



UTILIZANDO AS REDES NEURAIS ARTIFICIAIS SUPPORT VECTOR MACHINES E MULTI-LAYER PERCEPTRON PARA DIAGNÓSTICO PRECOCE DA DOENÇA DE ALZHEIMER

Trabalho de Conclusão de Curso

Engenharia da Computação

Dierson Alessandro Cruz de Oliveira
Orientador: Prof. Dr. Mêuser Jorge Silva Valença



**UNIVERSIDADE
DE PERNAMBUCO**

**Universidade de Pernambuco
Escola Politécnica de Pernambuco
Graduação em Engenharia de Computação**

Dierson Alessandro Cruz de Oliveira

**UTILIZANDO AS REDES NEURAIS
ARTIFICIAIS SUPPORT VECTOR
MACHINES E MULTI-LAYER
PERCEPTRON PARA DIAGNÓSTICO
PRECOCE DA DOENÇA DE
ALZHEIMER**

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

Recife, Junho de 2015.

De acordo

Recife

____/____/____

Orientador da Monografia

Dedico à minha mãe, Maria da Saúde Cruz de Oliveira, ao meu Pai, José Arnaldo da Cruz, a meus irmãos, Adson Diervo Cruz de Oliveira e Aila Maria Cruz de Oliveira, a minha namorada Talita Novaes e a o meus amigos Itúrbide Felipe, Thiago Britto, Karolyne Cavalcanti, Lara Dantas e Gearlles Ferreira e Felipe Mendonça.

Agradecimentos

Agradeço primeiramente aos meus pais, Maria S. de Oliveira e José A. Cruz, que sempre me apoiaram em todos os sentidos e graças os seus esforços cheguei até aqui e tenho capacidade de ir muito mais além

Agradeço também a meu irmão que sempre me deu muitos conselhos para vida. E a minha namorada que sempre me apoiou incondicionalmente, me dando forças para realizar minhas metas.

Aos meus amigos da universidade, Itúrbide Felipe, Felipe Mendonça, Gearlles Ferreira, Karolyne Cavalcanti, Lara Dantas, Thiago Britto, Nestor Jr., com os quais tive o prazer de compartilhar conhecimentos para vida dentro e fora da universidade.

Agradeço também aos professores que foram os principais contribuintes para minha aprendizagem que se estende da universidade para o mundo.

E agradeço a Deus por essa grande oportunidade cheias de desafios que me ajudaram a crescer e amadurecer muito.

Resumo

Diante de um crescente fenômeno do envelhecimento populacional, que atinge os países mais desenvolvidos, surge maior preocupação com a qualidade de vida dos idosos, pois devido a o aumento de expectativa de vida, aumenta-se também a incidência de doenças relacionadas à idade avançada, como as demências senís. A doença de Alzheimer (DA) é considerada a mais frequente, dentre as demências, atingindo mais de 1,2 milhões de brasileiros e 35,6 milhões em todo mundo, segundo um dos maiores levantamentos realizados até hoje sobre a DA, Alzheimer's Disease Internacional (ADI). Segundo a ADI, a incidência dobra a cada 20 anos. Desta forma, pode-se chegar a 65,7 milhões em 2030 e a 115,4 milhões em 2050. Embora seja uma doença neurodegenerativa progressiva e irreversível, se a DA for descoberta na fase inicial, algumas medidas podem ser adotadas e, usando medicamentos adequados, a progressão dos sintomas pode ser retardada. Existem alguns trabalhos relacionados a o diagnóstico precoce da DA com boas taxas de classificação como em Ray *et al.* e Moscato, os quais se basearam no uso de uma base de dados de 120 proteínas contidas no plasma e, utilizando classificadores contidos no software Weka, apresentaram taxas de acerto de 91% e 96% respectivamente. Desta forma, este trabalho de pesquisa busca utilizar duas redes neurais artificiais distintas como classificadores para realizar diagnóstico precoce de DA. São elas: Support Vector Machines (SVM) e Multilayer Perceptron (MLP). Posteriormente, serão realizadas comparações com os resultados obtidos dos classificadores.

Palavras-Chave: Doença de Alzheimer, Support Vector Machines, Multilayer Perceptron, Redes Neurais Artificiais

Abstract

Faced with a growing phenomenon of population aging, which affects the most developed countries, appears more concerned with the quality of life for elderly, because due to the increase in life expectancy also increases the incidence of age-related diseases such as senile dementias. Alzheimer's Disease (AD) is considered the most common among dementia, affecting over 1.2 million Brazilians and 35.6 million worldwide, according to one of the largest surveys to date on AD, Alzheimer's Disease International (ADI). According to ADI, incidence doubles every 20 years. In this way, one can reach 65.7 million in 2030 and 115.4 million in 2050. Although it is a progressive and irreversible neurodegenerative disease, if AD discovered in early stages, some measures can be adopted and, using appropriate medications, symptoms progression may be delayed. There are some works related to early diagnosis of AD with good rates as in Ray et al. and Moscato, which are based on the use of a database 120 proteins contained in plasma and, using software contained in the classifiers, showed success rates of 91% and 96% respectively. Thus, this research seeks to use two different artificial neural networks as classifiers to perform early diagnosis of AD: Support Vector Machines (SVM) and Multilayer Perceptron (MLP). Subsequently, comparisons will be made with the results of the classifiers.

Keywords: Alzheimer's Disease, Support Vector Machines, Multilayer Perceptron, Hopfield Network Recurrent, Recurrent Networks, Artificial Neural Networks

Sumário

| | |
|-------------------------------------|-----------|
| 1. Introdução | 1 |
| 1.1 Motivação | 1 |
| 1.2 Objetivos | 3 |
| 1.2.1 Objetivos Gerais | 3 |
| 1.2.2 Objetivos Específicos | 4 |
| 1.3 Estrutura da Monografia | 4 |
| 2. Fundamentação Teórica | 5 |
| 2.1. Doença de Alzheimer | 5 |
| 2.2. Redes Neurais Artificiais | 8 |
| 2.2.1. Multi-Layer Perceptron | 12 |
| 2.2.1.1. Algoritmo BackPropagation | 13 |
| 2.2.1.2. Validação Cruzada | 16 |
| 2.2.2. Support Vector Machines | 16 |
| 2.2.2.1. SVMs Lineares | 18 |
| 2.2.2.1.1. SVMs com Margens Rígidas | 18 |
| 2.2.2.1.2. SVMs com Margens Suaves | 20 |
| 2.2.2.2. SVMs Não-Lineares | 21 |
| 3. Metodologia | 23 |
| 3.1. Base de Dados | 23 |
| 3.1.1. Pré – Processamento | 24 |
| 3.2. Simulação dos Classificadores | 24 |
| 3.2.1. Simulação da RNA MLP | 25 |

| | |
|--|-----------|
| 3.2.2. Simulação da SVM | 25 |
| 3.3. Teste da Soma dos Postos Wilcoxon | 26 |
| 4. Resultados | 28 |
| 5. Conclusões e Trabalhos Futuros | 31 |
| Bibliografia | 33 |
| Apêndice A Resultado do Teste de Wilcoxon no Software | 35 |

Índice de Figuras

| | |
|---|----|
| Figura 1. a) Ressonância magnética de cérebro normal; b) Ressonância magnética de cérebro de paciente portador da doença de Alzheimer, exibindo atrofia cerebral difusa. [Fonte: Imagens retiradas do site “The Whole Brain Atlas” http://www.med.harvard.edu/aanlib/] | 7 |
| Figura 2. Representação do neurônio artificial proposto por McCulloch e Pitts. [Fonte: Elaboração própria] | 10 |
| Figura 3. Representação de Rede Neural Artificial Multi Layer Perceptron com uma camada intermediária (ou camada escondida) de neurônios. [Fonte: Elaboração própria] | 13 |
| Figura 4. Representação de uma classificação binária realizada por uma SVM de margem rígida [Fonte: Elaboração própria] | 19 |
| Figura 5. Representação de uma classificação binária realizada por uma SVM de margem suave [Fonte: Elaboração própria] | 21 |
| Figura 6. Representação de uma classificação binária realizada por uma SVM não linear [Fonte: Elaboração própria] | 22 |
| Figura 7. Execução do teste de Wilcoxon [Fonte: elaboração própria]..... | 35 |

Índice de Tabelas

| | | |
|------------------|---|----|
| Tabela 1. | Representação dos parâmetros utilizados na rede MLP para simulações dos dados normalizados [Fonte: elaboração própria]..... | 25 |
| Tabela 2. | Representação das funções Kernel utilizadas na rede SVM para simulações [Fonte: elaboração própria]..... | 26 |
| Tabela 3. | Formulação das hipóteses definidas na aplicação do teste da Soma dos Postos de Wilcoxon [Fonte: elaboração própria] | 27 |
| Tabela 4. | Representação da média das taxas de classificação para 30 simulações das SVMs com diferentes funções Kernel [Fonte: elaboração própria] | 28 |
| Tabela 5. | Representação da média das taxas de classificação para 30 simulações [Fonte: elaboração própria]..... | 29 |
| Tabela 6. | Formulação das hipóteses definidas na aplicação do teste da Soma dos Postos de Wilcoxon para os resultados da RNA MLP e SVM [Fonte: elaboração própria] | 30 |
| Tabela 7. | Representação das comparações das média das taxas de classificação com os disponíveis na literatura [Fonte: elaboração própria] | 30 |

Tabela de Símbolos e Siglas

ADI – Alzheimer’s Disease International

ABRAz – Associação Brasileira de Alzheimer

DA – Doença de Alzheimer

DMDA – Dia Mundial da Doença de Alzheimer

FBR – Função de Base Radial

MLP – Multi-Layer Perceptron

OMS – Organização Mundial de Saúde

RNA – Rede Neural Artificial

SVM– Support Vector Machines

TAE – Teoria do Aprendizado Estatístico

1. Introdução

1.1 Motivação

O aumento da expectativa de vida provoca um “envelhecimento populacional”, ou seja, um aumento de idosos na população. Isto acontece principalmente nos países mais desenvolvidos. Conseqüentemente, há um aumento na incidência de doenças que comprometem a qualidade de vida dos idosos, como as demências que provocam declínio de funções cognitivas e retenção de memórias. Segundo a ABRAz [1], a Doença de Alzheimer é a demência mais frequente no Brasil, em 2012 já havia mais de 1,2 milhões de casos no Brasil e, segundo levantamentos da Alzheimer’s Disease Internacional (ADI) [2] no mesmo ano, 35,6 milhões de casos em todo mundo.

A (DA) é uma enfermidade neurodegenerativa progressiva e irreversível. A perda de neurônios e a degeneração das sinapses nervosas causam atrofia cerebral de forma difusa afetando, principalmente, as áreas de coordenação de atividades intelectuais, provocando declínio de funções cognitivas, dificuldades progressiva de reter memórias e realizar atividades cotidianas. Desta forma, os pacientes portadores de DA possuem maior dependência da família e (ou) cuidadores profissionais para realização das necessidades diárias.

Na maioria dos casos, a DA inicia-se após os 65 anos de idade, mas em alguns casos pode haver manifestação da DA em indivíduos mais jovens, conhecida como início precoce. Os fatores de risco mais relevantes para o início tardio da DA são a idade e baixa escolaridade [3], mas existem outros fatores importantes, como fatores genéticos, histórico de traumatismo craniano com perda de consciência, descontrole de fatores de risco cardiovascular, sedentarismo, baixa atividade mental ao longo da vida [4]. Já para a DA de início precoce, os fatores de risco estão mais associados, em geral, a mutações genéticas ou a algumas síndromes como Síndrome de Down, por exemplo.

Segundo a ADI, a incidência da DA dobra a cada 20 anos, desta forma pode chegar a 65,7 milhões em 2030 e a 115,4 milhões em 2050 [2]. Cerca de 60% dos casos de DA se encontram nos países desenvolvidos e em 2050 deverá chegar em 72%[5]. A ADI afirma também que o custo total (direto e indireto) para cuidados dos portadores de DA no mundo podem chegar cerca de U\$ 400 bilhões anuais distribuídos em diversos fatores desde medicação até medidas necessárias para os cuidados [2].

Apesar de ser irreversível, através de medicamentos e cuidados é possível amenizar os efeitos da DA e quanto mais precocemente diagnosticado for o paciente, mais fácil será de realizar procedimentos e retardar a progressão da doença de modo que o portador possa ter uma vida mais saudável e com maior qualidade de vida.

Na literatura existem alguns trabalhos relacionados a diagnóstico precoce de DA. Ray et al. realizou experimentos utilizando uma base de dados com 120 proteínas que são encontradas no plasma, e em seu experimento, chegou à conclusão que através de uma combinação de 18, dessas 120 proteínas, era possível realizar o diagnóstico precoce de DA. Para o teste, ele utilizou um conjunto de dados contendo as proteínas de 92 pacientes que tinham, ou não, sido diagnosticados. Utilizando classificador chegou a taxas de 91% para esse conjunto de teste [6].

Posteriormente, Moscato realizou experimentos com a base de 120 proteínas com intuito de reduzir a quantidade de proteínas necessárias para diagnóstico precoce de DA. Em sua pesquisa, fez experiências com assinaturas com 18, 10, 6 e 5 proteínas e chegou à conclusão de que 5 proteínas é suficiente para classificação. Neste experimento chegou a taxa de 96% para conjunto de testes [6].

Para os dois experimentos mencionados, foram utilizados classificadores disponíveis pelo Software Weka. Neste trabalho serão realizados experimentos utilizando duas Redes Neurais Artificiais distintas como classificadores e a mesma base de dados contendo a assinatura de cinco proteínas definida por Moscato e considerada suficiente para classificação. Serão adotadas as mesmas estratégias utilizadas por Moscato em relação à divisão da base de dados em conjuntos de

treinamento, validação e de teste. As RNA's que serão utilizadas neste trabalho são: A rede *Multi Layer Perceptron* (MLP) e *Support Vector Machines* (SVM).

A rede MLP é uma generalização do *Perceptron*, contendo ao menos uma camada escondida responsável pela não linearidade da rede, tornando-a capaz de resolver problemas não linearmente separáveis, desde que seja utilizada uma função de ativação não linear na camada escondida, com alta capacidade de generalização (dependendo da configuração estabelecida para o problema em questão)[7].

As SVMs são rede neurais artificiais baseadas na teoria do aprendizado estatístico (TAE), a qual foi proposta inicialmente por Vapnik e Chervonenkis [8] e foi desenvolvida, mais detalhadamente, por Vapnik [9] [10]. As SVMs possuem alta capacidade de generalização e alto custo, pois, durante a fase de treinamento, buscam minimizar o erro de treinamento (Risco Empírico) e, ao mesmo tempo, reduzir a complexidade da rede obtida (Risco Estrutural). Desta forma o objetivo da aplicação de uma SVM não é só encontrar uma solução, mas também aperfeiçoar-la buscando um hiperplano que seja ótimo, ou seja, o que possui maior margem de separação de classes.

Desta forma, os resultados obtidos, após as aplicações das duas redes neurais artificiais para classificação, serão comparados entre eles e com os resultados dos experimentos de Moscato para assinatura de 5 proteínas.

1.2 Objetivos

Nesta seção serão apresentados os objetivos gerais e, posteriormente, os específicos para este trabalho.

1.2.1 Objetivos Gerais

Introduzir conceitos sobre as metodologias das RNA's (MLP e SVM) que serão aplicadas, verificar o desempenho de cada uma para classificação de DA utilizando assinatura de cinco proteínas proposta por Moscato. Os resultados das metodologias serão comparados entre eles e com o resultado obtido por Moscato.

1.2.2 Objetivos Específicos

- Realizar experimentos com a rede MLP explorando os parâmetros de configuração a fim de se obter as melhores taxas de acerto para classificação de DA;
- Realizar o mesmo procedimento com a rede SVM;
- Fazer comparações com os resultados obtidos pelos experimentos das redes MLP e SVM ;
- Comparar estes resultados com os obtidos por Moscato utilizando classificadores do software Weka.

1.3 Estrutura da Monografia

O capítulo 2 trata de uma fundamentação teórica que descreverá conceitos sobre a doença de Alzheimer, MLP, SVM. Estes conceitos servirão de base para compreensão dos experimentos que serão realizados neste trabalho.

O capítulo 3 mostra a metodologia adotada para os experimentos desde preparação da base de dados até aplicação de teste estatístico para validação dos resultados.

Logo em seguida, no capítulo 4, serão apresentados análises dos experimentos realizados e a análise estatística dos resultados obtidos para realizar comparação entre as metodologias adotadas e com resultados existentes na literatura.

Por fim, no capítulo 5 será exposta a conclusão do trabalho realizado com ênfase nos resultados do trabalho como um todo e nas dificuldades encontradas durante os experimentos e destacando os possíveis trabalhos futuros.

2. Fundamentação Teórica

2.1. Doença de Alzheimer

A Doença de Alzheimer (DA) é uma enfermidade neurodegenerativa progressiva e irreversível, caracterizada por degeneração neuronal, expressadas pelas perdas de neurônios e degeneração das sinapses nervosas, cursando com declínio de funções cognitivas, dificuldades progressiva de reter memórias recentes, na aprendizagem, fala e execução de atividades cotidianas.

Em 1907, o psiquiatra alemão Alois Alzheimer descreveu o primeiro caso em um congresso científico na Alemanha, relatando um caso de uma paciente, Auguste D., atendida, inicialmente aos 51 anos, com sintomas característicos de demência, como alterações de memória e linguagem, e, posteriormente, desorientação no tempo e no espaço, que se instalou piorando progressivamente até que, cerca de quatro anos e meio depois do início dos sintomas, a paciente foi a óbito. Durante a autópsia da paciente, Alois Alzheimer observou lesões neuronais, atrofia das células nervosas em várias regiões cerebrais com deformações de fibras e foram detectadas placas estranhas, alterações histológicas que hoje são conhecidas como características da doença, que, até então, não haviam sido descritas. Tais relatos originaram um novo tipo de demência denominada Mal de Alzheimer ou Doença de Alzheimer.

Os fatores de risco mais relevantes para o início tardio da DA são a idade e baixa escolaridade[3]. Outros fatores são também relacionados, como fatores genéticos (presença do alelo $\epsilon 4$ no gene ApoE), histórico de traumatismo craniano com perda de consciência, descontrole de fatores de risco cardiovascular (hipertensão, diabetes melito, dislipidemia), sedentarismo, baixa atividade mental ao longo da vida[4]. Na maioria dos casos, a doença inicia-se após os 65 anos de idade, conhecido como início tardio, mas em alguns casos pode haver manifestação da DA em indivíduos mais jovens, conhecida como início precoce. No caso de início precoce, ou pré-senil, os fatores de risco estão mais associados, em geral, a

mutações genéticas, como a síndrome de Down e o comprometimento das funções neurológicas ocorre de forma mais rápida, culminado num processo de evolução mais rápido da doença[11]. Independentemente de ser do início tardio ou precoce, a fisiopatologia da DA é semelhante: lesões extracelulares (placas senis ou neuríticas) e lesões intracelulares (emaranhados neurofibrilares).

A evolução dos sintomas da DA pode ser dividida em três etapas: leve, moderada e grave. No entanto esta classificação possui mais caráter didático, podendo haver sintomas em estágios diferentes, já que cada a DA pode se desenvolver de forma distinta em cada paciente [12].

A primeira fase é marcada pela perda de memória recente, desorientação no tempo e no espaço, dificuldade na elaboração de palavras e nas tomadas de decisões, redução de motivação e interesse por atividades e passatempos, agressividade e alguns sinais de depressão. Normalmente esta fase não é percebida e, muitas vezes, é confundida com o processo de envelhecimento natural, pois existem pontos em comuns, o que dificulta a distinção entre envelhecimento natural e DA. Essa fase é o momento ideal para intervenção terapêutica, assim, é fundamental estar atento a qualquer sintoma que sugira um quadro de demência, uma vez que isso possibilita buscar assistência médica precoce para avaliação clínica, diagnóstico e tratamento. A doença de Alzheimer possui evolução progressiva, portanto, realizar o diagnóstico precoce é um passo importante para promover qualidade de vida ao portador da doença [5].

Na segunda fase, a fase moderada, observa-se, além do agravamento dos sintomas apresentados na primeira fase, esquecimento de fatos mais importantes, como, por exemplo, nomes de pessoas próximas. Nesta fase existe uma maior dependência do portador, necessitando de cuidados e ajuda para realizar atividades cotidianas, incluindo higiene pessoal. Durante este estágio, nota-se alterações comportamentais, expressando maior irritabilidade, agressividade e inquietação, além de sentimentos fora de contexto como desconfiança e ciúmes. Também são observados casos de alucinações visuais e auditivas.

Na terceira fase, observa-se agravamento da retenção de memória. Nesta fase o paciente apresenta dificuldade de reconhecer parentes, amigos, locais. A

desorientação no espaço passa a atingir até mesmo a própria casa. O portador apresenta dificuldades de desempenhar, de forma independente, atividades básicas de vida diária essenciais para manutenção do seu bem-estar, como alimentar-se, vestir-se, andar, tomar banho, entre outras. Pode haver também incontinência urinária e fecal, tornando-o um ser sem autonomia e dependente de cuidadores.

O indivíduo com sintomas de DA deve ser submetido a uma avaliação médica minuciosa e levantamento de histórico pessoal e familiar, com realização de testes clínicos padronizados para estabelecer um diagnóstico provável ou possível da doença. Mesmo diante de avanços, o diagnóstico definitivo só é possível através de biopsia ou necropsia. Durante a avaliação diagnóstica, é ainda importante realizar exames de sangue e de imagem (tomografia ou ressonância magnética craniana) a fim de descartar outras causas de demências.

A análise de imagens de ressonância magnética (IRM) é muito importante, mas não suficiente, para diagnóstico da DA. Através da análise de IRM podemos notar alguns fenômenos que se tornaram muito importantes para o diagnóstico, por exemplo, alguns tipos de atrofia cerebral (Figura 1), número e tamanho das placas senis no cérebro, as deformações cerebrais em relação aos padrões neuroatômicos[13].

Os avanços na genética, da neuropatologia, da biologia molecular e da neuroimagem permitiram descobertas no campo da neurociência que, num futuro próximo, podem tornar possível diagnósticos mais precoces que as manifestações dos sinais clínicos da DA.

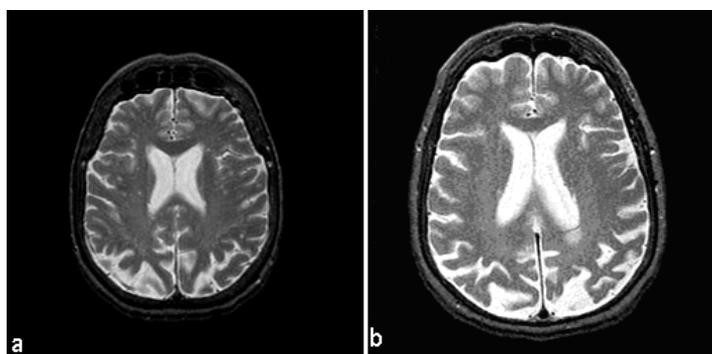


Figura 1. a) Ressonância magnética de cérebro normal; b) Ressonância magnética de cérebro de paciente portador da doença de Alzheimer, exibindo atrofia

cerebral difusa. [Fonte: Imagens retiradas do site “The Whole Brain Atlas”
<http://www.med.harvard.edu/aanlib/>]

O tratamento da doença de Alzheimer é realizado através de medicamentos (tratamento farmacológico) e cuidados (tratamentos não farmacológicos) com o intuito de amenizar os efeitos da doença, afinal, trata-se de uma doença progressiva. Dentre os tratamentos não farmacológicos estão: atividades cognitivas, social e física. Os portadores da DA necessitam de um cuidado integral para que, apesar de suas limitações, possam desfrutar da melhor qualidade de vida possível. Para isso, é necessário ações conjuntas de entidades internacionais e autoridades para redução do problema, com diagnóstico precoce e tratamento multiprofissional. É essencial uma política que dê suporte e estrutura para tratamento dos portadores de DA, incluindo acesso a medicamentos, investimentos em pesquisas para amenizar os sintomas e melhorar a qualidade de vida do portador.

Existem iniciativas e associações no mundo que objetiva dar assistência e apoio aos portadores de DA e seus familiares. A Alzheimer’s Disease Internacional, por exemplo, declarou o dia mundial da doença de Alzheimer, dia 21 de setembro, dia em que ocorrem diversos movimentos no mundo visando à conscientização através de palestras informativas para que se diminua o preconceito[2]. A Associação Brasileira de Alzheimer (ABRAz) é um outro grande exemplo. Trata-se de uma entidade que não possui fins lucrativos e é constituída por profissionais de saúde e familiares de portadores da DA. A ABRAz objetiva tornar público o conhecimento sobre a doença esclarecendo a importância do diagnóstico precoce além do apoio a familiares, cuidadores profissionais e demais profissionais na área de saúde para que sejam mais capazes de cuidar dos portadores de Alzheimer [1]. Ela também estimula pesquisas e trabalhos científicos na área.

2.2. Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) são modelos computacionais que se inspiram no funcionamento do cérebro humano para resolver problemas reais. Desta forma as RNA’s são constituídas por unidades de processamento conhecida como neurônio

artificial, uma metáfora a os neurônios biológicos que são lentos e simples (do ponto de vista computacional), mas constituem uma rede maciçamente paralela capaz de resolver problemas complexos de forma mais rápida do que qualquer computador existente, já que podem processar trilhões de operações simultaneamente.

O primeiro modelo de neurônio artificial foi elaborado pelo neurofisiologista McCulloch e pelo matemático Pitts em 1943 [14]. Neste trabalho eles propuseram a construção de uma máquina inspirada no funcionamento do cérebro humano, servindo de base para o surgimento da neuro computação. Na Figura 2 podemos visualizar o modelo artificial proposto por McCulloch e Pitts.

Para o entendimento do funcionamento de uma RNA é necessário entender o funcionamento de um neurônio artificial, e para isto é importante entender a Lei do Tudo ou Nada, ou seja, a emissão de sinal se baseia na regra de propagação função somatório e a função de ativação.

A emissão de sinal em um neurônio artificial está relacionada à Lei do Tudo ou Nada presente nos conceitos do neurônio biológico. Esta lei descreve a intensidade de estímulo, chamado de limiar excitatório. A Lei do Tudo ou Nada se refere ao comportamento desse estímulo, ou seja, se dispara, ou não, o estímulo nervoso. O disparo do estímulo só ocorre se a sua intensidade for superior a um limiar (Θ) ocorrendo sempre com a mesma intensidade (independente do quão superior seja a intensidade do limiar), caso contrário não ocorre emissão do sinal. Na Figura 2 o limiar está implícito e é representado por W_0 , ou seja, $W_0 = -\Theta$, e por esse motivo o $X_0=1$.

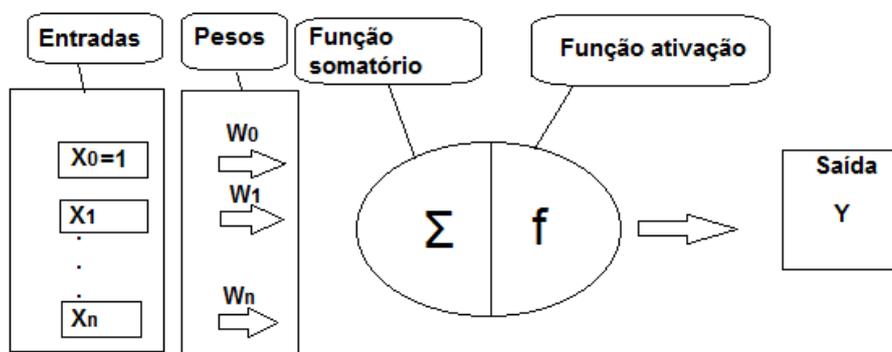


Figura 2. Representação do neurônio artificial proposto por McCulloch e Pitts.

[Fonte: Elaboração própria]

A uma função de somatório representa a soma do produto de cada entrada (X_i) pelo peso sináptico (W_i) atribuído a sua ligação. O resultado do somatório conhecido como entrada líquida (net_i). A equação 1 representa a função somatório descrita. Como podemos observar na Figura 2, a entrada X_0 é igual 1, isso ocorre porque nesta representação, o limiar θ está implícito em $\sum_{i=0}^n X_i \cdot W_i$.

Também podemos observar que o valor de X_0 é 1. Isto ocorre porque a Figura 2 é uma representação com limiar implícito, ou seja, $W_0 = -\theta$. Se representarmos o limiar explicitamente, a equação do net_i pode ser descrito como na equação 2. A representação implícita é usada por ser mais adequado para algoritmos de aprendizagem[7]. Porém não muda o comportamento pois na equação 1, que representa implicitamente o limiar, o X_0 é igual 1 e W_0 é igual a $-\theta$. Pela lei do Tudo ou Nada, só haverá condução de estímulo se $\sum_{j=0}^n X_j \cdot W_{ij}$ for maior que o limiar θ .

$$net_i = \sum_{j=0}^n X_j \cdot W_{ij} \quad (1)$$

$$net_i = \sum_{j=1}^n X_j \cdot W_{ij} - \theta \quad (2)$$

A função de ativação, descrita na equação 3, calcula a saída de um neurônio artificial usando como entrada a entrada líquida (net_i) calculada na equação 1. Na literatura existem diversos exemplos para função de ativação, porém as mais usadas são a linear, sigmoide logística e tangente hiperbólica[7], representadas pelas equações 4, 5 e 6 respectivamente.

$$f(net_i) = Y \quad (3)$$

$$f(net_i) = net_i \quad (4)$$

$$f(net_i) = \frac{1}{1 + e^{-net_i}} \quad (5)$$

$$f(net_i) = \frac{e^{net_i} - e^{-net_i}}{e^{net_i} + e^{-net_i}} \quad (6)$$

A primeira arquitetura de RNA foi o *Perceptron*, desenvolvido em 1958 por Frank Rosenblatt. O *Perceptron* é uma rede muito simples, composta por uma camada de entrada e uma camada de saída, sendo cada camada composta por um ou mais neurônios. Cada neurônio de entrada se conecta a cada neurônio de saída através de uma ligação ponderada onde o peso é o peso sináptico. A função de saída (ou função de ativação) estabelecida para o *Perceptron* é a função degrau, desta forma quando o neurônio está ativo (entrada líquida superior ao limiar), é emitido sinal 1 e quando está em repouso (entrada líquida inferior ao limiar) emite sinal 0. A equação 7 ilustra a expressão da função degrau.

$$f(net) = \begin{cases} 1, & \text{se } net \geq 0; \\ 0, & \text{se } net < 0; \end{cases} \quad (7)$$

Em 1960, Windrow e Hoff apresentaram regras de aprendizagem e aplicaram aos conceitos de *Perceptron* dando origem a ADALINE (Adaptive Linear Neuron). A regra de aprendizagem se baseava no método dos mínimos quadrados (conhecidas como regra delta) e são utilizadas até os dias atuais [7]. Além disso, a ADALINE possibilita o uso de uma função contínua (diferentemente da função degrau) para função de saída, permitindo resolver problemas contínuos, enquanto que *Perceptron* só resolve problemas discretos. As funções mais usadas são: Linear (equação 4), Sigmóide Logística (equação 5) e a tangente hiperbólica (equação 6).

A aprendizagem da RNA se dá através dos ajustes dos pesos sinápticos que ocorrem durante a fase de treinamento, quando os exemplos (conjuntos de entradas e saída) são apresentados. Como se conhece, na maioria dos casos, previamente a saída desejada de cada exemplo usado no treino, pode-se dizer que a aprendizagem da RNA é do tipo supervisionado. Porém existem alguns exemplos de RNA com aprendizagem não supervisionada, mas todas RNA's que serão apresentadas neste trabalho serão com aprendizagem supervisionada.

Embora sejam capazes de resolver diversos problemas, tanto *Perceptron* quanto ADALINE são incapazes de resolver problemas não linearmente separáveis, que são a maioria dos problemas reais. Isto ocorre devido as suas arquiteturas que só possuem uma camada de entrada e uma camada de saída.

2.2.1. Multi-Layer Perceptron

A rede *Multi Layer Perceptron* (MLP) possui uma arquitetura diferente do *Perceptron* e da ADALINE. Pode-se dizer que é um *Perceptron*, mas com muitas camadas, ou seja, a MLP apresenta pelo menos uma camada intermediária (escondida) de neurônios (entre a camada de entrada e de saída). Estas camadas são responsáveis pela não linearidade da rede, permitindo que a MLP seja capaz de resolver problemas não linearmente separáveis, desde que os neurônios desta camada possuam função de ativação não linear como a função sigmoide logística (equação 5) ou tangente hiperbólica (equação 6).

A presença de pelo menos uma camada intermediária permite uma RNA aproximar qualquer função contínua e se houver mais de uma camada aproxima qualquer função matemática [15].

A Figura 3 representa um exemplo de topologia da RNA MLP com uma camada escondida. Através da imagem podemos observar que a RNA MLP é composta por um conjunto de neurônios artificiais interconectados através de ligações que ponderam os sinais, mas diferentemente do *Perceptron* e da ADALINE, é constituída por pelo menos três camadas, são elas:

- **Camada de entrada:** onde cada neurônio representa uma variável de entrada do problema;
- **Ao menos uma camada intermediária (ou escondida):** camada responsável pela não linearidade da rede (dependendo da função de ativação utilizada) e pela aproximação de quaisquer funções matemáticas (dependendo da quantidade de camadas e das funções de ativação).
- **Camada de saída:** formada por um ou mais neurônios responsáveis pela resposta da rede e por representar a variável (saída) desejada.

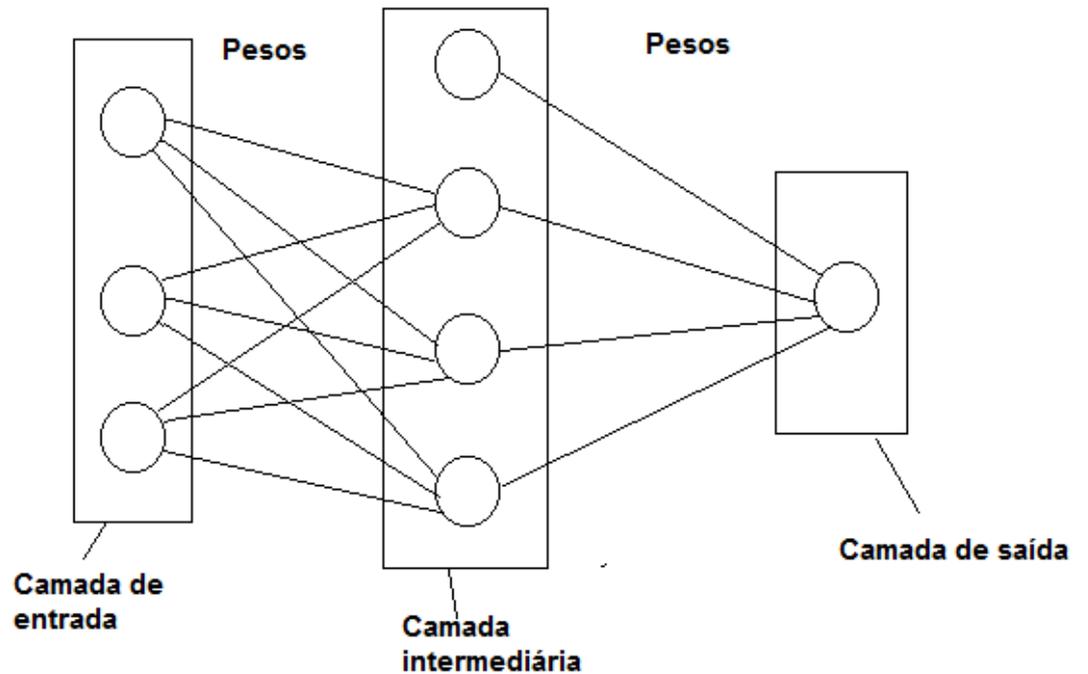


Figura 3. Representação de Rede Neural Artificial Multi Layer Perceptron com uma camada intermediária (ou camada escondida) de neurônios. [Fonte: Elaboração própria]

Durante a aprendizagem, os pesos são inicializados aleatoriamente e são ajustados durante a etapa de treino até obterem resultados otimizados para solucionar o problema em questão. Qualquer topologia de RNA utilizada necessita de um algoritmo de aprendizagem para a otimização dos pesos. Como a MLP possui camadas intermediárias, o processo de treinamento se torna mais complexo em relação à *Perceptron* e ADALINE por exemplo. O algoritmo de aprendizagem tradicionalmente usado na literatura para as redes MLP é o algoritmo *backpropagation*, que será mais detalhado na seção a seguir.

2.2.1.1. Algoritmo *Backpropagation*

O algoritmo *backpropagation* é uma generalização da regra delta, conhecida como técnica do gradiente descendente, utilizada por Widrow e Hoff na aprendizagem da rede ADALINE. O *backpropagation* foi desenvolvido por Paul Werbos em 1974[16].

No *Perceptron* e na *ADALINE*, a aprendizagem supervisionada se dá através dos ajustes dos pesos. Mesmo possuindo algoritmos de aprendizagem diferentes, ambos necessitam da saída desejada para cálculo do erro e a partir disto ajustam os pesos. Porém, na aprendizagem supervisionada só se tem conhecimento de exemplos de entradas e de saídas desejadas, mas, como não se conhece a “saída desejada” da camada escondida e, conseqüentemente, não se conhece o erro necessário para ajuste dos pesos na camada escondida, há uma maior dificuldade em otimizar os pesos para esta camada.

Para resolver este problema, o algoritmo *backpropagation* realiza uma propagação recursiva dos erros que são calculados a partir da saída desejada e da calculada. O funcionamento deste algoritmo pode ser dividido em dois passos, são eles:

- **Forward:** Nesse passo os sinais são propagados progressivamente no sentido entrada-saída passando pela camada escondida. Nesta etapa a saída é calculada e posteriormente calcula-se o erro.
- **Backward:** Nesse passo o erro, calculado na etapa anterior, é propagado recursivamente no sentido da camada da saída para a de entrada passando pela camada intermediária. Neste processo a rede ajusta seus pesos usando a regra delta generalizada.

Visando minimizar o erro entre saída desejada e saída calculada, usa-se a equação 8 na etapa *backward* para ajuste dos pesos.

$$W_{i,j}^m(t+1) = W_{i,j}^m(t) + \alpha \cdot \delta_i^m \cdot f^{m-1}(net_j^{m-1}) \quad (8)$$

Onde $W_{i,j}^m(t+1)$ é o peso ajustado, $W_{i,j}^m(t)$ é o peso antes do ajuste, α representa a taxa de aprendizagem, δ_i^m é o termo de sensibilidade representada pelas equações 9 e 10, sendo a primeira equação usada no ajuste dos neurônios da camada de saída, quando se conhece o erro entre a saída desejada (d_i) e saída calculada (y_i), e a segunda equação usada para ajustes dos pesos da camada escondida (quando não se conhece a saída desejada). O $f^{m-1}(net_j^{m-1})$ são sinais de entradas propagados pelos neurônios da camada anterior (camada “ $m - 1$ ”).

$$\delta_i^m = (d_i - y_i) \cdot f'_{(net_i)} \quad (9)$$

$$\delta_i^{m-1} = f'^{m-1}(net_j^{m-1}) \cdot \sum_{i=0}^N w_{ij}^m \cdot \delta_i^m \quad (10)$$

A equação de ajuste de pesos do algoritmo *backpropagation* representada pela equação 8 permite que hajam grandes possibilidades do algoritmo ficar preso nos mínimos locais, evitando que se encontre os valores ótimos para o peso. Isto ocorre devido à complexidade da função erro. Por este motivo, surgiram algumas propostas de modificações, uma delas consiste na inclusão de um novo termo, taxa de momento, representado por (β) dando origem a o algoritmo *backpropagation* modificado representado pela equação 11.

$$W_{i,j}^m(t+1) = W_{i,j}^m(t) + \alpha \cdot \delta_i^m \cdot f^{m-1}(net_j^{m-1}) + \beta \cdot \Delta w_{i,j}^m(t-1) \quad (11)$$

Estes passos do algoritmo *backpropagation* se repetem por muitos ciclos (iterações) para todos os exemplos utilizados para treinamento da rede. Após cada ciclo pode-se calcular o erro médio quadrático (EMQ) e, após o termino do treino, é possível gerar um gráfico (EMQ x iterações) contendo a curva chamada de curva de aprendizagem. As equações 12, 13 e 14 representam o erro médio quadrático (para N exemplos de um ciclo) e erro quadrático (para um exemplo n qualquer) e o erro médio instantâneo (para S neurônios na camada de saída) respectivamente.

$$EMQ = \frac{1}{N} \cdot \sum_{k=1}^N e_M^2(k) \quad (12)$$

$$e_i^2(n) = (d_i - y_i)^2 \quad (13)$$

$$e_M^2(n) = \frac{1}{S} \cdot \sum_{k=1}^S e_S^2(n) \quad (14)$$

Dependendo da quantidade de ciclos usados para treinamento, é possível que a rede “memorize” os exemplos, dando falsa impressão de aprendizagem, ou seja se fosse fazer testes com os mesmo exemplos usados para treinamento, poder-

se-ia obter resultados próximo de 100 % de acerto, mas com exemplos diferentes poderia dar valores bem baixos. Para evitar esse problema existem técnicas de validação como a validação cruzada que será apresentada mais detalhadamente da próxima seção.

2.2.1.2. Validação Cruzada

É uma técnica que tem como objetivo evitar que a rede “memorize” os exemplos perdendo a sua capacidade de generalização. Tal problema pode ser causado pelo excesso de treinos, chamado de *overfitting*.

Esta técnica consiste no cálculo do erro médio quadrático para um conjunto de dados que não é utilizado para ajuste dos pesos (equação 15). Desta forma, a cada ciclo o treinamento o erro é calculado para este conjunto chamado de conjunto de validação cruzada.

$$EMQ^{VC} = \frac{1}{N} \cdot \sum_{k=1}^N e_M^{VC}(k) \quad (15)$$

$$e_i^{VC}(n) = (d_i - y_i)^2 \quad (16)$$

$$e_M^{VC}(n) = \frac{1}{S} \cdot \sum_{k=1}^S e_S^{VC}(n) \quad (17)$$

2.2.2. Support Vector Machines

As Máquinas de Vetores de Suportes (denotada como SVMs, do inglês *Support Vector Machines*) são rede neurais artificiais baseadas na teoria do aprendizado estatístico (TAE), a qual foi proposta inicialmente em 1971 por Vapnik e Chervonenkis [8] e foi desenvolvida, mais detalhadamente, em 1995 e em 1998 por Vapnik[9][10].

O objetivo do desenvolvimento das SVMs é obter redes que, durante a fase de treinamento, busquem minimizar o erro de treinamento e, ao mesmo tempo, reduzir a complexidade da rede obtida.

O desejável no treinamento supervisionado de uma rede é minimizar o erro de generalização, ou seja, minimizar o erro ao apresentar novos exemplos à rede. Na TAE, este erro é conhecido como Risco Funcional e pode ser obtido a partir da distribuição de probabilidade da população [10]. Porém em problemas reais, em geral, se usa amostra da população em questão, desta forma a distribuição não é conhecida. Porém, é possível calcular uma função de erro conhecida como Risco Empírico, podendo ser o erro médio quadrático, por exemplo, mas, dependendo do tamanho da amostra, a minimização do Risco Empírico não significa necessariamente a minimização do Risco Funcional.

De acordo com a TAE, a capacidade máxima de generalização está relacionada com a minimização do Risco Funcional, e para esta minimização, é necessário minimizar de dois termos, do Risco Empírico (função erro de treinamento) e da complexidade da rede, conhecido como Risco Estrutural ($R_{estrutural}$) representado pela equação 18. Isto é possível graças a variável h presente na equação do Risco Estrutural. Esta variável representa a dimensão VC (Vapnik-Chervonenkis), ou seja, o número máximo de exemplos de treinamento que uma máquina de aprendizagem é capaz de classificar corretamente para todas as possíveis combinações binárias destes dados.

$$R_{estrutural} = \sqrt{\frac{h \cdot \left(\log\left(\frac{2N}{h}\right) + 1 \right) - \log\left(\frac{\delta}{4}\right)}{N}} \quad (18)$$

Através da equação 18, podemos observar que o Risco Estrutural ($R_{estrutural}$) depende da dimensão VC (denotada pela variável h), número de exemplos (N) e é limitado a uma probabilidade $1 - \delta$.

Através da minimização o Risco Estrutural e do Risco Empírico, é possível minimizar o Risco funcional, dando à rede a capacidade máxima de generalização. Neste processo, as SVMs buscam não só um hiperplano que satisfaça um problema, como o *Perceptron*, por exemplo, mas um hiperplano que seja ótimo, ou seja, o que possui maior margem de separação de classes.

Na obtenção do hiperplano ótimo, ou seja, a maior possível, ocorre o ajuste dos pesos (w), os quais são representados por vetores que partem da origem e são ortogonais a os hiperplanos que separam as duas classes. Na classificação de duas classes, cada uma das duas margens é delimitada por alguns elementos (os vetores de suporte) de uma das duas classes e os elementos que não estão contidos nas margens não são considerados na busca do hiperplano ótimo. Quanto à aplicação, as SVMs podem ser divididas em SVMs Lineares e SVMs Não Lineares.

2.2.2.1. SVMs Lineares

São usadas para resolver problemas linearmente separáveis, buscando a maior margem (solução ótima) entre duas classes. Existem dois tipos: SVMs com margens rígidas (máxima) e SVMs com margens suaves.

2.2.2.1.1. SVMs com Margens Rígidas

As SVMs de margens rígidas foram as primeiras formulações de SVM. A figura 5 ilustra um exemplo de classificação de duas classes (A e B).

Através da Figura 4 podemos notar a separação entre duas classes (A e B) através de um hiperplano cuja margem está a uma distância d do hiperplano e é a maior possível para a classificação. Os elementos que estão em vermelho (*vetores de suporte*), tanto da classe A quanto da classe B, definem a fronteira, ou seja, as margens devem passar por eles. Consequentemente não existe nenhum elemento de nenhuma classe entre as margens.

Por se tratar de uma busca por um hiperplano ótimo, a determinação dos pesos exige maior esforço computacional, pois se transforma num problema otimização com restrição de margens máximas e pode ser resolvida através do método de multiplicadores de Lagrange.

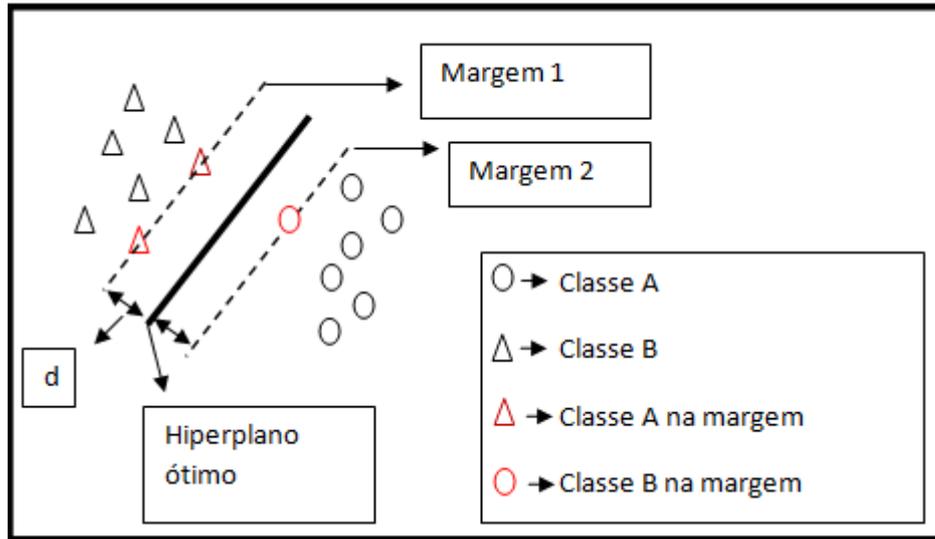


Figura 4. Representação de uma classificação binária realizada por uma SVM de margem rígida [Fonte: Elaboração própria]

Um conjunto de dados de treino pode ser representado por:

$$(x_1, y_1), \dots, (x_n, y_n), X \in R^d, Y \in \{-1, +1\} \quad (19)$$

Onde a variável X representa vetor de entrada e Y representa o valor de saída ou classificação do dado de entrada, podendo ser -1 para uma classe e $+1$ para outra classe. No exemplo da figura 5, poderia ser usado o valor -1 para classe A e $+1$ para a classe B.

Desta forma, a equação do hiperplano ótimo pode ser representada pela equação 20. Onde a variável w representa o peso e o valor de y é zero. Conseqüentemente, as margens 1 e 2 podem ser representadas pelas equações 21 e 22.

$$(w \cdot x) - b = 0 \quad (20)$$

$$(w \cdot x) - b = -1 \quad (21)$$

$$(w \cdot x) - b = +1 \quad (22)$$

Baseado nas equações 20, 21 e 22, podemos descrever a separação do hiperplano de acordo com o seguinte formato:

$$(w \cdot x_i) - b \geq +1 \text{ se } y_i = +1 \quad (23)$$

$$(w \cdot x_i) - b \leq -1 \text{ se } y_i = -1$$

Este formato pode ser representado por uma notação compacta de acordo com a expressão 24

$$y_i \cdot [(w \cdot x_i) - b] \geq 1 \quad (24)$$

A distância d entre as margens pode ser obtida através da distância euclidiana conforme a equação 25. Para maximizar a esta distância (d) entre as margens é necessário minimizar a função objetivo descrita na equação 26 obedecendo às condições descritas pela inequação 24[10]. A variável W representa o vetor de pesos e $\|W\|$ representa a norma euclidiana (equação 27) do vetor de pesos.

$$d = \frac{2}{\|W\|} \quad (25)$$

$$\phi(W) = \frac{\|W\|^2}{2} \quad (26)$$

$$\|W\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2} \quad (27)$$

Para informações mais detalhadas sobre a minimização por Lagrange pode ser encontrado em [9] e[10].

2.2.2.1.2. SVMs com Margens Suaves

Observando a Figura 5, podemos notar que se um problema for linearmente separável e contiver alguns ruídos, representados por elementos de uma classe fora de sua fronteira, seja entre as margens ou na região de outra classe, as SVMs com margens rígidas (máximas) não são capazes de resolver, pois não há tolerância nesta formulação. Para resolver problemas dessa natureza, existe uma formulação criada que tolera a presença de ruídos. Trata-se da SVM com margem suave. A

figura 6 exemplifica uma classificação usando SVM com margens suaves em um problema linearmente separável, mas com alguns ruídos.

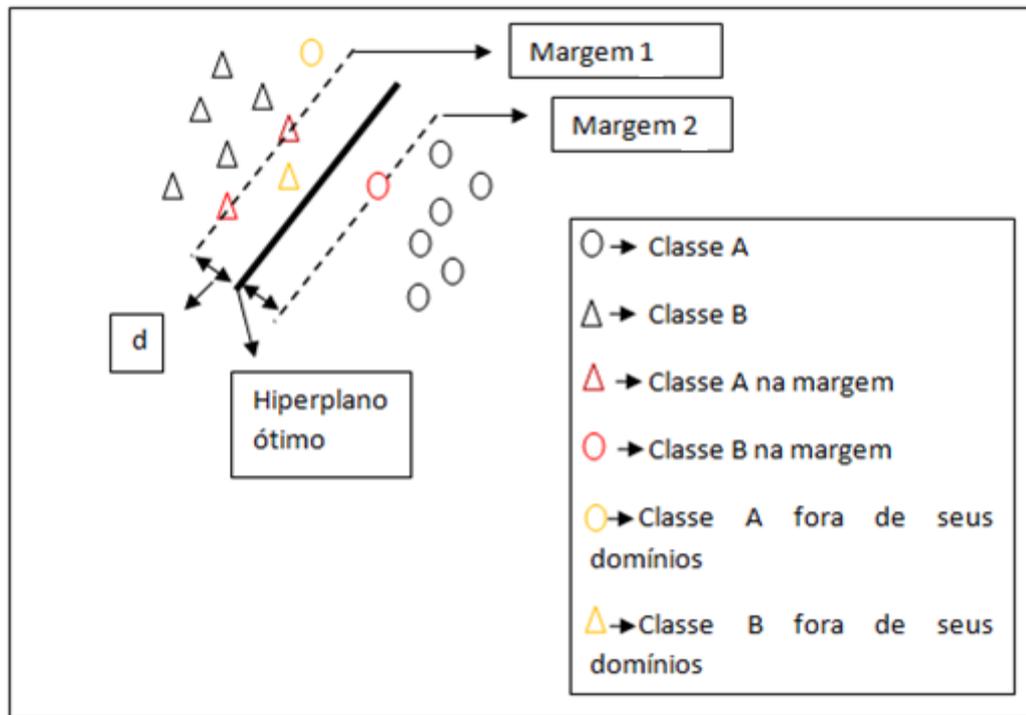


Figura 5. Representação de uma classificação binária realizada por uma SVM de margem suave [Fonte: Elaboração própria]

2.2.2.2. SVMs Não-Lineares

As SVMs de margens rígidas e de margens suaves são incapazes de resolver problemas não linearmente separáveis. Para estes tipos de problemas pode-se aplicar SVMs não lineares, que possuem funções kernel na camada escondida.

São exemplos de funções kernel: função Sigmoid Logística (equação 4), função Tangente Hiperbólica (equação 5) e a função de Base Radial (FBR) Gaussiana (equação 28).

A Figura 6 apresenta um exemplo de problema de classificação não linearmente separável em duas dimensões. Porém aplicando uma função kernel, é possível realizar um mapeamento para uma terceira dimensão provocando deslocamento de uma das classes de modo que se possam separá-las através de um hiperplano utilizando uma SVM linear.

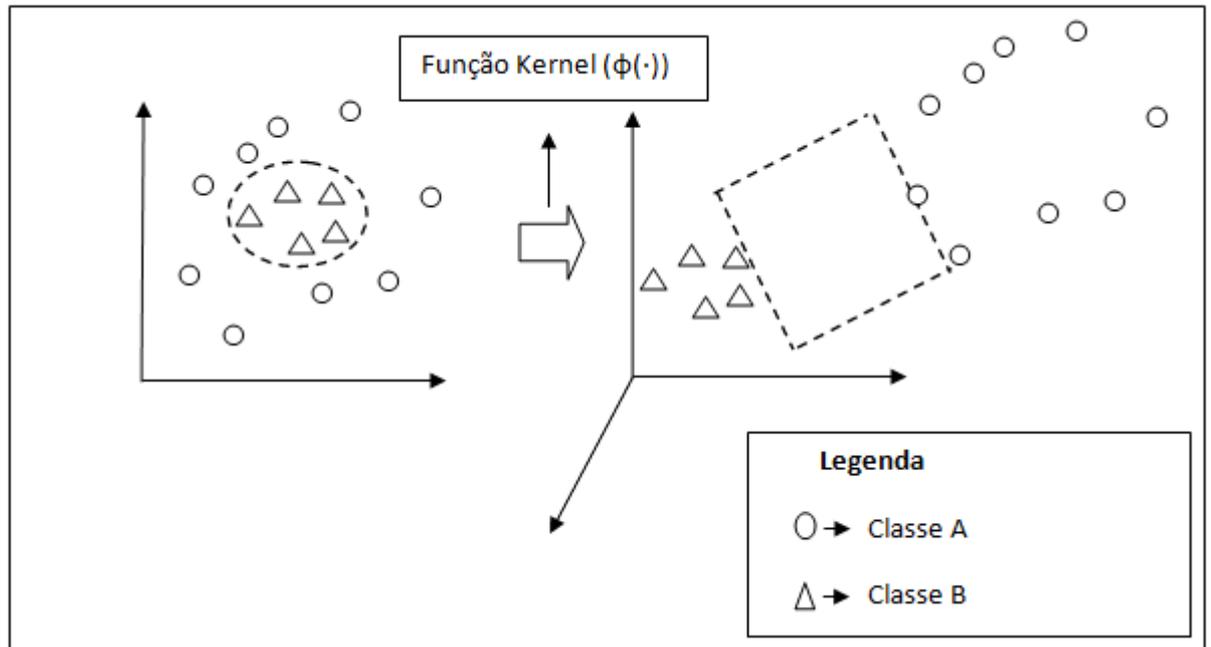


Figura 6. Representação de uma classificação binária realizada por uma SVM não linear [Fonte: Elaboração própria]

$$y = \exp\left(-\frac{\|x_j \cdot \mu_i\|^2}{2 \cdot \sigma_i^2}\right) \quad (28)$$

3. Metodologia

3.1. Base de Dados

Neste trabalho será utilizada a mesma base de dados que foi empregada no desenvolvimento da publicação de Moscato [6] e de Dantas L. em [17] e em [18]. Esta base foi utilizada para classificar se um paciente pode ou não ser diagnosticado com DA através de valores de 120 proteínas adquiridas a partir de análise de amostras sanguíneas de alguns pacientes.

Em sua publicação, Moscato dividiu a base em 3 conjuntos de amostras. O primeiro contendo 83 amostras sendo 68 usadas para treinamento dos classificadores utilizados, e 15 usadas no processo de critério de parada de treinamento, a validação cruzada (mencionada na seção 2.2.1.2), o segundo conjunto contendo 92 amostras de pacientes com ou sem DA e é utilizado para fase de testes.

Moscato realizou testes com o intuito de reduzir a quantidade de proteínas necessárias para classificação. Os testes foram realizados com assinaturas de 18, 10, 6 e 5 proteínas e, como já foi mencionado na introdução, concluiu que a assinatura com cinco proteínas é suficiente para classificação do diagnóstico de DA.

Desta forma, este trabalho tem como objetivo reproduzir o experimento de Moscato [6] e pelos trabalhos realizados por Dantas L. em [17] e em [18] para assinatura com 5 proteínas (IL-1^a, IL-3, EGF, TNF-a e G-CSF) para dois classificadores, MLP e SVM, para diagnosticar DA.

Para o experimento foi utilizado a mesma estratégia adotada por Moscato para assinatura de 5 proteínas, ou seja, em um total de 175 amostras serão usadas 83 para treinamento (sendo 15 delas para validação cruzada) e 92 para fase de testes. O objetivo é verificar as taxas de acertos de cada classificador e compara-las.

3.1.1. Pré – Processamento

Com o objetivo de se obter os melhores resultados, é fundamental realizar alguns procedimentos antes de utilizar os classificadores. É preciso estar atento para as diferenças de valores dos exemplos, pois segundo Valença [7], isto pode comprometer na capacidade de generalização da RNA na etapa de treino.

Para evitar perda de capacidade de generalização dos classificadores utilizados, neste trabalho foi usada uma técnica de normalização bastante utilizada na literatura, a transformação linear, representada pela equação 29, onde x_{min} representa o menor valor de entrada, x_{max} representa o maior valor e a e b são os valores mínimos e máximos, respectivamente, utilizados para normalização de um valor x.

$$y = \left((b - a) \cdot \left(\frac{x - x_{min}}{x - x_{max}} \right) \right) + a \quad (29)$$

Através desta transformação linear os valores são normalizados num intervalo entre a e b. Neste trabalho foram escolhidos os valores 0,15 e 0,85 para a e b respectivamente. Após a normalização, a base poderá ser utilizada para etapa de treinamento da RNA. Porém, após o término da fase de treino, é necessário desnormalizar, utilizando a mesma estratégia da equação 29, o valor para o intervalo inicial para calcular a taxa de acertos.

3.2. Experimentos dos Classificadores

Após a etapa de pré-processamento dos dados da base, descritas na seção anterior, inicia-se a aplicação dos classificadores. Para todos classificadores empregados, foi utilizada a mesma base de dados, contendo assinatura de 5 proteínas, seguindo as mesmas estratégias de Moscato em relação a separação da base em conjuntos de treino, validação e teste descritas na seção 3.1.

Para cada RNA aplicada, foram realizados 30 experimentos. Para isso utilizou-se a ferramenta Matlab, a qual permite criação de todas as redes neurais, que serão aplicadas neste trabalho, de forma configurável.

As próximas seções descreverão de forma mais detalhada os experimentos de cada rede neural empregada.

3.2.1. Experimento da MLP

Para realização dos experimentos com a MLP, foi necessário definir alguns parâmetros de configuração para o treinamento da rede. A Tabela 1 descrevem as configurações utilizadas para os 30 experimentos da rede MLP.

Tabela 1. Representação dos parâmetros utilizados na rede MLP para experimentos dos dados normalizados [Fonte: elaboração própria]

| Parâmetro | Valor |
|--|--------------------|
| Número de neurônios na camada de entrada | 5 |
| Número de camadas escondidas | 1 |
| Número de neurônios na camada escondida | 10 |
| Número de neurônios na camada de saída | 2 |
| Função de ativação dos neurônios da camada escondida | Sigmoide Logística |
| Função de ativação dos neurônios da camada de saída | Linear |
| Taxa de aprendizagem | 0,7 |
| Taxa de momento | 0,4 |

3.2.2. Experimento da SVM

Assim como no experimento da MLP, foi necessário definir parâmetros de configuração para inicialização e treinamento da SVM. Foram realizadas 3

experimentos com diferentes funções *kernel* para se comparar o resultados das taxas e definir a melhor função para o problema em questão.

A Tabela 2 descreve todas as configurações que foram utilizadas para as três funções, a Sigmoides Logística (equação 5), a Tangente Hiperbólica (equação 6) e a Base Radial Gaussiana (equação 28)

Tabela 2. Representação das funções *kernel* utilizadas na rede SVM para simulações [Fonte: elaboração própria]

| Simulação | Função |
|--------------------|------------------------------|
| Primeira simulação | Sigmoides Logística |
| Segunda simulação | Tangente Hiperbólica |
| Terceira simulação | Função Base Radial Gaussiana |

3.3. Teste da Soma dos Postos Wilcoxon

Nas seções 3.2.1 e 3.2.2 nota-se que foram realizadas 30 experimentos para cada classificador. Após armazenar os resultados, foi necessário realizar um teste estatístico para validação científica dos resultados. Como os dados não são normalmente distribuídos, foi utilizado o teste não paramétrico de Wilcoxon.

Para realização deste teste estatístico, foi utilizada a ferramenta de software matemático R, que contém a implementação deste e de outros testes estatísticos. A ferramenta R utiliza o valor de nível de significância (denotada por α) de 0,05 (valor padrão).

O teste de Wilcoxon analisa os resultados baseando-se a partir da mediana de duas amostras independentes. Para realização do teste foram utilizadas hipóteses formuladas, as quais estão representadas na Tabela 4.

Tabela 3. Formulação das hipóteses definidas na aplicação do teste da Soma dos Postos de Wilcoxon [Fonte: elaboração própria]

| Hipótese | Descrição |
|-----------------------------|---|
| Hipótese nula | As medianas das amostras são estatisticamente iguais |
| Hipótese Alternativa | As medianas das amostras analisadas não são estatisticamente iguais |

Desta forma, se o *p-value*, calculado durante o teste, for menor que α , então se rejeita a hipótese nula acusando diferença significativa entre as duas amostras.

4. Resultados

Este capítulo apresentará todos os resultados obtidos pelos experimentos realizados neste trabalho. Como foi definido na metodologia, capítulo 3, foi utilizada uma base contendo 5 proteínas definidas por Moscato como suficientes para classificação, além disso foram realizados um total de 30 experimentos para cada classificador.

Desta forma, o objetivo é verificar o desempenho da MLP e da SVM com base de 5 proteínas e comparar seus resultados.

Antes de realizar a comparação entre a MLP e a SVM, foram realizados experimentos utilizando SVM com três funções kernel diferentes como objetivo de se determinar a SVM com melhor desempenho. No entanto notou-se que todas as taxas dos 30 experimentos da SVM não variavam para cada função Kernel escolhida. Isto pode ser explicado pelo fato de a SVM realizar otimização durante o treinamento buscando melhor hiperplano de separação das classes (seção 2.2.2.). Desta forma não é necessário realizar teste estatístico para escolha, sendo a melhor taxa o suficiente para a escolha da SVM com melhor desempenho.

A Tabela 4 exibe as taxas dos experimentos realizados para SVM utilizando diferentes funções Kernel. Como podemos observar, a melhor taxa foi da SVM função Base Radial Gaussiana como função Kernel.

Tabela 4. Representação da média das taxas de classificação para 30 experimentos das SVMs com diferentes funções Kernel [Fonte: elaboração própria]

| Função Kernel usada na SVM | Média das taxas para 30 simulações |
|----------------------------|------------------------------------|
| Sigmoide Logística | 82,6% |
| FBR (Gaussiana) | 95,65% |
| Tangente Hiperbólica | 54,34% |

A Tabela 5 exibe a média das taxas para 30 experimentos tanto para MLP como para SVM com a função kernel RBF Gaussiana.

Tabela 5. Representação da média das taxas de classificação para 30 experimentos

[Fonte: elaboração própria]

| Classificador | Média das taxas para 30 simulações |
|----------------------|---|
| MLP | 90,8% |
| SVM | 95,65% |

Como foi mencionado no capítulo 3, para validação dos resultados dos 30 experimentos armazenados, foi realizado um teste estatístico (Soma dos Postos Wilcoxon). Mesmo com todas as taxas dos 30 experimentos da SVM possuindo o mesmo valor (95,65%), foi necessário realizar o teste estatístico da Soma dos Postos, já que para a MLP notou-se valores diferentes para cada uma das 30 simulações, o que pode ser explicado pela inicialização aleatória dos pesos sinápticos iniciais.

A tabela 6 exibe o teste estatístico de Wilcoxon realizado utilizando o Software R. Na realização do teste estatístico foi utilizado o nível de significância pré- definido 0,05 (capítulo 3). O resultado para o teste foi um *p-value* muito inferior a 0,05. Desta forma se descarta a hipótese nula e pode-se concluir que a SVM possui melhor desempenho.

Tabela 6. Formulação das hipóteses definidas na aplicação do teste da Soma dos Postos de Wilcoxon para os resultados da MLP e SVM [Fonte: elaboração própria]

| Hipótese | Descrição |
|-----------------------------|--|
| Hipótese nula | As redes MLP e SVM possuem o mesmo desempenho. |
| Hipótese Alternativa | A SVM possui melhor desempenho |

.A Tabela 7 exibe de forma resumida os resultados encontrados para as topologias MLP e SVM em comparação aos resultados encontrados por Moscato [6].

Tabela 7. Representação das comparações das média das taxas de classificação com os disponíveis na literatura [Fonte: elaboração própria]

| Classificador | Média das taxas para 30 simulações |
|----------------|------------------------------------|
| MLP | 90,8% |
| SVM | 95,65% |
| Moscato | 93% |

5. Conclusões e Trabalhos Futuros

A doença de Alzheimer é uma enfermidade neurodegenerativa, irreversível e progressiva que causa declínio das funções cognitivas provocando, principalmente, dificuldade de aprendizagem e perda de memória. Partindo da premissa que a população alvo da doença é constituída predominantemente por idosos acima de 65 anos, sendo a idade um fator de risco importante, é esperado que a incidência da DA aumente cada vez mais, devido ao aumento da expectativa de vida presenciada nos últimos anos, principalmente nos países desenvolvidos.

O fato de a DA constituir uma doença que compromete significativamente a vida dos portadores em vários aspectos – físicos, sociais - já é suficiente para entender a importância do tema abordado neste trabalho a fim de que o diagnóstico precoce seja alcançado. Pois intervenção terapêutica com fármacos e medidas não farmacológicas deve ser iniciada de forma imediata a fim de promover a melhor qualidade de vida possível amenizando os sintomas.

Este trabalho teve como objetivo realizar experimentos utilizando duas topologias diferentes de Redes Neurais Artificiais, A *Multi-Layer Perceptron* (MLP) e a *Support Vector Machine* (SVM) e, após as análises estatísticas para validações dos resultados, realizar comparações dos resultados obtidos.

Assim como a MLP, a SVM também é capaz de resolver problemas não linearmente separáveis, desde que se use uma função kernel, e mesmo que tenha maior custo computacional, a SVM possui a vantagem de realizar otimização durante a fase de treinamento de modo que se tenha o melhor hiperplano separador de classes.

Através dos experimentos, foi observado que a rede SVM obteve um melhor desempenho, tendo uma maior capacidade de generalização para este problema. No entanto foi necessário realizar testes com diferentes funções *kernel* para que se chegasse a este resultado, e a SVM com melhor desempenho utilizou a função *kernel* RBF Gaussiana.

Visando obter melhores resultados, para os trabalhos futuros, serão utilizadas técnicas de otimização para obtenção dos melhores valores para os parâmetros da rede Support Vector Machine.

Bibliografia

- [1] ABRAz, “ABRAz,” ABRAz, 06 Agosto 2012. [Online]. Available: <http://abraz.org.br/>. [Acesso em 30 Março 2015].
- [2] ADI, “ADI,” 6 Agosto 2012. [Online]. Available: <http://www.alz.co.uk/>. [Acesso em 30 Março 2015].
- [3] K. B. E. K. Hestad K, “Low blood pressure is a better predictor of cognitive deficits than the apolipoprotein e4 allele in the oldest old,” 2005.
- [4] A. P. ASSOCIATION, “APA,” 1995. [Online]. Available: <http://www.psychiatry.org/>. [Acesso em 10 Maio 2015].
- [5] L. Caixeta, Doença de Alzheimer, Porto Alegre: artmed, 2012.
- [6] M. Gómez Ravetti, “Identification of a 5-protein biomarker molecular signature for predicting alzheimer's disease.,” vol. III, 2008.
- [7] M. Valença, Fundamentos das Redes Neurais, Recife: Livro Rápido, 2013.
- [8] V. N. Vapnik e A. Y. Chervonenkis, “On the uniform convergence of relative frequencies of events to their probabilities,” *Theory Probab. Appl.*, n. 17, 1971.
- [9] V. Vapnik, “The Nature of Statistical Learning Theory,” 1995.
- [10] V. Vapnik, “Statistical Learning Theory,” 1998.
- [11] E. Giusti e V. Surdo, Cuidados Clínicos e Aconselhamento Familiar, Gryphus, 2010.

- [12] P. R. d. B. Marques, Demência tipo Alzheimer: diagnóstico, tratamento e aspectos sociais, Editora de Publicações Científicas Ltda, 1997.
- [13] M. Torabi, H. Moradzedeh, R. Varizi, S. M. J. Razavian, R. D. Ardekani, A. Taalimi e E. Fatemizadeh, "DEVELOPMENT OF ALZHEIMER'S DISEASE RECOGNITION USING SEMIAUTOMATIC ANALYSIS OF STATISTICAL PARAMETERS BASED ON FREQUENCY CHARACTERISTICS OF MEDICAL IMAGES," *IEEE International Conference on Signal Processing and Communications*, 2007.
- [14] W. S. McCulloch e W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, 1943.
- [15] S. Haykin, Redes Neurais: Princípios e Práticas, Bookman, 2007.
- [16] P. J. Werbos, "Beyond Regression: New tools for prediction and analysis in behavioral sciences," *Tese (Doutorado)- Harvard University*, 1974.
- [17] L. Dantas, "Using Neural Networks in the Identification of Signatures for Prediction of Alzheimer's Disease," *ICTAI*, pp. 238-242, 12 Novembro 2014.
- [18] L. Dantas, "UTILIZANDO RESERVOIR COMPUTING PARA REALIZAR DIAGNÓSTICO PRECOCE DE MAL DE ALZHEIMER," Recife, 2012.
- [19] A. P. ASSOCIATION. [Online].

Apêndice A

Resultado do Teste de Wilcoxon no Software

Neste Apêndice está contida a imagem referente ao resultado obtido após execução da Ferramenta R para teste estatístico de Wilcoxon das taxas da SVM e MLP, como foi descrito no Capítulo 4. Na imagem, MLP representa um vetor contendo as 30 taxas dos experimentos da MLP e SVM representa um vetor com as 30 taxas para SVM.

```
> wilcox.test(MLP, SVM)

      Wilcoxon rank sum test with continuity correction

data:  MLP and SVM
W = 60, p-value = 3.891e-10
alternative hypothesis: true location shift is not equal to 0
```

Figura 7. Execução do teste de Wilcoxon [Fonte: elaboração própria]