelos classificadores por meio do estabelecimento de um protocolo de avaliação que considere medidas clássicas usadas na área de aprendizado de máquina (*F-score* e acurácia) (FAWCETT, 2006), e a visão de um especialista em língua de sinais interessado no estudo da execução da expressão facial ou da interpretação da Libras, considerando erros de borda e erros de segmentação (MADEO, 2013).

1.5 Método

Este trabalho está organizado como uma pesquisa do tipo experimental, que contém pesquisa bibliográfica exploratória, levantamento bibliográfico, construção do *corpus* multiclasse, implementação e teste de algoritmos e avaliação de resultados referentes à construção de modelos classificadores para localização das EFGs na sentença em Libras.

Na pesquisa bibliográfica exploratória foram executadas leituras de bibliografia da área de língua de sinais, Libras, reconhecimento de padrões e técnicas de aprendizado de máquina (*Multilayer Perceptron* e mistura de especialistas). Esse estudo bibliográfico teve o objetivo de fundamentar a pesquisa já realizada. Também foi realizada uma revisão

⁶ <https://archive.ics.uci.edu/ml/datasets/Grammatical+Facial+Expressions>

de literatura para levantar o estado da arte em reconhecimento de expressões faciais gramaticas considerando o contexto de diferentes línguas de sinais.

Foi realizada a construção e atualização do *corpus Grammatical Facial Expressions Data Set* por meio de uma reorganização das sentenças e geração de uma nova rotulação para o modelo multiclasse. A proponente desta pesquisa possui conhecimentos básicos em Libras e realizou o trabalho de atualização do *corpus* junto do Prof. Dr. Felipe Barbosa (especialista em estudo da Libras na FFLCH/USP).

A construção dos classificadores foi realizada em dois momentos. Em um primeiro momento a técnica MLP foi aplicada no problema (em versões simplificadas do problema e na sua versão completa) para que fosse possível melhorar o conhecimento sobre a complexidade da resolução de um problema multiclasse nessa área. Em seguida foram implementadas as arquiteturas de mistura de especialistas. Os experimentos foram realizados com diferentes tipos de parametrizações.

A avaliação dos classificadores foi realizada via medidas tradicionalmente usadas na área de reconhecimento de padrões extraídas por meio da matriz de confusão (MONARD; BARANAUSKAS, 2003) e por estatística descritiva. Também foram analisados os tipos de erros: os erros da borda e os erros de segmentação.

1.6 Organização do documento

O presente trabalho de dissertação está dividido em cinco capítulos, considerando esta introdução. Os demais capítulos estão organizados da seguinte forma:

- o capítulo 2 contextualiza a Libras e sua importância para os estudos na área. Apresenta o estudo bibliográfico referente ao reconhecimento de expressões faciais gramaticais na Libras. O capítulo traz as informações referentes aos diferentes problemas estudados na língua de sinais, as técnicas que estão sendo utilizadas para o reconhecimento das EFGs, as bases de dados utilizadas, as diferentes língua de sinais para as quais já existem estudos nessa área, contribuições técnicas e científicas produzidas, as limitações dos estudos desenvolvidos e o que os estudos apontam como perspectivas;
- o capítulo 3 apresenta o estudo sobre *Multilayer Perceptron* e mistura de especialistas, bem como informações sobre como funcionam esses modelos classificadores;

- o capítulo 4 apresenta a abordagem de resolução da segmentação das EFGs, descreve a definição do problema, o conjunto de dados utilizados, a extensão do conjunto de dados e o pré-processamento aplicado a eles, e descreve as configurações dos experimentos realizados e os resultados e análises resultantes do estudo;
- o capítulo 5 apresenta as considerações finais deste trabalho, com as principais contribuições do estudo, propostas para trabalhos futuros e as limitações encontradas.

2 Expressões faciais gramaticais na língua brasileira de sinais

As expressões faciais fazem parte da comunicação humana, sendo que por meio delas é possível expressar emoções e modificar o discurso durante a fala. Na língua de sinais, as expressões faciais assumem um papel importante na construção da sintaxe e da semântica do discurso, portanto, são essenciais para dar sentido ao que é dito (SANDLER, 2010). A Libras foi reconhecida como a língua oficial para comunicação e expressão da comunidade surda¹ do Brasil pela Lei nº10.436², 24 de abril de 2002, “como meio legal de comunicação e expressão [...] em que o sistema linguístico de natureza visual-motora, com estrutura gramatical própria, constituem um sistema linguístico de transmissão de ideias e fatos, oriundos de comunidades de pessoas surdas do Brasil”. Visto isso, é perceptível a importância das EFs no contexto da língua de sinais.

2.1 Estrutura básica da Língua de Sinais

A Libras tem um sistema de sinais gestuais bastante rico e complexo. Existem parâmetros básicos como: configuração das mãos, pontos de articulação, movimento e expressões faciais e a execução combinada de variantes destes e de outros parâmetros promove o funcionamento da língua. De acordo com Ferreira-Brito (1990), há 46 configurações de mão, 6 tipos de orientações de mão, em torno de 40 locações no corpo, 16 locações no espaço neutro, 22 expressões não-manuais, uma lista de 35 possíveis movimentos internos da mão. Em relação ao movimento, Ferreira-Brito (1990) identifica 28 especificações, 17 direcionalidades, 5 maneiras e 2 tipos de frequência. A figura 4 revela que o espaço de enunciação é uma área que contém todos os pontos dentro do raio de alcance das mãos em que os sinais são articulados (QUADROS; KARNOPP, 2009).

Os autores Amaral et al. (2012), afirmaram que não há um levantamento exato sobre todos possíveis estados dos elementos constituintes da língua de sinais. Assim, é notável que esses elementos podem ser estudados de forma mais aprofundada para que se descubra todos aspectos envolvidos no estudo.

¹ O termo Surdo tem sido utilizado quando a pessoa com surdez é caracterizada como surdez profunda no âmbito da medicina, quando é leve ou moderada ainda persiste o termo Deficiente Auditivo, já na Comunidade Surda, o Surdo é aquele que é usuário de Libras e é pertencente a tal.

² http://www.planalto.gov.br/ccivil_03/leis/2002/110436.htm

Figura 4 – Espaço de realização dos sinais e suas quatro áreas principais de articulação

Fonte: Ferreira-Brito (1995)

No contexto das línguas de sinais, expressões faciais assumem um papel específico, constituindo-se como expressões faciais gramaticais. As EFGs estão relacionadas às estruturas específicas das línguas de sinais, tanto no nível morfológico quanto no nível sintático (QUADROS; KARNOPP, 2004). As EFGs podem modificar sinais atribuindo-lhes informações sintáticas ou morfológicas e em casos em que o traço mínimo é distintivo de valor, alterando os seus sentidos. Expressões faciais gramaticais são, portanto, marcações não-manuais e estão relacionadas às estruturas específicas das línguas, tanto no nível fonético-fonológico, morfológico ou no nível sintático (QUADROS; KARNOPP, 2004; FERREIRA-BRITO, 1990):

- nível fonético-fonológico: as EFGs são tradicionalmente descritas como um dos parâmetros formadores do sinal (item lexical das línguas de sinais) (XAVIER; BARBOSA, 2017) e podem ter caráter distintivo;
- nível morfológico: as EFGs são usadas como marcações morfêmicas, por exemplo, fazendo com que seja determinado um grau de intensidade para um adjetivo ou na construção de superlativos e comparativos de superioridade e inferioridade;
- nível sintático: as EFGs são responsáveis por determinar a construção de frases interrogativas, com polaridades definidas (negativas e afirmativas), referentes às condicionais relativas, na formação de sentenças de tópicos e com indicação de foco.

Para que a sentença construída em língua de sinais (por exemplo, a Libras) tenha o mesmo sentido que a mesma sentença dita em língua oral (por exemplo, o português), geralmente é necessário fazer uso das expressões faciais. Na figura 5, é apresentada à

construção de uma sentença usando EFGs. Se a sentença em questão é construída apenas com os sinais “Maria”, “gosta” e “abacaxi”, ela representaria apenas uma sequência de sinais. Já com o uso de EFGs, sentenças construídas com os mesmos sinais podem assumir sentidos diferentes resultando em, por exemplo:

- uma sentença interrogativa com topicalização “Maria, ela gosta de abacaxi?”, quando a EFG para indicar tópico é realizada junto com o sinal que expressa o sujeito sobre o qual a sentença se refere, neste caso “Maria”, e a EFG de indicação de uma interrogação que espera uma resposta afirmativa ou negativa é realizada junto com os sinais para “gostar” e “abacaxi”.
- uma sentença interrogativa negativa com topicalização “Maria, ela não gosta de abacaxi?”, quando além as EFG citadas no item anterior, ainda é realizada a EFG para negação junto do sinal para “gostar”.

Figura 5 – Modelagem de uma sentença em LS usando EFGs

Sinais em Libras:	Maria	gosta	abacaxi
Expressões faciais:	Tópico	Interrogativa (s/n)	
Português:	Maria, ela gosta de abacaxi?		
Sinais em Libras:	Maria	gosta	abacaxi
Expressões faciais:	Tópico	Interrogativa (s/n)	
		Negativa	
Português:	Maria, ela não gosta de abacaxi?		

Fonte: adaptado de Freitas (2015)

A fim de explicar em mais detalhes o uso das EFGs nos diferentes contextos de construção de frases, segue uma explicação resumida baseada nos exemplos apresentados por Quadros e Karnopp (2004) e Ferreira-Brito (1990)):

- interrogativa (qu): permite construções do tipo *quem, que, quando, por que, como, onde*. Por exemplo: QUANDO MARIA ESTUDAR (qu)³;
- interrogativas (s/n): formula questões que esperam como resposta um SIM ou NÃO. Por exemplo: MARIA GOSTAR VÔLEI (s/n);

³ Indica que a EFG é executada durante toda a frase.

- interrogativa (dúvida): expressa algum tipo de desconfiança; *que* ou *quem* aparecem em sentenças subordinadas sem a EFG interrogativa, utilizando a marcação própria da frase. Por exemplo: ESCOLA PROFESSOR ENSINAR LÍNGUA DE SINAIS (dúvida);
- negativa: frases que expressam ideias ou ações negativas⁴. Por exemplo: EU NÃO VOU (neg);
- afirmativa: frases que expressam ideias ou ações afirmativas. Por exemplo: EU VOU (afirm);
- condicional: frases que estabelecem uma condição para realizar alguma coisa. Por exemplo: SE CHOVER (condição)⁵ EU NÃO VOU FESTA (negativa);
- relativa: uma inserção dentro da frase para explicar, acrescentar informações ou encaixar outra informação relativa ao que está sendo dito. Por exemplo, MARIA CAIU CHÃO (relativa), ELA ESTÁ NO HOSPITAL;
- tópico: deslocamento de um constituinte da sentença para posição inicial. Por exemplo, COR (tópico) EU GOSTO AMARELO;
- foco: frases que introduzem uma informação nova no discurso para: (a) estabelecer contraste; (b) informar algo adicional; (c) enfatizar alguma coisa. Por exemplo: MARIA COMPROU CHOCOLATE (foco) NÃO (negativa) PAULO COMPROU CHOCOLATE.

Na figura 6 são apresentados dois quadros de vídeos extraídos da execução das frases EU VOU (à esquerda na figura) e EU NÃO VOU (à direita na figura), com o uso das EFGs afirmativa e negativa respectivamente. Note que em ambos os quadros o sinal manual executado é o mesmo, logo, o que diferencia o sentido desejado para o discurso são as EFGs.

⁴ Em Arrotéia (2005), o autor destaca que frases negativas podem aparecer com um elemento negativo explícito (*não, nada, nunca*) que pode estar incorporado aos sinais ou expressos apenas por meio da marcação-manual.

⁵ Neste caso, a EFG ocorre em uma parte da execução da frase.

Figura 6 – Quadro de vídeo contendo EFGs usadas na execução de uma sentença afirmativa (esquerda) e uma sentença negativa (direita)

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Assim, na língua de sinais essas expressões estão presentes no discurso da Libras e são utilizadas por pessoas fluentes. O quadro 2 destaca algumas das possibilidades discutidas em Ferreira-Brito (1995) com relação à configuração de elementos da face que permitem o estabelecimento das expressões faciais na Libras.

Quadro 2 – Possibilidades de movimentos importantes para a construção da EFG na fala em LS.

Rosto	
<i>Parte Superior</i>	<i>Parte Inferior</i>
franzidas	bochechas infladas
olhos arregalados	bochechas contraídas
lance dos olhos	lábios contraídos e projetados e sobrancelhas franzidas
sobrancelhas levantadas	correr da língua contra a parte inferior da bochecha
	apenas a bochecha direita inflada
	contração do lábio superior
	franzir do nariz
Cabeça	
balanceamento para frente e para trás (sim)	
balanceamento para os lados (não)	
inclinação para frente	
inclinação para o lado	
inclinação para trás	
Rosto e Cabeça	
cabeça projetada a frente, olhos levemente cerrados, sobrancelhas franzidas	
cabeça projetada para trás e olhos arregalados	

Fonte: Ferreira-Brito (1995)

Na figura 7(a), que representa o rosto na parte superior, é mostrado o intérprete com os olhos arregalados e sobrancelhas levantadas, já para parte inferior (figura 7(b)) o intérprete está com a bochechas infladas. A figura 7(c) que representa a cabeça temos

o intérprete realizando o balanceamento da cabeça para trás e por fim, a imagem 7(d) rosto e cabeça que o intérprete projeta a cabeça a frente e tem olhos levemente cerrados e sobrancelhas franzidas. Esses são alguns exemplos das possibilidades de movimentos da construção da EFG na fala.

Figura 7 – Exemplo de EFG da palavra gordão na fala em LS (bochechas infladas)

(a) Rosto: Parte Superior

(b) Rosto: Parte Inferior

(c) Cabeça

(d) Rosto e Cabeça

Fonte: Maria Eduarda de Araújo Cardoso, 2018

2.2 Segmentação de expressões faciais gramaticais: levantamento bibliográfico

O estudo discutido nesta pesquisa explorou o problema de segmentação de EFGs por meio da construção de classificadores implementados usando a rede neural artificial MLP e a mistura de especialistas. As expressões faciais gramaticais são importantes para interpretar sentenças de interrogação, negação, afirmação, dentre outras na língua de sinais. São as expressões faciais que colaboram para criar o sentido semântico da sentença, contribuem também para a formação da estrutura gramatical da língua e na formação dos sinais. No entanto, o reconhecimento automático das expressões faciais gramaticais exige cuidado devido à variação das EFGs com base no sinalizador e a co-ocorrência de sinais manuais e não manuais.

No contexto das EFGs, há uma série de trabalhos que estudam seu reconhecimento nas diferentes línguas de sinais. A fim de levantar o estado da arte nesta área, foi realizado um levantamento dos estudos referentes aos artigos publicados entre os anos de 2012 e 2018, que teve como objetivo o estudo específico de reconhecimento de expressões faciais gramaticais.

Essa seção tem como objetivo apresentar e discutir os resultados de uma revisão de literatura realizada sobre reconhecimento de padrões em EFGs. A revisão aborda a extração das características das expressões faciais, como são analisadas essas EFGs, bem como

usam as técnicas utilizadas considerando o aspecto temporal durante o reconhecimento das expressões faciais gramaticais, tendo em vista o discurso na língua de sinais.

A organização de apresentação das informações dos estudos realizados tem a seguinte ordem: apresentação do escopo dos estudos que foram levantados durante a realização do levantamento bibliográfico (Seção 2.2.1); apresentação da natureza dos dados e pré-processamento identificadas neste estudo (Seção 2.2.2); breve descrição de bases de dados utilizada pelos autores na seção 2.2.3; listagem das técnicas computacionais aplicadas na construção dos modelos que analisam as EFG, e os métodos de extração das características e metodologias de avaliação de desempenho nos estudos realizados, na seção 2.2.4, e na seção 2.2.5 por fim, os comentários sobre a revisão sistemática realizada.

2.2.1 Escopo dos estudos

Dentro do escopo dos estudos realizados há autores que trabalham em diferentes Línguas de Sinais. A língua de sinais Americana é a língua mais estudada entre elas. Os autores Kacorri e Huenerfauth (2016), Benitez-Quiroz, Wilbur e Martinez (2016), Benitez-Quiroz et al. (2014), Liu et al. (2014) e Liu et al. (2013) trabalham com essa língua. Já os autores Marshall et al. (2015) e Caridakis, Asteriadis e Karpouzis (2014) abordam a Língua Britânica de Sinais. Os trabalhos de Freitas et al. (2014), Uddin (2015), Bhuvan et al. (2016) e Freitas et al. (2017) trabalham com a Língua Brasileira de Sinais, e a Língua de Sinais Irlandesa é estudada por Kumar, Roy e Dogra (2018).

Os estudos objetivaram avaliar as expressões faciais gramaticais das línguas de sinais analisadas por áreas da face utilizadas para expressar as EFGs, como também a automação do reconhecimento dessas expressões durante um discurso de Língua de Sinais. Esses estudos também abordam diferentes tipos de expressões faciais gramaticais.

O trabalho de Kacorri e Huenerfauth (2016) apresenta uma nova metodologia para gerar movimentos de face e cabeça para animações de linguagem de sinais. Eles criaram conteúdos acessíveis para usuários surdos e investigaram a sintetização automática de animações da ASL, incluindo expressões faciais de *Interrogativa (qu)*, *Interrogativa (s/n)*, *Interrogativa (dúvida)*, *Tópico*, *Retórica* e *Negativa*. Também, analisaram os movimentos de cabeças gramaticalmente importantes. Esses autores tem vários trabalhos na área de Língua de Sinais ((KACORRI, 2013), (KACORRI; HUENERFAUTH, 2014), (KACORRI;

HUENERFAUTH, 2016)) referentes à Língua de Sinais Americana no contexto das expressões faciais gramaticais, sendo que o mais recente, de Kacorri (2016), aborda a síntese das expressões faciais ASL sintáticas, que são gramaticalmente necessárias e essenciais para o significado das frases, propondo explorar os aspectos metodológicos da avaliação de animações de linguagem gestual com expressões faciais e examinar modelagem orientada por dados de expressões faciais a partir de múltiplas gravações de sinalizador na ASL.

Liu et al. (2013) e Liu et al. (2014) propõem um sistema de reconhecimento automático para marcadores gramaticais não manuais com base em uma análise de multiescala, espaço-temporal, pose de cabeça e expressões faciais gramaticais. As expressões faciais gramaticais de *Interrogativa (qu)*, *Interrogativa (s/n)*, *Interrogativa (dúvida)*, *Tema*, *Negativa*, *Condicional* são as EFGs que esses autores estudaram.

O software *The Language Archive (ELAN)* é utilizado por Benitez-Quiroz et al. (2014) para análise sobre as anotações realizadas através de uma análise visual cuidadosa para identificar sinais não-manuais e os marcadores gramaticais que ocorrem em sequências de vídeos gravadas na Língua Americana de Sinais. Expressões faciais *Condicionais*, *Interrogativa(qu)*, *Interrogativa (s/n)*, *Interrogativa (dúvida)*, como também *Afirmativa* e *Negativa* são analisadas nesse estudo. Um estudo mais recente de Benitez-Quiroz, Wilbur e Martinez (2016) revela que a expressão facial gramatical da negação é comum em língua de sinais da Espanha, China e a Americana, e nesse estudo foi encontrado exatamente a mesma produção em marcadores gramaticais nessas diversas línguas de sinais.

Os estudos de Freitas et al. (2014) e Freitas et al. (2017) descreveram o reconhecimento das EFGs utilizados na Língua de Sinais Brasileira, e teve como objetivo apresentar um estudo que aplique o raciocínio indutivo para reconhecer padrões, como forma de estudar o problema envolvendo o reconhecimento automatizado de EFGs no nível sintático da língua empregando uma técnica de aprendizado de máquina. Os autores trabalharam com as seguintes EFGs: *Interrogativa (qu)*, *Interrogativa (s/n)*, *Interrogativa(dúvida)*, *Tópicos*, *Relativo*, *Afirmativa*, *Negativa* e *Condicional*.

Bhuvan et al. (2016) investigaram a seleção do algoritmo de aprendizado de máquina com melhor desempenho para cada expressão facial gramatical, tanto para modelos dependentes do usuário quanto para modelo independente do usuário. Após isso, identificaram os principais pontos faciais para detectar cada expressão com o melhor algoritmo de aprendizado de máquina para o modelo independente do usuário. Utilizaram as EFGs: *In-*

terrogativa (qu), *Interrogativa (s/n)*, *Interrogativa (dúvida)*, *Tópicos*, *Relativo*, *Afirmativa*, *Negativa* e *Condicional*.

O estudo de Uddin (2015) apresenta uma nova estrutura *Ada-Random Forests* para reconhecer as expressões faciais gramaticais. Para este estudo, as EFGs *Interrogativa (qu)*, *Interrogativa (s/n)*, *Interrogativa (dúvida)*, *Tópicos*, *Relativo*, *Afirmativa*, *Negativa* e *Condicional* foram utilizadas. Os autores Marshall et al. (2015) e Caridakis, Asteriadis e Karpouzis (2014) não citam com quais expressões faciais gramaticais trabalharam em seus estudos. O propósito dos autores Caridakis, Asteriadis e Karpouzis (2014) trata-se da incorporação de sugestões não-manuais no reconhecimento automático de língua de sinais e relata que expressões faciais, o olhar do olho e a cabeça do sinalizador são usados como modificadores para sinais específicos, e por isso, também devem ser incluídos na fase de reconhecimento automático. O estudo investiga a extração de características faciais gramaticais, visão do olho e sugestões de pose de cabeça. Já Marshall et al. (2015) analisa o conteúdo lexical. Para os experimentos foram utilizadas participantes crianças que receberam sentenças para a replicação exata dos itens lexicais independentemente da ordem, sendo eles: assinar ordem (se as frases foram repetidas com todos os elementos na mesma ordem que a sentença alvo, mesmo admitindo sinais omitidos), significado (o significado geral da sentença fosse reproduzido com precisão, mesmo que algumas omissões ou substituições de itens lexicais fossem feitas) e expressão facial gramatical (a falta de expressão facial alteraria o significado da frase ou mudanças de significado por falta de gramática facial).

Os autores Kumar, Roy e Dogra (2018) apresentaram uma nova estrutura multimodal para o sistema de reconhecimento da língua de sinais (SLR), incorporando a expressão facial com gestos de sinais usando dois sensores diferentes, a saber, o de movimento (*Leap Motion*) e o *Kinect*. Os gestos do sinal foram gravados usando o movimento *Leap* e, simultaneamente, um *Kinect* é usado para capturar os dados faciais do sinalizador. O reconhecimento foi realizado usando o *Hidden Markov Model* (HMM). Foi aplicada a abordagem de Combinação de Classificação Bayesiana Independente (IBCC) para combinar a decisão de diferentes modalidades para melhorar o desempenho do reconhecimento.

É notável que muitos trabalhos executam a análise, cujo o propósito é analisar áreas utilizadas da face durante a expressão facial gramatical e que o objetivo é o reconhecimento automático da Língua de Sinais (LIU et al., 2013).

Os estudos levantados apresentam algumas complexidades em suas análises. A rotulação humana é um viés no estudo de Freitas et al. (2014) Freitas et al. (2017) e também para Uddin (2015) e Bhuvan et al. (2016) pois, utilizaram o mesmo conjunto de dados. Outro ponto observado é que cada expressão tem um conjunto diferente de pontos faciais que representam fortemente uma expressão facial, uma vez que a importância dos pontos faciais varia de acordo com a expressão gramatical. Um ponto verificado é a expressão ênfase que depende mais da boca, das sobrancelhas esquerda e direita, enquanto a expressão condicional depende do contorno boca e face. Outra observação que pode ser feita facilmente ao analisar essas EFGs é que os pontos do rosto pertencentes à boca são os mais importantes em comparação com outros pontos faciais. Esta informação pode ser fundamental na concepção de várias aplicações que dependem de poucos pontos faciais críticos (BHUVAN et al., 2016). O número de expressões faciais é finito, pois o ser humano pode realizar apenas um número limitado de movimentos faciais. Existe um conjunto de expressões faciais dentro de cada língua de sinais que pode ser considerado como EFG (UDDIN, 2015). A abordagem automatizada de reconhecimento das EFGs em um discurso enfrenta dois desafios principais: as EFGs podem variar de pessoa para pessoa; durante um discurso pode ocorrer frequentes oclusões faciais (UDDIN, 2015).

Os sinais comunicativos podem ser difíceis de analisar afirmam Benitez-Quiroz, Wilbur e Martinez (2016), por isso, analisar uma expressão isoladamente é crucial para entender e analisar suas expressões faciais gramaticais. Já em outro estudo de Benitez-Quiroz et al. (2014) observaram que existem características discriminantes que podem ser usadas para separar *Condicionalis* de não *Condicionalis*, *Interrogativa (qu)*, *Interrogativa (s/n)*, *Interrogativa (dúvida)*, *Afirmativa* e *Negativa*.

Para avaliar o conteúdo léxico, Marshall et al. (2015) utilizaram crianças entre 8 e 11 anos para seus experimentos o que é um ponto complexo do trabalho, visto que as crianças nesse estudo em sua maioria esqueciam de fazer a EFG durante a realização da sentença.

Problemas de sincronização com características manuais, movimentos da face e fusão com técnicas de classificação precisam ser abordados e melhor investigadas, concluíram o estudo de Caridakis, Asteriadis e Karpouzis (2014). Dados de treinamento conduzidos por gravações de um sinalizador (feminino) e o um sinalizador diferente (masculino), há diferenças entre essas expressões faciais devido a aspectos idiossincráticos (maneira de ver, de sentir e de reagir, própria de cada pessoa) dos sinalizadores individuais o que pode

causar uma dificuldade nas análises segundo Kacorri e Huenerfauth (2016). Em (KUMAR; ROY; DOGRA, 2018) foi observado que as expressões faciais do sinalizador mudam com frequência para fornecer o significado e o sentido exatos ao gesto realizado. E é possível que uma expressão facial corresponda a muitos gestos de sinais, da mesma forma, os mesmos gestos de sinais podem corresponder a múltiplas expressões faciais.

2.2.2 Natureza dos dados

Nas análises dos estudos, foi possível notar um conjunto comum de informações referentes aos dados utilizados nos experimentos:

1. **natureza:** imagens ou vídeos;
2. **formato:** *pixels* ou geometria do rosto;
3. **origem:** estudos primários ou secundários.
4. **dimensionalidade:** 2D ou 3D;
5. **ambiente:** controlado ou informal;

De maneira resumida, a maioria dos estudos analisados utilizaram vídeos em suas análises devido ao escopo temporal do problema, pois as expressões faciais gramaticais fazem sentido na execução da frase, que acarreta uma sequência de ações, ocorrendo sequencialmente ou simultaneamente. Apenas um estudo aqui levantado abordou a oclusão da face. O estudo de (LIU et al., 2014) utilizou um rastreador de rosto 3D deformável baseado em um conjunto adaptativo de *Active Shape Models (ASMs)*⁶, que foi capaz de rastrear na presença de grandes movimentos de cabeça as oclusões do rosto ocasionadas pelas mãos que podem ocorrer durante a sinalização., tendo em vista que ao executar as frases em língua de sinais, é possível que a mão passe pela frente da face, causando uma oclusão na expressão facial e prejudicando a aquisição dos dados. Em trabalhos anteriores, os autores Nguyen e Ranganath (2012) e Ding e Martinez (2010) utilizam técnicas para os casos que ocorrem oclusão.

⁶ São modelos estatísticos da forma de objetos que se deformam iterativamente para caber em um exemplo do objeto em uma nova imagem, desenvolvida por Tim Cootes e Chris Taylor em 1995 (COOTES et al., 1995).

2.2.3 Bases de dados

Os autores Kumar, Roy e Dogra (2018), Freitas et al. (2017), Freitas et al. (2014), Benitez-Quiroz et al. (2014), Marshall et al. (2015) e Benitez-Quiroz, Wilbur e Martinez (2016) adquiriram seus próprios conjunto de dados, sendo assim dados primários. Esses conjuntos de dados estão disponíveis para uso de futuros estudos.

Em Kumar, Roy e Dogra (2018) foi construído um conjunto de dados contendo gestos de Língua de Sinais Indiana, capturados simultaneamente usando os sensores *Leap Motion* e *Kinect*. O conjunto de dados tem 51 gestos de palavras com sinais dinâmicos diferentes, dos quais 31 palavras de sinais foram realizadas usando as duas mãos e o restante foi realizado usando uma única mão (apenas a mão direita). Os gestos foram realizados por 10 sinalizadores diferentes que são os alunos de “*Anushruti*” (uma escola para deficientes auditivos do Instituto Indiano de Tecnologia, Roorkee, Índia). Cada entrada de sinal foi repetida oito vezes por cada sinalizador que perfaz um total de 4080 gestos de sinais diferentes para a construção do conjunto de dados. Do conjunto completo de dados, 50% são usados no treinamento, 25% para os dados de validação e o que resta dos 25% são usados para fins de teste. Os autores utilizaram frases que despertam emoções para analisar gramaticalmente as expressões.

O conjunto de Freitas et al. (2014) é composto por dezoito vídeos gravados usando o sensor *Microsoft Kinect*. Em cada vídeo, um usuário executa (cinco vezes), em frente ao sensor, cinco frases em Libras que exigem o uso de uma expressão facial gramatical. As imagens permitiram uma rotulação manual de cada arquivo por um especialista, fornecendo uma verdade solo para a classificação. O conjunto de dados está organizado em 36 arquivos: 18 arquivos *datapoints* (pontos cartesianos x, y, z) e 18 arquivos *targets* (rótulos binários). Os vídeos são gravados por dois usuários fluentes em Libras. Os autores Uddin (2015) e Bhuvan et al. (2016) também utilizaram esse conjunto de dados para seus experimentos.

O trabalho dos autores Marshall et al. (2015) utilizou as sentenças descritas em Herman, Holmes e Woll (1999) e gravou vídeos com onze crianças surdas. O estudo de Kacorri e Huenerfauth (2016) utilizou o conjunto de dados descrito em Kacorri et al. (2016), que consiste em gravações de vídeo de sinalizadores que executam ASL com anotações rotuladas no vídeo quando cada um dos cinco tipos de expressões faciais e gramaticais são

executadas. O conjunto consiste em 199 exemplos de expressões faciais realizadas por uma sinalizadora registrada na Universidade de Boston.

A aquisição dos dados do trabalho de Benitez-Quiroz, Wilbur e Martinez (2016) foi composto por 184 pessoas e, além de seu conjunto de dados primários, também utilizou o conjunto de dados ASL de Benitez-Quiroz et al. (2014), que inclui quinze sinalizadores de ASL. Este é o maior conjunto de anotações disponíveis para o estudo de marcadores não-manuais na ASL. Já Liu et al. (2013) e Liu et al. (2014) utilizaram o conjunto de dados coletados por Neidle (2000) e seu grupo de pesquisa. O *corpus* consistiu em 60 vídeos ASL com sentenças gravadas, nessas sentenças possuem 94 marcadores gramaticais não manuais. Caridakis, Asteriadis e Karpouzis (2014) utilizam o banco de dados de Cascia, Sclaroff e Athitsos (2000). O quadro 3 lista os artigos analisados em relação ao conjunto comum de informações levantadas.

Quadro 3 – Informações sobre os dados utilizados nas experimentações e aplicações dos estudos referentes as EFGs. Abreviações: Img - imagem; Víd - vídeo; Pix - pixel; Geo - geometria; 1^a - primeira; 2^a segunda; Con - controlado; Inf - informal

Trabalhos	Natureza		Formato		Origem		Dimensão		Ambiente	
	Img.	Víd.	Pix.	Geo.	1 ^a	2 ^a	2D	3D	Con.	Inf.
Kurman et al. (2018)		x	x		x			x	x	
Freitas et al. (2017)		x	x		x		x		x	
Kacorri et al (2016)		x	x		x		x		x	
Bhuvan et al. (2016)		x	x			x	x		x	
Benitez-Quiroz et al. (2016)	x		x		x	x	x			
Uddin (2015)		x	x			x	x		x	
Marshall et al. (2015)		x	x		x	x	x			x
Benitez-Quiroz et al. (2014)		x	x		x	x	x		x	
Freitas et al. (2014)		x	x		x		x		x	
Liu et al. (2014)		x	x			x		x	x	
Caridakis et al. (2014)	x		x			x		x		x
Liu et al. (2013)		x	x	x		x		x	x	

Fonte: Maria Eduarda de Araújo Cardoso, 2018

2.2.4 Técnicas utilizadas e métodos de avaliação de desempenho

Nesta seção um resumo de como foi realizado os trabalhos de cada estudo aqui levantado é apresentado, assim como cada técnica foi aplicada e a metodologia para analisar e avaliar os resultados. Na análise dos artigos é perceptível o uso de aprendizado máquina para resolver o problema de reconhecimento das EFGs na língua de sinais. As medidas

tradicionalmente usadas nessa área, como: *F-score*, acurácia, curva ROC também são utilizadas nos estudos aqui levantados. No quadro 4 está listado as técnicas e os métodos utilizados nos estudos.

Quadro 4 – Técnicas e métodos utilizados nos estudos levantados

Trabalhos	Técnicas	Método de avaliação
Bhuvan et al. (2016)	<i>Multilayer Perceptron</i> <i>BFGS</i> Bayes Network RBF Classifier Bootstrap Aggregating Adaboost Logit Boost Random Forest	Curva ROC
Benitez-Quiroz et al. (2014) Benitez-Quiroz et al. (2016)	FACS	Testes estatísticos
Kacorri et al(2016)	Continuous Profile Models Visage <i>FAP</i>	DTW
Uddin (2015)	Ada-Random Forests Ada-bost	<i>F-score</i>
Marshall et al. (2015)	Análise de vídeo	Testes estatísticos
Caradakis et al. (2014)	<i>Lucas-Kanade</i> Redes recorrentes	Acurácia
Liu et al. (2013) Liu et al. (2014)	<i>Conditional Random Field</i>	Acurácia
Freitas et al (2015) Freitas et al (2017)	<i>Multilayer Perceptron</i>	<i>F-score</i>

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Bhuvan et al. (2016) realizaram experimentações exaustivas com oito diferentes algoritmos de aprendizado de máquina para detectar nove diferentes tipos de expressões faciais gramaticais modelados como problema de classificação binária diferente para cada expressão. Os cenários foram modelados com testes dependentes e testes independentes do usuário (treina com o sinalizador a e testa com o sinalizador b). As redes neurais utilizadas para o reconhecimento das EFGs foram: *MultiLayer Perceptron* (MLP) com uma camada oculta, variando de dois a dez neurônios, a *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) ⁷ para minimizar o erro quadrático, a *Bayes Network* considerando apenas o gráfico acíclico, a *RBF Classifier* (RBFC) também para minimizar o erro quadrático, e a Floresta aleatória considerando uma floresta com 10 árvores sem limite para a profundidade máxima. A

⁷ É um algoritmo popular para estimação de parâmetros na aprendizagem de máquina, é um método iterativo para resolver problemas de otimização não linear irrestritos. (HEAD; ZERNER, 1985)

rede *Bootstrap Aggregating (Bagging)*, a *Adaboost* e a *Logit Boost* todas considerando árvores de decisão. A avaliação do estudo é feita considerando a Área sob Curvas de Características Operacionais do Receptor (ROC)⁸. Por meio desse experimento foi obtido como resultado o classificador ótimo para cada expressão e obteve área ROC maior que 0,95 para cada expressão. Na conclusão do estudo é mostrada a importância de cada ponto facial na detecção de cada tipo de expressão, e é indicado que um estudo futuro precisa ser conduzido para entender a importância das características geométricas formadas por um grupo de pontos faciais na detecção de uma expressão particular. O autor afirma que pode ser fundamental para pesquisas futuras e para várias aplicações utilizar pontos faciais como base para a tomada de decisões.

O trabalho de Benitez-Quiroz et al. (2014) investigou sobre as anotações que são tipicamente realizadas através de uma análise visual cuidadosa para identificar os marcadores gramaticais que ocorrem em grande quantidade de sequências de vídeo. A ferramenta *ELAN* foi projetada especificamente para este propósito. Neste estudo, é obtido uma representação linguística do rosto e uma abordagem computacional é então empregada para determinar a combinação dessas características consistentemente observadas em cada classe da expressão facial na sentença, mas não com outras. O modelo linguístico resultante prova ser capaz de discriminar entre nove classes diferentes de frases - *Condicional*, *Interrogativa (dúvida)*, *Interrogativa (qu)*, *Afirmativa* e *Interrogativa (s/n)* em sua polaridade *Afirmativa*. As análises descritas, sugerem fortemente que existem características discriminantes que podem ser usadas para separar essas EFGS. Além disso, o estudo mostrou que para cada uma destas sentenças, exceto as questões de punho negativo na ASL, os recursos discriminantes separam as estruturas negativas de suas contrapartes positivas. Com os resultados desse estudo os autores estudaram a expressão facial negativa para analisar seu comportamento isoladamente nas sentenças.

Assim, Benitez-Quiroz, Wilbur e Martinez (2016) afirmam que as expressões faciais de emoção possam ter evoluído a partir do desenvolvimento dos músculos faciais utilizados na regulação sensorial e posteriormente adaptados para expressar o julgamento moral negativo o qual inclui as expressões de raiva, desgosto e desprezo. O estudo baseou-se na hipótese de que essas expressões faciais de julgamento moral negativo evoluíram para uma expressão facial de negação regularmente utilizada como marcador gramatical na

⁸ ROC é uma métrica de desempenho econômica para avaliar o sistema classificador binário, criado por traçar a taxa positiva verdadeira contra a taxa de falso positivo.

linguagem humana, isto é, pessoas de culturas diferentes que expressam negação usam os mesmos músculos faciais que aqueles empregados para expressar julgamento moral negativo. A produção de EFGs são únicas (EKMAN; SORENSON; FRIESEN, 1969), o que significa que as articulações faciais (conhecidas como Unidades de Ação, AUs) são utilizadas para produzir essas expressões distintas entre si. Diferentes articulações faciais são identificadas com um número de AU distinto, e cada AU envolve uma ativação única dos músculos faciais.

O trabalho de Benitez-Quiroz, Wilbur e Martinez (2016), teve como objetivo analisar partes da face humana que se movimentam durante uma EFG de negação, utilizando imagens de 4000x3000 *pixels*. As imagens foram manualmente codificadas utilizando FACS (*Facial Action Coding System*)⁹ para identificar as ações utilizadas (AUs) por cada participante. O autor usa a demarcação de AUs para demarcar o ponto da face o qual está sendo movimentado. Exemplos de AU são: a demarcação para as sobrancelhas representa a AU4, para representar o queixo utilizou a AU17, para a marcação da boca AU24 e para bochechas AU14. Neste trabalho foi utilizado estatísticas para análise de resultados, contabilizando o uso de AUs de cada participante ao executar uma EFG na sentença.

Foram realizados quatro experimentos onde cada experimento analisava um conceito. O experimento 1 analisou a expressão não verbal de negação, o 2 a co-articulação da negação do rosto na fala, o experimento 3 o marcador gramatical (não manual) na assinatura e por último o experimento 4 a frequência de produção. Os resultados obtidos por meio desses experimentos demonstraram que os participantes utilizaram consistentemente as AUs: 4, 17 simultaneamente (sobrancelhas e queixo) e 14 ou 24, ou ambos juntos (bochechas e boca). A consistência do uso de AUs 4,17 e 14 ou 24 é comparável à consistência observada no uso de AU em expressões faciais universais de emoção (> 70%) em estudos realizados por Du, Tao e Martinez (2014), e assim é afirmado que uma expressão facial de negação que é produzida utilizando os mesmos AUs por pessoas de diferentes culturas, de forma que esta expressão é universal.

Para criar conteúdos acessíveis para usuários surdos, Kacorri e Huenerfauth (2016) investigou a sintetização automática de animações da Língua Sinais Americana, incluindo expressões faciais e movimentos de cabeças gramaticalmente importantes. Os experimentos

⁹ é um sistema para taxonomizar expressões faciais humanas, originalmente desenvolvido por Paul Ekman e Wallace Friesen em 1976 por meio dos estudos pioneiros do Dr. Carl-Herman Hjortsjö (FRIESEN; EKMAN, 1978)

foram realizados com base em gravações de sinalizadores que realizam vários tipos de movimentos sintáticos de face e cabeça (que incluem variação idiossincrática (expressão usada para fazer referência, muitas vezes, a situações incomuns ou até impróprias), foi utilizado o Modelos Contínuos de Perfil (CPMs), que alinha um conjunto de dados da série temporal relacionados, enquanto contabiliza mudanças na amplitude na identificação de um “rastreamento latente” essencial do desempenho, para uso na produção de animações ASL. A técnica de avaliação utilizada foi baseada em métricas e no estudo com usuários surdos que indicaram que essa abordagem era mais eficaz do que um método prévio para produzir animações.

Para a extração das informações de movimento de face e cabeça do vídeo, foi usado um rastreador de rosto chamado Visage ¹⁰ que foi usado para produzir um conjunto de parâmetros de animação facial MPEG4 ¹¹. Para produzir um conjunto de parâmetros de animação facial do FAP (*Facial Animation Parameters*) utilizando o padrão MPEG4 para cada quadro de vídeo. A técnica utilizada foi Modelo de Perfil Contínuo aplicado no modelo para alinhar o tempo e integrar coerentemente dados de séries temporais de múltiplas execuções de expressão facial na Língua Sinais Americana. Este modelo foi previamente avaliado em sinais de fala e em outros dados biológicos de séries temporais (LISTGARTEN et al., 2004). O CPM foi treinado e obtidos os traços latentes. As gravações no conjunto de treino têm 14 dimensões, correspondentes aos 14 traços faciais abordados no estudo. Para avaliar os resultados, foi utilizado o algoritmo *Dynamic time warping* (DTW) um algoritmo para comparar e alinhar duas séries temporais. Esses autores compararam seus resultados com trabalhos desenvolvidos por eles mesmos, e concluíram que obtendo o traço latente com o CPM, por meio de uma avaliação métrica e de um estudo de usuário experimental, é afirmado que as expressões faciais geradas pelos modelos CPM produzem expressões faciais de alta qualidade, expressões com mais marcações fortes o que facilita o reconhecimento dessas expressões em diversos sinalizadores.

Já em Uddin (2015) é apresentada uma nova estrutura *Ada-Random Forests* para reconhecer as expressões faciais gramaticais utilizadas na Língua Brasileira de Sinais. Nesta abordagem, um algoritmo de seleção de recurso *Ada-Boost* é aplicado para selecionar subconjuntos de recursos compactos dos inúmeros recursos extra-extraídos para reduzir o

¹⁰ <https://visagetechologies.com/products-and-services/visagesdk/facetrack>.

¹¹ é um padrão utilizado primeiramente para compressão de dados digitais de áudio e vídeo (AV) para cada quadro de vídeo: esses valores representam os movimentos de face ou cabeça do ser humano no vídeo.

tempo computacional, bem como para melhorar a taxa de reconhecimento do sistema. O objetivo do estudo foi classificar cada quadro como uma EFG particular, ou como uma expressão facial neutra na configuração de classificação binária por meio de um quadro de reconhecimento de duas camadas. Na primeira camada, as características relevantes e discriminativas são selecionadas a partir de recursos brutos disponíveis via *Ada-Boost*; enquanto na segunda camada, os classificadores de florestas aleatórias são construídos alimentando os recursos selecionados do *Ada-Boost* para reconhecer as expressões faciais gramaticais. Os resultados foram comparados a outros já publicados na literatura e concluem que a estrutura de *Ada-Random Forests* supera o classificador de *Multilayer Perceptron* (FREITAS et al., 2014). O resultado deste estudo teve um aumento de 7,53% de classificação desempenho acima do resultado da *Multilayer Perceptron*.

No trabalho de Marshall et al. (2015) é estudada a informação gramatical que é expressa por meio do movimento e mudanças de configurações das mãos e da face. A questão levantada nesse estudo é: como o impacto da linguagem afetará a aquisição desses tipos de dispositivos linguísticos em crianças, e foram investigados as habilidades de repetição de sentenças em um grupo de onze crianças surdas que exibem SLI¹² na língua de sinais britânica (BSL).

Cada sentença em língua de sinais recebeu uma pontuação final por critérios realizados na execução da LS, para levantar resultados de quanto cada um é importante na EFG. As crianças foram testadas por um intérprete nativo. As sentenças foram apresentadas uma vez às crianças em um computador *laptop* e todas as respostas foram gravadas em vídeo. Cada criança observou as instruções da tarefa e pôde fazer perguntas clarificadoras. Os participantes então assistiram a três frases de prática e depois de cada uma delas, foram encorajadas a repeti-la da forma mais exata possível. Mais 17 frases foram mostradas e as respostas foram gravadas em vídeo para posterior pontuação e análise. Esses pontos eram somados conforme as sentenças em LS iam sendo realizadas. As funções avaliadas são: o conteúdo léxico, ordem de sinalização, significado e expressão facial. A pontuação é realizada conforme a conformidade desses pontos. A pontuação geral é apresentada nos resultados do estudo em forma de gráfico com os valores de cada função realizada, também resultados como desvio padrão e média foram calculados. Os autores afirmaram que a expressão facial gramatical altera o significado da sentença, logo nessa função a maioria

¹² O SLI é um transtorno heterogêneo, com variabilidade individual considerável na gravidade e perfil das deficiências linguísticas.

dos usuários perdiam pontos significantes. Portanto, o usuário ao gesticular a frase em LS com omissões de expressão facial ou mudanças de significado por falta de gramática facial, além de perder pontos a frase não teria o significado correto.

Caridakis, Asteriadis e Karpouzis (2014) abordam a incorporação de pistas não-manuais no reconhecimento automático da língua de sinais, mais especificamente o olhar do olho, a pose da cabeça e as expressões faciais são os elementos discutidos em relação à sua função gramatical e a sintática dos mesmos, e uma maneira de incluí-los na fase de reconhecimento. Esses elementos são rastreados usando um rastreador iterativo, 3-pirâmide *Lucas-Kanade* (LUCAS; KANADE et al., 1981). O rastreamento de *Lucas-Kanade* é um dos rastreadores mais difundidos e usados em bibliografias, a escolha desse rastreador baseou-se no fato de que ele pode rastrear com precisão e eficácia características em uma grande variedade de circunstâncias afirmaram os autores.

Este estudo utiliza redes recorrentes para classificar as expressões faciais gramaticais. Este tipo de rede difere das redes convencionais de *feedforward* na medida em que uma primeira camada tem uma ligação recorrente. O atraso nessa conexão armazena valores do passo anterior que podem ser usados no tempo atual, fornecendo assim, o elemento de memória. Embora os autores tenham seguido uma abordagem que incluía apenas uma única camada de conexões recorrentes, na realidade, a rede possui a capacidade de aprender padrões com um comprimento maior e os valores atuais são afetados por todos os valores anteriores e não apenas pelo último, e assim foi adotada uma rede de duas camadas com *feedback* da saída da primeira camada para a entrada de primeira camada. A camada de entrada da rede utilizada possui 25 neurônios, a camada oculta possui 20 neurônios e a camada de saída tem tantos neurônios quanto as possíveis classes correspondentes às expressões faciais. Foi analisado os pontos da boca, olhos e ponta do nariz no reconhecimento. E aplicado a técnica de matriz de covariância para calcular o erro médio de cada EFG das sentenças. A matriz de erros é calculada com base nas estimativas do algoritmo de pose de cabeça e as variáveis do ângulo da pose da cabeça. Para testar a validade do sistema para estimar com precisão os valores de rotação de cabeças, uma série de experiências foram realizadas no conjunto de dados. Os autores não apresentam um resultado final de acurácia, mas apresenta uma tabela com a assertividade de cada expressão referente à rede utilizada. O estudo concluiu que o trabalho apresentado fornece uma base sólida para uma investigação mais aprofundada da incorporação de recursos não-manuais na incorporação automática de língua de sinais, pesquisar este aspecto do

reconhecimento de língua de sinais está longe de ser completo afirma os autores. Problemas de sincronização com características manuais e fusão com técnicas de classificação baseadas nesses recursos precisam ser abordados e investigados.

Liu et al. (2013) desenvolveram um sistema de reconhecimento automático para marcadores gramaticais não-manuais com base em análise multiescala e espaciotemporal de poses de cabeça e expressões faciais. Foi realizado o reconhecimento de gestos de sobrancelha e cabeça, bem como suas fases temporais. Os autores melhoraram a detecção de marcadores gramaticais não manuais em ASL, e propuseram uma estrutura hierárquica do CRF (é um modelo estocástico comumente usado para marcar e segmentar sequências de dados ou extrair informações de documentos) para reconhecer automaticamente os gestos relevantes das sequências de vídeo, incluindo suas fases temporais. Foi utilizado um conjunto adaptativo de rastreadores de rosto para obter pontos de referência e poses de cabeças com o modelo *Active Shape* (ASM)(COOTES et al., 1995). O artigo afirma que o uso desta técnica pode melhorar a detecção de marcadores gramaticais não-manuais em ASL em comparação com uma abordagem que usa apenas características de nível inferior para expressões faciais. Além disso, essa abordagem foi útil na diferenciação de outros usos dos gestos de sobrancelha de seu papel na sinalização de informações gramaticais. Os autores afirmam que os contornos temporais são diferentes para expressões faciais gramaticais e isso seria explorado em pesquisas futuras como foi em Liu et al. (2014).

A técnica que os autores Liu et al. (2014) desenvolveram foi um sistema de reconhecimento automático para marcadores gramaticais não-manuais. A análise leva em conta os componentes gestuais, tais como sobrancelhas levantadas ou abaixadas e diferentes tipos de movimentos periódicos da cabeça. Para avançar o estado da arte no reconhecimento de marcadores gramaticais não-manuais, os autores propuseram uma nova abordagem de aprendizado multiescala que explore características faciais espacialmente temporariamente de baixo nível e de alto nível. Os recursos de baixo nível são baseados em informações sobre geometria e aparência facial, bem como a pose da cabeça, e são obtidas por meio de um rastreamento de rosto baseado em modelo 3D deformável e preciso. Este rastreador 3D funcionou bem com os grandes movimentos de cabeças e oclusões do rosto pelas mãos que ocorrem durante a sinalização. Outra vantagem vista dessa abordagem foi o modelo de face 3D que elimina a necessidade de alinhamento de pose facial necessário nas abordagens 2D, que muitas vezes é uma fonte de erros significativos na estimativa de pose facial e reconhecimento de expressão.

Os recursos de alto nível baseiam-se na identificação de eventos gestuais, de duração variável, que constituem os componentes de marcadores linguísticos não-manuais. O reconhecimento dos marcadores foram eventos como sobrancelhas levantadas e abaixadas e movimento de cabeça. Esses eventos foram divididos em fases temporais. Para a análise foi separado o movimento de transição do início da parcela linguisticamente significativa do evento e separado o núcleo do evento do movimento de transição que ocorre quando os articuladores retornam ao ponto morto que seria a posição para o final do evento. Os autores relatam que essa partição é essencial para a localização temporalmente precisa dos marcadores gramaticais, o que não pode ser alcançado nesse nível de precisão com métodos prévios de visão por computador.

Para reconhecer expressões faciais gramaticais, uma estrutura hierárquica de Campo Aleatório Condicional (CRF) foi utilizada, para reconhecer as sobrancelhas levantadas ou abaixadas, assentamentos de cabeça em sequências de vídeos. O CRF é um modelo probabilístico proposto por Lafferty, McCallum e Pereira (2001) e tem sido amplamente utilizado para previsão estruturada, como segmentação de imagens, detecção de eventos e rastreamento de objetos. O modelo considera não apenas as dependências entre observações e estados, mas também interações entre estados. A técnica de matriz de confusão foi utilizada para avaliar não-marcadores das sentenças, e o resultado foi que: (1) menos regiões NM são incorretamente detectadas como não-marcadores gramaticais manuais; (2) a nova abordagem reconhece e distingue mais precisamente os cinco tipos de marcadores e aponta a melhoria na precisão temporal dos resultados de detecção obtidos pelo método proposto.

Para o reconhecimento das EFgs e dos gestos em Kumar, Roy e Dogra (2018) foi utilizado o *Hidden Markov Model* (HMM), o processo de reconhecimento é realizado separadamente para gestos manuais e expressão facial gramatical para diferentes estados de HMM e componentes de mistura de Gauss. Também foi aplicada a abordagem de Combinação de Classificação Bayesiana Independente (IBCC) para combinar a decisão de diferentes modalidades para melhorar o desempenho do reconhecimento da EFG. Os autores usaram 71 pontos da face para o reconhecimento da EFG e obtiveram 96,05% e 94,27% de taxa de acurácia no reconhecimento em gestos e expressões faciais gramaticais.

Os trabalhos de Freitas et al. (2014) e Freitas et al. (2017) utilizaram aprendizado de máquina para o desenvolvimento de métodos e técnicas empregados para implementar a aprendizagem indutiva. A técnica escolhida é MLP. Os experimentos cobriram nove tipos

de frases com um senso gramatical distinto, e o problema de reconhecimento foi modelado por meio de um conjunto de tarefas de classificação binária. Vale ressaltar que o estudo é o primeiro no reconhecimento automatizado de EFGs em Libras. Assim, as conclusões sobre a complexidade das expressões representam uma contribuição inédita afirma o estudo.

Após os experimentos com a MLP, foram usadas para os resultados matriz de confusão para que obtivesse a *F-score* de cada EFG. Os autores discutem os resultados em termos de *F-scores*, erros de borda e erros internos, mostrando a complexidade do problema de segmentação automática de cada EFG. As EFGs de maior complexidade de segmentação são as que envolvem movimentos repetitivos da cabeça (*Afirmativa* e *Negativa*), para as quais *F-scores* variando de 0.74 a 0.87 e 0.67 a 0.95 são obtidos, a depender das dificuldades impostas ao classificador pelo projeto dos experimentos. Na EFG de menor complexidade (*Foco*), os resultados variam de 0.88 a 0.96.

2.2.5 Considerações finais

Por meio do estudo realizado foram levantados doze artigos que abrangeram expressões faciais gramaticais na língua de sinais. Tais estudos trabalham com objetivos diversificados, como mostra o quadro 5:

Quadro 5 – Finalidade dos estudos levantados

Autores	Estudo
Bhuvan et al. (2016), Uddin (2015), Caridakis, Asteriadis e Karpouzis (2014), Liu et al. (2013), Liu et al. (2014), Freitas et al. (2014), Freitas et al. (2017)	Reconhecimento da expressão facial gramatical
Benitez-Quiroz, Wilbur e Martinez (2016), Kumar, Roy e Dogra (2018)	Análise da face na ocorrência de uma EFG
Marshall et al. (2015)	Prosódia da expressão facial gramatical

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A revisão da literatura sob o escopo em expressões faciais gramaticais teve o intuito de identificar as principais técnicas de extração de características, a importância da informação temporal e os métodos de avaliação para reconhecimento das EFG, também como as técnicas de avaliação de resultados. O levantamento também mostrou que tais estudos se basearam na identificação de desafios ao reconhecer automaticamente uma EFG, na complexidade no seu reconhecimento além da importância da EFG em frases da LSs.

Nos estudos realizados ainda são nítidos os problemas relacionados com dependências temporais, pois a importância da representação temporal está relacionada à execução da expressão facial durante a sinalização em uma sentença, sinalizando um ou mais sinais. E só alguns estudos abordaram a representação temporal. A oclusão também é um problema que ainda não foi analisado com precisão, como já dito nessa análise só um estudo abordou esse assunto.

3 Rede Neural Artificial Multilayer Perceptron e Mistura de Especialistas

Diferentes técnicas foram desenvolvidas para implementar o aprendizado de máquina, por exemplo: árvores de decisão, máquinas de vetores suporte, redes neurais artificiais e arquiteturas de comitê de máquinas. Este trabalho tem especial interesse na rede neural artificial *Multilayer Perceptron* (MLP) e na arquitetura de comitê de máquinas mistura de especialistas (ME). Os experimentos discutidos neste trabalho foram primeiramente executados usando MLP, técnica mais simples que a ME, para que fosse possível explorar o contexto de estudo - segmentação das EFGs, com mais agilidade, já que execuções desta arquitetura demandam menos tempo computacional do que as MEs. Diante de um conhecimento mais apurado sobre o problema de segmentação em questão, as MEs foram aplicadas.

As MLPs são extratoras de características, pois sua estrutura é composta por camadas ocultas que são formadas por neurônios *Perceptron* interconectados (HAYKIN, 2009). Esses neurônios são responsáveis por realizar, localmente, a discretização do erro envolvido na tarefa de predição. Por isso a escolha dessa rede. As MEs são estruturas capazes de explorar regiões diferentes do espaço de busca com estruturas de aprendizado especializadas (os especialistas). Diante da especificidade de caracterização de cada EFG, o uso de ME é motivado.

Este capítulo é destinado a apresentar os conceitos básicos referentes à MLP e à ME e está organizado da seguinte forma: a seção 3.1 apresenta conceitos gerais sobre redes neurais artificiais, e detalha a rede neural artificial MLP; a seção 3.2 conceitualiza a ME.

3.1 Multilayer Perceptron

Redes Neurais Artificiais (RNA) são modelos matemáticos cujo projeto é inspirado na estrutura neural de organismos inteligentes, capazes de adquirir conhecimento por meio da experiência. Fausett (1994) define que RNAs estão fundamentadas nos seguintes conceitos:

1. os neurônios são os elementos responsáveis por processar as informações;
2. a RNA é um modelo composto por uma camada de entrada (neurônios de entrada), que recebe a informação a ser processada, e uma camada de saída (neurônios de

- saída), que representa a resposta da rede, tipicamente uma classe à qual cada valor de entrada foi associado;
3. por meio de conexões, ou sinapses, os sinais são transmitidos entre as camadas de neurônios;
 4. cada conexão esta associada a um peso que influencia o sinal transmitido;
 5. a saída de cada neurônio é determinada por uma função de ativação aplicada à sua entrada.

Assim como a RNA, as estruturas referentes aos neurônios também são modelos matemáticos inspirados no modelo de funcionamento de um neurônio biológico. Várias destas estruturas organizadas dentro de uma arquitetura caracterizam, de forma simplificada, uma RNA.

As primeiras teorias desenvolvidas na área de redes neurais artificiais foram propostas por: McCulloch e Pitts (1943), Hebb (1949) e Rosenblatt (1958). Este autores introduziram, respectivamente, o primeiro modelo de neurônio artificial, a primeira estratégia de aprendizado, e o modelo do neurônio *Perceptron*.

De forma simplificada, uma RNA é composta por várias unidades simples (os neurônios), responsáveis pelo processamento paralelo e distribuído da informação, seguindo a metáfora de funcionamento do cérebro humano. Algoritmos de aprendizado são associados a essa estrutura de forma a dotá-la da capacidade de aprender e de tomar decisões. Os neurônios são conectados por meio de canais de comunicação (sinapses) associados a pesos (força da sinapse). Por meio desses canais de comunicação, o resultado das operações localmente executadas em cada unidade é transmitido pelo rede. O comportamento geral de uma RNA é decorrente, portanto, das interações entre as unidades de processamento. A figura 8 mostra a operação de uma unidade de processamento.

Figura 8 – Esquema de unidade McCulloch - Pitts

A figura 8 mostra os sinais que são apresentados como entrada para o neurônio $(x_1, x_2, \dots, x_n)$. Cada um desses sinais é multiplicado pelo peso sináptico $(w_1, w_2, \dots, w_n)$ associado à conexão entre as entradas e o neurônio. Essa multiplicação implementa a influência de cada entrada à saída que será emitida pelo neurônio. Na sequência é realizada a soma dos sinais ponderados, o que produz um nível de atividade no neurônio. Se esse nível de atividade exceder um certo limite, chamado de *threshold*, a unidade produz uma determinada resposta de saída (y) . Uma função de ativação $(\varphi.)$ relaciona o nível de atividade, o *threshold* e a resposta de saída.

As RNAs podem ser categorizadas em três tipos (HAYKIN, 2009): i) camada única; ii) multi-camadas; iii) recorrentes. Este trabalho aplica uma RNS multi-camadas.

O modelo de neurônio utilizado nas implementações deste trabalho é referente ao *Perceptron*. A equação que implementa o processamento do *Perceptron* é definida como (HAYKIN, 2009):

$$y = \varphi \left(\sum_{i=1}^n w_i x_i + b \right) = \varphi (w^T x + b),$$

em que n é o número de entradas do neurônio, w_i é o peso da conexão referente à entrada i , x_i é o valor da entrada i , b é um *bias*. O *bias* age exatamente como um peso em uma conexão de uma unidade cuja ativação é sempre 1. Se o valor do *bias* é positivo, o efeito é o de aumentar a força de influência das entradas para o neurônio. Se é negativo, o efeito é de diminuir esta força. O φ é a função de ativação do neurônio. A função de ativação é aquela que processa o sinal gerado pela combinação linear das entradas e dos pesos das sinapses para gerar o sinal de saída do neurônio. Os três tipos de função de ativação mais utilizados são: função limiar (*step*), função limiar por partes e função sigmóide (HAYKIN, 2009).

A *Multilayer Perceptron* é uma rede de aprendizado supervisionado, que implica em ser necessário fornecer os dados de entrada e comparar a saída gerada pela rede neural com a saída desejada para cada um dos dados. Esse processo é conhecido como treinamento e, por meio dele, é possível ajustar os parâmetros da arquitetura neural para que a saída gerada se aproxime do resultado desejado (FAUSETT, 1994).

Uma MLP é organizada em uma arquitetura em camadas. A primeira é conhecida como camada sensorial, por meio da qual um sinal de entrada (um dado) é recebido. A última é conhecida como camada de saída, por meio da qual se conhece a resposta da MLP para o dado de entrada. Entre elas encontram-se uma ou mais camadas ocultas.

As camadas ocultas e a camada de saída são formadas por neurônios do tipo *Perceptron*, os quais recebem um sinal ponderado, processam o sinal por meio de uma função de ativação e enviam um novo sinal para a próxima camada da MLP ou para o meio externo (se o neurônio fizer parte da camada de saída). A ponderação do sinal que circula pelos neurônios de uma MLP é realizada por meio de sinapses (ou pesos), que precisam ser calibradas de forma que a resposta na camada de saída da MLP, para uma determinada entrada na camada de entrada, seja a resposta desejada. A calibração dessas sinapses pode ser feita por meio de um algoritmo de minimização do erro total cometido pela MLP, conhecido como algoritmo de retropropagação (ou *Backpropagation*).

Em resumo, uma rede neural *Multilayer Perceptron* é tipicamente composta de:

- um conjunto de neurônios sensoriais (ou nós fonte) que constitui a camada de entrada da rede;
- uma ou mais camadas escondidas de neurônios (*Perceptron*) que fazem processamento de sinal a partir de funções de ativação não lineares (diferenciáveis em todos os seus pontos);
- a camada de saída da rede com neurônios que fazem processamento de sinal a partir de funções de ativação não lineares ou funções de ativação lineares (diferenciáveis em todos os seus pontos).

A figura 9 (TAFNER, 1998) segue o exemplo de uma arquitetura genérica de uma rede MLP, na qual os círculos maiores representam os neurônios do tipo *Perceptron*, os quadrados representam as unidades de entrada e as linhas representam as conexões existentes entre os neurônios responsáveis por transferir/propagar, e ponderar, um sinal pela rede. O tipo de propagação, ou transferência, de sinal em uma rede deste tipo é chamado de *feedforward* (propagação para frente), e assim, a informação é propagada de forma progressiva, pois cada neurônio é conectado a todos os neurônios da próxima camada.

O treinamento da MLP envolve três estágios: i) a passagem (*feedforward*) dos dados de treinamento; ii) o cálculo e retropropagação do erro associado a cada neurônio; iii) o ajuste de pesos. Os neurônios fazem dois tipos de processamento:

1. o processamento clássico: ativação do neurônio mediante entradas e uma função de ativação;

Figura 9 – Exemplo de uma rede MLP de múltiplas camadas (4 entradas, 1 camada escondida e 3 saídas)

Fonte: Tafner (1998)

2. o cálculo da informação de erro - computação do gradiente.

O treinamento da MLP otimiza o conjunto de pesos (W) encontrando valores que ajustam o mapeamento que resolve o problema de classificação. Ele é implementado em três estágios que se repetem até que o erro de classificação cometido na saída da rede alcance um limite mínimo desejado (SILVA; PERES; BOSCARIOLI, 2017):

- passagem dos dados de treinamento e cálculo das saídas dos neurônios da camada oculta e de saída seguindo: $z_h = f(v_h) = f(\sum_{j=1}^D (x_j * w_{hj}))$ para a camada oculta e $y_k = f(u_k) = f(\sum_{l=1}^h (z_l * w_{kl}))$ para a de saída;
- cálculo e retropropagação do erro associado aos neurônios, seguindo: $e_k = y_{ik} - \mathbf{y}_k$ como erro dos neurônios de saída, $\epsilon = \frac{1}{2} \sum_{p=1}^k e_p^2$ como erro total cometido pela rede neural, e $\delta_k = e_k * f'(u_k)$ como informação de erro na camada de saída, e $\delta_h = \sum_{p=1}^h (\delta_p * w_{ph} * f'(v_h))$ como informação de erro na camada escondida;
- o ajuste de pesos seguindo: $w_{hD} = w_{hD} + \eta * z_h * \delta_k$ para os pesos na camada de saída, e $w_{kh} = w_{kh} + \eta * x_D * \delta_h$ para os pesos na camada oculta.

em que x é um sinal de entrada, w são pesos associados a neurônios, D é a dimensão do dado de entrada, h é o número de neurônios na camada oculta, k é o número de neurônios na camada de saída, z é o sinal produzido em um neurônio da camada oculta, y

é o sinal produzido em um neurônio da camada de saída, $\mathbf{y}$ é uma saída desejada, f é a função de ativação e f' sua derivada e η é uma taxa de aprendizado.

Implementações da MLP necessitam considerar os seguintes parâmetros (BISHOP, 1995):

- **Número de camadas escondidas:** Geralmente definido empiricamente. Deve-se ter cuidado para não utilizar um número alto de neurônios, pois pode levar a rede a memorizar os dados de treinamento (*overfitting*¹), ao invés de extrair as características gerais que liberarão a generalização. Por outro lado, um número muito pequeno de neurônios pode forçar a rede a gastar tempo em excesso tentando encontrar uma representação ótima. Não há uma regra específica para a escolha do número de neurônios nessa camada, e tipicamente, a escolha do melhor número de neurônios é feita empiricamente e para cada problema sob estudo.
- **Taxa de aprendizado:** O parâmetro taxa de aprendizado tem grande influência durante o processo de treinamento da rede neural. Uma taxa de aprendizado muito baixa torna o aprendizado da rede muito lento, ao passo que uma taxa de aprendizado muito alta causa variações no treinamento e impede a convergência do processo de aprendizado. O valor desse parâmetro deve variar no intervalo $(0, 1]$.
- **Épocas:** São usadas como um possível critério de parada, onde é feita a determinação do momento em que o treinamento de uma rede neural deve ser encerrado. Um número excessivo de épocas pode levar a rede à perda do poder de generalização (*overfitting*). Por outro lado, com um número de épocas pequeno, a rede pode não atingir convergência e gerar *underfitting*.
- **Função de ativação:** Essa função precisa ser derivável em todos os seus pontos.
- **Bias:** Age como um peso em uma conexão de uma unidade cuja ativação é sempre 1. Ele pode ou não ser usado nos neurônios do tipo *Perceptron*.

3.2 Mistura de Especialistas

Misturas de especialistas (ME) vêm sendo aplicadas com sucesso junto a uma ampla gama de problemas em aprendizado de máquina (VILLANUEVA, 2006), (VILLANUEVA et

¹ *Overfitting*: quando o classificador tende a se adaptar a detalhes específicos da base de treinamento, o que pode causar uma redução da taxa de acerto em ambientes de teste do classificador (THEODORIDIS; KOUTROUMBAS et al., 2008)

al., 2011), (LIMA, 2012), (NGUYEN; CHAMROUKHI, 2018), (KARIMU; AZADI, 2018), (VERMA; VIG, 2018). É uma arquitetura modular para aprendizado supervisionado e foi devidamente formalizada por Jacobs et al. (1991). Os autores Jacobs et al. (1991) afirmam que se a retropropagação for usada para treinar uma única rede multicamadas para executar subtarefas diferentes em ocasiões diferentes, geralmente haverá fortes efeitos de interferência que levam ao aprendizado lento e à generalização deficiente. Esse autor afirma que se for possível saber de antemão que um conjunto de dados de treinamento pode ser naturalmente dividido em subconjuntos que correspondem a subtarefas distintas, a interferência pode ser reduzida usando um sistema composto por várias redes “especializadas” e uma rede de controle que decide quais especialistas serão usados para cada dado de treinamento. A ideia de Jacobs et al. (1991) é que a rede *gating* (rede de controle) aloque cada dado para um ou alguns especialistas e, se a saída estiver incorreta, as mudanças de peso sejam localizadas para esses especialistas. Portanto, não há interferência com os pesos de outros especialistas que se especializam em casos bem diferentes. Os especialistas são, portanto, locais, no sentido de que os pesos de um especialista são dissociados dos pesos de outros especialistas (HAMPSHIRE; WAIBEL, 1992). Além disso, muitas vezes serão locais no sentido de que cada especialista será alocado apenas a uma pequena região do espaço de possíveis vetores de entrada (JACOBS et al., 1991). Para ilustrar a maneira como um problema é resolvido mediante a aplicação de uma ME, considere o exemplo de Jacobs et al. (1991), no qual um problema de discriminação de vogais e teste com quatro usuários (NOWLAN, 1990) foi resolvido. Os dados consistiram no primeiro e segundo formantes² das vogais [i], [I], [a] e [A] de 75 usuários (homens, mulheres e crianças) proferidos em um contexto onde a palavra inicia com “h” e termina com “d”³. Os dados formam dois pares de classes sobrepostas, e diferentes especialistas aprendem a se concentrar em um par de classes ou no outro (Figura 10). O experimento foi realizado com a seguinte proporção: todos os modelos foram treinados com dados dos primeiros 50 usuários e testados com

² As características do som de cada segmento vocálico dependem da formação das cavidades supraglóticas que geram as frequências de ressonância no trato vocal que são denominadas formantes. A frequência da primeira formante (F1) e da segunda formante (F2) são essenciais para determinar a característica de uma vogal. Na produção das vogais há movimentos nos articuladores e os estudos de Lindblom e Sudbergn mostram que a formante F1 está interligada com a mandíbula e a formante F2 está relacionada com a língua. Entretanto, a faringe influencia em todas as formantes (GREGIO et al., 2006).

³ A lista que os autores Peterson e Barney (1952) utilizaram em seu trabalho continha dez palavras monossilábicas, cada palavra começando com [h] e terminando com [d] e diferindo apenas na vogal. Essas palavras eram: *heed, hid, head, had, hod, hawed, hood, who'd, hud, anal heard* (PETERSON; BARNEY, 1952)

dados restantes dos 25 usuários. Foram usados quatro e oito especialistas. Na figura 10, o eixo horizontal é o primeiro valor do formante⁴, e o eixo vertical é o segundo valor do formante (os valores do formante foram escalonados linearmente dividindo por um fator de 1000). Cada exemplo é rotulado com seu símbolo de vogal correspondente. As vogais [i] e [I] formam um par de classes sobrepostas, as vogais [a] e [A] formam o outro par.

Figura 10 – Problema de discriminação de vogais. A nuvem de dados a esquerda representam as vogais [i] minúsculo e [I] maiúsculo e a nuvem de dados a direita representam as vogais [a] minúsculo e [A] maiúsculo

Fonte: Jacobs et al. (1991)

As linhas identificadas como Net 0, 1 e 2 representam as linhas de decisão para três redes especialistas. De um lado dessas linhas, a saída do especialista correspondente é menor que 0,5, no outro lado a saída é maior que 0,5. Embora a mistura, neste caso, contivesse quatro especialistas, um desses especialistas não contribuiu significativamente para a mistura final, já que sua proporção de mistura p_i foi efetivamente 0 para todos os casos. A linha denominada Gate 0 indica que 2 é a decisão entre o especialista 0 e o especialista 2 feita pela rede de *gate*. À esquerda desta linha $p_2 > p_0$, à direita desta linha $p_0 > p_2$. O limite entre as classes [a] e [A] é formado pela combinação da parte esquerda da linha de decisão da Net 2 e a parte direita da linha de decisão da Net 0. Embora o sistema tenda a usar o mínimo de especialistas possível para resolver um problema, ele também é sensível a características específicas do problema, como o limite ligeiramente curvo entre as classes [a] e [A] (JACOBS et al., 1991).

Complementando, trata-se de uma estratégia que tenta solucionar problemas de classificação ou regressão com base em uma estratégia dividir-e-conquistar – dividindo o

⁴ Neste caso os valores do formante são os picos de energia em uma região do espectro sonoro de cada vogal.

problema entre vários especialistas. Nesta estratégia, determinados aspectos do problema são exclusivamente tratados por um especialista, ou pode haver uma sobreposição de tratamento a determinados aspectos. Durante a concepção da ME, conforme alguns especialistas vão ganhando mais importância, outros vão perdendo.

Na mistura de especialistas o espaço de entrada é automaticamente dividido em regiões, sendo que para cada região existe um único ou um subconjunto de especialistas mais indicados a agir. A arquitetura de mistura de especialistas emprega métodos probabilísticos de forma a promover uma abordagem sistemática capaz de definir automaticamente tanto a divisão de tarefas como a forma de combinar as soluções modulares, conforme elas vão sendo produzidas para cada proposta de divisão de tarefas. A divisão do espaço de entrada pode ser linear ou não-linear, também podendo ser gradual e contemplar sobreposições de regiões. Isso é controlado por meio da implementação de uma “rede *gating*” que define os coeficientes de uma combinação das saídas de cada especialista. A figura 11 ilustra uma arquitetura típica de uma ME, com m especialistas e portanto com uma combinação dos coeficientes $(g_1, g_2, \dots, g_m)$ e as saídas $(y_1, y_2, \dots, y_m)$. Os especialistas que compõem uma mistura podem ser RNAs como as MLPs ou modelos lineares. Neste trabalho foram usados modelos lineares.

Figura 11 – Estrutura típica de uma arquitetura de mistura de especialistas

Fonte: adaptado de Wagner et al. (2013)

Diferentes redes especialistas são apropriadas para diferentes regiões do espaço de entrada. Para cada entrada x , a arquitetura da ME requer um mecanismo capaz de identificar essa entrada, ou seja, o especialista ou combinação de especialistas mais capazes de produzir a saída correta, em termos probabilísticos, e isto é realizado por meio de uma

rede auxiliar, como já mencionado, a rede *gating*. A interpretação probabilística da rede *gating* é de um sistema que calcula, para cada especialista, a probabilidade dele gerar a saída desejada, e com base apenas no conhecimento da entrada x . Estas probabilidades são expressas pelos coeficientes g_i ($i=1,\dots,m$), e estes devem ser não-negativos e devem produzir sempre o valor unitário quando somados, para cada x . Estes coeficientes não são constantes, mas variam em função da entrada x . Segundo Lima (2004), caso os coeficientes g_i ($i=1,\dots,m$) sejam constantes e as redes especialistas atuam juntas, para todos os aspectos do problema. E, uma abordagem para garantir que os coeficientes g_i ($i=1,\dots,m$) respondam as restrições mencionadas, é preciso utilizar a função *softmax* (JACOBS et al., 1991). Essa função *softmax* define um conjunto de variáveis intermediárias ξ_i ($i=1,\dots,m$) como funções da entrada x e de um vetor de parâmetros v_i ($i=1,\dots,m$) na forma:

$$\xi_i = \xi_i(x, v_i) \quad (1)$$

e os coeficientes g_i ($i=1,\dots,m$) podem ser definidos em termos de ξ_i ($i=1,\dots,m$) como segue:

$$g_i = \frac{\exp(\xi_i)}{\sum_{k=1}^m \exp(\xi_k)} \quad (2)$$

Com base nessa definição, os coeficientes g_i ($i=1,\dots,m$) passam a respeitar as restrições impostas, isto é, são não-negativos e, somados, produzem sempre o valor unitário, para cada x .

O modelo de probabilidade adotado para a arquitetura de mistura de especialistas é explicado em seguida, seguindo o exposto em Lima (2004). Considere que um conjunto de dados de treinamento é gerado da seguinte forma: dada uma entrada x , um especialista i é escolhido com alguma probabilidade, válida no problema sob estudo. Dada a escolha do especialista e dada a entrada, a saída desejada y é supostamente gerada de acordo com a probabilidade aprendida pela ME. Ainda, segundo o mesmo autor, uma determinada saída da ME pode ser gerada de m maneiras diferentes, correspondendo às m maneiras diferentes de escolha do especialista que responde àquela entrada. Assim, a probabilidade total de geração de y a partir de x é dada pela soma sobre i , de acordo com (LIMA, 2004) na forma de:

$$P(y|x, \Theta^0) = \sum_{i=1}^m P(i|x, v^0) P(y|x, \Theta_i^0). \quad (3)$$

A densidade modelada na equação é conhecida como mistura de densidade ou função de verossimilhança (LIMA, 2004). Trata-se, de uma mistura de densidade no espaço de saída, condicionada à escolha da entrada, onde $P(i|x, v^0)$ é a probabilidade de escolher o especialista i dado a entrada x e o vetor de parâmetros v^0 , e $P(y|x, \Theta_1^0)$ é a probabilidade deste especialista gerar a saída y , dado a entrada x e o vetor de parâmetros Θ_1^0 (LIMA, 2004).

É tarefa da rede *gating* modelar a probabilidade $P(i|x, v^0)$. É possível parametrizar esta probabilidade via equações (2) e (1), fazendo a saída da rede *gating* g_i ($i=1, \dots, m$) ser igual a $P(i|x, v_i)$. A saída da mistura de densidades pode ser calculada por meio da média condicional. A média condicional é obtida por (LIMA, 2004):

$$\mu = \sum_{i=1}^m g_i \mu_i \quad (4)$$

em que μ_i é a média condicional associada com a distribuição de probabilidade $P(y|x, \Theta_1^0)$. A saída do especialista é uma combinação da escolha usual para a saída de uma arquitetura modular.

A figura 12 ilustra como a ME é aplicada nesse estudo. Cada pico da gaussiana apresentada como “Entrada x” é uma EFG diferente. Cada pico é a entrada da mistura de especialistas.

Figura 12 – Arquitetura de mistura de especialistas modelada ao problema

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Haykin (2009) defende que a ME possui um caráter dinâmico e que ele se deve ao fato de que as regiões de atuação a serem alocadas para os especialistas não são definidas *a priori*, e são implementadas de forma interativa e com garantia de convergência para um

mínimo local. Essa característica se faz interessante para o problema definido neste estudo, uma vez que trata-se de um problema de reconhecimento de padrões cuja complexidade varia a cada EFG, sendo que será necessário alocar um número diferente de especialistas para resolver o reconhecimento de cada uma delas.

Um dos fatores que pode influenciar o desempenho de uma mistura de especialista é a inicialização dos parâmetros de cada especialista. Em Yuan e Neubauer (2009) foi descrito um método de inicialização baseado no algoritmo *k-means*. Esse método foi aplicado neste trabalho e segue aqui descrito.

Para essa inicialização, os dados de treinamento são submetidos ao algoritmo de agrupamento *k-means*. O algoritmo criará grupos que são representados pelos seus centroides. Os vetores dos centroides são usados como inicialização de especialistas. Os especialistas são inicializados incrementalmente. Para o primeiro especialista, todos os dados de treinamento são considerados, e a inicialização é feita com base em um vetor central para o conjunto de dados completo. Para o segundo e terceiro especialistas, o conjunto de dados é submetido ao *k-means*, considerando $k = 2$, e os dois centroides resultantes são usados como inicialização. Esse procedimento deve ser repetido conforme o número de especialistas que se quer usar na mistura. Como exemplo, considere o caso em que se pretende usar dez especialistas. Para que todos sejam criados usando tal estratégia, o *k-means* será executado três vezes, com $k = 2$, $k = 3$ e $k = 4$. Assim, nove especialistas serão gerados usando o centroide dos grupos e um, o primeiro, será gerado a partir do conjunto de dados completo.

A intenção com esse método de inicialização é posicionar os diferentes especialistas em diferentes regiões do espaço dos dados de treinamento. Segundo Yuan e Neubauer (2009), esse método de inicialização nos fornece um ponto de partida significativo.

3.2.1 Considerações finais

Esse capítulo apresentou uma breve explicação sobre as técnicas de aprendizado máquina MLP e MEs, abrangendo conceitos como arquitetura, implementação e parametrização. Especificamente para o caso das MEs, um exemplo de aplicação e um método alternativa de inicialização foram apresentados.

4 Reconhecimento automático das expressões faciais gramaticais

Este trabalho visa o desenvolvimento de um estudo de reconhecimento de expressões faciais gramaticais da Libras, sob um modelo de classificação multiclasse, por meio da aplicação de técnicas de aprendizado de máquina, utilizando *Multilayer Perceptron* e mistura de especialistas. Este capítulo apresenta o processo de modelagem do problema de classificação e a aplicação de algoritmos de aprendizado de máquina (seção 4.1). Na seção 4.2 é descrito o conjunto de dados já disponível (rotulado sob um modelo de classes binárias), o novo conjunto de dados (alterado de forma que o modelo de rotulação permitisse a classificação multiclasse) e os procedimentos de pré-processamento que foram aplicados aos conjuntos de dados. Na seção 4.3 disserta sobre o pré-processamento, a seção 4.4 descreve a configuração dos experimentos e na seção 4.5 descreve a avaliação dos resultados. Além disso, nesta seção, é explicada a estratégia adotada na experimentação realizada sobre o conjuntos de dados, com uso de *Multilayer Perceptron* e mistura de especialistas. A seção 4.6 apresenta os resultados obtidos com os experimentos e as análises realizadas.

4.1 Definição do problema

Expressões faciais, além de possibilitar a expressividade da emoção do indivíduo, também são usadas na formação da estrutura gramatical da língua sinais, e nesse contexto, o seu papel é expressar a prosódia caracterizando o tempo verbal ou tipo de frases. Tais expressões faciais são chamadas gramaticais. Na língua de sinais há um conjunto finito de expressões faciais gramaticais. Assim, a fim de contribuir com o desenvolvimento de sistemas que sejam capazes de processar esse tipo de língua, é imprescindível que eles estejam aptos a localizar e classificar uma ocorrência de uma expressão facial gramatical usada em língua de sinais.

O problema tratado neste trabalho é caracterizado pela busca de um modelo classificador que seja capaz de interpretar uma sequência de quadros de um vídeo (Figura 13), no qual está gravada uma fala em Libras, e identificar os quadros que se referem à execução de uma (ou mais) expressões faciais. A solução para este problema contribui para a área de reconhecimento de padrões e para área de Linguística, evidenciando que é

possível localizar e classificar, automaticamente, as expressões faciais gramaticais dentro do contexto de fala em língua de sinais.

Figura 13 – Exemplo de uma sequência de quadros de vídeo capturadas pelo *Kinect*

Fonte: Freitas (2015)

O estudo discutido neste trabalho explora o problema de segmentação de EFGs por meio da construção de modelos para classificação. Para definir o problema, considere uma expressão facial gramatical EFG_i , com $i = \{1, \dots, n+1\}$, pertencente ao conjunto finito de $n+1$ expressões faciais gramaticais $EFG = \{EFG_1, EFG_2, \dots, EFG_n, EFG_{n+1}\}$, sendo EFG_{n+1} a expressão neutra. A figura 14 exemplifica uma face neutra com seus respectivos pontos e uma face ocorrendo a EFG e seus respectivos pontos. Uma EFG_i é descrita por um conjunto de d pontos $P = \{p_1, p_2, \dots, p_d\}$ extraídos da face humana e dispostos no espaço bidimensional (x, y) . Então, assumo um vídeo, que contém uma fala em Libras, como sendo uma sequência de quadros $S = \{q_1, q_2, \dots, q_t\}$ de tamanho t . Como entrada para um modelo de classificação binária ou multiclasse, é usada uma representação vetorial dos quadros desse vídeo, baseada no conjunto de pontos P ou em características derivadas dele, como o posicionamento dos elementos constituintes da face ou as relações entre eles.

O objetivo do classificador é analisar cada quadro como sendo referente à execução de uma EFG_i durante uma fala em Libras. Considerando que o classificador executa sua tarefa com sucesso, sequências de quadros são classificados como contendo a execução de uma EFG_i , fornecendo a solução para o problema de segmentação da EFG, e consequentemente, de sua classificação.

Figura 14 – Exemplo de face neutra e da execução de uma EFG, e os respectivos pontos (x, y) extraídos da face pelo *Kinect*

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os modelos de classificação construídos consideram o problema de segmentação das EFGs sob dois aspectos: atemporal e temporal. No primeiro caso, os classificadores recebem como entrada a informação que descreve a expressão facial presente em um quadro do vídeo. Neste caso, a informação analisada pelo classificador diz respeito às coordenadas espaciais de d pontos da face, ou às características extraídas destes pontos neste único quadro de vídeo. Assim, a representação vetorial para o quadro de vídeo q_t , considerando coordenadas espaciais (x, y) de d pontos, é da forma

$$\vec{x}_{q_t} = \{x_1, y_1, x_2, y_2, \dots, x_d, y_d\}. \quad (5)$$

O espaço de decisão criado possui $2d$ dimensões. Essa representação vetorial pode ser mapeada para uso de outras características descritivas do quadro de vídeo. Neste caso, a dimensionalidade do espaço de decisão passa a ser $\#cd$, em que $\#c$ representa a quantidade de características em uso.

Uma MLP básica não é capaz de processar a informação temporal, então é preciso inserir essa informação explicitamente na representação vetorial para tratar o problema sob o aspecto temporal. Para isso, considere a sequência S como um sinal variante no tempo, de maneira a considerar a informação sobre a movimentação dos elementos da face no tempo. Para embutir essa informação em uma representação vetorial, faz-se uso de uma organização de dados em janelas. Nesta estratégia, a representação vetorial para o quadro de vídeo q_t , considerando informação das coordenadas espaciais (x, y) de d pontos e uma janela de tamanho ímpar j é da forma

$$\vec{x}_{q_t}^j = \{x'_{q_t - \lfloor j/2 \rfloor}, \dots, x'_{q_t - 1}, x'_{q_t}, x'_{q_t + 1}, \dots, x'_{q_t + \lfloor j/2 \rfloor}\}, \quad (6)$$

em que x'_{q_t} é equivalente à representação apresentada na eq. 5, e o quadro de interesse de classificação é o quadro do meio da janela, considerando informação temporal como sendo a informação dos $\lfloor j/2 \rfloor$ quadros passados e dos $\lfloor j/2 \rfloor$ futuros¹. O espaço de decisão criado para o classificador com uso de dados janelados possui $2dj$ dimensões, ou $\#cd$ dimensões se um conjunto c de características é usado na descrição do quadro de vídeo. A representação com janela possui a desvantagem de excluir alguns quadros de vídeo do trabalho de classificação, uma vez que a falta da informação passada, ou futura, para

¹ Representações com quadro de interesse no início da janela ou no fim da janela também são possíveis. Em trabalhos prévios notou-se que usar o quadro de interesse no meio da janela melhora o resultado de classificação.

quadros no início do vídeo, ou no fim, impossibilita a criação da representação vetorial completa.

O modelo da resposta dos classificadores usa alfabeto bipolar. Para classificadores binários, a expressão neutra é reconhecida se as respostas são $[+1, -1]$ e a EFG se as respostas são $[-1, +1]$. Para o problema multiclasse, o formato da resposta é um conjunto de $n + 1$ saídas bipolares ordenadas na forma $O = [o_1, \dots, o_n, o_{n+1}]$, que representa o conjunto *EFG*, e assume um valor positivo (+1) na posição que indica a EFG reconhecida. Por exemplo, para a segmentação de duas EFGs mais a expressão neutra, se o conjunto de saídas do classificador é $O = [+1, -1, -1]$, o quadro de vídeo foi classificado como “expressão neutra” – sendo o primeiro valor no vetor de classes, o valor referente à expressão neutra e os outros dois referentes às duas EFGs conforme convenção estabelecida pelo projetista do classificador.

4.2 Conjuntos de dados

Esta seção tem o propósito de apresentar os conjuntos de dados a serem usados para analisar o problema de reconhecimento de expressão facial gramatical na Libras. O conjunto *Grammatical Facial Expressions Data Set* foi a base para a construção do novo conjunto com rotulação multiclasse, necessária para suportar a análise das EFGs considerando o problema multiclasse.

4.2.1 *Grammatical Facial Expressions Data Set*

Os experimentos usam dados provenientes do conjunto de dados *Grammatical Facial Expressions Data Set*² (FREITAS et al., 2014; LICHMAN, 2013), composto por dezoito vídeos gravados com o sensor *Microsoft Kinect*, usando uma taxa de captação de aproximadamente 30 quadros por segundo. Em cada vídeo um intérprete executa cinco repetições de cinco frases em Libras, as quais requerem a execução de pelo menos uma EFG dentre aquelas utilizadas na referida língua. O conjunto de dados conta com a atuação de dois intérpretes. A informação referente à expressão facial é armazenada por meio de 100 coordenadas espaciais (x, y, z) de pontos dos olhos, nariz, sobrancelhas, boca, contorno do rosto e íris. O conjunto de dados foi rotulado manualmente por especialistas em Libras,

² <https://archive.ics.uci.edu/ml/datasets/Grammatical+Facial+Expressions>

de forma a oferecer uma “verdade fundamental”³ referente à EFG em execução em cada quadro de vídeo. A rotulação está organizada de forma binária, ou seja, os quadros de vídeos são associados à expressão neutra (0 - classe negativa) ou a uma EFG (1 - classe positiva). Desta forma, uma sequência de quadros com rótulos 1 indica um segmento de vídeo no qual uma EFG está sendo usada. Um exemplo de frases executadas para cada uma das EFGs é apresentado no quadro 6. O conjunto completo de frases é encontrado em Freitas et al. (2014), Freitas et al. (2017).

Quadro 6 – Exemplos de frases usadas no conjunto de dados

EFG	Frase
Interrogativa (qu)	Onde você mora?
Interrogativa (s/n)	Você vai embora?
Interrogativa (dúvida)	Você vai EMBORA? (duvidando)
Negativa	Eu nunca fui preso.
Afirmativa	Eu trabalho lá.
Condicional	Se fizer sol, eu vou à praia.
Relativa	O Seli, escola de surdos, fica em São Paulo.
Tópicos	Frutas ... eu gosto de abacaxi.
Foco	Foi a WAINE quem fez.

Fonte: Adaptado de Freitas et al. (2014)

O conjunto de dados é organizado em 36 arquivos: 18 arquivos de dados e 18 arquivos de rótulos. O processo de rotulação foi realizado por mais de uma pessoa visto que o processo de análise de cada expressão facial gramatical é um processo subjetivo. O conjunto de dados foi rotulado por pessoas diferentes como em Freitas (2015), a figura 15 apresenta o intérprete A e o intérprete B na qual ambos estão realizando o mesmo sinal de afirmação contidos nas sentenças do conjunto de dados.

Figura 15 – Intérprete *a* e intérprete *b* sinalizando as frases do conjunto de dados (ambos intérpretes realizando a frase “Eu vou” (afirmativa))

Fonte: Freitas (2015)

³ Rotulações manuais realizadas sobre quadros de vídeos carregam algum grau de subjetividade e viés, principalmente porque embora seja natural para o especialista humano interpretar uma cena em um vídeo, não é natural fazer o mesmo tipo de interpretação quadro a quadro, ou decidir exatamente em qual quadro de vídeo determinada ação inicia ou termina.

A tabela 1 apresenta a quantidade de quadros existentes no conjunto de dados, por vídeo e para cada EFG. A tabela também apresenta informações sobre a rotulação binária dos quadros, sendo: “-” neutra e “+” expressão facial gramatical.

Tabela 1 – Quantidade total de quadros de vídeo de cada expressão; quadros positivos (contém expressão) e quadros negativos (não contém expressão)

EFG	#quadros					
	intérprete a			intérprete b		
	total	+	-	total	+	-
Afirmativa	1062	414	648	1074	528	546
Interrogativa (dúvida)	1312	491	821	1497	780	717
Interrogativa (qu)	1286	609	677	1328	549	779
Interrogativa (y/n)	1390	532	858	1738	715	1023
Foco	1403	330	1073	1344	531	813
Negativa	1124	528	596	1582	712	870

Fonte: Maria Eduarda de Araújo Cardoso, 2018

4.2.2 Extensão do conjunto de dados *Grammatical Facial Expressions Data Set*

Originalmente, o conjunto de dados está preparado para suportar experimentos de segmentação binária, separando a ocorrência de uma EFG da ocorrência de expressões neutras. Nele, há frases executadas para estudo das EFGs do tipo condicional, relativa e tópico, que envolvem também uma segunda EFG, porém a ocorrência dela está rotulada como classe negativa, da mesma forma que a expressão neutra. Devido a essa característica, as frases referentes a essas EFGs foram excluídas do escopo do estudo apresentado neste trabalho para os estudos de classificação multiclasse.

Para suportar os experimentos deste estudo, o conjunto original foi reorganizado. Para os experimentos com classificação binária, os rótulos foram substituídos por 1 (classe positiva) e -1 (classe negativa) adequando-os à imagem da função de ativação escolhida para respostas dos classificadores. Para as MEs foi usado 0 para a classe negativa devido à características de como ela foi implementada. Para os experimentos de classificação multiclasse, as frases da organização original foram combinadas, gerando novos conjuntos de dados, com rotulação multiclasse. Do ponto de vista de quadros de vídeo, essa combinação constitui uma concatenação de quadros, ou seja, uma sequência de quadros referentes a uma frase é concatenado a uma sequência de quadros referentes à outra frase. O quadro 7 mostra um exemplo da concatenação de frases com duas EFGs. A primeira coluna do

quadro ilustra a situação original, com as EFGs associadas a classes positivas e negativas. A segunda coluna do quadro ilustra a adaptação, de forma que as EFGs estão associadas a rótulos para problemas multiclasse.

Quadro 7 – Exemplo de adaptação do conjunto de dados para adequação à modelagem de problemas multiclasse

Dois conjuntos disjuntos de frases. Cada um deles forma um problema binário diferente junto com os quadros de expressão neutra (<i>classe -</i>).		Um único conjunto de frases Um problema multiclasse é formado juntamente com os quadros de expressão neutra (<i>classe 1</i>).
Afirmativa	Negativa	Afirmativa + Negativa
Eu vou.	Eu não vou.	Eu vou. (<i>classe 2</i>)
Eu quero.	Eu não fiz nada.	Eu quero. (<i>classe 2</i>)
...	...	...
(<i>classe +</i>)	(<i>classe +</i>)	Eu não vou. (<i>classe 3</i>)
		Eu não fiz nada. (<i>classe 3</i>)
		...

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Seguindo as combinações de EFGs, 15 conjuntos de dados reorganizados foram preparados para suportar a execução com experimentos de classificação com 3 classes, e um novo conjunto de dados reorganizado foi preparado para o suportar a execução de experimentos de classificação com 7 classes. Este último conjunto, portanto, conta com quadros com expressão neutra e com as EFGs: interrogativas (qu), (s/n) e (dúvida), negativa, afirmativa e foco. Para esses dois tipos de experimentos, a estratégia de rotulação considerou um vetor de rótulos com três componentes e com sete componentes, respectivamente. Para o caso do problema multiclasse ilustrado no quadro 7, a rotulação dos quadros é como segue (a rotulação para a Mistura de Especialistas usa 0 em vez de -1):

- *Experimento #1*

- quadros com expressão neutra: $\{\mathbf{1}, -1, -1\}$;
- quadros com EFG afirmativa: $\{-1, \mathbf{1}, -1\}$;
- quadros com EFG negativa: $\{-1, -1, \mathbf{1}\}$.

- *Experimento #2*

- quadros com expressão neutra: $\{\mathbf{1}, -1, -1, -1, -1, -1, -1\}$;
- quadros com EFG afirmativa: $\{-1, \mathbf{1}, -1, -1, -1, -1, -1\}$;
- quadros com EFG interrogativa (dúvida): $\{-1, -1, \mathbf{1}, -1, -1, -1, -1\}$.
- quadros com EFG foco: $\{-1, -1, -1, \mathbf{1}, -1, -1, -1\}$.

- quadros com EFG negativa: $\{-1, -1, -1, -1, \mathbf{1}, -1, -1\}$.
- quadros com EFG interrogativa (qu): $\{-1, -1, -1, -1, -1, \mathbf{1}, -1\}$.
- quadros com EFG interrogativa (s/n): $\{-1, -1, -1, -1, -1, -1, \mathbf{1}\}$.

Nas tabelas 2 e 3 são apresentadas as junções realizadas das EFGs. Os novos conjuntos de dados são então rotulados de forma a representar três classes distintas e sete classes distintas.

Tabela 2 – Total de quadros com as EFG concatenadas. Abreviações: Afirm. (afirmativa), Inter. (interrogativa)

EFG concatenadas	#quadros intérprete		EFG concatenadas	#quadros intérprete	
	a	b		a	b
Afirm. - Inter. (qu)	2.348	2.402	Inter. (qu) - Afirm.	2.348	2.402
Afirm. - Foco	2.465	2.418	Inter. (qu) - Foco	2.689	2.672
Afirm. - Inter. (y/n)	2.452	2.813	Inter. (qu) - Inter. (y/n)	2.676	3.067
Afirm. - Inter. (dúvida)	2.374	2.571	Inter. (qu) - Inter. (dúvida)	2.598	2.825
Afirm. - Negativa	2.186	2.656	Inter. (qu) - Negativa	2.410	2.910
Foco - Inter. (qu)	2.689	2.672	Negativa - Afirm.	2.186	2.656
Foco - Afirm.	2.465	2.418	Negativa - Foco	2.527	2.926
Foco - Inter. (y/n)	2.793	2.418	Negativa - Inter. (y/n)	2.514	3.321
Foco - Inter. (dúvida)	2.715	2.841	Negativa - Inter. (dúvida)	2.436	3.079
Foco - Negativa	2.527	2.926	Negativa - Inter. (qu)	2.410	2.910
Inter. (dúvida) - Inter. (qu)	2.598	2.825	Inter. (y/n) - Afirm.	2.452	2.813
Inter. (dúvida) - Afirm.	2.374	2.571	Inter. (y/n) - Foco	2.793	2.418
Inter. (dúvida) - Inter. (y/n)	2.702	3.236	Inter. (y/n) - Negativa	2.513	3.321
Inter. (dúvida) - Foco	2.715	2.841	Inter. (y/n) - Inter. (dúvida)	2.702	3.236
Inter. (dúvida) - Negativa	2.436	3.079	Inter. (y/n) - Inter. (qu)	2.676	3.067

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Tabela 3 – Total de quadros com todas EFGs combinadas

EFGs concatenadas	#quadros	
	intérprete a	intérprete b
Todas EFGs	6987	7852

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A rotulação do trabalho anterior e a nova rotulação para o presente trabalho não são livres de viés pois são rotuladas por seres humanos especialistas em língua de sinais, e assim existindo uma subjetividade, pois cada especialista rótula os dados considerando suas percepções particulares.

4.3 Pré-processamento

Os dados referentes às frases em Libras foram submetidos aos procedimentos de pré-processamento ilustrados na figura 16. Normalização e translação de dados foram realizadas para anular efeitos de variação de posicionamentos do sensor e do gesticulador. A partir dos dados normalizados, a seleção de atributos e extração de características foram realizadas com base no que é geralmente praticado na pesquisa na área. Desta forma, cinco conjuntos de dados foram obtidos. Na sequência, a representação janelada foi criada usando janelas de tamanho 3, 5 e 9, seguindo diretrizes do trabalho de Freitas (2015).

Figura 16 – Representação do pré-processamento dos dados

Fonte: Maria Eduarda de Araújo Cardoso, 2018

O procedimento de normalização posicionou as coordenadas capturadas para todos os quadros de vídeos em um hipercubo unitário. Já o procedimento de translação centralizou as coordenadas em relação ao ponto central referente à ponta do nariz do gesticulador. O efeito obtidos com a execução destes procedimentos pode ser observado na figura 17.

A seleção de atributos foi realizada com base em resultados observados na literatura correlata de análise de expressões faciais afetivas e gramaticais. Há exemplos de trabalhos que usam: um ponto em cada sobrancelha e quatro pontos na boca (CHANG; HUANG, 2010; WANG et al., 2010); e que acrescentam alguns pontos entre a boca e o nariz (NGUYEN; RANGANATH, 2012; DAHMANE; MEUNIER, 2012; YU; YOU; TSAI, 2012). Especificamente para o conjunto de dados usado neste trabalho, os autores em (FREITAS,

Figura 17 – Efeitos dos procedimentos de normalização e translação: quadros originais à esquerda; quadros normalizados e transladados à direita

Fonte: Maria Eduarda de Araújo Cardoso, 2018

2015) realizaram um estudo da correlação existente entre os 100 pontos extraídos da face durante a execução dos movimentos faciais necessários para realização das EFGs. Nesse estudo foi constatado que o melhor sub-conjunto de pontos a ser usado é composto por oito pontos conforme mostrado na figura 18 que foi acrescentado o ponto do nariz. Assim, para fins das experimentações aqui relatadas, além do uso de 100 pontos, foram também consideradas representações com os oito pontos citados, e com nove pontos, adicionando a ponta do nariz como meio de representar uma informação de referência para os demais.

Figura 18 – Indicação dos pontos usados nos subconjuntos com nove pontos

Fonte: Maria Eduarda de Araújo Cardoso, 2018

No que diz respeito à extração de características, foi realizado o cálculo das distâncias entre os pontos que descrevem a face, considerando o uso de oito e nove pontos. A figura 19 a informação fornecida pelas distâncias entre os pontos é ilustrada considerando: (acima)

distâncias extraídas para um quadro no qual ocorre a expressão neutra; (abaixo) distâncias extraídas para um quadro no qual ocorre uma EFG. Nos detalhes ampliados é possível observar que as relações entre os pontos se alteram quando a EFG é executada. A figura 20, a informação de distância entre pontos é exemplificada para cada uma das EFGs usadas nos experimentos deste artigo.

Figura 19 – Distância entre pontos da face: acima - um quadro com expressão neutra; (abaixo) um quadro com uma EFG

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Figura 20 – Quadros reais e distâncias entre pontos da face para a expressão neutra e EFGs estudadas nos experimentos deste estudo

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para uso desses dados em um classificador é necessário criar uma representação vetorial. A representação vetorial para um dado, nesse contexto, é a combinação dos atributos descritivos de cada quadro. Diferentes combinações geram diferentes representações vetoriais. A partir da realização do pré-processamento as representações vetoriais dos dados foram criadas considerando: coordenadas espaciais dos pontos da face ou distâncias entre elas para o caso do uso de oito e nove pontos, uso ou não de janelas, aplicação ou não de normalização e translação.

Após os dados serem pré-processados, foi necessário organizá-los de maneira a embutir na representação uma informação referente às características temporais. Como já mencionado, esse estudo trabalha com o problema atemporal e temporal. O caso atemporal pode ser identificado em um quadro de vídeo como ilustra a figura 21. Nessa figura, oito quadros de vídeo estão representando um trecho de vídeo. O quadro *p3*, em destaque, é usado para mostrar que a informação contida em cada quadro diz respeito à coordenadas espaciais de pontos representados no espaço de três dimensões. Um ponto dentro de um quadro diz respeito a uma região de interesse para representar o objeto sob análise, por exemplo, o canto esquerdo da boca de uma pessoa é um ponto de interesse e sua localização no quadro de vídeo é representada pelas coordenadas x , y e z .

Figura 21 – Exemplo de problema atemporal identificado em um quadro de vídeo

Fonte: Maria Eduarda de Araújo Cardoso, 2018

As técnicas para classificação utilizadas não conseguem processar a informação temporal, assim é preciso mostrá-la explicitamente na representação vetorial. Considerando a expressão facial gramatical como um sinal variante no tempo, ou seja, como uma série temporal, caracteriza-se a informação sobre a movimentação dos elementos da face no tempo. Uma organização de dados em janelas foi realizada para que fosse possível caracterizar a informação sobre a movimentação dos elementos da face no tempo. O quadro 8, apresenta um esquema que explica a construção de janelas de tamanhos diferentes.

Quadro 8 – Exemplo de janelas de tamanhos distintos

Tam.	Janela 1	Janela 2	...	Janela n
1	$quadro_1$	$quadro_2$	...	$quadro_n$
2	$quadro_1;quadro_2$	$quadro_2;quadro_3$	...	$quadro_{n-1};quadro_n$

Fonte: Adaptado de Madeo (2013)

Na figura 22 é ilustrada a estrutura da janela, incluindo as características de um quadro. Nesta figura estão os oitos pontos selecionados dos cem pontos extraídos da face humana por meio do sensor *Kinect*. Os pontos P_1 , P_2 , P_3 e P_4 são referentes à sobrancelha, os pontos P_5 , P_6 , P_7 e P_8 pertencem à boca. O quadro Q_{int} é o quadro 3 trecho de vídeo e representa o quadro de interesse para o classificador. Os quadros Q_{int-i} , ou seja, na figura os quadros 1 e 2, são os quadro anteriores ao quadro de interesse (informação do passado). Os quadros Q_{int+i} , ou seja, os quadros 4 e 5, são os quadros posteriores ao quadro de interesse (informação sobre o futuro).

Figura 22 – Exemplo de janela com 8 pontos centrado no quadro Q_{int}

Fonte: Maria Eduarda de Araújo Cardoso, 2018

No caso de representação de dados com janelamento, desconsiderando um pequeno trecho inicial e/ou final do vídeo, todos os quadros são apresentados às técnicas de classificação, acompanhados de quadros anteriores e/ou posteriores, representando informação passada e futura. A figura 23 ilustra como é feito o janelamento com a janela de tamanho 3, sendo as dez dimensões representadas pelas coordenadas X_1, Y_1 , X_2, Y_2 , X_3, Y_3 , X_4, Y_4 e X_5, Y_5 . Os quadros de vídeo são representados por F_1 , F_2 , F_3 , F_4 e F_5 , neste caso o vídeo possui 5 quadros de vídeo. Como a janela é de tamanho 3 foi janelado os três primeiros

quadros F_1 , F_2 e F_3 e assim sucessivamente vai janelando os quadros como ilustrado na figura 23. O quadro de interesse dentro de uma janela, escolhido neste trabalho, é o quadro do meio. Desta forma, o rótulo referente ao quadro do meio da janela passa a ser usado como rótulo para a janela.

Figura 23 – Exemplo de janelamento de tamanho 3 com rótulo no meio

		10 dimensões (coordenada x e y)												
		X ₁	Y ₁	X ₂	Y ₂	X ₃	Y ₃	X ₄	Y ₄	X ₅	Y ₅	Rótulos		
Quadros de vídeo	F ₁	0.1	0.2	0.3	0.1	0.4	0.2	0.3	0.1	0.2	0.3	1	0	0
	F ₂	0.1	0.4	0.3	0.1	0.2	0.3	0.1	0.2	0.4	0.1	1	0	0
	F ₃	0.1	0.1	0.2	0.3	0.2	0.3	0.1	0.2	0.3	0.1	0	1	0
	F ₄	0.5	0.5	0.2	0.3	0.1	0.5	0.1	0.1	0.2	0.3	0	1	0
	F ₅	0.1	0.4	0.3	0.1	0.1	0.4	0.3	0.1	0.1	0.4	0	0	1

	X ₁	Y ₁	X ₂	Y ₂	X ₃	Y ₃	X ₄	Y ₄	X ₅	Y ₅	X ₁	Y ₁	X ₂	Y ₂	X ₃	Y ₃	X ₄	Y ₄	X ₅	Y ₅	Rótulos		
F ₁ /F ₂ /F ₃	0.1	0.2	0.3	0.1	0.4	0.2	0.3	0.1	0.2	0.3	0.1	0.4	0.3	0.1	0.2	0.3	0.1	0.2	0.3	0.1	1	0	0
F ₂ /F ₃ /F ₄	0.1	0.4	0.3	0.1	0.2	0.3	0.1	0.2	0.4	0.1	0.5	0.5	0.2	0.3	0.1	0.5	0.1	0.1	0.2	0.3	0	1	0
F ₃ /F ₄ /F ₅	0.5	0.5	0.2	0.3	0.1	0.5	0.1	0.1	0.2	0.3	0.5	0.5	0.2	0.3	0.1	0.5	0.1	0.1	0.4	0.3	0	1	0
.....																							
F _{n-2} /F _{n-1} /F _n																							

Fonte: Maria Eduarda de Araújo Cardoso, 2018

4.4 Configuração dos experimentos

Foram realizados experimentos com a classificação binária e com a classificação multiclasse (duas estratégias diferentes). Todos envolveram treinamento de redes neurais artificiais *Multilayer Perceptron* e da mistura de especialistas. Para os experimentos, o problema que envolve as EFGs foi modelado da seguinte forma:

- os dados vêm da sinalização de uma frase em Libras e estão rotulados de forma a constituir um problema de classificação binária (uma EFG de interesse), e um problema de classificação multiclasse (duas EFG de interesse + neutra/seis EFGs de interesse + neutra);
- para entrada do classificador é utilizada uma representação vetorial referente a (x, y) de cada ponto, contendo informações extraídas de cada quadro do vídeo;
- o classificador analisa a informação referente a cada um dos quadros de entrada, e decide se ele é uma expressão neutra ou uma expressão facial gramatical;
- a saída do classificador MLP pertence ao conjunto $\{1, -1\}$, considerando 1 o quadro que ocorre a EFG e -1 sendo o quadro que não ocorre EFG (neutra) para o caso binário;

- a saída do classificador MLP pertence ao conjunto $\{1, -1, -1\}$ ou $\{1, -1, -1, -1, -1, -1, -1, -1\}$, considerando 1 o quadro que ocorre a EFG e -1 sendo o quadro que não ocorre EFG (neutra) para o caso multiclasse;
- a saída do classificador ME pertence ao conjunto $\{0, 1\}$, considerando 1 o quadro que ocorre a EFG e 0 sendo o quadro que não ocorre EFG (neutra).
- a saída do classificador ME pertence ao conjunto $\{1, 0, 0\}$ ou $\{1, 0, 0, 0, 0, 0, 0, 0\}$, considerando 1 o quadro que ocorre a EFG e -1 sendo o quadro que não ocorre EFG (neutra) para o caso multiclasse;

A figura 24 apresenta um diagrama que ilustra como foram realizadas as experimentações. De forma resumida, o conjunto de dados possui falas em Libras provenientes de dois intérpretes; essas falas foram representadas em termos de coordenadas espaciais e distâncias euclidianas; os dados foram rotulados para suportar classificação binária e multi-classe; representações atemporais e temporais foram criadas; modelos classificadores foram obtidos sobre essas representações usando MLP e mistura de especialistas, considerando experimentos binários, com três classes e com sete classes; a avaliação dos resultados foi feita usando medidas classicamente usadas em aprendizado de máquina e também fazendo uso de análises sobre tipos de erros cometidos pelos classificadores.

Para o treinamento foi estudado o uso de diferentes parâmetros e após esse estudo decidiu-se pelas seguintes variações:

- *Multilayer Perceptron*
 - número de neurônios na camada escondida: média aritmética calculada sobre o número de neurônios na camada de entrada e número de neurônios na camada de saída (CASTRO et al., 1998). Para variar esse número de neurônios e criar diferentes ambientes de experimentação, o resultado da média aritmética foi acrescido de 50% a mais de neurônios e decrescido em 50% a menos de neurônios (cf.tabela 4);
 - taxa de aprendizado inicial: 0,01, 0,1, e 0,5;
 - número de épocas: 500 e 3.000.
 - função de ativação: *tansig*
- *Mistura de especialistas*
 - Número de especialistas caso binário: 2, 3 e 4;

Figura 24 – Estrutura de como foi realizada as experimentações

Fonte: Maria Eduarda de Araújo Cardoso, 2018

- Número de especialistas caso multiclasse (Neutra, EFG1, EFG2): 3, 6 e 8;
- Número de especialistas caso multiclasse (neutra e todas EFGs): 7, 14 e 16;
- Taxa de aprendizado: 0,1 e 0,01;
- Número de iterações: 1.000;

Modelos classificadores MLP foram criados para todos os subconjuntos de dados criados, considerando diferentes configurações de parâmetros. Para MLP foi usando o gradiente descendente como método de otimização e taxa de aprendizado sem decaimento durante o processo de treinamento.

Tabela 4 – Número de neurônios na cada oculta da MLP para cada representação de dados e uso ou não de janelas: janelas de tamanho 3, 5, e 9. Rep: representação; tam: tamanho

Rep.	s/ janela	#neurônios ocultos								
		tam. 3			tam. 5			tam. 9		
100 pts	50 100 150	75	151	226	125	251	376	226	451	677
8 pts	4 8 12	6	13	16	11	21	32	18	37	56
9 pts	5 9 13	7	15	22	12	24	36	23	46	69
28 dist	7 14 21	11	22	33	18, 36, 54			63	127	191
36 dist	9 18 27	14	28	42	23, 46, 69			82	163	245

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para o treinamento das misturas de especialistas foram usados dois tipos de inicialização, a inicialização aleatória e a inicialização com o algoritmo *K-means*. Nos três experimentos, as técnicas foram treinadas em um subconjunto de dados de treino e tiveram seus desempenhos aferidos em um subconjunto de dados de teste utilizando a estratégia *holdout*⁴. Para construção destes subconjuntos, a seguinte estratégia foi aplicada: para cada EFG, os quadros dos vídeos referentes a três sentenças foram usados para treinamento dos classificadores e os quadros de vídeos referentes às duas sentenças restantes foram usados para teste (como ilustrado na figura 25).

Figura 25 – Estratégia adotada para construção de subconjuntos de dados para treino e teste dos classificadores

	#neurônios ocultos	
	Interrogativa (qu)	Interrogativa (s/n)
Treino ←	Quando a Waine pagou?	Waine comprou um carro?
	Porque a Waine pagou?	Isso é seu?
	O que é isso?	Você se formou?
Teste ←	Como faz isso?	Você gosta de mim?
	Onde você mora?	Você vai embora?

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os modelos classificadores foram obtidos nos seguintes contextos:

- Experimento #0: treinamento com os dados binários;
- Experimento #1: treinamento com os dados multiclasse com três classes, sendo EFG neutra, EFG_1 e EFG_2 ;
- Experimento #2: treinamento com os dados multiclasse com todas EFGs e a EFG neutral.

⁴ Este método consiste em dividir o conjunto total de dados em dois subconjuntos mutuamente exclusivos, um para treinamento (estimação dos parâmetros) e outro para teste (validação) (KOHAVI et al., 1995).

4.5 Avaliação dos resultados

Para a avaliação dos resultados produzidos pela *Multilayer Perceptron* e mistura de especialistas foram utilizadas as medidas da matriz de confusão. As medidas calculadas foram *F-score*, revocação e precisão e acurácia. Os erros de borda e de segmentação foram também calculados apresentando erros da borda, erros totais e erros na borda. De maneira mais detalhada, a forma de aferir o erro de classificação pode ser por meio do erro total cometido pelo classificador em termos de número de quadros classificados erroneamente e a porcentagem que ele representa do total de quadros apresentados no teste. Os erros na borda são medidos em termos de número de quadros de borda classificados erroneamente e quanto isso representa do total de quadros classificados errados no teste, por exemplo 11110000, se o classificador responder 10000000 ou 11110001 o classificador apresentou três erros de borda (os três primeiros rótulos do primeiro exemplo e os três últimos rótulos do segundo exemplo. Os erros de segmentação correspondem a grandes sequências de quadros classificados erroneamente.

4.6 Reconhecimento de expressões faciais gramaticais: resultados e análises

Esta seção é dedicada a discutir os resultados obtidos nos três experimentos executados neste estudo. Os resultados são apresentados em termos de: *F-score*, precisão e revocação para o caso dos classificadores binários (*Experimento #0*) e em termos de estatística descritiva para a acurácia no caso de classificadores multiclasse (*Experimento #1* e *Experimento #2*). Foi realizada também uma avaliação voltada ao entendimento dos tipos de erros de segmentação cometidos pelos classificadores multiclasse.

4.6.1 Resultados com o problema de classificação binária: Experimento #0

Este experimento tratou o problema de classificação binária, considerando uma EFG e a expressão neutra. Na maioria dos casos, os classificadores criados para as diferentes representações de dados não levaram a resultados com diferenças significativas, com exceção do uso de dados não janelados, cujos resultados são inferiores aqueles obtidos com dados janelados, principalmente para EFGs com movimento de cabeça (*Afirmativa* e *Negativa*). Esta constatação reforça a necessidade de uso de informação de caráter temporal no

tratamento do problema de segmentação de EFGs. Neste experimento com a *Multilayer Perceptron* foi utilizada a representação janelada mais simples e de menor dimensão: oito pontos normalizados e janela de tamanho 3. Foram realizadas dez execuções, com diferentes condições de inicialização aleatória de pesos e passagem aleatória de dados durante o treinamento das MLPs, usando os parâmetros que apresentaram os melhores resultados para a maioria das execuções preliminares que consideraram a parametrização apresentada na seção 4.5. Para esses experimentos foi utilizado a seguinte parametrização: 6 neurônios ocultos, taxa de aprendizado em 0,1 para testes dependentes do intérprete e para testes independentes do intérprete e 3.000 épocas de treinamento. Os resultados de *F-score*, precisão, revocação e acurácia média obtida com esses parâmetros são mostrados nas tabelas 5 e 6 que apresenta os resultados com os oito pontos normalizados, considerando as análises dependente e independente do intérprete. Para o experimento com as misturas de especialistas utilizou a mesma representação de pontos e janelas e também foram realizadas dez execuções, com diferentes condições de inicialização da ME, sendo elas: inicialização aleatória e inicialização com *K-means*. A parametrização utilizada para esses experimentos foi de 4 especialistas e 0,1 de taxa de aprendizado. Os resultados desses experimentos são apresentados nas tabelas 7, 8 e 9, apresentando também os resultados dos testes dependentes e independentes do intérprete.

Multilayer Perceptron

O experimento para a classificação binária foi realizado para verificar os resultados obtidos em trabalhos anteriores. Neste experimento, foram criados classificadores dedicados à análise de uma EFG contra a expressão neutra. Os resultados obtidos para o presente trabalho confirmam as principais conclusões obtidas em trabalhos anteriores.

As EFGs que envolvem movimentação da cabeça (negativa e afirmativa) são as mais difíceis de serem diferenciadas da expressão neutra. Os melhores classificadores são obtidos sobre os dados do intérprete A, com uso de janelas de tamanho 3, e alcançam $\{0,80, 78, 0,82, 0,79\}$ e $\{0,94, 0,98, 0,91, 0,94\}$, respectivamente como *F-score*, precisão, revocação e acurácia para as EFGs afirmativa e negativa. Já para o intérprete B os resultados são inferiores. Para a EFG negativa, os índices são $\{0,75, 0,88, 0,65, 0,73\}$, e

para a EFG afirmativa $\{0,53\ 0,60\ 0,48\ 0,65\}$. Os índices indicam que o classificador não foi capaz de extrair adequadamente os padrões.

Na análise visual das execuções das expressões, percebe-se que o intérprete A as executa de maneira mais marcada, com maior movimentação dos elementos da face. O intérprete B executa movimentos mais discretos, o que dificulta a extração de características discriminantes para indução de uma boa superfície de decisão pelo classificador. Para as demais EFGs, os valores de *F-score* variam de 0,95 a 0,99 no reconhecimento de expressões executados pelo intérprete A, e de 0,92 a 0,94 para as expressões executadas pelo intérprete B. Todos esses resultados (Tabela 5) levam em consideração a representação usando coordenadas (x, y) normalizadas de oito pontos da face, e testes dependentes do intérprete (treino com três sentenças executadas pelo intérprete e teste em duas diferentes sentenças executadas pelo mesmo intérprete).

Tabela 5 – *F-score*, precisão, revocação e acurácia de classificadores no problema de duas classes com **oito pontos** da face: uma EFG mais a expressão neutra. Análise dependente do intérprete. Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	<i>F-score</i>	precisão	revocação	acurácia
Intérprete A				
Afirmativa	0,80	0,78	0,82	0,79
Interrogativa (Dúvida)	0,95	0,95	0,96	0,94
Foco	0,99	0,99	0,99	0,99
Negativa	0,94	0,98	0,91	0,94
Interrogativa (Qu)	0,97	0,98	0,95	0,97
Interrogativa (s/n)	0,97	0,95	0,98	0,96
Intérprete B				
Afirmativa	0,53	0,60	0,48	0,65
Interrogativa (Dúvida)	0,94	0,94	0,95	0,95
Foco	0,93	0,94	0,93	0,93
Negativa	0,75	0,88	0,65	0,73
Interrogativa (Qu)	0,92	0,96	0,88	0,90
Interrogativa (s/n)	0,94	0,94	0,93	0,92

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para o caso de testes independentes do intérprete, i.e. treino com três sentenças executadas por um intérprete e teste em duas sentenças executadas pelo outro intérprete, os melhores resultados alcançados foram $\{0,83\ 0,94\ 0,73\ 0,84\}$ para *F-score*, precisão, revocação e acurácia na EFG *Interrogativa (Dúvida)*, no caso em que o treinamento foi realizado com sentenças executadas pelo intérprete A, e teste com sentenças executadas pelo intérprete B. As expressões *Interrogativa (Dúvida)* são marcadas pelo abaixamento das

sobrancelhas. Para as demais EFGs, apenas a *Afirmativa* e a *Interrogativa (qu)* trouxeram uma complexidade tal que os *F-scores* ficaram com resultados baixos. Já para o teste contrário, os melhores resultados foram obtidos com a EFG *Foco*: $\{0,90 \ 1,0 \ 0,82 \ 0,83\}$. As expressões faciais gramaticais do tipo *Foco* são executadas com uma intensidade maior, levantando as sobrancelhas e abrindo os olhos intensamente. Para as demais EFGs, os resultados foram satisfatórios, alcançando *F-scores* altos. A tabela 6 apresenta os resultados.

Tabela 6 – *F-score*, precisão, revocação e acurácia de classificadores no problema de duas classes com **oito pontos** da face: uma EFG mais a expressão neutra. Análise dependente do intérprete. Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	<i>F-score</i>	precisão	revocação	acurácia
Intérprete A - Intérprete B				
Afirmativa	0,00	0,00	0,00	0,66
Interrogativa (Dúvida)	0,83	0,94	0,73	0,84
Foco	0,65	0,48	1,00	0,72
Negativa	0,68	0,71	0,66	0,70
Interrogativa (Qu)	0,00	0,00	0,00	0,41
Interrogativa (s/n)	0,82	0,97	0,71	0,75
Intérprete B - Intérprete A				
Afirmativa	0,69	1,00	0,52	0,52
Interrogativa (Dúvida)	0,79	0,66	0,99	0,79
Foco	0,90	1,00	0,82	0,83
Negativa	0,75	1,00	0,60	0,63
Interrogativa (Qu)	0,83	0,98	0,71	0,79
Interrogativa (s/n)	0,75	1,00	0,60	0,61

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A figura 26 ilustra os resultados obtidos nos experimentos dependentes do intérprete. A EFG *Afirmativa* obteve o pior resultado para ambos os testes, e os melhores resultados foram obtidos com a EFG *Foco* para os testes dependentes do intérprete A e *Interrogativa (Dúvida)* para o intérprete B. Para esses testes o classificador teve um comportamento semelhante para ambos experimentos, o que não acontece para os experimentos com os testes independentes.

Figura 26 – F -score dos testes dependentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os gráficos da figura 27 apresentam o comportamento dos testes independentes do intérprete. Nota-se a instabilidade nas execuções. O classificador não alcança resultados bons treinando com os dados provenientes do intérprete A e testando com os dados provenientes do intérprete B. O classificador chega a apresentar F -score 0,00⁵ nas EFGs *Afir mativa* e *Interrogativa (Qu)*. Para treinos com os dados provenientes do intérprete B e teste com os dados provenientes do intérprete A, o classificador responde melhor. Porém, ainda chega a apresentar F -score 0,00 no caso da EFG *Interrogativa (Qu)*.

Figura 27 – F -score dos testes independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Analisando os gráficos fica visível que o classificador MLP alcança resultados melhores quando analisa os dados provenientes da fala do intérprete A. No entanto, para ambos os intérpretes, os modelos de classificação possuem mais dificuldade em reconhecer as EFGs de afirmação e negação, nos testes dependentes do usuário. Assim, confirma-se

⁵ Esse resultado nulo é obtido por que nenhum verdadeiro-positivo foi obtido.

que a movimentação da cabeça de um lado para o outro (negação) e para cima e para baixo (afirmação) pode dificultar o reconhecimento das EFGs.

Como comentado, o objetivo principal dos experimentos realizados com a classificação binária e o classificador baseado em MLP foi verificar resultados de trabalhos anteriores. Entretanto, além disso, com esses experimentos foi possível construir um conhecimento geral sobre a dificuldade inerente ao problema para então trabalhar com os experimentos usando a mistura de especialistas.

Mistura de especialistas com inicialização aleatória

Assim como para MLP, os classificadores construídos com a mistura de especialistas também apresentaram os resultados mais baixos para a EFG *Afirmativa*. Os melhores classificadores são obtidos sobre os dados do intérprete A, confirmando os resultados obtidos pela MLP.

Nesse experimento, o classificador obteve resultados mais baixos, inclusive para a EFG *Foco*. Comparando os resultados obtidos para a EFG *Foco* com a EFG *Afirmativa* para os testes com os dados provenientes do intérprete A, tem-se $\{0,85 \ 1,00 \ 0,74 \ 0,74\}$ e $\{0,77 \ 0,73 \ 0,82 \ 0,77\}$, respectivamente como *F-score*, precisão, revocação e acurácia. Para os dados provenientes do intérprete B os resultados obtidos são inferiores. Para a EFG *Afirmativa* e para a EFG *Negativa*, os índices são $\{0,59 \ 0,78 \ 0,48 \ 0,64\}$ e $\{0,63 \ 1,00 \ 0,46 \ 0,46\}$. Mais uma vez é confirmado que os índices indicam que o classificador não foi capaz de extrair adequadamente os padrões. A tabela 7 resume os resultados obtidos nos testes dependentes do usuário usando mistura de especialistas.

A tabela 8 apresenta os resultados com os testes independentes do intérprete. Para os treinos com os dados provenientes do intérprete A e teste com os dados do intérprete B os resultados apresentaram melhoras com a mistura de especialistas. Logo com a MLP as EFGs *Afirmativa* e *Interrogativa (Qu)* apresentaram respectivamente 0,00 para o *F-score* e para os experimentos com a ME apresentam 0,50 para a EFG *Afirmativa* e 0,74 para *Interrogativa (Qu)*. A mistura de especialistas para esta experimentação exhibe resultados melhores que a MLP.

Tabela 7 – *F-score*, precisão, revocação e acurácia de classificadores no problema de duas classes **com oito pontos** da face: uma EFG mais a expressão neutra. Análise dependente do intérprete. Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	<i>F-score</i>	precisão	revocação	acurácia
Intérprete A				
Afirmativa	0,77	0,73	0,82	0,77
Interrogativa (Dúvida)	0,95	0,93	0,96	0,93
Foco	0,85	1,00	0,74	0,74
Negativa	0,91	0,96	0,87	0,89
Interrogativa (Qu)	0,94	0,93	0,95	0,94
Interrogativa (s/n)	0,93	0,94	0,91	0,91
Intérprete B				
Afirmativa	0,59	0,78	0,48	0,64
Interrogativa (Dúvida)	0,94	0,94	0,94	0,95
Foco	0,92	0,96	0,89	0,92
Negativa	0,63	1,00	0,46	0,46
Interrogativa (Qu)	0,92	0,98	0,87	0,90
Interrogativa (s/n)	0,95	0,96	0,95	0,94

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Tabela 8 – *F-score*, precisão, revocação e acurácia de classificadores no problema de duas classes **com oito pontos** da face: uma EFG mais a expressão neutra. Análise dependente do intérprete. Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	<i>F-score</i>	precisão	revocação	acurácia
Intérprete A - Intérprete B				
Afirmativa	0,50	1,00	0,33	0,33
Interrogativa (Dúvida)	0,57	0,93	0,42	0,45
Foco	0,70	1,00	0,54	0,54
Negativa	0,63	1,00	0,46	0,46
Interrogativa (Qu)	0,74	1,00	0,59	0,59
Interrogativa (s/n)	0,76	1,00	0,61	0,61
Intérprete B - Intérprete A				
Afirmativa	0,69	1,00	0,52	0,52
Interrogativa (Dúvida)	0,93	0,88	0,99	0,92
Foco	0,83	0,98	0,90	0,84
Negativa	0,72	1,00	0,56	0,56
Interrogativa (Qu)	0,86	0,81	0,93	0,87
Interrogativa (s/n)	0,77	1,00	0,63	0,64

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para essa configuração de experimentos, que considera os testes independentes do intérprete, apenas a EFG *Afirmativa* do teste com os dados do intérprete B não alcançam resultados bons de *F-score*, e vale ressaltar que para a MLP a EFG obteve o mesmo

comportamento. A EFG *Afirmativa* para ambos intérpretes apresentaram resultados piores. A EFG *Foco*, a ME apresentou resultado menor do que a MLP, o que indica que a MLP consegue reconhecer os padrões dessa EFG com mais facilidade que a ME. A EFG *Interrogativa (Qu)* e *Interrogativa (Dúvida)*, para ambos intérpretes e técnicas de classificação foram as que geraram os melhores resultados, ambas EFGs são caracterizadas pelo abaixamento da sobrancelha o que pode ser um facilitador para reconhece-las, o que diferencia uma EFG da outra é a movimentação da cabeça.

Mistura de especialistas com inicialização com *K-means*

Para esses experimentos foi escolhido o melhor e o pior resultado da execução da mistura de especialistas com inicialização aleatória e executado com a mistura de especialistas com a inicialização com *K-means*. Para alguns casos houve melhora em seus resultados e para os demais resultados foi obtido números semelhantes à inicialização aleatória. A tabela 9 apresenta os resultados com a EFG *Afirmativa* (pior caso) e *Interrogativa (Dúvida)* (melhor caso) para o intérprete A, e os resultados com a EFG *Afirmativa* (pior caso) e a *Interrogativa (s/n)* (melhor caso) para o intérprete B. A EFG *Afirmativa* dos dados provenientes do intérprete A teve uma melhora com a inicialização com *K-means* o qual alcançou 0,82 e para a inicialização aleatória obteve 0,77 de *F-score*.

Tabela 9 – *F-score*, precisão, revocação e acurácia de classificadores no problema de duas classes com **oito pontos** da face: uma EFG mais a expressão neutra. Análise dependente e independente do intérprete.

EFGs	<i>F-score</i>	precisão	revocação	acurácia
Intérprete A				
Afirmativa	0,82	0,82	0,82	0,81
Interrogativa (Dúvida)	0,94	0,93	0,95	0,93
Intérprete B				
Afirmativa	0,50	1,00	0,33	0,33
Interrogativa (s/n)	0,94	0,95	0,93	0,92
Intérprete A - Intérprete B				
Afirmativa	0,00	0,00	0,00	0,65
Interrogativa (Dúvida)	0,57	1,00	0,40	0,40
Intérprete B - Intérprete A				
Afirmativa	0,74	1,00	0,58	0,62
Interrogativa (s/n)	0,75	1,00	0,60	0,60

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para os testes independentes com os dados do intérprete A e teste com os dados do intérprete B a EFG *Afirmativa* obteve um pior resultado comparado com a inicialização aleatória que teve como resultado 0,50 e para inicialização com *K-means* obteve 0,00 de *F-score*.

Os gráficos da figura 28 apresentam uma visão geral do desempenho de cada rede com seu pior e melhor caso, para casos dependentes e independentes. Para os casos dependentes nota-se que as redes MLP e ME têm resultados semelhantes. Com relação aos testes independentes com os dados provenientes do intérprete A testando com o intérprete B para a EFG *Afirmativa* (pior caso) a única rede que conseguiu um resultado mais alto foi a ME com inicialização aleatória, as outras redes apresentaram resultado 0,0 em seus respectivos resultados *F-score*. Do mesmo modo que para o caso da EFG dos dados provenientes do intérprete B testados com os dados do intérprete A a MLP obteve um resultado melhor com a EFG *Interrogativa (Dúvida)* atingindo 0,83 e a ME com inicialização aleatória e inicialização com *K-means* atingiram 0,57. Os demais resultados foram semelhantes em seus desempenhos apresentando resultados muito próximos. Para os casos dos experimentos com a classificação binária a ME com inicialização aleatória foi a que obteve resultados uniformes comparados as outras duas redes.

Figura 28 – *F-score* dos piores e melhores casos das redes MLP e ME, casos dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

4.6.2 Resultados com o problema de classificação multiclasse: Experimento #1

Multilayer Perceptron

Este experimento tratou o problema de classificação multiclasse, considerando duas EFGs e a expressão neutra. De forma geral, os classificadores criados para as diferentes

representações de dados não levaram a resultados com diferenças significativas, com exceção do uso de dados não janelados, cujos resultados são inferiores aqueles obtidos com dados janelados, principalmente para EFGs com movimento de cabeça. Esta constatação reforça a necessidade de uso de informação de caráter temporal no tratamento do problema de segmentação de EFGs.

Neste experimento são reportados os resultados obtidos com a representação janelada mais simples e de menor dimensão: oito pontos normalizados e distância dos pontos e janela de tamanho 3. Foram realizadas dez execuções com diferentes condições de inicialização aleatória de pesos, usando os parâmetros que apresentaram os melhores resultados para a maioria das execuções preliminares que consideraram a parametrização apresentada na seção 4.4. Os parâmetros escolhidos foram: 13 neurônios ocultos, taxa de aprendizado em 0,1 para testes dependentes e independentes do intérprete e 3.000 épocas de treinamento. Os resultados de acurácia média obtida é apresentada na tabela 10 (oito pontos).

Para ambos os intérpretes, os melhores resultados foram obtidos a partir da combinação das EFGs *Foco - Interrogativa (s/n)* que é caracterizada pelo uso das sobrancelhas. Por outro lado os piores resultados foram obtidos quando a combinação de EFG *Afirmativa* com *Negativa* foram combinadas.

Os resultados mostrados na tabela 10 confirmam a dificuldade do classificador em relação aos dados provenientes do intérprete B. A complexidade inerente ao movimento da cabeça é reforçada nos resultados da análise multiclasse, visto que os classificadores apresentam os menores desempenhos de acurácia quando as EFGs *Afirmativas* e *Negativas* estão envolvidas no problema de decisão, com destaque para a EFG *Afirmativa*. No estudo com as sentenças executadas pelo intérprete B, os classificadores mais instáveis são aqueles construídos para segmentação dessas EFGs. Os melhores resultados foram obtidos em problemas que envolveram combinações de EFGs nas quais o movimento da sobrancelha e da boca são mais acentuados (*Foco* e *Interrogativa (s/n)*). É notável que para os dois intérpretes quando a combinação de EFGs *Afirmativa* e *Negativa* é analisada representam a maior dificuldade uma vez que os resultados obtidos alcançam a média mais baixa. Já no caso da combinação de EFGs *Foco* e *Interrogativa (s/n)* a média apresentada é mais alta. Ao considerar testes realizados de maneira independente do intérprete, os classificadores multiclasse perdem desempenho.

A tabela 11 mostra os resultados obtidos com os testes independentes do intérprete. Para esses resultados apenas os treinos com os dados provenientes do intérprete B e teste

Tabela 10 – Acurácia de classificadores no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete. Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A				
Afirm. - Int. (Dúvida)	0,87	0,01	0,88	0,86
Afirm. - Foco	0,88	0,01	0,88	0,87
Afirm. - Negativa	0,84	0,01	0,85	0,83
Afirm.- Inter. (Qu)	0,89	0,01	0,90	0,87
Afirm. - Inter. (s/n)	0,89	0,01	0,89	0,87
Inter. (Dúvida) - Foco	0,96	0,00	0,96	0,96
Inter. (Dúvida) - Negativa	0,93	0,00	0,93	0,92
Inter. (Dúvida) - Inter. (Qu)	0,94	0,00	0,95	0,94
Inter. (Dúvida) - Inter. (s/n)	0,94	0,00	0,95	0,94
Foco - Negativa	0,95	0,01	0,95	0,91
Foco - Inter. (Qu)	0,95	0,01	0,97	0,94
Foco - Inter. (s/n)	0,97	0,00	0,97	0,96
Negativa - Inter. (Qu)	0,93	0,00	0,93	0,92
Negativa - Inter. (s/n)	0,93	0,00	0,93	0,92
Inte. (Qu) - Inter. (s/n)	0,95	0,01	0,96	0,94
Intérprete B				
Afirm. - Inter. (Dúvida)	0,77	0,02	0,80	0,74
Afirm. - Foco	0,70	0,01	0,72	0,68
Afirm. - Negativa	0,60	0,02	0,64	0,54
Afirm. - Inter. (Qu)	0,73	0,01	0,74	0,69
Afirm. - Inter. (s/n)	0,81	0,01	0,82	0,79
Inter. (Dúvida) - Foco	0,91	0,01	0,94	0,89
Inter. (Dúvida) - Negativa	0,80	0,01	0,83	0,78
Inter. (Dúvida) - Inter. (Qu)	0,89	0,01	0,90	0,88
Inter. (Dúvida) - Inter. (s/n)	0,91	0,01	0,93	0,90
Foco - Negativa	0,72	0,03	0,75	0,68
Foco - Inter. (Qu)	0,87	0,02	0,89	0,85
Foco - Inter. (s/n)	0,92	0,00	0,93	0,92
Negativa - Inter. (Qu)	0,75	0,02	0,77	0,71
Negativa - Inter. (s/n)	0,82	0,03	0,86	0,78
Inter. (Qu) - Inter. (s/n)	0,88	0,02	0,90	0,85

Fonte: Maria Eduarda de Araújo Cardoso, 2018

com os dados provenientes do intérprete A atingiram a acurácia média com resultados altos, sendo elas a combinação das EFGs *Afirmativa - Interrogativa (Dúvida)*, *Afirmativa - Foco*, *Interrogativa (Dúvida) - foco*, *Interrogativa (Dúvida) - Negativa*, *Interrogativa (Dúvida) - Interrogativa (s/n)*, *Foco - Interrogativa (Qu)*, *Foco - Interrogativa (s/n)* e *Interrogativa (Qu) - Interrogativa (s/n)* a partir desses resultados pode-se inferir que a rede tem certa facilidade em reconhecer os dados do intérprete A, que é uma sinalização com expressões

bem definidas (expressões marcadas). Para os testes com os dados provenientes do intérprete A e testados com os dados do intérprete B, nenhum resultado alcançou resultados altos atingindo uma mínima de 0,19 para a combinação de EFG *Foco - Interrogativa (s/n)* e a máxima 0,57 para *Afirmativa - Interrogativa (s/n)*.

Tabela 11 – Acurácia de classificadores no problema de três classes com **oito pontos da face**: duas EFGs mais a expressão neutra. Análise independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A (treino) - Intérprete B (teste)				
Afir. - Inter. (Dúvida)	0,48	0,07	0,59	0,38
Afir. - Foco	0,42	0,03	0,45	0,38
Afir. - Negativa	0,43	0,04	0,48	0,34
Afir. - Inter. (Qu)	0,37	0,06	0,46	0,30
Afir. - Inter. (s/n)	0,57	0,03	0,60	0,52
Inter. (Dúvida) - Foco	0,53	0,04	0,63	0,49
Inter. (Dúvida) - Negativa	0,44	0,04	0,55	0,41
Inter. (Dúvida) - Inter. (Qu)	0,54	0,07	0,64	0,41
Inter. (Dúvida) - Inter. (s/n)	0,55	0,05	0,64	0,50
Foco - Negativa	0,24	0,06	0,34	0,16
Foco - Inter. (Qu)	0,24	0,01	0,25	0,23
Foco - Inter. (s/n)	0,19	0,03	0,24	0,16
Negativa - Inter. (Qu)	0,33	0,04	0,40	0,28
Negativa - Inter. (s/n)	0,40	0,08	0,52	0,29
Inter. (Qu) - Inter. (s/n)	0,32	0,05	0,39	0,25
Intérprete B (treino) - Intérprete A (teste)				
Afirm. - Inter. (Dúvida)	0,66	0,11	0,75	0,38
Afirm. - Foco	0,65	0,0	0,65	0,65
Afirm. - Negativa	0,56	0,05	0,70	0,53
Afirm. - Inter. (Qu)	0,52	0,08	0,74	0,45
Afirm. - Inter. (s/n)	0,57	0,00	0,57	0,57
Inter. (Dúvida) - Foco	0,85	0,04	0,89	0,77
Inter. (Dúvida) - Negativa	0,66	0,06	0,73	0,54
Inter. (Dúvida) - Inter. (Qu)	0,43	0,04	0,47	0,35
Inter. (Dúvida) - Inter. (s/n)	0,63	0,15	0,77	0,37
Foco - Negativa	0,53	0,09	0,64	0,37
Foco - Inter. (Qu)	0,67	0,05	0,76	0,63
Foco - Inter. (s/n)	0,70	0,05	0,76	0,60
Negativa - Inter. (Qu)	0,48	0,07	0,65	0,40
Negativa - Inter. (s/n)	0,42	0,17	0,67	0,16
Inter. (Qu) - Inter. (s/n)	0,61	0,15	0,73	0,24

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os casos nos quais os resultados foram baixos, foi considerado que os classificadores não foram capazes de reconhecer o padrão de execução das EFGs. Nenhum caso de teste alcançou bons resultados para classificadores treinados com sentenças executadas pelo intérprete A e testados com sentenças executadas pelo intérprete B. Ainda, é notável que há uma instabilidade nos modelos obtidos, indicando a complexidade inerente ao problema e a necessidade de mais investigação sobre o problema.

Os erros de borda, erros gerais e erros na borda cometidos nesse experimento são ilustrados pelos gráficos nas figuras 29, 30 e 31. Esses erros foram calculados sobre o classificador que atingiu a acurácia mais alta em cada um dos problemas, e para delineamento da borda seis quadros foram considerados na transição entre as EFGs.

Figura 29 – Análise de erros na borda no problema de três classes: duas EFGs mais a expressão neutra com **oito pontos** da face. Análise em testes dependentes e independentes do intérprete

A análise dos erros de borda fornece uma visão do tipo e importância do erro cometido pelo classificador. Erros na borda podem ser considerados de menor importância do que erros cometidos dentro da ocorrência de uma EFG ou da expressão neutra, mas vez que a borda é sujeita à imprecisão da rotulação realizada pelos especialistas, que neste trabalho é usada como verdade fundamental para treino e avaliação dos classificadores. Sendo assim, quanto maior é a porcentagem de erros de classificação cometidos na borda e menor a quantidade de erros cometidos fora da borda, maior a robustez do classificador.

Analisando os erros de borda, nota-se que a EFG *Afirmativa* novamente aparece como um dificultador da segmentação de EFGs, principalmente para os testes com os dados provenientes do intérprete B. Para esse intérprete, a EFG negativa também aparece em contexto de classificadores que erram mais dentro das sequências de EFGs do que nas bordas. A análise dos resultados com os dados provenientes do intérprete B apresenta erros maiores em segmentações das EFGs e o para o caso dos dados provenientes do intérprete A apresenta maiores erros na borda.

Figura 30 – Análise de erros na borda no problema de três classes: duas EFGs mais a expressão neutra com **oito pontos** da face. Análise em testes dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A combinação das EFGs *Foco -Interrogativa (s/n)* atingiram 100% de erros na borda para os dados provenientes do intérprete A apresentando uma dificuldade para

esse caso. Os erros de segmentação e erros totais a EFG *Afirmativa* é que apresenta resultados mais altos para ambos intérpretes. Ao analisar os erros cometidos fora da borda, nota-se com mais evidência que os classificadores cometem erros importantes, dentro das sequências de quadros das EFGs ou expressões neutras o que mostra que a segmentação de EFGs apresenta um desafio para generalização dos modelos.

A figura 31 apresenta os resultados dos erros gerais e de borda do melhor caso *Foco - Interrogativa (s/n)* e o pior caso *Afirmativa - Negativa* para os dados provenientes do intérprete A. E melhor caso a combinação das EFGs *Foco - Interrogativa (s/n)* e pior caso *Afirmativa - Negativa* com os dados provenientes do intérprete B. Nesses resultados o melhor caso do intérprete A apresentou a menor taxa de erros.

Figura 31 – Análise de erros gerais e de borda no problema de três classes com **oito pontos** da face. Análise em testes dependentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A figura 32 apresenta os resultados dos testes independentes. Para os dados provenientes do intérprete A e teste com os dados do intérprete B a combinação das EFGs *Interrogativa (Dúvida) - Interrogativa (s/n)* representam o melhor caso e combinação de EFGs *Foco - Interrogativa (s/n)* pior caso. Para dos dados provenientes do intérprete B e teste com os dados do intérprete A o melhor caso foi *Interrogativa (Dúvida) - Foco* e pior caso *Negativa - Interrogativa (s/n)* que apresentou o maior número de erros gerais.

Ambos os treinamentos com intérpretes independentes apresentam um número alto de erros gerais comparados aos testes dependentes. A EFG *Foco - Interrogativa (s/n)* apresenta uma quantidade grande de erros de borda e erros gerais para os testes com

os dados provenientes do intérprete A testados com os dados do intérprete B. Para os dados provenientes do intérprete B testados com os dados do intérprete A os erros na borda são menores comparando os resultados. Em comparação dos erros gerais entre testes dependentes e independentes do intérprete é importante notar que para os testes dependentes o número maior de erros é 282 e esse mesmo chega a 655 (resultados nas tabelas 29 e 30 no apêndice) para os independentes. O que afirma que há uma complexidade muito grande para resultados com testes independentes.

Figura 32 – Análise de erros gerais e de borda no problema de três classes com **oito pontos** da face. Análise em testes independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Experimentos com a distância dos pontos também foram realizados e os resultados de acurácia média obtida são apresentados na tabela 12, considerando a análise dependente do intérprete. Os parâmetros escolhidos foram: 33 neurônios ocultos, taxa de aprendizado em 0,1 para testes dependentes do intérprete e para testes independentes do intérprete e 3.000 épocas de treinamento. Em comparação aos resultados anteriores com 8 e 9 pontos (resultados no apêndice A), percebe-se um aumento de 0,1 aos resultados das combinações das EFGs *Afirmativa - Interrogativa (Dúvida)*, *Afirmativa - Negativa* e um aumento de 0,2 nas combinações das EFGs *Afirmativa - Foco* para ambos intérpretes.

Tabela 12 – Acurácia de classificadores no problema de três classes com a **distância dos pontos**: duas EFGs mais a expressão neutra. Análise dependente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A				
Afirm. - Int. (Dúvida)	0,88	0,01	0,89	0,87
Afirm. - Foco	0,90	0,01	0,91	0,89
Afirm. - Negativa	0,85	0,00	0,86	0,84
Afirm.- Inter. (Qu)	0,88	0,01	0,89	0,87
Afirm. - Inter. (s/n)	0,88	0,01	0,89	0,87
Inter. (Dúvida) - Foco	0,96	0,00	0,97	0,96
Inter. (Dúvida) - Negativa	0,92	0,01	0,93	0,91
Inter. (Dúvida) - Inter. (Qu)	0,93	0,00	0,93	0,92
Inter. (Dúvida) - Inter. (s/n)	0,94	0,00	0,95	0,93
Foco - Negativa	0,95	0,01	0,95	0,94
Foco - Inter. (Qu)	0,96	0,00	0,97	0,96
Foco - Inter. (s/n)	0,96	0,00	0,97	0,95
Negativa - Inter. (Qu)	0,91	0,01	0,93	0,90
Negativa - Inter. (s/n)	0,91	0,01	0,93	0,90
Inte. (Qu) - Inter. (s/n)	0,95	0,01	0,96	0,93
Intérprete B				
Afirm. - Inter. (Dúvida)	0,73	0,01	0,76	0,71
Afirm. - Foco	0,71	0,02	0,74	0,69
Afirm. - Negativa	0,58	0,02	0,61	0,54
Afirm. - Inter. (Qu)	0,71	0,03	0,73	0,62
Afirm. - Inter. (s/n)	0,78	0,01	0,80	0,77
Inter. (Dúvida) - Foco	0,85	0,01	0,86	0,85
Inter. (Dúvida) - Negativa	0,74	0,02	0,78	0,70
Inter. (Dúvida) - Inter. (Qu)	0,87	0,01	0,89	0,86
Inter. (Dúvida) - Inter. (s/n)	0,88	0,01	0,90	0,86
Foco - Negativa	0,63	0,01	0,65	0,60
Foco - Inter. (Qu)	0,87	0,00	0,87	0,86
Foco - Inter. (s/n)	0,89	0,01	0,90	0,88
Negativa - Inter. (Qu)	0,76	0,01	0,79	0,74
Negativa - Inter. (s/n)	0,80	0,02	0,82	0,77
Inter. (Qu) - Inter. (s/n)	0,90	0,01	0,91	0,87

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A combinação das EFGs *Afirmativa - Interrogativa (Qu)*, *Interrogativa (Dúvida) - Negativa*, *Interrogativa (Dúvida) - Interrogativa (Qu)*, *Negativa - Interrogativa (Qu)*, *Negativa - Interrogativa (s/n)* para os resultados dos dados provenientes do intérprete A obtiveram uma queda de 0,1 a 0,2, logo para os dados provenientes do intérprete B todos

os resultados tiveram uma queda em seus valores para essa parametrização. A tabela 13 apresenta os resultados dos testes independentes do intérprete com a distância dos pontos.

Tabela 13 – Acurácia de classificadores no problema de três classes com a **distância dos pontos**: duas EFGs mais a expressão neutra. Análise independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A (treino) - Intérprete B (teste)				
Afir. - Inter. (Dúvida)	0,40	0,05	0,47	0,35
Afir. - Foco	0,47	0,07	0,57	0,37
Afir. - Negativa	0,31	0,05	0,37	0,22
Afir. - Inter. (Qu)	0,46	0,08	0,57	0,31
Afir. - Inter. (s/n)	0,46	0,17	0,60	0,10
Inter. (Dúvida) - Foco	0,51	0,06	0,59	0,39
Inter. (Dúvida) - Negativa	0,38	0,05	0,05	0,28
Inter. (Dúvida) - Inter. (Qu)	0,33	0,06	0,43	0,25
Inter. (Dúvida) - Inter. (s/n)	0,22	0,07	0,36	0,08
Foco - Negativa	0,36	0,03	0,41	0,31
Foco - Inter. (Qu)	0,27	0,07	0,47	0,23
Foco - Inter. (s/n)	0,27	0,09	0,40	0,16
Negativa - Inter. (Qu)	0,37	0,05	0,46	0,32
Negativa - Inter. (s/n)	0,27	0,01	0,29	0,25
Inter. (Qu) - Inter. (s/n)	0,29	0,02	0,31	0,26
Intérprete B (treino) - Intérprete A (teste)				
Afirm. - Inter. (Dúvida)	0,65	0,04	0,70	0,59
Afirm. - Foco	0,65	0,01	0,69	0,65
Afirm. - Negativa	0,46	0,07	0,53	0,35
Afirm. - Inter. (Qu)	0,56	0,08	0,71	0,51
Afirm. - Inter. (s/n)	0,57	0,00	0,57	0,57
Inter. (Dúvida) - Foco	0,79	0,03	0,84	0,73
Inter. (Dúvida) - Negativa	0,70	0,02	0,74	0,67
Inter. (Dúvida) - Inter. (Qu)	0,64	0,03	0,66	0,58
Inter. (Dúvida) - Inter. (s/n)	0,64	0,07	0,69	0,46
Foco - Negativa	0,45	0,10	0,59	0,25
Foco - Inter. (Qu)	0,68	0,08	0,75	0,52
Foco - Inter. (s/n)	0,72	0,07	0,78	0,53
Negativa - Inter. (Qu)	0,49	0,09	0,58	0,32
Negativa - Inter. (s/n)	0,24	0,01	0,26	0,23
Inter. (Qu) - Inter. (s/n)	0,57	0,14	0,74	0,24

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A análise desses resultados levaram a conclusão que para os dados provenientes do intérprete A testando com o B, os resultados considerando os pontos da distância eles sofrem uma queda para as EFGs *Afirmativa - Interrogativa (Dúvida)*, *Afirmativa - Negativa*, *Afirmativa - Interrogativa (s/n)*, *Interrogativa (Dúvida) - Foco*, *Interrogativa*

(Dúvida) - Negativa, Interrogativa (Dúvida) - Interrogativa (Qu), Interrogativa (Dúvida) - Interrogativa (s/n) e Negativa - Interrogativa (s/n). Para as demais EFGs houve uma melhora em seus resultados. Para o treino com os dados provenientes do intérprete B e teste com o A essa queda foi menor afetando apenas as combinações das EFGs *Afirmativa - Interrogativa (Dúvida)*, *Afirmativa - Negativa*, *Interrogativa (Dúvida) - Foco*, *Foco - Negativa* e *Negativa, Interrogativa (s/n)*. Para as demais EFGs, a melhora nos resultados foi de 0,1 a 0,4. As figuras 33 e 34 apresentam os erros na borda, i.e, erros de borda dividido pelos erros gerais.

Figura 33 – Análise de erros na borda no problema de três classes **com a distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete

Figura 34 – Análise de erros na borda no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para os resultados dos dados provenientes do intérprete A as combinações das EFGs *Foco - Interrogativa (Qu)* e *Foco - Interrogativa (s/n)* apresentam menos erros de borda e erros gerais, e as combinações das EFGs *Afirmativa - Negativa* comprovam a complexidade de reconhecimento da técnica para essas EFGs, apresentando maior índice de erros para ambos os intérpretes. Logo, os dados provenientes do intérprete B os erros foram menores que para o intérprete A em comparação aos erros na borda que cada um obteve.

A figura 35 apresenta os erros gerais e de borda para os casos dependentes do intérprete. O melhor caso dos dados provenientes do intérprete A foi a combinação das EFGs *Foco - Interrogativa (Qu)* e o pior caso foi a combinação das EFGs *Afirmativa - Negativa*. Para o intérprete B o melhor caso é a combinação da EFGs *Interrogativa (Qu) - Interrogativa (s/n)* e o pior caso *Afirmativa - Negativa*. Analisando o gráfico é perceptível que para o pior caso do intérprete B ele obteve o maior número de erros gerais.

Figura 35 – Análise de erros de borda e erros gerais no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A figura 36 apresenta os resultados dos erros de borda e erros gerais dos testes independentes, com os dados provenientes do intérprete A e testados com os dados provenientes do intérprete B e o caso contrário.

Figura 36 – Análise de erros de borda e erros gerais no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os dados provenientes do intérprete A e testados com os dados do intérprete B apresentam como melhor caso a combinação das EFGs *Interrogativa (Dúvida) - Foco* e para pior caso *Interrogativa (Dúvida) - Interrogativa (s/n)*. Para os dados provenientes do intérprete B e testados com os dados provenientes do intérprete A o melhor caso foi a combinação das EFGs *Interrogativa (Dúvida) - Foco* e o pior caso *Interrogativa (Dúvida) - Interrogativa (s/n)*. É notável que para os testes independentes os erros gerais e de borda aumentam assim como o resultado da média da acurácia obtém resultados baixos, o que confirma nesses testes de erros gerais e de borda os altos números de erros.

Mistura de especialistas com inicialização aleatória

Para esses experimentos são reportados os resultados obtidos com a representação janelada mais simples e de menor dimensão: oito pontos normalizados e também nove pontos normalizados (ponto do nariz), distância dos pontos e janela de tamanho 3 seguindo a mesma representação utilizada nos testes com a rede MLP. Foram realizadas dez execuções, com inicialização aleatória do número de especialistas, usando os parâmetros que apresentaram os melhores resultados para a maioria das execuções preliminares que consideraram a parametrização apresentada na seção 4.4. Os parâmetros escolhidos foram: 6 especialistas, alfa em 0,1 para testes dependentes do intérprete e para testes independentes do intérprete e 1.000 épocas de treinamento. Os resultados de acurácia média obtida com esses parâmetros são mostrados na tabela 14.

Diferente da MLP os resultados para essa representação foram inferiores em vários casos das combinações das EFGs. Para os dados provenientes do intérprete A a combinação das EFGs *Interrogativa (Dúvida) - Foco* apresenta uma média de acurácia baixa de 0,17 ocorrendo o pior caso para essa combinação. Para as combinações de EFGs para *Foco - Negativa* e *Interrogativa (Qu) - Interrogativa (s/n)* a média de acurácia foi alta de 0,95 correspondendo ao melhor caso. Para dos dados provenientes do intérprete A as combinações *Afirmativa - Negativa* e *Foco - Negativa* apresentam os piores casos, ressaltando a informação que essas EFGs *Afirmativa* e *Negativa* necessitam do movimento da cabeça para expressá-las e por isso levantasse a hipótese de que elas vão pior pela movimentação que elas exigem.

Tabela 14 – Acurácia de classificadores no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A				
Afirm. - Int. (Dúvida)	0,87	0,00	0,87	0,87
Afirm. - Foco	0,27	0,20	0,65	0,14
Afirm. - Negativa	0,84	0,01	0,85	0,84
Afirm.- Inter. (Qu)	0,87	0,08	0,91	0,65
Afirm. - Inter. (s/n)	0,89	0,01	0,89	0,89
Inter. (Dúvida) - Foco	0,17	0,03	0,19	0,13
Inter. (Dúvida) - Negativa	0,85	0,23	0,93	0,18
Inter. (Dúvida) - Inter. (Qu)	0,93	0,00	0,93	0,93
Inter. (Dúvida) - Inter. (s/n)	0,94	0,00	0,95	0,94
Foco - Negativa	0,95	0,01	0,95	0,94
Foco - Inter. (Qu)	0,53	0,39	0,96	0,13
Foco - Inter. (s/n)	0,22	0,26	0,95	0,12
Negativa - Inter. (Qu)	0,87	0,19	0,95	0,34
Negativa - Inter. (s/n)	0,78	0,30	0,93	0,18
Inte. (Qu) - Inter. (s/n)	0,95	0,00	0,96	0,95
Intérprete B				
Afirm. - Inter. (Dúvida)	0,66	0,20	0,77	0,28
Afirm. - Foco	0,41	0,18	0,71	0,24
Afirm. - Negativa	0,31	0,05	0,40	0,29
Afirm. - Inter. (Qu)	0,62	0,22	0,76	0,21
Afirm. - Inter. (s/n)	0,68	0,22	0,80	0,26
Inter. (Dúvida) - Foco	0,77	0,30	0,92	0,20
Inter. (Dúvida) - Negativa	0,44	0,21	0,76	0,24
Inter. (Dúvida) - Inter. (Qu)	0,90	0,00	0,90	0,89
Inter. (Dúvida) - Inter. (s/n)	0,93	0,01	0,93	0,92
Foco - Negativa	0,31	0,18	0,66	0,15
Foco - Inter. (Qu)	0,67	0,31	0,91	0,20
Foco - Inter. (s/n)	0,84	0,24	0,92	0,16
Negativa - Inter. (Qu)	0,60	0,19	0,71	0,19
Negativa - Inter. (s/n)	0,36	0,22	0,73	0,18
Inter. (Qu) - Inter. (s/n)	0,86	0,08	0,91	0,65

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A tabela 15 apresenta os resultados independentes do intérprete. Para o treinamento com os dados provenientes do intérprete A e teste com os dados do intérprete B. Os resultados dos dados provenientes do intérprete B testando com os dados do intérprete A os resultados que atingiram números favoráveis foram as combinações das EFGs *Afirmativa* - *Interrogativa (Dúvida)* e *Interrogativa (Dúvida)* - *Foco*.

Tabela 15 – Acurácia de classificadores no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A (treino) - Intérprete B (teste)				
Afir. - Inter. (Dúvida)	0,37	0,08	0,46	0,28
Afir. - Foco	0,31	0,08	0,44	0,19
Afir. - Negativa	0,33	0,03	0,39	0,30
Afir. - Inter. (Qu)	0,34	0,09	0,54	0,27
Afir. - Inter. (s/n)	0,24	0,04	0,36	0,23
Inter. (Dúvida) - Foco	0,33	0,14	0,58	0,20
Inter. (Dúvida) - Negativa	0,36	0,05	0,43	0,24
Inter. (Dúvida) - Inter. (Qu)	0,46	0,05	0,49	0,45
Inter. (Dúvida) - Inter. (s/n)	0,56	0,05	0,65	0,51
Foco - Negativa	0,36	0,03	0,41	0,31
Foco - Inter. (Qu)	0,23	0,02	0,25	0,20
Foco - Inter. (s/n)	0,24	0,05	0,35	0,16
Negativa - Inter. (Qu)	0,38	0,07	0,49	0,28
Negativa - Inter. (s/n)	0,45	0,04	0,47	0,35
Inter. (Qu) - Inter. (s/n)	0,29	0,11	0,60	0,22
Intérprete B (treino) - Intérprete A (teste)				
Afirm. - Inter. (Dúvida)	0,67	0,16	0,75	0,22
Afirm. - Foco	0,28	0,20	0,66	0,14
Afirm. - Negativa	0,27	0,09	0,51	0,21
Afirm. - Inter. (Qu)	0,40	0,14	0,51	0,22
Afirm. - Inter. (s/n)	0,50	0,15	0,57	0,20
Inter. (Dúvida) - Foco	0,74	0,24	0,90	0,13
Inter. (Dúvida) - Negativa	0,37	0,21	0,73	0,18
Inter. (Dúvida) - Inter. (Qu)	0,32	0,10	0,44	0,11
Inter. (Dúvida) - Inter. (s/n)	0,38	0,12	0,46	0,21
Foco - Negativa	0,30	0,20	0,67	0,15
Foco - Inter. (Qu)	0,49	0,21	0,70	0,13
Foco - Inter. (s/n)	0,51	0,22	0,77	0,21
Negativa - Inter. (Qu)	0,21	0,07	0,35	0,14
Negativa - Inter. (s/n)	0,23	0,13	0,58	0,16
Inter. (Qu) - Inter. (s/n)	0,56	0,19	0,72	0,21

Fonte: Maria Eduarda de Araújo Cardoso, 2018

As figuras 37 e 38 apresentam os erros na borda dos casos dependentes e independentes do intérprete. Analisando os gráficos é notória a porcentagem maior de erros para os dados provenientes do intérprete A para os testes dependentes, atingindo 96,88% de erros na borda para a combinação de EFGs *Foco - Interrogativa (s/n)*. Para os dados provenientes do intérprete B a combinação de EFG que obteve mais erros foi a combinação de EFGs *Interrogativa (Dúvida) - Interrogativa (s/n)* atingindo 87,85% de erros na borda.

Para o caso dos testes independentes os dados provenientes do intérprete B e testes com os dados do intérprete A apresentam maior número de erros na maioria dos casos, só não é maior nos casos das combinações de EFGs *Afirmativa - Interrogativa (qu)*, *Interrogativa (Dúvida) - Interrogativa (s/n)* e *Negativa - Interrogativa (Qu)*.

Figura 37 – Análise de erros na borda no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Figura 38 – Análise de erros na borda no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

As figuras 39 e 40 apresentam os erros gerais e de borda do classificador para dos casos dependentes e independentes do intérprete. Para os casos dependentes dos dados provenientes do intérprete A a combinação das EFGs *Interrogativa (Dúvida) - Foco* apresenta o pior caso, e nos erros gerais ela atinge 609 erros de 899 quadros resultando em um número alto de erros. O melhor caso dos dados provenientes do intérprete A é a combinação das EFGs *Interrogativa (Qu) - Interrogativa (s/n)* no qual o índices de erros são baixos. Os dados provenientes do intérprete B o melhor caso foi a combinação de EFGs *Interrogativa (Qu) - Interrogativa (s/n)* que apresenta 309 erros gerais de 922 quadros, os erros ainda são menores que a metade do número de quadros. O pior caso é a combinação das EFGs *Foco - Negativa* que para os erros de borda apresenta 123 erros de 245 sendo considerado alto esse número de erros, errando mais da metade dos quadros.

Figura 39 – Análise de erros de borda e erros gerais no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A figura 40 apresenta os erros gerais e de borda dos testes independentes do intérprete. Para os casos independentes do intérprete para os dados provenientes do intérprete A e teste com os dados do intérprete B o melhor caso foi a combinação de EFGS *Interrogativa (Dúvida)* - *Interrogativa (Qu)* e o pior caso *Negativa* - *Interrogativa (Qu)* o qual ambos casos apresentam em erros gerais altos apresentando mais de 400 erros gerais. Para os dados provenientes do intérprete B e teste com os dados do intérprete A o melhor caso foi a combinação de EFGS *Interrogativa (Dúvida)* - *Foco* e o pior caso *Negativa* - *Interrogativa (Qu)* e apresentam em ambos casos erros gerais e de borda consideravelmente baixos. Para o pior caso do intérprete A testando com o intérprete B os erros gerais apresentam 420 erros de 891, um número alto de erros comparados aos demais erros gerais apresentados pelos demais casos. O melhor caso dos dados do Intérprete B testando com o A os erros gerais apresentam 74 erros de 899 quadros obtendo um número baixo de erros.

Figura 40 – Análise de erros na borda e erros gerais no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise em testes independentes

Fonte: Maria Eduarda de Araújo Cardoso, 2018

As tabelas 16 e 17 apresentam os resultados dos testes dependentes e independentes da distância dos pontos da face com a mesma parametrização apresentada para oito pontos normalizados e a mesma configuração apresentada para a MLP. OS testes dependentes com os dados provenientes do intérprete A apresentaram a média de acurácia acima de 0,8 para todas combinações de EFGs. No caso dos testes dependentes com os dados provenientes do intérprete B as médias das acurácias não foram tão altas, no seu pior caso a combinação de EFGs *Afirmativa - Negativa* não atingiram resultados altos apresentando 0,5 de média. O melhor caso com os dados provenientes do intérprete B foi a combinação das EFGs *Foco - Interrogativa (s/n)* que atingiram 0,91. Os dados do intérprete A considerando o melhor caso foi maior com a combinação de EFGS *Foco - Interrogativa (Qu)* atingindo uma média da acurácia de 0,98.

A combinação de EFGS *Afirmativa- Negativa* mais uma vez não obteve bons resultados para ambos intérpretes, foi o pior caso para os dois intérpretes, o que afirma a complexidade dessas EFGs em seu reconhecimento.

Tabela 16 – Acurácia de classificadores no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A				
Afirm. - Int. (Dúvida)	0,88	0,01	0,89	0,87
Afirm. - Foco	0,91	0,00	0,91	0,91
Afirm. - Negativa	0,87	0,00	0,88	0,87
Afirm.- Inter. (Qu)	0,89	0,08	0,92	0,74
Afirm. - Inter. (s/n)	0,90	0,00	0,90	0,90
Inter. (Dúvida) - Foco	0,95	0,00	0,96	0,94
Inter. (Dúvida) - Negativa	0,92	0,00	0,92	0,91
Inter. (Dúvida) - Inter. (Qu)	0,93	0,01	0,94	0,91
Inter. (Dúvida) - Inter. (s/n)	0,93	0,01	0,94	0,92
Foco - Negativa	0,94	0,02	0,96	0,92
Foco - Inter. (Qu)	0,98	0,00	0,98	0,98
Foco - Inter. (s/n)	0,97	0,00	0,97	0,97
Negativa - Inter. (Qu)	0,91	0,01	0,93	0,89
Negativa - Inter. (s/n)	0,91	0,04	0,93	0,81
Inte. (Qu) - Inter. (s/n)	0,95	0,00	0,95	0,95
Intérprete B				
Afirm. - Inter. (Dúvida)	0,70	0,02	0,72	0,67
Afirm. - Foco	0,70	0,01	0,71	0,68
Afirm. - Negativa	0,50	0,03	0,56	0,49
Afirm. - Inter. (Qu)	0,72	0,03	0,77	0,70
Afirm. - Inter. (s/n)	0,78	0,01	0,79	0,76
Inter. (Dúvida) - Foco	0,88	0,00	0,88	0,87
Inter. (Dúvida) - Negativa	0,71	0,01	0,71	0,69
Inter. (Dúvida) - Inter. (Qu)	0,89	0,00	0,90	0,88
Inter. (Dúvida) - Inter. (s/n)	0,90	0,00	0,90	0,90
Foco - Negativa	0,62	0,02	0,63	0,57
Foco - Inter. (Qu)	0,86	0,02	0,89	0,85
Foco - Inter. (s/n)	0,91	0,00	0,92	0,90
Negativa - Inter. (Qu)	0,70	0,02	0,76	0,69
Negativa - Inter. (s/n)	0,76	0,00	0,77	0,76
Inter. (Qu) - Inter. (s/n)	0,88	0,05	0,92	0,73

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para os casos independentes os testes com os dados provenientes do intérprete A testando com o intérprete B nenhum resultado apresentou resultados altos, o melhor caso a combinação de EFGs *Interrogativa (Dúvida) - Foco* atingiu 0,51 de acurácia média e o pior caso atingiu 0,29 que foi a combinação de EFGs *Interrogativa (Dúvida) - Inter. (s/n)*. Os casos independentes com os dados provenientes do intérprete B testando com o

intérprete A cinco combinações de EFGs conseguiram atingir resultados altos, sendo elas: *Afirmativa - Interrogativa (Dúvida)*, *Afirmativa - Foco*, *Interrogativa (Dúvida) - Foco*, *Foco - Interrogativa (s/n)* e *Interrogativa (Qu) - Interrogativa (s/n)*, as demais combinações de EFGs não apresentaram resultados altos.

Tabela 17 – Acurácia de classificadores no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A (treino) - Intérprete B (teste)				
Afir. - Inter. (Dúvida)	0,47	0,05	0,51	0,38
Afir. - Foco	0,51	0,02	0,55	0,49
Afir. - Negativa	0,36	0,04	0,40	0,30
Afir. - Inter. (Qu)	0,38	0,07	0,52	0,28
Afir. - Inter. (s/n)	0,28	0,03	0,32	0,24
Inter. (Dúvida) - Foco	0,51	0,10	0,60	0,37
Inter. (Dúvida) - Negativa	0,43	0,01	0,44	0,42
Inter. (Dúvida) - Inter. (Qu)	0,34	0,08	0,48	0,26
Inter. (Dúvida) - Inter. (s/n)	0,29	0,04	0,32	0,23
Foco - Negativa	0,37	0,01	0,38	0,36
Foco - Inter. (Qu)	0,27	0,08	0,42	0,23
Foco - Inter. (s/n)	0,38	0,02	0,39	0,32
Negativa - Inter. (Qu)	0,36	0,02	0,40	0,34
Negativa - Inter. (s/n)	0,27	0,04	0,32	0,22
Inter. (Qu) - Inter. (s/n)	0,26	0,01	0,30	0,26
Intérprete B (treino) - Intérprete A (teste)				
Afirm. - Inter. (Dúvida)	0,63	0,01	0,67	0,62
Afirm. - Foco	0,65	0,00	0,65	0,65
Afirm. - Negativa	0,54	0,00	0,54	0,54
Afirm. - Inter. (Qu)	0,55	0,08	0,72	0,51
Afirm. - Inter. (s/n)	0,57	0,00	0,57	0,57
Inter. (Dúvida) - Foco	0,80	0,07	0,83	0,60
Inter. (Dúvida) - Negativa	0,56	0,03	0,63	0,55
Inter. (Dúvida) - Inter. (Qu)	0,40	0,10	0,59	0,34
Inter. (Dúvida) - Inter. (s/n)	0,33	0,11	0,64	0,20
Foco - Negativa	0,58	0,02	0,64	0,57
Foco - Inter. (Qu)	0,57	0,14	0,74	0,44
Foco - Inter. (s/n)	0,75	0,11	0,78	0,42
Negativa - Inter. (Qu)	0,48	0,11	0,52	0,48
Negativa - Inter. (s/n)	0,47	0,01	0,56	0,45
Inter. (Qu) - Inter. (s/n)	0,69	0,08	0,74	0,55

Fonte: Maria Eduarda de Araújo Cardoso, 2018

As figuras 41 e 42 apresentam os erros na borda para os casos dependentes e independentes do intérprete. Para os testes com os dados provenientes do intérprete A a

combinação de EFGS *Foco - Interrogativa (s/n)* atingiu 100% de erros na borda. Esses dados provenientes do intérprete A os resultados apresentaram erros altos na maioria das combinações de EFGs. Nos testes com os dados provenientes do intérprete B a maior porcentagem de erros foi 72,65% e para os testes independentes todos resultados foram semelhantes apresentando erros menor que 52%.

Figura 41 – Análise de erros na borda no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Figura 42 – Análise de erros na borda no problema de três classes com a **distância dos pontos**: duas EFGs mais a expressão neutra. Análise em testes dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

As figuras 43 e 44 apresentam os resultados dos erros gerais e erros de borda para os casos dependentes e independentes do intérprete. Os dados provenientes do intérprete A apresentou a combinação de EFGs *Afirmativa - Negativa* para o pior caso e para o melhor caso *Foco - Interrogativa (Qu)* que apresentaram um número de erros baixos para erros gerais e de borda. Logo para dados provenientes do intérprete B a combinação das EFGs *Afirmativa - Negativa* foi o pior caso e o melhor caso foi a combinação de EFGs *Foco - Interrogativa (s/n)*. O pior caso do intérprete B apresenta em erros gerais 340 erros de 775 quadros apresentando um número alto de erros comparados ao pior caso do intérprete A que obteve 84 erros de 675 quadros, considerado baixo se comparados ao pior caso do intérprete B. Para os demais casos apenas os erros de borda do intérprete B também apresentou um número alto de erros, obtendo 125 erros de 273 quadros.

Figura 43 – Análise de erros na borda e erros gerais no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para os casos independentes os testes com a combinação *Interrogativa (Dúvida)* - *Interrogativa (s/n)* foi o pior caso para ambos intérpretes. O melhor caso para ambos intérpretes foi a combinação das EFGs *Interrogativa (Dúvida)* - *Foco*.

Figura 44 – Análise de erros na borda e erros gerais no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes independentes

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para os dados provenientes do intérprete B o pior caso foi a combinação *Interrogativa (Dúvida) - Interrogativa (s/n)* que apresentou os maiores erros em relação os dados do intérprete A, que apresentou a combinação de EFGs *Interrogativa (Dúvida) - Interrogativa (s/n)* como o pior caso. O pior caso do intérprete B obteve 713 erros de 1218 quadros apresentando um número alto de erros.

Mistura de especialistas com inicialização com *k-means*

A tabela 18 apresenta os resultados do melhor caso e o pior caso treinados com a inicialização aleatória com oito pontos da face. E esses casos foram treinados com a mistura de especialistas com inicialização com *k-means*.

Tabela 18 – *F-score*, precisão, revocação e acurácia de classificadores no problema de duas classes com **oito pontos** da face: uma EFG mais a expressão neutra. Análise dependente e independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	max	min	# de borda	# erros gerais	% na borda
Intérprete A							
Inter. (Dúv.) - Foco	0,35	0,28	0,68	0,13	137 (50,18)	290 (32,26)	47,24
Inter. (Qu) - Inter. (s/n)	0,94	0,05	0,97	0,79	30 (12,24)	30 (3,33)	100,00
Intérprete B							
Afirmativa - Negativa	0,33	0,06	0,50	0,29	125 (45,79)	346 (44,65)	36,13
Inter. (dúv.) - Inter. (s/n)	0,68	0,32	0,93	0,23	71 (23,59)	81 (6,65)	87,65
Intér. A - Intér. B							
Inter.(dúv.) - Foco	0,21	0,05	0,34	0,20	122 (49,80)	416 (46,07)	29,33
Inter. (Qu) - Inter. (s/n)	0,25	0,01	0,26	0,22	133 (48,72)	654 (60,28)	20,34
Intér. B - Intér. A							
Afirmativa - Negativa	0,27	0,10	0,54	0,21	119 (48,57)	303 (44,89)	39,27
Inter. (dúv.) - Inter. (s/n)	0,28	0,12	0,45	0,19	97 (35,53)	314 (34,06)	30,89

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A combinação de EFGs *Interrogativa (Dúvida) - Foco* teve um aumento de 0,18 com a inicialização com *k-means* e a *Interrogativa (Qu) - Interrogativa (s/n)* manteve o mesmo valor para ambas inicializações, esses são os resultados apresentados pelos dados provenientes do intérprete A. Para os dados provenientes do intérprete B a combinação de EFGs *Afirmativa - Negativa* manteve o mesmo valor e a *Interrogativa (Qu) - Interrogativa (s/n)* teve uma queda de 0,25 em seu resultado, o que certifica que a inicialização aleatória para esse caso desempenha um melhor resultado. Para os casos independentes do intérprete

apenas a combinação de EFGs dos dados provenientes do Intérprete B e testado com os dados do intérprete A teve uma queda de 0,28. As demais combinações de EFGs ficaram semelhantes aos testes com a inicialização aleatória.

Os gráficos da figura 45 apresentam uma visão geral do desempenho de cada rede *Multilayer Perceptron* e mistura de especialistas com inicialização aleatória e com a mistura de especialistas com inicialização *K-means*. Cada rede apresenta seu pior caso e melhor caso, para casos dependentes e independentes do intérprete. Analisando os gráficos a rede MLP tem um desempenho melhor para o pior caso de ambos os testes. Para o melhor caso a combinação de EFGs *Interrogativa (Qu) - Interrogativa (s/n)* os resultados das redes respondem iguais para os testes dependentes do intérprete. Logo, para este caso de experimentos a MLP obteve resultados uniformes e em sua maioria melhores.

Figura 45 – A média das acurácias dos piores e melhores casos das redes MLP e ME, casos dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Resultados com o problema de classificação multiclasse: Experimento #2

Multilayer Perceptron

Este experimento considerou seis EFGs e a expressão neutra. Foi utilizada a mesma estratégia do experimento #1, apenas alterando o número de neurônios na camada escondida para 33 neurônios. A tabela 19 apresenta os resultados, e como pode ser observado, o problema de seis classes não representa um desafio complexo para o caso dos testes dependentes do usuários. Considerando a já conhecida dificuldade de análise das EFGs do intérprete B, os resultados obtidos são promissores. Além disso, os classificadores são robustos, apresentado pouca variabilidade de desempenho dentro das execuções realizadas.

Analisando os erros de borda, o potencial dos classificadores para segmentação das seis EFGs se confirma, pois a maioria dos erros cometidos, para o caso dos dois intérpretes, ocorre na borda (um erro de menor importância). Por outro lado, a capacidade de generalização dos modelos nos testes independentes do intérprete são insatisfatórias, com acurácias médias baixas.

Tabela 19 – Acurácia de classificadores no problema de sete classes com **oito pontos** da face e erros de borda e erros gerais: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ).

	μ	σ	max	min	# erros de borda	# erros gerais	% erros na borda
A	0,86	0,03	0,89	0,81	195 (25,10)	231 (9,31)	84,42
B	0,70	0,02	0,71	0,64	317 (39,38)	665 (24,03)	47,67
A (treino) - B (teste)	0,51	0,01	0,52	0,48	370 (45,96)	937 (33,86)	39,49
B (treino) - A (teste)	0,49	0,04	0,55	0,44	301 (38,74)	584 (23,55)	51,54

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Em questão de erros na borda o intérprete B apresenta menos erros em comparação ao intérprete A, uma diferença de 36,75%. Os erros gerais e erros de borda do intérprete B apresentam um número maior em relação ao intérprete A, chegando a 665 erros gerais enquanto o intérprete A obteve 231 erros quase a metade de erros do intérprete B. Para o caso dos testes independentes do intérprete ambos apresentaram um número muito alto de erros de borda e erros gerais, atingindo 937 erros para com os dados do intérprete A e teste com os dados do intérprete B.

A tabela 20 apresenta os resultados com as distância dos pontos com a mesma parametrização apresentada para os testes com oito pontos normalizados.

Tabela 20 – Acurácia de classificadores no problema de sete classes com **distância dos pontos** da face e erros de borda e erros gerais : seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ).

	μ	σ	max	min	# erros de borda	# erros gerais	% erros na borda
A	0,91	0,01	0,92	0,87	180 (23,17)	232 (9,35)	77,59
B	0,75	0,01	0,76	0,73	310 (38,51)	595 (21,50)	52,10
A (treino) - B (teste)	0,41	0,06	0,50	0,29	389 (48,32)	1268 (45,83)	30,68
B (treino) - A (teste)	0,52	0,07	0,59	0,37	304 (39,12)	681 (27,46)	44,64

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os resultados com a distância dos pontos apresentam uma melhora nos testes dependentes, atingindo uma acurácia de 0,91 para os dados provenientes do intérprete A

e 0,75 para os dados provenientes do intérprete B, inferindo que para o caso com todas as EFGs em uma sentença o classificador trabalhou melhor com a distância dos pontos. Para os testes independentes os resultados continuam apresentando resultados baixos.

Mistura de especialistas com inicialização aleatória

A tabela 21 apresenta os resultados com todas as EFGs treinadas pela mistura de especialistas com oito pontos normalizados. A *Multilayer Perceptron* para os casos dependentes obteve melhores resultados que a ME, apresentando números acima de 0,8 e também apresenta menos erros na borda. Para os casos independentes do intérprete ambos os casos não alcançaram resultados altos o que também aconteceu para a parametrização dos oito pontos. Podendo afirmar que a MLP ainda obteve resultados maiores que a ME.

Tabela 21 – Acurácia de classificadores no problema de sete classes com **oito pontos** da face e erros de borda e erros gerais: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ).

	μ	σ	max	min	# erros de borda	# erros gerais	% erros na borda
A	0,65	0,40	0,92	0,07	161 (20,72)	193 (7,78)	83,42
B	0,71	0,03	0,75	0,64	375 (46,58)	950 (34,33)	39,47
A (treino) - B (teste)	0,18	0,10	0,37	0,09	304 (37,76)	654 (23,64)	46,48
B (treino) - A (teste)	0,39	0,17	0,52	0,06	291 (37,45)	659 (26,57)	44,16

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os erros gerais dos dados provenientes do intérprete B alcançam 950 erros enquanto que para os dados provenientes do intérprete A alcançam 193 uma proporção de erros bem menor de erros para o intérprete A. Em termos de erros na borda os dados provenientes do intérprete A alcançam uma porcentagem maior de 83,42% se comparados aos dados provenientes do intérprete B que alcançam 39,47%. A tabela 22 apresenta as acurácias alcançadas pelo classificador com a distâncias dos pontos e os erros de borda.

Tabela 22 – Acurácia de classificadores no problema de sete classes com a **distância dos pontos** da face e erros de borda e erros gerais: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ).

	μ	σ	max	min	# erros de borda	# erros gerais	% erros na borda
A	0,81	0,03	0,87	0,78	215 (27,67)	284 (11,45)	75,70
B	0,69	0,04	0,75	0,61	311 (38,63)	687 (24,83)	45,27
A (treino) - B (teste)	0,21	0,10	0,45	0,13	417 (51,80)	1348 (48,72)	30,93
B (treino) - A (teste)	0,58	0,01	0,60	0,56	360 (46,33)	814 (32,82)	44,23

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Esses resultados apresentaram média de acurácia maiores para está configuração comparados com os resultados treinados com os oito pontos. Para este caso a MLP também desempenhou resultados melhores que a ME. Os testes independentes dos dados provenientes do intérprete A testando com os dados provenientes do intérprete B atingiu 1348 erros enquanto para o caso do intérprete B testando com os dados provenientes do intérprete A o número de erros foi 814, ambos testes apresentam um número alto de erros gerais, e para o caso dos testes dependentes esses números foram baixos.

Mistura de especialistas com inicialização com *K-means*

A tabela 23 apresenta os resultados com todas EFGs treinados com a inicialização com *K-means*. Os resultados foram menores comparados a ME com inicialização aleatória e também com a *Multilayer Perceptron*. Para os testes independentes, ambos obtiveram 0,18 de média da acurácia, e envolvem os dados provenientes do intérprete B que alcançam erros gerais altos atingindo 100% para o caso independente do intérprete.

Tabela 23 – Acurácia de classificadores no problema de sete classes com **oito pontos** da face e erros de borda e erros gerais: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ).

	μ	σ	max	min	# erros de borda	# erros gerais	% erros na borda
A	0,46	0,15	0,52	0,03	182 (24,10)	190 (5,01)	95,78
B	0,66	0,07	0,70	0,50	709 (88,07)	2750 (99,38)	25,78
A (treino) - B (teste)	0,18	0,08	0,30	0,06	377 (47,00)	811 (29,30)	46,48
B (treino) - A (teste)	0,18	0,17	0,56	0,04	777 (100,00)	2480 (100,00)	31,33

Para os dados provenientes do intérprete A os erros na borda atingem uma porcentagem alta de 95,78%, enquanto que para o intérprete B a porcentagem é de 25,78%, uma taxa pequena comparada a porcentagem que intérprete A obteve.

A figura 46 apresenta a média das acurácias com o desempenho de cada rede treinada com os oito pontos normalizados. Os gráficos ilustram que a MLP obteve melhores resultados na maioria dos casos e a mistura de especialistas com inicialização em *K-means* foi a que obteve o pior desempenho.

Figura 46 – A média das acurácias dos piores e melhores casos das redes MLP e ME, casos dependentes e independentes do intérprete

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Para os dados provenientes do intérprete A no caso dos testes dependentes e testes independentes com os dados do intérprete B e teste com os dados do intérprete A, eles obtiveram o mesmo comportamento a MLP alcançando os melhores resultados, a mistura com inicialização aleatória com resultados um pouco abaixo, e a mistura com inicialização com *K-means* com os piores resultados. Já os testes com os dados provenientes do intérprete B os resultados das três redes foram semelhantes e para os testes independentes com os dados do intérprete A e teste com os dados do intérprete B, a MLP foi o melhor resultado.

Resultados dos experimentos com os nove pontos estão disponíveis no Apêndice A. Também está disponível as tabelas com os erros de borda, erros gerais e erros na borda de todas as experimentações contidas neste trabalho.

Sumarizando os resultados obtidos pelos experimentos descritos neste estudo, é possível afirmar que: o uso de representações vetoriais baseadas em janelas, i.e., com representação vetorial para a informação relacionada ao tempo, traz resultados melhores para o problema de segmentação de EFGs; as EFGs de maior dificuldade para segmentação são aquelas que envolvem movimentos acentuados da cabeça (EFGs afirmativa e negativa) e,

aquelas que exigem movimentos mais acentuados de sobrancelha e boca geram problemas de segmentação mais fáceis para os classificadores; a generalização do problema de segmentação de EFGs para diferentes intérpretes tem alta complexidade. Esse tipo de generalização, que envolve o comportamento de diferentes indivíduos vêm sendo estudado na área de análise de gestos também em outros contextos (MADEO; PERES; LIMA, 2016; RAMAKRISHNAN, 2011) e conclusões similares têm sido obtidas.

5 Conclusão

Este estudo apresentou resultados referentes à segmentação de expressões faciais gramaticais utilizadas na Libras, empregando rede neural artificial MLP e mistura de especialistas como os classificadores responsáveis pela segmentação. Os experimentos foram realizados tanto em relação à definição do problema como um problema de classificação binária, quanto como um problema de classificação multiclasse. Para o caso de estudo com múltiplas classes, seis tipos de EFGs foram consideradas em duas abordagens: classificação com três classes, classificação com sete classes (incluindo a expressão neutra como classe). Os experimentos foram conduzidos principalmente na sua faceta de representação temporal, já que na faceta atemporal, experimentos preliminares indicaram um desempenho mais baixo de classificadores já para o caso binário.

Para que esse estudo pudesse ser concluído foram realizados estudos na área da língua de sinais, estudo das expressões faciais gramaticais, estudos das técnicas de aprendizado de máquina (rede neural artificial *Multilayer Perceptron* e abordagem de mistura de especialistas), como também a implementação dessas técnicas para a execução dos experimentos. Para as experimentações foram criados vários tipos de representações vetoriais com os cem pontos, oito pontos e nove pontos da face, com as coordenadas (x, y) e também com as distâncias entre os pontos. A distância entre pontos se destaca por atingir, em alguns casos, os melhores resultados.

Para os casos de testes dependentes do intérpretes ambas as técnicas de aprendizado de máquina utilizadas apresentaram resultados positivos, porém, para os testes independentes de intérprete, os resultados foram insatisfatórios. Esse fato demonstra a dificuldade de generalização inerente ao problema sob estudo. Especialmente no contexto do conjunto de dados usados nos experimentos deste trabalho, é conhecido que o intérprete B possui mais fluência na língua de sinais do que o intérprete A. Sendo assim, a sua gesticulação é mais natural e menos marcada, o que dificulta a discriminação entre classes para os classificadores.

Os resultados obtidos mostram que MLPs tem potencial para uso na segmentação de EFGs no contexto dependente do intérprete. Embora, ainda neste contexto, alguns resultados não alcançaram acurácias médias altas, uma exploração mais aprofundada desta

técnica, variando métodos de otimização e outras estratégias de controle de parâmetros, pode melhorar os resultados.

A mistura de especialistas foi usada em sua estrutura mais simples, com funções lineares, o que pode ser um limitador para o alcance de resultados satisfatórios. Especificamente para o caso da mistura de especialistas, foram testadas duas formas de inicialização de pesos para os especialistas: a aleatória e a baseada no algoritmo *k-means*. A alteração da inicialização, da aleatória para a baseada em *k-means*, já representou uma evolução para técnica, pois em alguns casos alcançou melhores resultados (pior caso dos testes binários). As EFGs *Afirmativa* e *Negativa*, para ambas as técnicas, representaram dificuldades, e a movimentação da cabeça que ambas as EFGs apresentam pode ser o fator que determinou os resultados mais baixos.

Com essas experimentações foi possível analisar que a *Multilayer Perceptron* trabalha bem tanto com a representação que usa pontos quanto com a representação que usa a distância entre pontos. Já a mistura de especialistas obtém melhores resultados com as distâncias dos pontos. O ambiente de exploração proporcionado pelo conjunto de dados limita a implementação de testes de classificadores, pois trata-se de um conjunto de dados com poucas instâncias para uso de estratégias de teste mais robustas – como validação cruzada; e aplicação de técnicas mais poderosas – como aprendizado profundo, *deep learning*.

5.1 Contribuições

Este trabalho apresenta como principal contribuição científica um estudo que mostrou a potencialidade de uso de duas técnicas de aprendizado de máquina aplicada no problema de segmentação de expressões faciais gramaticais da Libras, considerando a definição do problema num ambiente de múltiplas classes.

Além disso, este trabalho se estabelece como um estudo interdisciplinar, que levou em consideração a análise das EFGs sob um ponto de vista com embasamento linguístico, especialmente desenvolvimento para a Libras. Considerando o que foi possível levantar na literatura da área, esse é o primeiro trabalho que explorou o problema de segmentação de EFGs da Libras sob a perspectiva de classificadores multiclasse.

Como uma contribuição técnica destaca-se a extensão do conjunto de dados *Grammatical Facial Expressions Data Set* para incorporar rotulação multiclasse.

5.2 Limitações

O trabalho aqui desenvolvido possui algumas limitações. A primeira diz respeito ao próprio conjunto de dados usado para a experimentação. Embora seja o único conjunto de dados rotulado e disponível publicamente encontrado na pesquisa realizada, que permita o desenvolvimento de modelos de classificação para o problema de análise das expressões faciais gramaticais, é um conjunto de dados que possui poucas instâncias de dados, é limitado à gesticulação de sentenças isoladas, e considera apenas a gesticulação de dois intérpretes. Considerando as condições consideradas ideais para desenvolvimento de modelos de aprendizado indutivo, principalmente no que diz respeito à maximização do poder de generalização, seria necessário o uso de um conjunto de dados maior e mais diversificado. E ainda, considerando o real contexto de gesticulação de uma língua de sinais, o uso de sentenças isoladas também limita a plena representação das possibilidades de movimentos faciais usados em uma conversação natural.

Seguindo esta linha, a segunda limitação é apresentada. O problema aqui definido está limitado à análise de EFGs, mas é sabido que, durante a conversação, as pessoas se expressam também de forma afetiva, fazendo uso de expressões faciais afetivas. Desta forma, alcançar um modelo de classificação que de fato pudesse ser aplicado para aplicações reais implica em também implementar a capacidade de diferenciar EFGs e expressões faciais afetivas, ou compor os dois tipos de expressões e modelar a classificação da EFGs dentro desse contexto híbrido.

5.3 Trabalhos Futuros

Ao explorar o problema de reconhecimento de EFGs, o presente trabalho abre espaço para estudos mais específicos:

- considerando o atual conjunto de dados, realizar experimentações nas quais a fase de treinamento envolva EFGs executadas por ambos os intérpretes. Desta forma, o poder de generalização dos algoritmos pode ser testado com maior propriedade,

uma ideia seria juntar o conjunto de dados dos intérpretes A e B para outros testes independentes.

- construir um conjunto de dados, como extensão ao conjunto *Grammatical Facial Expressions Data Set*, ou complementar a ele, que possibilite estender o estudo aqui realizado. Algumas características são importantes nesta construção: aumentar o número de instância; aumentar o número de intérpretes; construir a gesticulação dentro de um trecho de discurso, não se limitando a sentenças isoladas, e construindo trechos de discurso que envolvam várias EFGs executadas em sequência e que envolvam também o uso de expressões faciais afetivas. A partir de um conjunto de dados estendido, será possível melhorar a variabilidade dos experimentos de forma a delinear conclusões mais específicas sobre as dificuldades inerentes ao problema de segmentação de EFGs.
- estabelecer, construir e submeter a testes, um esquema de representação dos movimentos da face ocorridos no decorrer dos quadros de vídeos. Na representação usada neste trabalho, o movimento da cabeça e dos elementos da face está representando pela variação nas coordenadas dos pontos e nas distâncias entre os pontos expressa em termos de janelas de quadros. Uma alternativa é calcular a variação que ocorre nas coordenadas e nas distâncias considerando dois quadros de vídeos distintos, porém ordenados no tempo.
- investigar a técnica mistura de especialistas implementada de outras maneiras. Uma opção seria implementar componentes da *Multilayer Perceptron* na ME. Uma das implementações mais aplicadas para a ME é a mistura de especialistas em MLP (MME) (MASOUDNIA; EBRAHIMPOUR, 2014). Na versão do MME, o MLP é usado para especialistas e redes de portas, em vez de redes lineares para melhorar o desempenho em relação a uma ME convencional.

Referências¹

- AGRIS, U. V.; KNORR, M.; KRAISS, K.-F. The significance of facial features for automatic sign language recognition. In: IEEE. *Proceedings of the 8th IEEE International Conference on Automatic Face Gesture Recognition. Amsterdam, The Netherlands: IEEE*, [S.l.], 2008. p. 1–6. Citado na página 29.
- AMARAL, W. M. do et al. Sistema de transição da língua brasileira de sinais voltado à produção de conteúdo sinalizado por avatares 3D. Campinas, SP, 2012. Citado na página 33.
- ARROTÉIA, J. *O papel da marcação não-manual nas sentenças negativas em Língua de Sinais Brasileira (LSB)*. Dissertação (Mestrado), 2005. Citado na página 36.
- BENITEZ-QUIROZ, C. F. et al. Discriminant features and temporal structure of nonmanuals in american sign language. *PloS One*, v. 9, n. 2, p. e86268, 2014. Citado 6 vezes nas páginas 39, 40, 42, 44, 45 e 47.
- BENITEZ-QUIROZ, C. F.; WILBUR, R. B.; MARTINEZ, A. M. The not face: A grammaticalization of facial expressions of emotion. *Cognition*, Elsevier, v. 150, p. 77–84, 2016. Citado 10 vezes nas páginas 24, 28, 39, 40, 42, 44, 45, 47, 48 e 54.
- BHUVAN, M. et al. Detection and analysis model for grammatical facial expressions in sign language. In: IEEE. *Region 10 Symposium – “Technologies for Smart Cities”*. [S.l.], 2016. p. 155–160. Citado 8 vezes nas páginas 28, 39, 40, 42, 44, 45, 46 e 54.
- BISHOP, C. M. *Neural networks for pattern recognition*. [S.l.]: Oxford University, 1995. Citado na página 61.
- CARIDAKIS, G.; ASTERIADIS, S.; KARPOUZIS, K. Non-manual cues in automatic sign language recognition. *Personal and ubiquitous computing*, Springer, v. 18, n. 1, p. 37–46, 2014. Citado 6 vezes nas páginas 39, 41, 42, 45, 51 e 54.
- CASCIA, M. L.; SCLAROFF, S.; ATHITSOS, V. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 4, p. 322–336, 2000. Citado na página 45.
- CASTRO, M. C. F. D. et al. A complex valued hebbian learning algorithm. In: *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*. [S.l.: s.n.], 1998. v. 2, p. 1235–1238. Citado na página 82.
- CHANG, C.-Y.; HUANG, Y.-C. Personalized facial expression recognition in indoor environments. In: IEEE. *International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2010. p. 1–8. Citado 2 vezes nas páginas 22 e 76.
- COOTES, T. F. et al. Active shape models-their training and application. *IEEE Computer Vision and Image Understanding*, Elsevier, v. 61, n. 1, p. 38–59, 1995. Citado 2 vezes nas páginas 43 e 52.

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

COSTA-VIEIRA, H. A.; SOUZA, W. C. de. O reconhecimento de expressões faciais e prosódia emocional: Investigação preliminar em uma amostra brasileira jovem. *Estudos de Psicologia*, Universidade Federal do Rio Grande do Norte, v. 19, n. 2, 2014. Citado na página 22.

DAHMANE, M.; MEUNIER, J. Sift-flow registration for facial expression analysis using gabor wavelets. In: IEEE. *11th Int. Conf. on Inf. Sci. Signal Processing and their Applications*. [S.l.], 2012. p. 175–180. Citado na página 76.

DEAF. *Sign language - World Federation of the Deaf*. 2015. Citado na página 27.

DING, L.; MARTINEZ, A. M. Features versus context: An approach for precise and detailed detection and delineation of faces and facial features. *Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 32, n. 11, p. 2022–2038, 2010. Citado na página 43.

DU, S.; TAO, Y.; MARTINEZ, A. M. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, National Academy Sciences, p. E1454–E1462, 2014. Citado na página 48.

EKMAN, P.; FRIESEN, W. V. *Unmasking the face: A guide to recognizing emotions from facial clues*. [S.l.]: Los Altos, California: Malor Books, 2003. 212 p. Citado na página 22.

EKMAN, P.; SORENSON, E. R.; FRIESEN, W. V. Pan-cultural elements in facial displays of emotion. *Science*, American Association for the Advancement of Science, v. 164, n. 3875, p. 86–88, 1969. Citado na página 48.

FAUSETT, L. V. *Fundamentals of neural networks: architectures, algorithms, and applications*. [S.l.]: Prentice-Hall Englewood Cliffs, 1994. v. 3. Citado 2 vezes nas páginas 56 e 58.

FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006. Citado na página 30.

FERREIRA-BRITO, L. Uma abordagem fonológica dos sinais da lscb. *Informativo Técnico-Científico do INES, Rio de Janeiro*, v. 1, n. 1, p. 20–43, 1990. Citado 3 vezes nas páginas 33, 34 e 35.

FERREIRA-BRITO, L. *Por uma gramática de línguas de sinais*. [S.l.]: Rio de Janeiro, RJ: Tempo Brasileiro, 1995. Citado 3 vezes nas páginas 24, 34 e 37.

FREITAS, F. et al. Grammatical facial expressions recognition with machine learning. In: AAAI. *27th Florida Artificial Intelligence Research Society Conference (FLAIRS)*, Palo Alto: The AAAI Press, [S.l.], 2014. p. 180–185. Citado 13 vezes nas páginas 26, 27, 30, 39, 40, 42, 44, 45, 50, 53, 54, 71 e 72.

FREITAS, F. A. et al. Grammatical facial expression recognition in sign language discourse: a study at the syntax level. *Information Systems Frontiers*, Springer, p. 1–17, 2017. Citado 9 vezes nas páginas 28, 39, 40, 42, 44, 45, 53, 54 e 72.

FREITAS, F. d. A. *Reconhecimento automático de expressões faciais gramaticais na língua brasileira de sinais*. Dissertação (Mestrado) — Universidade de São Paulo, 2015. Citado 9 vezes nas páginas 12, 26, 27, 30, 35, 69, 72, 76 e 77.

FRIESEN, E.; EKMAN, P. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto*, Consulting Psychologists, 1978. Citado na página 48.

GREGIO, F. N. et al. Configuração do trato vocal supraglótico na produção das vogais do português brasileiro: dados de imagens de ressonância magnética. *Mestrado em Linguística Aplicada e Estudos da Linguagem PUC/SP, São Paulo*, 2006. Citado na página 62.

HAMPSHIRE, J. B.; WAIBEL, A. The meta-pi network: Building distributed knowledge representations for robust multisource pattern recognition. *Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 14, n. 7, p. 751–769, 1992. Citado na página 62.

HAYKIN, S. *Neural networks and learning machines*. [S.l.]: Pearson Upper Saddle River, 2009. v. 3. Citado 4 vezes nas páginas 26, 56, 58 e 66.

HEAD, J. D.; ZERNER, M. C. A broyden—fletcher—goldfarb—shanno optimization procedure for molecular geometries. *Chemical physics letters*, Elsevier, v. 122, n. 3, p. 264–270, 1985. Citado na página 46.

HEBB, D. O. *The organization of behavior: A neuropsychological approach*. [S.l.]: John Wiley & Sons, 1949. Citado na página 57.

HERMAN, R.; HOLMES, S.; WOLL, B. *Assessing BSL development: Receptive skills test*. [S.l.]: Forest Books, Coleford., 1999. Citado na página 44.

JACK, R. E.; GARROD, O. G.; SCHYNS, P. G. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, Elsevier, v. 24, n. 2, p. 187–192, 2014. Citado na página 23.

JACOBS, R. A. et al. Adaptive mixtures of local experts. *Neural Computation*, v. 3, n. 1, p. 79–87, 1991. Citado 4 vezes nas páginas 26, 62, 63 e 65.

JORDAN, M. I.; JACOBS, R. A. Hierarchical mixtures of experts and the em algorithm. *Neural computation*, The MIT Press, v. 6, n. 2, p. 181–214, 1994. Citado na página 26.

KACORRI, H. Models of linguistic facial expressions for american sign language animation. *SIGACCESS Accessibility and Computing*, ACM, n. 105, p. 19–23, 2013. Citado 2 vezes nas páginas 23 e 39.

KACORRI, H. *Data-Driven Synthesis and Evaluation of Syntactic Facial Expressions in American Sign Language Animation*. Tese (Doutorado) — The City University of New York, 2016. Citado na página 40.

KACORRI, H.; HUENERFAUTH, M. Implementation and evaluation of animation controls sufficient for conveying asl facial expressions. In: ACM. *Proceedings of the 16th international SIGACCESS conference on Computers & accessibility*. [S.l.], 2014. p. 261–262. Citado 2 vezes nas páginas 23 e 39.

KACORRI, H.; HUENERFAUTH, M. Continuous profile models in asl syntactic facial expression synthesis. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. [S.l.: s.n.], 2016. v. 1, p. 2084–2093. Citado 6 vezes nas páginas 23, 39, 40, 43, 44 e 48.

KACORRI, H. et al. Centroid-based exemplar selection of asl non-manual expressions using multidimensional dynamic time warping and mpeg4 features. In: *Proceedings of the Workshop on the Representation and Processing of Sign Languages: Corpus Mining, Language Resources and Evaluation Conference 2016*. [S.l.: s.n.], 2016. Citado na página 44.

KARIMU, R. Y.; AZADI, S. Diagnosing the adhd using a mixture of expert fuzzy models. *International Journal of Fuzzy Systems*, Springer, v. 20, n. 4, p. 1282–1296, 2018. Citado na página 62.

KAUSAR, S.; JAVED, M. Y. A survey on sign language recognition. In: IEEE. *Frontiers of Information Technology*. [S.l.], 2011. p. 95–98. Citado na página 25.

KOHAVI, R. et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International Joint Conference on Artificial Intelligence*. [S.l.: s.n.], 1995. v. 14, n. 2, p. 1137–1145. Citado na página 84.

KUMAR, P.; ROY, P. P.; DOGRA, D. P. Independent bayesian classifier combination based sign language recognition using facial expression. *Information Sciences*, Elsevier, v. 428, p. 30–48, 2018. Citado 7 vezes nas páginas 24, 39, 41, 43, 44, 53 e 54.

LAFFERTY, J.; MCCALLUM, A.; PEREIRA, F. C. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001. Citado na página 53.

LICHMAN, M. *UCI Machine Learning Repository*. 2013. Disponível em: <http://archive.ics.uci.edu/ml>. Citado na página 71.

LIMA, C. A. d. M. *Comitê de Máquinas: uma abordagem unificada empregando máquinas de vetores-suporte*. Tese (Doutorado) — Universidade Estadual de Campinas, 2004. Citado 2 vezes nas páginas 65 e 66.

LIMA, N. H. C. *Classificação de padrões através de um comitê de máquinas aprimorado por aprendizagem por reforço*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2012. Citado na página 62.

LISTGARTEN, J. et al. Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms. *Clinical Cancer Research*, AACR, v. 10, n. 8, p. 2725–2737, 2004. Citado na página 49.

LIU, J. et al. Recognizing eyebrow and periodic head gestures using crfs for non-manual grammatical marker detection in asl. In: IEEE. *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*. [S.l.], 2013. p. 1–6. Citado 6 vezes nas páginas 39, 40, 41, 45, 52 e 54.

LIU, J. et al. Non-manual grammatical marker recognition based on multi-scale, spatio-temporal analysis of head pose and facial expressions. *Image and Vision Computing*, Elsevier, v. 32, n. 10, p. 671–681, 2014. Citado 7 vezes nas páginas 28, 39, 40, 43, 45, 52 e 54.

LUCAS, B. D.; KANADE, T. et al. An iterative image registration technique with an application to stereo vision. *Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*, Vancouver, BC, Canada, 1981. Citado na página 51.

LYONS, M. J. et al. The japanese female facial expression (JAFPE) database. *Proceedings of third international conference on automatic face and gesture recognition (pp. 14-16)*., 1998. Citado na página 23.

MADEO, R. C. et al. A committee machine implementing the pattern recognition module for fingerspelling applications. In: ACM. *Proceedings of the 2010 Symposium on Applied Computing. New York, NY, USA: ACM, 2010*. [S.l.], 2010. p. 954–958. Citado na página 25.

MADEO, R. C. B. *Máquinas de Vetores Suporte e a Análise de Gestos: incorporando aspectos temporais*. Tese (Doutorado) — Universidade de São Paulo, 2013. Citado 2 vezes nas páginas 30 e 80.

MADEO, R. C. B.; PERES, S. M.; LIMA, C. A. M. Gesture phase segmentation using Support Vector Machines. *Expert Systems with Applications*, Elsevier, v. 56, p. 100 – 115, 2016. Citado na página 124.

MARSHALL, C. et al. Sentence repetition in deaf children with specific language impairment in british sign language. *Language Learning and Development*, Taylor & Francis, v. 11, n. 3, p. 237–251, 2015. Citado 9 vezes nas páginas 28, 29, 39, 41, 42, 44, 45, 50 e 54.

MASOUDNIA, S.; EBRAHIMPOUR, R. Mixture of experts: a literature survey. *Artificial Intelligence Review*, Springer, v. 42, n. 2, p. 275–293, 2014. Citado na página 128.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 57.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas Inteligentes-Fundamentos e Aplicações*, v. 1, n. 1, 2003. Citado na página 31.

NEIDLE, C. J. *The syntax of American Sign Language: Functional categories and hierarchical structure*. [S.l.]: The MIT press, 2000. Citado na página 45.

NEIVA, D. H.; ZANCHETTIN, C. Gesture recognition: A review focusing on sign language in a mobile context. *Expert Systems with Applications*, v. 103, p. 159 – 183, 2018. Citado na página 25.

NEIVA, D. H.; ZANCHETTIN, C. Gesture recognition: A review focusing on sign language in a mobile context. *Expert Systems with Applications*, Elsevier, 2018. Citado na página 25.

NGUYEN, H. D.; CHAMROUKHI, F. Practical and theoretical aspects of mixture-of-experts modeling: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Wiley Online Library, p. e1246, 2018. Citado na página 62.

NGUYEN, T. D.; RANGANATH, S. Facial expressions in american sign language: Tracking and recognition. *Pattern Recognition*, Elsevier, v. 45, n. 5, p. 1877–1891, 2012. Citado 3 vezes nas páginas 28, 43 e 76.

NOWLAN, S. J. Maximum likelihood competitive learning. In: *Advances in neural information processing systems*. [S.l.: s.n.], 1990. p. 574–582. Citado na página 62.

- ONG, S. C.; RANGANATH, S. Automatic sign language analysis: A survey and the future beyond lexical meaning. *Transactions on Pattern Analysis & Machine Intelligence*, IEEE, n. 6, p. 873–891, 2005. Citado na página 25.
- PETERSON, G. E.; BARNEY, H. L. Control methods used in a study of the vowels. *The Journal of the acoustical society of America*, ASA, v. 24, n. 2, p. 175–184, 1952. Citado na página 62.
- PRETTE, A. D.; PRETTE, Z. A. D. *Psicologia das habilidades sociais na infância: teoria e prática*. [S.l.]: Editora Vozes Limitada, 2017. Citado na página 22.
- QUADROS, R. d.; KARNOPP, L. B. Língua de sinais brasileira: estudos lingüísticos. *Porto Alegre: Artmed*, v. 1, p. 222, 2004. Citado 2 vezes nas páginas 34 e 35.
- QUADROS, R. M. de; KARNOPP, L. B. *Língua de sinais brasileira: estudos lingüísticos*. [S.l.]: Artmed Editora, 2009. Citado na página 33.
- RAMAKRISHNAN, A. S. *Segmentation of Hand Gestures using Motion Capture Data*. 67 p. Dissertação (Mestrado) — University of California, 2011. Citado na página 124.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958. Citado 2 vezes nas páginas 26 e 57.
- SANDLER, W. Prosody and syntax in sign languages. *Transactions of the Philological Society*, Wiley Online Library, v. 108, n. 3, p. 298–328, 2010. Citado na página 33.
- SILVA, L. A. da; PERES, S. M.; BOSCARIOLI, C. *Introdução à mineração de dados: com aplicações em R*. [S.l.]: Elsevier Brasil, 2017. Citado na página 60.
- SOUZA, R. G.; PERES, S. M.; LIMA, C. A. M. Adaptação do jogo da forca para língua de sinais usando luvas coloridas e redes neurais multilayer perceptron. In: *XIV Simpósio Brasileiro de Jogos e Entretenimento Digital - SBGames*. [S.l.: s.n.], 2015. p. 283–286. Citado na página 25.
- STOKOE, W. C. Studies in linguistics: Occasional papers 8. In: LINSTOCK PRESS. *Sign Language Structure: An Outline of the Visual Communication System of the American Deaf*. [S.l.], 1960. Citado na página 24.
- TAFNER, M. A. Redes neurais artificiais: aprendizado e plasticidade. *Cérebro Mente*, São Paulo, v. 5, 1998. Citado 2 vezes nas páginas 59 e 60.
- THEODORIDIS, S.; KOUTROUMBAS, K. et al. Pattern recognition. *Transactions on Neural Networks*, IEEE Institute of Electrical and Eletronics, v. 19, n. 2, p. 376, 2008. Citado na página 61.
- TRACY, J. L.; RANDLES, D. Four models of basic emotions: a review of ekman and cordaro, izard, levenson, and panksepp and watt. *Emotion Review*, Sage Publications Sage UK: London, England, v. 3, n. 4, p. 397–405, 2011. Citado na página 22.
- UCHIBORI, A.; MATSUOKA, K. Split movement of wh-elements in japanese sign language: A preliminary study. *Lingua*, Elsevier, v. 183, p. 107–125, 2016. Citado na página 23.

UDDIN, M. T. An ada-random forests based grammatical facial expressions recognition approach. In: IEEE. *Informatics, Electronics & Vision (ICIEV), 2015 International Conference on*. [S.l.], 2015. p. 1–6. Citado 9 vezes nas páginas 24, 28, 39, 41, 42, 44, 45, 49 e 54.

VERMA, A.; VIG, L. Acf based feature extraction and mixture of expert cnns for pedestrian detection. *Artificial Intelligent Systems and Machine Learning*, v. 10, n. 4, p. 90–97, 2018. Citado na página 62.

VILLANUEVA, W. *Comitê de Máquinas em Predição de Séries Temporais*. Tese (Doutorado) — Dissertação de Mestrado, Unicamp, 2006. Citado na página 61.

VILLANUEVA, W. J. P. et al. *Síntese automática de redes neurais artificiais com conexões à frente arbitrárias*. [S.l.]: [Federal University of Campinas], 2011. Citado na página 62.

WAGNER, P. K. et al. Segmentação de unidades gestuais com multilayer perceptrons. 2013. Citado na página 64.

WANG, H. et al. Emotion detection via discriminative kernel method. In: ACM. *Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments*. [S.l.], 2010. p. 7. Citado na página 76.

WHISSELL, C. et al. A dictionary of affect in language: Iv. reliability, validity, and applications. *Perceptual and Motor Skills*, SAGE Publications, v. 62, n. 3, p. 875–888, 1986. Citado na página 23.

XAVIER, A. N.; BARBOSA, F. V. Variabilidade e estabilidade na produção de sinais da libras. *Domínios de Linguagem*, v. 11, n. 3, p. 983–1006, 2017. Citado 2 vezes nas páginas 23 e 34.

YU, Y.-C.; YOU, S. D.; TSAI, D.-R. Magic mirror table for social-emotion alleviation in the smart home. *IEEE Transactions on Consumer Electronics*, IEEE, v. 58, n. 1, 2012. Citado na página 76.

YUAN, C.; NEUBAUER, C. Variational mixture of gaussian process experts. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2009. p. 1897–1904. Citado na página 67.

Apêndice A – Resultados

Nesse apêndice serão apresentados os resultados obtidos com o uso de nove pontos da face (adição do ponto do nariz) para representação das EFGs, e também as tabelas de erros de borda e erros gerais. Os resultados com os nove pontos são inferiores ou iguais aos resultados obtidos com oito pontos, por isso é afirmado neste trabalho que os pontos da boca e da sobrancelha são suficientes para suportar o reconhecimento da expressão facial gramatical.

Multilayer Perceptron - Experimento #1

Os resultados alcançados com a MLP são mostrados na tabela 25, considerando a análise dependente do intérprete.

Tabela 24 – Acurácia de classificadores no problema de três classes: duas EFGs mais a expressão neutra. Análise dependente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A				
Afirm. - Int. (Dúvida)	0,87	0,00	0,87	0,86
Afirm. - Foco	0,88	0,00	0,88	0,87
Afirm. - Negativa	0,84	0,01	0,85	0,81
Afirm.- Inter. (Qu)	0,89	0,00	0,89	0,88
Afirm. - Inter. (s/n)	0,88	0,01	0,89	0,88
Inter. (Dúvida) - Foco	0,96	0,00	0,96	0,95
Inter. (Dúvida) - Negativa	0,93	0,00	0,93	0,93
Inter. (Dúvida) - Inter. (Qu)	0,95	0,00	0,96	0,94
Inter. (Dúvida) - Inter. (s/n)	0,94	0,00	0,95	0,94
Foco - Negativa	0,94	0,01	0,95	0,93
Foco - Inter. (Qu)	0,96	0,01	0,97	0,94
Foco - Inter. (s/n)	0,96	0,00	0,97	0,96
Negativa - Inter. (Qu)	0,92	0,01	0,94	0,91
Negativa - Inter. (s/n)	0,93	0,00	0,93	0,92
Inte. (Qu) - Inter. (s/n)	0,96	0,00	0,96	0,95

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Tabela 25 – Acurácia de classificadores no problema de três classes: duas EFGs mais a expressão neutra. Análise dependente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete B				
Afirm. - Inter. (Dúvida)	0,78	0,01	0,80	0,76
Afirm. - Foco	0,72	0,01	0,73	0,71
Afirm. - Negativa	0,61	0,03	0,64	0,56
Afirm. - Inter. (Qu)	0,73	0,02	0,77	0,69
Afirm. - Inter. (s/n)	0,82	0,01	0,84	0,80
Inter. (Dúvida) - Foco	0,92	0,01	0,93	0,89
Inter. (Dúvida) - Negativa	0,81	0,01	0,83	0,79
Inter. (Dúvida) - Inter. (Qu)	0,90	0,01	0,91	0,88
Inter. (Dúvida) - Inter. (s/n)	0,92	0,01	0,94	0,90
Foco - Negativa	0,73	0,01	0,75	0,70
Foco - Inter. (Qu)	0,88	0,02	0,90	0,85
Foco - Inter. (s/n)	0,92	0,01	0,93	0,91
Negativa - Inter. (Qu)	0,76	0,03	0,80	0,72
Negativa - Inter. (s/n)	0,82	0,02	0,83	0,78
Inter. (Qu) - Inter. (s/n)	0,89	0,01	0,90	0,88

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os resultados apresentados com o ponto do nariz (9 pontos) são semelhantes aos resultados dos resultados sem nariz (8 pontos). As combinações das EFGs *Afirmativa - Interrogativa (Dúvida)*, *Afirmativa - Interrogativa (Foco)*, *Afirmativa - Negativa*, *Afirmativa - Interrogativa (Qu)*, *Interrogativa (Dúvida) - Foco*, *Interrogativa (Dúvida) - Negativa*, *Interrogativa (Dúvida) - Interrogativa (s/n)* e *Negativa - Interrogativa (s/n)* dos dados provenientes do intérprete A apresentam a mesma média nas execuções e os demais resultados apresentam 0,1 de diferença. Já para os dados provenientes do intérprete B as combinações de EFGs *Afirmativa - Interrogativa (Qu)*, *Foco - interrogativa (s/n)* também apresentam médias iguais e as demais apresentam a diferença também de 0,1.

A tabela 26 retrata os erros de borda cometidos pelo classificador para os nove pontos utilizados para o treinamento dos testes dependentes.

Tabela 26 – Análise de erros de borda no problema de três classes **nove pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A			
Afirm. - Int. (Dúvida)	78 (28,57)	113 (14,04)	69,03
Afirm. - Foco	50 (18,32)	88 (10,81)	56,82
Afirm. - Negativa	68 (27,76)	96 (14,22)	70,83
Afirm. -Int. (Qu)	57 (23,27)	89 (11,34)	64,04
Afirm. - Int. (s/n)	67 (24,54)	97 (11,59)	69,07
Int. (Dúvida) - Foco	35 (12,82)	39 (4,34)	89,74
Int. (Dúvida) - Negativa	50 (20,41)	56 (7,37)	89,29
Int. (Dúvida) - Int. (Qu)	32 (13,06)	37 (4,25)	86,49
Int. (Dúvida) - Int. (s/n)	47 (17,22)	48 (5,21)	97,92
Foco - Negativa	32 (13,06)	35 (4,55)	91,43
Foco - Int. (Qu)	29 (11,84)	36 (4,10)	80,56
Foco - Int. (s/n)	40 (14,65)	41 (4,40)	97,56
Negativa - Int. (Qu)	43 (19,82)	51 (6,89)	84,31
Negativa. - Int. (s/n)	46 (18,78)	59 (7,45)	77,97
Int. (Qu) - Int. (s/n)	33 (13,47)	43 (4,77)	76,74
Intérprete B			
Afirm. - Int. (Dúvida)	78 (28,57)	179 (20,48)	43,58
Afirm. - Foco	60 (24,49)	141 (18,73)	42,55
Afirm. - Negativa	103 (37,73)	244 (31,48)	42,21
Afirm. -Int. (Qu)	71 (28,98)	159 (21,46)	44,65
Afirm.- Int. (s/n)	81 (26,91)	170 (15,92)	47,65
Int. (Dúvida) - Foco	50 (20,41)	74 (8,19)	67,57
Int. (Dúvida) - Negativa	77 (28,21)	169 (18,27)	45,56
Int. (Dúvida) - Int. (Qu)	51 (20,82)	80 (8,98)	63,75
Int. (Dúvida) - Int. (s/n)	70 (23,26)	87 (7,14)	80,46
Foco - Negativa.	73 (29,80)	172 (21,39)	42,44
Foco - Int. (Qu)	41 (18,89)	68 (8,83)	60,29
Foco - Int. (s/n)	52 (19,05)	72 (6,56)	72,22
Negativa. - Int. (Qu)	67 (27,35)	138 (17,42)	48,55
Negativa. - Int. (s/n)	103 (34,22)	229 (20,46)	44,98
Int. (Qu) - Int. (s/n)	79 (28,94)	122 (11,24)	64,75

Fonte: Maria Eduarda de Araújo Cardoso, 2018

O comportamento dos erros é semelhante aos dados dos oito pontos. A combinação das EFGs *Afirmativa - Foco* tem o resultado igual para ambos treinamentos com oito e nove pontos com testes dependentes.

A tabela 27 apresentam os resultados com os nove pontos com os testes independentes dos dados provenientes dos dois intérpretes.

Tabela 27 – Acurácia de classificadores no problema de três classes **nove pontos** da face: duas EFGs mais a expressão neutra. Análise independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A (treino) - Intérprete B (teste)				
Afir. - Inter. (Dúvida)	0,42	0,10	0,66	0,34
Afir. - Foco	0,41	0,03	0,47	0,35
Afir. - Negativa	0,43	0,05	0,48	0,31
Afir. - Inter. (Qu)	0,38	0,05	0,48	0,33
Afir. - Inter. (s/n)	0,55	0,06	0,59	0,40
Inter. (Dúvida) - Foco	0,57	0,06	0,73	0,50
Inter. (Dúvida) - Negativa	0,48	0,06	0,58	0,41
Inter. (Dúvida) - Inter. (Qu)	0,55	0,08	0,68	0,47
Inter. (Dúvida) - Inter. (s/n)	0,55	0,05	0,68	0,52
Foco - Negativa	0,24	0,02	0,26	0,22
Foco - Inter. (Qu)	0,27	0,06	0,43	0,24
Foco - Inter. (s/n)	0,20	0,03	0,27	0,18
Negativa - Inter. (Qu)	0,33	0,02	0,36	0,29
Negativa - Inter. (s/n)	0,43	0,07	0,55	0,27
Inter. (Qu) - Inter. (s/n)	0,26	0,05	0,36	0,19
Intérprete B (treino) - Intérprete A (teste)				
Afirm. - Inter. (Dúvida)	0,72	0,02	0,75	0,70
Afirm. - Foco	0,65	0,00	0,65	0,65
Afirm. - Negativa	0,56	0,04	0,67	0,54
Afirm. - Inter. (Qu)	0,52	0,04	0,65	0,51
Afirm. - Inter. (s/n)	0,57	0,00	0,57	0,57
Inter. (Dúvida) - Foco	0,85	0,04	0,89	0,77
Inter. (Dúvida) - Negativa	0,69	0,03	0,72	0,60
Inter. (Dúvida) - Inter. (Qu)	0,44	0,02	0,47	0,40
Inter. (Dúvida) - Inter. (s/n)	0,66	0,12	0,76	0,45
Foco - Negativa	0,53	0,08	0,63	0,36
Foco - Inter. (Qu)	0,66	0,03	0,72	0,63
Foco - Inter. (s/n)	0,75	0,05	0,93	0,70
Negativa - Inter. (Qu)	0,43	0,08	0,56	0,26
Negativa - Inter. (s/n)	0,49	0,12	0,56	0,17
Inter. (Qu) - Inter. (s/n)	0,55	0,21	0,85	0,21

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Esses resultados apresentados (tabela 27) mostram uma instabilidade nos resultados. Ainda que *Afirmativa - Negativa*, *Interrogativa (Dúvida) - Interrogativa (s/n)*, *Foco - Negativa* e *Negativa - Interrogativa (Qu)* com o treino com os dados provenientes do intérprete A apresente as médias iguais, os demais resultados expõe uma oscilação de até 0,8 de diferença aos demais. Os resultados para o intérprete B todas apresentam oscilações em suas médias.

A tabela 28 apresenta os resultados dos erros gerais e de borda dos testes independentes com os nove pontos. Nesses resultados a combinação de EFG *Afirmativa - Interrogativa (s/n)* apresentaram resultados iguais para os oito pontos utilizados para o treinamento do classificador.

Tabela 28 – Análise de erros de borda no problema de três classes **nove pontos** da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A (treino) - Intérprete B (teste)			
Afirm. - Int. (Dúvida)	139 (50,92)	295 (33,75)	47,12
Afirm. - Foco	96 (39,18)	177 (23,51)	54,24
Afirm. - Neg.	138 (50,55)	276 (35,61)	50,00
Afirm. - Int. (Qu)	125 (51,02)	387 (52,23)	32,30
Afirm. - Int. (s/n)	131 (43,52)	337 (31,55)	38,87
Int. (Dúvida) - Foco	90 (36,73)	184 (20,38)	48,91
Int. (Dúvida) - Negativa	122 (44,69)	367 (39,68)	33,24
Int. (Dúvida) - Int. (Qu)	109 (44,49)	235 (26,37)	46,38
Int. (Dúvida) - Int. (s/n)	154 (51,16)	395 (32,43)	38,99
Foco - Neg.	115 (46,94)	337 (41,92)	34,12
Foco - Int. (Qu)	104 (47,93)	434 (56,36)	23,96
Foco - Int. (s/n)	133 (48,72)	620 (56,52)	21,45
Neg. - Int. (Qu)	111 (45,31)	377 (47,60)	29,44
Neg. - Int. (s/n)	134 (44,52)	610 (54,51)	21,97
Int. (Qu) - Int. (s/n)	131 (47,99)	651 (60,00)	20,12
Intérprete B (treino) - Intérprete A (teste)			
Afirm. - Inter. (Dúvida)	80 (29,30)	169 (20,99)	47,34
Afirm. - Foco	137 (50,18)	287 (35,26)	47,74
Afirm. - Neg.	102 (41,63)	228 (33,78)	44,74
Afirm. - Inter. (Qu)	101 (41,22)	272 (34,65)	37,13
Afirm.- Int. (s/n)	137 (50,18)	362 (43,25)	37,85
Int. (Dúvida) - Foco	60 (21,98)	74 (8,23)	81,08
Int. (Dúvida) - Negativa	84 (34,29)	227 (29,87)	37,00
Int. (Dúvida) - Int. (Qu)	109 (44,49)	411 (47,24)	26,52
Int. (Dúvida) - Int. (s/n)	56 (20,51)	69 (7,48)	81,16
Foco - Negativa	79 (32,24)	252 (32,77)	31,35
Foco - Inter. (Qu)	88 (35,92)	245 (27,87)	35,92
Foco - Inter. (s/n)	39 (14,29)	50 (5,37)	78,00
Neg. - Inter. (Qu)	88 (40,55)	246 (33,24)	35,77
Neg. - Inter. (s/n)	110 (44,90)	346 (43,69)	31,79
Inter. (Qu) - Inter. (s/n)	61 (24,90)	178 (19,73)	34,27

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Agora serão apresentadas as tabelas dos erros de borda, dos erros na borda e dos erros gerais cometidos pelo classificador no treinamento dos dados provenientes do intérprete A e dos dados provenientes do intérprete B para oito pontos e as distâncias dos pontos. Os resultados são apresentados na tabela 29 com os testes dependentes do intérprete e a tabela 30 com os testes independentes do intérprete.

Tabela 29 – Análise de erros de borda no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A			
Afirm. - Int. (Dúvida)	71 (26,01)	104 (12,92)	68,27
Afirm. - Foco	51 (18,68)	84 (10,44)	60,71
Afirm. - Negativa	64 (26,12)	92 (13,63)	69,57
Afirm. -Int. (Qu)	52 (21,22)	90 (11,46)	57,78
Afirm. - Int. (s/n)	66 (24,18)	96 (11,47)	68,75
Int. (Dúvida) - Foco	33 (12,09)	36 (4,00)	91,67
Int. (Dúvida) - Negativa	51 (20,82)	58 (7,63)	87,93
Int. (Dúvida) - Int. (Qu)	39 (15,92)	45 (5,17)	86,67
Int. (Dúvida) - Int. (s/n)	49 (17,95)	53 (5,75)	92,45
Foco - Negativa	34 (13,88)	37 (4,81)	91,89
Foco - Int. (Qu)	24 (9,80)	29 (3,30)	82,76
Foco - Int. (s/n)	32 (11,72)	32 (3,44)	100,00
Negativa - Int. (Qu)	42 (19,35)	54 (7,30)	77,78
Negativa. - Int. (s/n)	51 (20,82)	76 (9,60)	67,11
Int. (Qu) - Int. (s/n)	35 (14,29)	43 (4,77)	81,40
Intérprete B			
Afirm. - Int. (Dúvida)	82 (30,04)	187 (21,40)	43,85
Afirm. - Foco	67 (27,35)	156 (20,72)	42,95
Afirm. - Negativa	105 (38,46)	282 (36,39)	37,23
Afirm. -Int. (Qu)	82 (33,47)	198 (26,72)	41,41
Afirm.- Int. (s/n)	91 (30,23)	200 (18,73)	45,50
Int. (Dúvida) - Foco	44 (17,96)	66 (7,31)	66,67
Int. (Dúvida) - Negativa	75 (27,47)	162 (17,51)	46,30
Int. (Dúvida) - Int. (Qu)	55 (22,45)	88 (9,88)	62,50
Int. (Dúvida) - Int. (s/n)	69 (22,92)	88 (7,22)	78,41
Foco - Negativa.	63 (25,71)	170 (21,14)	37,06
Foco - Int. (Qu)	41 (18,89)	73 (9,48)	56,16
Foco - Int. (s/n)	45 (16,48)	73 (6,65)	61,64
Negativa. - Int. (Qu)	75 (30,61)	152 (19,19)	49,34
Negativa. - Int. (s/n)	95 (31,56)	156 (13,94)	60,90
Int. (Qu) - Int. (s/n)	72 (26,37)	109 (10,05)	66,06

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Tabela 30 – Análise de erros de borda no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A (treino) - Intérprete B (teste)			
Afirm. - Int. (Dúvida)	151 (55,31)	430 (49,20)	35,12
Afirm. - Foco	101 (41,22)	198 (26,29)	51,01
Afirm. - Neg.	134 (49,08)	279 (36,00)	48,03
Afirm. - Int. (Qu)	141 (57,55)	394 (53,17)	35,79
Afirm. - Int. (s/n)	148 (49,17)	346 (32,40)	42,77
Int. (Dúvida) - Foco	96 (39,18)	326 (36,10)	29,45
Int. (Dúvida) - Negativa	128 (46,89)	313 (33,84)	40,89
Int. (Dúvida) - Int. (Qu)	106 (43,27)	261 (29,29)	40,61
Int. (Dúvida) - Int. (s/n)	152 (50,50)	416 (34,15)	36,54
Foco - Neg.	116 (47,35)	313 (38,93)	37,06
Foco - Int. (Qu)	105 (48,39)	437 (56,75)	24,03
Foco - Int. (s/n)	132 (48,35)	602 (54,88)	21,93
Neg. - Int. (Qu)	106 (43,27)	337 (42,55)	31,45
Neg. - Int. (s/n)	135 (44,85)	569 (50,85)	23,73
Int. (Qu) - Int. (s/n)	133 (48,72)	655 (60,37)	20,31
Intérprete B (treino) - Intérprete A (teste)			
Afirm. - Inter. (Dúvida)	135 (49,45)	369 (45,84)	36,59
Afirm. - Foco	132 (48,35)	273 (33,54)	48,35
Afirm. - Neg.	101 (41,22)	201 (29,78)	50,25
Afirm. - Inter. (Qu)	82 (33,47)	205 (26,11)	40,00
Afirm.- Int. (s/n)	137 (50,18)	362 (43,25)	37,85
Int. (Dúvida) - Foco	60 (21,98)	65 (7,23)	92,31
Int. (Dúvida) - Negativa	86 (35,10)	207 (27,24)	41,55
Int. (Dúvida) - Int. (Qu)	119 (48,57)	442 (50,80)	26,92
Int. (Dúvida) - Int. (s/n)	54 (19,78)	58 (6,29)	93,10
Foco - Negativa	76 (31,02)	219 (28,48)	34,70
Foco - Inter. (Qu)	95 (38,78)	294 (33,45)	32,31
Foco - Inter. (s/n)	81 (29,67)	205 (22,02)	39,51
Neg. - Inter. (Qu)	70 (32,26)	200 (27,03)	35,00
Neg. - Inter. (s/n)	108 (44,08)	259 (32,70)	41,70
Inter. (Qu) - Inter. (s/n)	122 (49,80)	498 (55,21)	24,50

Fonte: Maria Eduarda de Araújo Cardoso, 2018

A tabela 31 apresenta os resultados dos testes dependentes e a tabela 32 apresenta os resultados dos testes independentes do intérprete para as distância dos pontos.

Tabela 31 – Análise de erros de borda no problema de três classes com **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A			
Afirm. - Int. (Dúvida)	65 (23,81)	92 (11,43)	70,75
Afirm. - Foco	45 (16,48)	80 (9,83)	56,25
Afirm. - Negativa	62 (25,31)	100 (14,81)	62,00
Afirm. -Int. (Qu)	52 (21,22)	93 (11,85)	55,91
Afirm. - Int. (s/n)	58 (21,25)	94 (11,23)	61,70
Int. (Dúvida) - Foco	31 (11,36)	32 (3,56)	96,88
Int. (Dúvida) - Negativa	46 (18,78)	54 (7,11)	85,19
Int. (Dúvida) - Int. (Qu)	30 (15,51)	63 (7,24)	60,32
Int. (Dúvida) - Int. (s/n)	50 (18,32)	51 (5,53)	98,04
Foco - Negativa	28 (11,43)	36 (4,68)	77,78
Foco - Int. (Qu)	21 (8,57)	25 (2,84)	84,00
Foco - Int. (s/n)	27 (9,89)	29 (3,11)	93,10
Negativa - Int. (Qu)	36 (16,59)	59 (7,97)	61,02
Negativa. - Int. (s/n)	48 (19,59)	56 (7,07)	85,71
Int. (Qu) - Int. (s/n)	37 (15,10)	39 (4,32)	94,87
Intérprete B			
Afirm. - Int. (Dúvida)	115 (42,12)	245 (28,03)	46,94
Afirm. - Foco	77 (31,43)	171 (22,71)	45,03
Afirm. - Negativa	114 (41,76)	272 (35,10)	41,91
Afirm. -Int. (Qu)	81 (33,06)	195 (26,32)	41,54
Afirm.- Int. (s/n)	109 (36,21)	222 (20,79)	49,10
Int. (Dúvida) - Foco	69 (28,16)	145 (16,06)	47,59
Int. (Dúvida) - Negativa	97 (35,53)	222 (24,00)	43,69
Int. (Dúvida) - Int. (Qu)	57 (23,27)	91 (10,21)	62,64
Int. (Dúvida) - Int. (s/n)	85 (28,24)	138 (11,33)	61,59
Foco - Negativa.	99 (40,41)	238 (29,60)	41,60
Foco - Int. (Qu)	50 (23,04)	84 (10,91)	59,52
Foco - Int. (s/n)	77 (28,21)	130 (11,85)	59,23
Negativa. - Int. (Qu)	84 (34,29)	164 (20,71)	51,22
Negativa. - Int. (s/n)	117 (38,87)	226 (20,20)	51,77
Int. (Qu) - Int. (s/n)	67 (24,54)	100 (9,22)	67,00

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Tabela 32 – Análise de erros de borda no problema de três classes com a **distância dos pontos**: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A (treino) - Intérprete B (teste)			
Afirm. - Int. (Dúvida)	145 (53,11)	461 (52,75)	31,45
Afirm. - Foco	120 (48,98)	294 (39,04)	40,82
Afirm. - Neg.	133 (48,72)	369 (47,61)	36,04
Afirm. - Int. (Qu)	135 (55,10)	341 (46,02)	39,59
Afirm. - Int. (s/n)	151 (50,17)	420 (39,33)	35,95
Int. (Dúvida) - Foco	109 (44,49)	379 (41,97)	28,76
Int. (Dúvida) - Negativa	129 (47,25)	374 (40,43)	34,49
Int. (Dúvida) - Int. (Qu)	131 (53,47)	403 (45,23)	32,51
Int. (Dúvida) - Int. (s/n)	149 (49,50)	702 (57,64)	21,23
Foco - Neg.	93 (37,96)	303 (37,69)	30,69
Foco - Int. (Qu)	107 (49,31)	374 (48,57)	28,61
Foco - Int. (s/n)	136 (49,82)	641 (58,43)	21,22
Neg. - Int. (Qu)	107 (43,67)	360 (45,45)	29,72
Neg. - Int. (s/n)	142 (47,18)	555 (49,60)	25,59
Int. (Qu) - Int. (s/n)	133 (48,72)	654 (60,28)	20,34
Intérprete B (treino) - Intérprete A (teste)			
Afirm. - Inter. (Dúvida)	117 (42,86)	214 (26,58)	54,67
Afirm. - Foco	74 (27,11)	168 (20,64)	44,05
Afirm. - Neg.	100 (40,82)	204 (30,22)	49,02
Afirm. - Inter. (Qu)	98 (40,00)	226 (28,79)	43,36
Afirm.- Int. (s/n)	133 (48,72)	286 (34,17)	46,50
Int. (Dúvida) - Foco	78 (28,57)	139 (15,46)	56,12
Int. (Dúvida) - Negativa	97 (39,59)	198 (26,05)	48,99
Int. (Dúvida) - Int. (Qu)	101 (41,22)	298 (34,25)	33,89
Int. (Dúvida) - Int. (s/n)	122 (44,69)	294 (31,89)	41,50
Foco - Negativa	93 (37,96)	222 (28,87)	41,89
Foco - Inter. (Qu)	79 (32,24)	301 (34,24)	26,25
Foco - Inter. (s/n)	61 (22,34)	131 (14,07)	46,56
Neg. - Inter. (Qu)	96 (44,24)	302 (40,81)	31,79
Neg. - Inter. (s/n)	117 (47,76)	397 (50,13)	29,47
Inter. (Qu) - Inter. (s/n)	79 (32,24)	237 (26,27)	33,33

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Mistura de especialistas

Resultados dos erros de borda e segmentação treinados com oito pontos da face são apresentados nas tabelas 33 com os testes dependentes e 34 com os testes independentes do intérprete.

Tabela 33 – Análise de erros de borda no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A			
Afirm. - Int. (Dúvida)	68 (24,91)	101 (12,55)	67,33
Afirm. - Foco	137 (50,18)	287 (35,26)	47,74
Afirm. - Negativa	61 (24,90)	101 (14,96)	60,40
Afirm. -Int. (Qu)	54 (22,04)	69 (8,79)	78,26
Afirm. - Int. (s/n)	66 (24,18)	89 (10,63)	74,16
Int. (Dúvida) - Foco	136 (49,82)	609 (67,74)	22,33
Int. (Dúvida) - Negativa	48 (19,59)	55 (7,24)	87,27
Int. (Dúvida) - Int. (Qu)	35 (14,29)	39 (4,48)	89,74
Int. (Dúvida) - Int. (s/n)	44 (16,12)	48 (5,21)	91,67
Foco - Negativa	35 (14,29)	46 (5,98)	76,09
Foco - Int. (Qu)	22 (8,98)	32 (3,64)	68,75
Foco - Int. (s/n)	31 (11,36)	32 (3,44)	96,88
Negativa - Int. (Qu)	31 (14,29)	38 (5,14)	81,58
Negativa. - Int. (s/n)	46 (18,78)	54 (6,82)	85,19
Int. (Qu) - Int. (s/n)	34 (13,88)	40 (4,43)	85,00
Intérprete B			
Afirm. - Int. (Dúvida)	84 (30,77)	198 (22,65)	42,42
Afirm. - Foco	71 (28,98)	190 (25,23)	37,37
Afirm. - Negativa	137 (50,18)	465 (60,00)	29,46
Afirm. -Int. (Qu)	74 (30,20)	177 (23,89)	41,81
Afirm.- Int. (s/n)	99 (32,89)	216 (20,22)	45,83
Int. (Dúvida) - Foco	55 (22,45)	75 (8,31)	73,33
Int. (Dúvida) - Negativa	93 (34,07)	224 (24,22)	41,52
Int. (Dúvida) - Int. (Qu)	59 (24,08)	86 (9,65)	68,60
Int. (Dúvida) - Int. (s/n)	71 (23,59)	81 (6,65)	87,65
Foco - Negativa.	89 (36,33)	227 (28,23)	39,21
Foco - Int. (Qu)	43 (19,82)	72 (9,35)	59,72
Foco - Int. (s/n)	65 (23,81)	84 (7,66)	77,38
Negativa. - Int. (Qu)	87 (35,51)	229 (28,91)	37,99
Negativa. - Int. (s/n)	117 (38,87)	287 (25,65)	40,77
Int. (Qu) - Int. (s/n)	74 (27,11)	94 (8,66)	78,72

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Tabela 34 – Análise de erros de borda no problema de três classes com **oito pontos** da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A (treino) - Intérprete B (teste)			
Afirm. - Int. (Dúvida)	144 (52,75)	474 (54,23)	30,38
Afirm. - Foco	123 (50,20)	420 (55,78)	29,29
Afirm. - Neg.	131 (47,99)	373 (48,13)	35,12
Afirm. - Int. (Qu)	138 (56,33)	333 (44,94)	41,44
Afirm. - Int. (s/n)	136 (45,18)	497 (46,54)	27,36
Int. (Dúvida) - Foco	105 (42,86)	327 (36,21)	32,11
Int. (Dúvida) - Negativa	136 (49,82)	529 (57,19)	25,71
Int. (Dúvida) - Int. (Qu)	120 (48,98)	420 (47,14)	28,57
Int. (Dúvida) - Int. (s/n)	148 (49,17)	404 (33,17)	36,63
Foco - Neg.	123 (50,20)	403 (50,12)	30,52
Foco - Int. (Qu)	110 (50,69)	485 (62,99)	22,68
Foco - Int. (s/n)	134 (49,08)	639 (58,25)	20,97
Neg. - Int. (Qu)	107 (43,67)	351 (44,32)	30,48
Neg. - Int. (s/n)	140 (46,51)	553 (49,42)	25,32
Int. (Qu) - Int. (s/n)	140 (51,28)	431 (39,72)	32,48
Intérprete B (treino) - Intérprete A (teste)			
Afirm. - Inter. (Dúvida)	86 (31,50)	149 (18,51)	57,72
Afirm. - Foco	140 (51,28)	274 (33,66)	51,09
Afirm. - Neg.	112 (45,71)	294 (43,56)	38,10
Afirm. - Inter. (Qu)	123 (50,20)	384 (48,92)	32,03
Afirm.- Int. (s/n)	137 (50,18)	362 (43,25)	37,85
Int. (Dúvida) - Foco	54 (19,78)	74 (8,23)	72,97
Int. (Dúvida) - Negativa	99 (40,41)	208 (27,37)	47,60
Int. (Dúvida) - Int. (Qu)	95 (38,78)	271 (31,15)	35,06
Int. (Dúvida) - Int. (s/n)	92 (33,70)	309 (33,51)	29,77
Foco - Negativa	123 (50,20)	255 (33,16)	48,24
Foco - Inter. (Qu)	69 (28,16)	261 (29,69)	26,44
Foco - Inter. (s/n)	80 (29,30)	211 (22,66)	37,91
Neg. - Inter. (Qu)	76 (35,02)	267 (36,08)	28,46
Neg. - Inter. (s/n)	121 (49,39)	261 (32,95)	46,36
Inter. (Qu) - Inter. (s/n)	92 (37,55)	255 (28,27)	36,08

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Resultados dos erros de borda e segmentação treinados com as distâncias dos pontos da face são apresentados na tabela 35 com os testes dependentes do intérprete e a tabela 36 com os testes independentes do intérprete.

Tabela 35 – Análise de erros de borda no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A			
Afirm. - Int. (Dúvida)	65 (23,81)	89 (11,06)	73,03
Afirm. - Foco	48 (17,58)	72 (8,85)	66,67
Afirm. - Negativa	63 (25,71)	84 (12,44)	75,00
Afirm. -Int. (Qu)	47 (19,18)	60 (7,64)	78,33
Afirm. - Int. (s/n)	61 (22,34)	81 (9,68)	75,31
Int. (Dúvida) - Foco	36 (13,19)	38 (4,23)	94,74
Int. (Dúvida) - Negativa	52 (21,22)	63 (8,29)	82,54
Int. (Dúvida) - Int. (Qu)	47 (19,18)	55 (6,32)	85,45
Int. (Dúvida) - Int. (s/n)	51 (18,68)	53 (5,750)	96,23
Foco - Negativa	28 (11,43)	34 (4,42)	82,35
Foco - Int. (Qu)	19 (7,76)	20 (2,28)	95,00
Foco - Int. (s/n)	28 (10,26)	28 (3,01)	100,00
Negativa - Int. (Qu)	40 (18,43)	54 (7,30)	74,07
Negativa. - Int. (s/n)	50 (20,41)	56 (7,07)	89,29
Int. (Qu) - Int. (s/n)	39 (15,92)	44 (4,88)	(88,64)
Intérprete B			
Afirm. - Int. (Dúvida)	113 (41,39)	244 (27,92)	46,31
Afirm. - Foco	94 (38,37)	220 (29,22)	42,73
Afirm. - Negativa	125 (45,79)	340 (43,87)	36,76
Afirm. -Int. (Qu)	84 (34,29)	174 (23,48)	48,28
Afirm.- Int. (s/n)	105 (34,88)	224 (20,97)	46,88
Int. (Dúvida) - Foco	69 (28,16)	111 (12,29)	62,16
Int. (Dúvida) - Negativa	108 (39,56)	267 (28,86)	40,45
Int. (Dúvida) - Int. (Qu)	67 (27,35)	93 (10,44)	72,04
Int. (Dúvida) - Int. (s/n)	85 (28,24)	117 (9,61)	72,65
Foco - Negativa.	106 (43,27)	277 (34,45)	38,27
Foco - Int. (Qu)	51 (23,50)	88 (11,43)	57,95
Foco - Int. (s/n)	61 (22,34)	93 (8,48)	65,59
Negativa. - Int. (Qu)	80 (32,65)	193 (24,37)	41,45
Negativa. - Int. (s/n)	100 (33,22)	261 (23,32)	38,31
Int. (Qu) - Int. (s/n)	63 (23,08)	91 (8,39)	69,23

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Tabela 36 – Análise de erros de borda no problema de três classes com a **distância dos pontos** da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A (treino) - Intérprete B (teste)			
Afirm. - Int. (Dúvida)	137 (50,18)	430 (49,20)	31,86
Afirm. - Foco	113 (46,12)	283 (37,58)	39,93
Afirm. - Neg.	134 (49,08)	383 (49,42)	34,99
Afirm. - Int. (Qu)	137 (55,92)	359 (48,45)	38,16
Afirm. - Int. (s/n)	144 (47,84)	451 (42,23)	31,93
Int. (Dúvida) - Foco	96 (39,18)	356 (39,42)	26,97
Int. (Dúvida) - Negativa	127 (46,52)	460 (49,73)	27,61
Int. (Dúvida) - Int. (Qu)	114 (46,53)	432 (48,48)	26,39
Int. (Dúvida) - Int. (s/n)	146 (48,50)	713 (58,54)	20,48
Foco - Neg.	110 (44,90)	329 (40,92)	33,43
Foco - Int. (Qu)	105 (48,39)	441 (57,27)	23,81
Foco - Int. (s/n)	136 (49,82)	641 (58,43)	21,22
Neg. - Int. (Qu)	95 (38,78)	322 (40,66)	29,50
Neg. - Int. (s/n)	144 (47,84)	611 (54,60)	23,57
Int. (Qu) - Int. (s/n)	131 (47,99)	611 (56,31)	21,44
Intérprete B (treino) - Intérprete A (teste)			
Afirm. - Inter. (Dúvida)	126 (46,15)	267 (33,17)	47,19
Afirm. - Foco	136 (49,82)	269 (33,05)	50,56
Afirm. - Neg.	118 (48,16)	226 (33,48)	52,21
Afirm. - Inter. (Qu)	94 (38,37)	216 (27,52)	43,52
Afirm. - Int. (s/n)	137 (50,18)	362 (43,25)	37,85
Int. (Dúvida) - Foco	74 (27,11)	156 (17,35)	47,44
Int. (Dúvida) - Negativa	103 (42,04)	278 (36,58)	37,05
Int. (Dúvida) - Int. (Qu)	106 (43,27)	324 (37,24)	32,72
Int. (Dúvida) - Int. (s/n)	137 (50,18)	330 (35,79)	41,52
Foco - Negativa	85 (34,69)	190 (24,71)	44,74
Foco - Inter. (Qu)	66 (26,94)	227 (25,82)	29,07
Foco - Inter. (s/n)	77 (28,21)	202 (21,70)	38,12
Neg. - Inter. (Qu)	111 (51,15)	353 (47,70)	31,44
Neg. - Inter. (s/n)	104 (42,45)	231 (29,17)	45,02
Inter. (Qu) - Inter. (s/n)	79 (32,24)	236 (26,16)	33,47

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os resultados de acurácia média obtida com os mesmos parâmetros adicionando o pontos do nariz (9 pontos), são mostrados na tabela 37, considerando a análise dependente do intérprete. A tabela 38 apresentam os resultados para os casos independentes do intérprete.

Tabela 37 – Acurácia de classificadores no problema de três classes com **nove pontos** da face: duas EFGs mais a expressão neutra. Análise dependente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nos números para problemas de maior dificuldade; nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A				
Afirm. - Int. (Dúvida)	0,37	0,23	0,86	0,21
Afirm. - Foco	0,27	0,20	0,65	0,14
Afirm. - Negativa	0,67	0,26	0,87	0,25
Afirm.- Inter. (Qu)	0,91	0,01	0,92	0,90
Afirm. - Inter. (s/n)	0,87	0,08	0,90	0,65
Inter. (Dúvida) - Foco	0,26	0,22	0,68	0,13
Inter. (Dúvida) - Negativa	0,36	0,30	0,92	0,18
Inter. (Dúvida) - Inter. (Qu)	0,72	0,36	0,95	0,20
Inter. (Dúvida) - Inter. (s/n)	0,45	0,37	0,95	0,19
Foco - Negativa	0,35	0,30	0,94	0,15
Foco - Inter. (Qu)	0,48	0,38	0,96	0,13
Foco - Inter. (s/n)	0,35	0,30	0,95	0,12
Negativa - Inter. (Qu)	0,93	0,01	0,94	0,92
Negativa - Inter. (s/n)	0,85	0,23	0,94	0,24
Inte. (Qu) - Inter. (s/n)	0,96	0,00	0,96	0,95
Intérprete B				
Afirm. - Inter. (Dúvida)	0,46	0,21	0,78	0,28
Afirm. - Foco	0,34	0,14	0,70	0,24
Afirm. - Negativa	0,32	0,04	0,40	0,29
Afirm. - Inter. (Qu)	0,52	0,18	0,77	0,33
Afirm. - Inter. (s/n)	0,34	0,19	0,81	0,23
Inter. (Dúvida) - Foco	0,44	0,27	0,92	0,20
Inter. (Dúvida) - Negativa	0,39	0,20	0,81	0,24
Inter. (Dúvida) - Inter. (Qu)	0,75	0,31	0,90	0,17
Inter. (Dúvida) - Inter. (s/n)	0,30	0,12	0,52	0,23
Foco - Negativa	0,41	0,20	0,65	0,22
Foco - Inter. (Qu)	0,35	0,24	0,89	0,20
Foco - Inter. (s/n)	0,22	0,06	0,25	0,08
Negativa - Inter. (Qu)	0,34	0,22	0,77	0,19
Negativa - Inter. (s/n)	0,27	0,17	0,75	0,20
Inter. (Qu) - Inter. (s/n)	0,69	0,30	0,89	0,14

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Tabela 38 – Acurácia de classificadores no problema de três classes **nove pontos** da face: duas EFGs mais a expressão neutra. Análise independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ). Destaques em negrito nas EFGs para problemas de menor dificuldade

EFGs	μ	σ	máxima	miníma
Intérprete A (treino) - Intérprete B (teste)				
Afir. - Inter. (Dúvida)	0,31	0,05	0,40	0,28
Afir. - Foco	0,30	0,03	0,32	0,24
Afir. - Negativa	0,33	0,05	0,40	0,29
Afir. - Inter. (Qu)	0,40	0,09	0,57	0,28
Afir. - Inter. (s/n)	0,25	0,07	0,43	0,23
Inter. (Dúvida) - Foco	0,30	0,12	0,50	0,20
Inter. (Dúvida) - Negativa	0,28	0,05	0,33	0,24
Inter. (Dúvida) - Inter. (Qu)	0,48	0,08	0,65	0,32
Inter. (Dúvida) - Inter. (s/n)	0,34	0,15	0,59	0,23
Foco - Negativa	0,27	0,09	0,50	0,22
Foco - Inter. (Qu)	0,22	0,02	0,24	0,20
Foco - Inter. (s/n)	0,27	0,12	0,58	0,16
Negativa - Inter. (Qu)	0,31	0,07	0,36	0,19
Negativa - Inter. (s/n)	0,38	0,12	0,47	0,20
Inter. (Qu) - Inter. (s/n)	0,29	0,11	0,60	0,26
Intérprete B (treino) - Intérprete A (teste)				
Afirm. - Inter. (Dúvida)	0,32	0,17	0,57	0,21
Afirm. - Foco	0,42	0,26	0,74	0,14
Afirm. - Negativa	0,27	0,10	0,54	0,21
Afirm. - Inter. (Qu)	0,38	0,14	0,51	0,22
Afirm. - Inter. (s/n)	0,26	0,11	0,57	0,20
Inter. (Dúvida) - Foco	0,22	0,18	0,68	0,13
Inter. (Dúvida) - Negativa	0,34	0,20	0,66	0,18
Inter. (Dúvida) - Inter. (Qu)	0,31	0,06	0,41	0,20
Inter. (Dúvida) - Inter. (s/n)	0,27	0,18	0,60	0,10
Foco - Negativa	0,20	0,09	0,40	0,15
Foco - Inter. (Qu)	0,29	0,24	0,71	0,12
Foco - Inter. (s/n)	0,32	0,24	0,67	0,12
Negativa - Inter. (Qu)	0,28	0,14	0,52	0,14
Negativa - Inter. (s/n)	0,24	0,13	0,58	0,17
Inter. (Qu) - Inter. (s/n)	0,58	0,24	0,73	0,24

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Os resultados dos erros de borda e erros gerais com o ponto do nariz (9 pontos), são mostrados na tabela 39, considerando a análise dependente do intérprete. A tabela 40 apresentam os resultados para os casos independentes do intérprete.

Tabela 39 – Análise de erros de borda no problema de três classes com **nove pontos** da face: duas EFGs mais a expressão neutra. Análise em testes dependentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A			
Afirm. - Int. (Dúvida)	137 (50,18)	345 (42,86)	39,71
Afirm. - Foco	137 (50,18)	287 (35,26)	47,74
Afirm. - Negativa	61 (24,90)	88 (13,04)	69,32
Afirm. -Int. (Qu)	45 (18,37)	63 (8,03)	71,43
Afirm. - Int. (s/n)	65 (23,81)	84 (10,04)	77,38
Int. (Dúvida) - Foco	137 (50,18)	290 (32,26)	47,24
Int. (Dúvida) - Negativa	53 (21,63)	61 (8,03)	86,89
Int. (Dúvida) - Int. (Qu)	40 (16,33)	46 (5,29)	86,96
Int. (Dúvida) - Int. (s/n)	44 (16,12)	48 (5,21)	91,67
Foco - Negativa	34 (13,88)	44 (5,72)	77,27
Foco - Int. (Qu)	22 (8,98)	32 (3,64)	68,75
Foco - Int. (s/n)	31 (11,36)	32 (3,44)	96,88
Negativa - Int. (Qu)	39 (17,97)	47 (6,35)	82,98
Negativa. - Int. (s/n)	43 (17,55)	48 (6,06)	89,58
Int. (Qu) - Int. (s/n)	36 (14,69)	40 (4,43)	90,00
Intérprete B			
Afirm. - Int. (Dúvida)	83 (30,40)	192 (21,97)	43,23
Afirm. - Foco	77 (31,43)	191 (25,37)	40,31
Afirm. - Negativa	137 (50,18)	465 (60,00)	29,46
Afirm. -Int. (Qu)	71 (28,98)	169 (22,81)	42,01
Afirm.- Int. (s/n)	94 (31,23)	199 (18,63)	47,24
Int. (Dúvida) - Foco	53 (21,63)	75 (8,31)	70,67
Int. (Dúvida) - Negativa	76 (27,84)	175 (18,92)	43,43
Int. (Dúvida) - Int. (Qu)	59 (24,08)	87 (9,76)	67,82
Int. (Dúvida) - Int. (s/n)	151 (50,17)	585 (48,03)	25,81
Foco - Negativa.	95 (38,78)	253 (31,47)	37,55
Foco - Int. (Qu)	47 (21,66)	84 (10,91)	55,95
Foco - Int. (s/n)	136 (49,82)	641 (58,43)	21,22
Negativa. - Int. (Qu)	75 (30,61)	184 (23,23)	40,76
Negativa. - Int. (s/n)	108 (35,88)	265 (23,68)	40,75
Int. (Qu) - Int. (s/n)	87 (31,87)	123 (11,34)	70,73

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Tabela 40 – Análise de erros de borda no problema de três classes com **nove pontos** da face: duas EFGs mais a expressão neutra. Análise em testes independentes do intérprete

EFGs	erros de borda # (%)	erros gerais # (%)	erros na borda %
Intérprete A (treino) - Intérprete B (teste)			
Afirm. - Int. (Dúvida)	83 (30,40)	192 (21,97)	43,23
Afirm. - Foco	77 (31,43)	191 (25,37)	40,31
Afirm. - Neg.	137 (50,18)	465 (60,00)	29,46
Afirm. - Int. (Qu)	131 (53,47)	318 (42,91)	41,19
Afirm. - Int. (s/n)	130 (43,19)	337 (31,55)	38,58
Int. (Dúvida) - Foco	115 (46,94)	452 (50,06)	25,44
Int. (Dúvida) - Negativa	136 (49,82)	393 (42,49)	34,61
Int. (Dúvida) - Int. (Qu)	113 (46,12)	311 (34,90)	36,33
Int. (Dúvida) - Int. (s/n)	143 (47,51)	454 (37,27)	31,50
Foco - Neg.	123 (50,20)	403 (50,12)	30,52
Foco - Int. (Qu)	104 (47,93)	434 (56,36)	23,96
Foco - Int. (s/n)	137 (50,18)	456 (41,57)	30,04
Neg. - Int. (Qu)	108 (44,08)	356 (44,95)	30,34
Neg. - Int. (s/n)	135 (44,85)	548 (48,97)	24,64
Int. (Qu) - Int. (s/n)	140 (51,28)	431 (39,72)	32,48
Intérprete B (treino) - Intérprete A (teste)			
Afirm. - Inter. (Dúvida)	137 (50,18)	345 (42,86)	39,71
Afirm. - Foco	116 (42,49)	200 (24,57)	58,00
Afirm. - Neg.	123 (50,20)	310 (45,93)	39,68
Afirm. - Inter. (Qu)	123 (50,20)	384 (48,92)	32,03
Afirm.- Int. (s/n)	137 (50,18)	362 (43,25)	37,85
Int. (Dúvida) - Foco	137 (50,18)	290 (32,26)	47,24
Int. (Dúvida) - Negativa	97 (39,59)	258 (33,95)	37,60
Int. (Dúvida) - Int. (Qu)	99 (40,41)	296 (34,02)	33,45
Int. (Dúvida) - Int. (s/n)	137 (50,18)	365 (39,59)	37,53
Foco - Negativa	117 (47,76)	358 (46,55)	32,68
Foco - Inter. (Qu)	71 (28,98)	258 (29,35)	27,52
Foco - Inter. (s/n)	137 (50,18)	307 (32,98)	44,63
Neg. - Inter. (Qu)	109 (50,23)	352 (47,57)	30,97
Neg. - Inter. (s/n)	123 (50,200)	330 (41,67)	37,27
Inter. (Qu) - Inter. (s/n)	80 (32,65)	241 (26,72)	33,20

Fonte: Maria Eduarda de Araújo Cardoso, 2018

Multilayer Perceptron - Experimento #1

A tabela 41 apresenta a média das acurácias treinadas com a combinação de todas EFGs com os nove pontos e também apresenta os erros de borda e gerais para esses testes. São apresentados os testes dependentes e independentes do intérprete.

Tabela 41 – Acurácia de classificadores no problema de sete classes e erros de borda: seis EFGs mais a expressão neutra. Análise dependente do intérprete e independente do intérprete. Abreviações: μ - média; σ - desvio padrão (σ).

	μ	σ	max	min	# erros de borda	# erros gerais	% erros na borda
A	0,87	0,01	0,88	0,86	179 (23,04)	217 (8,75)	82,49
B	0,70	0,01	0,72	0,68	805 (100,00)	2767 (100,00)	29,09
A (treino) - B (teste)	0,53	0,03	0,58	0,49	375 (46,58)	917 (33,14)	40,89
B (treino) - A (teste)	0,53	0,06	0,64	0,44	777 (100,00)	2480 (100,00)	31,33

Fonte: Maria Eduarda de Araújo Cardoso, 2018

O intérprete B para ambos experimentos não consegue alcançar bons resultados, apresenta uma melhora quando treinado com nove pontos, mas comete mais erros gerais atingindo 100% para testes dependentes e independentes.