

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CÂMPUS CORNÉLIO PROCÓPIO
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

ANDRÉ LUIZ PRZYBYSZ

**CLASSIFICAÇÃO AUTOMÁTICA DE EMOÇÕES EM MÚSICAS
LATINAS UTILIZANDO DIFERENTES FONTES DE INFORMAÇÃO**

DISSERTAÇÃO – MESTRADO

CORNÉLIO PROCÓPIO

2016

ANDRÉ LUIZ PRZYBYSZ

**CLASSIFICAÇÃO AUTOMÁTICA DE EMOÇÕES EM MÚSICAS
LATINAS UTILIZANDO DIFERENTES FONTES DE INFORMAÇÃO**

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Universidade Tecnológica Federal do Paraná – UTFPR como requisito parcial para a obtenção do título de “Mestre em Informática”.

Orientador: Prof. Dr. Carlos N. Silla Jr.

CORNÉLIO PROCÓPIO

2016

Dados Internacionais de Catalogação na Publicação

- P973 Przybysz, André Luiz
Classificação automática de emoções em músicas latinas utilizando diferentes fontes de informação / André Luiz Przybysz. – 2016.
91 f. : il. ; 30 cm
- Orientador: Carlos Nascimento Silla Junior.
Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-graduação em Informática. Cornélio Procópio, 2016.
Bibliografia: p. 85-89.
1. Classificação - Multimeios. 2. Emoções. 3. Códigos telegráficos e cifrados. I. Silla Junior, Carlos Nascimento, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Informática. IV. Título.

CDD (22. ed.) 004



Título da Dissertação Nº 18:

“Classificação Automática de Emoções em Músicas Latinas Utilizando Diferentes Fontes de Informação”.

por

André Luiz Przybysz

Orientador: **Prof. Dr. Carlos Nascimento Silla Junior**

Esta dissertação foi apresentada como requisito parcial à obtenção do grau de MESTRE EM INFORMÁTICA – Área de Concentração: Computação Aplicada, pelo Programa de Pós-Graduação em Informática – PPGI – da Universidade Tecnológica Federal do Paraná – UTFPR – Câmpus Cornélio Procópio, às 14h do dia 17 de março de 2016. O trabalho foi APROVADO pela Banca Examinadora, composta pelos professores:

Prof. Dr. Carlos Nascimento Silla Junior
(Presidente – UTFPR-CP)

Prof. Dr. Estevam Rafael Hruschka Junior
(UFSCAR-SP)

Prof. Dr. Fábio Parra Furlanete
(UEL-PR)

Prof. Dr. Yandre Maldonado e Gomes da Costa
(UEM-PR)

Visto da coordenação:

André Takeshi Endo

Coordenador do Programa de Pós-Graduação em Informática
UTFPR Câmpus Cornélio Procópio

A Folha de Aprovação assinada encontra-se na Coordenação do Programa.

À minha família.

AGRADECIMENTOS

A Deus, por ser o meu guia em todos os momentos da vida.

À minha mãe, pelo seu amor incalculável, pelo seu exemplo de sabedoria e por sempre acreditar em mim.

Ao meu pai, por me ensinar que, independente das adversidades, sempre teremos um caminho a seguir, assim como maiores forem as adversidades, maiores também serão as nossas conquistas.

À minha esposa, por nunca deixar de acreditar nas promessas de Deus para as nossas vidas. Por ser paciente nas horas difíceis e por sempre estar ao meu lado.

Ao meu filho Samuel, por ser a minha maior motivação para continuar lutando, por ser a razão dos meus melhores sorrisos, por me dar a oportunidade de ensiná-lo o que realmente tem valor nessa vida.

Ao meu orientador Carlos N. Silla Jr., por me dar a oportunidade de dar a volta por cima, por todas as valiosas críticas construtivas e por acreditar em mim.

À UTFPR-CP, pelas oportunidades nesses treze anos de instituição.

“Tudo tem o seu tempo determinado, e há tempo para todo o propósito
debaixo do céu.”

(Eclesiastes 3:1)

RESUMO

PRZYBYSZ, André Luiz. CLASSIFICAÇÃO AUTOMÁTICA DE EMOÇÕES EM MÚSICAS LATINAS UTILIZANDO DIFERENTES FONTES DE INFORMAÇÃO. 92 f. Dissertação – Mestrado – Programa de Pós-graduação em Informática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2016.

Com a quantidade de músicas disponíveis *on-line*, tem ocorrido uma crescente expansão nas pesquisas de informação musical e nas investigações de recuperação para sistemas automatizados. O campo de *Music Information Retrieval* (MIR) olha para vários aspectos relacionados à como organizar, categorizar e acessar coleções musicais. O desenvolvimento de novos métodos e a criação de novas representações musicais podem contribuir na precisão dos classificadores para o reconhecimento das emoções, uma vez que estes estão entre os maiores desafios da área de *Music Emotion Recognition* (MER). Esse trabalho investiga, implementa e combina três diferentes fontes de informações (cifras, áudio e letras) para a classificação automática de emoções em músicas. A metodologia de trabalho é descrita por meio das seguintes atividades: criação da base de dados *Multimodal Latin Music Mood Database* (MLMMD), pré-processamento dos dados, mineração e combinação dos dados. Por meio dos procedimentos foi realizada uma análise dos diferentes resultados utilizando as diferentes fontes de informação de forma individual e combinada. Primeiramente, foi possível observar que o método combinatório multimodal *early fusion* foi melhor do que os demais (*no fusion* e *late fusion*). Segundo, o classificador *Support Vector Machine* (SVM) mostrou uma média melhor que os demais.

Palavras-chave: Classificação automática. Emoções. Cifras. Letras. Áudio. Unimodal. Multimodal.

ABSTRACT

PRZYBYSZ, André Luiz. EMOTIONS AUTOMATIC CLASSIFICATION LATIN SONGS USING DIFFERENT SOURCES OF INFORMATION. 92 f. Dissertação – Mestrado – Programa de Pós-graduação em Informática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2016.

With the growing amount of music available online, there has been an increasing expansion in research of musical information and the recovery investigations to automated systems. The Music Information Retrieval (MIR) field looks at various aspects related to how to organize, categorize, and access music collections. The development of new methods and the creation of new musical representations can contribute to the accuracy of classifiers for recognition of emotions, since these are among the greatest challenges in the area of Music Emotion Recognition (MER). This work investigates, implements and combines three different sources of information (cifras, audio and lyrics) for automatic emotion classification in songs. The following activities have been used to develop this work: database definition Multimodal Latin Music Mood Database (MLMMD), preprocessing of the different types of data, mining and combination of the different types of data. Through the procedures applied it was possible to carry out an analysis of the different results. First, it was observed that the multimodal early fusion method was better than the others other approaches (no fusion and late fusion). Second, the Support Vector Machine (SVM) showed an overall average better than the other classifiers.

Keywords: Automatic classification. Emotions. Cifras. Lyrics. Audio. Unimodal. Multimodal.

LISTA DE FIGURAS

FIGURA 1	– Modelo de emoções proposto por Hevner.	19
FIGURA 2	– Representação bidimensional Valence-Arousal.	21
FIGURA 3	– Exemplo da notação neumática.	29
FIGURA 4	– Trecho de uma música com os símbolos criados a partir de Guido d'Arezzo.	31
FIGURA 5	– Trecho da música Xote dos Milagres - Falamansa.	33
FIGURA 6	– Diagrama das etapas SSD para a extração das características do áudio. .	36
FIGURA 7	– Regra do maior valor.	38
FIGURA 8	– Regra do produto.	39
FIGURA 9	– Regra da soma.	39
FIGURA 10	– Esquema geral do Early Fusion proposto.	40
FIGURA 11	– Esquema geral do Late Fusion proposto.	41

LISTA DE TABELAS

TABELA 1	– Exemplo de etiquetas de emoções do guia All Music.	20
TABELA 2	– Visão geral das características acústicas na classificação de emoções. ..	23
TABELA 3	– Sumário dos trabalhos relacionados.	28
TABELA 4	– Notação alfabética utilizando letras latinas.	29
TABELA 5	– Exemplos de opções de notações em cifras.	32
TABELA 6	– Nomenclatura de acordes e cifras utilizados nesse trabalho.	32
TABELA 7	– Exemplo do descritor presença e ausência dos acordes.	33
TABELA 8	– Exemplo do descritor tom dos acordes.	34
TABELA 9	– Exemplo do descritor quantidade de repetições dos acordes.	34
TABELA 10	– Exemplo do descritor frequência dos acordes.	35
TABELA 11	– Exemplo do descritor transição dos acordes.	35
TABELA 12	– <i>Multimodal Latin Music Mood Database</i> com três emoções.	42
TABELA 13	– <i>Multimodal Latin Music Mood Database</i> com Seis Emoções.	42
TABELA 14	– Resultados sem combinações para três emoções.	45
TABELA 15	– Sumário dos resultados utilizando o método <i>no fusion</i> com três emoções.	49
TABELA 16	– Melhores resultados obtidos utilizando o método <i>no fusion</i> para representações individuais com três emoções.	50
TABELA 17	– Resultados sem combinações para seis emoções.	50
TABELA 18	– Sumário dos resultados utilizando o método <i>no fusion</i> com seis emoções.	55
TABELA 19	– Melhores resultados obtidos utilizando o método <i>no fusion</i> para representações individuais com seis emoções.	56
TABELA 20	– Resultados das combinações unimodais para três emoções.	57
TABELA 21	– Sumário dos resultados Unimodais com três emoções.	61
TABELA 22	– Resultados das combinações unimodais para três emoções.	61
TABELA 23	– Sumário dos resultados unimodais com seis emoções.	65
TABELA 24	– Resultados das combinações multimodais para três emoções.	67
TABELA 25	– Sumário dos resultados multimodais com três emoções.	73
TABELA 26	– Resultados das combinações multimodais para seis emoções.	74
TABELA 27	– Sumário dos resultados multimodais com seis emoções.	82

LISTA DE SIGLAS

1-NN	<i>One Nearest Neighbors.</i>
3-NN	<i>Three Nearest Neighbors.</i>
5-NN	<i>Five Nearest Neighbors.</i>
ANEW	<i>Affective norms for English words.</i>
ARFF	<i>Attribute-Relation File Format.</i>
BME	<i>Bibliotecas de Músicas Emergentes.</i>
BOW	<i>Bag of Words.</i>
F-M	<i>F-Measure</i>
GMM	<i>Gaussian Mixture Model</i>
IEEE	<i>Institute of Electrical and Electronics Engineers.</i>
k-NN	<i>k-Nearest Neighbors.</i>
LMCD	<i>Latin Music Cifras Database.</i>
LMCMD	<i>Latin Music Cifras Mood Database.</i>
LMD	<i>Latin Music Database.</i>
LMMD	<i>Latin Music Mood Database.</i>
MLMMD	<i>Multimodal Latin Music Mood Database.</i>
LSA	<i>Latent Semantic Analysis.</i>
MER	<i>Music Emotion Recognition.</i>
MIDI	<i>Musical Instrument Digital Interface.</i>
MIR	<i>Music Information Retrieval.</i>
MIREX	<i>Music Information Retrieval Evaluation Exchange.</i>
PAD	<i>Pleasure, Arousal and Dominance.</i>
PANAS	<i>Positive and Negative Affect Schedule.</i>
PLN	<i>Processamento de Linguagem Natural.</i>
PLSA	<i>Probabilistic Latent Semantic Analysis.</i>
SSD	<i>Statistical Spectrum Descriptor.</i>
SMC	<i>Systems, Man, and Cybernetics.</i>
SMO	<i>Sequential Minimal Optimization.</i>
SVM	<i>Support Vector Machine.</i>
TAE	<i>Teoria do Aprendizado Estatístico.</i>
TF-IDF	<i>Term Frequency-Inverse Document Frequency.</i>
V-A	<i>Valence-Arousal.</i>
VSM	<i>Vector Space Model.</i>
WEKA	<i>Waikato Environment for Knowledge Analysis.</i>

SUMÁRIO

1 INTRODUÇÃO	14
1.1 CONTEXTUALIZAÇÃO	14
1.2 MOTIVAÇÃO	15
1.3 OBJETIVOS	15
1.3.1 Objetivo Geral	15
1.3.2 Objetivos Específicos	16
1.4 HIPÓTESE	16
1.5 ORGANIZAÇÃO DA DISSERTAÇÃO	16
2 CLASSIFICAÇÃO DE EMOÇÕES EM MÚSICA	18
2.1 PSICOLOGIA DA EMOÇÃO	18
2.2 REPRESENTAÇÕES DAS EMOÇÕES	19
2.3 <i>FRAMEWORK</i> PARA RECONHECIMENTO DE EMOÇÕES	20
2.4 TRABALHOS RELACIONADOS	22
2.4.1 Utilizando Áudio	22
2.4.2 Utilizando Letras	23
2.4.3 Utilizando Informações Culturais	24
2.4.4 Combinando Áudio e Imagens	25
2.4.5 Combinando Áudio e Letras das Músicas	25
2.4.6 Combinando Áudio, MIDI e Letras	27
2.4.7 Combinando Áudio e Tags	27
2.4.8 Sumário dos Trabalhos Relacionados	28
3 METODOLOGIA	29
3.1 EXTRAÇÃO DE CARACTERÍSTICAS	29
3.1.1 Através das Cifras	29
3.1.2 Presença e Ausência dos Acordes	32
3.1.3 Tom dos Acordes	33
3.1.4 Quantidade de Repetições dos Acordes	34
3.1.5 Frequência dos Acordes	34
3.1.6 Transição dos Acordes	35
3.1.7 Através do Áudio	35
3.1.8 Através das Letras	36
3.1.9 Remoção de Stopwords	36
3.1.10 N-Grams	37
3.1.11 Stemming	37
3.2 CLASSIFICAÇÃO	37
3.3 COMBINAÇÃO DE CLASSIFICADORES	38
3.4 PROTOCOLO EXPERIMENTAL	40
3.4.1 Base de Dados	40
3.4.2 Pré-Processamento dos Dados	42
3.4.3 Vetores de Características	42
3.4.4 Mineração dos Dados	43
3.4.5 Métricas de Avaliação	43
4 EXPERIMENTOS SEM COMBINAÇÕES (NO FUSION)	45
4.1 RESULTADOS UTILIZANDO <i>NO FUSION</i> COM TRÊS EMOÇÕES	45
4.1.1 Sumário dos Resultados <i>No Fusion</i> com Três Emoções	49

4.2	RESULTADOS UTILIZANDO <i>NO FUSION</i> COM SEIS EMOÇÕES	50
4.2.1	Sumário dos Resultados <i>No Fusion</i> com Seis Emoções	54
4.3	COMPARAÇÕES EXPERIMENTAIS	56
5	EXPERIMENTOS COM COMBINAÇÕES UNIMODAIS	57
5.1	RESULTADOS UNIMODAIS COM TRÊS EMOÇÕES	57
5.1.1	Sumário dos Resultados Unimodais com Três Emoções	60
5.2	RESULTADOS UNIMODAIS COM SEIS EMOÇÕES	60
5.2.1	Sumário dos Resultados Unimodais com Seis Emoções	65
5.3	COMPARAÇÕES EXPERIMENTAIS	66
6	EXPERIMENTOS COM COMBINAÇÕES MULTIMODAIS	67
6.1	RESULTADOS MULTIMODAIS COM TRÊS EMOÇÕES	67
6.1.1	Sumário dos Resultados Multimodais com Três Emoções	72
6.2	RESULTADOS MULTIMODAIS COM SEIS EMOÇÕES	73
6.2.1	Sumário dos Resultados Multimodais com Seis Emoções	81
6.3	COMPARAÇÕES EXPERIMENTAIS	82
7	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	84
	REFERÊNCIAS	86
	ANEXO A - LISTAS DE <i>STOPWORDS</i>	91

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Com o aumento de grandes bases musicais, como por exemplo *Allmusic.com*, *Last.fm* (HENNING; REICHEL, 2008), *The Million Song Dataset* (BERTIN-MAHIEUX et al., 2011), dentre outras, e do fácil acesso à música digital ao longo das últimas décadas, tem ocorrido uma rápida expansão de informação musical e de investigações de recuperação para sistemas automatizados.

O campo de *Music Information Retrieval* (MIR) (TYPKE et al., 2005) olha para vários aspectos relacionados à como organizar, categorizar e acessar coleções musicais. Nesse contexto, a tarefa de *Music Emotion Recognition* (MER) (HU et al., 2008) vem recebendo cada vez mais atenção. Pesquisas já realizadas utilizaram diferentes tipos de informação como o áudio das músicas (LU et al., 2006; LAURIER et al., 2007; EEROLA et al., 2009), as letras das músicas (HU et al., 2009; YANG; LEE, 2009; ZANEN; KANTERS, 2010; MIHALCEA; STRAPPARAVA, 2012), ou informações culturais (BISCHOFF et al., 2009; YANG; HU, 2012).

Alguns autores investigam a combinação de diferentes fontes de informação como o áudio e letras (HU et al., 2009; HU; DOWNIE, 2010b; MIHALCEA; STRAPPARAVA, 2012; LAURIER et al., 2008; HU; DOWNIE, 2010a), áudio, letras e representação simbólica (PANDA et al., 2013), a combinação de áudio e tags (TURNBULL et al., 2009).

Sistemas computacionais para o reconhecimento de emoções por meio da música podem basear-se em um determinado modelo de emoção (OLIVEIRA; CARDOSO, 2010); no entanto, tais representações continuam a ser um tópico das pesquisas psicológicas (KIM et al., 2010).

Esse trabalho investiga, implementa e combina três diferentes fontes de informação para a classificação automática de emoções. As fontes de informações utilizadas são: o conteúdo do sinal de áudio, as letras das músicas e suas Cifras que contêm informações simbólicas sobre a melodia das músicas.

1.2 MOTIVAÇÃO

Assim como as demais áreas do reconhecimento de padrões, conhecer a emoção musical continua a ser um problema desafiador principalmente devido às ambiguidades inerentes das emoções humanas (HU et al., 2008). Estudos como o de (KIM et al., 2010) indicam que sistemas de classificação automática de emoções têm se tornado uma área crescente de pesquisa.

Por intermédio da revisão bibliográfica realizada nesse trabalho foi possível perceber que os sistemas para reconhecimento de emoções em músicas utilizam normalmente uma única fonte de informação. Essa representação pode ser baseada no conteúdo do sinal de áudio (LU et al., 2006; LAURIER et al., 2007; EEROLA et al., 2009; SONG et al., 2012), mediante às informações presentes nas letras das músicas (YANG; LEE, 2009; ZANEN; KANTERS, 2010; MIHALCEA; STRAPPARAVA, 2012; KUMAR; MINZ, 2013) ou baseada em informações culturais (BISCHOFF et al., 2009; YANG; HU, 2012). No entanto, combinações multimodais (utilizando diferentes fontes de informação) ainda são pouco exploradas, sendo que os trabalhos existentes na literatura utilizam a combinação do áudio com as letras das músicas (LAURIER et al., 2008; HU et al., 2009; HU; DOWNIE, 2010b, 2010a).

A principal contribuição desse trabalho é explorar o uso de diferentes fontes de informação e suas combinações para a tarefa de classificação automática de emoções em músicas.

Em relação à classificação automática de emoções em músicas, as lacunas hoje existentes na comunidade científica ocorrem em todos os níveis do processo. Ou seja, desde as representações emocionais, os métodos de anotações, bem como os métodos para caracterizar a seleção e o aprendizado de máquina e os métodos de combinações. Essas lacunas existentes justificam o trabalho e serão elencados nos objetivos propostos.

1.3 OBJETIVOS

1.3.1 OBJETIVO GERAL

O objetivo desse trabalho é investigar, implementar e combinar três fontes de informações (cifras, áudio e letras) para a classificação automática de emoções em músicas latinas. No referencial bibliográfico, esse tema também pode ser encontrado como reconhecimento de emoções musicais multimodais (DUNKER et al., 2008; PANDA et al., 2013), ou combinações híbridas para o reconhecimento de emoções (BISCHOFF et al., 2009).

1.3.2 OBJETIVOS ESPECÍFICOS

Os objetivos específicos desse trabalho são:

- criar uma versão Multimodal da *Latin Music Mood Database* (LMMD) (SANTOS; SILLA JR., 2015);
- combinar as diferentes fontes de informação (cifras, áudio e letras) utilizando métodos de combinação unimodais e multimodais;
- avaliar os resultados obtidos pelos diferentes métodos na base criada.

1.4 HIPÓTESE

Acredita-se que os métodos combinatórios (*early fusion* e *late fusion*) tendem a promover uma classificação mais precisa das emoções, considerando diferentes fontes de informação musical e diferentes modelos emocionais.

1.5 ORGANIZAÇÃO DA DISSERTAÇÃO

Esse trabalho é dividido em 7 capítulos, além do capítulo inicial (capítulo 1), de introdução, a qual esclarece a contextualização, a motivação e os objetivos propostos, bem como são apresentadas os demais capítulos:

- Capítulo 2: é apresentada a fundamentação teórica, que envolve a apresentação de alguns dos modelos de emoções existentes na literatura. Além disso, são apresentados os trabalhos relacionados, os quais por sua vez auxiliaram na identificação dos modelos e métodos propostos. Esses ajudaram a fundamentar o estado da arte na classificação automática de emoções por meio da música;
- Capítulo 3: é apresentada a metodologia de pesquisa, além dos protocolos experimentais que foram utilizados nos experimentos;
- Capítulo 4: são exibidos os resultados experimentais, por intermédio do método sem fusão *no fusion*, para três e seis emoções.
- Capítulo 5: são mostrados os resultados experimentais unimodais, mediante os métodos de combinação (*early fusion* e *late fusion*), também para três e seis emoções.

- Capítulo 6: são expostos os resultados experimentais multimodais, por intermédio dos métodos de combinação (*early fusion* e *late fusion*), também para três e seis emoções.
- Capítulo 7: são apresentadas as considerações finais e os trabalhos futuros.

2 CLASSIFICAÇÃO DE EMOÇÕES EM MÚSICA

Nesse capítulo, são apresentadas a fundamentação teórica desse trabalho na área de classificação de emoções em música e os trabalhos relacionados da área de pesquisa de Recuperação de Informações Musicais.

2.1 PSICOLOGIA DA EMOÇÃO

Ao longo do último século, tem havido vários desenvolvimentos importantes, os quais abrangem múltiplas abordagens para a quantificação e qualificação das emoções ligadas à música (HEVNER, 1936; KATAYOSE et al., 1988; DODDS; DANFORTH, 2010; BRINKER et al., 2012).

Ao realizar quaisquer medições de emoções, deve-se considerar a fonte da emoção que está sendo medida. Uma investigação sobre essas medições (KIM et al., 2010) indica uma importante distinção entre a própria percepção da emoção expressa pela música e a emoção induzida pela música.

(KIM et al., 2010) descreve os primeiros estudos realizados baseados na resposta psicológica.

“Os primeiros estudos da resposta psicológica consideraram o peso emocional da música como foco de estímulos, estes respondem afetivamente encontrando a música também para ser sensível ao ambiente e aos contextos de escuta. Ao estudar as diferenças entre percepções e induções de emoções, demonstraram que tanto podem estas estar sujeitas apenas ao contexto social, bem como estão sujeitas a experiência da audição (como público e local), mas também a uma motivação pessoal (ou seja, utilização da música para o relaxamento, estimulação, etc.)” (KIM et al., 2010)

Outros estudos também abordam o tema (OLIVEIRA; CARDOSO, 2010; DODDS; DANFORTH, 2010). Em (OLIVEIRA; CARDOSO, 2010) são descritos os desafios no reconhecimento automático de emoções e as pesquisas conduzidas pelos autores que descrevem todos os estágios do reconhecimento automático de emoções. Ainda em (OLIVEIRA; CARDOSO, 2010) é apresentado o modelo de representação de emoções *valence-arousal* aplicado às experimentações realizadas. Em (DODDS; DANFORTH, 2010) é descrita a importância de se quantificar os estados emocionais em diversas áreas sociais. São utilizadas as letras das músicas para identificar os níveis de felicidade. São analisadas diferentes tendências, nas diferentes épocas (1960-1990), como épocas de crise ou de ascensão (2005-2009). Nesse estudo,

assim como o realizado por (OLIVEIRA; CARDOSO, 2010) é utilizado o modelo de representação de emoções *valence-arousal*. Ainda em (DODDS; DANFORTH, 2010) os autores consideram que, caso as interações sociais e as experiências pessoais continuarem a crescer, esses fornecerão conjuntos de dados cada vez mais ricos, levarão a novas oportunidades, bem como a necessidades de novas investigações científicas.

2.2 REPRESENTAÇÕES DAS EMOÇÕES

Sistemas de recuperação de informações musicais tendem a usar descrições categóricas ou modelos paramétricos de emoções para a classificação ou reconhecimento. Cada representação é suportada por um grande corpo de apoio à investigação da psicologia, alguns desses estudos podem ser verificados no capítulo 3 (Trabalhos Relacionados).

Abordagens categóricas envolvem a descoberta e a organização de um conjunto de descritores emocionais (tags) com base em sua relevância e um pouco da música em questão.

Um dos primeiros estudos que relatam um conjunto de descritores emocionais foi realizado por Hevner, publicado em 1936, em que foram utilizados inicialmente 66 rótulos, os quais foram distribuídos em oito grupos. A figura 1 apresenta esses descritores emocionais.

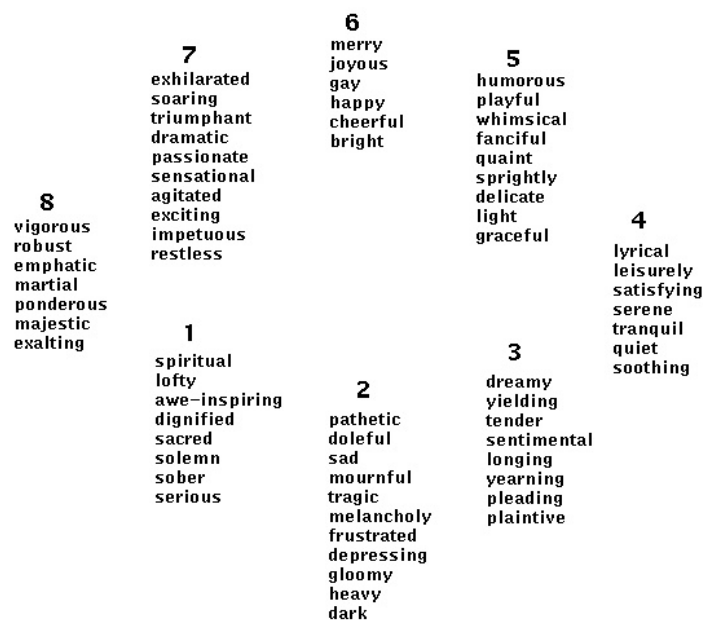


Figura 1: Modelo de emoções proposto por Hevner.

Fonte: (HEVNER, 1936).

As recentes avaliações para a classificação automática de emoções em músicas, por meio da *Music Information Retrieval Evaluation Exchange* (Mirex) categorizaram as músicas

em um dos cinco grupos de emoções. Esses grupos de emoções são apresentados na tabela 1. As cinco categorias são obtidas mediante a realização de agrupamentos de uma matriz de co-ocorrência de rótulos de emoções para músicas populares do guia *Allmusic.com*.

Tabela 1: Exemplo de etiquetas de emoções do guia All Music.

CLUSTERS	ADJETIVOS DE HUMOR
Grupo 1	apaixonado (<i>passionate</i>), vibrante (<i>rousing</i>), confiante (<i>confident</i>), violento (<i>boisterous</i>), turbulento (<i>rowdy</i>).
Grupo 2	brincalhão (<i>rollicking</i>), alegre (<i>cheerful</i>), divertido (<i>fun</i>), doce (<i>sweet</i>), amável (<i>amiable</i>)/boa índole (<i>good natured</i>).
Grupo 3	letrado (<i>literate</i>), comovente (<i>poignant</i>), saudoso (<i>wistful</i>), agridoce (<i>bittersweet</i>), outonal (<i>autumnal</i>).
Grupo 4	engraçado (<i>humorous</i>), bobo (<i>silly</i>), extrovertido (<i>campy</i>), peculiar (<i>quirky</i>), lunático (<i>whimsical</i>), espirituoso (<i>witty</i>), irônico (<i>wry</i>).
Grupo 5	agressivo (<i>aggressive</i>), impetuoso (<i>fiery</i>), tenso/ansioso (<i>tense/anxious</i>), intenso (<i>intense</i>), volátil (<i>volatile</i>), instintivo (<i>visceral</i>).

Fonte: Adaptado de (KIM et al., 2010).

Outras pesquisas sugerem que as emoções podem ser escaladas e medidas por descritores contínuos ou por métricas simples multidimensionais. Em (THAYER, 1989) foram estudadas as dimensões da excitação (arousal) e estabeleceram uma notação sobre as quais conjuntos de descritores de emoções podem ser organizados em modelos de baixa dimensionalidade, Valence-Arousal (V-A).

A figura 2 apresenta a representação bidimensional V-A, (THAYER, 1989; BRINKER et al., 2012) descrevem que as emoções existem em um plano ao longo de eixos independentes de excitação (*arousal*), variando do maior para o menor, e valência entre positivo-negativo (uma avaliação de polaridade).

Uma ferramenta de avaliação relacionada, mas categórica para a emoção (afeto) é a *Positive and Negative Affect Schedule* (PANAS) (WATSON; CLARK, 1999), a qual afirma que todas as emoções discretas e seus rótulos associados existem como incidências de um efeito positivo ou negativo semelhante a valência. Nesse caso, no entanto, positivo e negativo são tratados como categorias separadas ao contrário da abordagem paramétrica de V-A.

2.3 FRAMEWORK PARA RECONHECIMENTO DE EMOÇÕES

O reconhecimento das emoções pode ser visto como um problema de classificação de múltiplas classes, considerando as várias partes de uma música, ou de uma única classe, considerando a música inteira. Exemplos: anotações por seções de uma música (refrão, verso),

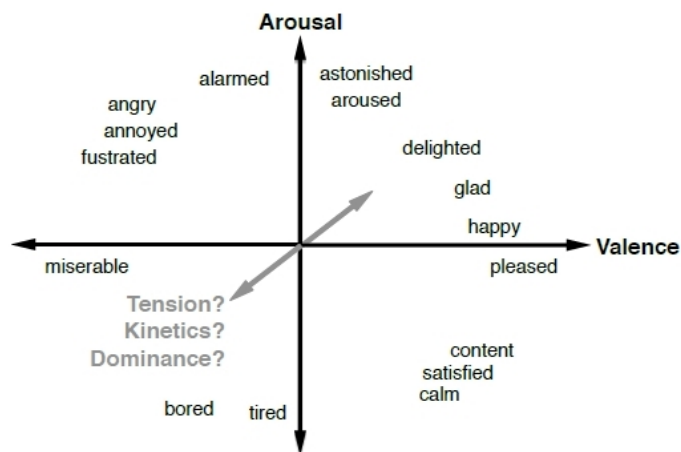


Figura 2: Representação bidimensional Valence-Arousal.

Fonte: (KIM et al., 2010).

por um clipe de tamanho fixo (EX: 40 segundos de uma música), analisando a música como um todo.

As representações das emoções podem ser obtidas a partir de diversas formas. Assim, pode-se:

- solicitar aos ouvintes humanos que avaliem a pertinência de uma emoção para uma determinada música;
- analisar formas contextuais de meta-dados em forma de texto, possibilitando incluir *text-mining*, web-documentos (por exemplo, biografias de artistas, comentários de álbuns) (HU; DOWNIE, 2010a) ou uma grande coleção de rótulos sociais (a que se refere a uma nuvem de tags), e a análise de letras (como análise de sentimento);
- também analisar o conteúdo do áudio, usando o processamento de sinal e o aprendizado de máquina supervisionado para anotar automaticamente as emoções em músicas ou partes (LU et al., 2006; LAURIER et al., 2007; EEROLA et al., 2009). Com base em métodos de conteúdos, pode-se utilizar para análise outras formas relacionadas aos dados multimídia, como vídeos de músicas e até mesmo fotografias (DUNKER et al., 2008). Além disso, várias fontes de dados, como por exemplo letras e áudio, podem ser combinadas para determinar o conteúdo emocional da música (LAURIER et al., 2008; HU et al., 2009; HU; DOWNIE, 2010b, 2010a).

2.4 TRABALHOS RELACIONADOS

A tarefa de reconhecimento das emoções a partir da música vem ganhando destaque mesmo antes dos sistemas de recuperações de informações musicais (HEVNER, 1936). Diversos estudos já foram realizados sobre o assunto e diante desses aspectos, (KIM et al., 2010) resumiram três fatores principais desse crescimento:

- referenciar a música como uma linguagem de emoções, definindo a capacidade do ser humano em categorizar a música em termos de suas associações emocionais;
- os desafios no desenvolvimento de sistemas computacionais para o reconhecimento das emoções baseados em modelos de emoções;
- as lacunas ainda sem respostas relacionadas aos modelos de emoções e aos métodos de classificação em sistemas computacionais para o reconhecimento das emoções. Esse último também foi abordado por (YANG; CHEN, 2012; STURM, 2013, 2014).

Os diferentes trabalhos utilizam diversas fontes de informação. Essas fontes podem ser utilizadas de forma isolada ou combinada: áudio, letras, informações culturais, imagens e tags.

2.4.1 UTILIZANDO ÁUDIO

Estudos (LU et al., 2006; EEROLA et al., 2009) mostram a importância da utilização do áudio na detecção das emoções. Em (LU et al., 2006) foi proposto um *framework* hierárquico para automatizar a tarefa de detecção de emoções por meio de informações acústicas de uma música. Segundo (LU et al., 2006), a estrutura hierárquica possui a vantagem de enfatizar as características de um clipe de música. Foram analisados a partir do áudio três conjuntos de recursos: intensidade, timbre e ritmo. As características relacionadas à intensidade são representadas pela energia de cada sub-banda, já as características representadas pelo timbre foram compostas por formas espectrais e por características retiradas dos contrastes espectrais. As características do ritmo indicam que três aspectos os quais estão intimamente relacionadas à emoção do indivíduo: a força, a regularidade do ritmo e o próprio ritmo. Já em (EEROLA et al., 2009) foram desenvolvidos recursos para a extração de informações de áudio para a emoção, denominado *MIRtoolbox*. Essas características são agregadas a partir de uma vasta gama de domínios, incluindo a dinâmica, o timbre, a harmonia, o ritmo e as articulações. Os autores ainda citam que outras abordagens adotaram características generalizadas de extração,

por meio de compilações de vários conjuntos de recursos e empregando redução nas técnicas de dimensionalidade.

(KIM et al., 2010) fornece uma análise desses recursos nas emoções, citando que podem ser representados a partir do tempo, do timbre, da harmonia e da sonoridade. Uma visão geral das características mais comuns utilizando os sinais de áudio para o reconhecimento de emoções podem ser verificadas na tabela 2.

Tabela 2: Visão geral das características acústicas na classificação de emoções.

TIPO	CARACTERÍSTICAS
Dinâmicos	Potência RMS.
Timbre	MFCCs, forma espectral, contraste espectral.
Harmonia	Rugosidade, mudança harmônica, tonalidade.
Registro	Chromagram, centroide croma e desvio.
Rítmo	Força Ritmo, regularidade, tempo, batida.
Articulações	Densidade do evento, inclinação de ataque, tempo de ataque.

Fonte: Adaptado de (KIM et al., 2010).

2.4.2 UTILIZANDO LETRAS

Estudos (YANG; LEE, 2009; ZAAANEN; KANTERS, 2010; MIHALCEA; STRAPPA-RAVA, 2012) também mostram a importância das letras na detecção das emoções.

Em (YANG; LEE, 2009) foi citado que grandes repositórios de músicas *on-line* têm sido criados por empresas, no entanto esses não apresentam conteúdos baseados em métodos de recuperação. Uma exceção é o site *Allmusic.com* que oferece uma navegação por emoção musical, o qual utilizou especialistas humanos para classificar milhares de músicas em 183 emoções. Diante disso, (YANG; LEE, 2009) utilizaram técnicas de aprendizado de máquina ao invés de especialistas humanos para extrair emoções em músicas. A classificação dos autores é baseada em um modelo psicológico de emoção que é estendido a 23 categorias de emoções específicas. Os resultados (65,00%) mostram que a mineração em letras de emoções em músicas são promissoras.

Em (ZAAANEN; KANTERS, 2010) foram apresentados resultados da investigação em utilizar partes de letras de músicas em um sistema de classificação automática de emoções. Usando um conjunto de letras, correspondentes a uma emoção rotulada por um usuário, os autores construíram classificadores para categorizar letras de cânticos em emoções. Ao comparar o desempenho de diferentes *frameworks* (*arousal*, *valence*, *Thayer*, *fine-grained*) de emoções, (ZAAANEN; KANTERS, 2010) examinaram em que medida a parte linguística da música revelava as informações adequadas para se atribuir uma categoria de emoção e em que aspectos de

emoções esses podem ser melhor classificados. Os resultados mostraram que as métricas orientadas a palavras fornecem uma valiosa fonte de informação para a classificação automática de emoção, com base apenas nas letras. Métricas as quais termos frequentes e valores *Term Frequency-Inverse Document Frequency* (TF-IDF) são usados para medir a relevância das palavras para as diferentes classes das emoções. (ZANEN; KANTERS, 2010) ainda citaram que as previsões sobre a valência, tensões e combinações de aspectos levaram a um desempenho semelhante (*arousal=77,23%*, *valence=76,29%*, *Thayer=75,85%*, *fine-grained=70,89%*).

2.4.3 UTILIZANDO INFORMAÇÕES CULTURAIS

No trabalho de (BISCHOFF et al., 2009; YANG; HU, 2012) foram utilizadas informações culturais na detecção de emoções.

Em (BISCHOFF et al., 2009) foi mencionado que a percepção musical está altamente ligada às emoções e o ao seu contexto cultural. O trabalho de (BISCHOFF et al., 2009) avalia as necessidades de informações dos usuários e a música indexando recursos por meio do desenvolvimento de algoritmos de classificação de músicas baseadas em emoções e temas, levando igualmente em consideração as dimensões temáticas das músicas e usam dados sociais do portal musical *Last.fm*. Para suporte aos métodos de classificações, os autores exploram ambos os recursos, áudio e anotações dos usuários realizadas de forma colaborativa, fundindo-os para melhor desempenho global. Comparações realizadas com o *Allmusic.com* mostraram que ambos os tipos de informações são complementares e deveriam ser fundidas para maior precisão nas classificações.

(YANG; HU, 2012) descreveram que na maioria dos estudos existentes na classificação de emoções musicais, foram utilizadas músicas ocidentais e citam ainda que poucas pesquisas investigaram se as categorias de emoções, características de áudio e modelos de classificação já desenvolvidos a partir dessas músicas ocidentais são aplicáveis também em músicas não ocidentais. Os autores procuram responder essa questão a partir de um estudo comparativo sobre canções inglesas e chinesas. Especificamente, um conjunto de 500 músicas chinesas foram rotuladas, utilizando uma representação de emoções existentes desenvolvidas para músicas em inglês (*Russel's Model*). Seis conjuntos de características de áudio foram utilizados na música ocidental (por exemplo, timbre e ritmo) os quais foram extraídos das músicas em chinês e em inglês. As músicas e performance de classificação das emoções baseadas nestes conjuntos de recursos foram comparados. Além disso, experiências foram realizadas para testar a generalização da classificação, modelos totalmente em inglês e músicas chinesas. Os resultados de (YANG; HU, 2012) indicam a aplicabilidade transcultural dos resultados investigados sobre a

classificação das emoções musicais.

2.4.4 COMBINANDO ÁUDIO E IMAGENS

Em (DUNKER et al., 2008) fora descrito que informações emocionais são frequentemente utilizadas em termos de pesquisas ou em propriedades de navegação dentro de arquivos multimídia, recuperação de sistemas ou *players* multimídia. A maioria dessas aplicações envolvem os usuários ou peritos para marcar objetos multimídias com anotações de emoções. Dentro da comunidade científica diferentes abordagens para conteúdos baseados em música, foto ou técnicas multimodais podem ser encontradas e com uma vasta gama de definições ou modelos de emoções utilizadas. Diante disso, o trabalho de (DUNKER et al., 2008) teve como objetivo revisar os modelos comuns de emoções, a fim de avaliar sua flexibilidade e então apresentar um *framework* genérico multimodal de classificação *Gaussian Mixture Model (GMM)*, do qual foram utilizadas várias características audiovisuais e múltiplos classificadores para apresentar uma nova música e foto, avaliadas e referenciadas em uma classificação de emoções. A classificação mediante o *framework* apresentado é a base para diferentes aplicações, como por exemplo rotulação de músicas automáticas ou *music players*. O conjunto referenciado pode ser usado para comparações de diferentes algoritmos de vários grupos de pesquisas.

2.4.5 COMBINANDO ÁUDIO E LETRAS DAS MÚSICAS

Uma combinação bastante utilizada é a junção do áudio com as letras das músicas. (LAURIER et al., 2008; HU et al., 2009; HU; DOWNIE, 2010b, 2010a) utilizaram esta combinação na detecção das emoções.

(LAURIER et al., 2008) descrevem que a emoção de uma música é expressa por meio de recursos musicais, mas uma parte relevante também parece ser transmitida pelas letras. Dessa forma, foram avaliados cada um dos fatores independentes e explorou-se a possibilidade de combinar ambos, usando processamento de linguagem natural e técnicas de MIR. (LAURIER et al., 2008) introduziram um método baseado nas diferenças entre os modelos PLN e classificadores baseados em áudio. Além disso, foi integrado esse novo método em um sistema multimodal (áudio + letras) por meio da regra do voto da maioria, a qual permitiu melhoria no desempenho global. Os autores ainda demonstraram que as letras e as informações de áudio são complementares e podem ser combinadas para aperfeiçoar um sistema de classificação. A combinação obteve um resultado de (91,71%), e os resultados sem combinações alcançaram: áudio=91,40% e letras=79,70%.

Em (HU et al., 2009) foi analisado o papel das letras de músicas e foi relatado que

por intermédio dessa abordagem, pode-se melhorar a classificação das emoções se comparadas ao áudio. Os autores propuseram um novo método, construído a partir de uma base de 5.558 músicas em 18 categorias de emoções, baseando-se em tags sociais. Estas refletiram uma perspectiva centrada no usuário. (HU et al., 2009) descrevem que são levadas em consideração palavras como: introdução, interlúdio, verso, dentre outras. O recurso da letra que obteve o melhor desempenho foi comparado com um sistema de áudio, baseado em liderança. Ao combinar fontes de letra e áudio, conjuntos de recursos híbridos foram construídos com base em três métodos de seleção de atributos. Os resultados mostraram padrões em desacordo com resultados obtidos em estudos anteriores, ou seja, características de áudio nem sempre apresentam melhores resultados que as letras e que combinando letras e áudio podem melhorar o desempenho de muitas categorias de emoções, mas não todas elas, relatam os autores. Os resultados mostraram uma taxa média de acerto de 60,00%.

Em (HU; DOWNIE, 2010b) os autores se baseiam e estendem trabalhos anteriores sobre classificações de emoções multimodais (HU et al., 2009), para analisar em profundidade os tipos de recursos que demonstram fornecer melhorias estatisticamente significativas na classificação de emoções. O conjunto de dados utilizados neste estudo corresponde à 5.296 músicas (com letras e áudios), divididas em 18 categorias de emoções derivadas das *tags* geradas pelos usuários da *Last.fm*. Essas 18 categorias mostraram uma consistência notável se comparada ao modelo popular de Russell (RUSSELL, 1980). Em sete categorias as características baseadas nas letras superaram significativamente as características espectrais de áudio. Apenas uma característica foi superada pelo áudio. Uma análise detalhada mostrou que os tipos de elementos textuais indicam uma forte e evidente associação semântica entre os termos extraídos e as categorias.

De acordo com (HU; DOWNIE, 2010a) emoções são tipos de metadados e de pontos de acesso em músicas de Bibliotecas de Músicas Emergentes (BME) em repositórios de músicas online. (HU; DOWNIE, 2010a) apresentam uma investigação abrangente sobre a utilidade das letras de música na classificação de emoções por meio da avaliação e comparação de uma ampla gama de recursos, incluindo as características linguísticas textuais das letras, assim como também as estilísticas do texto. Em seguida, foram combinadas as melhores características das letras com os atributos extraídos do áudio, usando para isso a fusão de dois métodos. Os resultados apresentados mostraram que a combinação das letras e do áudio superaram os sistemas que utilizaram apenas as características do áudio. Adicionalmente, examinando as curvas de aprendizagem, os autores mostraram que em um sistema híbrido (áudio + letra) são necessárias menos amostras de treinamento para alcançar resultados iguais ou melhores do que a exatidão da classificação de sistemas usando letras ou áudio. Essas experiências foram conduzidas utili-

zando uma base de dados de 5.296 músicas (com letra e áudio para cada), com 18 categorias de emoções derivadas de *tags* sociais. Os resultados dos experimentos de (HU; DOWNIE, 2010a) mostraram uma média aproximada das combinações de 63,00%.

Em (MIHALCEA; STRAPPARAVA, 2012) foram utilizadas 100 músicas rotuladas para as emoções: raiva, repulsa, medo, alegria, tristeza e surpresa. Foram demonstradas e descritas que a utilização conjunta entre letras e música trazem melhorias significativas sobre cada um dos textos individuais e sobre os classificadores musicais, com redução das taxas de erro em até 31%.

2.4.6 COMBINANDO ÁUDIO, MIDI E LETRAS

Outra combinação já utilizada foi a junção do áudio, do MIDI e das letras das músicas na detecção das emoções (PANDA et al., 2013).

Em (PANDA et al., 2013) foram inseridas metodologias para a criação automática de uma emoção em uma base de dados musicais multimodal, denominada *Allmusic.com*, já citada em outros trabalhos (YANG; LEE, 2009; BISCHOFF et al., 2009). Os arquivos e as letras corresponderam a um sub-conjunto de amostras de áudio recolhidas em MIDI. O conjunto de dados foi organizado em uma estrutura de 5 *clusters*, definidos pelas emoções. A partir dos dados do áudio, 177 características padrões e 98 características melódicas foram extraídas. Foram experimentadas várias estratégias de aprendizagem e de seleção de atributos supervisionados para avaliar a abordagem multimodal proposta. Empregando as características de áudio por padrão, o melhor desempenho alcançado foi de 44,3% (F-média). Com a abordagem multimodal, os resultados melhoraram para 61,1%, utilizando-se apenas 19 características multimodais. Recursos melódicos de áudio foram importantes para essa melhoria.

2.4.7 COMBINANDO ÁUDIO E TAGS

Em (TURNBULL et al., 2009) foi mencionado que para anotar músicas é importante considerar tanto as características do áudio como as do contexto social. Foram exploradas técnicas para relacionar e combinar múltiplas fontes de informações para fins de construção de um sistema de recuperação de músicas por consultas textuais. Foram consideradas duas representações tiradas das características do áudio (timbre e harmonia), além de duas fontes sociais (*tags* sociais e web documentos). Em seguida, foram comparados três algoritmos que combinam essas fontes de informações: CSA (ZADROZNY; ELKAN, 2002), RankBoost (FREUND et al., 2003) e KC-SVM (LANCKRIET et al., 2004). Também foi demonstrado empiricamente que esses algoritmos são superiores àqueles que utilizam fontes de informações individuais (em um

cenário de 400 músicas de treinamento utilizando o SVM de um único *kernel* e 50 músicas para validar os resultados). Foi utilizada a validação cruzada em dez *folds* para cada rótulo e depois calculados sobre o conjunto de 72 tags.

2.4.8 SUMÁRIO DOS TRABALHOS RELACIONADOS

A tabela 3 sumariza os trabalhos relacionados, apresentando uma visão geral das representações, técnicas e modelos utilizados.

Tabela 3: Sumário dos trabalhos relacionados.

Representação	Trabalho	Base Utilizada	Modelo Emocional	Nº Rótulos
Áudio	(LU et al., 2006)	Músicas de Johann Strauss	Modelo Bidimensional	5
	(EEROLA et al., 2009)	<i>Allmusic/Lastfm</i>	Emoções Básicas	
Letras	(YANG; LEE, 2009)	<i>Allmusic/Lastfm=1.032</i>	Emoções Básicas	
	(ZAAANEN; KANTERS, 2010)	Cânticos Líricos	V-A	
	(MIHALCEA; STRAPPARAVA, 2012)	100 Músicas	Emoções Básicas	6
Inf. Culturais	(BISCHOFF et al., 2009)	<i>Lastfm</i>	Emoções Básicas	178
	(YANG; HU, 2012)	Músicas Ocidentais e Não Oci.	<i>Russel's Model</i>	
Áudio e Imagens	(DUNKER et al., 2008)	Diversas Fontes	Adaptação de V-A	12
Áudio e Letras	(LAURIER et al., 2008)	<i>Lastfm=1.000</i>	Emoções Básicas	4
	(HU et al., 2009)	5.558	Emoções Básicas	135
	(HU et al., 2009)	5.296	Emoções Básicas	135
	(HU; DOWNIE, 2010a)	5.296	Emoções Básicas	135
Áudio, MIDI e Letras	(PANDA et al., 2013)	178	Emoções Básicas	
Áudio e Tags	(TURNBULL et al., 2009)	<i>Lastfm</i>	Emoções Básicas	

Fonte: Autoria própria.

Como conclusão, após analisar os trabalhos relacionados, notou-se que as combinações entre as representações são válidas e já foram utilizadas pela comunidade científica para a detecção das emoções. Entretanto, percebeu-se alguns fatores interessantes se comparados ao trabalho proposto:

- nas pesquisas analisadas não foram encontrados estudos utilizando músicas latinas;
- não foram encontrados estudos utilizando a progressão harmônica das músicas, que podem ser obtidas usando as cifras;
- não foram realizadas pela comunidade científica combinações envolvendo as cifras com qualquer outra representação musical.

não indicava um som preciso. Ela possuía um número muito elevado de símbolos e não se sabe de quem nasceu a ideia ou a época exata do seu aparecimento.

“No século XI surgiu o grande Guido d’Arezzo (monje do convento de Pomposa), que imaginou um novo sistema indicando com precisão a altura relativa dos sons. Dele nasceu a ideia da criação de pauta.” (PRIOLLI, 1989)

O sistema de linhas por meio de uma pauta, criado por Guido d’Arezzo, pôs fim à incerteza dos sinais neumáticos. Estudos de (PRIOLLI, 1989) mencionam ainda que Guido d’Arezzo juntou primeiramente mais uma linha às duas já existentes e verificando o bom resultado, juntou mais uma linha, dando a cada linha uma nota fixa. Na mesma época foi adicionada mais uma linha (5ª linha).

Outra contribuição importante de Guido d’Arezzo, segundo (PRIOLLI, 1989), além da pauta, foi a criação da letra-clave, que mais tarde foi definida como clave. Na clave podemos especificar três letras C (Dó), F (Fá), e G (Sol) essas letras definem a intensidade do som (agudo ou grave) na partitura musical (por meio da pauta), sendo que a partir dessas letras definimos com quais instrumentos musicais devemos tocar uma determinada partitura musical.

Além da pauta, das notas e da clave, (PRIOLLI, 1989) cita que sempre com a ideia de facilitar a leitura musical, Guido d’Arezzo modificou também a configuração dos sinais neumáticos, conforme a seguinte citação:

“...Guido d’Arezzo modificou também a configuração dos sinais neumáticos, dando-lhes uma figura mais ou menos uniforme, com a forma de pontos negros quadrados ou losangos, indicando precisamente o lugar do som correspondente. Esta notação ficou sendo chamada notação negra ou quadrada. Nesta ocasião surgiu também a música proporcionada (compasso), marcando a época da música medida. Para sílabas longas eram usadas notas de forma quadrada (chamadas longas) e para sílabas breves usavam notas em forma de losango (chamadas breves). Surgiu então um outro sistema: notação colorida, com notas vermelhas (valores breves) e negras (valores longos).” (PRIOLLI, 1989)

Podemos notar com a citação que por intermédio do compasso, podemos marcar o tempo de uma música, mediante as representações longas e breves. (PRIOLLI, 1989) ainda afirma que na opinião de vários musicólogos, foi para evitar a utilização de diferentes cores que surgiu a ideia de substituir notas pretas por brancas para representar os valores longos e a utilização das notas pretas para representar valores breves. (PRIOLLI, 1989) ainda relata que tais acontecimentos ocorreram em princípios do século XIV.

“Na idade Média o tempo musical era denominado: perfeito (correspondente ao atual compasso ternário) e imperfeito (correspondente ao atual compasso binário). Eis aí a origem dos nossos compassos simples e compostos.” (PRIOLLI, 1989)

Nota-se com a citação de (PRIOLLI, 1989) que as notações musicais foram evoluindo com o tempo, conforme os musicistas achassem necessidade para aprimorarem a linguagem musical.

Na figura 4 podemos visualizar a pauta, a clave, as notas musicais, os compassos e outros símbolos musicais criados a partir de Guido d'Arezzo e adaptado para os tempos modernos por musicistas. O trecho musical é parte da Sonata Número 1 de Bach.



Figura 4: Trecho de uma música com os símbolos criados a partir de Guido d' Arezzo.

Fonte: Adaptado de (BONTE, 2015).

(PRIOLLI, 1989) menciona ainda que a conceitualização de acorde dá-se ao conjunto de sons ouvidos simultaneamente.

“Dá-se o nome de acorde ao conjunto de sons ouvidos simultaneamente, e cujas relações de altura são determinadas pelas leis da natureza. Os acordes são formados por grupos de 3, 4 e 5 sons.” (PRIOLLI, 1989)

(CHEDIAK, 1984) define o conceito de cifras, baseando-se no acordes.

“Cifras são símbolos criados para representar acordes, e são compostas de letras, números e sinais. Em cifra substituímos os nomes Lá, Si, Dó, Ré, Mi, Fá, Sol pelas sete primeiras letras, maiúsculas, do alfabeto, respectivamente. O acorde maior é representado apenas pela letra (exemplo: A), quando menor coloca-se um “m” (minúsculo) ao lado direito da letra (exemplo: Am). Os sinais de alteração sustenido (#) e bemol (b) aparecem imediatamente depois da letra maiúscula indicando a nota fundamental alterada, também, aparecem antes do número que indica o grau a ser alterado.” (CHEDIAK, 1984)

(CHEDIAK, 1984) também menciona que a cifra estabelece o tipo de acorde, as eventuais alterações e a inversão do mesmo, porém destaca que a notação das cifras não está mundialmente padronizada.

“A notação de cifras ainda não está mundialmente padronizada, razão que encontramos um mesmo acorde anotado de maneiras diferentes, como é o caso do acorde de sétima maior, diminuto e outros.” (CHEDIAK, 1984)

A tabela 5 apresenta um exemplo de diferentes notações utilizadas nas cifras e a tabela 6 apresenta a notação das cifras utilizadas nesse trabalho.

Tabela 5: Exemplos de opções de notações em cifras.

Recomendadas	Aceitáveis	Evitadas
C		CM
Cm	C -	
C(#5)	C +, C 5 +	
C ^o , C5 ^{dim}	C ^{o7} , C ^{dim7}	

Fonte: Adaptado de (CHEDIAK, 1984).

Tabela 6: Nomenclatura de acordes e cifras utilizados nesse trabalho.

Nota	Maior	Menor	Sustenido	Bemol	Quarta	Quinta	Sétima	Nona	Diminuto
Lá	A	Am	A#	Ab	A4	A5	A7	A9	A ^o
Si	B	Bm	B#	Bb	B4	B5	B7	B9	B ^o
Dó	C	Cm	C#	Cb	C4	C5	C7	C9	C ^o
Ré	D	Dm	D#	Db	D4	D5	D7	D9	D ^o
Mi	E	Em	E#	Eb	E4	E5	E7	E9	E ^o
Fá	F	Fm	F#	Fb	F4	F5	F7	F9	F ^o
Sol	G	Gm	G#	Gb	G4	G5	G7	G9	G ^o

Fonte: Autoria própria.

A seguir serão apresentados os descritores propostos para as cifras para classificações automáticas das emoções. Os descritores a seguir foram propostos originalmente por (CORASSA et al., 2013) para a classificação automática de gêneros musicais e esses vão ser utilizados nesse trabalho para a classificação de emoções.

3.1.2 PRESENÇA E AUSÊNCIA DOS ACORDES

Esse descritor, originalmente desenvolvido por (CORASSA et al., 2013), contém um vetor com 63 atributos para cada música pertencente a *Latin Music Cifras Database* (LMCD). Os 63 atributos contêm especificamente os 63 acordes correspondentes, conforme as variações

das 7 notas musicais (A, B, C, D, E, F, G) e dos 9 tipos de acordes (maior, menor, sustenido, bemol, quarta, quinta, sétima, nona, diminuto) utilizados nesse trabalho. Caso o descritor encontre um acorde, dentre os 63, é anotado o valor 1 (um) em sua respectiva posição no vetor, ou 0 (zero) caso não encontre. Esse procedimento é repetido até que a música termine.

A tabela 7 mostra um exemplo do descritor presença e ausência dos acordes, com base na figura 5 que apresenta o trecho musical de uma das músicas contidas na *Latin Music Cifras Mood Database* (LMCMD).

Falamansa
Xote Dos Milagres
 Tom: G
 Intro 3x: G D Am C
 G D Am
 Escrevi seu nome na areia
 G D Am
 O sangue que corre em mim sai da tua veia
 G D Em C D
 Veja só você é a única que não me dá valor

Figura 5: Trecho da música Xote dos Milagres - Falamansa.

Fonte: Adaptado de Cifras Club.

Tabela 7: Exemplo do descritor presença e ausência dos acordes.

Música	C	D	E	F	G	A	B	Cm	Dm	Em	Fm	Gm	Am	Bm	...
Xote dos Milagres	1	1	0	0	1	0	0	0	0	1	0	0	1	0	...

Fonte: Adaptado de (CORASSA et al., 2013).

3.1.3 TOM DOS ACORDES

No cabeçalho de cada cifra é especificado o tom da música, o qual por sua vez define a altura sonora em que a música deverá ser tocada, sendo a música influenciada diretamente pelo seu tom musical, ou pela sua nota tônica.

A partir do tom de uma música definimos a sua musicalidade, como exemplo, a maneira como serão tocados os acordes, os solos ou até mesmo a sua transposição, caso seja necessário alterar a altura sonora da música. No descritor é realizada a leitura da cifra e anotado o seu tom musical, ou, caso não o encontre, caso não esteja especificado na cifra, é procurado na música o acorde com maior número de repetições, sendo este então definido como o tom dessa música.

A tabela 8 mostra um exemplo do descritor tom dos acordes, com base na figura 5 que apresenta o trecho musical de uma das músicas contidas na *Latin Music Cifras Mood Database* (LMCMD).

Tabela 8: Exemplo do descritor tom dos acordes.

Música	Tom
Xote dos Milagres	G

Fonte: Adaptado de (CORASSA et al., 2013).

3.1.4 QUANTIDADE DE REPETIÇÕES DOS ACORDES

Esse descritor conta os diferentes acordes durante a execução musical e incrementa com o valor 1 (um) cada acorde encontrado, dessa forma no final da música teremos a soma total de cada um dos acordes. Caso não seja encontrado por um acorde específico na música, dos 63 utilizados nesse trabalho, aquele é preenchido com o valor 0 (zero).

A tabela 9 mostra um exemplo do descritor quantidade de repetição dos acordes, com base na figura 5 que apresenta o trecho musical de uma das músicas contidas na *Latin Music Cifras Mood Database* (LMCMD).

Tabela 9: Exemplo do descritor quantidade de repetições dos acordes.

Música	C	D	E	F	G	A	B	Cm	Dm	Em	Fm	Gm	Am	Bm	...
Xote dos Milagres	1	4	0	0	3	0	0	0	0	1	0	0	2	0	...

Fonte: Adaptado de (CORASSA et al., 2013).

3.1.5 FREQUÊNCIA DOS ACORDES

Nesse descritor é considerado as inúmeras vezes em que um acorde pode ser utilizado na mesma música, considerando que podemos ter cifras curtas e cifras longas. É estabelecido dessa forma uma proporção com que o acorde se repete, estabelecendo a frequência de cada acorde. O cálculo desta proporção por acorde pode ser visto através da equação (1).

$$S_n = \frac{\sum_i^n \text{acorde}}{TOTAL_{\text{acordes}}} \quad (1)$$

A tabela 10 exhibe um exemplo do descritor frequência dos acordes, com base na figura 5 que apresenta o trecho musical de uma das músicas contidas na *Latin Music Cifras Mood Database* (LMCMD).

Tabela 10: Exemplo do descritor frequência dos acordes.

Música	C	D	E	F	G	...	Em	Fm	Gm	Am	Bm	...
Xote dos Milagres	0.09	0.36	0	0	0.27	...	0.09	0	0	0.18	0	...

Fonte: Adaptado de (CORASSA et al., 2013).

3.1.6 TRANSIÇÃO DOS ACORDES

Nesse descritor é analisada a transição dos acordes, ou progressão dos acordes em uma música. Com a transição é possível verificar o encaixe harmonioso dos acordes em uma música. O descritor considera as transições existentes entre cada um dos acordes pertencentes a uma música, verificando o seu avanço ou a sua regressão. Dessa forma o descritor quando encontra uma transição entre acordes preenche com o valor 1 (um) e quando não encontra preenche com o valor 0 (zero), percorrendo entre todas as transições de acordes possíveis.

A tabela 11 expõe um exemplo do descritor transição dos acordes, com base na figura 5 que apresenta o trecho musical de uma das músicas contidas na *Latin Music Cifras Mood Database* (LMCMD).

Tabela 11: Exemplo do descritor transição dos acordes.

Música	C→D	C→E	C→F	...	D→Am	...	Em→C	...
Xote dos Milagres	1	0	0	...	1	...	1	...

Fonte: Adaptado de (CORASSA et al., 2013).

3.1.7 ATRAVÉS DO ÁUDIO

Uma abordagem comumente utilizada para o processamento do áudio é a *Statistical Spectrum Descriptor* (SSD). Nesse descritor, são considerados o ritmo e o timbre para a extração das características musicais. Mediante esse descritor é possível encontrar músicas com características semelhantes. O descritor utilizado nesse trabalho, originalmente desenvolvido por (LIDY; RAUBER, 2005), denominado jAudio, contém um vetor com 161 atributos.

A figura 6 exibe um diagrama mostrando a extração das características do áudio por meio do SSD. Nas informações do SSD são calculados os padrões dos ritmos (de acordo com a ocorrência de batidas ou variação rítmica de energia em uma faixa específica) e medidas estatísticas, capazes de descrever o conteúdo do áudio a partir de valores ultrassonoros de cada banda crítica (média, mediana, variância, assimetria, curtose, mineração e *max-value*) (LIDY; RAUBER, 2005).

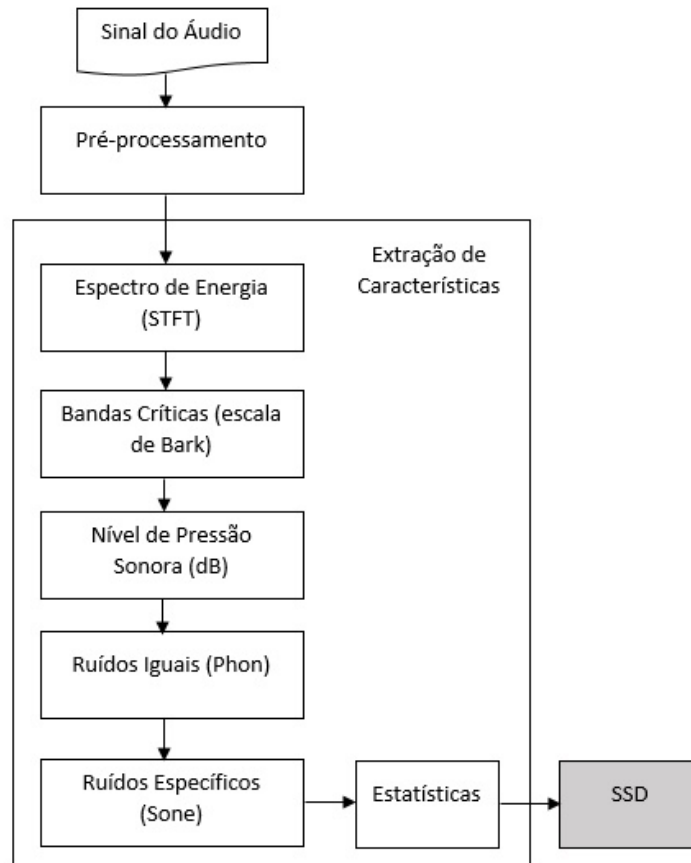


Figura 6: Diagrama das etapas SSD para a extração das características do áudio.

Fonte: Adaptado de (LIDY; RAUBER, 2005).

3.1.8 ATRAVÉS DAS LETRAS

A seguir serão apresentados os descritores propostos às letras para as classificações automáticas das emoções.

3.1.9 REMOÇÃO DE STOPWORDS

De acordo com (EL-KHAIR, 2006), as *Stopwords* são palavras muito comuns, que possuem poucos significados semânticos, servindo apenas de função sintática, mas podem afetar a eficácia da recuperação, porque estas têm uma frequência muito alta e tendem a diminuir o impacto nas diferenças de frequência entre as palavras menos comuns. A remoção das *Stopwords* segundo (EL-KHAIR, 2006), podem alterar o comprimento do documento e desta forma afetar o processo de ponderação. Dessa forma, são criadas listas pré-definidas, chamadas lista de *stopwords*. Nesse trabalho foram utilizadas três listas: inglês, português e espanhol (ver Anexo A), obtidas na internet por meio de ranks.nl.

3.1.10 N-GRAMS

De acordo com (CAVNAR; TRENKLE, 1994) a categorização de textos é uma tarefa fundamental no processamento de documentos, o que permite automatizar o tratamento dos fluxos dos documentos eletrônicos. No entanto, de acordo com (CAVNAR; TRENKLE, 1994) uma dificuldade em lidar com algumas classes de documentos é a presença de diferentes tipos de erros textuais, citam-se erros ortográficos e gramaticais. Desta forma, a categorização de texto devem funcionar corretamente em todas as entradas, tolerando estes tipos de problemas. Uma maneira de contornar esses problemas é utilizando o conceito de *n-gram*. A literatura define *n-gram* como sendo qualquer conjunto de co-ocorrência de termos com o comprimento N (CAVNAR; TRENKLE, 1994). Por exemplo, na palavra alegria têm-se: *bi-grams*: (_a, al, le, eg, gr, ri, a_), *tri-grams*: (_al, egr, gri, ri_, ia_). Ainda de acordo com (CAVNAR; TRENKLE, 1994), para ajudar com correspondência de início e fim de palavra utiliza-se o caractere underline (“_”).

3.1.11 STEMMING

Essa técnica permite a redução de ruídos em classificações textuais, permitindo uma melhor eficácia nos resultados. Em (SEBASTIANI, 2002; LIMA et al., 2014) a técnica *stemming* é utilizada no domínio de classificação de texto, reduzindo a dimensionalidade de uma palavra, como exemplo cita “programação” e “programar”, o resultado do algoritmo resultante seria “programa”. Essa redução de dimensionalidade também é utilizada nesse trabalho, com o intuito de melhorar a eficiência dos resultados.

3.2 CLASSIFICAÇÃO

Após a extração das características das diferentes fontes de informação é possível utilizar um algoritmo de aprendizado de máquina para realizar a classificação. Os algoritmos de classificação normalmente possuem duas etapas, a etapa de treinamento e a etapa de teste. Na etapa de treinamento o algoritmo recebe como entrada o conjunto de dados com o rótulo (Classe) conhecido. Na etapa de teste o algoritmo utiliza o modelo criado na etapa de treinamento para fazer a classificação. Em todos os experimentos realizados neste trabalho são utilizados os algoritmos: *Support Vector Machines* (SVM) (CORTES; VAPNIK, 1995), *k-Nearest Neighbours* (k-NN) (KELLER et al., 1985; WEINBERGER et al., 2005) e Árvore de Decisão (J48) (QUINLAN, 1996).

3.3 COMBINAÇÃO DE CLASSIFICADORES

No trabalho de (KITTLER et al., 1998) foi desenvolvido um *framework* comum para a combinação de classificadores o qual utiliza representações de padrões distintas e mostra que muitos sistemas podem ser considerados como um caso especial de classificações combinadas, em que todas as representações padrões são usadas em conjunto para tomar uma decisão. A ideia principal é que a saída (*scores*) dos diferentes classificadores possam ser avaliadas e combinadas de forma a criar um sistema reconhecimento de padrões mais robusto. As regras de combinação utilizadas neste trabalho são:

- Regra do Maior Valor: Para a regra do maior valor simplesmente contam-se os votos recebidos por esta hipótese nos classificadores individuais. A classe que recebe maior número de votos é então selecionada como a de consenso (a maioria) na decisão. A figura 7 mostra um exemplo com a regra do maior valor.
- Regra do Produto: Para a regra do produto os scores de cada classificador para cada classe são multiplicados e a classe com o maior valor é retornada como sendo a classe prevista pelo sistema. A figura 8 exibe um exemplo com a regra do produto.
- Regra da Soma: Para a regra da soma os scores de cada classificador, para cada classe, são somados e a classe com o maior valor é retornada como sendo a classe prevista pelo sistema. A figura 9 exibe um exemplo com a regra da soma.

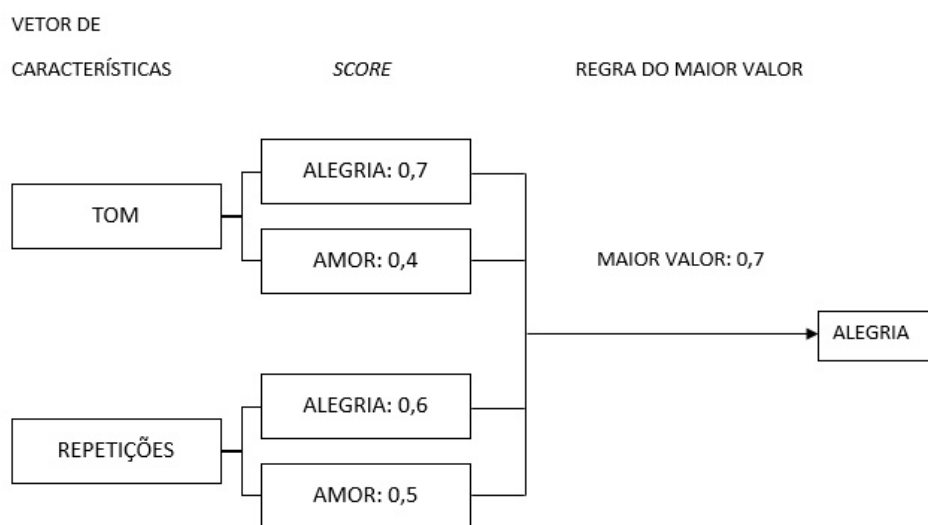


Figura 7: Regra do maior valor.

Fonte: Autoria própria.

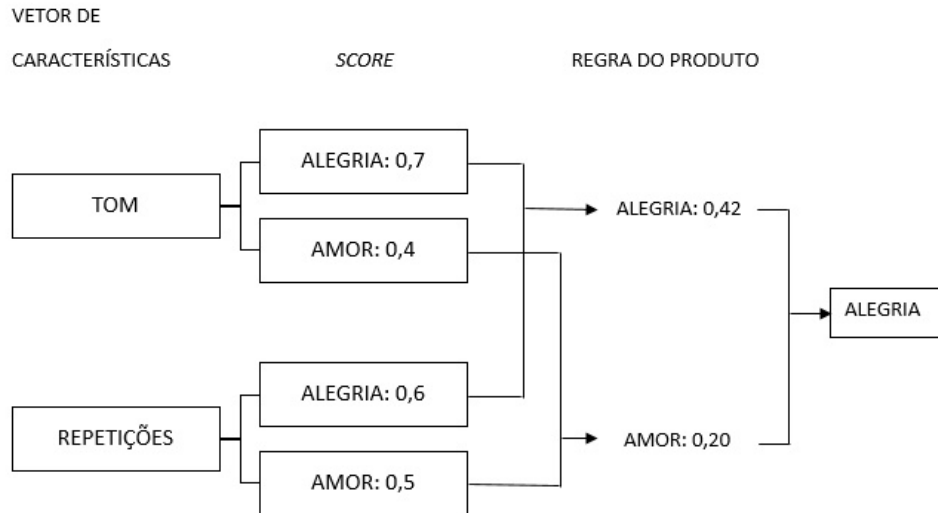


Figura 8: Regra do produto.

Fonte: Autoria própria.

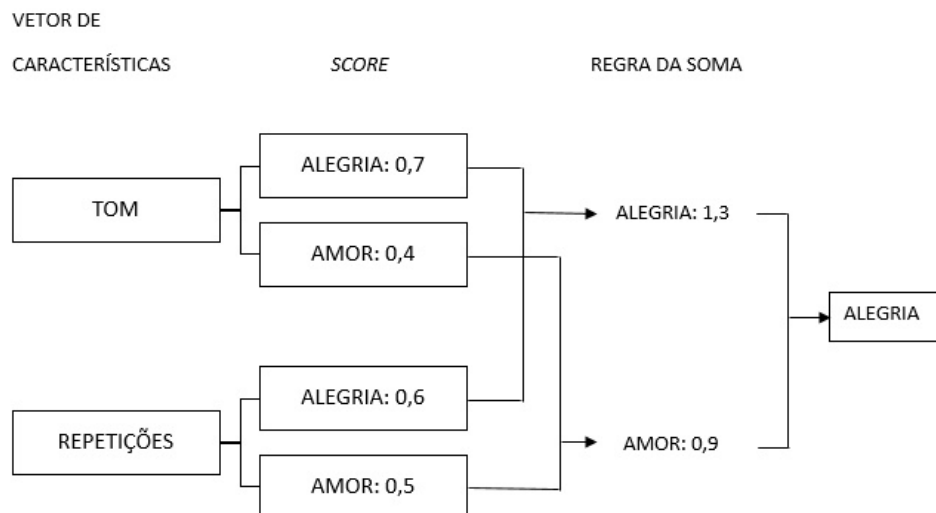


Figura 9: Regra da soma.

Fonte: Autoria própria.

No trabalho de (SNOEK et al., 2005) as abordagens que utilizam a estratégia definida em (KITTLER et al., 1998) são descritas como métodos de combinação de fusão tardia (*Late Fusion*), além disso também é apresentada a abordagem de combinação de fusão prematura (*early fusion*) que consiste em combinar os vetores de característica em uma única representação.

Nesse trabalho são utilizados métodos de combinação de fusão prematura e fusão tardia. Além disso são realizadas combinações unimodais (isto é, extraindo características da mesma fonte de informação) e multimodais (isto é, extraindo características de diferentes fon-

tes de informação). A figura 10 apresenta o fluxo das extrações unimodais propostas nesse trabalho. Os recursos extraídos por meio das cifras, do áudio e das letras são combinados a partir de uma única representação, gerando um único arquivo *Attribute-Relation File Format* (ARFF), que será classificado por intermédio dos algoritmos de classificação supervisionados.

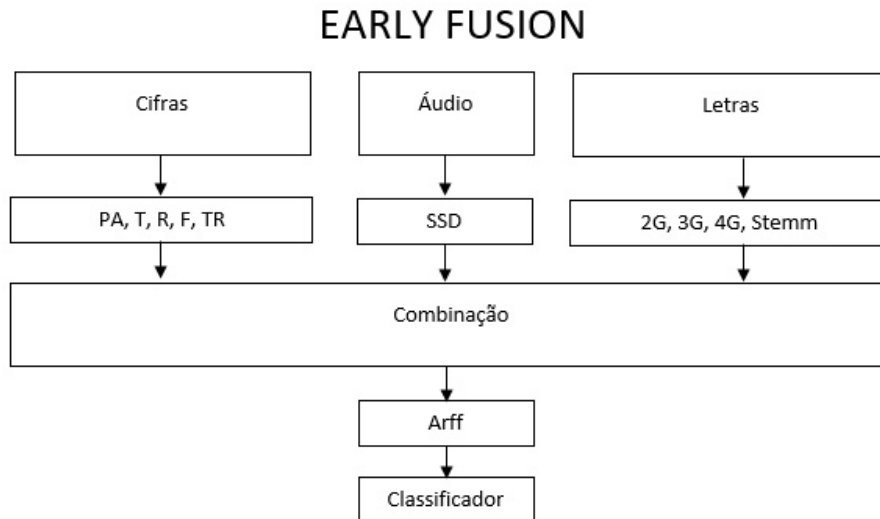


Figura 10: Esquema geral do Early Fusion.

Fonte: Autoria própria.

PA = Presença e ausência; T = Tom; R = Repetição; F = Frequência; TR = Transição;

2-G = 2-Grams; 3-G = 3-Grams; 4-G = 4-Grams;

A figura 11 apresenta a fusão tardia das extrações unimodais propostas. Os recursos extraídos mediante as cifras, do áudio e das outras representações irão gerar vários arquivos ARFFs, um arquivo para cada representação, que após passarão por uma classificação isolada em cada um dos algoritmos propostos. Somente após a classificação individualizada das representações utilizadas, será realizada a combinação.

3.4 PROTOCOLO EXPERIMENTAL

3.4.1 BASE DE DADOS

A *Latin Music Database* (LMD), (SILLA JR. et al., 2008) é uma base de dados que contém 3.139 composições musicais no formato MP3, originados a partir de composições de 501 artistas. A partir da LMD foi criada a *Latin Music Mood Database* (LMMD), uma base de dados de emoções criada por meio das 3.139 composições no formato MP3 da LMD (SANTOS; SILLA JR., 2015). A *Latin Music Mood Database* (LMMD) foi desenvolvida no laboratório de computação e tecnologia musical da UTFPR-CP. A LMMD possui duas versões. A primeira

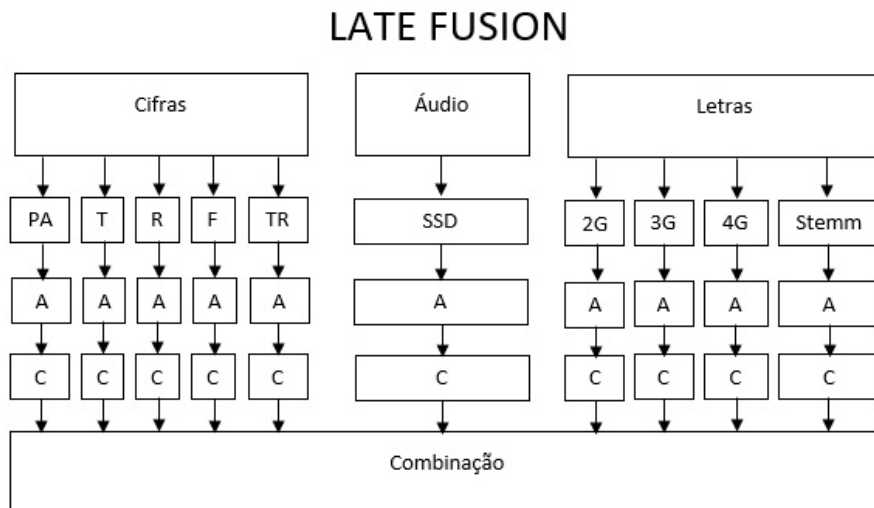


Figura 11: Esquema geral do Late Fusion.

Fonte: Autoria própria.

PA = Presença e ausência; T = Tom; R = Repetição; F = Frequência; TR = Transição;

2-G = 2-Grams; 3-G = 3-Grams; 4-G = 4-Grams; A = Arff; C = Classificador.

versão possui seis rótulos: alegria, amor, decepção, entusiasmo, paixão e tristeza. A segunda versão possui três rótulos: positiva, negativa e neutra. Após a criação das duas versões da base, foi observado que existe um mapeamento direto entre as duas versões. A emoção positiva possui todas músicas rotuladas como alegria, excitação/entusiasmo. A emoção neutra possui todas as músicas rotuladas com amor e paixão e a emoção negativa possui todas as músicas rotuladas com decepção e tristeza.

A criação da *Multimodal Latin Music Mood Database (MLMMD)* ocorreu a partir da criação de uma base para cada representação: cifras, áudio e letras. As cifras foram selecionadas mediante um processo manual de pesquisas em bases nacionais de cifras *on-line* para cada uma das 3.139 músicas existentes na LMD, no entanto, foram encontradas somente 784 cifras. Para o áudio foram selecionadas as músicas em mp3 da base LMD que equivalassem às mesmas músicas encontradas por intermédio das cifras. Para as letras foram obtidas as letras equivalentes às 784 músicas das cifras e do áudio.

Na tabela 12 é possível verificar base multimodal utilizada para a classificação com três emoções e na tabela 13 é possível verificar a mesma base com seis emoções.

Tabela 12: *Multimodal Latin Music Mood Database* com três emoções.

Base de Dados 1	MLMMD	LMMD
Positiva	239	922
Neutra	427	1.647
Negativa	118	570
Total	784	3.139

Fonte: Autoria própria.

Tabela 13: *Multimodal Latin Music Mood Database* com Seis Emoções.

Base de Dados 2	MLMMD	LMMD
Alegria	117	437
Amor	216	775
Decepção	15	99
Entusiasmo	122	485
Paixão	211	872
Tristeza	103	471
Total	784	3.139

Fonte: Autoria própria.

3.4.2 PRÉ-PROCESSAMENTO DOS DADOS

Na etapa de pré-processamento dos dados foram necessários realizar ajustes da base utilizada (LMMD), para os experimentos, tanto para a representação das cifras, quanto na representação das letras. Na construção da base das cifras foram refinadas as informações musicais individualmente, dessa forma, pesquisando-as em diferentes *websites* de repositórios de cifras. Cada website possui um padrão específico na forma de apresentar cada cifra musical, informações que não fossem referentes à estrutura harmônica das músicas foram retiradas, como exemplo, citam-se palavras como “refrão”, “intro”, dentre outras. Estas poderiam gerar problemas durante a extração dos descritores.

Na etapa de pré-processamento das letras para a base LMMD foram retiradas caracteres de pontuação, acentos, caracteres especiais, quebra de palavras com hífen e padronização das letras por minúsculo. As palavras *stopwords* utilizadas nesse trabalho são palavras com pouco significado semântico ao texto, inclusas em uma lista pré-definida. Nesse trabalho foram utilizadas três listas de *stopwords*.

3.4.3 VETORES DE CARACTERÍSTICAS

A partir do extrator de características foram geradas as características individuais para cada representação (para três e seis emoções) da MLMMD:

- cifras: presença e ausência de acordes, tom da música, número de repetições dos acordes, frequência dos acordes e transição dos acordes;
- áudio: ritmo e timbre das músicas;
- letras: *2-Grams*, *3-Grams*, *4-Grams* e *Stemm* dos textos.

Para avaliar o impacto das combinações: unimodais e multimodais, nas três representações, foram utilizados os métodos *Early Fusion* e *Late Fusion*, que se diferenciam na maneira de integrar os resultados de extração das características (fusão prematura e da fusão tardia).

3.4.4 MINERAÇÃO DOS DADOS

Para realização dos experimentos foi utilizada a ferramenta WEKA, (HALL et al., 2009) que implementa os algoritmos de mineração de dados propostos: SVM, k-NN (k=1,3,5) e J48.

Para que o WEKA possa fazer a leitura dos dados esses precisam estar em um formato específico (ARFF), diante disso, foi criado um arquivo nesse formato para todas as representações propostas: cifras, áudio e letras e nas versões *no fusion* e *early fusion*.

Para realizar as etapas de treinamento e teste, inerente à tarefa de aprendizado de máquina, por intermédio do WEKA foi utilizado o procedimento de validação cruzada em dez *folds*.

3.4.5 MÉTRICAS DE AVALIAÇÃO

Apresentam-se a seguir as medidas de desempenho, segundo (WITTEN; FRANK, 2005).

Acurácia: é a média de percentual de amostras positivas e negativas corretamente classificadas. É calculada a partir da razão de amostras corretamente classificadas pelo número de amostras disponíveis (WITTEN; FRANK, 2005).

Recall: é a medida de porcentagem de amostras positivas classificadas corretamente sobre o total de amostras positivas (WITTEN; FRANK, 2005).

Precisão: é a proporção de exemplos que uma classe sobre o total de exemplos classificados como aquela classe (WITTEN; FRANK, 2005).

F-Measure: a medida *F-Measure* é definida como a média ponderada harmônica dos

valores de precisão (*Precision*) e (*Recall*) (WITTEN; FRANK, 2005), dada por meio da equação (2).

$$F - measure = 2 \times Precision \times Recall / (Precision + Recall) \quad (2)$$

Média: é a média aritmética das médias ponderadas harmônicas, obtida mediante o somatório das médias ponderadas da *F-Measure* individuais de cada classe, dividido pela quantidade de classes (três ou seis).

Além disso, considerando o desbalanceamento de classes existentes na base, são sempre apresentadas a média aritmética e a média ponderada (*weighted average*). As análises apresentadas são realizadas utilizando a média aritmética, contudo dado o desbalanceamento de classes achamos interessante apresentar também a média ponderada.

4 EXPERIMENTOS SEM COMBINAÇÕES (*NO FUSION*)

Esse capítulo apresenta os experimentos sem combinações (*no fusion*) com três e seis emoções, obtidos a partir dos descritores analisados das seguintes representações: cifras, áudio e letras. Parte dos resultados apresentados nesse capítulo, utilizando apenas as cifras, foram apresentados na *International Conference on Systems, Man, and Cybernetics* da IEEE (IEEE SMC 2015) (PRZYBYSZ et al., 2015).

4.1 RESULTADOS UTILIZANDO *NO FUSION* COM TRÊS EMOÇÕES

Os experimentos desse capítulo têm como objetivo analisar a partir de três representações musicais: cifras, áudio e letras, qual apresenta melhores resultados e se alguma das três representações destaca-se sobre as demais no reconhecimento de emoções em músicas. Nesse capítulo não são utilizadas estratégias de combinações e analisam-se os desempenhos dos classificadores: SVM, IBk(k=1,3,5) e J48 de forma individual (por emoção) e geral (média de classificação).

A tabela 14 está organizada de forma a apresentar os resultados detalhados dos experimentos para as representações que contêm informações das cifras, do áudio e das letras para três emoções. Por meio do símbolo de # é especificada a numeração do experimento realizado, seguido pela fonte de informação (cifras, áudio ou letras), pelas características analisadas para cada fonte de informação e pelos algoritmos utilizados para classificar as classes de emoções. Foram utilizadas duas médias, a média ponderada da *F-measure* (MP-F) e a média aritmética da *F-measure* (MA-F). A escolha pela utilização de duas médias ocorreu devido o desbalançamento das classes, aspecto que pode estar prejudicando o desempenho dos classificadores na MLMMD. A análise dos resultados é realizada por intermédio da média aritmética da *F-measure* (MA-F).

Tabela 14 – Resultados sem combinações para três emoções. (continua)

#	Fonte de Informação	Características	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
1	Cifras	Presença e ausência	Positiva	24,70%	36,10%	35,60%	37,50%	39,70%
			Neutra	69,20%	62,70%	64,10%	65,70%	65,00%
			Negativa	01,70%	27,00%	22,70%	15,70%	05,00%
			MP-F	45,50%	49,20%	49,20%	49,60%	48,30%
			MA-F	31,86%	41,93%	40,80%	39,63%	36,56%

Tabela 14 – Resultados sem combinações para três emoções.

(continuação)

#	Fonte de Informação	Características	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
2	Cifras	Tom	Positiva	14,00%	11,50%	10,30%	10,30%	00,00%
			Neutra	68,90%	68,60%	68,10%	68,10%	70,60%
			Negativa	00,00%	00,00%	00,00%	00,00%	00,00%
			MP-F	41,90%	40,90%	40,30%	40,30%	38,50%
			MA-F	27,63%	26,70%	26,13%	26,13%	23,53%
3	Cifras	Repetição	Positiva	32,00%	45,30%	38,50%	37,80%	45,30%
			Neutra	71,50%	64,40%	63,10%	65,20%	65,40%
			Negativa	00,00%	19,80%	17,60%	14,30%	12,00%
			MP-F	48,70%	51,70%	48,80%	49,20%	51,20%
			MA-F	34,50%	43,16%	39,73%	39,10%	40,90%
4	Cifras	Frequência	Positiva	16,20%	40,20%	38,90%	35,90%	40,50%
			Neutra	68,40%	58,50%	60,80%	63,70%	65,60%
			Negativa	00,00%	14,60%	38,90%	15,90%	12,20%
			MP-F	42,40%	46,30%	46,50%	48,10%	49,90%
			MA-F	28,20%	37,76%	46,20%	38,50%	39,43%
5	Cifras	Transição	Positiva	46,20%	46,20%	43,30%	41,10%	38,40%
			Neutra	70,60%	68,10%	71,40%	71,10%	70,90%
			Negativa	13,70%	17,70%	11,90%	07,10%	00,00%
			MP-F	54,60%	53,80%	53,90%	52,30%	50,30%
			MA-F	43,50%	44,00%	42,20%	39,76%	36,43%
6	Áudio	SSD	Positiva	46,10%	52,10%	50,10%	50,90%	44,40%
			Neutra	71,00%	61,00%	63,10%	64,90%	58,90%
			Negativa	01,70%	17,80%	14,90%	13,00%	22,00%
			MP-F	53,00%	51,80%	51,90%	52,90%	49,00%
			MA-F	39,60%	43,63%	42,70%	42,93%	41,76%
7	Letras	2-Grams	Positiva	69,00%	61,20%	65,80%	67,50%	55,30%
			Neutra	74,80%	70,70%	71,90%	73,40%	64,80%
			Negativa	25,90%	30,80%	18,30%	11,70%	22,90%
			MP-F	65,70%	61,80%	62,00%	62,30%	55,60%
			MA-F	56,56%	54,23%	52,00%	50,86%	47,66%
8	Letras	3-Grams	Positiva	72,50%	55,90%	52,70%	46,00%	55,80%
			Neutra	74,10%	68,30%	70,40%	71,80%	65,90%
			Negativa	28,80%	17,80%	18,10%	16,70%	25,70%
			MP-F	66,80%	57,00%	57,10%	55,60%	56,80%
			MA-F	58,46%	47,33%	47,06%	44,83%	49,13%

Tabela 14 – Resultados sem combinações para três emoções. (continuação)

#	Fonte de Informação	Características	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
9	Letras	4-Grams	Positiva	75,10%	32,20%	09,30%	01,70%	57,80%
			Neutra	78,30%	68,90%	70,70%	70,30%	65,40%
			Negativa	22,90%	14,90%	12,70%	00,00%	27,40%
			MP-F	69,00%	49,60%	43,20%	38,80%	57,40%
			MA-F	58,76%	38,66%	30,90%	24,00%	50,20%
10	Letras	Stemm	Positiva	00,00%	00,00%	00,00%	00,00%	00,00%
			Neutra	70,50%	70,50%	70,50%	70,50%	70,50%
			Negativa	00,00%	00,00%	00,00%	00,00%	00,00%
			MP-F	38,40%	38,40%	38,40%	38,40%	38,40%
			MA-F	23,50%	23,50%	23,50%	23,50%	23,50%

Fonte: Autoria própria.

Nota-se que em todos os experimentos para três emoções que ao analisar as emoções, individualmente, a emoção neutra foi a mais fácil de ser classificada por qualquer um dos classificadores empregados.

O Experimento # 1: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando a presença e a ausência dos acordes) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (J48=39,70%), neutra (SVM=69,20%) e negativa(1-NN= 27,00%). Os melhores resultados para essa representação foram obtidos pelo classificador 1-NN com MA-F=41,93% .

O Experimento # 2: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando o tom dos acordes) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=14,00%), neutra (J48=70,60%) e negativa(SVM, 1-NN, 3-NN, 5-NN, J48= 00,00%). Todos os classificadores falharam ao reconhecer a emoção negativa. Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F= 27,63%.

O Experimento # 3: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando a repetição dos acordes) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (1-NN,J48=45,30%), neutra (SVM=71,50%) e negativa(1-NN=19,80%). Os melhores resultados para essa representação foram obtidos pelo classificador 1-NN com MA-F=43,16%.

O Experimento # 4: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando a frequência dos acordes) para cada emoção. Os melhores resultados desse

experimento por emoção (em negrito na tabela) foram: positiva (J48=40,50%), neutra (SVM=68,40%) e negativa(3-NN=38,90%). Os melhores resultados para essa representação foram obtidos pelo classificador 3-NN com MA-F=46,20%.

O Experimento # 5: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando a transição dos acordes) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM,1-NN=46,20%), neutra (3-NN=71,40%) e negativa(1-NN=17,70%). Os melhores resultados para essa representação foram obtidos pelo classificador 1-NN com MA-F=44,00%.

O Experimento # 6: apresenta os melhores resultados da classificação sem combinação do áudio (utilizando o SSD da música) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (1-NN=52,10%), neutra (SVM=71,00%) e negativa (J48= 22,00%). Os melhores resultados para essa representação foram obtidos pelo classificador 1-NN com MA-F=43,63%.

O Experimento # 7: apresenta os melhores resultados da classificação sem combinação das letras (utilizando 2-Grams) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=69,00%), neutra (SVM=74,80%) e negativa (1-NN=30,80%). Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=56,56%.

O Experimento # 8: apresenta os melhores resultados da classificação sem combinação das letras (utilizando 3-Grams) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=72,50%), neutra (SVM=74,10%) e negativa (SVM=28,80%). Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=58,46%.

O Experimento # 9: apresenta os melhores resultados da classificação sem combinação das letras (utilizando 4-Grams) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=75,10%), neutra (SVM=78,30%) e negativa (J48=27,40%). Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=58,76%.

O Experimento # 10: apresenta os resultados da classificação sem combinação das letras (utilizando o *Stemm*) para cada emoção. A análise dos resultados desse experimento, mostram que claramente todos os classificadores falharam na etapa de treinamento em criar um modelo capaz de distinguir a classe majoritária (Emoção: Neutra) das demais. Por esse motivo todos os classificadores apresentaram o mesmo resultado nesse experimento.

4.1.1 SUMÁRIO DOS RESULTADOS *NO FUSION* COM TRÊS EMOÇÕES

A tabela 15 apresenta um sumário com os resultados da média aritmética da *F-measure* obtidos a partir do método *no fusion* para cada abordagem das cifras, do áudio e das letras com os diferentes classificadores e utilizando três emoções.

A análise da tabela 15 mostra alguns resultados interessantes. Em primeiro lugar, os melhores resultados foram obtidos pela representação das letras (2-Grams, 3-Grams e 4-Grams), o classificador SVM apresentou um melhor resultado em 4-Grams=58,76%, seguido pelo 2-Grams=IBk(1-NN=56,56%, 3-NN=52,00%, 5-NN=50,86%) e 4-Grams=J48(50,20%). Em segundo lugar, encontramos resultados bem próximos entre as representações das cifras e do áudio, exemplo (por meio da transição dos acordes): SVM=43,50% e 39,60%, 1-NN=44,00% e 43,63%, 3-NN=42,20% e 42,70%, 5-NN=39,76% e 42,93%. Dessa forma, ao se analisar a representação das letras para o método sem combinações, conclui-se que estes descritores são úteis para a detecção das emoções na base utilizada, uma vez que ofereceu os melhores resultados globais para as três emoções, seguidos pelo áudio e transição dos acordes.

Tabela 15: Sumário dos resultados utilizando o método *no fusion* com três emoções.

Conjunto de Características	SVM	1-NN	3-NN	5-NN	J48
Presença e ausência dos acordes	31,86%	41,93%	40,80%	39,63%	36,56%
Tom da música	27,63%	26,70%	26,13%	26,13%	23,53%
Número de repetições dos acordes	34,50%	43,16%	39,73%	39,10%	40,90%
Frequência dos acordes	28,20%	37,76%	46,20%	38,50%	39,43%
Transição dos acordes	43,50%	44,00%	42,20%	39,76%	36,43%
SSD	39,60%	43,63%	42,70%	42,93%	41,76%
2-Grams	56,56%	54,23%	52,00%	50,86%	47,66%
3-Grams	58,46%	47,33%	47,06%	44,83%	49,13%
4-Grams	58,76%	38,66%	30,90%	24,00%	50,20%
Stemm	23,50%	23,50%	23,50%	23,50%	23,50%

Fonte: Autoria própria.

A tabela 16 apresenta os melhores resultados individuais utilizando o método *no fusion* obtidos para MLMMD com três emoções, a partir das representações das cifras, do áudio e das letras. Analisando os resultados por meio dos experimentos realizados é possível obter os melhores resultados individuais da classificação das três representações para cada emoção: positiva (4-Grams=75,10%), neutra (2-Grams=74,80%) e negativa (Transição=44,00%). Novamente, a melhor representação sem combinação dos experimentos realizados para a base MLMMD foi a letra. Ao analisar as emoções individuais, a emoção neutra foi a mais fácil de ser classificada por qualquer uma das representações e dos classificadores empregados para as três emoções.

Tabela 16: Melhores resultados obtidos utilizando o método *no fusion* para representações individuais com três emoções.

Emoção	Cifras					Áudio		Letras		
	PA	T	R	F	TR	SSD	2-Gram	3-Gram	4-Gram	Stemm
Positiva	39,70%	14,00%	45,30%	40,50%	46,20%	52,10%	69,00%	72,50%	75,10%	00,00%
Neutra	69,20%	70,60%	71,50%	68,40%	71,40%	71,00%	74,80%	74,10%	78,30%	70,50%
Negativa	27,00%	00,00%	19,80%	38,90%	44,00%	22,00%	30,80%	28,80%	27,40%	38,40%

PA = Presença e ausência; T = Tom; R = Repetição; F = Frequência; TR = Transição

Fonte: Autoria própria.

4.2 RESULTADOS UTILIZANDO *NO FUSION* COM SEIS EMOÇÕES

A tabela 17 está organizada de forma a apresentar os resultados detalhados dos experimentos para as representações que contêm informações das cifras, do áudio e das letras para três emoções. Por meio do símbolo de # é especificada a numeração do experimento realizado, seguido pela fonte de informação (cifras, áudio ou letras), pelas características analisadas para cada fonte de informação e pelos algoritmos utilizados para classificar as classes de emoções. Foram utilizadas duas médias, a média ponderada da *F-measure* (MP-F) e a média aritmética da *F-measure* (MA-F). A análise dos resultados é realizada por intermédio da média aritmética da *F-measure* (MA-F).

Tabela 17 – Resultados sem combinações para seis emoções.

(continua)

#	Fonte de Informação	Características	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
1	Cifras	Presença e ausência	Alegria	13,30%	17,70%	14,90%	18,50%	17,90%
			Amor	42,80%	32,60%	36,70%	39,50%	29,10%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	11,80%	17,30%	14,60%	14,60%	14,50%
			Paixão	33,90%	28,30%	27,10%	30,60%	27,00%
			Tristeza	01,70%	16,80%	11,11%	09,10%	12,20%
			MP-F	24,90%	24,20%	23,30%	25,40%	21,80%
			MA-F	17,25%	18,78%	19,44%	18,71%	16,78%
2	Cifras	Tom	Alegria	10,50%	10,50%	08,40%	08,40%	10,50%
			Amor	42,90%	42,20%	41,40%	42,00%	42,20%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	08,30%	08,10%	05,50%	05,50%	08,10%
			Paixão	45,20%	45,80%	45,30%	44,60%	45,80%
			Tristeza	01,60%	01,60%	01,60%	01,60%	01,60%

Tabela 17 – Resultados sem combinações para seis emoções.

(continuação)

#	Fonte de Informação	Características	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
			MP-F	27,10%	27,00%	25,90%	25,90%	27,00%
			MA-F	18,08%	18,03%	17,03%	17,01%	18,03%
3	Cifras	Repetição	Alegria	09,90%	20,20%	22,60%	18,90%	20,70%
			Amor	31,30%	31,10%	30,10%	32,30%	32,20%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	05,10%	23,70%	17,00%	23,20%	19,30%
			Paixão	37,40%	36,90%	32,00%	34,40%	35,00%
			Tristeza	00,00%	18,40%	11,80%	07,60%	14,40%
			MP-F	21,00%	27,60%	24,50%	25,60%	26,30%
			MA-F	13,95%	21,71%	18,91%	19,40%	20,26%
			Alegria	00,00%	18,20%	19,60%	19,60%	18,00%
			Amor	35,30%	34,00%	35,20%	36,40%	26,00%
4	Cifras	Frequência	Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	03,10%	24,20%	15,20%	20,00%	14,80%
			Paixão	34,90%	13,35%	31,00%	38,20%	34,60%
			Tristeza	00,00%	12,90%	12,50%	09,60%	12,50%
			MP-F	19,60%	26,50%	25,00%	27,60%	23,10%
			MA-F	12,21%	17,10%	18,91%	20,63%	17,65%
			Alegria	26,10%	26,70%	22,90%	25,90%	24,60%
			Amor	43,70%	26,70%	29,30%	26,40%	43,50%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	20,10%	24,40%	19,80%	23,20%	17,10%
5	Cifras	Transição	Paixão	27,40%	37,90%	38,70%	39,00%	25,30%
			Tristeza	08,10%	06,90%	01,60%	00,00%	01,50%
			MP-F	27,50%	26,20%	25,20%	25,20%	25,30%
			MA-F	20,90%	20,43%	18,71%	19,08%	18,66%
			Alegria	17,40%	23,30%	27,10%	29,10%	23,70%
			Amor	34,30%	31,50%	35,90%	38,20%	30,00%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	38,10%	36,00%	27,10%	31,10%	28,50%
			Paixão	26,30%	31,90%	24,30%	33,20%	32,20%
			Tristeza	00,00%	19,00%	03,10%	06,10%	26,70%
6	Áudio	SSD	MP-F	25,10%	28,80%	25,10%	29,50%	28,40%
			MA-F	19,35%	23,61%	19,58%	19,91%	23,51%
			Alegria	41,20%	33,70%	39,50%	36,60%	21,40%
			Amor	33,70%	39,50%	36,60%	21,40%	
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	38,10%	36,00%	27,10%	31,10%	28,50%
7	Letras	2-Grams	Alegria	41,20%	33,70%	39,50%	36,60%	21,40%

Tabela 17 – Resultados sem combinações para seis emoções.					(continuação)			
#	Fonte de Informação	Características	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
8	Letras	3-Grams	Amor	46,10%	49,50%	52,40%	50,00%	30,20%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	54,20%	37,50%	28,90%	25,60%	34,70%
			Paixão	40,00%	31,30%	28,10%	22,10%	26,70%
			Tristeza	27,70%	22,00%	14,30%	09,60%	16,10%
			MP-F	41,70%	35,80%	34,30%	30,40%	26,20%
			MA-F	34,86%	29,00%	27,20%	23,98%	21,51%
			Alegria	47,30%	36,70%	39,00%	42,70%	17,30%
			Amor	47,50%	50,50%	49,50%	49,80%	37,20%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	61,00%	29,70%	13,30%	04,70%	41,40%
			Paixão	39,20%	28,40%	16,20%	24,10%	27,50%
			Tristeza	31,60%	16,90%	16,00%	13,50%	26,70%
			MP-F	44,44%	33,90%	28,00%	29,10%	30,20%
MA-F	37,76%	27,03%	22,33%	22,46%	25,01%			
9	Letras	4-Grams	Alegria	46,20%	18,90%	15,80%	06,30%	28,20%
			Amor	48,10%	43,90%	43,70%	43,20%	39,70%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	59,10%	20,00%	01,60%	00,00%	40,30%
			Paixão	43,30%	19,30%	04,50%	02,80%	30,40%
			Tristeza	26,40%	10,90%	03,60%	00,00%	25,40%
			MP-F	44,50%	24,70%	16,30%	13,60%	33,00%
			MA-F	37,18%	18,83%	11,53%	08,71%	27,33%
10	Letras	Stemm	Alegria	00,00%	00,00%	00,00%	00,00%	00,00%
			Amor	43,20%	43,20%	43,20%	43,20%	43,20%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	00,00%	00,00%	00,00%	00,00%	00,00%
			Paixão	00,00%	00,00%	00,00%	00,00%	00,00%
			Tristeza	00,00%	00,00%	00,00%	00,00%	00,00%
			MP-F	11,90%	11,90%	11,90%	11,90%	11,90%
			MA-F	07,20%	07,20%	07,20%	07,20%	07,20%

Fonte: Autoria própria.

Nota-se que todos os classificadores falharam no reconhecimento da emoção decepção.

O Experimento # 1: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando a presença e a ausência dos acordes) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (5-NN=18,50%), amor (SVM=42,80%), entusiasmo (1-NN= 17,30%), paixão (SVM=33,90%) e tristeza (1-NN= 16,80%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada por qualquer um dos classificadores empregados. Os melhores resultados para essa representação foram obtidos pelo classificador 3-NN com MA-F=19,44%.

O Experimento # 2: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando o tom dos acordes) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (SVM, 1-NN, J48=10,50%), amor (SVM=42,90%), entusiasmo (SVM=08,30%), paixão (1-NN=45,80%) e tristeza (SVM, 1-NN, 3-MM, 5-NN, J48=01,06%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=18,08%.

O Experimento # 3: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando a repetição dos acordes) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=22,60%), amor (5-NN=32,30%), entusiasmo (1-NN=23,70%), paixão (SVM= 37,40%) e tristeza (1-NN=18,40%). Ao analisar as emoções, individualmente, o entusiasmo foi a mais fácil de ser classificada por qualquer um dos classificadores empregados. Os melhores resultados para essa representação foram obtidos pelo classificador 1-NN com MA-F=21,71%.

O Experimento # 4: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando a frequência dos acordes) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN, 5-NN=19,60%), amor (5-NN=36,40%), entusiasmo (1-NN=24,20%), paixão (5-NN=38,20%) e tristeza (1-NN=12,90%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essa representação foram obtidos pelo classificador 5-NN com MA-F=20,63%.

O Experimento # 5: apresenta os melhores resultados da classificação sem combinação das cifras (utilizando a transição dos acordes) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (1-NN=26,70%), amor (SVM= 43,70%), entusiasmo (1-NN=24,40%), paixão (5-NN= 39,00%) e tristeza (SVM=08,10%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de sere classificada. Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=20,90%.

O Experimento # 6: apresenta os melhores resultados da classificação sem combinação do áudio (utilizando o SSD da música) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (5-NN=29,10%), amor (5-NN=38,20%), entusiasmo (SVM= 38,10%), paixão (5-NN= 33,20%) e tristeza (5-NN=26,70%). Ao analisar as emoções, individualmente, o amor e o entusiasmo foram as mais fáceis de serem classificadas. Os melhores resultados para essa representação foram obtidos pelo classificador 1-NN com MA-F=23,61%.

O Experimento # 7: apresenta os melhores resultados da classificação sem combinação das letras (utilizando 2-Grams) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (SVM=41,20%), amor (3-NN=52,40%), entusiasmo (SVM=54,20%), paixão (SVM= 40,00%) e tristeza (SVM=27,70%). Ao analisar as emoções, individualmente, o entusiasmo foi a mais fácil de ser classificada. Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=34,86%.

O Experimento # 8: apresenta os melhores resultados da classificação sem combinação das letras (utilizando 3-Grams) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (SVM=47,30%), amor (1-NN=50,50%), entusiasmo (SVM=61,00%), paixão (SVM= 39,20%) e tristeza (SVM=31,60%). Ao analisar as emoções, individualmente, o entusiasmo foi a mais fácil de ser classificada. Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=37,76%.

O Experimento # 9: apresenta os melhores resultados da classificação sem combinação das letras (utilizando 4-Grams) para cada emoção. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (SVM=46,20%), amor (SVM=48,10%), entusiasmo (SVM=59,10%), paixão (SVM= 43,30%) e tristeza (SVM=26,40%). Ao analisar as emoções, individualmente, o entusiasmo foi a mais fácil de ser classificada. Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=37,18%.

O Experimento # 10: apresenta os resultados da classificação sem combinação das letras (utilizando o *Stemm*) para cada emoção. A análise dos resultados desse experimento, mostram que claramente todos os classificadores falharam na etapa de treinamento em criar um modelo capaz de distinguir a classe Amor das demais. Por esse motivo todos os classificadores apresentaram o mesmo resultado nesse experimento.

4.2.1 SUMÁRIO DOS RESULTADOS *NO FUSION* COM SEIS EMOÇÕES

A tabela 18 apresenta um sumário com os resultados da média aritmética da *F-measure* obtidos a partir do método *no fusion* para cada abordagem das cifras, do áudio e das letras com

os diferentes classificadores e utilizando seis emoções.

A análise da tabela 18 mostra alguns resultados interessantes. Em primeiro lugar, os melhores resultados da média aritmética da *F-measure* foram obtidos pela representação das letras (2-Grams e 3-Grams) em todos os classificadores: o classificador SVM apresentou um melhor resultado em 3-Grams=37,76%, seguido pelo 2-Grams=IBk(1-NN=29,00%, 3-NN=27,20%, 5-NN=23,98%) e 4-Grams=J48(55,30%). Em segundo lugar, encontramos resultados bem próximos entre as representações das cifras e do áudio, (a partir da transição dos acordes): SVM=20,90% e 19,35%, 1-NN=20,43% e 23,61%, 3-NN=18,71% e 19,58%, 5-NN=19,08% e 19,91%. Ao se analisar as representações das cifras, do áudio e das letras para o método sem combinações, conclui-se que esses descritores não obtiveram resultados satisfatórios para as seis emoções.

Tabela 18: Sumário dos resultados utilizando o método *no fusion* com seis emoções.

Conjunto de Características	SVM	1-NN	3-NN	5-NN	J48
Presença e ausência dos acordes	17,25%	18,78%	19,44%	18,71%	16,78%
Tom da música	18,08%	18,03%	17,03%	17,01%	18,03%
Número de repetições dos acordes	13,95%	21,71%	18,91%	19,40%	20,26%
Frequência dos acordes	12,21%	17,10%	18,91%	20,63%	17,65%
Transição dos acordes	20,90%	20,43%	18,71%	19,08%	18,66%
SSD	19,35%	23,61%	19,58%	19,91%	23,51%
2-Grams	34,86%	29,00%	27,20%	23,98%	21,51%
3-Grams	37,76%	27,03%	22,33%	22,46%	25,01%
4-Grams	37,18%	18,83%	11,53%	08,71%	27,33%
Stemm	07,20%	07,20%	07,20%	07,20%	07,20%

Fonte: Autoria própria.

A tabela 19 apresenta os melhores resultados individuais utilizando o método *no fusion* obtidos para MLMMD com seis emoções, a partir das representações das cifras, do áudio e das letras. Analisando os resultados por intermédio dos experimentos realizados é possível obter os melhores resultados individuais da classificação das três representações para cada emoção: alegria (3-Grams=47,30%), amor (2-Grams=52,40%), entusiasmo (3-Grams=61,00%), paixão (Tom=45,80%) e tristeza (3-Grams=31,60%). Novamente, a melhor representação sem combinação dos experimentos realizados para a base MLMMD foi extraindo características (2-grams, 3-grams e 4-grams) das letras das músicas. Ao analisar as emoções individuais, as emoções amor, entusiasmo e paixão foram as mais fáceis de serem classificadas por qualquer uma das representações e dos classificadores empregados para as seis emoções. No entanto, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção.

Tabela 19: Melhores resultados obtidos utilizando o método *no fusion* para representações individuais com seis emoções.

Emoção	Cifras					Áudio		Letras			
	PA	T	R	F	TR	SSD	2-Gram	3-Gram	4-Gram	Stemm	
Alegria	18,50%	10,50%	22,60%	19,60%	26,70%	29,10%	41,20%	47,30%	46,20%	00,00%	
Amor	42,80%	42,90%	32,30%	36,40%	43,70%	38,20%	52,40%	50,50%	48,10%	43,20%	
Decepção	00,00%	00,00%	00,00%	00,00%	00,00%	00,00%	00,00%	00,00%	00,00%	00,00%	
Entusiasmo	11,80%	08,30%	23,70%	24,20%	24,20%	38,10%	54,20%	61,00%	59,10%	00,00%	
Paixão	33,90%	45,80%	37,40%	38,20%	39,00%	33,20%	40,00%	39,20%	43,30%	00,00%	
Tristeza	16,80%	01,60%	18,40%	12,90%	08,10%	26,70%	27,70%	31,60%	26,40%	00,00%	

PA = Presença e ausência; T = Tom; R = Repetição; F = Frequência; TR = Transição

Fonte: Autoria própria.

4.3 COMPARAÇÕES EXPERIMENTAIS

Nesse capítulo foram apresentados os resultados experimentais envolvendo três representações: cifras, áudio e letras, para a detecção automática de emoções em músicas. Nos trabalhos relacionados, não foi encontrado nenhum trabalho que utilize a cifra como fonte de informação. Utilizando o áudio como fonte de informação, no trabalho de (LU et al., 2006), foi utilizado um modelo bidimensional de emoções, para uma base contendo 250 músicas de Johann Strauus, com cinco rótulos: contentamento, depressão, exuberante, ansioso e frenético. Os resultados mostraram uma média de 77,80% de acertos, para contentamento=71,30%, depressão=94,50%, exuberante=71,20%, ansioso/frenético=77,80%.

No trabalho de (YANG; LEE, 2009), as letras das músicas foram utilizadas como fonte de informação juntamente com um modelo de emoções básicas para classificar 23 emoções específicas, para uma base contendo 1.032 músicas retiradas da *allmusic.com*. Os resultados mostraram uma média de 67,00% de acertos. Nenhum dos trabalhos analisados corroboraram com os experimentos realizados, os trabalhos analisados utilizaram bases de dados e modelos de emoções diferentes dos utilizados nesse trabalho, não permitindo assim uma comparação direta entre os mesmos.

5 EXPERIMENTOS COM COMBINAÇÕES UNIMODAIS

Os experimentos desse capítulo têm como objetivo analisar as combinações unimodais utilizando os métodos de *early fusion* e *late fusion*, isto é, usando diferentes procedimentos de extração de características da mesma fonte de informação. Pretende-se comparar os resultados das representações unimodais e os resultados obtidos no capítulo 5, onde foram realizados experimentos com as representações individuais sem nenhum tipo de combinação.

O tamanho do vetor unimodal para a representação das cifras é de 3.969 atributos, já o tamanho do vetor unimodal para a representação das letras é de 18.267 atributos.

5.1 RESULTADOS UNIMODAIS COM TRÊS EMOÇÕES

A tabela 20 está organizada de forma a apresentar os métodos das combinações *early fusion* e *late fusion* utilizando uma única fonte de informação com diferentes métodos de extração de características (unimodal), obtidos a partir das informações das cifras e das letras. Para as cifras combinam-se as seguintes características: presença e ausência de acordes, tom da música, número de repetições dos acordes, frequência dos acordes e transição dos acordes. Para as letras combinam-se as seguintes características: (2-Grams, 3-Grams, 4-Grams e Stem) com *stopwords*, ambas representações para três emoções. Por meio do símbolo de # é especificada a numeração do experimento realizado, seguido pela fonte de informação (cifras, áudio ou letras), pelas características analisadas para cada fonte de informação e pelos algoritmos utilizados para classificar as classes de emoções. Foram utilizadas duas médias, a média ponderada da *F-measure* (MP-F) e a média aritmética da *F-measure* (MA-F). A escolha pela utilização de duas médias ocorreu devido o desbalanceamento das classes, aspecto que pode estar prejudicando o desempenho dos classificadores na MLMMD. A análise dos resultados é realizada por intermédio da média aritmética da *F-measure* (MA-F).

Tabela 20 – Resultados das combinações unimodais para três emoções. (continua)

#	Fonte de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
1	Cifras	<i>Early fusion</i>	Positiva	44,80%	38,60%	41,80%	41,10%	38,50%
			Neutra	67,70%	57,50%	59,50%	63,00%	63,60%
			Negativa	11,40%	17,90%	16,90%	14,90%	07,10%
			MP-F	52,30%	45,80%	47,70%	49,10%	47,40%
			MA-F	41,30%	38,00%	39,40%	39,66%	36,40%

Tabela 20 – Resultados das combinações unimodais para três emoções. (continuação)

#	Fonte de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
2	Cifras	<i>Late fusion - Max</i>	Positiva	45,60%	42,90%	46,00%	41,40%	42,90%
			Neutra	70,10%	63,60%	68,50%	70,00%	63,60%
			Negativa	11,70%	14,30%	02,60%	00,00%	14,30%
			MP-F	54,10%	50,10%	52,00%	51,10%	50,10%
			MA-F	42,46%	40,26%	39,03%	37,13%	40,26%
3	Cifras	<i>Late fusion - Prod</i>	Positiva	23,60%	40,70%	42,30%	38,70%	40,70%
			Neutra	70,90%	67,10%	68,80%	70,30%	67,10%
			Negativa	00,00%	10,60%	04,40%	01,70%	10,60%
			MP-F	46,10%	50,80%	51,30%	50,70%	50,80%
			MA-F	31,50%	39,46%	38,50%	36,90%	39,46%
4	Cifras	<i>Late fusion - Sum</i>	Positiva	17,90%	43,40%	42,00%	40,60%	43,40%
			Neutra	71,60%	70,50%	70,10%	70,90%	70,50%
			Negativa	00,00%	00,00%	01,60%	00,00%	00,00%
			MP-F	44,70%	51,90%	51,60%	51,30%	51,90%
			MA-F	41,76%	37,96%	37,90%	37,16%	37,96%
5	Letras	<i>Early fusion</i>	Positiva	74,00%	38,10%	18,70%	03,30%	58,50%
			Neutra	77,00%	70,00%	70,90%	70,00%	64,70%
			Negativa	25,80%	16,80%	14,00%	04,30%	21,20%
			MP-F	68,40%	52,20%	46,40%	39,80%	56,30%
			MA-F	58,93%	41,63%	34,53%	25,86%	49,13%
6	Letras	<i>Late fusion - Max</i>	Positiva	00,00%	00,00%	40,40%	36,00%	45,60%
			Neutra	70,70%	70,70%	70,80%	72,20%	61,30%
			Negativa	00,00%	00,00%	03,20%	01,70%	24,30%
			MP-F	38,80%	38,80%	51,70%	50,90%	51,10%
			MA-F	23,56%	23,56%	38,13%	36,63%	43,73%
7	Letras	<i>Late fusion - Prod</i>	Positiva	70,10%	70,10%	51,50%	49,00%	35,20%
			Neutra	78,90%	78,90%	74,10%	74,30%	71,60%
			Negativa	00,00%	00,00%	04,70%	00,00%	13,30%
			MP-F	64,70%	64,70%	57,10%	55,80%	52,00%
			MA-F	49,66%	49,66%	43,43%	41,10%	40,03%
8	Letras	<i>Late fusion - Sum</i>	Positiva	69,00%	69,00%	52,30%	47,20%	58,60%
			Neutra	77,70%	77,70%	74,80%	74,20%	71,10%
			Negativa	11,70%	11,70%	06,40%	00,00%	19,80%
			MP-F	65,50%	65,50%	58,00%	55,10%	59,80%
			MA-F	52,80%	49,46%	44,50%	40,46%	49,83%

Nota-se que em todos os experimentos para três emoções que ao analisar as emoções, individualmente, a emoção neutra foi a mais fácil de ser classificada por qualquer um dos classificadores empregados.

O Experimento # 1: apresenta os resultados da classificação unimodal utilizando o método de combinação *early fusion* com as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=44,80%), neutra (SVM=67,70%) e negativa (1-NN=17,90%). Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=41,30%.

O Experimento # 2: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Max* com as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (3-NN=46,00%), neutra (SVM=70,10%) e negativa (1-NN, J48=14,30%). Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=42,46%.

O Experimento # 3: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Prod* com as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (3-NN=42,30%), neutra (SVM=70,90%) e negativa (1-NN, J48=10,60%). Os melhores resultados para essa representação foram obtidos pelos classificadores 1-NN e J48 ambos com MA-F=39,46%.

O Experimento # 4: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Sum* com as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (1-NN, J48=43,40%), neutra (SVM=71,60%). No entanto, ela também mostra que todos classificadores falharam ao reconhecer a emoção negativa. Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=41,76%.

O Experimento # 5: apresenta os resultados da classificação unimodal utilizando o método de combinação *early fusion* com as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=74,00%), neutra (SVM=77,00%) e negativa (SVM=25,80%). Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=58,93%.

O Experimento # 6: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Max* com as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (J48=45,60%), neutra (5-NN=72,20%) e negativa (J48= 24,30%). Os melhores resultados para essa representação foram obtidos pelo classificador J48 com MA-F=43,73%.

O Experimento # 7: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Prod* com as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM, 1-NN=70,10%), neutra (SVM, 1-NN=78,90%) e negativa (J48=13,30%). Os melhores resultados para essa representação foram obtidos pelos classificadores SVM e 1-NN ambos com MA-F=49,66%.

O Experimento # 8: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Sum* com as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM, 1-NN=69,00%), neutra (SVM, 1-NN=77,70%) e negativa (SVM, 1-NN=11,70%). Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=52,80%.

5.1.1 SUMÁRIO DOS RESULTADOS UNIMODAIS COM TRÊS EMOÇÕES

A tabela 21 apresenta o sumário com os resultados da média aritmética da *F-measure* obtidos a partir do método *early fusion e late fusion* para as combinações unimodais das cifras e das letras com os diferentes classificadores e utilizando três emoções. A análise das tabelas mostram alguns resultados interessantes. Em primeiro lugar, verificou-se um certo equilíbrio entre os melhores resultados da média aritmética da *F-measure*, utilizando o método *early fusion* as cifras obtiveram melhores resultados nos classificadores: 3-NN=39,40%, 5-NN=39,66% e as letras obtiveram melhores resultados entre os classificadores: SVM=58,93%, 1-NN=41,63% e J48=49,13%. Utilizando o método *late fusion* as cifras obtiveram melhores resultados nos classificadores: SVM-MAX=42,26%, 1-NN-MAX=40,26%, 3-NN-MAX=39,03%, 5-NN-MAX=37,13% e as letras obtiveram melhores resultados entre os classificadores: J48-MAX=43,73%, SVM-PROD= 49,66%, 1-NN-PROD= 49,66%, 3-NN-PROD=43,43%, 5-NN-PROD=41,10%, J48-PROD=40,03%, SVM-SUM= 52, 80%, 1-NN-SUM=49,46%, 3-NN-SUM=44,50%, 5-NN-SUM= 40,46% e J48-SUM=49,83%. Em segundo lugar, ao se analisar os métodos de combinações unimodais *early e late fusion* conclui-se que as letras apresentam descritores que são úteis para a detecção das emoções na base MLMMD.

5.2 RESULTADOS UNIMODAIS COM SEIS EMOÇÕES

A tabela 22 está organizada de forma a apresentar os métodos das combinações *early fusion e late fusion* utilizando uma única fonte de informação com diferentes métodos de extração de características (unimodal), obtidos a partir das informações das cifras e das letras. Para as cifras combinam-se as seguintes características: presença e ausência de acordes, tom da música, número de repetições dos acordes, frequência dos acordes e transição dos acordes.

Tabela 21: Sumário dos resultados Unimodais com três emoções.

Fonte de Informação	Combinações	SVM	1-NN	3-NN	5-NN	J48
Cifras	<i>Early Fusion</i>	41,30%	38,00%	39,40%	39,66%	36,40%
	<i>Late Fusion (Max)</i>	42,46%	40,26%	39,03%	37,13%	40,26%
	<i>Late Fusion (Prod)</i>	41,76%	37,96%	37,90%	37,16%	37,96%
	<i>Late Fusion (Sum)</i>	31,50%	39,46%	38,50%	36,90%	39,46%
Letras	<i>Early Fusion</i>	58,93%	41,63%	34,53%	25,86%	49,13%
	<i>Late Fusion (Max)</i>	23,56%	23,56%	38,13%	36,63%	43,73%
	<i>Late Fusion (Prod)</i>	52,80%	49,46%	44,50%	40,46%	49,83%
	<i>Late Fusion (Sum)</i>	49,66%	49,66%	43,43%	41,10%	40,03%
<i>Best No Fusion SVM 4-Grams</i>		58,76%				

Fonte: Autoria própria.

Para as letras combinam-se as seguintes características: (2-Grams, 3-Grams, 4-Grams e Stem) com *stopwords*, ambas representações para três emoções. Por meio do símbolo de # é especificada a numeração do experimento realizado, seguido pela fonte de informação (cifras, áudio ou letras), pelas características analisadas para cada fonte de informação e pelos algoritmos utilizados para classificar as classes de emoções. Foram utilizadas duas médias, a média ponderada da *F-measure* (MP-F) e a média aritmética da *F-measure* (MA-F). A análise dos resultados é realizada por intermédio da média aritmética da *F-measure* (MA-F).

Tabela 22 – Resultados das combinações unimodais para seis emoções. (continua)

#	Fonte de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
1	Cifras	<i>Early fusion</i>	Alegria	24,10%	19,30%	21,70%	24,20%	18,80%
			Amor	41,70%	34,50%	37,20%	36,10%	34,20%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	23,00%	18,30%	15,90%	20,90%	13,90%
			Paixão	35,60%	30,20%	28,20%	31,40%	28,40%
			Tristeza	07,50%	23,80%	11,00%	04,30%	12,30%
			MP-F	29,20%	26,50%	25,00%	25,80%	23,60%
			MA-F	21,98%	21,01%	19,00%	19,48%	17,93%
2	Cifras	<i>Late fusion - Max</i>	Alegria	09,40%	16,40%	16,70%	14,50%	16,40%
			Amor	31,10%	29,00%	38,40%	35,20%	29,00%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	07,40%	16,10%	21,30%	18,80%	16,10%
			Paixão	35,40%	31,70%	39,20%	41,10%	31,70%
			Tristeza	00,00%	15,00%	12,90%	08,00%	15,00%
			MP-F	20,30%	23,30%	28,50%	26,70%	23,30%

Tabela 22 – Resultados das combinações unimodais para seis emoções. (continuação)

#	Fonte de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48			
			MA-F	13,88%	18,03%	21,41%	19,60%	18,03%			
3	Cifras	<i>Late fusion - Prod</i>	Alegria	03,30%	12,60%	18,00%	16,70%	12,60%			
			Amor	46,10%	23,80%	40,70%	41,70%	23,80%			
			Decepção	00,00%	02,60%	00,00%	00,00%	02,60%			
			Entusiasmo	11,00%	11,30%	18,90%	10,90%	11,30%			
			Paixão	42,30%	26,80%	42,20%	39,20%	26,80%			
			Tristeza	00,00%	12,10%	11,30%	14,10%	12,10%			
			MP-F	26,00%	18,90%	29,50%	27,80%	18,90%			
			MA-F	17,11%	14,86%	21,85%	20,43%	14,86%			
			4	Cifras	<i>Late fusion - Sum</i>	Alegria	03,20%	17,60%	12,40%	12,40%	17,60%
						Amor	43,50%	36,20%	41,50%	41,50%	36,20%
Decepção	00,00%	00,00%				00,00%	00,00%	00,00%			
Entusiasmo	06,70%	13,60%				14,00%	14,00%	13,60%			
Paixão	42,50%	37,30%				41,20%	41,20%	37,30%			
Tristeza	00,00%	11,70%				06,90%	06,90%	11,70%			
MP-F	24,60%	26,00%				27,20%	27,20%	26,00%			
MA-F	15,98%	19,40%				19,33%	19,33%	19,40%			
5	Letras	<i>Early fusion</i>				Alegria	43,80%	22,60%	27,80%	19,00%	30,40%
						Amor	48,50%	43,60%	45,70%	45,00%	39,70%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%			
			Entusiasmo	60,50%	26,00%	01,60%	01,60%	45,70%			
			Paixão	42,80%	24,10%	08,50%	10,60%	31,80%			
			Tristeza	32,50%	12,00%	09,90%	05,40%	22,10%			
			MP-F	45,10%	27,50%	20,60%	19,00%	34,00%			
			MA-F	38,01%	21,38%	15,58%	13,60%	28,28%			
			6	Letras	<i>Late fusion - Max</i>	Alegria	00,00%	00,00%	29,90%	20,60%	14,80%
						Amor	38,80%	38,80%	44,90%	44,20%	28,40%
Decepção	00,00%	00,00%				00,00%	00,00%	00,00%			
Entusiasmo	00,00%	00,00%				16,30%	13,80%	38,30%			
Paixão	31,50%	31,50%				16,10%	14,30%	32,70%			
Tristeza	00,00%	00,00%				09,50%	04,90%	22,10%			
MP-F	18,80%	18,80%				24,60%	21,50%	27,90%			
MA-F	11,71%	11,71%				19,45%	16,30%	22,71%			
7	Letras	<i>Late fusion - Prod</i>				Alegria	40,00%	40,00%	42,40%	42,40%	00,00%
						Amor	48,90%	48,90%	52,20%	50,40%	09,90%

Tabela 22 – Resultados das combinações unimodais para seis emoções. (continuação)

#	Fonte de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
			Decepção	00,00%	00,00%	00,00%	00,00%	03,80%
			Entusiasmo	51,10%	51,10%	24,40%	19,30%	08,80%
			Paixão	43,80%	43,80%	24,90%	24,20%	16,40%
			Tristeza	06,60%	06,60%	08,50%	04,70%	11,40%
			MP-F	39,80%	39,80%	31,80%	29,80%	10,20%
			MA-F	31,73%	31,73%	25,40%	23,50%	08,38%
			Alegria	42,60%	42,60%	40,80%	38,80%	20,40%
			Amor	49,00%	49,00%	51,00%	50,10%	39,80%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
8	Letras	<i>Late fusion - Sum</i>	Entusiasmo	52,40%	52,40%	17,40%	12,80%	49,10%
			Paixão	45,20%	45,20%	22,70%	21,30%	31,30%
			Tristeza	12,30%	12,30%	08,70%	08,80%	27,80%
			MP-F	41,60%	41,60%	29,60%	27,80%	34,00%
			MA-F	33,58%	33,58%	23,43%	21,96%	28,06%

Fonte: Autoria própria.

Nota-se que todos os classificadores falharam no reconhecimento da emoção decepção.

O Experimento # 1: apresenta os resultados da classificação unimodal utilizando o método de combinação *early fusion* com as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (5-NN=24,20%), amor (SVM= 41,70%), entusiasmo (SVM= 23,00%), paixão (SVM=35,60%) e tristeza (1-NN=23,80%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=21,98%.

O Experimento # 2: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Max* com as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=16,70%), amor (3-NN=38,40%), entusiasmo (3-NN=21,30%), paixão (3-NN= 39,20%) e tristeza (1-NN, J48=15,00%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essa representação foram obtidos pelo classificador 3-NN com MA-F=21,41%.

O Experimento # 3: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Prod* com as cifras. Os melhores resultados desse experimento

por emoção (em negrito na tabela) foram: alegria (3-NN=18,00%), amor (SVM=46,10%), entusiasmo (3-NN=18,90%), paixão (SVM= 42,30%) e tristeza (5-NN=14,10%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essa representação foram obtidos pelo classificador 3-NN com MA-F=21,85%.

O Experimento # 4: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Sum* com as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (1-NN, J48=17,60%), amor (SVM=43,50%), entusiasmo (3-NN, 5-NN=14,00%), paixão (SVM=42,50%) e tristeza (1-NN, J48=11,70%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essa representação foram obtidos pelos classificadores SVM e J48 ambos com MA-F=19,40%.

O Experimento # 5: apresenta os resultados da classificação unimodal utilizando o método de combinação *early fusion* com as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (SVM=43,80%), amor (SVM=48,50%), entusiasmo (SVM=60,50%), paixão (SVM= 42,80%) e tristeza (SVM=32,50%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Os melhores resultados para essa representação foram obtidos pelo classificador SVM com MA-F=38,01%.

O Experimento # 6: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Max* com as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=29,90%), amor (3-NN=44,90%), entusiasmo (J48=38,30%), paixão (J48=32,70%) e tristeza (J48=22,10%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essa representação foram obtidos pelo classificador J48 com MA-F=22,71%.

O Experimento # 7: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Prod* com as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN, 5-NN=42,40%), amor (3-NN=52,20%), entusiasmo (SVM, 1-NN=51,10%), paixão (SVM, 1-NN=43,80%) e tristeza (J48=11,40%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essa representação foram obtidos pelos classificadores SVM e 1-NN ambos com MA-F=31,73%.

O Experimento # 8: apresenta os resultados da classificação unimodal utilizando o método de combinação *late fusion - Sum* com as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (SVM, 1-NN=42,60%), amor

(3-NN=51,00%), entusiasmo (SVM, 1-NN=52,40%), paixão (SVM, 1-NN=45,20%) e tristeza (J48=27,80%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Os melhores resultados para essa representação foram obtidos pelos classificadores SVM e 1-NN ambos com MA-F=33,58%.

5.2.1 SUMÁRIO DOS RESULTADOS UNIMODAIS COM SEIS EMOÇÕES

A tabela 23 apresenta o sumário com os resultados da média aritmética da *F-measure* obtidos a partir do método *early fusion* e *late fusion* para as combinações unimodais das cifras e das letras com os diferentes classificadores e utilizando seis emoções. A análise das tabelas mostram alguns resultados interessantes. Em primeiro lugar, de modo geral, verificou-se um certo equilíbrio entre os resultados da média aritmética da *F-measure*, utilizando o método *early fusion* as cifras obtiveram melhores resultados nos classificadores: 3-NN=19,00%, 5-NN=19,48% e as letras obtiveram melhores resultados entre os classificadores: SVM=38,01%, 1-NN=21,38% e J48=28,28%. Utilizando o método *late fusion* as cifras obtiveram melhores resultados nos classificadores e nas regras de decisão: SVM-MAX=13,88%, 1-NN-MAX=18,03%, 3-NN-MAX=21,41%, 5-NN-MAX=19,60%, J48-PROD=14,86%, e as letras obtiveram melhores resultados entre os classificadores e as regras de decisão: J48-MAX=22,71%, SVM-PROD=31,73%, 1-NN-PROD=31,73%, 3-NN-PROD=25,40%, 5-NN-PROD=23,50%, SVM-SUM=33,58%, 1-NN-SUM=33,58%, 3-NN-SUM=23,43%, 5-NN-SUM=21,96% e J48-SUM=28,06%. Em segundo lugar, ao se analisar os métodos de combinações unimodais *early* e *late fusion* das cifras e das letras, conclui-se que as combinações realizadas não obtiveram resultados satisfatórios para as seis emoções.

Tabela 23: Sumário dos resultados unimodais com seis emoções.

Fonte de Informação	Combinações	SVM	1-NN	3-NN	5-NN	J48
Cifras	<i>Early Fusion</i>	21,98%	21,01%	19,00%	19,48%	17,93%
	<i>Late Fusion</i> (Max)	13,88%	18,03%	21,41%	19,60%	18,03%
	<i>Late Fusion</i> (Prod)	17,11%	14,86%	21,85%	20,43%	14,86%
	<i>Late Fusion</i> (Sum)	15,98%	19,40%	19,33%	19,33%	19,40%
Letras	<i>Early Fusion</i>	38,01%	21,38%	15,58%	13,60%	28,28%
	<i>Late Fusion</i> (Max)	11,71%	11,71%	19,45%	16,30%	22,71%
	<i>Late Fusion</i> (Prod)	31,73%	31,73%	25,40%	23,50%	08,38%
	<i>Late Fusion</i> (Sum)	33,58%	33,58%	23,43%	21,96%	28,06%
<i>Best No Fusion SVM 3-Grams</i>		37,76%				

Fonte: Autoria própria.

5.3 COMPARAÇÕES EXPERIMENTAIS

Se comparado ao método *no fusion* as combinações unimodais tiveram um desempenho inferior em quase todos os classificadores, embora tenha se mantido um certo equilíbrio de resultados nas combinações unimodais e sem combinações. O melhor resultado para as três emoções ocorreu utilizando o método *early fusion* das letras SVM=58,93%, bem como o melhor resultado para seis emoções SVM=38,01%. Comparando os melhores resultados unimodais com os resultados sem combinações, temos, sem combinações com SVM-4-Grams=58,76% para três emoções e SVM-3-Grams=37,76% para seis emoções. Não foram encontrados trabalhos utilizando a representação das cifras em combinações unimodais. Não foram encontrados trabalhos com combinações unimodais envolvendo as letras.

6 EXPERIMENTOS COM COMBINAÇÕES MULTIMODAIS

Os experimentos desse capítulo tem como objetivo analisar as combinações, multimodais *early fusion* e *late fusion* a partir de três fontes de informações musicais: cifras (presença e ausência, tom, repetição, frequência e transição), áudio e letras (2-Grams, 3-Grams, 4-Grams e Stemm) e com *stopwords*, com três e seis emoções. Pretende-se comparar os resultados das representações multimodais com as representações unimodais e os resultados obtidos sem combinações.

O tamanho do vetor de características multimodal para cada combinação das representações é de: Áudio+Cifras 4.131 atributos; Áudio+Letras 18.428 atributos; Cifras+Letras 22.237 atributos; Áudio+Cifras+Letras 22.445 atributos.

6.1 RESULTADOS MULTIMODAIS COM TRÊS EMOÇÕES

A tabela 24 está organizada de forma a apresentar os métodos de combinações *early fusion* e *late fusion* utilizando três fontes de informações com diferentes métodos de extração de características (multimodal), obtidos a partir das informações do áudio com as cifras, das letras com o áudio, das cifras com as letras e da junção do áudio com as cifras e com as letras, para as três emoções. Por meio do símbolo de # é especificada a numeração do experimento realizado, seguido pela fonte de informação (cifras, áudio ou letras), pelas características analisadas para cada fonte de informação e pelos algoritmos utilizados para classificar as classes de emoções. Foram utilizadas duas médias, a média ponderada da *F-measure* (MP-F) e a média aritmética da *F-measure* (MA-F). A escolha pela utilização de duas médias ocorreu devido o desbalanceamento das classes, aspecto que pode estar prejudicando o desempenho dos classificadores na MLMMD. A análise dos resultados é realizada por intermédio da média aritmética da *F-measure* (MA-F).

Tabela 24 – Resultados das combinações multimodais para três emoções.

(continua)

#	Fontes de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
1	Áudio e Cifras	<i>Early fusion</i>	Positiva	52,00%	45,50%	44,20%	45,20%	48,40%
			Neutra	68,30%	58,50%	58,60%	60,90%	59,30%
			Negativa	14,60%	16,70%	18,20%	45,20%	09,40%
			MP-F	55,20%	48,20%	48,10%	49,20%	48,50%
			MA-F	44,96%	40,23%	40,33%	50,43%	39,03%

Tabela 24 – Resultados das combinações multimodais para três emoções. (continuação)

#	Fontes de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
2	Áudio e Cifras	<i>Late fusion - Max</i>	Positiva	45,60%	46,40%	51,00%	46,80%	41,70%
			Neutra	70,10%	61,90%	69,10%	69,30%	58,60%
			Negativa	11,70%	16,60%	01,40%	00,00%	22,60%
			MP-F	54,10%	50,60%	53,70%	52,30%	48,20%
			MA-F	42,46%	41,63%	40,50%	38,70%	40,96%
3	Áudio e Cifras	<i>Late fusion - Prod</i>	Positiva	14,30%	48,70%	48,20%	42,30%	34,60%
			Neutra	70,80%	68,00%	70,70%	70,10%	66,20%
			Negativa	00,00%	10,30%	01,60%	03,30%	11,50%
			MP-F	43,20%	53,70%	53,70%	51,90%	48,60%
			MA-F	28,36%	42,33%	40,16%	38,56%	37,43%
4	Áudio e Cifras	<i>Late fusion - Sum</i>	Positiva	09,10%	48,80%	46,90%	43,40%	46,60%
			Neutra	70,80%	68,10%	71,60%	71,40%	71,30%
			Negativa	00,00%	08,90%	00,00%	00,00%	01,60%
			MP-F	41,60%	53,60%	53,60%	52,50%	53,60%
			MA-F	26,63%	41,93%	39,50%	38,26%	39,83%
5	Letras e Áudio	<i>Early fusion</i>	Positiva	73,80%	34,40%	10,80%	04,80%	58,60%
			Neutra	75,70%	70,30%	70,90%	70,30%	64,90%
			Negativa	30,50%	18,80%	13,50%	07,10%	21,40%
			MP-F	68,30%	51,60%	43,90%	40,90%	56,50%
			MA-F	60,00%	41,16%	31,73%	27,40%	48,30%
6	Letras e Áudio	<i>Late fusion - Max</i>	Positiva	00,00%	44,80%	48,40%	44,50%	41,30%
			Neutra	70,70%	59,80%	70,70%	71,60%	59,60%
			Negativa	00,00%	24,30%	03,10%	00,00%	22,50%
			MP-F	38,80%	50,10%	54,10%	52,90%	48,60%
			MA-F	23,56%	42,96%	40,73%	38,70%	41,13%
7	Letras e Áudio	<i>Late fusion - Prod</i>	Positiva	72,40%	59,90%	57,20%	53,90%	21,00%
			Neutra	79,50%	73,00%	75,20%	74,70%	71,80%
			Negativa	00,00%	13,20%	03,30%	00,00%	05,00%
			MP-F	65,80%	60,30%	59,20%	57,50%	46,50%
			MA-F	50,63%	48,70%	45,23%	42,86%	32,60%
8	Letras e Áudio	<i>Late fusion - Sum</i>	Positiva	71,80%	58,50%	55,30%	51,80%	60,00%
			Neutra	78,90%	73,20%	75,20%	74,60%	71,30%
			Negativa	07,90%	09,70%	01,70%	00,00%	17,50%
			MP-F	66,40%	59,40%	58,40%	56,80%	60,00%
			MA-F	52,86%	47,13%	44,06%	42,13%	49,60%

Tabela 24 – Resultados das combinações multimodais para três emoções. (continuação)

#	Fontes de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
9	Cifras e Letras	<i>Early fusion</i>	Positiva	73,90%	37,30%	09,40%	03,30%	59,80%
			Neutra	76,40%	70,60%	70,50%	70,50%	67,30%
			Negativa	29,70%	15,30%	10,50%	04,70%	22,70%
			MP-F	68,60%	52,10%	42,80%	40,10%	58,30%
			MA-F	60,00%	41,06%	30,13%	26,16%	49,93%
10	Cifras e Letras	<i>Late fusion - Max</i>	Positiva	30,90%	46,40%	47,80%	46,20%	43,20%
			Neutra	71,70%	62,00%	71,20%	73,20%	59,50%
			Negativa	00,00%	17,50%	01,60%	00,00%	20,00%
			MP-F	48,80%	50,70%	53,90%	54,30%	48,80%
			MA-F	34,20%	41,96%	40,20%	39,80%	40,90%
11	Cifras e Letras	<i>Late fusion - Prod</i>	Positiva	41,60%	57,10%	50,10%	51,00%	21,20%
			Neutra	74,20%	73,30%	74,20%	75,20%	71,00%
			Negativa	00,00%	11,90%	50,10%	00,00%	04,90%
			MP-F	53,40%	59,40%	56,80%	56,80%	46,20%
			MA-F	38,60%	47,43%	58,13%	42,06%	32,36%
12	Cifras e Letras	<i>Late fusion - Sum</i>	Positiva	31,90%	56,40%	49,90%	47,40%	58,60%
			Neutra	73,30%	73,20%	75,00%	74,90%	75,80%
			Negativa	00,00%	10,30%	00,00%	00,00%	04,90%
			MP-F	50,00%	58,90%	56,40%	55,60%	60,20%
			MA-F	35,06%	46,63%	41,63%	40,76%	46,43%
13	Áudio, Cifras e Letras	<i>Early fusion</i>	Positiva	74,00%	35,00%	10,90%	04,10%	59,60%
			Neutra	77,40%	70,60%	70,90%	70,50%	67,10%
			Negativa	30,40%	19,10%	12,20%	07,80%	23,60%
			MP-F	69,30%	52,00%	43,80%	40,80%	58,20%
			MA-F	60,60%	41,56%	31,33%	27,46%	50,10%
14	Áudio, Cifras e Letras	<i>Late fusion - Max</i>	Positiva	30,90%	46,40%	50,30%	49,90%	41,50%
			Neutra	71,70%	62,00%	70,90%	72,50%	59,70%
			Negativa	00,00%	17,50%	01,60%	00,00%	10,90%
			MP-F	48,80%	50,70%	54,50%	55,10%	48,50%
			MA-F	34,20%	41,96%	40,93%	40,80%	37,36%
15	Áudio, Cifras e Letras	<i>Late fusion - Prod</i>	Positiva	41,00%	58,00%	53,40%	51,80%	14,90%
			Neutra	74,10%	73,30%	74,70%	75,20%	71,20%
			Negativa	00,00%	07,80%	01,70%	00,00%	03,40%
			MP-F	53,20%	59,10%	57,60%	57,10%	42,90%
			MA-F	38,36%	46,36%	43,26%	42,33%	29,83%

Tabela 24 – Resultados das combinações multimodais para três emoções. (continuação)

#	Fontes de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
16	Áudio, Cifras e Letras	<i>Late fusion - Sum</i>	Positiva	29,20%	58,60%	53,10%	50,60%	59,10%
			Neutra	73,20%	73,80%	75,50%	75,40%	75,50%
			Negativa	00,00%	07,80%	00,00%	00,00%	04,90%
			MP-F	49,10%	59,50%	57,70%	56,80%	60,20%
			MA-F	34,13%	46,73%	42,86%	42,00%	46,50%

Fonte: Autoria própria.

Nota-se que em todos os experimentos para três emoções que ao analisar as emoções, individualmente, a emoção neutra foi a mais fácil de ser classificada por qualquer um dos classificadores empregados.

O Experimento # 1: apresenta os resultados da classificação multimodal utilizando o método de combinação *early fusion* com o áudio e as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=52,00%), neutra (SVM=68,30%) e negativa (5-NN=45,20%). Os melhores resultados para essas representações foram obtidos pelo classificador 5-NN com MA-F=50,43%.

O Experimento # 2: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Max* com o áudio e as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (3-NN=51,00%), neutra (SVM=70,10%) e negativa (J48=22,60%). Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=42,46%.

O Experimento # 3: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Prod* com o áudio e as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (1-NN=48,70%), neutra (SVM=70,80%) e negativa (J48=11,50%). Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=42,33%.

O Experimento # 4: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Sum* com o áudio e as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (1-NN=48,80%), neutra (3-NN=71,60%) e negativa (1-NN=08,90%). Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=41,93%.

O Experimento # 5: apresenta os resultados da classificação multimodal utilizando o método de combinação *early fusion* com as letras e o áudio. Os melhores resultados desse

experimento por emoção (em negrito na tabela) foram: positiva (SVM=73,80%), neutra (SVM=75,70%) e negativa (SVM=30,50%). Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=60,00%.

O Experimento # 6: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Max* com as letras e o áudio. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (3-NN=48,40%), neutra (5-NN=71,60%) e negativa (1-NN=24,30%). Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=42,96%.

O Experimento # 7: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Prod* com as letras e o áudio. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=72,40%), neutra (SVM=79,50%) e negativa (1-NN=13,20%). Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=50,63%.

O Experimento # 8: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Sum* com as letras e o áudio. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=71,80%), neutra (SVM=78,90%) e negativa (J48=17,50%). Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=52,86%.

O Experimento # 9: apresenta os resultados da classificação multimodal utilizando o método de combinação *early fusion* com as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=73,90%), neutra (SVM=76,40%) e negativa (SVM=29,70%). Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=60,00%.

O Experimento # 10: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Max* com as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (3-NN=47,80%), neutra (5-NN=73,20%) e negativa (J48=20,00%). Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=41,96%.

O Experimento # 11: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Prod* com as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (1-NN=57,10%), neutra (5-NN=75,20%) e negativa (3-NN=50,10%). Os melhores resultados para essas representações foram obtidos pelo classificador 3-NN com MA-F=58,13%.

O Experimento # 12: apresenta os melhores resultados da classificação multimodal *late fusion - Sum* com as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (J48=58,60%), neutra (J48=75,80%) e negativa (1-NN=10,30%). Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=46,63%.

O Experimento # 13: apresenta os resultados da classificação multimodal utilizando o método de combinação *early fusion* com o áudio, as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (SVM=74,00%), neutra (SVM=77,40%) e negativa (SVM=30,40%). Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=60,60%.

O Experimento # 14: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Max* com o áudio, as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (3-NN=50,30%), neutra (5-NN=72,50%) e negativa (1-NN=17,50%). Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=41,96%.

O Experimento # 15: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Prod* com o áudio, as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (1-NN=58,00%), neutra (5-NN=75,20%) e negativa (1-NN=07,80%). Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=46,36%.

O Experimento # 16: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Sum* com o áudio, as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: positiva (J48=59,10%), neutra (3-NN, J48=75,50%) e negativa (1-NN=07,80%). Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=46,73%.

6.1.1 SUMÁRIO DOS RESULTADOS MULTIMODAIS COM TRÊS EMOÇÕES

A tabela 25 apresenta o sumário com os resultados da média aritmética da *F-measure* obtidos a partir do método *early fusion e late fusion* para as combinações multimodais: do áudio com as cifras, das letras com o áudio, das cifras com as letras e das cifras com o áudio e com as letras e com os diferentes classificadores e utilizando três emoções. A análise da tabela 25 mostra alguns resultados interessantes. Em primeiro lugar, verificou-se um certo equilíbrio entre os melhores resultados da média aritmética da *F-measure* entre todas as combinações multimodais. A combinação das cifras com o áudio e com as letras obteve o melhor

resultado *early fusion* SVM=60,60%, seguido pelas letras com o áudio e das cifras com as letras SVM=60,00%. A combinação das cifras com as letras obteve o melhor resultado *late fusion* 1-NN-PROD=58,13% , seguida pela combinação das cifras com o áudio e com as letras: 1-NN-SUM=46,73%, J48-SUM=46,50% e 1-NN-PRD=43,26%. Em segundo lugar, a combinação das cifras com as letras e das letras com o áudio, obtiveram resultados próximos aos das combinações das cifras com o áudio e com as letras, tanto no *early fusion* como no *late fusion*. Os três resultados citados, apresentaram melhores resultados no SVM se comparados com o método unimodal e sem combinação. Em terceiro lugar, analisando por intermédio de um contexto geral as combinações multimodais a partir do classificador SVM melhoraram os resultados unimodais e sem combinações, sendo um bom classificador para as três emoções e para a base MLMMD.

Tabela 25: Sumário dos resultados multimodais com três emoções.

Fontes de Informação	Combinações	SVM	1-NN	3-NN	5-NN	J48
Áudio + Cifras	<i>Early Fusion</i>	44,96%	40,23%	40,33%	50,43%	39,03%
	<i>Late Fusion</i> (Max)	42,46%	41,63%	40,50%	38,70%	40,96%
	<i>Late Fusion</i> (Prod)	28,36%	42,33%	40,16%	38,56%	37,43%
	<i>Late Fusion</i> (Sum)	26,33%	41,93%	39,50%	38,20%	39,83%
Letras + Áudio	<i>Early Fusion</i>	60,00%	41,16%	31,73%	27,04%	48,30%
	<i>Late Fusion</i> (Max)	23,56%	42,96%	40,73%	38,70%	41,13%
	<i>Late Fusion</i> (Prod)	50,63%	48,70%	45,23%	42,86%	32,60%
	<i>Late Fusion</i> (Sum)	52,86%	47,13%	44,06%	42,13%	49,60%
Cifras + Letras	<i>Early Fusion</i>	60,00%	41,06%	30,13%	26,16%	49,93%
	<i>Late Fusion</i> (Max)	34,20%	41,96%	40,20%	39,80%	40,90%
	<i>Late Fusion</i> (Prod)	38,60%	47,43%	58,13%	42,06%	32,36%
	<i>Late Fusion</i> (Sum)	35,06%	46,63%	41,63%	40,76%	46,43%
Cifras + Áudio + Letras	<i>Early Fusion</i>	60,60%	41,56%	31,33%	27,46%	50,10%
	<i>Late Fusion</i> (Max)	34,20%	41,96%	40,93%	40,80%	37,36%
	<i>Late Fusion</i> (Prod)	38,60%	46,36%	43,26%	42,33%	29,83%
	<i>Late Fusion</i> (Sum)	34,14%	46,73%	42,86%	42,00%	46,50%
<i>Best No Fusion</i> SVM 4-Grams		58,76%				
<i>Best Unimodal EF</i> SVM Letras		58,93%				

Fonte: Autoria própria.

6.2 RESULTADOS MULTIMODAIS COM SEIS EMOÇÕES

A tabela 26 está organizada de forma a apresentar os métodos de combinações *early fusion* e *late fusion* utilizando três fontes de informações com diferentes métodos de extração de características (multimodal), obtidos a partir das informações do áudio com as cifras, das letras com o áudio, das cifras com as letras e da junção do áudio com as cifras e com as letras,

para as três emoções. Por meio do símbolo de # é especificada a numeração do experimento realizado, seguido pela fonte de informação (cifras, áudio ou letras), pelas características analisadas para cada fonte de informação e pelos algoritmos utilizados para classificar as classes de emoções. Foram utilizadas duas médias, a média ponderada da *F-measure* (MP-F) e a média aritmética da *F-measure* (MA-F). A escolha pela utilização de duas médias ocorreu devido o desbalanceamento das classes, aspecto que pode estar prejudicando o desempenho dos classificadores na MLMMD. A análise dos resultados é realizada por intermédio da média aritmética da *F-measure* (MA-F).

Tabela 26 – Resultados das combinações multimodais para seis emoções.

(continua)

#	Fontes de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
1	Áudio e Cifras	<i>Early fusion</i>	Alegria	21,80%	20,10%	23,80%	26,90%	19,60%
			Amor	39,70%	31,60%	37,20%	38,50%	33,60%
			Decepção	00,00%	00,00%	00,10%	00,00%	06,90%
			Entusiasmo	34,00%	24,10%	20,90%	29,00%	23,90%
			Paixão	41,00%	33,90%	24,50%	30,50%	30,10%
			Tristeza	13,20%	15,60%	09,90%	04,40%	17,80%
			MP-F	32,20%	26,60%	25,10%	27,90%	26,50%
			MA-F	24,95%	20,88%	17,75%	21,55%	21,98%
2	Áudio e Cifras	<i>Late fusion - Max</i>	Alegria	09,40%	18,80%	21,90%	19,80%	20,70%
			Amor	31,10%	32,40%	39,70%	36,60%	29,60%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	07,40%	23,90%	24,80%	25,40%	27,80%
			Paixão	35,40%	38,40%	38,40%	39,30%	36,90%
			Tristeza	00,00%	21,40%	12,30%	09,80%	19,60%
			MP-F	20,30%	28,60%	29,80%	28,70%	28,10%
			MA-F	13,88%	22,48%	22,85%	21,81%	22,43%
3	Áudio e Cifras	<i>Late fusion - Prod</i>	Alegria	03,40%	18,40%	18,70%	14,90%	04,90%
			Amor	44,70%	39,10%	39,80%	40,20%	10,40%
			Decepção	00,00%	00,00%	00,00%	00,00%	03,40%
			Entusiasmo	12,20%	27,20%	27,20%	24,10%	11,60%
			Paixão	42,00%	38,90%	41,80%	36,70%	12,50%
			Tristeza	00,00%	15,90%	15,90%	10,70%	10,70%
			MP-F	25,70%	30,20%	31,20%	28,20%	10,30%
			MA-F	17,05%	23,25%	23,90%	21,10%	08,91%
4	Áudio e Cifras	<i>Late fusion - Sum</i>	Alegria	05,10%	19,70%	16,90%	14,30%	17,20%
			Amor	44,40%	37,60%	40,00%	41,60%	33,80%

Tabela 26 – Resultados das combinações multimodais para seis emoções. (continuação)

#	Fontes de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	08,30%	30,50%	24,80%	20,30%	22,00%
			Paixão	42,00%	38,60%	41,40%	43,10%	37,40%
			Tristeza	00,00%	12,90%	07,10%	04,60%	16,50%
			MP-F	25,20%	30,00%	29,30%	27,70%	27,40%
			MA-F	16,63%	23,21%	21,70%	20,65%	21,15%
			Alegria	48,40%	19,50%	24,20%	04,40%	30,40%
			Amor	48,20%	46,50%	45,80%	43,70%	39,00%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
5	Letras e Áudio	<i>Early fusion</i>	Entusiasmo	57,50%	27,20%	07,90%	03,20%	45,70%
			Paixão	41,40%	24,70%	09,90%	08,40%	32,40%
			Tristeza	29,60%	16,00%	10,20%	06,80%	22,30%
			MP-F	44,50%	28,70%	21,50%	16,40%	34,10%
			MA-F	37,51%	22,31%	16,33%	11,08%	28,30%
			Alegria	00,00%	25,40%	30,90%	23,70%	16,60%
			Amor	38,80%	30,40%	45,60%	45,60%	26,40%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
6	Letras e Áudio	<i>Late fusion - Max</i>	Entusiasmo	00,00%	26,10%	30,10%	27,50%	34,10%
			Paixão	31,50%	28,90%	19,50%	18,40%	32,20%
			Tristeza	00,00%	24,10%	08,30%	04,90%	21,50%
			MP-F	18,80%	27,20%	27,90%	25,70%	26,70%
			MA-F	11,71%	22,48%	22,40%	20,01%	21,80%
			Alegria	38,00%	36,40%	43,20%	35,40%	00,00%
			Amor	51,70%	48,10%	51,40%	50,10%	04,50%
			Decepção	00,00%	00,00%	00,00%	00,00%	03,40%
7	Letras e Áudio	<i>Late fusion - Prod</i>	Entusiasmo	49,10%	26,00%	29,80%	22,40%	01,50%
			Paixão	42,80%	32,60%	26,30%	26,70%	09,40%
			Tristeza	03,50%	23,90%	10,60%	06,50%	05,30%
			MP-F	39,30%	34,40%	33,30%	30,20%	04,80%
			MA-F	30,85%	27,83%	26,88%	27,96%	04,10%
			Alegria	39,40%	34,60%	40,50%	41,00%	21,70%
			Amor	50,60%	47,90%	51,60%	49,90%	36,40%
8	Letras e Áudio	<i>Late fusion - Sum</i>	Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	48,30%	28,20%	27,30%	20,00%	46,40%
			Paixão	42,90%	31,50%	24,30%	24,40%	31,50%

Tabela 26 – Resultados das combinações multimodais para seis emoções. (continuação)

#	Fontes de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
9	Cifras e Letras	<i>Early fusion</i>	Tristeza	03,50%	25,40%	04,60%	06,60%	23,80%
			MP-F	39,10%	34,40%	31,20%	29,90%	32,30%
			MA-F	30,78%	33,66%	24,71%	23,65%	26,63%
			Alegria	51,70%	22,70%	25,60%	27,30%	32,60%
			Amor	51,50%	48,60%	45,20%	45,50%	39,80%
			Decepção	00,00%	00,00%	00,00%	00,00%	07,70%
			Entusiasmo	58,40%	24,50%	01,60%	01,60%	44,50%
			Paixão	45,90%	24,30%	09,20%	11,50%	28,80%
			Tristeza	31,00%	19,50%	10,10%	10,30%	23,30%
			MP-F	47,40%	29,70%	20,30%	21,30%	33,70%
			MA-F	39,75%	23,26%	15,28%	16,03%	29,45%
			10	Cifras e Letras	<i>Late fusion - Max</i>	Alegria	09,40%	18,60%
Amor	31,10%	32,10%				45,20%	44,80%	27,20%
Decepção	00,00%	00,00%				00,00%	00,00%	00,00%
Entusiasmo	07,40%	23,40%				22,50%	16,10%	29,60%
Paixão	35,40%	38,90%				27,00%	21,70%	36,10%
Tristeza	00,00%	20,40%				09,10%	06,10%	19,70%
MP-F	20,30%	28,40%				28,30%	24,20%	26,90%
MA-F	13,88%	22,23%				21,93%	18,23%	21,50%
Alegria	14,40%	28,30%				29,10%	32,30%	00,00%
Amor	49,10%	48,80%				49,10%	48,90%	03,70%
Decepção	00,00%	00,00%				00,00%	00,00%	03,40%
11	Cifras e Letras	<i>Late fusion - Prod</i>				Entusiasmo	20,90%	31,30%
			Paixão	42,60%	36,30%	34,00%	27,60%	05,30%
			Tristeza	00,00%	22,50%	09,00%	04,70%	01,80%
			MP-F	30,00%	35,10%	30,70%	28,50%	03,00%
			MA-F	21,16%	27,86%	23,28%	21,76%	02,61%
			Alegria	09,80%	31,70%	25,40%	29,90%	19,10%
			Amor	47,10%	48,40%	48,10%	48,30%	40,70%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	17,20%	33,50%	22,22%	14,70%	42,40%
			Paixão	41,70%	36,30%	30,70%	28,50%	33,60%
			Tristeza	00,00%	18,60%	03,20%	00,00%	19,10%
			MP-F	28,00%	35,30%	28,80%	27,20%	32,30%
MA-F	19,30%	28,41%	21,60%	20,23%	25,81%			

Tabela 26 – Resultados das combinações multimodais para seis emoções.

(continuação)

#	Fontes de Informação	Combinações	Emoção / Médias	SVM	1-NN	3-NN	5-NN	J48
13	Áudio, Cifras e Letras	<i>Early fusion</i>	Alegria	47,80%	20,70%	23,80%	15,00%	31,90%
			Amor	51,30%	48,60%	45,20%	45,10%	39,90%
			Decepção	00,00%	00,00%	00,00%	00,00%	07,70%
			Entusiasmo	59,20%	24,80%	01,60%	04,80%	44,40%
			Paixão	45,00%	24,10%	11,60%	08,90%	29,20%
			Tristeza	26,70%	18,70%	13,60%	12,70%	23,30%
			MP-F	46,10%	29,30%	21,20%	19,50%	33,70%
			MA-F	38,33%	22,81%	15,96%	14,41%	29,40%
14	Áudio, Cifras e Letras	<i>Late fusion - Max</i>	Alegria	09,40%	18,60%	27,80%	22,40%	17,70%
			Amor	31,10%	32,10%	45,10%	45,60%	26,10%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	07,40%	23,40%	26,10%	23,50%	31,70%
			Paixão	35,40%	38,90%	29,50%	23,40%	36,40%
			Tristeza	00,00%	20,40%	08,00%	06,20%	17,70%
			MP-F	20,30%	28,40%	29,40%	26,40%	27,00%
			MA-F	13,88%	22,23%	22,75%	20,18%	21,60%
15	Áudio, Cifras e Letras	<i>Late fusion - Prod</i>	Alegria	08,30%	31,30%	27,90%	36,90%	00,00%
			Amor	48,60%	50,50%	49,50%	59,30%	00,09%
			Decepção	00,00%	00,00%	00,00%	00,00%	03,30%
			Entusiasmo	22,00%	32,40%	30,10%	22,80%	00,00%
			Paixão	43,10%	38,10%	36,00%	29,70%	03,70%
			Tristeza	00,00%	23,10%	09,30%	06,30%	01,80%
			MP-F	29,40%	36,70%	33,10%	31,00%	01,60%
			MA-F	20,33%	29,23%	25,46%	25,83%	01,48%
16	Áudio, Cifras e Letras	<i>Late fusion - Sum</i>	Alegria	06,70%	31,30%	30,50%	34,60%	25,90%
			Amor	46,00%	48,80%	49,00%	48,90%	39,10%
			Decepção	00,00%	00,00%	00,00%	00,00%	00,00%
			Entusiasmo	21,80%	33,80%	30,20%	18,40%	41,70%
			Paixão	41,50%	36,10%	32,90%	30,50%	36,60%
			Tristeza	00,00%	19,80%	01,70%	00,00%	14,10%
			MP-F	28,00%	35,50%	31,40%	29,20%	32,80%
			MA-F	19,33%	28,30%	24,05%	22,06%	26,23%

Fonte: Autoria própria.

O Experimento # 1: apresenta os resultados da classificação multimodal utilizando o método de combinação *early fusion* com o áudio e as cifras. Os melhores resultados desse

experimento por emoção (em negrito na tabela) foram: alegria (5-NN=26,90%), amor (SVM=39,70%), entusiasmo (SVM=34,00%), decepção (J48=06,90%), entusiasmo (SVM=34,00%), paixão (1-NN=33,90%) e tristeza (J48=17,80%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=24,95%.

O Experimento # 2: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Max* com o áudio e as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=21,90%), amor (3-NN=39,70%), entusiasmo (J48=27,80%), paixão (5-NN=39,30%) e tristeza (1-NN=21,40%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os melhores resultados para essas representações foram obtidos pelo classificador 3-NN com MA-F=22,85%.

O Experimento # 3: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Prod* com o áudio e as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=18,70%), amor (SVM=44,70%), decepção (J48=03,40%), entusiasmo (1-NN, 3-NN=27,20%), paixão (SVM=42,00%) e tristeza (1-NN, 3-NN=15,90%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essas representações foram obtidos pelo classificador 3-NN com MA-F=23,90%.

O Experimento # 4: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Sum* com o áudio e as cifras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=16,90%), amor (SVM=44,40%), entusiasmo (1-NN=30,50%), paixão (5-NN=43,10%) e tristeza (J48=16,50%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=23,21%.

O Experimento # 5: apresenta os resultados da classificação multimodal utilizando o método de combinação *early fusion* com as letras e o áudio. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (SVM=48,40%), amor (SVM=48,20%), entusiasmo (SVM=57,50%), paixão (SVM=41,40%) e tristeza (SVM=29,60%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os

melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=37,51%.

O Experimento # 6: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Max* com as letras e o áudio. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=30,90%), amor (3-NN, 5-NN= 45,60%), entusiasmo (J48=34,10%), paixão (J48= 32,20%) e tristeza (1-NN=24,10%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=22,48%.

O Experimento # 7: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Prod* com as letras e o áudio. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=43,20%), amor (SVM=51,70%), decepção (J48=03,40%), entusiasmo (SVM= 49,10%), paixão (SVM= 42,80%) e tristeza (1-NN=23,90%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=30,85%.

O Experimento # 8: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Sum* com as letras e o áudio. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (5-NN=41,00%), amor (3-NN=51,60%), entusiasmo (SVM=48,30%), paixão (SVM= 42,90%) e tristeza (1-NN=25,40%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=33,66%.

O Experimento # 9: apresenta os resultados da classificação multimodal utilizando o método de combinação *early fusion* com as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (SVM=51,70%), amor (SVM= 51,50%), decepção (J48=07,70%), entusiasmo (SVM= 58,40%), paixão (SVM=45,90%) e tristeza (SVM= 31,00%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F=39,75%.

O Experimento # 10: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Max* com as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=27,80%), amor (3-NN=45,20%), entusiasmo (J48=29,60%), paixão (1-NN= 38,90%) e tristeza (1-NN=20,40%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=22,23%.

O Experimento # 11: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Prod* com as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (5-NN=32,30%), amor (SVM=49,10%), decepção (J48=03,40%), entusiasmo (1-NN= 31,30%), paixão (SVM= 22,50%) e tristeza (1-NN=22,50%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Os melhores resultados para essas representações foram para essas representações pelo classificador 1-NN com MA-F=27,86%.

O Experimento # 12: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Sum* com as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (1-NN=31,70%), amor (1-NN=48,40%), entusiasmo (J48=42,40%), paixão (SVM = 41,70%) e tristeza (J48=19,10%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=28,41%.

O Experimento # 13: apresenta os resultados da classificação multimodal utilizando o método de combinação *early fusion* com o áudio, as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (SVM=47,80%), amor (SVM=51,30%), decepção (J48=07,70%), entusiasmo (SVM= 59,20%), paixão (SVM= 45,00%) e tristeza (SVM= 26,70%). Ao analisar as emoções, individualmente, o amor e a paixão foram as mais fáceis de serem classificadas. Os melhores resultados para essas representações foram obtidos pelo classificador SVM com MA-F= 38,33%.

O Experimento # 14: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Max* com o áudio, as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (3-NN=27,80%), amor (5-NN= 45,60%), entusiasmo (J48=31,70%), paixão (1-NN= 38,90%) e tristeza (1-NN=

20,40%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os melhores resultados para essas representações foram obtidos pelo classificador 3-NN com MA-F=22,75%.

O Experimento # 15: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Prod* com o áudio, as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (5-NN=36,90%), amor (1-NN=50,50%), decepção (J48=03,30%), entusiasmo (1-NN= 32,40%), paixão (SVM= 43,10%) e tristeza (1-NN= 23,10%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=29,23%.

O Experimento # 16: apresenta os resultados da classificação multimodal utilizando o método de combinação *late fusion - Sum* com o áudio, as cifras e as letras. Os melhores resultados desse experimento por emoção (em negrito na tabela) foram: alegria (5-NN=34,60%), amor (3-NN= 49,00%), entusiasmo (J48=41,70%), paixão (SVM = 41,50%) e tristeza (1-NN= 19,80%). Ao analisar as emoções, individualmente, o amor foi a mais fácil de ser classificada. Porém, ela também mostra que todos os classificadores falharam ao reconhecer a emoção decepção. Os melhores resultados para essas representações foram obtidos pelo classificador 1-NN com MA-F=28,30%.

6.2.1 SUMÁRIO DOS RESULTADOS MULTIMODAIS COM SEIS EMOÇÕES

A tabela 27 apresenta o sumário com os resultados da média aritmética da *F-measure* obtidos a partir do método *early fusion e late fusion* para as combinações multimodais: do áudio com as cifras, das letras com o áudio, das cifras com as letras e das cifras com o áudio e com as letras e com os diferentes classificadores e utilizando seis emoções. A análise da tabela 27 mostra alguns resultados interessantes. Em primeiro lugar, verificou-se um certo equilíbrio entre os melhores resultados da média aritmética da *F-measure* entre todas as combinações multimodais. A combinação das cifras com as letras obteve o melhor resultado *early fusion SVM=39,75%*, seguido pela combinação das cifras com o áudio e com as letras *SVM=38,33%*. A combinação das letras como áudio obteve o melhor resultado *late fusion 1-NN-SUM=33,66%*, seguida pelo *SVM-PROD=30,85%* e *SVM-SUM=30,78%*. Em segundo lugar, a combinação das cifras com as letras e das letras com o áudio, obtiveram resultados próximos aos das combinações das cifras com o áudio e com as letras, tanto no *early fusion* como no *late fusion*. Os resultados das combinações cifras e letras e cifras, áudio e letras apresentaram os melhores resultados utilizando o

classificador SVM, se comparados com o método unimodal e sem combinação. A combinação das letras e do áudio utilizando o classificador SVM, além dos dois já citados (cifras e letras; áudio, cifras e letras), perde apenas para a combinação unimodal das letras, 3-Grams=37,76%. Em terceiro lugar, analisando por meio de um contexto geral as combinações multimodais através do classificador SVM melhoraram os resultados unimodais e sem combinações, sendo um bom classificador para as seis emoções e para a base MLMMD.

Tabela 27: Sumário dos resultados multimodais com seis emoções.

Fontes de Informação	Combinações	SVM	1-NN	3-NN	5-NN	J48
Áudio + Cifras	<i>Early Fusion</i>	24,95%	20,88%	17,75%	21,55%	21,98%
	<i>Late Fusion (Max)</i>	13,88%	22,48%	22,85%	21,81%	22,43%
	<i>Late Fusion (Prod)</i>	17,05%	23,25%	23,90%	21,10%	08,91%
	<i>Late Fusion (Sum)</i>	16,63%	23,21%	21,70%	20,65%	21,15%
Letras + Áudio	<i>Early Fusion</i>	37,51%	22,31%	16,33%	11,08%	28,30%
	<i>Late Fusion (Max)</i>	11,71%	22,48%	22,40%	20,01%	21,80%
	<i>Late Fusion (Prod)</i>	30,85%	27,83%	26,88%	27,96%	04,10%
	<i>Late Fusion (Sum)</i>	30,78%	33,66%	24,71%	23,65%	26,63%
Cifras + Letras	<i>Early Fusion</i>	39,75%	23,26%	15,28%	16,03%	29,45%
	<i>Late Fusion (Max)</i>	13,88%	22,23%	21,93%	18,23%	21,50%
	<i>Late Fusion (Prod)</i>	21,16%	27,86%	23,28%	21,76%	02,61%
	<i>Late Fusion (Sum)</i>	28,41%	28,41%	21,60%	20,23%	25,81%
Cifras + Áudio + Letras	<i>Early Fusion</i>	38,33%	22,81%	15,96%	14,41%	29,40%
	<i>Late Fusion (Max)</i>	13,88%	22,23%	22,75%	20,18%	21,60%
	<i>Late Fusion (Prod)</i>	20,33%	29,23%	25,46%	25,83%	01,48%
	<i>Late Fusion (Sum)</i>	19,33%	28,30%	24,05%	22,06%	26,23%
<i>Best No Fusion SVM 3-Grams</i>		37,76%				
<i>Best Unimodal EF SVM Letras</i>		38,01%				

Fonte: Autoria própria.

6.3 COMPARAÇÕES EXPERIMENTAIS

Nesse capítulo foram apresentados os resultados experimentais envolvendo quatro combinações multimodais: áudio e cifras; letras e áudio; cifras e letras; cifras, áudio e letras, para detecção automática de emoções em músicas. Não foram encontrados trabalhos com representações multimodais envolvendo as cifras em outros trabalhos de MER.

Utilizando a combinação do áudio com as letras, em (LAURIER et al., 2008), os autores utilizaram um modelo de emoções básicas, para uma base contendo 1.000 músicas da *Last Fm*, com quatro rótulos: irritado, feliz, triste, relaxado. Os resultados mostraram as seguintes médias de acertos por emoções, irritado-SVM=98,30%, feliz-SVM=86,80%, triste-SVM=92,80%, relaxado-SVM=91,70%. A combinação do trabalho de (LAURIER et al., 2008)

ocorreu a partir de duas abordagens: a primeira utilizando previsões separadas, por intermédio do voto da maioria; a segunda combinando todas as características em um mesmo arquivo, com um vetor composto pelas características do áudio e das letras. (LAURIER et al., 2008) citam ainda que a segunda abordagem apresentou resultados melhores.

Ainda utilizando a combinação do áudio com as letras, em (HU et al., 2009), os autores utilizaram uma base de 5.558 músicas, retiradas da *Last Fm*, com 135 rótulos em 18 categorias de emoções. Os resultados mostraram uma média aproximada de 61,00% de acerto nas categorias. A combinação do trabalho de (HU et al., 2009) ocorreu combinando todas as características em um mesmo arquivo, no entanto não foi explicitado a regra de combinação.

Ainda utilizando a combinação do áudio com as letras, em (MIHALCEA; STRAPPA-RAVA, 2012), os autores utilizaram um modelo de emoções básicas com seis rótulos: raiva, repulsa, medo, alegria, tristeza, surpreso, para uma base contendo 100 músicas. Os resultados mostraram uma média de 54,39% de acertos.

Não foram encontradas as outras combinações utilizadas nesse trabalho em outras pesquisas.

Dos trabalhos multimodais analisados, os trabalhos de (LAURIER et al., 2008) e (HU et al., 2009) corroboraram com os experimentos multimodais realizados nesse capítulo. No entanto, houve muita diferença nos resultados alcançados em (LAURIER et al., 2008). Já (HU et al., 2009) obtiveram resultados semelhantes e também utilizaram a combinação do áudio com as letras, mantendo todas as características juntas.

7 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Nesse trabalho foi apresentado um estudo sobre a classificação automática de emoções em músicas latinas utilizando a *Multimodal Latin Music Mood Database*. Para realizar os experimentos foram utilizadas três fontes de informações: cifras, áudio e letras de forma individual e com combinações unimodal e multimodal com os métodos de combinação de classificadores de *early fusion* e *late fusion*.

Foram realizados diversos experimentos para verificar a hipótese desse trabalho: que existem diferentes fontes de informações para a classificação automática de emoções em músicas, bem como diferentes maneiras de combiná-las. No entanto, qual fonte de informação ou combinação se sobressai sobre as demais em MER e em quais situações? Para responder essa pergunta, foram realizadas uma série de experimentos computacionais, utilizando duas versões da *Multimodal Latin Music Mood Database*, uma com três emoções (positiva, neutra e negativa) e outra com seis emoções (alegria, amor, decepção, entusiasmo, paixão e tristeza).

De modo geral, por meio dos experimentos com três emoções constatou-se um certo equilíbrio nos resultados das três representações: sem fusão, unimodal e multimodal. Porém, a representação multimodal com a combinação das características extraídas das letras, áudio e cifras obteve o melhor resultado geral com o classificador SVM e o método de combinação *early fusion* (média=60,60%), resultado bem semelhante ao trabalho de (HU et al., 2009), utilizando a combinação do áudio com as cifras SVM=61,00%. Considerando a combinação unimodal, os melhores resultados também foram obtidos utilizando o classificador SVM com o método de combinação *early fusion* e as características extraídas das letras das músicas (média=58,93%). Para os resultados sem combinação, o melhor resultado obtido foi utilizado o classificador SVM com a representação 4-grams extraída das letras das músicas (média=58,76%).

De modo geral, da mesma forma que ocorreu nos experimentos com três emoções, os experimentos com seis emoções mantiveram um certo equilíbrio nos resultados em todas as representações: sem fusão, unimodal e multimodal. Com a combinação das características extraídas das letras, áudio e cifras obteve o melhor resultado geral com o classificador SVM e o método de combinação *early fusion* (média=39,75%). Considerando a combinação unimodal, os melhores resultados também foram obtidos utilizando o classificador SVM com o método de combinação *early fusion* e as características extraídas das letras das músicas (média=38,01%). Para os resultados sem combinação, o melhor resultado obtido foi utilizado o classificador SVM com a representação 3-grams extraída das letras das músicas (média=37,76%).

Finalmente, respondendo à hipótese, conclui-se que os melhores resultados foram obtidos por meio do método combinatório multimodal *early fusion*, utilizando o classificador SVM, tanto para três, quanto para seis emoções. Como futuras direções de pesquisa pretende-se verificar o impacto do desbalanceamento das classes, aspecto que pode estar prejudicando o desempenho dos classificadores na MLMMD. Outra direção futura seria utilizar a abordagem multimodal apresentada nesse trabalho em outras bases de dados.

REFERÊNCIAS

- BERTIN-MAHIEUX, T.; ELLIS, D. P.; WHITMAN, B.; LAMERE, P. The million song dataset. In: **Proceedings of the 12th International Society for Music Information Retrieval (ISMIR) Conference**. Miami, Florida: University of Miami, 2011. p. 591–596.
- BISCHOFF, K.; FIRAN, C. S.; PAIU, R.; NEJDL, W.; LAURIER, C.; SORDO, M. Music mood and theme classification—a hybrid approach. In: **Proceedings of the 10th International Society for Music Information Retrieval (ISMIR) Conference**. Kobe, Japan: McGill University, 2009. p. 657–662.
- BONTE, T. **MuseScore: Open source music notation and composition software**. 2015. Disponível em: <<http://pt.slideshare.net/thomasbonte/musescore-at-fosdem-2009>>. Acesso em: 25 jan. 2015.
- BRINKER, B. den; DINTHER, R. van; SKOWRONEK, J. Expressed music mood classification compared with valence and arousal ratings. **EURASIP Journal on Audio, Speech and Music Processing, Springer**, v. 2012, n. 1, p. 1–14, 2012.
- CAVNAR, W. B.; TRENKLE, J. M. N-gram-based text categorization. **Ann Arbor MI**, v. 48113, n. 2, p. 161–175, 1994.
- CHEDIAK, A. **Dicionário de acordes cifrados**. São Paulo: Irmãos Vitale, 1984.
- CORASSA, R.; SANTOS, C. L. D.; SILLA JR., C. N. Classificação automática de gêneros musicais utilizando cifras das músicas. In: **Proceedings of the Brazilian Symposium on Computer Music**. Brazil, São Paulo: , 2013. v. 1, n. 1, p. 178–188.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.
- DODDS, P. S.; DANFORTH, C. M. Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. **Journal of Happiness Studies**, v. 11, n. 4, p. 441–456, 2010.
- DUNKER, P.; NOWAK, S.; BEGAU, A.; LANZ, C. Content-based mood classification for photos and music: a generic multi-modal classification framework and evaluation approach. In: **Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval**. New York, NY, USA: ACM, 2008. v. 1, p. 97–104.
- EEROLA, T.; LARTILLOT, O.; TOIVIAINEN, P. Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In: **Proceedings of the 10th International Society for Music Information Retrieval (ISMIR) Conference**. Kobe, Japan: McGill University, 2009. p. 621–626.
- EL-KHAIR, I. A. Effects of stop words elimination for arabic information retrieval: a comparative study. **International Journal of Computing & Information Sciences**, v. 4, n. 3, p. 119–133, 2006.
- FREUND, Y.; IYER, R.; SCHAPIRE, R. E.; SINGER, Y. An efficient boosting algorithm for combining preferences. **The Journal of machine learning research**, v. 4, n. 1, p. 933–969, 2003.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: an update. **ACM SIGKDD explorations newsletter**, v. 11, n. 1, p. 10–18, 2009.

HENNING, V.; REICHEL, J. Mendeley-a last. fm for research? In: **Proceedings of the IEEE Fourth International Conference on eScience**. Indianapolis, Indiana, USA: Institute of Electrical and Electronics Engineers (IEEE), 2008. p. 327–328.

HEVNER, K. Experimental studies of the elements of expression in music. **The American Journal of Psychology**, v. 48, n. 2, p. 246–268, 1936.

HU, X.; DOWNIE, J. S. Improving mood classification in music digital libraries by combining lyrics and audio. In: **Proceedings of the 10th annual joint conference on Digital libraries**. Gold Coast, Queensland, Australia: ACM, 2010. p. 159–168.

HU, X.; DOWNIE, J. S. When lyrics outperform audio for music mood classification: A feature analysis. In: **Proceedings of the 11th International Society for Music Information Retrieval (ISMIR) Conference**. Utrecht, Netherlands: , 2010. p. 619–624.

HU, X.; DOWNIE, J. S.; LAURIER, C.; BAY, M.; EHMANN, A. F. The 2007 mirex audio mood classification task: Lessons learned. In: **Proceedings of the 9th International Society for Music Information Retrieval (ISMIR) Conference**. Philadelphia, USA: Drexel University, 2008. p. 462–467.

HU, Y.; CHEN, X.; YANG, D. Lyric-based song emotion detection with affective lexicon and fuzzy clustering method. In: **Proceedings of the 10th International Society for Music Information Retrieval (ISMIR) Conference**. Kobe, Japan: McGill University, 2009. p. 123–128.

KATAYOSE, H.; IMAI, M.; INOKUCHI, S. Sentiment extraction in music. In: **Proceedings of the 9th IEEE International Conference on Pattern Recognition**. Rome: Institute of Electrical and Electronics Engineers (IEEE), 1988. p. 1083–1087.

KELLER, J. M.; GRAY, M. R.; GIVENS, J. A. A fuzzy k-nearest neighbor algorithm. **Systems, Man and Cybernetics, IEEE Transactions on**, IEEE, n. 4, p. 580–585, 1985.

KIM, Y. E.; SCHMIDT, E. M.; MIGNECO, R.; MORTON, B. G.; RICHARDSON, P.; SCOTT, J.; SPECK, J. A.; TURNBULL, D. Music emotion recognition: A state of the art review. In: **Proceedings of the 11th International Society for Music Information Retrieval (ISMIR) Conference**. Utrecht, Netherlands: , 2010. p. 255–266.

KITTLER, J.; HATEF, M.; DUIN, R. P.; MATAS, J. On combining classifiers. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 20, n. 3, p. 226–239, 1998.

KUMAR, V.; MINZ, S. Mood classification of lyrics using sentiwordnet. In: **Proceedings of the IEEE International Conference on Computer Communication and Informatics**. 2013. p. 1–5.

LANCKRIET, G. R.; CRISTIANINI, N.; BARTLETT, P.; GHAOUI, L. E.; JORDAN, M. I. Learning the kernel matrix with semidefinite programming. **The Journal of Machine Learning Research**, v. 5, n. 1, p. 27–72, 2004.

- LAURIER, C.; GRIVOLLA, J.; HERRERA, P. Multimodal music mood classification using audio and lyrics. In: **17th International Conference on Machine Learning and Applications**. San Diego, USA: IEEE, 2008. p. 688–693.
- LAURIER, C.; HERRERA, P.; MANDEL, M.; ELLIS, D. Audio music mood classification using support vector machine. **Music Information Retrieval Evaluation Exchange, MIREX task on Audio Mood Classification**, p. 2–4, 2007.
- LIDY, T.; RAUBER, A. Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In: **Proceedings of the 6th International Society for Music Information Retrieval (ISMIR) Conference**. London, UK: University of London, 2005. p. 34–41.
- LIMA, A. A. de; NUNES, R. M.; RIBEIRO, R. P.; SILLA JR., C. N. Nordic music genre classification using song lyrics. In: **19th International Conference on Application of Natural Language to Information Systems**. Montpellier, France: , 2014. v. 8455, p. 89–100.
- LU, L.; LIU, D.; ZHANG, H.-J. Automatic mood detection and tracking of music audio signals. **IEEE Transactions on Audio, Speech, and Language Processing**, IEEE, v. 14, n. 1, p. 5–18, 2006.
- MICHELS, U. Atlas de música. Alianza editorial, Madrid, Espana, 1985.
- MIHALCEA, R.; STRAPPARAVA, C. Lyrics, music, and emotions. In: **Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning**. Jeju Island, Korea: Association for Computational Linguistics, 2012. p. 590–599.
- OLIVEIRA, A. P.; CARDOSO, A. A musical system for emotional expression. **Knowledge-Based Systems**, Elsevier, v. 23, n. 8, p. 901–913, 2010.
- PANDA, R.; MALHEIRO, R.; ROCHA, B.; OLIVEIRA, A.; PAIVA, R. Multi-modal music emotion recognition: A new dataset, methodology and comparative analysis. In: **Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research**. Marseille, France: University of Coimbra, 2013.
- PRIOLLI, M. L. d. M. Princípios básicos da música para a juventude. volumes i e ii. Casa Oliveira de Música Ltda, Rio de Janeiro, 1989.
- PRZYBYSZ, A. L.; CORASSA, R.; SANTOS, C. L.; SILLA JR., C. N. Latin music mood classification using cifras. In: **Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC)**. 2015. p. 1682–1686.
- QUINLAN, J. R. Improved use of continuous attributes in c4. 5. **Journal of artificial intelligence research**, p. 77–90, 1996.
- RUSSELL, J. A. A circumplex model of affect. **Journal of personality and social psychology**, American Psychological Association, v. 39, n. 6, p. 1161, 1980.
- SANTOS, C. L. D.; SILLA JR., C. N. The latin music mood database. **EURASIP Journal on Audio, Speech, and Music Processing**, Springer, v. 2015, n. 1, p. 1–11, 2015.
- SEBASTIANI, F. Machine learning in automated text categorization. **ACM computing surveys**, ACM, v. 34, n. 1, p. 1–47, 2002.

- SILLA JR., C. N.; KOERICH, A. L.; KAESTNER, C. A. The latin music database. In: **Proceedings of the 9th International Society for Music Information Retrieval (ISMIR) Conference**. 2008. p. 451–456.
- SNOEK, C. G.; WORRING, M.; SMEULDERS, A. W. Early versus late fusion in semantic video analysis. In: **Proceedings of the 13th annual ACM international conference on Multimedia**. 2005. p. 399–402.
- SONG, Y.; DIXON, S.; PEARCE, M. Evaluation of musical features for emotion classification. In: **Proceedings of the 13th International Society for Music Information Retrieval (ISMIR) Conference**. 2012. p. 523–528.
- STURM, B. L. Evaluating music emotion recognition: Lessons from music genre recognition? In: **Proceedings of the IEEE International Conference on MULTimedia and Expo Workshops**. 2013. p. 1–6.
- STURM, B. L. The state of the art ten years after a state of the art: Future research in music information retrieval. **Journal of New Music Research**, Taylor & Francis, v. 43, n. 2, p. 147–172, 2014.
- THAYER, R. E. **The biopsychology of mood and arousal**. : Oxford University Press, 1989.
- TURNBULL, D. R.; BARRINGTON, L.; LANCKRIET, G.; YAZDANI, M. Combining audio content and social context for semantic music discovery. In: **Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval**. 2009. p. 387–394.
- TYPKE, R.; WIERING, F.; VELTKAMP, R. C. A survey of music information retrieval systems. In: **Proceedings of the 6th International Society for Music Information Retrieval (ISMIR) Conference**. 2005. p. 153–160.
- WATSON, D.; CLARK, L. A. The panas-x: Manual for the positive and negative affect schedule-expanded form. 1999.
- WEINBERGER, K. Q.; BLITZER, J.; SAUL, L. K. Distance metric learning for large margin nearest neighbor classification. In: **Advances in neural information processing systems**. 2005. p. 1473–1480.
- WITTEN, I. H.; FRANK, E. **Data Mining: Practical machine learning tools and techniques**. San Francisco: Morgan Kaufmann, 2005.
- YANG, D.; LEE, W.-S. Music emotion identification from lyrics. In: **Proceedings of the IEEE International Symposium on Multimedia**. 2009. p. 624–629.
- YANG, Y.-H.; CHEN, H. H. Machine recognition of music emotion: A review. **ACM Transactions on Intelligent Systems and Technology (TIST)**, ACM, v. 3, n. 3, p. 40, 2012.
- YANG, Y.-H.; HU, X. Cross-cultural music mood classification: A comparison on english and chinese songs. In: **Proceedings of the 13th International Society for Music Information Retrieval (ISMIR) Conference**. 2012. p. 19–24.
- ZAAANEN, M. V.; KANTERS, P. Automatic mood classification using tf* idf based on lyrics. In: **Proceedings of the 11th International Society for Music Information Retrieval (ISMIR) Conference**. 2010. p. 75–80.

ZADROZNY, B.; ELKAN, C. Transforming classifier scores into accurate multiclass probability estimates. In: **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**. 2002. p. 694–699.

ANEXO A - LISTAS DE *STOPWORDS*

A seguir são apresentadas as listas pré-definidas das palavras com pouco significado semântico no idioma português, inglês e espanhol. As listas utilizadas nesse trabalho são de domínio público, elaboradas por desenvolvedores web, para redirecionar conteúdos web nos motores de pesquisas. As listas foram retiradas de ranks.nl, disponíveis em (www.ranks.nl/stopwords).

Português: a, agora, ainda, alguém, algum, alguma, algumas, alguns, ampla, amplas, amplo, amplos, ante, antes, ao, aos, após, aquela, aquelas, aquele, aqueles, aquilo, as, até, através, cada, coisa, coisas, com, como, contra, contudo, da, daquele, daqueles, das, de, dela, delas, dele, deles, depois, dessa, dessas, desse, desses, desta, destas, deste, deste, destes, deve, devem, devendo, dever, deverá, deverão, deveria, deveriam, devia, deviam, disse, disso, disto, dito, diz, dizem, do, dos, e, é, ela, elas, ele, eles, em, enquanto, entre, era, essa, essas, esse, esses, esta, está, estamos, estão, estas, estava, estavam, estávamos, este, estes, estou, eu, fazendo, fazer, feita, feitas, feito, feitos, foi, for, foram, fosse, fossem, grande, grandes, há, isso, isto, já, la, lá, lhe, lhes, lo, mas, me, mesma, mesmas, mesmo, mesmos, meu, meus, minha, minhas, muita, muitas, muito, muitos, na, não, nas, nem, nenhum, nessa, nessas, nesta, nestas, ninguém, no, nos, nós, nossa, nossas, nosso, nossos, num, numa, nunca, o, os, ou, outra, outras, outro, outros, para, pela, pelas, pelo, pelos, pequena, pequenas, pequeno, pequenos, per, perante, pode, pude, podendo, poder, poderia, poderiam, podia, podiam, pois, por, porém, porque, posso, pouca, poucas, pouco, poucos, primeiro, primeiros, própria, próprias, próprio, próprios, quais, qual, quando, quanto, quantos, que, quem, são, se, seja, sejam, sem, sempre, sendo, será, serão, seu, seus, si, sido, só, sob, sobre, sua, suas, talvez, também, tampouco, te, tem, tendo, tenha, ter, teu, teus, ti, tido, tinha, tinham, toda, todas, todavia, todo, todos, tu, tua, tuas, tudo, última, últimas, último, últimos, um, uma, umas, uns, vendo, ver, vez, vindo, vir, vos, vós.

Inglês: a, able, about, across, after, all, almost, also, am, among, an, and, any, are, as, at, be, because, been, but, by, can, cannot, could, dear, did, do, does, either, else, ever, every, for, from, get, got, had, has, have, he, her, hers, him, his, how, however, i, if, in, into, is, it, its, just, least, let, like, likely, may, me, might, most, must, my, neither, no, nor, not, of, off, often, on, only, or, other, our, own, rather, said, say, says, she, should, since, so, some, than, that, the, their, them, then, there, these, they, this, tis, to, too, twas, us, wants, was, we, were, what, when, where, which, while, who, whom, why, will, with, would, yet, you, your.

Espanhol: a, acá, adelante, además, ahí, ahora, al, alguna, algunas, alguno, algunos, as, así, aún, bastante, bien, cada, como, con, cosa, cosas, cual, cuál, cuales, cuáles, cuando, da, de, debe, debemos, del, dentro, desde, después, donde, dos, durante, e, el, ellos, en, entonces, entre, era, eran, es, esa, esas, ese, eso, esos, esta, está, estaba, estaban, estamos, están, estar,

estas, este, esto, estos, frente, fue, fuera, fueron, ha, haber, había, hace, hacen, hacer, hacia, haciendo, han, hasta, hay, hecho, hoy, la, las, le, les, lo, los, me, mientras, misma, mismas, mismo, mismos, mucha, muchas, mucho, muchos, muy, ni, no, nos, nuestra, nuestras, nuestro, nuestros, o, ó, os, otra, otras, otro, otros, para, pero, podemos, podría, por, porque, puede, pueden, pues, que, qué, se, sea, según, segundo, ser, sería, si, sido, siendo, sobre, solamente, son, su, tal, también, tan, tanto, tenemos, tener, tenían, tiene, tienen, toda, todas, todavía, todo, todos, través, un, una, uno, va, vamos, van, veces, vemos, ver, vez, y, ya.