

BRUNO KOSTIUK

**CLASSIFICAÇÃO MULTI-RÓTULO DE
EMOÇÕES EM VÍDEOS MUSICAIS
UTILIZANDO CONJUNTOS DE
CARACTERÍSTICAS DE ÁUDIO E VÍDEO**

Curitiba - PR, Brasil

2019

BRUNO KOSTIUK

**CLASSIFICAÇÃO MULTI-RÓTULO DE EMOÇÕES EM
VÍDEOS MUSICAIS UTILIZANDO CONJUNTOS DE
CARACTERÍSTICAS DE ÁUDIO E VÍDEO**

apresentada ao Programa de Pós-Graduação
em Informática da Pontifícia Universidade Ca-
tólica do Paraná como requisito parcial para
obtenção do título de mestre em Informática.

Pontifícia Universidade Católica do Paraná - PUCPR

Programa de Pós-Graduação em Informática - PPGIa

Orientador: CARLOS NASCIMENTO SILLA JUNIOR

Coorientadora: XIAO HU

Colaborador: ALCEU DE SOUZA BRITTO JUNIOR

Curitiba - PR, Brasil

2019

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Edilene de Oliveira dos Santos CRB 9 /1636

K
2019 Kostiuk, Bruno
Classificação multi-rótulo de emoções em vídeos musicais utilizando conjuntos de características de áudio e vídeo / Bruno Kostiuk ; orientador, Carlos Nascimento Silla Junior ; colaboradora, Xiao Hu ; coorientador, Alceu de Souza Britto Junior.
– 2019
131 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2019
Bibliografia: f.93-95

1. Informática. 2. Vídeos musicais. 3. Emoções na música. 4. Classificação. 5. Rótulos. 6. C++ (Linguagem de programação de computador). I. Silla Junior, Carlos Nascimento. II. Hu, Xiao. II. Britto Junior. Alceu de Souza. IV. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática. IV. Título.

CDD 20. ed. – 004

DECLARAÇÃO


Declaro para os devidos fins que o aluno **BRUNO KOSTIUK**, defendeu sua dissertação de Mestrado intitulada “**CLASSIFICAÇÃO MULTI-RÓTULO DE EMOÇÕES EM VÍDEOS MUSICAIS UTILIZANDO CONJUNTOS DE CARACTERÍSTICAS DE ÁUDIO E VÍDEO**”, na área de concentração Ciência da Computação, no dia 13 de junho de 2019, no qual foi aprovado.

Declaro ainda que foram feitas todas as alterações solicitadas pela Banca Examinadora, cumprindo todas as normas de formatação definidas pelo Programa.

Por ser verdade, firmo a presente declaração.

Curitiba, 12 de setembro de 2019.





Prof. Dr. Emerson Cabrera Paraiso
Coordenador do Programa de Pós-Graduação em Informática
Pontifícia Universidade Católica do Paraná

Agradecimentos

Agradeço primeiramente a minha noiva Camila Camargo Juliani que me apoiou em todos os momentos desta etapa, me apoiando tanto emocionalmente quanto me auxiliando na correção de documentos de apresentações.

Agradeço a meu orientador Dr. Carlos Nascimento Silla Junior por me auxiliar e orientar durante todo andamento do trabalho. Agradeço também a minha Coorientadora Dr. Xiao Hu e ao nosso colaborador Dr. Alceu de Souza Britto Junior por suas contribuições que certamente agregaram valor a este trabalho. Agradeço também aos membros da banca, Dr. Jean Paul Barddal e Dr. Danilo Sipoli Sanches por aceitarem fazer parte da banca mesmo com o tempo curto para a mesma.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Tanto o vídeo quanto a música são utilizados para transmitir emoções. No entanto, apesar de sua importância em diversas aplicações, existem poucos trabalhos que tratam da questão de classificação de emoções em vídeos musicais, os mais recentes trabalhos utilizam trechos de vídeos provindos de filmes ou apenas o áudio de vídeos musicais. A principal razão é possivelmente a falta de bancos de dados disponíveis com rotulação para o vídeo musical completo. Neste trabalho é proposto um método para classificação multi-rótulo de emoções em vídeos musicais completos utilizando características extraídas do áudio (extraídas através da biblioteca Yaafe) e do vídeo (extraídas através das bibliotecas OpenCV e Keras). Para avaliar o método proposto são realizados experimentos no intuito de avaliar a performance das modalidades individuais (apenas áudio, apenas vídeo) e combinadas (utilizando o método de fusão tardia). Para realizar os experimentos foi necessário estender a base de dados CAL500 para incluir os vídeos musicais oficiais, visto que originalmente a CAL500 possuía apenas o áudio e os rótulos de emoções de cada música. A justificativa para realizar a extensão da CAL500 utilizando apenas os vídeos musicais oficiais é que estes foram desenvolvidos para transmitir a mesma emoção da música. Nos experimentos realizados os melhores resultados (Micro-averaged AUC de 74,40%) foram obtidos através da fusão tardia utilizando as predições obtidas através do áudio em conjunto com as predições obtidas através do Fluxo Óptico. Utilizando apenas o áudio o melhor resultado obtido foi de 72,44% e utilizando apenas as imagens vídeo musical o melhor resultado obtido foi de 71,76%.

Palavras-chave: Emoção em Vídeos Musicais, Classificação, Emoção em Vídeo, Emoção em Áudio, Fusão de Bases, Classificação Multi-Rótulo, Combinação de Áudio e Vídeo.

Abstract

Video as well as music are used to convey emotions. However, in spite of its importance in several applications, there are few works that deal with the issue of emotion classification in music videos, the latest work uses stretches of videos from movies or just the audio of music videos. The main reason is possibly the lack of available databases with labeling for the full music video. This work proposes a method for multi-label classification of emotions in complete music videos using features extracted from the audio (extracted through the Yaafe library) and video (extracted through the OpenCV and Keras library). To evaluate the proposed method, experiments will be carried out to evaluate the performance of the individual modalities (audio only, video only) and combined (using the late fusion method). To perform the experiments it was necessary to extend the CAL500 database to include the official music videos, since originally the CAL500 has only the audio and emotion labels of each song. The justification for the extension of the CAL500 using only official music videos is that they were developed to convey the same emotion as the music. In the experiments performed the best results (Micro-averaged AUC 74.40%) were obtained through the late fusion using the predictions obtained through the audio in conjunction with the predictions obtained through the Optical Flow. Using only audio the best result was 72.44% and using only the music video images the best result was 71.76%.

Keywords: Emotion in Music Videos, Classification, Video Emotion, Emotion in Audio, Merger of Bases, Multi-Label Classification, Audio and Video Combination.

Lista de ilustrações

Figura 1 – Sinal modulado de exemplo.	25
Figura 2 – Transformada de Hilbert usando a figura 1.	26
Figura 3 – Exemplificação do Tremolo utilizando o envelope exemplificado na figura 2.	26
Figura 4 – Exemplificação do Grain utilizando o envelope exemplificado na figura 2.	27
Figura 5 – Histograma RGB de um frame escolhido aleatoriamente na base de dados	32
Figura 6 – Exemplo de processo de cálculo da diferença acumulada entre cinco frames de vídeo	33
Figura 7 – Exemplo de saída do processo de Fluxo Óptico	34
Figura 8 – Arquitetura do VGG16	35
Figura 9 – Operação do LBP. O Pixel C que se encontra no meio e seus P vizinhos a uma distância R	35
Figura 10 – Exemplo do algoritmo HOMER em uma base com 8 rótulos.	39
Figura 11 – Base de Exemplo 1, arquivo ARFF multi-rótulo com características de Fluxo Óptico.	42
Figura 12 – Base de Exemplo 2, arquivo ARFF multi-rótulo com características de Áudio.	42
Figura 13 – Fusão precoce entre as bases de exemplo 1(figura 11) e 2(figura 12). . .	43
Figura 14 – Exemplo de arquivo de predições.	44
Figura 15 – Modelo exemplificando a criação da base CAL500 Estendido.	52
Figura 16 – Modelo exemplificando o processo de extração de características auditivas.	55
Figura 17 – Modelo exemplificando a extração de características de Fluxo Óptico. .	58
Figura 18 – Modelo exemplificando o processo de extração de características visuais.	59
Figura 19 – Modelo exemplificando o processo de geração de predições auditivas e visuais e avaliação métrica das mesmas.	62
Figura 20 – Gráfico ROC de exemplo.	64
Figura 21 – Treinamento do modelo visual de classificação.	69
Figura 22 – Teste e geração de predições auditivas e visuais.	70
Figura 23 – Geração da predição visual médias.	72
Figura 24 – Replicação das predições auditivas	75

Lista de tabelas

Tabela 1 – Exemplo de base de dados Multi-Rótulo	37
Tabela 2 – Exemplo de resultados da aplicação do BR sobre a base de dados da tabela 1	37
Tabela 3 – Aplicação da transformação LP no exemplo de base da tabela 1	39
Tabela 4 – Exemplo de resultados da aplicação do RakEL sobre a base de dados da tabela 1.	41
Tabela 5 – Comparativo entre classificadores utilizados.	50
Tabela 6 – Comparativo entre emoções e fontes dos dados utilizados.	50
Tabela 7 – Separação real dos vídeos (representados pelos seus Ids) por fold de forma estratificada.	53
Tabela 8 – Distribuição dos rótulos por fold.	54
Tabela 9 – Informações sobre arquivos gerados.	61
Tabela 10 – Quantidade de combinações realizadas no processo de fusão tardia. . .	65
Tabela 11 – Quantidade de características extraídas da base CAL500 Estendida. . .	68
Tabela 12 – Resultados obtidos através da métrica Micro-averaged AUC para cada classificador vs cada método de extração de características visuais utilizando uma média entre as predições das 220 imagens do vídeo musical.	72
Tabela 13 – Resultados obtidos através da métrica Micro-averaged AUC para cada classificador vs cada método de extração de características visuais utilizando as predições das 220 imagens do vídeo musical.	73
Tabela 14 – Resultados obtidos através da métrica Micro-averaged AUC para cada classificador comparando as extrações de características auditivas e visuais.	75
Tabela 15 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e Histograma de Cores.	78
Tabela 16 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e Diferença Acumulada.	79
Tabela 17 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e Fluxo Óptico.	80
Tabela 18 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e VGG16.fc7+ImageNet.	81
Tabela 19 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e LBP.	82
Tabela 20 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de predições do Histograma de Cores.	83

Tabela 21 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de predições da Diferença Acumulada.	84
Tabela 22 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de predições do Fluxo Óptico.	85
Tabela 23 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de predições do VGG16.fc7+ImageNet.	86
Tabela 24 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de predições do LBP.	87
Tabela 26 – Informações sobre os vídeos musicais.	125
Tabela 27 – Informações sobre os rótulos de emoções.	129
Tabela 28 – Rotulação da base.	129

Lista de abreviaturas e siglas

abs	Valor absoluto
ac	Auto-Correlação ou Correlação Automática
ARFF	Attribute-Relation File Format
AUC	Área abaixo da Curva
BR	Binary Relevance
BR k NN	Binary Relevance k -Nearest-Neighbor
BN	Baeyesian Network
BPMLL	Back-Programation Multi-Label Learning
C++	Linguagem de programação C++
C4.5	Classificador de único-rótulo
CAL500	Base de dados utilizada no projeto
CK+	Base de dados CK+
CLR	Calibrated Label Ranking
DEAP	Base de dados DEAP
dct	Transformação Discreta de Cosseno
EEG	Eletroencefalograma
en	Energia
exp	Exponencial
fc7	Fully Connected 7
fft	Transformada rápida de Furrier
FP	False Positive ou Falso Positivo
HOMER	Hierarchy Of Multi-label classifiERs
J48	Versão java do C4.5

k NN	k -Nearest-Neighbor
LBP	Local Binary Patterns
log	Função logarítmica
LP	Label Powerset
MBD	Método Baysiano Dinâmico
MFCC	Coefficiente de Cepstrum de Frequência Mel
ML k NN	Multi-Label k -Nearest-Neighbor
MTV	Music TeleVision
OBSI	Intensidade de Sinal da Banda de Oitava
OBSIR	Taxa de Intensidade de Sinal da Banda de Oitava
ORL	Base de dados ORL
Ra k EL	Random k labELsets
RGB	Red, Green and Blue
SOM	Self-Organization Map
stepSize	Janela de Tempo
STFT	Transformada de Furrier de curto prazo
SVM	Suport Vector Machine
TP	True Positive ou Positivo Verdadeiro
VGG16	Rede convolucional muito profunda para classificação de imagens em larga escala
ZCR	Zero-Crossing Rate

Lista de símbolos

σ	Letra grega Sigma, utilizada para representar o desvio padrão.
λ	Lambda, utilizada para representar os rótulos de uma base.
μ	Mu, utilizado para representar um meta-rótulo.
Δ	Delta, utilizado para representar a função base das estatísticas de forma espectral.
\subseteq	É subconjunto próprio de.
Σ	Somatório, utilizado para realizar somas sequenciais.
\in	Pertence, utilizado para operações matemáticas.
\notin	Não pertence, utilizado para operações matemáticas.
\neg	Negação, utilizado para expressar a ausência de um rótulo.
a_k	Probabilidade de k .
b_k	Valor em binário do dado k obtido através da STFT.
b_f	Valor em binário da frequência f obtido através da STFT.
C	Pixel central do processo de LBP.
fa	Frame de Áudio.
f	Frequência.
f_c	Frequência de corte.
f_k	Frequência de k .
f_{max}	Frequência máxima.
f_{min}	Frequência mínima.
f_{sr}	Frequência de Nyquist.
\bar{F}	Frame de vídeo médio.
F	Frame de vídeo.
h	Classificador.

Hz	Valor em Hertz, podendo variar para kHz (kilo Hertz) e GHz (Giga Hertz).
i	Variável, contador.
I	Conjunto de rótulos Irrelevantes.
k	Representação de Dados.
L	Conjunto de rótulos.
$L(x)$	Conjunto de rótulos de x .
n	Frame de vídeo médio de exemplo.
N_F	Quantidade de Frames de vídeo.
N_{fa}	Quantidade de Frames de áudio.
N_k	Quantidade de dados.
N_P	Quantidade de vizinhos.
P	Pixels vizinhos do processo de LBP.
$P()$	Probabilidade a Posteriori.
R	Distância entre pixels do processo de LBP.
t	Frame de vídeo de exemplo.
T	Conjunto de rótulos Relevantes.
\vec{u}	Vetor u .
u	Posição u .
v	Posição v .
x	Instância da base.
X_{fa}	Vetor de Frames de áudio.
y	Espectro de Magnitude.
w	Janela de análise Hamming ou Hanning.
z	Vizinhos da classe.

Sumário

1	INTRODUÇÃO	21
1.1	Definição do Problema	21
1.2	Objetivo Principal	22
1.3	Questões de Pesquisa	22
1.4	Estrutura do Trabalho	23
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Conceitos de Áudio	25
2.2	Conceitos de Vídeo	32
2.2.1	Histograma de cores	32
2.2.2	Diferença acumulada	33
2.2.3	Fluxo Óptico	34
2.2.4	VGG16.fc7 + ImageNet	34
2.2.5	Local Binary Patterns (LBP)	35
2.3	Classificação Multi-rótulo	36
2.3.1	Binary Relevance (BR)	36
2.3.2	BRkNN	37
2.3.3	Calibrated Label Ranking (CLR)	38
2.3.4	HOMER	38
2.3.5	Label Powerset (LP)	39
2.3.6	MLkNN	40
2.3.7	RakEL	40
2.4	Fusão Multi-rótulo	41
2.4.1	Fusão Precoce	41
2.4.2	Fusão Tardia	43
3	ESTADO DA ARTE	45
3.1	Utilizando apenas o áudio	45
3.2	Utilizando o áudio e imagens faciais	46
3.3	Utilizando o áudio e EEG	46
3.4	Utilizando o áudio e o vídeo	47
3.5	Comparativos	49
4	METODOLOGIA DE PESQUISA	51
4.1	Preparo da base de dados	51
4.2	Extração de características e geração de arquivos	54

4.2.1	Extração de Características Auditivas	54
4.2.2	Extração de Características Visuais	56
4.2.3	Geração de Arquivos	59
4.3	Classificação	61
4.3.1	Geração de Predições	62
4.3.2	Avaliação Métrica	63
4.4	Fusão de bases	64
5	EXPERIMENTOS	67
5.1	Configuração Experimental	67
5.2	Primeiro Experimento	71
5.3	Segundo Experimento	76
6	CONCLUSÃO	93
	REFERÊNCIAS	95
	APÊNDICES	99
	APÊNDICE A – BASE CAL500 - COLETA DE VÍDEOS MUSICAIS	101
	APÊNDICE B – BASE CAL500 ESTENDIDA	125

1 Introdução

Tanto o vídeo quanto a música tem o objetivo de transmitir para o público emoções sejam elas felizes, tristes, raivosas entre outras. Essas emoções podem aparecer isoladas ou em conjunto com outras emoções pois nem sempre o contexto dos vídeos ou músicas apresenta apenas um único tipo de emoção, como por exemplo, apenas momentos felizes. Grande parte dos trabalhos relacionados a detecção ou classificação de emoções em arquivos multimídia utilizam dados musicais como fonte principal de informações para a extração de características. Porém os vídeos musicais são comuns e fornecem diversas informações adicionais sobre as músicas. Partindo da premissa que os vídeos musicais são criados para transmitir a mesma mensagem e emoções presentes na música, eles possuem diversas informações adicionais que podem ser utilizadas para realizar a extração de características e posterior classificação de emoções. Tomando como base a ideia de que um vídeo musical é criado para representar visualmente uma música, transmitindo assim as mesmas emoções presentes em uma música. Nesse contexto as emoções podem aparecer isoladas ou em conjuntos com outras emoções, porém alguns autores ainda utilizam métodos de classificação pertencentes ao ambiente de único-rótulo, gerando assim resultados que possivelmente não são os mais adequados para detectar ou classificar emoções.

1.1 Definição do Problema

A transmissão de emoções através de arquivos multimídia pode repercutir de maneira direta na vida das pessoas. (TROHIDIS, 2008) cita que os humanos, por natureza, são emocionalmente afetados pela música. (HATAMIKIA, 2014) diz que as emoções têm um papel fundamental na gestão da vida diária de um indivíduo e afetam diretamente a cognição, percepção, atenção, raciocínio, tomada de decisão e memória. À medida que as bases de dados crescem em tamanho e número, a recuperação da música pela emoção está se tornando uma tarefa importante para várias aplicações, como seleção de músicas em dispositivos móveis, sistemas de recomendação musical, programas de TV, rádio e musicoterapia (TROHIDIS, 2008).

O processo de classificação de emoções em vídeos musicais apresenta peculiaridades desde o momento de extração de características, onde as características extraídas podem levar a resultados inadequados ou ruins. Nem sempre os métodos de extração de características mais robustos conseguem fornecer informações relevantes para o processo de detecção de emoções, por isso é necessário realizar uma extração de características através de métodos robustos e simples, a fim de se comparar qual método é mais apropriado para

alcançar o resultado desejado. A quantidade de características extraídas também pode impactar nos resultados, pois além de aumentar a quantidade de informações que podem ser irrelevantes aumenta também o tempo de classificação.

Outro desafio está na combinação de informações auditivas e visuais, assim como na extração de características, este procedimento pode levar a resultados piores do que na classificação utilizando apenas características auditivas ou visuais de forma isolada. A seleção da forma de combinação de bases pode impactar nos resultados da classificação e também no tempo de classificação e obtenção de resultados. A classificação por sua vez pode variar de classificador para classificador, onde cada classificador pode ser melhor que outro dependendo do tipo de característica utilizada. Há necessidade de selecionar mais de um classificador para que seja realizada a comparação de resultados gerados por eles. Tais resultados devem ser avaliados através de métricas pertencentes ao ambiente multi-rótulo, pois as métricas de ambiente de único-rótulo não se aplicam no ambiente multi-rótulo.

Para realizar uma classificação multi-rótulo de vídeos musicais é necessário uma base de dados rotulada de forma que cada instância da base possa receber mais de um rótulo e que a base seja relacionada a vídeos musicais. Para garantir que os dados da base sejam reais é necessário que a base tenha sido rotulada por espectadores humanos, ou seja, que a base possua as informações de emoções que o espectador humano sentiu e identificou durante o processo de classificação. É necessário também que cada vídeo musical seja rotulado por mais de um espectador para que a classificação apresente melhor confiança.

1.2 Objetivo Principal

O objetivo principal deste trabalho é realizar a classificação multi-rótulo de emoções em vídeos musicais utilizando características visuais e auditivas.

Os objetivos específicos deste trabalho são:

- Estender a base de dados CAL500 com vídeos musicais;
- Realizar e avaliar experimentos utilizando diferentes tipos de características visuais para a classificação multi-rótulos de emoções em vídeos musicais;
- Realizar e avaliar experimentos utilizando a combinação das diferentes características visuais com características auditivas.

1.3 Questões de Pesquisa

A execução do trabalho terá além de seu objetivo, questões a serem respondidas com base na classificação de emoções em vídeos musicais utilizando características auditivas e

visuais. As questões a serem respondidas estão abaixo listadas:

1. A utilização de informações visuais é relevante no processo de classificação de vídeos musicais?
2. A classificação de vídeos musicais por emoção utilizando apenas características visuais apresenta resultados relevantes?
3. Métodos mais robustos de extração de características visuais são melhores do que métodos simples?
4. A combinação de informações visuais combinadas com as informações auditivas melhora as predições em relação ao uso de apenas informações auditivas?
5. A combinação de informações visuais combinadas com as informações auditivas melhora as predições em relação ao uso de apenas informações visuais?
6. Quais os melhores classificadores para a classificação de emoções em vídeos musicais no ambiente multi-rótulo?

1.4 Estrutura do Trabalho

O Capítulo 2 fornece uma breve fundamentação para entender melhor as técnicas utilizadas para extração de características auditivas (seção 2.1) e visuais (seção 2.2), apresenta também os classificadores multi-rótulo utilizados (seção 2.3) e os métodos de fusão de bases em ambiente multi-rotulo (seção 2.4). O Capítulo 3 apresenta os trabalhos relacionados ao tema proposto, juntamente com uma breve explicação sobre como este se difere dos demais trabalhos. O Capítulo 4 explica a metodologia de pesquisa utilizada, explicando cada uma das etapas que serão seguidas. Após isso, o Capítulo 5 mostra os Experimentos realizados juntamente com os resultados obtidos e o Capítulo 6 apresenta a conclusão do trabalho e os trabalhos futuros. No apêndice é apresentada a fonte de aquisição para cada vídeo musical da base CAL500 Estendida juntamente com os rótulos para cada vídeo musical.

2 Fundamentação teórica

Durante a pesquisa foi necessária a compreensão de conceitos distintos para cada passo da execução dos experimentos. Tais conceitos foram separados em Conceitos de Áudio, Conceitos de Vídeo, Classificação Multi-rótulo e Fusão Multi-rótulo.

2.1 Conceitos de Áudio

Para a extração de características do Áudio foi utilizada a biblioteca Yaafe¹ (MATHIEU, 2010), a qual está disponível para download gratuito no site oficial da biblioteca. Foram utilizadas 26 características disponíveis na biblioteca Yaafe e são descritas a seguir:

1. **Envelope:** Extrai o envelope de amplitude usando a transformada de Hilbert, filtragem de passa baixa e decimação. O envelope é a magnitude do sinal analítico calculado por Hilbert a qual captura as características de variação lenta do sinal, enquanto a fase contém a informação de alta frequência. O envelope pode ser exemplificado através das figuras 1 e 2, onde é aplicada a transformada de Hilbert em um sinal modulado em amplitude de banda lateral dupla onde: a frequência da portadora é de 1 kHz; a frequência de modulação é de 50 Hz; a profundidade de modulação é de 100%; e a taxa de amostragem é de 10 kHz.

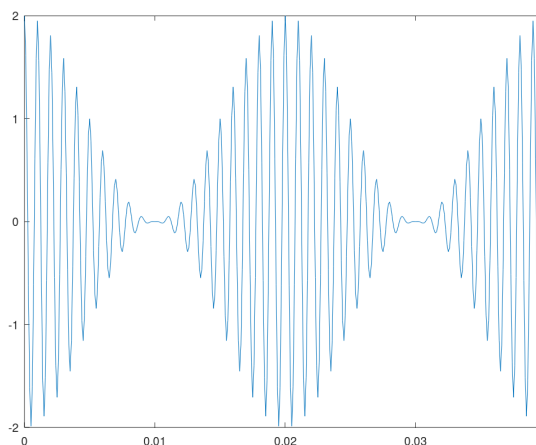


Figura 1 – Sinal modulado de exemplo.

Fonte: Autoria própria

¹ YAAFE. *Yaafe - audio features extraction*. Disponível em: <<http://yaafe.sourceforge.net/>>. Acesso em: 01 jun. 2019.

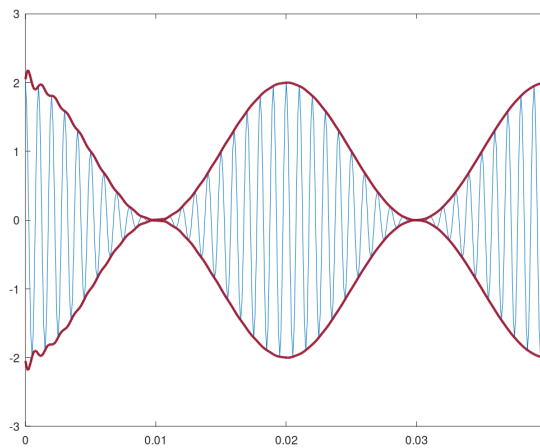


Figura 2 – Transformada de Hilbert usando a figura 1.

Fonte: Autoria própria

- 2. Modulação de Amplitude:** Utiliza da característica de envelope para descrever o Tremolo (faixa de frequência entre 4 e 8 Hz) e o Grain (entre 10 e 40 Gz), os quais são exemplificados nas figuras 3 e 4, para cada um desses intervalos é calculado: a frequência de energia máxima no intervalo; diferença da energia desta frequência e a energia média de todas as frequências; diferença da energia desta frequência e a energia média da faixa; e produto dos dois primeiros valores (ERONEN, 2001).

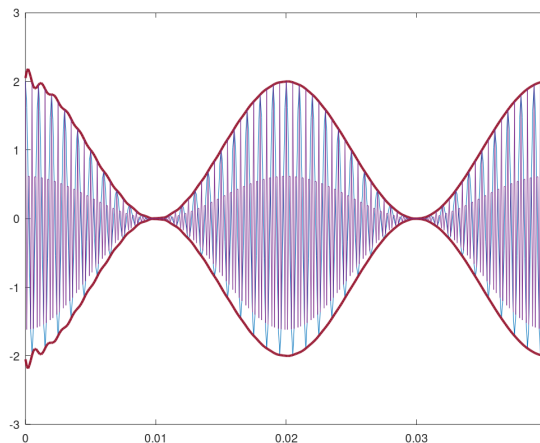


Figura 3 – Exemplificação do Tremolo utilizando o envelope exemplificado na figura 2.

Fonte: Autoria própria

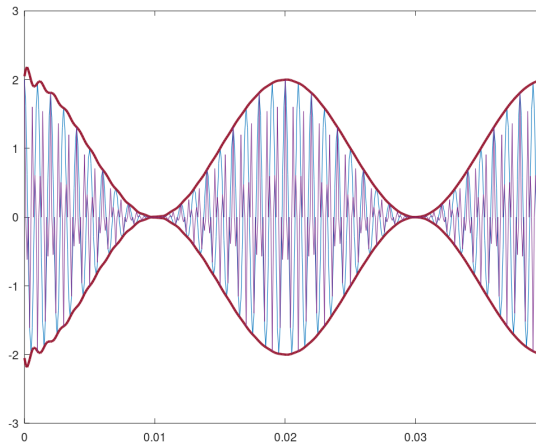


Figura 4 – Exemplificação do Grain utilizando o envelope exemplificado na figura 2.

Fonte: Autoria própria

3. **Frames de Áudio:** Segmenta o sinal de entrada em quadros. O primeiro quadro tem zeros na metade esquerda, de modo que é centrado no tempo 0s, depois os quadros consecutivos são igualmente espaçados. Consequentemente, o quadro fa (a partir de 0) é centralizado na amostra $fa * stepSize$ (por default usamos o valor de 512 para o $stepSize$).
4. **Correlação Automática:** Calcula o valor de correlação automática (*Autocorrelation*) ac em cada frame de áudio através da equação 2.1:

$$ac(fa) = \sum_{i=0}^{N_{fa}-fa-1} X_{fa}(i)X_{fa}(i+fa) \quad (2.1)$$

Onde fa representa o frame de áudio atual, N_{fa} representa a quantidade total de frames de áudio do vetor de frames X_{fa} .

5. **Domínio Complexo na Detecção de Início:** O domínio complexo na detecção de início (*Complex Domain Onset Detection*) detecta o começo usando um método de fluxo espectral de domínio complexo, o qual gera uma função de detecção que é nítida na posição de início em qualquer lugar do áudio. É baseado na combinação dos métodos de detecção de início baseado em energia (*Energy-Based Onset Detection*) e detecção de início baseada em fase (*Phase-based onset detection*), descrita em (DUXBURY, 2003).
6. **Energia:** Calcula a energia (*Energy*) com o raiz média do quadrado de um frame de

áudio através da equação 2.2:

$$en = \sqrt{\frac{\sum_{i=0}^{N_{fa}-1} X_{fa}(i)^2}{N_{fa}}} \quad (2.2)$$

Onde N_{fa} representa a quantidade total de frames de áudio do vetor de frames X_{fa} .

7. **Espectro de Magnitude:** Calcula o espectro de magnitude (*Magnitude Spectrum*) dos frames de áudio usando uma janela de análise (*Hanning* ou *Hamming*) através da equação 2.3:

$$y = abs(fft(w * fa)) \quad (2.3)$$

Onde um espectro de magnitude y é obtido através do valor absoluto (*abs*) do retorno da transformada rápida de Fourier (*fft*), a qual recebe como parâmetros os frames de áudio (fa) e a janela de análise (*Hanning* ou *Hamming*)(w).

8. **Estatística de Forma Espectral:** Calculam estatísticas de forma do Espectro de Magnitude (*Spectral Shape Statistics*). As estatísticas de forma são centroides, espalhamento, assimetria e curtose, definidas pelas equações 2.5, 2.6, 2.7 e 2.8 respectivamente:

$$\Delta_i = \frac{\sum_{k=1}^{N_k} f_k^i * a_k}{\sum_{k=1}^{N_k} a_k} \quad (2.4)$$

$$centroide = \Delta_1 \quad (2.5)$$

$$espalhamento = \sqrt{\Delta_2 - \Delta_1^2} \quad (2.6)$$

$$assimetria = \frac{2\Delta_1^3 - 3\Delta_1\Delta_2 + \Delta_3}{espalhamento^3} \quad (2.7)$$

$$curtose = \frac{-3\Delta_1^4 + 6\Delta_1\Delta_2 - 4\Delta_1\Delta_3 + \Delta_4}{espalhamento^4} - 3 \quad (2.8)$$

A equação 2.4 representa a função base para o cálculo das estatísticas de forma espectral (Δ_i), onde f_k representa a frequência em que um determinado dado k foi observado e a_k representa a probabilidade de observar um determinado dado k . N_k representa a quantidade total de dados que foram observados.

O centroide é calculado através da fórmula de Δ_i utilizando $i = 1$ (descrito na equação 2.5). Definimos o espalhamento espectral como o espalhamento do espectro em torno de seu valor médio, ou seja, a variância da distribuição. A assimetria

retorna uma medida da assimetria de uma distribuição em torno de seu valor médio. A curtose retorna uma medida do nivelamento de uma distribuição em torno de seu valor médio.

9. **Estatísticas de Forma Envelope:** As estatísticas de forma envelope (*Envelope Shape Statistics*) calculam o centroide, espalhamento, assimetria e curtose (equação 2.5, equação 2.6, equação 2.7 e equação 2.8) do envelope de amplitude para cada frame de áudio.
10. **Sonoridade:** Os coeficientes de sonoridade (*Loudness*) são a energia em cada banda de Bark, normalizada pela soma total. A escala de Bark varia de 1 a 24 e corresponde a 24 bandas críticas de audição. As frequências base da escala de Bark de audiometria são, em Hz: 20, 100, 200, 300, 400, 510, 630, 770, 920, 1.080, 1.270, 1.480, 1.720, 2.000, 2.320, 2.700, 3.150, 3.700, 4.400, 5.300, 6.400, 7.700, 9.500, 12.000, 15.500 (MOORE, 1997).
11. **Espectro Mel:** Calcula o Espectro Mel (*Mel Spectrum*) com base no banco de filtros Mel. O banco de filtros Mel é construído com 40 filtros com espaçamento em *log* de acordo com a escala de mel. A escala de frequências de mel é um espaçamento de frequência linear abaixo de 1000 Hz e um espaçamento logarítmico acima de 1000 Hz. Como ponto de referência, o tom de um tom de 1kHz, 40dB acima do limiar de audição perceptual, é definido como 1000 mels. Portanto, podemos usar a equação 2.9 para calcular os mels para uma determinada frequência f em Hz:

$$melfreq = 1127 * \log_e \left(1 + \frac{f}{700} \right) \quad (2.9)$$

Cada filtro é um filtro triangular com altura $2/(f_{max} - f_{min})$.

12. **MFCC:** O MFCC ou Mel-Frequencies Cepstrum Coefficients (*Coefficientes de Cepstrum de Frequência Mel*) utiliza o cálculo do Espectro Mel para gerar os MFCCs, que são calculados usando a equação 2.10:

$$mfcc = dct(\log(abs(fft(Hanning(i).fa)).MelFilterBank)) \quad (2.10)$$

Onde o banco de filtros de mel (*MelFilterBank*, obtido através da equação 2.9) é combinado com a transformada rápida de Fourier (*fft*) utilizando uma janela *Hanning* de tamanho i e os frames de áudio (fa), e o *log* do valor absoluto (*abs*) dessa combinação é utilizado para o cálculo da transformação discreta de cosseno (*Discrete Cosine Transform - dct*).

13. **Redução Espectral:** Calcula a redução espectral (*Spectral Decrease*) com base na equação 2.11, onde o objetivo é calcular a média do conjunto de declives entre

a frequência k e frequência 1. Por isso, enfatiza as encostas das frequências mais baixas.

$$S_{decrease} = \frac{1}{\sum_{k=2}^k b_k} \sum_{k=2}^k \frac{b_k - b_1}{k - 1} \quad (2.11)$$

Onde b_k representa o valor em binário k da magnitude da Transformada de Fourier de Curto Prazo - STFT (*Short-term Fourier Transform*) (PEETERS, 2011).

14. **Nivelamento Espectral:** Calcula o nivelamento espectral (*Spectral Flatness*) global usando a razão entre a média geométrica e a média aritmética.

$$S_{flatness} = \frac{\exp\left(\frac{1}{N_k} \sum_k \log(b_k)\right)}{\frac{1}{N_k}} \quad (2.12)$$

Onde b_k representa o valor em binário k da magnitude da STFT, \exp representa o cálculo do exponencial e N_k representa a quantidade total de dados.

15. **Nivelamento Espectral por Banda:** Calcula o nivelamento espectral por banda (*Spectral Flatness Per Band*) com espaçamento em \log de 1/4 de oitava, faixas de frequência sobrepostas logaritmicamente espaçadas, conforme proposto no padrão MPEG7 (VOZÁRIKOVÁ, 2011).
16. **Fluxo Espectral:** Calcula o fluxo espectral (*Spectral Flux*) entre os frames de áudio consecutivos através da equação 2.13:

$$S_{flux} = \frac{\sum_k (b_k(fa) - b_k(fa - 1))^2}{\sqrt{\sum_k b_k(fa - 1)^2} \sqrt{\sum_k b_k(fa)^2}} \quad (2.13)$$

Onde b_k representa o valor em binário k da magnitude da STFT, fa representa o frame de áudio atual e $fa - 1$ o frame de áudio anterior.

17. **Deslocamento Espectral:** O deslocamento espectral (*Spectral Rolloff*) é a frequência, de modo que a energia contida está abaixo de 95%, assim como descrito na equação 2.14. É utilizado para distinguir a voz falada de outros efeitos sonoros.

$$\sum_{f=0}^{f_c} b_f^2 = 0.95 \sum_{f=0}^{f_{sr}/2} b_f^2 \quad (2.14)$$

Onde b_f representa o valor em binário f da magnitude da STFT, f_c representa frequência de corte e f_{sr} é a frequência de Nyquist. No caso de sons harmônicos, pode-se mostrar experimentalmente que o deslocamento espectral está relacionado à frequência de corte harmônico/ruído (PEETERS, 2011).

18. **Encosta Espectral:** A encosta espectral (*Spectral Slope*) é calculada pela regressão linear da amplitude espectral através da equação 2.15:

$$S_{slope} = \frac{N_k \sum_k f_k b_k - \sum_k f_k \sum_k b_k}{N_k \sum_k f_k^2 - (\sum_k b_k)^2} \quad (2.15)$$

Onde b_k representa o valor em binário k da magnitude da STFT, f_k representa a frequência de k e N_k a quantidade total de dados.

19. **Variação Espectral:** A variação espectral (*Spectral Variation*) é calculada através da correlação normalizada do espectro entre frames de áudio consecutivos através da equação 2.16:

$$S_{var} = 1 - \frac{\sum_k b_k(fa - 1)b_k(fa)}{\sqrt{\sum_k b_k(fa - 1)^2} \sqrt{\sum_k b_k(fa)^2}} \quad (2.16)$$

Onde b_k representa o valor em binário k da magnitude da STFT, fa representa o frame de áudio atual e $fa - 1$ o frame de áudio anterior.

20. **Estatística de Forma Temporal:** A estatística de forma temporal (*Temporal Shape Statistics*) calcula as estatísticas de forma de sinal de frame de áudio, ou seja, com base na aparição dos frames de áudio em forma temporal.
21. **ZCR:** ZCR (*Zero-crossing Rate*) ou taxa de cruzamento zero é um correlato do centroide espectral, também chamado de medida de frequência dominante em um sinal. É definido como o número de cruzamentos de zero no domínio do tempo dentro de uma região definida de sinal, dividido pelo número de amostras dessa região (GOUYON, 2000).
22. **OBSI:** O OBSI (*Octave band signal intensity*) calcula a intensidade do sinal da banda de oitava usando um banco de filtros triangular de oitava. A ideia deste descritor é captar brevemente a distribuição de potência dos diferentes harmônicos do som, sem ter que recorrer a um passo de detecção da frequência fundamental (ESSID, 2005).
23. **OBSIR:** O OBSIR (*Octave band signal intensity ratio*) calcula o *log* de razão do OBSI entre a oitava consecutiva.
24. **Nitidez Perceptual:** Calcula a nitidez dos coeficientes de sonoridade (item 10), é o equivalente perceptual ao centroide espectral, mas calculado usando o volume específico das bandas de Bark (PEETERS, 2004).
25. **Propagação Perceptiva:** Calcula a propagação dos coeficientes de sonoridade (item 10) através da distância do maior valor de intensidade específica até o volume total (PEETERS, 2004).

26. **Fator de Crista Espectral por Banda:** Calcula o fator de crista espectral por faixa espaçada de \log de 1/4 de oitava. É calculado pela relação entre o valor máximo dentro da faixa e a média aritmética do valor do espectro de energia (PEETERS, 2004).

2.2 Conceitos de Vídeo

Para a extração de características de vídeo foi utilizada a biblioteca gratuita em linguagem C++ OpenCV² (BRADSKI, 2000) e da biblioteca gratuita Keras³ (KETKAR, 2017) em linguagem Python. As características extraídas são detalhadas abaixo.

2.2.1 Histograma de cores

O procedimento conhecido como histograma de cores analisa as ocorrências de cada nível de cor em uma determinada imagem. Os níveis de cor são basicamente a intensidade de cada cor para cada pixel, onde em um ambiente RGB (Red, Green e Blue) a variação de intensidade pode gerar qualquer cor utilizando apenas as três cores primárias Vermelho, Verde e Azul. Os níveis de cor no ambiente RGB podem variar de 0 a 255, sendo 0 as cores mais escuras e 255 as mais claras. Um exemplo de histograma é apresentado na figura 5.

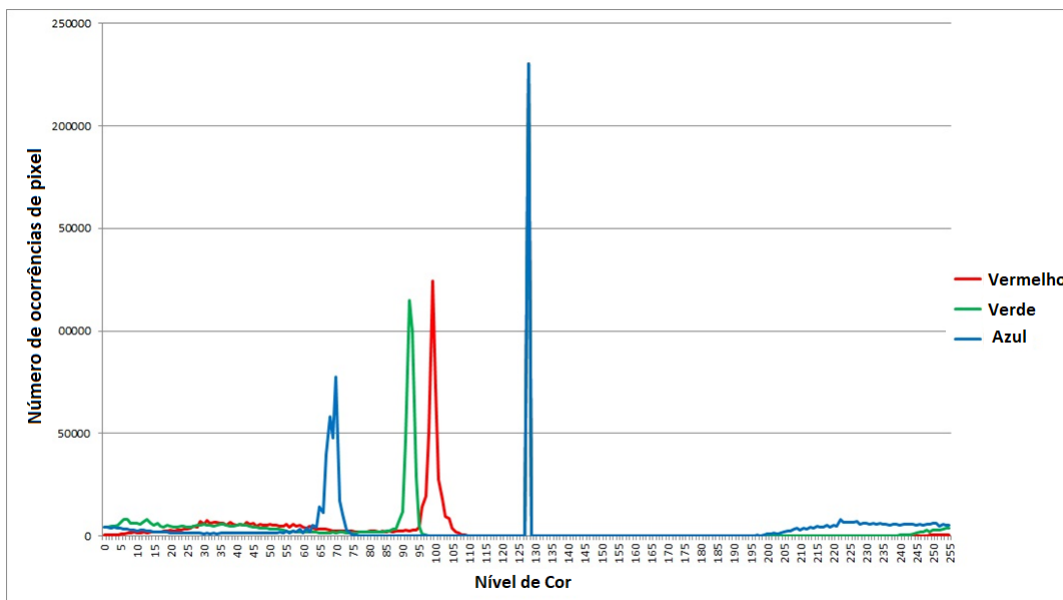


Figura 5 – Histograma RGB de um frame escolhido aleatoriamente na base de dados

Fonte: Autoria própria

² OPENCV. *OpenCV Library*. Disponível em: <<https://opencv.org/>>. Acesso em: 01 jun. 2019.

³ KERAS. *Keras: The Python Deep Learning library*. Disponível em: <<https://keras.io/>>. Acesso em: 01 jun. 2019.

2.2.2 Diferença acumulada

A diferença acumulada é um processo criado para a pesquisa, visando identificar variações entre os frames de vídeo (imagens) para contabilizar o movimento do vídeo comparando os pixels dos frames de vídeo adjacentes. Os valores de cada nível de cor de cada pixel de cada frame de vídeo são comparados aos valores dos frames de vídeo adjacentes (frame de vídeo anterior e frame de vídeo posterior), o valor da diferença é acumulado até atingir os n quadros da divisão, após isso, um frame de vídeo de diferença acumulada é obtido, o qual contém os valores de diferença acumulada entre os frames de vídeo para cada nível de cor e para cada pixel. O procedimento é representado pela fórmula 2.17.

$$\bar{F} = \frac{\sum_{i=1}^{N_F-1} |F_{i+1} - F_i|}{(N_F - 1)} \quad (2.17)$$

Onde F_i representa o frame de vídeo atual e F_{i+1} representa o próximo frame de vídeo, N_F representa o número de frames na divisão e \bar{F} representa o frame de diferença acumulada. A figura 6 apresenta um exemplo do processo para cinco quadros, onde a diferença entre os quadros t e $t+1$, $t+1$ e $t+2$, $t+2$ e $t+3$, $t+3$ e $t+4$ é calculada, e obtemos os quadros n , $n+1$, $n+2$ e $n+3$, que serão representados por um quadro médio que chamamos de frame de saída média, que representa a diferença acumulada entre os quadros t e $t+4$.

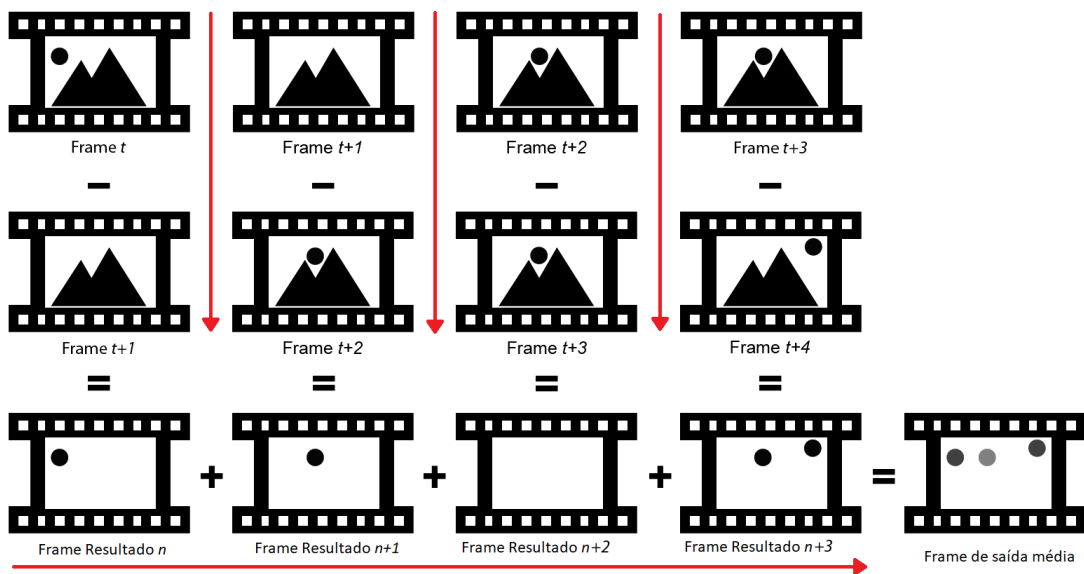


Figura 6 – Exemplo de processo de cálculo da diferença acumulada entre cinco frames de vídeo

Fonte: Autoria própria

2.2.3 Fluxo Óptico

O processo do fluxo óptico descreve o movimento aparente em uma sequência de frames de vídeo. A ideia por trás do fluxo óptico é encontrar para cada pixel um vetor $\vec{u} = (u, v)$ que diga quão rápido o pixel se move pelo frame de vídeo e em que direção o mesmo se move (NETO, 2011). A figura 7 apresenta um exemplo de saída do processo de fluxo óptico, onde os pontos representam os pixels e os vetores representam a direção para onde o pixel se moveu, sendo que vetores maiores representam uma maior movimentação, e conseqüentemente uma maior velocidade do pixel.

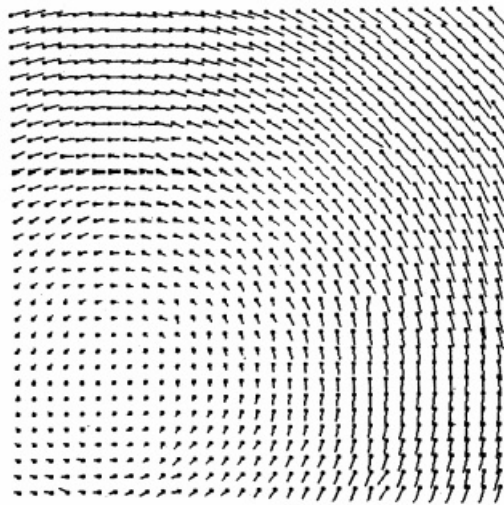


Figura 7 – Exemplo de saída do processo de Fluxo Óptico

Fonte: (HORN, 1981)

2.2.4 VGG16.fc7 + ImageNet

O processo denominado de VGG16.fc7 + ImageNet é a combinação do uso de uma rede convolucional profunda para reconhecimento de imagens em grande escala (VGG16 - (SIMONYAN, 2014)) com um banco de imagens hierárquico de grande escala (ImageNet - (DENG, 2009)). A rede VGG16 realiza em cima de uma imagem redimensionada para 224x224, vários procedimentos de convolução a fim de refinar e extrair informações da imagem. A figura 8 mostra a arquitetura da rede, onde a camada utilizada é a "fully connected"(fc7) com tamanho 1 x 1 x 4096, a qual fornece 4096 características para a imagem de entrada. O banco de imagens ImageNet é utilizado como entrada para o VGG16, a qual realiza o treinamento da rede através das imagens do banco.

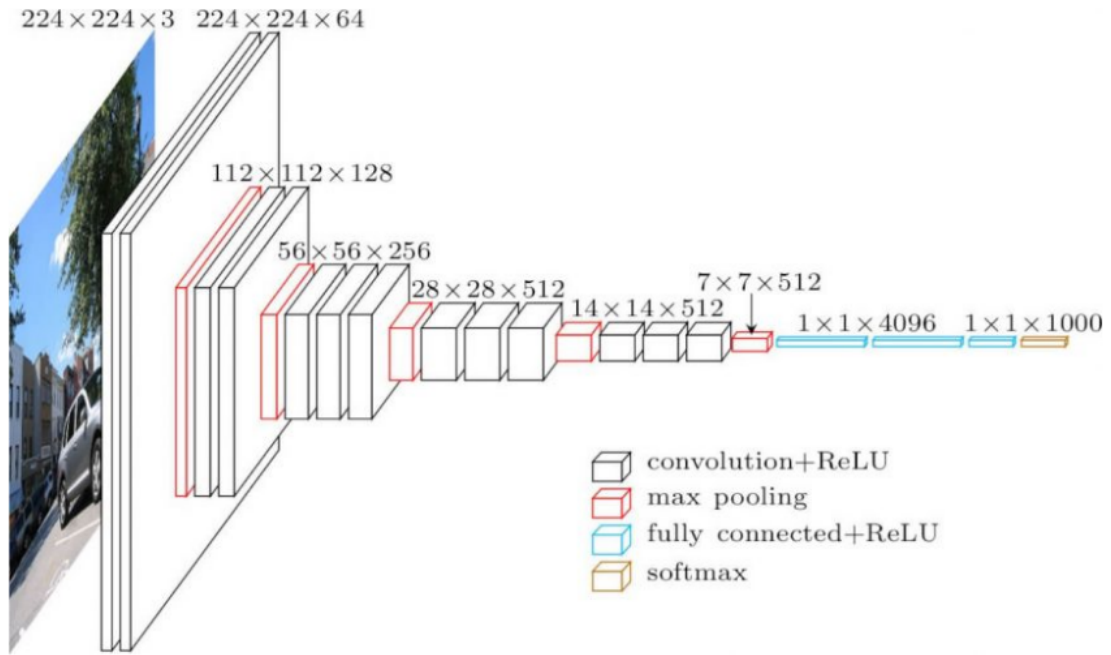


Figura 8 – Arquitetura do VGG16

Fonte: <https://neurohive.io/en/popular-networks/vgg16/>

2.2.5 Local Binary Patterns (LBP)

O *Local Binary Patterns* ou LBP apresentado por (OJALA, 2002) e descrito em (COSTA, 2012) é um descritor de textura que leva em consideração cada pixel C e seus P vizinhos igualmente espaçados a uma distância R assim como é apresentado na figura 9

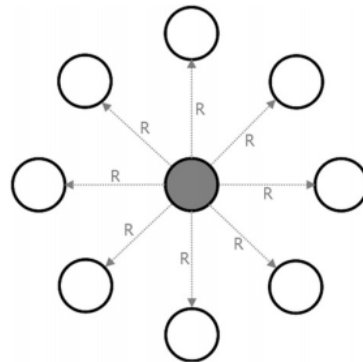


Figura 9 – Operação do LBP. O Pixel C que se encontra no meio e seus P vizinhos a uma distância R

Fonte: (COSTA, 2012)

A execução do LBP é realizada com base no pixel C , onde ele compara o valor do pixel com o valor de seus P vizinhos individualmente, começando pelo pixel superior esquerdo e seguindo em sentido anti-horário a fim de passar pelos oito vizinhos, quando o valor do pixel vizinho é superior ao valor do pixel C ele recebe o valor "1", caso

contrário recebe o valor "0", gerando assim uma sequência binária de oito dígitos (exemplo: "00101101") que é atribuído ao valor do pixel central. Este procedimento é realizado em toda a imagem gerando assim uma nova imagem LBP, na qual é realizado o procedimento de histograma onde é gerado um vetor com 256 posições os quais contêm o somatório de ocorrências de um determinado valor na imagem LBP.

Em cima do vetor de histograma é realizado o procedimento chamado padrão uniforme que é utilizado para reduzir o comprimento do vetor de recursos, onde a ideia é motivada pelo fato de alguns padrões binários ocorrerem mais comumente em imagens de textura do que outros. Um padrão binário local é chamado de uniforme se o padrão binário contiver no máximo duas transições 0-1 ou 1-0. Exemplo: "00100000"(possui 2 transições) é um padrão uniforme e "01010100"(possui 6 transições) não é um padrão uniforme. Com base no padrão uniforme os vetores que não são uniformes são armazenados em uma única posição do histograma reduzindo assim o tamanho do histograma de 256 para 59 posições (58 posições com vetores uniformes e 1 posição com vetores não uniformes).

2.3 Classificação Multi-rótulo

A classificação multi-rótulo é realizada quando as instâncias de uma base possuem ou podem possuir mais de um rótulo. Como exemplo podemos levantar uma base musical onde uma música pode receber vários rótulos em conjunto (Feliz; Feliz e Romântico; Triste e Romântico; Triste e Agressivo; Triste, agressivo e Romântico), diferente da classificação único-rótulo, onde uma instância pode apresentar apenas um rótulo (Gato, Cachorro, Papagaio). Por se diferenciar dos métodos de classificação de único-rótulo é necessário que os classificadores sejam pertencentes ao ambiente multi-rótulo, e nesta seção são descritos alguns classificadores multi-rótulo.

2.3.1 Binary Relevance (BR)

O *Binary Relevance* (Relevância Binária) descrito em (TSOUMAKAS, 2009) é uma transformação de problema popular, onde para cada rótulo é criado uma nova base binária de único-rótulo, onde basicamente é descrito se cada instância apresenta ou não aquele rótulo como classificação. Após a transformação da base multi-rótulo em uma base único-rótulo, é aplicada a classificação único-rótulo tradicional. O procedimento realizado pela transformação do Binary Relevance é exemplificado abaixo. Dado uma base multi-rótulo de exemplo possuindo 4 rótulos distintos: λ_1 , λ_2 , λ_3 e λ_4 (tabela 1).

Exemplo	Atributos	Rótulos
1	x_1	$\{\lambda_2, \lambda_4\}$
2	x_2	$\{\lambda_2, \lambda_3\}$
3	x_3	$\{\lambda_1\}$
4	x_4	$\{\lambda_1, \lambda_3, \lambda_4\}$

Tabela 1 – Exemplo de base de dados Multi-Rótulo

Após aplicar a transformação do Binary Relevance serão criadas quatro novas bases, onde cada base representa um rótulo distinto (tabela 2).

Exemplo	Rótulos
1	$\neg\lambda_1$
2	$\neg\lambda_1$
3	λ_1
4	λ_1

(a)

Exemplo	Rótulos
1	λ_2
2	λ_2
3	$\neg\lambda_2$
4	$\neg\lambda_2$

(b)

Exemplo	Rótulos
1	$\neg\lambda_3$
2	λ_3
3	$\neg\lambda_3$
4	λ_3

(c)

Exemplo	Rótulos
1	λ_4
2	$\neg\lambda_4$
3	$\neg\lambda_4$
4	λ_4

(d)

Tabela 2 – Exemplo de resultados da aplicação do BR sobre a base de dados da tabela 1

Na tabela 2 são apresentadas as quatro novas bases de único-rótulo, onde cada base possui característica binária, ou seja, cada base representa apenas um rótulo e em suas instâncias são descritas as presenças (λ) ou ausências ($\neg\lambda$) dos rótulos. Após as bases único-rótulos serem geradas a classificação de único-rótulo será executada em cada uma das bases.

2.3.2 BR k NN

O BR k NN proposto em (SPYROMITROS, 2008) é uma adaptação do algoritmo de único-rótulo k NN (k -Nearest Neighbor) e conceitualmente equivalente ao uso do BR (seção 2.3.1) em conjunto com o algoritmo k NN, no entanto o BR k NN tem um desempenho mais alto do que o uso do BR e k NN emparelhados, para uma base com L rótulos o BR em conjunto com o k NN executa L vezes o procedimento do k NN já o BR k NN executa apenas uma vez pois elimina cálculos redundantes estendendo o algoritmo k NN para que predições independentes sejam feitas para cada rótulo, o tornando L vezes mais rápido do que o BR em conjunto com o k NN.

2.3.3 Calibrated Label Ranking (CLR)

O *Calibrated Label Ranking* (CLR ou Classificação de Rótulo Calibrada) proposto em (FÜRNKRANZ, 2008), separa os L rótulos de uma base em rótulos relevantes (T) e rótulos irrelevantes (I), onde $T \cup I = L$, separando-os através de um calibrador de rótulos λ_0 , tal que $T > \lambda_0 > I$. Para exemplificar o processo realizado pelo CLR utilizaremos uma base L com 18 rótulos.

Dada uma base $L = \{\lambda_1, \lambda_2, \dots, \lambda_{17}, \lambda_{18}\}$ e um calibrador λ_0 , é realizado a calibração, onde a base será dividida em duas com base no calibrador gerando assim um ranking calibrado.

$$\lambda_1 > \dots > \lambda_9 > \lambda_0 > \lambda_{10} > \dots > \lambda_{18}$$

que induz um ranking entre os rótulos

$$\lambda_1 > \dots > \lambda_9 > \lambda_{10} > \dots > \lambda_{18}$$

e uma bipartição

$$T = \{\lambda_1, \dots, \lambda_9\} \text{ e } I = \{\lambda_{10}, \dots, \lambda_{18}\}$$

de forma direta.

Após separar os rótulos em relevantes (T) e irrelevantes (I), é realizado um procedimento semelhante ao realizado pelo BR (seção 2.3.1), realizando assim uma classificação binária de único-rótulo utilizando apenas os rótulos que são considerados como relevantes (T).

2.3.4 HOMER

O algoritmo HOMER (*Hierarchy Of Multi-label classifiERs*) proposto em (TSOU-MAKAS, 2008) segue o design do paradigma de dividir para conquistar. A ideia principal é a transformação de uma tarefa multi-rótulo sob um grande conjunto de rótulos L em uma hierarquia em forma de árvore de tarefas, onde os rótulos são separados em grupos menores para que a classificação ocorra em cada um destes grupos menores para depois unir o todo. A figura 10 apresenta um exemplo do procedimento realizado pelo HOMER onde a partir de um conjunto de oito rótulos ($\lambda_1, \dots, \lambda_8$) são gerados conjuntos menores de rótulos, até que ao chegar nas folhas da árvore, onde são apresentados os rótulos de forma isolada. Após a geração da árvore de rótulos são realizadas as classificações dos nós da árvore através de um classificador multi-rótulo (h_i) onde são gerados os méta-rótulos (μ_i).

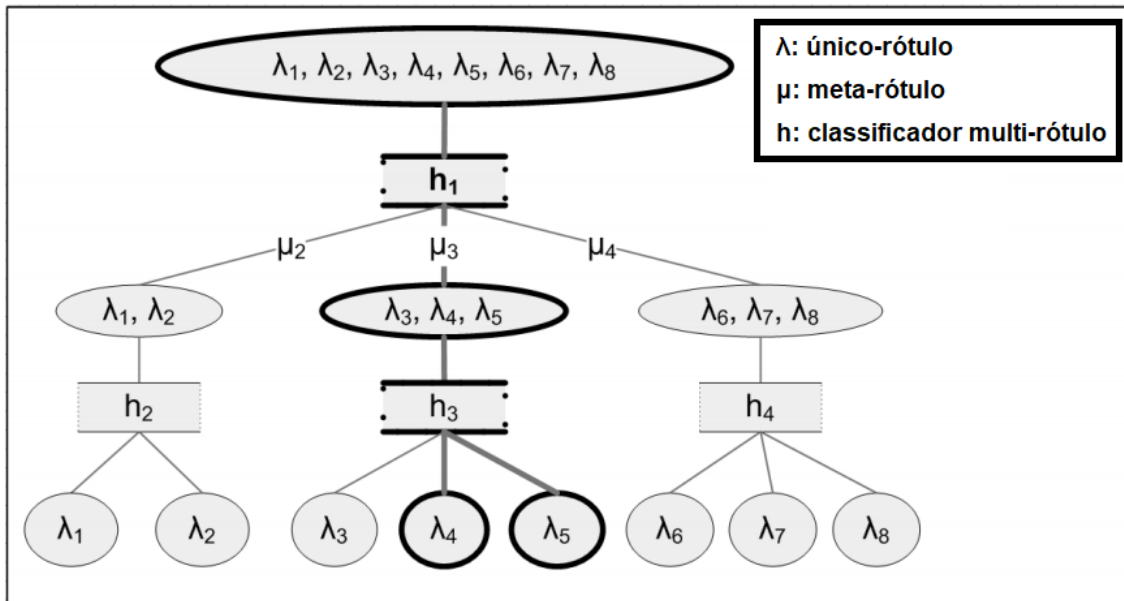


Figura 10 – Exemplo do algoritmo HOMER em uma base com 8 rótulos.

Fonte: (TSOUMAKAS, 2008)

2.3.5 Label Powerset (LP)

O *Label Powerset* (LP ou Conjunto de Força de Rótulos) descrito em (TSOUMAKAS, 2009) é um método de transformação de problemas simples, porém eficaz. Ele considera que cada conjunto exclusivo de rótulos que existem em um conjunto de treinamento multi-rótulos como uma das classes de uma nova tarefa de classificação de único-rótulo. Baseado na base de exemplo da tabela 1, após a transformação via LP a base multi-rótulo se transforma na base único-rótulo apresentada na tabela 3.

Exemplo	Rótulos
1	$\lambda_{2,4}$
2	$\lambda_{2,3}$
3	λ_1
4	$\lambda_{1,3,4}$

Tabela 3 – Aplicação da transformação LP no exemplo de base da tabela 1

Na tabela 3 podemos destacar que os rótulos λ_1 , λ_2 , λ_3 e λ_4 da base multi-rótulo original (tabela 1) não fazem parte da nova base, onde foram substituídos pelos rótulos únicos $\lambda_{2,4}$, $\lambda_{2,3}$, λ_1 e $\lambda_{1,3,4}$, formando assim uma nova base de único-rótulo na qual será realizada uma classificação em ambiente de único-rótulo através de um classificador de único-rótulo.

2.3.6 MLkNN

O MLkNN ou *Multi-Label kNN* proposto em (ZHANG; ZHOU, 2007) e descrito também em (LIU, 2015) é uma adaptação do algoritmo kNN para o ambiente multi-rótulo e se baseia no algoritmo tradicional kNN e no princípio de MAP (*Maximum a Posteriori*) da estatística Bayesiana. A ideia principal do MLkNN é que os rótulos de uma instância dependem do número de vizinhos que possuem rótulos idênticos. Dada uma instância x com um conjunto de rótulos desconhecidos $L(x) \subseteq L$, o MLkNN identifica primeiro os N_P vizinhos mais próximos nos dados de treinamento e conta o número de vizinhos pertencentes a cada classe (ou seja, uma variável j de 0 a N_P). Em seguida o princípio de MAP é usado para determinar o conjunto de rótulos para a instância de teste. A probabilidade a posteriori de um rótulo $\lambda_i \in L(x)$ é representada por:

$$P(\lambda_i \in L(x)|j) = \frac{P(j|\lambda_i \in L(x)).P(\lambda_i \in L(x))}{P(j)} \quad (2.18)$$

Onde z é o número de vizinhos pertencentes a cada classe ($0 \leq j \leq N_P$). Então, para cada rótulo $\lambda_i \in L(x)$, o algoritmo constrói um classificador h_i usando a regra:

$$h_i(x) = \begin{cases} 1 & P(\lambda_i \in L(x)|j) > P(\lambda_i \notin L(x)|j) \\ 0 & \text{de outra forma} \end{cases} \quad (2.19)$$

Onde é comparada a probabilidade do rótulo λ_i pertencer ($\lambda_i \in L(x)$) ou não pertencer ($\lambda_i \notin L(x)$) ao conjunto $L(x)$ e sendo $0 \leq j \leq N_P$. Se $h_i(x) = 1$, significa que o rótulo λ_i está no conjunto de rótulos real enquanto se for 0 significa que não está no conjunto.

2.3.7 RakEL

O RakEL (*Random k labELsets*) proposto em (TSOUMAKAS; VLAHAVAS, 2007) é um classificador baseado no LP (seção 2.3.5), onde seleciona aleatoriamente um rótulo λ_i dentro de L , e treina um classificador em ambiente de único-rótulo a partir do conjunto de dados gerados pelo LP, contendo o rótulo selecionado λ_i . Após as execuções, ele toma a decisão por procuração de voto para determinar a previsão. A tabela 4 apresenta um exemplo de execução do RakEL quando selecionado aleatoriamente os rótulos λ_2 e λ_3 .

Onde (a) representa a base gerada quando o rótulo λ_2 é selecionado e (b) representa a base gerada quando o rótulo λ_3 é selecionado.

Exemplo	Rótulos
1	$\lambda_{2,4}$
2	$\lambda_{2,3}$

(a)

Exemplo	Rótulos
2	$\lambda_{2,3}$
4	$\lambda_{1,3,4}$

(b)

Tabela 4 – Exemplo de resultados da aplicação do Ra k EL sobre a base de dados da tabela 1.

2.4 Fusão Multi-rótulo

O procedimento de fusão/combinção de bases é um artifício utilizado para melhorar as predições através do acúmulo de informações, porém tal procedimento pode ser realizado apenas no caso das informações presentes nas bases serem oriundas da fonte de informações, como exemplo deste trabalho do mesmo vídeo musical. Além da necessidade da fonte de informações ser a mesma, é necessário garantir que as instâncias que serão combinadas (tomando como base a combinação de arquivos ARFF) representam um mesmo elemento na base, ou seja, as informações contidas nos dois arquivos devem ser referentes a exatamente a mesma fonte de informação, onde a diferença entre as bases é basicamente o tipo de informação extraída. Caso os requisitos sejam atendidos, a fusão de bases pode ser realizada para dois ou mais arquivos e podem ocorrer de duas maneiras distintas, antes da classificação (Fusão Precoce ou *Early Fusion*) ou após a classificação (Fusão Tardia ou *Late Fusion*), as quais são descritas de forma detalhada em (SNOEK, 2005).

2.4.1 Fusão Precoce

Na fusão precoce as informações ou características presentes nas bases são combinadas de forma que o resultado desta combinação apresente as informações presentes nas bases combinadas. Na combinação de arquivos ARFF é realizada uma concatenação dos atributos de cada instância e uma replicação dos rótulos, os quais são idênticos para as bases, tomando como base o requisito de que a única diferença entre as bases são as informações extraídas e que os rótulos são provindos da fonte de informações. O procedimento de fusão precoce é exemplificado abaixo, onde os arquivos ARFF das figuras 11 e 12 são combinados e o resultado é apresentado na figura 13.

```

1 % 1. Title: Detection of emotions in music–video
2 %
3 % 2. Sources:
4 % (a) Creator: B. Kostiuk
5 % (b) Fundation: Pontificia Universidade Catolica do Parana – PUCPR
6 %
7 % 3. Type: Optical Flow Video Features Multi–Label File
8
9 @RELATION Fluxo_Optico_Video_Features_Mulan
10 @ATTRIBUTE FluxoOp1 NUMERIC
11 @ATTRIBUTE FluxoOp2 NUMERIC
12 @ATTRIBUTE FluxoOp3 NUMERIC
13 @ATTRIBUTE HAPPY {0, 1}
14 @ATTRIBUTE SAD {0, 1}
15 @ATTRIBUTE BIZARRE_WEIRD {0, 1}
16 @ATTRIBUTE ANGRY_AGRESSIVE {0, 1}
17
18 @data
19 05,15,25,1,0,1,0
20 15,25,35,1,0,0,1
21 25,35,45,0,1,1,0
22 35,45,55,0,1,0,1
23 45,55,65,0,0,0,1
24 55,65,75,1,1,1,1

```

Figura 11 – Base de Exemplo 1, arquivo ARFF multi-rótulo com características de Fluxo Óptico.

```

1 % 1. Title: Detection of emotions in music–video
2 %
3 % 2. Sources:
4 % (a) Creator: B. Kostiuk
5 % (b) Fundation: Pontificia Universidade Catolica do Parana – PUCPR
6 %
7 % 3. Type: Audio Features Multi–Label File
8
9 @RELATION Audio_Features_Mulan
10 @ATTRIBUTE Audio1 NUMERIC
11 @ATTRIBUTE Audio2 NUMERIC
12 @ATTRIBUTE Audio3 NUMERIC
13 @ATTRIBUTE Audio4 NUMERIC
14 @ATTRIBUTE Audio5 NUMERIC
15 @ATTRIBUTE HAPPY {0, 1}
16 @ATTRIBUTE SAD {0, 1}
17 @ATTRIBUTE BIZARRE_WEIRD {0, 1}
18 @ATTRIBUTE ANGRY_AGRESSIVE {0, 1}
19
20 @data
21 10,20,30,40,50,1,0,1,0
22 20,30,40,50,60,1,0,0,1
23 30,40,50,60,70,0,1,1,0
24 40,50,60,70,80,0,1,0,1
25 50,60,70,80,90,0,0,0,1
26 60,70,80,90,00,1,1,1,1

```

Figura 12 – Base de Exemplo 2, arquivo ARFF multi-rótulo com características de Áudio.

```

1 % 1. Title: Detection of emotions in music–video
2 %
3 % 2. Sources:
4 % (a) Creator: B. Kostiuk
5 % (b) Fundation: Pontificia Universidade Catolica do Parana – PUCPR
6 %
7 % 3. Type: Audio Features fusion Optical Flow Features Multi–Label File
8
9 @RELATION Audio_Fusion_FluxoOp_Features_Mulan
10 @ATTRIBUTE FluxoOp1 NUMERIC
11 @ATTRIBUTE FluxoOp2 NUMERIC
12 @ATTRIBUTE FluxoOp3 NUMERIC
13 @ATTRIBUTE Audio1 NUMERIC
14 @ATTRIBUTE Audio2 NUMERIC
15 @ATTRIBUTE Audio3 NUMERIC
16 @ATTRIBUTE Audio4 NUMERIC
17 @ATTRIBUTE Audio5 NUMERIC
18 @ATTRIBUTE HAPPY {0, 1}
19 @ATTRIBUTE SAD {0, 1}
20 @ATTRIBUTE BIZARRE_WEIRD {0, 1}
21 @ATTRIBUTE ANGRY_AGRESSIVE {0, 1}
22
23 @data
24 05,15,25,10,20,30,40,50,1,0,1,0
25 15,25,35,20,30,40,50,60,1,0,0,1
26 25,35,45,30,40,50,60,70,0,1,1,0
27 35,45,55,40,50,60,70,80,0,1,0,1
28 45,55,65,50,60,70,80,90,0,0,0,1
29 55,65,75,60,70,80,90,00,1,1,1,1

```

Figura 13 – Fusão precoce entre as bases de exemplo 1(figura 11) e 2(figura 12).

Após a fusão precoce, o arquivo de saída (figura 13) será utilizado no ambiente de classificação.

2.4.2 Fusão Tardia

Na fusão tardia o dados que são combinados são as predições dos classificadores, onde a lista de predições de cada instância de uma determinada base é combinada com a lista de predições de cada instância de outra base. As predições geradas pelos classificadores são dividias em três partes, a primeira e mais importante é a confiança (*Confidence*) onde é apresentado um vetor com as taxas de confiança que cada um dos rótulos pertence aquela instância, a qual varia de 0 a 1, onde 1 representa 100% de certeza de que aquele rótulo pertence a instância e 0 representa 100% de certeza de que aquele rótulo NÃO pertence a instância. Com base no vetor de confiança gerado o vetor de ranking, onde os rótulos são ordenados de acordo com sua taxa de confiança, onde os rótulos com maior taxa de confiança ficam nas melhores posições do ranking, e com base no ranking é gerado o vetor de bipartição (*Bipartion*) onde basicamente os rótulos são divididos em dois grupos, os presentes (taxa de confiança maior ou igual a 50%) e os não presentes (taxa de confiança

menor que 50%) na instância. A fusão tardia trabalha utilizando os vetores de confiança onde os vetores são combinados e é gerado um novo vetor de confiança, o qual é utilizado para a geração do novo vetor de bipartição. Após a combinação é gerado um arquivo (figura 14) contendo os vetores de bipartição e de confiança, e esse arquivo é utilizado pela biblioteca Mulan para gerar os resultados métricos. O novo vetor de ranking é gerado automaticamente pela biblioteca Mulan.

```
1 Bipartion: [false , true , true , true] Confidences: [0.041667, 0.979167,  
0.589286, 0.983333]  
2 Bipartion: [false , true , false , true] Confidences: [0.055254, 0.960648,  
0.126701, 0.983333]  
3 Bipartion: [false , false , false , true] Confidences: [0.041667, 0.479167,  
0.134158, 0.625000]  
4 Bipartion: [false , true , false , true] Confidences: [0.041667, 0.979167,  
0.089286, 0.500000]  
5 Bipartion: [false , false , false , true] Confidences: [0.067552, 0.479167,  
0.101786, 1.000000]  
6 Bipartion: [false , true , false , true] Confidences: [0.067529, 0.966095,  
0.136352, 0.513228]
```

Figura 14 – Exemplo de arquivo de predições.

3 Estado da arte

A detecção de emoções em arquivos multimídia pode ser dividida em grupos relacionados a origem de seus dados, os quais podem ser dados sonoros coletados do áudio do arquivo multimídia, dados neurais que são coletados diretamente do espectador através de eletroencefalograma (EEG), dados faciais que são coletados através de gravações das reações do espectador e dados de vídeo que são coletados através do próprio arquivo multimídia. Como forma de busca por trabalhos relacionados a detecção de emoção em vídeos musicais, foi realizado um levantamento de trabalhos no IEEE. Com o intuito de analisar qualquer trabalho envolvendo emoções em arquivos multimídia, foi utilizada a string de busca: *((movie OR video OR audio OR music-video OR videoclip) AND (emotion OR emotional OR emotions) AND (classifier OR detection OR classification or recognition))*. A busca realizada retornou um total de 872 trabalhos, os quais foram filtrados através do resumo e selecionados com base na utilização do áudio e do vídeo na detecção de emoções. Os trabalhos selecionados são apresentados abaixo.

3.1 Utilizando apenas o áudio

(TROHIDIS, 2008) classifica músicas por emoção utilizando os dados sonoros, ou seja, apenas o arquivo de áudio, os quais são classificados de forma multi-rótulo em seis rótulos de emoção ((Espanto/Surpresa, Feliz/Satisfeito, Relaxante/Calmo, Quietos, Triste, Zangado/Medo). Em seu trabalho foram utilizados 100 arquivos musicais para cada um dos 7 gêneros: Clássica, reggae, Rock, Pop, Hip-Hop, Techno e Jazz, selecionados a partir de 223 álbuns musicais, após a seleção em cada música foram retirados os 30 primeiros segundos e os 30 últimos segundos e convertidos em arquivos wave. A rotulação das músicas foi realizada através de três estudantes de música, onde cada um as classificou de forma multi-rótulo para cada uma das emoções acima listadas. Após o preparo da base foram extraídas características de Ritmo (através da quantidade de batidas por minuto) e de Timbre (através do MFCC). Como método de classificação, foram utilizadas transformadas multi-rótulo, as quais transformam a base multi-rótulo em uma ou mais bases de único-rótulo. Foram utilizados os classificadores BinaryRelevance (BR), LabelPowerset (LP) e RakEL (Random k labELsets) em conjunto com SVM (Support Vector Machine), e foi utilizado também o classificador ML k NN que é uma variação do k NN para ambiente multi-rótulo. O autor faz uma comparação entre os resultados dos classificadores e destaca o classificador RakEL, o qual obteve 82,36% de acerto através da métrica Micro AUC e também obteve taxas de acerto superiores em outras métricas comparadas aos outros classificadores utilizados. Também foi realizada comparação entre os rótulos, onde a emoção

“Quieto” apresentou melhor taxa de acerto.

3.2 Utilizando o áudio e imagens faciais

(SENG, 2016) utiliza da combinação de Áudio e de imagens faciais dos espectadores para classificar emoções através da fusão tardia das predições obtidas em cada ambiente. Ele utiliza as bases ORL e Yale que são rotuladas em 6 emoções (bravo, feliz, triste, desgosto, surpreso e medo), CK+ rotulada em 7 emoções (raiva, desprezo, desgosto, medo, felicidade, tristeza e surpresa), RML e eNTERFACE'05 que são rotuladas em 7 emoções (raiva, nojo, medo, tristeza, surpresa, felicidade e neutra). O autor propõe um método para classificação de emoções em ambiente único-rótulo que supere os métodos existentes. Após os experimentos seu método demonstra melhorias quanto aos demais, alcançando taxa de predição de 98,50% para a base ORL, 99,50% para a base Yale, 96,11% para a base CK+, 86,67% para a base eNTERFACE'05 e 90,83% para a base RML. Conseguiu mostrar também a importância da fusão de informações visuais e sonoras no processo de classificação de emoções em arquivos multimídia.

Assim como Seng, (NOROOZI, 2017) utiliza a combinação de áudio com imagens da Face dos ouvintes, onde realiza uma fusão tardia das predições obtidas através das características extraídas do áudio (Pitch, Intensidade, duração, distribuição de energia espectral, MFCC, ZCD e parâmetros de energia do banco de filtros) com as predições obtidas através da geometria facial do espectador e com a predição obtida através de uma rede neural em conjunto com o SVM. Após a fusão das predições o mesmo obtém as novas predições para as 7 emoções utilizadas (raiva, nojo, medo, tristeza, surpresa, felicidade e neutra). O autor trabalha no ambiente de único-rótulo, onde cada emoção é classificada de forma binária, e é considerado se há ou não tal emoção. Ele utiliza para seu trabalho as bases SAVEE, RML e eNTERFACE'05, as quais compara os resultados dos seus métodos com: Método Bayesiano Dinâmico (MBD), Método Gaussiano com PCA (MG+PCA), Rede neural ARTMAP e Fonema Específico (FE), a fim de provar a eficácia dos métodos propostos através da taxa de classificação. Como resultado ele obteve uma melhora comparado aos métodos tradicionais com taxas de 99,72% de acerto para a base SAVEE, 98,73% de acerto para a base eNTERFACE'05 e 100% para a base RML.

3.3 Utilizando o áudio e EEG

(HATAMIKIA, 2014) por outro lado classifica emoções em vídeos musicais em ambiente de único-rótulo baseado em características coletadas por EEG e por características extraídas dos arquivos de áudio que foram apresentados aos espectadores. As características utilizadas são fornecidas pela base rotulada DEAP, a qual trechos de música são apresentados aos espectadores e os dados do EEG são coletados e armaze-

dados juntamente com a rotulação recebida pelo ouvinte, os quais classificaram com notas de 1 a 5 para valência e excitação, o DEAP também fornece os arquivos de áudio que foram apresentados aos espectadores. O Autor realiza uma classificação baseado nas características extraídas, onde o mesmo realiza a classificação baseado apenas em uma característica de cada vez e por último une todas para realizar uma classificação geral. Ele faz a comparação entre as características de: Entropia Espectral, Entropia aproximada, Dimensão Fractal Petrosiana e Dimensão Fractal Katz's. As emoções foram divididas em 4 sub-emoções, representando Alta-Valência-Alta-Excitação, Alta-Valência-Baixa-Excitação, Baixa-Valência-Alta-Excitação e Baixa-Valência-Baixa-Excitação e a classificação se deu pelo algoritmo SOM (self-organization map) e avaliados através das métricas de Precisão e F1-score. Os resultados mostraram que apesar do uso de todas as características receberam nota superior ao uso delas isoladas, a taxa de melhoria ficou entre 5% e 9%.

3.4 Utilizando o áudio e o vídeo

(WANG, 2013) realiza a classificação de vídeos utilizando características de áudio e vídeo extraídas do próprio arquivo multimídia os combinando através de uma fusão precoce de bases. Para o projeto ele utilizou de 72 vídeos coletados na internet, os quais foram classificados por emoção de forma multi-rótulo por 50 estudantes, os quais rotularam os vídeos em feliz, bravo, triste, medo, desgosto e surpreso. Para a classificação, o mesmo transformou a base em uma base de único-rótulo, onde cada rótulo foi separado dos demais e classificado de forma binária (possui ou não determinado rótulo) através do SVM, após realizar as predições foram utilizados conceitos de probabilidade condicional e uma rede bayesiana para se obter uma classificação multi-rótulo da base. Como resultados principais ele obteve uma taxa de classificação de 63,3% com uma média de falso-positivo de 34,2% e mostrou também como é possível transformar um problema de único-rótulo em um problema multi-rótulo.

(WANG, 2015) por sua vez propõe um método para classificação multi-rótulo utilizando de quatro bases rotuladas por emoções. As bases escolhidas foram: Memorability (que é uma base de imagens rotulada em 12 emoções: assustador, excitante, engraçado, envolvente, pacífico, excitante, interessante, misterioso, estranho, impressionante, feliz e triste), Music Emotion (que é uma base de áudio rotulada em 6 emoções: espanto, feliz, relaxante, quieto, triste e bravo), NVIE vídeo (base com trechos de vídeos rotulados em 6 emoções: felicidade, raiva, tristeza, medo, nojo, surpresa) e FilmStim (base com trechos de filmes rotulados em 6 emoções: raiva, nojo, tristeza, medo, diversão e ternura). No caso da Memorability e Music Emotion, as características foram disponibilizadas junto com a base, e não houve necessidade de extrair mais características, já a NVIE e a FilmStim não disponibilizam as características, sendo assim foram extraídas características de áudio e vídeo para realizar a classificação. Para efetuar as classificações cada base foi separada em

duas sub-bases, uma de treino e outra de teste, onde para cada sub-base foram efetuadas as extrações de características e criados os modelos para predição. Através da base de treino foi gerado um modelo para classificação e aplicado na base de teste para realizar as classificações. Como forma de avaliação do modelo criado, foram comparados os resultados do modelo com os classificadores multi-rótulo: Bayesian Network (BN), Multi-Label Back-Programation (BPMLL) e Multi-Label k -nearest neighbor (ML k NN), onde para cada base foram aplicados todos os classificadores e avaliados através das métricas de precisão, recall e F1. Para as bases Memorability e Music Emotion o método proposto apresentou melhores resultados comparados com os classificadores multi-rótulo acima listados, porém para a base NVIE vídeo o mesmo apresentou avaliação inferior ao ML k NN em algumas métricas e para a base FilmStim o mesmo apresentou avaliação inferior ao BPMLL para algumas métricas. Apesar de não ser unanime a melhoria na classificação, o método proposto mostrou ser eficiente para classificar emoções em bases multi-rótulo.

(CHEN, 2012) propôs uma classificação de vídeos coletados da Televisão Musical (Musical Television - MTV) em seis emoções (feliz, triste, bravo, excitado, nervoso e calmo). Cada vídeo coletado foi dividido em trechos de 16 segundos onde cada trecho recebeu uma rotulação, após isso utilizou características provindas do áudio e as combinou através de uma fusão precoce com as características provindas do vídeo. Antes da combinação de características foi realizada uma seleção de características a fim de maximizar as predições e diminuir o tempo de execução do classificador. Foi utilizado um método de classificação hierárquico em ambiente de único-rótulo, onde inicialmente as emoções foram separadas em dois grupos quanto a sua excitação: alta excitação e baixa excitação, e classificados através do SVM com base nesses dois grupos. Para cada grupo destes, foram separados internamente em outros subgrupos, o grupo de baixa excitação foi dividido novamente em Triste e Calmo, onde foi aplicado uma classificação através do SVM para estas duas emoções. Já o grupo de alta excitação foi separado em dois subgrupos: Feliz_Excitado e Bravo_Nervoso, onde foi aplicado uma classificação através do SVM para estes dois. Dentro de cada um destes dois grupos foi aplicado novamente uma classificação através do SVM, para classificar em Feliz ou Excitado e Bravo ou Nervoso. Como características dos arquivos multimídia o autor utilizou o ZCR, Duração, Centroides espectral e propagação espectral, Tempo, desvio padrão de vetores de movimento e temperatura de cor. Como resultados obteve uma taxa de acerto de 73,3% com base na matriz de confusão utilizando o esquema hierárquico.

(THOMAS, 2017) utilizando a base MediaEval 2015 (que consiste em 10.900 trechos de vídeo extraídos de 199 filmes) e uma rede neural, classificou trechos de 8 a 12 segundos por emoção em ambiente de único-rótulo de acordo com valência (positiva, neutra ou negativa) e excitação (calmo, neutro ou ativo). Para seu experimento ele separou a base em treinamento e teste, contendo 6.144 e 4.175 vídeos respectivamente, dos quais foram extraídas características visuais através da rede pré-treinada VGG19 utilizando

o ImageNet, as características extraídas foram as contidas na camada fc7 do ImageNet. As características de áudio foram extraídas e classificadas de duas formas: na primeira através do MFCC, ZCR e Fluxo Espectral os quais foram classificados através do SVM e do k NN; na segunda foram utilizadas redes neurais (AlexNet, VGGnet e GoogleNet) utilizando o espectrograma do áudio em forma de imagem. Também foram combinadas as características de áudio e vídeo através da fusão tardia e da fusão precoce. Como resultado para a classificação baseada no áudio se obteve taxas de acurácia para a valência de: 27,63% para o SVM; 30,76% para o k NN; 37,62% para a rede neural; e 33,77% para a rede neural usando o SVM. Em relação a excitação se obteve: 21,36% para o SVM; 20,54% para o k NN; 45,02% para a rede neural; e 48,34% para a rede neural usando o SVM. Em relação as combinações de áudio e vídeo com base na valência, se obteve: 44,64% para a rede neural utilizando apenas o vídeo; 37,62% para a rede neural utilizando apenas o áudio; 43,52% para a rede neural utilizando a fusão tardia; e 44,15% para a rede neural utilizando a fusão precoce. Para a combinação com base na excitação se obteve: 48,20% para a rede neural utilizando apenas o vídeo; 45,02% para a rede neural utilizando apenas o áudio; 53,61% para a rede neural utilizando a fusão tardia; e 54,45% para a rede neural utilizando a fusão precoce.

Neste projeto utilizamos as características extraídas a partir do áudio e do próprio vídeo musical, ao contrário de (CHEN, 2012), (THOMAS, 2017) e (HATAMIKIA, 2014) utilizamos o ambiente multi-rótulo para classificar vídeos musicais completos. Como diferencial ao trabalho de (WANG, 2015) e (WANG, 2013), foi utilizado apenas vídeos musicais, e não dados provindos de imagens ou filmes. Realizamos também uma combinação de predições através da fusão tardia, a fim de determinar a eficácia da utilização de características visuais no processo de classificação multi-rótulo utilizando características visuais e auditivas dos vídeos musicais. Como procedimento de extração de características de áudio, foram utilizadas todas as características disponíveis na biblioteca Yaafe. As características visuais de Histograma, Diferença Acumulada e Fluxo Óptico foram extraídas através da biblioteca OpenCV em C++ e as características visuais provindas da camada fc7 do VGG16, obtidas através da rede neural pré-treinada ImageNet foram coletadas através da biblioteca Keras em Python. Os procedimentos de classificação foram realizados em ambiente multi-rótulo através da biblioteca multi-rótulo Mulan. Como base foi utilizada uma extensão da base de dados CAL500, a qual possui rotulação de 18 emoções.

3.5 Comparativos

Como forma de comparação entre os trabalhos acima listados, a tabela 5 mostra a diferença entre os classificadores e ambientes utilizados em cada trabalho, já a tabela 6 mostra as emoções utilizadas e a fonte dos dados utilizados, os quais podem ser oriundos do Áudio (A), imagens faciais (F), coletados através de EEG (E) ou coletados a partir do

próprio arquivo de vídeo (V).

Trabalho	Classificadores	Ambiente
(TROHIDIS, 2008)	BR(SVM), LP(SVM), RakEL(SVM) e ML k NN	Multi-rótulo
(CHEN, 2012)	SVM	Único-rótulo
(WANG, 2013)	SVM+BN	Multi-rótulo
(HATAMIKIA, 2014)	SOM	Único-rótulo
(WANG, 2015)	BN, BPMLL, ML k NN e TRBM	Multi-rótulo
(SENG, 2016)	(BDPCA+LSLDA+OKL-RBF), LRC, 2DNPP, 2DLDA	Único-rótulo
(NOROOZI, 2017)	MBD, MG+PCA, ARTMAP, FE, CNN,	Único-rótulo
(THOMAS, 2017)	CNN+SVM, SVM, k NN	Único-rótulo

Tabela 5 – Comparativo entre classificadores utilizados.

Trabalho	Emoções	Fonte dos dados			
		A	F	E	V
(TROHIDIS, 2008)	Espanto/Surpresa, Feliz/Satisfeito, Quietos, Relaxante/Calmo, Triste, Zangado/Medo	X			
(CHEN, 2012)	Feliz, Triste, Bravo, Excitado, Nervoso e Calmo	X			X
(WANG, 2013)	Feliz, Bravo, Triste, Medo, Desgosto e Surpreso	X			X
(HATAMIKIA, 2014)	Valência e Excitação	X		X	
(WANG, 2015)	Memorability: Assustador, Excitante, Engraçado, Envolvente, Pacífico, Excitante, Interessante, Feliz, Estranho, Impressionante, Misterioso e Triste Music Emotion: Espanto, Feliz, Relaxante, Quietos, Triste e Bravo NVIE vídeo: Felicidade, Raiva, Tristeza, Medo, Nojo, Surpresa FilmStim: Raiva, Nojo, Tristeza, Medo, Diversão e Ternura	X			X
(SENG, 2016)	ORL e Yale: Bravo, Feliz, Triste, Desgosto, Surpreso e Medo CK+: Raiva, Desprezo, Desgosto, Medo, Felicidade, Tristeza e Surpresa RML e eINTERFACE'05: Raiva, Nojo, Medo, Tristeza, Surpresa, Felicidade e Neutra	X	X		
(NOROOZI, 2017)	Raiva, Nojo, Medo, Tristeza, Surpresa, Felicidade e Neutra	X	X		
(THOMAS, 2017)	Valência e Excitação	X			X

Tabela 6 – Comparativo entre emoções e fontes dos dados utilizados.

4 Metodologia de Pesquisa

O processo de classificação de Vídeos Musicais por emoção em ambiente multi-rótulo foi dividido em quatro etapas: preparo da base de dados, extração de características e geração de arquivos, classificação e fusão de bases. Cada etapa é de suma importância para o resultado final do projeto, sendo assim cada etapa foi tratada de forma rigorosa.

4.1 Preparo da base de dados

Neste projeto foi utilizada a base de dados CAL500 descrita em (TURNBULL, 2007), que é um banco de dados de música catalogado bem conhecido, possuindo os arquivos de áudio e os rótulos para cada arquivo. No CAL500 o banco de dados foi classificado por pelo menos 3 ouvintes humanos em Emoção, Gênero, Instrumentos, Som, Uso e Vocais. O banco de dados apresenta 502 músicas populares ocidentais compostas entre 1950 e 2007 por 500 artistas diferentes, escolhidos para cobrir uma grande quantidade de variação acústica enquanto ainda representam alguns gêneros familiares e artistas populares. A principal característica da base para a escolha no projeto é que a mesma utiliza a música completa e a rotulação é realizada por humanos. Algumas bases utilizam apenas trechos de áudio ou vídeo, podendo assim gerar uma classificação não dominante para o vídeo todo. Outras bases contendo áudio e vídeo não são provindas de vídeos musicais (como a FilmStim e a NVIE) ou são provindas apenas de trechos musicais e sem vídeo (como a Music Emotion). (KOELSTRA, 2012) possuem ainda bases das quais a catalogação foi gerada a partir de sensores colocados em espectadores de áudios e vídeos. Neste projeto foi utilizado apenas as informações de emoção da base, as quais são separadas em 18 rótulos de emoção: *Angry_Agressive*, *Arousing_Awakening*, *Bizarre_Weird*, *Calming_Soothing*, *Carefree_Lighthearted*, *Cheerful_Festive*, *Emotional_Passionate*, *Exciting_Thrilling*, *Happy*, *Sad*, *Laidback_Mellow*, *Light_Playfull*, *Loving_Romantic*, *Pleasant_Comfortable*, *Positive_Optimistic*, *Powerful_Strong*, *Tender_Soft* e *Touching_Loving* (que são todos os rótulos de emoção da base). Os quais foram avaliadas em uma escala de 1 a 5 por cada ouvinte humano (sendo 1 pouca presença da emoção e 5 muita presença da emoção) com base em suas emoções, gerando assim uma base desbalanceada.

Como a CAL500 é uma base musical e não visual, foi realizada uma extensão procurando vídeos musicais de cada música apresentada na base. Durante a coleta dos vídeos musicais como regra principal foram utilizados de apenas vídeos musicais oficiais, pois são criados para transmitir as mesmas emoções das músicas originais, garantindo assim que o vídeo musical expresse a mesma emoção que a música e para garantir que o vídeo é

realmente oficial foram utilizadas fontes distintas para obtenção dos mesmos. O VEVO¹ e o MTV² foram escolhidos como as principais fontes de conteúdo por não permitir que os usuários façam upload de arquivos e o conteúdo de ambos os sites vem dos produtores responsáveis pelos vídeos musicais e pelo próprio site. Também foram selecionados apenas vídeos musicais que foram criados e editados para a música, descartando assim vídeos oficiais de apresentações e shows. Após o processo de coleta de vídeos um banco de dados com 140 vídeos musicais foi gerado, a qual chamamos de CAL500 Estendida. Muitos vídeos (164) foram descartados por serem provindos de shows ao vivo ou apresentações, por só estarem disponíveis em fontes de acesso público para upload (74) como Youtube, Vimeo entre outros, ou por não possuírem vídeos musicais (122), já que o banco de dados possui músicas antigas muitos deles não possuem vídeos musicais, e por possuírem o mesmo artista de outro elemento da base (2). A figura 15 apresenta o diagrama explicando como ocorreu a geração da CALL500 Estendida.

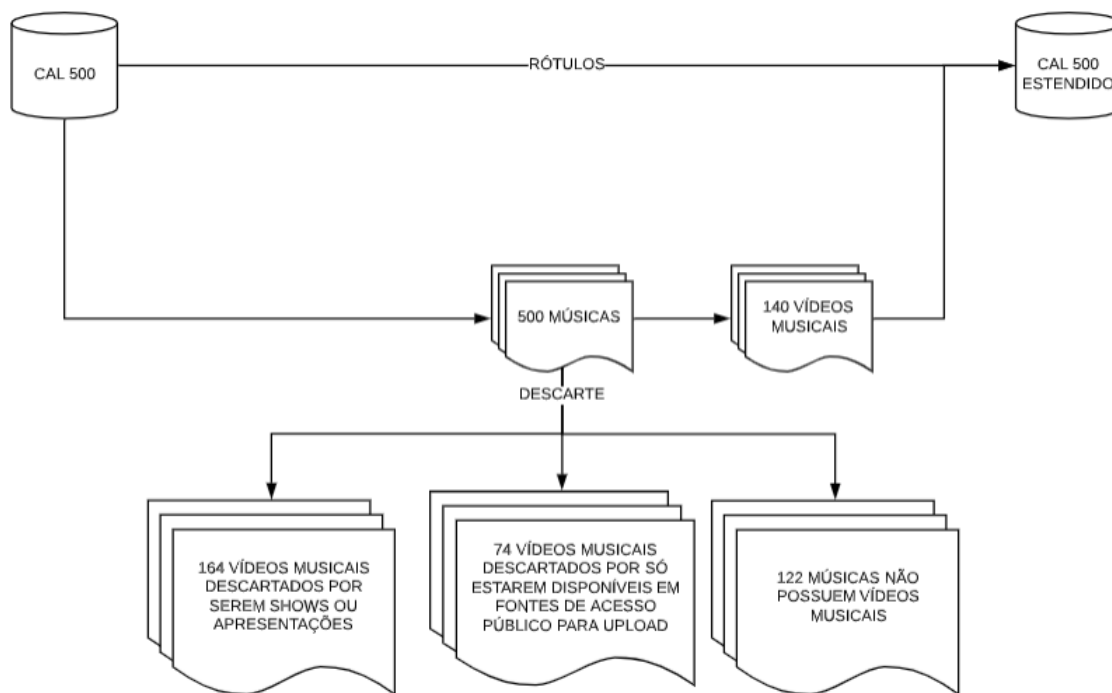


Figura 15 – Modelo exemplificando a criação da base CAL500 Estendido.

Fonte: Autoria própria

Utilizando a base CAL500 Estendida foi realizado um procedimento para identificar as emoções com maior presença em cada vídeo musical, onde com base nas avaliações provindas da CAL500 original foi realizada uma média aritmética entre avaliações da

¹ VEVO. *VEVO*. Disponível em: <<https://www.vevo.com/>>. Acesso em: 01 jul. 2019.

² MTV. *MTV database*. Disponível em: <<http://www.mtv.co.uk/database/videos>>. Acesso em: 01 jun. 2019.

mesma música (cada música possui no mínimo 3 avaliações de usuários distintos), com base nessa média foi estabelecido um limiar (três) que foi utilizado para definir se uma determinada música possui ou não uma emoção. Se a média aritmética entre as avaliações de uma música para uma determinada emoção λ forem iguais ou superiores que o limiar ($\bar{\lambda} \geq 3.0$) foi considerado que essa música apresenta a emoção λ , caso contrário a música não apresenta a emoção. Ficando assim a base CAL500 Estendida com 140 vídeos musicais avaliados de forma binária (0 e 1) para as 18 emoções originais da base CAL500.

Para garantir que durante os testes a base CAL500 Estendida fosse utilizada da forma mais balanceada possível, foi realizado o procedimento de estratificação em ambiente multi-rótulo. O procedimento de estratificação da base foi realizado através da biblioteca Mulan utilizando a função "*stratify*", onde os 140 vídeos da base CAL500 Estendida (Apêndice (B)) foram separados com base em seus rótulos em 10 folds contendo 14 vídeos cada, os quais são apresentados na tabela 7. O procedimento de estratificação garante que nenhum fold fique com desfalque de rótulos e que a divisão fique da forma mais balanceada possível, balanceamento o qual é mostrado na tabela 8. A saída do procedimento de estratificação foram 10 arquivos (um por fold) contendo o Id dos vídeos e seus respectivos rótulos, arquivos os quais foram utilizados como base para a geração de todos os arquivos ARFF.

Fold1	Fold2	Fold3	Fold4	Fold5	Fold6	Fold7	Fold8	Fold9	Fold10
7	15	2	14	19	20	47	6	4	13
57	22	5	56	44	63	65	9	17	43
87	73	39	60	79	117	76	10	85	108
107	160	53	104	132	138	97	31	86	122
128	272	55	143	149	141	126	84	99	175
154	295	67	194	177	163	133	144	205	190
202	329	224	196	193	164	137	168	235	206
209	331	285	234	252	214	178	170	271	257
255	335	326	306	254	232	269	197	307	287
380	385	356	336	303	296	286	228	332	361
412	415	359	339	308	368	294	265	382	431
414	417	402	389	311	388	343	302	419	433
424	461	407	404	439	392	370	304	441	474
465	484	488	445	496	452	403	353	471	482

Tabela 7 – Separação real dos vídeos (representados pelos seus Ids) por fold de forma estratificada.

Emoções	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Total
λ_1	2	2	7	1	1	4	2	4	2	3	28
λ_2	9	12	7	10	11	10	11	11	9	6	96
λ_3	3	5	5	2	1	1	2	3	2	3	27
λ_4	5	8	7	7	8	7	6	4	6	8	66
λ_5	7	11	2	10	9	8	7	8	1	6	69
λ_6	6	10	3	9	9	9	10	7	4	3	70
λ_7	14	8	9	12	11	12	11	6	10	10	103
λ_8	8	11	8	7	8	9	9	7	7	5	79
λ_9	8	12	3	8	10	11	10	9	6	9	86
λ_{10}	5	9	5	7	6	7	5	5	2	5	56
λ_{11}	7	8	2	10	7	8	8	7	2	6	65
λ_{12}	7	4	7	5	6	7	5	2	5	5	53
λ_{13}	8	13	8	11	13	11	11	12	8	12	107
λ_{14}	7	8	3	9	10	12	9	10	6	7	81
λ_{15}	13	10	10	12	13	11	11	8	12	9	109
λ_{16}	5	1	6	4	3	4	4	5	4	6	42
λ_{17}	5	5	5	7	6	6	5	4	4	5	52
λ_{18}	8	5	6	5	7	6	2	4	5	5	53

Tabela 8 – Distribuição dos rótulos por fold.

4.2 Extração de características e geração de arquivos

O procedimento de extração de características e geração de arquivos foi dividido em três partes, a primeira está relacionada a extração de características auditivas o qual foi realizado através da biblioteca Yaafe em ambiente Linux, a segunda se refere a extração de características visuais, as quais foram realizadas em linguagens e ambientes variados, a terceira e última se refere a geração de todos os arquivos ARFF utilizando as características extraídas juntamente com as informações providas do procedimento de preparo da base (subseção 4.1).

4.2.1 Extração de Características Auditivas

As características auditivas foram extraídas em uma máquina virtual Ubuntu Linux 14.04.05 através da biblioteca Yaafe em linguagem Python 2.7. Os arquivos de áudio foram extraídos dos vídeos musicais e padronizados para o formato ".wav" com taxa de amostragem 44.100. As características extraídas foram: Envelope, Modulação de Amplitude, Frames de Áudio, Correlação Automática, Domínio Complexo na Detecção de Início, Energia, Espectro de Magnitude, Estatísticas de Forma Espectral, Estatísticas de Forma Envelope, Sonoridade, Espectro Mel, MFCC, Redução Espectral, Nivelamento Espectral, Nivelamento Espectral por Banda, Fluxo Espectral, Deslocamento Espectral, Encosta Espectral, Variação Espectral, Estatística de Forma Temporal, ZCR, OBSI, OBSIR,

Nitidez Perceptual, Propagação Perceptiva e Fator de Crista Espectral por Banda, os quais são descritos na seção 2.1. Como o foco principal do trabalho é mostrar o impacto no uso de informações visuais no processo de classificação de vídeos musicais, as características de áudio não sofreram nenhum tipo de refinamento ou seleção. A figura 16 apresenta o processo realizado para extração de características auditivas onde com base nos 140 vídeos musicais da base é realizada a extração do arquivo de áudio de cada vídeo musical e os arquivos de áudio servem como entrada para o procedimento e extração de características auditivas.

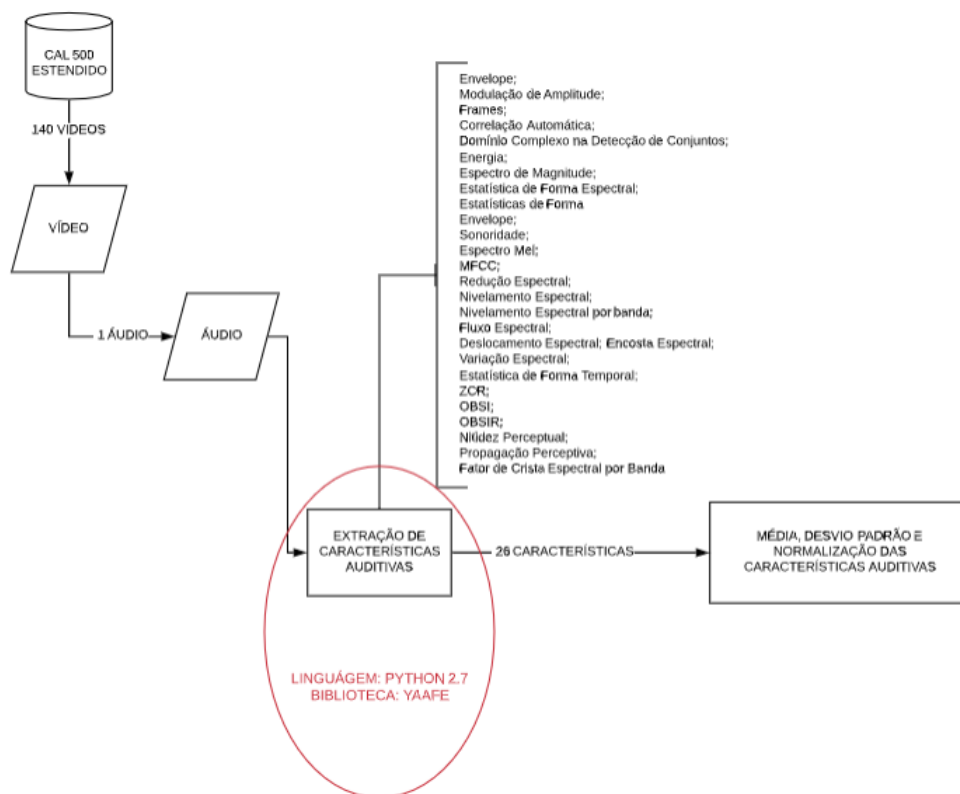


Figura 16 – Modelo exemplificando o processo de extração de características auditivas.

Fonte: Autoria própria

Para cada arquivo de áudio foram executados todos os 26 procedimentos acima listados, onde foram gerados 26 arquivos contendo em cada linha os Ids dos arquivos de áudio e os valores obtidos através da característica que representa o arquivo, gerando assim um arquivo contendo 140 linhas e valores separados por ',' (ponto e vírgula). Como as características auditivas muitas vezes dependem do tamanho do áudio (tempo em segundos) não há um padrão de quantidade de características extraídas por cada procedimento, por isso foi padronizado que o primeiro valor de cada linha do arquivo representa o Id do arquivo de áudio e os demais representam as características extraídas. Os 26 arquivos gerados foram salvos para posteriormente serem utilizados no procedimento de geração de arquivos (subseção 4.2.3).

4.2.2 Extração de Características Visuais

Antes de realizar os procedimentos de extração de características visuais alguns procedimentos foram realizados a fim de evitar perda de informações e padronizar as informações. Primeiramente os arquivos de vídeos musicais foram convertidos para formato mp4 e sua altura e largura permaneceram as mesmas dos vídeos musicais originais evitando assim a perda de informações presentes nos mesmos. Outro cuidado que se teve foi em relação a obtenção de imagens dos vídeos (também chamado de frames de vídeo) onde foram extraídos exatamente 220 imagens por vídeo musical, as quais foram armazenadas e nomeadas de forma a identificar o vídeo musical de origem e seu frame no andamento do vídeo (ex: "002_Imagem_Frame_2040", onde o número "002" representa o Id do vídeo musical e o "2040" representa a imagem extraída). O valor 220 foi definido com base na média total de imagens dividido pelo número de imagens por segundo (apresentado na equação 4.1), para que as amostras tenham a mesma proporção de dados, ou seja, que cada vídeo tenha o mesmo peso na base. As imagens extraídas foram utilizadas para realizar todos os procedimentos de extração de características, onde em alguns deles foi utilizado também os demais frames do vídeo musical para coleta de mais informações. As características visuais extraídas foram: Histograma de Cores, Diferença Acumulada, Fluxo Óptico, VGG16.fc7+ImageNet e LBP, as quais são descritas na seção 2.2.

$$\frac{\left(\frac{TotalImagens}{TotalVideos}\right)}{ImagensPorSegundo} = \frac{\left(\frac{924.000}{140}\right)}{30} = \frac{6.600}{30} = 220 \quad (4.1)$$

O primeiro tipo de característica visual extraída foi obtido através do procedimento de Histograma de Cores, realizada em ambiente Windows 7 através da biblioteca OpenCV 3.2 em linguagem C++. Para o procedimento foram utilizadas as 220 imagens de cada um dos 140 vídeos musicais (totalizando 30.800 imagens), onde para cada imagem foi analisado o valor de cada pixel para cada espaço de cor (RGB), onde a quantidade de pixels varia de vídeo para vídeo, pois cada vídeo possui dimensões (altura e largura) distintas. Para cada imagem foi gerado um vetor com 768 posições (256 posições para cada cor do espaço RGB) contendo os valores de ocorrência de cada nível de cor para cada cor, vetor o qual foi armazenado em um arquivo juntamente com o nome da imagem do qual foi extraído, possibilitando assim saber de qual exatamente de qual imagem aquelas características foram extraídas. O arquivo gerado contém o nome da imagem seguido pelas 768 características separados por ';' para cada uma das 30.800 imagens da base.

O segundo tipo de característica visual foi obtido através do procedimento de Diferença Acumulada, o qual foi realizada assim como o Histograma de Cores através da biblioteca OpenCV 3.2. Para o procedimento foram utilizadas além das 220 imagens de cada vídeo musical, o próprio vídeo musical, do qual foram extraídas todas as imagens durante o procedimento. Como as imagens de um mesmo vídeo possuem as mesmas dimensões,

foram gerados três vetores contendo altura*largura posições, cada um representando uma cor distinta no espaço RGB, os vetores foram utilizados como acumuladores de diferença, onde após realizar o procedimento de diferença entre duas imagens sequenciais, os valores de cada pixel para cada cor da saída dessa diferença foram acumulados (somados) até chegar em uma "imagem chave" (imagem pertencente as 220 extraídas de cada vídeo musical). Após foi realizado o procedimento utilizando a imagem chave no qual calcula uma média aritmética em cada posição dos vetores com base na quantidade de cálculos de diferença realizados, posteriormente é realizado um procedimento de Histograma em cada um dos vetores, gerando assim três vetores com 256 posições. Esses vetores são salvos sequencialmente em um arquivo juntamente com o nome da imagem chave utilizada, gerando assim um arquivo semelhante ao gerado através do procedimento de Histograma de Cores descrito anteriormente contendo um identificador da imagem e 768 características para cada uma das 30.800 imagens da base.

O Fluxo Óptico foi o terceiro procedimento de extração de características a ser realizado, o qual assim como o Histograma de Cores e a Diferença Acumulada ocorreu através da biblioteca OpenCV 3.2 em ambiente Windows 7 e linguagem C++. Para o procedimento do Fluxo Óptico foram utilizados além das 220 imagens da base as imagens adjacentes a elas (anterior e posterior), as quais foram utilizadas para determinar a movimentação do pixel ao decorrer do vídeo. Após determinada a direção do pixel (obtida através do vetor traçado entre o ponto de origem e o ponto de destino do pixel) foram levantadas nove características com base em sua direção e velocidade (calculada com base no tamanho vetorial do vetor traçado) que são apresentadas na figura 17. As características de direção se baseiam no ângulo do vetor tendo como ponto de origem a posição de origem do pixel e foram normalizadas em oito ângulos distintos: 0° ou 360° , 45° , 90° , 135° , 180° , 225° , 270° e 315° . Com base em cada ângulo foi realizado um histograma para cada sequência de imagem, onde um vetor com nove posições foi gerado e utilizado como acumulador, onde cada uma das oito primeiras posições representam a quantidade de pixels que se moveram para cara ângulo e a última posição recebe uma média de velocidades dos pixels da imagem, representando assim a intensidade de movimento da imagem. O vetor gerado foi salvo em um arquivo contendo o nome da imagem chave (imagem pertencente as 220 imagens da base e conseqüentemente a imagem central do procedimento) e as nove características extraídas, ficando o arquivo assim com 30.800 linhas representando cada uma das 220 imagens para cada vídeo musical.

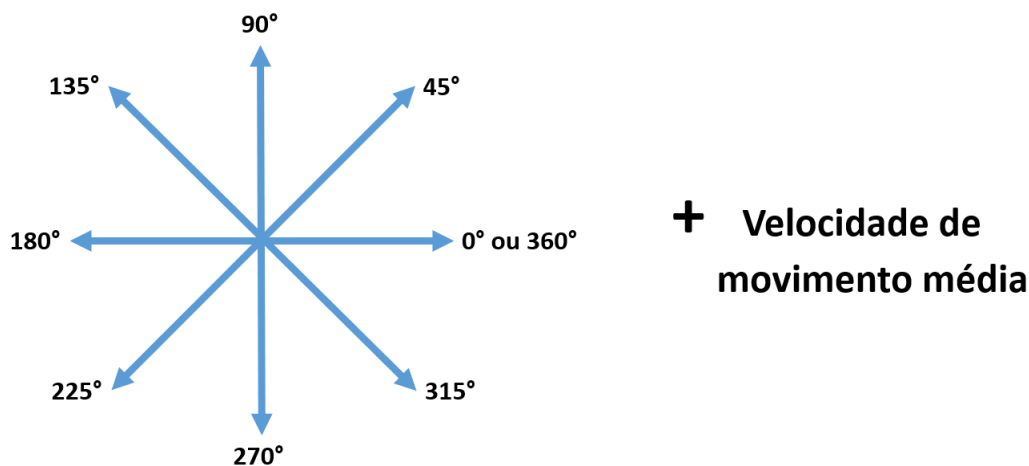


Figura 17 – Modelo exemplificando a extração de características de Fluxo Óptico.

Fonte: Autoria própria

Outro procedimento para extração de características realizado é o que chamamos de VGG16.fc7+Imagenet que foi realizado em uma máquina virtual Ubuntu Linux 14.04.05 em linguagem Python 2.7 através da combinação das bibliotecas TensorFlow e Keras. Para o procedimento foram utilizadas apenas as 220 imagens selecionadas para cada vídeo, onde o próprio algoritmo redimensionou todas as imagens independentes de seus tamanhos para 224x224. Ao rodar o algoritmo com uma imagem de entrada é realizado vários procedimentos convolucionais descritos na subseção 2.2.4 e ao chegar na camada fc7 um vetor com 4.096 características é gerado, este vetor juntamente com o nome da imagem são salvos em um arquivo, o qual possui o nome da imagem e 4.096 características para cada uma das 30.800 imagens selecionadas.

O último procedimento de extração de características realizado é o LBP, o qual foi realizado em ambiente Windows 7 através da linguagem Matlab. Para o procedimento assim como o VGG16.fc7+Imagenet foram utilizadas apenas as 220 imagens de cada vídeo musical, as quais foram transformadas em imagens em nível de cinza (imagens preto e branco) através da função *rgb2gray* e em seguida aplicado o procedimento de LBP, o qual retorna um vetor de histograma normalizado seguindo o padrão uniforme contendo 59 posições. Foi gerado um arquivo contendo o nome das imagens juntamente com as 59 características normalizadas extraídas da mesma, arquivo o qual possui tais informações para todas as 30.800 imagens utilizadas.

A figura 18 apresenta os procedimento de extração de todas as características visuais onde se tem como entrada os 220 frames de cada um dos 140 vídeos musicais, os quais são utilizados pelos cinco métodos de extração de características visuais.

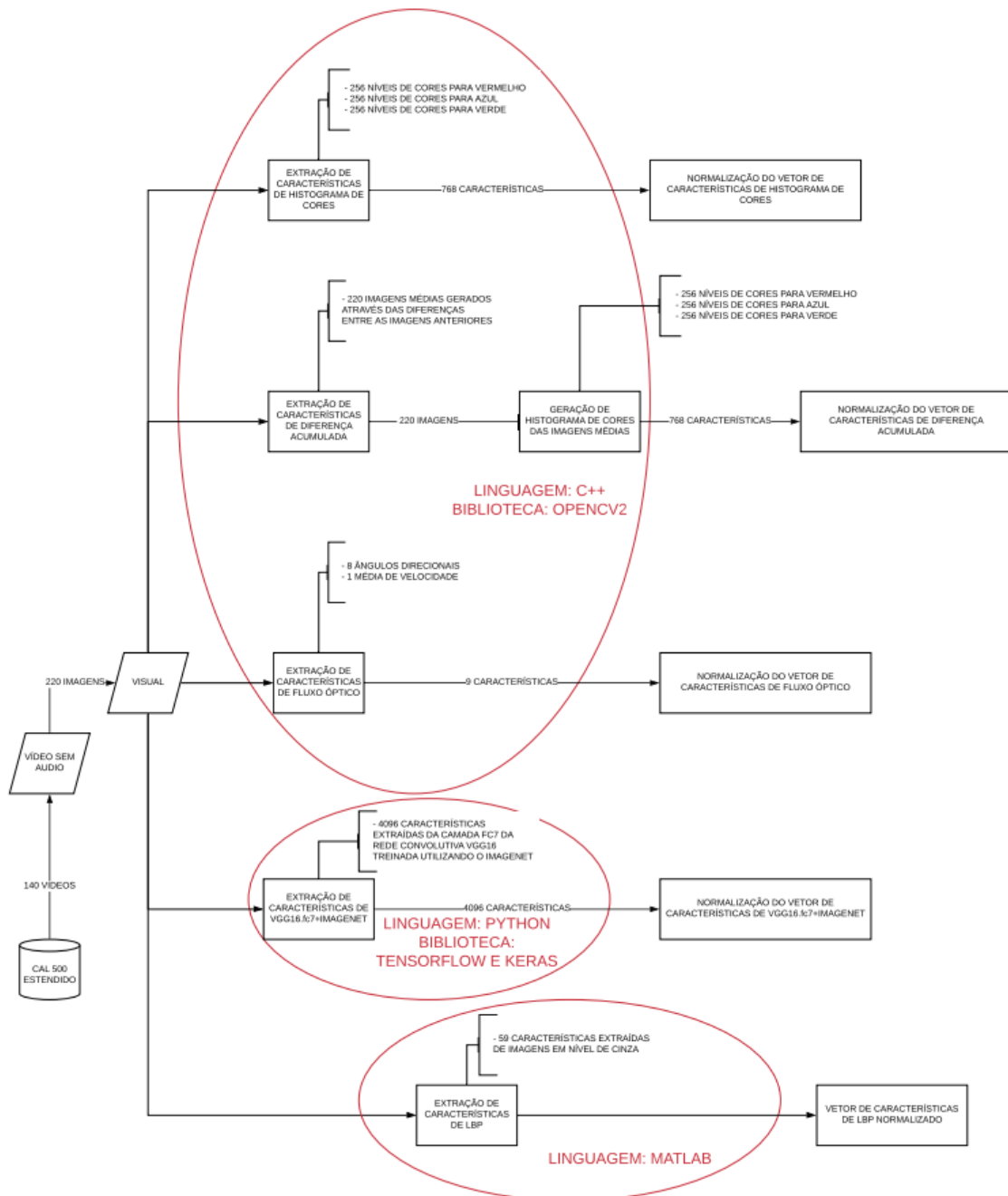


Figura 18 – Modelo exemplificando o processo de extração de características visuais.

Fonte: Autoria própria

4.2.3 Geração de Arquivos

A geração de arquivos foi realizada em ambiente Windows 7 em linguagem C++, onde se teve como entrada os arquivos contendo as características auditivas, características visuais e os dez arquivos de estratificação representando cada um dos folds. Os arquivos aqui gerados seguem o padrão ARFF Multi-rótulo (que podem ser exemplificados através da figura 13) o qual se difere de um arquivo ARFF comum através de seus rótulos, em um ARFF de único-rótulo os rótulos ou classes são definidos em uma única variável chamada

de Classe (@ATTRIBUTE Classe), a qual contém todas as possibilidades de ocorrência para aquela base (ex: @ATTRIBUTE Classe Gato, Cachorro, Periquito), já o arquivo multi-rótulo apresenta uma variável para cada rótulo da base, podendo esse atributo receber o valor de 0 ou 1 representando a ausência ou presença do rótulo respectivamente. Além do arquivo ARFF é necessário gerar um arquivo xml contendo apenas os rótulos da base. A geração de todos os arquivos multi-rótulo teve como base a estratificação apresentada na tabela 7, onde para cada fold foram gerados arquivos de teste e arquivos de treino de forma a garantir que instâncias de um mesmo vídeo musical não estejam no teste e no treino simultaneamente, os arquivos xml foram todos gerados de forma idêntica, devido ao fato de apresentarem apenas os nomes dos rótulos e os rótulos serem os mesmos para todo o experimento, já geração de arquivos ARFF se deu de forma única para cada tipo de procedimento de extração de características.

Para a geração dos arquivos ARFF referentes as características auditivas foi necessário inicialmente realizar alguns procedimentos em cima do arquivo com características extraídas, primeiramente foi realizado o cálculo da média e do desvio padrão para as características presentes em cada linha de cada um dos 26 arquivos que representam cada uma das 26 características auditivas. Com base nos Ids das músicas foram gerados para cada música um vetor de características com 52 posições (26 médias e 26 desvios padrões), o qual foi normalizado entre 0 e 1 com base nos valores máximo e mínimo. Após isso foram gerados os arquivos de teste e treino para cada fold apresentando sequencialmente o vetor com 52 características normalizadas com os 18 rótulos presentes nos arquivos de estratificação.

A geração dos arquivos ARFF referentes as características de Histograma de Cores e Diferença Acumulada ocorreram da mesma maneira, onde inicialmente foi realizado uma normalização dos vetores de características de 768 posições, normalizando os valores entre 0 e 1 com base nos valores máximo e mínimo, após isso foram gerados os arquivos ARFF de teste e treino para cada fold contendo sequencialmente o vetor com 768 características normalizadas com os 18 rótulos presentes nos arquivos de estratificação.

Os arquivos ARFF referentes as características de Fluxo Óptico e VGG16.fc7+ImageNet foram gerados utilizando da mesma forma que os arquivos de Histograma de Cores e Diferença Acumulada, realizando uma normalização entre os valores e sequencialmente gerando os arquivos de teste e treino para cada fold. Já os arquivos ARFF referentes as características de LBP foram gerados sem a realização de uma normalização, pois os mesmos já são normalizados pelo algoritmo do LBP.

Ao término da geração de arquivos seis tipos de arquivos foram gerados, sendo eles referentes as características de Áudio, Histograma de Cores, Diferença Acumulada, Fluxo Óptico, VGG16.fc7+Imagenet e LBP, onde cada tipo de arquivo foi dividida em 10 folds em arquivos de teste e treino. As informações sobre os tipos de arquivos são apresentados

na tabela 9.

Descritor do Arquivo	Rótulos	Atributos	Instâncias	Características totais
Áudio	18	52	140	7.280
Histograma de Cores	18	768	30.800	23.654.400
Diferença Acumulada	18	768	30.800	23.654.400
Fluxo Óptico	18	9	30.800	277.200
VGG16.fc7+ImageNet	18	4.096	30.800	126.156.800
LBP	18	59	30.800	1.817.200

Tabela 9 – Informações sobre arquivos gerados.

4.3 Classificação

Neste trabalho foi utilizado apenas a classificação multi-rótulo devido ao fato de emoções em músicas e vídeos quase nunca estarem sozinhas, normalmente as emoções são apresentadas em conjuntos e para os vídeos musicais ocorre o mesmo, podendo assim um vídeo musical apresentar emoções como: feliz e romântico; triste e calmo; triste, calmo e bizarro e assim por diante. O processo de classificação pode ser dividido em duas etapas distintas, a primeira etapa é a classificação multi-rótulo através dos classificadores: *Binary Relevance* (BR), *BRkNN*, *Calibrated Label Ranking* (CLR), *HOMER*, *Label Powerset* (LP), *MLkNN* e *RakEL* (descritos na seção 2.3) os quais tem como saída as predições para cada instância da base, esta etapa podemos chamar de Geração de Predições. A segunda etapa nomeada de Avaliação Métrica é obtida através da utilização das predições para avaliar de forma métrica os procedimentos de classificação multi-rótulo. A figura 19 apresenta o procedimento de geração de predições e Avaliação Métrica destas predições, onde a partir do uso dos 20 arquivos (10 arquivos de teste e 10 arquivos de treino) para cada tipo de característica extraída é realizada a geração de predições e avaliação métrica para cada classificador.

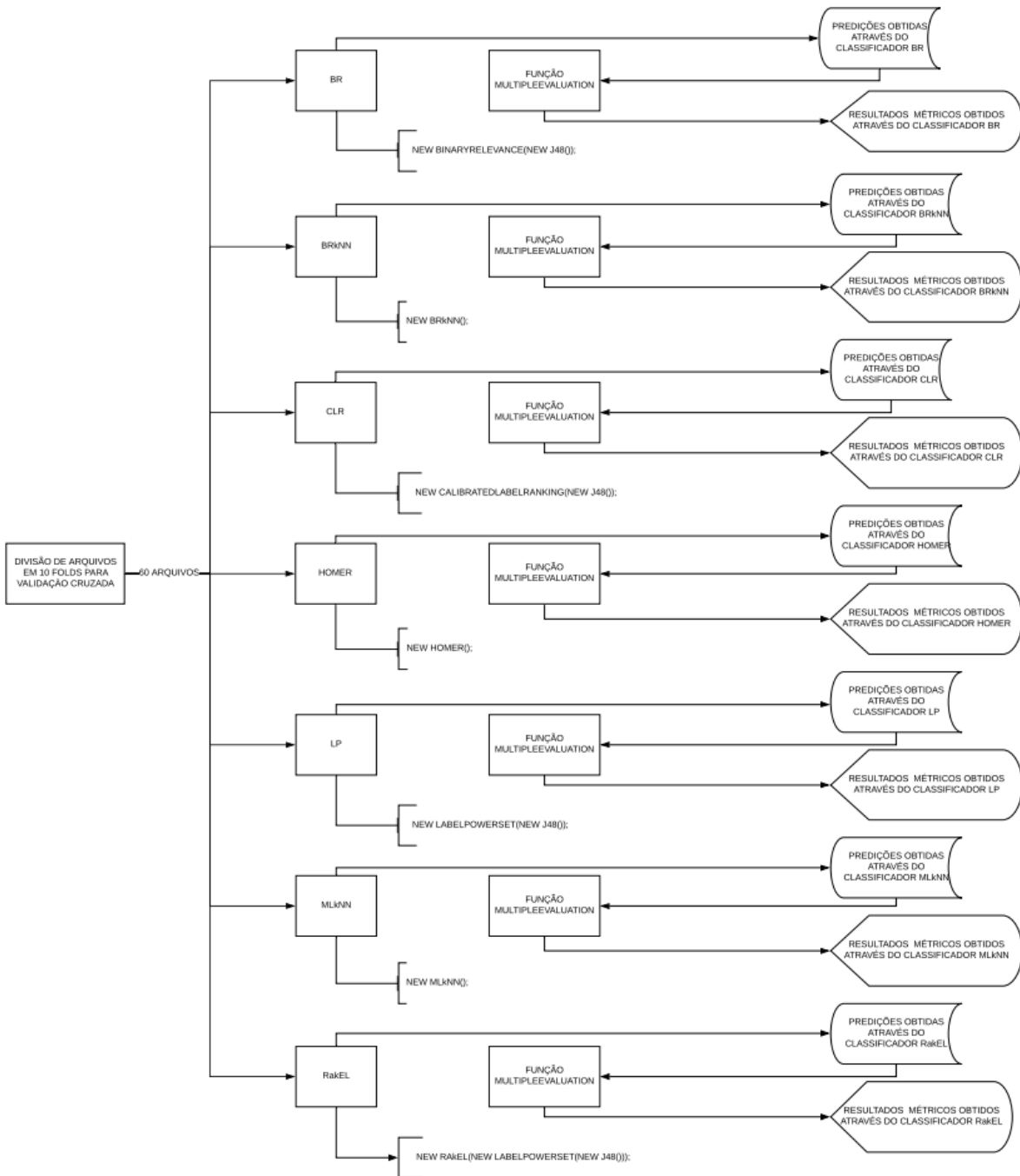


Figura 19 – Modelo exemplificando o processo de geração de predições auditivas e visuais e avaliação métrica das mesmas.

Fonte: Autoria própria

4.3.1 Geração de Predições

A geração das predições através dos classificadores multi-rótulo é realizada através do método de validação cruzada, onde foram utilizados os arquivos gerados no processo de geração de arquivos (descritos na subseção 4.2.3). Para cada uma das seis bases (Áudio, Histograma de Cores, Diferença Acumulada, Fluxo Óptico, VGG16.fc7+Imagenet e LBP)

os arquivos de teste e treino para cada um dos 10 folds foram utilizados pelos classificadores multi-rótulo, os quais realizaram as classificações de forma padronizada. Devido ao fato do classificador HOMER utilizar como atributo interno fixo o classificador de único-rótulo J48 (versão em java do classificador C4.5 que utiliza de árvore de decisão para realizar suas predições), todos os classificadores que necessitam como parâmetro algum tipo de classificador de único-rótulo utilizarão o J48. Devido ao tempo necessário para realizar as classificações e ao curto prazo do projeto, as classificações multi-rótulo se darão apenas através dos sete classificadores multi-rótulo anteriormente descritos e com o classificador de único-rótulo J48 como parâmetro quando necessário. Como padrão também podemos ressaltar o uso da validação cruzada entre os 10 folds, onde cada fold é testado com base em um treinamento realizado através do uso dos 9 outros folds.

Devido ao fato dos arquivos de cada fold serem gerados com base no procedimento de estratificação, as instâncias de cada fold não foram randomizadas novamente, facilitando assim os procedimentos de Avaliações Métricas e também os procedimentos de Fusão de Bases (subseção 4.4). Após a geração de predições para cada Fold, os resultados foram agrupados em um arquivo de saída contendo as predições (Confiança, Bipartição e Ranking) que são descritas na subseção 2.4.2.

4.3.2 Avaliação Métrica

A etapa de avaliação métrica é realizada através do uso da função *calculateStatistics()* que realiza o cálculo de todas as métricas multi-rótulo disponíveis na biblioteca Mulan com base nas predições obtidas através do procedimento de geração de predições (descrito na subseção 4.3.1) e das predições geradas através da fusão tardia (subseção 4.4). Como os dois procedimentos geram as predições no mesmo formato não foi necessário realizar adaptações para realizar o procedimento métrico. Os resultados apresentados neste projeto utilizam apenas a *Métrica Micro-averaged AUC* (Micro AUC) devido ao grande número de resultados provindos da fusão de bases, os resultados para outras métricas podem ser solicitados por e-mail.

O procedimento do cálculo do *Micro-averaged AUC* (*Micro-averaged Area Under a Curve*) descrito em (FAWCETT, 2006) e avaliado em (WU; ZHOU, 2017) é baseado no conceito de ROC (*Receiver operating characteristics*) onde inicialmente é realizado a plotagem de um gráfico bidimensional onde os eixos X e Y são as taxas de FP (*False Positive*) e TP (*True Positive*) respectivamente, descrevendo assim as compensações relativas entre os benefícios (TP) e custos (FP). O cálculo da taxa TP e da taxa FP são descritos através das equações 4.2 e 4.3 respectivamente

$$\text{taxa TP} \approx \frac{\text{Positivos corretamente classificados}}{\text{Total de positivos}} \quad (4.2)$$

$$\text{taxa FP} \approx \frac{\text{Negativos incorretamente classificados}}{\text{Total de negativos}} \quad (4.3)$$

A figura 20 apresenta a Curva ROC gerada a partir de uma classe e sua Área Abaixo da Curva (AUC) com base nos rótulos relevantes e irrelevante desta classe, esta área é que define o valor da métrica AUC. Com base no valor da métrica AUC de cada rótulo é calculado o Micro-averaged AUC.

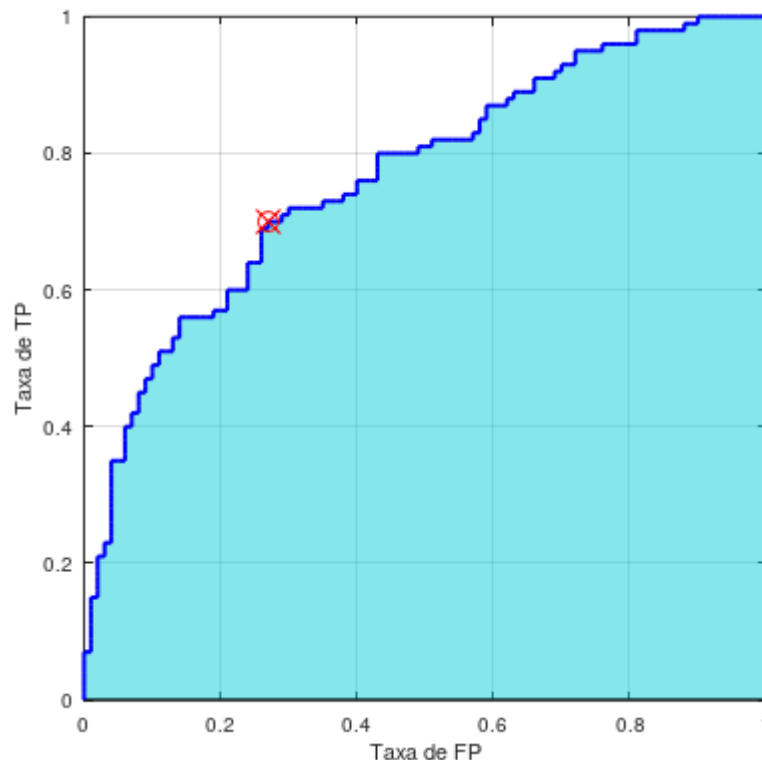


Figura 20 – Gráfico ROC de exemplo.

Fonte: Autoria própria

4.4 Fusão de bases

Apesar do método de fusão precoce ser apresentado na fundamentação teórica, devido ao tempo limitado do projeto foram realizados apenas os procedimentos de fusão tardia pois a fusão precoce exige a realização de novas classificações, aumentando assim o tempo necessário para concluir os procedimentos. O procedimento de fusão tardia utiliza dos arquivos de predições gerados através do procedimento de geração de predições (subseção 4.3.1) para realizar sua combinação, como saída do procedimento de fusão tardia se tem os arquivos de predições de fusão, que seguem o mesmo modelo dos arquivos de entrada do procedimento. A fusão tardia foi realizada utilizando todos os recursos e combinações disponíveis, combinando assim as sete predições de áudio (provindas de cada um dos sete

classificadores) com as sete predições de vídeo (com exceção do VGG16.fc7+ImageNet que foi realizada somente quatro classificações) de três formas distintas (Soma, Média e Produto), assim como é apresentado na tabela 10.

Combinação de Áudio com	Qtd. Arq. Auditivos	Qtd. Arq. Visuais	Qtd. Formas de Combinação	Total de Combinações
Histograma de Cores	7	7	3	147
Diferença Acumulada	7	7	3	147
Fluxo Óptico	7	7	3	147
VGG16.fc7+ImageNet	7	4	3	84
LBP	7	7	3	147
Total	7	32	3	672

Tabela 10 – Quantidade de combinações realizadas no processo de fusão tardia.

Para cada uma das 672 combinações realizadas foi executado o processo de Avaliação Métrica (subseção 4.3.2) o qual gerou os resultados de combinação apresentados na seção 5.3. Um detalhe importante a ser ressaltado é a forma quantitativa da combinação entre as bases, devido ao fato da base auditiva apresentar apenas 140 instâncias e as bases visuais apresentarem 30.800 instâncias foram realizadas combinações utilizando primeiramente a replicação das predições auditivas para cada uma das 220 instâncias representantes daquele mesmo vídeo musical e em um segundo momento a realização de uma média de predições das 220 instâncias de um mesmo vídeo musical, gerando assim uma predição para o vídeo musical completo a qual foi combinada com as predições auditivas sem a necessidade de uma replicação.

5 Experimentos

Para demonstrar o impacto no uso de informações visuais no processo de classificação de vídeos musicais por emoção e para responder as questões levantadas na seção 1.3, foram realizados experimentos seguindo os padrões definidos na metodologia de pesquisa (capítulo 4). A configuração experimental, juntamente com todas as informações técnicas dos experimentos é apresentada na seção 5.1. Os experimentos foram divididos em duas etapas, onde a primeira foi realizada utilizando apenas fontes de características isoladas sem qualquer fusão de dados e a segunda etapa foi realizada utilizando o conceito de fusão tardia através da combinação de informações visuais e auditivas.

5.1 Configuração Experimental

Para o procedimento foram utilizadas duas máquinas físicas e uma máquina virtual. A extração de áudio foi realizada na máquina virtual Ubuntu Linux 14.04.05 através da biblioteca Yaafe e da linguagem Python 2.7. A extração das características de Histograma de Cores, Diferença Acumulada, Fluxo Óptico e a geração de todos os arquivos no formato de relação de atributos (ARFF) foi feita em uma máquina física no ambiente Windows 7, através da biblioteca visual OpenCV 3.2 e da linguagem C ++. As características extraídas através do VGG16.fc7+ImageNet foram extraídas através da biblioteca Keras e da linguagem Python, na máquina virtual Ubuntu Linux 14.04.05 e as características extraídas através do LBP foram obtidas através da linguagem Matlab em ambiente Windows 7.

A quantificação das características extraídas são apresentadas na tabela 11, onde foram utilizados todos os rótulos da base CAL500 para todos os arquivos ARFF gerados, os atributos são as características extraídas de cada arquivo descritas na seção 4.2, as instâncias representam as amostras que foram utilizadas onde para o áudio foi utilizada toda a música, ou seja, um áudio por vídeo musical totalizando 140 instâncias, já para as características visuais foram utilizado os 220 frames de vídeo extraídos totalizando 30.800 (ou 140 vídeos musicais x 220 frames de vídeo). As características totais representam a quantidade total de características presentes em cada arquivo ARFF (desconsiderando na contagem as informações de rotulação e identificação do vídeo musical).

Método de extração	Rótulos	Atributos	Instâncias	Características totais
Áudio	18	52	140	7.280
Histograma de Cores	18	768	30.800	23.654.400
Diferença Acumulada	18	768	30.800	23.654.400
Fluxo Óptico	18	9	30.800	277.200
VGG16.fc7+ImageNet	18	4.096	30.800	126.156.800
LBP	18	59	30.800	1.817.200

Tabela 11 – Quantidade de características extraídas da base CAL500 Estendida.

Os procedimentos de gerações de predições e estratificação foram realizados através da biblioteca MULAN no ambiente Windows 7 com Java 1.8 -Xmx14G e Weka 3.7.10 e são descritos nas seções 4.3.1 e 4.1 respectivamente.

As figuras 21 e 22 apresentam o processo de geração de predições visuais e auditivas, onde na figura 21 são apresentados os treinamentos do Modelo Visual de Classificação Multi-rótulo (através das instâncias visuais e rótulos relacionadas a cada imagem do vídeo musical) e do Modelo Auditivo de Classificação Multi-rótulo (através da instância auditiva e rótulos do áudio do vídeo musical). Os modelos são treinados utilizando todos os vídeos da base de treinamento.

A figura 22 apresenta o processo de geração das predições visuais e auditivas através do uso das instâncias visuais e auditivas, as quais são apresentadas para o procedimento de Predição de classes que com base no Modelo de Classificação Multi-rótulo nos fornece as predições para cada rótulo e o valor de predição para cada rótulo

Com base nessa separação os arquivos de teste e treinamento foram gerados separadamente para Áudio, Histograma de Cores, Diferença Acumulada, Fluxo Óptico, VGG16.fc7+ImageNet e LBP (descritos na seção 4.2.3), os quais foram utilizados pelos sete classificadores descritos na seção 4.3, que retornaram as predições para cada instância. As predições foram agrupadas e comparadas com a rotulação original para definir um percentual de sucesso através da etapa de Avaliação Métrica. Para definir o percentual de sucesso e realizar comparações foi utilizada a métrica Micro-averaged AUC, que é implementada na biblioteca MULAN e descrita na seção de avaliação métrica (seção 4.3.2).

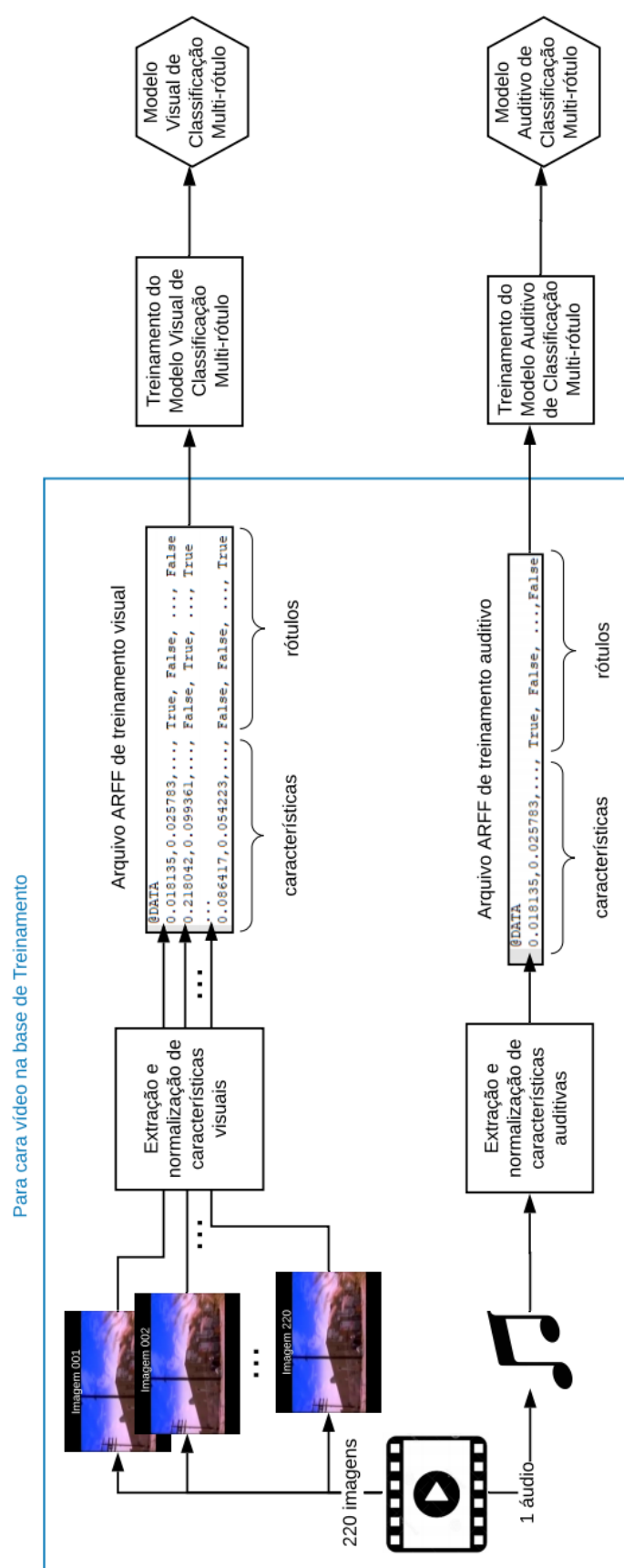


Figura 21 – Treinamento do modelo visual de classificação.

Fonte: Autoria própria

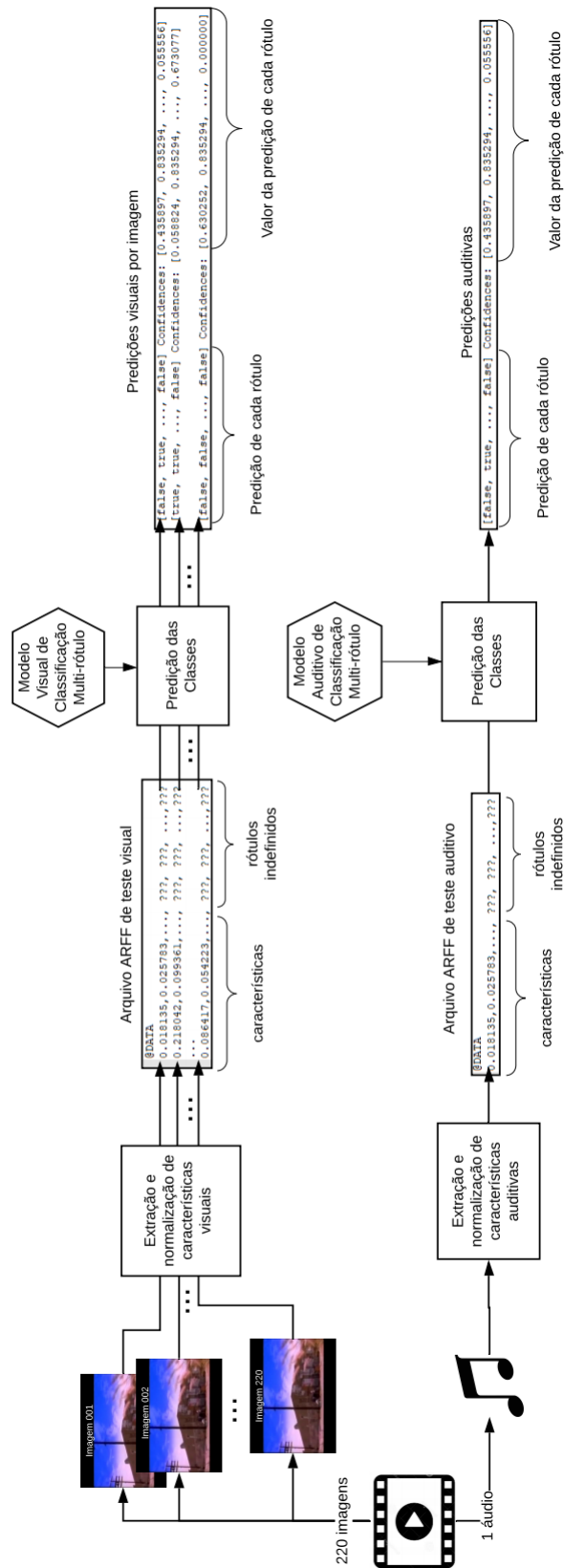


Figura 22 – Teste e geração de previsões auditivas e visuais.

Fonte: Autoria própria

5.2 Primeiro Experimento

O primeiro experimento realizado teve como objetivo responder as questões:

- 1 A utilização de informações visuais é relevante no processo de classificação de vídeos musicais?
- 2 A classificação de vídeos musicais por emoção utilizando apenas características visuais apresenta resultados relevantes?
- 3 Métodos mais robustos de extração de características visuais são melhores do que métodos simples?

Neste experimento foram utilizadas apenas as bases de forma isolada, nos possibilitando assim realizar comparações entre a classificação utilizando apenas as características extraídas do vídeo e permitindo dizer se é possível classificar vídeos musicais por emoção utilizando apenas características visuais e qual a relevância das características visuais no processo. Comparando os resultados dos classificadores que utilizaram apenas características visuais podemos obter uma comparação dos métodos de extração de características.

As tabelas 12 e 13 mostram os resultados para cada método de extração de características visuais e cada classificador. Devido ao grande volume de características do método VGG16.fc7+ImageNet, não foi possível realizar a execução da classificação utilizando os classificadores BR, HOMER e RaKEL. A tabela 12 apresenta os resultados obtidos através da realização da média entre as predições das instâncias que representam um mesmo vídeo, gerando assim um resultado que representa o vídeo musical completo. Já a tabela 13 apresenta os resultados obtidos através das predições de cada um dos 220 frame de vídeo ou imagens extraídos do vídeo musical gerando assim um resultado baseado nos frames do vídeo. Os valores destacados em negrito são os melhores resultados para cada método de extração de características. A figura 23 apresenta a geração da predição visual médias, onde as predições das 220 imagens de um mesmo vídeo são combinadas para gerar uma predição visual média para cada vídeo musical, que posteriormente (Segundo Experimento) serão combinadas com as predições auditivas de forma direta através da fusão tardia (Late Fusion).

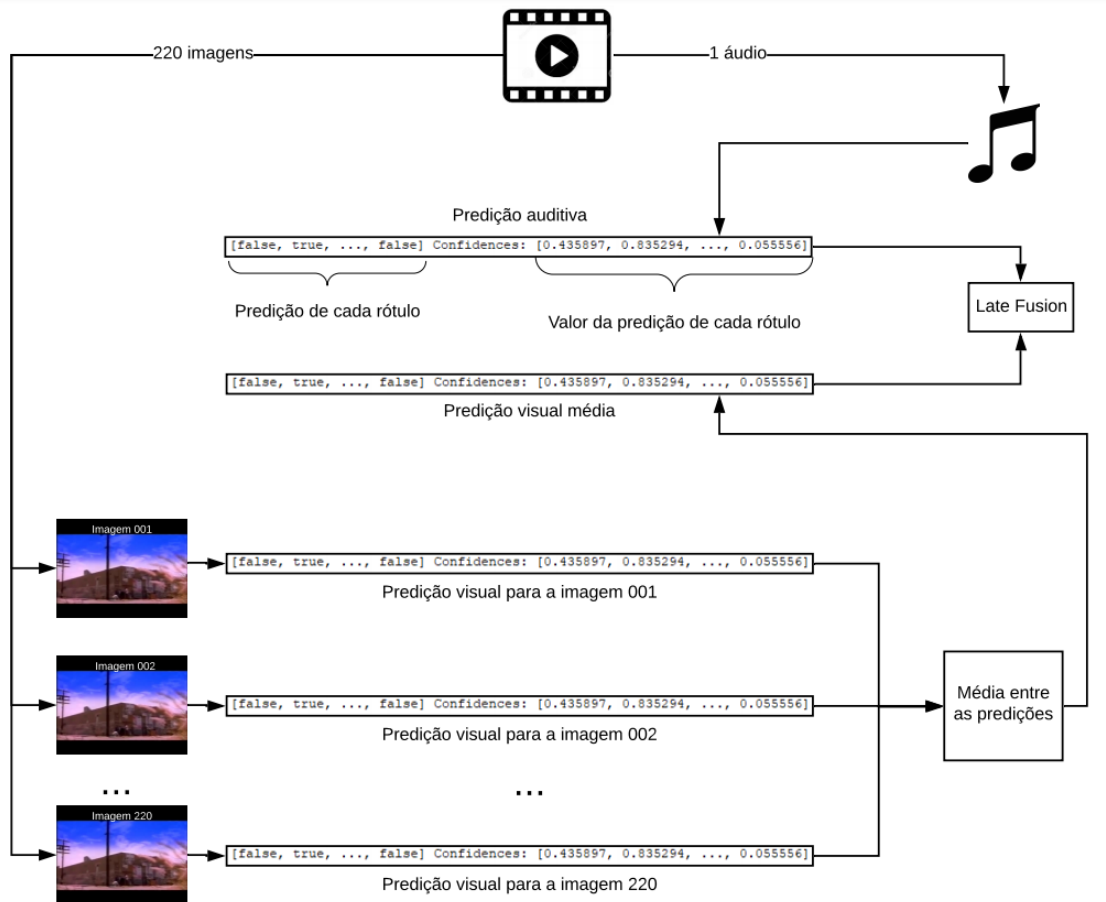


Figura 23 – Geração da previsão visual médias.

Fonte: Autoria própria

Média por Vídeo	Diferença Acumulada		Fluxo Óptico		Histograma de Cores		VGG16.fc7 +ImageNet		LBP	
	média	σ	média	σ	média	σ	média	σ	média	σ
BR	0,6931	0,0659	0,7074	0,0669	0,6768	0,0665	—	—	0,6596	0,0614
BR k NN	0,6983	0,0627	0,7173	0,0677	0,6671	0,0538	0,7002	0,0596	0,6991	0,0562
CLR	0,6997	0,0587	0,7064	0,0632	0,6882	0,0600	0,6970	0,0590	0,6989	0,0553
HOMER	0,6962	0,0683	0,6703	0,0651	0,6802	0,0613	—	—	0,6906	0,0498
LP	0,6820	0,0751	0,6824	0,0655	0,6552	0,0667	0,6946	0,0630	0,6760	0,0674
ML k NN	0,7003	0,0615	0,7061	0,0646	0,6735	0,0535	0,7017	0,0596	0,6997	0,0538
Ra k EL	0,6968	0,0684	0,7176	0,0697	0,6834	0,0700	—	—	0,7005	0,0594

Tabela 12 – Resultados obtidos através da métrica Micro-averaged AUC para cada classificador vs cada método de extração de características visuais utilizando uma média entre as previsões das 220 imagens do vídeo musical.

Uso das 220 Imagens	Diferença Acumulada		Fluxo Óptico		Histograma de Cores		VGG16.fc7 +ImageNet		LBP	
	média	σ	média	σ	média	σ	média	σ	média	σ
BR	0,5613	0,0192	0,6917	0,0597	0,5632	0,0266	—	—	0,6299	0,0499
BR k NN	0,6341	0,0427	0,6560	0,0467	0,6007	0,0368	0,6310	0,0391	0,6306	0,0364
CLR	0,6622	0,0462	0,6938	0,0587	0,6604	0,0511	0,6619	0,0462	0,6627	0,0430
HOMER	0,5640	0,0227	0,6535	0,0601	0,5672	0,0267	—	—	0,5643	0,0179
LP	0,5579	0,0238	0,5296	0,0098	0,5627	0,0277	0,5569	0,0190	0,5515	0,0201
ML k NN	0,6528	0,0452	0,6934	0,0598	0,6103	0,0375	0,6240	0,0369	0,6313	0,0339
Ra k EL	0,6229	0,0421	0,6341	0,0420	0,6235	0,0491	—	—	0,6259	0,0374

Tabela 13 – Resultados obtidos através da métrica Micro-averaged AUC para cada classificador vs cada método de extração de características visuais utilizando as predições das 220 imagens do vídeo musical.

Analisando a tabela 12 podemos notar que os melhores resultados utilizando características visuais e aplicando uma média de predições para cada vídeo foram obtidos utilizando as características extraídas através do Fluxo Óptico, com exceção quando combinado com o classificador HOMER onde apresentou o pior resultado (67,03%) em relação a outros métodos de extração de características visuais utilizando o mesmo classificador. Analisando a tabela 13 podemos notar que assim como na tabela 12 os melhores resultados foram obtidos através da classificação utilizando as características visuais de Fluxo Óptico, com exceção quando combinado com o classificador LP onde apresentou o pior resultado (52,96%) em relação a outros métodos de extração de características visuais utilizando o mesmo classificador. Realizando uma breve comparação entre as duas tabelas podemos notar que a aplicação de uma média nas predições obtidas para cada uma das 220 imagens de cada vídeo musical maximizou os resultados obtidos, apesar do ganho ser pequeno ele ainda está presente nos resultados. Como as taxas de acerto das duas tabelas utilizando características visuais estão próximas a 70.00%, podemos considerar que a classificação de vídeos musicais por emoção utilizando apenas características visuais apresenta resultados relevantes, respondendo assim a questão 2. Por apresentar resultados relevantes na classificação utilizando apenas informações visuais, podemos reconhecer que o vídeo assim como o áudio, é relevante para o processo de classificação de emoções em vídeos musicais, respondendo a questão 1.

Ao realizar comparações entre os resultados que foram obtidos através do uso de informações visuais podemos notar que os resultados obtidos através das características de VGG16.fc7+ImageNet, considerado um método robusto para extrair características, são estatisticamente os mesmos aos resultados obtidos através dos outros métodos de extração de características visuais, os quais são considerados métodos simples, nos possibilitando dizer que para os experimentos realizados, um método robusto para extração de características não é melhor que um método simples para extração de características, respondendo

assim a questão 3.

Os melhores resultados para cada método de extração utilizando a média entre as predições de um mesmo vídeo musical (tabela 12) foram: 70,03% para características extraídas da Diferença Acumulada combinadas com o classificador *MLkNN*; 71,76% para características extraídas do Fluxo Óptico combinadas com o classificador *RakEL*; 68,82% para características extraídas do Histograma de Cores combinadas com o classificador *CLR*; 70,17% para características extraídas do VGG16.fc7+ImageNet combinadas com o classificador *MLkNN*; e 70,05% para características extraídas do LBP combinadas com o classificador *RakEL*. Podemos destacar que neste experimento os melhores resultados foram obtidos através dos classificadores *CLR*, *MLkNN* e *RakEL*.

Os melhores resultados para cada método de extração utilizando as 220 imagens do vídeo musical (tabela 13) foram: 66,22% para características extraídas da Diferença Acumulada combinadas com o classificador *CLR*; 69,38% para características extraídas do Fluxo Óptico combinadas com o classificador *CLR*; 66,04% para características extraídas do Histograma de Cores combinadas com o classificador *CLR*; 66,19% para características extraídas do VGG16.fc7+ImageNet combinadas com o classificador *CLR*; e 66,27% para características extraídas do LBP combinadas com o classificador *CLR*. Podemos destacar que neste experimento os melhores resultados foram obtidos através do classificador *CLR*.

Para verificar a eficácia do uso de informações visuais no processo de classificação de emoções em vídeos musicais, foi realizada uma comparação entre a classificação utilizando apenas características de áudio e a classificação utilizando apenas características de vídeo. Para isso foi realizada a comparação entre a classificação utilizando o método de extração de características visuais que apresentou o melhor resultado (Fluxo Óptico) e a classificação utilizando o método de extração de características auditivas. Para o segundo experimento foi necessário a replicação das predições auditivas para cara predição visual do mesmo vídeo musical, ou seja, cada predição de áudio foi clonada 220 vezes para que a fusão tardia (Late Fusion) realizada no segundo experimento ocorra de forma uniforme. Devido a esta replicação foi realizada uma avaliação métrica utilizando a base de áudio replicada com o intuito de provar que a replicação do mesmo não interfere nas avaliações, o procedimento de replicação do áudio é apresentado na figura 24. Como os experimentos utilizando apenas características visuais foram conduzidos de duas formas distintas, os resultados provindos de cada uma das formas foram comparados juntamente com os resultados provindos da utilização de predições auditivas originais e predições auditivas de forma replicada.

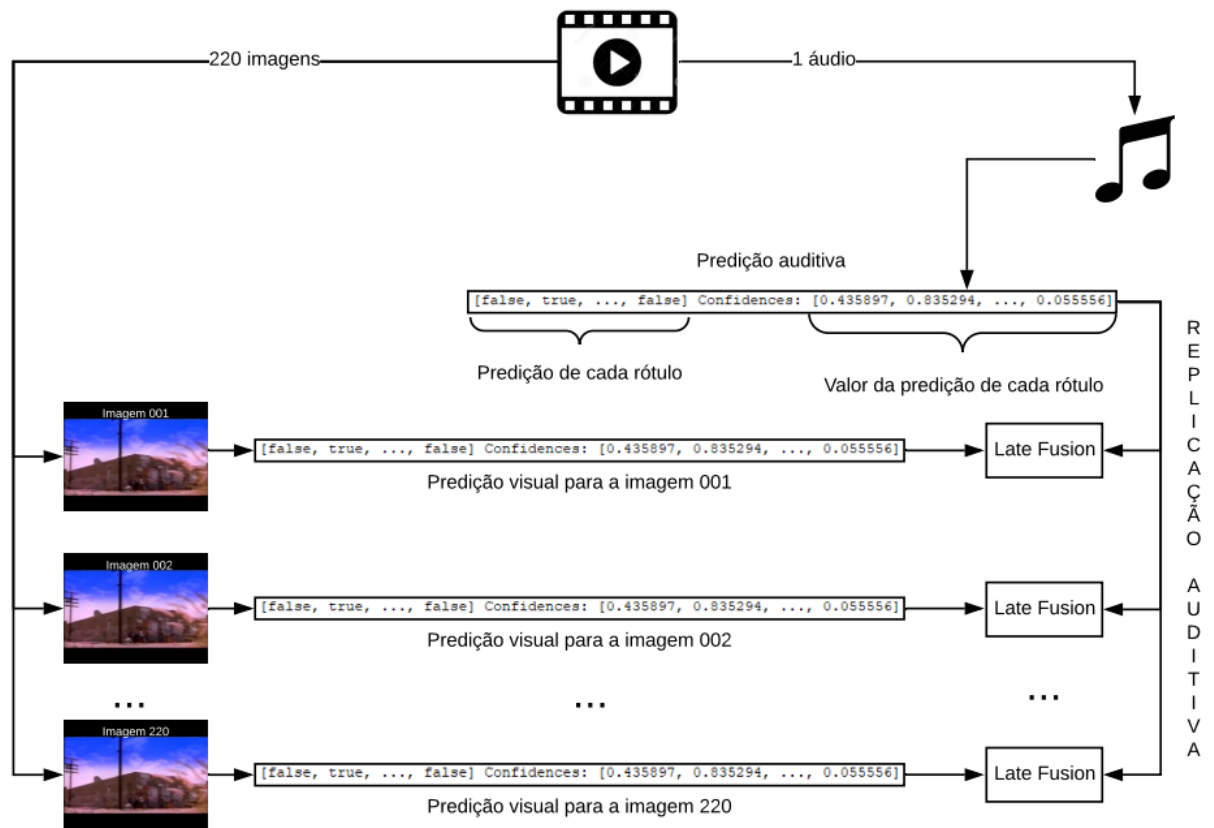


Figura 24 – Replicação das previsões auditivas

Fonte: Autoria própria

	Áudio replicado		Áudio 1 por vídeo		Fluxo Óptico 220 imagens		Fluxo Óptico um por vídeo	
	média	σ	média	σ	média	σ	média	σ
BR	0,6403	0,0344	0,6403	0,0344	0,6917	0,0597	0,7074	0,0669
BR k NN	0,7062	0,0643	0,7062	0,0643	0,6560	0,0467	0,7173	0,0677
CLR	0,7244	0,0645	0,7244	0,0645	0,6938	0,0587	0,7064	0,0632
HOMER	0,6408	0,0449	0,6408	0,0449	0,6535	0,0601	0,6703	0,0651
LP	0,5187	0,0546	0,5187	0,0546	0,5296	0,0098	0,6824	0,0655
ML k NN	0,7080	0,0509	0,7080	0,0509	0,6934	0,0598	0,7061	0,0646
Ra k EL	0,6910	0,0530	0,6910	0,0530	0,6341	0,0420	0,7176	0,0697

Tabela 14 – Resultados obtidos através da métrica Micro-averaged AUC para cada classificador comparando as extrações de características auditivas e visuais.

A tabela 14 mostra primeiramente que a replicação das previsões auditivas para as 220 imagens do vídeo musical não apresenta alterações em seus resultados, com base nos resultados apresentados na tabela serem idênticos. Podemos ressaltar também que a classificação utilizando apenas características visuais está estatisticamente no mesmo patamar que os resultados obtidos através do uso de apenas características auditivas. Os melhores resultados apresentados na tabela para cada classificador (valores em negrito) são:

72,44% para a classificação utilizando características de áudio tanto na forma replicada como na forma normal combinadas com o classificador CLR; 69,38% para a classificação utilizando as características visuais com base nas predições das 220 imagens do vídeo musical através do classificador CLR; e 71,76% para a classificação utilizando as características visuais com base na média das 220 predições de cada imagem combinadas com o classificador RakEL.

5.3 Segundo Experimento

O objetivo do segundo experimento é responder as questões da seção 1.3:

- 4 A combinação de informações visuais combinadas com as informações auditivas melhora as predições em relação ao uso de apenas informações auditivas?
- 5 A combinação de informações visuais combinadas com as informações auditivas melhora as predições em relação ao uso de apenas informações visuais?

Após realizar o primeiro experimento as predições de cada instância geradas pelos classificadores foram utilizadas pelo método de fusão tardia (descrito na seção 4.4), onde as confidências preditivas da classificação utilizando características auditivas foram combinadas de duas formas, a primeira é obtida através da combinação das predições visuais com as predições auditivas através da replicação das predições auditivas para cada predição visual que representa o mesmo vídeo musical, a segunda é obtida através da realização da média entre as 220 predições visuais de cada vídeo musical e em seguida a combinação com as predições auditivas de forma direta. Os resultados obtidos através da replicação das predições auditivas são apresentadas em cinco tabelas com base no tipo de fonte de características visuais utilizado, os quais são apresentados da seguinte forma: fusão tardia entre predições auditivas replicadas para predições visuais de Histograma de Cores (15); fusão tardia entre predições auditivas replicadas para predições visuais de Diferença Acumulada (16); fusão tardia entre predições auditivas replicadas para predições visuais de Fluxo Óptico (17); fusão tardia entre predições auditivas replicadas para predições visuais de VGG16.fc7+ImageNet (18); e fusão tardia entre predições auditivas replicadas para predições visuais de LBP (19).

Os resultados obtidos através da realização de uma média entre as predições visuais combinadas com as predições auditivas de forma direta através da fusão tardia são apresentados da seguinte forma: fusão tardia entre as médias das predições visuais de Histograma de Cores e predições auditivas de forma direta (20); fusão tardia entre as médias das predições visuais de Diferença Acumulada e predições auditivas de forma direta (21); fusão tardia entre as médias das predições visuais de Fluxo Óptico e predições auditivas de forma direta (22); fusão tardia entre as médias das predições visuais de

VGG16.fc7+ImageNet e predições auditivas de forma direta (23); fusão tardia entre as médias das predições visuais de LBP e predições auditivas de forma direta (24). Os melhores resultados de cada tabela são apresentados em negrito.

Áudio replicado	Fusão com Histograma de Cores	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	0,6386	0,0336	0,6037	0,0226	0,6382	0,0348
	BR \backslash NN	0,6620	0,0385	0,6527	0,0343	0,6607	0,0393
	CLR	0,6856	0,0429	0,6710	0,0394	0,6809	0,0436
	HOMER	0,6382	0,0333	0,6062	0,0228	0,6383	0,0348
	LP	0,6368	0,0364	0,6124	0,0288	0,6377	0,0358
	ML \backslash NN	0,6652	0,0381	0,6549	0,0346	0,6641	0,0393
	Ra \backslash EL	0,6724	0,0432	0,6536	0,0372	0,6697	0,0440
BR \backslash NN	BR	0,6635	0,0417	0,6139	0,0242	0,6597	0,0415
	BR \backslash NN	0,6685	0,0425	0,6675	0,0403	0,6724	0,0446
	CLR	0,7102	0,0535	0,7076	0,0524	0,7044	0,0532
	HOMER	0,6636	0,0403	0,6181	0,0240	0,6596	0,0405
	LP	0,6460	0,0426	0,6181	0,0327	0,6563	0,0447
	ML \backslash NN	0,6792	0,0444	0,6737	0,0414	0,6772	0,0448
	Ra \backslash EL	0,6898	0,0524	0,6774	0,0480	0,6869	0,0515
CLR	BR	0,6732	0,0472	0,6201	0,0312	0,6727	0,0449
	BR \backslash NN	0,6878	0,0516	0,6836	0,0513	0,6922	0,0483
	CLR	0,7222	0,0596	0,7218	0,0610	0,7223	0,0549
	HOMER	0,6729	0,0462	0,6240	0,0300	0,6724	0,0441
	LP	0,6598	0,0484	0,6281	0,0397	0,6749	0,0502
	ML \backslash NN	0,6910	0,0515	0,6853	0,0510	0,6955	0,0482
	Ra \backslash EL	0,7021	0,0579	0,6907	0,0556	0,7049	0,0539
HOMER	BR	0,6431	0,0375	0,6112	0,0291	0,6428	0,0376
	BR \backslash NN	0,6663	0,0465	0,6527	0,0403	0,6644	0,0490
	CLR	0,6903	0,0487	0,6733	0,0442	0,6850	0,0499
	HOMER	0,6428	0,0359	0,6097	0,0269	0,6430	0,0370
	LP	0,6436	0,0419	0,6189	0,0315	0,6416	0,0414
	ML \backslash NN	0,6694	0,0466	0,6557	0,0405	0,6677	0,0490
	Ra \backslash EL	0,6758	0,0485	0,6531	0,0391	0,6735	0,0488
LP	BR	0,5774	0,0394	0,5748	0,0328	0,5682	0,0323
	BR \backslash NN	0,6094	0,0414	0,6118	0,0442	0,6012	0,0399
	CLR	0,6548	0,0512	0,6542	0,0508	0,6459	0,0505
	HOMER	0,5783	0,0370	0,5767	0,0318	0,5721	0,0297
	LP	0,5740	0,0341	0,5737	0,0355	0,5556	0,0262
	ML \backslash NN	0,6157	0,0417	0,6195	0,0447	0,6088	0,0400
	Ra \backslash EL	0,6270	0,0467	0,6269	0,0483	0,6195	0,0456
ML \backslash NN	BR	0,6631	0,0437	0,6122	0,0270	0,6605	0,0438
	BR \backslash NN	0,6679	0,0447	0,6635	0,0418	0,6676	0,0458
	CLR	0,7033	0,0531	0,7020	0,0521	0,6983	0,0529
	HOMER	0,6630	0,0420	0,6172	0,0266	0,6604	0,0426
	LP	0,6424	0,0411	0,6145	0,0319	0,6559	0,0440
	ML \backslash NN	0,6720	0,0452	0,6660	0,0422	0,6715	0,0460
	Ra \backslash EL	0,6829	0,0535	0,6729	0,0495	0,6795	0,0529
Ra \backslash EL	BR	0,6607	0,0444	0,6213	0,0329	0,6588	0,0424
	BR \backslash NN	0,6834	0,0496	0,6780	0,0465	0,6829	0,0464
	CLR	0,7120	0,0547	0,7050	0,0530	0,7073	0,0503
	HOMER	0,6605	0,0429	0,6227	0,0317	0,6589	0,0414
	LP	0,6558	0,0438	0,6305	0,0357	0,6613	0,0440
	ML \backslash NN	0,6863	0,0494	0,6793	0,0464	0,6860	0,0463
	Ra \backslash EL	0,6961	0,0536	0,6809	0,0483	0,6944	0,0493

Tabela 15 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e Histograma de Cores.

Áudio replicado	Fusão com Diferença Acumulada	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	0,6377	0,0328	0,6049	0,0227	0,6370	0,0341
	BRkNN	0,6765	0,0431	0,6627	0,0385	0,6746	0,0438
	CLR	0,6857	0,0430	0,6706	0,0401	0,6824	0,0438
	HOMER	0,6383	0,0350	0,6076	0,0269	0,6385	0,0362
	LP	0,6403	0,0387	0,6212	0,0338	0,6404	0,0375
	MLkNN	0,6823	0,0426	0,6677	0,0388	0,6802	0,0434
	RakEL	0,6727	0,0443	0,6537	0,0414	0,6702	0,0449
BRkNN	BR	0,6607	0,0421	0,6142	0,0232	0,6566	0,0419
	BRkNN	0,6927	0,0503	0,6929	0,0492	0,6937	0,0520
	CLR	0,7112	0,0543	0,7093	0,0529	0,7056	0,0550
	HOMER	0,6623	0,0434	0,6189	0,0265	0,6582	0,0435
	LP	0,6472	0,0456	0,6305	0,0388	0,6608	0,0488
	MLkNN	0,7068	0,0531	0,7045	0,0520	0,7025	0,0536
	RakEL	0,6896	0,0536	0,6792	0,0500	0,6863	0,0526
CLR	BR	0,6705	0,0461	0,6210	0,0289	0,6699	0,0435
	BRkNN	0,7090	0,0564	0,7079	0,0573	0,7126	0,0536
	CLR	0,7218	0,0596	0,7219	0,0604	0,7229	0,0549
	HOMER	0,6721	0,0481	0,6253	0,0328	0,6721	0,0452
	LP	0,6624	0,0499	0,6441	0,0454	0,6801	0,0498
	MLkNN	0,7176	0,0577	0,7172	0,0587	0,7215	0,0540
	RakEL	0,7014	0,0583	0,6916	0,0572	0,7043	0,0541
HOMER	BR	0,6437	0,0385	0,6138	0,0289	0,6427	0,0402
	BRkNN	0,6803	0,0481	0,6636	0,0431	0,6785	0,0493
	CLR	0,6916	0,0493	0,6743	0,0453	0,6865	0,0494
	HOMER	0,6436	0,0396	0,6130	0,0296	0,6434	0,0411
	LP	0,6464	0,0476	0,6250	0,0385	0,6422	0,0464
	MLkNN	0,6872	0,0489	0,6700	0,0443	0,6845	0,0501
	RakEL	0,6764	0,0500	0,6550	0,0436	0,6740	0,0504
LP	BR	0,5751	0,0334	0,5735	0,0283	0,5669	0,0253
	BRkNN	0,6363	0,0456	0,6360	0,0454	0,6320	0,0410
	CLR	0,6562	0,0473	0,6558	0,0462	0,6490	0,0438
	HOMER	0,5769	0,0357	0,5750	0,0299	0,5705	0,0272
	LP	0,5709	0,0301	0,5713	0,0304	0,5420	0,0317
	MLkNN	0,6490	0,0469	0,6502	0,0469	0,6447	0,0424
	RakEL	0,6268	0,0439	0,6259	0,0445	0,6203	0,0422
MLkNN	BR	0,6598	0,0415	0,6118	0,0235	0,6573	0,0413
	BRkNN	0,6905	0,0512	0,6881	0,0498	0,6891	0,0520
	CLR	0,7046	0,0524	0,7033	0,0515	0,7002	0,0531
	HOMER	0,6615	0,0429	0,6176	0,0278	0,6589	0,0436
	LP	0,6423	0,0417	0,6261	0,0364	0,6591	0,0446
	MLkNN	0,6993	0,0519	0,6973	0,0510	0,6975	0,0527
	RakEL	0,6836	0,0524	0,6744	0,0488	0,6805	0,0524
RakEL	BR	0,6587	0,0430	0,6218	0,0302	0,6571	0,0409
	BRkNN	0,7025	0,0547	0,6960	0,0520	0,7014	0,0510
	CLR	0,7126	0,0547	0,7059	0,0523	0,7094	0,0499
	HOMER	0,6599	0,0447	0,6240	0,0334	0,6588	0,0427
	LP	0,6594	0,0470	0,6429	0,0415	0,6661	0,0459
	MLkNN	0,7091	0,0548	0,7023	0,0527	0,7079	0,0509
	RakEL	0,6965	0,0542	0,6827	0,0509	0,6946	0,0500

Tabela 16 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e Diferença Acumulada.

Áudio replicado	Fusão com Fluxo Óptico	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	0,6928	0,0450	0,6741	0,0411	0,6909	0,0447
	BR \hat{k} NN	0,6849	0,0443	0,6700	0,0407	0,6816	0,0442
	CLR	0,7000	0,0474	0,6826	0,0444	0,6995	0,0480
	HOMER	0,6863	0,0412	0,6739	0,0374	0,6844	0,0437
	LP	0,6396	0,0349	0,6310	0,0333	0,6377	0,0352
	ML \hat{k} NN	0,6923	0,0445	0,6741	0,0400	0,6907	0,0439
	Ra \hat{k} EL	0,6774	0,0433	0,6583	0,0400	0,6746	0,0437
BR \hat{k} NN	BR	0,7214	0,0578	0,7194	0,0575	0,7173	0,0574
	BR \hat{k} NN	0,7067	0,0531	0,7062	0,0527	0,7046	0,0526
	CLR	0,7205	0,0556	0,7166	0,0547	0,7176	0,0550
	HOMER	0,7106	0,0499	0,7084	0,0498	0,7055	0,0489
	LP	0,6635	0,0470	0,6565	0,0464	0,6677	0,0482
	ML \hat{k} NN	0,7209	0,0576	0,7190	0,0574	0,7162	0,0571
	Ra \hat{k} EL	0,6958	0,0517	0,6851	0,0482	0,6923	0,0495
CLR	BR	0,7310	0,0628	0,7309	0,0639	0,7359	0,0575
	BR \hat{k} NN	0,7207	0,0602	0,7196	0,0612	0,7225	0,0552
	CLR	0,7281	0,0616	0,7255	0,0627	0,7331	0,0577
	HOMER	0,7213	0,0570	0,7206	0,0573	0,7241	0,0558
	LP	0,6809	0,0510	0,6780	0,0530	0,6915	0,0509
	ML \hat{k} NN	0,7305	0,0630	0,7302	0,0641	0,7357	0,0574
	Ra \hat{k} EL	0,7066	0,0578	0,6969	0,0564	0,7089	0,0526
HOMER	BR	0,7006	0,0509	0,6797	0,0464	0,6957	0,0529
	BR \hat{k} NN	0,6889	0,0497	0,6713	0,0453	0,6849	0,0504
	CLR	0,7046	0,0537	0,6831	0,0482	0,7020	0,0552
	HOMER	0,6850	0,0494	0,6677	0,0467	0,6835	0,0509
	LP	0,6437	0,0449	0,6307	0,0405	0,6386	0,0441
	ML \hat{k} NN	0,6992	0,0511	0,6785	0,0471	0,6943	0,0530
	Ra \hat{k} EL	0,6807	0,0499	0,6588	0,0433	0,6782	0,0496
LP	BR	0,6730	0,0577	0,6731	0,0575	0,6667	0,0562
	BR \hat{k} NN	0,6513	0,0487	0,6500	0,0472	0,6433	0,0459
	CLR	0,6893	0,0580	0,6888	0,0581	0,6831	0,0582
	HOMER	0,6355	0,0609	0,6253	0,0590	0,6423	0,0565
	LP	0,5455	0,0328	0,5456	0,0320	0,5191	0,0376
	ML \hat{k} NN	0,6725	0,0585	0,6730	0,0584	0,6661	0,0550
	Ra \hat{k} EL	0,6359	0,0437	0,6357	0,0430	0,6284	0,0425
ML \hat{k} NN	BR	0,7130	0,0557	0,7117	0,0554	0,7111	0,0565
	BR \hat{k} NN	0,7028	0,0527	0,7008	0,0523	0,6992	0,0529
	CLR	0,7119	0,0565	0,7097	0,0565	0,7108	0,0573
	HOMER	0,7058	0,0510	0,7057	0,0492	0,7038	0,0560
	LP	0,6576	0,0400	0,6555	0,0420	0,6672	0,0412
	ML \hat{k} NN	0,7122	0,0549	0,7108	0,0547	0,7104	0,0557
	Ra \hat{k} EL	0,6889	0,0516	0,6802	0,0486	0,6857	0,0510
Ra \hat{k} EL	BR	0,7190	0,0582	0,7111	0,0559	0,7176	0,0530
	BR \hat{k} NN	0,7111	0,0564	0,7039	0,0543	0,7079	0,0509
	CLR	0,7226	0,0595	0,7119	0,0584	0,7224	0,0558
	HOMER	0,7063	0,0571	0,6981	0,0529	0,7052	0,0522
	LP	0,6690	0,0462	0,6632	0,0453	0,6703	0,0449
	ML \hat{k} NN	0,7185	0,0578	0,7104	0,0560	0,7167	0,0524
	Ra \hat{k} EL	0,7006	0,0547	0,6867	0,0513	0,6976	0,0496

Tabela 17 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e Fluxo Óptico.

Áudio replicado	Fusão com VGG16.fc7 + ImageNet	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	—	—	—	—	—	—
	BRkNN	0,6766	0,0435	0,6634	0,0386	0,6740	0,0446
	CLR	0,6855	0,0408	0,6704	0,0377	0,6813	0,0410
	HOMER	—	—	—	—	—	—
	LP	0,6375	0,0334	0,6138	0,0268	0,6378	0,0329
	MLkNN	0,6730	0,0417	0,6592	0,0369	0,6708	0,0433
	RakEL	—	—	—	—	—	—
BRkNN	BR	—	—	—	—	—	—
	BRkNN	0,6895	0,0494	0,6880	0,0476	0,6913	0,0498
	CLR	0,7115	0,0532	0,7092	0,0523	0,7063	0,0511
	HOMER	—	—	—	—	—	—
	LP	0,6426	0,0370	0,6218	0,0297	0,6585	0,0422
	MLkNN	0,6880	0,0482	0,6811	0,0445	0,6857	0,0476
	RakEL	—	—	—	—	—	—
CLR	BR	—	—	—	—	—	—
	BRkNN	0,7067	0,0553	0,7038	0,0554	0,7107	0,0516
	CLR	0,7219	0,0577	0,7221	0,0590	0,7224	0,0526
	HOMER	—	—	—	—	—	—
	LP	0,6583	0,0423	0,6357	0,0365	0,6780	0,0444
	MLkNN	0,6990	0,0524	0,6918	0,0512	0,7025	0,0496
	RakEL	—	—	—	—	—	—
HOMER	BR	—	—	—	—	—	—
	BRkNN	0,6812	0,0482	0,6655	0,0441	0,6793	0,0516
	CLR	0,6909	0,0482	0,6751	0,0445	0,6863	0,0501
	HOMER	—	—	—	—	—	—
	LP	0,6447	0,0402	0,6196	0,0310	0,6414	0,0409
	MLkNN	0,6776	0,0469	0,6625	0,0433	0,6759	0,0501
	RakEL	—	—	—	—	—	—
LP	BR	—	—	—	—	—	—
	BRkNN	0,6329	0,0367	0,6322	0,0375	0,6268	0,0352
	CLR	0,6542	0,0462	0,6530	0,0456	0,6470	0,0444
	HOMER	—	—	—	—	—	—
	LP	0,5687	0,0279	0,5683	0,0274	0,5405	0,0320
	MLkNN	0,6265	0,0361	0,6280	0,0382	0,6222	0,0348
	RakEL	—	—	—	—	—	—
MLkNN	BR	—	—	—	—	—	—
	BRkNN	0,6873	0,0480	0,6833	0,0466	0,6866	0,0483
	CLR	0,7039	0,0513	0,7028	0,0511	0,7008	0,0500
	HOMER	—	—	—	—	—	—
	LP	0,6386	0,0356	0,6197	0,0301	0,6585	0,0392
	MLkNN	0,6808	0,0462	0,6730	0,0439	0,6802	0,0468
	RakEL	—	—	—	—	—	—
RakEL	BR	—	—	—	—	—	—
	BRkNN	0,7002	0,0511	0,6934	0,0493	0,6990	0,0475
	CLR	0,7119	0,0539	0,7055	0,0523	0,7082	0,0479
	HOMER	—	—	—	—	—	—
	LP	0,6558	0,0405	0,6347	0,0347	0,6637	0,0407
	MLkNN	0,6942	0,0492	0,6855	0,0473	0,6932	0,0462
	RakEL	—	—	—	—	—	—

Tabela 18 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e VGG16.fc7+ImageNet.

Áudio replicado	Fusão com LBP	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	0,6836	0,0492	0,6357	0,0368	0,6833	0,0505
	BRkNN	0,6765	0,0404	0,6631	0,0373	0,6760	0,0409
	CLR	0,6872	0,0411	0,6717	0,0378	0,6845	0,0415
	HOMER	0,6385	0,0304	0,6090	0,0223	0,6389	0,0317
	LP	0,6372	0,0326	0,6170	0,0267	0,6377	0,0324
	MLkNN	0,6766	0,0392	0,6635	0,0366	0,6761	0,0400
	RakEL	0,6742	0,0403	0,6559	0,0358	0,6723	0,0410
BRkNN	BR	0,7040	0,0532	0,6563	0,0354	0,7013	0,0544
	BRkNN	0,6886	0,0461	0,6876	0,0451	0,6914	0,0462
	CLR	0,7107	0,0530	0,7082	0,0520	0,7062	0,0509
	HOMER	0,6611	0,0401	0,6212	0,0249	0,6567	0,0390
	LP	0,6428	0,0384	0,6232	0,0324	0,6569	0,0416
	MLkNN	0,6937	0,0466	0,6899	0,0452	0,6921	0,0454
	RakEL	0,6908	0,0498	0,6805	0,0457	0,6883	0,0478
CLR	BR	0,7166	0,0605	0,6581	0,0420	0,7179	0,0570
	BRkNN	0,7057	0,0539	0,7031	0,0539	0,7096	0,0508
	CLR	0,7218	0,0576	0,7220	0,0585	0,7226	0,0531
	HOMER	0,6707	0,0441	0,6288	0,0285	0,6706	0,0415
	LP	0,6584	0,0435	0,6396	0,0393	0,6763	0,0453
	MLkNN	0,7059	0,0530	0,7031	0,0529	0,7101	0,0497
	RakEL	0,7034	0,0546	0,6943	0,0525	0,7056	0,0507
HOMER	BR	0,6816	0,0494	0,6373	0,0377	0,6809	0,0511
	BRkNN	0,6785	0,0455	0,6625	0,0405	0,6768	0,0473
	CLR	0,6914	0,0469	0,6744	0,0426	0,6867	0,0483
	HOMER	0,6441	0,0366	0,6142	0,0264	0,6440	0,0380
	LP	0,6435	0,0426	0,6214	0,0347	0,6405	0,0435
	MLkNN	0,6793	0,0454	0,6639	0,0404	0,6776	0,0469
	RakEL	0,6769	0,0465	0,6560	0,0396	0,6742	0,0470
LP	BR	0,6417	0,0634	0,6331	0,0486	0,6359	0,0569
	BRkNN	0,6339	0,0423	0,6330	0,0414	0,6280	0,0398
	CLR	0,6574	0,0473	0,6564	0,0459	0,6489	0,0441
	HOMER	0,5773	0,0334	0,5777	0,0281	0,5718	0,0262
	LP	0,5634	0,0338	0,5640	0,0330	0,5391	0,0334
	MLkNN	0,6339	0,0399	0,6346	0,0397	0,6288	0,0380
	RakEL	0,6287	0,0439	0,6286	0,0427	0,6211	0,0405
MLkNN	BR	0,7041	0,0539	0,6559	0,0400	0,7029	0,0543
	BRkNN	0,6866	0,0460	0,6835	0,0441	0,6871	0,0466
	CLR	0,7040	0,0504	0,7029	0,0501	0,7016	0,0501
	HOMER	0,6602	0,0378	0,6210	0,0237	0,6576	0,0381
	LP	0,6374	0,0366	0,6210	0,0320	0,6556	0,0387
	MLkNN	0,6873	0,0449	0,6836	0,0430	0,6875	0,0453
	RakEL	0,6844	0,0484	0,6764	0,0447	0,6828	0,0484
RakEL	BR	0,7017	0,0598	0,6481	0,0477	0,7009	0,0563
	BRkNN	0,6993	0,0513	0,6922	0,0490	0,6990	0,0475
	CLR	0,7126	0,0538	0,7055	0,0517	0,7090	0,0491
	HOMER	0,6591	0,0401	0,6256	0,0288	0,6582	0,0378
	LP	0,6555	0,0423	0,6374	0,0367	0,6626	0,0411
	MLkNN	0,6993	0,0501	0,6921	0,0480	0,6992	0,0462
	RakEL	0,6973	0,0515	0,6846	0,0475	0,6955	0,0473

Tabela 19 – Resultados obtidos através da métrica Micro-averaged AUC para fusão tardia entre as bases de Áudio replicada e LBP.

Áudio	Fusão com média de predições do Histograma de Cores	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	0,6901	0,0461	0,6724	0,0410	0,6856	0,0472
	BR \hat{k} NN	0,6833	0,0419	0,6690	0,0369	0,6802	0,0421
	CLR	0,6920	0,0445	0,6748	0,0405	0,6877	0,0446
	HOMER	0,6905	0,0450	0,6719	0,0382	0,6842	0,0488
	LP	0,6784	0,0462	0,6642	0,0424	0,6741	0,0470
	ML \hat{k} NN	0,6850	0,0422	0,6701	0,0371	0,6818	0,0424
	Ra \hat{k} EL	0,6905	0,0475	0,6739	0,0414	0,6879	0,0495
BR \hat{k} NN	BR	0,7173	0,0579	0,7157	0,0579	0,7120	0,0602
	BR \hat{k} NN	0,7120	0,0542	0,7117	0,0534	0,7087	0,0582
	CLR	0,7204	0,0567	0,7190	0,0562	0,7162	0,0583
	HOMER	0,7184	0,0562	0,7167	0,0546	0,7130	0,0610
	LP	0,7099	0,0616	0,7073	0,0619	0,7034	0,0618
	ML \hat{k} NN	0,7137	0,0540	0,7130	0,0531	0,7112	0,0574
	Ra \hat{k} EL	0,7205	0,0603	0,7192	0,0594	0,7169	0,0615
CLR	BR	0,7285	0,0646	0,7288	0,0673	0,7307	0,0594
	BR \hat{k} NN	0,7230	0,0605	0,7237	0,0622	0,7288	0,0556
	CLR	0,7328	0,0627	0,7333	0,0641	0,7357	0,0574
	HOMER	0,7299	0,0606	0,7303	0,0607	0,7352	0,0580
	LP	0,7221	0,0667	0,7202	0,0689	0,7269	0,0637
	ML \hat{k} NN	0,7243	0,0600	0,7248	0,0612	0,7299	0,0542
	Ra \hat{k} EL	0,7305	0,0654	0,7309	0,0665	0,7364	0,0602
HOMER	BR	0,6964	0,0525	0,6758	0,0441	0,6900	0,0542
	BR \hat{k} NN	0,6922	0,0491	0,6718	0,0407	0,6864	0,0549
	CLR	0,7005	0,0507	0,6800	0,0449	0,6941	0,0527
	HOMER	0,6963	0,0513	0,6744	0,0440	0,6892	0,0540
	LP	0,6870	0,0517	0,6676	0,0455	0,6819	0,0543
	ML \hat{k} NN	0,6942	0,0486	0,6737	0,0409	0,6883	0,0536
	Ra \hat{k} EL	0,6973	0,0528	0,6760	0,0444	0,6942	0,0561
LP	BR	0,6650	0,0607	0,6671	0,0621	0,6540	0,0673
	BR \hat{k} NN	0,6570	0,0585	0,6573	0,0591	0,6476	0,0607
	CLR	0,6725	0,0581	0,6733	0,0581	0,6644	0,0625
	HOMER	0,6708	0,0556	0,6706	0,0568	0,6599	0,0678
	LP	0,6455	0,0515	0,6448	0,0526	0,6378	0,0649
	ML \hat{k} NN	0,6619	0,0579	0,6636	0,0589	0,6529	0,0591
	Ra \hat{k} EL	0,6692	0,0608	0,6716	0,0633	0,6622	0,0627
	Ra \hat{k} EL	0,6287	0,0439	0,6286	0,0427	0,6211	0,0405
ML \hat{k} NN	BR	0,7106	0,0592	0,7096	0,0589	0,7067	0,0599
	BR \hat{k} NN	0,7045	0,0522	0,7047	0,0514	0,7033	0,0550
	CLR	0,7135	0,0555	0,7127	0,0552	0,7108	0,0558
	HOMER	0,7104	0,0565	0,7097	0,0549	0,7066	0,0590
	LP	0,7034	0,0586	0,7013	0,0580	0,7002	0,0608
	ML \hat{k} NN	0,7063	0,0523	0,7058	0,0516	0,7043	0,0549
	Ra \hat{k} EL	0,7118	0,0608	0,7110	0,0600	0,7090	0,0604
Ra \hat{k} EL	BR	0,7178	0,0589	0,7095	0,0586	0,7131	0,0555
	BR \hat{k} NN	0,7110	0,0557	0,7047	0,0534	0,7106	0,0516
	CLR	0,7199	0,0570	0,7126	0,0553	0,7162	0,0515
	HOMER	0,7182	0,0576	0,7097	0,0553	0,7148	0,0546
	LP	0,7091	0,0582	0,6999	0,0569	0,7062	0,0542
	ML \hat{k} NN	0,7132	0,0557	0,7063	0,0534	0,7117	0,0515
	Ra \hat{k} EL	0,7187	0,0598	0,7110	0,0565	0,7189	0,0535

Tabela 20 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de predições do Histograma de Cores.

Áudio	Fusão com média de predições da Diferença Acumulada	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	0,6926	0,0469	0,6740	0,0428	0,6902	0,0492
	BR \hat{k} NN	0,6941	0,0462	0,6754	0,0408	0,6932	0,0469
	CLR	0,6950	0,0440	0,6766	0,0406	0,6933	0,0452
	HOMER	0,6936	0,0494	0,6753	0,0451	0,6928	0,0496
	LP	0,6831	0,0467	0,6671	0,0447	0,6783	0,0442
	ML \hat{k} NN	0,6945	0,0448	0,6760	0,0399	0,6932	0,0449
	Ra \hat{k} EL	0,6944	0,0488	0,6762	0,0443	0,6944	0,0488
BR \hat{k} NN	BR	0,7221	0,0619	0,7206	0,0623	0,7160	0,0655
	BR \hat{k} NN	0,7242	0,0604	0,7229	0,0599	0,7200	0,0633
	CLR	0,7236	0,0583	0,7222	0,0580	0,7175	0,0605
	HOMER	0,7234	0,0620	0,7221	0,0618	0,7185	0,0655
	LP	0,7202	0,0681	0,7193	0,0683	0,7121	0,0715
	ML \hat{k} NN	0,7238	0,0595	0,7225	0,0593	0,7202	0,0609
	Ra \hat{k} EL	0,7240	0,0640	0,7230	0,0638	0,7175	0,0660
CLR	BR	0,7317	0,0669	0,7318	0,0681	0,7374	0,0627
	BR \hat{k} NN	0,7337	0,0636	0,7337	0,0644	0,7379	0,0605
	CLR	0,7334	0,0628	0,7338	0,0633	0,7367	0,0578
	HOMER	0,7320	0,0668	0,7315	0,0683	0,7371	0,0619
	LP	0,7304	0,0701	0,7299	0,0710	0,7336	0,0638
	ML \hat{k} NN	0,7342	0,0630	0,7345	0,0640	0,7384	0,0582
	Ra \hat{k} EL	0,7330	0,0670	0,7329	0,0687	0,7380	0,0623
HOMER	BR	0,7031	0,0529	0,6805	0,0473	0,6986	0,0549
	BR \hat{k} NN	0,7028	0,0519	0,6808	0,0468	0,7001	0,0529
	CLR	0,7044	0,0515	0,6832	0,0472	0,6996	0,0521
	HOMER	0,7033	0,0547	0,6805	0,0501	0,7007	0,0552
	LP	0,6968	0,0574	0,6786	0,0523	0,6871	0,0548
	ML \hat{k} NN	0,7027	0,0512	0,6813	0,0473	0,6989	0,0546
	Ra \hat{k} EL	0,7035	0,0547	0,6821	0,0487	0,7000	0,0543
LP	BR	0,6775	0,0595	0,6786	0,0616	0,6682	0,0614
	BR \hat{k} NN	0,6784	0,0602	0,6772	0,0602	0,6746	0,0528
	CLR	0,6788	0,0573	0,6812	0,0554	0,6734	0,0527
	HOMER	0,6794	0,0601	0,6786	0,0597	0,6748	0,0594
	LP	0,6441	0,0576	0,6429	0,0573	0,6430	0,0592
	ML \hat{k} NN	0,6795	0,0590	0,6803	0,0595	0,6747	0,0530
	Ra \hat{k} EL	0,6801	0,0609	0,6807	0,0614	0,6730	0,0593
ML \hat{k} NN	BR	0,7145	0,0593	0,7135	0,0586	0,7118	0,0629
	BR \hat{k} NN	0,7164	0,0580	0,7154	0,0571	0,7152	0,0604
	CLR	0,7164	0,0554	0,7157	0,0547	0,7143	0,0582
	HOMER	0,7159	0,0611	0,7150	0,0601	0,7144	0,0645
	LP	0,7153	0,0601	0,7141	0,0588	0,7100	0,0641
	ML \hat{k} NN	0,7163	0,0570	0,7155	0,0566	0,7151	0,0583
	Ra \hat{k} EL	0,7173	0,0612	0,7162	0,0606	0,7148	0,0636
Ra \hat{k} EL	BR	0,7216	0,0608	0,7129	0,0582	0,7195	0,0579
	BR \hat{k} NN	0,7217	0,0592	0,7148	0,0563	0,7231	0,0561
	CLR	0,7219	0,0573	0,7144	0,0542	0,7209	0,0538
	HOMER	0,7222	0,0612	0,7139	0,0588	0,7208	0,0577
	LP	0,7132	0,0615	0,7084	0,0604	0,7103	0,0548
	ML \hat{k} NN	0,7222	0,0583	0,7146	0,0557	0,7219	0,0544
	Ra \hat{k} EL	0,7223	0,0611	0,7145	0,0587	0,7223	0,0572

Tabela 21 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de predições da Diferença Acumulada.

Áudio	Fusão com média de previsões do Fluxo Óptico	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	0,6974	0,0457	0,6777	0,0411	0,6947	0,0445
	BR k NN	0,7004	0,0472	0,6806	0,0417	0,6979	0,0471
	CLR	0,7049	0,0482	0,6840	0,0440	0,7047	0,0485
	HOMER	0,6915	0,0431	0,6742	0,0370	0,6885	0,0443
	LP	0,6644	0,0371	0,6582	0,0374	0,6582	0,0381
	ML k NN	0,6965	0,0452	0,6775	0,0407	0,6934	0,0435
	Ra k EL	0,7018	0,0483	0,6817	0,0433	0,6991	0,0483
BR k NN	BR	0,7258	0,0596	0,7235	0,0596	0,7217	0,0587
	BR k NN	0,7293	0,0607	0,7280	0,0606	0,7246	0,0595
	CLR	0,7259	0,0575	0,7241	0,0572	0,7233	0,0569
	HOMER	0,7154	0,0510	0,7137	0,0514	0,7105	0,0487
	LP	0,7156	0,0635	0,7158	0,0633	0,7060	0,0612
	ML k NN	0,7229	0,0591	0,7209	0,0587	0,7183	0,0580
	Ra k EL	0,7308	0,0622	0,7295	0,0621	0,7255	0,0608
CLR	BR	0,7339	0,0637	0,7339	0,0650	0,7400	0,0585
	BR k NN	0,7378	0,0651	0,7373	0,0660	0,7429	0,0593
	CLR	0,7331	0,0632	0,7319	0,0648	0,7395	0,0589
	HOMER	0,7248	0,0574	0,7257	0,0585	0,7292	0,0563
	LP	0,7288	0,0655	0,7284	0,0657	0,7285	0,0600
	ML k NN	0,7324	0,0632	0,7321	0,0647	0,7380	0,0576
	Ra k EL	0,7383	0,0666	0,7377	0,0683	0,7440	0,0605
HOMER	BR	0,7057	0,0521	0,6842	0,0474	0,7001	0,0554
	BR k NN	0,7103	0,0545	0,6873	0,0496	0,7050	0,0557
	CLR	0,7094	0,0549	0,6863	0,0484	0,7066	0,0575
	HOMER	0,6926	0,0537	0,6738	0,0491	0,6928	0,0548
	LP	0,6830	0,0502	0,6757	0,0482	0,6726	0,0497
	ML k NN	0,7025	0,0508	0,6813	0,0475	0,6965	0,0542
	Ra k EL	0,7107	0,0570	0,6869	0,0506	0,7058	0,0583
LP	BR	0,6805	0,0617	0,6811	0,0609	0,6756	0,0594
	BR k NN	0,6873	0,0619	0,6881	0,0621	0,6838	0,0606
	CLR	0,6979	0,0616	0,6974	0,0620	0,6917	0,0620
	HOMER	0,6482	0,0654	0,6364	0,0610	0,6535	0,0618
	LP	0,5898	0,0536	0,5906	0,0533	0,6212	0,0532
	ML k NN	0,6786	0,0604	0,6784	0,0608	0,6725	0,0568
	Ra k EL	0,6906	0,0628	0,6917	0,0625	0,6860	0,0610
ML k NN	BR	0,7169	0,0574	0,7158	0,0570	0,7150	0,0580
	BR k NN	0,7215	0,0578	0,7203	0,0574	0,7194	0,0583
	CLR	0,7170	0,0585	0,7156	0,0589	0,7153	0,0590
	HOMER	0,7096	0,0522	0,7097	0,0513	0,7067	0,0570
	LP	0,7127	0,0526	0,7123	0,0526	0,7093	0,0514
	ML k NN	0,7143	0,0560	0,7131	0,0555	0,7120	0,0564
	Ra k EL	0,7223	0,0603	0,7215	0,0600	0,7206	0,0604
Ra k EL	BR	0,7226	0,0585	0,7145	0,0564	0,7213	0,0531
	BR k NN	0,7256	0,0597	0,7174	0,0571	0,7230	0,0539
	CLR	0,7268	0,0607	0,7162	0,0593	0,7282	0,0569
	HOMER	0,7113	0,0584	0,7030	0,0547	0,7113	0,0539
	LP	0,7027	0,0551	0,7001	0,0542	0,6965	0,0507
	ML k NN	0,7206	0,0579	0,7127	0,0557	0,7182	0,0525
	Ra k EL	0,7272	0,0617	0,7182	0,0593	0,7252	0,0558

Tabela 22 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de previsões do Fluxo Óptico.

Áudio	Fusão com média de predições do VGG16.fc7+ImageNet	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	—	—	—	—	—	—
	BR k NN	0,6976	0,0475	0,6774	0,0418	0,6936	0,0503
	CLR	0,6945	0,0424	0,6761	0,0385	0,6907	0,0415
	HOMER	—	—	—	—	—	—
	LP	0,6854	0,0418	0,6708	0,0381	0,6782	0,0418
	ML k NN	0,6985	0,0470	0,6787	0,0416	0,6945	0,0497
	Ra k EL	—	—	—	—	—	—
BR k NN	BR	—	—	—	—	—	—
	BR k NN	0,7252	0,0609	0,7234	0,0602	0,7213	0,0619
	CLR	0,7226	0,0572	0,7207	0,0569	0,7191	0,0559
	HOMER	—	—	—	—	—	—
	LP	0,7210	0,0606	0,7185	0,0604	0,7159	0,0594
	ML k NN	0,7273	0,0608	0,7247	0,0601	0,7242	0,0618
	Ra k EL	—	—	—	—	—	—
CLR	BR	—	—	—	—	—	—
	BR k NN	0,7355	0,0636	0,7348	0,0644	0,7424	0,0610
	CLR	0,7326	0,0610	0,7329	0,0623	0,7373	0,0550
	HOMER	—	—	—	—	—	—
	LP	0,7337	0,0626	0,7330	0,0640	0,7365	0,0568
	ML k NN	0,7369	0,0631	0,7353	0,0644	0,7414	0,0619
	Ra k EL	—	—	—	—	—	—
HOMER	BR	—	—	—	—	—	—
	BR k NN	0,7060	0,0520	0,6848	0,0486	0,7016	0,0580
	CLR	0,7024	0,0509	0,6832	0,0464	0,6983	0,0544
	HOMER	—	—	—	—	—	—
	LP	0,6966	0,0477	0,6800	0,0427	0,6878	0,0506
	ML k NN	0,7066	0,0518	0,6842	0,0482	0,7025	0,0577
	Ra k EL	—	—	—	—	—	—
LP	BR	—	—	—	—	—	—
	BR k NN	0,6786	0,0516	0,6787	0,0528	0,6722	0,0531
	CLR	0,6761	0,0554	0,6764	0,0561	0,6720	0,0537
	HOMER	—	—	—	—	—	—
	LP	0,6480	0,0516	0,6472	0,0516	0,6461	0,0507
	ML k NN	0,6815	0,0533	0,6828	0,0547	0,6779	0,0570
	Ra k EL	—	—	—	—	—	—
ML k NN	BR	—	—	—	—	—	—
	BR k NN	0,7169	0,0561	0,7164	0,0562	0,7158	0,0577
	CLR	0,7144	0,0550	0,7140	0,0543	0,7150	0,0535
	HOMER	—	—	—	—	—	—
	LP	0,7145	0,0552	0,7132	0,0543	0,7123	0,0564
	ML k NN	0,7177	0,0567	0,7163	0,0563	0,7173	0,0575
	Ra k EL	—	—	—	—	—	—
Ra k EL	BR	—	—	—	—	—	—
	BR k NN	0,7230	0,0570	0,7151	0,0546	0,7235	0,0522
	CLR	0,7209	0,0557	0,7128	0,0541	0,7197	0,0493
	HOMER	—	—	—	—	—	—
	LP	0,7151	0,0552	0,7098	0,0532	0,7108	0,0476
	ML k NN	0,7249	0,0565	0,7159	0,0554	0,7255	0,0523
	Ra k EL	—	—	—	—	—	—

Tabela 23 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de predições do VGG16.fc7+ImageNet.

Áudio	Fusão com média de predições do LBP	Média		Produto		Soma	
		média	σ	média	σ	média	σ
BR	BR	0,6880	0,0497	0,6578	0,0459	0,6879	0,0507
	BR k NN	0,6968	0,0439	0,6777	0,0399	0,6971	0,0432
	CLR	0,6957	0,0431	0,6763	0,0386	0,6959	0,0415
	HOMER	0,6937	0,0421	0,6744	0,0390	0,6933	0,0441
	LP	0,6794	0,0409	0,6668	0,0384	0,6762	0,0393
	ML k NN	0,6966	0,0431	0,6783	0,0399	0,6967	0,0420
	Ra k EL	0,6965	0,0441	0,6778	0,0404	0,6970	0,0434
BR k NN	BR	0,7055	0,0531	0,6767	0,0462	0,7030	0,0540
	BR k NN	0,7230	0,0569	0,7216	0,0570	0,7198	0,0527
	CLR	0,7226	0,0575	0,7210	0,0572	0,7199	0,0535
	HOMER	0,7210	0,0574	0,7195	0,0583	0,7173	0,0542
	LP	0,7154	0,0603	0,7133	0,0601	0,7089	0,0567
	ML k NN	0,7230	0,0565	0,7221	0,0567	0,7194	0,0515
	Ra k EL	0,7231	0,0588	0,7218	0,0588	0,7202	0,0559
CLR	BR	0,7168	0,0601	0,6789	0,0523	0,7183	0,0571
	BR k NN	0,7328	0,0619	0,7341	0,0630	0,7387	0,0561
	CLR	0,7336	0,0610	0,7339	0,0623	0,7385	0,0551
	HOMER	0,7318	0,0611	0,7323	0,0624	0,7350	0,0552
	LP	0,7282	0,0651	0,7279	0,0661	0,7315	0,0588
	ML k NN	0,7333	0,0611	0,7346	0,0623	0,7388	0,0550
	Ra k EL	0,7339	0,0635	0,7348	0,0652	0,7393	0,0572
HOMER	BR	0,6921	0,0529	0,6603	0,0458	0,6917	0,0544
	BR k NN	0,7026	0,0485	0,6820	0,0419	0,6992	0,0499
	CLR	0,7034	0,0489	0,6821	0,0440	0,6998	0,0499
	HOMER	0,7001	0,0491	0,6780	0,0431	0,6967	0,0515
	LP	0,6908	0,0487	0,6741	0,0454	0,6830	0,0509
	ML k NN	0,7030	0,0478	0,6825	0,0418	0,6998	0,0488
	Ra k EL	0,7047	0,0497	0,6820	0,0442	0,7005	0,0501
LP	BR	0,6575	0,0668	0,6596	0,0589	0,6553	0,0622
	BR k NN	0,6770	0,0565	0,6762	0,0558	0,6771	0,0508
	CLR	0,6797	0,0553	0,6806	0,0552	0,6760	0,0503
	HOMER	0,6757	0,0541	0,6752	0,0531	0,6741	0,0500
	LP	0,6264	0,0650	0,6264	0,0635	0,6323	0,0582
	ML k NN	0,6773	0,0538	0,6773	0,0540	0,6756	0,0491
	Ra k EL	0,6785	0,0612	0,6795	0,0608	0,6784	0,0555
ML k NN	BR	0,7037	0,0552	0,6756	0,0507	0,7031	0,0553
	BR k NN	0,7151	0,0536	0,7153	0,0532	0,7179	0,0519
	CLR	0,7152	0,0538	0,7152	0,0542	0,7166	0,0516
	HOMER	0,7137	0,0527	0,7133	0,0522	0,7138	0,0535
	LP	0,7086	0,0556	0,7074	0,0560	0,7090	0,0520
	ML k NN	0,7157	0,0527	0,7157	0,0526	0,7180	0,0507
	Ra k EL	0,7168	0,0558	0,7166	0,0558	0,7180	0,0548
Ra k EL	BR	0,7056	0,0615	0,6707	0,0590	0,7052	0,0576
	BR k NN	0,7216	0,0568	0,7137	0,0550	0,7242	0,0509
	CLR	0,7214	0,0561	0,7141	0,0537	0,7209	0,0510
	HOMER	0,7193	0,0563	0,7112	0,0540	0,7204	0,0510
	LP	0,7093	0,0567	0,7033	0,0553	0,7070	0,0493
	ML k NN	0,7216	0,0559	0,7146	0,0542	0,7239	0,0501
	Ra k EL	0,7219	0,0580	0,7155	0,0561	0,7234	0,0521

Tabela 24 – Resultados obtidos através da métrica Micro-averaged AUC para fusão entre as bases de Áudio e média de predições do LBP.

Analisando os resultados apresentados na tabela 15 podemos notar que os melhores resultados obtidos (combinando predições auditivas replicadas e predições de Histograma de Cores) são os resultados provindos da utilização do classificador CLR para o Áudio através do método de fusão tardia de Soma, sendo que quando combinadas com o classificador CLR para o Histograma de Cores se obtém o melhor resultado da tabela (72,23%), o qual é inferior ao melhor resultado da utilização de base isolada de áudio (72,44%) e superior ao melhor resultado da utilização da base isolada de Histograma de Cores (66,04%) que são apresentados nas tabelas 14 e 13 respectivamente. Comparando os métodos de fusão tardia (média, produto e soma) podemos notar que o método de média apresenta os melhores resultados quando comparado com os demais métodos, porém os melhores resultados ainda assim foram obtidos através do método de soma. Um dado importante a se ressaltar é que durante a primeira etapa os melhores resultados para a classificação utilizando características de Histograma de Cores e a classificação utilizando a replicação das predições auditivas foram obtido através do classificador CLR, o qual obteve também o melhor resultado para a fusão tardia.

A análise dos resultados apresentados na tabela 16 mostra que o melhor resultado obtido (quando se combina predições de Áudio replicadas e predições de Diferença Acumulada) foi de 72,29%, o qual foi obtido usando uma combinação do classificador CLR com características de Áudio e o classificador CLR com características de Diferença Acumulada. Assim como na classificação utilizando características de Histograma de Cores, os melhores resultados foram obtidos através do uso do classificador CLR para áudio e do método de fusão tardia da soma. Comparando os métodos de fusão tardia podemos notar que assim como na tabela 15 o método da média se destaca, porém os melhores resultados são apresentados através do método de soma. Ao comparar com os resultados obtidos na primeira etapa, podemos notar que as taxas de acerto foram inferiores para áudio replicado (72,44%) e superiores para Diferença Acumulada (66,22%) os quais são apresentados nas tabelas 14 e 13 respectivamente.

Analisando os resultados apresentados na tabela 17, podemos destacar os resultados obtidos combinando a classificação pelo classificador CLR utilizando características de Áudio que posteriormente sofre uma replicação de predições com a classificação pelo classificador BR utilizando características de Fluxo Óptico, que foi o melhor resultado obtido durante segundo experimento utilizando a replicação de predições auditivas (73,59%). Podemos destacar também que a quantidade de características e atributos para este método é muito inferior as características e atributos de outros métodos de extração de informações visuais (apresentados na tabela 11). Ao contrário dos resultados apresentados nas tabelas 15 e 16, os resultados obtidos foram melhores que os resultados obtidos através da utilização de bases isoladas.

Os resultados apresentados na tabela 18 mostram que os resultados obtidos

através da fusão de predições auditivas de forma replicada com predições visuais de VGG16.fc7+ImageNet (72,24%) apresentam assim como na fusão de predições auditivas replicadas com predições de Histograma de Cores e predições de Diferença Acumulada, os resultados são inferiores ao uso de características isoladas auditivas (72,44%) e superiores a utilização de características isoladas visuais de VGG16.fc7+ImageNet (66,19%). Vale ressaltar que dentre os métodos de extração de características visuais, o VGG16.fc7+ImageNet apresenta maior quantidade de características e atributos (tabela 11), mostrando assim que o processo de extração de características utilizando uma rede pré-treinada nem sempre é a melhor opção quando se trata de classificação multi-rótulo de emoções.

Os resultados apresentados na tabela 19 mostram que os resultados obtidos através da fusão de predições auditivas de forma replicada com predições visuais de LBP (72,26%) apresentam assim como nas fusões de predições anteriores (com exceção a fusão utilizando predições provindas de características de Fluxo Óptico), os resultados são inferiores ao uso de características isoladas auditivas (72,44%) e superiores a utilização de características isoladas visuais de LBP (66,27%).

Ao comparar os resultados das tabelas de fusão tardia utilizando a replicação de predições auditivas e as predições visuais (tabelas 15, 16, 17, 18 e 19) podemos ressaltar os resultados obtidos através do uso do classificador CLR, o qual apresentou os melhores resultados tanto no uso de características visuais de forma isolada quanto na fusão tardia de predições auditivas replicadas com predições visuais. Apesar de estatisticamente os resultados não apresentarem variações significativas, podemos notar uma leve melhoria quando utilizada a combinação de predições auditivas replicadas com predições visuais de Fluxo Óptico.

Analisando os resultados apresentados na tabela 20 podemos verificar que os resultados obtidos através da fusão tardia entre as predições auditivas e da média de predições visuais de Histograma de Cores foram superiores aos resultados obtidos através da classificação utilizando apenas características auditivas de forma isolada (72,44%) e superiores aos resultados obtidos através da classificação utilizando apenas características visuais de Histograma de Cores de forma isolada (68,82%), comparando com os resultados da tabela 15 podemos notar uma melhoria em sua melhor avaliação. Com base nestes dados é possível afirmar que o uso da predição média entre os frames visuais combinados com as predições auditivas apresentou melhores resultados (73,64%) quando comparado com a utilização da combinação das predições auditivas de forma replicada combinadas com as predições visuais (72,23%).

A tabela 21 apresenta os resultados da fusão tardia entre as predições auditivas e a média de predições visuais de Diferença Acumulada, dos quais o melhor resultado foi apresentado através da combinação da classificação utilizando o classificador CLR para as características auditivas e da média de predições obtidas através do classificador MLkNN

utilizando informações visuais de Diferença Acumulada. Quando realizada a comparação entre as tabelas 21 e 16 podemos notar que assim como no caso do Histograma de Cores, os resultados obtidos através da realização da média de predições foi superior aos resultados utilizando a replicação do áudio (72,29%) e também foi superior quando comparado com as classificações utilizando características de forma isolada.

A tabela 22 apresenta os resultados da fusão tardia entre as predições auditivas e a média de predições visuais de Fluxo Óptico, dos quais o melhor resultado foi apresentado através da combinação da classificação utilizando o classificador CLR para as características auditivas e da média de predições obtidas através do classificador RakEL utilizando informações visuais de Fluxo Óptico (74,40%). Devemos ressaltar que o melhor resultado apresentado nesta tabela é também o melhor resultado apresentado em todos os experimentos (tanto de forma isolada quanto através das fusões tardias).

Os resultados apresentados na tabela 23 se referem a combinação de predições auditivas com a média de predições visuais de VGG16.fc7+ImageNet, onde o melhor resultado obtido foi através da combinação de predições auditivas obtidas através do classificador CLR com a média de predições obtidas através do classificador BKkNN utilizando características extraídas através do VGG16.fc7+ImageNet (74,24%).

Analisando os resultados da tabela 24 podemos ressaltar o resultado obtido através da combinação das predições auditivas provindas do classificador CLR com a média de predições visuais obtidas através do classificador RakEL (73,93%). Os quais assim como os demais casos em que se utilizou a média de predições visuais, foi melhor do que a classificação utilizando características de forma isolada e também melhores que os resultados obtidos através da fusão tardia com base na replicação de informações auditivas.

Com base em todas as tabelas as quais apresentam resultados oriundos da utilização da média de predições visuais (tabelas 20, 21, 22, 23 e 24) podemos concluir que o classificador CLR combinado com informações auditivas ainda é o mais relevante, e o classificador RakEL foi o que apresentou melhores resultados para estes casos em específico.

Quando comparamos todos os resultados obtidos no segundo experimento, podemos notar que a utilização do Fluxo Óptico para a classificação de vídeos musicais é o mais relevante entre os demais métodos de extração de características visuais utilizados neste trabalho. Podemos ressaltar também que apesar dos resultados de combinação serem estatisticamente iguais aos resultados da classificação utilizando as bases de forma isolada, houve uma pequena melhora nos resultados. Com base nos experimentos podemos responder também as questões anteriormente levantadas: respondendo a questão 4: Sim e não, após realizar a combinação de informações visuais com informações auditivas através do método de fusão tardia, quando combinadas utilizando a replicação das predições auditivas os resultados foram inferiores aos resultados obtidos através do uso de apenas informações auditivas de forma isolada, porém ao realizar a combinação com a média de predições

visuais os resultados obtidos foram superiores aos utilizando apenas informações auditivas de forma isolada, onde podemos ressaltar o melhor resultado obtido durante o experimento apresentado na tabela 22 (74,40%); respondendo a questão 5: Sim, como as classificações utilizando características auditivas de forma isolada apresentam melhores predições quando comparado a classificações utilizando características visuais de forma isolada, a combinação dos dois maximizou os resultados quando se comparado a classificação utilizando apenas características visuais de forma isolada.

6 Conclusão

Neste trabalho foi realizado um estudo para classificar vídeos musicais por emoção em ambiente multi-rótulo. Para a classificação multi-rótulo foram utilizadas características visuais e auditivas. Para extrair as características visuais foram utilizados as imagens (frames) dos vídeos musicais. As características auditivas também foram extraídas a partir dos vídeos musicais. Para realizar a classificação multi-rótulo de emoções em vídeos musicais foi expandida a base CAL500, que possui originalmente apenas músicas com 18 rótulos de emoção.

Como contribuição científica o trabalho apresenta o impacto do uso de informações visuais no processo de classificação de emoções em vídeos musicais, comparando a classificação utilizando apenas o áudio com classificações utilizando características visuais isoladas (seção 5.2). Também foi realizada a combinação de classificações utilizando características visuais e auditivas (seção 5.3).

Os experimentos (descritos no capítulo 5) permitiram responder as questões levantadas na seção 1.3: Q1 - A utilização de informações visuais é relevante no processo de classificação de vídeos musicais?; Q2 - A classificação de vídeos musicais por emoção utilizando apenas características visuais apresenta resultados relevantes?; Q3 - Métodos mais robustos de extração de características visuais são melhores do que métodos simples?; Q4 - A combinação de informações visuais combinadas com as informações auditivas melhora as predições em relação ao uso de apenas informações auditivas?; Q5 - A combinação de informações visuais combinadas com as informações auditivas melhora as predições em relação ao uso de apenas informações visuais?; e Q6 - Quais os melhores classificadores para a classificação de emoções em vídeos musicais no ambiente multi-rótulo? Nos permitiram também atingir o objetivo desejado (descrito na seção 1.2), que é realizar a classificação multi-rótulo de emoções em vídeos musicais utilizando características visuais e auditivas, mostrando que a utilização de informações visuais possui grande relevância no processo de classificação de emoções em vídeos musicais.

Através dos experimentos realizados (seções 5.2 e 5.3) foi possível responder as questões de pesquisa Q1, Q2, Q3, Q4 e Q5. Após a análise dos resultados foi possível concluir que as informações visuais possuem grande importância no processo de classificação de emoções em vídeos musicais, possibilitando realizar as classificações utilizando apenas informações visuais e atingir resultados relevantes.

Os experimentos utilizando a combinação de características visuais e auditivas apresentam melhora nas classificações quando comparados com os experimentos utilizando apenas características visuais ou auditivas. Dentre todos os experimentos realizados, os

melhores resultados foram obtidos utilizando a combinação de características auditivas e características visuais de fluxo óptico (74.40%). Com base nesse resultado é possível destacar a importância das informações de movimento para classificação de emoções em vídeos musicais.

Em relação a Q6, quais os melhores classificadores para classificar emoções em ambiente multi-rótulo? Não foi possível destacar um classificador em especial, pois em diversos momentos os melhores resultados foram obtidos através de classificadores diferentes.

Como trabalho futuro pretende-se criar bases de vídeos musicais rotuladas por emoção por ouvintes humanos. Outra possibilidade é fragmentar os vídeos e definir as emoções presentes ao longo do tempo. Além disso pretende-se explorar outras formas de extrair características, selecionar as imagens dos vídeos musicais e fusão de classificadores.

Referências

- BRADSKI, G.; ; KAEHLER, A. Opencv. *Dr. Dobb's journal of software tools*, 2000. Citado na página 32.
- CHEN, O. T.-C.; GU, J. J.; CHEN, C.-C.; LU, P.-T. Affective perception of musical television. In: *Proceedings of the 4th International Conference on Awareness Science and Technology (iCAST)*. [S.l.: s.n.], 2012. p. 104–108. Citado 3 vezes nas páginas 48, 49 e 50.
- COSTA, Y. M.; OLIVEIRA, L.; KOERICH, A. L.; GOUYON, F.; MARTINS, J. Music genre classification using lbp textural features. *Signal Processing*, Elsevier, v. 92, n. 11, p. 2723–2737, 2012. Citado na página 35.
- DENG, J.; ; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2009. p. 248–255. Citado na página 34.
- DUXBURY, C.; BELLO, J. P.; DAVIES, M.; SANDLER, M. et al. Complex domain onset detection for musical signals. In: QUEEN MARY UNIVERSITY LONDON. *Proceedings Digital Audio Effects Workshop (DAFx)*. [S.l.], 2003. v. 1, p. 6–9. Citado na página 27.
- ERONEN, A. Automatic musical instrument recognition. *Mémoire de DEA, Tempere University of Technology*, p. 178, 2001. Citado na página 26.
- ESSID, S. *Classification automatique des signaux audio-fréquences: reconnaissance des instruments de musique*. Tese (Doutorado) — Université Pierre et Marie Curie-Paris VI, 2005. Citado na página 31.
- FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006. Citado na página 63.
- FÜRNKRANZ, J.; HÜLLERMEIER, E.; MENCÍA, E. L.; BRINKER, K. Multilabel classification via calibrated label ranking. *Machine learning*, Springer, v. 73, n. 2, p. 133–153, 2008. Citado na página 38.
- GOUYON, F.; PACHET, F.; DELERUE, O. et al. On the use of zero-crossing rate for an application of classification of percussive sounds. In: *Proceedings of the Co-Operation and Scientific Transfer (COST) G-6 conference on Digital Audio Effects (DAFX-00)*, Verona, Italy. [S.l.: s.n.], 2000. p. 26. Citado na página 31.
- HATAMIKIA, S.; ; NASRABADI, A. M. Recognition of emotional states induced by music videos based on nonlinear feature extraction and som classification. In: *Proceedings of the 21th Iranian Conference on Biomedical Engineering (ICBME)*. [S.l.: s.n.], 2014. p. 333–337. Citado 4 vezes nas páginas 21, 46, 49 e 50.
- HORN, B. K.; ; SCHUNCK, B. G. Determining optical flow. *Artificial intelligence*, Elsevier, v. 17, n. 1-3, p. 185–203, 1981. Citado na página 34.
- KERAS. *Keras: The Python Deep Learning library*. Disponível em: <<https://keras.io/>>. Acesso em: 01 jun. 2019. Citado na página 32.

- KETKAR, N. Introduction to keras. In: *Deep Learning with Python*. [S.l.]: Springer, 2017. p. 97–111. Citado na página 32.
- KOELSTRA, S.; MUHL, C.; SOLEYMANI, M.; LEE, J.-S.; YAZDANI, A.; EBRAHIMI, T.; PUN, T.; NIJHOLT, A.; PATRAS, I. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, IEEE, v. 3, n. 1, p. 18–31, 2012. Citado na página 51.
- LIU, C.; ; CAO, L. A coupled k-nearest neighbor algorithm for multi-label classification. In: SPRINGER. *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*. [S.l.], 2015. p. 176–187. Citado na página 40.
- MATHIEU, B.; ESSID, S.; FILLON, T.; PRADO, J.; RICHARD, G. Yaafe, an easy to use and efficient audio feature extraction software. In: *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*. [S.l.: s.n.], 2010. p. 441–446. Citado na página 25.
- MOORE, B. C.; GLASBERG, B. R.; BAER, T. A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, Audio Engineering Society, v. 45, n. 4, p. 224–240, 1997. Citado na página 29.
- MTV. *MTV database*. Disponível em: <<http://www.mtv.co.uk/database/videos>>. Acesso em: 01 jun. 2019. Citado na página 52.
- NETO, V. d. O.; ; GOMES, D. Comparação de métodos para localização de fluxo óptico em sequências de imagens. *Programa de Pós-Graduação em Ciência da Computação*, 2011. Citado na página 34.
- NOROOZI, F.; MARJANOVIC, M.; NJEGUS, A.; ESCALERA, S.; ANBARJAFARI, G. Audio-visual emotion recognition in video clips. *IEEE Transactions on Affective Computing*, IEEE, 2017. Citado 2 vezes nas páginas 46 e 50.
- OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE, n. 7, p. 971–987, 2002. Citado na página 35.
- OPENCV. *OpenCV Library*. Disponível em: <<https://opencv.org/>>. Acesso em: 01 jun. 2019. Citado na página 32.
- PEETERS, G. A large set of audio features for sound description (similarity and classification) in the cuidado project. *CUIDADO IST Project Report*, v. 54, n. 0, p. 1–25, 2004. Citado 2 vezes nas páginas 31 e 32.
- PEETERS, G.; GIORDANO, B. L.; SUSINI, P.; MISDARIIS, N.; MCADAMS, S. The timbre toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, ASA, v. 130, n. 5, p. 2902–2916, 2011. Citado na página 30.
- SENG, K.; ANG, L.-M.; OOI, C. A combined rule-based and machine learning audio-visual emotion recognition approach. *IEEE Transactions on Affective Computing*, IEEE, 2016. Citado 2 vezes nas páginas 46 e 50.
- SIMONYAN, K.; ; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 34.

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*. [S.l.: s.n.], 2005. p. 399–402. Citado na página 41.

SPYROMITROS, E.; TSOUMAKAS, G.; VLAHAVAS, I. An empirical study of lazy multilabel classification algorithms. In: *Proceedings of the Hellenic conference on artificial intelligence*. [S.l.: s.n.], 2008. p. 401–406. Citado na página 37.

THOMAS, T.; DOMÍNGUEZ, M.; PTUCHA, R. Deep independent audio-visual affect analysis. In: *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. [S.l.: s.n.], 2017. p. 1417–1421. Citado 3 vezes nas páginas 48, 49 e 50.

TROHIDIS, K.; TSOUMAKAS, G.; KALLIRIS, G.; VLAHAVAS, I. P. Multi-label classification of music into emotions. In: *Proceedings of the International Society for Music Information Retrieval (ISMIR)*. [S.l.: s.n.], 2008. v. 8, p. 325–330. Citado 3 vezes nas páginas 21, 45 e 50.

TSOUMAKAS, G.; KATAKIS, I.; VLAHAVAS, I. Effective and efficient multilabel classification in domains with large number of labels. In: *Proceedings of the ECML/PKDD 2008 Workshop on Mining Multidimensional Data (MMD'08)*. [S.l.: s.n.], 2008. v. 21, p. 53–59. Citado 2 vezes nas páginas 38 e 39.

TSOUMAKAS, G.; KATAKIS, I.; VLAHAVAS, I. Mining multi-label data. In: *Data mining and knowledge discovery handbook*. [S.l.]: Springer, 2009. p. 667–685. Citado 2 vezes nas páginas 36 e 39.

TSOUMAKAS, G.; VLAHAVAS, I. Random k-labelsets: An ensemble method for multilabel classification. In: *European conference on machine learning*. [S.l.: s.n.], 2007. p. 406–417. Citado na página 40.

TURNBULL, D.; BARRINGTON, L.; TORRES, D.; LANCKRIET, G. Towards musical query-by-semantic-description using the cal500 data set. In: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. [S.l.: s.n.], 2007. p. 439–446. Citado na página 51.

VEVO. VEVO. Disponível em: <<https://www.vevo.com/>>. Acesso em: 01 jul. 2019. Citado na página 52.

VOZÁRIKOVÁ, E.; JUHÁR, J.; ČIŽMÁR, A. Acoustic events detection using mfcc and mpeg-7 descriptors. In: *Proceedings of the International Conference on Multimedia Communications, Services and Security*. [S.l.: s.n.], 2011. p. 191–197. Citado na página 30.

WANG, S.; WANG, J.; WANG, Z.; JI, Q. Multiple emotion tagging for multimedia data by exploiting high-order dependencies among emotions. *IEEE Transactions on Multimedia*, IEEE, v. 17, n. 12, p. 2185–2197, 2015. Citado 3 vezes nas páginas 47, 49 e 50.

WANG, Z.; WANG, S.; HE, M.; LIU, Z.; JI, Q. Emotional tagging of videos by exploring multiple emotions' coexistence. In: *Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. [S.l.: s.n.], 2013. p. 1–6. Citado 3 vezes nas páginas 47, 49 e 50.

WU, X.-Z.; ZHOU, Z.-H. A unified view of multi-label performance measures. In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. [S.l.: s.n.], 2017. p. 3780–3788. Citado na página 63.

YAAFE. *Yaafe - audio features extraction*. Disponível em: <<http://yaafe.sourceforge.net/>>. Acesso em: 01 jun. 2019. Citado na página 25.

ZHANG, M.-L.; ZHOU, Z.-H. Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, Elsevier, v. 40, n. 7, p. 2038–2048, 2007. Citado na página 40.

Apêndices

APÊNDICE A – Base CAL500 - Coleta de vídeos musicais

Id	Artista	Composição	Tipo de Vídeo	Fonte
1	10cc	For You And I	Não possui vídeo	
2	2Pac	Trapped	Oficial ¹	Youtube
3	Fifth Dimension	One Less Bell To Answer	Apresentação	Youtube
4	A Tribe Called Quest	Bonita Applebum	Oficial	Vevo
5	Aaron Neville	Tell It Like It Is	Oficial	Vevo
6	ABBA	SOS	Oficial	Vevo
7	ABC	Poison Arrow	Oficial	Vevo
8	AD/DC	Dirty Deeds Done Dirt Cheap	Apresentação	Vevo
9	Adam & The Ants	Prince Charming	Oficial	Vevo
10	Adam Ant	Wonderful	Oficial	MTV
11	Adverts	Gary Gilmore's Eyes	Não Oficial	Youtube
12	Aerobic Jonquil	Sweat Machine	Não possui vídeo	
13	Aerosmith	Dude (Looks Like A Lady)	Oficial	Vevo
14	Aimee Mann	Wise Up	Oficial ²	Vimeo
15	Air	Sexy Boy	Oficial	MTV
16	Al Green	Sha-la-la (Make me happy)	Apresentação	Youtube
17	Alanis Morissette	Thank U	Oficial	MTV
18	Alice Cooper	Elected	Não Oficial	Youtube

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
19	Alice In Chains	No Excuses	Oficial	Vevo
20	Alicia Keys	Fallin'	Oficial	Vevo
21	Allman Brothers Band	Melissa	Apresentação	Youtube
22	Altered Images	Don't Talk to Me About Love	Oficial	Vevo
23	American Music Club	Jesus Hands	Apresentação	Youtube
24	Andrews Sisters	Boogie Woogie Bugle Boy	Não Oficial	Youtube
25	Ani DiFranco	Crime for Crime	Apresentação	Youtube
26	The Animals	I'm Crying	Apresentação	Youtube
27	Anita Baker	Caught Up In The Rupture	Apresentação	MTV
28	Antiguru	Peering	Não possui vídeo	
29	Antonio Carlos Jobim	Wave	Não Oficial	Youtube
30	Anup	Life Glides	Não possui vídeo	
31	Aphex Twin	Come to Daddy	Oficial	MTV
32	Apples in stereo	Glowworm	Não Oficial	Youtube
33	Aretha Franklin	Don't Play That Song	Apresentação	Youtube
34	Arlo Guthrie	Alices Restaurant Massacre	Não Oficial	Youtube
35	Art Tatum	Weep For Me	Não possui vídeo	
36	Artemis	Don't Look Back	Não possui vídeo	
37	Arthur Alexander	You Don't Care	Não possui vídeo	
38	Arthur Yoria	At Least You've Been Told	Não possui vídeo	
39	Ashford & Simpson	Solid	Oficial	Vevo
40	Association	Windy	Apresentação	Youtube

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
41	Atomic Opera	Water Grave	Não possui vídeo	
42	B.B. King	Sweet Little Angel	Apresentação	Vevo
43	Backstreet Boys	As Long As You Love Me	Oficial	Vevo
44	Badly Drawn Boy	All Possibilities	Oficial	MTV
45	The Band	King Harvest (Has Surely Come)	Não Oficial	Youtube
46	Barbara Leoni	Don't Rain On My Parade	Não possui vídeo	
47	Barenaked Ladies	It's All Been Done	Oficial	Vevo
48	Barry Manilow	Mandy	Apresentação	UOL
49	Barry White	Can't get enough of your Love, Babe	Apresentação	Youtube
50	Bauhaus	Ziggy Stardust	Não Oficial	Youtube
51	Bay City Rollers	Saturday Night	Não Oficial	Youtube
52	The Beach Boys	I Get Around	Não Oficial	Youtube
53	The Beatles	Strawberry Fields Forever	Oficial	Vevo
54	Beatundercontrol	This is Beatundercontrol	Não possui vídeo	
55	The Beautiful South	One Last Love Song	Oficial	MTV
56	Beck	Where It's At	Oficial	Vevo
57	Bee Gees	Stayin' Alive	Oficial	Vevo
58	Belief Systems	Skunk Werks	Não possui vídeo	
59	Belle & Sebastian	Like Dylan in the Movies	Não Oficial	Youtube
60	Ben Folds Five	Brick	Oficial	Vevo
61	Beth Quist	Survival	Não possui vídeo	
62	Big Star	In The Street	Apresentação	Youtube

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
63	Billie Holiday & Tony Bennett	God Bless The Child	Oficial	Vevo
64	Billy Bragg	Jeane	Apresentação	Youtube
65	Billy Joel	We Didn't Start the Fire	Oficial	Vevo
66	Birthday Party	Mr. Clarinet	Não possui vídeo	
67	Björk	Army Of Me	Oficial	Vevo
68	The Black Crowes	Thorn in my Pride	Apresentação	Youtube
69	Black Flag	Six Pack	Apresentação	Youtube
70	Black Sabbath	Black Sabbath	Apresentação	MTV
71	Blind Faith	Had to cry Today	Apresentação	Youtube
72	Blind Melon	No Rain	Não Oficial	Youtube
73	Blondie	Rupture	Oficial	Vevo
74	Blood Sweat	Sometimes in Winter	Apresentação	Youtube
75	Blue Öyster Cult	Burnin' for You	Apresentação	Youtube
76	Blur	Country House	Oficial	MTV
77	Bo Diddley	You Can't Judge A Book By The Cover	Apresentação	Youtube
78	Bob Dylan	I'll Be Your Baby Tonight	Apresentação	Youtube
79	Bob Marley	Three Little Birds	Oficial ³	Youtube
80	Bob Seger	Turn The Page	Não Oficial	Vimeo
81	Bobby Brown	My Prerogative	Apresentação	Vevo
82	Bobby Fuller Four	I Fought The Law	Apresentação	Youtube
83	Bobby Womack	Woman's Gotta Have It	Apresentação	Youtube
84	Bomb The Bass	Bug Powder Dust	Oficial	MTV

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
85	Bon Jovi	Livin' On A Prayer	Oficial	Vevo
86	Bonnie Tyler	Total Eclipse of the Heart	Oficial	Vevo
87	The Boo Radleys	Wake up Boo!	Oficial	Vevo
88	Boogie Down Productions	The Bridge is Over	Não Oficial	Youtube
89	Booker T. & The MG's	Time Is Tight	Não Oficial	Youtube
90	Boston	More Than a Feeling	Apresentação	Vevo
91	Bots	Take the Power Back	Não possui vídeo	
92	Brad Sucks	Overreacting	Não Oficial	Youtube
93	Bread	If	Apresentação	Youtube
94	The Breeders	Cannonball	Não Oficial	Youtube
95	Brenton Woods	Lovey Dovey Kind of Love	Apresentação	Youtube
96	Brian Eno	Here Come the Warm Jets	Não Oficial	Youtube
97	Britney Spears	I'm A Slave 4 U	Oficial	Vevo
98	Bruce Springsteen	Badlands	Apresentação	Vimeo
99	Bryan Adams	Cuts Like a Knife	Oficial	MTV
100	Buddy Holly	Peggy Sue	Não Oficial	Youtube
101	Buena Vista Social Club	El Cuarto de Tula	Não Oficial	Youtube
102	Buffalo Springfield	For What it's Worth	Apresentação	Youtube
103	Buffalo Springfield	Mr Soul	Apresentação	Youtube
104	The Buggles	Video Killed The Radio Star	Oficial	Vevo
105	Built to Spill	I Would Hurt A Fly	Apresentação	Youtube
106	Burnshee Thornsides	Goodbye On a Beautiful Day	Não possui vídeo	

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
107	Bush	Comedown	Oficial	Vevo
108	Busta Rhymes	Woo-Hah! Got You All In Check	Oficial	MTV
109	Buzzcocks	Everybody's Happy Nowadays	Apresentação	Youtube
110	Byrds	Wasn't Born To Follow	Não Oficial	Youtube
111	C. Layne	Devil and The Woman	Não possui vídeo	
112	Cab Calloway	Minnie the Moocher	Apresentação	Youtube
113	Cake	Perhaps, Perhaps, Perhaps	Não Oficial	Youtube
114	Camper Van Beethoven	Take The Skinheads Bowling	Não Oficial	Youtube
115	Canned Heat	On the Road Again	Apresentação	Youtube
116	Captain Beefheart	Safe As Milk	Apresentação	Youtube
117	The Cardigans	Lovefool	Oficial	Vevo
118	Cargo Cult	Garden	Não possui vídeo	
119	Carl Perkins	Matchbox	Apresentação	Vevo
120	Carly Simon	You're So Vain	Não Oficial	Vimeo
121	Carole King	You've Got A Friend In Me	Apresentação	Vevo
122	Carpenters	Rainy Days And Mondays	Oficial	Vevo
123	The Cars	Good Times Roll	Apresentação	Youtube
124	Cat Power	He War	Não Oficial	Youtube
125	Catherine Wheel	Black Metallic	Apresentação	Youtube
126	C+C Music Factory	Gonna Make You Sweat (Everybody Dance Now)	Oficial	Vevo
127	Chad & Jeremy	Before and After	Apresentação	Youtube
128	Chantal Kreviazuk	Surrounded	Oficial	Vevo

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
129	Charles Mingus	Mood Indigo	Não possui vídeo	
130	Charlie Parker	Ornithology	Não possui vídeo	
131	Charlie Rich	Behind Closed Doors	Apresentação	Vevo
132	Cheap Trick	Dream Police	Oficial	Vevo
133	The Chemical Brothers	Setting Sun	Oficial	Vevo
134	Cheryl ann Fulton	Marsh of Rhuddlan	Não possui vídeo	
135	Chet Baker	These foolish Things	Não possui vídeo	
136	Chi-Lates	Stoned Out of My Mind	Não possui vídeo	
137	Chic	Le Freak	Oficial	MTV
138	Chicago	If You Leave Me Now	Oficial	MTV
139	The Chills	I Love My Leather Jacket	Não Oficial	Youtube
140	Chris Juergensen	Prospects	Não possui vídeo	
141	Christina Aguilera	Genie In A Bottle	Oficial	Vevo
142	Chuck Berry	Roll Over Beethoven	Apresentação	Youtube
143	Chumbawamba	Tubthumping	Oficial	Vevo
144	The Church	Under The Milky Way	Oficial	Vevo
145	Cilla Black	Alfie	Apresentação	Vevo
146	Clarence Ashley	The House Carpenter	Não possui vídeo	
147	The Clash	Lost In The Supermarket	Não Oficial	Youtube
148	The Coasters	Poison Ivy	Apresentação	Youtube
149	Coldplay	Clocks	Oficial	Vevo
150	The Contours	Do You Love Me	Não Oficial	Youtube

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
151	Count Basie	Lester Leaps In	Não possui vídeo	
152	Counting Crows	Speedway	Não Oficial	Youtube
153	Cowboy Junkies	Postcard Blues	Não possui vídeo	
154	The Cranberries	Linger	Oficial	Vevo
155	Cream	Tales Of Brave Ulysses	Não Oficial	Dailymotion
156	Creedence Clearwater	Revival Travelin Band	Apresentação	Youtube
157	Crosby, Stills & Nash	Guinnevere	Apresentação	Youtube
158	Crosby, Stills & Nash	Teach Your Children	Artista Duplicado	Youtube
159	Curandero	Aras	Não possui vídeo	
160	The Cure	Just Like Heaven	Oficial	Vevo
161	Curtis Mayfield	Move On Up	Apresentação	Youtube
162	Cyndi Lauper	Money Changes Everything	Apresentação	Vevo
163	Cypress Hill	Lowrider	Oficial	Vevo
164	Daft Punk	Da Funk	Oficial	MTV
165	The Darkness	I Believe In A Thing Called Love	Apresentação	Vevo
166	Dave Matthews Band	Ants Marching	Apresentação	Youtube
167	David Bowie	Song for Bob Dylan	Apresentação	Youtube
168	De La Soul	Eye Know	Oficial	Youtube
169	Dead Kennedys	Chemical Warfare	Apresentação	Youtube
170	Def Leppard	Pour Some Sugar On Me	Oficial	Vevo
171	Dennis Brown	Tribulation	Não possui vídeo	
172	Depeche Mode	World In My Eyes	Apresentação	Vevo

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
173	Devo	Girl U Want	Apresentação	Vevo
174	Diana Ross and the Supremes	Where did our Love go	Apresentação	Youtube
175	Dido	Here with Me	Oficial	Vevo
176	Dione Warwick	Walk On By	Apresentação	Youtube
177	Dire Straits	Money For Nothing	Oficial	Vevo
178	DJ Jazzy Jeff & The Fresh Prince	Summertime	Oficial	Vevo
179	DJ Markitos	Sunset 138 bpm remix	Não possui vídeo	
180	Django Reinhardt	Brazil	Não possui vídeo	
181	Domased	Too Slow	Não possui vídeo	
182	Donovan	Catch The Wind	Apresentação	Youtube
183	The Doobie Brothers	China Grove	Apresentação	Vevo
184	The Doors	Touch Me	Apresentação	Vevo
185	Dr. Dre Featuring Snoop Dogg	Nothin' But a G Thang	Não Oficial	Youtube
186	Dr Kuch	Persuaders Dub	Não possui vídeo	
187	Drevo	Our Watcher Apresentação us the way	Não possui vídeo	
188	Drop Trio	Slap Jack	Apresentação	Vimeo
189	Duke Ellington and His Orchestra	Caravan	Apresentação	Youtube
190	Duncan Sheik	Barely Breathing	Oficial	Vevo
191	Duran Duran	Come Undone	Apresentação	Youtube
192	Eagles	Tequila Sunrise	Apresentação	Youtube
193	Earth, Wind & Fire	September	Oficial	Vevo
194	Eels	Cancer For The Cure	Oficial	Vevo

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
195	Electric Frankenstein	Teenage Shutdown	Não possui vídeo	
196	Elliott Smith	Baby Britain	Oficial	Vevo
197	Elton John	Tiny Dancer	Oficial	Vevo
198	Elvis Costello	Less Than Zero	Apresentação	Youtube
199	Elvis Presley	Heartbreak Hotel	Apresentação	Vimeo
200	Eminem	My Fault	Apresentação	Youtube
201	Emma's Mini	Lost	Não possui vídeo	
202	Erasure	Chains of Love	Oficial	Vevo
203	Eric Clapton	Wonderful Tonight	Apresentação	Vevo
204	Etherine	Never Leave	Não possui vídeo	
205	Eurythmics	Sweet Dreams (Are Made Of This)	Oficial	Vevo
206	Evanescence	My Immortal	Oficial	Vevo
207	Everly Brothers	Take a Message to Mary	Não Oficial	Youtube
208	Faith Hill	Let's Make Love	Não Oficial	Youtube
209	Faith No More	Epic	Oficial	Vevo
210	Falik	Bliss	Não possui vídeo	
211	Fiona Apple	Love Ridden	Apresentação	Youtube
212	Fleetwood Mac	Say You Love Me	Apresentação	Youtube
213	The Flying Burrito Brothers	Break My Mind	Não possui vídeo	
214	Foo Fighters	Big Me	Oficial	Vevo
215	Four Stones	Brilliant Day Eine Kleine Mix	Não possui vídeo	
216	Frank Sinatra	Fly Me To The Moon	Apresentação	Youtube

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
217	Frank Zappa	Whats the Ugliest Part of Your Body	Não possui vídeo	
218	Franz Ferdinand	Come on Home	Não Oficial	Youtube
219	Garbage	Hammering in my Head	Apresentação	Youtube
220	Gene Clark	The True One	Apresentação	Youtube
221	Genesis	Cuckoo Cocoon	Não possui vídeo	
222	George Harrison	All Things Must Pass	Apresentação	Youtube
223	Gerns	Lexicon Devil	Apresentação	Youtube
224	Gin Blossoms	Hey Jealousy	Oficial	Vevo
225	Glen Bledsoe	P Pop	Não possui vídeo	
226	Glenn Miller	In The Mood	Não Oficial	Youtube
227	Gloria Gaynor	I Will Survive	Não Oficial	Youtube
228	The Go-Gos	Vacation	Oficial	Vevo
229	Gram Parsons	\$1000 Wedding	Apresentação	Youtube
230	Grateful Dead	High Time	Apresentação	Youtube
231	Grayson Wray	Heaven's the Place	Não possui vídeo	
232	Green Day	Longview	Oficial	Vevo
233	Guided by Voices	Kicker Of Elves	Não possui vídeo	
234	Guns N' Roses	November Rain	Oficial	Vevo
235	Daryl Hall & John Oates	Private Eyes	Oficial	Vevo
236	Howlin Wolf	Moanin at Mignight	Não possui vídeo	
237	Human Response	Angels	Não possui vídeo	
238	Hybris	Hate	Não possui vídeo	

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
239	Ike & Tina Turner	River Deep Mountain	Não Oficial	Youtube
240	Indiginus	Dusty Lands	Não possui vídeo	
241	Interpol	Stella Was A Diver And She Was Always Down	Não Oficial	Vimeo
242	Introspekt	TBD	Não possui vídeo	
243	Ivilion	D B L	Não possui vídeo	
244	Jacklopes	Rotgut	Não possui vídeo	
245	The Jackson 5	A B C	Apresentação	MTV
246	Jacques Brel	Les Vieux	Apresentação	Youtube
247	Jade Leary	Going In	Não possui vídeo	
248	Jag	Jag's Rag	Não possui vídeo	
249	James Brown	Give It Up or Turn it Loose	Não Oficial	Youtube
250	James Taylor	Fire & Rain	Apresentação	Youtube
251	Jamie Janover	Event Horizon	Não possui vídeo	
252	Jamiroquai	Little L	Oficial	Vevo
253	Jan & Dean	Surf City	Não Oficial	Youtube
254	Jane's Addiction	Been Caught Stealing	Oficial	Vevo
255	Janet Jackson	Miss You Much	Oficial	Vevo
256	Jay Kishor	Raga Malgunji Jor	Não possui vídeo	
257	Jeff Buckley	Last Goodbye	Oficial	Vevo
258	Jefferson Starship	Somebody To Love	Apresentação	Youtube
259	Jerry Lee Lewis	Great Balls	Não Oficial	Youtube
260	Jesse Manno	Alf Leyla Wa Leyla	Não possui vídeo	

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
261	Jewel	Enter From The East	Apresentação	Youtube
262	Jimi Hendrix	Highway Chile	Não possui vídeo	
263	John Cale	Pablo Picasso	Apresentação	Vimeo
264	John Coltrane	Giant Steps	Não possui vídeo	
265	John Lee Hooker	Boom Boom	Oficial	MTV
266	John Lennon	Imagine	Não Oficial	Youtube
267	Johnny Cash	The Man Comes Around	Não possui vídeo	
268	Joram	Solipsism	Não possui vídeo	
269	Joy Division	Love Will Tear Us Apart	Oficial	Vevo
270	Junior Murvin	Police and Thieves	Apresentação	Youtube
271	Kansas	Carry On Wayward Son	Oficial	Vevo
272	Kanye West	Gold Digger	Oficial	Vevo
273	Kenji Williams	I'm Alive	Não possui vídeo	
274	King Crimson	Thela Hun Ginjeet	Apresentação	Vimeo
275	Kiss	Deuce	Apresentação	Vevo
276	Kokoon	Order	Não possui vídeo	
277	Kool & The Gang	Funky Stuff	Apresentação	Youtube
278	Kourosh Zolani	Peaceful Planet	Não possui vídeo	
279	Kraftwerk	Spacelab	Apresentação	Youtube
280	Kris Kristofferson	Best Of All Possible Worlds	Apresentação	Youtube
281	Lambert Hendricks and Ross	Gimme That Wine	Apresentação	Youtube
282	Led Zeppelin	Immigrant Song	Apresentação	Youtube

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
283	Leonard Cohen	Suzanne	Apresentação	Vevo
284	Lisa Debenedictis	Fruitless	Não possui vídeo	
285	Live	Lightning Crashes	Oficial	Vevo
286	Liz Phair	Supernova	Oficial	Vevo
287	LL Cool J	Mama Said Knock You Out	Oficial	Vevo
288	Los Lobos	Corrido 1	Não Oficial	Youtube
289	Lou Reed	Walk On The Wild Side	Não Oficial	Youtube
290	Louis Armstrong	Hotter Than That	Não possui vídeo	
291	Love	You Set The Scene	Não possui vídeo	
292	Love and Rockets	Moirror People	Não Oficial	Youtube
293	Lynyrd Skynyrd	Sweet Home Alabama	Apresentação	Vevo
294	Marcy Gray	I Try	Oficial	Vevo
295	Madness	Baggy Trousers	Oficial	MTV
296	Madonna	Ray Of Light	Oficial	Vevo
297	The Mamas & The Papas	World of Love	Apresentação	Youtube
298	Manassas	Bound To Fall	Apresentação	Youtube
299	The Marvelettes	Please Mr. Postman	Apresentação	Youtube
300	Marvin Gaye	Whats Going on	Apresentação	Youtube
301	Mary Wells	My Guy	Apresentação	Youtube
302	Massive Attack	Risingson	Oficial	Vevo
303	Mazzy Star	Fade Into You	Oficial	Vevo
304	MC Hammer	U Can't Touch This	Oficial	Vevo

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
305	Memories of Tomorrow	Transcend Fate of A DJ	Não possui vídeo	
306	Men At Work	Who Can It Be Now?	Oficial	Vevo
307	Metallica	One	Oficial	MTV
308	Michael Jackson	Billie Jean	Oficial	Vevo
309	Michael Masley	Advice From The Angel Of Thresholds	Não possui vídeo	
310	Miles Davis	Blue In Green	Não possui vídeo	
311	Moby	Porcelain	Oficial	Vevo
312	Modest Mouse	What People Are Made Of	Apresentação	Youtube
313	The Monkees	A Little bit Me, A Little Bit You	Não Oficial	Youtube
314	Monoide	Golden Key	Não possui vídeo	
315	Morrissey	Everyday Is Like Sunday	Não Oficial	Youtube
316	Mose Allison	Monsters of the ID	Não possui vídeo	
317	Mott The Hoople	Roll Away The Stone	Apresentação	Youtube
318	Mr Epic	Ruff and Tumble	Não possui vídeo	
319	Mr Gelatin	Knysna Mushrooms	Não possui vídeo	
320	Mrdc	Leaving	Não possui vídeo	
321	Muddy Waters	I'm Ready	Apresentação	Youtube
322	Muddy Waters	Mannish Boy	Apresentação	Youtube
323	My Bloody Valentine	When You Sleep	Apresentação	Youtube
324	Myles Cochran	Getting Stronger	Não possui vídeo	
325	Napoleon Blown Aparte	Higher Education	Não possui vídeo	
326	Natalie Imbruglia	Torn	Oficial	Vevo

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
327	Neil Young	Razor Love	Apresentação	Youtube
328	Neil Young and Crazy Horse	Western Hero	Não possui vídeo	
329	Nelly	Country Grammar	Oficial	Vevo
330	Neutral Milk Hotel	Where You'll Find Me Now	Não possui vídeo	
331	New Order	Blue Monday	Oficial	Vevo
332	Nine Inch Nails	Head Like A Hole	Oficial	Vevo
333	Nirvana	Aneurysm	Apresentação	Vevo
334	No Doubt	Artificial Sweetener	Apresentação	Youtube
335	No Doubt	Simple Kind Of Life	Oficial	Vevo
336	Norah Jones	Don't Know Why	Oficial	Vevo
337	Norine Braun	Spanish Banks	Não possui vídeo	
338	Nova Express	I'm Alive	Não possui vídeo	
339	Oasis	Supersonic	Oficial	Vevo
340	O'Jays	Livin for The Weekend	Apresentação	Youtube
341	The Olivia Tremor Control	I Have Been Floated	Apresentação	Youtube
342	Otis Redding	Mr. Pitiful	Apresentação	Vevo
343	OutKast	Ms. Jackson	Oficial	Vevo
344	Panacea	Dragaicuta	Não possui vídeo	
345	Pantera	Becoming	Não Oficial	Youtube
346	Paul McCartney	Ebony and Ivory	Não Oficial	Youtube
347	Pearl Jam	Yellow	Apresentação	Youtube
348	Pet Shop Boys	Being Boring	Não Oficial	Vimeo

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
349	Pink Floyd	Echoes	Não Oficial	UOL
350	Pixies.	Wave of Mutilation	Apresentação	Youtube
351	Pizzle	What's Wrong With My Footm	Não possui vídeo	
352	P J Harvey	Dry	Apresentação	Youtube
353	The Police	Every Little Thing She Does Is Magic	Oficial	Vevo
354	Portishead	All Mine	Não Oficial	Youtube
355	The Pretenders	Day After Day	Não Oficial	Youtube
356	Primus	Jerry Was A Race Car Driver	Oficial	Vevo
357	Processor	Nibtal 7	Não possui vídeo	
358	Propellerheads	Take California	Não possui vídeo	
359	The Psychedelic Furs	Love My Way	Oficial	Vevo
360	Psychotropic	Dead Slow Day	Não possui vídeo	
361	Queen and Five	We Will Rock You	Oficial	Vevo
362	R.E.M.	Camera	Não possui vídeo	
363	Radiohead	Karma Police	Apresentação	Youtube
364	Rage Against the Machine	Maggie's Farm	Não Oficial	Youtube
365	Ramones	I Just Want To Have Something To Do	Não Oficial	Youtube
366	Randy Newman	Sail Away	Apresentação	Youtube
367	Ray Charles	Hit The Road Jack	Não Oficial	Youtube
368	Red Hot Chili Peppers	Give It Away	Oficial	MTV
369	Replacements	Answering Machine	Apresentação	Youtube
370	Rick James	Super Freak	Oficial	Vevo

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
371	Robert Johnson	Sweet Home Chicago	Não Oficial	Youtube
372	Rocket City	Riot Mine Tonite	Não possui vídeo	
373	The Rolling Stones	Little by Little	Apresentação	Youtube
374	The Ronettes	Walking In The Rain	Apresentação	Youtube
375	Roots of Rebellion	Legend	Não possui vídeo	
376	Roxy Music	Love Is The Drug	Apresentação	Vevo
377	Roy Orbison	Running Scared	Apresentação	Vevo
378	Ruben Gonzales	El Cumbanchero	Apresentação	Youtube
379	Rufus Wainwright	Cigarettes and Chocolate Milk	Apresentação	Youtube
380	Sade	Smooth Operator	Oficial	Vevo
381	Santana & Dave Matthews	Love of My Life	Apresentação	Youtube
382	Sarah McLachlan	Possession	Oficial	Vevo
383	Saros	Prelude	Não possui vídeo	
384	Scott Hill	Silk Road	Não possui vídeo	
385	Screaming Trees	Nearly Lost You	Oficial	Vevo
386	Sebadoh	Soul and Fire	Apresentação	Youtube
387	Seismic Anomaly	Wrecking Ball	Não possui vídeo	
388	Sex Pistols	Pretty Vacant	Oficial	Vevo
389	Shakira	The One	Oficial	MTV
390	Shane Jackman	Set Fire to The City	Não possui vídeo	
391	Sheryl Crow	I Shall Believe	Apresentação	Dailymotion
392	The Shins	New Slang	Oficial	MTV

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
393	Shira Kammen	The Water in Wide	Não possui vídeo	
394	Shiva In Exile	Hollow Earth	Não possui vídeo	
395	Shuggie Otis	Sweet Thang	Apresentação	Youtube
396	Paul Simon	The Only Living Boy In New York	Apresentação	Vevo
397	Sinead O'Connor	No Man's Woman	Não Oficial	Youtube
398	Sir Mix-A-Lot	Baby Got Back	Não Oficial	Youtube
399	Skitzo	Last Depression	Não possui vídeo	
400	Sly and the Family Stone	Just Like a Baby	Não possui vídeo	
401	Small Faces	Lazy Sunday	Não Oficial	Youtube
402	The Smashing Pumpkins	Rocket	Oficial	Vevo
403	The Smithereens	Behing The Wall Of Sleep	Oficial	Vevo
404	The Smiths	How Soon is Now	Oficial	MTV
405	Smokey Robinson & Jessie J	Cruisin'	Apresentação	Vevo
406	Smokey Robinson and Miracles	Ooo Baby Baby	Não Oficial	Youtube
407	Snoop Dogg	Gin and Juice	Oficial	Vevo
408	Solace	Laz (7/8)	Não possui vídeo	
409	Somadrone	Coda	Não possui vídeo	
410	Sonic Youth	Teenage Riot	Não Oficial	Youtube
411	Sonny Rollins	Stroke Rode	Apresentação	Youtube
412	Soul II Soul and Caron Wheeler	Keep On Movin'	Oficial	Vevo
413	Soulprint	Crawlspac	Não possui vídeo	
414	Soundgarden	Black Hole Sun	Oficial	Vevo

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
415	The Specials	Gangsters	Oficial	Vevo
416	Spencer Davis Group	Gimme Some Lovin'	Apresentação	Youtube
417	Spice Girls	Stop	Oficial	Vevo
418	Spinecar	Stay	Não possui vídeo	
419	Spiritualized	Stop Your Crying	Oficial	Vevo
420	Squarepusher	A Journey to Reedham (7am mix)	Apresentação	Youtube
421	Squeeze	Pulling Mussels (From The Shell)	Apresentação	Vevo
422	Stan Getz	Corcovado Quiet Nights of Quiet Stars	Não possui vídeo	
423	The Standells	Dirty Water / Why Pick On Me	Apresentação	Youtube
424	Starship	Nothing's Gonna Stop Us Now	Oficial	Vevo
425	Steely Dan	Rikki Don't Lose that Number	Apresentação	Youtube
426	Steppenwolf	Born to be wild	Não Oficial	Vimeo
427	Stereolab	Cybele's Reverie	Apresentação	Youtube
428	Stevie Ray Vaughan and Double Trouble	Pride and Joy	Apresentação	Vevo
429	Stevie Wonder	For Once In My Life	Apresentação	Youtube
430	Sting	Big Lie Small World	Não possui vídeo	
431	The Stone Roses	I Wanna Be Adored	Oficial	Vevo
432	The Stooges & Iggy	Dirt	Não Oficial	Youtube
433	The Stranglers	Golden Brown	Oficial	MTV
434	Strawbs	New World	Apresentação	Youtube
435	Strojovna 07	130	Não possui vídeo	
436	The Style Council	Headstart For Happiness	Apresentação	Youtube

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
437	The Sundays	Heres Where The Story Ends	Não Oficial	Vimeo
438	Superchunk	Slack Motherfucker	Apresentação	Youtube
439	Sweet	Fox On The Run	Oficial	Vevo
440	Syd Barrett	Effervescing Elephant	Não Oficial	Vimeo
441	Sylvester	You Make Me Feel(Mighty Real)	Oficial	Vevo
442	Syreeta	What Love Has Joined Together	Não possui vídeo	
443	T-Rex	Children of The Revolution	Apresentação	Vevo
444	Talking Heads	And She Was	Não Oficial	Youtube
445	Tears For Fears	Everybody Wants To Rule The World	Oficial	Vevo
446	Teenage Fanclub	The Concept	Apresentação	MTV
447	Television	Venus	Apresentação	Youtube
448	Temptations	Since I Lost My Baby	Não possui vídeo	
449	Thelonious Monk	Epsitrophy	Apresentação	Youtube
450	They Might Be Giants	I Should be Allowed to Think	Não possui vídeo	
451	Thin Lizzy	Don't Believe A Word	Apresentação	Vevo
452	Third Eye Blind	Semi-Charmed Life	Oficial	Vevo
453	Throwing Muses	Hate My Way	Apresentação	Youtube
454	Thursday Group	Like White on Rice	Não possui vídeo	
455	Tilopa	Kyo Rei	Não possui vídeo	
456	Tim Buckley	Morning Glory	Não Oficial	Youtube
457	Tim Hardin	Don't Make Promises	Não possui vídeo	
458	Tim Rayborn	Yedi Tekrar	Não possui vídeo	

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
459	Todd Rundgren	Bang the Drum all day	Apresentação	Vevo
460	Tom Paul	Little Part of me	Não possui vídeo	
461	Tom Petty and Heartbreakers	I Won't Back Down	Oficial	Vevo
462	Tom Petty and Heartbreakers	Don't Come Around Here No More	Artista Duplicado	Vevo
463	Tom Waits	Time	Apresentação	Youtube
464	Tommy James and The Shondells	I Think We're Alone now	Apresentação	Youtube
465	Tori Amos	Professional Window	Oficial	MTV
466	Touching Grace	Wild Spring Apples	Não possui vídeo	
467	Traffic	Pearly Queen	Apresentação	Youtube
468	Tricky	Christiansands	Apresentação	MTV
469	The Troggs	Wild Thing	Apresentação	Vevo
470	The Turtles	Elenore	Apresentação	Youtube
471	U2	Hold Me, Thrill Me, Kiss Me, Kill Me ⁴	Oficial	Vimeo
472	Ultravox	Dancing With Tears In My Eyes	Não Oficial	Youtube
473	Uncle Tupelo	The Long cut	Apresentação	Youtube
474	Urge Overkill	Sister Havana	Oficial	Vevo
475	Utopia Banished	By Mourning	Não possui vídeo	
476	Van Halen	Ain't Tankin' Bout Love	Não Oficial	Youtube
477	Van Morrison	And It Stoned Me	Não Oficial	Youtube
478	The Vapors	Turning Japanese	Não Oficial	Youtube
479	The Velvet Underground	New Age	Apresentação	Youtube
480	Version	Universal Humans	Não possui vídeo	

Continua na próxima página

Id	Artista	Composição	Tipo de Vídeo	Fonte
481	Very Large Array	Psychedelic Baby	Não possui vídeo	
482	Violent Femmes	Blister In The Sun	Oficial	MTV
483	War	All Day Music	Não possui vídeo	
484	Weezer	Buddy Holly	Oficial	Vevo
485	Wes Montgomery	Bumpin'	Não possui vídeo	
486	The West Exit	Nocturne	Não possui vídeo	
487	The White Stripes	Hotel Yorba	Apresentação	Youtube
488	Whitney Houston	How Will I Know	Oficial	Vevo
489	The Who	Bargain	Apresentação	MTV
490	Wicked Allstars	Happy	Não possui vídeo	
491	Wicked Boy	Pressure	Não possui vídeo	
492	Wilco	KingPin	Apresentação	Youtube
493	Williamson	What's on The Ceiling Beats Whats on TV	Não possui vídeo	
494	XTC	Love at First Sight	Apresentação	Youtube
495	Yakshi	Chandra	Não possui vídeo	
496	Yeah Yeah Yeahs	Maps	Oficial	Vevo
497	Yes	Leave It	Não Oficial	Youtube
498	Yo La Tengo	Tom Courtenay	Apresentação	Youtube
499	Young MC	Bust A Move	Não Oficial	Youtube
500	The Rascals	Baby Let's Wait	Apresentação	Youtube
501	Zapp and Roger	Dance Floor	Apresentação	Youtube
502	The Zombies	Beechwood Park	Apresentação	Youtube

¹ Faixa bônus do filme Tupac: Resurrection

² Usado no filme Magnólia onde as músicas da artista são utilizadas para expressar as cenas.

³ Informado como Oficial no site oficial do Artista <http://www.bobmarley.com/media/videos/music-videos/three-little-birds/>

⁴ Informado como Oficial no site oficial do Artista <http://www.u2.com/media/player/63/1>

APÊNDICE B – Base CAL500 Estendida

Tabela 26 – Informações sobre os vídeos musicais.

Id	Artista	Música
2	2Pac	Trapped
4	A Tribe Called Quest	Bonita Applebum
5	Aaron Neville	Tell It Like It Is
6	ABBA	SOS
7	ABC	Poison Arrow
9	Adam & The Ants	Prince Charming
10	Adam Ant	Wonderful
13	Aerosmith	Dude (Looks Like A Lady)
14	Aimee Mann	Wise Up
15	Air	Sexy Boy
17	Alanis Morissette	Thank U
19	Alice In Chains	No Excuses
20	Alicia Keys	Fallin'
22	Altered Images	Don't Talk to Me About Love
31	Aphex Twin	Come to Daddy
39	Ashford & Simpson	Solid
43	Backstreet Boys	As Long As You Love Me
44	Badly Drawn Boy	All Possibilities
47	Barenaked Ladies	It's All Been Done
53	The Beatles	Strawberry Fields Forever
55	The Beautiful South	One Last Love Song
56	Beck	Where It's At
57	Bee Gees	Stayin' Alive
60	Ben Folds Five	Brick
63	Billie Holiday & Tony Bennett	God Bless The Child
65	Billy Joel	We Didn't Start the Fire
67	Björk	Army Of Me
73	Blondie	Rupture
76	Blur	Country House
79	Bob Marley	Three Little Birds

Continua na próxima página

Id	Artista	Música
84	Bomb The Bass	Bug Powder Dust
85	Bon Jovi	Livin' On A Prayer
86	Bonnie Tyler	Total Eclipse of the Heart
87	The Boo Radleys	Wake up Boo!
97	Britney Spears	I'm A Slave 4 U
99	Bryan Adams	Cuts Like a Knife
104	The Buggles	Video Killed The Radio Star
107	Bush	Comedown
108	Busta Rhymes	Woo-Hah! Got You All In Check
117	The Cardigans	Lovefool
122	Carpenters	Rainy Days And Mondays
126	C+C Music Factory	Gonna Make You Sweat (Everybody Dance Now)
128	Chantal Kreviazuk	Surrounded
132	Cheap Trick	Dream Police
133	The Chemical Brothers	Setting Sun
137	Chic	Le Freak
138	Chicago	If You Leave Me Now
141	Christina Aguilera	Genie In A Bottle
143	Chumbawamba	Tubthumping
144	The Church	Under The Milky Way
149	Coldplay	Clocks
154	The Cranberries	Linger
160	The Cure	Just Like Heaven
163	Cypress Hill	Lowrider
164	Daft Punk	Da Funk
168	De La Soul	Eye Know
170	Def Leppard	Pour Some Sugar On Me
175	Dido	Here with Me
177	Dire Straits	Money For Nothing
178	DJ Jazzy Jeff & The Fresh Prince	Summertime
190	Duncan Sheik	Barely Breathing
193	Earth, Wind & Fire	September
194	Eels	Cancer For The Cure
196	Elliott Smith	Baby Britain

Continua na próxima página

Id	Artista	Música
197	Elton John	Tiny Dancer
202	Erasure	Chains of Love
205	Eurythmics	Sweet Dreams (Are Made Of This)
206	Evanescence	My Immortal
209	Faith No More	Epic
214	Foo Fighters	Big Me
224	Gin Blossoms	Hey Jealousy
228	The Go-Gos	Vacation
232	Green Day	Longview
234	Guns N' Roses	November Rain
235	Daryl Hall & John Oates	Private Eyes
252	Jamiroquai	Little L
254	Jane's Addiction	Been Caught Stealing
255	Janet Jackson	Miss You Much
257	Jeff Buckley	Last Goodbye
265	John Lee Hooker	Boom Boom
269	Joy Division	Love Will Tear Us Apart
271	Kansas	Carry On Wayward Son
272	Kanye West	Gold Digger
285	Live	Lightning Crashes
286	Liz Phair	Supernova
287	LL Cool J	Mama Said Knock You Out
294	Marcy Gray	I Try
295	Madness	Baggy Trousers
296	Madonna	Ray Of Light
302	Massive Attack	Risingson
303	Mazzy Star	Fade Into You
304	MC Hammer	U Can't Touch This
306	Men At Work	Who Can It Be Now?
307	Metallica	One
308	Michael Jackson	Billie Jean
311	Moby	Porcelain
326	Natalie Imbruglia	Torn
329	Nelly	Country Grammar
331	New Order	Blue Monday

Continúa na próxima página

Id	Artista	Música
332	Nine Inch Nails	Head Like A Hole
335	No Doubt	Simple Kind Of Life
336	Norah Jones	Don't Know Why
339	Oasis	Supersonic
343	OutKast	Ms. Jackson
353	The Police	Every Little Thing She Does Is Magic
356	Primus	Jerry Was A Race Car Driver
359	The Psychedelic Furs	Love My Way
361	Queen and Five	We Will Rock You
368	Red Hot Chili Peppers	Give It Away
370	Rick James	Super Freak
380	Sade	Smooth Operator
382	Sarah McLachlan	Possession
385	Screaming Trees	Nearly Lost You
388	Sex Pistols	Pretty Vacant
389	Shakira	The One
392	The Shins	New Slang
402	The Smashing Pumpkins	Rocket
403	The Smithereens	Behing The Wall Of Sleep
404	The Smiths	How Soon is Now
407	Snoop Dogg	Gin and Juice
412	Soul II Soul and Caron Wheeler	Keep On Movin'
414	Soundgarden	Black Hole Sun
415	The Specials	Gangsters
417	Spice Girls	Stop
419	Spiritualized	Stop Your Crying
424	Starship	Nothing's Gonna Stop Us Now
431	The Stone Roses	I Wanna Be Adored
433	The Stranglers	Golden Brown
439	Sweet	Fox On The Run
441	Sylvester	You Make Me Feel(Mighty Real)
445	Tears For Fears	Everybody Wants To Rule The World
452	Third Eye Blind	Semi-Charmed Life
461	Tom Petty and Heartbreakers	I Won't Back Down
465	Tori Amos	Professional Window

Continua na próxima página

Id	Artista	Música
471	U2	Hold Me, Thrill Me, Kiss Me, Kill Me
474	Urge Overkill	Sister Havana
482	Violent Femmes	Blister In The Sun
484	Weezer	Buddy Holly
488	Whitney Houston	How Will I Know
496	Yeah Yeah Yeahs	Maps

Tabela 27 – Informações sobre os rótulos de emoções.

Representação	Emoção Original da CAL500	Total de Ocorrências
λ_1	'EMOTION-ANGRY_/_AGRESSIVE'	28
λ_2	'EMOTION-AROUSING_/_AWAKENING'	96
λ_3	'EMOTION-BIZARRE_/_WEIRD'	27
λ_4	'EMOTION-CALMING_/_SOOTHING'	66
λ_5	'EMOTION-CAREFREE_/_LIGHTHEARTED'	69
λ_6	'EMOTION-CHEERFUL_/_FESTIVE'	70
λ_7	'EMOTION-EMOTIONAL_/_PASSIONATE'	103
λ_8	'EMOTION-EXCITING_/_THRILLING'	79
λ_9	'EMOTION-HAPPY'	86
λ_{10}	'EMOTION-LAID-BACK_/_MELOW'	56
λ_{11}	'EMOTION-LIGHT_/_PLAYFUL'	65
λ_{12}	'EMOTION-LOVING_/_ROMANTIC'	53
λ_{13}	'EMOTION-PLEASANT_/_COMFORTABLE'	107
λ_{14}	'EMOTION-POSITIVE_/_OPTIMISTIC'	81
λ_{15}	'EMOTION-POWERFUL_/_STRONG'	109
λ_{16}	'EMOTION-SAD'	42
λ_{17}	'EMOTION-TENDER_/_SOFT'	52
λ_{18}	'EMOTION-TOUCHING_/_LOVING'	53

Tabela 28 – Rotulação da base.

Id	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
2	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0
4	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	1	1
6	0	1	0	0	1	1	0	1	1	0	1	0	1	1	1	0	0	0
7	0	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	0	0
9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
10	0	1	0	1	0	1	1	1	1	0	1	1	1	1	0	0	0	1
13	0	1	0	0	1	1	1	1	1	0	1	0	0	1	1	0	0	0

Continua na próxima página

Id	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
14	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	1	0
15	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
17	0	0	0	1	0	0	1	0	0	0	0	1	1	1	1	1	0	1
19	0	1	0	1	0	0	1	0	1	0	0	0	1	0	1	1	1	1
20	0	1	0	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1
22	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1
31	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
39	0	1	0	0	0	1	1	0	1	0	0	1	1	1	1	0	0	1
43	0	1	0	1	0	0	1	0	1	0	0	1	1	1	1	1	1	1
44	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
47	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0
53	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
55	0	0	0	1	0	0	1	0	0	1	0	1	1	0	0	1	1	0
56	0	1	1	1	1	1	1	1	1	0	1	0	0	1	1	0	0	0
57	0	1	0	0	1	1	1	1	1	0	1	0	1	1	1	0	0	1
60	0	0	0	1	0	0	1	0	0	1	1	1	1	0	1	1	1	1
63	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1
65	0	1	1	0	1	1	1	1	1	0	1	0	1	1	1	0	0	0
67	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
73	0	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	1	0
76	0	1	0	1	1	1	1	0	1	1	1	0	1	1	0	0	1	0
79	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1
84	1	1	0	0	0	0	1	1	0	0	0	0	1	1	1	1	0	0
85	0	1	0	0	0	0	1	1	1	0	0	0	0	1	1	0	0	0
86	0	0	0	1	0	0	1	1	0	0	0	1	1	0	1	1	1	1
87	0	1	1	0	1	1	1	1	1	0	1	0	1	1	0	0	0	1
97	0	1	0	1	1	1	1	1	1	0	1	1	1	0	1	0	1	0
99	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0
104	0	1	0	0	1	1	0	0	1	1	1	0	1	1	1	0	0	0
107	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0
108	1	1	0	0	0	0	0	1	1	0	1	0	1	0	1	1	0	0
117	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
122	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0	1	1	1
126	0	1	0	0	1	1	1	1	1	0	1	0	1	1	1	0	0	0
128	0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1	0
132	0	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	0	0

Continua na próxima página

Id	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
133	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0
137	0	1	0	0	1	1	0	1	1	0	1	0	1	1	1	0	0	0
138	0	0	0	1	0	0	1	0	0	1	0	1	1	1	1	1	1	1
141	0	1	0	1	1	1	1	1	1	0	1	1	0	1	1	0	0	0
143	0	1	0	0	1	1	0	1	1	0	1	0	1	1	1	0	0	0
144	0	0	0	1	1	0	0	0	0	1	1	0	1	0	0	1	1	0
149	0	1	0	1	1	1	1	0	1	0	0	1	1	1	1	0	1	1
154	0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1	1
160	0	1	0	0	1	0	0	0	0	1	0	1	1	0	1	0	0	1
163	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1	0	0	0
164	0	1	0	0	0	1	0	0	1	0	1	0	1	1	1	0	0	0
168	0	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0	0	1
170	1	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0	0	0
175	0	0	0	1	1	0	1	0	1	1	0	1	1	1	1	0	1	1
177	0	1	0	0	1	1	0	1	1	0	1	0	1	1	1	0	0	0
178	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
190	0	0	0	0	1	0	0	0	1	1	1	0	1	1	0	0	0	0
193	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
194	0	1	1	0	1	1	1	1	1	0	1	0	0	1	1	0	0	0
196	0	1	0	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1
197	0	0	0	1	1	0	1	0	1	1	0	0	1	1	0	0	1	1
202	0	1	0	0	1	1	1	1	1	0	1	1	0	1	1	0	0	1
205	0	1	1	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0
206	0	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	1	1
209	1	1	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0
214	0	0	0	1	1	0	1	0	1	1	1	1	1	1	0	0	1	1
224	1	1	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0
228	0	1	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0	0
232	1	1	0	0	1	1	1	1	0	0	0	0	0	1	1	0	0	0
234	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0
235	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0	0	1	0
252	0	1	0	0	1	1	1	1	1	0	1	0	1	1	1	0	0	0
254	1	1	0	0	1	1	1	1	1	0	1	0	1	1	1	0	0	0
255	0	1	0	1	1	1	1	1	1	0	1	1	0	1	1	0	1	1
257	0	0	0	1	1	0	1	0	1	0	1	1	1	1	1	0	0	1
265	0	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0	0	0

Continua na próxima página

Id	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
269	0	1	0	1	0	0	1	0	1	1	1	1	1	0	0	0	0	0
271	0	1	0	0	0	1	1	1	1	0	0	0	1	1	1	0	0	0
272	0	1	0	1	1	1	1	1	1	1	0	0	1	0	1	0	1	0
285	1	1	1	1	0	0	1	1	0	0	0	1	1	0	1	1	0	1
286	1	1	0	0	0	0	1	1	0	0	0	0	1	0	1	0	0	0
287	1	1	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0
294	0	0	0	1	0	1	1	0	0	1	0	1	1	1	1	1	1	1
295	0	1	1	0	1	1	0	1	1	0	1	0	1	1	0	0	0	0
296	0	1	0	0	1	1	0	1	1	0	1	0	1	1	1	0	0	0
302	1	1	0	0	0	0	0	1	0	1	0	0	1	0	1	1	1	0
303	0	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	1	1
304	0	1	0	0	1	0	0	1	1	0	0	0	1	1	1	0	0	0
306	0	1	0	0	1	1	1	1	1	1	1	0	1	1	1	0	0	0
307	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
308	0	1	0	0	0	0	1	1	0	0	0	0	1	0	1	0	0	1
311	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	1	0
326	1	0	0	1	0	0	1	1	0	0	0	1	1	0	1	1	1	1
329	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	0	0	0
331	0	1	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0
332	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
335	0	1	0	1	1	0	1	0	1	0	0	1	1	1	1	1	1	1
336	0	0	0	1	1	0	1	0	0	1	0	1	1	1	1	1	1	1
339	0	1	0	0	1	1	1	1	0	0	1	0	1	0	1	0	0	0
343	0	1	0	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1
353	0	1	0	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1
356	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
359	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0
361	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0
368	1	1	0	0	1	1	1	1	1	0	1	0	1	1	1	0	0	0
370	0	1	0	0	1	1	1	1	1	0	0	1	0	1	1	0	0	0
380	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
382	0	1	0	1	0	1	1	0	1	0	0	1	1	0	1	1	1	1
385	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0
388	1	1	1	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0
389	0	1	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1
392	0	0	0	1	0	0	1	0	0	1	0	1	1	0	0	1	1	1

Continua na próxima página

Id	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
402	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
403	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
404	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0
407	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
412	0	0	0	1	1	0	1	0	1	1	1	1	1	1	1	0	1	1
414	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0
415	0	1	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0
417	0	1	0	1	1	1	0	1	1	1	1	1	1	1	0	0	0	1
419	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1
424	0	1	0	0	1	1	1	1	1	0	0	1	1	1	1	0	0	1
431	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0
433	0	0	1	1	1	0	1	0	1	1	1	0	1	0	0	1	1	0
439	0	1	0	0	1	1	0	1	1	0	0	0	1	1	1	0	0	0
441	0	1	0	0	1	1	0	1	1	0	1	0	1	1	1	0	0	1
445	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1
452	0	1	0	0	1	1	1	1	1	0	1	0	1	1	1	0	0	0
461	0	0	0	1	0	1	1	0	1	1	0	0	1	1	1	0	0	0
465	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
471	0	1	0	0	0	0	1	1	0	0	0	0	1	1	1	0	0	0
474	0	1	0	0	0	0	1	1	0	0	0	0	1	1	1	0	0	0
482	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0
484	0	1	0	0	1	1	0	1	1	1	1	0	1	1	1	0	0	0
488	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1
496	0	0	0	1	0	0	1	0	0	1	0	1	1	1	0	0	1	1