

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

Ângeluz da Costa Canena

**REDES NEURAIS ARTIFICIAIS DINÂMICAS APLICADAS A
DETECÇÃO AUTOMÁTICA DE EVENTOS
COMPORTAMENTAIS**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Engenharia Elétrica
Orientador: José Marino Neto, Dr.
Co-orientador: Fernando Mendes de Azevedo, Dr.

Florianópolis

2012

Catálogo na fonte pela Biblioteca Universitária
da
Universidade Federal de Santa Catarina

C221r Canena, Ângeluz da Costa
Redes neurais artificiais dinâmicas aplicadas a detecção automática de eventos comportamentais [dissertação] / Ângeluz da Costa Canena ; orientador, José Marino Neto. - Florianópolis, SC, 2012.
99 p.: il., grafs., tabs.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia Elétrica.

Inclui referências

1. Engenharia elétrica. 2. Redes neurais (Computação). 3. Reconhecimento de padrões. 4. Engenharia biomédica. I. Marino Neto, José. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

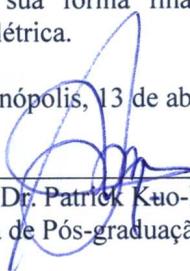
CDU 621.3

Ângeluz da Costa Canena

**REDES NEURAIIS ARTIFICIAIS DINÂMICAS APLICADAS A
DETECCÃO AUTOMÁTICA DE EVENTOS
COMPORTAMENTAIS**

Esta Dissertação foi julgada adequada para obtenção do Título de “Mestre”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica.

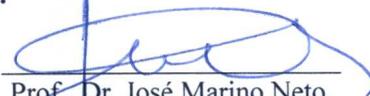
Local Florianópolis, 13 de abril de 2012.



Prof., Dr. Patrick Kuo-Peng,

Coordenador do Programa de Pós-graduação em Engenharia Elétrica

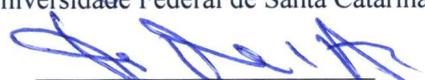
Banca Examinadora:



Prof., Dr. José Marino Neto,

Orientador

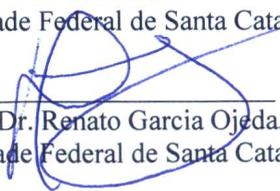
Universidade Federal de Santa Catarina



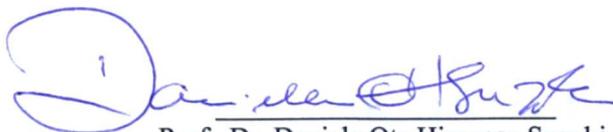
Prof., Dr. Fernando Mendes de Azevedo,

Co-Orientador

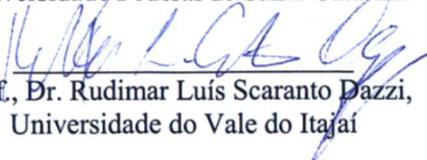
Universidade Federal de Santa Catarina



Prof., Dr. Renato Garcia Ojeda,
Universidade Federal de Santa Catarina



Prof., Dr. Daniela Ota Hisayasu Suzuki,
Universidade Federal de Santa Catarina



Prof., Dr. Rudimar Luis Scaranto Dazzi,
Universidade do Vale do Itajaí

AGRADECIMENTOS

Agradeço a minha família, que tanto amo, pela participação e apoio em minha formação pessoal, profissional e acadêmica.

Agradeço ao meu orientador prof. José Marino-Neto, pela notável dedicação à minha formação acadêmica e em fazer-me pensar como cientista. Ao meu co-orientador prof. Fernando M. de Azevedo, por sua participação na minha formação e preparo em relação para execução deste estudo, além da participação na formação pessoal através de opiniões partilhadas sobre a vida.

A todos os demais professores do IEB-UFSC, agradeço por seus esforços em transferir seus conhecimentos e experiências sob perceptível intuito de propiciar a mais completa formação.

Aos meus amigos e colegas Myla Bittencourt, Luiz Tauffer, Christine Boos, Geovani Scolaro, Aline Oliveira, Eduardo Silva, Glauco Cardozo, Ana Emília, Yasser Ismail, Marina, William Castañeda, Francisco dos Santos e Carlos Crispim, pelos momentos de descontração, companheirismo e conselhos sobre a vida. Ressaltando ao Tiago dos Santos, Fernanda Possamai, Fernando Melleu, Natália Saretta, e Amauri Balotin, pelo companheirismo em finais de semana e dia-a-dia, tanto dentro como fora do laboratório.

Aos demais amigos que tive pouco contato nestes últimos dois anos, mas que participaram do meu preparo para completar os mais diversos objetivos pessoais e profissionais.

Ao IEB-UFSC e CNPq, por proporcionar infraestrutura e recursos a esta formação.

Em especial agradeço a Deus pela minha vida, por moldar meu caráter, pelo amparo diante das adversidades e proporcionar o encontro às pessoas anteriormente citadas.

RESUMO

Registros comportamentais realizados de forma automatizada reduzem a influência de erros causados por fadiga e alerta sob os quais um observador humano está sujeito. O método de registro automático estudado em (CRISPIM, 2011) apresentou-se eficiente, embora não diferencie a ordem temporal de posturas a formar padrões de comportamento e suas diferentes durações. Neste trabalho investigamos o uso de uma rede neural artificial anteroalimentada com atrasos de tempo focados na entrada, através da qual foi possível observar o efeito de diferentes atrasos na detecção das diferentes categorias de comportamento (I – Imobilidade, L – Locomoção, V – Exploração vertical e A – Auto-limpeza). Além disso, foi feita uma comparação entre a detecção automática pelos classificadores treinados e a classificação por observadores humanos. Exceto pelo desempenho na detecção da categoria “A”, os valores médios deste foram altos (Índices Kappa >0,75 e AUC >0,8) no conjunto de amostras de validação (pré-processadas – balanceadas e sem transições). Já no teste uma grande redução no desempenho foi observada, especialmente no índice Kappa. Neste os maiores desempenhos médios Kappa/AUC foram 28,4/68,7% na categoria “I” (comprimento de memória – sessenta unidades), 30,7/71,5% para categoria “L” (comprimento de memória – quinze unidades), 23,8/75,4% na categoria “V” (comprimento de memória – trinta unidades) e 2,7/48,9% na categoria “A” (comprimento de memória – quinze unidades). Estes resultados apresentaram diferenças estatisticamente significantes com $p < 0,05$ entre desempenhos de comprimentos de memória para as categorias “I” e “L”, não tendo sido encontrada tal diferença entre memória nas categorias “E” e “A”. A confiabilidade computada para observadores humanos foi muito superior com valores médios do índice Kappa intra/inter observador de 78,9/78,9% categoria “I”, 62,3/52,1% para categoria “L”, 80,9/74,1% para categoria “E” e 89,4/91,2% para “A”. Diferenças estatisticamente significantes para $p < 0,05$ foram encontradas nos desempenhos inter-observador entre todas as categorias comportamentais, apontando maior acerto em relação à categoria “A” e maior dificuldade de registro na categoria “L”. Estes resultados demonstram que tal rede neural dinâmica é útil à classificação de padrões comportamentais, mas são necessárias adaptações no método a fim de possibilitar um sistema mais confiável em relação aos humanos. Além disto, a categoria comportamental em que foi encontrado o maior desempenho por observadores humanos, foi

a qual a rede neural desempenhou pior. Já a que a rede neural teve maior desempenho foi a qual observadores humanos tiveram seu pior. O que demonstra significativa diferença entre as informações recebidas e processamento destas por cada um destes. Também foi demonstrado que para nas categorias “I” e “L” há influência da profundidade da memória de curto-prazo sobre o desempenho de detecção.

Palavras-chave: bioengenharia, eventos comportamentais, reconhecimento de atividade, reconhecimento de padrões, redes neurais artificiais com atrasos de tempo.

ABSTRACT

Behavioral records done by some automatic method have reduced influence of errors by tiredness and awareness about what humans are liable. The automatic recording method studied in (CRISPIM, 2011) showed efficient results, besides it could not differentiate temporal order of poses that form a pattern of behavior and its duration. At this work we investigate the use of a focused time lagged feedforward neural network, through what became able to observe the influence of different delays on the detection of the various categories of behavior (I – Immobility, L – Locomotion, V – Vertical exploration and A – Grooming). Not just that, we made a comparison between the automatic detection made by the trained classifiers and the one made by humans. Except by the detection performance on the category “A”, the mean values of it was high (Kappa $>0,75$ and AUC $>0,8$ index) on the validation set (preprocessed – balanced and without transitions). On the test set otherwise, a great reduction in the performance was noticed, especially on Kappa index. The biggest mean values of Kappa/AUC indexes was 28,4/68,7% at category “I” (memory depth – sixty units), 30,7/71,5% for the “L” category (memory depth – fifteen units), 23,8/75,4% on category “V” (memory depth – thirty units) and 2,7/48,9% for “A” category (memory depth – fifteen units). These results demonstrated significant statistical difference for $p<0,05$ between performances of memory depths for the categories “I” and “L”, but not for the other ones. The reliability computed for human observers was quite superior with mean values of Kappa index intra/inter observer of 78,9/78,9% category “I”, 62,3/52,1% for category “L”, 80,9/74,1% for “E” and 89,4/91,2% for “A”. Significant statistical difference for $p<0,05$ was found at inter-observer performances between all behavioral categories, pointing a greater hit for “A” category and hardness at “L” category record. Accordingly to these results this neural network is useful to behavioral pattern classification, but it is still necessary adaptations on the method to allow a more reliable system. Beyond it, the behavioral category where was found the best performance by humans, was the worst for the neural network. On the other hand, the category where the neural network got his best performance was the worst for the humans. This shows significant difference between the received information and processing between both human and machine. For the categories “I” and “L” was demonstrated influence of the short-term memory depth at the detection performance.

Keywords: activity recognition, behavioral events, bioengineering, pattern recognition, time lagged feedforward neural networks.

LISTA DE FIGURAS

Figura 1 – Disposição de pontos luminosos: (A) Contornos externos de um sujeito andando e correndo e (B) Pontos luminosos correspondentes.	30
Figura 2 – Etapas de um sistema de visão artificial.	31
Figura 3 – Espaços de decisão bidimensionais: combinações de três variáveis descritoras de imagem (Distância percorrida, Pixeis alterados e Área).	36
Figura 4 – Espaços de decisão tridimensional aplicando três variáveis descritoras de imagem: Distância percorrida, Pixeis alterados e Área.	37
Figura 5 – Comparação de desempenho entre três classificadores utilizando curvas ROC.	41
Figura 6 – Modelo básico para processamento temporal.	44
Figura 7 – Diagrama genérico para memória de curto-prazo.	44
Figura 8 – TLFN focada com uma entrada (<i>bias</i> omitido).	45
Figura 9 – Filtro neural focado com múltiplas entradas.	46
Figura 10 – Rede neural recorrente com realimentação local e global.	47
Figura 11 – Componentes do processo de segmentação: a) Plano de fundo; b) Cena sob análise; c) Resultado da segmentação com limiar baixo; d) Resultado da segmentação com limiar intermediário; e) Resultado da segmentação com limiar alto.	51
Figura 12 – Sequencias de quadros exemplo de imobilidade e locomoção.	53
Figura 13 – Sequencias de quadros exemplo de Exploração Vertical e Auto-limpeza.	54
Figura 14 – Método de estimação do comprimento do animal: (A) Cálculo da reta para um ângulo do intervalo; e (B) Exemplo de conjunto de retas calculadas pelo método.	57
Figura 15 – Aplicação da TLFN focada desenvolvida no experimento computacional de predição de um passo. a) Sobreposição da forma de onda atual (traçado contínuo) e predita (traçado tracejado). b) Forma de onda do erro de predição.	61
Figura 16 – Apresentação de exemplos sem transição entre comportamentos.	62
Figura 17 – Composição percentual dos comportamentos no conjunto de validação cruzada usado nas redes para detecção de locomoção, antes da sub-amostragem.	63
Figura 18 – Desempenho de validação na detecção da categoria comportamental “Imobilidade”, dois neurônios na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.	69
Figura 19 – Desempenho de validação na detecção da categoria comportamental “Imobilidade”, taxa de aprendizagem fixa na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.	70
Figura 20 – Comparativo de desempenho de validação na detecção da categoria comportamental “Imobilidade”, melhores parâmetros taxa de aprendizagem e neurônios na camada oculta. Média e erro padrão da média dos índices a) Kappa	

e b) AUC. Valores marcados com “*” representam valores com diferença estatisticamente significativa para $p < 0,05$ em relação ao valor de desempenho para comprimento de trinta valores.....	71
Figura 21 – Desempenho de validação na detecção da categoria comportamental “Locomoção”, dois neurônios na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.	72
Figura 22 – Desempenho de validação na detecção da categoria comportamental “Locomoção”, taxa de aprendizagem fixa na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.....	73
Figura 23 – Comparativo de desempenho de validação na detecção da categoria comportamental “Locomoção”, melhores parâmetros taxa de aprendizagem e neurônios na camada oculta. Média e erro padrão da média dos índices a) Kappa e b) AUC.....	74
Figura 24 – Desempenho de validação na detecção da categoria comportamental “Exploração Vertical”, dois neurônios na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.....	75
Figura 25 – Desempenho de validação na detecção da categoria comportamental “Exploração Vertical”, taxa de aprendizagem fixa na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.....	76
Figura 26 – Comparativo de desempenho de validação na detecção da categoria comportamental “Exploração Vertical”, melhores parâmetros taxa de aprendizagem e neurônios na camada oculta. Média e erro padrão da média dos índices a) Kappa e b) AUC.....	77
Figura 27 – Desempenho de validação na detecção da categoria comportamental “Auto-limpeza”, dois neurônios na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.	78
Figura 28 – Desempenho de validação na detecção da categoria comportamental “Auto-limpeza”, taxa de aprendizagem fixa na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.....	79
Figura 29 – Comparativo de desempenho de validação na detecção da categoria comportamental “Auto-limpeza”, melhores parâmetros taxa de aprendizagem e neurônios na camada oculta. Média e erro padrão da média dos índices a) Kappa e b) AUC.....	80
Figura 30 – Comparativo de desempenho de teste na detecção da categoria comportamental “Imobilidade”. Média e erro padrão da média dos índices a) Kappa e b) AUC. Valores marcados com “*” representam diferença estatisticamente significativa com $p < 0,05$ em relação ao valor do comprimento de memória de sessenta valores.	81
Figura 31 – Comparativo de desempenho de teste na detecção da categoria comportamental “Locomoção”. Média e erro padrão da média dos índices a)	

Kappa e b) AUC. Valores marcados com “*” indicam diferença estatística em relação ao valor para o comprimento de memória de quinze valores.	82
Figura 32 – Comparativo de desempenho de teste na detecção da categoria comportamental “Exploração Vertical”. Média e desvio padrão dos índices a) Kappa e b) AUC.	82
Figura 33 – Comparativo de desempenho de teste na detecção da categoria comportamental “Auto-limpeza”. Média e desvio padrão dos índices a) Kappa e b) AUC.	83
Figura 34 – Desempenho de classificação intra-observador para cada categoria comportamental. Valor médio e erro padrão da média do índice Kappa. Valores marcados com “*” possuem diferença estatisticamente significativa com $p < 0,05$ em relação ao valor registrado para a categoria comportamental “Imobilidade”.	84
Figura 35 – Desempenho de classificação inter-observador para cada categoria comportamental. Valor médio e erro padrão da média do índice Kappa. Valores marcados com “*” possuem diferença estatisticamente significativa com $p < 0,05$ em relação ao valor registrado para a categoria comportamental “Imobilidade”.	86

LISTA DE QUADROS

Quadro 1 – Aplicações de redes neurais artificiais no reconhecimento de padrões em sinais biológicos.	35
Quadro 2 – Catálogo de comportamentos: definições e abreviações.	52
Quadro 3 – Parâmetros da rede, inicialização e treinamento.	59

LISTA DE TABELAS

Tabela 1 – Concordância em problema com “N” categorias	42
Tabela 2 – Conversão do registro em saída esperada para treinamento supervisionado.....	55
Tabela 3 – Melhores taxas de aprendizagem e neurônios na camada oculta na detecção de imobilidade para cada comprimento de memória	71
Tabela 4 – Melhores taxas de aprendizagem e neurônios na camada oculta na detecção de locomoção para cada comprimento de memória.....	74
Tabela 5 – Melhores taxas de aprendizagem e neurônios na camada oculta na detecção de exploração vertical para cada comprimento de memória	77
Tabela 6 – Melhores taxas de aprendizagem e neurônios na camada oculta na detecção de auto-limpeza para cada comprimento de memória.....	80
Tabela 7 – Parte do registro comportamental realizado pelos observadores, usado na avaliação da confiabilidade inter-observador. Onde “L” é locomoção, “H” é exploração horizontal e “V” é exploração vertical.	87
Tabela 8 – Comparativo do desempenho (índice Kappa médio) de classificação entre Redes neurais artificiais (automática) e Observadores humanos (manual).	89

LISTA DE ABREVIATURAS E SIGLAS

ECG – Eletrocardiograma

EEG – Eletroencefalograma

FBR – Funções de Base Radial

HM-SVM – *Hidden Markov Support Vector Machine*

MLP – *Multi Layer Perceptron*

PMC – Perceptron de Múltiplas Camadas

PNN – *Probabilistic Neural Network*

RNA – Rede Neural Artificial

SPECT – *Single-Photon Emission Computed Tomography*

SVM – *Support Vector Machine*

TLFN – *Time Lagged Feed-forward Network*

SUMÁRIO

1	INTRODUÇÃO	26
1.1	OBJETIVOS.....	28
1.1.1	Objetivos Específicos.....	28
2	FUNDAMENTAÇÃO TEÓRICA	29
2.1	RECONHECIMENTO DE MOVIMENTOS BIOLÓGICOS	29
2.1.1	Visão em seres biológicos	29
2.1.2	Visão computacional	30
2.2	REDES NEURAIAS ARTIFICIAIS	32
2.2.1	Reconhecimento de padrões	33
2.2.2	Aprendizagem em conjuntos de treino desbalanceados	38
2.2.3	Validação cruzada na seleção e comparação de modelos.....	39
2.2.4	Desempenho de detecção e classificação	39
2.3	INCORPORANDO TEMPO EM MODELOS CONEXIONISTAS ...	43
2.3.1	Redes anteroalimentadas atrasadas no tempo	44
2.3.2	Redes neurais recorrentes.....	46
3	MATERIAIS E MÉTODOS	49
3.1	Aquisição dos dados	49
3.1.1	Vídeos dos experimentos.....	49
3.1.1.1	Animais	49
3.1.1.2	Vídeos	49
3.1.1.3	Experimentos.....	50
3.1.2	ETHOWATCHER® – calibração e <i>activity tracking</i>	50
3.1.3	Associação da etografia com <i>activity tracking</i>	51
3.1.4	Descritores do objeto segmentado.....	56
3.1.4.1	Número de pixels do animal.....	56
3.1.4.2	Número de pixels alterados a cada quadro	56
3.1.4.3	Distância percorrida pelo centroide.....	56
3.1.4.4	Comprimento médio do animal	57
3.2	TLFN focada – Treinamento e validação	57
3.2.1	Parâmetros da rede e características do treinamento	59
3.2.1.1	Codificação e validação do algoritmo	60
3.2.2	Dados de entrada e pré-processamento	61

3.2.3	Estimação e comparação de desempenho	63
3.2.4	Seleção da taxa de aprendizagem	64
3.2.5	Número de neurônios na camada oculta.....	64
3.3	TLFN focada – Teste	65
3.4	Registro etográfico – Observador humano	65
3.4.1	Treinamento dos observadores.....	66
3.4.2	Finalidade dos vídeos.....	66
3.4.3	Ferramentas e métrica de desempenho.....	67
4	RESULTADOS.....	68
4.1	Desempenho de detecção – TLFN focada	68
4.1.1	Validação.....	68
4.1.1.1	Imobilidade.....	68
4.1.1.2	Locomoção	72
4.1.1.3	Exploração vertical	74
4.1.1.4	Auto-limpeza	77
4.1.2	Teste.....	80
4.1.2.1	Imobilidade.....	81
4.1.2.2	Locomoção	81
4.1.2.3	Exploração vertical	82
4.1.2.4	Auto-limpeza	83
4.2	Desempenho do observador humano	83
4.2.1	Concordância intra-observador	84
4.2.2	Concordância inter-observador	85
5	DISCUSSÃO	90
6	CONCLUSÃO	93

1 INTRODUÇÃO

O comportamento de um indivíduo é influenciado por diversos fatores, como a interpretação da situação e ambiente na qual se encontra, experiências em sua memória, fatores genéticos, exposição a fármacos e seu estado de nutrição. Assim, um registro dos comportamentos que tal indivíduo realiza, contém informações dos fatores que o influenciaram ao longo de um período de observação.

Para avaliação do efeito de tratamentos e condicionamentos em animais de laboratório, disciplinas como a neurociência e farmacologia analisam o registro comportamental destes animais. Entretanto, quando tais registros são feitos por um observador humano, ficam sujeitos a influência de sua experiência e estados de alerta (consciência) e atenção (concentração), resultando em uma imprecisão nos registros.

Esta fragilidade nos registros manuais originou esforços em sua automação. Dentre as metodologias de automação desenvolvidas, a detecção de comportamentos em vídeos digitais apresentou bons resultados. Exemplo disto é a sequência de estudos realizados no laboratório de Bioengenharia do IEB-UFSC (Instituto de Engenharia Biomédica da Universidade Federal de Santa Catarina). Utilizando técnicas de processamento de imagem, um sistema de rastreamento e análise de comportamentos, a partir de vídeos de animais de laboratório, foi desenvolvido e aplicado em experimentos com pombos e ratos (BOSE, 2003). Através de dados extraídos como a distância percorrida pelo animal durante o experimento, o sistema desenvolvido neste trabalho permitiu diferenciar um grupo de ratos tratados com Cafeína (psicoestimulante) de outro administrado com Haloperidol (antipsicótico). A partir dos algoritmos de processamento de imagem empregados por (BOSE, 2003), em Pederiva (2005) um sistema para auxílio ao registro comportamental foi desenvolvido e nomeado “Etógrafo”. O módulo para rastreamento deste sistema que incorporou os algoritmos de processamento de imagem comentados foi denominado “Espião”. Ademais um usuário pode criar ou carregar um catálogo de categorias comportamentais e fazer o registro dos comportamentos realizados pelo animal observado. Como resultado o usuário tem uma planilha com informações como duração, latência e frequência dos comportamentos registrados. Necessitando dos dados brutos de cada quadro das características extraídas através das técnicas de processamento de imagem empregadas em Bose (2003), o ETHOWATCHER® foi desenvolvido (CRISPIM, 2011). Em seu estudo

Crispim (2011) utilizou redes neurais artificiais estáticas do tipo perceptron multicamadas para detectar alguns comportamentos pré-catalogados a partir de vídeos de ratos de laboratório em arenas de campo aberto. Além disto, foram evidenciados descritores morfológicos e cinemáticos relevantes a detecção destes comportamentos sob as condições em questão.

Os eventos comportamentais abordados em Crispim (2011) foram locomoção, exploração vertical, imobilidade e auto-limpeza. Não somente estes, mas a maioria dos comportamentos conhecidos são caracterizados por um período de atividade (ainda que de curta duração) e não por uma postura somente. A fim de aglutinar a informação da atividade distribuída no tempo, em apenas um instante, Crispim (2011) utilizou parâmetros estatísticos de dados anteriores e adiante ao instante observado. Apesar de ter demonstrado eficiência na detecção destes comportamentos, o método empregado, apresenta o inconveniente de não considerar a ordem em que as poses observadas ocorrem. Além disto, um número fixo de quadros para todos os comportamentos foi utilizado, sendo que para alguns comportamentos, a observação de poucas poses torna possível caracterizá-lo, enquanto que para outros, um período de observação mais longo é necessário.

Para obter um classificador automático de comportamentos capaz de considerar a ordem das poses apresentadas e ainda, investigar a questão do número mais apropriado de posturas para discriminação dos mesmos, foi empregada uma rede neural artificial dinâmica em tal tarefa neste trabalho.

As redes neurais artificiais dinâmicas destacam-se das estáticas por apresentar memória associada a seus pesos e unidades neuronais, representadas na forma de unidades de atraso (HAYKIN, 2005). Cada unidade de atraso armazena informação de um instante de tempo e quando nova informação é inserida na rede, os atrasos mais profundos recebem o valor armazenado nos atrasos mais superficiais (mais próximos da entrada sem atraso), de forma que a informação da ordem de apresentação dos dados é preservada.

O uso de uma rede neural artificial dinâmica possibilita então, desenvolver uma ferramenta que reduza a possibilidade de erros de detecção provocados por fadiga e ambiguidade de definições das categorias comportamentais, também capaz de conduzir experimentos de longa duração ininterruptamente. Além disto, ao usar atributos morfológicos e cinemáticos da imagem que são ausentes das descrições habituais do comportamento, revelar aspectos adicionais das categorias comportamentais examinadas.

1.1 OBJETIVOS

Investigar o uso de uma rede neural artificial dinâmica, na detecção e classificação automática de comportamentos em animais em campo aberto.

1.1.1 Objetivos Específicos

- Estudar e apontar dentre as topologias de redes neurais artificiais dinâmicas, as potencialmente úteis à detecção e classificação automática de comportamentos;
- Simular e avaliar o efeito de diferentes tamanhos de memória de redes neurais artificiais com atrasos de tempo focados na classificação de diferentes categorias de comportamento;
- Aplicar índices de desempenho de ferramentas de diagnóstico binário na comparação entre o desempenho da solução automatizada de diferenciação dos movimentos locomotores, não locomotores e de imobilidade, e o desempenho por observação direta de humanos no registro de categorias comportamentais.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 RECONHECIMENTO DE MOVIMENTOS BIOLÓGICOS

As principais atividades de sobrevivência executadas por diversas espécies do reino animal como detecção de predadores, seleção de presa e cortejo, bem como o aprendizado de ações motoras por imitação, envolvem o reconhecimento de movimentos biológicos complexos (GIESE e POGGIO, 2003). A vasta gama de possíveis aplicações de autômatos com a habilidade de realizar tal tarefa, bem como possibilidade de desenvolvimento de tratamentos a patologias relacionadas à cognição visual, tem impulsionado estudos do treinamento de máquinas para o reconhecimento de atividades motoras tanto de humanos, quanto de animais de laboratório.

2.1.1 Visão em seres biológicos

Para Marr (1982), a visão é um processo composto de duas partes, mensura e inferência. O processo de mensura compreende a recepção e conversão do sinal luminoso em sinais elétricos, que apesar de modulado por experiência, contexto e atenção, fornece apenas indícios do que está sendo visto (CAVANAGH, 2011).

Já a parte da inferência no processamento visual, refere-se a processos de análises de cena baseadas em decisões, que combinam conhecimentos prévios armazenados na memória com o sinal na entrada da retina para formar representações do que está sendo visto (CAVANAGH, 2011). No processo de inferência estariam então relacionadas à assimilação de formas bidimensionais com modelos tridimensionais armazenados previamente, bem como a assimilação de padrões de movimentos às características do objeto sob movimentação.

Em estudos usando pontos luminosos, como Johansson (1973), foi demonstrado que humanos são capazes de reconhecer movimentos biológicos com grande robustez. Neste experimento, atores em um ambiente escuro com dez bulbos luminosos, dispostos como demonstrado na Figura 1, foram filmados realizando movimentos como caminhar, correr e dançar. Assistindo os vídeos, sujeitos foram capazes não somente de identificar a atividade realizada imediatamente, mas também identificar que foram realizados por atores humanos.

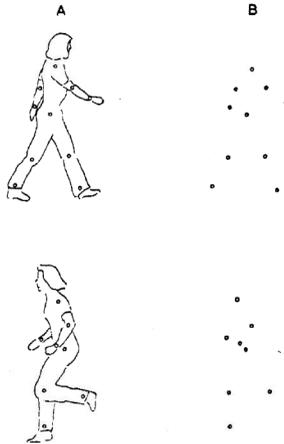


Figura 1 – Disposição de pontos luminosos: (A) Contornos externos de um sujeito andando e correndo e (B) Pontos luminosos correspondentes.

Fonte: (JOHANSSON, 1973).

Além deste, conforme relatado em Giese e Poggio (2003), outros trabalhos semelhantes ao de Johansson (1973) usando pontos luminosos revelaram a capacidade de reconhecimento também de características como sexo e estado emocional.

Estas inferências que ocorrem durante o processamento visual, tanto de associação de formas bidimensionais a modelos tridimensionais, quanto de padrões de movimento ao objeto sob atividade, permitem uma extensão da aplicação dos conceitos de auto e hetero associatividade.

2.1.2 Visão computacional

Analogamente, e até mesmo pela própria inspiração biológica de grande parte dos sistemas desenvolvidos, a visão em sistemas computacionais também pode ser dividida em etapa de mensura e inferência. Em Marques e Vieira (1999) os sistemas de visão artificial são subdivididos em cinco etapas, apresentadas na Figura 2.

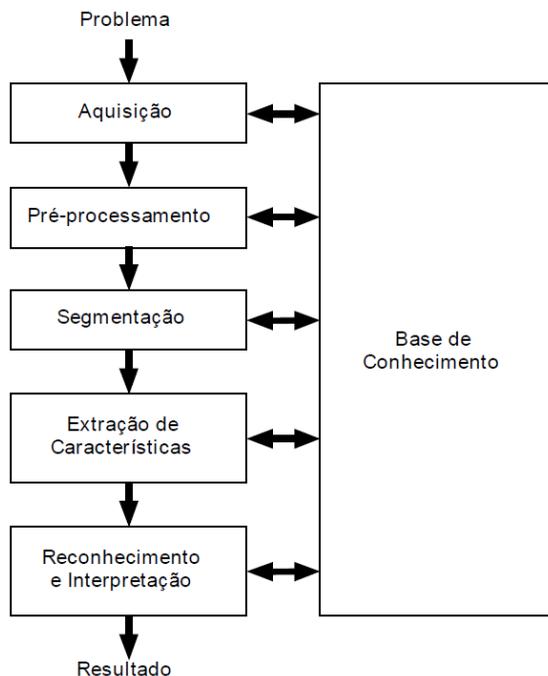


Figura 2 – Etapas de um sistema de visão artificial.
 Fonte: (MARQUES e VIEIRA, 1999).

Das etapas propostas na Figura 2, podemos enquadrar o processo de mensura como o conjunto das etapas de aquisição e pré-processamento. Por sua vez, a inferência seria uma composição das etapas de segmentação, extração de características, e também reconhecimento e interpretação. Para cada uma das etapas de processamento digital que compõem os sistemas de visão artificial (ou computacional) diferentes técnicas podem ser empregadas. Como mencionado em Crispim (2011), qual técnica será mais adequada aplicar a cada etapa dependerá das características do domínio de aplicação do sistema de visão. Ainda Crispim (2011) *apud* (FRÖLICH *et. al.*, 2008), (HEEREN e COOLS, 2000), (BOSE, 2003), (LIND *et. al.*, 2005), (SPINK *et. al.*, 2001), (ZURN, JIANG e MOTAI, 2005), (METRIS), (SILVA *et. al.*, 2007), (YU e TAN, 2009), (TWINING, TAYLOR e COURTNEY, 2001) e (JHUANG *et. al.*, 2010), indica que em estudos relacionados à análise de atividade animal e detecção automática de comportamentos é predominante o uso de técnicas de limiarização e

subtração de imagens na etapa de segmentação, visto que a maioria dos trabalhos rastreia apenas um objeto.

A segmentação em um sistema de visão artificial é o processo pelo qual é separado o(s) objeto(s) de interesse do restante da imagem (MARQUES e VIEIRA, 1999). A técnica de subtração de imagens consiste em subtrair ponto a ponto a imagem em análise (contendo o objeto de interesse) de uma imagem referência (contendo apenas o plano de fundo). Esta técnica apresenta como restrições significativo contraste entre o objeto e o plano de fundo, posição relativa fixa entre câmera e plano de fundo, bem como plano de fundo estático (CRISPIM, 2011) *apud* (GONZALEZ e WOODS, 2000) e (LIND *et. al.*, 2005). Neste trabalho foi empregada a técnica de subtração de imagens complementada por limiarização, mesma técnica usada em Crispim (2011), visto que são trabalhos que as condições atendem às restrições impostas pela técnica.

Em trabalhos onde as restrições mencionadas não são atendidas, como o rastreamento de múltiplos objetos, ou quando não há significativo contraste entre o objeto e o plano de fundo, ou ainda há alteração do plano de fundo durante o experimento, outras técnicas como o fluxo óptico (DOLLAR *et. al.*, 2005), adaptações do algoritmo *K-means* e do *CamShift* são aplicáveis (WANG *et. al.*, 2009).

2.2 REDES NEURAIS ARTIFICIAIS

Para Boden (1996) o conexionismo é uma abordagem para modelagem da percepção e cognição através do emprego explícito de mecanismos e forma de processamento como os quais se acredita ocorrer no cérebro. As redes neurais artificiais (também conhecidos como neurocomputadores e processadores paralelamente distribuídos) são modelos conexionistas, tentam imitar funções realizadas por redes neurais biológicas a partir da associação de modelos matemáticos do neurônio biológico. Como exposto em Haykin (2005) e Boden (1996), a principal propriedade de uma rede neural é a habilidade de aprender sob o ambiente em que está inserida e melhorar seu desempenho por um método de aprendizagem. Por ambiente entendem-se os exemplos que alimentam as entradas da rede e aprendizagem as modificações em seus parâmetros livres.

Os algoritmos de redes neurais são caracterizados principalmente pela regra de aprendizado empregada e por sua estrutura (arquitetura). Quanto à arquitetura temos como principais variações o número de

camadas empregadas, se todos os neurônios de uma camada conectam-se ou não a todos os da próxima e se há ou não conexões recorrentes. Conforme o método de aprendizagem aplicado, certas tarefas de aprendizagem são possíveis para uma rede neural em questão. De modo geral os métodos de aprendizado são enquadrados em dois paradigmas, o aprendizado com ou sem um professor. Na aprendizagem com um professor (aprendizagem supervisionada) considera-se que para um dado conjunto de exemplos apresentados durante um treinamento à rede neural, o professor que possui conhecimento sobre o ambiente (dados de entrada) determina qual a saída ideal a ser apresentada pela rede neural. A partir da diferença entre o ideal (saída desejada) e a saída apresentada pela rede os pesos livres da rede são atualizados, ocorrendo o aprendizado. Um dos modelos de aprendizado sem um professor é a aprendizagem não supervisionada (ou auto-organizada). Neste paradigma não há uma referência externa à rede quanto à classificação correta ou incorreta, uma medida independente da tarefa é fornecida para ajuste dos pesos. Uma regra de aprendizagem baseada neste paradigma é a aprendizagem competitiva, onde apenas um neurônio vencedor tem seus pesos ajustados de forma a aproximar cada vez mais a resposta de um dado neurônio aos padrões de entrada com características semelhantes (HAYKIN, 2005).

2.2.1 Reconhecimento de padrões

Conforme Haykin (2005) reconhecimento de padrões é o processo pelo qual um padrão recebido é atribuído a uma classe dentre um número predeterminado de classes (ou categorias). E uma interessante definição para sinais biológicos ou biossinais é apresentada em Escabi (2005) como sendo registros no espaço, tempo ou espaço e tempo de eventos biológicos. O uso de algoritmos de redes neurais no reconhecimento (classificação) de padrões em sinais biológicos é encontrado em muitos trabalhos, abrangendo aplicações no auxílio a diagnóstico em sistemas de imagem médica, sinais elétricos cerebrais, cardíacos, bem como sinais acústicos da voz (palavras, conversação, fonemas). Além destes, muitos trabalhos foram realizados nas últimas décadas sobre o reconhecimento de padrões em sinais biológicos comportamentais e de atividade em humanos e animais. O Quadro 1 apresenta alguns trabalhos onde redes neurais artificiais foram utilizadas no reconhecimento de padrões em sinais biológicos, especificando o tipo de rede, sinal adquirido e a classificação aplicada.

Autor	Tipo e aplicação do classificador
(WATROUS e SHASTRI, 1987)	Classificação das palavras “no” e “go” a partir dos sinais acústicos de dezesseis canais de áudio utilizando uma rede neural com conexões recorrentes locais nas unidades da camada oculta e saída.
(YUHAS, GOLDSTEIN e SEJNOWSKI, 1989)	Classificação de nove diferentes vogais a partir da combinação do espectro de frequências do sinal acústico e a imagem da boca do locutor aplicando uma rede neural do tipo MLP.
(GONZALEZ e ARNALDO, 1993)	Rede neural MLP na classificação de seis categorias comportamentais de um rato a partir do espectro de frequências de um sinal de transdutor capacitivo.
(PAGE <i>et. al.</i> , 1996)	Detecção do mal de Alzheimer em imagens obtidas através de SPECT da ganglia basal e cerebelo aplicando rede neural MLP.
(SAHINER <i>et. al.</i> , 1996)	Detecção de tecido cancerígeno e normal em região de interesse de mamogramas utilizando uma rede neural convolutiva.
(ROUSSEAU <i>et. al.</i> , 2000)	Classificação de atividade de rato em arena de campo aberto em dez categorias comportamentais a partir de descritores morfológicos e cinemáticos de sequencias de imagens utilizando rede TLFN focada com MLP na parte estática.
(YU <i>et. al.</i> , 2005)	Classificação de três categorias do padrão locomotor em humanos utilizando análise de componentes principais de um sinal descritor de imagem através de uma rede MLP.
(PALODETO e MARQUES,	Rede neural MLP utilizada para classificar o tipo de batida do coração (normal, contração ventricular

2007)	prematura ou bloqueio de ramo) a partir da transformada <i>wavelet</i> de um sinal de ECG.
(JHUANG <i>et. al.</i>, 2010)	Classificação de oito categorias comportamentais de ratos em gaiola utilizando uma combinação de descritores de movimento com descritores morfológicos e cinemáticos aplicados a uma rede HM-SVM.
(SCOLARO e AZEVEDO, 2010)	Classificação de sinais de EEG em epileptiformes ou não epileptiformes utilizando uma TLFN focada com MLP na parte estática.
(ZADEH, KHAZAEI e RANAEE, 2010)	Classificação de arritmias em sinal de ECG utilizando comparativamente redes neurais MLP, RBF, PNN e SVM.
(BOOS <i>et. al.</i>, 2011)	Detecção da ausência e presença de paroxismos em sinais de EEG a partir de descritores morfológicos através de uma rede neural do tipo MLP.
(CRISPIM, 2011)	Detecção de quatro categorias comportamentais usando uma rede do tipo MLP alimentada por parâmetros estatísticos de descritores morfológicos e cinemáticos a partir de sequências de imagens de vídeos de ratos em arenas de campo aberto.

Quadro 1 – Aplicações de redes neurais artificiais no reconhecimento de padrões em sinais biológicos.

Neste trabalho a tarefa requerida à máquina de aprendizado é a detecção de padrões comportamentais, associando um padrão não visto da classe a ser detectada a tal classe. Sendo um problema de reconhecimento de padrões com duas classes (pertencer ou não à classe em questão). Neste tipo de tarefa, cada padrão (ou exemplo) é representado por um ponto em um espaço de decisão multidimensional (uma dimensão para cada entrada do modelo) (HAYKIN, 2005). Por exemplo, uma rede neural sob a tarefa de reconhecimento dos padrões comportamentais de um objeto em que o ambiente é formado por variáveis que descrevem sequências de imagens do mesmo. Cada variável descritora representaria uma dimensão no espaço de decisão,

sob quais as classes seriam distribuídas. Na Figura 3 apresentamos exemplos de espaços bidimensionais ilustrando combinações das variáveis descritoras da imagem (Distância percorrida, Pixels alterados e Área) para o exemplo de reconhecimento de padrões sugerido.

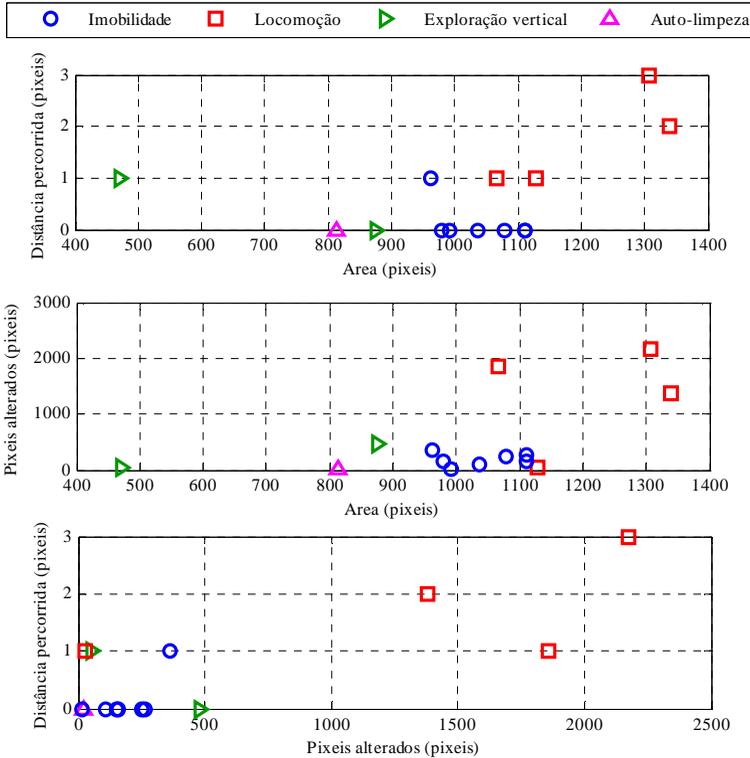


Figura 3 – Espaços de decisão bidimensionais: combinações de três variáveis descritoras de imagem (Distância percorrida, Pixels alterados e Área).

Pelo exemplo da Figura 3, nota-se no gráfico inferior um exemplo sob o qual a combinação de descritores pixels alterados e distância percorrida seriam insuficientes para diferenciar a classe “Auto-limpeza” da classe “Imobilidade”. O que não ocorre em os outros dois exemplos onde as dimensões da entrada são distância percorrida por área e pixels alterados e área. Ainda sob o exemplo de reconhecimento de padrões mencionado, a Figura 4 demonstra uma representação de um espaço de

decisão tridimensional onde são combinadas as três variáveis descritoras.

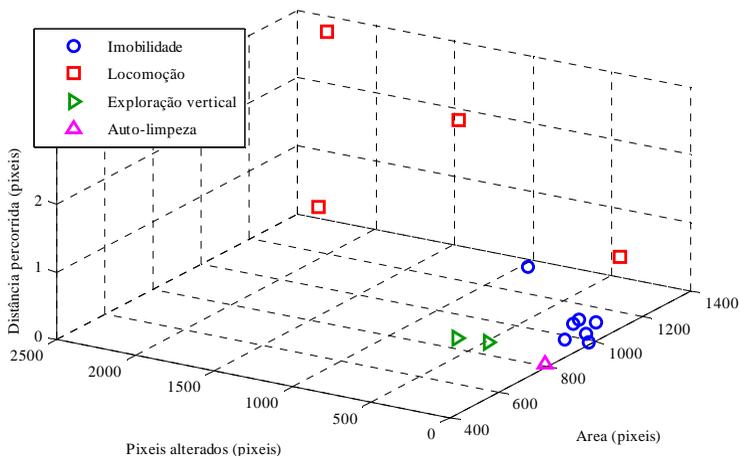


Figura 4 – Espaços de decisão tridimensional aplicando três variáveis descritoras de imagem: Distância percorrida, Pixéis alterados e Área.

Quanto maior a dimensionalidade dos padrões de entrada, mais informações sobre o ambiente. Entretanto, mais parâmetros livres podem ser necessários para compor a função que separe corretamente as classes. É válido ressaltar que o número de dimensões dos dados de entrada de uma rede neural não tem limites, os exemplos abordados apresentaram até três dimensões por ser de mais fácil representação gráfica. Adiante na seção 2.3 serão discutidas redes neurais que possuem como dimensões de entrada não somente um conjunto de variáveis, mas também estas variáveis atrasadas no tempo.

O processo de treinamento define a forma que as fronteiras de decisão irão tomar (regra sobre qual a rede neural define a classe a qual um vetor de entrada pertence). Sendo este um processo estatístico determinado pela variabilidade inerente nos dados que formam cada uma das classes (HAYKIN, 2005). Nesta mesma literatura são sugeridas duas abordagens ao reconhecimento de padrões utilizando redes neurais artificiais. A primeira abordagem sugerida é associar uma rede neural não supervisionada (encarregada de extrair características de um conjunto de entradas reduzindo a dimensionalidade dos dados de entrada) a uma rede neural supervisionada (com função de classificar os

dados comprimidos nas classes desejadas). Outra abordagem é uma máquina de aprendizado de múltiplas camadas com treinamento supervisionado, onde as camadas intermediárias da rede assumem o papel de redução de dimensionalidade. Esta segunda abordagem foi empregada neste estudo, conforme detalhes na seção 3.2.

2.2.2 Aprendizagem em conjuntos de treino desbalanceados

Em uma tarefa de detecção de padrão, quando há mais exemplos não pertencentes à classe a ser detectada no conjunto de treino do que exemplos desta, diz-se que o conjunto de treino está desbalanceado. Condição esta, que implica em prejuízo de convergência para a maioria dos algoritmos de aprendizagem de máquina. Quando um conjunto de treino significativamente desbalanceado é aplicado, a fronteira de decisão entre classes tende a aproximar-se da classe mais representada no conjunto, levando a um aumento na classificação incorreta da classe em minoria. Além disto, a maioria dos algoritmos de aprendizagem baseia-se em medidas de desempenho, de modo que se a métrica usada para mensurá-lo não leva em conta a proporção das classes, os erros e acertos da classe em minoria não são adequadamente representados (NGUYEN, BOUZERDOUM e PHUNG, 2009).

O desbalanceamento de um conjunto de treino pode ter origem, na proporção desigual dos exemplos das classes, na escassez de exemplos de forma geral, bem como na separabilidade dos dados em relação às classes. Entre as abordagens para solução do desbalanceamento, uma das mais comuns é mudar a amostragem de forma a minimizar a má distribuição dos exemplos no conjunto de treino. As formas mais básicas de realizar esta tarefa é através da sub-amostragem (do inglês *under-sampling*) e da sobre-amostragem (do inglês *over-sampling*). A sub-amostragem consiste em extrair um conjunto menor de dados da classe em maioria e preservar os dados da classe em minoria. Já a sobre-amostragem, consiste em repetir as amostras sob minoria e preservar as amostras sob maioria (NGUYEN, BOUZERDOUM e PHUNG, 2009). Neste trabalho, adotamos a sub-amostragem, descrita em detalhes no capítulo 2.6.2.1.1.

2.2.3 Validação cruzada na seleção e comparação de modelos

A generalização em um processo de aprendizagem consiste em mapear corretamente um exemplo não apresentado durante o treinamento sob as possíveis saídas. Esta característica é influenciada pelo conjunto de treinamento (quanto a tamanho e o quão representativo este é da classe em questão), a arquitetura de rede (quanto a parâmetros e algoritmo), bem como a complexidade do problema a ser mapeado (HAYKIN, 2005).

Como na maioria das aplicações o conjunto de dados é limitado, a capacidade de generalização de um modelo é avaliada sob um conjunto de dados dividido entre treino e validação (exemplos não apresentados durante o treino). A estratégia de validação feita com apenas uma divisão do conjunto é denominada *hold-out*, mas esta não é uma estratégia estatisticamente atraente. Uma melhor estimativa da capacidade de generalização é fornecida por estratégias de validação cruzada, onde mais de uma divisão é realizada no conjunto de dados e o desempenho resultante é expresso pela média de cada *hold-out*. Um dos primeiros métodos de validação cruzada a ser usado foi o *leave-one-out*, que consiste em dividir o conjunto de dados um número de vezes igual ao número de exemplos no conjunto e validar em cada um. Este método, apesar de fornecer uma ótima estimativa de desempenho, é muito custoso computacionalmente. Uma alternativa ao *leave-one-out*, menos custosa computacionalmente é o *V-fold CV* (Validação cruzada de “V” dobras), que consiste em dividir o conjunto de dados em “V” subconjuntos de mesma cardinalidade e validar com cada um (ARLOT e CELISSE, 2010). Neste trabalho foi empregado o mesmo método de validação cruzada que em Crispim (2011), *V-fold* com dez dobras, descrito com mais detalhes na seção 3.2.2.

2.2.4 Desempenho de detecção e classificação

Os índices usados na avaliação do desempenho de detectores ou classificadores binários são equacionados a partir de quatro medidas principais. Estas medidas são o número de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. Para mensurá-las é necessário tomar uma referência, se existente um “padrão ouro”. A partir de tal referência, considera-se verdadeiro quando o detector sob avaliação concorda com a referência e falso caso contrário.

Respectivamente, classifica-se como positivo ou negativo quando o evento procurado foi dito como encontrado ou não (CASSON, LUNA e VILLEGAS, 2009), (CRISPIM, 2011) e (SOVIERZOSKI, ARGOUD e AZEVEDO, 2008).

A sensibilidade mede o quanto um classificador binário é capaz de identificar como positivos os padrões positivos apresentados e a especificidade, o quanto é capaz de identificar como negativos os padrões negativos (SOVIERZOSKI, ARGOUD e AZEVEDO, 2008). Estes índices são calculados através da Equação 2.1 e da Equação 2.2, respectivamente:

$$\text{sensibilidade} = \frac{VP}{VP+FN} \quad (2.1)$$

$$\text{especificidade} = \frac{TN}{TN+FP} \quad (2.2)$$

Onde “VP” é o total de verdadeiros positivos, “VN” o de verdadeiros negativos, “FP” total de falsos positivos e “FN” falsos negativos.

A curva ROC (*Receiver Operating Characteristics*) é um gráfico cartesiano que representa a relação entre a sensibilidade e especificidade de um classificador. Assim, sua análise demonstra o quanto este é hábil na classificação a que está sendo empregado. Os pontos que formam a curva ROC são pares de sensibilidade (ordenadas) e 1-especificidade (abscissas). Em redes neurais artificiais estes vários pares são obtidos pela aplicação de limiares sob a saída da mesma, que variam do valor de saída mínimo ao máximo. Quanto mais limiares aplicados, maior a quantidade de pontos na curva e conseqüentemente sua exatidão (CRISPIM, 2011) e (SOVIERZOSKI, ARGOUD e AZEVEDO, 2008). Na Figura 5 um exemplo de comparação entre classificadores utilizando curvas ROC é apresentado.

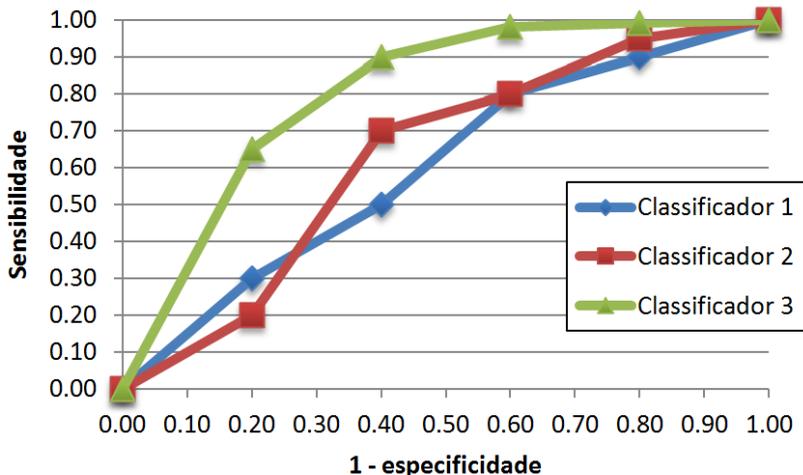


Figura 5 – Comparação de desempenho entre três classificadores utilizando curvas ROC.

No caso da comparação do desempenho entre os classificadores um e dois apresentados na Figura 5, onde as curvas ROC são muito próximas, mede-se a área sobre a curva ROC. Esta medida conhecida como AUC (*Area Under the Curve*) também varia de zero (desempenho mínimo) a um (desempenho máximo) e reúne o desempenho do classificador para os vários limiares testados (CRISPIM, 2011).

Como dito anteriormente, para calcular os índices descritos é necessário tomar uma referência, tornando-os especialmente recomendados sob a existência de um “padrão ouro” de resposta. Não havendo tal, a confiabilidade de um processo de classificação pode ser medida indiretamente em termos de reprodutibilidade (SOVIERZOSKI, ARGOUD e AZEVEDO, 2008), (CRISPIM, 2011) e (LE, 2003).

Um índice estatístico indicado para avaliar a concordância entre observadores (classificadores) é o Kappa, o qual possibilita avaliar a concordância entre dois observadores para mais de duas categorias. Este índice é influenciado não somente pela proporção de concordância geral entre os observadores, mas também pela proporção esperada da concordância ocorrer por chance (MARTIN e BATESON, 2007) e (CRISPIM, 2011).

Tabela 1 – Concordância em problema com “N” categorias

Categorias	1	2	N	Total
1	a_{11}	a_{12}	a_{1N}	L_1
2	a_{21}	a_{22}	a_{2N}	L_2
N	a_{N1}	a_{N2}	a_{NN}	L_N
Total	C_1	C_2	C_N	TR

Na Tabela 1, os totais em “L” e “C” representam a soma dos “a” em cada linha e coluna respectivamente. O valor em “TR” é a soma dos totais em “C” ou “L”, representando o total de registros e “N” o total de categorias do registro. Tomando como exemplo a Tabela 1, a proporção de concordância entre os observadores pode ser calculada pela Equação 2.3 e a proporção de concordância esperada por chance pela Equação 2.4 (FISHER e ET. AL, 2004):

$$PA = \sum_{i=1}^N \frac{a_{ii}}{TR} \quad (2.3)$$

$$PC = \sum_{i=1}^N \frac{C_i \cdot L_i}{TR^2} \quad (2.4)$$

Definidas a proporção de concordância e a proporção de concordância esperada por chance, o índice Kappa pode ser calculado pela Equação 2.5 (FISHER e ET. AL, 2004), (CRISPIM, 2011), (MARTIN e BATESON, 2007) e (LE, 2003):

$$Kappa = \frac{PA-PC}{1-PC} \quad (2.5)$$

O valor do Kappa varia entre zero e um, (LE, 2003) da às seguintes orientações para avaliação deste índice:

- Se $Kappa > 0,75$ a classificação apresenta excelente reprodutibilidade;
- Se $0,40 \leq Kappa \leq 0,75$ boa reprodutibilidade;
- $Kappa < 0,40$ reprodutibilidade pobre, sugere-se a necessidade de avaliações múltiplas.

Neste trabalho foram empregados os índices AUC e Kappa como indicadores do desempenho de generalização da rede neural, bem como

para avaliar a concordância entre classificadores humanos. A aplicação destes índices está descrita em detalhes na seção 3.2.2.

2.3 INCORPORANDO TEMPO EM MODELOS CONEXIONISTAS

Os modelos conexionistas conhecidos podem realizar diversas tarefas, conforme seu método de aprendizado. Dentre estas tarefas há a associação de padrões, o reconhecimento de padrões, a aproximação de funções, o controle de um processo, a filtragem e o “*beamforming*”, um processo de filtragem espacial. Memória e aprendizado estão intrinsecamente conectados, de modo que para que algo seja aprendido é necessária memória para armazenar a informação e, por sua vez, para que algo seja armazenado na memória um processo de aprendizado é necessário. A memória pode ser classificada como de curto ou longo prazo, conforme o tempo de retenção da informação, sendo a memória de curto prazo uma compilação que representa o estado atual do meio e que é atualizado a cada mudança deste, já a memória de longo prazo um conjunto de informações que fica armazenado por um longo período ou permanentemente (HAYKIN, 2005).

Em modelos conexionistas a representação do tempo pode ser de forma dita explícita ou implícita. A representação é dita explícita quando alcançada através da “paralelização do tempo”, ou seja, representando o tempo como uma nova dimensão da entrada. Já a representação é considerada implícita quando o tempo é representado pelo efeito que tem sobre o processamento (ELMAN, 1990). A representação do tempo em um modelo conexionista é o que o torna dinâmico e apto ao processamento de padrões ou series temporais. Redes neurais artificiais estáticas possuem apenas memória considerada como de longo prazo, obtida por um processo de aprendizagem supervisionado, no qual a informação fica armazenada nos pesos sinápticos da rede. Já as redes neurais dinâmicas possuem memórias de curto prazo obtidas através de atrasos de tempo em conexões anteroalimentadas (do inglês *feedforward*) ou conexões recorrentes, correspondendo respectivamente à representação explícita e implícita do tempo (HAYKIN, 2005).

2.3.1 Redes anteroalimentadas atrasadas no tempo

As redes alimentadas adiante atrasadas no tempo (do inglês *Time Lagged Feed-forward Network* ou TLFN) ou simplesmente redes neurais atrasadas no tempo RNAT, utilizam a representação explícita do tempo e são inerentemente estáveis quando sua memória de curto-prazo também o é. Em sua forma mais simples, a TLFN focada é exemplificada pelo modelo da Figura 6.

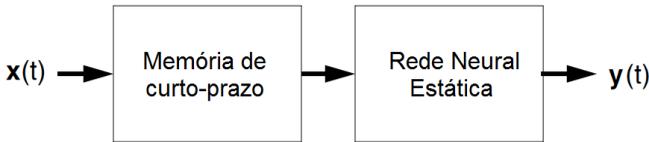


Figura 6 – Modelo básico para processamento temporal.

Fonte: Modificado de (MOZER, 1994).

A memória de curto-prazo pode assumir diversas formas, assim um diagrama genérico para este bloco seria o demonstrado na Figura 7.

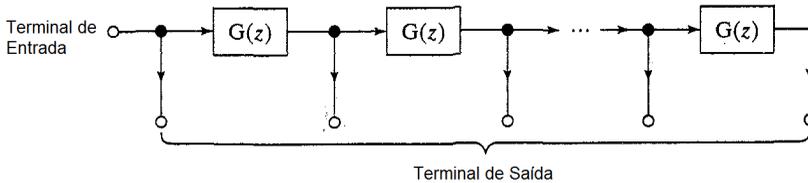


Figura 7 – Diagrama genérico para memória de curto-prazo.

Fonte: Modificado de (HAYKIN, 2005).

Conforme a taxonomia empregada em Mozer (1994), a memória de curto-prazo pode ter a forma de memória de linha de atrasos, memória de traço exponencial, memória gamma e outros, dependendo da função de transferência empregada no núcleo gerador “ $G(z)$ ”. Sendo que a forma mais comumente empregada é a memória de linha de atrasos (do inglês *Tapped Delay Line*), onde “ $G(z)$ ” é igual a “ z^{-1} ”, por sua simplicidade e baixo custo computacional.

Utilizando uma memória de linha de atrasos e um perceptron de múltiplas camadas a TLFN focada assume a forma apresentada na Figura 8. É importante notar que mais de uma variável de entrada pode ser empregada, cada qual armazenada em sua linha de atrasos.

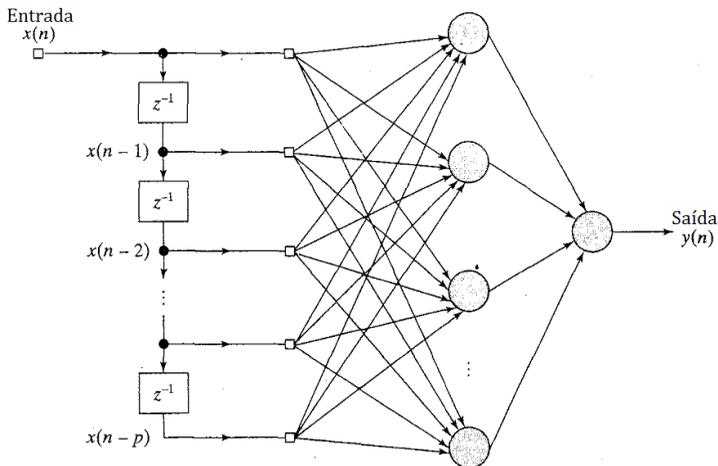


Figura 8 – TLFN focada com uma entrada (*bias* omitido).

Fonte: Modificado de (HAYKIN, 2005).

Cada conexão da rede neural apresentada na Figura 8 possui peso próprio, logo a equação da saída de um neurônio na primeira camada oculta em função das entradas é expressa por:

$$y_j(n) = \varphi\left(\sum_{i=1}^{m_0} \sum_{l=0}^p w_{ji}(l)x_i(n-l) + b_j\right) \quad (2.6)$$

Onde:

y = valor de saída;

φ = função de ativação do neurônio;

w = peso da conexão;

x = variável de entrada;

n = instante de tempo atual;

b = desvio (do inglês *bias*);

m_0 = total de entradas;

p = total de unidades de atraso por entrada;

j = índice do neurônio;

i = índice da variável de entrada;

l = índice da unidade de atraso.

Conforme Haykin (2005) *apud* (SANDBERG e XU, 1997), o teorema do mapeamento míope universal assegura que esta arquitetura é capaz de aproximar arbitrariamente bem qualquer mapa dinâmico míope

(memória com decaimento uniforme) invariante a deslocamento (sendo $y(n)$ saída para uma entrada $x(n)$ no tempo n e n_0 um inteiro, a saída para $x(n-n_0)$ é $y(n-n_0)$).

A TLFN focada é uma rede neural simples de projetar e suficientemente capaz para o processamento de mapas dinâmicos em ambientes estacionários (invariantes a deslocamento), além de poder ser treinada pelo algoritmo de retropropagação (do inglês *backpropagation*) padrão. Entretanto, para processamento de mapas dinâmicos em ambientes não estacionários (variantes a deslocamento) a memória de curto-prazo precisa estar em camadas internas da rede como na TLFN distribuída (HAYKIN, 2005).

A rede neural TLFN distribuída baseia-se no filtro neural focado de múltiplas entradas, apresentado na Figura 9. Neste tipo de rede cada unidade de neurônio da mesma é um filtro neural como este. Outra forma de entender a estrutura de uma TLFN distribuída é substituir todas as conexões sinápticas de uma rede neural estática por um filtro de resposta infinita ao impulso (FIR – *Finite Impulse Response filter*). Apesar de ser uma rede neural com maior poder de processamento em relação à TLFN focada, a TLFN distribuída requer o uso de algoritmos de treino mais elaborados como retropropagação temporal (do inglês *Temporal Backpropagation*) (HAYKIN, 2005).

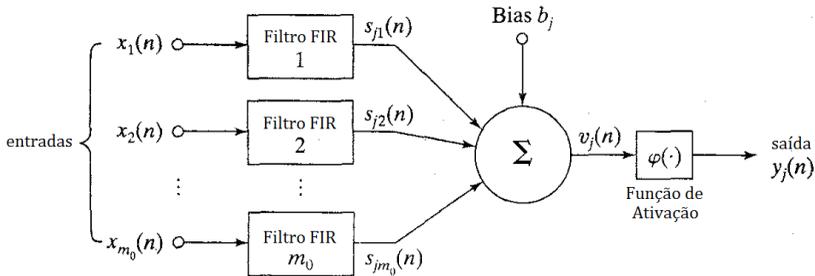


Figura 9 – Filtro neural focado com múltiplas entradas.

Fonte: Modificado de (HAYKIN, 2005).

2.3.2 Redes neurais recorrentes

Conforme Elman (1990), a representação explícita do tempo apresenta três desvantagens principais que são a necessidade de armazenar sequencias de entrada, restrição da duração dos padrões e fraca capacidade de distinguir posição temporal relativa da absoluta. O

comprimento de memória (número de unidades de atraso) em uma TLFN deve ser suficientemente grande para representar o maior padrão a ser armazenado, dificultando ainda a representação de padrões de comprimentos diferentes. Além disto, instancias defasadas no tempo de vetores de entrada de um mesmo padrão (classe) como “1 0 0”, “0 1 0” e “0 0 1” são espacialmente diferentes. Esta limitação pode ser superada através de treinamento, possibilitando a rede considerar as três amostras como o mesmo padrão. Entretanto, tal estrutura temporal não será generalizada para novos padrões. Estas questões sugerem o uso da representação implícita do tempo.

Em Haykin (2005) a representação implícita é sugerida através de dois modos principais, realimentação local ou global sendo ambas consideradas recorrentes. A realimentação é considerada local quando ocorre da saída de um dado neurônio para ele mesmo e global ao retornar para outro neurônio, a Figura 10 exemplifica ambas as situações.

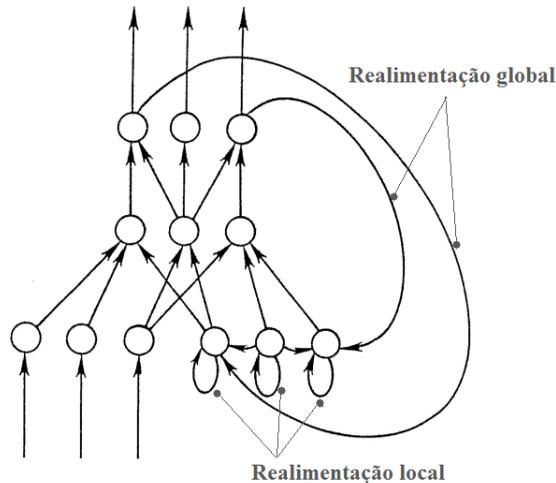


Figura 10 – Rede neural recorrente com realimentação local e global.

Fonte: Modificado de (JORDAN, 1986).

As redes neurais recorrentes permitem o aprendizado da estrutura temporal relativa dos padrões, entretanto requerem um cuidado adicional com relação à estabilidade da rede. Redes neurais alimentadas adiante são inerentemente estáveis, quando incorporadas realimentações o sistema pode ficar instável e prejudicar o aprendizado. No entanto, se

aplicadas corretamente realimentações podem reduzir a necessidade de memória de curto-prazo e oferecer melhores soluções para certas aplicações como a equalização adaptativa de canais de comunicação e o processamento de voz (HAYKIN, 2005).

3 MATERIAIS E MÉTODOS

3.1 AQUISIÇÃO DOS DADOS

Os dados aplicados às redes neurais artificiais deste estudo foram extraídos de quatro vídeos de experimentos com ratos em arena de campo aberto gravados durante o estudo (CRISPIM, 2011). A partir deste estudo, foram também selecionados os descritores de imagem relevantes a detecção dos eventos comportamentais.

3.1.1 Vídeos dos experimentos

Foram utilizados neste trabalho quatro dos vídeos de experimentos capturados durante o estudo (CRISPIM, 2011), cada vídeo de um rato tratado com veículo.

3.1.1.1 Animais

Os ratos experimentados eram da espécie *Rattus norvegicus*, linhagem *wistar*, todos fêmeas com massa corporal entre 250-300g. Estes animais foram condicionados e tratados conforme o protocolo PP00254 aprovado em 16/07/2008 pela Comissão de Ética no Uso de Animais (Ofício nº 128/CEUA/PRPe/2008).

3.1.1.2 Vídeos

Os vídeos foram gravados com uma câmera filmadora SONY DCR-SR45 (configurada para adquirir vídeos a 30 quadros por segundo), posicionada perpendicular em relação ao centro da arena. Após aquisição, os vídeos foram convertidos do formato MPEG2 para AVI (usando o software VirtualDub versão 1.78).

3.1.1.3 Experimentos

Foram realizadas três rodadas de experimentos com seis animais experimentalmente ingênuos sob um esquema amostral de quadrado latino.

Cada rodada consistindo da injeção de veículo (salina) em dois animais, cafeína 2mg/kg em outros dois animais e cafeína 6mg/kg nos dois restantes. Após injeção os animais aguardavam por 10 minutos fora do ambiente de teste (arena de campo aberto pintada com tinta acrílica preta). Terminado o intervalo, os sujeitos eram colocados na arena e filmados por no mínimo 10 minutos.

3.1.2 ETHOWATCHER[®] – calibração e *activity tracking*

O módulo de *activity tracking* do software ETHOWATCHER[®] (ferramenta computacional desenvolvida no laboratório de Bioengenharia do Instituto de Engenharia Biomédica e laboratório de Neurofisiologia Comparada sediados na Universidade Federal de Santa Catarina, disponível no site <http://www.ethowatcher.ufsc.br/>, descrições em (CRISPIM *et. al.*, 2008), (CRISPIM *et. al.*, 2010) e pormenorizada em (CRISPIM *et. al.*, 2012) foi usado para processar os vídeos e extrair dados que descrevem as posturas e movimentos realizados pelo objeto (rato). Neste processamento, a cada quadro do vídeo o objeto de interesse é segmentado em relação à cena de fundo e do mesmo, são extraídos valores de variáveis morfológicas e cinemáticas.

A segmentação do objeto se dá pela subtração do plano de fundo por imagem estática, aplicação de limiar e filtro de faixa de intensidade de cor, conforme descrito em Crispim (2011). Todos os três procedimentos podem ser configurados pelo usuário da ferramenta, sendo os dois primeiros imprescindíveis.

Na Figura 11, são apresentadas as componentes do processo de segmentação por subtração do plano de fundo. Na Figura 11 “a”, a imagem estática (plano de fundo) e em “b” a cena sob análise com o objeto. Em “c”, “d” e “e”, são demonstrados resultados da segmentação com diferentes limiares de intensidade de cor, respectivamente do limiar mais baixo ao mais alto. É notório que um baixo limiar deixa muito ruído que afetarão medidas da forma e posição do objeto, de mesmo modo que um limiar excessivamente alto cortará partes do objeto, causando efeitos semelhantes.

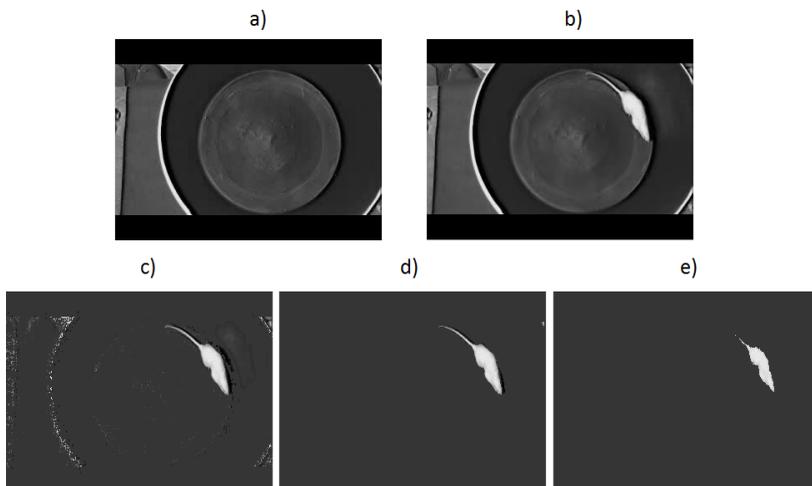


Figura 11 – Componentes do processo de segmentação: a) Plano de fundo; b) Cena sob análise; c) Resultado da segmentação com limiar baixo; d) Resultado da segmentação com limiar intermediário; e) Resultado da segmentação com limiar alto.

3.1.3 Associação da etografia com activity tracking

Para cada uma das quatro filmagens de experimentos, foi feito um registro dos comportamentos realizados pelo animal, conforme o catálogo de categorias estabelecido no Quadro 2. As definições das categorias deste quadro foram elaboradas com base em Bindra e Spinner (1958), bem como em Crispim (2011).

Categoria	Código	Definição
Imobilidade	I	Quando o animal está sentado ou deitado sem movimento de qualquer de seus membros ou cauda, podendo haver tremores e acomodação da cabeça se estiver dormindo.
Locomoção	L	Movimento contínuo e coordenado das patas dianteiras e traseiras do animal, afastando o corpo do animal de sua posição de origem.
Exploração Vertical	V	Movimentos do focinho do animal no sentido ventral/dorsal acompanhado da retirada das suas patas dianteiras do solo.
Auto-limpeza	A	Caracterizado por três movimentos principais:

		limpeza caudal e ventral, através do encurvamento do dorso animal, aproximando o plano coronal do plano transversal, conseqüentemente sua cabeça ao ventre; limpeza do dorso, através do movimento de aproximação do plano sagital ao transversal, usando a boca e/ou patas; limpeza da cabeça, com movimento de raspagem de suas patas sobre sua cabeça e focinho.
Exploração Horizontal	H	Movimentos do focinho do animal no sentido ventral/dorsal quando suas patas estão apoiadas no solo, podendo haver movimentos de cabeça e pescoço em sentido de aproximação do plano sagital ao transversal, bem como movimentos dos mesmos na aproximação do plano coronal ao transversal. São também incluídos nesta categoria eventos que não se encaixam na descrição das demais categorias.

Quadro 2 – Catálogo de comportamentos: definições e abreviações.

Um exemplo das categorias, imobilidade e locomoção são demonstrados na Figura 12; e na Figura 13 exemplos das categorias exploração vertical e auto-limpeza. Como a categoria exploração horizontal foi definida de forma muito abrangente, incluindo eventos não encaixados em qualquer das outras categorias, exemplos da mesma não foram apresentados.

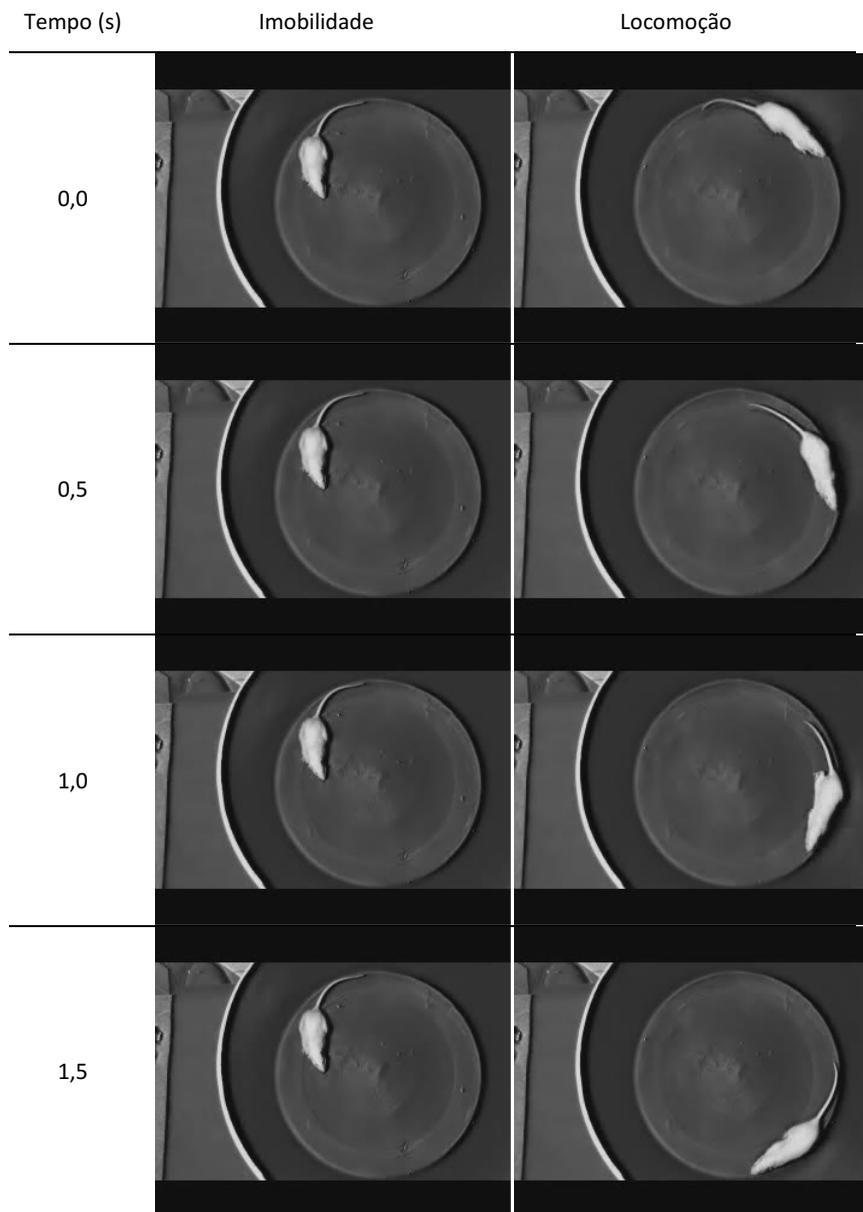


Figura 12 – Sequencias de quadros exemplo de imobilidade e locomoção.

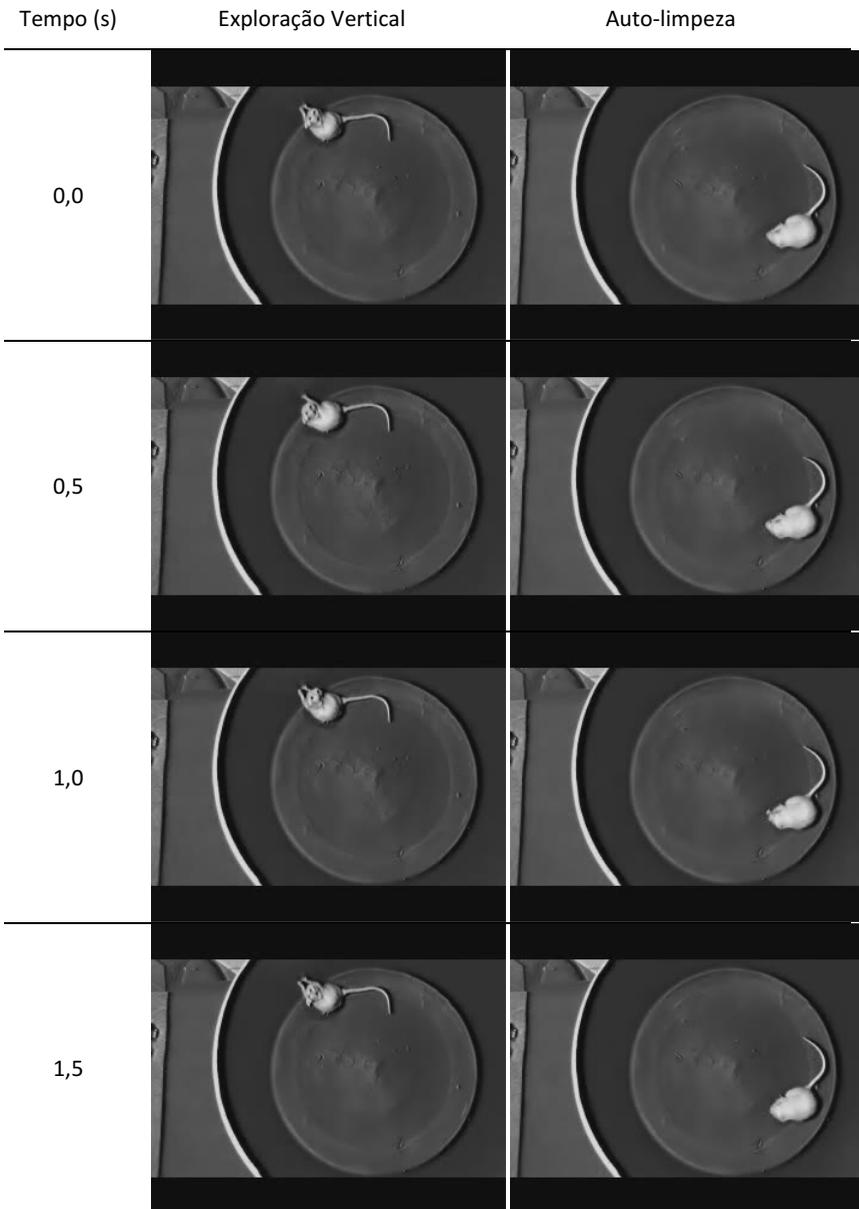


Figura 13 – Sequencias de quadros exemplo de Exploração Vertical e Auto-limpeza.

Com base nos registros comportamentais, foram estabelecidas as saídas esperadas da rede neural artificial durante o aprendizado supervisionado. Cada padrão de entrada da rede neural artificial é formado por dados de “n” quadros do vídeo (dados armazenados nas linhas de atrasos) e cada quadro, por sua vez, relacionado à categoria de comportamento que o observador humano registrou.

O valor máximo foi atribuído à saída esperada de uma dada categoria, quando o quadro correspondente ao quadro atual observado pela rede, foi classificado sob tal categoria pelo observador humano e valor mínimo para a saída esperada das demais categorias. Tal atribuição de valores é exemplificada pela Tabela 2, onde um registro hipotético de uma transição entre os comportamentos “Locomoção” e “Exploração Vertical” é apresentado. Neste exemplo são demonstradas as saídas convertidas conforme a finalidade da rede neural alvo do treinamento. Em uma rede treinada para detecção de “Locomoção”, o valor máximo “1” é atribuído aos quadros em que foi registrada “Locomoção” e “-1” onde foram registradas “Exploração Vertical” como demais categorias (caso houvesse no registro).

Tabela 2 – Conversão do registro em saída esperada para treinamento supervisionado

Quadro	Comportamento Registrado	Saída convertida		
		Locomoção	Exploração Vertical	Outras Categorias
433	Locomoção	1	-1	-1
434	Locomoção	1	-1	-1
435	Locomoção	1	-1	-1
436	Exploração Vertical	-1	1	-1
437	Exploração Vertical	-1	1	-1

NOTAS: Saída máxima “1” e saída mínima “-1”.

No catálogo de comportamentos foram definidas cinco categorias. No entanto, foram estudadas redes neurais artificiais para detecção de somente quatro destas. A categoria comportamental

“Exploração Horizontal” foi definida principalmente a fim de que no decorrer do registro, ao observar atividades que não se enquadram nas demais categorias ainda houvesse registro. Por tal razão, ao treinar redes para a detecção das outras categorias “Exploração Horizontal” sempre seria convertida em valor mínimo (não pertence a categoria procurada).

3.1.4 Descritores do objeto segmentado

Dentre os descritores reportados pelo *activity tracking*, foram usados como parâmetros de entrada da rede neural os dados dos descritores identificados em Crispim (2011) como estatisticamente significantes para a classificação dos eventos comportamentais, sendo eles o “número de pixels do animal”, “distância percorrida pelo centroide (em pixels)”, “comprimento médio do animal (em pixels)” e “número de pixels alterados a cada quadro”.

3.1.4.1 Número de pixels do animal

Do processo de segmentação de objeto, descrita no item 3.1.2, resultam pixels na cor definida como plano de fundo e pixels mais claros, que são contatos para definir o número de pixels que pertencem ao objeto (animal) (CRISPIM, 2011).

3.1.4.2 Número de pixels alterados a cada quadro

Diferente das técnicas usuais de computação de pixels alterados a cada quadro, que calculavam a diferença da área (número de pixels do objeto) a cada quadro. O descritor número de pixels alterados a cada quadro utilizado neste trabalho é computado com base na variação de intensidade de cor dos pixels do objeto (animal) no quadro analisado em relação a seu precedente (CRISPIM, 2011).

3.1.4.3 Distância percorrida pelo centroide

A distância percorrida pelo centroide do objeto é calculada como, a distância euclidiana em pixels entre a estimativa de sua coordenada cartesiana no quadro sob análise e sua coordenada estimada no quadro

anterior. Onde as coordenadas no plano cartesiano do centroide do objeto são calculadas, respectivamente, pela média das ordenadas e média das abscissas, de cada um dos pixels considerados como parte do objeto após sua segmentação (CRISPIM, 2011).

3.1.4.4 Comprimento médio do animal

Retas com inclinação no intervalo de -1 a -89° e 1 a 89° em relação ao eixo das ordenadas, são calculadas do centroide do objeto até alcançar pixels da imagem que estão na cor do plano de fundo (que não fazem parte do objeto). Uma média das dez maiores retas calculadas define o descritor comprimento médio do animal (CRISPIM, 2011).

A Figura 14 adiante ilustra a metodologia empregada no cálculo do descritor comprimento médio do animal.

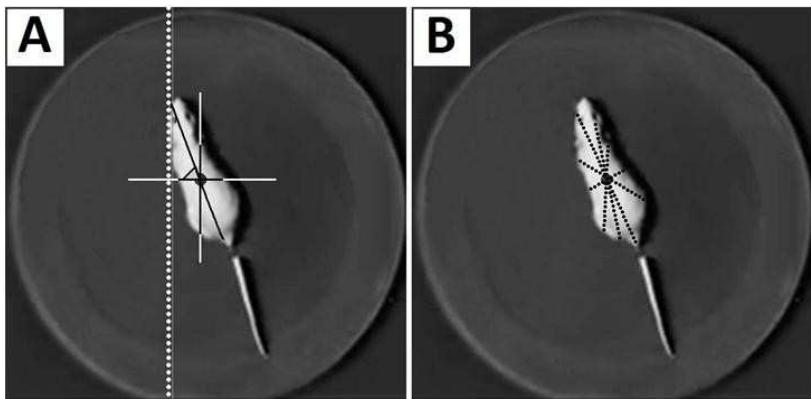


Figura 14 – Método de estimação do comprimento do animal: (A) Cálculo da reta para um ângulo do intervalo; e (B) Exemplo de conjunto de retas calculadas pelo método.

Fonte: (CRISPIM, 2011).

3.2 TLFN FOCADA – TREINAMENTO E VALIDAÇÃO

Conforme exposto na fundamentação teórica, a rede neural artificial do tipo TLFN (*Time Lagged Feedforward Network*) focada, é uma rede neural artificial com propriedade dinâmica aportada pelos seus atrasos explicitamente dispostos em suas entradas. Diferente das redes neurais artificiais do tipo TLFN distribuídas, a representação do tempo

empregada na arquitetura focada a prepara apenas para lidar com ambientes estacionários (invariantes no tempo), ou sob variação negligenciável, o que é uma desvantagem em relação à arquitetura de tempo distribuído. Sua vantagem, porém é a facilidade de desenvolvimento advinda da possibilidade de treiná-la com o algoritmo de retropropagação clássico, desde que a parte estática da rede seja do tipo perceptron ou perceptron de múltiplas camadas (HAYKIN, 2005). Esta arquitetura foi escolhida por ser a mais simples dentre as adequadas ao problema da detecção de comportamentos, além de permitir o estudo do comprimento de memória e estrutura dos padrões comportamentais. A parte estática da rede neural utilizada foi o perceptron com múltiplas camadas, as entradas, uma camada oculta e a camada de saída. Foram feitas simulações com memórias de quatorze, vinte e nove, e também cinquenta e nove atrasos na camada de entrada. Para cada comprimento de memória foi feita a escolha da taxa de aprendizagem e número de neurônios na camada oculta. Isto a fim de que ao comparar a influência destes diferentes comprimentos de memória no desempenho de detecção, cada um estivesse aproximadamente em sua parametrização mais adequada. Determinados os comprimentos de memória e demais parâmetros para cada comportamento, estes detectores foram testados e seus desempenhos comparados aos desempenhos da classificação de comportamentos por humanos.

Estes comprimentos de memória foram escolhidos por totalizarem um período de observação de aproximadamente meio, um e dois segundos, respectivamente. O intervalo de um segundo apresentou-se adequado à detecção dos comportamentos estudados na metodologia empregada em Crispim (2011). Assim, os outros dois comprimentos foram escolhidos em medidas das quais acreditávamos serem significativamente acima e abaixo deste intervalo a fim de identificar a influência deste parâmetro.

O vídeo de experimento do rato quatro tratado com veículo foi separado para avaliar o comportamento da rede na classificação automática dos comportamentos em situação real (conjunto de teste). Os três outros vídeos de experimentos (ratos um, dois e três tratados com veículo), foram usados na validação cruzada com 10-dobras para seleção e comparação do desempenho dos modelos (diferentes configurações de taxa de aprendizagem, neurônios na camada oculta e atrasos nas entradas). Em cada época do treinamento uma dobra era usada como conjunto de validação e as demais como conjunto de treino.

3.2.1 Parâmetros da rede e características do treinamento

O algoritmo de treino utilizado nas simulações foi a retropropagação do erro com termo momento, descrito em Haykin (2005). Além disto, o critério de parada utilizado no treinamento da rede foi baseado na generalização, proporcionado pelo o método de treinamento com parada antecipada.

Este método de treino consiste em a cada época do treino, apresentar os padrões para a rede, atualizar os pesos, apresentar os padrões de validação e fazer a mensura do erro médio quadrático na época. Inicialmente os erros médios quadráticos dos conjuntos de treino e validação diminuem a cada época, no entanto, a partir de um ponto o erro médio quadrático de treino diminui e o de validação começa a aumentar, quando o treino é encerrado para manter a capacidade de generalização da rede (HAYKIN, 2005). Este procedimento foi realizado durante o treinamento com cada conjunto de dobras da validação cruzada, medindo não somente o valor médio na época do erro médio quadrático, como também o índice AUC e índices Kappa para trinta limiares aplicados. Além disto, o treinamento foi interrompido ao atingir três mil épocas.

Os parâmetros fixos nos treinamentos das redes estão demonstrados no Quadro 3 abaixo.

Parâmetros	Camada de Entrada	Camada Oculta	Camada de Saída
Unidades	4	Variável	1
Função de ativação	-	$y(x) = 1,7159.tanh((2/3).x)$	
Parâmetros Gerais			
Atualização dos pesos	Sequencial (Estocástico)		
Pesos iniciais	Aleatórios entre -0,01 e 0,01		
Termo momento	0,8		

Quadro 3 – Parâmetros da rede, inicialização e treinamento.

Conforme heurística de treino sugerida em LeCun (1998) e Haykin (2005), as taxas de aprendizagem foram definidas de forma independente para cada neurônio, em valor inversamente proporcional à raiz quadrada do número de entradas para tal neurônio. Como os

neurônios de uma mesma camada possuem o mesmo número de entradas e atrasos, a diferença na taxa de aprendizagem ficou apenas entre diferentes camadas. A função de ativação e suas constantes também foram definidas baseadas nas heurísticas das mesmas literaturas, já o termo momento foi definido empiricamente. Foi utilizada apenas uma unidade de saída, que corresponderia à existência ou não do padrão comportamental procurado. A cada época do treinamento, os padrões apresentados durante o treino foram embaralhados, mas mantendo a ordem dos quadros na janela de cada padrão.

3.2.1.1 Codificação e validação do algoritmo

As redes neurais artificiais aplicadas neste trabalho foram desenvolvidas na linguagem de programação C/C++ e validadas através do uso destas no experimento computacional 13.5 proposto em Haykin (2005).

Neste experimento é sugerida a aplicação de uma TLFN focada na predição da série temporal da Equação 3.1, que representa um sinal modulado em frequência.

$$x(n) = \text{sen}(n + \text{sen}(n^2)) \quad (3.1)$$

Onde “n” é um número inteiro representando tempo discreto. A rede neural desenvolvida foi configurada, treinada e validada com as condições e características indicadas no problema. Os resultados obtidos são apresentados na Figura 15, sendo que o valor médio do erro quadrático na validação apresentada é $35,1 \cdot 10^{-3}$.

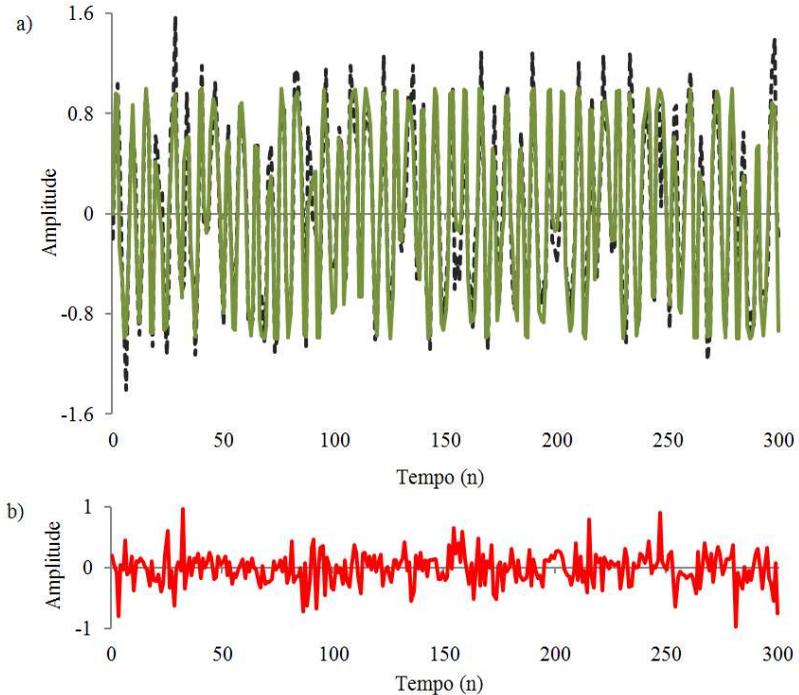


Figura 15 – Aplicação da TLFN focada desenvolvida no experimento computacional de predição de um passo. a) Sobreposição da forma de onda atual (traçado contínuo) e predita (traçado tracejado). b) Forma de onda do erro de predição.

3.2.2 Dados de entrada e pré-processamento

As variáveis de ambiente das redes neurais treinadas neste trabalho foram os descritores apresentados na seção 3.1.4 e seus valores armazenados nas memórias de entrada. A fim de tornar adequada uma futura comparação entre a rede utilizada neste trabalho e a abordagem empregada em Crispim (2011), os dados de entrada utilizados no treinamento foram extraídos dos mesmos trechos dos vídeos dos ratos um, dois e três. Ainda em Crispim (2011) foram utilizados não somente estes três ratos como também o rato quatro no treinamento e validação, mas neste trabalho este ficou de fora do treinamento a fim de ser utilizado inteiramente no teste como dito anteriormente. Além de utilizar os mesmos trechos dos vídeos, foram também utilizados os mesmos

parâmetros de calibração do módulo *activity tracking* do software ETHOWATCHER[®]. Os trechos de vídeos utilizados no treinamento não incluíram as transições entre os comportamentos, tendo sido incluídos apenas os trechos onde claramente os comportamentos estudados estavam sendo realizados. Esta não inclusão de transições é exemplificada na Figura 16. Neste exemplo a rede neural teria linhas de atrasos com cinco valores (quatro unidades de atraso por variável de entrada), as linhas tracejadas delimitando os dados de entrada da rede e linhas contínuas enquadrando o quadro atual observado e sua respectiva classificação no registro. Supondo uma apresentação ordenada dos exemplos, os dados e registro em verde seriam apresentados primeiramente, em seguida seria apresentado o exemplo enquadrado em preto e por fim o exemplo em vermelho (havendo pulado a transição, quadros de oito a onze).

Quadro	Variáveis Descritoras Observadas				Registro
1	1275	2	88	1066	Locomoção
2	1250	0	85	1186	Locomoção
3	1240	2	85	1191	Locomoção
4	1253	1	82	1236	Locomoção
5	1219	2	79	1208	Locomoção
6	1203	1	76	1097	Locomoção
7	1199	1	74	1134	Locomoção
8	1139	3	64	1181	Exploração Vertical
9	1098	1	68	1182	Exploração Vertical
10	1119	4	61	1155	Exploração Vertical
11	1070	0	58	1048	Exploração Vertical
12	1059	1	60	892	Exploração Vertical
13	1057	1	66	546	Exploração Vertical

Figura 16 – Apresentação de exemplos sem transição entre comportamentos.

O não uso das transições no treinamento supervisionado teve o intuito de evitar que durante este tal supervisor direcionasse os pesos da rede em padrões mais duvidosos (padrões classificados erroneamente com maior frequência pelo próprio agente supervisor).

O número de exemplos de cada categoria comportamental no conjunto de dados usado na validação cruzada é significativamente desigual, pois além do animal observado executar mais vezes um comportamento em relação a outros, também suas durações são diferentes. Na Figura 17 é ilustrada a composição percentual dos comportamentos no conjunto de dados da validação cruzada usado para treinar as redes de detecção de locomoção com memória de trinta valores. Com base na distribuição dos dados para cada classe e na revisão da seção 2.4.1, que pondera sobre conjuntos de treino desbalanceados, foi adotada uma metodologia de sub-amostragem. Entre a sub-amostragem e a sobre-amostragem, a primeira foi escolhida em vista do conjunto de treino possuir uma quantidade massiva de exemplos, tornando o treino mais rápido após a sub-amostragem.

Antes de separar as dobras, foi retirada aleatoriamente uma quantidade de exemplos de cada conjunto de dados de treino e validação, suficiente para que 50% dos exemplos no conjunto pertençam à categoria a ser detectada pela rede em questão, ficando os outros 50% dos dados distribuídos entre classes consideradas atividade de fundo por tal rede.

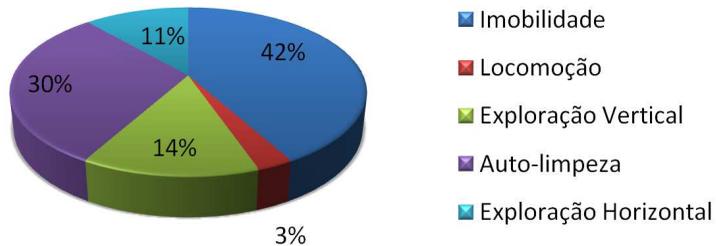


Figura 17 – Composição percentual dos comportamentos no conjunto de validação cruzada usado nas redes para detecção de locomoção, antes da sub-amostragem.

3.2.3 Estimação e comparação de desempenho

Embora para cada uma das redes treinadas tenha sido medido o desempenho em relação aos índices Kappa e AUC, o segundo foi utilizado como referência para determinar quais os melhores parâmetros de cada detector. Este índice foi escolhido como referência por ser uma

ponderação dos desempenhos medidos ao aplicar cada um dos limiares, diferente do Kappa que neste trabalho foi utilizado o maior valor encontrado sob os limiares aplicados. Como dito na seção 2.2.4 estes índices de desempenho necessitam de uma referência, seja um padrão “ouro” (índice AUC) ou outro observador (índice Kappa). Neste trabalho foi utilizado como referência o registro comportamental realizado pelo autor do catálogo apresentado na seção 3.1.3, também utilizado para treinar outros observadores.

3.2.4 Seleção da taxa de aprendizagem

Como sugerido em LeCun (1998) e Haykin (2005), as taxas de aprendizagem devem ser inversamente proporcionais ao número de conexões de um neurônio. Assim, tanto a taxa de aprendizagem na camada oculta quanto na camada de saída, foram definidas sob a seguinte equação:

$$\eta = fator \cdot \left(\frac{1}{\sqrt{m_o \cdot (p+1)}} \right) \quad (3.2)$$

Onde:

- η = taxa de aprendizagem no neurônio;
- *fator* = constante de redução da taxa de aprendizagem;
- m_o = número de neurônios na camada anterior;
- p = número de atrasos em cada entrada;

Os valores para o “*fator*” foram definidos inicialmente de forma empírica, começando com 10^{-6} e decrescendo uma casa decimal a cada simulação até encontrar redução no valor médio do desempenho (em relação ao valor médio do índice AUC). Quando encontrada redução no desempenho já no primeiro decréscimo do “*fator*”, o processo inverso foi realizado (acrécimo do “*fator*” em uma casa decimal a cada simulação). Nestas simulações foram utilizados dois neurônios na camada oculta.

3.2.5 Número de neurônios na camada oculta

Fixando a taxa de aprendizagem na camada oculta e acrescentando neurônios a cada nova simulação o número de neurônios

foi definido. O “*fator*” utilizado na Equação 3.2 foi o que resultou no maior valor médio do índice AUC para a rede de cada categoria. Assim, o limite no número de neurônios a ser adicionado foi definido no ponto em que não foi observado ganho no desempenho com o aumento do número de neurônios. A taxa de aprendizagem na camada de saída, por sua vez, variou em função do número de neurônios na camada oculta conforme expresso pela Equação 3.2 com valor fixo para “*fator*”.

3.3 TLFN FOCADA – TESTE

O teste consistiu não somente em avaliar as redes treinadas (em esquema de validação cruzada 10-dobras com vídeos dos ratos um, dois e três tratados com veículo) na detecção dos comportamentos imobilidade, locomoção, exploração vertical e auto-limpeza ao expô-las a um vídeo de experimento completo (rato quatro, também tratado com veículo), mas também comparar o efeito sobre o desempenho destes classificadores ao usar os diferentes comprimentos de memória.

Nesta etapa, para cada categoria comportamental e cada comprimento de memória foram tomadas as dez redes (10-dobras) treinadas com a melhor combinação de taxa de aprendizagem e neurônios na camada oculta e aplicadas à detecção dos comportamentos no vídeo de teste (extração de dados dos vídeos de teste, apresentação às redes, propagação e classificação). Novamente foram registrados os índices de desempenho Kappa e AUC, determinados a partir da aplicação de trinta limiares.

3.4 REGISTRO ETOGRÁFICO – OBSERVADOR HUMANO

Erros na medida resultante de uma etapa de coleta de dados de um estudo comportamental são compostos não somente por influência do método aplicado, mas também por efeitos inerentes ao observador.

Conforme Lehner (1996) a confiabilidade de um estudo etológico é determinada por quão este é reprodutível. Já a reprodutibilidade seria composta por quão confiável é que um observador seja capaz de repetir suas medidas (confiabilidade intra-observador) e quão confiável é que outros observadores (confiabilidade inter-observador) também as obtenham ao repetir sob as mesmas condições. Nem sempre um observador vê, ouve ou percebe o que gostaria durante um experimento. Suas observações são influenciadas por fatores como experiência no uso

das metodologias empregadas, equipamentos usados e estado de humor. Comparações de registros em tempos diferentes realizados por um mesmo observador em relação a um mesmo experimento permitem determinar o quanto este é estável (preciso) em suas medidas e, por consequência o quanto afetará a confiabilidade do estudo em que está envolvido. Ainda que um observador seja preciso em suas medidas (medindo sempre da mesma forma) pode apresentar-se inexato (e.g. sempre atribuindo comportamentos de uma categoria a outra ou ignorando um mesmo trecho de dado evento). A diferente atenção empregada por cada observador, suas capacidades de ver e ouvir, bem como seus níveis de treinamento faz com os erros cometidos em cada trecho sejam diferentes. Assim, a confiabilidade inter-observador acaba sofrendo grande influência por erros de exatidão (desvio médio em relação ao ponto de acerto ideal) dos observadores. Como consequência, ao medir as confiabilidades intra e inter observador é obtida uma boa indicação da precisão e exatidão do procedimento de coleta de dados aplicado.

3.4.1 Treinamento dos observadores

Participaram deste estudo quatro estudantes, dois de graduação e dois de pós-graduação. Inicialmente (antes deste estudo), nenhum dos quatro possuía qualquer experiência no registro comportamental de ratos.

O observador considerado mais experiente neste estudo era um dos estudantes de pós-graduação. Este recebeu tal experiência ao elaborar o catálogo definido na seção 3.1.3 e fazer o registro comportamental de doze experimentos (não utilizados neste trabalho, mas adquiridos no mesmo protocolo experimental), supervisionado por observadores já experientes. Após tal experiência este observador treinou os demais observadores. Tal treinamento consistiu na apresentação do catálogo, orientações sobre o mesmo, bem como a execução supervisionada do registro de um experimento com o rato número um.

3.4.2 Finalidade dos vídeos

Todos os quatro observadores realizaram dois registros do vídeo do rato número quatro (o mesmo utilizado no teste com a rede neural

artificial), com um intervalo de aproximadamente duas semanas. Estes registros foram utilizados para avaliar ambas as confiabilidades intra e inter observadores. Em Lehner (1996) é relatado que em um período de até dois dias, a percepção de um observador quanto a comportamentos observados não é alterada.

Quanto aos outros três vídeos (ratos número um, dois e três) utilizados para treinamento e validação das redes neurais artificiais, apenas um registro comportamental foi realizado (pelo observador mais experiente).

Todos os registros comportamentais realizados compreenderam todo o período em que o animal esteve na arena, aproximadamente nove minutos e quarenta e cinco segundos por vídeo.

3.4.3 Ferramentas e métrica de desempenho

Para realizar o registro cada observador baseou-se nas definições descritas no catálogo da seção 3.1.3 e utilizou o módulo *ethography* do ETHOWATCHER[®].

Como dito na seção 2.2.4, a métrica definida para avaliar a concordância (desempenho) de observadores humanos foi o índice Kappa (medido separadamente para cada classe comportamental). Um software desenvolvido para realizar avaliações de confiabilidade entre registros comportamentais foi utilizado para medir a concordância intra-observador de cada um, bem como todas as seis possíveis combinações de concordância inter-observador. Tal software foi desenvolvido por estudantes do próprio Laboratório de Bioengenharia do Instituto de Engenharia Biomédica (IEB-UFSC, Departamento de Engenharia Elétrica, EEL) da Universidade Federal de Santa Catarina.

4 RESULTADOS

4.1 DESEMPENHO DE DETECÇÃO – TLFN FOCADA

Nesta seção são apresentados os resultados da validação e teste das redes neurais treinadas para detectar as categorias comportamentais estudadas: imobilidade, locomoção, exploração vertical e auto-limpeza. Como mencionado anteriormente na seção 3.2, o processo de treino e validação ocorreu com partes dos vídeos de três ratos e o teste foi realizado sob o vídeo de um quarto rato. Na seção 4.2 será apresentado o resultado da avaliação do desempenho de observadores humanos na classificação dos comportamentos, também sob o vídeo deste quarto rato.

4.1.1 Validação

As medidas de desempenho apresentadas nesta seção como índice Kappa e AUC correspondem à média do maior valor do índice encontrado durante a validação para cada uma das dez dobras (a cada época do treinamento um valor de AUC é calculado, bem como um valor de Kappa para cada limiar aplicado).

4.1.1.1 Imobilidade

Conforme a metodologia descrita nas seções 3.2.4 e 3.2.5, as primeiras simulações foram com dois neurônios ocultos para memórias com quinze, trinta e sessenta valores, a fim de definir as taxas de aprendizagem mais adequadas. Nestas simulações as redes treinadas para detecção da categoria comportamental “Imobilidade” resultaram nos desempenhos indicados na Figura 18. Tais resultados indicam que a probabilidade de convergência para um ponto de maior desempenho médio na validação com relação à medida AUC ocorre ao utilizar um “*fator*” de 10^{-6} para um comprimento de memória de meio segundo (quatorze unidades de atraso mais valor atual) sob as condições estudadas. Já para um comprimento de memória de um segundo (vinte e nove unidades de atraso mais o valor atual) o “*fator*” que proporciona melhor desempenho é 10^{-5} . Por fim, o “*fator*” que aparentemente proporciona melhor desempenho quando utilizado um comprimento de

memória de dois segundos (amostras em sessenta instantes de tempo) foi 10^{-7} .

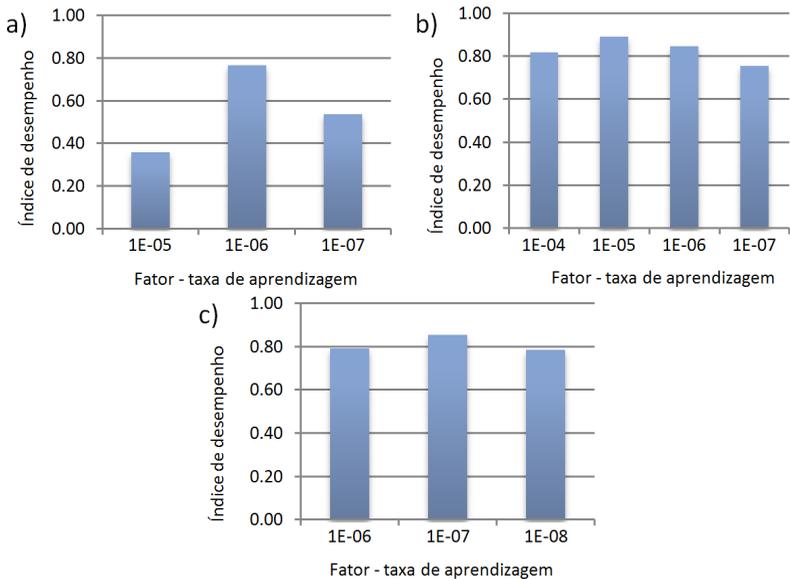


Figura 18 – Desempenho de validação na detecção da categoria comportamental “Imobilidade”, dois neurônios na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.

Seguindo com a metodologia definida, os valores de “*fator*” (constante que modifica a taxa de aprendizagem) que resultaram em um maior desempenho foram usados para definir o número de neurônios na camada oculta. Assim, o “*fator*” foi mantido fixo em 10^{-6} , 10^{-5} e 10^{-7} , respectivamente para redes com memória de quinze, trinta e sessenta valores. O resultado da simulação para definir o número de neurônios na camada oculta é demonstrado na Figura 19. Neste grupo de resultados o número de neurônios ocultos aproximadamente mais adequado em relação ao desempenho pelo índice AUC é encontrado como sendo dois, três e dois respectivamente para comprimentos de memória de quinze, trinta e sessenta valores. Um fato interessante a se observar ao analisar os resultados obtidos na Figura 18 e Figura 19 é que não somente o maior desempenho médio para detecção do comportamento estudado foi encontrado utilizando um comprimento de memória de trinta valores, mas também com tal comprimento foi observado menor perda de

desempenho em taxas de aprendizagem em torno do valor apontado como mais adequado.

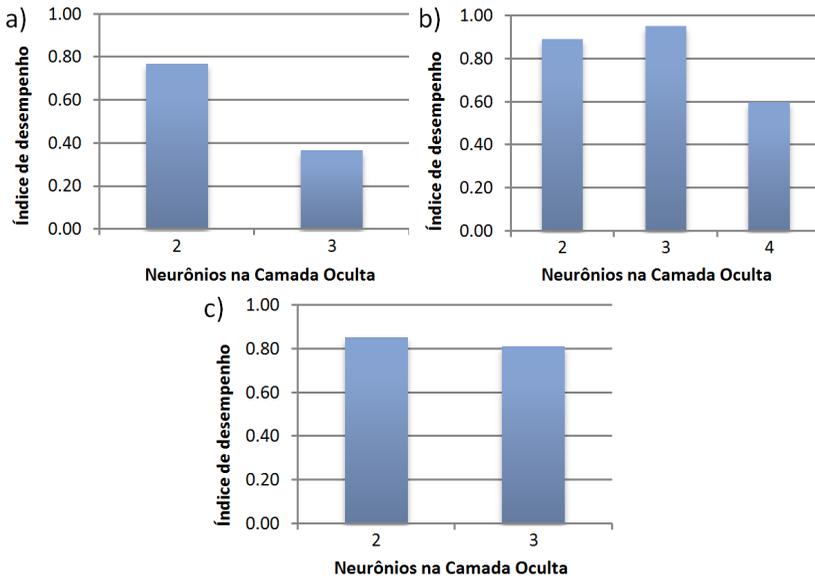


Figura 19 – Desempenho de validação na detecção da categoria comportamental “Imobilidade”, taxa de aprendizagem fixa na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.

A partir dos resultados das simulações apresentados é possível comparar o efeito do comprimento de memória para as redes treinadas na detecção de padrões da categoria “Imobilidade”. Um comparativo do desempenho dos diferentes comprimentos de memória (cada qual em sua melhor parametrização em relação à taxa de aprendizagem e neurônios na camada oculta) foi feito através das médias e erros padrão da média dos índices Kappa e AUC, ilustrado na Figura 20. As combinações dos parâmetros apresentados na Figura 20 são resumidas pela Tabela 3. A fim de determinar se um ou outro comprimento de memória resultava em desempenho estatisticamente maior ou menor, foi primeiramente realizado um teste One way ANOVA (Análise de variância de uma via). O resultado deste teste entre os comprimentos de memória e o índice Kappa foi $F(2,27)=12,52$ com $p<0,001$, para comprimentos de memória contra o índice AUC o resultado foi $F(2,27)=4,52$ com $p=0,020$. O teste estatístico *post hoc* de Duncan para

$p < 0,05$ dos comprimentos de memória em relação ao índice Kappa apontou diferença estatística entre os comprimentos de trinta valores e quinze valores com $p < 0,001$, bem como entre comprimentos de trinta e sessenta valores com $p = 0,003$. O mesmo teste, mas realizado entre comprimentos de memória e o índice AUC apontou diferença apenas entre comprimentos de memória de trinta e quinze valores com $p = 0,008$.

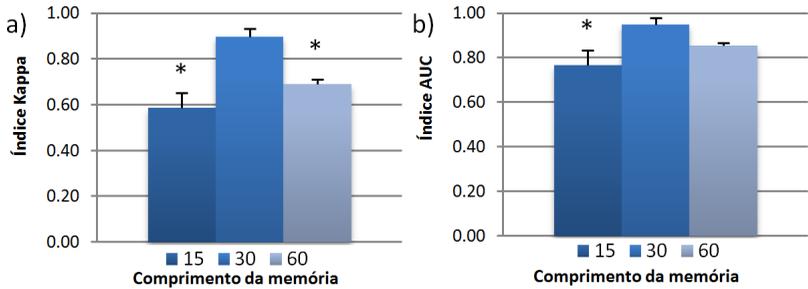


Figura 20 – Comparativo de desempenho de validação na detecção da categoria comportamental “Imobilidade”, melhores parâmetros taxa de aprendizagem e neurônios na camada oculta. Média e erro padrão da média dos índices a) Kappa e b) AUC. Valores marcados com “*” representam valores com diferença estatisticamente significativa para $p < 0,05$ em relação ao valor de desempenho para comprimento de trinta valores.

Tabela 3 – Melhores taxas de aprendizagem e neurônios na camada oculta na detecção de imobilidade para cada comprimento de memória

Comprimento de memória	Taxa de aprendizagem		Neurônios Oculcos
	Camada oculta	Camada de saída	
15	$1,29 \cdot 10^{-7}$	$7,07 \cdot 10^{-7}$	2
30	$9,13 \cdot 10^{-7}$	$5,77 \cdot 10^{-6}$	3
60	$6,45 \cdot 10^{-9}$	$7,07 \cdot 10^{-8}$	2

A média de desempenho tanto do índice Kappa quanto do índice AUC foram maiores ao utilizar o comprimento de memória de trinta valores. Embora com sessenta valores tenha apresentado boa convergência e resultado estatisticamente equivalente para $p < 0,05$ em relação ao desempenho AUC.

4.1.1.2 Locomoção

A sequencia de simulações descrita para reconhecimento da categoria comportamental “Imobilidade” foi realizada também para redes treinadas para detecção da categoria comportamental “Locomoção”, sendo o resultado das diferentes taxas de aprendizagem demonstrado pela Figura 21. Nos resultados apresentados nesta figura os valores de “*fator*” que resultaram em melhores desempenhos em relação ao valor médio do índice AUC foram 10^{-5} , 10^{-5} e 10^{-7} respectivamente para comprimentos de memória de quinze, trinta e sessenta valores. Novamente nota-se uma perda menor de desempenho nas taxas de aprendizagem próximas as indicadas como mais adequadas para o comprimento de memória de trinta e sessenta valores, embora desta vez o maior desempenho médio tenha sido encontrado ao utilizar quinze valores no comprimento de memória.

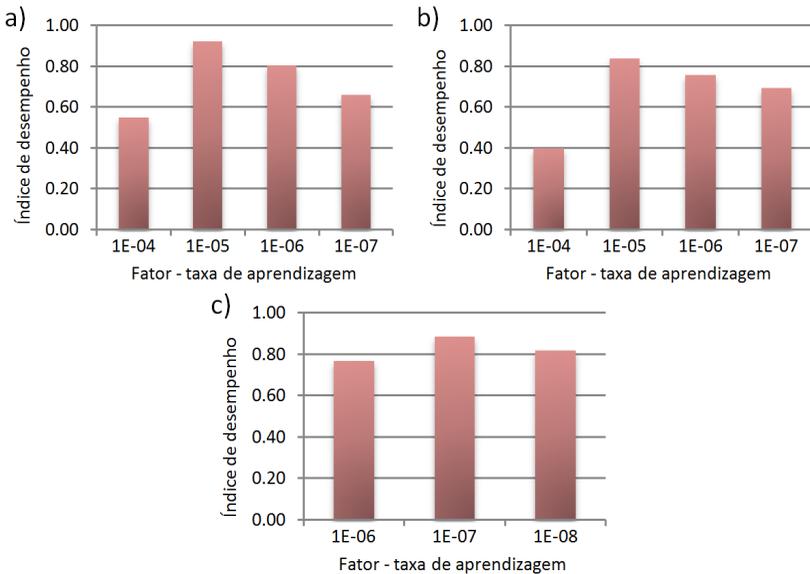


Figura 21 – Desempenho de validação na detecção da categoria comportamental “Locomoção”, dois neurônios na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.

Assim como para detecção de “Imobilidade” as taxas de aprendizagem que apresentaram melhor desempenho na simulação com

dois neurônios foram utilizadas para encontrar o número de neurônios mais adequado. O resultado da simulação com diferentes neurônios e “*fator*” da taxa de aprendizagem fixa é demonstrado na Figura 22.

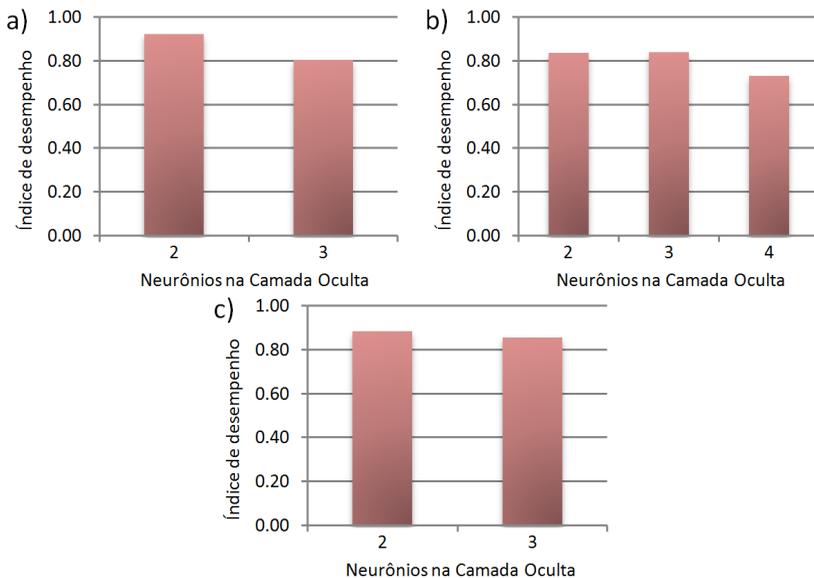


Figura 22 – Desempenho de validação na detecção da categoria comportamental “Locomoção”, taxa de aprendizagem fixa na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.

Novamente, combinando os resultados da Figura 21 e Figura 22 um comparativo entre os comprimentos de memória foi realizado. Tal comparativo é apresentado na Figura 23 e o resumo dos parâmetros aplicados na Tabela 4. Tanto utilizando quinze, trinta ou sessenta valores no comprimento de memória o valor médio de desempenho em relação ao índice AUC foi próximo, sendo que com quinze valores o maior valor foi encontrado. Entretanto, conforme o teste One way ANOVA não houve diferença estatisticamente significativa entre os desempenhos de nenhum dos comprimentos de memória para $p < 0,05$. Os resultados do teste foi $F(2,27)=0,38$ com $p=0,687$ e $F(2,27)=0,50$ com $p=0,613$, respectivamente para o índice Kappa e AUC em relação aos comprimentos.

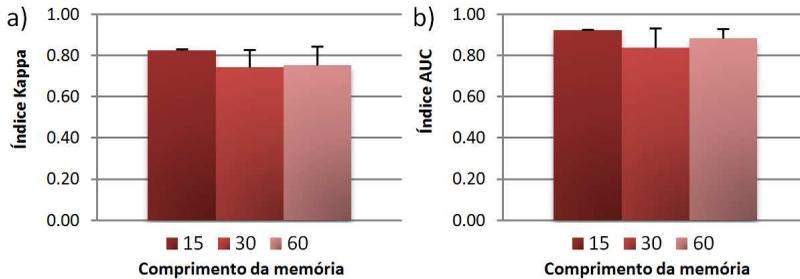


Figura 23 – Comparativo de desempenho de validação na detecção da categoria comportamental “Locomoção”, melhores parâmetros taxa de aprendizagem e neurônios na camada oculta. Média e erro padrão da média dos índices a) Kappa e b) AUC.

Tabela 4 – Melhores taxas de aprendizagem e neurônios na camada oculta na detecção de locomoção para cada comprimento de memória

Comprimento de memória	Taxa de aprendizagem		Neurônios Ocultos
	Camada oculta	Camada de saída	
15	$1,29 \cdot 10^{-6}$	$7,07 \cdot 10^{-6}$	2
30	$9,13 \cdot 10^{-7}$	$5,77 \cdot 10^{-6}$	3
60	$6,45 \cdot 10^{-9}$	$7,07 \cdot 10^{-8}$	2

4.1.1.3 Exploração vertical

Para as redes treinadas para detecção da categoria “Exploração Vertical” o resultado das diferentes taxas de aprendizagem foi o graficamente exposto na Figura 24. Conforme o resultado exposto observa-se um valor comum de taxa de aprendizagem a proporcionar melhor desempenho para ambos os comprimentos de memória, sendo o correspondente ao “fator” 10^{-7} .

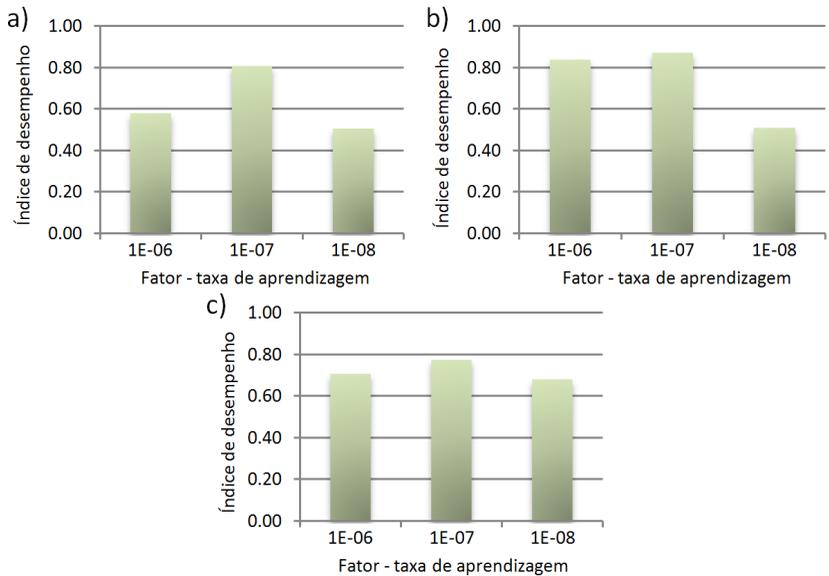


Figura 24 – Desempenho de validação na detecção da categoria comportamental “Exploração Vertical”, dois neurônios na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.

Por sua vez, fixada a taxa de aprendizagem o resultado da simulação com diferentes neurônios é representado na Figura 25. Assim como para as redes treinadas para a detecção de “Locomoção” a melhor generalização ocorreu com apenas dois neurônios para ambos os comprimentos de memória.

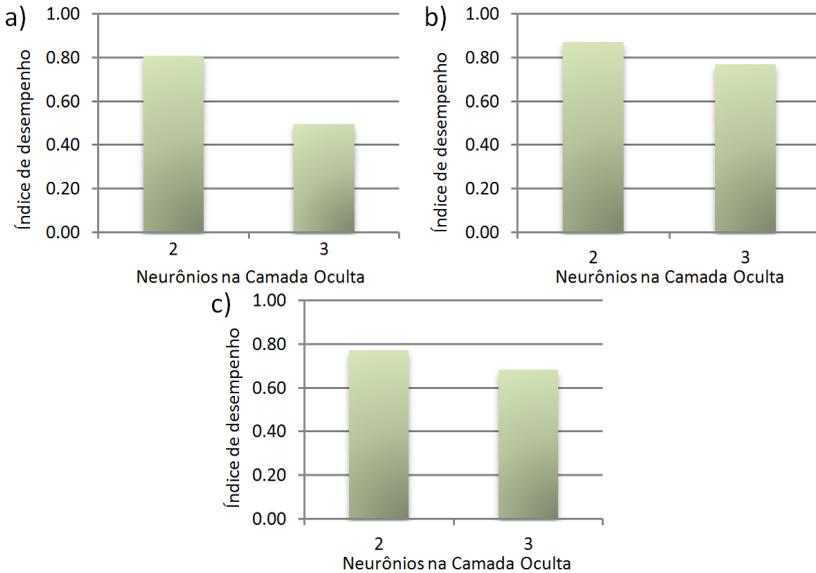


Figura 25 – Desempenho de validação na detecção da categoria comportamental “Exploração Vertical”, taxa de aprendizagem fixa na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.

Na detecção de “Exploração Vertical” a combinação dos melhores parâmetros resultou nos desempenhos comparados na Figura 26. Tais parâmetros são resumidos pela Tabela 5. Na detecção deste comportamento a diferença de desempenho pelos diferentes comprimentos de memória foi ainda menor do que na detecção dos comportamentos anteriormente apresentados, tendo sido o maior valor do desempenho médio encontrado ao utilizar trinta valores na memória. Mais uma vez o teste One way ANOVA apontou não haver diferença estatística significativa para $p < 0,05$. Os resultados do teste foram $F(2,27)=1,12$ com $p=0,341$ para o índice Kappa e $F(2,27)=0,72$ com $p=0,495$ para o AUC, ambos em relação aos diferentes comprimentos de memória.

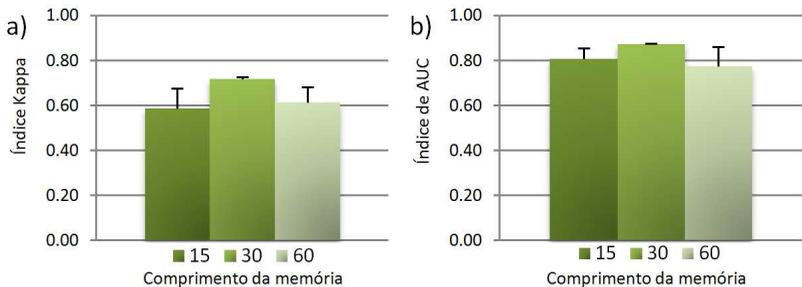


Figura 26 – Comparativo de desempenho de validação na detecção da categoria comportamental “Exploração Vertical”, melhores parâmetros taxa de aprendizagem e neurônios na camada oculta. Média e erro padrão da média dos índices a) Kappa e b) AUC.

Tabela 5 – Melhores taxas de aprendizagem e neurônios na camada oculta na detecção de exploração vertical para cada comprimento de memória

Comprimento de memória	Taxa de aprendizagem		Neurônios Ocultos
	Camada oculta	Camada de saída	
15	$1,29 \cdot 10^{-8}$	$7,07 \cdot 10^{-8}$	2
30	$9,13 \cdot 10^{-9}$	$7,07 \cdot 10^{-8}$	2
60	$6,45 \cdot 10^{-9}$	$7,07 \cdot 10^{-8}$	2

4.1.1.4 Auto-limpeza

Como aos demais comportamentos, foram realizadas simulações para definição da taxa de aprendizagem na detecção da categoria comportamental “Auto-limpeza” as quais resultaram nos desempenhos apresentados na Figura 27. Para este comportamento os valores para o “fator” que resultaram em taxas de aprendizagem mais apropriadas em relação ao valor médio do índice AUC foram 10^{-6} , 10^{-6} e 10^{-7} respectivamente para quinze, trinta e sessenta valores no comprimento de memória.

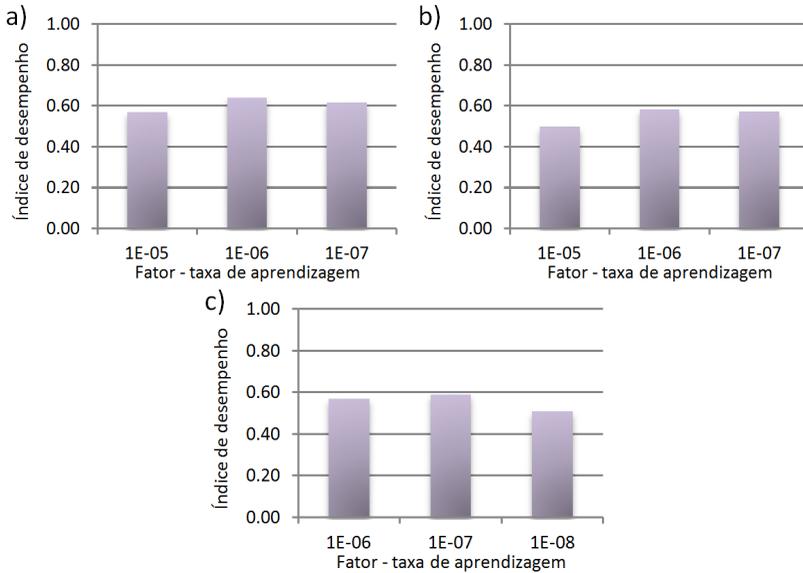


Figura 27 – Desempenho de validação na detecção da categoria comportamental “Auto-limpeza”, dois neurônios na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.

Novamente, determinados os valores de “*fator*” que correspondem a taxas de aprendizagem mais adequadas este foi fixado para determinar o parâmetro número de neurônios na camada oculta. As simulações com diferentes números de neurônios resultaram nos desempenhos da Figura 28. Este grupo de resultados parece apontar uma maior necessidade de neurônios para a detecção desta categoria comportamental, visto que resultados melhores foram encontrados ao utilizar três e quatro neurônios na camada oculta para comprimentos de memória de trinta e sessenta valores respectivamente.

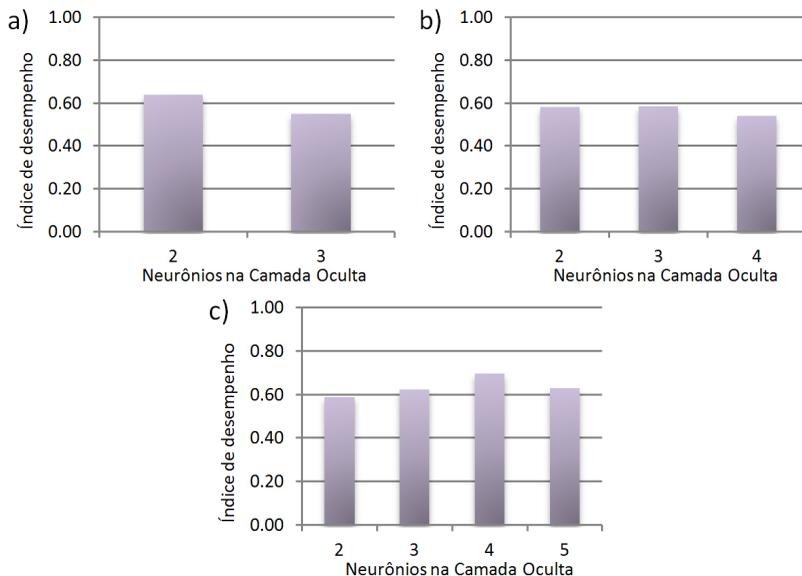


Figura 28 – Desempenho de validação na detecção da categoria comportamental “Auto-limpeza”, taxa de aprendizagem fixa na camada oculta e memórias com: a) quinze; b) trinta e c) sessenta valores. Medidas de desempenho pelo índice AUC.

Combinando os parâmetros (taxas de aprendizagem e número de neurônios na camada oculta) que indicaram o melhor desempenho (resumidos na Tabela 6), o efeito dos diferentes comprimentos de memória foi avaliado e ilustrado na Figura 29. Apesar do comprimento de memória ter tido pouca influência no desempenho médio encontrado, as redes treinadas com sessenta valores desempenharam melhor em relação a este critério. Entretanto, estatisticamente não houve diferença sobre os desempenhos por influência dos comprimentos de memória para $p < 0,05$. O teste One way ANOVA resultou em $F(2,27)=0,96$ com $p=0,395$ para o índice Kappa e $F(2,27)=0,78$ com $p=0,468$ para o índice AUC, ambos testados contra os comprimentos de memória.

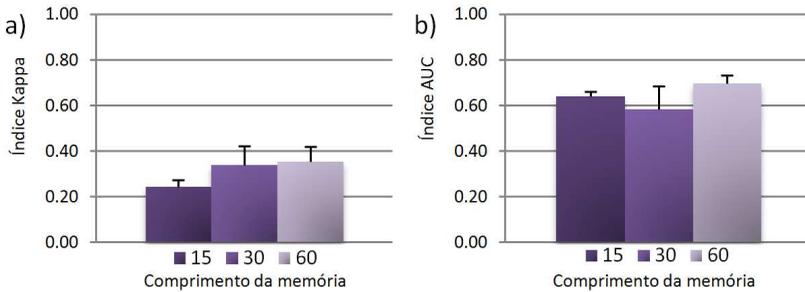


Figura 29 – Comparativo de desempenho de validação na detecção da categoria comportamental “Auto-limpeza”, melhores parâmetros taxa de aprendizagem e neurônios na camada oculta. Média e erro padrão da média dos índices a) Kappa e b) AUC.

Tabela 6 – Melhores taxas de aprendizagem e neurônios na camada oculta na detecção de auto-limpeza para cada comprimento de memória

Comprimento de memória	Taxa de aprendizagem		Neurônios Ocultos
	Camada oculta	Camada de saída	
15	$1,29 \cdot 10^{-7}$	$7,07 \cdot 10^{-7}$	2
30	$9,13 \cdot 10^{-8}$	$5,77 \cdot 10^{-7}$	3
60	$6,45 \cdot 10^{-9}$	$5,00 \cdot 10^{-8}$	4

4.1.2 Teste

Esta seção traz o resultado da aplicação de algumas das redes treinadas com dados de vídeos dos ratos um, dois e três na classificação de dados do vídeo do rato quatro sob mesmo tratamento (veículo). As redes neurais utilizadas são as dez redes (10-dobras) treinadas com as configurações da Tabela 3 a Tabela 6 (combinações de parâmetros que apresentaram melhor desempenho durante a validação).

Novamente as medidas expressas como Kappa e AUC são a média resultante destes índices para cada uma das dez dobras. Como a cada rede testada são calculados índices de desempenho Kappa para cada limiar, o maior valor para cada rede foi utilizado.

Da Figura 30 a Figura 34, são apresentados respectivamente os desempenhos das redes neurais testadas na detecção das categoriais

comportamentais “Imobilidade”, “Locomoção”, “Exploração vertical” e “Auto-limpeza”.

4.1.2.1 Imobilidade

A fim de diferenciar quais comprimentos de memória proporcionaram um melhor ou pior desempenho, foram feitos testes estatísticos. O teste One way ANOVA do desempenho Kappa em relação aos comprimentos de memória resultou em $F(2,27)=3,83$ com $p=0,034$, e do desempenho AUC em $F(2,27)=3,39$ com $p=0,048$. Na análise *post hoc* de Duncan foi encontrada diferença estatisticamente significativa para $p<0,05$ no desempenho Kappa entre comprimentos de memória de quinze e sessenta com $p=0,018$, bem como entre trinta e sessenta com $p=0,049$. Esta mesma análise, mas quanto ao índice AUC resultou em diferença entre comprimentos de quinze e sessenta valores com $p=0,021$.

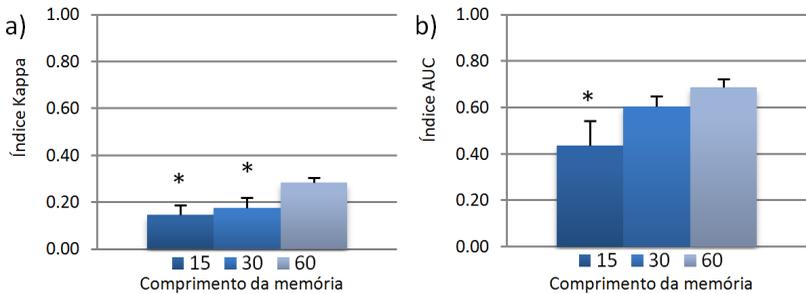


Figura 30 – Comparativo de desempenho de teste na detecção da categoria comportamental “Imobilidade”. Média e erro padrão da média dos índices a) Kappa e b) AUC. Valores marcados com “*” representam diferença estatisticamente significativa com $p < 0,05$ em relação ao valor do comprimento de memória de sessenta valores.

4.1.2.2 Locomoção

Os testes estatísticos realizados para os dados do comportamento “Imobilidade” foram feitos para o comportamento “Locomoção”. O resultado do primeiro teste (One way ANOVA) em relação ao índice Kappa foi $F(2,27)=13,07$ com $p < 0,001$, e $F(2,27) = 3,64$ com $p = 0,040$ em relação ao índice AUC. O teste de Duncan (análise *post hoc*) resultou em diferença estatística com $p<0,05$ em entre todas as combinações entre comprimentos de memória quanto ao índice Kappa,

sendo o maior valor de “p” entre comprimentos de trinta e sessenta ($p=0,022$). Já quanto ao índice AUC foi encontrada diferença com $p<0,05$ apenas entre desempenhos de comprimentos de quinze e sessenta ($p=0,020$).

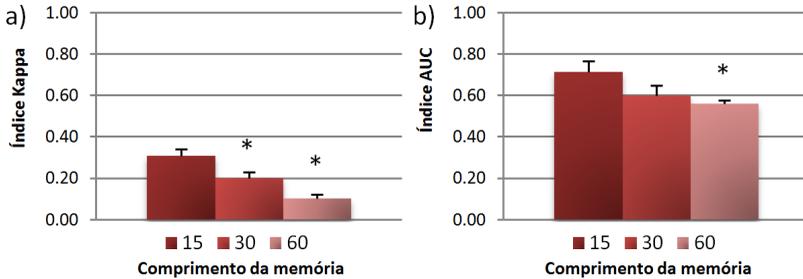


Figura 31 – Comparativo de desempenho de teste na detecção da categoria comportamental “Locomoção”. Média e erro padrão da média dos índices a) Kappa e b) AUC. Valores marcados com “*” indicam diferença estatística em relação ao valor para o comprimento de memória de quinze valores.

4.1.2.3 Exploração vertical

Quanto ao comportamento “Exploração vertical”, a análise One way ANOVA para os desempenhos Kappa e AUC apontou não haver diferença estatisticamente significativa entre cada comprimento de memória para $p<0,05$. Sendo os resultados $F(2,27)=0,78$ com $p=0,469$ e $F(2,27)=2,28$ com $p=0,122$, respectivamente para índices Kappa e AUC em relação aos comprimentos.

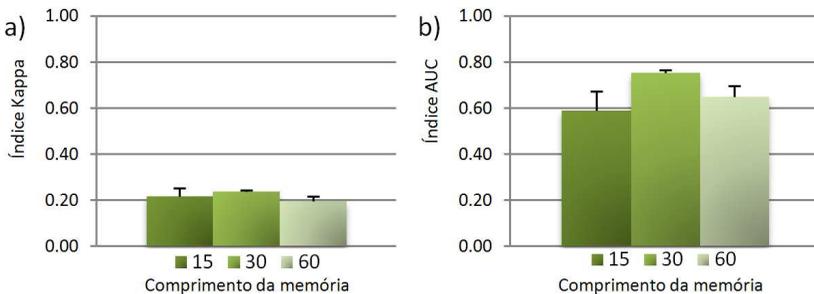


Figura 32 – Comparativo de desempenho de teste na detecção da categoria comportamental “Exploração Vertical”. Média e desvio padrão dos índices a) Kappa e b) AUC.

4.1.2.4 Auto-limpeza

Para o comportamento “Auto-limpeza” não foi diferente em relação à “Exploração vertical”. Sendo o resultado do teste One way ANOVA $F(2,27)=0,93$ com $p=0,408$ e $F(2,27)=2,01$ com $p=0,153$, respectivamente para índices Kappa e AUC com relação aos comprimentos de memória.

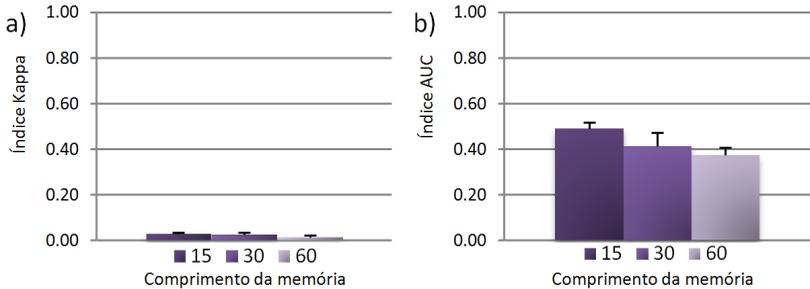


Figura 33 – Comparativo de desempenho de teste na detecção da categoria comportamental “Auto-limpeza”. Média e desvio padrão dos índices a) Kappa e b) AUC.

Estes resultados demonstram que na detecção da categoria “Imobilidade” o aumento do comprimento de memória traz maiores desempenhos, ao contrário das redes treinadas para detecção das categorias “Locomoção”. Já na detecção da categoria “Exploração vertical” e “Auto-limpeza” os comprimentos de memória estudados não demonstraram influência sobre o desempenho. Interessante notar neste conjunto de resultados é que os comprimentos de memória que apresentaram melhor desempenho na detecção de cada categoria comportamental não necessariamente foi o que melhor desempenhou sobre o conjunto de teste. Além disso, nota-se clara redução nos desempenhos de detecção a ambos os comprimentos de memória entre todas as categorias, em especial no desempenho sob o índice Kappa.

4.2 DESEMPENHO DO OBSERVADOR HUMANO

Esta seção demonstra o resultado da medida de concordância entre os registros comportamentais tanto entre um mesmo observador (registrados em períodos diferentes) quanto entre diferentes

observadores quanto ao vídeo do rato quatro, tratado com veículo (salina).

4.2.1 Concordância intra-observador

Primeiramente a confiabilidade de cada um dos quatro observadores que participaram do estudo foi avaliada, através da mensura de concordância entre seus registros quanto a um mesmo experimento em datas distintas. Na Figura 34 são exibidos os valores médios e erro padrão da média do desempenho (índice Kappa) dos observadores na classificação de cada um dos comportamentos. Foi feita uma análise One way ANOVA (Análise de variância de uma via) para verificar a existência de diferença estatisticamente significativa entre os valores do índice Kappa registrados para os diferentes comportamentos, resultando em $F(3,12)=4,49$ com $p=0,025$. Encontrada diferença estatística significativa no grupo, uma análise *post-hoc* de Duncan foi realizada para determinar entre quais comportamentos havia tal diferença. Esta análise demonstrou diferenças entre os valores do Kappa para “Locomoção” e “Imobilidade” com $p=0,049$, “Locomoção” e “Exploração vertical” com $p=0,036$, bem como “Locomoção” e “Auto-limpeza” com $p=0,006$.

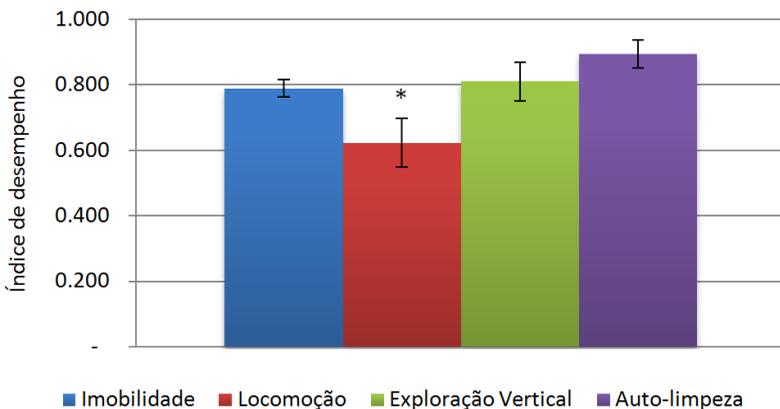


Figura 34 – Desempenho de classificação intra-observador para cada categoria comportamental. Valor médio e erro padrão da média do índice Kappa. Valores marcados com “*” possuem diferença estatisticamente significativa com $p<0,05$ em relação ao valor registrado para a categoria comportamental “Imobilidade”.

Além da análise de variância realizada para verificar diferença entre os comportamentos, foi também realizado o teste One way ANOVA do desempenho em relação aos observadores. O resultado desta análise foi $F(3,12)=1,21$ com $p=0,348$, ou seja, para $p<0,05$ não houve diferença estatisticamente significativa de desempenho entre os observadores.

Os índices de concordância apresentados na Figura 34 demonstram que em média os observadores que realizaram o registro possuem excelente capacidade de reproduzi-lo, sendo a maior dificuldade encontrada em reproduzir o registro da categoria “Locomoção”. O valor médio do índice Kappa calculado ao considerar todas as categorias juntas foi de 0,743 com desvio padrão de $\pm 0,081$.

4.2.2 Concordância inter-observador

A fim de determinar a acurácia dos observadores em registrar as categorias comportamentais definidas, foi medida a concordância entre os observadores. A Figura 35 apresenta os valores médios e erro padrão da média dos desempenhos dos observadores que participaram do estudo em relação a cada categoria comportamental. Novamente foi realizado o teste One way ANOVA para verificar diferença estatisticamente significativa entre índices Kappa com relação às classes comportamentais e com relação aos observadores. Entre os observadores, mais uma vez não foi encontrada diferença estatisticamente significativa para $p<0,05$ sendo o resultado $F(5,18)=0,07$ com $p=0,996$. Já entre os comportamentos o resultado foi $F(3,20)=127,68$ com $p<0,001$. Em análise *post hoc* (teste de Duncan) foi encontrada diferença estatisticamente significativa para $p<0,05$ entre todas as categorias comportamentais, sendo a de maior valor de “p” entre “Imobilidade” e “Exploração vertical” com $p=0,030$.

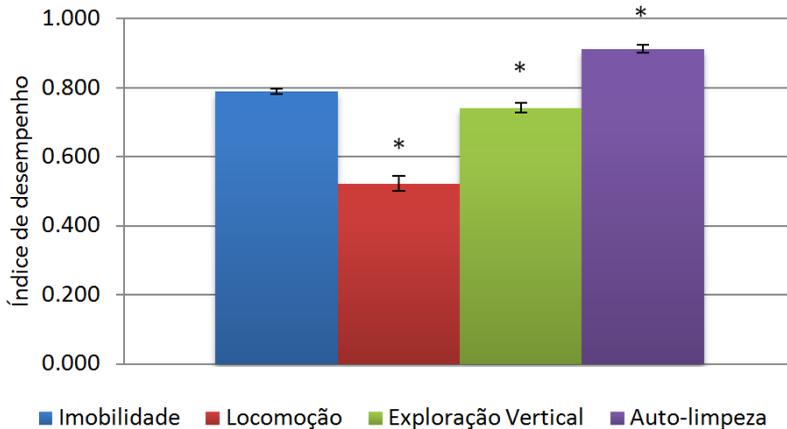


Figura 35 – Desempenho de classificação inter-observador para cada categoria comportamental. Valor médio e erro padrão da média do índice Kappa. Valores marcados com “*” possuem diferença estatisticamente significativa com $p < 0,05$ em relação ao valor registrado para a categoria comportamental “Imobilidade”.

Exceto pela categoria “Auto-limpeza”, o valor médio do desempenho inter-observadores foi menor que o valor intra-observador para cada categoria. Em especial, um valor intensamente menor foi encontrado na categoria “Locomoção” (16,3% de redução) atingindo valor considerado de pobre reprodutibilidade. O índice de desempenho médio para registro de “Exploração vertical” também foi razoavelmente menor (8,4% de redução) em relação à concordância intra-observador. Na concordância inter-observador o valor médio do índice Kappa considerando todas as categorias juntas foi de 0,720 com desvio padrão de $\pm 0,019$.

Como exemplo, um trecho do registro comportamental realizado pelos quatro observadores é apresentado na Tabela 7. Neste exemplo, o registro do observador mais experiente é o da coluna mais à esquerda. Este trecho reflete as partes onde costumam ocorrer mais discordâncias entre os observadores, transições entre comportamentos. O animal que estava se locomovendo, da uma parada e inicia uma exploração vertical. Embora todos concordassem que o animal estava se locomovendo, nenhum dos observadores concordou com relação ao momento em que o animal parou. O primeiro registrou fim da locomoção entre 16,2 e 16,3 segundos, o segundo em 16,1 segundos, o quarto em 16,3 e o terceiro sequer notou uma exploração horizontal que os outros registraram. Para o terceiro o fim da locomoção ocorreu em 16,7s, quando registrou início

da exploração vertical. Ao longo do registro vários trechos como este são encontrados, especialmente envolvendo transições com locomoção.

Tabela 7 – Parte do registro comportamental realizado pelos observadores, usado na avaliação da confiabilidade inter-observador. Onde “L” é locomoção, “H” é exploração horizontal e “V” é exploração vertical.

Tempo (s)	Registros				Tempo (s)	Registros			
15.7	L	L	L	L	16.7	H	V	L	H
15.7	L	L	L	L	16.7	H	V	L	H
15.8	L	L	L	L	16.8	H	V	L	H
15.8	L	L	L	L	16.8	H	V	L	H
15.8	L	L	L	L	16.8	H	V	L	H
15.9	L	L	L	L	16.9	H	V	L	H
15.9	L	L	L	L	16.9	H	V	L	H
15.9	L	L	L	L	16.9	H	V	L	H
16.0	L	L	L	L	17.0	H	V	L	H
16.0	L	L	L	L	17.0	H	V	H	H
16.0	L	L	L	L	17.0	H	V	H	H
16.1	L	L	L	L	17.1	H	V	H	H
16.1	L	H	L	L	17.1	H	V	H	H
16.1	L	H	L	L	17.1	H	V	H	H
16.2	L	H	L	L	17.2	H	V	H	V
16.2	L	H	L	L	17.2	H	V	H	V
16.2	L	H	L	L	17.2	H	V	H	V
16.3	H	H	L	L	17.3	H	V	H	V
16.3	H	H	L	L	17.3	H	V	H	V
16.3	H	H	L	H	17.3	H	V	H	V
16.4	H	H	L	H	17.4	H	V	H	V
16.4	H	H	L	H	17.4	H	V	H	V
16.4	H	H	L	H	17.4	V	V	H	V
16.5	H	H	L	H	17.5	V	V	H	V
16.5	H	H	L	H	17.5	V	V	H	V
16.5	H	H	L	H	17.5	V	V	H	V
16.6	H	H	L	H	17.6	V	V	H	V
16.6	H	H	L	H	17.6	V	V	H	V
16.6	H	V	L	H	17.6	V	V	H	V
16.7	H	V	L	H	17.7	V	V	H	V

Em resumo, os maiores desempenhos médios Kappa/AUC no conjunto de amostras de validação (pré-processadas – balanceadas e sem transições) foram 89,7/95,0% (Imobilidade – memória trinta valores), 82,5/92,3% (Locomoção – memória quinze valores), 71,6/87,1% (Exploração vertical – memória quinze valores) e 35,4/69,7% (Auto-limpeza – memória sessenta valores). No teste uma grande redução no desempenho foi observada, especialmente no índice Kappa.

Comparando o maior desempenho médio Kappa/AUC de cada categoria na validação contra o teste, foi observada redução de 68,3/27,7% (Imobilidade – memória sessenta valores), 62,8/22,5% (Locomoção – memória quinze valores), 66,8/13,5% (Exploração vertical – memória trinta valores) e 92,4/29,8% (Auto-limpeza – memória quinze valores). No teste foram encontradas diferenças estatisticamente significantes com $p < 0,05$ entre desempenhos resultantes dos diferentes comprimentos de memória nas categorias “Imobilidade” e “Locomoção”, não tendo sido encontrada tal diferença quanto as categorias “Exploração vertical” e “Auto-limpeza”. O que indica que eventos da categoria “Imobilidade” são melhor representados por sequencias de dados mais longas (sessenta valores – aprox. dois segundos), já a “Locomoção” o inverso (quinze valores – aprox. meio segundo), e por fim “Exploração vertical” e “Auto-limpeza” aparentemente não são melhor caracterizados por nenhum dos tamanhos de sequencia estudados. Ainda que o desempenho AUC apresentado pelas redes neurais testadas tenha sido bem razoável, está muito aquém da confiabilidade inter-observador desempenhada por observadores humanos. O valor médio do índice de concordância Kappa intra/inter observador entre humanos foi 78,9/78,9% (Imobilidade), 62,3/52,1% (Locomoção), 80,9/74,1% (Exploração vertical) e 89,4/91,2% (Auto-limpeza). Foram encontradas diferenças estatisticamente significantes para $p < 0,05$ nos desempenhos intra-observador entre a categoria comportamental “Locomoção” e as demais, apontando especial dificuldade dos observadores humanos em registrar precisamente tal categoria. Para a concordância inter-observador houve diferença estatisticamente significativa entre todas as categorias comportamentais com $p < 0,05$, apontando maior acerto pelos observadores em relação à categoria “Auto-limpeza” e maior erro na categoria “Locomoção”.

A Tabela 8 traz um comparativo entre os desempenhos de classificação pela rede neural artificial estudada e o método usual de classificação (observadores humanos).

Tabela 8 – Comparativo do desempenho (índice Kappa médio) de classificação entre Redes neurais artificiais (automática) e Observadores humanos (manual).

Classe Comportamental	Observadores Humanos		Redes Neurais Artificiais	
	Intra-obs.	Inter-obs.	Intra-obs.	Inter-obs.
Imobilidade	78,87%	78,89%	100,00 %	28,41 %
Locomoção	62,26%	52,14%	100,00 %	30,74 %
Exploração Vertical	80,94%	74,10%	100,00 %	23,78 %
Auto-limpeza	89,37%	91,24%	100,00 %	2,68 %

Na Tabela 8 são expostos valores médios do índice Kappa por ser comum a ambas as formas de classificação, permitindo a comparação. O desempenho Intra-observador da Rede Neural Artificial estudada é sempre 100% por apresentar sempre a mesma resposta para um dado exemplo de entrada, visto que depois de treinada seus pesos não são mais alterados.

5 DISCUSSÃO

O objetivo desta dissertação foi investigar uma rede neural dinâmica no reconhecimento de comportamentos de animais em uma arena de campo aberto, abordando questões como profundidades (comprimento) de memória adequadas à rede neural e quão melhor ou pior é o desempenho do método automático em relação ao método usual de registro comportamental (observador humano).

Dentre as topologias de redes neurais dinâmicas estudadas, foi escolhida a TLFN focada por sua simplicidade e aplicabilidade à tarefa requerida. Os experimentos e métodos de processamento de imagem aplicados foram os mesmos que em Crispim (2011), ficando como principais diferenças a separação das amostras entre os conjuntos de treino, validação e teste, bem como o pré-processamento das amostras e método de reconhecimento de padrões empregados. Foram utilizados neste trabalho quatro vídeos (um de cada animal) processados no software ETHOWATCHER[®]. Dentre os descritores morfológicos e cinemáticos extraídos foram utilizados a “área do animal”, “comprimento”, “distância percorrida pelo centroide” e “número de pixels alterados a cada quadro” todos medidos em pixels. Os descritores extraídos dos vídeos dos ratos de um a três foram utilizados no treinamento e validação da rede neural. Na expectativa de um melhor aprendizado, os dados utilizados no treinamento/validação de cada rede neural foram pré-processados. Estes conjuntos de dados foram balanceados em relação a cada categoria e trechos de transições foram retirados (onde o supervisor da rede também cometeria mais erros). O vídeo do quarto rato por sua vez, foi utilizado para teste da rede neural e avaliações de confiabilidade dos registros comportamentais realizados por humanos.

Os comprimentos de memória que demonstraram serem significativamente influentes em relação ao desempenho Kappa no processo de validação e resultaram na melhor detecção sob a mesma, foram trinta valores para detecção de “Imobilidade” e quinze valores para detecção de “Locomoção” (não tendo causado influência sob as categorias “Exploração vertical” e “Auto-limpeza”). Neste grupo de resultados os maiores desempenhos médios de classificação foram bastante razoáveis se comparados com o desempenho de observadores humanos (inter-observadores). Nesta última categoria o desempenho não foi satisfatório nem mesmo durante a validação. De modo geral as categorias apresentaram seus melhores resultados com dois neurônios na

camada oculta. Porém, os resultados das redes treinadas para detecção da categoria “Auto-limpeza” parecem apontar uma maior necessidade de neurônios, visto que os melhores desempenhos foram encontrados com três e quatro neurônios na camada oculta para comprimentos de memória de trinta e sessenta valores respectivamente.

Os desempenhos encontrados na validação dão uma indicação de quais comprimentos de memória e quais as categorias comportamentais mais redes convergiram para pontos de maior desempenho (aprenderam mais) em relação aos dados utilizados. No entanto, para saber quais comprimentos de memória e qual a estimativa de desempenho de uma aplicação destas redes em situação prática o teste fornece a melhor referência. Os comprimentos de memória onde houve melhor desempenho para cada categoria foram diferentes no teste em relação à validação, sendo sessenta valores para “Imobilidade” e quinze para “Locomoção”. Novamente não houve diferença estatisticamente significativa para o uso de diferentes comprimentos de memória na detecção das categorias “Exploração vertical” e “Auto-limpeza” (onde o desempenho foi insignificante a todos os comprimentos de memória estudados) com $p < 0,05$.

O desempenho AUC de um classificador totalmente ineficiente é 50% e eficiência máxima 100%, de modo que exceto pelo classificador “Auto-limpeza” houve razoável generalização das redes neurais em relação ao conjunto de teste para os melhores comprimentos de memória. Entretanto, apesar do desempenho encontrado entre as redes neurais ser melhor do que a chance, o mesmo foi fraco em relação ao alto desempenho de classificação encontrado entre observadores humanos. Para redes neurais artificiais o desempenho intra-observador é máximo (100%), pois não há variação entre suas observações para um mesmo experimento. Entretanto, comparando os desempenhos médios no teste das redes neurais (desempenho inter-observador) contra o desempenho inter-observador médio por humanos foram observados desempenhos 64,0%, 41,0%, 67,9% e 97,1% menores para redes neurais das respectivas categorias “Imobilidade”, “Locomoção”, “Exploração vertical” e “Auto-limpeza”. O que se traduz em uma boa precisão, mas baixa exatidão por parte das redes neurais em detectar tais comportamentos sob tal comparação.

Nos desempenhos inter-observador foram encontradas diferenças estatisticamente significantes para $p < 0,05$ entre ambas as categoria comportamentais. O que reflete dificuldade dos observadores humanos em registrar corretamente “Locomoção”, ao contrário da grande facilidade em relação à “Auto-limpeza”. A concordância é diretamente

proporcional a duração dos trechos concordantes e inversamente proporcional aos trechos discordantes. Também, são observadas discordâncias em maior frequência nas transições entre comportamentos. Somando estas duas afirmações é possível estimar que comportamentos em que os eventos são frequentemente de duração maior (como auto-limpeza e imobilidade), menos erros de concordância sejam encontrados em relação a categorias em que os eventos são geralmente de duração mais curta como a “Locomoção”.

Para as redes neurais artificiais a categoria mais “difícil” de detectar (concordar com o humano – menor desempenho) foi a “Auto-limpeza” tanto durante validação quanto no teste. Estas diferenças (ou até mesmo contraste) entre as facilidades e dificuldades de registro comportamental entre humanos e redes neurais artificiais com relação a cada categoria, indica haver realmente grande discrepância em relação ao que cada grupo (humanos ou redes neurais) está vendo e como está aprendendo. Mudanças como o acréscimo de outros descritores de imagem, aplicação de técnicas mais sofisticadas de processamento de imagem ou até mesmo outras técnicas de classificação de padrões (e. g. uso de redes neurais artificiais com dinâmica intrínseca) devem impactar significativamente neste quadro.

Embora entre os comprimentos de memória estudados não tenha havido influência estatisticamente significativa no desempenho de detecção das categorias “Exploração vertical” e “Auto-limpeza”, não quer dizer que não sofram influência por parte deste parâmetro. Estudando outros valores de comprimentos de memória além dos estudados, influência estatisticamente significativa deste parâmetro pode ser revelada.

O pré-processamento (retirada das transições entre comportamentos e sub-amostragem) dos dados de treinamento, apesar de provavelmente ter melhorado o desempenho durante a validação pode ter prejudicado a generalização sob o conjunto de teste (composto por várias transições entre comportamentos). Esta escolha dos exemplos de treino pode ter alterado-o a ponto de não representar adequadamente o conjunto de teste, mas melhorar o desempenho de validação ao ser formado por exemplos mais claros das categorias comportamentais. Ainda assim, os resultados indicam que o conjunto de treino/validação represente razoavelmente o conjunto de teste, visto que em geral os comprimentos de memória que melhor desempenharam no treino/validação também o fizeram no teste.

6 CONCLUSÃO

Este trabalho demonstrou que redes neurais artificiais dinâmicas são úteis na classificação de padrões comportamentais. O uso de índices de desempenho de classificadores binários permitiu a comparação da solução automatizada com a manual (observação e registro por humanos), na qual foi encontrada maior exatidão (reprodutibilidade entre observadores diferentes).

A categoria comportamental em que foi encontrado o maior desempenho por parte de observadores humanos, foi na qual a rede neural artificial desempenhou pior. Já a categoria em que a rede neural teve maior desempenho foi na qual os observadores humanos tiveram o seu pior. Estas diferenças entre categorias em que cada um dos envolvidos na comparação desempenhou melhor demonstraram que há significativa diferença entre as informações recebidas e forma de processamento desta por cada um destes. Apesar de a metodologia aplicada ter resultado em um classificador útil, modificações são necessárias para alcançar uma maior concordância com registros de humanos.

O uso de uma rede neural artificial anteroalimentada com atrasos de tempo focados na entrada possibilitou responder a questão sobre a influência do comprimento de memória na detecção de eventos comportamentais. Foi demonstrado que para certas categoriais diferentes profundidades da memória de curto-prazo são melhores ou piores para o reconhecimento das mesmas, resultando em melhor ou pior aprendizado. Entretanto, para outras esta influência não foi encontrada em relação aos comprimentos de memória estudados.

Para trabalhos futuros aplicando metodologias semelhantes é interessante investigar o uso de comprimentos de memória menores e maiores, a fim de obter maior clareza em relação à influência deste parâmetro sobre a detecção das categorias “Exploração vertical” e “Auto-limpeza”. Além disto, seria interessante investigar a influência do uso de filmes completos ou sem as transições entre comportamentos diferentes no treinamento da rede neural artificial estudada. Em relação a trabalhos aplicando outras redes neurais artificiais, é interessante investigar o uso de redes neurais artificiais intrinsecamente dinâmicas na qual outra forma de processamento temporal poderia resultar em um aprendizado diferente sobre as categorias estudadas. Por fim, estudos da aplicação de outros descritores de imagem como o fluxo óptico poderiam resultar em melhor detecção de categorias comportamentais

como a “Auto-limpeza” onde o desempenho de detecção no teste foi quase nulo.

REFERÊNCIAS

ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. **Statistics Surveys**, v. 4, p. 40-79, 2010.

BINDRA, D.; SPINNER, N. Response to different degrees of novelty: the incidence of various activities. **J. Exp. Anal. Behav.**, n. 1, p. 341-350, 1958.

BODEN, M. A. **Artificial Intelligence – Handbook of Perception and Cognition**. 2. ed. San Diego: Academic Press, 1996.

BOOS, C. F.; AZEVEDO, F. M.; SCOLARO, G. R.; PEREIRA, M. C. V. Pereira. Automatic Detection of Paroxysms in EEG Signals Using Morphological Descriptors and Artificial Neural Networks. In: LASKOYSKI, A. N. **Biomedical Engineering: Trends in Electronics, Communications and Software**. Rijeka, Croácia: InTech, 2011.

BOSE, R. C. **Proposta para um Sistema de Rastreamento e Análise de Movimentos em Animais de Laboratório**. Dissertação de Mestrado em Engenharia Elétrica, Instituto de Engenharia Biomédica, Universidade Federal de Santa Catarina. Florianópolis. 2003.

CASSON, A. J.; LUNA, E.; VILLEGAS, E. R. Performance metrics for the accurate characterization of interictal spike detection algorithms. **Elsevier - Journal of Neuroscience Methods**, v. 177, n. 2, p. 479-487, março 2009.

CAVANAGH, P. Visual cognition. **Vision research**, v. 51, n. 13, p. 1538-1551, 2011.

CRISPIM, C. F. **Uso de descritores morfológicos e cinemáticos na identificação automática de comportamentos de animais de laboratório**. Tese de Doutorado em Engenharia Elétrica, Instituto de Engenharia Biomédica, Universidade Federal de Santa Catarina. Florianópolis. 2011.

CRISPIM, C. F.; SANTOS, T. S.; HOELLER, A. A.; MARINO-NETO, J. ETÓGRAFO 2: Etografia, Análise de Locomoção e Classificação Automática de Comportamentos através de Recursos Estatísticos. **I IBRO/LARC Congress of Neurosciences of Latin America, The Caribbean and Iberian Peninsula**, Buzios, 2008. v. 1, p. 1386.

CRISPIM, C. F.; SANTOS, T. S.; MELLEU, F. F.; GARCIA, V. A.; PEDERIVA, C. N.; BOSE, R. C.; MARINO-NETO, J. EthoWatcher: uma ferramenta para o registro comportamental e a análise automática da atividade

de animais de laboratório. **XXXIV Congresso Anual da Sociedade Brasileira de Neurociência e Comportamento**, Caxambu, 2010.

CRISPIM, C. F.; PEDERIVA, C. N.; BOSE, R. C.; GARCIA, V. A.; LINO-DE-OLIVEIRA, C.; MARINO-NETO, J. ETHOWATCHER: validation of a tool for behavioral and video-tracking analysis in laboratory animals. **Elsevier - Computers in Biology and Medicine**, v. 42, n. 2, p. 257-264, 2012.

DOLLÁR, P.; RABAUD, V.; COTTRELL, G.; BELONGIE, S. Behavior Recognition via Sparse Spatio-Temporal Features. **Proceedings of 2nd Joint IEEE International Workshop on VS-PETS**, Beijing, 2005. p. 15-16.

ELMAN, J. L. Finding structure in time. **Cognitive Science**, v. 14, p. 179-211, 1990.

ESCABI, M. A. Biosignal Processing. In: ENDERLE, J. D.; BLANCHARD, S. M.; BRONZINO, J. **Introduction to Biomedical Engineering**. 2 ed. Burlington, MA: Elsevier Academic Press, 2005. p. 549-625.

FISHER, L. D.; VAN-BELLE, G.; HEAGERTY, P. J.; LUMLEY, T. **Biostatistics a Methodology for the Health Sciences**. 2 ed. Canada: John Wiley & Sons, 2004.

FRÖLICH, H.; HOENSELAAR, A.; EICHNER, J.; ROSENBROCK, H.; BIRK, G.; ZELL, A. Automated classification of the behavior of rats in the forced swimming test with support vector machines. **Neural Networks**, v. 21, p. 92-101, 2008.

GIESE, M. A.; POGGIO, T. Neural mechanisms for the recognition of biological movements. **Nature reviews - Neuroscience**, v. 4, n. 3, p. 179-192, março 2003.

GONZALEZ, L. P.; ARNALDO, C. M. Classification of drug-induced behaviors using a multi-layer feed-forward neural network. **Computer Methods and Programs in Biomedicine**, v. 40, n. 3, p. 167-173, julho 1993.

GONZALEZ, R. C.; WOODS, R. E. **Processamento de Imagens Digitais**. 1 ed. São Paulo: Blücher, 2000.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. 2 ed. Índia: Pearson Prentice Hall, 2005.

HEEREN, D. J.; COOLS, A. R. Classifying postures of freely moving rodents with the help of Fourier descriptors and a neural network. **Behavior Research Methods, Instruments, & Computers**, v. 32, n. 1, p. 56-62, 2000.

JHUANG, H.; GARROTE, E.; YU, X.; KHILNANI, V.; POGGIO, T.; STEELE, A. D.; SERRE, T. Automated home-cage behavioural phenotyping of mice. **Nature Communications**, v. 1, n. 68, setembro 2010.

JOHANSSON, G. Visual perception of biological motion and a model for its analysis. **Perception & Psychophysics**, v. 14, n. 2, p. 201-211, 1973.

JORDAN, M. I. **Serial order: A parallel distributed processing approach**. Institute for Cognitive Science, University of California. San Diego, p. 40. 1986.

LE, C. T. **Introductory Biostatistics**. Canada: John Wiley & Sons, 2003.

LECUN, Y. Efficient BackProp. In: ORR, G. B.; MÜLLER, K. R. **Neural Networks: Tricks of the trade**. [S.l.]: Springer, 1998.

LEHNER, P. N. **Handbook of Ethological Methods**. 2 ed. New York: Cambridge University Press, 1996.

LIND, N. M.; VINTHER, M.; HEMMINGSEN, R. P.; HANSEN, A. K. Validation of a digital video tracking system for recording pig locomotor behaviour. **Journal of Neuroscience Methods**, v. 143, p. 123-132, 2005.

MARQUES, O.; VIEIRA, H. **Processamento Digital de Imagens**. Rio de Janeiro: Brasport, 1999.

MARR, D. **Vision**. San Francisco, CA: W H Freeman, 1982.

MARTIN, P.; BATESON, P. **Measuring Behaviour**. New York: Cambridge University Press, 2007.

METRIS, B. V. Laboras Product Information. **Metris - Animal Behavior Analysis Solution**. Disponível em:
<http://www.metris.nl/en/products/laboras/laboras_information>. Acesso em: 30 dezembro 2011.

MOZER, M. C. Neural net architectures for temporal sequence processing. In: WEIGEND, A.; GERSHENFELD, N. **Time Series Prediction: Forecasting the Future and Understanding the Past**. Redwood, California: Addison-Wesley Publishing, 1994. p. 243-264.

NGUYEN, G. H.; BOUZERDOUM, A.; PHUNG, S. Learning pattern classification tasks with imbalanced data sets. In: YIN, P. **Pattern recognition**. Vukovar, Croatia: In-Teh, 2009. p. 193-208.

PAGE, M. P. A.; HOWARD, R. J.; O'BRIEN, J. T.; BUXTON-THOMAS, M. S.; PICKERING, A. D. Use of neural networks in brain SPECT to diagnose Alzheimer's disease. **Journal of Nuclear Medicine**, v. 37, n. 2, p. 195-200, 1996.

PALODETO, V.; MARQUES, J. L. B. Methodology for Classification and Analysis of Neonate and Adult ECG., v. 14, p. 1214-1217, 2007.

PEDERIVA, C. N. **Etógrafo: um sistema de auxílio ao Registro e à análise de dados comportamentais**. Dissertação de Mestrado em Engenharia Elétrica, Instituto de Engenharia Biomédica, Universidade Federal de Santa Catarina. Florianópolis. 2005.

ROUSSEAU, J. B. I.; VAN-LOCHEM, P. B.; GISPEN, W. H.; SPRUIJT, B. M. Classification of rat behavior with an image-processing method and a neural network. **Behavior research methods, instruments & computers: a journal of Psychonomic Society**, v. 32, n. 1, p. 63-71, fevereiro 2000.

SAHINER, B.; HEANG-PING, C.; PETRICK, N.; DATONG, W.; HELVIE, M. A.; ADLER, D. D.; GOODSITT, M. M. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. **Medical Imaging, IEEE Transactions on**, v. 15, n. 5, p. 598-610, outubro 1996.

SANDBERG, I. W.; XU, L. Uniform Approximation of Multidimensional Myopic Maps. **IEEE Transactions on Circuits and Systems – Fundamental Theory and Applications**, v. 44, n. 6, p. 477-485, junho 1997.

SCOLARO, G. R.; AZEVEDO, F. M. Classification of Epileptiform Events in Raw EEG Signals using Neural Classifier. **Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on**, v. 5, p. 368-372, julho 2010.

SILVA, J. A.; GONÇALVES, W. N.; MACHADO, B. B.; PISTORI, H.; SOUZA, A. S. Modelos Ocultos de Markov Aplicados na Identificação de Comportamento de Camundongos. **Hifen**, v. 31, n. 59, 2007.

SOVIERZOSKI, M. A.; ARGOUD, F. I. M.; AZEVEDO, F. M. Performance Evaluation of an ANN FF Classifier of Raw EEG Data using ROC Analysis. **Proceedings of the 2008 International Conference on BioMedical Engineering and Informatics**, v. 1, p. 332-336, 2008.

SPINK, A. J.; TEGELENBOSCH, R. A.; BUMA, M. O.; NOLDUS, L. P. The EthoVision video tracking system: A tool for behavioral phenotyping of transgenic mice. **Physiology & Behavior**, v. 73, p. 731-744, 2001.

TWINING, C. J.; TAYLOR, C. J.; COURTNEY, P. Robust tracking and posture description for laboratory rodent using active shape models. **Behavior Research Methods, Instruments, & Computers**, v. 33, n. 3, p. 381-391, 2001.

WANG, Z.; YANG, X.; XU, Y.; YU, S. CamShift guided particle filter for visual tracking. **Pattern Recognition Letters**, v. 30, n. 4, 2009.

WATROUS, R. L.; SHASTRI, L. Learning phonetic features using connectionist networks: An experiment in speech recognition. **Proceedings of the 10th international joint conference on Artificial intelligence (IJCAI)**, v. 2, p. 851-854, 1987.

YU, H.; SUN, G.; SONG, W.; LI, X. Human motion recognition based on neural network. **Communications, Circuits and Systems, 2005 International Conference on**, v. 2, p. 2, maio 2005.

YU, J.; TAN, J. Object density-based image segmentation and its applications in biomedical image analysis. **Computer methods and programs in biomedicine**, v. 96, p. 193-204, 2009.

YUHAS, B. P.; GOLDSTEIN, M. H.; SEJNOWSKI, T. Integration of acoustic and visual speech signals using neural networks. **Communications Magazine, IEEE**, v. 27, n. 11, p. 65-71, novembro 1989.

ZADEH, A. E.; KHAZAEI, A.; RANAEE, V. Classification of the electrocardiogram signals using supervised classifiers and efficient features. **Computer Methods and Programs in Biomedicine**, v. 99, p. 179-194, 2010.

ZURN, J. B.; JIANG, X.; MOTAI, Y. Video-Based Rodent Activity Measurement Using Near-Infrared Illumination. **Proceedings of the IEEE Instrumentation and Measurement Technology Conference**, Ottawa-Canada, 2005. 1928-1931.