



INTERFACE CÉREBRO-MÁQUINA:
RECONHECIMENTO DE SINAIS
CEREBRAIS ATRAVÉS DE *RESERVOIR*
COMPUTING

Trabalho de Conclusão de Curso
Engenharia de Computação

Felipe Costa Farias

Orientador: Prof. Dr. Carmelo José Albanez Bastos Filho



Felipe Costa Farias

*INTERFACE CÉREBRO-MÁQUINA:
RECONHECIMENTO DE SINAIS
CEREBRAIS ATRAVÉS DE RESERVOIR
COMPUTING*

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco - Universidade de Pernambuco

Orientador:

Prof. Dr. Carmelo José Albanez Bastos Filho

UNIVERSIDADE DE PERNAMBUCO
ESCOLA POLITÉCNICA DE PERNAMBUCO
GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

Recife - PE, Brasil

30 de maio de 2014

MONOGRAFIA DE FINAL DE CURSO

Avaliação Final (para o presidente da banca)*

No dia 30 de 5 de 2014, às 11:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente FELIPE COSTA FARIAS, orientado pelo professor Carmelo José Albanez Bastos Filho, sob título Interface Cérebro-Máquina: Reconhecimento de sinais cerebrais através de resevoir computing, a banca composta pelos professores:

Diego José Rátiva Millan

Carmelo José Albanez Bastos Filho

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

Aprovada Aprovada com Restrições* Reprovada

e foi-lhe atribuída nota: 9,5 (nove e meio)

*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 7 dias para entrega da versão final da monografia a contar da data deste documento.

DIEGO JOSÉ RÁTIVA MILLAN

CARMELO JOSÉ ALBANEZ BASTOS FILHO

Dedico este trabalho aos meus familiares e amigos que sempre me incentivaram na caminhada acadêmica.

Agradecimentos

Agradeço a Deus por todas as pessoas, oportunidades e conquistas que Ele me presenteou. Aos meus familiares: meus pais Flávio Farias e Cláudia Farias, pelos ensinamentos, amor, carinho e esforços que tiveram para a minha educação e caráter. A minha irmã, Flávia Farias, pela convivência e amor por mim. A minhas tias, Célia Costa e Celi Costa, pelo amor e incentivo nos estudos. A minha noiva, Camila Bacelar, pelo amor, paciência e força para continuar. Ao meu orientador, Carmelo Bastos, pelo tempo e sabedoria dedicada em minha orientação. E a todos os professores e funcionários da POLI.

Resumo

A Interface Cérebro-Máquina utiliza sinais cerebrais para controlar dispositivos externos através da tradução desses sinais em comandos. A ICM tem grande utilidade no auxílio de pessoas com os mais diversos tipos de deficiências. Este trabalho se propôs a estudar a influência temporal, espacial e combinacional dos sinais adquiridos através de EEGs. Para isto, foram utilizados 3 classificadores de padrões instantâneos: Perceptron de Múltiplas Camadas com o algoritmo *backpropagation*, Máquinas de Aprendizado Extremo e Máquinas de Vetor de Suporte com *kernel* de base radial, e 1 classificador de padrões temporais: *Reservoir Computing* com o algoritmo de treinamento *Echo State Network*. Foram propostas 4 configurações diferentes para avaliar as taxas de acerto dos classificadores, que utilizaram 14 canais de EEG para realizar a classificação de movimentos imaginados do punho esquerdo ou direito de forma *offline* de 5 sujeitos diferentes. A escolha do EEG para efetuar a classificação se deu devido ao fato de ser um procedimento não invasivo e de baixo custo. Utilizou-se o banco de dados *eegmmidb* para realizar as simulações. Foram feitos testes estatísticos para a comparação das técnicas. A partir da análise estatística de Wilcoxon, foi sugerido que a influência temporal dos sinais de EEG merece ser considerada durante a classificação. Percebeu-se também que, se os classificadores fossem treinados para sujeitos específicos, em alguns casos, as taxas de classificação melhoravam consideravelmente.

Palavras-chave: *Reservoir Computing, Perceptron de Múltiplas Camadas, Rede de Estado de Eco, Máquinas de Vetor de Suporte, Máquinas de Aprendizado Extremo, Interface Cérebro-Máquina, Eletroencefalograma.*

Abstract

The Brain Machine Interface uses brain signals to control external devices by translating these signals to commands. The BMI is very useful in helping people with various types of disabilities. This paper aimed to study the temporal, spatial and combinational influence of the signals acquired through EEGs. In order to do this it were used 3 classifiers of instant patterns: Multilayer Perceptron with the backpropagation algorithm, Extreme Learning Machine and Support Vector Machine with the Radial Basis Function kernel, and 1 classifier of temporal patterns: Reservoir Computing, with the Echo State Network algorithm. It were proposed 4 different configurations to evaluate the accuracy rates of the classifiers, that used 14 channels of EEG to perform the movement imagery classification of left and right fists in an offline way of 5 different subjects. The choice of EEG to accomplish the classification was due to the fact that this is a non-invasive and inexpensive procedure. The eegmidb database was used to perform the simulations. Statistical tests were performed to compare the techniques. From the Wilcoxon statistical analysis, it was suggested that the temporal influence of EEG signals is worth to consider during the classification. It was also noticed that, if the classifiers were trained for specific subjects, in some cases, classification rates improved considerably.

Keywords: Reservoir Computing, Multilayer Perceptron, Echo State Network, Support Vector Machine, Extreme Learning Machine, Brain Machine Interface, Electroencephalogram.

Sumário

Lista de Figuras	p. x
Lista de Tabelas	p. xi
Lista de Abreviaturas e Siglas	p. xi
1 Introdução	p. 1
1.1 Qualificação do Problema	p. 1
1.2 Objetivos	p. 1
1.2.1 Objetivos Específicos	p. 2
1.3 Estrutura da Monografia	p. 2
2 Referencial Teórico	p. 3
2.1 Aquisição de Sinais Biológicos	p. 3
2.1.1 Eletromiograma (EMG)	p. 3
2.1.2 Eletrocardiograma (ECG)	p. 4
2.1.3 Eletrooculografia (EOG)	p. 4
2.1.4 Eletrocorticografia (ECOG)	p. 4
2.1.5 Eletroencefalografia (EEG)	p. 5
2.2 Interface Cérebro-Máquina	p. 6
2.3 Classificadores de Padrões Instantâneos	p. 9
2.3.1 Perceptron de Múltiplas Camadas	p. 9
2.3.2 Máquinas de Aprendizado Extremo	p. 12

2.3.3	Máquinas de Vetor de Suporte	p. 13
2.4	Classificadores de Padrões Temporais	p. 15
2.4.1	<i>Reservoir Computing</i>	p. 15
2.4.1.1	Modelo Básico	p. 16
2.4.1.2	Parâmetros da ESN	p. 18
2.4.1.2.1	Tamanho do <i>Reservoir</i>	p. 18
2.4.1.2.2	Esparsidade do <i>Reservoir</i>	p. 18
2.4.1.2.3	Raio Espectral	p. 18
2.4.1.2.4	Taxa de Vazamento	p. 19
2.4.1.2.5	Número de Passos do Aquecimento	p. 19
2.4.1.3	Treinamento da ESN	p. 19
3	Propostas e Arranjo Experimental	p. 20
3.1	Entradas dos Classificadores	p. 20
3.2	Pré-processamento	p. 20
3.3	Saídas dos Classificadores	p. 21
3.4	Utilização dos EEGs	p. 21
3.5	Configurações dos Classificadores	p. 22
3.5.1	Configurações MLP	p. 22
3.5.2	Configurações ELM	p. 23
3.5.3	Configurações SVM	p. 23
3.5.4	Configurações ESN	p. 24
3.6	Banco de Dados	p. 24
3.6.1	Protocolo de Aquisição dos Dados	p. 24
3.6.2	Preparação do Banco de Dados	p. 26
3.7	Simulações	p. 27

4 Resultados	p. 28
4.0.1 Configurações MLP	p. 28
4.0.2 Configurações ELM	p. 29
4.0.3 Configurações SVM	p. 30
4.0.4 Configurações ESN	p. 31
4.1 Análise Estatística	p. 32
4.1.1 Teste de Wilcoxon	p. 32
5 Considerações Finais	p. 33
5.1 Conclusões	p. 33
5.2 Trabalhos Futuros	p. 34
Referências	p. 35

Lista de Figuras

1	Eletromiograma	p. 3
2	Eletrocardiograma	p. 4
3	Eletrooculografia	p. 4
4	Instalação da Matriz de Eletrodos para Realização do ECOG	p. 5
5	Eletroencefalografia	p. 5
6	Modelo de Sistema ICM Genérico	p. 9
7	Arquitetura da MLP com uma Camada Intermediária	p. 11
8	Fronteiras de Decisão	p. 15
9	Estrutura de uma ESN Básica	p. 17
10	Tangente Hiperbólica	p. 21
11	Disposição dos Eletrodos no Sistema 10-10	p. 22
12	<i>Boxplots</i> das Configurações da MLP.	p. 29
13	<i>Boxplots</i> das Configurações da ELM.	p. 30
14	<i>Boxplots</i> das Configurações da SVM.	p. 31
15	<i>Boxplots</i> das Configurações da ESN.	p. 32

Lista de Tabelas

1	Resultados Globais da MLP	p. 28
2	Parâmetros das Configurações da MLP	p. 28
3	Resultados Específicos por Sujeitos da MLP	p. 28
4	Resultados Globais da ELM	p. 29
5	Parâmetros das Configurações da ELM	p. 29
6	Resultados Específicos por Sujeitos da ELM	p. 29
7	Resultados Globais da SVM	p. 30
8	Parâmetros das Configurações da SVM	p. 30
9	Resultados Específicos por Sujeitos da SVM	p. 30
10	Resultados Globais da ESN	p. 31
11	Parâmetros das Configurações da ESN	p. 31
12	Resultados Específicos por Sujeitos da ESN	p. 31
13	Teste de Wilcoxon	p. 32
14	Significado dos Símbolos do Teste de Wilcoxon	p. 32

Lista de Abreviaturas e Siglas

IBGE	<i>Instituto Brasileiro de Geografia e Estatística</i>
ICM	<i>Interface Cérebro-Máquina</i>
EEG	<i>Eletroencefalograma</i>
RC	<i>Reservoir Computing</i>
ESN	<i>Echo State Network</i>
MLP	<i>Multilayer Perceptron</i>
SLFN	<i>Single Layer Feedforward Network</i>
ELM	<i>Extreme Learning Machine</i>
SVM	<i>Support Vector Machine</i>
EMG	<i>Eletromiograma</i>
ECG	<i>Eletrocardiograma</i>
EOG	<i>Eletrooculografia</i>
ECOG	<i>Eletrocorticografia</i>
SNC	<i>Sistema Nervoso Central</i>
BMI	<i>Brain Machine Interface</i>
BCI	<i>Brain Computer Interface</i>
SNP	<i>Sistema Nervoso Periférico</i>
AVC	<i>Acidente Vascular Cerebral</i>
PEV	<i>Potencial Evocado Visual</i>
ELA	<i>Esclerose Lateral Amiotrófica</i>
ER	<i>Eventos Relacionados</i>
SNR	<i>Signal-to-Noise Rate</i>
RNA	<i>Rede Neural Artificial</i>
VC	<i>Vapnik-Chervonenkis</i>
RNR	<i>Rede Neural Recorrente</i>
AM	<i>Aprendizagem de Máquina</i>
ESN	<i>Echo State Network</i>
LSM	<i>Liquid State Machine</i>
EMQ	<i>Erro Médio Quadrático</i>

1 *Introdução*

1.1 **Qualificação do Problema**

De acordo com o censo demográfico 2010 (IBGE. . . , 2010), realizado pelo IBGE (Instituto Brasileiro de Geografia e Estatística), cerca de 45,6 milhões de pessoas – correspondendo a 23,9% da população brasileira – declararam-se portadoras de alguma deficiência, ocorrendo 35.774.392 casos de deficiência visual, 9.717.318 de deficiência auditiva, 13.265.599 de deficiência motora e 2.611.536 de deficiência mental/intelectual. Portanto, seria de grande valia qualquer assistência que reduzisse os impactos causados por essas deficiências nas vidas pessoas, aumentando assim a qualidade de vida de portadores dos mais diversos tipos de deficiências.

Com o avanço nas áreas de Reconhecimento de Padrões, Processamento de Sinais e Neuroengenharia, tornou-se possível a concepção de sistemas que realizem a interação de humanos com máquinas de maneira natural. A partir da análise de sinais biológicos, pode-se extrair informações que possibilitem a realização de tarefas cujas pessoas com determinadas deficiências teriam dificuldade, ou até mesmo seriam impossibilitadas de fazê-las.

Neste trabalho, a busca pela solução dos problemas mencionados acima foi através da Interface Cérebro-Máquina (ICM), utilizando sinais extraídos de eletroencefalogramas (EEG).

1.2 **Objetivos**

O objetivo deste trabalho foi analisar a importância temporal, espacial e combinacional de sinais cerebrais, especificamente adquiridos através de EEGs, através do uso de técnicas inteligentes para a classificação de tarefas imaginadas.

1.2.1 Objetivos Específicos

- Avaliou-se o uso das seguintes técnicas para classificar os sinais de EEG:
 - *Reservoir Computing*(RC) com o algoritmo de aprendizagem para *Echo State Network* (ESN).
 - Rede Neural Perceptron de Múltiplas Camadas (MLP, *Multilayer Perceptron*) com o algoritmo de treinamento *backpropagation*.
 - Rede Neural Alimentada Adiante de Camada Única (SLFN, *Single Layer Feedforward Networks*) com o algoritmo de treinamento da Máquina de Aprendizado Extremo (ELM, *Extreme Learning Machine*).
 - Máquina de Vetor de Suporte (SVM, *Support Vector Machine*) utilizando a Função de Base Radial (RBF, *Radial Basis Function*).
- Avaliou-se a influência temporal, espacial e combinacional a partir dos resultados obtidos das técnicas citadas acima.
- Realizou-se comparações entre os resultados dos algoritmos através do teste estatístico de Wilcoxon.

1.3 Estrutura da Monografia

Este trabalho está dividido em 5 capítulos. No Capítulo 2 é discutida a fundamentação teórica acerca de aquisição de biosinais, Interface Cérebro-Máquina, com seus conceitos e aplicações, o funcionamento e descrição de parâmetros dos classificadores: Perceptron de Múltiplas Camadas, Máquina de Aprendizado Extremo, Máquinas de Vetor de Suporte e *Reservoir Computing*. No Capítulo 3, é apresentado os modelos com configurações propostas para realizar os testes, as entradas, o pré-processamento e saídas dos sistemas de ICM, a forma que foi utilizada o EEG para realizar a classificação dos movimentos imaginados, o banco de dados utilizado e a forma como as simulações foram conduzidas. Em seguida, no Capítulo 4 são apresentados os resultados, análises estatísticas e discussões em cima das mesmas. Finalizando com o Capítulo 5, onde são concluídas as proposições que este trabalho se propôs a analisar e listados possíveis trabalhos futuros.

2 *Referencial Teórico*

2.1 Aquisição de Sinais Biológicos

Sinais biológicos, também conhecidos como biosinais, são sinais que podem ser mensurados e monitorados a partir de seres biológicos. A aquisição de biosinais normalmente são leituras nas variações de correntes elétricas produzidas por um tecido especializado, órgão ou sistema, que são capturadas por eletrodos colocados em regiões previamente estabelecidas.

Entre as formas mais comuns da aquisição de biosinais elétricos, destacam-se:

2.1.1 Eletromiograma (EMG)

É a representação da corrente elétrica gerada na membrana das fibras musculares, associada a contração do músculo, que se propaga pelos tecidos até chegar aos eletrodos de superfície ou em agulhas que penetram o músculo. Um sistema de aquisição do sinal eletromiográfico utilizando eletrodos de superfície é mostrado na Figura 1.



Figura 1: Eletromiograma
[Fonte: Extraído de (LAB...,)]

2.1.2 Eletrocardiograma (ECG)

É o registro gráfico das atividades elétricas do coração por meio dos eletrodos posicionados no corpo do sujeito. Pode ser visto um sistema que captura os sinais eletrocardiográficos na Figura 2.



Figura 2: Eletrocardiograma
[Fonte: Extraído de (CAEP,)]

2.1.3 Eletrooculografia (EOG)

É a medição do potencial existente entre a córnea e a retina através do uso de dois eletrodos em volta do olho e mais um para o sinal de referência (terra), está associada ao movimento do olho. A Figura 3 mostra um sistema que captura os sinais eletrooculográficos.



Figura 3: Eletrooculografia
[Fonte: Extraído de (EOG...,)]

2.1.4 Eletrocorticografia (ECOG)

Utiliza-se uma matriz de eletrodos em contato direto com a superfície do cérebro para registrar atividades elétricas do córtex cerebral, embora seja um método considerado

bastante invasivo, o ruído do sinal adquirido é bem pequeno. A Figura 4 mostra a instalação dos eletrodos para a captação dos sinais eletrocorticográficos.

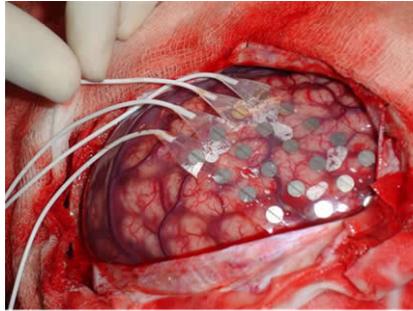


Figura 4: Instalação da Matriz de Eletrodos para Realização do ECOG

[Fonte: Extraído de (CORTICOGRAFIA...,)]

2.1.5 Eletroencefalografia (EEG)

Usa eletrodos em contato com o couro cabeludo a fim de perceber a atividade elétrica de regiões do cérebro. Ele vem sendo utilizado para estudar a organização do cérebro em processos cognitivos, como: percepção, memória, atenção, linguagens, etc (TEPLAN, 2002). A Figura 5 mostra a utilização de um capacete com eletrodos posicionados para a aquisição dos sinais eletroencefalográficos que, diferentemente do ECOG, é um método de aquisição não invasivo, com uma relação sinal-ruído baixa.



Figura 5: Eletroencefalografia

[Fonte: Extraído de (EEG...,)]

2.2 Interface Cérebro-Máquina

De acordo com o presente entendimento, o papel do Sistema Nervoso Central (SNC) é responder a ocorrências no ambiente ou no corpo produzindo saídas apropriadas. As saídas naturais do SNC são neuromusculares ou hormonais (HE, 2013). A Interface Cérebro-Máquina (ICM), do inglês *BMI - Brain Machine Interface*, também conhecido como *BCI - Brain Computer Interface*, são sistemas que utilizam medições da atividade do SNC e as converte em saídas artificiais que substitui, restaura, aprimora, complementa ou melhora as saídas naturais do SNC, alterando assim as interações entre o SNC e seu ambiente interno ou externo (WOLPAW; WOLPAW, 2012). A efetividade da ICM depende da capacidade do usuário produzir sinais cerebrais que reflitam a intenção e que o sistema ICM possa decodificar com confiança e precisão em comandos de saída que atendam àquela intenção (WOLPAW et al., 2002; CURRAN; STOKES, 2003).

Para o correto entendimento da ICM, é necessário saber que o SNC é composto pelo cérebro e pela medula espinhal, diferindo do Sistema Nervoso Periférico (SNP), o qual é composto de nervos periféricos, gânglios e receptores sensoriais. As estruturas do SNC estão localizadas dentro das coberturas das meninges, e têm a função de integrar as mais diversas entradas sensoriais para produzir saídas motoras efetivas. Já as estruturas do SNP não estão dentro das meninges e servem principalmente para trazer as entradas sensoriais para SNC e transportar as saídas motoras a partir deste (HE, 2013).

As atividades do SNC incluem fenômenos eletrofisiológicos, neuroquímicos e metabólicos (como potenciais de ação neuronal, potenciais sinápticos, liberação de neurotransmissores e consumo do oxigênio), que ocorrem continuamente no SNC. Esses fenômenos podem ser monitorados através da medição de campos elétricos ou magnéticos, oxigenação da hemoglobina, ou outros parâmetros através de sensores colocados no couro cabeludo, na superfície do cérebro, ou até mesmo dentro do cérebro. Um sistema ICM compreende várias etapas: registro dos sinais, pré-processamento, extração características e, por fim, classificação para a geração de saídas artificiais que atuam sobre o ambiente ou sobre o próprio corpo, como ilustrado na Figura 6.

Os referidos sistemas podem:

- **Substituir** saídas naturais que foram perdidas por conta de lesões ou doenças. Assim, alguém com membros paralisados poderia usar a ICM para operar uma cadeira de rodas.

- **Restaurar** saídas naturais perdidas. Portanto, alguém que tenha perdido a função da bexiga por conta de esclerose múltipla poderia usar a ICM para estimular o nervo periférico, controlando a bexiga para a produção da urina.
- **Aprimorar** a saída do SNC. Deste modo, alguém que esteja fazendo alguma tarefa que necessita de atenção máxima por um longo período de tempo, por exemplo: dirigindo um carro, poderia usar a ICM para detectar falhas na atenção, produzindo alguma saída para alertar o usuário a restaurar a atenção necessária.
- **Complementar** a saída natural do SNC. Desta forma, alguém poderia utilizar a ICM para controlar um terceiro braço (robótico).
- **Melhorar** a saída natural do SNC. Nestas condições, um usuário que tem o movimento dos braços comprometido por um Acidente Vascular Cerebral (AVC) que danificou o córtex sensorimotor, utilizaria a ICM para mensurar os sinais das áreas danificadas, excitando os músculos e controlando uma órtese que melhoraria o movimento deste braço.

Segundo He (2013), Existem três tipos de ICMs conhecidas:

- **ICM Dependente:** Depende das atividades nas saídas naturais do SNC para a geração da atividade cerebral que levará as informações para o sistema ICM. Um exemplo de sistema desse tipo foi proposto por Sutter (1992), que apresentou uma matriz bi-dimensional com letras que piscam, uma de cada vez, e o usuário seleciona a letra apenas fixando o olhar na mesma, então o Potencial Evocado Visual (PEV), sinal adquirido na área do córtex visual, fica muito maior quando a letra em que o usuário está prestando atenção pisca do que se comparado com as outras letras piscando. Neste caso, o canal de saída do cérebro é o EEG, mas a geração da atividade que o EEG representa depende da direção do olhar e, portanto, dos músculos extra-oculares e dos nervos cranianos que os ativam. Embora a ICM dependente não dá ao cérebro uma nova via de comunicação, pode ser muito útil.
- **ICM Independente:** Não depende das saídas naturais do SNC, atividade muscular não é necessária para gerar os sinais cerebrais. É essencialmente um método alternativo para a detecção de mensagens carregadas, sendo bastante útil para pessoas deficientes.
- **ICM Híbrida:** utiliza sinais de característica dependente e independente, por exemplo: PEV e EEG para realizar alguma atividade.

A ICM vem ganhando muita atenção dos pesquisadores nos últimos anos devido a utilidade que esses sistemas podem ter para pessoas com deficiência. Especialmente em pacientes que possuem deficiências neuromusculares, como: Esclerose Lateral Amiotrófica (ELA), uma doença que acaba destruindo completamente o sistema motor central e periférico, normalmente de causa desconhecida. Embora não afetem com tanta intensidade as funções cognitivas e sensoriais, atualmente essa é uma doença ainda sem cura (SMITH, 1992); Síndrome do Encarceramento, do inglês *locked-in*, uma doença neurológica rara que ocorre a paralisia de todos os músculos do corpo; AVC, que na maioria dos casos leva à dificuldade de mobilização de um dos lados do corpo; Lesão Medular, que compromete a função motora; etc.

Por meio da ICM, Donchin, Spencer e Wijesinghe (2000) conseguiram fazer pessoas operarem um processador de textos que permite o usuário selecionar uma letra a partir de uma grade bi-dimensional com todas as letras do alfabeto; Kennedy et al. (2000) proveram aos usuários que possuíam Síndrome do Encarceramento um cursor bi-dimensional, habilitando o controle das coordenadas X e Y do *mouse*; Murugappan (2010) classificou o estado emocional através da transformada *wavelet* discreta dos sinais de EEG; Tavella et al. (2010) controlaram uma neuroprótese da mão; Huang et al. (2012) controlaram uma cadeira de rodas elétrica, baseando-se nos Eventos Relacionados (ER) extraídos de EEGs; Lee et al. (2012) controlaram a direção de robôs; Nicoletis (2012), em uma das mais recentes pesquisas, está tentando fazer um paraplégico andar através de um exoesqueleto.

O modelo de um sistema ICM genérico pode ser visto na Figura 6.

Segundo Wolpaw et al. (2002), a maioria das pesquisas em ICM utilizam características eletrofisiológicas dos sinais que representam eventos cerebrais que são bem definidas anatomicamente e fisiologicamente. Estas incluem os: *ritmos* refletindo oscilações em circuitos neurais particulares (por exemplo, ritmos *mu* ou *beta* do córtex motor; Potenciais Evocados por estímulos particulares (por exemplo, P300 e PEV). As características dos sinais da ICM não são canais de saída do cérebro naturais ou normais. Elas são canais artificiais de saída criadas pelos sistemas ICM.

O treino dos usuários é um dos fatores mais importantes que afetam a ICM. Como qualquer outro sistema de comunicação, o desempenho da ICM é totalmente dependente da relação sinal-ruído (SNR, *Signal-to-Noise Rate*). Os usuários tendem a diferir na proeminência e estabilidade de características específicas dos sinais e nas suas capacidades para inicialmente demonstrar ou adquirir e, subsequentemente, manter o controle sobre essas características. Portanto, métodos ICM devem ser testados em mais de um sujeito,

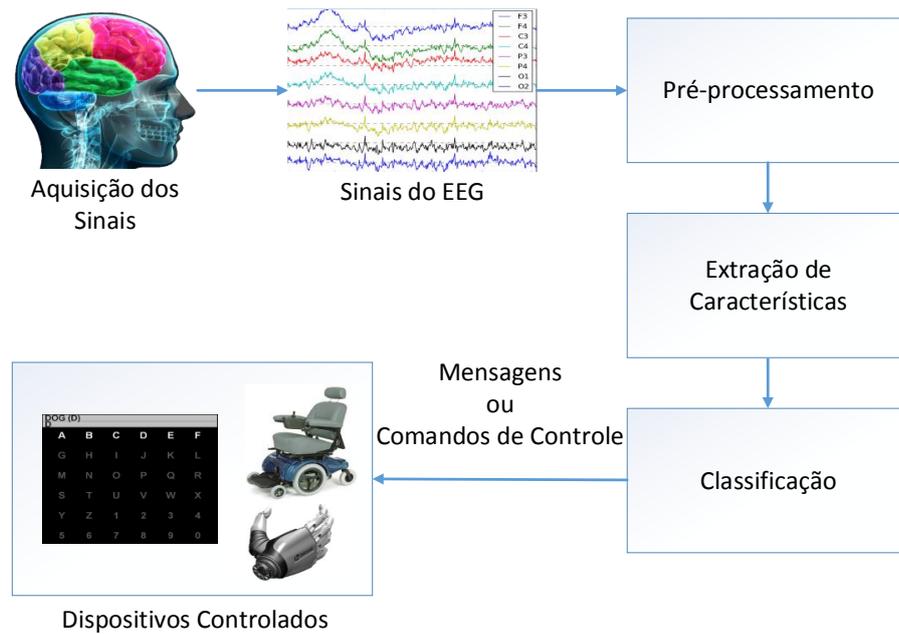


Figura 6: Modelo de Sistema ICM Genérico

além disso, é importante que os dados sejam adquiridos varias vezes e em diferentes períodos. Assim, sistemas ICM devem ser capazes de identificar, acomodar e encorajar as características dos sinais que melhor atendem às necessidades do usuário.

Análises *offline* dos dados de ICM normalmente sugerem métodos que devem funcionar melhor na análise *online*, embora não possam prever os efeitos de curto e longo prazo entre os métodos e o *feedback* do usuário. Por isso, os métodos que aparentemente são promissores na análise *offline* devem ser testados intensivamente nas análises *online*.

2.3 Classificadores de Padrões Instantâneos

Neste trabalho serão considerados classificadores de padrões instantâneos aqueles que classificam padrões baseando-se apenas na combinação das entradas no instante atual.

2.3.1 Perceptron de Múltiplas Camadas

As Redes Perceptron de Múltiplas Camadas (MLP, *Multilayer Perceptron*) são sistemas conexionistas que possuem em sua estrutura um conjunto de unidades de processamento (neurônios artificiais) interconectadas, onde para cada unidade de processamento existe uma função de ativação (normalmente não-linear). As MLPs são muito utilizadas

devido ao fato delas possuírem a capacidade de aprender o comportamento de funções, esse aprendizado é armazenado nos pesos presentes nas interconexões entre as unidades de processamento VALENCA (2010).

A MLP é uma das arquiteturas de Redes Neurais Artificiais (RNA) mais utilizadas pela simplicidade e poder que a mesma detém. As características principais deste tipo de rede são:

- Rede do tipo *feedforward*, ou seja, que propaga os sinais em um único sentido durante a utilização da rede, após o treinamento.
- Apresenta pelo menos três camadas:
 1. Uma **Camada de Entrada**: cada neurônio representa uma variável de entrada para a resolução do problema.
 2. Uma ou mais **Camada(s) Intermediária(s)**: são responsáveis pela não linearidade da rede. Os neurônios que fazem parte desta camada normalmente possuem funções de ativação sigmoidais, as funções mais utilizadas na literatura são: tangente hiperbólica e sigmoide logística. A presença desse tipo de camada faz com que as RNAs resolvam problemas não linearmente separáveis, sendo possível aproximar qualquer função contínua, com uma única camada, e qualquer função matemática, com mais de uma camada (HAYKIN, 2007).
 3. Uma **Camada de Saída**: Essa é a camada que representa a resposta do problema para as variáveis de entradas apresentadas. Os neurônios desta camada também possuem funções de ativação sigmoidais ou simplesmente uma função linear.
- Possui Aprendizado Supervisionado: a MLP precisa ser apresentada para um conjunto de entradas e saídas conhecidas durante o seu aprendizado para que seja possível calcular o erro obtido pelo sistema, o qual é usado para reajustar os parâmetros e tentar atingir um ponto de aprendizagem satisfatório para a resolução do problema.

A Figura 7 mostra a arquitetura de uma MLP com: N neurônios na camada de entrada e 1 neurônio que representa o limiar (*bias*), 1 camada intermediária com M neurônios e 1 neurônio representando o limiar (*bias*), e P neurônios na camada de saída.

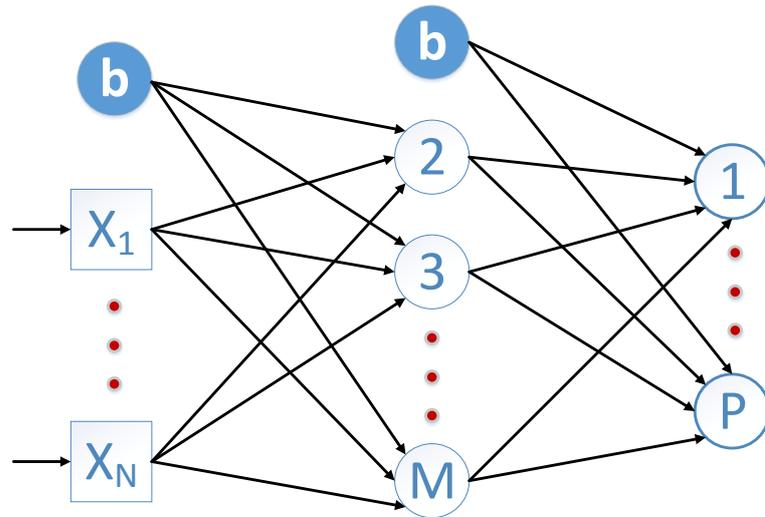


Figura 7: Arquitetura da MLP com uma Camada Intermediária

A rede é inicializada com os seus pesos aleatórios, normalmente próximos de 0, com o objetivo de não saturar a saída das funções de ativação (VALENCA, 2010). Para que as redes neurais aprendam, é necessário algum algoritmo de treinamento. Um dos mais usados é conhecido como: *backpropagation*, que ajusta os pesos da rede através de uma propagação recursiva dos erros. Basicamente, este algoritmo é constituído de 2 passos (VALENCA, 2010):

1. Propagam-se as entradas no sentido progressivo (a partir da camada de entrada no sentido da camada de saída), calcula-se a saída e o erro, mantendo os pesos fixos.
2. Propaga os erros no sentido contrário (da camada de saída para a camada de entrada) e os pesos são ajustados através da Equação 2.1.

$$w_{ij}^m(\text{novo}) = w_{ij}^m(\text{antigo}) + \alpha \delta_i^m f^{m-1}(\text{net}_j^{m-1}), \quad (2.1)$$

onde: w_{ij}^m é o peso que liga o neurônio emissor j da camada m ao neurônio i , α é a taxa de aprendizagem, f^{m-1} são os sinais emitidos pelos neurônios da camada anterior e δ_i^m é a sensibilidade.

Caso o neurônio de índice j da Equação 2.1 seja da camada intermediária, δ_i é calculado conforme a Equação 2.2.

$$\delta_j^{m-1} = f'^{m-1}(net_j^{m-1}) \sum_{i=1}^N w_{ij}^m \delta_i^m, \quad (2.2)$$

onde δ_i^m é a sensibilidade do neurônio i da camada m , w_{ij}^m é o peso da conexão entre os neurônios i e j , N é o número de neurônios na camada.

O cálculo da sensibilidade dos neurônios pertencentes à camada de saída é calculados a partir da Equação 2.3

$$\delta_i^m = (d_i - y_i) f'(net_i), \quad (2.3)$$

onde d_i é a saída desejada, y_i é a saída obtida e $f'(net_i)$ é a derivada da função de ativação dos neurônios da camada de saída.

Outro ponto muito importante no treinamento de RNAs é o critério de parada, essa estratégia é tomada para que se evite o superajustamento dos pesos. Para maiores detalhes de implementação do *backpropagation*, consultar (VALENCA, 2010; HAYKIN, 2007).

2.3.2 Máquinas de Aprendizado Extremo

A Máquina de Aprendizado Extremo (ELM, *Extreme Learning Machine*) é uma abordagem de treinamento para Redes Neurais Alimentadas Adiante de Camada (intermediária) Única (SLFN, *Single Layer Feedforward Networks*), ou seja, redes que apresentam as camadas: *entrada*, *oculta*, *saída*. Esta técnica trata a SLFN como um sistema linear e seus pesos da camada de *saída* – que ligam a camada *oculta* com a de *saída* – podem ser analiticamente determinados através da inversão de matrizes (HUANG; ZHU; SIEW, 2006).

Para N exemplos (x_i, t_i) , onde $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$ e $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathbb{R}^m$, SLFNs padrão com K nós ocultos e funções de ativação $g(x)$ são modeladas matematicamente através da Equação 2.4

$$\sum_{i=1}^K \beta_i g_i(\mathbf{x}_j) = \sum_{i=1}^K \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{o}_j, \quad j = 1, \dots, K, \quad (2.4)$$

onde $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ é um vetor de pesos que conecta o i -ésimo nó oculto e os nós de entrada, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ é o vetor de pesos que conecta o i -ésimo nó oculto com os nós de saída e b_i é o limiar do i -ésimo nó oculto. $\mathbf{w}_i \cdot \mathbf{x}_j$ denota o produto interno entre \mathbf{w}_i e \mathbf{x}_j .

As equações acima podem ser representadas matricialmente através da seguinte equa-

ção:

$$\mathbf{H}\beta = \mathbf{T} \quad (2.5)$$

$$\mathbf{H}_{N \times K} = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_K \cdot x_1 + b_K) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_K \cdot x_N + b_K) \end{bmatrix}_{N \times K} \quad (2.6)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_K^T \end{bmatrix}_{K \times m}, \quad \mathbf{T} = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}, \quad (2.7)$$

onde \mathbf{H} é conhecida como a matriz de saída da camada oculta. A i -ésima coluna de \mathbf{H} é a saída do i -ésimo nó da camada oculta que se relaciona com as entradas $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$

Se $K = N$, então a matriz \mathbf{H} é quadrada e inversível quando o vetor de pesos \mathbf{w}_i e os *bias* b_i são escolhidos aleatoriamente. Nesta condição, a SLFN pode aproximar os exemplos de treinamento com o erro zero. Na prática, $K \ll N$, fazendo com que \mathbf{H} não seja uma matriz quadrada e possa ser que não exista $\mathbf{w}_i, b_i, \beta_i$ ($i = 1, \dots, K$) que satisfaça $\mathbf{H}\beta = \mathbf{T}$ (HUANG; ZHU; SIEW, 2006). Dessa forma, a solução para esse problema é calcular $\hat{\beta} = \mathbf{H}^\dagger \mathbf{T}$, onde \mathbf{H}^\dagger é a matriz *inversa generalizada* de Moore-Penrose (BEN-ISRAEL; GREVILLE, 2003).

O treinamento da ELM está descrito no Algoritmo 1.

Algoritmo 1: Passos do treinamento da ELM

Data: Dado um conjunto de treinamento

$\mathfrak{N} = \{(x_i, t_i) | x_i \in \mathbb{R}^n, t_i \in \mathbb{R}^m, (i = 1, \dots, N)\}$, a função de ativação $g(x)$, infinitamente diferenciável, e o número de nós na camada oculta K

- 1 Gerar aleatoriamente os pesos \mathbf{w}_i e os *bias* $b_i, i = 1, \dots, K$;
 - 2 Calcular a matriz de saída da camada oculta (\mathbf{H});
 - 3 Calcular os pesos da camada de saída $\hat{\beta} = \mathbf{H}^\dagger \mathbf{T}$
-

2.3.3 Máquinas de Vetor de Suporte

As Máquinas de Vetor de Suporte (SVM, *Support Vector Machine*) têm uma abordagem desenvolvida baseada na teoria do aprendizado estatístico (VAPNIK; CHERVONENKIS, 1971; VAPNIK, 1995; VAPNIK, 1998). Ela busca dois objetivos: minimizar o erro de treinamento e diminuir a complexidade estrutural do modelo a ser obtido (VALENCA, 2010).

A teoria do aprendizado estatístico afirma que o erro de generalização, conhecido como *Risco Funcional*, pode ser obtido caso seja conhecida a distribuição de probabilidade da população Vapnik (1998). Porém, geralmente em problemas reais, apenas existem amostras da população, o erro calculado a partir dessas amostras é conhecido como *Risco Empírico*. A dimensão *Vapnik-Chervonenkis* (VC) é a representação do número máximo de exemplos de treinamento que uma máquina de aprendizagem é capaz de classificar corretamente, para todas as possíveis combinações binárias desses dados (VALENCA, 2010).

De acordo com Vapnik (1998), qualquer máquina de aprendizagem, durante o aprendizado supervisionado, consegue obter a capacidade máxima de generalização quando minimiza-se o *Risco Funcional*, que equivale a minimizar o *Risco Empírico* e a complexidade do modelo, conhecida como *Risco Estrutural*.

A abordagem inicial das SVMs, que tratavam apenas de problemas linearmente separáveis, ficou conhecida como SVMs de *margens rígidas*. Mais tarde, surgiram as SVMs de *margens suaves*, que lidavam com treinamentos com erro ou ruidosos. Posteriormente, surgiu as SVMs que tratavam de problemas não lineares através da utilização de *kernels* (VALENCA, 2010).

Segundo (RUSSELL; NORVIG, 2003), as SVMs são muito utilizadas devido às seguintes características:

- Criam separadores de margens máximas, ou seja, fronteiras de decisão com a maior distância possível entre as classes diferentes, ajudando assim na generalização do modelo.
- Criam hiperplanos de separação lineares, mas elas também podem colocar os dados em um espaço de alta dimensão através da utilização de funções de *kernel*. A SVM se beneficia do fato que, normalmente, dados que não são facilmente separáveis no espaço original de entrada, são facilmente separáveis em dimensões mais altas.
- São métodos *não-paramétricos*, ou seja, elas retêm exemplos de treinamento. Na prática, elas apenas retêm uma pequena fração de exemplos de treinamento ao final do processo de aprendizagem. Portanto, as SVMs combinam as vantagens de modelos *paramétricos* e *não-paramétricos*: elas têm a flexibilidade de representar funções complexas, mas são resistentes ao *superajustamento*.

A Figura 8 mostra as fronteiras de decisões de uma SVM que aceita vetores de entrada

$x_i (i = 1, \dots, n)$, onde n é o número de exemplos, e provê saídas $y_i (-1$ ou $+1)$.

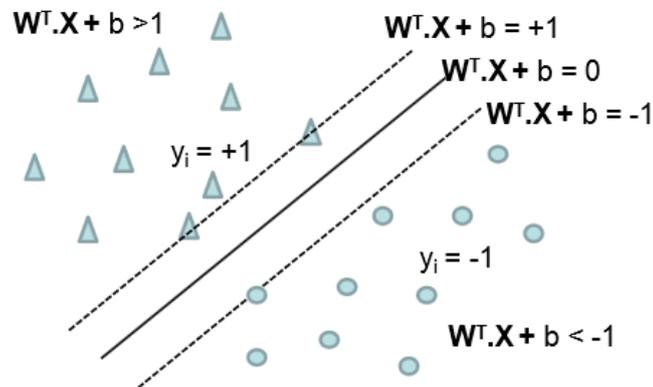


Figura 8: Fronteiras de Decisão

[Fonte: Extraído de (VALENCA, 2010)]

Maiores detalhes sobre as teorias e cálculos realizados durante o treinamento das SVMs podem ser vistos em (BURGES, 1998).

2.4 Classificadores de Padrões Temporais

Neste trabalho serão considerados classificadores de padrões temporais aqueles que classificam padrões baseando-se na combinação das entradas no instante atual e nos instantes anteriores.

2.4.1 Reservoir Computing

Outro tipo de RNA é a Rede Neural Recorrente (RNR). Neste tipo de rede, existem situações de recorrência entre os neurônios. Essa recorrência entre os neurônios faz com que as RNRs possam preservar no seus estados internos uma transformação não linear do histórico de entradas, ou seja, uma *memória dinâmica*, sendo assim capaz de processar informações de contexto temporal. Matematicamente, isso significa que as RNRs podem ser vistas como *sistemas dinâmicos*, enquanto as redes *feedforward*, como a MLP, são *funções* (LUKOSEVICIUS; JAEGER, 2009). Embora elas apresentem uma ferramenta genérica poderosa, o seu uso não era muito comum devido a dificuldade que existia para o treinamento dessas arquiteturas. Também acredita-se que as RNR são os modelos de Aprendizagem de Máquina (AM) que mais se assemelham ao cérebro biológico (LUKOSEVICIUS, 2012).

O algoritmo *backpropagation* para o treinamento de redes estáticas foi estendido para

as RNRs (WERBOS, 1990), mas obtiveram sucesso parcial. Foram propostas novas alternativas para o entendimento, treinamento e uso de RNRs através das técnicas: Rede de Estado de Eco (ESN, *Echo State Network*) (JAEGER, 2001; JAEGER, 2007) na AM e Máquina de Estado Líquido (LSM, *Liquid State Machine*) (MAASS; NATSCHLAGER; MARKRAM, 2002) no campo da Neurociência. Foi mostrado que as RNRs frequentemente funcionam bem o bastante mesmo sem o ajustamento de todos os pesos da rede. Na ESN clássica, a camada de recorrência (chamada de *reservoir*) é gerada aleatoriamente, e apenas a camada de saída é treinada. Esta abordagem levou a excelentes performances em alguns *benchmarks* (JAEGER, 2001; JAEGER, 2004; TRIEFENBACH et al., 2010).

A tendência começada pelas ESNs e LSMs se tornaram conhecidas como *Reservoir Computing* (RC) (VERSTRAETEN et al., 2007).

2.4.1.1 Modelo Básico

As redes ESN são aplicadas em tarefas temporais supervisionadas de AM, onde para um dado sinal de entrada de treinamento $\mathbf{u}(n) \in \mathbb{R}^{N_u}$, um sinal de saída alvo $y^{alvo}(n) \in \mathbb{R}^{N_y}$ é conhecido. Aqui, $n = 1, \dots, T$ é o tempo discreto e T é o número de pontos de dados que faz parte do conjunto de treinamento. A objetivo é aprender um modelo com saída $\mathbf{y}(n) \in \mathbb{R}^{N_y}$, onde $\mathbf{y}(n)$ corresponde ao valor mais próximo possível de $y^{alvo}(n)$, minimizando o a medida do erro $E(\mathbf{y}, \mathbf{y}^{alvo})$, e, o mais importante: generalizando bem para dados que não foram apresentados. a medida do erro E é tipicamente o erro médio quadrático (EMQ) apresentado na Equação 2.8.

$$E(\mathbf{y}, \mathbf{y}^{alvo}) = \frac{1}{N} \sum_{i=1}^{N_y} \sqrt{\frac{1}{T} \sum_{n=1}^T (y_i(n) - y_i^{target}(n))^2} \quad (2.8)$$

As ESNs utilizam as seguintes equações de atualização, com as unidades de valores contínuos, de tempo discreto e com vazamento integrado:

$$\tilde{\mathbf{x}}(n) = \tanh(W^{entrada}[\mathbf{1}; \mathbf{u}(n)] + \mathbf{W}\mathbf{x}(n-1)) \quad (2.9)$$

$$\mathbf{x}(n) = (1 - \alpha)\mathbf{x}(n-1) + a\tilde{\mathbf{x}}(n), \quad (2.10)$$

onde n é o tempo discreto, $\mathbf{u}(n) \in \mathbb{R}^{N_u}$ é o sinal de entrada, $\mathbf{x}(n) \in \mathbb{R}^{N_x}$ é o vetor de ativação dos neurônios do *reservoir* e $\tilde{\mathbf{x}}(n) \in \mathbb{R}^{N_x}$ é sua atualização, todos no tempo n , a função $\tanh(\cdot)$ é aplicada elemento por elemento, $[\cdot; \cdot]$ é a concatenação vertical entre

vetores (ou matrizes), $\mathbf{W}^{entrada} \in \mathbb{R}^{N_x \times (1+N_u)}$ e $\mathbf{W} \in \mathbb{R}^{N_x \times N_x}$ são, respectivamente, as matrizes de peso da *entrada* e do *reservoir*, $\alpha \in (0, 1]$ é a *taxa de vazamento*. O modelo também é frequentemente usado sem a *taxa de vazamento*, que é o caso especial quando $\alpha = 1$ e, portanto, $\tilde{\mathbf{x}}(n) \equiv \mathbf{x}(n)$. A camada de saída linear é definida como:

$$\mathbf{y}(n) = \mathbf{W}^{saída}[1; \mathbf{u}(n); \mathbf{x}(n)], \quad (2.11)$$

onde $\mathbf{y}(n) \in \mathbb{R}^{N_y}$ é a saída da ESN, e $\mathbf{W}^{saída} \in \mathbb{R}^{(1+N_u+N_x)}$ é a matriz de pesos de *saída*.

A Figura 9 mostra a estrutura de uma ESN básica.

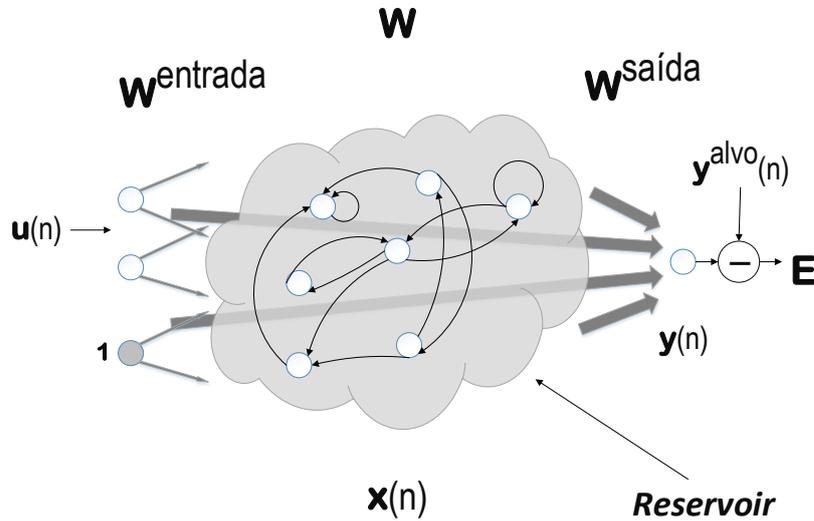


Figura 9: Estrutura de uma ESN Básica
[Fonte: Elaboração própria.]

Os passos para a utilização da ESN são descritos no Algoritmo 2:

Algoritmo 2: Passos da utilização da ESN

- 1 Gerar o *reservoir* ($\mathbf{W}^{entrada}, \mathbf{W}, \alpha$), de forma aleatória com os pesos entre $[-0.5, 0.5]$;
 - 2 Realizar a fase de “aquecimento”;
 - 3 Apresentar os dados de treinamento $\mathbf{u}(n)$ ao *reservoir* e coletar os estados das ativações $\mathbf{x}(n)$;
 - 4 Computar os pesos da camada de saída $\mathbf{W}^{saída}$ do *reservoir* utilizando regressão linear, minimizando o erro médio quadrático da saída $\mathbf{y}(n)$ da rede com relação a saída desejada $\mathbf{y}^{alvo}(n)$;
 - 5 Utilizar a ESN treinada em novos dados $\mathbf{u}(n)$, computando as saídas $\mathbf{y}(n)$ empregando os pesos da camada de saída $\mathbf{W}^{saída}$;
-

2.4.1.2 Parâmetros da ESN

O *reservoir* é definido pela tupla $(\mathbf{W}^{entrada}, \mathbf{W}, \alpha)$. As matrizes $\mathbf{W}^{entrada}$ e \mathbf{W} são geradas aleatoriamente de acordo com alguns parâmetros que serão explicados a seguir e α é um parâmetro livre conhecido como *taxa de vazamento*. Os parâmetros globais que precisam ser definidos são: O *tamanho* N_x , *esparsidade* de \mathbf{W} , *raio espectral* de \mathbf{W} , *taxa de vazamento* e o *número de passos do aquecimento*.

2.4.1.2.1 Tamanho do *Reservoir* Trata-se do número de unidades N_x que pertencem ao *reservoir*. Geralmente, quanto maior o *reservoir*, mais fácil é de se obter performance, desde que medidas apropriadas de regularizações sejam tomadas para evitar o superajustamento. Quanto maior o espaço dos sinais $x(n)$ do *reservoir*, mais fácil é a busca por uma combinação linear dos sinais para aproximar $y^{alvo}(n)$.

2.4.1.2.2 Esparsidade do *Reservoir* Parâmetro que define a quantidade de células da matriz serem iguais a zero. Útil para aumentar a velocidade de computação dos pesos, caso representações de matrizes esparsas possam ser utilizadas.

2.4.1.2.3 Raio Espectral É definido como o máximo autovalor absoluto da matriz \mathbf{W} desejado. Pode ser interpretado como o determinante do quão rápido a influência de uma entrada deixa de existir no *reservoir* com o tempo. A matriz de pesos \mathbf{W} é gerada aleatoriamente e computado o seu raio espectral $\rho(\mathbf{W})$, então divide-se \mathbf{W} por $\rho(\mathbf{W})$ para obter uma matriz com o raio espectral unitário, então essa nova matriz de pesos é escalada com o raio espectral pré-estabelecido. Para a ESN funcionar corretamente, o *reservoir* deve garantir a *Propriedade do Estado de Eco*, do inglês *Echo State Property*, que pode ser descrita como o desvanecimento de memória da entrada: a trajetória do estado do *reservoir* deve convergir dado a mesma entrada, independentemente das entradas anteriores (LUKOSEVICIUS; JAEGER; SCHRAUWEN, 2012), ou seja, para uma entrada $u(n)$ longa o bastante, o estado do *reservoir* $x(n)$ não deve depender das condições do estado antes da entrada $x(n)$ (LUKOSEVICIUS, 2012). Tipicamente, isso é assegurado através da escolha do *raio espectral*, na maioria das situações, a escolha de $\rho(\mathbf{W}) < 1$ faz com que essa propriedade seja assegurada. Para se guiar, $\rho(\mathbf{W})$ deve ser maior para situações em que a saída dependa mais de entradas passadas, e menor se a saída depender mais do histórico recente das entradas. que necessitam de um histórico de entradas maiores para

2.4.1.2.4 Taxa de Vazamento A *Taxa de Vazamento* α pode ser interpretada como a velocidade de atualização da dinâmica do sistema, ou ainda como a dependência do estado atual com o estado anterior. Esse parâmetro governará o quanto do estado anterior o estado atual vai resgatar, como pode ser visto na Equação 2.10.

2.4.1.2.5 Número de Passos do Aquecimento Inicialmente no treinamento, os estados do RC assumem o valor 0. Verstraeten (2009) sugere que uma fase de “aquecimento” no RC seja realizada com a finalidade de esquecer o seu estado inicial. Nesta etapa do treinamento não é levado em conta os estados gerados para encontrar os pesos da camada de saída. Este parâmetro indicará quantos passos serão utilizados para a fase de “aquecimento”.

2.4.1.3 Treinamento da ESN

Normalmente a camada de saída da ESN é linear e *feedforward*, então a Equação 2.11 pode ser reescrita da seguinte forma:

$$\mathbf{Y} = \mathbf{W}^{saída} \mathbf{X}, \quad (2.12)$$

onde $\mathbf{Y} \in \mathbb{R}^{N_y \times T}$ são todos os $\mathbf{y}(n)$ e $\mathbf{X} \in \mathbb{R}^{(1+N_u+N_x) \times T}$ são todos os $[1; \mathbf{u}(n); \mathbf{x}(n)]$. A matriz \mathbf{X} também é conhecida como *Matriz de Design*.

Achar os pesos de $\mathbf{W}^{saída}$ é equivalente a encontrar as raízes do sistemas de equações lineares, normalmente sobredeterminados, ou seja, possuem mais equações do que incógnitas:

$$\mathbf{Y}^{alvo} = \mathbf{W}^{saída} \mathbf{X} \quad (2.13)$$

Para a resolução do Sistema de Equações 2.13, uma das técnicas mais aplicadas para o treinamento da ESN é conhecida como *Regressão Ridge*, do inglês *Ridge Regression*, ou *Regularização Tikhonov*, do inglês *Tikhonov Regularization*:

$$\mathbf{W}^{saída} = \mathbf{Y}^{alvo} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T + \beta \mathbf{I})^{-1}, \quad (2.14)$$

onde β é o *coeficiente de regularização*, que controla o que deve ser mais importante: menor erro de treinamento ou dificultar que o superajustamento aconteça.

3 *Propostas e Arranjo Experimental*

A hipótese considerada neste trabalho é de que as atividades cerebrais mensuradas através de EEGs são melhores classificadas se as informações temporais, espaciais e combinacionais dos eletrodos forem utilizadas simultaneamente. Para analisar a influência temporal, espacial e combinacional dos sinais, foram efetuadas simulações com classificadores de padrões instantâneos e de padrões temporais com diversas configurações.

3.1 Entradas dos Classificadores

Para realizar a classificação dos sinais adquiridos durante o EEG, os classificadores foram estruturados para que os n sinais cerebrais no instante t fossem a entrada, por exemplo: se cada tarefa imaginada pelo sujeito dura 10s e foram amostradas em 100Hz, então existirão 1000 pontos para cada canal; se for utilizado 5 canais para a classificação, implica dizer que para cada tarefa serão analisados 5000 pontos.

3.2 Pré-processamento

Sabe-se que é muito importante fazer a normalização dos dados para que as técnicas de aprendizagem de máquina funcionem corretamente. A realização da normalização é uma forma de assegurar que variáveis em intervalos numéricos diferentes tenham a mesma atenção durante o treinamento (VALENCA, 2010). As variáveis devem ser normalizadas tendo como base as funções de ativação do modelo utilizado. Neste trabalho, todas as técnicas utilizaram a função *tangente hiperbólica* formulada pela Equação 3.1, apresentada na Figura 10.

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3.1)$$

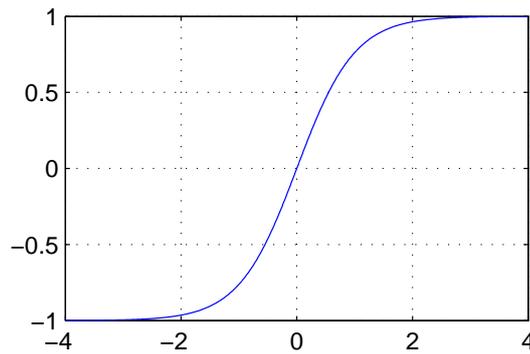


Figura 10: Tangente Hiperbólica

Uma vez que a *tangente hiperbólica* está confinada no intervalo $(-1, 1)$, a normalização dos dados foi feita para que os valores das variáveis ficassem dentro do intervalo $[-0.85, 0.85]$ utilizando a Equação 3.2.

$$y = (0.85 - (-0.85)) \frac{x_i - x_{min}}{x_{max} - x_{min}} + (-0.85) \quad (3.2)$$

3.3 Saídas dos Classificadores

As saídas estão representadas através de vetores onde a posição do vetor que indica a classe atual tem o valor 1 e o restante das posições assumem o valor 0. Durante o treinamento dos classificadores, as saídas representam a classe da tarefa no instante t , aproveitando o exemplo da Seção 3.1, se existem 2 possíveis classes, então elas são representadas por $[0, 1]$ e $[1, 0]$. Se na entrada dos classificadores estiverem sendo apresentados os 1000 instantes (5000 pontos) que representam a tarefa $[0, 1]$, então para cada instante apresentado, a saída desejada é $[0, 1]$. Os classificadores mapeiam uma classe para cada instante analisado, por isso, faz-se necessário que haja um mecanismo de decisão para julgar qual a classe da tarefa realizada. Neste trabalho assumiu-se que a classe correta é aquela que teve a maior média durante a análise da tarefa em questão.

3.4 Utilização dos EEGs

Os sinais cerebrais extraídos através dos EEGs, assim como na maioria dos biosinais, são *não estacionários, não lineares e ruidosos*, no inglês, diz-se que eles são sinais '**3N**' - *Nonstationary, Nonlinear, Noisy*. Não estacionário significa que as características estatísticas mudam com o tempo (KLONOWSKI, 2009). Para o registro de sinais cerebrais

a partir de EEGs, existem alguns sistemas que definem as posições em que os eletrodos devem se encontrar. Um desses sistemas é conhecido como *sistema 10-10* (JURCAK; TSUZUKI; DAN, 2007) que pode ser visto na Figura 11.

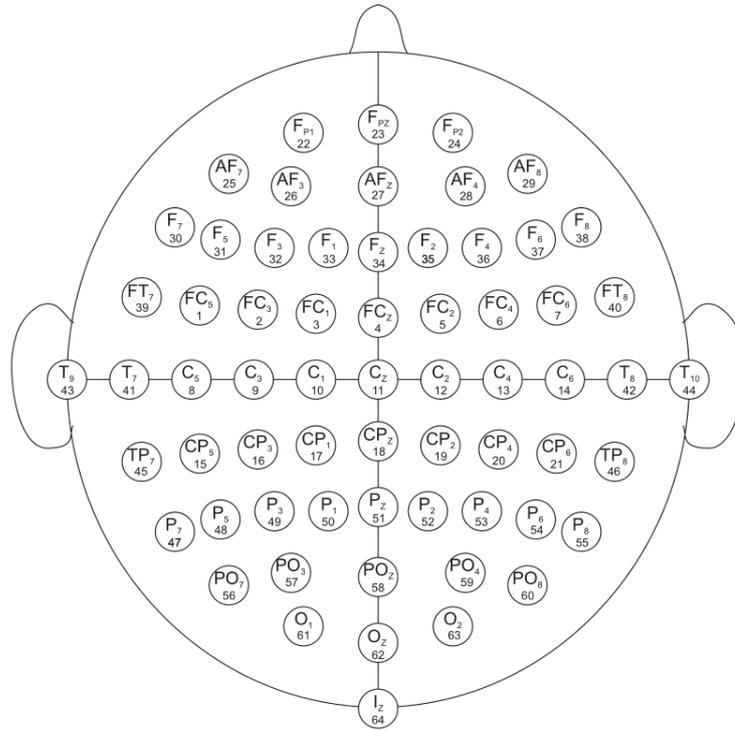


Figura 11: Disposição dos Eletrodos no Sistema 10-10
[Fonte: Extraído de (PhysioBank..., 2014)]

Para este trabalho, utilizou-se os canais presentes no equipamento de aquisição de EEGs comercial “*Emotiv EPOC*” (EMOTIV..., 2014): AF3, AF4, F3, F4, F7, F8, FC5, FC6, P7, P8, T7, T8, O1 e O2.

As entradas para os classificadores são os valores de tensão em cada instante de tempo dos canais citados acima.

3.5 Configurações dos Classificadores

3.5.1 Configurações MLP

A MLP foi utilizada devido a sua popularização em inúmeros trabalhos como classificador. A função de ativação dos neurônios foi a *tangente hiperbólica* descrita na Equação 3.1, o algoritmo de treinamento utilizado foi o *backpropagation*, o critério de parada do

treinamento foram: máximo de 10 épocas seguidas que pioraram os resultados ou 1000 iterações.

Para buscar uma boa configuração da MLP, analisou-se os seguintes parâmetros:

- 1 Camada Oculta
 - Quantidade de Neurônios = [5, 10, 15, 20, 25, 50, 100];
- 2 Camadas Ocultas
 - Quantidade de Neurônios = [5, 10, 15, 20, 25, 50, 100];

Totalizando 14 configurações diferentes para a MLP.

3.5.2 Configurações ELM

A ELM foi utilizada devido a sua boa capacidade de generalização e rápida velocidade de treinamento. A função de ativação dos neurônios foi a *tangente hiperbólica* descrita na Equação 3.1.

Para buscar uma boa configuração da ELM, analisou-se os seguintes parâmetros:

- Quantidade de Neurônios da Camada Escondida = [5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 450, 500, 750, 1000];

Totalizando 27 configurações diferentes para a ELM.

3.5.3 Configurações SVM

A SVM implementada neste trabalho utilizou o *kernel* conhecido como *Radial Basis Function* (RBF) definido pela como Equação 3.3.

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (3.3)$$

Para buscar uma boa configuração da SVM, analisou-se os seguintes parâmetros:

- $C = [2^{-5}, 2^{-1}, 2^3, 2^7, 2^{11}, 2^{15}]$;
- $\gamma = [2^{-15}, 2^{-11}, 2^{-7}, 2^{-3}, 2, 2^3]$;

Utilizando todas as combinações desses parâmetros, existem 36 configurações diferentes para a SVM. A escolha desses parâmetros exponencialmente crescentes foi sugerida por Hsu, Chang e Lin (2003).

3.5.4 Configurações ESN

A proposta de utilizar a ESN se torna interessante para esse tipo de problema, devido ao fato desta conseguir modelar sistemas dinâmicos. A função de ativação dos neurônios foi a *tangente hiperbólica* descrita na Equação 3.1.

Para buscar uma boa configuração da ESN, analisou-se os seguintes parâmetros:

- Tamanho do *Reservoir* (N_x) = [25, 50, 100, 250, 500, 750];
- Esparsidade do *Reservoir* = [0.1, 0.5, 1];
- Raio Espectral $\rho(\mathbf{W})$ = [0.1, 0.5, 0.7, 1];
- Taxa de Vazamento (α) = [0.05, 0.1, 0.5, 0.9];
- Número de Passos do Aquecimento = [100];

Utilizando todas as combinações desses parâmetros, existem 288 configurações diferentes para a RC.

3.6 Banco de Dados

O banco de dados utilizado foi o *eegmmidb - EEG Motor Movement/Imagery Dataset* (EEG..., 2009), capturado utilizando o sistema *BCI2000* (SCHALK et al., 2004), disponível através do *PhysioBank* (PhysioBank..., 2014).

O *eegmmidb* é composto por mais de 1500 registros de um e dois minutos de EEGs, obtidos a partir de 109 voluntários.

3.6.1 Protocolo de Aquisição dos Dados

Sujeitos realizam diferentes diferentes tarefas motoras/imaginadas enquanto 64 canais de EEG era registrado utilizando o *BCI2000* no sistema internacional 10-10 visto na Figura 11. Cada sujeito realizou 14 sessões experimentais: 2 ensaios de um minuto de

sessões *baseline* (um minuto com os olhos abertos, um minuto com os olhos fechados), e 3 ensaios de dois minutos de cada uma das seguintes tarefas:

1. Um alvo aparece no lado esquerdo ou direito da tela. O sujeito abre e fecha o punho correspondente até que o alvo desapareça. Em seguida, o sujeito relaxa.
2. Um alvo aparece no lado esquerdo ou direito da tela. O sujeito imagina abrindo e fechando o punho correspondente até que o alvo desapareça. Em seguida, o sujeito relaxa.
3. Um alvo aparece na parte superior ou inferior da tela. O sujeito abre e fecha ambos os punhos (se o alvo está no topo) ou ambos os pés (se o alvo está na parte inferior) até que o alvo desapareça. Em seguida, o sujeito relaxa.
4. Um alvo aparece na parte superior ou inferior da tela. O sujeito imagina abrindo e fechando ambos os punhos (se o alvo está no topo) ou ambos os pés (se o alvo está na parte inferior) até que o alvo desapareça. Em seguida, o sujeito relaxa.

Em resumo, as sessões experimentais são:

1. *Baseline*, olhos abertos.
2. *Baseline*, olhos fechados.
3. Tarefa 1 (abre e fecha o punho esquerdo ou direito).
4. Tarefa 2 (imagina abrindo e fechando o punho esquerdo ou direito).
5. Tarefa 3 (abre e fecha ambos os punhos ou ambos os pés).
6. Tarefa 4 (imagina abrindo e fechando ambos os punhos ou ambos os pés).
7. Tarefa 1.
8. Tarefa 2.
9. Tarefa 3.
10. Tarefa 4.
11. Tarefa 1.
12. Tarefa 2.

13. Tarefa 3.

14. Tarefa 4.

Os dados são providos no formato EDF+ (contendo 64 sinais de EEG, cada um amostrado em 160 Hz, e um canal de anotação). Cada anotação include um dos três códigos ($T0$, $T1$, ou $T2$):

- **T0** corresponde ao descanso, com duração de 4,2 segundos.
- **T1** corresponde ao início do movimento real ou imaginário, com duração de 4,1 segundos, do:
 - Punho esquerdo (nas sessões 3, 4, 7, 8, 11, e 12)
 - Ambos os punhos (nas sessões 5, 6, 9, 10, 13 e 14)
- **T2** corresponde ao início do movimento real ou imaginário, com duração de 4,1 segundos, do:
 - Punho direito (nas sessões 3, 4, 7, 8, 11 e 12)
 - Ambos os pés (nas sessões 5, 6, 9, 10, 13 e 14)

3.6.2 Preparação do Banco de Dados

Neste trabalho, dos 64 canais disponíveis, utilizou-se apenas os 14 canais citados na Seção 3.4.

Dos 109 sujeitos, **5** foram escolhidos ao acaso (1, 23, 34, 58 e 77).

Dos quatro movimentos, apenas foram utilizados dois: os movimentos **imaginados** dos punhos **esquerdo** ou **direito**, ou seja, apenas as partes das sessões **4**, **8** e **12** (Tarefa 2) que continham o código **T1** ou **T2** foram utilizadas.

Cada movimento dura 4,1 segundos e possui 656 pontos por cada canal do EEG, ou seja, cada movimento é representado por $14 \times 656 = \mathbf{9184}$ pontos. São 45 movimentos por sujeito, totalizando $45 \times 9184 = \mathbf{413280}$.

3.7 Simulações

Para as simulações, foi utilizado a validação cruzada *k-fold*, que consiste em separar os dados em k subconjuntos mutualmente exclusivos (os *folds*) de tamanhos aproximadamente iguais. O classificador utiliza $k - 1$ subconjuntos para o treinamento, e o subconjunto restante é utilizado para o teste. Este procedimento é repetido k vezes e, desta maneira, cada um dos k subconjunto é utilizado para o teste uma vez. Por fim, calculam-se a média e desvio padrão para os k experimentos.

Para cada sujeito foram feitas **30** execuções dos classificadores: **10** simulações utilizando a validação **3-fold**. Então calculou-se a média e o desvio padrão de cada sujeito.

4 Resultados

4.0.1 Configurações MLP

A Tabela 1 mostra os 3 melhores resultados das configurações da MLP apresentadas na Seção 3.5.1 com as médias, desvios padrão entre parênteses e a média geral entre os sujeitos daquela configuração. A Tabela 2 mostra a quantidade de neurônios usados nas configurações. Os *boxplots* das configurações para os 5 sujeitos são apresentados nos gráficos da Figura 12. A Tabela 3 mostra a maior média, com seu respectivo desvio padrão, das configurações específicas por sujeito.

Tabela 1: Resultados Globais da MLP

Configuração	Sujeitos					Geral
	S1	S23	S34	S58	S77	
1	0,57(0,12)	0,50(0,10)	0,60(0,10)	0,51(0,07)	0,58(0,10)	0,55(0,10)
2	0,55(0,11)	0,49(0,11)	0,53(0,11)	0,52(0,06)	0,53(0,10)	0,53(0,10)
3	0,50(0,09)	0,50(0,11)	0,58(0,09)	0,52(0,08)	0,55(0,09)	0,53(0,09)

Tabela 2: Parâmetros das Configurações da MLP

Configuração	Número de Camadas	Número de Neurônios
1	2	[50, 50]
2	2	[20, 20]
3	1	[100]

Tabela 3: Resultados Específicos por Sujeitos da MLP

Sujeitos				
S1	S23	S34	S58	S77
0,57(0,12)	0,52(0,05)	0,60(0,10)	0,54(0,07)	0,58(0,10)

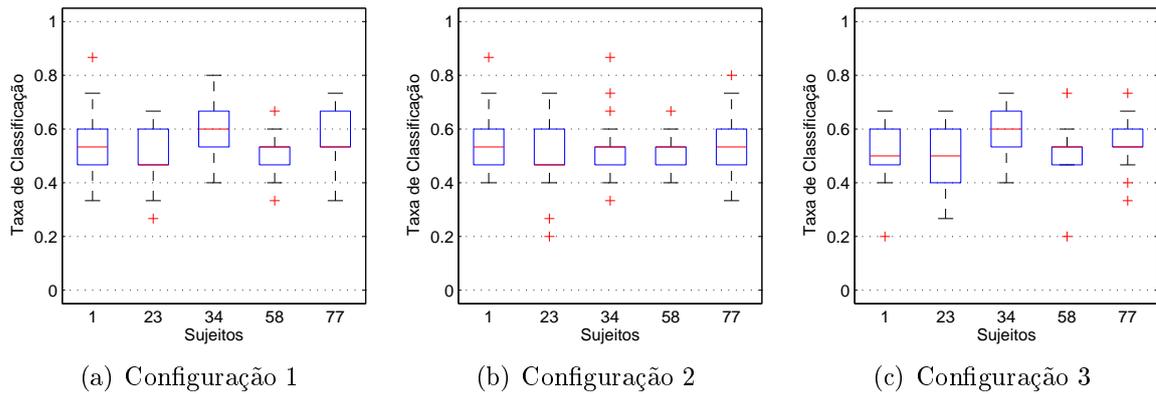


Figura 12: *Boxplots* das Configurações da MLP.

4.0.2 Configurações ELM

A Tabela 4 mostra os 3 melhores resultados das configurações da ELM apresentadas na Seção 3.5.2 com as médias, desvios padrão entre parênteses e a média geral entre os sujeitos daquela configuração. A Tabela 5 mostra a quantidade de neurônios usados nas configurações. Os *boxplots* das configurações para os 5 sujeitos são apresentados nos gráficos da Figura 13. A Tabela 6 mostra a maior média, com seu respectivo desvio padrão, das configurações específicas por sujeito.

Tabela 4: Resultados Globais da ELM

Configuração	Sujeitos					Geral
	S1	S23	S34	S58	S77	
1	0,64(0,09)	0,64(0,07)	0,63(0,10)	0,64(0,13)	0,66(0,09)	0,65(0,10)
2	0,63(0,10)	0,63(0,12)	0,66(0,10)	0,64(0,10)	0,69(0,10)	0,65(0,11)
3	0,65(0,10)	0,64(0,13)	0,65(0,11)	0,66(0,13)	0,66(0,14)	0,65(0,13)

Tabela 5: Parâmetros das Configurações da ELM

Configuração	Número de Neurônios
1	500
2	300
3	1000

Tabela 6: Resultados Específicos por Sujeitos da ELM

Sujeitos				
S1	S23	S34	S58	S77
0,67(0,12)	0,65(0,15)	0,71(0,10)	0,66(0,14)	0,69(0,10)

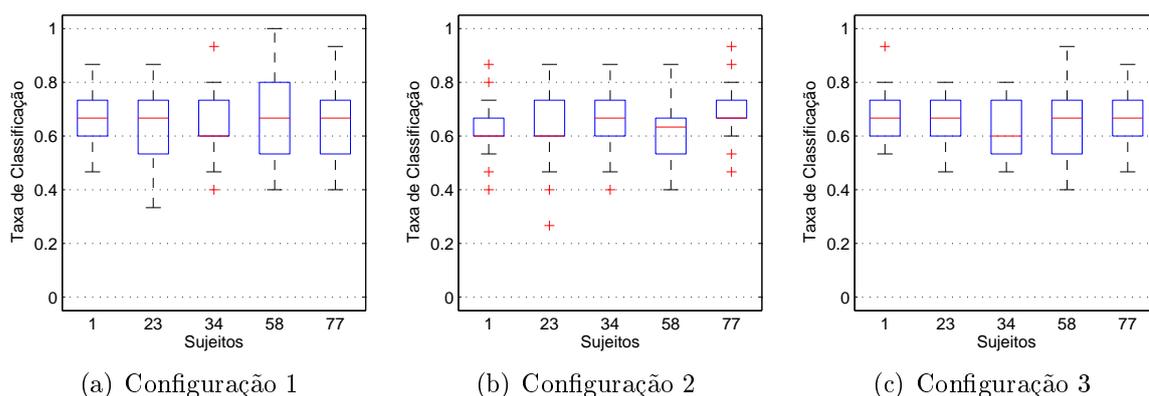


Figura 13: *Boxplots* das Configurações da ELM.

4.0.3 Configurações SVM

A Tabela 7 mostra os 3 melhores resultados das configurações da SVM apresentadas na Seção 3.5.3 com as médias, desvios padrão entre parênteses e a média geral entre os sujeitos daquela configuração. A Tabela 8 mostra os parâmetros utilizados nas configurações. Os *boxplots* das configurações para os 5 sujeitos são apresentados nos gráficos da Figura 14. A Tabela 9 mostra a maior média, com seu respectivo desvio padrão, das configurações específicas por sujeito.

Tabela 7: Resultados Globais da SVM

Configuração	Sujeitos					Geral
	S1	S23	S34	S58	S77	
1	0,68(0,09)	0,61(0,09)	0,65(0,11)	0,67(0,11)	0,68(0,10)	0,65(0,10)
2	0,67(0,09)	0,59(0,11)	0,65(0,11)	0,68(0,10)	0,66(0,10)	0,65(0,10)
3	0,65(0,10)	0,59(0,12)	0,66(0,09)	0,69(0,11)	0,66(0,10)	0,65(0,10)

Tabela 8: Parâmetros das Configurações da SVM

Configuração	C	γ
1	2^{11}	2^3
2	2^3	2^3
3	2^3	2

Tabela 9: Resultados Específicos por Sujeitos da SVM

Sujeitos				
S1	S23	S34	S58	S77
0,68(0,09)	0,61(0,09)	0,72(0,10)	0,70(0,12)	0,70(0,11)

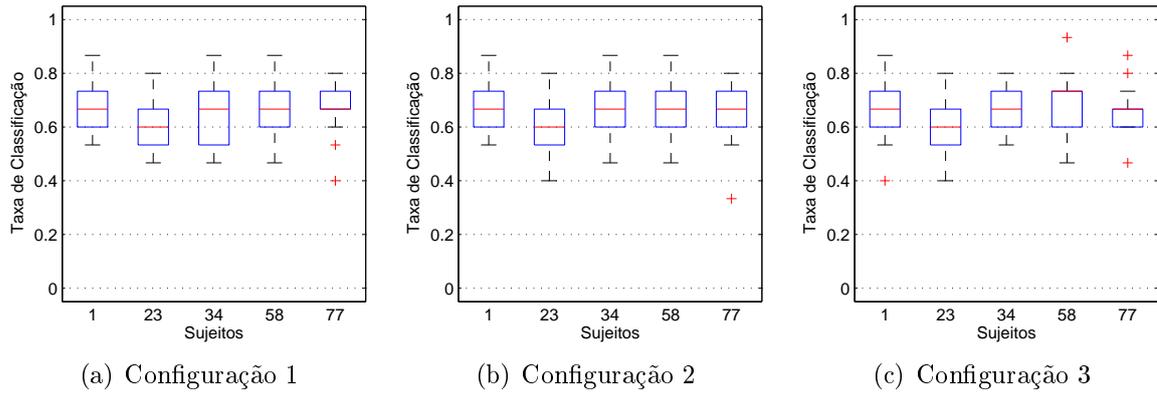


Figura 14: *Boxplots* das Configurações da SVM.

4.0.4 Configurações ESN

A Tabela 10 mostra os 3 melhores resultados das configurações da ESN apresentadas na Seção 3.5.4 com as médias, desvios padrão entre parênteses e a média geral entre os sujeitos daquela configuração. A Tabela 11 mostra os parâmetros: *número de passos do aquecimento* ($N.P.A$), *tamanho* (N_x), *taxa de vazamento* (α), *raio espectral* ($\rho(\mathbf{W})$) e *esparsidade* (esp). Os *boxplots* das configurações para os 5 sujeitos são apresentados nos gráficos da Figura 15. A Tabela 12 mostra a maior média, com seu respectivo desvio padrão, das configurações específicas por sujeito.

Tabela 10: Resultados Globais da ESN

Configuração	Sujeitos					Geral
	S1	S23	S34	S58	S77	
1	0,67(0,11)	0,77(0,09)	0,70(0,10)	0,64(0,10)	0,68(0,10)	0,69(0,10)
2	0,63(0,10)	0,75(0,10)	0,69(0,11)	0,69(0,10)	0,68(0,10)	0,69(0,11)
3	0,63(0,13)	0,79(0,10)	0,70(0,10)	0,64(0,11)	0,66(0,14)	0,68(0,12)

Tabela 11: Parâmetros das Configurações da ESN

Configuração	$N.P.A$	N_x	α	$\rho(\mathbf{W})$	esp
1	100	500	0,5	1	0
2	100	250	0,1	1	0,9
3	100	250	0,1	0,7	0,5

Tabela 12: Resultados Específicos por Sujeitos da ESN

Sujeitos				
S1	S23	S34	S58	S77
0,67(0,11)	0,81(0,09)	0,73(0,12)	0,69(0,10)	0,72(0,10)

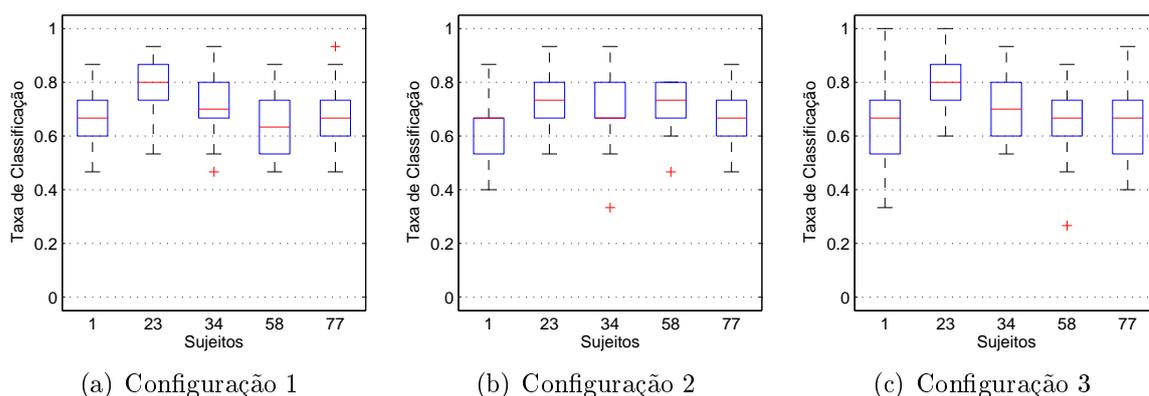


Figura 15: *Boxplots* das Configurações da ESN.

4.1 Análise Estatística

4.1.1 Teste de Wilcoxon

Para realizar o teste de Wilcoxon, serão utilizadas as configurações dos classificadores que obtiveram a maior média global e menor desvio padrão. Sendo assim, serão utilizados os resultados das configurações de número 1 de cada classificador. A Tabela 13 mostra o Teste de Wilcoxon, com 5% de significância estatística, onde o classificador da linha (C_i) é comparado com o da coluna (C_j) através do símbolo presente na célula ($cel_{i,j}$). Os significados dos símbolos podem ser vistos na Tabela 14.

Tabela 13: Teste de Wilcoxon

	MLP	ELM	SVM	ESN
MLP	—	▽	▽	▽
ELM	▲	—	=	▽
SVM	▲	=	—	▽
ESN	▲	▲	▲	—

Tabela 14: Significado dos Símbolos do Teste de Wilcoxon

Símbolo	Significado
▲	C_i tem desempenho superior ao C_j
=	C_i tem desempenho equivalente ao C_j
▽	C_i tem o desempenho inferior ao C_j

De acordo com o teste estatístico utilizado, pode-se afirmar que o desempenho da ESN foi superior a todos os outros classificadores, a MLP foi o classificador que teve o pior desempenho, enquanto que a SVM e a ELM tiveram desempenhos equivalentes entre si.

5 *Considerações Finais*

5.1 Conclusões

Os sinais cerebrais adquiridos através de EEGs são muito úteis por se tratarem de procedimentos não invasivos e de baixo custo para o desenvolvimento da ICM. Este trabalho analisou a importância temporal, espacial e combinacional dos sinais, utilizando 14 eletrodos, para a classificação de movimentos imaginados dos punhos esquerdos ou direitos. Foram realizadas 4 proposições, sendo 3 de *classificadores de padrões instantâneos* e 1 de um *classificador de padrões temporais*.

Nas propostas dos classificadores de padrões instantâneos, procurou-se utilizar a MLP com o algoritmo de treinamento *backpropagation*, a ELM e a SVM com *kernel* RBF, de forma a estudar os resultados das classificações utilizando apenas as características espaciais e combinacionais dos sinais. Enquanto que na proposta do classificador de padrão temporal, utilizou-se a RC através do algoritmo ESN, com a finalidade de estudar os resultados das classificações utilizando características temporais, espaciais e combinacionais dos sinais.

De acordo com o teste de Wilcoxon com 5% de significância estatística, o classificador de padrões temporais (ESN) teve o melhor desempenho entre os classificadores testados, sugerindo que a característica temporal deve ser levada em consideração para a análise de sinais cerebrais. Acredita-se que o desvio padrão poderia ser diminuído se uma quantidade maior de exemplos por sujeito fosse disponibilizada durante o treinamento. Além disso, pode-se perceber que através das tabelas com as melhores configurações específicas por sujeito que, caso sejam utilizados parâmetros específicos por sujeitos, as taxas de acerto sobem consideravelmente.

5.2 Trabalhos Futuros

Como trabalhos futuros, é são pretendidos:

- Utilizar algoritmos de busca para encontrar os parâmetros dos classificadores.
- Utilizar algoritmos de busca para encontrar o conjunto de eletrodos e intervalos de frequências que maximizem a taxa de classificação das tarefas.
- Testar os algoritmos propostos em sistemas ICM *online*.
- Utilizar outros bancos de dados com uma quantidade maior de exemplos por sujeitos.
- Implementar RC com outras arquiteturas para a classificação de biosinais.
- Extrair características dos sinais.
- Melhorar o pré-processamento dos sinais.

Referências

- BEN-ISRAEL, A.; GREVILLE, T. *Generalized Inverses: Theory and Applications*. Springer, 2003. (CMS Books in Mathematics). ISBN 9780387002934. Disponível em: <<http://books.google.com.br/books?id=o3-97W8vCdIC>>.
- BURGES, C. J. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, v. 2, n. 2, p. 121–167, 1998.
- CAEP. Disponível em: <<http://www.caepcampinas.com.br/shared/images/e-infra-eletro/saladeeletrocardiograma04.jpg>>.
- CORTICOGRAFIA.JPG. Disponível em: <<http://jonlieffind.com/wp-content/uploads/2012/12/corticografia.jpg>>.
- CURRAN, E. A.; STOKES, M. J. Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain and Cognition*, v. 51, n. 3, p. 326–336, abr. 2003. ISSN 0278-2626.
- DONCHIN, E.; SPENCER, K. M.; WIJESINGHE, R. The mental prosthesis: assessing the speed of a p300-based brain-computer interface. *Rehabilitation Engineering, IEEE Transactions on*, v. 8, n. 2, p. 174–179, 2000.
- EEG Motor Movement/Imagery Dataset. 2009. Acesso em: 18 abr. 2014. Disponível em: <<http://www.physionet.org/pn4/eegmidb/>>.
- EEG.JPG. Disponível em: <<http://braincoreny.com/wp-content/uploads/2014/04/eeg.jpg>>.
- EMOTIV EPOC. 2014. Acesso em: 26 mai. 2014. Disponível em: <<http://emotiv.wikia.com/wiki/Emotiv\EPOC>>.
- EOG Spordabilim. Disponível em: <http://www.spordabilim.com/imgs/iki_img.jpg>.
- HAYKIN, S. *Neural Networks: A Comprehensive Foundation (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2007. ISBN 0131471392.
- HE, B. (Ed.). *Neural Engineering*. Boston, MA: Springer US, 2013. ISBN 978-1-4614-5226-3, 978-1-4614-5227-0.
- HSU, C.-W.; CHANG, C.-C.; LIN, C.-J. *A practical guide to support vector classification*. [s.n.], 2003. Disponível em: <<https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf>>.

- HUANG, D. et al. Electroencephalography (EEG)-Based brain computer interface (BCI): a 2-d virtual wheelchair control based on event-related Desynchronization/Synchronization and state control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, v. 20, n. 3, p. 379–388, maio 2012. ISSN 1534-4320.
- HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing*, v. 70, n. 1-3, p. 489–501, dez. 2006. ISSN 09252312.
- IBGE Dados Estatísticos sobre deficiências no Brasil. 2010. Acesso em: 30 abr. 2014. Disponível em: <<http://7a12.ibge.gov.br/vamos-conhecer-o-brasil/nosso-povo/caracteristicas-da-populacao>>.
- JAEGER, H. The “echo state“ approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, v. 148, p. 34, 2001.
- JAEGER, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, v. 304, n. 5667, p. 78–80, abr. 2004. ISSN 0036-8075, 1095-9203. Disponível em: <<http://www.sciencemag.org/cgi/doi/10.1126/science.1091277>>.
- JAEGER, H. Echo state network. *Scholarpedia*, v. 2, n. 9, p. 2330, 2007. ISSN 1941-6016. Disponível em: <http://www.scholarpedia.org/article/Echo_state_network>.
- JURCAK, V.; TSUZUKI, D.; DAN, I. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*, v. 34, n. 4, p. 1600–1611, fev. 2007. ISSN 1053-8119.
- KENNEDY, P. R. et al. Direct control of a computer from the human central nervous system. *Rehabilitation Engineering, IEEE Transactions on*, v. 8, n. 2, p. 198–202, 2000.
- KLONOWSKI, W. Everything you wanted to ask about EEG but were afraid to get the right answer. *Nonlinear Biomedical Physics*, v. 3, n. 1, p. 2, 2009. ISSN 1753-4631.
- LAB. Biom. UFSC. Disponível em: <http://biomec.paginas.ufsc.br/?page_id=171>.
- LEE, P.-L. et al. A brain-wave-actuated small robot car using ensemble empirical mode decomposition-based approach. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, v. 42, n. 5, p. 1053–1064, set. 2012. ISSN 1083-4427.
- LUKOSEVICIUS, M. A practical guide to applying echo state networks. In: MONTAVON, G.; ORR, G. B.; MULLER, K.-R. (Ed.). *Neural Networks: Tricks of the Trade*. [S.l.]: Springer Berlin Heidelberg, 2012, (Lecture Notes in Computer Science, 7700). p. 659–686. ISBN 978-3-642-35288-1, 978-3-642-35289-8.
- LUKOSEVICIUS, M.; JAEGER, H. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, v. 3, n. 3, p. 127–149, ago. 2009. ISSN 1574-0137.
- LUKOSEVICIUS, M.; JAEGER, H.; SCHRAUWEN, B. Reservoir computing trends. *KI-Kunstliche Intelligenz*, v. 26, n. 4, p. 365–371, 2012.

- MAASS, W.; NATSCHLAGER, T.; MARKRAM, H. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.*, v. 14, n. 11, p. 2531–2560, nov. 2002. ISSN 0899-7667.
- MURUGAPPAN, M. Classification of human emotion from EEG using discrete wavelet transform. *Journal of Biomedical Science and Engineering*, v. 03, n. 04, p. 390–396, 2010. ISSN 1937-6871, 1937-688X.
- NICOLELIS, M. A. Mind in motion. *Scientific American*, v. 307, n. 3, p. 58–63, 2012.
- PhysioBank, PhysioToolkit, and PhysioNet. 2014. Acesso em: 18 abr. 2014. Disponível em: <<https://circ.ahajournals.org/content/101/23/e215.full.pdf+html>>.
- RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 2. ed. [S.l.]: Pearson Education, 2003. ISBN 0137903952.
- SCHALK, G. et al. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Trans Biomed Eng*, v. 51, n. 6, p. 1034–1043, jun. 2004. ISSN 0018-9294. PMID: 15188875.
- SMITH, R. *Handbook of Amyotrophic Lateral Sclerosis*. [S.l.]: Taylor & Francis, 1992. (Neurological Disease and Therapy). ISBN 9780824786106.
- SUTTER, E. E. The brain response interface: communication through visually-induced electrical brain responses. *Journal of Microcomputer Applications*, v. 15, n. 1, p. 31–45, jan. 1992. ISSN 0745-7138.
- TAVELLA, M. et al. Towards natural non-invasive hand neuroprostheses for daily living. In: *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. [S.l.]: IEEE, 2010. p. 126–129.
- TEPLAN, M. Fundamentals of EEG measurement. *Measurement science review*, v. 2, n. 2, p. 1–11, 2002.
- TRIEFENBACH, F. et al. Phoneme recognition with large hierarchical reservoirs. In: *NIPS*. [s.n.], 2010. p. 2307–2315. Disponível em: <<https://papers.nips.cc/paper/4056-phoneme-recognition-with-large-hierarchical-reservoirs.pdf>>.
- VALENCA, M. *FUNDAMENTOS DAS REDES NEURAIS*. [S.l.]: MEUSER VALENCA, 2010. ISBN 9788577163427.
- VAPNIK, V. N. *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995. ISBN 0-387-94559-8.
- VAPNIK, V. N. *Statistical Learning Theory*. [S.l.]: Wiley-Interscience, 1998.
- VAPNIK, V. N.; CHERVONENKIS, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, SIAM, v. 16, n. 2, p. 264–280, 1971. Disponível em: <<http://link.aip.org/link/?TPR/16/264/1>>.
- VERSTRAETEN, D. *Reservoir Computing: computation with dynamical systems*. Tese (Doutorado) — Ghent University, 2009. Disponível em: <<https://biblio.ugent.be/publication/779431>>.

- VERSTRAETEN, D. et al. An experimental unification of reservoir computing methods. *Neural Networks*, v. 20, n. 3, p. 391–403, abr. 2007. ISSN 08936080. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S089360800700038X>>.
- WERBOS, P. J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, v. 78, n. 10, p. 1550–1560, 1990.
- WOLPAW, J.; WOLPAW, E. W. *Brain-Computer Interfaces: Principles and Practice*. [S.l.]: Oxford University Press, 2012. ISBN 9780195388855.
- WOLPAW, J. R. et al. Brain-computer interfaces for communication and control. *Clinical neurophysiology*, v. 113, n. 6, p. 767–791, 2002.