



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2014/05.01.23.23-TDI

AVALIAÇÃO DE MÉTODOS DE SELEÇÃO DE AMOSTRAS PARA REDUÇÃO DO TEMPO DE TREINAMENTO DO CLASSIFICADOR SVM

Peterson Lima Sarmiento

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelo Dr. Luciano Vieira Dutra, aprovada em 28 de maio de 2014.

URL do documento original:

<<http://urlib.net/8JMKD3MGP5W34M/3G8F49P>>

INPE
São José dos Campos
2014

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):**Presidente:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Membros:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Amauri Silva Montes - Coordenação Engenharia e Tecnologia Espaciais (ETE)

Dr. André de Castro Milone - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr. Joaquim José Barroso de Castro - Centro de Tecnologias Espaciais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2014/05.01.23.23-TDI

AVALIAÇÃO DE MÉTODOS DE SELEÇÃO DE AMOSTRAS PARA REDUÇÃO DO TEMPO DE TREINAMENTO DO CLASSIFICADOR SVM

Peterson Lima Sarmento

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelo Dr. Luciano Vieira Dutra, aprovada em 28 de maio de 2014.

URL do documento original:

<<http://urlib.net/8JMKD3MGP5W34M/3G8F49P>>

INPE
São José dos Campos
2014

Dados Internacionais de Catalogação na Publicação (CIP)

Sarmiento, Peterson Lima.

Sa74a Avaliação de métodos de seleção de amostras para redução do tempo de treinamento do classificador SVM / Peterson Lima Sarmiento. – São José dos Campos : INPE, 2014.

xxii + 104 p. ; (sid.inpe.br/mtc-m21b/2014/05.01.23.23-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014.

Orientador : Dr. Luciano Vieira Dutra.

1. SVM. 2. Classificação supervisionada. 3. Multiedição.
4. Condensação. 5. Líder. I.Título.

CDU 528.854



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Mestre** em

Computação Aplicada

Dr. Fernando Manuel Ramos

Presidente / INPE / SJC Campos - SP

Dr. Luciano Vieira Dutra

Orientador(a) / INPE / SJC Campos - SP

Dr. Sidnei João Siqueira Sant'Anna

Membro da Banca / INPE / São José dos Campos - SP

Dr. Thales Sehn Korting

Membro da Banca / INPE / São José dos Campos - SP

Dr. Alexandre Noma

Convidado(a) / UFABC / Santo André - SP

Dr. Márcio Porto Basgalupp

Convidado(a) / UNIFESP / São Paulo - SP

Este trabalho foi aprovado por:

maioria simples

unanimidade

Aluno (a): **Peterson Lima Sarmento**

São José dos Campos, 28 de Maio de 2014

“Nada é tão difícil que não possa ser feito, nem tão fácil que possa ser copiado. A única pessoa que pode ajudar-lhe a conseguir seus objetivos é você mesmo, não deixando de esquecer-se daqueles que te serviram como apoio, nem dos que utilizarão seus frutos para, também, ajudá-los a trilhar seu caminho”.

O autor

A minha esposa Denise, a meus filhos Eliane e Felipe, a meus pais Dalila e Patrício e aos meus irmãos.

Aos colegas da Memória Técnico-Científica e aos alunos e amigos da Pós-Graduação do INPE.

AGRADECIMENTOS

A minha esposa Denise, pela compreensão, paciência e companheirismo, que enfrentamos juntos no dia-a-dia para superar este desafio. Ao Orientador e Professor Luciano Vieira Dutra pelo aprendizado e pela orientação do trabalho. Ao Professor Guaraci José Erthal que, além de paciência, ajudou a construir este trabalho contribuindo com seu conhecimento e experiência. Ao Professor Camilo que contribuiu na parte estatística e ao meu amigo Rogério Galante Negri, que permitiu a utilização das imagens de seu trabalho e que ajudou na programação em IDL.

RESUMO

O classificador supervisionado Máquina de Vetores Suporte (SVM, do inglês *Support Vector Machine*), tem sido utilizado em diferentes aplicações devido ao seu bom desempenho. O SVM é um classificador supervisionado que, para a construção do modelo de classificação, utiliza basicamente os dados próximos à fronteira de separação entre as classes no espaço de atributos. Uma desvantagem deste classificador é o aumento no tempo de treinamento à medida que o tamanho do conjunto de treinamento por classe aumenta. Este trabalho objetiva diminuir o tempo de treinamento do SVM através da redução no tamanho do conjunto de treinamento, procurando manter a acurácia da classificação. Para tanto, sete técnicas de redução, são avaliadas quanto ao tempo de processamento e acurácia da classificação. Foram realizados testes com imagens sintéticas e imagens reais obtidas de sensores óticos (IKONOS) e de radar (PALSAR). O treinamento realizado com o conjunto de dados completos é utilizado como referência para as comparações. Os resultados mostram que a aplicação do algoritmo sequencial Líder pode ser uma opção para reduzir o conjunto de treinamento. A melhor redução de custo computacional foi de 85%, comparado com a classificação utilizando o conjunto completo de amostras de treinamento. A seleção aleatória e o CNN mostraram excelentes resultados para redução do custo computacional, mas o Teste de hipótese aplicado não indica estas aplicações para alguns casos. Desta forma a aplicação fica a critério do usuário.

Palavras chave: Classificação supervisionada, Treinamento, SVM, k -NN, Seleção de amostras de treinamento, Tempo de processamento, Multiedição, Condensação, Agrupamento sequencial Líder.

ANALYSIS OF SELECTION METHODS SAMPLES FOR REDUCING THE TIME TRAINING OF THE SVM CLASSIFIER

The supervised classifier Support Vector Machine (SVM), has been used in different applications due to its good performance. The SVM is a supervised classifier that for building the classification model uses basically near the decision boundary the classes in the feature space data. A disadvantage of this classifier is increased training time as the size of the training set of class increases. The objective of this work is reduce the SVM training time by reducing the size of the training set, trying to maintain the accuracy of classification. For this, seven reduction techniques are evaluated for processing time and classification accuracy. These methods are tested on simulated images as well as on real images (IKONOS and PALSAR). The results show that the images applied to the leader sequence algorithm may be an option to reduce the training set. The best computational cost was reduced by 85% compared to classification using the full set. The random selection and CNN showed excellent results as reducing the computational cost, but the hypothesis test applied not indicate its application in some cases. This way the application will be up to the user.

Keywords: Supervised classifier, Training, SVM, k -NN, Selecting samples training, processing time, computational cost, Multiediting, Condensing, Leader sequence.

LISTA DE FIGURAS

Pág.

Figura 2.1 -Exemplo com duas classes linearmente separáveis e duas soluções lineares possíveis com as respectivas margens.....	5
Figura 2.2 – Modelo linear para o SVM: (a) caso separável e (b) caso não separável.....	6
Figura 2.3- Exemplo de mapeamento dos padrões para um espaço de maior dimensão.....	8
Figura 2.4 – Superfície de decisão para o método OAO.....	11
Figura 2.5 – Superfície de decisão para o método OAA.....	11
Figura 2.6 – Modelos possíveis para m problemas binários, com p núcleos e q penalidades.....	13
Figura 2.7 – Exemplificação para o método k -NN.....	15
Figura 2.8 – Algoritmo k -NN.....	15
Figura 2.9 – Algoritmo RENN.....	16
Figura 2.10 – Algoritmo ME.....	17
Figura 2.11 – Exemplo com duas classes e dois atributos. A curva pontilhada mostra a fronteira de decisão de Bayes.....	17
Figura 2.12 – Amostras classificadas erroneamente.....	18
Figura 2.13 – Algoritmo ME – Dados selecionados após a última iteração.....	18
Figura 2.14 – Algoritmo CNN.....	19
Figura 2.15 – Aplicação do CNN aos dados gerados pelo ME apresentados na Figura 2.13.....	20
Figura 2.16 – Algoritmo RND.....	21
Figura 2.17 – Algoritmo Líder.....	22
Figura 2.18 – Número total de grupos com todas as amostras.....	23
Figura 2.19 – Características dos grupos gerados pelo algoritmo Líder.....	24
Figura 2.20 – Processos de redução do conjunto de amostras.....	25
Figura 4.1 – Imagens Sintéticas e regiões da coleta de amostras.....	39
Figura 4.2 – Amostras das imagens sintéticas no espaço de atributos.....	39
Figura 4.3 – Imagem ALOS PALSAR em amplitude.....	40
Figura 4.4 – Imagem IKONOS MS - região sul de São José dos Campos.....	41
Figura 4.5 – Fluxo para seleção da penalidade C	43
Figura 4.6 – Fluxo para redução do conjunto de treinamento.....	44
Figura 5.1 – Amostras das imagens sintéticas sem redução.....	47
Figura 5.2 – Amostras das imagens sintéticas reduzidas pelo RENN.....	48
Figura 5.3– Amostras das imagens sintéticas reduzidas pelo ME.....	48
Figura 5.4– Amostras das imagens sintéticas reduzidas pelo CNN.....	49
Figura 5.5– Amostras das imagens sintéticas reduzidas pelo RENN+CNN.....	49
Figura 5.6– Amostras das imagens sintéticas reduzidas pelo ME+CNN.....	50
Figura 5.7– Amostras das imagens sintéticas reduzidas pelo LID_C_10.....	51
Figura 5.8– Amostras das imagens sintéticas reduzidas pelo LID_H_10.....	51
Figura 5.9– Amostras das imagens sintéticas reduzidas pelo LID_HC_10.....	52
Figura 5.10– Amostras das imagens sintéticas reduzidas pelo LID_C_30.....	52
Figura 5.11– Amostras das imagens sintéticas reduzidas pelo LID_H_30.....	53
Figura 5.12– Amostras das imagens sintéticas reduzidas pelo LID_HC_30.....	53
Figura 5.13– Amostras das imagens sintéticas reduzidas pelo LID_C_50.....	54
Figura 5.14– Amostras das imagens sintéticas reduzidas pelo LID_H_50.....	54
Figura 5.15– Amostras das imagens sintéticas reduzidas pelo LID_HC_50.....	55
Figura 5.16– Amostras das imagens sintéticas reduzidas pelo LID_C_70.....	55
Figura 5.17– Amostras das imagens sintéticas reduzidas pelo LID_H_70.....	56
Figura 5.18– Amostras das imagens sintéticas reduzidas pelo LID_HC_70.....	56
Figura 5.19– Acurácia vs. variação do C – imagens sintéticas.....	57
Figura 5.20– Acurácia vs. $\log_2 C$ vs. tamanho da amostra - 1DP.....	58

Figura 5.21– Acurácia vs. $\log_2 C$ vs. tamanho da amostra - 2DP.....	59
Figura 5.22– Acurácia vs. $\log_2 C$ vs. tamanho da amostra - 3DP.....	59
Figura 5.23– Acurácia vs. tamanho da amostra - imagem 1DP.....	61
Figura 5.24– Acurácia vs. tamanho da amostra - imagem 2DP.....	61
Figura 5.25– Acurácia vs. tamanho da amostra - imagem 3DP.....	61
Figura 5.26– Tempo Global vs. Tamanho da amostra – 1DP.....	62
Figura 5.27– Tempo Global vs. Tamanho da amostra – 2DP.....	63
Figura 5.28– Tempo Global vs. Tamanho da amostra – 3DP.....	63
Figura 5.29 – Acurácia em relação ao custo computacional - imagem sintética 1DP	64
Figura 5.30 – Acurácia em relação ao custo computacional - imagem sintética 2DP	65
Figura 5.31 – Acurácia em relação ao custo computacional - imagem sintética 2DP	66
Figura 5.32 – Amostras de treinamento da imagem PALSAR.....	66
Figura 5.33 – Amostras de treinamento por pares de classes - IKONOS	67
Figura 5.34 – Acurácia vs. C - Imagem ALOS PALSAR.	68
Figura 5.35 – Acurácia dos métodos de redução para a imagem ALOS PALSAR	68
Figura 5.36 – Amostras de treinamento do IKONOS MS	69
Figura 5.37 – Acurácia vs. C - Imagem IKONOS.....	70
Figura 5.38 – Acurácia dos métodos de redução para a imagem IKONOS MS.....	71

LISTA DE TABELAS

Pág.

Tabela 2.1 – Matriz de Confusão para c classes	26
Tabela 2.2 – Exemplo da aplicação do Teste de Wilcoxon – Imagem IKONOS	31
Tabela 4.1 – Dados das imagens sintéticas.....	38
Tabela 4.2 - Classes de interesse e tamanhos de amostra.....	40
Tabela 4.3 – Classes de interesse e tamanhos de amostra	41
Tabela 4.4 – Classes de interesse e tamanhos de amostra por classe.....	42
Tabela 5.1 – Acurácias e desvios-padrão das imagens sintéticas	60
Tabela 6.1 – Redução do Tempo Global (%) – Teste de Wilcoxon	74
Tabela 6.2 – Redução do Tempo Global (%) – Conjunto completo -0,5%	76

LISTA DE SIGLAS E ABREVIATURAS

INPE	Instituto Nacional de Pesquisas Espaciais
SVM	Support Vector Machine
k -NN	k -Nearest Neighbor
IDL	Interactive Data Language
ENN	Edited Nearest Neighbor
CNN	Condensed Nearest Neighbor
SMO	Sequential Minimal Optimization
ME	Multiedit
RENN	Repeat Edited Nearest Neighbor
OAO	One against one
OAA	One against all

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO.....	1
2 FUNDAMENTAÇÃO TEÓRICA	3
2.1 O classificador SVM.....	3
2.1.1 Modelo SVM linearmente separável	4
2.1.2 Modelo SVM não linearmente separável.....	6
2.1.3 Modelo SVM não linear	8
2.2 Classificação Multiclasses	10
2.2.1 OAO – Método de decomposição um contra um.....	10
2.2.2 OAA – Método de decomposição um contra todos	11
2.3 Seleção de modelos e parâmetros	12
2.4 Alternativa para solução do Problema Quadrático.....	13
2.5 ALGORITMOS DE REDUÇÃO.....	14
2.5.1 Classificador vizinho mais próximo	14
2.5.2 Edição de Wilson Repetida - RENN.....	16
2.5.3 Edição múltipla de Devijver e Kittler – ME	16
2.5.4 Condensação de Hart – CNN.....	19
2.5.5 Redução encadeada.....	20
2.5.6 Redução por Seleção Aleatória – RND	21
2.5.7 Algoritmo Sequencial Líder - LID.....	21
2.6 Estrutura de utilização dos algoritmos	25
2.7 Método de Avaliação	26
2.8 Teste de hipótese.....	27
2.8.1 Teste de Wilcoxon	28
3 REVISÃO BIBLIOGRÁFICA	33
4 METODOLOGIA	37
4.1 Computador utilizado e Implementação	37
4.2 Imagens sintéticas	38
4.3 Imagem do sensor ALOS PALSAR.....	39
4.4 Imagem do sensor IKONOS MS	40
4.5 Estrutura dos experimentos computacionais	42
4.5.1 Coleta e Seleção de amostras.....	42
4.5.2 Seleção da penalidade C	43
4.5.3 Redução do conjunto de treinamento.....	44
5 EXPERIMENTO COMPUTACIONAL E ANÁLISE DOS RESULTADOS	47
5.1 Acurácia vs. Valor de C	56
5.2 Análise das imagens sintéticas	57

5.3	Classificação SVM imagens sintéticas – conjunto de treinamento completo	60
5.4	Aplicação dos métodos de redução e classificação SVM das imagens sintéticas	63
5.5	Experimentos sobre a imagem do sensor ALOS PALSAR	66
5.6	Experimentos sobre a imagem do sensor IKONOS MS	69
6	CONCLUSÃO.....	73
	REFERÊNCIAS BIBLIOGRÁFICAS	78
	APÊNDICE A – CLASSIFICAÇÃO DAS IMAGENS SINTÉTICAS.....	81
A.1	– Imagem sintética 1DP.....	81
A.2	– Imagem sintética 2DP.....	85
A.2	– Imagem sintética 3DP.....	89
	APÊNDICE B – RESULTADOS DO TESTE DE WILCOXON Z CALCULADO	93
	APÊNDICE C – RESULTADOS DOS TEMPOS GLOBAIS	94
	APÊNDICE D – CLASSIFICAÇÃO IMAGEM ALOS PALSAR	95
	APÊNDICE E – CLASSIFICAÇÃO IMAGEM IKONOS MS	98
	APÊNDICE G – ESPAÇO ATRIBUTOS 2D – ALOS PALSAR	103

1 INTRODUÇÃO

Na literatura de Reconhecimento de Padrões há uma série de técnicas de classificação supervisionada bem conhecidas, podendo-se citar, por exemplo, os classificadores: *k*-vizinhos mais próximos (*k*-NN, do inglês *k-nearest neighbor*), máxima probabilidade a posteriori, perceptron multicamadas, árvore de decisão, e máquina de vetores suporte (SVM, do inglês *Support Vector Machine*) (WEBB; COPSEY, 2011).

Dentre esses classificadores, o SVM apresenta características como boa capacidade de generalização e convexidade da função objetivo (WEBB; COPSEY, 2011). Sendo um classificador supervisionado, o SVM exige, na fase de treinamento, um conjunto de dados rotulados. Apesar das qualidades mencionadas, o procedimento de treinamento não é adequado para trabalhar com grandes conjuntos de dados por exigir a solução de um problema de programação quadrática para construir o modelo do classificador, o que implica em altos custos de memória e tempo de processamento.

Alternativas têm sido propostas para viabilizar a utilização do SVM para grandes conjuntos de dados, podendo ser divididos em dois grupos: o primeiro propõe modificações no próprio algoritmo de treinamento do SVM e o segundo procura selecionar amostras representativas do conjunto de treinamento a serem utilizadas no algoritmo de treinamento padrão do SVM, pois mesmo no caso de utilização do conjunto completo de amostras, somente as amostras que estão na fronteira de separação das classes fazem parte da construção do modelo do classificador.

A elaboração deste trabalho é direcionada seguindo a segunda abordagem, reduzindo o tamanho do conjunto de treinamento, procurando manter as amostras que são representativas para a construção do modelo do classificador. Para realizar esta tarefa foram implementados alguns algoritmos com características e capacidades de redução diversificada, com o objetivo de reduzir o tempo de treinamento do classificador do classificador. O Teste de Wilcoxon foi utilizado para avaliar os resultados obtidos pelos conjuntos reduzidos, com aqueles obtidos com o conjunto completo. O modelo mais simples do classificador SVM envolve a solução de um problema quadrático, com custo

computacional relacionado com o tamanho do conjunto de treinamento. Para modelos mais completos, necessita-se, também de definição e sintonia de outros parâmetros, o que pode aumentar ainda mais este custo computacional.

Na literatura há uma série de métodos usados para reduzir o custo computacional do classificador k -NN (WILSON; MARTINEZ, 2000). Dentre eles podem ser citadas as propostas de Wilson (1972), de Devijver e Kittler (1982) e de Hart (1968). Os algoritmos de agrupamento (hierárquicos, sequenciais e baseados em otimização de funções) também podem ser utilizados para redução de dados (ROMERO, 2011).

Os classificadores SVM e k -NN são semelhantes em relação à importância dada para as amostras próximas à fronteira de separação entre classes no espaço de atributos. As amostras que não estão localizadas nessa região não são utilizadas ou têm menor importância para os classificadores. Tanto o k -NN quanto a modelagem do SVM são sensíveis ao tamanho do conjunto de amostras de treinamento.

O k -NN, sendo um classificador *lazy learning*, não constrói um modelo para o classificador, mas utiliza diretamente as amostras de treinamento para classificar um dado objeto. Já o SVM faz uso de um modelo (função) para realizar a classificação, cujos parâmetros são obtidos a partir do conjunto de treinamento.

O objetivo principal deste trabalho constitui-se da aplicação e avaliação dos métodos para reduzir o tamanho do conjunto de dados visando diminuir o tempo total de treinamento e, concomitantemente, manter ou melhorar a acurácia de classificação do classificador SVM ao escolher as amostras mais representativas.

De modo geral, são relatados resultados sobre redução do conjunto de treinamento, o diferencial deste trabalho é um detalhamento da análise do tempo de redução do conjunto de treinamento adicionado ao tempo de estimação dos parâmetros e um modo de avaliar se as acurácias obtidas são significativamente aceitáveis.

2 FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo, será apresentado o classificador SVM, os métodos utilizados para realizar a classificação nos casos com mais de duas classes e a análise sobre o custo computacional.

2.1 O classificador SVM

Dentre o conjunto existente de técnicas para fins de classificação, algumas se destacam, entre elas estão as máquinas de vetores suporte (SVM), devido, principalmente, ao resultado satisfatório obtido ao final do processo, e pela robustez que apresenta, mesmo com poucos dados de treinamento, se comparado a outros classificadores.

As técnicas de aprendizado de máquina fazem uso do princípio de inferência chamado indução, do qual é possível obter conclusões genéricas a partir de entradas fornecidas. Esse aprendizado indutivo pode ser de dois tipos: supervisionado ou não supervisionado. No aprendizado supervisionado há a necessidade de um conjunto de amostras (rotuladas) que fornece informações para que a máquina possa aprender. Essas amostras estão contidas em um conjunto de entrada, com sua respectiva saída conhecida. O algoritmo, a partir desses exemplos, extrai o conhecimento necessário, com o objetivo de poder generalizar para futuras entradas as saídas esperadas.

O modelo básico do classificador supervisionado SVM consiste em resolver o problema de classificação binário (duas classes: ω_1 e ω_2) definindo, no espaço de atributos, uma superfície de decisão linear (hiperplano) dada por:

$$f_{12}(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 = 0 \quad (2.1)$$

onde $\mathbf{x} \in \mathbb{R}^p$ é o vetor de atributos que descreve o objeto analisado, \mathbf{w} define a orientação do hiperplano de separação, w_0 fornece a posição do hiperplano e p o número de atributos.

Sendo o SVM um classificador supervisionado, seu treinamento é realizado com base em um conjunto de n amostras rotuladas dado por $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$, onde $y_i \in \{+1, -1\}$ e com $(\mathbf{x}, +1) \in \omega_1$ e $(\mathbf{x}, -1) \in \omega_2$. A partir do conjunto de treinamento, os parâmetros do modelo (\mathbf{w} e w_0) são calculados e a regra de classificação pode ser posta como:

$$\text{Se } f_{12}(\mathbf{x}) \geq 0 \text{ então } \mathbf{x} \in \omega_1 \text{ senão } \mathbf{x} \in \omega_2 \quad (2.2)$$

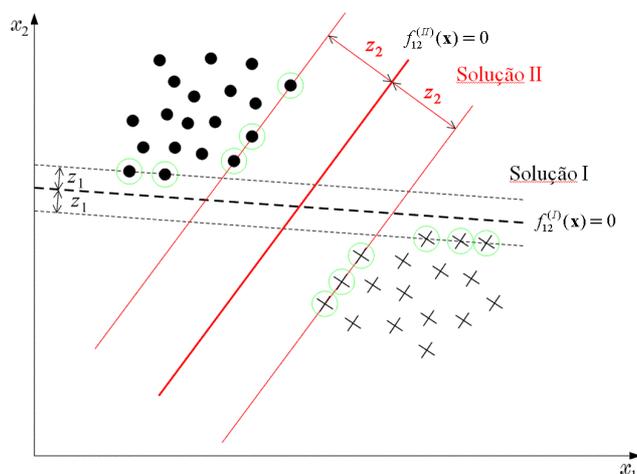
O modelo linear para o SVM é aplicado a dois casos: separável e não separável. No caso separável, existe uma solução que permite classificar os dados de treinamento sem a ocorrência de erros. No caso não separável não é possível realizar a classificação sem ocorrência de erros, mas ainda assim o modelo linear é considerado adequado. Há ainda o modelo do SVM mais geral para tratar o caso em que as superfícies de decisão são não lineares. Todos estes modelos tem em comum a característica de gerarem soluções que maximizam a margem entre as classes no espaço de atributos. A margem é definida como a maior região localizada na fronteira de decisão, definida pela maior distância entre as amostras pertencentes a ω_1 e as amostras pertencentes a ω_2 e o hiperplano de separação das classes. As próximas subseções abordam cada um destes modelos.

2.1.1 Modelo SVM linearmente separável

O modelo do SVM para o caso linearmente separável define como solução ótima, o hiperplano que maximiza a *margem* entre as duas classes.

A Figura 2.1 ilustra o conceito de margem para o caso linearmente separável, exibindo duas soluções para um problema de classificação com dois atributos (x_1 e x_2) e duas classes linearmente separáveis (“•” e “×”). Neste caso as superfícies de decisão ($f_{12}^{(I)}(\mathbf{x})=0$ e $f_{12}^{(II)}(\mathbf{x})=0$) e os limites das margens são retas, sendo a solução (II) aquela que maximiza a margem ($2z_2$).

Figura 2.1 -Exemplo com duas classes linearmente separáveis e duas soluções lineares possíveis com as respectivas margens.



Fonte: Adaptada de Theodoridis e Koutroumbas (2008).

Os pontos de ambas as classes localizados sobre os limites da margem, como mostra a Figura 2.1, são designados como *vetores de suporte* e são os pontos utilizados para a definição da superfície de decisão do classificador, no caso em que as classes são linearmente separáveis.

A distância entre os pontos localizados nos limites da margem e o hiperplano de decisão é o elemento essencial para encontrar a solução desejada (de máxima margem). A distância entre uma amostra \mathbf{x}_i qualquer e o hiperplano de separação $f_{12}(\mathbf{x}) = 0$ é calculada por $|f_{12}(\mathbf{x}_i)| / \|\mathbf{w}\|$.

Se reescalarmos \mathbf{w} de maneira que $f_{12}(\mathbf{x}_i) = +1$ e $f_{12}(\mathbf{x}_j) = -1$, para $\mathbf{x}_i \in \omega_1$ e $\mathbf{x}_j \in \omega_2$, sendo estas as amostras mais próximas do hiperplano $f_{12}(\mathbf{x}) = 0$, são obtidas as seguintes relações:

- (i) a margem tem largura $\frac{2}{\|\mathbf{w}\|}$;
- (ii) se $\mathbf{x}_i \in \omega_1$, então $f_{12}(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i + w_0 \geq +1$;
- (iii) se $\mathbf{x}_j \in \omega_2$, então $f_{12}(\mathbf{x}_j) = \mathbf{w}^T \mathbf{x}_j + w_0 \leq -1$.

Essas relações permitem concluir que a maximização da margem é obtida com a

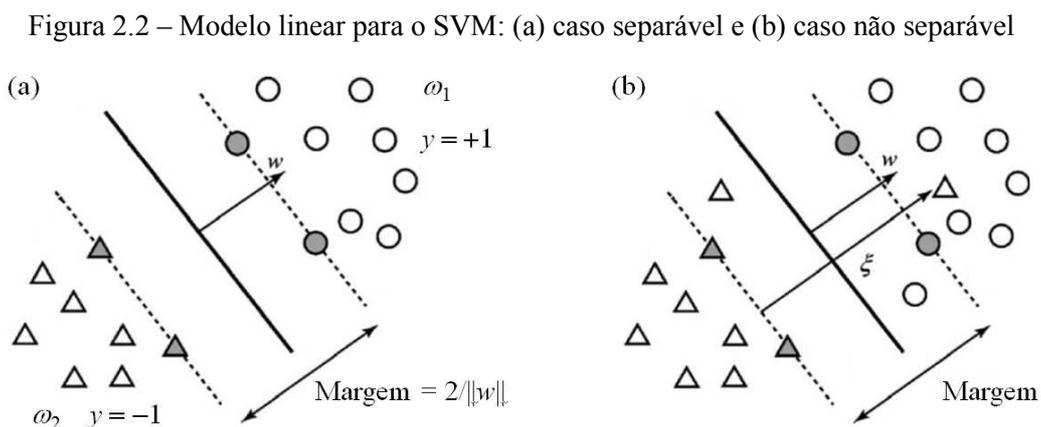
determinação de \mathbf{w} e w_0 que minimizam $\|\mathbf{w}\|$ desde que as relações (ii) e (iii) sejam satisfeitas. Trata-se de um problema de otimização que pode ser posto como:

$$\begin{aligned} & \min \left(\frac{1}{2} \|\mathbf{w}\|^2 \right) \\ & \mathbf{w}, w_0 \\ & \text{sujeito a: } y_i (\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1, \forall i = 1, \dots, n \end{aligned} \tag{2.3}$$

A solução deste problema permite criar a função discriminante $f_{12}(\mathbf{x})$ dada pela Equação 2.1 e construir o classificador apresentado em (2.2).

2.1.2 Modelo SVM não linearmente separável

É muito comum os dados não serem linearmente separáveis, ou seja, não existir solução sem erros de classificação sobre os dados de treinamento. A Figura 2.2 (a) mostra o caso linearmente separável, onde foi possível separar sem erros todas as amostras de cada classe e a Figura 2.2 (b) mostra o caso em que os dados não são linearmente separáveis. Neste caso a solução desejada deve ser obtida pela maximização da margem e pela penalização dos pontos dentro da margem, ou classificados erroneamente.



Fonte: Adaptada de Foody (2006).

Para tanto as seguintes restrições são impostas: $y_i (\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1 - \xi_i$, $\forall i = 1, \dots, n$ e onde $\xi_i \geq 0$ é a variável que quantifica a penalidade associada ao exemplo \mathbf{x}_i . A solução é obtida pela minimização da seguinte função objetivo:

$$\min_{\mathbf{w}, w_0, \xi_i} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \right) \quad (2.4)$$

$$\text{sujeito a: } \xi_i \geq 0 \text{ e } y_i (\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1 - \xi_i, \forall i = 1, \dots, n$$

onde o parâmetro $C > 0$ controla o compromisso entre a margem de separação e as penalizações ocorridas. Tal problema de otimização pode ser resolvido com a introdução da função Lagrangiana e posterior solução do problema dual, que implica em maximizar as variáveis α_i e minimizar \mathbf{w} e w_0 . O problema pode ser posto como

$$\max_{\alpha} \left(\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j) \right) \quad (2.5)$$

$$\text{sujeito a } \sum_{i=1}^n \alpha_i y_i = 0 \text{ e } 0 \leq \alpha_i \leq C, \forall i = 1, \dots, n$$

onde α_i é o multiplicador de Lagrange associado a i -ésima restrição. A solução do problema será dada por:

$$\mathbf{w} = \sum_{i \in SV1} \alpha_i y_i \mathbf{x}_i \text{ e } w_0 = \frac{1}{|SV2|} \sum_{i \in SV2} \alpha_i (y_i - \mathbf{w}^T \mathbf{x}_i) \quad (2.6)$$

onde $SV1$ é o conjunto de índices para todos os vetores de suporte, $SV2$ é o conjunto de índices para os vetores de suporte que estão sobre os limites da margem e $|SV2|$ a cardinalidade de $SV2$. O problema apresentado em (2.5) mostra claramente a dependência com relação ao número de amostras (n). Finalmente o classificador será dado pela seguinte regra:

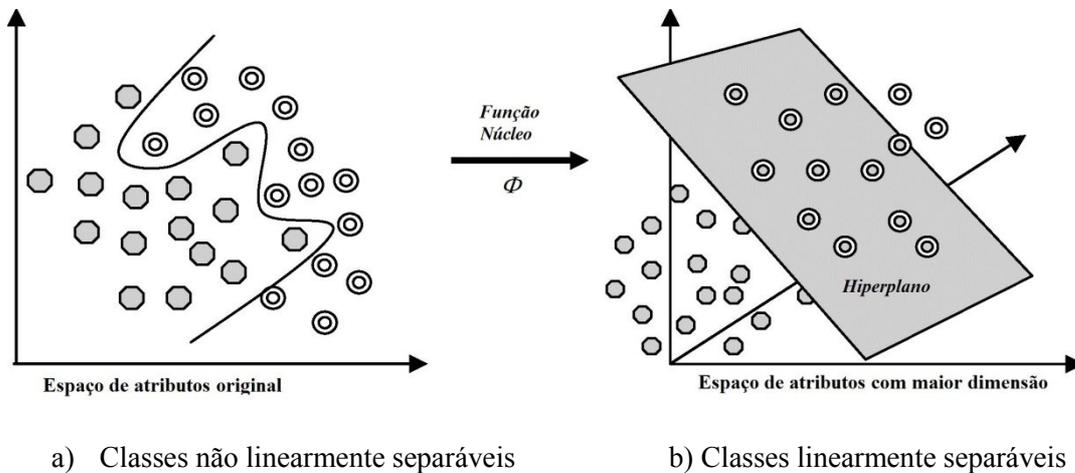
$$\text{Se } f_{12}(\mathbf{x}) = (\mathbf{w}^T \mathbf{x} + w_0) \geq 0 \text{ então } \mathbf{x} \in \omega_1 \text{ senão } \mathbf{x} \in \omega_2 \quad (2.7)$$

Percebe-se, a partir de (2.6) e (2.7), que a expressão final do classificador não depende, diretamente, do número de vetores de suporte obtido no treinamento, ou seja, \mathbf{w} e w_0 são expressos em valores explícitos.

2.1.3 Modelo SVM não linear

Nas duas seções anteriores foram descritos modelos lineares para o SVM. Porém, existem situações em que modelos não lineares são mais adequados para solucionar o problema de classificação. A solução, no caso do SVM, consiste em realizar o mapeamento dos padrões, não linearmente separáveis no espaço original, para outro espaço de atributos onde o problema se torna linear. A Figura 2.3 ilustra a situação, onde a função $\Phi(\cdot)$ realiza o mapeamento entre os espaços.

Figura 2.3- Exemplo de mapeamento dos padrões para um espaço de maior dimensão.



Fonte: Adaptada de Kavzoglu e Colkesen (2009).

A função discriminante que no caso linear é dada pela Equação (2.1) passa a ser:

$$f_{12}(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + w_0 \quad (2.8)$$

Desta maneira, as formulações (2.5) e (2.6) com a aplicação de um mapeamento para uma dimensão maior passam a ser (2.9) e (2.10), respectivamente:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \quad (2.9)$$

$$\text{sujeito a} \quad \sum_{i=1}^n \alpha_i y_i = 0 \text{ e } 0 \leq \alpha_i \leq C, \forall i = 1, \dots, n$$

$$\mathbf{w} = \sum_{i \in SV1} \alpha_i y_i \Phi(\mathbf{x}_i) \text{ e } w_0 = \frac{1}{|SV2|} \left\{ \sum_{i \in SV2} y_i - \sum_{i \in SV1, j \in SV2} \alpha_i y_i \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \right\} \quad (2.10)$$

Em geral o mapeamento $\Phi(\cdot)$ é desconhecido. A solução reside em substituir o produto escalar $\Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$ por uma função núcleo (do inglês *kernel function*) dada por:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \quad (2.11)$$

que realiza, ao mesmo tempo, o mapeamento e o produto escalar no espaço mapeado.

Substituindo a função núcleo no problema formulado em (2.8) obtêm-se uma nova expressão para (2.9) e (2.10). A nova função discriminante passa a ser:

$$f_{12}(\mathbf{x}) = \sum_{i \in SV1} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + w_0 \quad (2.12)$$

Contrariamente ao modelo linear, cujos parâmetros são calculados explicitamente, a Equação 2.12 mostra a dependência dos vetores de suporte, que implicitamente permitem calcular o vetor de pesos da função discriminante.

Há vários tipos de função núcleo que podem se utilizadas no SVM. As mais comumente utilizadas são a linear, a polinomial e a função de base radial (RBF, do inglês *Radial Basis Function*) definidas respectivamente pelas Equações 2.13-2.15 abaixo:

$$K_{\text{Linear}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j) \quad (2.13)$$

$$K_{\text{Polinomial}}(\mathbf{x}_i, \mathbf{x}_j) = ((\mathbf{x}_i^T \mathbf{x}_j) + 1)^q \quad (2.14)$$

$$K_{\text{RBF}}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-|\mathbf{x}_i - \mathbf{x}_j|^2 / \sigma^2) \quad (2.15)$$

Para ser considerada função núcleo, uma função deve satisfazer as condições do teorema de Mercer (WEBB;COPSEY, 2011).

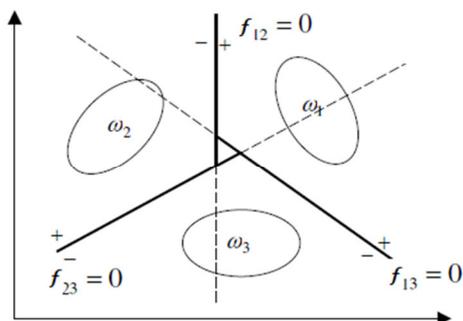
2.2 Classificação Multiclasses

O SVM é um modelo de classificação binária. No caso mais geral em que há mais de duas classes, vários métodos têm sido propostos para resolver esta questão. As abordagens mais tradicionais permitem a criação de um classificador multiclasse combinando vários classificadores binários, como nos casos do SMO (PLATT, 1999) e do SVM light (JOACHIMS, 1999). Alguns autores também propuseram métodos que consideram todas as classes ao mesmo tempo (HSU, 2002). Dentre os métodos baseados em classificações binárias, podem ser citados (HSU, 2002): um contra um (OAO, do inglês *One Against One*), um contra todos (OAA, do inglês *One Against All*) e Grafo acíclico direcionado (DAGSVM, do inglês *Directed Acyclic Graph Support Vector Machine*).

2.2.1 OAO – Método de decomposição um contra um

O método de decomposição OAO, consiste na construção do classificador SVM para cada par de classes. Desta maneira, para um dado problema de c classes, obtêm-se $\frac{1}{2}c(c-1)$ classificadores binários gerados. Cada um deles responsável por diferenciar um par de classes (ω_1 e ω_2), em que $\omega_1 \neq \omega_2$. Assim, uma amostra pode ser classificada em uma dada classe diversas vezes. Normalmente, a classificação de uma amostra desconhecida é feita de acordo com a votação máxima, onde cada classificador binário SVM vota em uma classe. A classe ganhadora será a majoritária, e no caso de empate a que tiver maior valor absoluto da função discriminante. A Figura 2.4 mostra a superfície de decisão para um problema com três classes e dois atributos.

Figura 2.4 – Superfície de decisão para o método OAO

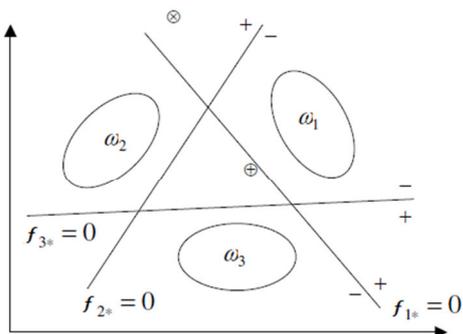


Fonte: Adaptada de Abe (2003).

2.2.2 OAA – Método de decomposição um contra todos

No método de decomposição de classes OAA, dado um problema com c classes, c classificadores binários são gerados. Cada um desses classificadores é treinado de forma a distinguir uma das classes das demais. A classe ganhadora será a majoritária, e no caso de empate a que tiver maior valor absoluto no resultado da função discriminante pelo conjunto das outras classes. Na Figura 2.5 a superfície de decisão com $f_{1*}=0$ separa a classe ω_1 das classes ω_2 e ω_3 , e assim sucessivamente para as classes ω_2 e ω_3 . A Figura 2.5 mostra as superfícies de decisão para um problema com três classes e dois atributos.

Figura 2.5 – Superfície de decisão para o método OAA



Fonte: Adaptada de Abe (2003).

A decomposição OAA pode apresentar desvantagens quando a proporção de amostras de uma classe for relativamente pequena comparada à quantidade de amostras formada pela união das demais classes.

Esse tipo de desbalanceamento pode dificultar a indução de um classificador geral que apresente bom desempenho no reconhecimento da classe considerada.

2.3 Seleção de modelos e parâmetros

Na resolução do problema quadrático, a medida que a quantidade de amostras de treinamento envolvidas aumenta, aumenta exponencialmente o custo computacional. Como visto na subseção 2.2, referente à classificação com mais de duas classes, têm-se no mínimo segundo a literatura atual duas opções de resolução, OAO ou OAA. Contudo, para realizar o treinamento do SVM com uma solução mais refinada, exige-se também a estimação da penalidade (C), além da seleção da função núcleo e dos parâmetros nela envolvidos.

Nota-se então que quanto mais refinada, ou seja, mais calibrada seja a utilização das variáveis envolvidas para o cálculo da função do hiperplano separador das classes, maior será o custo computacional para resolvê-lo.

Seguindo este raciocínio e destacando algumas variáveis comentadas nos capítulos anteriores, têm-se os seguintes elementos que fazem parte do processo de definição do modelo do classificador SVM:

M – Problema de classificação binário. No caso de três ou mais classes, m problemas serão gerados, sendo o valor de m dependente do método de decomposição utilizado.

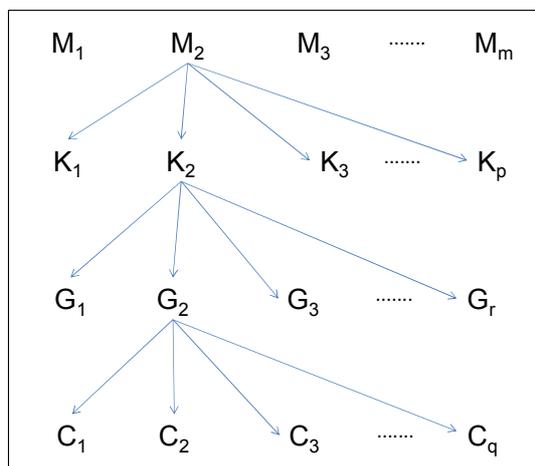
K – Função núcleo, que realiza o mapeamento dos dados para outro espaço, além do produto escalar neste novo espaço. Há várias opções de função núcleo (Linear, RBF, Polinomial, ...).

G - Parâmetro da função núcleo. Cada função núcleo tem pelo menos um parâmetro (exceto o núcleo linear) e que deve ser estimado.

C – Penalidade: O melhor valor de C depende da quantidade de amostras utilizadas, da função núcleo (*Kernel*) adotada, da quantidade de classes envolvidas no problema e do método proposto para tratar o caso com mais de duas classes. Desta maneira, encontrar um valor adequado demanda tempo e experimentos.

A Figura 2.6 mostra um esquema que pode possuir m classificadores binários gerados pelo método multiclases adotado, p tipos de funções núcleo, r valores de parâmetros de função núcleo e q valores para a penalidade (C). O custo computacional total para avaliar estas variáveis envolveriam $m \times p \times r \times q$ problemas quadráticos.

Figura 2.6 – Modelos possíveis para m problemas binários, com p núcleos e q penalidades



Fonte: Produção do autor.

2.4 Alternativa para solução do Problema Quadrático

O método adotado para solução de problemas de otimização quadrática pertence à biblioteca IMSL (*International Mathematics and Statistics Library*). Sua descrição é apresentada em POWEL (1985). Este método está implementado em linguagem C (IMSL C *Numerical Library*), mas está disponível como uma função da biblioteca IDL (IMSL_QUADPROG) no módulo *Advanced Math and Stats*.

A adoção deste método foi considerado após análise realizada sobre os resultados utilizando o classificador SVM com o método SMO. O SMO apresentou instabilidades no tempo gasto para solucionar o problema. Essas instabilidades poderiam comprometer os resultados para várias execuções, já que o foco do trabalho é justamente o tempo de treinamento. O método para solução do problema quadrático disponível no IMSL não é tão rápido quanto os outros métodos (SMO e SVM^{LIGHT}), porém ele não apresenta oscilações.

A função IMS_QUADPROG foi encapsulada num procedimento que restringe as suas funcionalidades àquelas que são necessárias apenas para a solução do problema de treinamento do SVM. Desta forma, a visão que o usuário tem é a de uma função que resolve o problema apresentado na Equação 2.10, incluindo a função núcleo como apresentada na Equação 2.12.

2.5 ALGORITMOS DE REDUÇÃO

Nesta seção serão apresentados o classificador k -vizinhos mais próximos (k -NN), devido à utilização em alguns algoritmos, os métodos de redução do conjunto de treinamento por edição de Wilson repetida (RENN), edição múltipla de Devijver e Kittler (ME), condensação de Hart (CNN), Seleção aleatória (RND), e o algoritmo de agregação sequencial (LID).

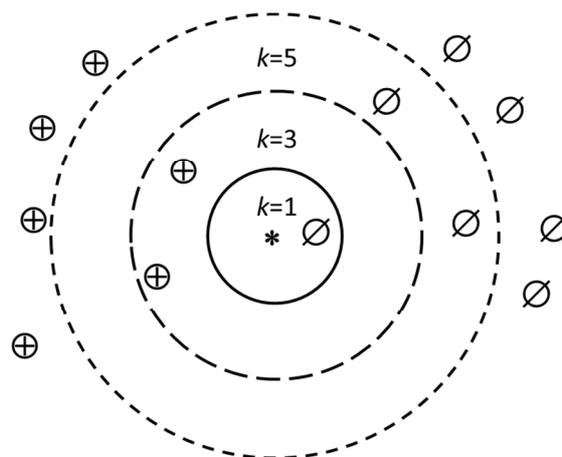
2.5.1 Classificador vizinho mais próximo

O algoritmo k -vizinhos mais próximos (k -NN) é um método para classificar objetos com base em exemplos de treinamento mais próximos no domínio do espaço de atributos (WEBB;COPSEY, 2011). O algoritmo de vizinho mais próximo está entre os mais simples algoritmos de aprendizagem, pois um objeto é classificado pelo voto da maioria de seus vizinhos, sendo atribuído à classe mais comum entre os seus k vizinhos mais próximos.

Para exemplificar o método, toma-se como exemplo a classificação do objeto no centro do círculo da Figura 2.7. Se $k = 1$, a classe escolhida será a \emptyset . Se $k = 3$, será a classe \oplus . Se $k = 5$ será a classe \emptyset novamente.

No caso de empate adota-se $k-1$.

Figura 2.7 – Exemplificação para o método k -NN.



Fonte: produção do autor

Os passos principais do k -NN são mostrados na Figura 2.8.

Figura 2.8 – Algoritmo k -NN.

<p>Entrada: conjuntos S_T (treinamento) e S_C (a classificar). Saída: conjunto S_C classificado. Passo 1 – Para cada amostra \mathbf{x} do conjunto S_C: Encontrar os k vizinhos mais próximos de \mathbf{x}, em S_T. Classificar \mathbf{x} para a classe majoritária entre os k pontos selecionados. Passo 2 – Retornar o conjunto S_C classificado.</p>
--

A medida de proximidade comumente utilizada é a distância Euclidiana. O valor de k deve ser escolhido de forma a maximizar a acurácia de classificação. Alguns dos algoritmos de redução dos dados de treinamento adotados neste trabalho utilizam o método k -NN para a seleção de amostras.

2.5.2 Edição de Wilson Repetida - RENN

O algoritmo de edição de Wilson (RENN, do inglês *Repeat Editing Nearest Neighbor*) é um dos primeiros métodos de edição propostos na literatura, para reduzir o conjunto de treinamento. Isto é realizado pela regra do vizinho mais próximo, através da eliminação de amostras classificadas erroneamente. O modelo proposto por Wilson (1972) pode ser resumido da seguinte maneira: se uma amostra é erroneamente classificada pelo método k -NN, onde k representa o número de vizinhos mais próximos, ela será removida do conjunto de treinamento. O método de estimativa de erro utilizado neste algoritmo corresponde ao *leaving-one-out*: um ponto é separado do conjunto para teste e o restante é utilizado para gerar o modelo. O algoritmo RENN é descrito na Figura 2.9.

Figura 2.9 – Algoritmo RENN.

Entrada: conjunto de treinamento S , número de vizinhos k . Saída: conjunto editado $S_{\text{RENN}} \subseteq S$. Passo 1 – Classificar cada amostra x do conjunto S com o k -NN. Passo 2 – Eliminar de S todas as amostras classificadas incorretamente. Passo 3 – Repetir os Passos 1 e 2 até que não haja eliminação de amostras. Passo 4 – Retornar S_{RENN} .
--

2.5.3 Edição múltipla de Devijver e Kittler – ME

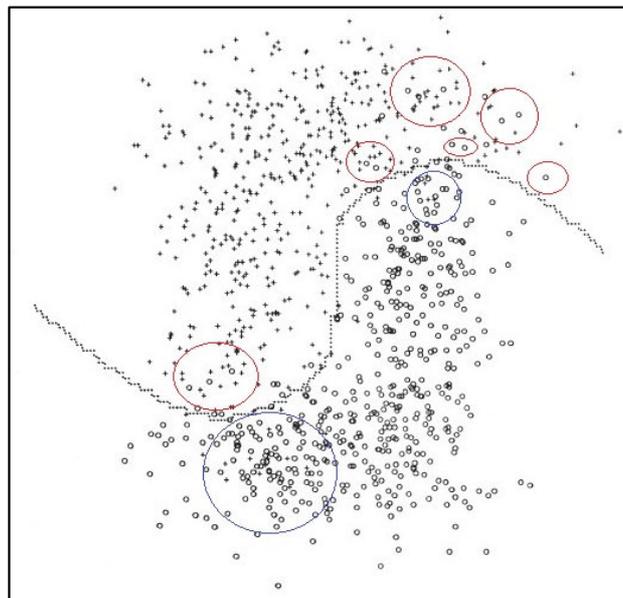
O algoritmo de edição múltipla (ME, do inglês *Multiedit*) de Devijver e Kittler (1982) é similar ao algoritmo de edição de Wilson, diferenciando-se pela maneira que as amostras são selecionadas para o treinamento e para a classificação. O método consiste em particionar o conjunto de treinamento em B blocos e, após o particionamento, classificar cada bloco b , usando o bloco $((b + 1) \text{ módulo } B)$ para o treinamento do classificador. Assim, a técnica ME permite fazer uso repetido da edição de dados para cada bloco. O algoritmo ME é descrito na Figura 2.10. O número de blocos é definido pelo usuário. Devijver e Kittler sugerem utilizar $B=3$ e $k=1$.

Figura 2.10 – Algoritmo ME.

Entrada: Conjunto de treinamento S e número de blocos B .
Saída: Conjunto $S_{ME} \subseteq S$.
Passo 1 – Particionar aleatoriamente S em B subconjuntos, $S_{(0)} \dots S_{(B-1)}$, com $B \geq 3$.
Passo 2 – Classificar $S_{(i)}$ utilizando o método 1-NN usando $S_{((i+1) \bmod B)}$ como um conjunto de treino, para todo $i = 0, \dots, B-1$.
Passo 3 – Descartar todas as amostras erroneamente classificadas no passo 2.
Passo 4 – Constituir um novo conjunto S , a partir conjuntos editados $S_{(0)}, \dots, S_{(B-1)}$.
Passo 5 – Encerramento: Se as últimas I iterações não produzirem nenhuma edição, então sair com o conjunto $S_{ME} = S$. Caso contrário, retornar ao passo 1.

O objetivo do algoritmo ME é eliminar do conjunto de treinamento, os pontos que são classificados erroneamente com base nas classes dos seus vizinhos mais próximos. A Figura 2.11 é um exemplo com duas classes e dois atributos que apresenta as amostras classificadas incorretamente, segundo a fronteira de decisão de Bayes. Considera-se aqui o classificador de Bayes que posiciona a fronteira de decisão de modo a minimizar a probabilidade de erro de classificação (WEBB; COPSEY, 2012).

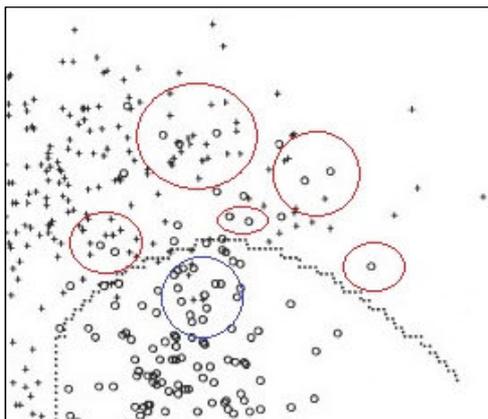
Figura 2.11 – Exemplo com duas classes e dois atributos. A curva pontilhada mostra a fronteira de decisão de Bayes.



Fonte: Adaptada de Devijver e Kittler (1982).

A Figura 2.12 apresenta um detalhe da região superior da Figura 2.11, onde é possível perceber alguns dos pontos de ambas as classes e que são classificados erroneamente pelo classificador Bayesiano. Esses pontos estão indicados pelas curvas em cor.

Figura 2.12 – Amostras classificadas erroneamente

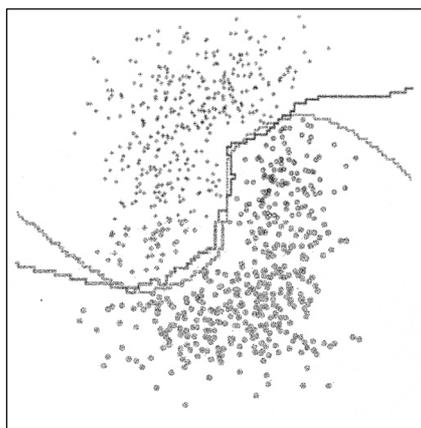


Fonte: Adaptada de Devijver e Kittler (1982).

Uma exemplificação do algoritmo de multiedição, é mostrado na Figura 2.13, que é resultado da aplicação do algoritmo sobre o conjunto de amostras da figura 2.11.

Figura 2.13 – Algoritmo ME – Dados selecionados após a última iteração

Linha sólida mostra a fronteira de decisão para o classificador 1-NN.



Fonte: Devijver e Kittler (1982)

2.5.4 Condensação de Hart – CNN

O método CNN (CNN, do inglês *Condensed Nearest Neighbor*) proposto por Hart (1968) é similar à edição de Wilson. O conceito de classificação de uma amostra de acordo com a sua vizinhança permanece, diferindo apenas na seleção das amostras que farão parte do conjunto de treinamento condensado. Se uma amostra é classificada incorretamente, de acordo com a sua vizinhança, pressupõe-se que ela esteja próxima da fronteira de decisão, pois é justamente nesta área que há, geralmente, a mistura entre as classes.

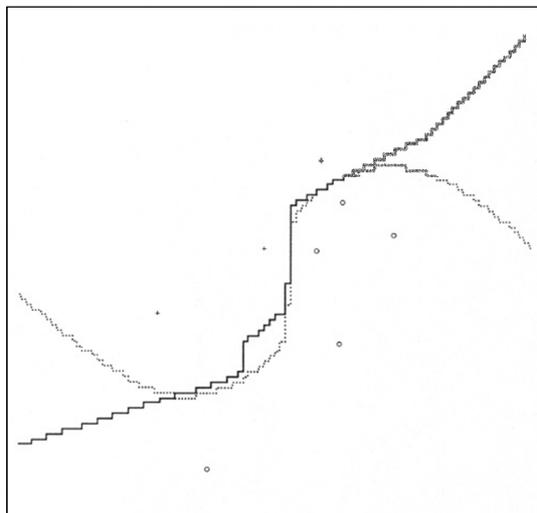
Para realizar a redução primeiramente transfere-se uma amostra para o conjunto S_{CNN} . A seguir percorre-se o conjunto de treinamento S classificando as demais amostras do restante do conjunto S utilizando S_{CNN} , como referência. Cada vez que uma amostra é classificada incorretamente, ela é transferida de S para S_{CNN} . Desta maneira o conjunto final condensado S_{CNN} terá somente as amostras classificadas incorretamente, ou seja, aquelas que estão próximas da fronteira de decisão. É possível perceber que o algoritmo é dependente da ordem em que os padrões são analisados. O algoritmo CNN é descrito na Figura 2.14.

Figura 2.14 – Algoritmo CNN

<p>Entrada: Conjunto de treinamento S Saída: Subconjunto condensado $S_{CNN} \subseteq S$ Passo 1 - Criar o conjunto S_{CNN} vazio. Passo 2 - Transferir aleatoriamente uma amostra do conjunto S para S_{CNN}. Passo 3 - Para cada x de S, fazer: Classificar x pela regra 1-NN usando o conjunto S_{CNN}. Se x é classificada incorretamente então transferir x de S para S_{CNN}. Passo 4 - Repetir o passo 3 até que: $S = \{\}$, ou não há transferências para S_{CNN}.</p>

A Figura 2.15 mostra uma aplicação do método CNN sobre um conjunto de treinamento editado pelo algoritmo ME.

Figura 2.15 – Aplicação do CNN aos dados gerados pelo ME apresentados na Figura 2.13



Fonte: Devijver e Kittler (1982)

Uma vantagem do CNN é a sua tendência a reduzir substancialmente o tamanho do conjunto de amostras. Por outro lado, esta pode ser uma desvantagem se o conjunto resultante não for representativo para caracterizar a fronteira de decisão do classificador.

2.5.5 Redução encadeada

Alguns métodos de redução descritos anteriormente, podem ser executados em sequência, como foi realizado por Devijver e Kittler encadeando o ME com o CNN, cujos resultados são apresentados nas Figuras 2.13 e 2.15, respectivamente.

Neste trabalho também serão estudados dois modelos em cascata: um aplica a redução pelo RENN seguido do CNN e o outro modelo aplica a redução ME seguida da redução pelo CNN. Ambos possuem potencial para reduzir consideravelmente o conjunto de treinamento com um acréscimo no tempo total de processamento para as reduções de dados.

2.5.6 Redução por Seleção Aleatória – RND

Adotou-se o método de seleção aleatória pela rapidez e facilidade de implementação. Este método será utilizado como referência, de conjunto reduzido, para comparação com os outros métodos. O método consiste em selecionar aleatoriamente uma fração do conjunto de dados de treinamento e utilizá-la para treinamento do classificador SVM.

Espera-se que este método seja um limitante inferior no tempo de processamento, devido a rapidez da seleção das amostras, e um limitante superior para o erro de classificação.

Figura 2.16 – Algoritmo RND

Entrada: Conjunto de treinamento S e percentual a selecionar.
Saída: Conjunto $S_{RND} \subseteq S$.
Passo 1 – Calcular o número de amostras (r) a partir do percentual definido.
Passo 2 – Selecionar aleatoriamente r amostras de S e colocá-las em S_{RND} .

2.5.7 Algoritmo Sequencial Líder - LID

Uma alternativa aos modelos descritos anteriormente, são os algoritmos de agrupamento (do inglês, *clustering algorithms*). Um dos mais simples e rápidos é o algoritmo de agrupamento sequencial, também conhecido como algoritmo Líder (HARTIGAN, 1975) e que é utilizado neste trabalho.

O objetivo do algoritmo sequencial Líder é a criação de grupos, a partir dos quais serão selecionadas as amostras que formarão o conjunto de treinamento reduzido. Os grupos são gerados sequencialmente. A primeira amostra retirada do conjunto é definida como centro (Líder) do primeiro grupo. Sequencialmente cada amostra x de S ou é alocada ao grupo mais próximo, utilizando a distância Euclidiana, ou passa a ser o centro de um novo grupo, dependendo de um limiar que define a extensão do grupo. O algoritmo é mostrado na Figura 2.17.

Figura 2.17 – Algoritmo Líder

Entrada: Conjunto de treinamento S , número máximo de grupos m e distância d
 Saída: Conjunto de grupos $\{S_i\}$ e de centros $\{c_i\}$, $i = 1, \dots, q \leq m$, $c_i \in S_i$ e $S_i \cap S_j = \emptyset \forall i \neq j$
 Passo 1 – Escolher aleatoriamente uma amostra x de S
 Passo 2 – Incluir x em S_1 e defini-lo como seu centro c_1 ($q=1$)
 Passo 3 – Enquanto houver x em S não classificado faça
 Escolher aleatoriamente uma amostra x de S
 Procurar o centro c_j mais próximo a x (distância euclidiana)
 Se distância $(x, c_j) \leq d$ então Incluir x em S_j
 Senão
 Se $q < m$ então $q = q+1$; Incluir x em S_q ; $c_q = x$
 Senão Incluir x em S_j

Alguns comentários sobre o método:

- É rápido, pois cada amostra é analisada somente uma vez;
- O resultado depende da ordem em que as amostras são analisadas;
- Não há método eficiente e universal para identificar o número de grupos;
- É sensível a valores atípicos (*outliers*), podendo gerar grupos isolados.

Parâmetros do algoritmo Líder

Os parâmetros de entrada do algoritmo são: o número máximo de grupos m e a distância máxima d para atribuição de elementos ao grupo. O objetivo do parâmetro m é impor um limitante superior para o número de grupos. Uma estimativa para d pode ser obtida com a equação 2.16:

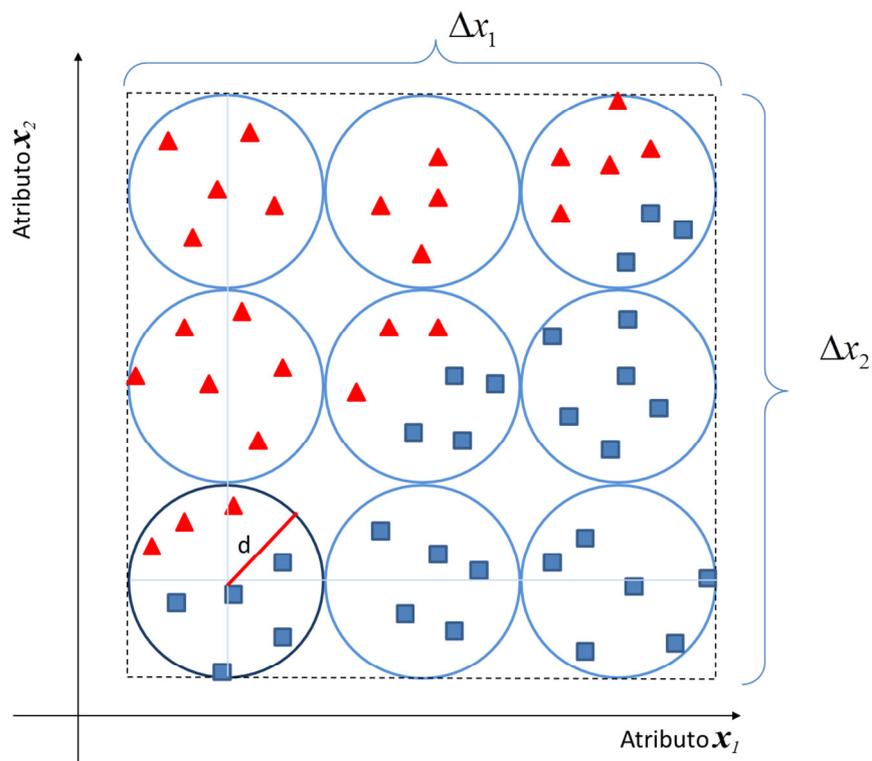
$$d = \frac{1}{2} \left(\frac{V_{hc}}{q} \right)^{\frac{1}{p}} \quad (2.16)$$

Onde q é a estimativa inicial do número de grupos desejado, p é número de atributos da amostra e V_{hc} é o volume do menor hiperparalelepípedo que contém todas as amostras de treinamento, calculado pela Equação 2.17.

$$V_{hc} = \prod_{i=1}^p \{\max(x_i) - \min(x_i)\} \quad (2.17)$$

O cálculo da distância foi selecionado visando a cobertura total das amostras no espaço de atributos e de uma maneira que tivesse o menor impacto no custo computacional. Um exemplo gráfico da estimação dos grupos para $p = 2$ atributos e $q = 9$ grupos é exibido na Figura 2.18.

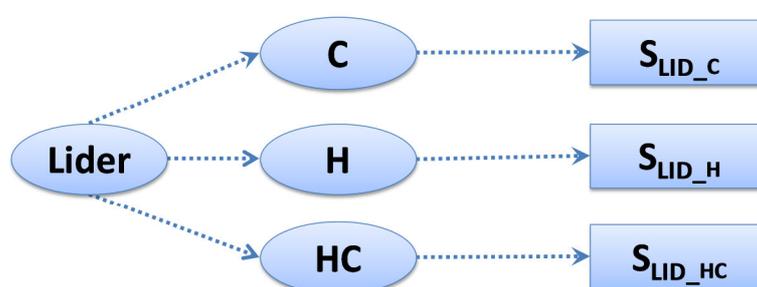
Figura 2.18 – Número total de grupos com todas as amostras.



Características dos grupos gerados pelo Algoritmo Líder

Analisando os grupos gerados pelo algoritmo Líder, nota-se que eles possuem características que permitem obter variações no conjunto de treinamento reduzido. Foram selecionadas três variações para realizar a composição do conjunto de treinamento reduzido. Elas são apresentadas na Figura 2.19 e estão associadas às características de homogeneidade e heterogeneidade dos grupos formados.

Figura 2.19 – Características dos grupos gerados pelo algoritmo Líder.



As três variações exploradas são as seguintes:

a) Centros dos Grupos (C): O conjunto de treinamento reduzido é composto apenas pelos centros de grupos e pode ser definido como:

$$S_{Lid_C} = \{c_i\}, i = 1, \dots, q \leq m, c_i \in S_i$$

b) Grupos Heterogêneos (H): O conjunto de treinamento reduzido é composto pelos grupos que possuem amostras de classes diferentes. Os grupos heterogêneos tendem a estar próximos da fronteira entre as classes. O conjunto pode ser definido como:

$$S_{Lid_H} = \{S_k\}, \forall k \mid \exists x_i, x_j \in S_k, x_i \in \omega_i \text{ e } x_j \in \omega_j, i \neq j$$

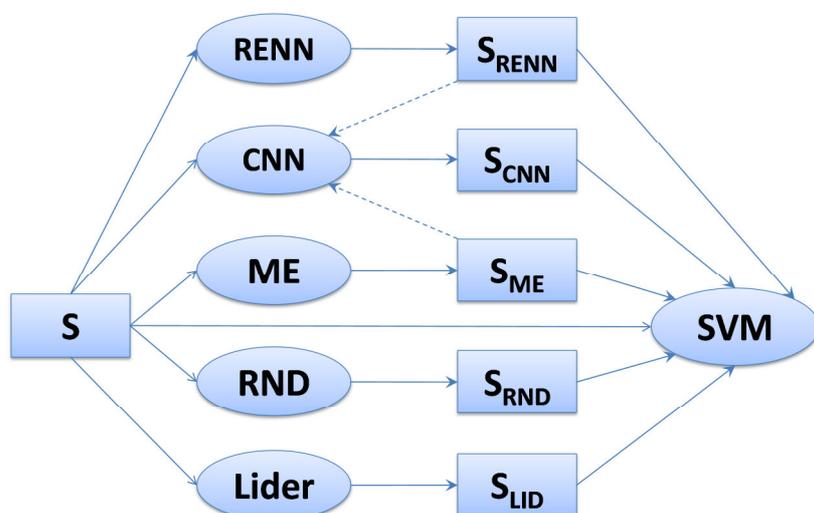
c) Grupos Heterogêneos e Centros (HC): O conjunto de treinamento reduzido é composto pelos grupos que possuem amostras de classes diferentes e pelos centros dos grupos homogêneos. Dessa maneira consegue-se cobrir a região interna e a fronteira das classes. O conjunto pode ser definido como:

$$S_{Lid_HC} = S_{Lid_H} \cup S_{Lid_C}$$

2.6 Estrutura de utilização dos algoritmos

No desenvolvimento do trabalho foram implementados cinco algoritmos de redução de dados, que foram testados individualmente, e em dois arranjos em cascata. O diagrama descrito na Figura 2.20 mostra os módulos de redução de dados e as possíveis combinações pretendidas neste trabalho para realizar a redução do conjunto de amostras inicial S e utilizá-lo como conjunto de treinamento para construir o classificador SVM. As setas indicam o fluxo dos processos.

Figura 2.20 – Processos de redução do conjunto de amostras



RENN – Edição de Wilson Repetida

CNN – Condensação de Hart

ME – Multiedição de Devijver e Kittler

RND – Seleção aleatória

LÍDER– Seleção por agrupamento sequencial

SVM – Classificador SVM

S – Conjunto de amostras

S_x – Conjunto obtido pelo alg. X

A linha tracejada indica a aplicação do algoritmo de redução CNN após a redução pelo algoritmo RENN ou ME. Estes procedimentos são nomeados como RENN+CNN e ME+CNN, respectivamente. Na análise dos resultados também foram considerados, para o algoritmo LÍDER, os três possíveis grupos gerados pelo algoritmo.

2.7 Método de Avaliação

Um dos métodos de avaliação conhecidos e que serve para analisar a qualidade da classificação é realizada a partir da construção da matriz de erros ou matriz de confusão. (STEHMAN; FOODY, 2009). Esta matriz apresenta a concordância/discordância entre a decisão (classe) fornecida por um especialista (referência), e aquela dada pelo programa computacional (classificação), para cada valor observado do conjunto de validação ou de teste.

As seguintes observações devem ser feitas sobre este método de avaliação:

- cada amostra pertence a apenas uma das classes disponíveis;
- não se considera a opção de rejeição pelo classificador (amostra não classificada);
- classes muito similares segundo os atributos disponíveis podem gerar um número grande de erros prejudicando a avaliação global da classificação.

A Tabela 2.1, apresenta o arranjo dos elementos na matriz de confusão para o problema de avaliação com c classes.

Tabela 2.1 – Matriz de Confusão para c classes

		REFERÊNCIA				Total
		1	2	...	c	
CLASSIFICAÇÃO	1	x_{11}	x_{12}	...	x_{1c}	x_{1+}
	2	x_{21}	x_{22}	...		x_{2+}
	⋮	⋮
	c	x_{c1}	x_{c2}	...	x_{cc}	x_{c+}
	Total	x_{+1}	x_{+2}	...	x_{+c}	n

Nesta tabela têm-se os seguintes elementos:

x_{ij} é número de amostras da classe j (referência), classificadas na classe i (classificação);

x_{ii} é número de amostras da classe i corretamente classificadas;

x_{+j} é número total de amostras da classe j segundo a referência;

x_{i+} é número total de amostras da classe i segundo o classificador;

n é o número total de amostras, do conjunto analisado.

Uma medida de concordância que pode ser obtida a partir da matriz de confusão, e que será utilizada neste trabalho, é a exatidão global, formulada pela equação 2.18.

$$\text{Exatidão Global} = \frac{1}{n} \sum_{k=1}^c x_{kk} \quad (2.18)$$

A acurácia da classificação utilizando os conjuntos de treinamento reduzidos, calculada pela exatidão global, é analisada posteriormente, por meio de teste de hipótese, comparada com a acurácia da classificação utilizando os dados completos do conjunto de treinamento.

2.8 Teste de hipótese

O teste de hipótese é um procedimento estatístico, baseado na análise de uma amostra, de modo que possa verificar se os dados são compatíveis com alguma hipótese pré-estabelecida, sendo constituído por duas hipóteses, a hipótese nula H_0 e a hipótese alternativa H_1 , que estabelece:

Hipótese nula (H_0): é a hipótese que traduz a ausência do efeito que se quer verificar.

Hipótese alternativa (H_1): é a hipótese que se quer verificar.

A utilização do teste de hipótese para este trabalho visa analisar a acurácia da classificação utilizando dados de treinamento do conjunto de amostras reduzidos pelos métodos de redução, comparada a acurácia obtida pelo classificador SVM utilizando o conjunto de dados de treinamento completo. O caso proposto, traduz-se em:

H_0 = Acurácia obtida pela classificação SVM utilizando o método de redução é igual à acurácia obtida pelo SVM utilizando os dados do conjunto de treinamento completo;

H_1 = Acurácia obtida pela classificação SVM utilizando o método de redução é menor que a acurácia obtida pelo SVM utilizando os dados do conjunto de treinamento completo.

H_{1*} = Acurácia obtida pela classificação SVM utilizando o método de redução é maior que a acurácia obtida pelo SVM utilizando os dados do conjunto de treinamento completo.

A ocorrência da hipótese alternativa H_{1*} é um resultado positivo, ou seja, houve melhora da acurácia quando o método de redução é utilizado. Desta maneira, interessam os resultados onde não se rejeita H_0 e H_{1*} .

A escolha do teste de hipótese adequado depende das seguintes características:

- Para cada método de redução são executadas 30 simulações
- As amostras são pareadas, pois as amostras do conjunto de teste são as mesmas utilizadas para o cálculo da acurácia em todos os modelos de redução.

Neste caso foi adotado o teste de Wilcoxon, descrito na seção seguinte.

2.8.1 Teste de Wilcoxon

O teste de Wilcoxon é um método estatístico não paramétrico para comparação de amostras pareadas (SIEGEL, 1978). Ele pode ser utilizado para os casos em que não seja possível a aplicação do Teste t para amostras pareadas. Um dos motivos da escolha foi referente à baixa quantidade de amostras (execuções) e a característica dos dados de serem pareados.

O teste é realizado sobre dois conjuntos A e B, o primeiro contendo os resultados das execuções do classificador SVM utilizando o conjunto de treinamento completo de amostras, e o segundo contendo os resultados das execuções do classificador SVM utilizando conjunto de treinamento reduzido pelo método de redução proposto. O teste é baseado na comparação sequencial par a par entre os elementos de cada conjunto, ou seja, o primeiro resultado do conjunto A é comparado com o primeiro resultado do

conjunto B, e assim sucessivamente para todos os elementos. Este teste, em vez de considerar somente o sinal das diferenças entre os pares, característica do Teste dos sinais, considera também o valor absoluto dessas diferenças, atribuindo maior ponderação a um par que tenha maior diferença entre as condições, do que a um par em que essa diferença seja pequena.

Um resumo do procedimento, de acordo com Siegel (1978), pode ser dado pelos seguintes passos:

- 1 – Para cada par, determinar a diferença (d_i), com sinal;
- 2 – Atribuir um posto a essas diferenças, independentemente do sinal. No caso de empate entre as diferenças, atribuir a média dos postos empatados;
- 3 – Atribuir a cada posto o sinal + ou sinal – da diferença que ela representa;
- 4 – Considerando que T^+ é a soma dos postos com sinal positivo e T^- é a soma dos postos com sinal negativo, atribuir a T o menor valor entre $|T^+|$ e $|T^-|$;
- 5 – Determinar N a partir da contagem do total de diferenças, descartando as diferenças nulas.
- 6 – Para $N > 25$, calcular o valor de z pela equação 2.19. Determinar sua probabilidade associada, sob H_0 , mediante referência a uma tabela de probabilidades para uma Distribuição Normal.

Se existirem apenas diferenças aleatórias, tal como é postulado pela hipótese nula, então haverá aproximadamente o mesmo número de ordens superiores e de ordens inferiores tanto para as diferenças positivas como negativas. Caso se verificar uma preponderância de baixos resultados para um dos lados, isso significa a existência de muitos resultados elevados para o outro lado, indicando uma diferença em favor de uma das situações, superior àquilo que seria de esperar se os resultados se devessem ao acaso.

O teste padrão de Wilcoxon é bilateral, para o problema unilateral deve ser observada a seguinte regra:

Para grandes amostras ($N > 25$), a estatística T pode ser aproximada para uma normal. Nesse caso, utiliza-se a estatística z , calculada através da equação 2.19:

$$z = \frac{T - \mu_T}{\sigma_T} = \frac{T - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}} \quad (2.19)$$

Esta aproximação foi utilizada neste trabalho, uma vez que a quantidade de experimentos realizados (N) foi de 30 execuções para cada método de redução.

Na Tabela 2.2 têm-se um exemplo de aplicação do Teste de Wilcoxon, considerando $\alpha = 5\%$, Teste Unilateral e z crítico = -1,645. Utilizando $T=148,5$ (menor entre $T+$ e $T-$) e $N=28$ (quantidade de diferenças não nulas) para calcular z .

Tabela 2.2 – Exemplo da aplicação do Teste de Wilcoxon – Imagem IKONOS

	Acurácia		diferença (<i>d</i>)	Posto de <i>d</i>	Posto com sinal menos frequente	Posto com sinal mais frequente
	SVM	LIDER_H_10				
1	97,83	97,90	-0,07	-12,5		-12,5
2	97,60	97,70	-0,10	-18,5		-18,5
3	97,50	97,53	-0,03	-2,5		-2,5
4	97,53	97,23	0,30	27	27	
5	97,47	97,67	-0,20	-25		-25
6	97,37	97,43	-0,07	-7,5		-7,5
7	97,40	97,23	0,17	22	22	
8	97,57	97,67	-0,10	-18,5		-18,5
9	97,43	97,30	0,13	20,5	20,5	
10	97,70	97,80	-0,10	-15,5		-15,5
11	97,40	97,50	-0,10	-15,5		-15,5
12	97,23	97,20	0,03	2,5	2,5	
13	97,50	97,63	-0,13	-20,5		-20,5
14	97,27	97,23	0,03	6	6	
15	97,37	97,37	0,00			
16	97,60	97,40	0,20	23	23	
17	97,50	97,30	0,20	25	25	
18	97,37	97,43	-0,07	-7,5		-7,5
19	97,40	97,50	-0,10	-15,5		-15,5
20	97,57	97,60	-0,03	-2,5		-2,5
21	97,67	97,67	0,00			
22	97,33	97,63	-0,30	-28		-28
23	97,57	97,60	-0,03	-2,5		-2,5
24	97,90	97,97	-0,07	-10		-10
25	97,70	97,77	-0,07	-10		-10
26	97,60	97,63	-0,03	-5		-5
27	97,47	97,40	0,07	10	10	
28	97,33	97,53	-0,20	-25		-25
29	97,67	97,77	-0,10	-15,5		-15,5
30	97,40	97,33	0,07	12,5	12,5	
					<i>T+</i> =148,5	<i>T-</i> =-257,5

Resolvendo o exemplo a partir da Equação 2.19 tem-se o seguinte valor de *z*.

$$z = \frac{148,5 - \frac{28(28+1)}{4}}{\sqrt{\frac{28(28+1)(2 \cdot 28 + 1)}{24}}} \quad z = -1,2410$$

Como o valor de *z* calculado em módulo é menor que o valor de *z* crítico, não se rejeita a hipótese de que a acurácia obtida com aplicação do método Lider_H_10 é igual a acurácia obtida utilizando o conjunto completo de treinamento.

3 REVISÃO BIBLIOGRÁFICA

Treinar um classificador SVM envolve a solução de um problema de programação quadrática (QP do inglês *Quadratic Programming*), que exige grande quantidade de memória e tempo de treinamento para aplicações em larga escala (WANG; NESKOVIC; COOPER, 2006). Por outro lado, a função de decisão do classificador SVM depende apenas de um subconjunto dos dados de treinamento, cujos elementos são chamados de vetores de suporte. Portanto, se os padrões que correspondem aos vetores de suporte são conhecidos antecipadamente, a mesma solução pode ser obtida através da resolução de um problema QP muito menor que envolve apenas os vetores de suporte. A dificuldade então se resume em como selecionar exemplos de treinamento que possam ser vetores de suporte.

Inicialmente, duas abordagens podem ser mencionadas para reduzir o custo do QP sem reduzir o conjunto de treinamento: fragmentação (do inglês *Chunking*) e decomposição (do inglês *Decomposition*).

A primeira abordagem (JOACHINS, 1999), citada em Koggalage e Halgamuge (2004), é uma heurística que utiliza blocos de amostras. Um bloco é um número pré-definido de amostras, que é consideravelmente menor do que o número total de amostras do conjunto de treinamento. Cada bloco de amostras é utilizado para identificar os vetores de suporte. As amostras que são vetores de suporte são mantidas e as amostras que não são vetores de suporte são descartadas. O processo continua com diferentes blocos até que todas as amostras de treinamento sejam utilizadas. Embora esta seja uma forma de reduzir o custo de memória e tempo, ainda é muito restritivo para conjuntos de dados grandes.

A segunda abordagem quebra o problema inicial em uma série de subproblemas menores, resolvidos por um método tradicional. Como exemplo, pode ser citado o SMO (do inglês *Sequential Minimal Optimization*) (PLATT, 1999), que trabalha com problemas de tamanho dois, ou seja, a cada iteração dois multiplicadores de Lagrange

são considerados na otimização. Neste caso, o problema pode ser resolvido analiticamente.

Uma abordagem alternativa, para reduzir o custo computacional, procura selecionar um subconjunto representativo, com o uso de técnicas para reduzir o número de amostras. Recentemente, alguns autores propuseram métodos que selecionam amostras perto da fronteira de decisão com base nas propriedades da vizinhança, tais como os vizinhos mais próximos (k -NN), a entropia das classes dos vizinhos e a probabilidade de uma amostra estar localizada próxima à região de decisão (SHIN;CHO, 2003). Outros utilizam o algoritmo de agrupamento k -médias para selecionar as amostras do conjunto de treinamento (KOGGALAGE; HALGAMUGE, 2004).

Nota-se então que a característica de vizinhança presente nos dados do conjunto de amostras pode ser utilizada na análise. A proximidade entre as amostras define algumas características relacionadas à fronteira de separação entre as classes. Alguns trabalhos que envolvem esta aplicação utilizam classificadores MLP (do inglês *Multilayer Perceptron*) e SVM.

Por exemplo, são apresentados alguns métodos exploram a proximidade dos dados, utilizando o método de edição pelo vizinho mais próximo (ENN, do inglês *Edit Nearest Neighbor*), apresentado primeiramente por Wilson (1972) e abordado neste trabalho, e a edição utilizando um centróide (NCNE, do inglês *Nearest Centroid Neighborhood Edit*), que é uma versão da edição de Wilson considerando uma distribuição simétrica em torno da amostra analisada (ALEJO, 2010).

Alejo também propõe mais dois métodos que utilizam grafos com a idéia de edição de Wilson, descartando as amostras que são classificadas incorretamente em relação a seus vizinhos, utilizando o Grafo de Gabriel (GG) ou o Grafo relativo à vizinhança (RNG, do inglês *Relative Neighborhood Graph*).

Romero (2011) propôs a aplicação de um pré-processamento sobre o conjunto de treinamento, utilizando um algoritmo de agrupamento, para selecionar amostras próximas à fronteira de decisão. Alejo (2010), por outro, explora a eliminação dos

outliers, através das variações da edição de Wilson (1972), podendo melhorar a acurácia da classificação.

Su (2009) realizou a redução do conjunto de treinamento com agrupamento hierárquico aglomerativo e utilizou-o no SVM para classificação de vegetação semiárida.

As alternativas e os métodos de redução são utilizados com a principal intenção de reduzir o custo computacional ou adequação para uma grande quantidade de amostras do conjunto de treinamento. As aplicações e utilização são abrangentes na Literatura, devendo ser observado o caso de aplicação do classificador e a eficiência para cada método.

A utilização dos métodos de redução nos conjuntos de treinamento é promissora e pode levar a um bom resultado de classificação.

4 METODOLOGIA

O objetivo principal do trabalho é realizar a redução dos dados do conjunto de treinamento para treinar o classificador SVM, comparando a acurácia da classificação utilizando o conjunto reduzido de amostras de treinamento perante acurácia obtida utilizando o conjunto completo. Para analisar a eficiência da redução dos dados, foi utilizado o tempo de treinamento do classificador, o tempo gasto para reduzir o conjunto completo de treinamento e a classificação das amostras de teste. Foram utilizadas três imagens sintéticas, com características distintas de espalhamento entre as classes. Para a aplicação em imagens reais, foi utilizada uma imagem ótica de alta resolução e uma imagem de radar, todas descritas posteriormente.

Outro ponto importante diz respeito à seleção do modelo e dos parâmetros do SVM, tais como o método proposto para o caso multiclases, a função núcleo utilizada, a quantidade inicial de amostras do conjunto de treinamento e os valores utilizados para o parâmetro C . Cada um destes temas exige esforço computacional e tempo para serem calibrados e analisados, levando muitas vezes o usuário a optar por uma arbitrariedade, que pode não ser adequada para o caso.

4.1 Computador utilizado e Implementação

1 – Para a realização dos experimentos foi utilizado um computador com sistema operacional Windows 7, versão 64 bits, processador Intel Core i7 3930k 3.20 GHz e memória RAM de 64 Gb.

2 – Implementação dos algoritmos de redução de dados, de treinamento do SVM e de classificação do SVM em linguagem IDL (do inglês *Interactive Data Language*) (EXELIS, 2010).

4.2 Imagens sintéticas

Foram utilizadas três imagens sintéticas, cada uma com tamanho de 512×512 pixels, três atributos de cores (RGB) e duas classes. As classes seguem a distribuição Normal, com matrizes de covariância diagonais e iguais para todas as classes. Os valores dos desvios-padrões e das médias, para as duas classes e para as três imagens, são apresentados na Tabela 4.1.

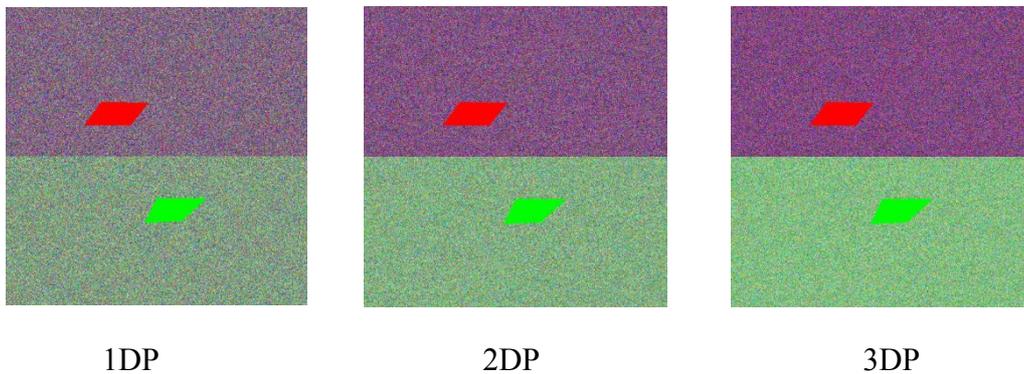
As posições dos centros das classes diferem apenas no valor do atributo (G), que tem desvio-padrão 18. As imagens foram geradas com três valores de distância entre centros que equivalem a 1, 2 e 3 desvios-padrões, daqui por diante referidas como 1DP, 2DP e 3DP, respectivamente.

O conjunto de amostras iniciais das imagens sintéticas contém 3.130 amostras da classe 1, polígono destacado em vermelho, e 3.033 amostra da classe 2, polígono destacado em verde. As imagens e as respectivas áreas do conjunto de amostras são apresentadas na Figura 4.1.

Tabela 4.1 – Dados das imagens sintéticas

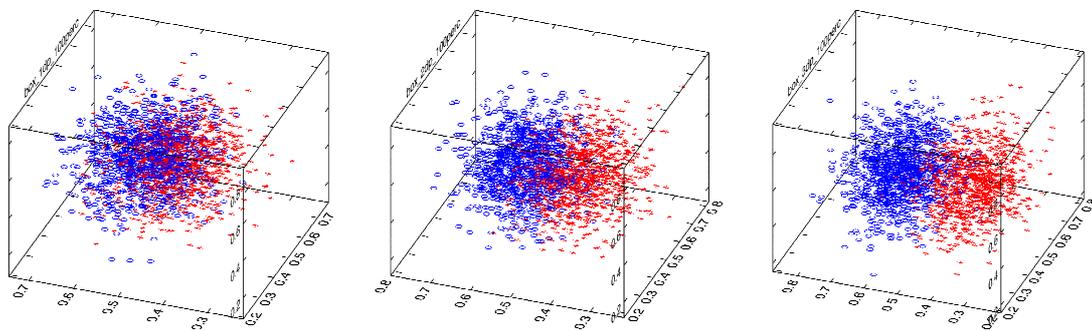
Imagem	Atributo	CLASSE 1	CLASSE 2	
		Média	Média	DP
1DP	1 (R)	128	128	25
	2 (G)	119	137	18
	3 (B)	128	128	30
2DP	1 (R)	128	128	25
	2 (G)	110	146	18
	3 (B)	128	128	30
3DP	1 (R)	128	128	25
	2 (G)	101	155	18
	3 (B)	128	128	30

Figura 4.1 – Imagens Sintéticas e regiões da coleta de amostras



Na Figura 4.2 abaixo, observa-se o comportamento das amostras de treinamento no espaço de atributos, correspondentes às imagens sintéticas, sendo a imagem à esquerda de alta mistura, ao centro a de média mistura e à direita a de baixa mistura.

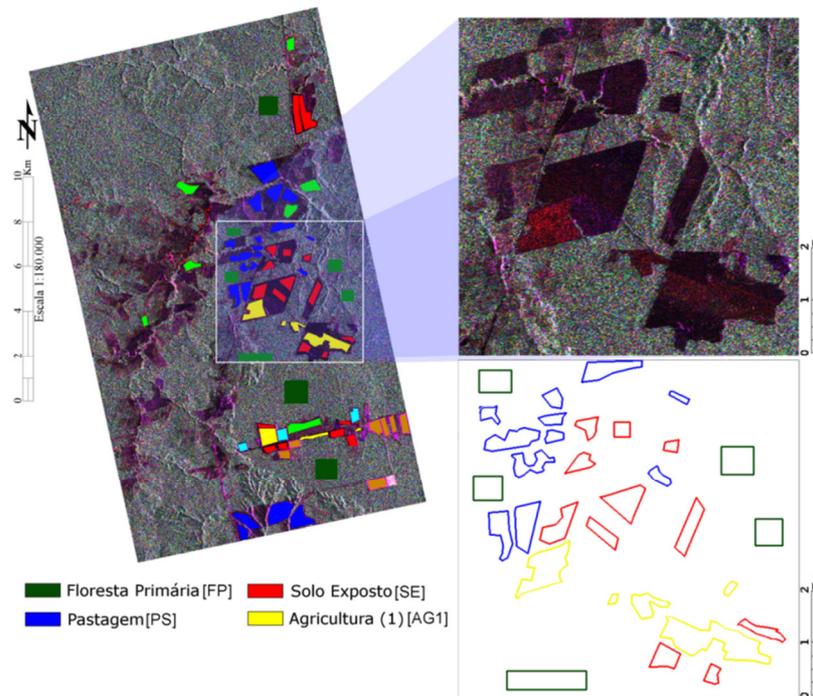
Figura 4.2 – Amostras das imagens sintéticas no espaço de atributos



4.3 Imagem do sensor ALOS PALSAR

A imagem adotada é da região da Floresta Nacional de Tapajós-PA, obtida pelo sensor ALOS PALSAR. A Figura 5.33 mostra a imagem de trabalho, a região de estudo, e as amostras que foram coletadas para treinamento e teste do classificador. A composição colorida foi obtida seguindo a configuração: (HH)R (HV)G (VV)B.

Figura 4.3 – Imagem ALOS PALSAR em amplitude



Fonte: adaptado de Negri (2013)

Para este experimento foram utilizadas quatro classes, indicadas na Figura 4.2 e detalhadas na tabela 4.2 abaixo.

Tabela 4.2 - Classes de interesse e tamanhos de amostra

Classes	Amostras Iniciais
Floresta Primária (FP)	14.136
Agricultura tipo 1 (AG1)	14.678
Pastagem (PS)	14.395
Solo Exposto (SE)	14.207

4.4 Imagem do sensor IKONOS MS

A imagem adotada é de uma área urbana, especificamente da parte da região sul da cidade de São José dos Campos, Estado de São Paulo. A imagem do sensor MS do

satélite IKONOS-II tem resolução de quatro metros. A subimagem utilizada neste experimento tem 600 x 600 pixels, e com as bandas do vermelho, verde e azul. Ela é mostrada na Figura 5.37 em composição natural e ao lado as amostras das classes.

Figura 4.4 – Imagem IKONOS MS - região sul de São José dos Campos



Fonte: adaptado de Negri (2013)

A tabela 4.3 mostra as classes da imagem IKONOS, detalha as classes utilizadas para este estudo e a quantidade de amostras do conjunto inicial.

Tabela 4.3 – Classes de interesse e tamanhos de amostra

Classe	Cor	Amostras Iniciais
Vegetação Alta (ARVORE)	■	5828
Vegetação Baixa (GRAMA)	■	4119
Telhado Cerâmico (TELHADO)	■	4026
Telhado Concreto/Metal (TELHADO_2)	■	4261
Solo Exposto (SOLO)	■	4299
Asfalto (VIAS)	■	4026

4.5 Estrutura dos experimentos computacionais

Os experimentos foram realizados em cinco etapas. A Coleta e seleção das amostras, a sintonização da penalidade C , a aplicação dos métodos de redução, a classificação utilizando o conjunto de amostras reduzidas e avaliação dos resultados.

4.5.1 Coleta e Seleção de amostras

Para a realização dos experimentos foi necessário dividir o conjunto inicial de amostras, tanto das imagens sintéticas, quanto das reais, em três subconjuntos, definidos como conjunto de treinamento, conjunto de validação e conjunto de teste.

A divisão foi realizada da seguinte forma:

Tabela 4.4 – Classes de interesse e tamanhos de amostra por classe

<i>Imagem</i>	Treinamento	Validação	Teste
Sintética	2.000	200	200
PALSAR	1.000	500	500
IKONOS	1.000	500	500

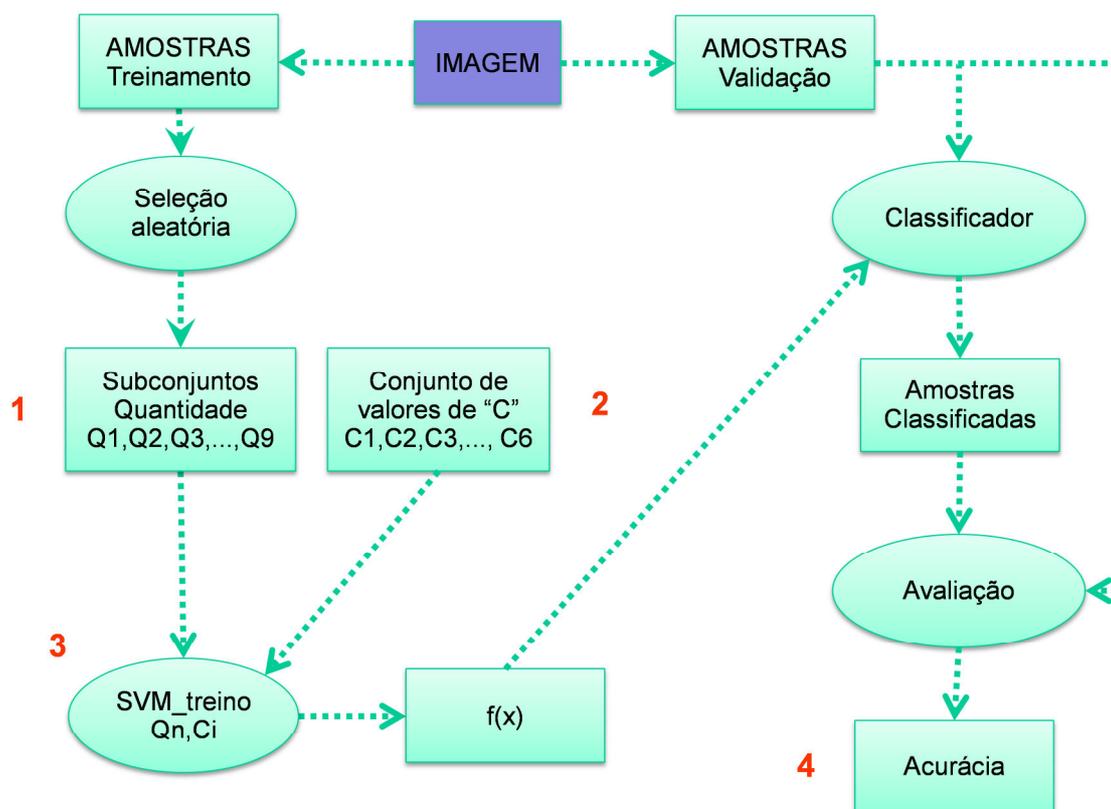
O conjunto de amostras de validação é utilizado para selecionar a penalidade C , e o conjunto de amostras de teste para avaliar os resultados obtidos pelos métodos de redução.

Os métodos de redução foram aplicados sobre o conjunto de treinamento e posteriormente utilizados para treinar o classificador SVM.

4.5.2 Seleção da penalidade C

No modelo de classificação adotado foi considerado a utilização da penalidade C . A seleção do parâmetro foi realizada conforme o fluxograma da figura abaixo.

Figura 4.5 – Fluxo para seleção da penalidade C



1 – Com finalidade analisar o comportamento do classificador foram adotados os seguintes tamanhos de quantidade de amostras de treinamento por classe:

$$Q = \{200, 300, 400, 500, 600, 700, 800, 900, 1000\}.$$

Cada valor de Q é retirado do conjunto de treinamento, com reposição, e utilizado para treinar o classificador SVM.

2 – Foram selecionados seis valores de C para realizar os experimentos, sendo:

$$C = \{2^{-3}, 2^0, 2^3, 2^6, 2^9, 2^{12}\}$$

3 – Realização de 30 execuções para cada Q_n, C_i .

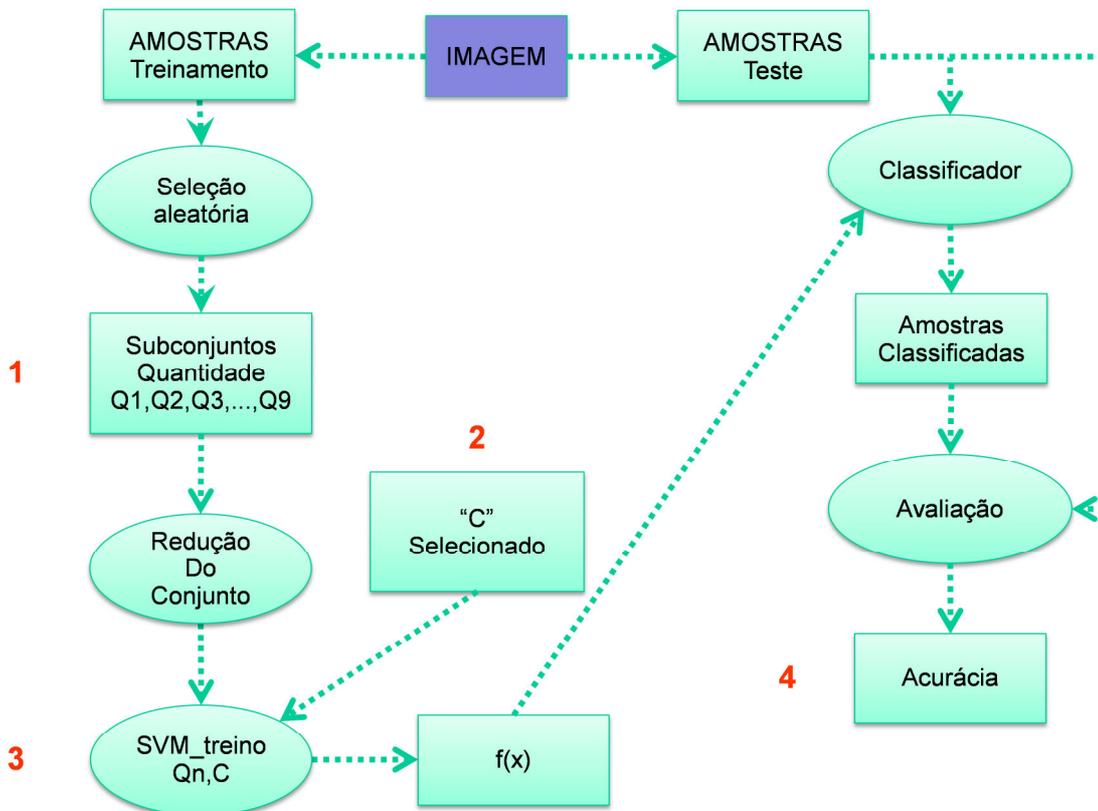
4 – Apuração, em cada execução, do tempo de estimação dos parâmetros do classificador e acurácia da classificação das amostras de validação.

Foi selecionado o valor de C que proporcionou a maior acurácia média para cada quantidade de amostra do conjunto de treinamento.

4.5.3 Redução do conjunto de treinamento

No modelo de classificação adotado, a redução do conjunto de treinamento é um processo independente. O fluxograma abaixo mostra os processos realizados.

Figura 4.6 – Fluxo para redução do conjunto de treinamento



1 – Foram adotados diferentes tamanhos do conjunto de treinamento com a finalidade analisar o comportamento dos algoritmos de redução. Os tamanhos da quantidade de amostras de treinamento são mostrados abaixo:

$$Q = \{200, 300, 400, 500, 600, 700, 800, 900, 1000\}$$

2 – Utilizado o valor de C selecionado no fluxo da figura 4.5

3 – Realizações de 30 execuções para cada Q_n, C .

4 – Apuração, em cada execução, do tempo de estimação dos parâmetros do classificador e acurácia da classificação das amostras de validação.

A partir dos resultados obtidos nos experimentos são então realizadas:

- Análise quantitativa e qualitativa da redução dos dados nos testes efetuados, verificando percentual de redução, tempo de redução dos dados de treinamento, tempo de treinamento do SVM e tempo global (tempo de redução adicionado ao tempo de treinamento utilizando o conjunto de dados reduzido);
- Análise da acurácia versus o tempo global do classificador SVM;
- Aplicação do teste de WILCOXON sobre o resultado dos experimentos para as imagens sintéticas.

5 EXPERIMENTO COMPUTACIONAL E ANÁLISE DOS RESULTADOS

Os experimentos e resultados foram divididos em três seções, considerando as aplicações em imagens sintéticas e imagens de Sensoriamento Remoto obtidas pelos sensores ALOS PALSAR (radar) e IKONOS II MS (ótica).

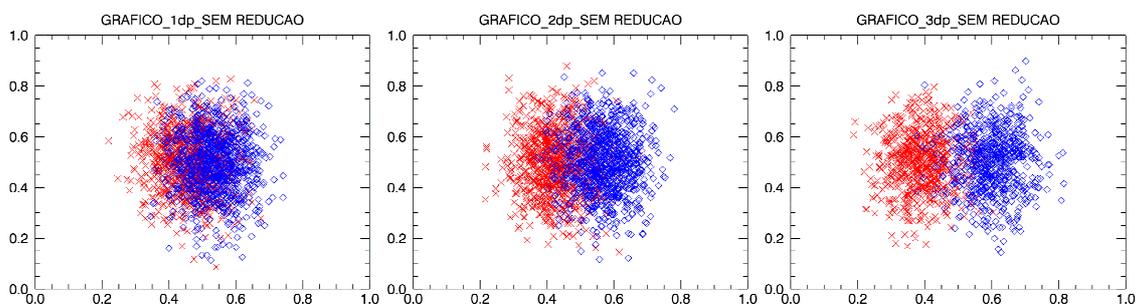
Em todas as execuções foram adotados, método OAO para o caso multiclases e função núcleo linear. Também foi considerado $k=3$ para o RENN, $k=1$ para o CNN e $k=1$ e $B=3$ para o ME. Os métodos em cascata seguem a mesma configuração. Foi adotado um $\alpha = 0,05$ para o Teste de Wilcoxon, e os valores de z calculado para cada um dos métodos de redução aplicados às imagens sintéticas e às imagens reais estão descritos no Apêndice B, e o resultado do tempo global (redução + treinamento) está descrito no Apêndice C.

Para efeitos de análise dos dados reduzidos, somente serão mostrados os gráficos em 2D para os atributos 2 (atributo que determina a mistura entre as 2 classes) e 3 das amostras do conjunto de treinamento. Na Figura 5.1, o gráfico à esquerda representa as amostras com alta mistura entre as classes (Imagem 1DP), ao centro a de média mistura (Imagem 2DP) e à direita a de baixa mistura (Imagem 3DP).

As Figuras 5.2-5.18 seguem a mesma ilustração da Figura 5.1

1 – Conjunto completo das amostras de treinamento das imagens sintéticas.

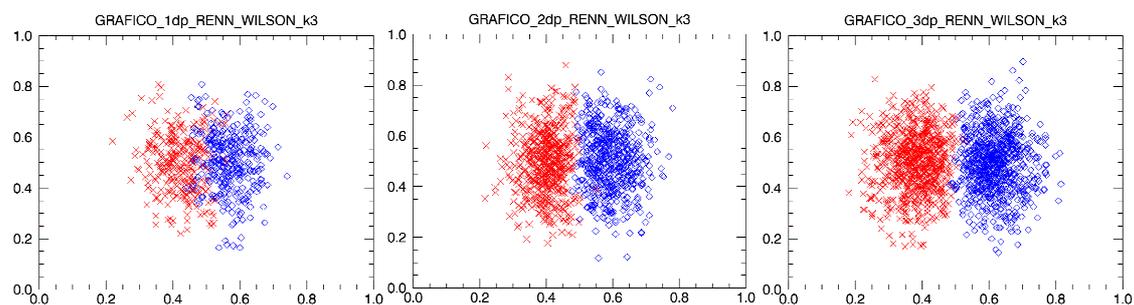
Figura 5.1 – Amostras das imagens sintéticas sem redução.



2 – Dados reduzidos pelo RENN

O modelo RENN retira amostras próximas da fronteira das classes. Na Figura 5.2 nota-se que esta região, após a redução, mostra maior separabilidade entre as classes, comparada à Figura 5.1.

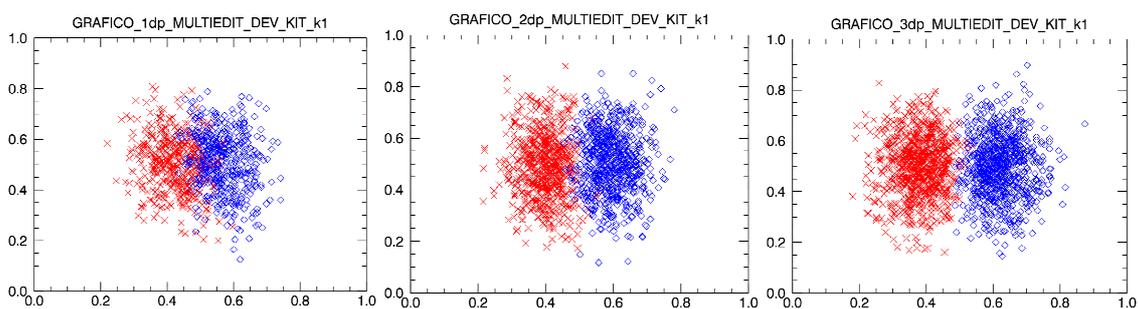
Figura 5.2 – Amostras das imagens sintéticas reduzidas pelo RENN.



3 – Dados reduzidos pelo modelo Multiedit (ME)

O modelo ME obteve resultados próximos à aplicação do RENN. Contudo, nota-se que o RENN descartou um pouco mais de amostras.

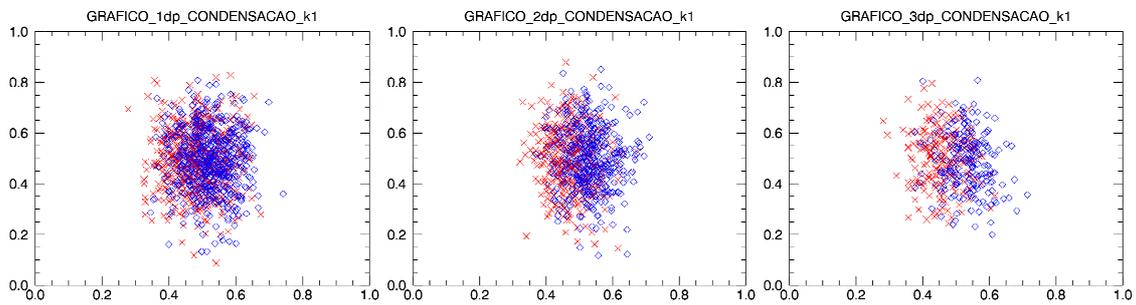
Figura 5.3– Amostras das imagens sintéticas reduzidas pelo ME.



4– Dados reduzidos pelo modelo CNN

O modelo CNN, inverso aos modelos RENN e ME, procura manter as amostras que realmente estão na fronteira de decisão de separação das classes. Assim pode ser verificado nos resultados da Figura 5.4 que restaram as amostras situadas nesta região.

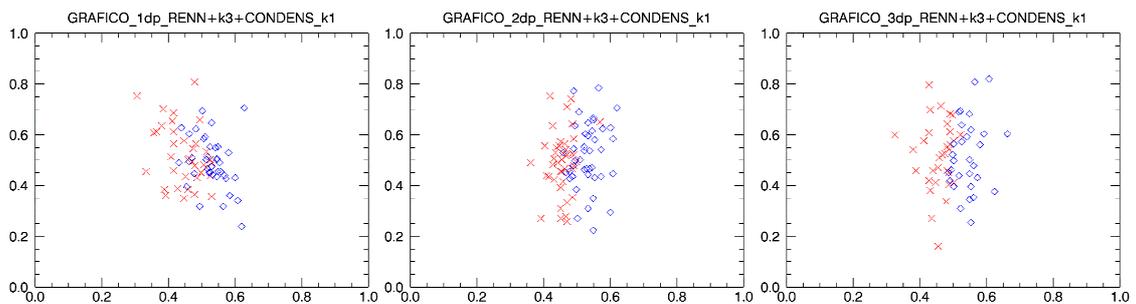
Figura 5.4– Amostras das imagens sintéticas reduzidas pelo CNN.



5– Dados reduzidos pelo modelo RENN+CNN

Utilizando uma redução sequencial com dois métodos, percebeu-se que a capacidade do modelo em reduzir dados. Na primeira etapa o RENN retirou algumas amostras da fronteira e posteriormente o CNN selecionou esta região de fronteira, descartando as amostras mais distantes. Porém esta quantidade pode não ser adequada para estimar os parâmetros do classificador SVM.

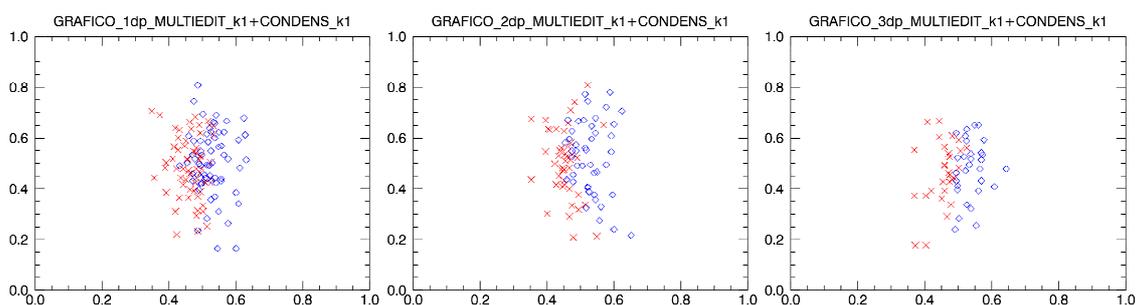
Figura 5.5– Amostras das imagens sintéticas reduzidas pelo RENN+CNN.



6– Dados reduzidos pelo modelo ME+CNN

Similar à aplicação do RENN+CNN, nota-se na Figura 5.6 a alta redução das amostras iniciais, comparada à Figura 5.1. Deve-se observar o resultado, que pode não representar o conjunto inicial.

Figura 5.6– Amostras das imagens sintéticas reduzidas pelo ME+CNN.



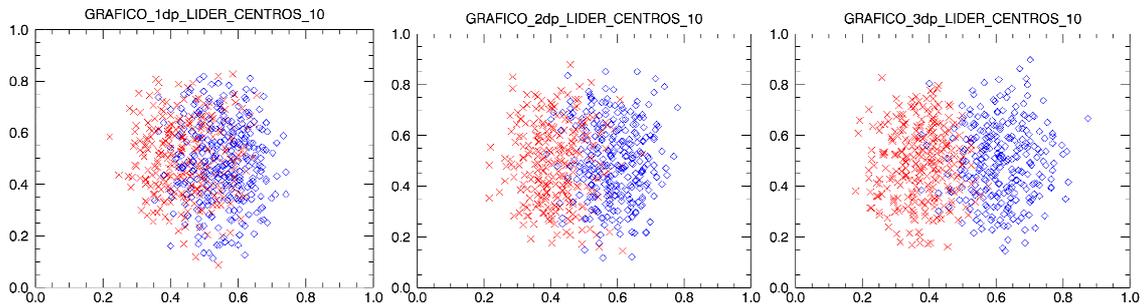
Para os resultados do com o algoritmo de redução Líder, foram utilizadas as três variações descritas na Seção 4.3.2. Devido à necessidade de uma estimativa inicial do número de grupos, foram considerados os percentuais de 10%, 30%, 50% e 70%, ou seja, se a amostra inicial completa tem 2.000 amostras, têm-se, para o caso do percentual de 10%, uma estimativa inicial de 200 grupos, para o caso em que se utiliza 30%, têm-se uma estimativa inicial de 600 grupos, e assim sucessivamente.

O objetivo é realizar o estudo para quantidades de grupos variados que possibilitem análise da estimativa inicial de pequenos a grandes grupos.

7– Dados reduzidos pelo modelo Líder Centros 10%

Utilizando somente os centros dos grupos gerados, o modelo procura manter a característica geral das amostras no espaço de atributos. Na Figura 5.7 pode ser verificado que a distribuição das amostras manteve-se similar aos dados originais, porém em uma quantidade menor.

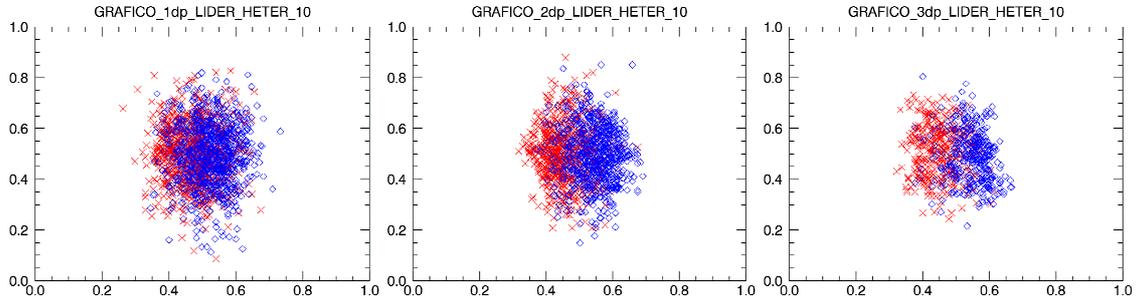
Figura 5.7– Amostras das imagens sintéticas reduzidas pelo LID_C_10.



8– Dados reduzidos pelo modelo Líder Heterogêneos 10%

Neste modelo de redução são mantidos os grupos que possuem amostras de mais de uma classe. Similarmente ao modelo CNN, pode-se notar nos resultados abaixo que a aplicação do modelo manteve as amostras que estão localizadas na fronteira das classes.

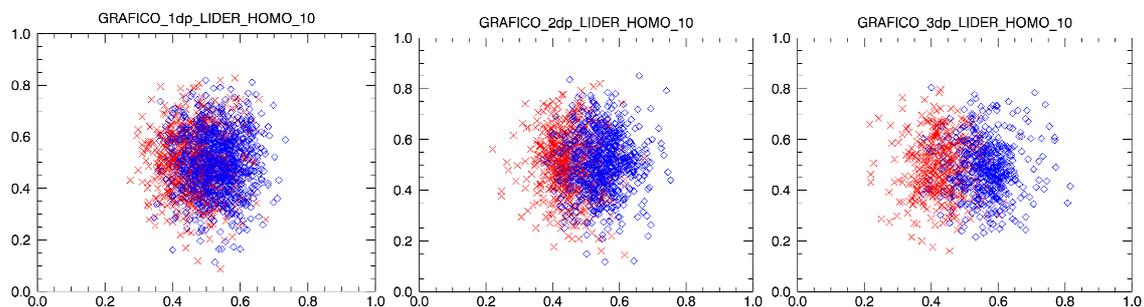
Figura 5.8– Amostras das imagens sintéticas reduzidas pelo LID_H_10.



9– Dados reduzidos pelo modelo Líder Heterogêneos + Centros 10%

Neste modelo de redução são mantidos os grupos que possuem amostras de mais de uma classe, e os centros dos grupos homogêneos. Nos resultados abaixo a aplicação do modelo manteve as amostras localizadas na fronteira das classes, e mais algumas ao redor, deixando o resultado similar a amostra completa.

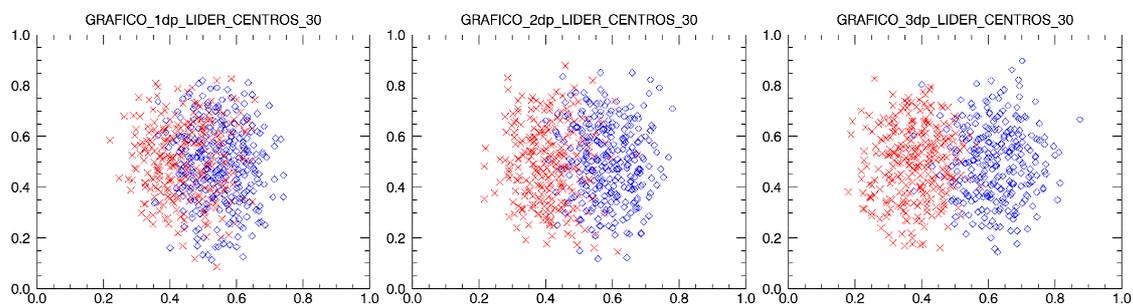
Figura 5.9– Amostras das imagens sintéticas reduzidas pelo LID_HC_10.



10– Dados reduzidos pelo modelo Líder Centros 30%

Similar à aplicação com 10% das amostras, diferindo na estimativa inicial, que neste caso é de 30% das amostras iniciais.

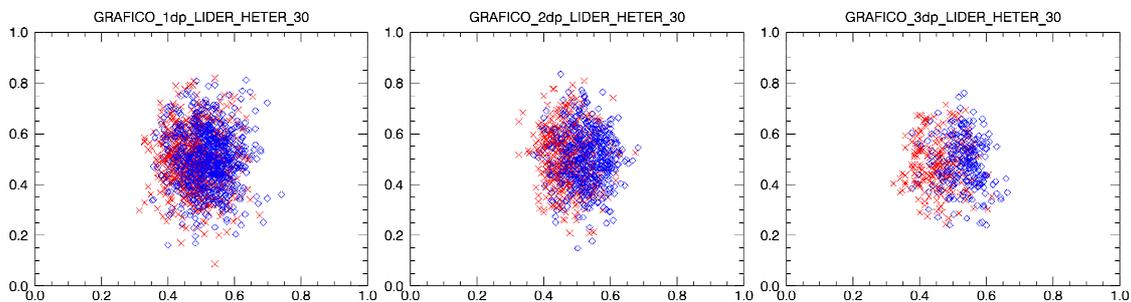
Figura 5.10– Amostras das imagens sintéticas reduzidas pelo LID_C_30.



11– Dados reduzidos pelo modelo Líder Heterogêneos 30%

Similar à aplicação com 10% das amostras, difere em dois aspectos, o primeiro pela estimativa inicial de grupos que é de 30%, a segunda por possuir menos amostras no conjunto reduzido.

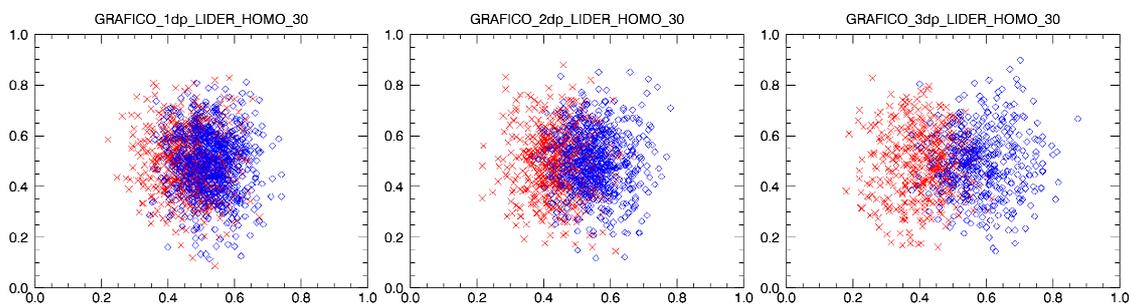
Figura 5.11– Amostras das imagens sintéticas reduzidas pelo LID_H_30.



12– Dados reduzidos pelo modelo Líder Heterogêneos + Centros 30%

Similar ao modelo com 10% das amostras. Neste resultado o número de amostras aumenta e a distribuição fica mais próxima da distribuição das amostras originais.

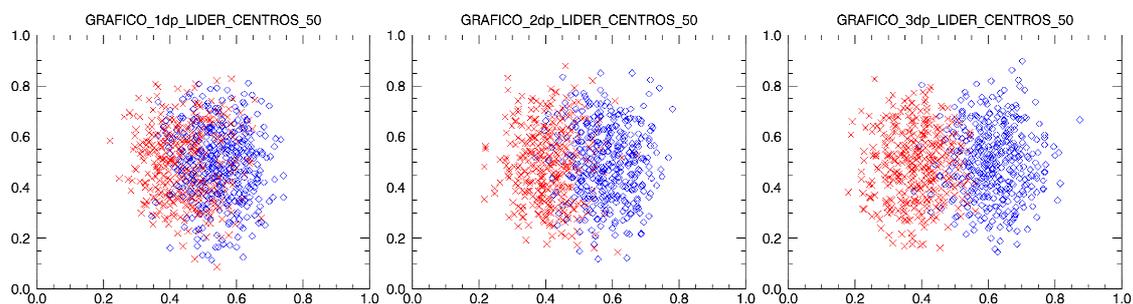
Figura 5.12– Amostras das imagens sintéticas reduzidas pelo LID_HC_30.



13– Dados reduzidos pelo modelo Líder Centros 50%

Similar às aplicações anteriores com 10% e 30% das amostras como estimativa inicial do número de grupos. Este modelo difere somente na quantidade final, que neste caso é de 50% das amostras iniciais.

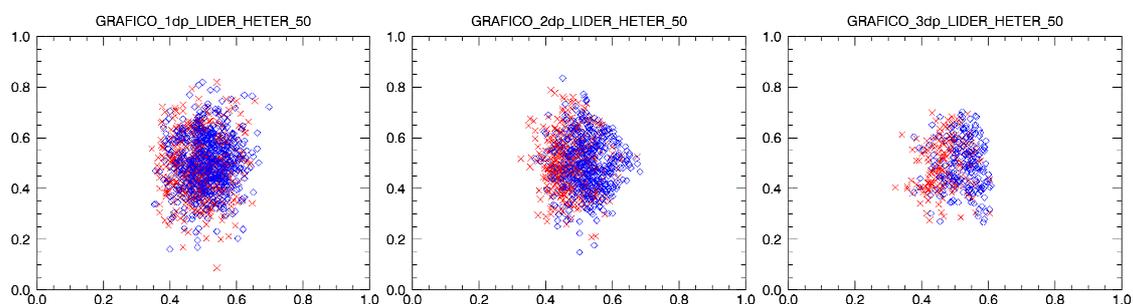
Figura 5.13– Amostras das imagens sintéticas reduzidas pelo LID_C_50.



14– Dados reduzidos pelo modelo Líder Heterogêneos 50%

Similar à aplicação com 10% e 30% das amostras iniciais, diferindo por ter menos amostras reduzidas. Comportamento visível na imagem com baixa mistura.

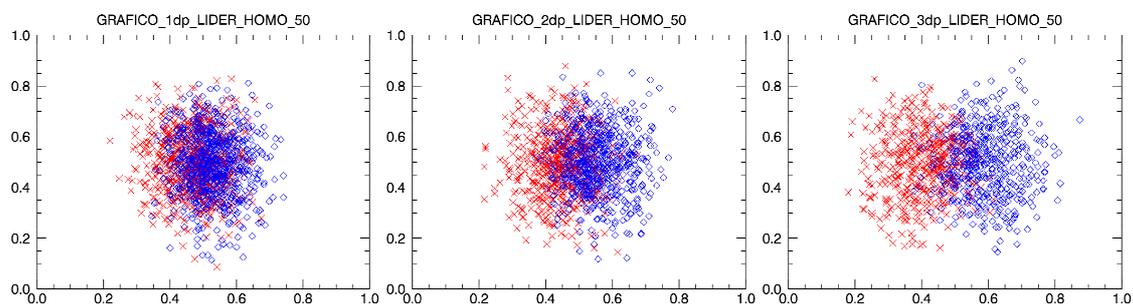
Figura 5.14– Amostras das imagens sintéticas reduzidas pelo LID_H_50.



15– Dados reduzidos pelo modelo Líder Heterogêneos + Centros 50%

Similar à aplicação com 10% e 30% das amostras iniciais. A maioria das amostras da fronteira de decisão foi mantida e a distribuição das amostras está similar a amostra completa.

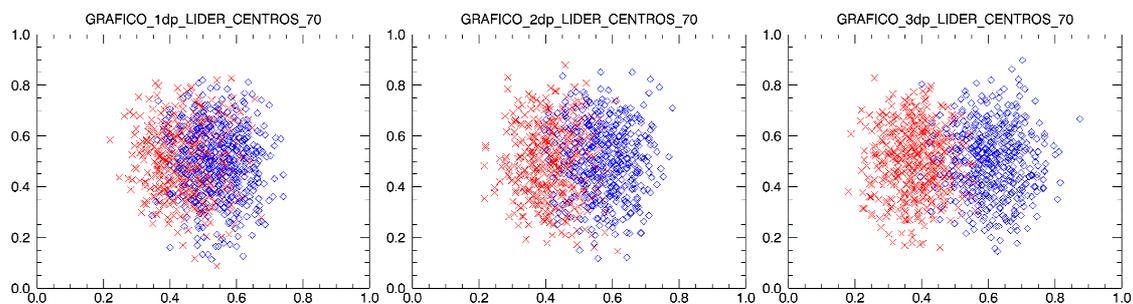
Figura 5.15– Amostras das imagens sintéticas reduzidas pelo LID_HC_50.



16– Dados reduzidos pelo modelo Líder Centros 70%

Similar às aplicações anteriores com 10%, 30% e 50% das amostras como estimativa inicial do número de grupos, este modelo difere somente na quantidade final, que neste caso é de 70% das amostras iniciais. Nota-se também que a distribuição das amostras foi mantida.

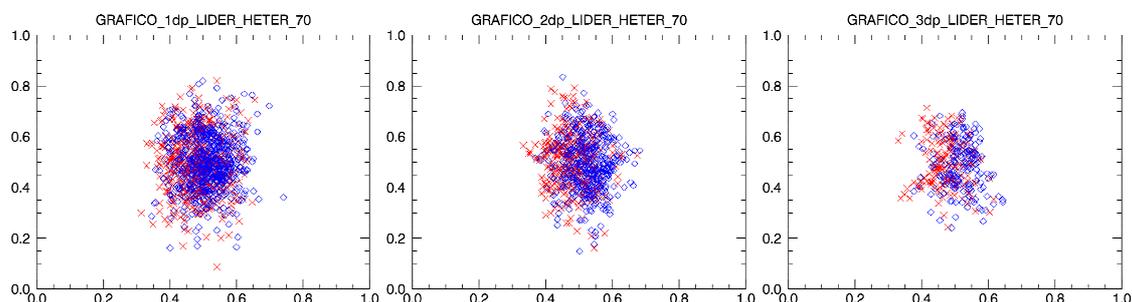
Figura 5.16– Amostras das imagens sintéticas reduzidas pelo LID_C_70.



17– Dados reduzidos pelo modelo Líder Heterogêneos 70%

Similar à aplicação com 10%, 30% e 50% das amostras iniciais.

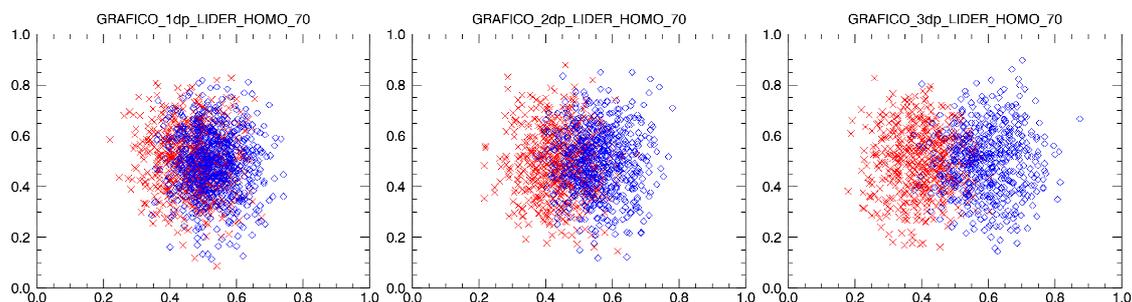
Figura 5.17– Amostras das imagens sintéticas reduzidas pelo LID_H_70.



18– Dados reduzidos pelo modelo Líder Heterogêneos + Centros 70%

Similar à aplicação com 10%, 30% e 50% das amostras iniciais. A maioria das amostras da fronteira de decisão foi mantida e a distribuição das amostras está similar a amostra completa.

Figura 5.18– Amostras das imagens sintéticas reduzidas pelo LID_HC_70.

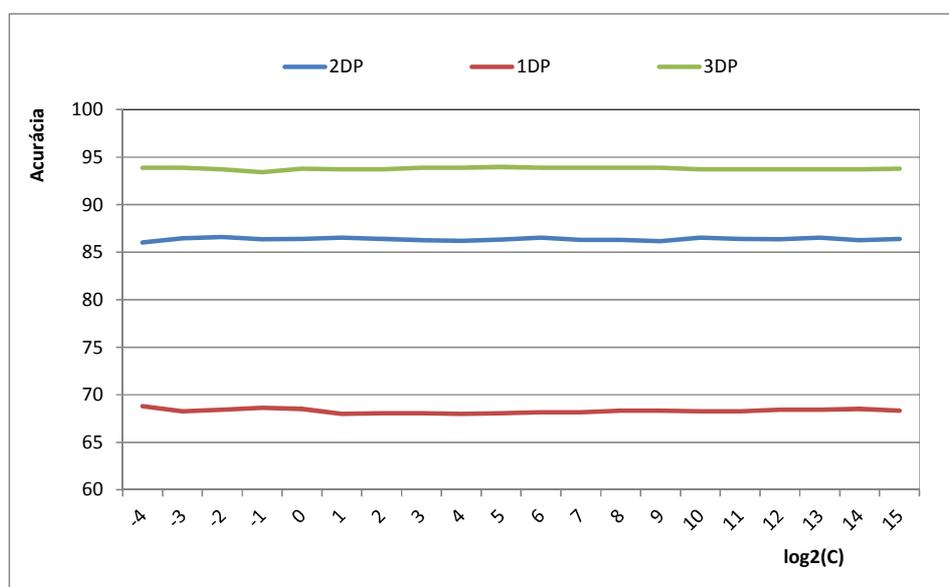


5.1 Acurácia vs. Valor de C

Um dos principais problemas na classificação utilizando o SVM é a utilização de um valor de C adequado. Em um primeiro experimento, foram realizadas 10 classificações do conjunto de amostras de validação das imagens sintéticas que contém 200 amostras por classe, variando o valor de C entre 2^{-4} e 2^{15} . O objetivo foi utilizar um valor de C para realizar os experimentos referentes à classificação das imagens sintéticas e apurar a acurácia destas classificações.

Notou-se que as variações da acurácia perante o valor de C , somente são perceptíveis em valores muito baixos, conforme Figura 5.19.

Figura 5.19– Acurácia vs. variação do C – imagens sintéticas.



Diante do exposto, decidimos por utilizar seis valores de C , que abrangessem o intervalo da Figura 5.19, para realizar os experimentos e avaliar a acurácia da classificação destas imagens, considerando:

$$\text{Valores de } C = \{2^{-3}, 2^0, 2^3, 2^6, 2^9, 2^{12}\}$$

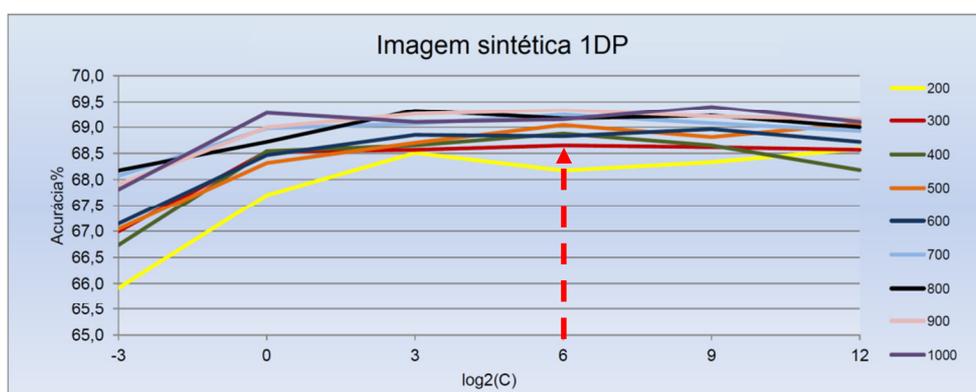
5.2 Análise das imagens sintéticas

Um dos principais problemas identificado no treinamento e classificação do SVM, se refere à calibração, ou melhor ajuste da penalidade C . Na utilização das imagens sintéticas, nota-se que, para valores maiores que 8, a influência da penalidade não demonstra grande impacto na acurácia. Porém, valores baixos de C (como 0,125 e 1) podem influenciar na acurácia da classificação. Analisando as Figuras 5.20, 5.21 e 5.22, referente à classificação das amostras de validação para às imagens sintéticas de 1DP,

2DP e 3DP, respectivamente, nota-se este comportamento de baixa acurácia em duas situações: a primeira, para os valores baixos de C e a segunda para quantidades baixas de amostras, que podem ser verificados nas faixas de 200 e 300 amostras por classe nas cores amarela e vermelha, respectivamente.

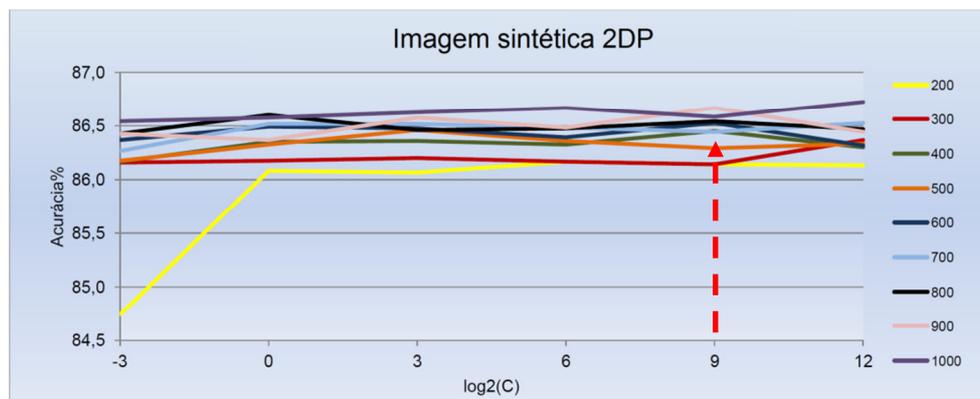
Uma questão importante e que precisou ser identificada para análise dos resultados foi sobre a escolha de um valor de C apropriado para as classificações das imagens. A realização de um teste de calibração da penalidade levaria mais tempo, sendo que o proposto para o trabalho foi escolher um valor que causasse menor impacto na variação da acurácia. Analisando o gráfico da Figura 5.20, nota-se que valores de C maiores que 8 estão seguindo um espalhamento próximo, porém, realçando o valor de $C = 64$, destacado através da seta pontilhada, percebe-se um “leve” estreitamento das acurácias neste intervalo, para os resultados das imagens sintéticas 1DP, conforme Figura 5.20.

Figura 5.20– Acurácia vs. $\log_2 C$ vs. tamanho da amostra - 1DP.



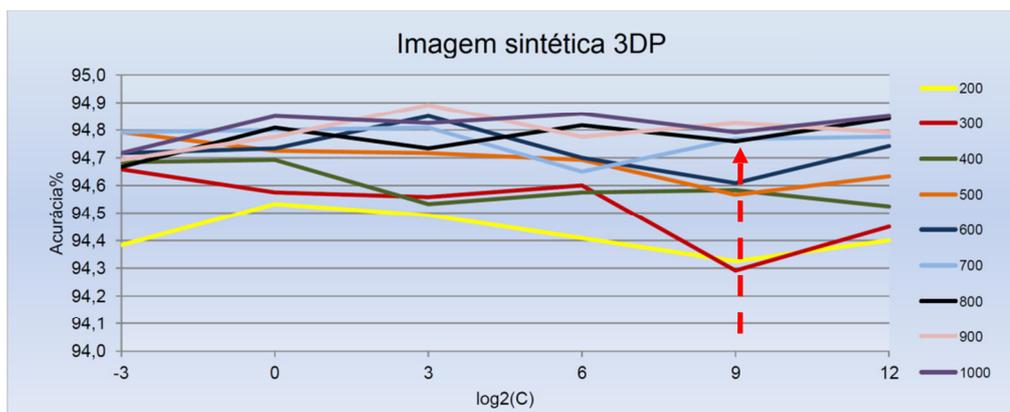
Analisando a Figura 5.21, referente à imagem 2DP, nota-se o mesmo estreitamento da acurácia para valores de C maiores que 8, entretanto, as acurácias obtidas utilizando $C = 512$ estão um pouco acima das demais.

Figura 5.21– Acurácia vs. $\log_2 C$ vs. tamanho da amostra - 2DP.



Para a imagem sintética 3DP, nota-se que o valor de $C = 8$ contém a acurácia mais alta, porém um valor de $C = 512$ tem menor espalhamento nas faixas de maior quantidade de amostras, conforme Figura 5.22.

Figura 5.22– Acurácia vs. $\log_2 C$ vs. tamanho da amostra - 3DP.



Retornando ao tema da calibração do C , nota-se que a necessidade de calibrar o valor de acordo com a imagem utilizada, e que a característica do espalhamento das amostras deve ser levada em consideração.

Para analisar os métodos de redução foram utilizados os valores de C iguais a 64 para a imagem 1DP, e 512 para as imagens 2DP e 3DP, considerados aparentemente adequados como para ser utilizado modelo de classificação do SVM.

5.3 Classificação SVM imagens sintéticas – conjunto de treinamento completo

Foi utilizado o classificador SVM e o conjunto de treinamento completo para classificar as imagens sintéticas e obter uma referência para analisar as classificações que utilizam o conjunto de treinamento reduzido. Nas Figuras 5.23, 5.24 e 5.25, são mostrados os gráficos com a acurácia obtida para nove tamanhos diferentes de conjunto de treinamento [200, 300, ..., 900, 1000) para as imagens sintética 1DP, 2DP e 3DP. Os resultados mostram uma melhora, mesmo que às vezes pouco perceptível, à medida que se aumenta a quantidade de amostras no conjunto de treinamento.

A Tabela 5.1 contém a média das acurácias e os desvios-padrão obtidos após a realização dos 30 experimentos para cada faixa de quantidade, considerando valores de $C = 64$, para a imagem 1DP e $C = 512$ para as imagens 2DP e 3DP. As demais acurácias das três imagens para todas as variações de C adotadas, estão descritas no Apêndice A.

Tabela 5.1 – Acurácias e desvios-padrão das imagens sintéticas

C	ACURÁCIA	200	300	400	500	600	700	800	900	1000
64	1 DP	68,2 ± 1,4	68,7 ± 1,1	68,9 ± 1,1	69 ± 1	68,8 ± 0,9	69,2 ± 1	69,2 ± 0,8	69,3 ± 0,7	69,2 ± 0,9
512	2 DP	86,1 ± 0,4	86,1 ± 0,4	86,5 ± 0,3	86,3 ± 0,5	86,5 ± 0,3	86,4 ± 0,3	86,5 ± 0,3	86,7 ± 0,3	86,6 ± 0,4
512	3 DP	94,3 ± 0,4	94,3 ± 0,4	94,6 ± 0,4	94,6 ± 0,3	94,6 ± 0,4	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,2

As Figuras 5.23, 5.24 e 5.25 fazem referência a Tabela 5.1 e são mostradas para evidenciar a relação entre o tamanho da amostra e a acurácia.

Figura 5.23– Acurácia vs. tamanho da amostra - imagem 1DP.

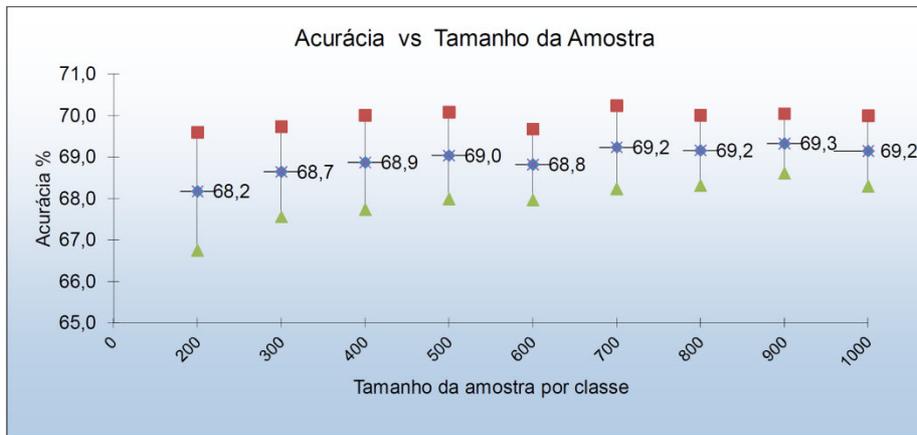


Figura 5.24– Acurácia vs. tamanho da amostra - imagem 2DP.

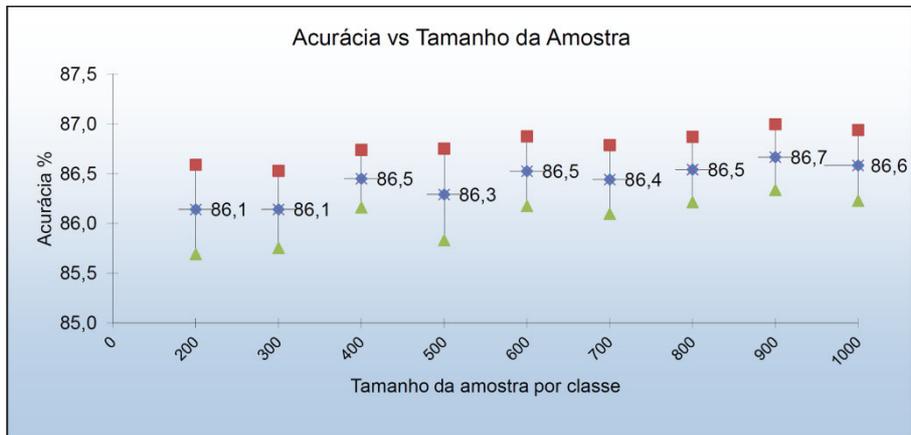
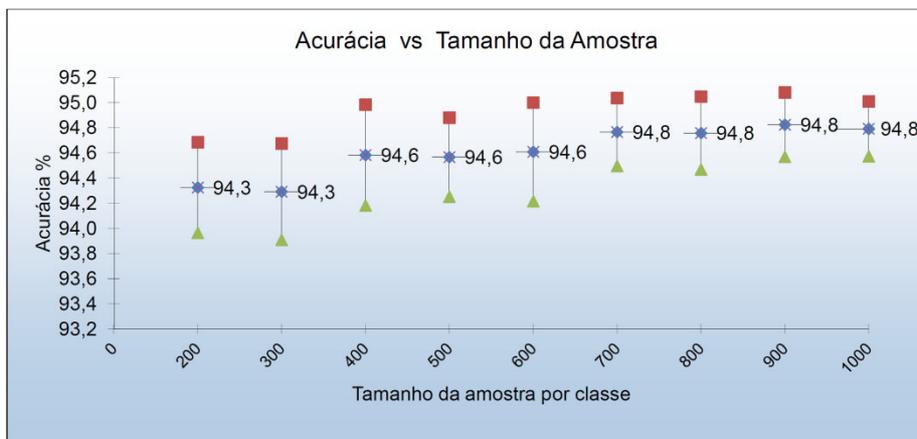


Figura 5.25– Acurácia vs. tamanho da amostra - imagem 3DP.



Em relação ao custo computacional para realizar a classificação SVM, verifica-se através dos gráficos das Figuras 5.26, 5.27 e 5.28, referente às imagens 1DP, 2DP e 3DP, respectivamente, que o espalhamento entre as classes no espaço de atributos, pode influenciar o aumento/diminuição do tempo global (redução+treinamento) necessário para treinar e estimar os parâmetros do classificador. A complexidade da resolução de um problema quadrático simples equivale a ordem $O(M^3)$ a medida que se aumenta a quantidade de amostras do conjunto de treinamento (M).

Uma redução de 1000 para 900 amostras por classe pode proporcionar uma redução entre 25% e 35% do tempo computacional de treinamento também pode ser observada nas Figuras 5.26, 5.27 e 5.28.

Figura 5.26– Tempo Global vs. Tamanho da amostra – 1DP.

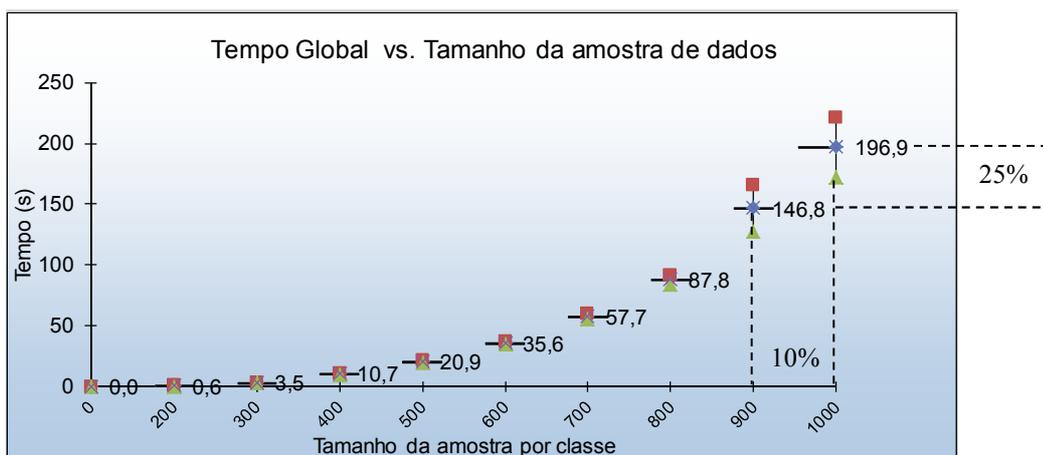


Figura 5.27– Tempo Global vs. Tamanho da amostra – 2DP.

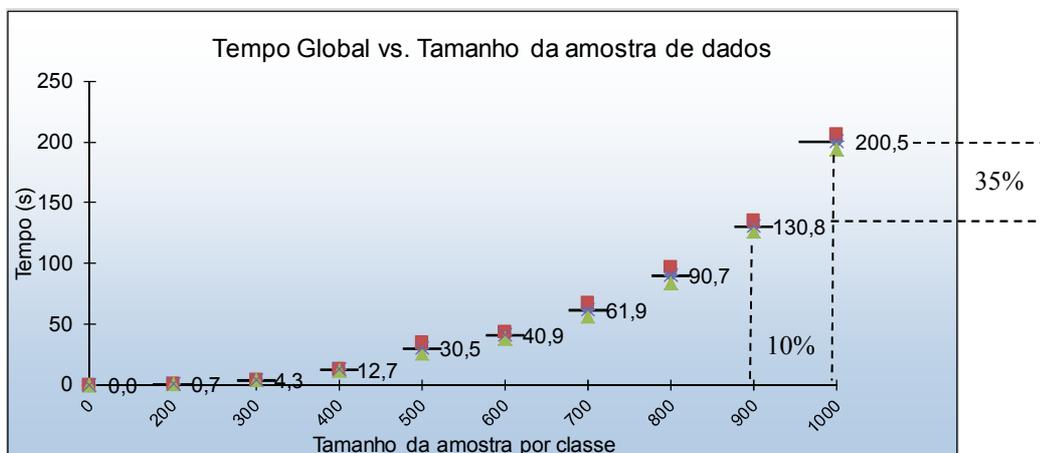
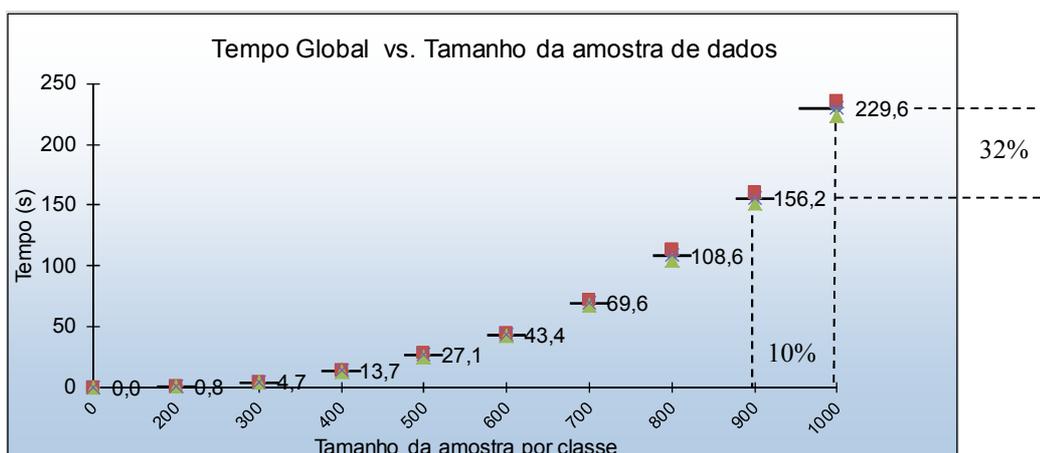


Figura 5.28– Tempo Global vs. Tamanho da amostra – 3DP.



5.4 Aplicação dos métodos de redução e classificação SVM das imagens sintéticas

Para cada imagem sintética foram realizadas 30 execuções para cada um dos 21 conjuntos obtidos pelos métodos de redução e cada um dos seis valores de C adotados no trabalho. Desta maneira têm-se $30 \times 21 \times 6 = 3780$ resultados pra cada imagem sintética. Foi definida a função núcleo Linear como parâmetro do classificador SVM.

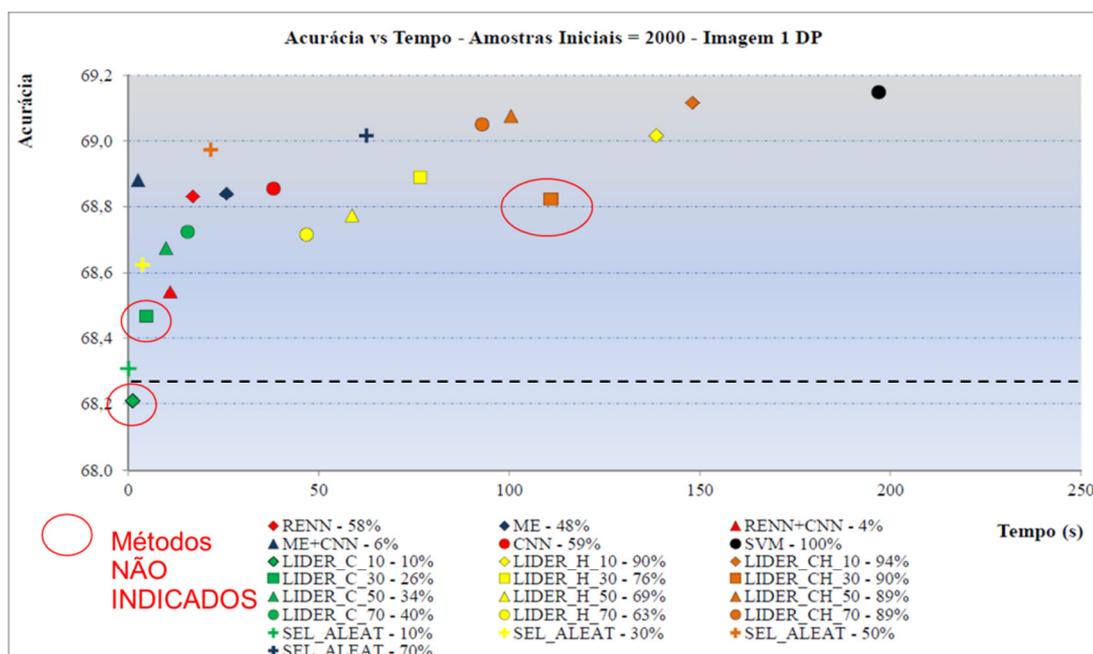
Após a aplicação dos métodos de redução nas três imagens sintéticas, foram elaborados três gráficos, onde cada gráfico mostra a acurácia média com seus respectivos desvios padrão inferior e superior e o custo computacional de cada método, que engloba o

tempo de treinamento do classificador adicionado ao tempo de redução do conjunto de amostra completo, obtidos a partir das 30 execuções sobre o conjunto de teste. Acima da linha tracejada estão os métodos com hipótese nula não rejeitada, ou seja, aqueles que apresentaram acurácias significativamente iguais à acurácia do SVM utilizando o conjunto completo de treinamento.

O teste de Wilcoxon, por ser um teste não paramétrico, não faz relação com a média, portanto podem haver métodos que, mesmo possuindo acurácia mais baixa, sejam aceitos no teste de hipótese. Na legenda de cada gráfico consta o nome do método de redução e o percentual de amostras restantes após aplicação do método

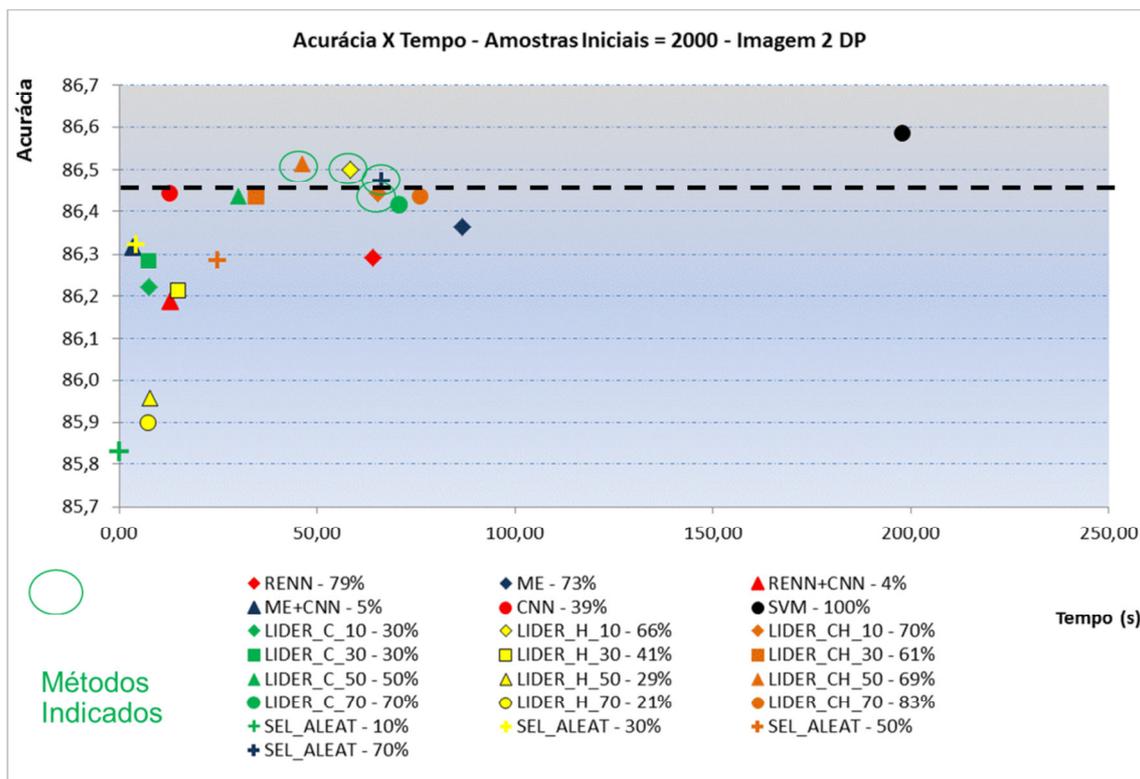
A Figura 5.30, referente à imagem sintética 1DP, relaciona a acurácia média com o tempo global das execuções obtida para cada método de redução. É importante ressaltar que os métodos LÍDER_C_30 e LÍDER_CH_30, mesmo estando acima da linha tracejada, tiveram a hipótese nula rejeitada e não são indicados como métodos de redução adequados.

Figura 5.29 – Acurácia em relação ao custo computacional - imagem sintética 1DP



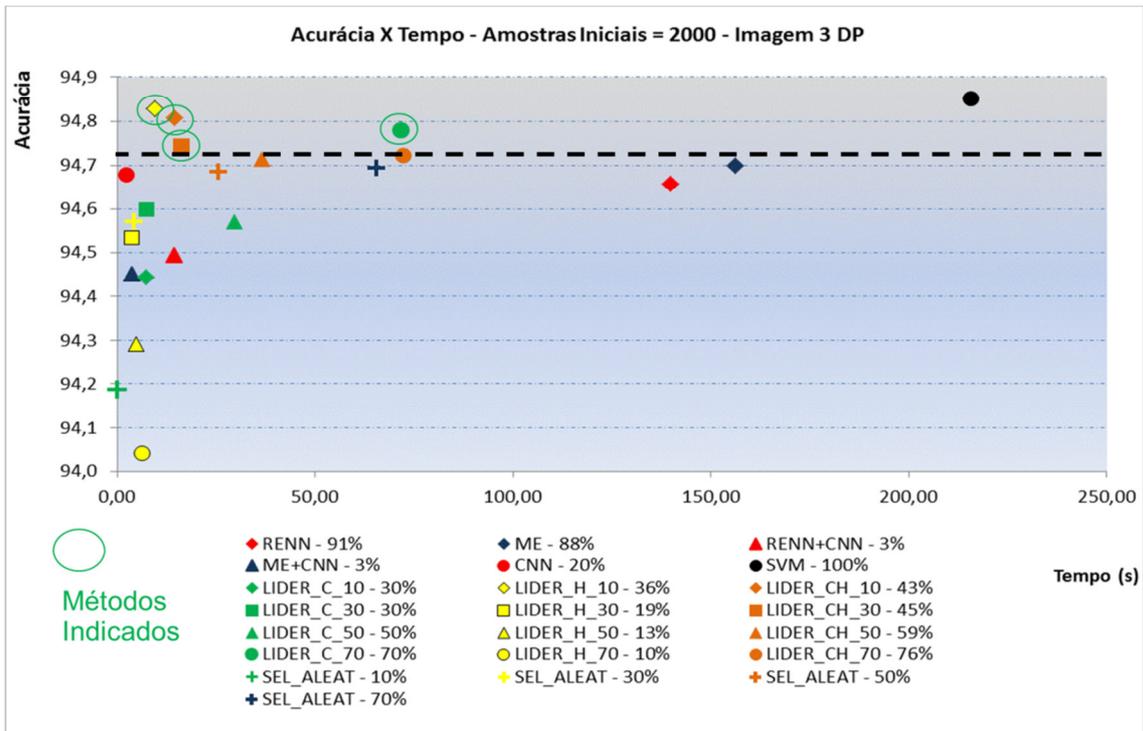
O próximo gráfico representado pela Figura 5.30 é referente à imagem sintética 2DP, similar ao gráfico anterior, com as respectivas acurácias médias dos métodos de redução e com a linha tracejada que separa os 4 métodos não rejeitados pela hipótese nula, LÍDER_H_10, LÍDER_CH_10, LÍDER_CH_50 e SEL_ALEAT_70%, dos demais métodos que tiveram a hipótese nula rejeitada.

Figura 5.30 – Acurácia em relação ao custo computacional - imagem sintética 2DP



No terceiro gráfico representado pela Figura 5.31, referente à imagem sintética 3DP e similar aos gráficos anteriores, nota-se que os quatro métodos de redução, LÍDER_C_70, LÍDER_H_10, LÍDER_CH_10 e LÍDER_CH_30, não tiveram a hipótese nula rejeitada, de acordo com o teste de Wilcoxon, diferente dos demais métodos.

Figura 5.31 – Acurácia em relação ao custo computacional - imagem sintética 2DP

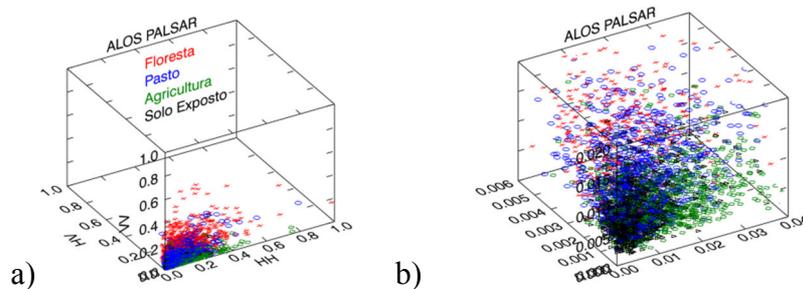


5.5 Experimentos sobre a imagem do sensor ALOS PALSAR

Neste experimento foi utilizada uma imagem de radar da região da Floresta Nacional de Tapajós-PA, descrita na seção 4.3.

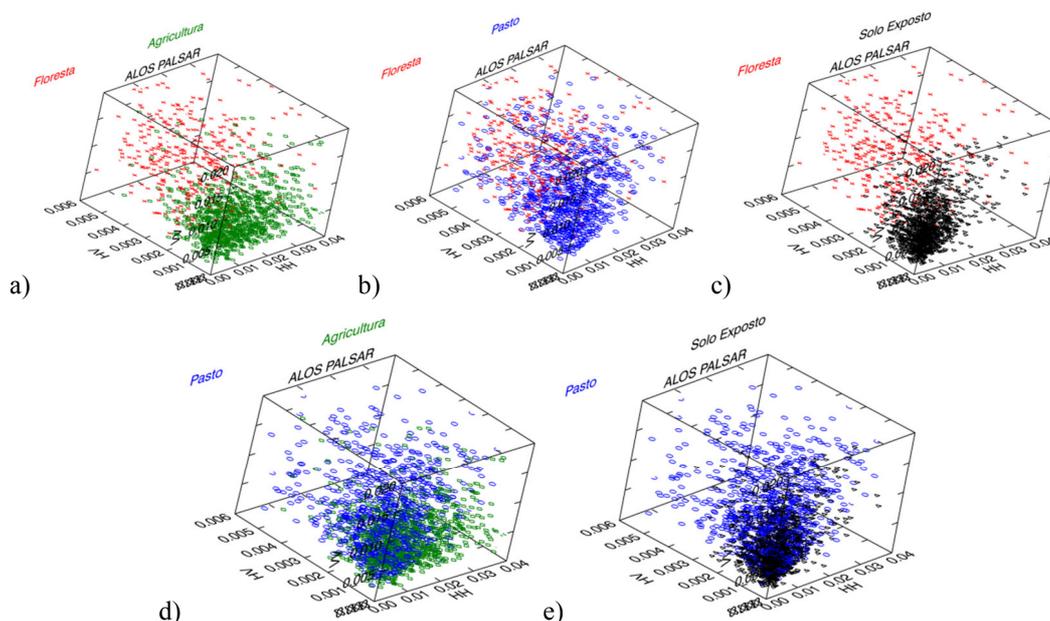
Analisando o conjunto de treinamento selecionado conforme tabela 4.4, nota-se uma concentração das amostras em um dos vértices do cubo que representa o espaço de atributos, como mostra a Figura 5.34 (a). A Figura 5.34 (b) mostra um subespaço do domínio da imagem que concentra a maior parte dos dados.

Figura 5.32 – Amostras de treinamento da imagem PALSAR



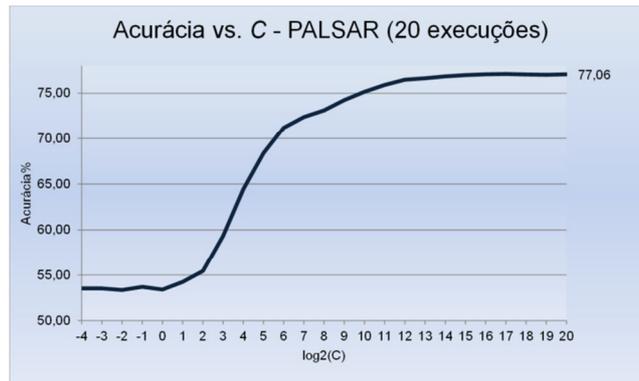
Analisando o conjunto de amostras por par de classes, verifica-se a distinção entre os pares de classes no espaço de atributos. Através dos gráficos é possível visualizar a possibilidade de aplicação dos métodos de redução, notando que em alguns pares, como exemplo Floresta e Agricultura ou Floresta e Solo Exposto, há uma separabilidade maior entre as classes. A Figura 5.35 mostra a distribuição das amostras de treinamento no espaço de atributos para os principais pares de classes.

Figura 5.33 – Amostras de treinamento por pares de classes - IKONOS



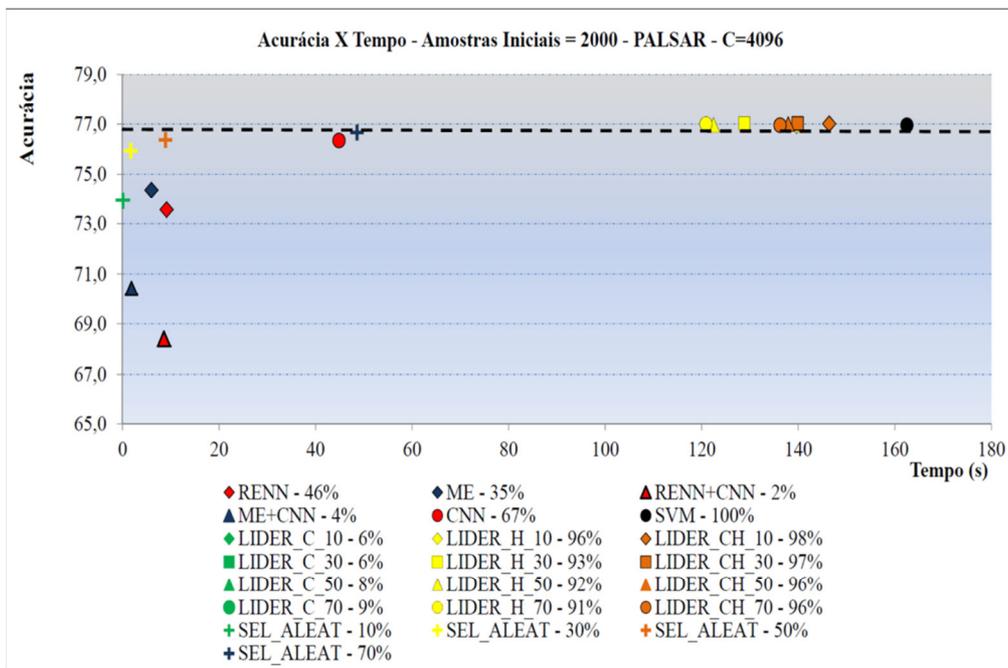
Para a seleção do parâmetro C foram realizadas 20 execuções utilizando um conjunto de treinamento com quantidade de 500 amostras por classe, retiradas com reposição para cada execução, e para avaliação e acurácia da classificação foi utilizado o conjunto de validação do conjunto selecionado conforme tabela 4.4. Foi utilizada somente esta quantidade inicial de amostras, pois o propósito deste caso é analisar o comportamento da acurácia mediante a variação da penalidade C . Analisando a Figura 5.34 nota-se grande variação da acurácia entre os valores de C entre 1 ($\text{Log}_2 0$) e 4096 ($\text{Log}_2 12$). Desta maneira, adotou-se para utilização nos experimentos de redução o valor de $C = 4096 (\text{Log}_2 12)$.

Figura 5.34 – Acurácia vs. C - Imagem ALOS PALSAR.



Os resultados obtidos para a imagem PALSAR são mostrados na Figura 5.35, que relaciona a acurácia da classificação com o tempo de processamento para cada um dos métodos. A linha tracejada separa os métodos não rejeitados no teste de hipótese dos demais. Nota-se que os oito métodos de redução não rejeitados são variações do Líder, exceto o Líder_Centros. A seleção aleatória 70% (SEL_ALEAT_70) foi rejeitada no teste de Wilcoxon, porém a média das acurácia utilizando este método ficou com diferença menor que 0,5%.

Figura 5.35 – Acurácia dos métodos de redução para a imagem ALOS PALSAR

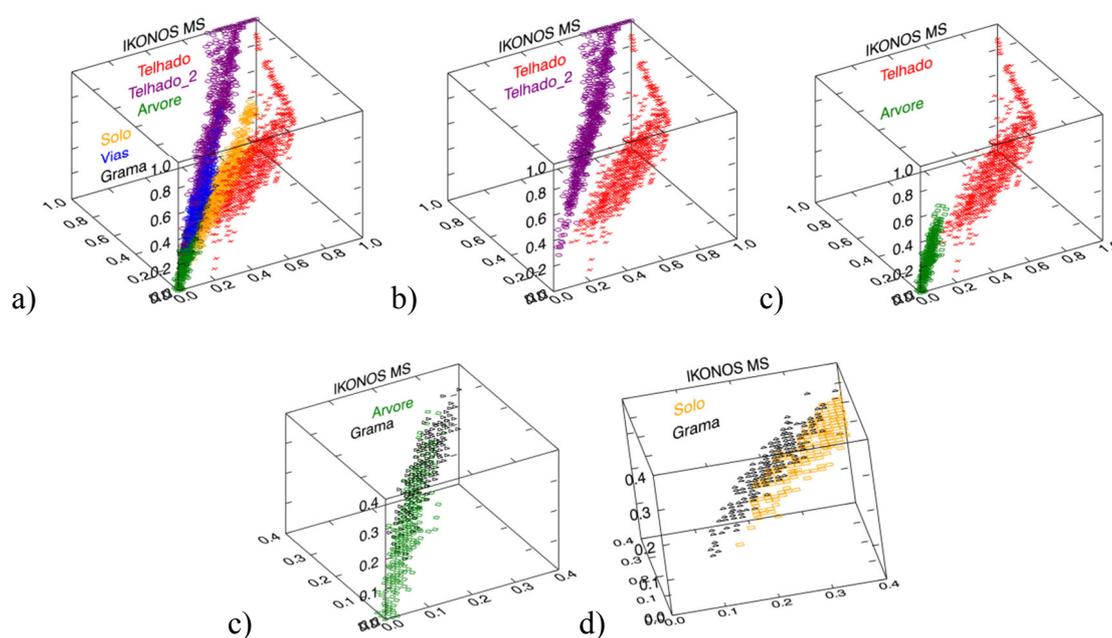


O CNN apresentou uma diferença com menos de 1% em relação ao SVM utilizando o conjunto completo, podendo ser uma opção a critério do usuário. Os métodos RENN e ME ficaram muito abaixo dos demais métodos, inclusive da seleção aleatória.

5.6 Experimentos sobre a imagem do sensor IKONOS MS

O conjunto de amostras de treinamento da imagem IKONOS, tem maior distribuição na diagonal principal, como pode ser observado na Figura 5.38(a). A separabilidade entre as classes pode ser observada no conjunto completo e, também, nos pares de classes principais. O objetivo é analisar a possibilidade de redução entre os pares de classes, conforme mostra as Figuras 5.38(b)-5.38(d). Neste aspecto, as classes grama, árvore e solo apresentam menor separabilidade.

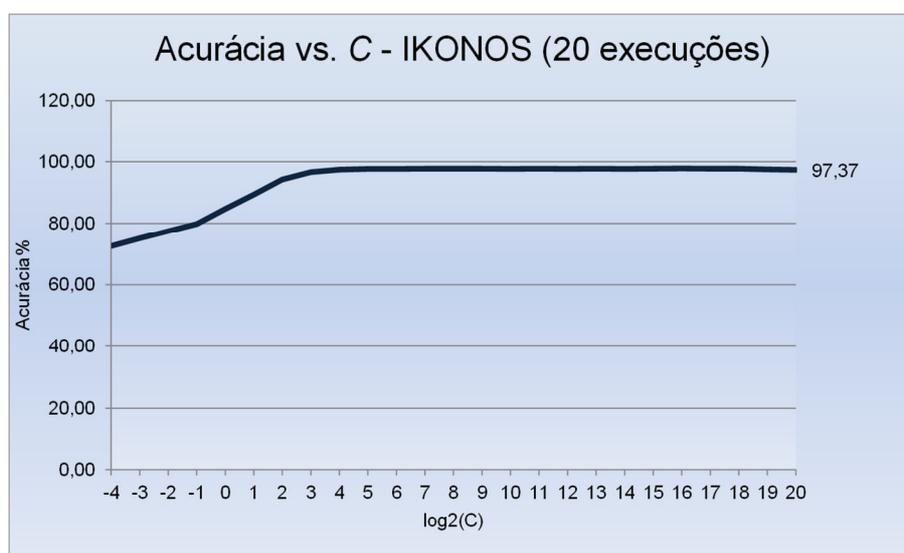
Figura 5.36 – Amostras de treinamento do IKONOS MS



Para a seleção do parâmetro C foram realizadas 20 execuções utilizando um conjunto de treinamento com quantidade de 500 amostras por classe, retiradas com reposição para cada execução, e para avaliação e acurácia da classificação foi utilizado o conjunto de

validação do conjunto selecionado conforme tabela 4.4. Foi utilizada somente esta quantidade inicial de amostras, pois o propósito deste caso é analisar o comportamento da acurácia mediante a variação da penalidade C . Analisando a Figura 5.37 fica evidente que, para esta aplicação, que os valores de C maiores que 16 ($\text{Log}_2 4$) não afetam a acurácia da classificação. Desta maneira, adotou-se para utilização nos experimentos de redução o valor de $C = 4096$ ($\text{Log}_2 12$).

Figura 5.37 – Acurácia vs. C - Imagem IKONOS

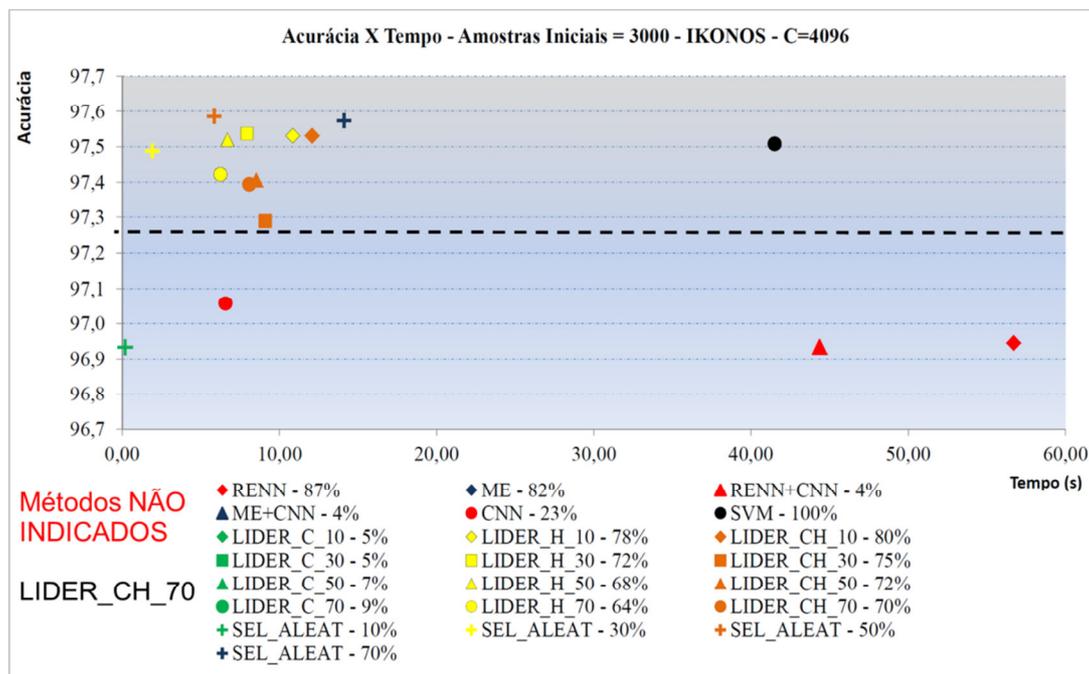


Foram realizadas 30 execuções para os experimentos com redução do conjunto de treinamento com quantidade de 500 amostras por classe, retiradas com reposição para cada execução e, para a avaliação e acurácia da classificação foi utilizado o conjunto de teste do conjunto selecionado conforme tabela 4.4. Foi analisada somente esta quantidade inicial de amostras, pois o propósito deste caso é analisar a redução e a acurácia da classificação, e não a variação para tamanhos diferentes de conjunto de treinamento.

Na Figura 5.38 dez métodos de redução proporcionam uma acurácia igual ou maior que a acurácia obtida utilizando o conjunto completo de treinamento. Neste caso, nota-se que a seleção aleatória (30%, 50% e 70%) proporcionou a maior acurácia dentre todos

os métodos de redução. As variações do Líder foram favoráveis em relação à acurácia e ao tempo global. Os métodos CNN e LID_CH_70 foram rejeitados no teste de hipótese, ambos com diferença de menos de 0,5% comparado aos SVM utilizando o conjunto completo. Uma comparação entre LID_CH_50/LID_CH_70/LID_H_30 e a SEL_ALEAT_70 mostra que tamanhos equivalentes da quantidade do conjunto de treinamento podem manter ou melhorar a acurácia, mas o custo computacional é melhor quando se utiliza um algoritmo de redução e não simplesmente uma seleção aleatória de dados. O RENN não obteve uma acurácia que fosse aceitável e seu tempo global ficou acima do SVM, provavelmente pelo fato de haver classes com maior separabilidade, e neste caso o método perde sua eficiência. O método Líder não gerou o número de centros necessários, para cobrir a região das amostras de treinamento, talvez pelo fato de como as amostras estão distribuídas no espaço de atributos (Figura 5.36).

Figura 5.38 – Acurácia dos métodos de redução para a imagem IKONOS MS



6 CONCLUSÃO

Os resultados obtidos mostraram que a redução de dados pode ser considerada uma opção viável para reduzir o custo computacional. Alguns métodos foram bem eficazes, chegando até a superar a acurácia do SVM completo, mostrando que a redução do conjunto de treinamento pode, em alguns casos, melhorar a acurácia da classificação.

Em relação ao tempo global, notou-se que apenas um dos métodos, o RENN aplicado na imagem IKONOS, apresentou tempo superior em relação ao SVM completo. A próxima etapa se resume em escolher o método com melhor acurácia, porém levando em consideração os comentários a seguir.

Antes de aplicar a redução, sugerimos que seja feita uma análise das amostras do conjunto de treinamento, no espaço de atributos, em relação à separabilidade das classes, pois cada método tem sua própria característica de redução, que pode ser adequada ou não para a imagem adotada.

A sintonia do parâmetro de penalização C é uma ação recomendada, pois não há garantia de que o valor de C adequado para o conjunto completo também seja adequado para o conjunto reduzido.

Os métodos aplicados às imagens 1DP, 2DP e 3DP foram eficazes, respectivamente em 11, 4 e 4 dos 21 métodos analisados. Para a imagem IKONOS houveram 10 métodos eficazes para redução e para imagem PALSAR 8 métodos que proporcionam uma eficácia na redução.

Na Tabela é apresentado o percentual de redução do custo computacional (tempo de redução do conjunto de treinamento + tempo de treinamento do classificador) para cada um dos métodos de redução, comparados com o tempo de treinamento utilizando o conjunto completo. Em azul têm-se os métodos com reduções maiores que 80% e que não foram rejeitados no teste de hipótese, em verde os métodos que não foram rejeitados no teste de hipótese e em branco os demais métodos.

Tabela 6.1 – Redução do Tempo Global (%) – Teste de Wilcoxon

■ Método NÃO Inferior E Redução >= 80%
 ■ Método NÃO Inferior
 Método Inferior

#	1DP	2DP	3DP	IKONOS	PALSAR
RENN	91,5	67,6	35,2	-36,6	94,4
MULTIEDIT	87,3	56,2	27,6	35,5	96,3
RENN_CONDENS	94,8	93,4	93,3	-6,9	94,7
ME_CONDENS	98,7	98,2	98,2	77,5	98,8
CONDENS	81,1	93,5	98,9	84,2	72,4
LIDER_C_10	96,6	96,3	96,6	96,8	99,6
LIDER_C_30	96,6	96,3	96,6	96,8	99,6
LIDER_C_50	86,6	84,8	86,3	95,8	99,5
LIDER_C_70	66,3	64,2	66,7	94,2	99,4
LIDER_H_10	23,8	70,5	95,6	73,9	14,1
LIDER_H_30	72,0	92,5	98,4	80,9	20,7
LIDER_H_50	87,7	96,0	97,8	83,9	24,7
LIDER_H_70	92,7	96,4	97,1	84,9	25,6
LIDER_CH_10	23,5	66,9	93,3	70,9	9,9
LIDER_CH_30	63,3	82,6	92,5	78,1	14,0
LIDER_CH_50	70,2	76,7	83,0	79,5	15,1
LIDER_CH_70	62,2	61,7	66,5	80,5	16,2
SELEC_ALEAT_FX_10	99,95	99,9	99,95	99,6	99,9
SELEC_ALEAT_FX_30	98,2	97,9	98,0	95,4	99,0
SELEC_ALEAT_FX_50	89,3	87,5	88,2	85,9	94,5
SELEC_ALEAT_FX_70	69,3	66,6	69,6	66,0	70,1

Os métodos Lider Heterogeneo (H_10;CH_10;CH_30) podem ser considerados ideais para quase todas as imagens, considerando os resultados do Teste de Wilcoxon. Esses métodos tiveram acurácia da classificação maior ou igual ao SVM completo nas imagens reais, porém são métodos que não tem um potencial de redução constante, necessitando de uma análise mais criteriosa em relação à imagem adotada para redução.

Na imagem IKONOS comparando os métodos Lider_H_50 e Selec_Aleat_50, ambos com quantidade de amostras próximas, pode ser verificado que o algoritmo Líder mostra vantagem quando aplicado sobre o conjunto.

Líder é eficaz nas imagens reais, porém, com baixa redução do custo computacional.

Uma seleção aleatória (70%) pode ser indicada, considerando a imagem adotada pelo usuário.

Na análise individual dos métodos em destaque da Tabela 6.1, pode-se citar o CNN, que é um dos métodos com maior percentual de redução do custo computacional, chegando a mais de 80% em quatro das cinco imagens analisadas.

Todos os métodos possuem potencial de redução de acordo com a imagem em que foram aplicados. Isto pode ser observado comparando-se os métodos aplicados na imagem 1DP, com menor percentual de redução do custo computacional, e os métodos aplicados na imagem 3DP, com maior percentual.

Completando a análise da Tabela 6.1, fica em destaque a Seleção Aleatória que teve bons resultados na imagem 1DP e na imagem IKONOS. Entretanto se for considerada somente a Seleção Aleatória de 70% das amostras (SEL_ALEAT_70), nota-se que o método foi aceito nas imagens 1DP, 2DP e IKONOS.

Uma avaliação somente dos resultados, sem considerar o teste de hipótese, pode levar à aceitabilidade da maioria dos métodos de redução. Isso pode ser observado na Tabela 6.2, Os métodos destacados em azul possuem diferença de menos de 0,5% (meio ponto percentual) comparado ao método SVM utilizando a o conjunto completo de amostras e redução do custo computacional com mais de 80%. Destacados em verde estão os métodos que possuem diferença de menos de 0,5% (meio ponto percentual) comparado ao método SVM, e em branco os demais métodos.

Tabela 6.2 – Redução do Tempo Global (%) – Conjunto completo -0,5%

■ Método NÃO Inferior E Redução >= 80%
 ■ Método NÃO Inferior
 Método Inferior

#	1DP	2DP	3DP	IKONOS	PALSAR
RENN	91,5	67,6	35,2	-36,6	94,4
MULTIEDIT	87,3	56,2	27,6	35,5	96,3
RENN_CONDENS	94,8	93,4	93,3	-6,9	94,7
ME_CONDENS	98,7	98,2	98,2	77,5	98,8
CONDENS	81,1	93,5	98,9	84,2	72,4
LIDER_C_10	96,6	96,3	96,6	96,8	99,6
LIDER_C_30	96,6	96,3	96,6	96,8	99,6
LIDER_C_50	86,6	84,8	86,3	95,8	99,5
LIDER_C_70	66,3	64,2	66,7	94,2	99,4
LIDER_H_10	23,8	70,5	95,6	73,9	14,1
LIDER_H_30	72,0	92,5	98,4	80,9	20,7
LIDER_H_50	87,7	96,0	97,8	83,9	24,7
LIDER_H_70	92,7	96,4	97,1	84,9	25,6
LIDER_CH_10	23,5	66,9	93,3	70,9	9,9
LIDER_CH_30	63,3	82,6	92,5	78,1	14,0
LIDER_CH_50	70,2	76,7	83,0	79,5	15,1
LIDER_CH_70	62,2	61,7	66,5	80,5	16,2
SELEC_ALEAT_FX_10	99,95	99,9	99,95	99,6	99,9
SELEC_ALEAT_FX_30	98,2	97,9	98,0	95,4	99,0
SELEC_ALEAT_FX_50	89,3	87,5	88,2	85,9	94,5
SELEC_ALEAT_FX_70	69,3	66,6	69,6	66,0	70,1

Com esta análise os métodos como o RENN, ME, CNN, SEL_ALEAT_30%, SEL_ALEAT_50% e SEL_ALEAT_70% podem ser utilizados, juntamente com os métodos mais consistentes como as variações do LÍDER que tratam os grupos heterogêneos, que justamente são os métodos que estão em destaque na Tabela 6.1.

Na comparação dos resultados é importante relatar que alguns métodos não são eficazes, como nos casos do LÍDER_H_70 e a seleção aleatória de 10% do conjunto de amostras de treinamento (SEL_ALEAT_10%).

As considerações finais sugerem o algoritmo do tipo Líder como o mais indicado para todas as situações, mostrando superioridade em ambas às avaliações. A seleção aleatória também apresentou bons resultados, nas duas avaliações, devendo ficar a critério do usuário a aplicação do método.

O CNN foi o método com maior redução de custo computacional. Não foi indicado segundo o Teste de hipótese, mas pode ser admitido na segunda avaliação.

Analisando a acurácia obtida para a imagem IKONOS, percebe-se que o conjunto completo de treinamento gasta, aproximadamente, 40 segundos para treinar o classificador. Com a aplicação dos métodos de redução, o custo do treinamento pode ser reduzido em mais de cinco vezes. As variações do Líder, que utilizam os grupos heterogêneos, e a Seleção aleatória foram considerados métodos eficazes para redução.

Em relação à imagem PALSAR, destacam-se, também, as variações do Líder que utilizam os grupos heterogêneos. Nestes métodos a acurácia foi maior ou igual ao SVM completo, com um custo reduzido.

A utilização dos métodos de redução como redutor do custo computacional é factível e pode ser aplicada com eficiência. Recomenda-se a aplicação nos casos em que o usuário deseja realizar vários experimentos sobre uma imagem, com o intuito de obter a maior acurácia possível e explorar as características da imagem adotada.

Nos casos mais simples, onde uma execução é o suficiente para classificar a imagem, deve ser observada a quantidade de classes e a quantidade de amostras envolvidas no problema. Neste caso o ganho de tempo pode ser útil.

REFERÊNCIAS BIBLIOGRÁFICAS

ABE, S. Analysis of multiclass support vector machines. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE FOR MODELLING CONTROL AND AUTOMATION, 2005, Vienna, Austria. **Proceedings...** Vienna, 2003. p.385-396.

ALEJO, R., SOTOCA, J.M., VALDOVINOS, R.M., TORIBIO, P. Edited Nearest Neighbor Rule for Improving Neural Networks Classifications. In: INTERNATIONAL CONFERENCE ON ADVANCES IN NEURAL NETWORKS, 2010, Shanghai, China. **Proceedings...** China, 2010. p.303-3010.

DEVIJVER, P. A. KITTLER, J. **Pattern recognition**: a statistical approach. London: Prentice Hall, 1982. 480p.

EXELIS. **Exelis**: visual information solutions. 2010. Acesso em 27 março 2014. Disponível em: <<http://www.exelisvis.com/>>.

HART, P.E. The condensed nearest neighbor. **IEEE Transactions on Information Theory**, v. 14, n. 3, p. 515–516, 1968.

HARTIGAN, J. **Clustering algorithms**. New York: John Wiley & Sons, 1975. 366p.

HSU, C.-W., LIN, C.-J. A comparison of methods for multiclass support vector machines. **IEEE Transactions on Neural Networks**, v. 13 n. 2, p. 415-425, 2002.

JOACHIMS, T. Making large-Scale SVM Learning practical. In: SCHOLKOPF, B.; BURGESS, C.J.C.; SMOLA, A.J. (ed.). **Advances in Kernel methods: support vector learning**. Cambridge, MA: MIT Press, 1999. cap. 11, p.169-184.

KOGGALAGE, R.; HALGAMUGE, S. Reducing the Number of Training Samples for Fast Support Vector Machine Classification. **Neural Information Processing: Letters and Reviews**, v. 2, n. 3, 2004.

NEGRI, R. G. **Máquina de vetores de suporte adaptativa ao contexto: formalização e aplicações em sensoriamento remoto**. 2013. 138 p. (sid.inpe.br/mtc-m19/2013/07.10.14.32-TDI). Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2013.

PLATT, J. Fast training of support vector machine using sequential minimal optimization, In: SCHOLKOPF, B.; BURGESS, C.J.C.; SMOLA, A.J. (ed.) **Advances in Kernel methods: support vector learning**. Cambridge, MA: MIT Press, 1999. cap. 12, p.185-209.

POWELL, M.J.D. On the quadratic-programming algorithm of Goldfarb and Idnani. **Mathematical Programming Study**, v. 25, p. 46-61, 1985.

ROMERO, E. Using the leader algorithm with support vector machines for large data sets. **Lecture Notes in Computer Science**, v. 6791, p 225–232, 2011. International Conference Artificial Neural Networks, 21, Espoo, Finland: Springer, 2011.

SHIN, H. J.; CHO, S. Z. Fast pattern selection for support vector classifiers.
In: PACIFIC-ASIA CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA
MINING, 7., 2003, Seoul, Korea. **Proceedings...** Berlin: Springer, 2003. p.376–387.

SIEGEL, S. **Nonparametric statistics for the behavioral sciences**. New York:
McGraw-Hill, 1956. 312p.

STEHMAN, S.V.; FOODY, G.M. Accuracy assessment. In: Warner, T.A., Nellis, M.D.,
Foody, G.M. (ed.) **The SAGE handbook of remote sensing**. New York: SAGE
Publications Ltd, 2009. cap. 21, p.297-309.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition**. 3. ed. New York:
Academic Press, 2008. 837p.

WANG, J.; NESKOVIC, P.; COOPER, L.N. Selecting data for fast support vector
machines training. In: CHEN, K., WANG, L. (ed.) **Trends in Neural Computation**.
Berlin: Springer Verlag, 2006. cap. 3, p.61-84.

WEBB, A. R.; COPSEY, K. D. **Statistical pattern recognition**. 3. ed. New York: John
Wiley & Sons, 2011. 672p.

WILSON, D.R., MARTINEZ, T.R. Reduction techniques for instance-based learning
algorithms. **Machine Learning**, v. 38 n. 3, p. 257-286, 2000.

WILSON, D. L. Asymptotic properties of nearest neighbor rules using edited data.
IEEE Transactions on Systems Man and Cybernetics, v. 2, n. 3, p. 408-421, 1972.

APÊNDICE A – CLASSIFICAÇÃO DAS IMAGENS SINTÉTICAS

Neste apêndice são apresentadas, para cada método, as acurácias médias (e desvios-padrões) obtidas para 30 execuções, nove tamanhos de amostra e seis valores da penalidade C .

A.1 – Imagem sintética IDP

Tabela A.1 – Acurácia x C x Tamanho da amostra – SVM completo

ACURÁCIA	200	300	400	500	600	700	800	900	1000
0,125	65,9 ± 1,9	67 ± 1,8	66,7 ± 1,3	67 ± 1,8	67,1 ± 1,3	68,1 ± 1,4	68,2 ± 1,2	67,9 ± 1,2	67,8 ± 0,8
1	67,7 ± 1,3	68,5 ± 1,4	68,5 ± 1	68,3 ± 1,4	68,5 ± 1	69 ± 0,9	68,7 ± 0,9	69 ± 1,2	69,3 ± 0,8
8	68,5 ± 1,7	68,6 ± 1,6	68,7 ± 1,1	68,7 ± 1,3	68,9 ± 1,2	69,1 ± 1,2	69,3 ± 0,9	69,3 ± 0,9	69,1 ± 0,8
64	68,2 ± 1,4	68,7 ± 1,1	68,9 ± 1,1	69 ± 1	68,8 ± 0,9	69,2 ± 1	69,2 ± 0,8	69,3 ± 0,7	69,2 ± 0,9
512	68,3 ± 1,6	68,6 ± 1,1	68,7 ± 1,4	68,8 ± 0,8	69 ± 0,9	69,1 ± 1	69,2 ± 0,9	69,2 ± 0,8	69,4 ± 0,7
4096	68,6 ± 1,5	68,6 ± 0,9	68,2 ± 1,4	69,1 ± 1	68,7 ± 0,9	68,9 ± 1,1	69 ± 1	69,2 ± 0,7	69,1 ± 1

Tabela A.2 – Acurácia vs. Tamanho da amostra – $C = 0,125$

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	65,9 ± 1,9	67 ± 1,8	66,7 ± 1,3	67 ± 1,8	67,1 ± 1,3	68,1 ± 1,4	68,2 ± 1,2	67,9 ± 1,2	67,8 ± 0,8
RENN	58,6 ± 8	60,1 ± 7	61,3 ± 6,6	60,7 ± 6,9	62,2 ± 5,6	63 ± 5,7	64,2 ± 4,5	65,3 ± 2,5	65,8 ± 2,5
ME	60,1 ± 7	58,6 ± 7,7	59,4 ± 6,6	62,9 ± 4,4	65 ± 3,3	66,9 ± 1,2	67,1 ± 1,7	67,2 ± 1,3	67,3 ± 0,9
RENN+CNN	56,7 ± 7,1	57,4 ± 8,2	54,9 ± 6,8	57,5 ± 7,9	55,8 ± 7,3	58,4 ± 8	61,3 ± 7,2	59,4 ± 7,9	57,7 ± 7,6
ME+CNN	55,6 ± 7,8	56 ± 6,7	59,2 ± 7,8	54,5 ± 6,7	55,4 ± 7,5	57,1 ± 8	55,9 ± 7,9	59,5 ± 7,6	56,6 ± 7,9
CNN	58,5 ± 8,2	58,9 ± 8,1	59,9 ± 7,8	59,2 ± 7,7	59,3 ± 7,8	61,5 ± 7,2	59,4 ± 7,9	61 ± 7,5	62,2 ± 6,9
LÍDER_C_10	55,2 ± 7,2	56,2 ± 7,5	58,7 ± 7,3	59,5 ± 7,6	58,9 ± 7,6	58,2 ± 7,4	59,4 ± 7,4	59,4 ± 7,4	58,1 ± 7,9
LÍDER_C_30	58,5 ± 7,3	58 ± 7,9	59,3 ± 7,4	60,3 ± 6,9	59,7 ± 7,1	61,5 ± 6,2	61,3 ± 6,2	62,8 ± 4,8	63,7 ± 3,4
LÍDER_C_50	58,3 ± 7,5	61,7 ± 6,7	60,2 ± 7,5	60,4 ± 7	62,4 ± 5,6	61,9 ± 5,2	63,6 ± 3,8	62,9 ± 4,3	64,1 ± 2,9
LÍDER_C_70	59,4 ± 6,9	57,5 ± 7,7	57,4 ± 7,7	61,8 ± 6,7	61,1 ± 5,9	63,9 ± 4	64,4 ± 2,7	65,4 ± 2	64,9 ± 1,6
LÍDER_H_10	61,2 ± 7,2	60,8 ± 7,3	60,3 ± 7,3	61,9 ± 6,4	63,5 ± 5,6	61,9 ± 6,1	65,3 ± 3,1	66,1 ± 3,5	66,7 ± 1,8
LÍDER_H_30	56 ± 7,6	56,4 ± 7,7	60,6 ± 7,4	59 ± 8,2	62,1 ± 7	60,6 ± 7,9	56 ± 7,6	57,3 ± 8,4	59,1 ± 7,1
LÍDER_H_50	56,4 ± 7,2	56,7 ± 7,9	56,5 ± 7,8	55,5 ± 7,5	58,5 ± 8,2	56,4 ± 8,1	56,2 ± 7,9	59,8 ± 8,2	59,4 ± 8,5
LÍDER_H_70	55,4 ± 7,5	56,5 ± 7,7	56,5 ± 7,2	56,7 ± 8	58,5 ± 8,3	57,6 ± 8,5	58,2 ± 8,4	60 ± 8	58,9 ± 8
LÍDER_HC_10	58,7 ± 8	62,6 ± 6,7	63,9 ± 5	63 ± 6,3	62,8 ± 5,1	63,9 ± 3,4	65,5 ± 2,4	66,5 ± 2,3	67 ± 1,3
LÍDER_HC_30	59,6 ± 7,6	58,1 ± 7,9	59,2 ± 7,8	59,4 ± 7	61,7 ± 5,9	61,7 ± 5,9	64,1 ± 3,2	64,8 ± 3,7	66,4 ± 1,7
LÍDER_HC_50	59,6 ± 7,2	60,3 ± 7,6	59,5 ± 7,5	58,4 ± 7,3	61,1 ± 6,1	63,1 ± 5,1	63,5 ± 4,7	64,8 ± 3	66,1 ± 2,3
LÍDER_HC_70	57,4 ± 7,2	60,6 ± 7,7	60,3 ± 7,5	60 ± 6,8	62,3 ± 5,8	63 ± 4,5	64,1 ± 4,4	66 ± 1,5	66,1 ± 1,7
SELEÇÃO_10%	63 ± 3,5	64,2 ± 3,1	65,2 ± 3,5	65,2 ± 3,1	65,7 ± 2,9	66,2 ± 2,5	65 ± 2,7	65,8 ± 2,1	65,8 ± 2,1
SELEÇÃO_30%	64,7 ± 2,1	65,9 ± 2	66,8 ± 1,6	66,4 ± 2	66,1 ± 2,4	66,1 ± 1,9	66,5 ± 1,9	66,3 ± 2,2	66,1 ± 1,7
SELEÇÃO_50%	65,1 ± 2,4	66,1 ± 2,3	66,2 ± 3,6	66,8 ± 1,9	66,9 ± 1,7	66,2 ± 3	67,2 ± 2	66,5 ± 1,3	67,2 ± 1,5
SELEÇÃO_70%	66,2 ± 2,6	65,1 ± 3,5	66,4 ± 1,6	66,3 ± 1,8	66,8 ± 1,9	67 ± 1,2	66,8 ± 1,5	67,4 ± 1,4	67,8 ± 1,5

Tabela A.3 – Resultado dos métodos - Acurácia vs. Tamanho da amostra – C = 1

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	67,7 ± 1,3	68,5 ± 1,4	68,5 ± 1	68,3 ± 1,4	68,5 ± 1	69 ± 0,9	68,7 ± 0,9	69 ± 1,2	69,3 ± 0,8
RENN	65,6 ± 3,4	67,4 ± 1,9	67,8 ± 1,7	67,8 ± 1,4	68,1 ± 1,3	68,6 ± 0,9	68 ± 1,6	68,5 ± 1,5	68,5 ± 0,9
ME	67,2 ± 2,4	67,3 ± 1,5	68 ± 1,5	67,7 ± 1,4	68 ± 1,2	68,7 ± 1,2	68,2 ± 1,4	68,8 ± 1,5	68,7 ± 1,4
RENN+CNN	56,1 ± 7,4	57,5 ± 7,8	58,3 ± 7,7	57 ± 7,7	59,8 ± 7,3	56,5 ± 7,9	55 ± 7	56,6 ± 7,7	56,2 ± 7,2
ME+CNN	59,8 ± 8	58,1 ± 8	57,9 ± 7,7	58,4 ± 7,7	58,5 ± 7,8	54,7 ± 7,2	55,9 ± 7,2	58,1 ± 7,9	58 ± 7,6
CNN	61,5 ± 6,3	65,4 ± 3,6	67 ± 1,7	67,3 ± 1,5	68 ± 1,3	68,3 ± 1,3	68,1 ± 1,2	68,5 ± 1,1	68,7 ± 1
LÍDER_C_10	58,4 ± 6,9	56,3 ± 7,2	58,7 ± 6,3	59,4 ± 6	63 ± 5,3	64,6 ± 4,9	65,7 ± 2	65,8 ± 2	66,2 ± 1,9
LÍDER_C_30	64,1 ± 4,9	66,7 ± 1,6	66,9 ± 1,5	67 ± 1,5	67,5 ± 1,5	68,3 ± 1,2	67,4 ± 1,2	68 ± 1,2	68,2 ± 1,3
LÍDER_C_50	66,5 ± 1,8	66,5 ± 1,7	66,8 ± 1,1	67,2 ± 1,3	67,6 ± 1,6	68,3 ± 1,1	67,6 ± 1,4	68,3 ± 1,5	67,9 ± 1,2
LÍDER_C_70	66,3 ± 1,7	66,9 ± 1,3	67,4 ± 1,3	67,8 ± 1,6	67,5 ± 1,4	68,4 ± 1,2	67,8 ± 1	68 ± 1,5	68,1 ± 1,2
LÍDER_H_10	67 ± 2,7	68,1 ± 1,5	68,7 ± 0,9	68,5 ± 1,4	68,2 ± 1,2	68,6 ± 1	68,4 ± 1,2	68,8 ± 1	69 ± 0,8
LÍDER_H_30	60,2 ± 7	65,9 ± 4	66,8 ± 3	67,3 ± 2,4	68,1 ± 1,5	68,1 ± 1,6	68 ± 1,2	68,2 ± 1,5	68,7 ± 1,2
LÍDER_H_50	57,2 ± 6,9	62,7 ± 5,5	62,9 ± 6,5	65,7 ± 4,2	66,5 ± 2,8	67,6 ± 1,6	67,8 ± 1,7	68 ± 1,2	68,8 ± 1,2
LÍDER_H_70	54,4 ± 7	58,7 ± 6,8	61,5 ± 6,5	61,9 ± 6,1	65,6 ± 4,4	66,4 ± 3,9	66,5 ± 3,4	67,7 ± 1,6	68,3 ± 1,4
LÍDER_HC_10	66,9 ± 2,2	68,1 ± 1,4	68,4 ± 1,1	68,2 ± 1,4	68,3 ± 1	68,9 ± 1	68,5 ± 1	68,9 ± 1,1	69 ± 0,9
LÍDER_HC_30	66,1 ± 1,7	67,5 ± 1,5	68,1 ± 1,3	67,9 ± 1,4	68,7 ± 1	68,7 ± 1,1	68 ± 1,1	68,7 ± 1,5	69 ± 0,9
LÍDER_HC_50	66 ± 3,1	68 ± 1,5	68,1 ± 1,2	67,6 ± 1,7	68,2 ± 1,4	68,7 ± 1,2	68 ± 1,1	68,6 ± 1,3	69 ± 0,9
LÍDER_HC_70	67 ± 1,8	67,6 ± 1,6	68,2 ± 1,4	68 ± 1,4	68,1 ± 1,2	68,8 ± 0,8	68,1 ± 1,1	68,4 ± 1,2	68,7 ± 0,8
SELEÇÃO_10%	63,2 ± 3,8	63,6 ± 3,7	64,3 ± 3,3	65,6 ± 2,3	65,6 ± 2,3	65,4 ± 2,4	67,2 ± 2,1	66,7 ± 2,4	66,8 ± 1,9
SELEÇÃO_30%	65 ± 3,4	66,6 ± 2,4	66,8 ± 1,8	67,7 ± 1,6	67,7 ± 1,7	67,9 ± 1,7	68,4 ± 1,4	67,6 ± 1,7	67,7 ± 1,5
SELEÇÃO_50%	66,1 ± 2,4	67,5 ± 1,9	67,8 ± 1,7	68 ± 1,8	67,8 ± 1,2	68,3 ± 1,5	68,5 ± 1,2	68,9 ± 1,1	68,2 ± 1,1
SELEÇÃO_70%	67,1 ± 2,2	68,2 ± 1,7	68,3 ± 1,2	68,1 ± 1,2	67,9 ± 1,3	68,6 ± 1,1	68,5 ± 1,2	68,6 ± 1	68,5 ± 0,9

Tabela A.4 – Acurácia vs. Tamanho da amostra – C = 8

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	68,5 ± 1,7	68,6 ± 1,6	68,7 ± 1,1	68,7 ± 1,3	68,9 ± 1,2	69,1 ± 1,2	69,3 ± 0,9	69,3 ± 0,9	69,1 ± 0,8
RENN	67,5 ± 1,5	67,6 ± 2,2	68,2 ± 1,6	68,6 ± 1,5	68,8 ± 1,2	68,7 ± 1,2	68,5 ± 1,1	68,9 ± 1,1	68,8 ± 1
ME	67,7 ± 1,8	68,2 ± 1,2	68 ± 1,7	68,2 ± 1,3	68,3 ± 1,5	68,5 ± 1,3	68,7 ± 1,1	68,9 ± 1	68,6 ± 1,1
RENN+CNN	64,4 ± 3,5	62,9 ± 5,2	65,9 ± 2,8	66,6 ± 2,5	66,8 ± 2,2	67,6 ± 1,8	67,7 ± 1,8	67,8 ± 1,7	67,5 ± 1,8
ME+CNN	66 ± 2,3	65,4 ± 4	66,8 ± 3,2	67,1 ± 2,6	67,6 ± 1,5	68 ± 1,4	68,3 ± 1,9	68,2 ± 1,8	67,7 ± 1,5
CNN	67,9 ± 1,5	68 ± 1,8	68,5 ± 1	67,8 ± 1,3	68,6 ± 1,5	68,9 ± 1,4	68,7 ± 0,9	68,6 ± 1	68,6 ± 0,8
LÍDER_C_10	63,8 ± 4,5	65,8 ± 2,5	66 ± 3,2	66,6 ± 2	67 ± 1,8	67,2 ± 2	67,6 ± 1,6	66,8 ± 1,5	67,6 ± 1,8
LÍDER_C_30	67,2 ± 2,1	67,4 ± 2,3	67,7 ± 1,7	67,8 ± 1,9	68,2 ± 1,5	68,2 ± 1,2	68,6 ± 1,2	68,3 ± 1,5	68,5 ± 1,1
LÍDER_C_50	67,5 ± 1,7	68 ± 1,9	67,9 ± 1,7	67,6 ± 1,6	68,5 ± 1,3	68,3 ± 1,3	68,2 ± 1,2	68,8 ± 0,9	68,2 ± 0,8
LÍDER_C_70	67,8 ± 1,7	67,7 ± 2,3	68,1 ± 1,4	67,9 ± 1,7	68,4 ± 1,1	68,5 ± 1,1	68,7 ± 1,1	68,7 ± 1,2	68,8 ± 1
LÍDER_H_10	68,6 ± 1,7	68,4 ± 1,5	68,5 ± 1,4	68,5 ± 1,5	68,6 ± 1,1	69,1 ± 1	69,1 ± 0,9	69,1 ± 1	68,8 ± 0,8
LÍDER_H_30	67,6 ± 1,9	68,1 ± 1,5	68,7 ± 1,3	68,5 ± 1,2	68,5 ± 1,3	68,8 ± 1,2	68,7 ± 1,1	68,9 ± 1,2	68,5 ± 1,2
LÍDER_H_50	67,1 ± 2,3	67,6 ± 1,7	68,6 ± 1,3	68,1 ± 1,5	67,8 ± 1,6	68,6 ± 1,3	68,5 ± 1	68,2 ± 1,1	68,8 ± 1
LÍDER_H_70	67 ± 3,6	67,3 ± 2	68,2 ± 1,4	68,2 ± 1,4	68,2 ± 1,6	68 ± 1,4	68,2 ± 1,3	68,2 ± 1,1	68,7 ± 1
LÍDER_HC_10	68,2 ± 1,3	68,3 ± 1,4	68,8 ± 1	68,7 ± 1,6	68,7 ± 1	68,9 ± 1,2	69,1 ± 1	69,1 ± 0,8	69 ± 0,8
LÍDER_HC_30	68 ± 1,6	68,3 ± 1,6	68,8 ± 1,3	68,5 ± 1,4	68,7 ± 1,1	69 ± 1,1	69,1 ± 1,1	69,2 ± 0,9	69,3 ± 0,8
LÍDER_HC_50	67,9 ± 1,7	68,4 ± 1,8	68,4 ± 1,2	68,3 ± 1,4	68,8 ± 1,1	68,7 ± 1	69,2 ± 1	69 ± 0,8	69,2 ± 0,7
LÍDER_HC_70	68 ± 1,8	68,3 ± 1,6	68,3 ± 1,3	68,1 ± 1,7	68,5 ± 1,3	68,8 ± 1,3	69 ± 1,1	68,9 ± 1	69,1 ± 0,7
SELEÇÃO_10%	64,6 ± 4	66,2 ± 2,5	65,8 ± 2,8	66,7 ± 2	65,9 ± 2,3	67,3 ± 1,9	67,5 ± 1,9	68,4 ± 1,7	68 ± 1,6
SELEÇÃO_30%	67,1 ± 1,9	67,4 ± 1,8	67,6 ± 1,6	67,8 ± 2	68,6 ± 1,5	68,5 ± 1,5	68,2 ± 1,1	68,2 ± 1,2	68,4 ± 1,1
SELEÇÃO_50%	68,1 ± 1,7	68,1 ± 1,3	68,4 ± 1,4	68,2 ± 1,4	68,5 ± 1,5	68,7 ± 1,1	68,7 ± 1,3	69 ± 0,8	68,8 ± 1,2
SELEÇÃO_70%	67,8 ± 2	68,3 ± 1,5	68,6 ± 1,4	68,6 ± 1,1	68,5 ± 1,1	68,8 ± 1	68,8 ± 0,9	69 ± 1	69,1 ± 0,9

Tabela A.5 – Acurácia vs. Tamanho da amostra – $C = 64$

<i>ACURÁCIA</i>	200	300	400	500	600	700	800	900	1000
SVM	68,2 ± 1,4	68,7 ± 1,1	68,9 ± 1,1	69 ± 1	68,8 ± 0,9	69,2 ± 1	69,2 ± 0,8	69,3 ± 0,7	69,2 ± 0,9
RENN	67,5 ± 1,9	67,7 ± 1,8	68,2 ± 1,3	68,5 ± 1,5	68,4 ± 1,4	68,4 ± 1,2	68,8 ± 1,4	68,9 ± 1	68,8 ± 1,2
ME	67,5 ± 2	67,7 ± 1,9	68,3 ± 1,2	68,4 ± 1,6	68,4 ± 1,6	69 ± 1,4	68,8 ± 1,4	68,9 ± 1	68,8 ± 1
RENN+CNN	67,1 ± 1,6	67,5 ± 1,7	68,1 ± 1,9	68,2 ± 1,8	68,2 ± 1,6	68 ± 1,5	68,7 ± 1,2	68,3 ± 1,5	68,5 ± 1,3
ME+CNN	67,1 ± 2	67,6 ± 2	67,7 ± 1,6	67,8 ± 1,7	68,1 ± 1,4	68,6 ± 1,1	68,7 ± 1,3	68,3 ± 1,5	68,9 ± 1,3
CNN	68,2 ± 1,7	68,4 ± 1,4	68,6 ± 1,3	68,9 ± 1,3	68,7 ± 1,1	68,7 ± 1,3	69,1 ± 1,1	69,2 ± 0,9	68,9 ± 0,9
LÍDER_C_10	65,5 ± 2,4	66,4 ± 2,3	67,1 ± 2,1	67,3 ± 2,1	67,6 ± 2	67,4 ± 2,2	67,4 ± 1,9	67,7 ± 1,3	68,2 ± 1,5
LÍDER_C_30	67,8 ± 1,7	67,4 ± 1,9	68,2 ± 1,3	68,3 ± 1,4	68,1 ± 1,4	68,1 ± 1,1	68,3 ± 1,2	68,5 ± 1	68,5 ± 1,2
LÍDER_C_50	68,1 ± 1,4	68 ± 1,7	68,2 ± 1,4	68,9 ± 1,3	68,5 ± 1,1	68,5 ± 1,2	68,5 ± 1,4	68,6 ± 1,1	68,7 ± 1,1
LÍDER_C_70	67,8 ± 1,7	68,2 ± 1,4	68,2 ± 1,4	68,5 ± 1,4	68,3 ± 1,3	68,4 ± 1,4	68,8 ± 0,9	68,7 ± 1	68,7 ± 1,3
LÍDER_H_10	68,5 ± 1,5	68,5 ± 1,4	68,7 ± 1,1	68,7 ± 1,3	68,8 ± 0,7	69,1 ± 1	68,9 ± 0,8	69,3 ± 0,9	69 ± 0,9
LÍDER_H_30	68,5 ± 1,4	68,4 ± 1,3	68,6 ± 1,1	68,2 ± 1,3	68,6 ± 1	68,7 ± 1,1	68,6 ± 1	69,1 ± 0,9	68,9 ± 1
LÍDER_H_50	67,4 ± 2	67,6 ± 1,6	68,8 ± 1,2	68,2 ± 1,3	68,6 ± 0,9	68,6 ± 1,1	68,6 ± 1,1	68,9 ± 1,1	68,8 ± 1,1
LÍDER_H_70	67,1 ± 1,8	67,9 ± 1,8	68,3 ± 1,3	68,1 ± 1,9	68,1 ± 1,4	68,6 ± 1,3	68,9 ± 1,1	68,6 ± 1,3	68,7 ± 1,2
LÍDER_HC_10	68,3 ± 1,6	68,5 ± 1,2	68,9 ± 1	68,9 ± 1,1	68,9 ± 0,8	69,1 ± 1	69 ± 0,8	69,1 ± 0,8	69,1 ± 0,9
LÍDER_HC_30	68,4 ± 1,7	68,5 ± 1,3	68,6 ± 1,4	68,6 ± 1,2	68,7 ± 1,2	68,9 ± 1	69 ± 0,9	69,2 ± 0,9	68,8 ± 0,9
LÍDER_HC_50	68,1 ± 1,4	68,5 ± 1,3	68,6 ± 1,3	68,9 ± 1,2	68,7 ± 1	69,1 ± 1	68,8 ± 1,1	69,2 ± 0,8	69,1 ± 0,9
LÍDER_HC_70	68 ± 1,2	68,5 ± 1,1	68,8 ± 1,2	69 ± 1,2	68,6 ± 1,1	69,1 ± 0,9	69,1 ± 1,1	69,2 ± 0,8	69,1 ± 0,8
SELEÇÃO_10%	65,4 ± 3,4	65,9 ± 2,6	67,9 ± 2,2	67,9 ± 1,8	67,4 ± 2,1	67,3 ± 2,1	67,9 ± 2,1	67,6 ± 2	68,3 ± 1,7
SELEÇÃO_30%	67,3 ± 2,1	67,6 ± 1,7	68 ± 1,7	68,5 ± 1,8	68,1 ± 1,3	68,4 ± 1,7	68,3 ± 1,5	68,5 ± 1,1	69 ± 1,1
SELEÇÃO_50%	68,1 ± 1,7	68 ± 1,7	68,3 ± 1,4	68,8 ± 1,1	68,5 ± 1,2	68,4 ± 1,2	68,8 ± 1,1	68,7 ± 1,1	68,6 ± 0,9
SELEÇÃO_70%	68,4 ± 1,5	68,4 ± 1,1	68,7 ± 1,1	69,1 ± 1	68,6 ± 1,2	68,7 ± 1,1	68,4 ± 0,9	69,1 ± 1	69 ± 0,8

Tabela A.6 – Acurácia vs. Tamanho da amostra – $C = 512$

<i>ACURÁCIA</i>	200	300	400	500	600	700	800	900	1000
SVM	68,3 ± 1,6	68,6 ± 1,1	68,7 ± 1,4	68,8 ± 0,8	69 ± 0,9	69,1 ± 1	69,2 ± 0,9	69,2 ± 0,8	69,4 ± 0,7
RENN	67,4 ± 1,9	67,9 ± 1,5	68,1 ± 1,8	68,2 ± 1,7	68,7 ± 1,1	68,9 ± 1,3	68,3 ± 1,5	68,7 ± 1	68,3 ± 1,1
ME	67,4 ± 2,3	68,4 ± 1,5	68,1 ± 1,5	68,3 ± 1,5	68,9 ± 1,2	68,7 ± 1,4	68,5 ± 1,3	68,6 ± 0,9	68,6 ± 1
RENN+CNN	66,3 ± 2,6	67,2 ± 2,2	68,1 ± 1,4	68 ± 1,5	68,9 ± 1,2	68,5 ± 1,5	67,8 ± 1,4	68,3 ± 1,3	68,9 ± 1,5
ME+CNN	67,1 ± 1,8	67,4 ± 2	68,1 ± 1,5	67,7 ± 1,4	68,1 ± 1,2	68,4 ± 1,5	68,3 ± 1,5	68,4 ± 1,3	68,2 ± 1,4
CNN	68,1 ± 1,7	68,3 ± 1,2	68,4 ± 1,3	68,7 ± 1,2	68,5 ± 1,2	68,7 ± 1,2	69,2 ± 0,9	69,1 ± 0,9	68,9 ± 0,8
LÍDER_C_10	65 ± 3,3	66,6 ± 2,2	67,1 ± 2,1	67,2 ± 1,9	67,3 ± 2,2	67,5 ± 1,5	67,9 ± 1,6	67,7 ± 1,5	67,7 ± 1,9
LÍDER_C_30	67,9 ± 1,7	67,8 ± 1,5	67,8 ± 1,8	68,2 ± 1,4	67,8 ± 1,8	68,7 ± 1,3	68,5 ± 1,1	68,1 ± 1,3	68,7 ± 1,1
LÍDER_C_50	68,1 ± 1,9	68,2 ± 1,8	68,3 ± 1,6	68,6 ± 1,2	68,5 ± 1	68 ± 1,4	68,7 ± 1,3	68,3 ± 1,1	68,6 ± 1,3
LÍDER_C_70	68,1 ± 1,9	68,3 ± 1,3	68,2 ± 1,8	68,9 ± 1	68,4 ± 1,4	68,8 ± 1,1	68,8 ± 1,2	68,4 ± 1,1	68,6 ± 1,1
LÍDER_H_10	68,1 ± 1,5	68,6 ± 1,2	68,6 ± 1,4	68,5 ± 1	68,7 ± 0,9	69,1 ± 0,9	68,8 ± 1	69,2 ± 0,8	69,3 ± 0,8
LÍDER_H_30	67,9 ± 1,6	68,4 ± 1,2	68,7 ± 1,1	68,3 ± 1,3	68,4 ± 1,1	68,7 ± 1,1	68,4 ± 1,1	68,8 ± 0,9	68,9 ± 0,9
LÍDER_H_50	68 ± 1,8	68,3 ± 1,2	68,4 ± 1,5	68,4 ± 1,3	68,3 ± 1,5	68,3 ± 0,9	68,2 ± 1,4	68,6 ± 1,2	68,8 ± 1
LÍDER_H_70	67,8 ± 2	68,3 ± 1,3	67,6 ± 1,4	68,5 ± 1,6	68,4 ± 1,4	68,7 ± 1,2	68,3 ± 1,2	68,9 ± 1,2	68,6 ± 0,7
LÍDER_HC_10	68,2 ± 1,5	68,4 ± 1,3	68,6 ± 1,4	68,8 ± 0,9	68,6 ± 1	69 ± 1	68,8 ± 1,1	69,2 ± 1	69,2 ± 0,8
LÍDER_HC_30	67,9 ± 1,6	68,7 ± 1,3	68,3 ± 1,4	68,6 ± 0,9	68,7 ± 1	69 ± 1,3	69 ± 1	69,2 ± 1	69,1 ± 0,8
LÍDER_HC_50	68 ± 1,8	68,7 ± 1,2	68,2 ± 1,5	68,7 ± 0,9	68,7 ± 1	68,7 ± 1	68,8 ± 1,2	68,9 ± 0,9	69,3 ± 0,8
LÍDER_HC_70	68,1 ± 1,7	68,5 ± 1,2	68,4 ± 1,7	68,7 ± 1	68,5 ± 0,9	68,9 ± 1,1	69 ± 1,1	69,1 ± 1	69,4 ± 0,8
SELEÇÃO_10%	66,5 ± 2,8	66,1 ± 2,2	67,3 ± 2,2	67 ± 2,6	67,9 ± 1,6	67,9 ± 1,6	66,8 ± 2	67,6 ± 1,6	67,8 ± 1,5
SELEÇÃO_30%	67,1 ± 1,9	67,9 ± 1,8	68,5 ± 1,5	68,7 ± 1,3	68,5 ± 1,5	68,5 ± 1,5	68,5 ± 1,5	68,4 ± 1,4	68,7 ± 1,3
SELEÇÃO_50%	68,3 ± 1,4	68,6 ± 1,8	68,1 ± 1,4	68,3 ± 1,3	68,8 ± 1	68,7 ± 1,1	68,8 ± 1	68,8 ± 1,1	68,6 ± 0,9
SELEÇÃO_70%	68 ± 1,6	68,3 ± 1,5	68,3 ± 1,4	68,3 ± 1,3	68,9 ± 1	68,8 ± 1,2	68,7 ± 1,4	69,1 ± 0,9	69,3 ± 0,8

Tabela A.7 – Acurácia vs. Tamanho da amostra – $C = 4096$

<i>ACURÁCIA</i>	200	300	400	500	600	700	800	900	1000
SVM	68,6 ± 1,5	68,6 ± 0,9	68,2 ± 1,4	69,1 ± 1	68,7 ± 0,9	68,9 ± 1,1	69 ± 1	69,2 ± 0,7	69,1 ± 1
RENN	68,2 ± 2,3	67,6 ± 1,7	67,3 ± 1,5	68,5 ± 1,2	68,2 ± 1,4	68,4 ± 1,4	68,5 ± 1,4	68,8 ± 1,2	68,7 ± 1,4
ME	67,2 ± 2	68,1 ± 1,5	67,5 ± 2,1	68,3 ± 1,7	68,6 ± 1,2	68,4 ± 1,5	68,8 ± 0,9	68,8 ± 1,2	68,7 ± 1,3
RENN+CNN	67,3 ± 2,3	67,1 ± 2	67,4 ± 1,7	68,1 ± 1,7	68,4 ± 1,6	68,4 ± 1,6	68 ± 1,7	69 ± 1,2	68,4 ± 1,5
ME+CNN	67,5 ± 2,2	67,3 ± 1,3	67,8 ± 1,4	68 ± 1,4	68,2 ± 1,6	68,1 ± 1,5	68,4 ± 1,3	68,6 ± 1,2	68,8 ± 1,3
CNN	68,2 ± 1,8	68,2 ± 1,7	68 ± 1,5	68,9 ± 1,1	68,4 ± 1,1	68,3 ± 1,2	68,8 ± 0,9	68,8 ± 1	68,9 ± 1,1
LÍDER_C_10	65,8 ± 3,4	66 ± 3,3	66,8 ± 1,9	67,4 ± 2,1	67,5 ± 1,7	67,6 ± 1,8	67,3 ± 1,5	68,2 ± 1,3	68,3 ± 1,2
LÍDER_C_30	67,4 ± 1,9	67,9 ± 1,5	67,5 ± 1,6	68,4 ± 1,7	68,1 ± 1,3	68,1 ± 1,3	68,1 ± 1,1	68,6 ± 1,1	68,5 ± 1,1
LÍDER_C_50	67,8 ± 1,8	67,9 ± 1,5	67,5 ± 1,7	68,7 ± 1,2	68,3 ± 1	68,6 ± 1,3	68,3 ± 1	68,5 ± 1	68,4 ± 1,2
LÍDER_C_70	68 ± 1,6	68,1 ± 1,3	68 ± 1,4	68,4 ± 1,4	68,1 ± 1,3	68,4 ± 1	68,6 ± 1,2	68,8 ± 1	68,7 ± 1,1
LÍDER_H_10	68,6 ± 1,6	68,5 ± 1	68,6 ± 1,3	68,8 ± 1,1	68,6 ± 0,8	68,7 ± 1,1	68,9 ± 0,9	69 ± 1	69 ± 0,9
LÍDER_H_30	68,4 ± 1,6	68,4 ± 1,2	68,1 ± 1,5	68,7 ± 1,3	68,4 ± 1,3	68,6 ± 1,2	68,4 ± 1,2	68,8 ± 1,1	68,8 ± 1,1
LÍDER_H_50	67,7 ± 1,4	67,9 ± 1,6	68,5 ± 1,6	68,9 ± 1,3	68,3 ± 1,2	68,7 ± 1,2	68,5 ± 1,2	68,8 ± 1,1	68,7 ± 1
LÍDER_H_70	67,7 ± 3,8	68,1 ± 1,5	67,5 ± 1,6	68,1 ± 1,7	68,6 ± 1,3	68,4 ± 1,4	68,5 ± 1,4	68,6 ± 1,1	68,3 ± 1,1
LÍDER_HC_10	68,5 ± 1,4	68,5 ± 1	68,2 ± 1,3	68,9 ± 1	68,6 ± 1	68,7 ± 1,1	68,7 ± 1,2	69,1 ± 0,9	69 ± 0,9
LÍDER_HC_30	68,7 ± 1,3	68,5 ± 1,3	68,1 ± 1,5	68,8 ± 1,2	68,5 ± 0,9	68,8 ± 1,1	69,1 ± 0,9	69 ± 0,9	69,1 ± 1
LÍDER_HC_50	68,5 ± 1,5	68,1 ± 1,2	68,1 ± 1,3	68,9 ± 1,3	68,5 ± 0,9	68,8 ± 1,1	69 ± 1,2	68,9 ± 0,8	68,9 ± 1
LÍDER_HC_70	68 ± 1,7	68,3 ± 1,2	68,1 ± 1,3	68,5 ± 1,4	68,5 ± 0,7	68,7 ± 1,1	68,6 ± 0,9	69,1 ± 0,9	69 ± 0,8
SELEÇÃO_10%	65,6 ± 3,9	66,5 ± 3,4	67,1 ± 2,2	67,4 ± 2,4	68,1 ± 1,4	67,9 ± 1,6	68,3 ± 1,9	67,8 ± 1,8	68,2 ± 1,7
SELEÇÃO_30%	67,3 ± 2,1	68,2 ± 1,4	68 ± 1,8	68,4 ± 1,1	68,1 ± 1,5	68,3 ± 1,4	68,3 ± 1,5	68,5 ± 1,4	69 ± 1,4
SELEÇÃO_50%	68,2 ± 1,2	68,3 ± 1,4	68,7 ± 1,2	68,4 ± 1,3	68,7 ± 1,1	68,6 ± 1	68,7 ± 1,1	68,8 ± 1,1	68,8 ± 0,9
SELEÇÃO_70%	68,3 ± 1,6	68,5 ± 1,4	68,5 ± 1,4	68,5 ± 1	68,5 ± 0,9	68,9 ± 1,1	68,7 ± 1,2	68,8 ± 0,9	69,1 ± 0,9

A.2 – Imagem sintética 2DP

Tabela A.8 – Acurácia x C x Tamanho da amostra – SVM completo

ACURÁCIA	200	300	400	500	600	700	800	900	1000
0,125	84,7 ± 2	86,2 ± 0,6	86,2 ± 0,5	86,2 ± 0,6	86,4 ± 0,4	86,3 ± 0,4	86,4 ± 0,3	86,4 ± 0,3	86,5 ± 0,4
1	86,1 ± 0,4	86,2 ± 0,4	86,4 ± 0,5	86,3 ± 0,4	86,5 ± 0,4	86,5 ± 0,3	86,6 ± 0,4	86,4 ± 0,4	86,6 ± 0,4
8	86,1 ± 0,4	86,2 ± 0,4	86,4 ± 0,4	86,5 ± 0,3	86,5 ± 0,5	86,5 ± 0,4	86,5 ± 0,4	86,6 ± 0,5	86,6 ± 0,4
64	86,2 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,4 ± 0,4	86,4 ± 0,4	86,5 ± 0,4	86,5 ± 0,3	86,5 ± 0,4	86,7 ± 0,4
512	86,1 ± 0,4	86,1 ± 0,4	86,5 ± 0,3	86,3 ± 0,5	86,5 ± 0,3	86,4 ± 0,3	86,5 ± 0,3	86,7 ± 0,3	86,6 ± 0,4
4096	86,1 ± 0,5	86,4 ± 0,5	86,3 ± 0,4	86,3 ± 0,4	86,3 ± 0,3	86,5 ± 0,3	86,5 ± 0,3	86,4 ± 0,4	86,7 ± 0,3

Tabela A.9 – Acurácia vs. Tamanho da amostra – $C = 0,125$

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	84,7 ± 2	86,2 ± 0,6	86,2 ± 0,5	86,2 ± 0,6	86,4 ± 0,4	86,3 ± 0,4	86,4 ± 0,3	86,4 ± 0,3	86,5 ± 0,4
RENN	80,6 ± 7,5	82,6 ± 4,8	85,5 ± 1,1	85,7 ± 0,7	86,2 ± 0,4	85,9 ± 0,5	86,3 ± 0,4	86,3 ± 0,3	86,3 ± 0,3
ME	81,9 ± 5,2	84,5 ± 3,1	86 ± 0,7	85,9 ± 0,6	86,3 ± 0,4	86,2 ± 0,4	86,2 ± 0,3	86,4 ± 0,4	86,4 ± 0,4
RENN+CNN	64,3 ± 15,2	66,8 ± 16,2	65,6 ± 16,1	65 ± 16,4	66,1 ± 16,7	65,1 ± 16,6	63 ± 16,2	60,7 ± 15,6	69,7 ± 16,5
ME+CNN	64,1 ± 16,5	64,1 ± 16,4	61,9 ± 15,9	68,6 ± 15,8	62,4 ± 15,6	59,4 ± 14,7	67,6 ± 16,9	59,8 ± 14,3	66,5 ± 16,9
CNN	67,3 ± 16,6	63,3 ± 16,6	74,4 ± 14	67,9 ± 16,3	75,4 ± 14,5	66,6 ± 17,2	71,6 ± 16,8	77,3 ± 12,8	74 ± 13,8
LÍDER_C_10	64,7 ± 16,1	67,2 ± 16,5	77,7 ± 12,9	74,3 ± 15	67 ± 16,3	73,4 ± 14,6	72,7 ± 15,4	70,8 ± 16,2	67,2 ± 16,2
LÍDER_C_30	73,2 ± 15,6	71,3 ± 15,8	78,4 ± 11,4	75 ± 13,7	80,2 ± 7,6	81,5 ± 5,1	84,3 ± 2,5	84 ± 2,9	85 ± 1,7
LÍDER_C_50	70,5 ± 16,1	71,9 ± 15,5	78,6 ± 9,7	79,1 ± 7,7	83,4 ± 3,6	85,1 ± 1,3	85,4 ± 1	85,4 ± 1,3	85,6 ± 0,7
LÍDER_C_70	74,7 ± 14,1	80,7 ± 8,4	79,7 ± 6,7	83 ± 3,4	85,5 ± 1	86 ± 0,8	85,9 ± 0,7	85,9 ± 0,7	86,2 ± 0,5
LÍDER_H_10	67,6 ± 16	68,6 ± 15,2	73,2 ± 13,5	74,5 ± 11,2	75,4 ± 11	76,5 ± 11,1	81,7 ± 4,9	83,6 ± 4,4	84,1 ± 3,3
LÍDER_H_30	61,2 ± 15,3	64,5 ± 16,9	67,6 ± 16,8	71,3 ± 15,5	73,9 ± 14,9	67,6 ± 16,8	67,4 ± 16,4	69,6 ± 15,1	75,5 ± 13,9
LÍDER_H_50	63,8 ± 16,2	69,4 ± 16,1	70,9 ± 16,3	67,6 ± 16,9	60,5 ± 15,4	70,4 ± 16,1	69,6 ± 16,4	64,8 ± 17	71,3 ± 16,6
LÍDER_H_70	59,8 ± 15,3	60,1 ± 14,7	64,2 ± 16,6	67 ± 16,6	65,7 ± 17,1	68,8 ± 16,9	66,7 ± 17	61,4 ± 15,4	69,6 ± 16,4
LÍDER_HC_10	67,2 ± 14,9	68,4 ± 15,1	72,4 ± 12,6	73,2 ± 10,1	79,4 ± 7,2	82,5 ± 4,8	84,1 ± 2,7	84,4 ± 2,3	85,4 ± 1,4
LÍDER_HC_30	70,4 ± 15,1	71,6 ± 14	75,3 ± 11,2	78,4 ± 9,4	81,9 ± 5,5	84,2 ± 3	84,1 ± 3,1	84,2 ± 2,2	85,7 ± 1,1
LÍDER_HC_50	67,6 ± 15,9	74,5 ± 13,5	78,4 ± 8,6	83 ± 3,7	84 ± 3	85,4 ± 1,2	85,7 ± 0,8	85,7 ± 1	85,9 ± 0,9
LÍDER_HC_70	75,9 ± 13,7	75,6 ± 12,7	80,2 ± 7,3	84,1 ± 2,9	85,3 ± 1,6	86 ± 0,5	85,9 ± 0,8	86 ± 0,5	85,9 ± 0,7
SELEÇÃO_10%	81,3 ± 3,7	83,1 ± 2,6	83,6 ± 1,8	83 ± 2,1	83,6 ± 2,1	84,3 ± 1,6	84,2 ± 1,8	84,6 ± 1,9	84,6 ± 2,1
SELEÇÃO_30%	84 ± 2,4	84,3 ± 1,6	85 ± 1,1	84,8 ± 1,5	85,4 ± 1,1	84,8 ± 1,2	85,5 ± 0,8	85,9 ± 0,5	85,8 ± 0,9
SELEÇÃO_50%	84,8 ± 1,5	84,5 ± 1,3	84,9 ± 1,3	85,9 ± 0,6	86,1 ± 0,6	86,2 ± 0,7	86,1 ± 0,5	86,2 ± 0,4	86,1 ± 0,4
SELEÇÃO_70%	84,6 ± 1,4	85,2 ± 1,2	85,9 ± 0,7	86,1 ± 0,5	86,2 ± 0,3	86,1 ± 0,5	86,2 ± 0,4	86,1 ± 0,4	86,2 ± 0,3

Tabela A.10 – Acurácia vs. Tamanho da amostra – $C = 1$

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	86,1 ± 0,4	86,2 ± 0,4	86,4 ± 0,5	86,3 ± 0,4	86,5 ± 0,4	86,5 ± 0,3	86,6 ± 0,4	86,4 ± 0,4	86,6 ± 0,4
RENN	85,7 ± 0,7	86,1 ± 0,4	86,2 ± 0,4	86,4 ± 0,4	86,4 ± 0,4	86,3 ± 0,4	86,4 ± 0,5	86,4 ± 0,3	86,4 ± 0,3
ME	85,9 ± 0,5	86,2 ± 0,4	86,1 ± 0,5	86,3 ± 0,4	86,2 ± 0,4	86,3 ± 0,3	86,4 ± 0,4	86,3 ± 0,4	86,4 ± 0,3
RENN+CNN	64,7 ± 16,2	69,4 ± 16,3	65,3 ± 16,8	58,6 ± 14,6	70,5 ± 15,9	65,6 ± 17	58,9 ± 14,4	60,1 ± 14,5	63,3 ± 16,4
ME+CNN	65,5 ± 16,1	66 ± 16,4	64 ± 16,4	63,7 ± 16,1	63,9 ± 16	65,5 ± 16,5	64,6 ± 15,6	65,6 ± 16	69,9 ± 15,5
CNN	77,1 ± 10	81,5 ± 7,3	84,8 ± 1,8	85,9 ± 0,4	85,8 ± 0,7	86 ± 0,5	86,1 ± 0,5	86 ± 0,4	86,2 ± 0,3
LÍDER_C_10	67,9 ± 14,6	80,1 ± 8	83,2 ± 3	84,2 ± 1,9	85 ± 1	85,4 ± 1,1	85,3 ± 0,9	85,3 ± 1,2	85,9 ± 0,9
LÍDER_C_30	84,6 ± 1,9	85,6 ± 1	85,7 ± 0,9	85,7 ± 0,7	86,1 ± 0,5	86 ± 0,7	86,2 ± 0,4	86,1 ± 0,4	86,1 ± 0,5
LÍDER_C_50	85,4 ± 1	85,8 ± 0,8	86 ± 0,7	86,2 ± 0,5	86,3 ± 0,4	86,1 ± 0,4	86,1 ± 0,3	86,2 ± 0,3	86,3 ± 0,4
LÍDER_C_70	85,9 ± 0,8	85,8 ± 0,8	86,2 ± 0,5	86,2 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,3 ± 0,3	86,2 ± 0,4	86,3 ± 0,3
LÍDER_H_10	84,6 ± 2,2	85,8 ± 0,8	85,9 ± 0,6	86,1 ± 0,5	86,3 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,3 ± 0,3	86,4 ± 0,4
LÍDER_H_30	76,3 ± 9,7	80,6 ± 6,4	83,9 ± 3,1	85,3 ± 0,9	85,9 ± 0,6	85,8 ± 0,7	85,9 ± 0,6	86 ± 0,4	86,1 ± 0,3
LÍDER_H_50	66,1 ± 15,2	75,1 ± 10,7	74,5 ± 11,5	82,9 ± 3,6	84,9 ± 1,5	85,4 ± 1	85,8 ± 0,5	85,8 ± 0,6	86,1 ± 0,4
LÍDER_H_70	67,4 ± 15,2	65,3 ± 14	77,3 ± 9,4	78,8 ± 7,7	81,7 ± 5,6	83,9 ± 2,4	84,9 ± 1,7	85,5 ± 1,1	85,7 ± 0,7
LÍDER_HC_10	85,7 ± 0,8	85,8 ± 0,7	86,1 ± 0,5	86,3 ± 0,4	86,3 ± 0,3	86,4 ± 0,4	86,4 ± 0,3	86,3 ± 0,4	86,5 ± 0,4
LÍDER_HC_30	85,2 ± 1,1	85,8 ± 0,6	86 ± 0,5	86,3 ± 0,4	86,3 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,2 ± 0,4	86,3 ± 0,4
LÍDER_HC_50	85,8 ± 1	85,9 ± 0,5	86,1 ± 0,4	86,1 ± 0,3	86,2 ± 0,5	86,3 ± 0,4	86,3 ± 0,4	86,2 ± 0,4	86,4 ± 0,3
LÍDER_HC_70	85,7 ± 0,9	86 ± 0,7	86 ± 0,5	86,2 ± 0,4	86,2 ± 0,3	86,3 ± 0,4	86,4 ± 0,4	86,3 ± 0,4	86,4 ± 0,5
SELEÇÃO_10%	81,4 ± 3,2	83,6 ± 2,1	84 ± 2	85 ± 1	85,2 ± 1,1	85,2 ± 0,9	85,6 ± 1	85,8 ± 0,7	85,7 ± 0,7
SELEÇÃO_30%	85,2 ± 1	85,6 ± 0,7	85,9 ± 0,4	86,1 ± 0,5	86 ± 0,5	86 ± 0,4	86,2 ± 0,3	86,2 ± 0,3	86,3 ± 0,4
SELEÇÃO_50%	85,8 ± 0,6	86 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,2 ± 0,6	86,2 ± 0,4	86,2 ± 0,5	86,3 ± 0,4	86,6 ± 0,4
SELEÇÃO_70%	86 ± 0,6	86 ± 0,3	86,3 ± 0,6	86,3 ± 0,5	86,2 ± 0,5	86,3 ± 0,3	86,4 ± 0,3	86,4 ± 0,3	86,4 ± 0,3

Tabela A.11 – Acurácia vs. Tamanho da amostra – $C = 8$

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	86,1 ± 0,4	86,2 ± 0,4	86,4 ± 0,4	86,5 ± 0,3	86,5 ± 0,5	86,5 ± 0,4	86,5 ± 0,4	86,6 ± 0,5	86,6 ± 0,4
RENN	86 ± 0,6	86 ± 0,6	86,1 ± 0,5	86,2 ± 0,4	86,4 ± 0,3	86,3 ± 0,4	86,3 ± 0,3	86,3 ± 0,3	86,3 ± 0,4
ME	85,9 ± 0,6	85,9 ± 0,6	86,1 ± 0,4	86,1 ± 0,4	86,2 ± 0,3	86,3 ± 0,4	86,2 ± 0,4	86,3 ± 0,5	86,3 ± 0,3
RENN+CNN	82,7 ± 3,7	84,4 ± 1,5	84,5 ± 2	84,8 ± 1,3	85,3 ± 1,2	85,5 ± 0,8	85,3 ± 0,7	85,6 ± 0,8	85,8 ± 0,8
ME+CNN	83,6 ± 2,2	84,5 ± 1,6	84,3 ± 2,2	85,4 ± 1,2	85,4 ± 0,9	85,6 ± 0,6	85,7 ± 0,7	85,7 ± 0,7	85,9 ± 0,8
CNN	85,4 ± 0,7	85,9 ± 0,5	86,3 ± 0,4	86,2 ± 0,3	86,2 ± 0,4	86,3 ± 0,5	86,1 ± 0,4	86,3 ± 0,4	86,4 ± 0,3
LÍDER_C_10	84,2 ± 1,4	85 ± 1,3	84,8 ± 1,2	85,4 ± 0,8	85,5 ± 0,8	85,8 ± 0,6	86 ± 0,6	85,7 ± 0,8	85,7 ± 0,8
LÍDER_C_30	85,6 ± 0,9	85,9 ± 0,6	86 ± 0,4	86 ± 0,6	86,2 ± 0,6	86,2 ± 0,4	86,3 ± 0,4	86,1 ± 0,6	86,2 ± 0,4
LÍDER_C_50	85,8 ± 0,6	85,9 ± 0,5	86,1 ± 0,4	86,3 ± 0,4	86,2 ± 0,3	86,2 ± 0,4	86,3 ± 0,4	86,3 ± 0,4	86,3 ± 0,4
LÍDER_C_70	85,9 ± 0,6	85,9 ± 0,7	86,1 ± 0,5	86,1 ± 0,4	86,4 ± 0,4	86,3 ± 0,4	86,2 ± 0,3	86,4 ± 0,3	86,3 ± 0,3
LÍDER_H_10	86 ± 0,5	86,1 ± 0,5	86,4 ± 0,4	86,5 ± 0,4	86,3 ± 0,5	86,4 ± 0,4	86,4 ± 0,3	86,6 ± 0,4	86,5 ± 0,3
LÍDER_H_30	85,4 ± 0,8	85,7 ± 0,8	85,9 ± 0,7	86,2 ± 0,4	86,1 ± 0,3	86,3 ± 0,4	86 ± 0,5	86,2 ± 0,4	86,3 ± 0,4
LÍDER_H_50	84,5 ± 1,9	85,3 ± 1	85,9 ± 0,6	86 ± 0,5	86,1 ± 0,4	86,2 ± 0,4	86,1 ± 0,4	86,2 ± 0,5	86,2 ± 0,4
LÍDER_H_70	84,4 ± 1,6	84,9 ± 0,8	85,4 ± 1,1	86,1 ± 0,6	85,7 ± 0,6	86 ± 0,6	85,8 ± 0,7	86,1 ± 0,6	86,1 ± 0,3
LÍDER_HC_10	86 ± 0,5	86,2 ± 0,4	86,3 ± 0,2	86,4 ± 0,3	86,4 ± 0,4	86,5 ± 0,4	86,4 ± 0,4	86,6 ± 0,4	86,5 ± 0,4
LÍDER_HC_30	86 ± 0,8	85,9 ± 0,6	86,2 ± 0,3	86,3 ± 0,4	86,4 ± 0,3	86,3 ± 0,5	86,3 ± 0,4	86,5 ± 0,4	86,5 ± 0,4
LÍDER_HC_50	85,9 ± 0,6	86 ± 0,6	86,2 ± 0,4	86,4 ± 0,4	86,3 ± 0,4	86,4 ± 0,4	86,3 ± 0,3	86,4 ± 0,3	86,5 ± 0,4
LÍDER_HC_70	86,1 ± 0,5	86 ± 0,5	86,2 ± 0,4	86,3 ± 0,4	86,4 ± 0,4	86,2 ± 0,3	86,4 ± 0,4	86,5 ± 0,5	86,5 ± 0,3
SELEÇÃO_10%	84,2 ± 2	84,6 ± 1,3	85 ± 1,4	85,4 ± 0,9	85,1 ± 0,9	85,5 ± 0,7	85,9 ± 0,6	85,8 ± 0,5	85,8 ± 0,7
SELEÇÃO_30%	85,3 ± 0,9	85,8 ± 0,7	85,9 ± 0,7	86,2 ± 0,5	86 ± 0,5	86,2 ± 0,5	86,3 ± 0,4	86,3 ± 0,5	86,3 ± 0,4
SELEÇÃO_50%	85,9 ± 0,6	86,1 ± 0,6	86 ± 0,5	86,2 ± 0,3	86,2 ± 0,4	86,4 ± 0,5	86,4 ± 0,5	86,3 ± 0,4	86,4 ± 0,4
SELEÇÃO_70%	86 ± 0,5	86,3 ± 0,5	86,3 ± 0,5	86,3 ± 0,4	86,3 ± 0,4	86,5 ± 0,6	86,3 ± 0,3	86,5 ± 0,3	86,5 ± 0,4

Tabela A.12 – Acurácia vs. Tamanho da amostra – C = 64

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	86,2 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,4 ± 0,4	86,4 ± 0,4	86,5 ± 0,4	86,5 ± 0,3	86,5 ± 0,4	86,7 ± 0,4
RENN	86 ± 0,5	86 ± 0,6	86,2 ± 0,5	86,1 ± 0,5	86,1 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,3 ± 0,5	86,2 ± 0,4
ME	86 ± 0,6	86 ± 0,6	86,2 ± 0,6	86 ± 0,5	86,1 ± 0,4	86,4 ± 0,4	86,3 ± 0,4	86,2 ± 0,4	86,2 ± 0,3
RENN+CNN	85,4 ± 1	85,7 ± 0,8	85,9 ± 0,7	86 ± 0,4	86 ± 0,6	86,1 ± 0,4	86,2 ± 0,3	86,1 ± 0,4	86,2 ± 0,4
ME+CNN	85,2 ± 0,9	85,5 ± 0,9	86 ± 0,8	85,8 ± 0,6	86 ± 0,5	86,1 ± 0,4	86,3 ± 0,4	86,1 ± 0,4	86,1 ± 0,4
CNN	85,9 ± 0,5	86,2 ± 0,5	86,3 ± 0,3	86,2 ± 0,3	86,4 ± 0,3	86,4 ± 0,4	86,4 ± 0,3	86,2 ± 0,4	86,4 ± 0,4
LÍDER_C_10	84,6 ± 1,1	84,9 ± 1,2	85,2 ± 0,7	85,1 ± 1	85,7 ± 0,7	85,6 ± 0,8	86 ± 0,8	85,7 ± 0,8	86,1 ± 0,5
LÍDER_C_30	85,6 ± 0,8	85,7 ± 0,8	86 ± 0,6	86,1 ± 0,6	86,1 ± 0,5	86,3 ± 0,4	86,2 ± 0,4	86,2 ± 0,5	86,2 ± 0,3
LÍDER_C_50	85,8 ± 0,7	86 ± 0,5	85,9 ± 0,6	86,2 ± 0,5	86,2 ± 0,5	86,2 ± 0,4	86,3 ± 0,4	86,3 ± 0,3	86,2 ± 0,5
LÍDER_C_70	85,8 ± 0,6	86 ± 0,6	86,1 ± 0,5	86,2 ± 0,7	86,2 ± 0,4	86,2 ± 0,5	86,3 ± 0,4	86,4 ± 0,4	86,3 ± 0,4
LÍDER_H_10	86,2 ± 0,4	86 ± 0,6	86,2 ± 0,3	86,3 ± 0,4	86,4 ± 0,4	86,4 ± 0,3	86,4 ± 0,3	86,4 ± 0,4	86,5 ± 0,4
LÍDER_H_30	85,8 ± 0,7	85,7 ± 0,8	86,1 ± 0,4	86,1 ± 0,3	86,1 ± 0,4	86,3 ± 0,4	86,2 ± 0,2	86,3 ± 0,4	86,3 ± 0,4
LÍDER_H_50	85,5 ± 1	85,9 ± 0,6	85,9 ± 0,6	86,1 ± 0,6	86,1 ± 0,5	86,1 ± 0,4	86,1 ± 0,4	86,2 ± 0,3	86,2 ± 0,4
LÍDER_H_70	85,5 ± 0,9	85,9 ± 0,8	85,7 ± 0,7	85,9 ± 0,5	86,1 ± 0,5	86 ± 0,5	86,2 ± 0,6	86,2 ± 0,3	86,2 ± 0,5
LÍDER_HC_10	86,3 ± 0,5	86,1 ± 0,5	86,3 ± 0,4	86,3 ± 0,3	86,4 ± 0,4	86,3 ± 0,4	86,4 ± 0,4	86,5 ± 0,4	86,6 ± 0,5
LÍDER_HC_30	86,1 ± 0,4	86 ± 0,6	86,2 ± 0,4	86,3 ± 0,5	86,3 ± 0,4	86,3 ± 0,5	86,3 ± 0,4	86,4 ± 0,3	86,5 ± 0,4
LÍDER_HC_50	86 ± 0,5	86,1 ± 0,5	86,2 ± 0,4	86,1 ± 0,6	86,4 ± 0,4	86,3 ± 0,4	86,3 ± 0,4	86,3 ± 0,5	86,5 ± 0,4
LÍDER_HC_70	86,1 ± 0,6	86,2 ± 0,5	86,3 ± 0,3	86,2 ± 0,5	86,3 ± 0,4	86,4 ± 0,5	86,4 ± 0,4	86,4 ± 0,4	86,4 ± 0,3
SELEÇÃO_10%	84,2 ± 1,9	84,9 ± 1,2	85,3 ± 1	85,3 ± 0,9	85,6 ± 0,8	85,8 ± 0,7	85,8 ± 0,5	85,7 ± 0,8	86,1 ± 0,5
SELEÇÃO_30%	85,5 ± 0,9	85,7 ± 0,6	86 ± 0,7	86,2 ± 0,5	86,1 ± 0,5	86,1 ± 0,4	86,3 ± 0,4	86,2 ± 0,5	86,2 ± 0,4
SELEÇÃO_50%	86 ± 0,7	85,9 ± 0,5	86,1 ± 0,4	86,1 ± 0,4	86,2 ± 0,4	86,2 ± 0,4	86,4 ± 0,3	86,3 ± 0,3	86,4 ± 0,3
SELEÇÃO_70%	86 ± 0,6	86,1 ± 0,5	86,1 ± 0,3	86,3 ± 0,4	86,4 ± 0,4	86,4 ± 0,5	86,3 ± 0,3	86,4 ± 0,4	86,4 ± 0,2

Tabela A.13 – Acurácia vs. Tamanho da amostra – C = 512

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	86,1 ± 0,4	86,1 ± 0,4	86,5 ± 0,3	86,3 ± 0,5	86,5 ± 0,3	86,4 ± 0,3	86,5 ± 0,3	86,7 ± 0,3	86,6 ± 0,4
RENN	85,9 ± 0,7	85,8 ± 0,5	86,2 ± 0,4	86,1 ± 0,5	86,2 ± 0,4	86,3 ± 0,4	86,4 ± 0,4	86,3 ± 0,3	86,2 ± 0,3
ME	85,8 ± 0,6	85,9 ± 0,6	86 ± 0,5	86,2 ± 0,3	86,2 ± 0,3	86,2 ± 0,4	86,3 ± 0,3	86,2 ± 0,4	86,3 ± 0,4
RENN+CNN	85,5 ± 0,9	85,5 ± 0,9	86 ± 0,4	86 ± 0,5	86 ± 0,6	86,1 ± 0,3	86,3 ± 0,4	86,2 ± 0,4	86,2 ± 0,4
ME+CNN	85,3 ± 0,8	85,6 ± 0,8	85,8 ± 0,6	85,9 ± 0,4	85,8 ± 0,6	86,1 ± 0,5	86,2 ± 0,5	86,2 ± 0,4	86,3 ± 0,4
CNN	85,9 ± 0,5	86 ± 0,7	86,1 ± 0,5	86,2 ± 0,4	86,4 ± 0,4	86,3 ± 0,4	86,4 ± 0,5	86,4 ± 0,3	86,5 ± 0,3
LÍDER_C_10	83,8 ± 2	85,1 ± 1	85,2 ± 1,2	85,5 ± 0,7	85,4 ± 0,9	85,6 ± 0,7	85,7 ± 0,6	85,8 ± 0,8	86 ± 0,7
LÍDER_C_30	85,3 ± 1	85,8 ± 0,8	86 ± 0,5	86 ± 0,7	86 ± 0,6	86,2 ± 0,4	86,2 ± 0,5	86,3 ± 0,3	86,3 ± 0,3
LÍDER_C_50	85,8 ± 0,9	85,9 ± 0,6	86,2 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,2 ± 0,5	86,4 ± 0,4	86,5 ± 0,4	86,5 ± 0,3
LÍDER_C_70	85,9 ± 0,7	86 ± 0,5	86,3 ± 0,3	86,3 ± 0,4	86,2 ± 0,4	86,2 ± 0,4	86,3 ± 0,3	86,3 ± 0,3	86,4 ± 0,3
LÍDER_H_10	86,1 ± 0,5	86,1 ± 0,5	86,3 ± 0,4	86,2 ± 0,4	86,4 ± 0,3	86,4 ± 0,3	86,5 ± 0,3	86,6 ± 0,4	86,5 ± 0,3
LÍDER_H_30	85,8 ± 0,7	85,8 ± 0,7	86 ± 0,5	86,2 ± 0,4	86,3 ± 0,3	86,2 ± 0,4	86,3 ± 0,4	86,3 ± 0,4	86,3 ± 0,5
LÍDER_H_50	85,8 ± 0,6	85,7 ± 0,7	86 ± 0,7	86,1 ± 0,5	86,1 ± 0,4	86,1 ± 0,6	86,3 ± 0,4	86,3 ± 0,4	86,3 ± 0,3
LÍDER_H_70	85,4 ± 1,3	85,4 ± 0,8	85,9 ± 0,6	86 ± 0,5	86 ± 0,6	85,9 ± 0,6	86,2 ± 0,4	86,3 ± 0,3	86 ± 0,4
LÍDER_HC_10	86 ± 0,6	86 ± 0,4	86,3 ± 0,3	86,3 ± 0,4	86,5 ± 0,4	86,3 ± 0,4	86,6 ± 0,4	86,7 ± 0,3	86,6 ± 0,4
LÍDER_HC_30	86,1 ± 0,6	86 ± 0,5	86,2 ± 0,4	86,4 ± 0,4	86,4 ± 0,4	86,3 ± 0,3	86,4 ± 0,4	86,4 ± 0,3	86,4 ± 0,3
LÍDER_HC_50	86,1 ± 0,7	86,1 ± 0,6	86,3 ± 0,4	86,2 ± 0,5	86,4 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,4 ± 0,4	86,5 ± 0,3
LÍDER_HC_70	86 ± 0,6	86 ± 0,5	86,4 ± 0,4	86,2 ± 0,5	86,4 ± 0,4	86,4 ± 0,3	86,4 ± 0,4	86,5 ± 0,4	86,4 ± 0,4
SELEÇÃO_10%	84,2 ± 2	84,9 ± 1,2	85,1 ± 1,2	85,5 ± 0,8	85,1 ± 0,9	85,5 ± 0,9	85,8 ± 0,6	85,6 ± 0,7	85,9 ± 0,6
SELEÇÃO_30%	85,5 ± 1	85,6 ± 0,8	85,9 ± 0,6	86,2 ± 0,3	86,1 ± 0,6	86,2 ± 0,5	86,2 ± 0,3	86,1 ± 0,5	86,2 ± 0,4
SELEÇÃO_50%	85,8 ± 0,7	86,2 ± 0,5	86,1 ± 0,5	86,2 ± 0,4	86,1 ± 0,3	86,2 ± 0,3	86,3 ± 0,4	86,3 ± 0,5	86,4 ± 0,3
SELEÇÃO_70%	86 ± 0,4	86,2 ± 0,5	86,1 ± 0,4	86,1 ± 0,4	86,3 ± 0,3	86,4 ± 0,4	86,4 ± 0,3	86,5 ± 0,3	86,4 ± 0,4

Tabela A.14 – Acurácia vs. Tamanho da amostra – $C = 4096$

<i>ACURÁCIA</i>	200	300	400	500	600	700	800	900	1000
SVM	86,1 ± 0,5	86,4 ± 0,5	86,3 ± 0,4	86,3 ± 0,4	86,3 ± 0,3	86,5 ± 0,3	86,5 ± 0,3	86,4 ± 0,4	86,7 ± 0,3
RENN	85,8 ± 0,8	85,9 ± 0,7	86,1 ± 0,4	86,1 ± 0,4	86,2 ± 0,5	86,2 ± 0,3	86,3 ± 0,4	86,4 ± 0,3	86,4 ± 0,3
ME	85,7 ± 0,8	86 ± 0,5	86,2 ± 0,4	86,3 ± 0,4	86,2 ± 0,4	86,2 ± 0,5	86,1 ± 0,5	86,2 ± 0,3	86,4 ± 0,3
RENN+CNN	85,5 ± 1,1	85,5 ± 0,7	86,1 ± 0,5	85,9 ± 0,7	86,2 ± 0,5	86,1 ± 0,5	86,1 ± 0,5	86,2 ± 0,4	86,3 ± 0,4
ME+CNN	85,4 ± 0,7	85,6 ± 0,9	85,9 ± 0,6	85,9 ± 0,6	86,1 ± 0,5	85,8 ± 0,5	86 ± 0,4	86 ± 0,6	86,3 ± 0,4
CNN	86,1 ± 0,6	86,2 ± 0,6	86,3 ± 0,4	86,2 ± 0,4	86,2 ± 0,4	86,4 ± 0,3	86,3 ± 0,4	86,4 ± 0,4	86,4 ± 0,3
LÍDER_C_10	84,4 ± 1,7	85 ± 1,3	85,4 ± 0,9	85,5 ± 0,9	85,5 ± 0,8	85,6 ± 0,8	85,7 ± 0,8	85,7 ± 0,6	85,9 ± 0,6
LÍDER_C_30	85,3 ± 1	85,9 ± 0,6	86 ± 0,5	86,1 ± 0,4	86,1 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,3 ± 0,4	86,4 ± 0,4
LÍDER_C_50	86 ± 0,6	86,1 ± 0,5	86,2 ± 0,4	86,2 ± 0,4	86,3 ± 0,3	86,3 ± 0,4	86,3 ± 0,4	86,2 ± 0,4	86,3 ± 0,3
LÍDER_C_70	85,9 ± 0,5	86 ± 0,5	86,1 ± 0,5	86,3 ± 0,4	86,2 ± 0,4	86,4 ± 0,4	86,4 ± 0,4	86,4 ± 0,4	86,4 ± 0,3
LÍDER_H_10	86,1 ± 0,5	86,1 ± 0,4	86,2 ± 0,4	86,3 ± 0,4	86,3 ± 0,3	86,5 ± 0,3	86,5 ± 0,4	86,4 ± 0,4	86,5 ± 0,3
LÍDER_H_30	85,8 ± 0,6	86 ± 0,6	86,1 ± 0,6	86,1 ± 0,4	86,1 ± 0,4	86,2 ± 0,4	86,2 ± 0,4	86,4 ± 0,4	86,4 ± 0,4
LÍDER_H_50	85,7 ± 0,7	85,8 ± 0,5	86 ± 0,5	85,9 ± 0,5	86,2 ± 0,5	86,3 ± 0,4	86,2 ± 0,4	86,4 ± 0,5	86,3 ± 0,5
LÍDER_H_70	85,4 ± 0,9	85,7 ± 0,7	85,6 ± 0,7	86 ± 0,6	86,1 ± 0,5	86,2 ± 0,4	86,1 ± 0,5	86,1 ± 0,4	86,2 ± 0,5
LÍDER_HC_10	86,1 ± 0,6	86,2 ± 0,5	86,3 ± 0,4	86,3 ± 0,3	86,4 ± 0,3	86,6 ± 0,3	86,5 ± 0,4	86,3 ± 0,4	86,6 ± 0,4
LÍDER_HC_30	86,1 ± 0,7	86,3 ± 0,4	86,3 ± 0,4	86,2 ± 0,5	86,4 ± 0,4	86,4 ± 0,4	86,4 ± 0,4	86,4 ± 0,4	86,6 ± 0,4
LÍDER_HC_50	86 ± 0,6	86,4 ± 0,4	86,3 ± 0,4	86,3 ± 0,4	86,4 ± 0,4	86,4 ± 0,4	86,3 ± 0,4	86,3 ± 0,4	86,5 ± 0,4
LÍDER_HC_70	85,9 ± 0,7	86,2 ± 0,4	86,3 ± 0,3	86,3 ± 0,4	86,4 ± 0,4	86,3 ± 0,4	86,4 ± 0,4	86,5 ± 0,4	86,4 ± 0,4
SELEÇÃO_10%	83,9 ± 1,9	85 ± 1,1	85 ± 1,4	85,5 ± 0,9	85,3 ± 0,9	85,7 ± 0,8	85,8 ± 0,7	85,9 ± 0,7	85,8 ± 0,8
SELEÇÃO_30%	85,5 ± 0,9	85,8 ± 0,7	86 ± 0,7	86 ± 0,6	85,9 ± 0,5	86,2 ± 0,4	86,2 ± 0,3	86,2 ± 0,4	86,3 ± 0,4
SELEÇÃO_50%	85,9 ± 0,7	86,2 ± 0,4	86,1 ± 0,5	86,2 ± 0,5	86,1 ± 0,4	86,3 ± 0,4	86,3 ± 0,5	86,5 ± 0,4	86,3 ± 0,4
SELEÇÃO_70%	85,9 ± 0,5	86,2 ± 0,6	86,2 ± 0,6	86,3 ± 0,4	86,4 ± 0,3	86,4 ± 0,4	86,4 ± 0,3	86,3 ± 0,3	86,5 ± 0,5

A.2 – Imagem sintética 3DP

Tabela A.15 – Acurácia x C x Tamanho da amostra – SVM completo

ACURÁCIA	200	300	400	500	600	700	800	900	1000
0,125	94,4 ± 0,4	94,7 ± 0,4	94,7 ± 0,3	94,8 ± 0,2	94,7 ± 0,3	94,8 ± 0,3	94,7 ± 0,3	94,7 ± 0,2	94,7 ± 0,2
1	94,5 ± 0,3	94,6 ± 0,3	94,7 ± 0,3	94,7 ± 0,2	94,7 ± 0,3	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,2	94,8 ± 0,2
8	94,5 ± 0,5	94,6 ± 0,3	94,5 ± 0,4	94,7 ± 0,3	94,8 ± 0,3	94,8 ± 0,3	94,7 ± 0,3	94,9 ± 0,2	94,8 ± 0,2
64	94,4 ± 0,5	94,6 ± 0,3	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,6 ± 0,4	94,8 ± 0,3	94,8 ± 0,3	94,9 ± 0,3
512	94,3 ± 0,4	94,3 ± 0,4	94,6 ± 0,4	94,6 ± 0,3	94,6 ± 0,4	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,2
4096	94,4 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,6 ± 0,3	94,7 ± 0,2	94,8 ± 0,2	94,8 ± 0,2	94,8 ± 0,2	94,8 ± 0,2

Tabela A.16 – Acurácia vs. Tamanho da amostra – C = 0,125

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	94,4 ± 0,4	94,7 ± 0,4	94,7 ± 0,3	94,8 ± 0,2	94,7 ± 0,3	94,8 ± 0,3	94,7 ± 0,3	94,7 ± 0,2	94,7 ± 0,2
RENN	94,1 ± 0,7	94,6 ± 0,4	94,5 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,6 ± 0,3	94,7 ± 0,2
ME	94,3 ± 0,7	94,6 ± 0,4	94,7 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,2	94,6 ± 0,3	94,6 ± 0,3
RENN+CNN	68,8 ± 20,6	66 ± 20,1	70,8 ± 21,2	66,6 ± 20,7	71,9 ± 21	67,4 ± 20,4	67,8 ± 20,8	72 ± 20,9	65,1 ± 20,4
ME+CNN	63,8 ± 19,9	68,5 ± 20,3	71,4 ± 20,6	79,1 ± 19,5	70,4 ± 20,8	67,5 ± 20,3	73,6 ± 21	75,1 ± 20,9	74,1 ± 21,5
CNN	68,7 ± 20,6	72,2 ± 21,2	72 ± 21	70,6 ± 21,2	78 ± 20,4	79,2 ± 19,6	73,1 ± 20,7	78,1 ± 20,2	73,8 ± 21,1
LÍDER_C_10	75,8 ± 20,4	79,7 ± 18,7	83,6 ± 17,3	81 ± 19,1	82,3 ± 18,3	80,4 ± 18,8	80,6 ± 18,9	80,6 ± 18,6	89,7 ± 8,1
LÍDER_C_30	80,2 ± 18,7	84,5 ± 15,9	91,1 ± 3,9	93,7 ± 1,2	93,9 ± 0,9	94,3 ± 0,3	94,4 ± 0,6	94,6 ± 0,4	94,6 ± 0,4
LÍDER_C_50	83 ± 15	91,2 ± 5,1	93,7 ± 0,9	94,2 ± 0,6	94,4 ± 0,4	94,6 ± 0,4	94,7 ± 0,3	94,6 ± 0,3	94,7 ± 0,2
LÍDER_C_70	89,2 ± 6,7	92,6 ± 3	94,3 ± 0,4	94,3 ± 0,4	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3
LÍDER_H_10	64,7 ± 19,8	72,3 ± 21,2	72,9 ± 20,7	70,6 ± 20,1	69,7 ± 18,7	84,3 ± 13,2	75,8 ± 16,7	83,5 ± 14,5	87,5 ± 11,7
LÍDER_H_30	66,6 ± 20,8	68,9 ± 20,7	71,1 ± 21,5	71 ± 21,4	65,4 ± 20,6	66,8 ± 21	69,5 ± 21,3	75,4 ± 21,1	74 ± 21,4
LÍDER_H_50	68,1 ± 20,1	62,7 ± 19,7	65 ± 20,1	68,2 ± 21,2	69,6 ± 21,5	71,1 ± 21,5	67 ± 21,3	73,9 ± 21,3	72,8 ± 21,8
LÍDER_H_70	62,6 ± 19,6	65 ± 20,2	66,2 ± 20,4	65,3 ± 20,6	69,5 ± 21,3	71,5 ± 21,9	66,2 ± 20,3	69,7 ± 21,5	68,4 ± 21,4
LÍDER_HC_10	72,8 ± 19,3	75,3 ± 18,1	74,3 ± 17,8	76,2 ± 15,1	83,5 ± 11,2	85,9 ± 8,5	88,5 ± 5,8	90,7 ± 4,4	91,5 ± 3,6
LÍDER_HC_30	84,2 ± 15,7	78,5 ± 17,2	90,4 ± 4,6	92,7 ± 2	93,4 ± 1,2	94,1 ± 0,8	94,4 ± 0,5	94 ± 0,8	94,3 ± 0,6
LÍDER_HC_50	87,6 ± 10,7	92,2 ± 3,1	93,7 ± 1,1	94,3 ± 0,5	94,5 ± 0,3	94,6 ± 0,4	94,6 ± 0,4	94,4 ± 0,4	94,6 ± 0,4
LÍDER_HC_70	88 ± 10,9	92,6 ± 2,9	94,2 ± 0,5	94,5 ± 0,5	94,5 ± 0,3	94,6 ± 0,4	94,7 ± 0,2	94,6 ± 0,4	94,7 ± 0,4
SELEÇÃO_10%	91,7 ± 2,9	92,7 ± 2	92,9 ± 1,4	93,1 ± 1,4	93,2 ± 1,5	93,4 ± 1,2	93,5 ± 1	93,7 ± 0,9	93,7 ± 0,8
SELEÇÃO_30%	93,5 ± 1,1	93,8 ± 1	93,8 ± 1	94,2 ± 0,5	94,3 ± 0,5	94,4 ± 0,4	94,6 ± 0,4	94,7 ± 0,4	94,5 ± 0,3
SELEÇÃO_50%	93,5 ± 1,1	94 ± 1	94,5 ± 0,4	94,6 ± 0,4	94,5 ± 0,3	94,5 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3
SELEÇÃO_70%	94 ± 0,7	94,4 ± 0,4	94,5 ± 0,3	94,6 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,2

Tabela A.17 – Acurácia vs. Tamanho da amostra – $C = 1$

<i>ACURÁCIA</i>	200	300	400	500	600	700	800	900	1000
SVM	94,5 ± 0,3	94,6 ± 0,3	94,7 ± 0,3	94,7 ± 0,2	94,7 ± 0,3	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,2	94,8 ± 0,2
RENN	94,4 ± 0,4	94,6 ± 0,4	94,6 ± 0,3	94,7 ± 0,4	94,8 ± 0,3	94,8 ± 0,2	94,8 ± 0,3	94,7 ± 0,2	94,8 ± 0,3
ME	94,4 ± 0,5	94,5 ± 0,4	94,6 ± 0,3	94,6 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,8 ± 0,3	94,6 ± 0,3	94,9 ± 0,3
RENN+CNN	74,6 ± 20,6	76,8 ± 19,5	69,2 ± 21	71,7 ± 20,7	73,4 ± 20,9	60,9 ± 18,2	65,2 ± 19,7	71,6 ± 21,1	69,2 ± 20,5
ME+CNN	81,5 ± 17,9	70,8 ± 21,2	66,2 ± 20,2	71,8 ± 20,8	63,7 ± 19,8	67,1 ± 20,2	71 ± 20,4	69,6 ± 20,9	71,4 ± 20,1
CNN	81,9 ± 16,6	89 ± 6,3	90 ± 7,3	91,6 ± 3,3	93,3 ± 1,6	94 ± 0,7	93,7 ± 1,2	94,2 ± 0,4	94,3 ± 0,5
LÍDER_C_10	85,9 ± 10,5	93 ± 1,8	92,8 ± 2,8	93,9 ± 0,7	94,1 ± 0,5	94,2 ± 0,5	94,2 ± 0,5	94,2 ± 0,5	94,2 ± 0,4
LÍDER_C_30	94 ± 0,6	94,4 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,6 ± 0,3	94,5 ± 0,4	94,6 ± 0,4	94,5 ± 0,4	94,7 ± 0,3
LÍDER_C_50	94,3 ± 0,5	94,4 ± 0,4	94,4 ± 0,3	94,5 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,8 ± 0,3	94,7 ± 0,3	94,8 ± 0,3
LÍDER_C_70	94,5 ± 0,5	94,5 ± 0,3	94,5 ± 0,3	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,2
LÍDER_H_10	91,7 ± 3,2	93,2 ± 1,6	94,1 ± 0,7	94,4 ± 0,5	94,4 ± 0,5	94,4 ± 0,4	94,5 ± 0,4	94,5 ± 0,5	94,7 ± 0,3
LÍDER_H_30	66,6 ± 19,3	78 ± 15,9	84,4 ± 10,3	89,4 ± 9,9	87,9 ± 9,5	89,9 ± 6,6	92,7 ± 2,4	92,1 ± 3,7	93,1 ± 2,8
LÍDER_H_50	71,2 ± 21,2	69,7 ± 19,1	73,8 ± 18,7	75,2 ± 17,4	78,6 ± 17,4	86,9 ± 9,8	87,9 ± 6,8	88,8 ± 8	90,8 ± 5,4
LÍDER_H_70	65,8 ± 19,7	67,6 ± 20,2	66,1 ± 18,5	71,9 ± 19,4	75,8 ± 19,5	79,8 ± 15	84,2 ± 12,3	80 ± 13,2	87,8 ± 10
LÍDER_HC_10	92,9 ± 1,9	94,1 ± 0,5	94,3 ± 0,6	94,4 ± 0,5	94,5 ± 0,4	94,5 ± 0,3	94,6 ± 0,3	94,6 ± 0,3	94,5 ± 0,3
LÍDER_HC_30	93,9 ± 0,6	94,2 ± 0,4	94,3 ± 0,4	94,2 ± 0,5	94,5 ± 0,4	94,7 ± 0,4	94,8 ± 0,3	94,7 ± 0,3	94,7 ± 0,4
LÍDER_HC_50	94,3 ± 0,5	94,3 ± 0,4	94,4 ± 0,4	94,6 ± 0,4	94,8 ± 0,3	94,7 ± 0,4	94,7 ± 0,3	94,6 ± 0,4	94,7 ± 0,3
LÍDER_HC_70	94,4 ± 0,5	94,4 ± 0,4	94,5 ± 0,3	94,6 ± 0,4	94,8 ± 0,2	94,7 ± 0,3	94,8 ± 0,3	94,7 ± 0,3	94,7 ± 0,3
SELEÇÃO_10%	92,2 ± 2,1	93 ± 1,3	93,8 ± 1	94 ± 0,6	94,2 ± 0,5	94,3 ± 0,4	94,3 ± 0,4	94,4 ± 0,4	94,3 ± 0,5
SELEÇÃO_30%	94,2 ± 0,4	94,4 ± 0,5	94,4 ± 0,5	94,3 ± 0,5	94,4 ± 0,5	94,5 ± 0,4	94,6 ± 0,4	94,6 ± 0,3	94,5 ± 0,3
SELEÇÃO_50%	94,4 ± 0,4	94,4 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,7 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,6 ± 0,3
SELEÇÃO_70%	94,4 ± 0,3	94,5 ± 0,4	94,6 ± 0,3	94,6 ± 0,4	94,6 ± 0,3	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,2	94,8 ± 0,2

Tabela A.18 – Acurácia vs. Tamanho da amostra – $C = 8$

<i>ACURÁCIA</i>	200	300	400	500	600	700	800	900	1000
SVM	94,5 ± 0,5	94,6 ± 0,3	94,5 ± 0,4	94,7 ± 0,3	94,8 ± 0,3	94,8 ± 0,3	94,7 ± 0,3	94,9 ± 0,2	94,8 ± 0,2
RENN	94,4 ± 0,4	94,4 ± 0,4	94,4 ± 0,3	94,6 ± 0,4	94,7 ± 0,4	94,6 ± 0,3	94,6 ± 0,3	94,8 ± 0,3	94,8 ± 0,2
ME	94,4 ± 0,4	94,4 ± 0,4	94,4 ± 0,5	94,6 ± 0,4	94,7 ± 0,3	94,6 ± 0,3	94,7 ± 0,3	94,7 ± 0,4	94,8 ± 0,3
RENN+CNN	91,5 ± 4,9	93,2 ± 1,8	93,2 ± 1,4	93,4 ± 1,4	93,7 ± 0,9	94 ± 0,7	93,8 ± 0,8	94,1 ± 0,6	94,3 ± 0,4
ME+CNN	90,1 ± 5,5	93,3 ± 1,3	93,5 ± 1,4	93,6 ± 1,2	93,8 ± 0,9	94 ± 0,6	94 ± 0,7	94,2 ± 0,5	94,2 ± 0,6
CNN	94,2 ± 0,5	94,2 ± 0,4	94,4 ± 0,5	94,5 ± 0,4	94,4 ± 0,4	94,6 ± 0,3	94,6 ± 0,3	94,6 ± 0,4	94,7 ± 0,3
LÍDER_C_10	93,3 ± 1,1	93,9 ± 0,8	93,9 ± 0,8	94,1 ± 0,7	94,3 ± 0,4	94,3 ± 0,5	94,2 ± 0,4	94,5 ± 0,4	94,4 ± 0,4
LÍDER_C_30	94,1 ± 0,6	94,3 ± 0,5	94,3 ± 0,4	94,6 ± 0,4	94,5 ± 0,3	94,5 ± 0,4	94,6 ± 0,4	94,6 ± 0,3	94,6 ± 0,3
LÍDER_C_50	94,3 ± 0,4	94,4 ± 0,5	94,4 ± 0,4	94,5 ± 0,3	94,6 ± 0,3	94,5 ± 0,3	94,6 ± 0,3	94,5 ± 0,3	94,6 ± 0,3
LÍDER_C_70	94,4 ± 0,4	94,4 ± 0,4	94,5 ± 0,3	94,6 ± 0,4	94,6 ± 0,3	94,7 ± 0,3	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3
LÍDER_H_10	94,1 ± 0,5	94,4 ± 0,4	94,5 ± 0,4	94,6 ± 0,4	94,7 ± 0,4	94,7 ± 0,4	94,6 ± 0,4	94,7 ± 0,4	94,7 ± 0,3
LÍDER_H_30	92,9 ± 3	93,8 ± 2	94,2 ± 0,4	94,1 ± 0,6	94,2 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,4
LÍDER_H_50	92,4 ± 2,1	91,4 ± 6,8	93,7 ± 1,1	94 ± 1	94,1 ± 0,4	94,2 ± 0,6	94,2 ± 0,5	94,5 ± 0,4	94,4 ± 0,5
LÍDER_H_70	85,4 ± 13,5	91,7 ± 5,4	93,4 ± 1,3	93,7 ± 0,9	93,6 ± 1	93,9 ± 0,8	94 ± 0,7	94,4 ± 0,4	94,3 ± 0,4
LÍDER_HC_10	94,3 ± 0,4	94,4 ± 0,4	94,5 ± 0,5	94,6 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,8 ± 0,3
LÍDER_HC_30	94,2 ± 0,4	94,4 ± 0,4	94,4 ± 0,4	94,6 ± 0,4	94,5 ± 0,3	94,6 ± 0,3	94,6 ± 0,3	94,7 ± 0,2	94,7 ± 0,2
LÍDER_HC_50	94,5 ± 0,4	94,4 ± 0,5	94,6 ± 0,4	94,6 ± 0,3	94,6 ± 0,3	94,7 ± 0,2	94,6 ± 0,3	94,6 ± 0,3	94,7 ± 0,3
LÍDER_HC_70	94,4 ± 0,4	94,5 ± 0,3	94,4 ± 0,4	94,6 ± 0,4	94,7 ± 0,4	94,7 ± 0,3	94,6 ± 0,4	94,7 ± 0,4	94,8 ± 0,3
SELEÇÃO_10%	93,7 ± 0,9	93,9 ± 0,7	94 ± 0,7	94,2 ± 0,5	94,2 ± 0,3	94,2 ± 0,4	94,1 ± 0,5	94,3 ± 0,5	94,5 ± 0,5
SELEÇÃO_30%	94,2 ± 0,5	94,3 ± 0,4	94,3 ± 0,4	94,6 ± 0,4	94,6 ± 0,3	94,4 ± 0,3	94,5 ± 0,5	94,6 ± 0,3	94,5 ± 0,4
SELEÇÃO_50%	94,3 ± 0,5	94,4 ± 0,3	94,6 ± 0,4	94,6 ± 0,4	94,7 ± 0,3	94,6 ± 0,4	94,6 ± 0,3	94,7 ± 0,4	94,7 ± 0,3
SELEÇÃO_70%	94,6 ± 0,5	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,4	94,7 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,8 ± 0,2	94,8 ± 0,2

Tabela A.19 – Acurácia vs. Tamanho da amostra – $C = 64$

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	94,4 ± 0,5	94,6 ± 0,3	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,6 ± 0,4	94,8 ± 0,3	94,8 ± 0,3	94,9 ± 0,3
RENN	94,3 ± 0,5	94,3 ± 0,4	94,4 ± 0,3	94,4 ± 0,4	94,5 ± 0,3	94,5 ± 0,3	94,8 ± 0,4	94,7 ± 0,3	94,8 ± 0,3
ME	94,2 ± 0,5	94,3 ± 0,4	94,4 ± 0,4	94,4 ± 0,4	94,7 ± 0,4	94,5 ± 0,4	94,8 ± 0,3	94,6 ± 0,3	94,5 ± 0,3
RENN+CNN	94,1 ± 0,7	94,1 ± 0,6	94,3 ± 0,5	94,4 ± 0,4	94,4 ± 0,4	94,4 ± 0,3	94,5 ± 0,3	94,5 ± 0,4	94,4 ± 0,3
ME+CNN	94,1 ± 0,6	94,1 ± 0,4	94,1 ± 0,5	94,2 ± 0,6	94,3 ± 0,4	94,3 ± 0,5	94,4 ± 0,4	94,4 ± 0,3	94,5 ± 0,4
CNN	94,2 ± 0,5	94,4 ± 0,4	94,5 ± 0,3	94,5 ± 0,4	94,6 ± 0,4	94,6 ± 0,4	94,7 ± 0,3	94,6 ± 0,3	94,5 ± 0,4
LÍDER_C_10	93 ± 1,4	94 ± 0,8	94 ± 0,6	93,9 ± 0,6	94,1 ± 0,5	94,1 ± 0,5	94,2 ± 0,5	94,3 ± 0,5	94,1 ± 0,5
LÍDER_C_30	94,2 ± 0,5	94,3 ± 0,5	94,3 ± 0,4	94,3 ± 0,4	94,4 ± 0,5	94,4 ± 0,4	94,6 ± 0,4	94,5 ± 0,3	94,5 ± 0,4
LÍDER_C_50	94,3 ± 0,5	94,4 ± 0,4	94,5 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,4	94,6 ± 0,4	94,6 ± 0,4
LÍDER_C_70	94,2 ± 0,5	94,4 ± 0,4	94,4 ± 0,4	94,4 ± 0,4	94,6 ± 0,4	94,5 ± 0,4	94,5 ± 0,3	94,7 ± 0,3	94,5 ± 0,3
LÍDER_H_10	94,3 ± 0,5	94,5 ± 0,3	94,5 ± 0,4	94,5 ± 0,4	94,5 ± 0,3	94,6 ± 0,3	94,8 ± 0,2	94,7 ± 0,3	94,7 ± 0,3
LÍDER_H_30	94 ± 0,7	94,1 ± 0,5	94,1 ± 0,5	94,3 ± 0,4	94,3 ± 0,4	94,4 ± 0,4	94,6 ± 0,3	94,5 ± 0,3	94,5 ± 0,3
LÍDER_H_50	93,4 ± 1,3	93,9 ± 0,7	94,1 ± 0,5	94,3 ± 0,5	94,3 ± 0,4	94,4 ± 0,4	94,4 ± 0,5	94,4 ± 0,3	94,4 ± 0,4
LÍDER_H_70	93 ± 2,4	94 ± 0,6	94,2 ± 0,5	94,1 ± 0,6	94,2 ± 0,5	94,3 ± 0,4	94,2 ± 0,4	94,2 ± 0,5	94,4 ± 0,5
LÍDER_HC_10	94,3 ± 0,5	94,4 ± 0,4	94,5 ± 0,4	94,6 ± 0,3	94,5 ± 0,4	94,6 ± 0,3	94,7 ± 0,2	94,7 ± 0,3	94,8 ± 0,3
LÍDER_HC_30	94,3 ± 0,6	94,5 ± 0,3	94,5 ± 0,3	94,5 ± 0,4	94,5 ± 0,3	94,6 ± 0,3	94,6 ± 0,3	94,7 ± 0,3	94,7 ± 0,3
LÍDER_HC_50	94,3 ± 0,5	94,5 ± 0,4	94,5 ± 0,5	94,5 ± 0,4	94,7 ± 0,3	94,6 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,6 ± 0,3
LÍDER_HC_70	94,2 ± 0,5	94,4 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,6 ± 0,3	94,5 ± 0,3	94,6 ± 0,3	94,6 ± 0,3	94,6 ± 0,3
SELEÇÃO_10%	93,3 ± 1,4	93,8 ± 0,8	93,8 ± 0,7	93,9 ± 0,5	94,1 ± 0,5	94,1 ± 0,6	94,3 ± 0,4	94,3 ± 0,4	94,1 ± 0,5
SELEÇÃO_30%	93,9 ± 0,6	94,2 ± 0,4	94,3 ± 0,4	94,4 ± 0,4	94,3 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,5 ± 0,3	94,5 ± 0,4
SELEÇÃO_50%	94,3 ± 0,4	94,2 ± 0,4	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,3	94,5 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,2
SELEÇÃO_70%	94,3 ± 0,4	94,5 ± 0,4	94,6 ± 0,4	94,5 ± 0,4	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,8 ± 0,3

Tabela A.20 – Acurácia vs. Tamanho da amostra – $C = 512$

ACURÁCIA	200	300	400	500	600	700	800	900	1000
SVM	94,3 ± 0,4	94,3 ± 0,4	94,6 ± 0,4	94,6 ± 0,3	94,6 ± 0,4	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,3	94,8 ± 0,2
RENN	94,2 ± 0,5	94,3 ± 0,5	94,4 ± 0,5	94,4 ± 0,5	94,5 ± 0,4	94,4 ± 0,3	94,6 ± 0,4	94,6 ± 0,4	94,7 ± 0,3
ME	94,3 ± 0,5	94,1 ± 0,4	94,4 ± 0,3	94,3 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,7 ± 0,4
RENN+CNN	93,7 ± 0,9	94 ± 0,6	94 ± 0,5	94,3 ± 0,6	94,4 ± 0,5	94,4 ± 0,4	94,5 ± 0,4	94,4 ± 0,3	94,5 ± 0,3
ME+CNN	93,9 ± 0,6	93,8 ± 0,8	94 ± 0,7	94,1 ± 0,5	94,3 ± 0,5	94,3 ± 0,3	94,3 ± 0,3	94,4 ± 0,4	94,4 ± 0,4
CNN	94,2 ± 0,4	94,3 ± 0,4	94,4 ± 0,5	94,5 ± 0,3	94,5 ± 0,3	94,5 ± 0,4	94,5 ± 0,3	94,6 ± 0,4	94,7 ± 0,3
LÍDER_C_10	93 ± 1,4	93,3 ± 1,2	93,7 ± 0,8	93,8 ± 0,8	93,9 ± 0,6	93,9 ± 0,8	94,1 ± 0,4	94,1 ± 0,6	94,1 ± 0,5
LÍDER_C_30	94 ± 0,6	94,2 ± 0,4	94,2 ± 0,5	94,4 ± 0,5	94,4 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,3
LÍDER_C_50	94,3 ± 0,5	94,3 ± 0,4	94,3 ± 0,4	94,4 ± 0,4	94,4 ± 0,4	94,4 ± 0,3	94,4 ± 0,4	94,7 ± 0,4	94,6 ± 0,3
LÍDER_C_70	94,3 ± 0,5	94,3 ± 0,5	94,4 ± 0,5	94,2 ± 0,4	94,4 ± 0,3	94,5 ± 0,3	94,6 ± 0,3	94,5 ± 0,2	94,6 ± 0,3
LÍDER_H_10	94,2 ± 0,4	94,3 ± 0,4	94,5 ± 0,4	94,6 ± 0,4	94,5 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,4	94,7 ± 0,2
LÍDER_H_30	94,1 ± 0,6	94,1 ± 0,5	94,4 ± 0,4	94,5 ± 0,5	94,4 ± 0,4	94,4 ± 0,4	94,3 ± 0,4	94,6 ± 0,5	94,5 ± 0,4
LÍDER_H_50	93,8 ± 0,9	93,5 ± 1	94 ± 0,5	94,3 ± 0,4	94,3 ± 0,5	94,3 ± 0,4	94,4 ± 0,4	94,4 ± 0,5	94,4 ± 0,4
LÍDER_H_70	93,4 ± 1,2	93,5 ± 1,2	94,1 ± 0,4	94,1 ± 0,7	94,2 ± 0,5	94,2 ± 0,5	94,3 ± 0,4	94,4 ± 0,4	94,4 ± 0,4
LÍDER_HC_10	94,2 ± 0,3	94,4 ± 0,6	94,5 ± 0,4	94,6 ± 0,4	94,7 ± 0,3	94,5 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3
LÍDER_HC_30	94,2 ± 0,4	94,3 ± 0,4	94,5 ± 0,4	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,3	94,6 ± 0,3	94,7 ± 0,3	94,7 ± 0,3
LÍDER_HC_50	94,3 ± 0,5	94,3 ± 0,5	94,4 ± 0,4	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,4	94,6 ± 0,4	94,5 ± 0,3	94,6 ± 0,4
LÍDER_HC_70	94,3 ± 0,5	94,3 ± 0,4	94,5 ± 0,4	94,4 ± 0,4	94,6 ± 0,4	94,6 ± 0,3	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3
SELEÇÃO_10%	92,3 ± 2,3	93,4 ± 1	93,5 ± 0,7	93,8 ± 0,7	94 ± 0,7	94 ± 0,6	94,1 ± 0,5	94,2 ± 0,5	94,1 ± 0,5
SELEÇÃO_30%	93,9 ± 0,7	94,3 ± 0,6	94,3 ± 0,6	94,4 ± 0,4	94,3 ± 0,4	94,4 ± 0,4	94,4 ± 0,4	94,4 ± 0,5	94,4 ± 0,3
SELEÇÃO_50%	94,3 ± 0,4	94,1 ± 0,5	94,3 ± 0,5	94,3 ± 0,4	94,5 ± 0,4	94,7 ± 0,4	94,5 ± 0,4	94,5 ± 0,4	94,7 ± 0,3
SELEÇÃO_70%	94,4 ± 0,4	94,3 ± 0,4	94,5 ± 0,5	94,4 ± 0,4	94,5 ± 0,3	94,5 ± 0,3	94,7 ± 0,4	94,7 ± 0,3	94,7 ± 0,3

Tabela A.21 – Acurácia vs. Tamanho da amostra – $C = 4096$

<i>ACURÁCIA</i>	200	300	400	500	600	700	800	900	1000
SVM	94,4 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,6 ± 0,3	94,7 ± 0,2	94,8 ± 0,2	94,8 ± 0,2	94,8 ± 0,2	94,8 ± 0,2
RENN	94,2 ± 0,7	94,4 ± 0,6	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,4	94,6 ± 0,4	94,6 ± 0,4	94,6 ± 0,3	94,7 ± 0,4
ME	94,1 ± 0,6	94,2 ± 0,5	94,5 ± 0,5	94,5 ± 0,4	94,5 ± 0,4	94,5 ± 0,5	94,6 ± 0,4	94,6 ± 0,3	94,6 ± 0,4
RENN+CNN	93,7 ± 0,8	94,1 ± 0,8	94,4 ± 0,5	94,3 ± 0,4	94,3 ± 0,4	94,4 ± 0,4	94,7 ± 0,5	94,4 ± 0,4	94,5 ± 0,4
ME+CNN	93,7 ± 1,3	94 ± 0,5	94,2 ± 0,4	94,2 ± 0,5	94,3 ± 0,6	94,3 ± 0,3	94,6 ± 0,5	94,4 ± 0,4	94,4 ± 0,5
CNN	94,3 ± 0,4	94,3 ± 0,5	94,4 ± 0,3	94,4 ± 0,4	94,5 ± 0,4	94,6 ± 0,3	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3
LÍDER_C_10	92,9 ± 1,3	93,1 ± 1,4	93,3 ± 1,1	93,8 ± 0,9	94,1 ± 0,6	94,1 ± 0,5	94,1 ± 0,5	93,9 ± 0,5	94,1 ± 0,4
LÍDER_C_30	93,9 ± 0,6	94,3 ± 0,5	94,2 ± 0,4	94,3 ± 0,3	94,5 ± 0,3	94,5 ± 0,4	94,5 ± 0,3	94,5 ± 0,5	94,5 ± 0,4
LÍDER_C_50	94,2 ± 0,4	94,3 ± 0,5	94,3 ± 0,4	94,5 ± 0,5	94,5 ± 0,4	94,5 ± 0,3	94,5 ± 0,3	94,6 ± 0,4	94,6 ± 0,3
LÍDER_C_70	94,1 ± 0,5	94,3 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,5 ± 0,3	94,5 ± 0,4	94,6 ± 0,3	94,5 ± 0,3	94,6 ± 0,4
LÍDER_H_10	94,3 ± 0,5	94,2 ± 0,6	94,5 ± 0,3	94,6 ± 0,4	94,5 ± 0,4	94,6 ± 0,2	94,7 ± 0,3	94,7 ± 0,3	94,8 ± 0,3
LÍDER_H_30	94,1 ± 0,6	94,1 ± 0,6	94,3 ± 0,4	94,4 ± 0,5	94,5 ± 0,4	94,4 ± 0,5	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,3
LÍDER_H_50	93,8 ± 0,9	93,9 ± 0,8	94,1 ± 0,7	94,1 ± 0,4	94,3 ± 0,5	94,5 ± 0,3	94,3 ± 0,3	94,4 ± 0,4	94,5 ± 0,3
LÍDER_H_70	93,6 ± 0,9	93,9 ± 0,7	93,9 ± 0,9	94,1 ± 0,5	94,2 ± 0,7	94,2 ± 0,5	94,3 ± 0,6	94,3 ± 0,5	94,3 ± 0,3
LÍDER_HC_10	94,3 ± 0,4	94,4 ± 0,5	94,5 ± 0,4	94,5 ± 0,4	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,8 ± 0,2	94,8 ± 0,3
LÍDER_HC_30	94,2 ± 0,5	94,3 ± 0,4	94,5 ± 0,5	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,6 ± 0,4	94,7 ± 0,3	94,6 ± 0,2
LÍDER_HC_50	94,3 ± 0,4	94,3 ± 0,4	94,6 ± 0,4	94,6 ± 0,4	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,6 ± 0,3	94,7 ± 0,3
LÍDER_HC_70	94,3 ± 0,6	94,4 ± 0,4	94,5 ± 0,4	94,5 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,7 ± 0,3	94,8 ± 0,3	94,7 ± 0,3
SELEÇÃO_10%	91,4 ± 2,1	92,7 ± 1,4	93,4 ± 1	93,9 ± 0,7	93,8 ± 0,8	93,9 ± 0,7	94,3 ± 0,4	94 ± 0,6	94,3 ± 0,6
SELEÇÃO_30%	93,7 ± 0,8	94 ± 0,5	94,4 ± 0,4	94,3 ± 0,4	94,4 ± 0,4	94,4 ± 0,4	94,3 ± 0,4	94,4 ± 0,4	94,5 ± 0,4
SELEÇÃO_50%	94,1 ± 0,4	94,3 ± 0,6	94,3 ± 0,4	94,4 ± 0,4	94,5 ± 0,4	94,6 ± 0,3	94,6 ± 0,4	94,6 ± 0,3	94,6 ± 0,3
SELEÇÃO_70%	94,3 ± 0,5	94,2 ± 0,5	94,4 ± 0,4	94,6 ± 0,4	94,7 ± 0,3	94,6 ± 0,3	94,7 ± 0,3	94,6 ± 0,3	94,8 ± 0,3

APÊNDICE B – RESULTADOS DO TESTE DE WILCOXON Z CALCULADO

Neste apêndice são apresentados os valores de z calculado pelo Teste de Wilcoxon. Considerando 1000 amostras por classe para treinamento das amostras sintéticas e 500 amostras de treinamento para as imagens reais. Considerado 200 amostras por classe para teste em todos os casos.

Tabela B.1 – Valores de z calculados no Teste de Wilcoxon

#	1DP	2DP	3DP	IKONOS	PALSAR
RENN	1,229657	2,745873	0,133694	4,67930	4,78214
MULTIEDIT	1,206885	2,499053	0,339378	4,78214	4,78214
RENN_CONDENS	2,146129	2,725305	2,211096	4,65980	4,78214
ME_CONDENS	0,686126	2,025981	3,486334	4,72043	4,78214
CONDENS	1,855312	0,668471	0,174831	3,82571	4,46519
LÍDER_C_10	2,551538	3,054398	4,576456	4,78214	4,78214
LÍDER_C_30	2,897509	2,684168	2,540190	4,78214	4,78214
LÍDER_C_50	2,151900	0,154263	1,614615	4,78214	4,78214
LÍDER_C_70	2,058379	1,141543	1,779161	4,78214	4,78214
LÍDER_H_10	0,591952	0,380514	0,380514	-1,24104	0,18380
LÍDER_H_30	1,434600	1,306090	2,416780	-1,13857	-2,43261
LÍDER_H_50	1,693762	2,252233	2,828147	-0,42127	-0,55192
LÍDER_H_70	1,753825	4,206225	3,835995	0,01201	-1,03791
LÍDER_CH_10	0,182490	0,318809	0,750744	-3,53167	-1,76938
LÍDER_CH_30	2,594700	1,408931	0,071989	0,17779	-3,16540
LÍDER_CH_50	0,497778	1,120974	1,203248	1,67143	-1,93025
LÍDER_CH_70	0,592057	0,894723	1,367795	2,95157	-0,42127
SELEC_ALEAT_FX_10	1,764785	3,835995	4,473614	4,22734	4,78214
SELEC_ALEAT_FX_30	0,336336	3,157240	3,486334	4,70305	4,78214
SELEC_ALEAT_FX_50	2,139099	1,552910	0,359946	-1,39865	4,45305
SELEC_ALEAT_FX_70	0,713566	0,771313	0,627334	-1,73802	4,14440

APÊNDICE C – RESULTADOS DOS TEMPOS GLOBAIS

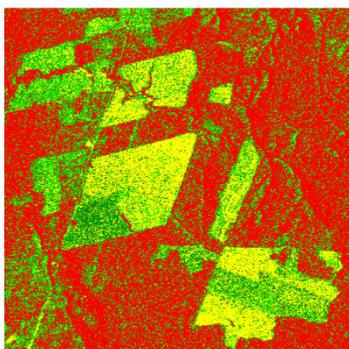
Neste apêndice são apresentados os tempos globais das imagens sintéticas, considerando $C=64$ para a imagem 1DP e $C=512$ para as imagens 2DP, 3DP, IKONOS e PALSAR.

Tabela C.1 – Tempo global dos métodos utilizados

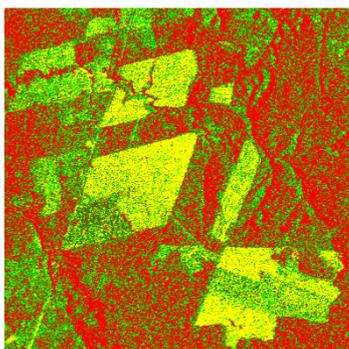
<i>Tempo Global (redução+treinamento)</i>	1DP	2DP	3DP	IKONOS	PALSAR
SVM	196,9 ± 24,8	200,5 ± 6	229,6 ± 6,1	223,7 ± 7,5	997,6 ± 26
RENN	17 ± 4,4	65,8 ± 6,6	149,9 ± 9,8	272,5 ± 27	39,8 ± 6,3
ME	25,8 ± 3,7	87,9 ± 5,2	164,9 ± 7,4	148,9 ± 7,4	62,8 ± 4,6
RENN+CNN	11 ± 1,8	13,4 ± 1,3	14,5 ± 1,9	190 ± 33,2	35,8 ± 2
ME+CNN	2,6 ± 0,3	3,5 ± 0,3	3,9 ± 0,4	33,8 ± 6,6	7,5 ± 0,1
CNN	38,1 ± 6,5	12,7 ± 1,3	2,6 ± 0,5	27,9 ± 0,8	290,3 ± 15,3
LÍDER_C_10	1,2 ± 0,2	1,2 ± 0	1,2 ± 0	2,1 ± 0,1	1,6 ± 0,7
LÍDER_C_30	4,8 ± 1,3	4,7 ± 0,8	7,1 ± 0,8	4,4 ± 0,1	2,9 ± 1,3
LÍDER_C_50	9,9 ± 1,9	10,4 ± 1,3	15,1 ± 1,6	6,1 ± 0,2	4,1 ± 2
LÍDER_C_70	15,6 ± 3,5	16,5 ± 2,2	22,3 ± 2,1	8,3 ± 0,3	5,1 ± 2,8
LÍDER_H_10	138,6 ± 17,4	70,2 ± 8,1	14,5 ± 3	50,7 ± 11,5	866,3 ± 30,4
LÍDER_H_30	76,6 ± 9,7	27,5 ± 3,8	4,1 ± 0,8	27 ± 3,8	774,3 ± 78,7
LÍDER_H_50	58,7 ± 8,3	17,9 ± 2,4	4,1 ± 0,3	21,3 ± 2,6	741,7 ± 76,9
LÍDER_H_70	46,8 ± 6,8	14,1 ± 2,3	4,5 ± 0,2	19,3 ± 1,2	738,4 ± 84,7
LÍDER_HC_10	148,1 ± 20	83,6 ± 7,9	20,5 ± 3,9	61,4 ± 8,6	859,2 ± 35,1
LÍDER_HC_30	110,9 ± 14,2	58,2 ± 5	21,7 ± 2,1	40,3 ± 0,9	820,7 ± 23,2
LÍDER_HC_50	100,5 ± 11,9	54,8 ± 4,5	28 ± 2,4	31,5 ± 2,5	822,1 ± 19,5
LÍDER_HC_70	92,9 ± 13,3	54 ± 4,3	32,4 ± 3,1	34,4 ± 1,4	802,1 ± 22,2
SELEÇÃO_10%	0,1 ± 0	0,1 ± 0	0,1 ± 0	0,8 ± 0,1	0,5 ± 0
SELEÇÃO_30%	3,8 ± 0,3	4,2 ± 0,3	4,7 ± 0,1	9,9 ± 1,2	18,7 ± 0,9
SELEÇÃO_50%	21,6 ± 0,8	25,6 ± 0,7	27 ± 1,6	44,2 ± 3,7	107,9 ± 3
SELEÇÃO_70%	62,5 ± 2,4	69,8 ± 6,4	70,1 ± 2,2	104,6 ± 1,6	299,4 ± 0,2

APÊNDICE D – CLASSIFICAÇÃO IMAGEM ALOS PALSAR

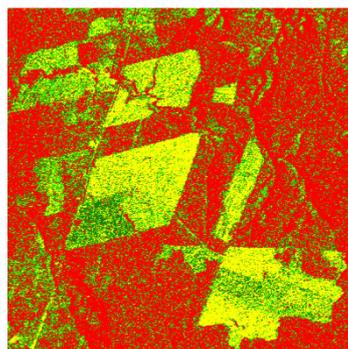
Neste apêndice são apresentadas as imagens classificadas para a imagem de radar ALOS PALSAR.



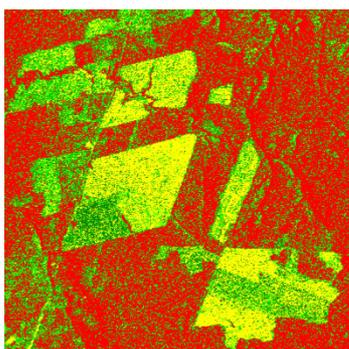
Sem Redução



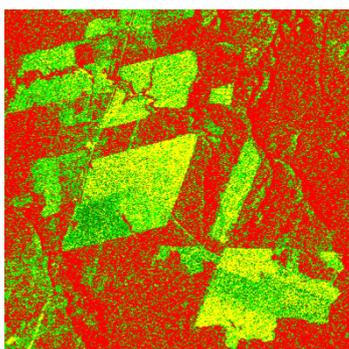
RENN



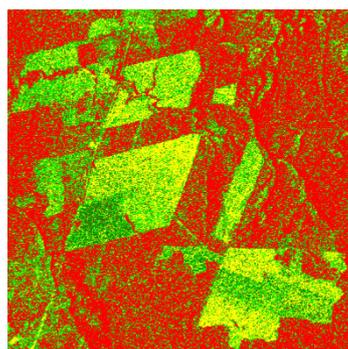
ME



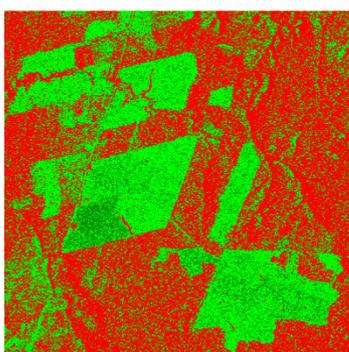
CNN



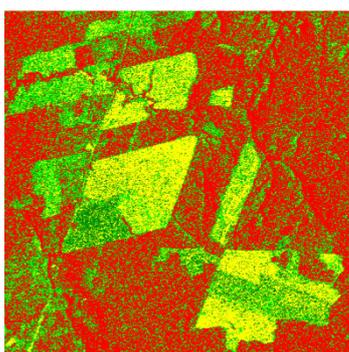
RENN+CNN



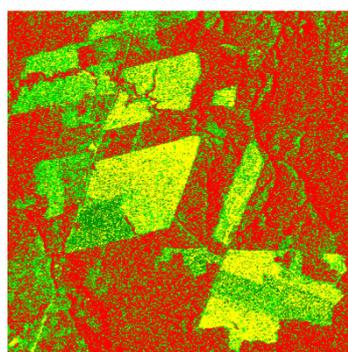
ME+CNN



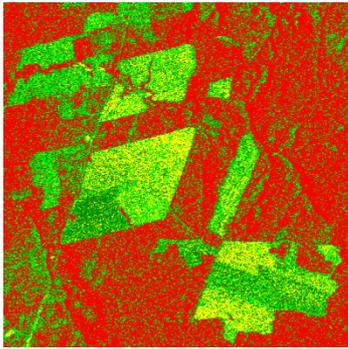
Líder Centros 10%



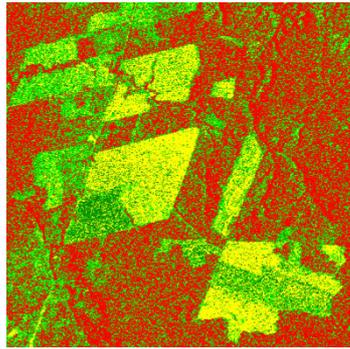
Líder H 10%



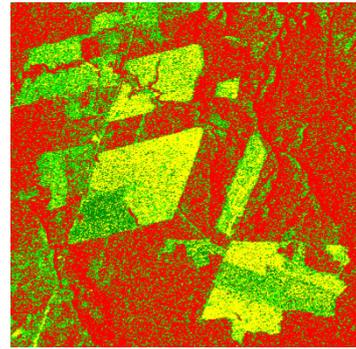
Líder HC 10%



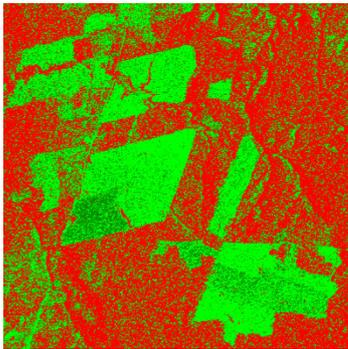
Líder Centros 30%



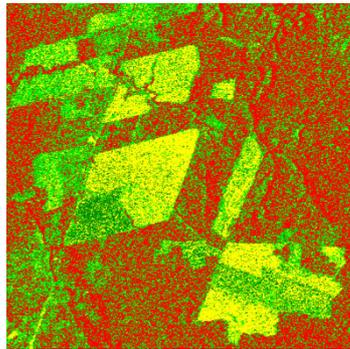
Líder H 30%



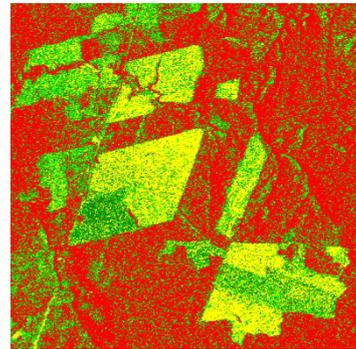
Líder HC 30%



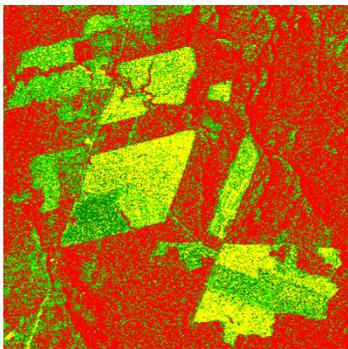
Líder Centros 50%



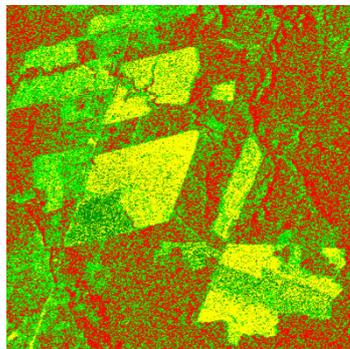
Líder H 50%



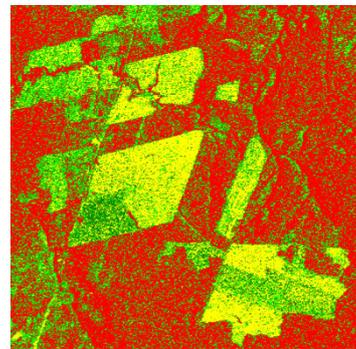
Líder HC 50%



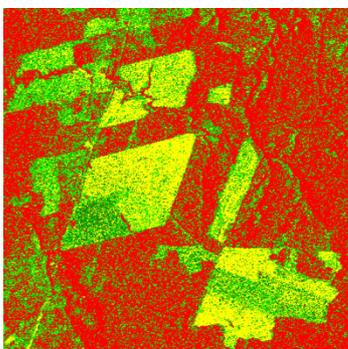
Líder Centros 70%



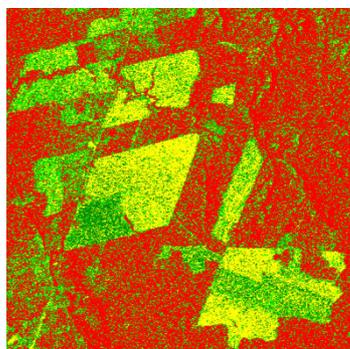
Líder H 70%



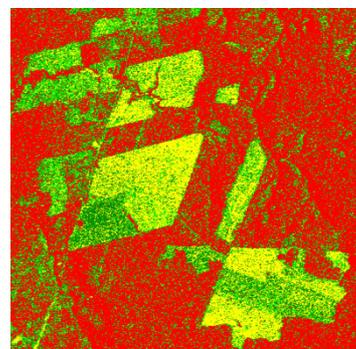
Líder HC 70%



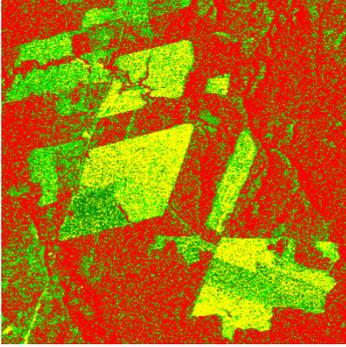
Seleção Aleatória 10%



Seleção Aleatória 30%



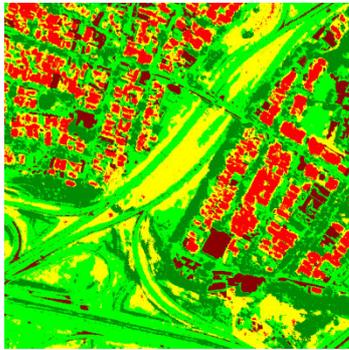
Seleção Aleatória 50%



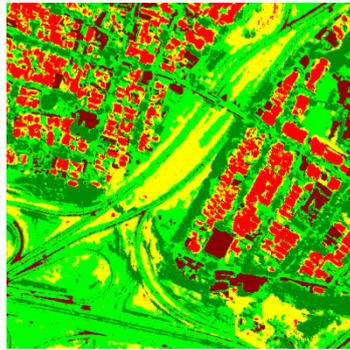
Seleção Aleatória 70%

APÊNDICE E – CLASSIFICAÇÃO IMAGEM IKONOS MS

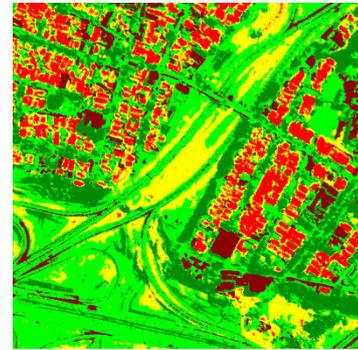
Neste apêndice são apresentadas as imagens classificadas para a imagem ótica IKONOS MS.



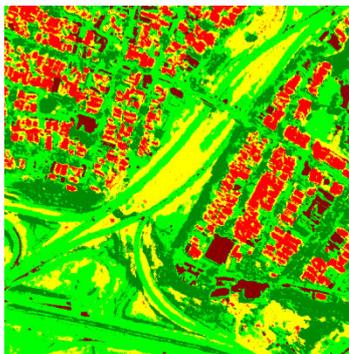
Sem Redução



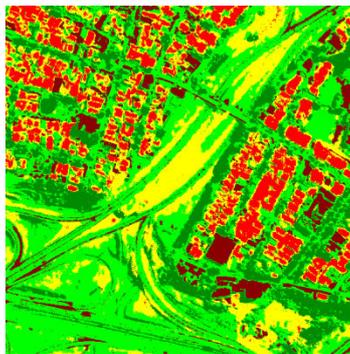
RENN



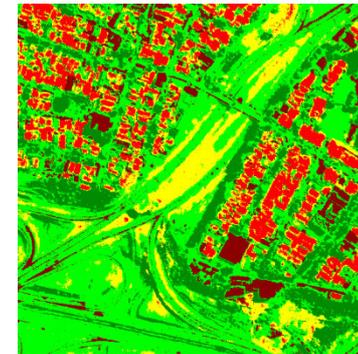
ME



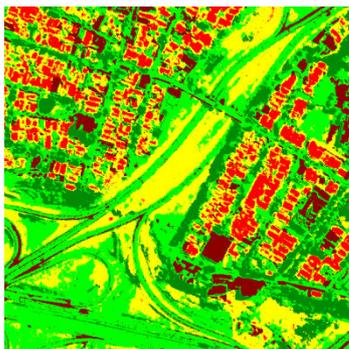
CNN



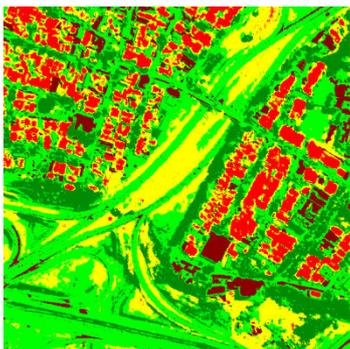
RENN+CNN



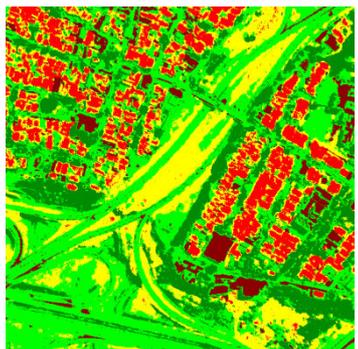
ME+CNN



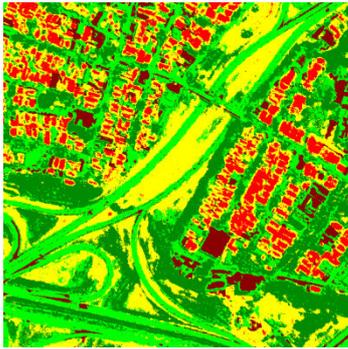
Líder Centros 10%



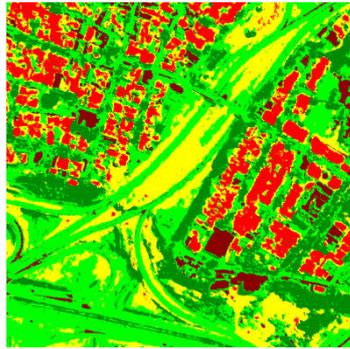
Líder H 10%



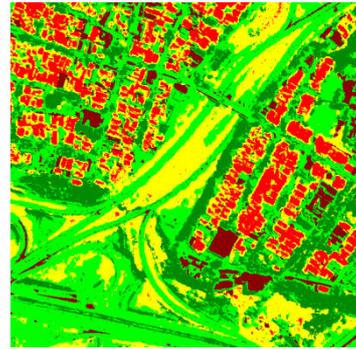
Líder HC 10%



Líder Centros 30%



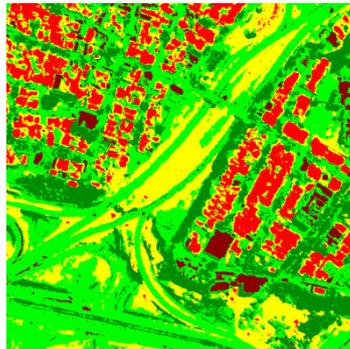
Líder H 30%



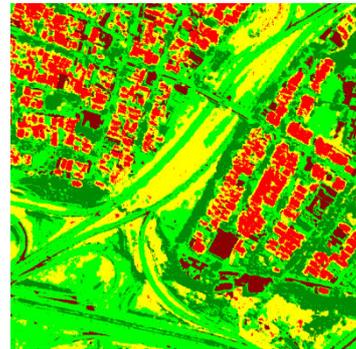
Líder HC 30%



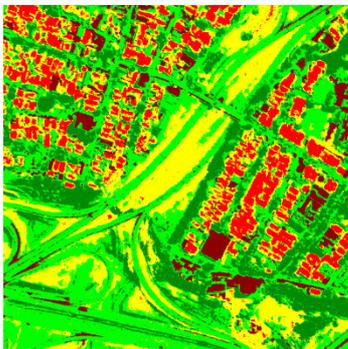
Líder Centros 50%



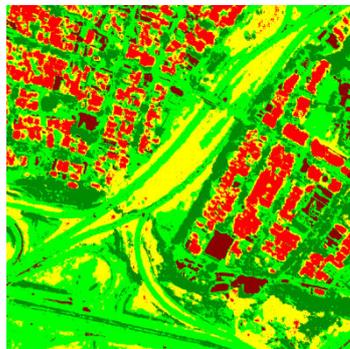
Líder H 50%



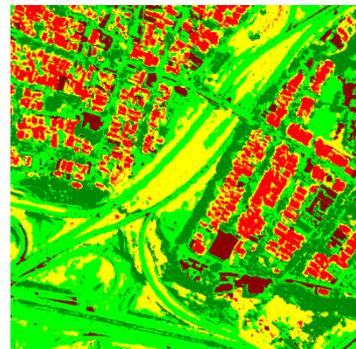
Líder HC 50%



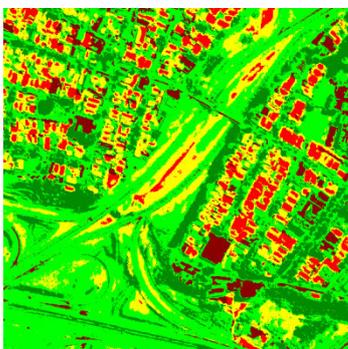
Líder Centros 70%



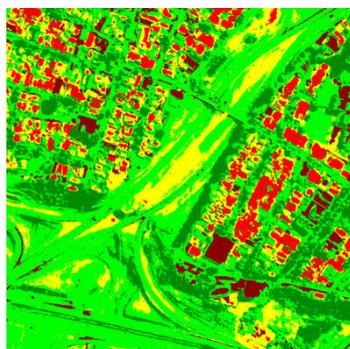
Líder H 70%



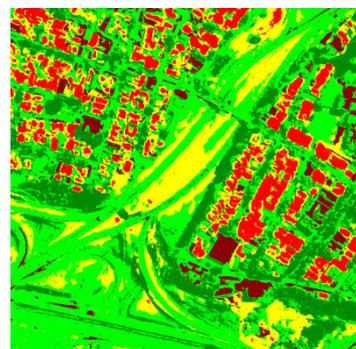
Líder HC 70%



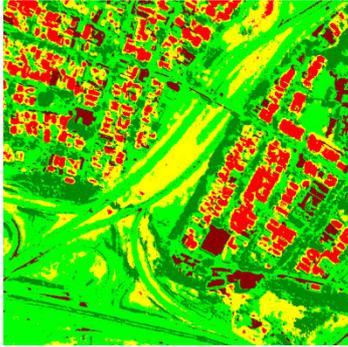
Seleção Aleatória 10%



Seleção Aleatória 30%



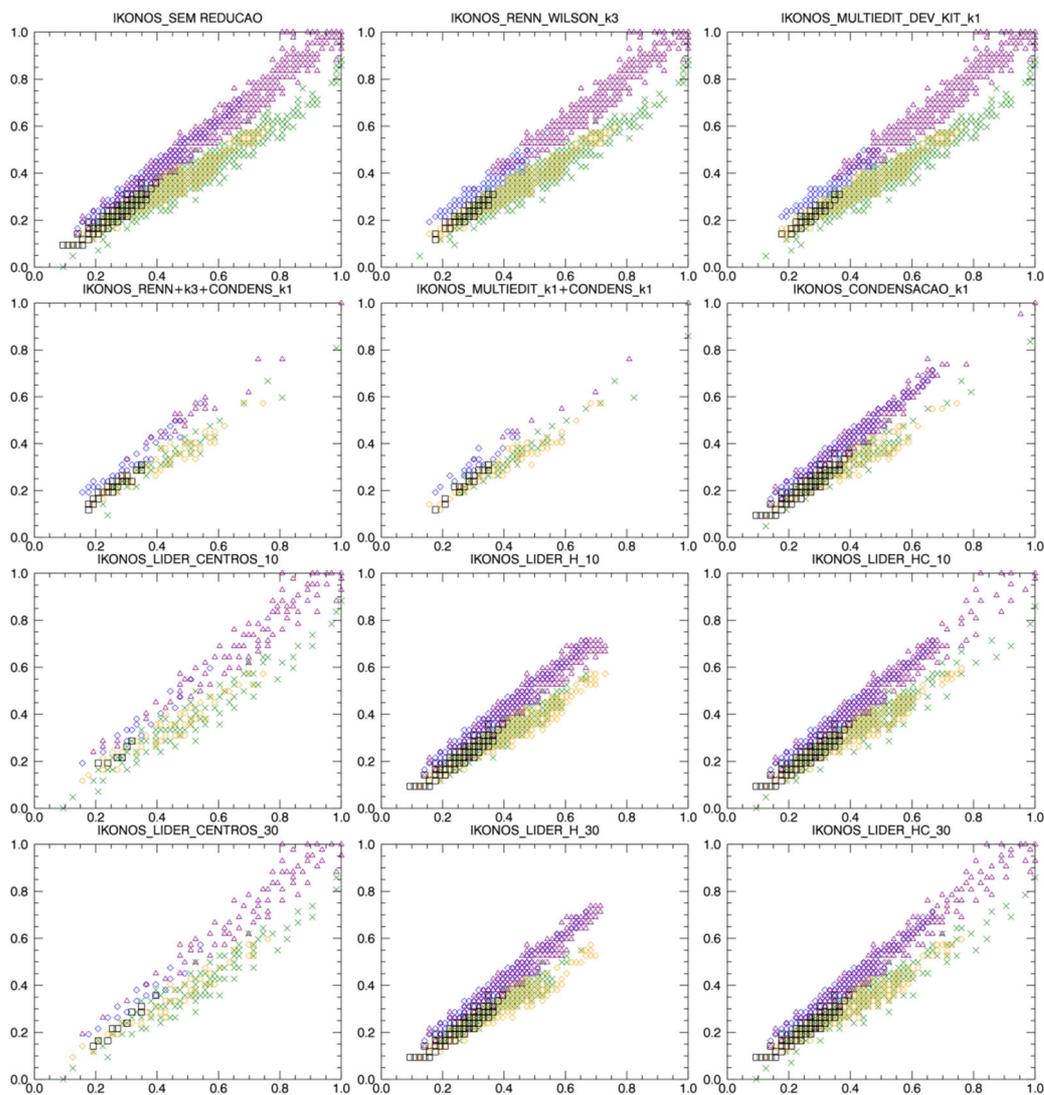
Seleção Aleatória 50%

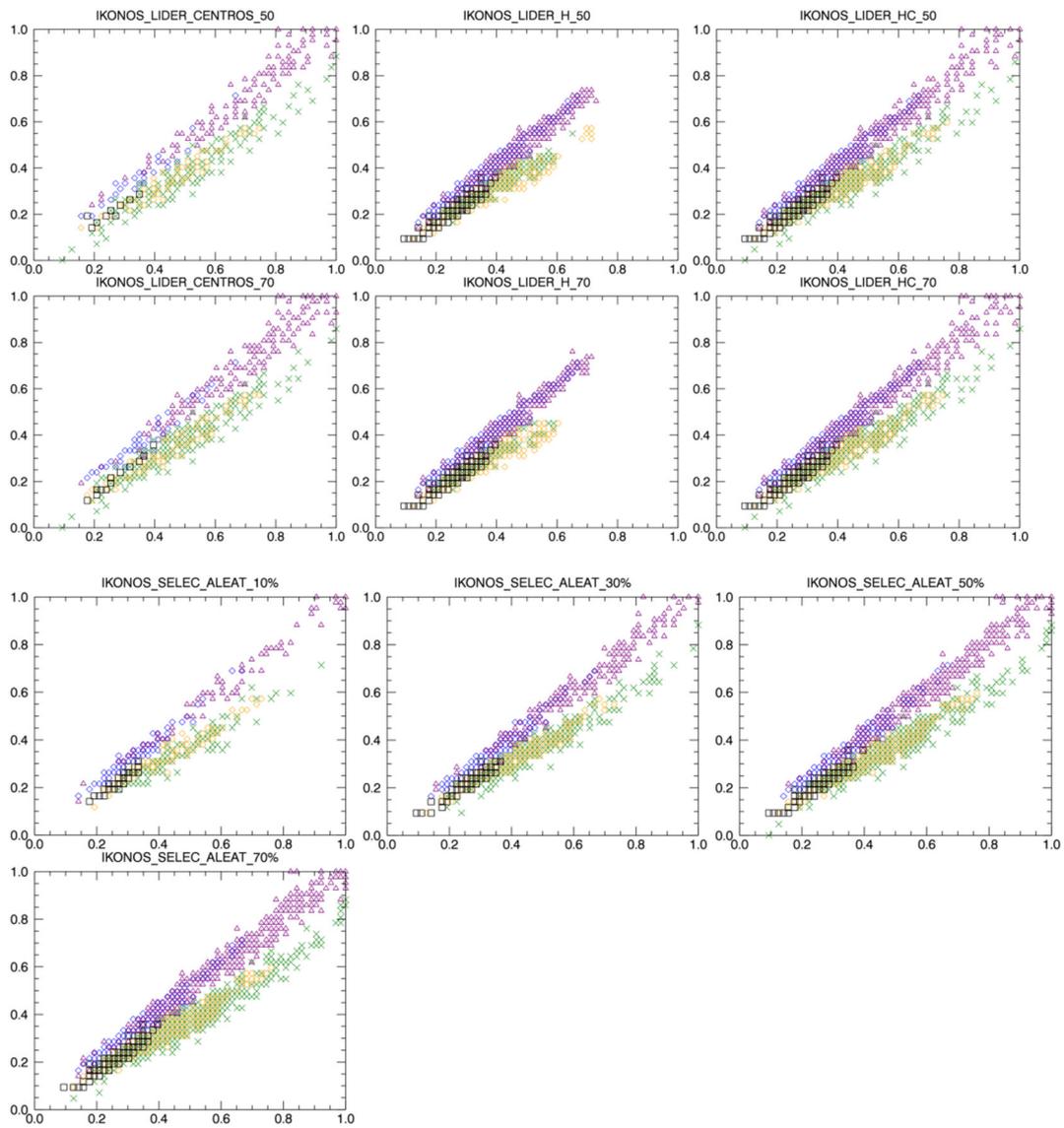


Seleção Aleatória 70%

APÊNDICE F – ESPAÇO DE ATRIBUTOS 2D – IKONOS MS

Neste apêndice são apresentadas os gráficos das amostras reduzidas em 2D da imagem IKONOS MS.





APÊNDICE G – ESPAÇO DE ATRIBUTOS 2D – ALOS PALSAR

Neste apêndice são apresentadas os gráficos das amostras reduzidas em 2D da imagem ALOS PALSAR.

