



# UTILIZAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA REALIZAR DIAGNÓSTICOS DE DIABETES

**Trabalho de Conclusão de Curso**

**Engenharia da Computação**

**Jonathan da Silva Bandeira**  
**Orientador: Prof. Mêuser Jorge Silva Valença**



UNIVERSIDADE  
DE PERNAMBUCO

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

**JONATHAN DA SILVA BANDEIRA**

**UTILIZAÇÃO DE REDES NEURAIS  
ARTIFICIAIS PARA REALIZAR  
DIAGNÓSTICOS DE DIABETES**

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

**Recife, dezembro de 2017.**

## MONOGRAFIA DE FINAL DE CURSO

### Avaliação Final (para o presidente da banca)\*

No dia 19 de dezembro de 2017, às 9:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente JONATHAN DA SILVA BANDEIRA, orientado pelo professor Méuser Jorge Silva Valença, sob título Utilização de Redes Neurais Artificiais para realizar diagnóstico de diabetes, a banca composta pelos professores:

Sérgio Mário Lins Galdino  
Méuser Jorge Silva Valença

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: 10,0 ( dez )

\*Obrigatório o preenchimento do campo abaixo com comentários para o autor

O discente terá 07 dias para entrega da versão final da monografia a contar da data deste documento.

SÉRGIO MÁRIO LINS GALDINO

MÉUSER JORGE SILVA VALENÇA

\* Este documento deverá ser encadernado juntamente com a monografia em versão final.

*Dedico este trabalho aos meus pais, Lúcia Maria Bandeira e José da Silva Bandeira Filho, ao meu orientador, Mêuser Valença, e aos colegas que contribuíram e me apoiaram durante este projeto, tornando possível a sua execução e conclusão.*

# Agradecimentos

Agradeço primeiramente à Deus e aos meus pais, José da Silva Bandeira Filho e Lúcia Maria Bandeira por tornarem este momento possível e por me apoiarem sempre e incondicionalmente.

Meu muito obrigado aos colegas de curso e de trabalho, professores, amigos e familiares que dividiram comigo experiências e momentos únicos e inesquecíveis ao longo desta jornada de aprendizado e crescimento pessoal e profissional.

Um agradecimento especial ao meu orientador, Prof<sup>o</sup>. Dr. Mêuser Valença, e aos colegas: Caio Vinícius, Pedro Rudá, Everalda Soares, Cauanne Linhares e Natália Guimarães pelas sugestões, inspirações, motivação, suporte técnico e apoio na elaboração deste documento e no desenvolvimento desta pesquisa. A conclusão deste trabalho só foi possível graças a todos vocês, muito obrigado. Foi uma honra trabalhar com todos.

# Resumo

Nos últimos anos, o crescimento e envelhecimento populacional, bem como a intensa urbanização e aumento da obesidade e sedentarismo na sociedade têm sido fatores cruciais para o aumento de casos de diabetes *mellitus*, doença que hoje já afeta pouco mais de 13 milhões de pessoas (cerca de 6,9%) da população adulta no Brasil, segundo a Sociedade Brasileira de Diabetes (SBD), e 422 milhões de adultos (8,5%) em todo planeta, segundo dados da Organização Mundial de Saúde (OMS). A Federação Internacional de Diabetes revela uma tendência crescente destes números e estima que até 2035 haja cerca de 19,2 milhões de diabéticos no Brasil e 471 milhões de diabéticos no mundo. A doença é crônica e consiste em uma série de distúrbios metabólicos que, se não tratados adequadamente, podem trazer complicações fatais para o paciente, tais como: infarto, danos aos vasos sanguíneos, cegueira e AVC. O alto teor de glicose no sangue é um dos indicadores da doença, no entanto, as diversas causas da enfermidade podem dificultar ou atrasar o diagnóstico clínico final. Segundo a SBD, o quanto antes um diagnóstico de comprovação da doença ou de um estado de risco (pré-diabetes) for feito, mais chances têm os pacientes de manter um estilo de vida controlado e com menos complicações. No caso dos pré-diabéticos, um diagnóstico antecipado pode fazer com que estas pessoas possam se prevenir de modo mais eficaz. A proposta deste trabalho é averiguar se uma técnica de aprendizado de máquina é capaz de realizar um diagnóstico rápido e preciso da ausência ou presença de diabetes em uma população de indivíduos. Para tal, utilizou-se a técnica de Redes Neurais Artificiais (RNAs), já utilizada em outros trabalhos similares para definir classificadores e a base de dados da população indígena de Pima, onde há alta taxa de incidência da doença. Se optou por uma análise comparativa, considerando duas topologias de RNAs: *Multi-Layer Perceptron* e *Reservoir Computing*. Desta forma, também é possível analisar qual das arquiteturas apresenta o melhor desempenho na resolução do problema.

**Palavras-chave:** Diabetes, Classificação, *Multi-Layer Perceptron*, *Reservoir Computing*, Redes Neurais Artificiais.

# Abstract

In recent years, population growth and aging, as well as the intense urbanization and increase of obesity and sedentary lifestyle in society have been crucial factors for increase cases of diabetes mellitus, a disease that today affects slightly more than 13 million people according to the Brazilian Society of Diabetes (SBD), and 422 million adults (8.5%) worldwide, according to the World Health Organization (WHO). The International Diabetes Federation shows a growing trend in these numbers and estimates that by 2035 there are about 19.2 million diabetics in Brazil and 471 million diabetics in the World. The disease is chronic and consists of a series of metabolic disorders that if not properly treated can produce fatal complications for the patient, such as: infarction, damage to blood vessels, blindness and stroke. The high glucose content in the blood is one of the indicators of the disease, however, the various causes of the disease may hamper or delay the final clinical diagnosis. According to the SBD, if an earlier diagnosis of disease or a state of risk (pre-diabetes) is made, patients will be able to maintain a controlled lifestyle with fewer complications. In case of pre-diabetics, an early diagnosis can make these people to prevent themselves against the disease more effective. The purpose of this study is to determine if a machine learning technique is able to perform a quick and accurate diagnosis of the absence or presence of diabetes in a population of individuals. For this, we used the technique of Artificial Neural Networks (ANNs), already used in similar works as classifiers and Pima indians database, where there is a high incidence rate of diabetes. We chose a comparative analysis, considering two topologies of ANNs: Multi-Layer Perceptron and Reservoir Computing. In this way, it is also possible to analyze which of the architectures presents the best performance in solving this problem.

**Keywords:** Diabetes, Reservoir Computing, Multi-Layer Perceptron, Artificial Neural Networks, Classification

# Sumário

<b>Capítulo 1 Introdução</b>	<b>1</b>
1.1 Motivação	1
1.2 Objetivos	3
1.2.1 Objetivo geral	3
1.2.2 Objetivos específicos	3
1.3 Estrutura da monografia	4
<b>Capítulo 2 Fundamentação Teórica</b>	<b>5</b>
2.1 <i>Diabetes Mellitus</i>	5
2.2 Redes Neurais	10
2.2.1 Rede Neural <i>Multi-Layer Perceptron</i> (MLP)	14
2.2.2 <i>Reservoir Computing</i>	16
<b>Capítulo 3 Metodologia</b>	<b>18</b>
3.1 Base de Dados	18
3.1.1 Pré-processamento da base de dados	20
3.2 Aplicação da topologia MLP	23
3.2.1 Treinamento	23
3.2.2 Experimentos	25
3.3 Aplicação da topologia RC	25
3.3.1 Configuração	26



3.3.2	Treinamento	27
3.3.3	Experimentos	28
3.4	Testes estatísticos	29
3.4.1	Teste t-Student	29
3.4.2	Teste Shapiro-Wilk	30
3.4.3	Teste F	30
3.4.4	Teste da Soma dos postos de Wilcoxon	31
<b>Capítulo 4</b>	<b>Análise dos Resultados</b>	<b>32</b>
<b>Capítulo 5</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>35</b>
	<b>Referências</b>	<b>37</b>
<b>Apêndice A</b>	<b>Tabela de resultados individuais</b>	<b>40</b>
<b>Apêndice B</b>	<b>Script utilizado para análises estatísticas</b>	<b>41</b>
<b>Apêndice C</b>	<b>Resultados das análises estatísticas</b>	<b>42</b>
<b>Anexo A</b>	<b>Pseudocódigo do <i>Reservoir Computing</i></b>	<b>43</b>

# Índice de Figuras

Figura 1.	Principais tipos de diabetes mellitus.....	8
Figura 2.	Modelo geral de um neurônio artificial .....	11
Figura 3.	Funções de ativação.....	13
Figura 4.	Modelo geral de uma RNA MLP .....	15
Figura 5.	Topologia Reservoir Computing .....	17

# Índice de Tabelas

<b>Tabela 1.</b>	Outros tipos específicos de DM	9
<b>Tabela 2.</b>	Exemplo da base de dados PID	19
<b>Tabela 3.</b>	Relação de valores nulos na PID	21
<b>Tabela 4.</b>	Parametrização da rede neural MLP	25
<b>Tabela 5.</b>	Parametrização do RC	28
<b>Tabela 6.</b>	Hipóteses do teste t-Student	29
<b>Tabela 7.</b>	Hipóteses do teste Shapiro-Wilk	30
<b>Tabela 8.</b>	Hipóteses do teste F	30
<b>Tabela 9.</b>	Hipóteses do teste da soma dos postos de Wilcoxon	31
<b>Tabela 10.</b>	Média dos resultados dos experimentos com as topologias MLP e RC	32
<b>Tabela 11.</b>	Hipóteses consideradas no teste Shapiro-Wilk para a MLP	33
<b>Tabela 12.</b>	Hipóteses consideradas no teste Shapiro-Wilk para o RC	33
<b>Tabela 13.</b>	Hipóteses consideradas no teste F para MLP e RC	33
<b>Tabela 14.</b>	Hipóteses consideradas no teste t-Student	34
<b>Tabela 15.</b>	Comparação das médias dos testes estatísticos obtidos com a literatura	34

# Tabela de Símbolos e Siglas

<b>SIGLA</b>	<b>SIGNIFICADO</b>
A.C.	Antes de Cristo
ADA	American Diabetes Association
ANN	Artificial Neural Network
AVC	Acidente Vascular Cerebral
DM	Diabetes Mellitus
DM1	Diabetes Mellitus Tipo 1
DM2	Diabetes Mellitus Tipo 2
DM 1A	Diabetes Mellitus Tipo 1 Autoimune
DM 1B	Diabetes Mellitus Tipo 1 Idiopática
ESN	Echo State Network
IDE	Integrated Development Environment
LSM	Liquid State Machine
MCP	Modelo de McCulloch e Pitts
MLP	Multi-Layer Perceptron
OMS	Organização Mundial de Saúde
PID	Pima Indians Diabetes
PNS	Pesquisa Nacional de Saúde
RC	Reservoir Computing
RNA	Rede Neural Artificial
RNP	Rede Neural Pulsada
RNR	Rede Neural Recorrente
SBD	Sociedade Brasileira de Diabetes
SVM	Support Vector Machine
WHO	World Health Organization

# Capítulo 1

## Introdução

Neste capítulo são expostos a motivação para realização desta pesquisa, seus objetivos e a estruturação e organização geral deste documento.

### 1.1 Motivação

O atual ritmo de crescimento e envelhecimento populacional, aliado a intensa urbanização e aumento da obesidade e sedentarismo na sociedade, têm se mostrado como fatores cruciais para o desenvolvimento de doenças ligadas a distúrbios metabólicos, tais como o *diabetes mellitus* (DM) [18].

Há relatos documentados de sintomas que remetem ao DM desde antes de Cristo [4]. No entanto, a doença só foi melhor compreendida à partir do momento em que o médico Harold Percival Himsworth classificou a enfermidade quanto à forma de resposta de um organismo ao tratamento com insulina, em 1936 [9]. Atualmente, o padrão de classificação da doença adotado pela Organização Mundial de Saúde (OMS) é mais robusto e baseia-se na etiologia, ou seja, nas causas e origens da doença e não mais no tipo de tratamento. O novo padrão considera quatro classes clínicas ao invés das tradicionais duas estabelecidas por Harold P. Humsworth.

Diabetes, para muitos médicos, não é apenas uma doença, mas um conjunto de distúrbios metabólicos que apresenta como principal problema a hiperglicemia (alto teor de glicose no sangue), seja ela resultante de uma falha na secreção da insulina, seja na ação da insulina no organismo ou em ambas [24]. O mal tratamento destes distúrbios pode acarretar outros problemas, tais como complicações cardiovasculares, Acidente Vascular Cerebral (AVC) e infarto [25]. Na maioria dos países desenvolvidos, a doença também já figura entre a quarta e oitava posição como uma das principais causas de óbito devido às complicações e agravantes que ela provoca.

A enfermidade, desconsiderando as particularidades de cada tipo, pode atingir qualquer faixa etária, apesar de mais comum na faixa dos 45 aos 64 anos e em mulheres (cerca de 7,0% dos casos), segundo a Pesquisa Nacional de Saúde (PNS) realizada no Brasil em 2013 [10]. Hoje, o diabetes já atinge cerca de 13 milhões (6,9%) e 422 milhões (8,5%) de pessoas na população adulta do Brasil e do mundo, respectivamente, segundo dados da Sociedade Brasileira de Diabetes (SBD) [18] e da OMS [1]. Considerando as pesquisas do ano de 1980 da OMS, que registravam cerca de 108 milhões de diabéticos no mundo, os casos quadruplicaram. No Brasil, a situação não é melhor. Em 2013, a PNS registrou 9 milhões de diabéticos (6,2%) na população adulta do país. Isto implica em um aumento de 4 milhões de casos da doença no país em um período de apenas quatro anos [20]. A Federação Internacional de Diabetes, em decorrência destes números, já estima um crescimento da incidência da doença para cerca de 19,2 milhões e 471 milhões de casos de diabetes no Brasil e no mundo, respectivamente, até o ano de 2035 [11].

Sendo uma doença crônica, o DM consiste em um quadro clínico que não tem cura e é irreversível. No entanto, existem tratamentos que permitem um convívio aceitável com a doença e um controle do quadro clínico para que não acarrete em complicações severas para o indivíduo diabético. Além disto, caso seja diagnosticada precocemente, alguns casos da doença (DM tipo 2) apresentam um estágio pré-diabético antes de avançarem para a fase diabética propriamente dita e nestes casos, o quadro clínico pode ser revertido. Mesmo um diagnóstico prévio de um quadro clínico já considerado diabético já é extremamente útil para recomendação de um tratamento mais adequado para um indivíduo com a doença. Levando em consideração que os muitos tipos de classificação e particularidades da DM podem dificultar ou atrasar um diagnóstico, se faz necessária a elaboração de modelos ágeis para classificar e mesmo prever diagnósticos relacionados a esta enfermidade.

Na literatura, um modelo para classificação e diagnóstico da presença ou ausência de DM em um grupo de indivíduos utilizando uma rede neural *Multi-Layer Perceptron* (MLP) foi desenvolvido por Marcos Rogério Ferrara [6]. No trabalho dele, o modelo proposto apresentou uma média de acurácia de cerca de 85% em 40 execuções. O trabalho de Ferrara foi aplicado à mesma base de dados adotada neste trabalho e é utilizado como um dos principais parâmetros comparativos durante as análises de resultados desta pesquisa. Outro trabalho relevante e motivador desta pesquisa foi a monografia de Lara Dantas [3], que apesar de aplicar as técnicas de RNA para o problema do mal de Alzheimer, abordou, com ótimos resultados, a utilização

das mesmas topologias que foram adotadas nesta pesquisa para um problema e área de aplicação similares.

Fundamentando-se no que foi exposto, este trabalho visa contribuir para solução do problema aqui elucidado, adotando a topologia de RNA MLP, dita com um bom desempenho para problemas de classificação, além de ser bem conhecida e utilizada para fins semelhantes [27]. Para efeitos comparativos e de inovação, a RNA *Reservoir Computing* (RC) foi adotada, pois além de já se fazer presente em outras publicações que utilizam uma abordagem comparativa de topologias de RNA, é uma arquitetura de rede neural do tipo recorrente, ou seja, possui *feedback* (realimentação) [28]. A grande variedade de comportamentos dinâmicos e não linearidade apresentados por uma rede recorrente, tal qual a RC, é objeto de estudo desta pesquisa, pois pode trazer melhoras nos resultados obtidos por uma rede do tipo *feedforward*, como uma rede MLP, por exemplo. Quando treinada, esta rede neural fornece a mesma resposta de saída para um dado sinal de entrada. Além disto, a utilização de RC não é muito comum, na literatura, para problemas envolvendo o auxílio ao diagnóstico de diabetes, tornando-a um dos diferenciais deste trabalho.

## 1.2 Objetivos

### 1.2.1 Objetivo geral

O objetivo central deste trabalho é avaliar o desempenho obtido na classificação da presença ou ausência de DM em um grupo de indivíduos, utilizando as topologias MLP e RC e comparando os resultados tanto entre as metodologias adotadas, quanto com os resultados advindos de outros trabalhos da literatura.

### 1.2.2 Objetivos específicos

Como metas específicas deste trabalho, destacam-se:

1. Realizar o procedimento de classificação de DM com cada uma das topologias escolhidas;
2. Alterar parâmetros de configuração da MLP e RC afim de obter as respectivas configurações que fornecem melhores resultados;
3. Comparar resultados finais de ambas;
4. Fazer comparação com resultados presentes na literatura;

5. Promover testes estatísticos objetivando demonstrar a confiabilidade do modelo de solução proposto;
6. Expor conclusões e considerações relacionadas às novas metas e trabalhos futuros que poderão ser alcançados após o término deste trabalho.

### 1.3 Estrutura da monografia

Esta monografia está dividida em cinco capítulos: Introdução, Fundamentação Teórica, Metodologia, Análises de Resultados e Conclusões.

O capítulo de Fundamentação Teórica apresenta as descrições conceituais essenciais para o entendimento pleno desta pesquisa. Além disto, traz maior aprofundamento na contextualização realizada neste capítulo introdutório enquanto trata dos conceitos referentes à doença de DM e às redes neurais MLP e RC, que são as redes neurais utilizadas neste trabalho.

Já o capítulo da Metodologia é responsável por trazer os materiais e métodos que foram utilizados para obtenção dos resultados analisados ao fim deste trabalho, relatando em detalhes todo o procedimento da pesquisa. Aqui se encontram descritos desde o pré-processamento e transformações realizadas na base de dados até a obtenção dos resultados com uso das técnicas aplicadas.

No capítulo de Análises de Resultados estão expostas as comparações dos resultados obtidos entre as diferentes topologias de RNA utilizadas e para com os resultados advindos da literatura. Além disto, neste capítulo também se encontram os testes estatísticos aplicados para comparar a eficácia dos resultados provenientes das diferentes técnicas abordadas neste trabalho.

Por fim, o capítulo de Conclusões é direcionado à análise final do autor desta monografia, tendo como base as análises dos resultados obtidos e as dificuldades encontradas para elaboração e execução deste projeto. Este capítulo apresenta ainda uma seção dedicada à listar as melhorias e trabalhos futuros derivados e inspirados pelo término desta pesquisa.



# Capítulo 2

## Fundamentação Teórica

Neste capítulo estão descritos conceitos fundamentais para o entendimento deste trabalho.

### 2.1 *Diabetes Mellitus*

*Diabetes Mellitus* é uma doença crônica que consiste em um conjunto de distúrbios metabólicos caracterizados principalmente pelo alto teor de glicose no sangue de um indivíduo, a chamada hiperglicemia. Em geral, os sintomas incluem aumento de sede e fome, ganho ou perda de peso repentina, fadiga frequente, problemas de cicatrização e circulação, pressão desregulada, tonturas, aumento da vontade de urinar e complicações na visão. Se não for adequadamente tratada, os distúrbios causados pela doença podem acarretar em severos agravantes para o indivíduo. Estes agravantes incluem: cegueira, AVC, infarto, danos aos vasos sanguíneos, problemas renais e úlceras nos pés [25]. Exames de rotina para avaliar os anticorpos relacionados à produção de insulina e para avaliar as taxas de glicose no sangue, alimentação saudável e balanceada, tratamento de estresse e prática regular de atividades físicas são recomendações gerais para prevenção da doença. Para os casos onde um indivíduo atesta a existência de ocorrência(s) de DM em sua família, também recomenda-se um constante e rigoroso acompanhamento médico [18].

Em 1500 A.C. (Antes de Cristo), estudiosos hindus documentaram descrições sobre uma doença, até então desconhecida, que causava um grande aumento na sede e na frequência do ato de urinar. Também se percebeu que formigas e mariposas eram atraídas pelos dejetos expelidos pelas pessoas com os sintomas citados [4].

Em 250 A.C. ouviu-se pela primeira vez o termo *diabetes*, de origem grega, ao qual se atribuiu o significado de sifão, baseando-se no fato de que os líquidos ingeridos eram rapidamente expelidos do organismo, ou seja, algo que passava rápido. Porém, o termo

(considerando seu significado grego) só foi oficialmente documentado e reconhecido em 1425 em um texto traduzido e publicado em inglês, por Adams [15]. Antes disto, o termo foi utilizado pela primeira vez fora da área médica por Columella [16], tendo sido empregado em latim.

O nome *mellitus*, por sua vez, surgiu em 1670 após Thomas Willis redescobrir o sabor adocicado da urina (notado pelos estudiosos hindus em 1500 A.C.). O significado da palavra *mellitus* se origina do latim e é o equivalente à palavra doce, em tradução livre. A descoberta de Willis permitiu que se diferenciasssem as duas classes de diabetes: *mellitus* e *insipidus*. O termo *mellitus* foi, no entanto, acrescentado pela primeira vez por Cullen [2], no século XVIII.

Em 1921, descobriu-se a importância da insulina no tratamento da doença. No ano de 1936, com base nesta e em descobertas anteriores recentes (tal como o papel do pâncreas no organismo e sua influência na doença), Harold P. Himsworth [9] identificou e distinguiu o DM em dois tipos, classificados quanto à forma de tratamento da enfermidade: diabetes sensível à insulina ou insulino dependente e diabetes não sensível à insulina ou não insulino dependente.

Atualmente, a classificação do DM proposta pela OMS e pela ADA se baseia no aspecto etiológico da enfermidade, ou seja, nas origens e causas da doença. Em virtude disto, hoje são consideradas quatro classes clínicas: DM tipo 1 (conhecida anteriormente como insulino dependente), DM tipo 2 (chamada primeiramente de não insulino dependente ou insulino independente), DM gestacional e o grupo de outros casos específicos de DM.

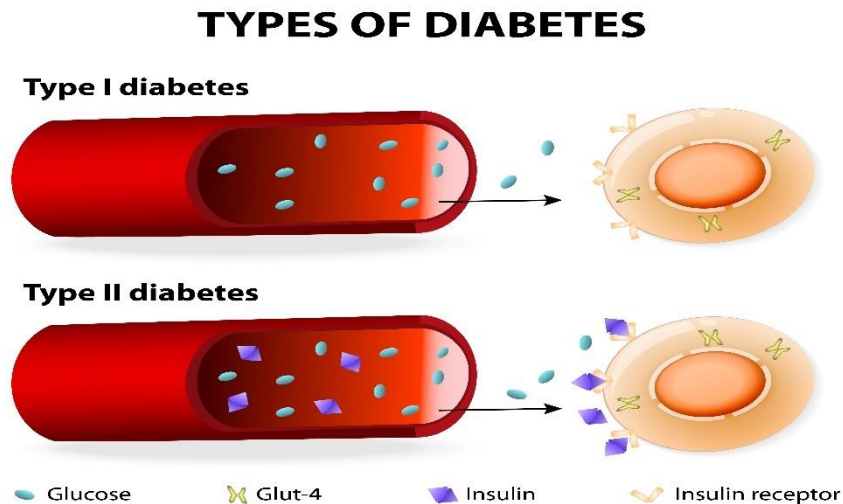
DM1 e DM2 são os dois principais tipos de classificação da doença e ocorrem por motivos distintos. De acordo com a Figura 1, o DM1 ocorre devido ao pâncreas de um indivíduo, responsável pela produção de insulina que controla o nível de glicose do indivíduo, ter as suas células beta destruídas. Isto acarreta na deficiência da produção de insulina, causando o aumento do teor de glicose no sangue (hiperglicemia) e fazendo com que o indivíduo portador da doença precise de um tratamento com insulina exógena, tornando-se dependente da insulina para sobreviver. O diabetes tipo 1 ainda se subdivide em outros dois tipos: DM 1A ou autoimune e DM 1B ou idiopática [18].

A forma autoimune do DM1 corresponde à uma taxa entre 5% e 10% dos casos de DM existentes. Nesta forma da DM1, as células beta pancreáticas são destruídas por um processo imunológico do próprio organismo, que para eliminar as células consideradas uma ameaça, forma os anticorpos: anti-insulina, antidescarboxilase do ácido glutâmico (GAD 65),

antitirosina-fosfatases (IA2 e IA2B) e antitransportador de Zinco (Znt) [25]. Há como prever o caso clínico autoimune mediante exames de sangue com o propósito de verificar a ação dos anticorpos citados. Essa previsão pode ser de meses ou até anos antes de um agravamento da situação clínica e em indivíduos que já apresentem hiperglicemia, há entre 85% e 90% de chance de se verificar esta ação [18]. Por outro lado, a forma idiopática é menos comum e não possui origem ou causa conhecida tal qual a forma autoimune. Por conta disto, indivíduos com esta forma de DM1 tendem a desenvolver a cetoacidose diabética, situação grave da doença em que o indivíduo apresenta desidratação severa, vômitos, dificuldade respiratória e sonolência, podendo entrar em estado de coma. Se faz necessário um tratamento intensivo para evitar fatalidades nestes casos. Nem todas as clínicas são capazes de identificar a ação dos anticorpos da forma autoimune, o que torna a identificação das subcategorias de DM1 nem sempre possível.

O DM2, como ilustrado na Figura 1, ocorre porque a ação da insulina nas células do organismo é dificultada e normalmente é associada ao aumento de peso (obesidade) e ao sedentarismo ou a questões genéticas e ambientais [25]. Isso é chamado de resistência à insulina e caracteriza-se pela dificuldade dos receptores de insulina em distribuir a insulina produzida. Em virtude disto, há um aumento na produção de insulina para tentar estabilizar os níveis de glicose no organismo, sendo este o estágio pré-diabético. Quando não é mais possível manter os níveis de glicose, surge o diabetes. Os sintomas no DM2 se desenvolvem de maneira mais lenta comparado ao DM1 e os tratamentos com insulina não são essenciais para sobrevivência, mas podem ser empregados para melhorar o controle metabólico. O tratamento mais comum do DM2 envolve controle com medicamentos, prática de atividades físicas regularmente e dieta recomendada e aprovada por um endocrinologista e um nutricionista. Se não tratada, entretanto, a doença traz os mesmos agravantes do tipo 1 de DM. O tipo 2 é a manifestação da enfermidade com mais casos, sendo estes cerca de 90%.

Apesar da possibilidade de ocorrerem em qualquer faixa etária, o DM1 costuma acometer crianças e jovens adultos enquanto o DM2 é mais comum em adultos à partir dos cinquenta anos, mas também tem se mostrado mais frequente em jovens e crianças, devido à alimentação mais gordurosa e desbalanceada e ao sedentarismo.



**Figura 1.** Principais tipos de *diabetes mellitus*

(Fonte: [www.dreamstime.com](http://www.dreamstime.com) )

O terceiro tipo de diabetes é o DM gestacional, que se trata de qualquer intolerância à glicose, de magnitude variável, com diagnóstico no início ou durante o período de gestação [18]. O DM gestacional ocorre de forma similar ao DM2, por conta da resistência à insulina e diminuição da função das células beta. Atinge entre 1% e 14% de todas as gestações no mundo e no Brasil, cerca de 7% das gestações são complicadas pela hiperglicemia gestacional. Após o parto, durante as primeiras quatro a seis semanas, deve-se reavaliar o diagnóstico das pacientes para averiguar se elas ainda apresentam DM e intolerância à glicose. Normalmente após o parto a tolerância à glicose retorna aos padrões normais, entretanto, deve-se manter as pacientes em observação e sendo acompanhadas, pois há o risco de 10% a 63% do desenvolvimento futuro de DM2, que varia de acordo com cada país do mundo, durante os 5 a 16 anos posteriores ao parto [18].

O grupo de outros tipos específicos de DM caracteriza-se por apresentar diversas formas menos comuns de DM nos quais os processos causadores podem ser identificados. A Tabela 1 mostra os grupos de processos causadores conhecidos e instâncias específicas de cada grupo.

**Tabela 1.** Outros tipos específicos de DM

(Fonte: Adaptado de OLIVEIRA [18])

<b>Processos causadores</b>	<b>Instâncias das causas</b>
Defeitos genéticos nas funções das células beta	MODY 1 (defeitos no gene HNF4A) MODY 2 (defeitos no gene GCK) MODY 3 (defeitos no gene HNF1A) MODY 4 (defeitos no gene IPF1) MODY 5 (defeitos no gene HNF1B) MODY 6 (defeitos no gene NEUROD1) Diabetes neonatal transitório Diabetes neonatal permanente DM mitocondrial
Defeitos genéticos na ação da insulina	Resistência à insulina do tipo A Leprechaunismo Síndrome de Rabson-Mendenhall DM lipoatrófico
Doenças do pâncreas exócrino	Pancreatite Pancreatectomia ou Trauma Neoplasia Pancreatopatia fibrocalculosa
Endocrinopatias	Acromegalia Síndrome de Cushing Glucagomia Somatostinoma
Introduzido por medicamentos ou agentes químicos	Determinadas toxinas Pentamidina Ácido nicotínico Glicocorticoides Hormônio tireoidiano Interferona Diazóxido
Infecções	Rubéola congênita Citomegalovírus
Formas incomuns de DM autoimune	Síndrome de Stiff-Man Anticorpos antirreceptores de insulina
Outras síndromes genéticas associadas	Síndrome de Down Síndrome de Wolfram Síndrome de Turner Síndrome de Klinefelter Síndrome de Prader-Wil Síndrome de Laurence-Moon-Biedl Distrofia miotônica

Para mais informações detalhadas sobre o diabetes: descrição clínica, etiologia da doença, tratamentos, tipologia e pesquisas de incidência, recomenda-se a leitura do documento de diretrizes de diabetes, elaborado pela SBD, mais recente. Para este trabalho, foi considerado o documento de 2015-2016 [18].

## 2.2 Redes Neurais

Redes Neurais Artificiais (RNA) são modelos computacionais inspirados no sistema nervoso central biológico, em particular os neurônios do cérebro, capazes de realizarem o aprendizado de máquina bem como o reconhecimento de padrões. O principal propósito de uma RNA é simular o comportamento conexionista de uma cadeia de neurônios objetivando aprender com seu ambiente e melhorar seu desempenho. Para isto, realiza-se o procedimento de treino, um processo iterativo de ajuste aplicados aos pesos (sinapses) das conexões entre os neurônios de uma RNA. Existem ainda as redes neurais que não possuem pesos. Estas apresentam como principal diferença com relação às redes convencionais a forma de armazenar informação [7]. Quando uma rede neural consegue generalizar uma solução para uma determinada classe de problemas, afirma-se que houve aprendizado.

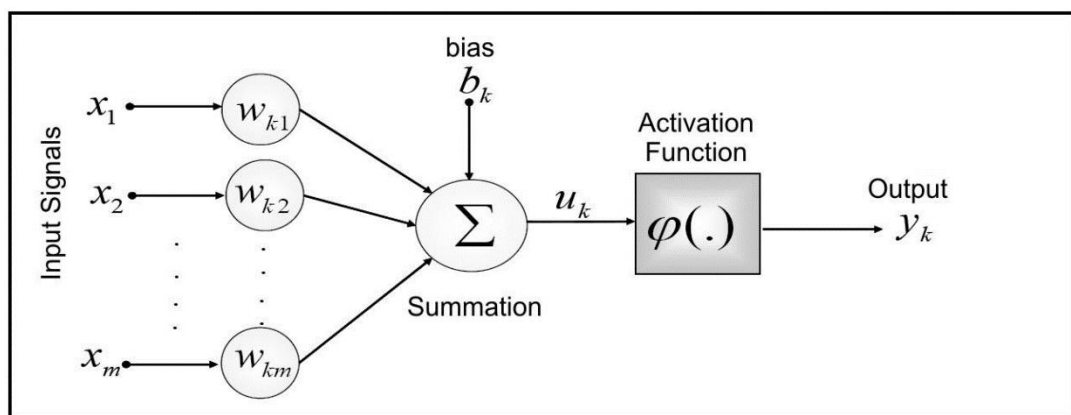
O processo de aprendizado de uma RNA se verifica a partir da modificação dos pesos. Os pesos iniciais, normalmente gerados aleatoriamente, são modificados iterativamente por um algoritmo de treinamento que segue um dos seguintes paradigmas:

- **Aprendizado Supervisionado:** é apresentado um conjunto de treino, consistindo de entradas e correspondentes saídas desejadas.
- **Aprendizado por Reforço:** para cada entrada apresentada, é produzida uma indicação (reforço) sobre a adequação das saídas correspondentes produzidas pela rede.
- **Aprendizado Não-Supervisionado:** a rede tem seus pesos ajustados sem a utilização de pares de entradas-saídas desejadas e sem indicações sobre a adequação de suas saídas correspondentes produzidas pela rede.

Para elaboração e configuração de uma RNA, deve-se conhecer as características dos neurônios, a topologia da rede e as suas regras de treinamento. Uma RNA bem elaborada e configurada traz além da aprendizagem alguns outros benefícios, tais como: alto poder de generalização, flexibilidade, tolerância à falhas, auto-organização e processamento paralelo de informações [8].

Na prática, a RNA pode ser aplicada para solucionar problemas de aproximação de funções, classificações, reconhecimentos de padrões e previsões de séries temporais. Além destes exemplos mais clássicos, aplicações cada vez mais comuns nas quais se podem encontrar redes neurais são: sensoriamento remoto, robótica, auxílio à diagnósticos médicos e biometria.

O primeiro modelo artificial de um neurônio biológico foi proposto por Warren McCulloch e Walter Pitts em 1943 [17]. Este modelo foi pioneiro e fundamental para o desenvolvimento dos modelos e aplicações que surgiram em seguida e para os avanços que culminaram no que se conhece atualmente. A Figura 2 apresenta um modelo geral de um neurônio artificial, com base no proposto por McCulloch e Pitts.



**Figura 2.** Modelo geral de um neurônio artificial  
(Fonte: Adaptado de HAYKIN [8] )

O primeiro elemento do neurônio consiste em um conjunto de sinapses ou elos de conexão, caracterizados cada qual por um peso  $W_{kj}$ , onde  $k = 1, \dots, m$  e corresponde ao número da camada na rede neural e  $j = 1, 2, \dots, n$  é o número do neurônio na camada  $k$ . Para o modelo de McCulloch e Pitts e para a topologia *perceptron* [27],  $k=1$ , pois há apenas uma camada. As equações descritas nesta seção levam em consideração o modelo da Figura 2, com  $k=1$ . O

produto de um sinal de entrada pelo seu respectivo peso sináptico indica a influência deste sinal na saída [8].

O segundo elemento é um somatório para os sinais de entrada, ponderado pelas sinapses do neurônio (pesos) como uma combinação linear [8]. Também é considerado o fator  $b_k$ , o *bias*, conhecido como limiar ou *threshold*. O limiar determina a emissão do sinal por um neurônio em analogia à Lei do Tudo ou Nada [27]. Considerando o limiar explícito como mostrado na figura, a equação resultante da combinação linear pode ser descrita como na Equação 2.1:

$$u_k = \sum_{j=1}^n (w_{kj} \cdot x_j - b_k) \quad (2.1)$$

Nos algoritmos de aprendizagem, no entanto, considera-se que a apresentação do limiar de forma implícita é mais adequada e para isto, duas modificações são realizadas: Considera-se o limiar como um novo sinal de entrada  $\mathbf{X}_0 = + \mathbf{1}$  e adiciona-se um novo peso sináptico correspondente a este sinal, descrito por:  $\mathbf{W}_{k0} = - \mathbf{b}_k$ . A equação resultante pode então ser reescrita como mostra a Equação 2.2 [27].

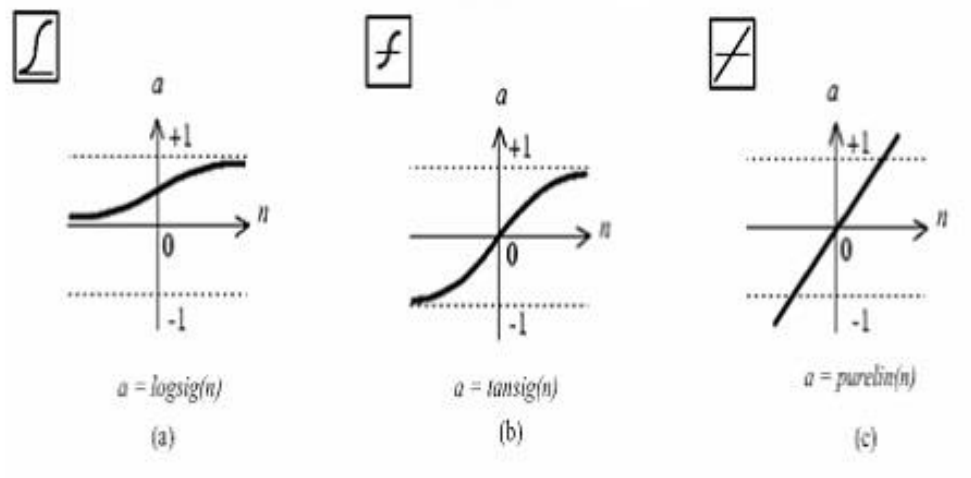
$$u_k = \sum_{j=0}^n (w_{kj} \cdot x_j) \quad (2.2)$$

O terceiro elemento é uma função de ativação, responsável por restringir a amplitude de saída de um neurônio e desta forma, limitar o intervalo de saída para um valor finito, sendo descrita pela Equação 2.3 [8].

$$y_k = \varphi (u_k) \quad (2.3)$$

A Figura 3 apresenta as três funções de ativação mais conhecidas e utilizadas, sendo: (a) a função Sigmóide Logística; (b) a função Tangente Hiperbólica; e (c) a função de ativação Linear.





**Figura 3.** Funções de ativação

(Fonte: Neural Network Toolbox )

Na função sigmoide logística (Equação 2.4), onde  $a$  representa o parâmetro de inclinação da curva, a saída pode assumir valores entre 0 e 1, enquanto na função da tangente hiperbólica (Equação 2.5), assume valores entre -1 e 1, sendo  $b$  os limites superiores e inferiores da função e  $a$ , a inclinação da curva. Na função de ativação linear (Equação 2.6), um número real  $A$  define a saída linear para os valores de entrada.

$$f(u_k) = \frac{1}{1 + e^{-au_k}} \quad (2.4)$$

$$f(u_k) = a \cdot \frac{e^{bu_k} - e^{-bu_k}}{e^{bu_k} + e^{-bu_k}} \quad (2.5)$$

$$f(u_k) = A \cdot u_k \quad (2.6)$$

Outro conceito importante a ser destacado é a arquitetura de uma RNA, ou seja, como estão organizados os neurônios em uma rede neural. As topologias *Perceptron* e *Adaline*, com uma única camada, só são capazes de resolver problemas linearmente separáveis. Esta limitação foi superada com o surgimento de redes neurais com múltiplas camadas, tal qual a MLP. A RNA MLP é parte de um grupo de arquitetura conhecida como *feedforward*, onde os neurônios são agrupados em camadas, os sinais percorrem a rede em uma única direção, da entrada para a saída, e os neurônios de uma mesma camada não possuem conexões entre si. Há ainda a arquitetura de Rede Neural Recorrente (RNR), que possui realimentação, de modo que as saídas de alguns neurônios de uma camada alimentam neurônios da mesma camada (inclusive a si próprios) ou de camadas anteriores. Nesta arquitetura, os sinais percorrem a rede em duas direções, a memória é dinâmica e há capacidade de representar estados em sistemas dinâmicos, sendo uma arquitetura apropriada para resolução de problemas que envolvem processamento temporal [27]. A Rede de Hopfield [27], a *Echo State Network* (ESN) [12] e a rede RC, detalhada na subseção 2.2.2 e no Capítulo 3, são exemplos de redes recorrentes.

### 2.2.1 Rede Neural *Multi-Layer Perceptron* (MLP)

Trata-se de uma topologia *Perceptron* com múltiplas camadas, organizadas e descritas da seguinte maneira:

- **Camada de Entrada:** é a camada onde os padrões de entrada do problema são apresentados à rede.
- **Camadas Intermediárias ou Escondidas:** São as camadas onde a maior parte do processamento é realizado utilizando as conexões ponderadas. Estas camadas podem ser consideradas como extratoras de características. Uma rede MLP possui ao menos uma camada escondida e isto propicia a resolução de problemas não linearmente separáveis. Com uma camada escondida, uma rede MLP pode realizar aproximações no âmbito de qualquer função contínua e à partir de duas camadas escondidas, a aproximação é válida no âmbito de qualquer função matemática.

- **Camada de Saída:** é a camada onde o resultado final, a saída da rede neural, é apresentado.

A Figura 4 expõe a arquitetura de uma rede neural MLP. Também é possível observar na imagem que para uma dada saída obtida, há uma saída desejada correspondente. Isto indica que o aprendizado e treinamento de uma MLP são supervisionados.

Para treinar uma MLP, normalmente utiliza-se o algoritmo de *backpropagation*, criado por Paul Werbos em 1974 [30], ou variações deste. Este algoritmo de treinamento é um dos mais populares e utilizados na literatura para encontrar os pesos ótimos da MLP e se encontra descrito no capítulo 3 deste trabalho.

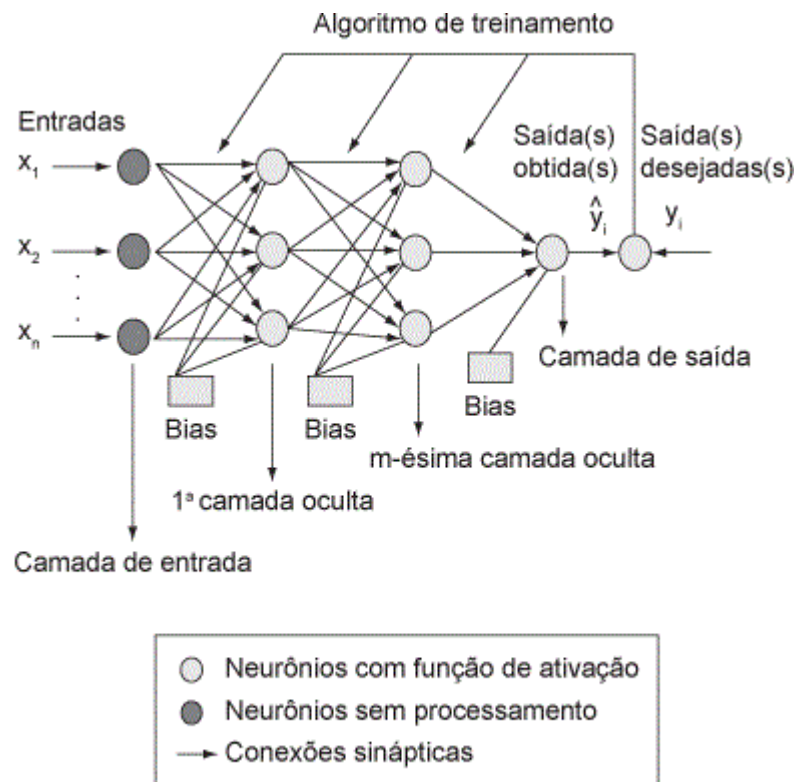


Figura 4. Modelo geral de uma RNA MLP  
(Fonte: Adaptado de SOARES [23])

### 2.2.2 *Reservoir Computing*

A primeira geração de redes neurais artificiais caracterizou-se por apresentar neurônios com saída binária, como por exemplo, o modelo de McCulloch & Pitts (MCP) de 1943. E apesar do modelo MCP não ser tão próximo do comportamento do neurônio biológico real como o modelo de Hodgkin-Huxley, de 1952, o primeiro era **praticável** computacionalmente enquanto o segundo ainda não. As limitações da primeira geração da RNA se deveram ao fato da possibilidade de resolução apenas de problemas discretos e linearmente separáveis [8].

Na segunda geração, surgiram as redes neurais artificiais com saídas mapeadas por funções contínuas como a sigmóide logística e a tangente hiperbólica. Os modelos desta geração já eram capazes de computar funções com entradas e saídas analógicas e suportavam algoritmos de treinamento baseados em gradiente, como por exemplo, o *backpropagation*. A RNA MLP é um dos mais conhecidos exemplos desta geração, que tinha como limitação o comportamento estático, de modo que após o treinamento, para um dado padrão de entrada, a saída correspondente é a mesma, invariante com o tempo [8].

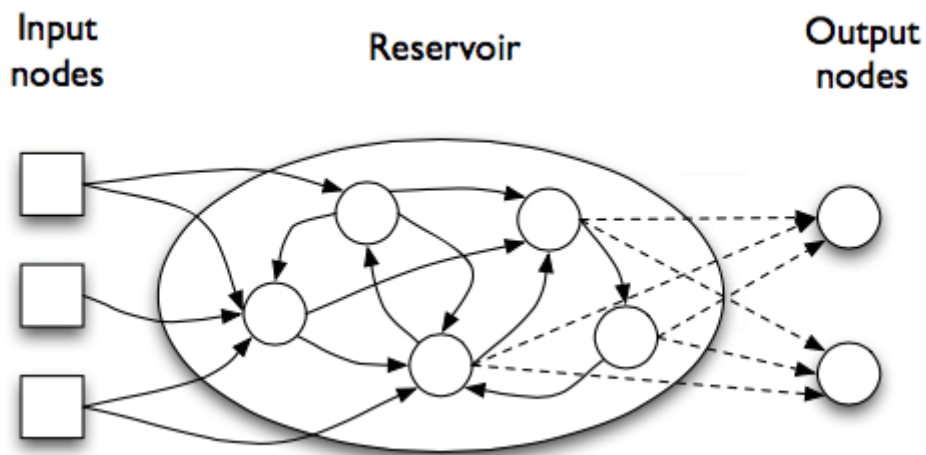
A terceira geração trouxe a proposta das Redes Neurais Pulsadas (RNP) ou *Spiking Neural Networks*. Os neurônios nesta geração são mais próximos do comportamento do neurônio biológico, trazendo para a prática o modelo proposto por Hodgkin e Huxley anos antes e outros modelos mais simples como o Integra e Dispara [21]. Estes modelos possuem representação temporal já que são baseados em pulsos gerados no tempo. A *Liquid State Machine* (LSM), criada por Wolfgang Maass [14] em 2001 é um tipo de RNP.

Tanto a LSM quanto a já citada ESN são redes recorrentes que possuem a propriedade *Echo State Property*, que permite à rede armazenar informações anteriores devido às suas conexões recorrentes. No entanto, a memória destas redes, chamada de *fading memory*, elimina gradualmente as informações mais antigas, atuando de forma similar ao cérebro humano [3].

O termo *Reservoir Computing* (RC) foi utilizado pela primeira vez por Verstraeten em 2007 e unificou os conceitos apresentados pela LSM e pela ESN [28]. Um RC é um *framework* para utilização de uma RNR com objetivo de realizar uma modelagem caixa-preta para sistemas dinâmicos (não-lineares).

Como se pode ver na Figura 5, o RC divide-se em três partes principais:

- **Camada de Entrada:** é a camada que representa as variáveis de entrada do problema.
- **Reservoir:** camada que consiste em uma RNR dinâmica e seus pesos internos fixos e aleatórios.
- **Readout:** é a camada de saída linear da rede. Em um RC, apenas os pesos desta camada são treinados.



**Figura 5.** Topologia *Reservoir Computing*  
(Fonte: Adaptado de SOARES [23] )

Como já mencionado, o grande benefício do RC é que em comparação às outras técnicas de RNR é a simplicidade do treinamento. Em um RC, o treinamento ocorre apenas na camada de *readout*. No entanto, é necessário que se ajustem os parâmetros e a topologia da rede para viabilizar a elaboração de um *reservoir* ótimo e adequado à aplicação. Mais informações com relação à estrutura do RC e seu método de treinamento se encontram no capítulo 3 deste trabalho.

# Capítulo 3

## Metodologia

Neste capítulo se encontra detalhada a metodologia desta pesquisa, compreendendo a descrição da base de dados e as operações realizadas nela, a aplicação das topologias MLP e RC com o intuito de identificar e classificar os padrões presentes na base de dados e, por fim, a apresentação dos testes estatísticos utilizados na análise estatística dos resultados.

### 3.1 Base de Dados

Há grande diferença no predomínio de DM entre os diversos países, classes sociais e grupos étnicos. Registraram-se taxas mais elevadas para Nauru, na Oceania, e para os índios Pima, no Arizona, Estados Unidos, onde quase metade da população adulta apresenta DM [18].

A base de dados utilizada neste trabalho é a *Pima Indians Diabetes* (PID) [19], frequentemente adotada na maioria dos trabalhos disponíveis na literatura que aplicam SVM ao diabetes, como pode ser visto nos trabalhos de Ferrara [6] e Vieira [29]. Elaborada por Knower e outros pesquisadores da Universidade da Califórnia, esta base de dados vem sendo utilizada como recurso para diversas análises e experimentos, principalmente por se tratar de uma das poucas bases de dados abertas que tratam do tema de DM [13].

A PID foi criada a partir da coleta de informações da tribo indígena de Pima (localizada nos arredores de Phoenix, Arizona, nos Estados Unidos) e contém informações referentes à presença de DM nas mulheres da tribo, que possuíam em média 21 anos ou mais na época e faziam parte de um dos grupos com a maior taxa de incidência de DM2 no mundo.

A PID possui 768 registros, dos quais 500 (cerca de 65,1%) correspondem à pacientes não diabéticas e 268 (34,9%) são de pacientes classificadas como diabéticas, ou seja, pouco mais de um terço dos diagnósticos apenas deste espaço amostral utilizado comprovam a alta incidência da doença na tribo. Para realizar o treinamento dos classificadores, 384 (50%)

registros foram utilizados. Os registros restantes foram divididos igualmente em 192 (25%) registros e, em seguida, utilizados no procedimento de validação cruzada (um dos critérios de parada do treinamento dos classificadores) e na realização de testes.

Segundo Valença [27], a utilização do método de validação cruzada como critério de parada de classificadores e técnicas de SVM (baseadas em aprendizado) é eficiente, pois evita que ocorra o superajustamento desses modelos.

**Tabela 2.** Exemplo da base de dados PID

(Fonte: elaboração própria)

I	II	III	IV	V	VI	VII	VIII	IX
6	148	72	35	80	33,6	0,627	50	1
1	85	66	29	80	26,6	0,351	31	0

A Tabela 2 apresenta uma amostra das instâncias da PID, onde pode-se ler, da esquerda para direita, as seguintes informações:

- I. A primeira coluna é o **número de gestações**, quantidade de vezes que a paciente comprovadamente engravidou.
- II. A segunda coluna é a **taxa de glicose no plasma sanguíneo após duas horas do teste de tolerância oral de glicose**. Segundo a ADA, uma taxa de glicose no sangue igual ou superior a 180 mg/dl (miligramas por decilitro) já comprova a presença de DM, porém nem todos os pacientes diagnosticados como diabéticos apresentam essa característica.
- III. A terceira coluna é a **pressão sanguínea diastólica ou mínima**, medida em mmHg (milímetros de mercúrio) e correspondente ao último batimento regular audível.

- IV. A quarta coluna é a **espessura da dobra da pele do tríceps**, considerada dentro do padrão até cerca de 3,5 cm. Os pacientes diagnosticados com diabetes geralmente apresentam uma diminuição da flexibilidade deste músculo.
- V. A quinta coluna diz respeito à **taxa de insulina no soro sanguíneo**.
- VI. A sexta coluna traz o **Índice de Massa Corporal (IMC)** das pacientes avaliadas.
- VII. A sétima coluna representa a **função de pedigree** ou a tendência que um paciente tem de contrair diabetes com base na sua genética.
- VIII. A oitava coluna diz respeito à **idade** das pacientes, calculada a partir da subtração da data do exame pela data de nascimento de cada paciente.
- IX. Por fim, a nona e última coluna corresponde ao diagnóstico, representado como um número binário, onde 1 representa a presença de diabetes e 0 a ausência da doença.

### 3.1.1 Pré-processamento da base de dados

A base de dados apresentou um problema que dificulta tanto a leitura e interpretação dos dados presentes na mesma quanto a sua utilização para identificar alguns padrões. Este problema é a ausência de informações. Visto que as informações ausentes se encontravam especialmente nos registros pertencentes às pacientes diabéticas, eliminar estes registros tornaria a análise desbalanceada, pois haveria significativa diminuição dos casos da classe de diabéticos. Conseqüentemente, isto traria problemas no reconhecimento dos padrões para o classificador. Por este motivo, decidiu-se preencher as lacunas onde se identificou a ausência de informação (estas lacunas se encontravam preenchidas com o valor zero, o que com exceção da primeira e da última coluna, não faz sentido no restante do contexto da aplicação, onde se mostrou equivalente ao valor NULL em bancos de dados).

A Tabela 3 mostra a quantidade de valores nulos identificados por coluna na base de dados.



**Tabela 3.** Relação de valores nulos na PID  
(Fonte: elaboração própria)

Nome da coluna	Quantidade de valores nulos
Número de gestações	Irrelevante, zero é um valor válido
Taxa de glicose no plasma sanguíneo	5 casos
Pressão sanguínea diastólica	35 casos
Espessura da dobra da pele do tríceps	227 casos
Taxa de insulina no soro sanguíneo	374 casos
IMC	11 casos
Função pedigree	Nenhum
Idade	Nenhum
Diagnóstico	Irrelevante, zero é um valor válido

Foram consideradas quatro abordagens para se realizar o preenchimento dos valores nulos: a utilização da média dos valores não nulos de uma coluna (I), a utilização da mediana dos valores não nulos de uma coluna (II), o valor mais frequente de uma coluna (III) e a aproximação de um padrão conhecido (IV). Essas abordagens constituem uma técnica de tratamento de dados faltantes chamada de imputação simples [5]. Com exceção da terceira abordagem, que retornou como valor mais frequente para algumas colunas o número zero, as demais foram utilizadas de acordo com a análise realizada por coluna, que identificou a melhor abordagem a ser utilizada em cada caso, de tal modo que:

- A coluna referente à **taxa de insulina no soro sanguíneo** teve seus valores nulos preenchidos com a média (80), visto que a mediana (30,5) se mostrou como um valor bem menor com relação a maioria dos demais registros da coluna. O valor da média, além de se aproximar mais do restante dos dados desta coluna, tornou-a mais concisa.
- Os valores nulos da coluna de **Índice de Massa Corporal** também foram preenchidos com a média (27,9). Neste caso, optou-se pela média devido a natureza contínua da maioria dos dados da coluna. Além disto, a mediana foi um valor muito próximo da média, não apresentando nenhuma outra consideração significativa.
- As colunas **da espessura da dobra do músculo do tríceps** e da **pressão sanguínea diastólica** tiveram os valores nulos preenchidos com os valores da

mediana (23 e 72, respectivamente). De maneira similar ao caso anterior, não houve muita diferença entre os valores da média e da mediana. No entanto, os valores da mediana são equivalentes a outros registros reais presentes em ambas as colunas e por isto, foram os escolhidos.

- A coluna da **taxa de glicose no sangue** foi a única na qual se utilizou uma aproximação de um padrão conhecido para preencher os valores nulos da coluna. Isto foi realizado, principalmente, por causa da quantidade de registros nulos (apenas cinco). Para os três casos onde o diagnóstico atestava a presença de DM, os valores nulos foram preenchidos com o número 180, tendo em vista que, um valor igual ou superior a 180 mg/dl já comprova a existência da doença e, portanto, é condizente com o diagnóstico. Os dois últimos valores nulos, que atestavam a ausência da enfermidade no diagnóstico, foram preenchidos com dois números da própria coluna da taxa de glicose (75 e 95) que corresponderam aos registros completos com alto grau de similaridade ou equivalência nos valores das demais colunas. Em outras palavras, as duas linhas com informações mais parecidas com as linhas que continham um valor nulo tiveram o valor da taxa de glicose replicado.

Após o tratamento da base de dados, ainda é necessário normalizar os dados antes de iniciar o procedimento de treinamento dos classificadores, pois um intervalo muito discrepante de valores pode prejudicar os cálculos realizados pela técnica de aprendizado de máquina, tendo em vista que as informações estariam em diferentes medidas de escala.

O processo de normalização utilizado é a transformação linear, um dos mais comuns e utilizados na literatura. A fórmula se encontra descrita na Equação 3.1.

$$y = (b - a) \cdot \frac{x - x_{max}}{x_{max} - x_{min}} + a \quad (3.1)$$

Para as constantes a e b presentes na equação, que representam os valores máximos e mínimos que os dados devem assumir, foram adotados os valores de 0,15 e 0,85, respectivamente.

Finalizando a etapa de normalização, os dados estão preparados para atuar como informações de entrada para a técnica de classificação, devendo ser desnormalizados antes de se aferir o percentual de acurácia do procedimento.

Caso as análises fossem sobre uma base de dados de série temporais para realização de previsões, seria necessário ainda defasar os dados normalizados, como atestam os recentes trabalhos de Caio Vinícius [26] e Everalda Soares [23]. Esta observação será utilizada em um dos trabalhos futuros derivados desta pesquisa, que tem como objetivo usar uma série temporal de taxas de glicemia de um grupo de pacientes para realizar uma previsão de suas taxas futuras.

## 3.2 Aplicação da topologia MLP

Nesta seção se encontra detalhada a utilização da MLP, subdividida em dois principais tópicos: treinamento e experimentos.

### 3.2.1 Treinamento

O algoritmo utilizado para treinar a MLP foi o *backpropagation*, que realiza uma propagação recursiva dos erros. O algoritmo consiste em duas fases, sendo a primeira fase *forward* e a segunda *backward*.

Em sua primeira fase, os sinais se propagam progressivamente da camada de entrada para a camada de saída, calculando-se o sinal de saída e o erro. Os pesos nesta fase permanecem fixos [27]. Nesta etapa, calcula-se:

- **Entradas líquidas para a camada escondida ( $net_i^m$ ):** A quantidade calculada varia de acordo com o número de neurônios na camada intermediária e o cálculo consiste em um somatório dos sinais advindos da camada de entrada que se conectam ao neurônio da camada intermediária, ponderados pelos seus respectivos pesos iniciais. A letra  $i$  representa o  $i$ -ésimo neurônio de uma camada e  $m$  corresponde a  $m$ -ésima camada de uma rede neural.
- **Função de saída para os neurônios da camada escondida ( $y_i^m$ ):** O cálculo é realizado para cada neurônio da camada intermediária e varia de acordo com a função de ativação utilizada. As fórmulas para as funções de ativação mais

usadas se encontram descritas no Capítulo 2. Para este trabalho, a função de ativação escolhida para a camada escondida foi a sigmoide logística.

- **Entradas líquidas para a camada de saída ( $net_i^m$ ):** É o mesmo que para camada intermediária, porém considerando os neurônios da camada de saída e os sinais advindos da camada intermediária.
- **Sinal de saída ( $y_i^m$ ):** Calcula-se a função de saída para os neurônios da camada de saída. É o mesmo que para camada intermediária, mas considerando que aqui os resultados compõem o sinal de saída da rede neural. A função de ativação utilizada neste trabalho para a camada de saída também foi a sigmoide logística.
- **Erros da camada de saída:** Calculados à partir da subtração dos valores desejados pelos resultados calculados (Equação 3.5).

A fase *backward* objetiva realizar a retropropagação do erro de tal forma que se possa realizar o reajuste dos pesos. Por se tratar de um aprendizado supervisionado, o ajuste dos pesos deve ser realizado de maneira que seja garantido que o erro entre os resultados desejados e calculados seja mínimo. Nesta etapa se calculam: os valores de sensibilidade ( $\delta$ ), necessários para o ajuste dos pesos, para os neurônios da camada intermediária (Equação 3.2) e para os neurônios da camada de saída (Equação 3.3), bem como os ajustes dos pesos (Equação 3.4) [27].

$$\delta_j^{m-1} = f'^{m-1}(net_j^{m-1}) \sum_{i=1}^{N_{neurônios}} (w_{ij}^m \cdot \delta_i^m) \quad (3.2)$$

$$\delta_i^m = f'(net_i^m) \cdot e_i(n) \quad (3.3)$$

$$w_{ij}^m(novo) = w_{ij}^m(antigo) + \alpha \delta_i^m = f^{m-1}(net_j^{m-1}) \quad (3.4)$$

$$e_i(n) = d_i - y_i \quad (3.5)$$

Os cálculos dos valores de sensibilidade para as camadas intermediárias são feitos recursivamente (daí o nome *backpropagation*) e a função de saída calculada na Equação 3.2 e na Equação 3.3 é a derivada da função de ativação. Já o termo  $\alpha$ , na Equação 3.4, representa a taxa de aprendizagem [27].

### 3.2.2 Experimentos

O modelo de MLP utilizado foi desenvolvido e validado previamente no trabalho de Santos de Oliveira [22].

Primeiramente, foram gerados aleatoriamente os pesos iniciais que ponderam as camadas de entrada, intermediárias e de saída. Os valores gerados compreenderam um intervalo entre 0 e 1. Após gerar os pesos iniciais, a rede neural foi parametrizada, como mostrado na Tabela 4.

**Tabela 4.** Parametrização da rede neural MLP

(Fonte: elaboração própria)

Parâmetro	Valor
Número de neurônios na camada de entrada	8
Número de neurônios na camada de saída	1
Número de neurônios na camada intermediária	6
Função de ativação dos neurônios da camada de saída	Sigmóide logística
Função de ativação dos neurônios da camada intermediária	Sigmóide logística
Taxa de aprendizagem ( $\alpha$ )	0,85
Momento ( $\beta$ )	0,15

Foram realizados 40 experimentos e ao final destes, as taxas de acerto foram registradas para posteriores análises comparativas e estatísticas. Os resultados registrados se encontram no Apêndice A deste trabalho.

## 3.3 Aplicação da topologia RC

Nesta seção se encontram os detalhes da aplicação do *framework* de *Reservoir Computing*, subdivididos em três partes: configuração, treinamento e experimentos.

### 3.3.1 Configuração

Para construir um RC, primeiramente deve-se criar a camada de entrada com uma quantidade de neurônios que seja equivalente a quantidade de variáveis de entrada do problema acrescido do neurônio de limiar.

Em seguida, a quantidade de neurônios no *reservoir* é definida empiricamente, variando de acordo com a aplicação, de modo que seja mais adequada para tal.

Para as conexões entre a camada de entrada e o *reservoir*, gera-se uma matriz de pesos ( $W_{in}$ ). Esta matriz terá como número de linhas a quantidade de neurônios na camada de entrada e como número de colunas a quantidade de neurônios no *reservoir*. Os pesos são aleatórios e estão contidos em um intervalo entre -1 e 1.

Logo na sequência, deve ser gerada outra matriz, correspondente aos pesos que ponderam as conexões entre os neurônios dentro do *reservoir* ( $W_{res}$ ). Estes pesos também são gerados aleatoriamente e estão contidos no intervalo entre -1 e 1. Nesta etapa também é definida a taxa de conectividade do *reservoir*, que determina qual a porcentagem de neurônios com pesos diferentes de zero. A taxa de conectividade também é escolhida empiricamente, variando de acordo com a melhor performance possível para a aplicação.

A matriz  $W_{res}$  deve possuir um raio espectral igual a 1 para quando for inicializada ter seus valores alterados da seguinte maneira:

- É inicialmente decomposta em valores singulares;
- Deve ter seus valores alterados até que o valor máximo da diagonal principal da matriz de autovalores seja menor ou igual a 1.

Por fim, a geração da matriz correspondente aos pesos que ponderam as conexões do *reservoir* para a camada de saída ( $W_{out}$ ) é realizada depois, durante o processo de treinamento, mediante o cálculo da pseudo-inversa de Moore-Penrose [3].

### 3.3.2 Treinamento

O pré-processamento da base de dados deve ser realizado antes de começar o treinamento do RC. Além disto, é necessário dividir a base em três subconjuntos: treinamento, validação cruzada e testes.

O conjunto de treinamento é utilizado para atualizar os estados dos neurônios do *reservoir* e para e ajustar a camada de *readout*, correspondente aos pesos que conectam o *reservoir* à camada de saída.

O conjunto de validação cruzada é utilizado como critério de parada do treinamento da RNA. Uma das motivações para utilização deste recurso é evitar tanto o superajustamento da rede neural quanto o risco da mesma ficar presa à mínimos locais.

Por sua vez, o conjunto de testes é utilizado para calcular a taxa de classificação da rede neural. Neste conjunto, não se devem alterar os estados dos neurônios do *reservoir* e nem a matriz de pesos  $W_{res}$ .

Os estados dos neurônios do *reservoir* são inicializados com o valor zero e armazenados na matriz  $M_{est}$ . Para que os valores finais obtidos pela rede neural não sejam influenciados pela inicialização dos estados, antes do procedimento de treinamento são destinados alguns ciclos para o “aquecimento” do *reservoir*, processo conhecido como *warm up*, que tem por finalidade realizar atualizações nos estados dos neurônios do RC.

Após o *warm up*, inicia-se o procedimento de treinamento do RC com a atualização dos estados do *reservoir*, tendo como base o conjunto de treinamento passado como parâmetro.

Durante o treinamento, apenas a matriz  $W_{out}$  é atualizada enquanto as matrizes  $W_{in}$  e  $W_{res}$ , geradas durante a elaboração do RC, não devem ser ajustadas. A matriz  $W_{out}$ , por se tratar de uma matriz não quadrada e seu determinante se aproximar de zero, é obtida por meio do cálculo da pseudo-inversa de Monre-Penrose.

Após cada ciclo de treinamento, é iniciado um ciclo de validação cruzada. A alternância da execução destes ciclos perdura até que algum dos critérios de parada seja atingido e o treinamento seja encerrado. Durante a validação cruzada, a matriz de pesos  $W_{out}$  deve continuar a ser atualizada.

A fase de testes finaliza o treino do RC. Nesta fase, nem as matrizes de pesos, tampouco a matriz de estados devem ser alteradas. O conjunto dedicado aos testes é apresentado ao algoritmo que calcula os erros de predição para previsão ou classificação, bem como a taxa de acerto em caso de classificação [3].

### 3.3.3 Experimentos

O código da aplicação utilizado foi desenvolvido por Lara Coutinho [3] e aprimorado por outros estudantes de mestrado em Engenharia da Computação da Universidade de Pernambuco. O pseudocódigo, proposto por Lara, que contempla o comportamento do *reservoir* se encontra na seção de anexos deste trabalho.

Foram realizados 40 experimentos e os resultados de saída obtidos foram agrupados em duas categorias: diabéticos (1) e não diabéticos (0). Foi definido um limiar de valor 0,5. As saídas cujo valor estivesse contido no intervalo entre 0 e 0,5 (mais próximos de zero) foram classificadas como 0. Já as saídas contidas no intervalo acima de 0,5 e menor ou igual a 1, foram classificadas como 1. Após uma comparação com a coluna de valores desejados, a taxa de acerto foi calculada. Os resultados individuais de cada experimento se encontram no Apêndice A, juntamente aos resultados individuais da MLP.

Para a realização dos experimentos, o RC foi parametrizado conforme mostra a Tabela 5.

**Tabela 5.** Parametrização do RC

(Fonte: elaboração própria)

Parâmetro	Valor
Quantidade de neurônios na camada de entrada	8
Quantidade de neurônios no RC	20
Quantidade de neurônios na camada de saída	1
Função de ativação nos neurônios da camada de saída	Linear
Função de ativação nos neurônios do RC	Sigmóide logística
Quantidade de ciclos de <i>warm up</i>	10
Conectividade do RC	30%



## 3.4 Testes estatísticos

Para validar os resultados obtidos, foi realizada uma sequência de testes estatísticos. Neste trabalho, foi utilizado o software matemático R, por meio da IDE *Rstudio*, para análise e execução dos testes considerados: t-Student, Shapiro-Wilk, Teste F e .Soma dos postos de Wilcoxon. Por padrão, o R já contém as implementações dos testes citados e utiliza um nível de significância ( $\alpha$ ) previamente definido no valor de 0,05.

A análise dos resultados em qualquer um dos testes citados exige que sejam formuladas hipóteses nulas e alternativas. A análise é feita mediante a verificação comparativa do *p-value* obtido na realização de um teste com o nível de significância ( $\alpha$ ).

Esta seção descreve brevemente cada um dos testes analisados.

### 3.4.1 Teste t-Student

Faz a análise a partir das médias dos valores de duas amostras independentes. Para se realizar este teste, é necessário validar se as amostras são normalmente distribuídas e também se suas variâncias são estatisticamente iguais. Isto pode ser avaliado com a aplicação do teste Shapiro-Wilk e do teste F, respectivamente.

O objetivo deste teste é avaliar se as médias das amostras são estatisticamente iguais ou se há diferença entre elas. Em caso de diferença, a amostra que apresenta a maior média de acerto é considerada a de melhor desempenho. A Tabela 6 apresenta as hipóteses nula e alternativa deste teste.

**Tabela 6.** Hipóteses do teste t-Student

[Fonte: elaboração própria]

Hipótese	Descrição
Hipótese nula	As médias das amostras analisadas são estatisticamente iguais
Hipótese alternativa	As médias das amostras analisadas não são estatisticamente iguais

Caso o *p-value* seja menor que o nível de significância, a hipótese nula é rejeitada e as médias das amostras são consideradas estatisticamente diferentes. Neste trabalho, considerando

as redes neurais RC e MLP, caso a hipótese nula seja rejeitada, a topologia com maior média de taxa de acerto é considerada a abordagem com melhor desempenho.

### 3.4.2 Teste Shapiro-Wilk

O teste de Shapiro-Wilk é usado para verificar se as amostras são normalmente distribuídas. A Tabela 7 mostra as hipóteses nula e alternativa consideradas para este teste.

**Tabela 7.** Hipóteses do teste Shapiro-Wilk

(Fonte: elaboração própria)

Hipótese	Descrição
Hipótese nula	A amostra é normalmente distribuída
Hipótese alternativa	A amostra não é normalmente distribuída

Caso o *p-value* obtido seja menor que o nível de significância, a hipótese nula é rejeitada e a amostra não é considerada normalmente distribuída.

### 3.4.3 Teste F

O objetivo do teste F ou teste é verificar se as variâncias das amostras utilizadas são estatisticamente iguais ou não. As hipóteses para este teste constam na Tabela 8.

**Tabela 8.** Hipóteses do teste F

(Fonte: elaboração própria)

Hipótese	Descrição
Hipótese nula	As amostras possuem variâncias estatisticamente iguais
Hipótese alternativa	As amostras não possuem variâncias estatisticamente iguais

Caso o *p-value* calculado seja menor que o nível de significância adotado, a hipótese nula é rejeitada e as amostras não possuem variâncias estatisticamente iguais.

### 3.4.4 Teste da Soma dos postos de Wilcoxon

Trata-se de um teste não-paramétrico que não exige que as amostras sejam consideradas normalmente distribuídas ou que possuam variâncias estatisticamente iguais.

Avalia a mediana, ao invés da média, para verificar se as amostras são estatisticamente iguais ou se há diferença significativa entre elas. As amostras devem ser distribuídas simetricamente para que seja aplicado corretamente. A Tabela 9 descreve as hipóteses consideradas neste teste.

**Tabela 9.** Hipóteses do teste da soma dos postos de Wilcoxon

(Fonte: elaboração própria)

Hipótese	Descrição
Hipótese nula	As medianas das amostras analisadas são estatisticamente iguais
Hipótese alternativa	As medianas das amostras analisadas não são estatisticamente iguais

Caso o *p-value* seja menor que o nível de significância, a hipótese nula é rejeitada e comprova-se que há diferença significativa entre as amostras.

# Capítulo 4

## Análise dos Resultados

Este capítulo apresenta os resultados obtidos após a aplicação da metodologia descrita no Capítulo 3. As comparações realizadas neste capítulo consideram os resultados presentes na literatura com a aplicação de MLP no trabalho de Ferrara e os resultados obtidos nesta pesquisa, mediante as análises das aplicações da MLP e do RC, comprovadas estatisticamente.

A Tabela 10 mostra a taxa média de acerto obtida após a realização de 40 experimentos com a MLP e a RC para classificação da base PID, utilizando as parametrizações descritas no capítulo anterior.

**Tabela 10.** Média dos resultados dos experimentos com as topologias MLP e RC

(Fonte: elaboração própria)

Hipótese	Descrição
MLP	85%
RC	80%

A partir dos registros dos resultados de cada experimento, realizaram-se os testes estatísticos. Primeiramente, foi aplicado o teste Shapiro-Wilk para verificar se as amostras eram consideradas normais. Para os resultados da MLP, o *p-value* foi igual a 0,0717, sendo maior que o nível de significância adotado (0,05) e portanto, atestando que os dados são normalmente distribuídos. A amostra do RC apresentou um *p-value* de 0,06741 e desta forma, também comprovou que os dados da amostra podem ser considerados normalmente distribuídos. As hipóteses consideradas para este teste se encontram na Tabela 11 (hipóteses para a topologia MLP) e na Tabela 12 (hipóteses para a topologia RC).

**Tabela 11.** Hipóteses consideradas no teste Shapiro-Wilk para a MLP  
(Fonte: elaboração própria)

Hipótese	Descrição
Hipótese nula	A amostra com taxas de acerto da MLP é normalmente distribuída (p-value > 0,05)
Hipótese alternativa	A amostra com taxas de acerto da MLP não é normalmente distribuída (p-value ≤ 0,05)

**Tabela 12.** Hipóteses consideradas no teste Shapiro-Wilk para o RC  
(Fonte: elaboração própria)

Hipótese	Descrição
Hipótese nula	A amostra com taxas de acerto do RC é normalmente distribuída (p-value > 0,05)
Hipótese alternativa	A amostra com taxas de acerto do RC não é normalmente distribuída (p-value ≤ 0,05)

O segundo teste estatístico aplicado foi o teste F, pois assim poderia se optar pela escolha de um teste paramétrico ou não paramétrico após verificar se a variância das amostras eram estatisticamente iguais ou não. Com um *p-value* igual a 0,2246, ficou estabelecido que as variâncias das amostras são estatisticamente iguais. A Tabela 13 descreve as hipóteses consideradas neste teste.

**Tabela 13.** Hipóteses consideradas no teste F para MLP e RC  
(Fonte: elaboração própria)

Hipótese	Descrição
Hipótese nula	As variâncias das amostras das taxas de resultados da MLP e do RC são estatisticamente iguais (p-value > 0,05)
Hipótese alternativa	As variâncias das amostras das taxas de resultados da MLP e do RC são estatisticamente iguais (p-value ≤ 0,05)

Com ambas as amostras sendo normalmente distribuídas e possuindo variâncias estatisticamente iguais, pôde-se utilizar o teste paramétrico t-Student. As hipóteses consideradas para o teste se encontram na Tabela 14.

**Tabela 14.** Hipóteses consideradas no teste t-Student  
(Fonte: elaboração própria)

Hipótese	Descrição
Hipótese nula	As médias das amostras da MLP e do RC são estatisticamente iguais ( $p\text{-value} > 0,05$ )
Hipótese alternativa	As médias das amostras da MLP e do RC não são estatisticamente iguais ( $p\text{-value} \leq 0,05$ )

O teste t apresentou um *p-value* muito menor que o nível de significância adotado. Isto implica dizer que há uma diferença significativa nas medidas das amostras. A maior média obtida pelo teste corresponde à topologia MLP que tem, comprovadamente, o melhor desempenho na classificação dos casos de diabetes, considerando os dados da PID.

O teste da Soma dos Postos de Wilcoxon ainda poderia ser executado para tornar a análise mais criteriosa, no entanto, como os dados não foram considerados distribuídos simetricamente, não se aplicou.

Por fim, notou-se que a topologia MLP utilizada manteve a mesma média de resultados encontrada na literatura, conforme mostrado na Tabela 15. No entanto, a variação dos resultados foi menor, o que significa que o tratamento realizado na base de dados diminuiu o impacto causado no desempenho quando aleatoriamente se seleciona uma faixa de registros onde anteriormente havia valores nulos.

**Tabela 15.** Comparação das médias dos testes estatísticos obtidos com a literatura  
(Fonte: elaboração própria)

Hipótese	Descrição
MLP	85%
MLP (Marcos Ferrara)	85%
RC	80%

Da mesma maneira que os resultados individuais dos experimentos feitos com a RNA, os testes estatísticos (scripts e execuções) também estão no apêndice deste trabalho.

# Capítulo 5

## Conclusões e Trabalhos Futuros

Tanto no Brasil quanto no restante do mundo, se vivencia uma epidemia de diabetes *mellitus* atualmente. O diabetes é um conjunto de distúrbios metabólicos que se não forem tratados adequadamente podem acarretar em complicações até mesmo fatais para os portadores da doença. Uma vez diagnosticado como diabético, o indivíduo não pode reverter este quadro clínico. Por conta disto, é essencial que se realize um diagnóstico preciso e pontual tanto para os pacientes em estado de risco (pré-diabéticos), que podem adotar medidas preventivas contra a enfermidade, quanto para os portadores da doença, que podem iniciar um tratamento adequado o mais breve possível.

A proposta deste trabalho consistiu em utilizar diferentes topologias de redes neurais (MLP e RC) de forma comparativa para averiguar qual delas ofereceria um melhor desempenho no diagnóstico (classificação) de pacientes diabéticos e não diabéticos, levando em consideração as informações contidas na clássica base de dados *Pima Indians Diabetes*. Para efeitos de comparação, também foram analisados os resultados presentes na literatura, neste caso, o trabalho de Marcos Ferrara, que trouxe uma proposta de pesquisa similar.

Por fim, pôde-se concluir que os resultados obtidos para a topologia MLP confirmaram o desempenho da rede já presente na literatura (média de taxa de acerto igual a cerca de 85%) enquanto os resultados utilizando o RC demonstraram um desempenho um pouco inferior (média de taxa de acerto igual a cerca de 80%). Isso pode ser justificado pelo fato do RC ser mais indicado para sistemas dinâmicos por sua capacidade de memorização e pela MLP ter melhor capacidade de aproximar funções não lineares. Apesar disto, observou-se que o número significativo de casos falsos positivos e falsos negativos contribuiu para a diminuição da taxa de acerto em ambas as abordagens. O tratamento de imputação simples realizado na base de dados também se mostrou útil para melhor reconhecimento dos padrões trabalhados.

Ao longo desta pesquisa, algumas dificuldades e possibilidades de melhorias foram identificadas e novos objetivos foram estabelecidos. Visando atingir estas metas, alguns trabalhos futuros podem ser destacados:

- Realização de revisão bibliográfica da literatura com o objetivo de investigar técnicas para o tratamento de dados ausentes (*missing data*) que melhor se adequem ao problema proposto;
- Obtenção ou elaboração de uma base de dados mais completa e com informações significativas para se realizar o tratamento dos falsos positivos e falsos negativos desta aplicação. Isto poderá trazer melhorias para as taxas de acerto dos classificadores e aumentar a confiabilidade do modelo de aprendizado de máquina utilizado;
- Criação de um modelo de previsão de taxas de glicemia, visando promover um diagnóstico precoce da doença. Outra proposta válida é o estudo e a predição, por meio de técnicas de aprendizado de máquina, da atuação dos anticorpos causadores de DM1 no organismo de um indivíduo;
- Integração dos modelos preditivos e de classificação com o objetivo de criar um sistema completo de diagnóstico de diabetes, capaz de prever as taxas glicêmicas de um indivíduo e classificar sua condição previamente.



---

# Referências

- [1] AGÊNCIA ESTADO, publicação de 06/04/2016. Disponível em <[https://www.em.com.br/app/noticia/nacional/2016/04/06/interna\\_nacional,750581/oms-diz-que-diabete-atinge-8-5-da-populacao-do-mundo-e-pede-acao-cont.shtml](https://www.em.com.br/app/noticia/nacional/2016/04/06/interna_nacional,750581/oms-diz-que-diabete-atinge-8-5-da-populacao-do-mundo-e-pede-acao-cont.shtml)>  
Último acesso em: 05/12/2017.
- [2] ARDUINO, F., Diabetes e suas complicações, 2.ed. Rio de Janeiro, Guanabara **Koogan**, 1973, p. 1-4
- [3] COUTINHO, L.D. Utilizando *Reservoir Computing* para realizar diagnóstico precoce de mal de Alzheimer. **Trabalho de Conclusão de Curso**. Pernambuco, Brasil, 2012.
- [4] DAS, A.K; SHAH, S. *History of Diabetes: From Ants to Analogs*.  
Último acesso em: 05/12/2017. Disponível em <[http://www.japi.org/special\\_issue\\_april\\_2011/01\\_Diabetic\\_History.pdf](http://www.japi.org/special_issue_april_2011/01_Diabetic_History.pdf)>
- [5] DIAS, A. J. R., ALBIERI, S. Uso de imputação em pesquisas domiciliares. VIII Encontro Nacional de Estudos Populacionais. **Anais...** Volume 1: Informação Demográfica, Fecundidade, Demográfica Histórica, p. 11:26, São Paulo: ABEP,1992.
- [6] FERRARA, MR. Utilização de Redes Neurais Artificiais para identificação do Diabetes *Mellitus*. **Trabalho de Conclusão de Curso**. Marília, Brasil, 2005.
- [7] GONÇALVES, R.M. ; COELHO, L.S. ; KRUEGER, C.P. ; HECK, B. Modelagem preditiva de linha de costa utilizando redes neurais artificiais. **Bol. Ciênc. Geod. vol.16 no.3**. Curitiba, Julho/Setembro. 2010
- [8] HAYKIN, S. Redes Neurais: princípios e práticas. Bookman, 2007.
- [9] HIMSWORTH, HP; LOND, M.R.C.P. *Its Differentiation into Insulin-sensitive and Insulin-insensitive types\**. **Diabetic Medicine**, 28:1440-1444.

- 
- [10] IBGE. Pesquisa Nacional de Saúde 2013: percepção do estado de saúde, estilos de vida e doenças crônicas. Brasil, grandes regiões e unidades de federação. Rio de Janeiro, IBGE, 2014, 180p.
- [11] INTERNATIONAL DIABETES FEDERATION. IDF Diabetes Atlas, 6ed. Brussels: International Diabetes Federation, 2014. Último acesso em 05/12/2017. Disponível em: <<http://www.idf.org/diabetesatlas>>
- [12] JAEGER, H.; *Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach*. **GMD Report 159, German National Research Center for Information Technology**, 2002 (48 pp.)
- [13] KNOWLER, W.C. ; PETTITT, D.J. ; SAAD, M.F ; BENNETT, P.H. (1990). *Diabetes mellitus in the pima indians: Incidence, risk factors and pathogenesis*. **Diabetes Metab. Rev.**, 6: 1-27.
- [14] MAASS, W.; *Liquid State Machines: Motivation, Theory, and Applications*. Janeiro, 2010.
- [15] MAJOR, R.H.; *Classic descriptions of disease, Vol. 1*. **Oxford, Blackwell Scientific Publications**, 1954, p. 236.
- [16] MARCOVECCHIO, E. - *Dizionario etimologico storico dei termini medici*. Firenze, Ed. Festina Lente, 1993.
- [17] MCCULLOCH, W.; PITTS, W. *A logical calculus of the ideas immanent in nervous activity*. **Bulletin of Mathematical Biophysics**, p. p115-133, 1943.
- [18] OLIVEIRA, J.E.P. ; VENCIO, S. Diretrizes da Sociedade Brasileira de Diabetes. São Paulo, **AC Farmacêutica**, 2015-2016.
- [19] *PIMA INDIANS DIABETES (PID)*. Último acesso em: 05/12/2017, Disponível em <<https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes>>
- [20] PORTAL BRASIL, 2015. Último acesso em: 05/12/2017. Disponível em <<http://www.brasil.gov.br/saude/2015/07/diabetes-atinge-9-milhoes-de-brasileiros>>

- 
- [21] ZANONA, Q.K.; Modelo do tipo Integra e Dispara de neurônios procurando estabelecer parâmetros para modelagem de integração hipocampo-cortical. **Trabalho de Conclusão de Curso**. Rio Grande do Sul, Porto Alegre, 2009.
- [22] SANTOS DE OLIVEIRA, M. S. dos. Uma Aplicação para Previsão de Recursos Hídricos Baseada em Redes Neurais e Enxames de Partículas. **Dissertação (Mestrado)**. Universidade de Pernambuco, 2012.
- [23] SOARES, E.S.M. *Multi-Layer Perceptron e Reservoir Computing* aplicadas em um processo de chuva x vazão. **Trabalho de Conclusão de Curso**. Pernambuco, Brasil, 2016.
- [24] SOCIEDADE BRASILEIRA DE DIABETES, 2017. Último acesso em: 05/12/2017. Disponível em: <http://www.diabetes.org.br/publico/diabetes/oque-e-diabetes>.
- [25] SOCIEDADE BRASILEIRA DE ENDOCRINOLOGIA E METABOLOGIA. Último acesso em: 05/12/2017. Disponível em: <https://www.endocrino.org.br/o-que-e-diabetes/>
- [26] SILVA, C.V.S. Utilização de Redes Neurais no auxílio da previsão de geração solar. **Trabalho de Conclusão de Curso**. Pernambuco, Brasil, 2016.
- [27] VALENÇA, M.J.S. Fundamentos das Redes Neurais: exemplos em JAVA. 2ed. Olinda, Livro rápido, 2009.
- [28] VERSTRAETEN, D. *Reservoir Computing: computation with dynamical systems*. Ph.D dissertation, 2009
- [29] VIEIRA, J.P.; PARREIRA, F.J.; SILVEIRA, S.R. Desenvolvimento de um Sistema para Auxílio ao Diagnóstico de Diabetes empregando Redes Neurais Artificiais (SADD). **Revista de Sistemas e Computação**, Salvador, v.6, n.1, p. 26-38. 2016.
- [30] WERBOS J., P. *Beyond Regression: New tools for prediction and analysis in behavioral sciences*. Tese (Doutorado). **Harvard University**, 1974.

# Apêndice A

## Tabela de resultados individuais

Experimento	Teste MLP	Teste RC
1	0,868794326	0,791666667
2	0,84751773	0,786458333
3	0,843971631	0,802083333
4	0,868794326	0,8125
5	0,861702128	0,807291667
6	0,84751773	0,822916667
7	0,85106383	0,817708333
8	0,84751773	0,822916667
9	0,858156028	0,791666667
10	0,85106383	0,796875
11	0,858156028	0,796875
12	0,836879433	0,791666667
13	0,843971631	0,807291667
14	0,833333333	0,802083333
15	0,84751773	0,786458333
16	0,868794326	0,8125
17	0,840425532	0,8125
18	0,858156028	0,802083333
19	0,840425532	0,796875
20	0,84751773	0,807291667
21	0,85106383	0,78125
22	0,879432624	0,791666667
23	0,840425532	0,807291667
24	0,85106383	0,796875
25	0,843971631	0,796875
26	0,858156028	0,802083333
27	0,854609929	0,833333333
28	0,85106383	0,796875
29	0,861702128	0,8125
30	0,854609929	0,833333333
31	0,872340426	0,807291667
32	0,843971631	0,802083333
33	0,85106383	0,791666667
34	0,843971631	0,8125
35	0,843971631	0,802083333
36	0,865248227	0,817708333
37	0,875886525	0,84375
38	0,843971631	0,8125
39	0,84751773	0,807291667
40	0,865248227	0,796875

## Apêndice B

### Script utilizado para análises estatísticas

```
dados<-read.csv(file.choose(),header=FALSE,sep = ";")
```

```
mlp<-c(dados[,1])
```

```
rc<-c(dados[,2])
```

```
shapiro.test(rc)
```

```
shapiro.test(mlp)
```

```
var.test(mlp,rc)
```

```
t.test(mlp,rc)
```

# Apêndice C

## Resultados das análises estatísticas

```
> shapiro.test(mlp)
```

```
shapiro-wilk normality test
```

```
data: mlp  
W = 0.94929, p-value = 0.0717
```

```
> shapiro.test(rc)
```

```
shapiro-wilk normality test
```

```
data: rc  
W = 0.94851, p-value = 0.06741
```

```
> var.test(mlp,rc)
```

```
F test to compare two variances
```

```
data: mlp and rc  
F = 0.67523, num df = 39, denom df = 39, p-value = 0.2246  
alternative hypothesis: true ratio of variances is not equal to 1  
95 percent confidence interval:  
 0.3571296 1.2766735  
sample estimates:  
ratio of variances  
 0.6752317
```

```
> t.test(mlp,rc)
```

```
welch Two Sample t-test
```

```
data: mlp and rc  
t = 17.26, df = 75.175, p-value < 2.2e-16  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 4.21793 5.31857  
sample estimates:  
mean of x mean of y  
 85.29650 80.52825
```

# Anexo A

## Pseudocódigo do *Reservoir Computing*

### Algoritmo 1: Pseudocódigo do RC

```
1 Definir a quantidade de neurônios na camada de entrada ;
2 Definir a quantidade de neurônios no reservóir ;
3 Definir a quantidade de neurônios na camada de saída ;
4 Gerar aleatoriamente os pesos da matriz  $W_{in}$  entre -1 e 1;
5 Gerar aleatoriamente os pesos da matriz  $W_{res}$  entre -1 e 1;
6 Normalizar os pesos de  $W_{res}$  para que o raio espectral da matriz fique menor ou
igual a 1;
7 while até o final da quantidade de ciclos de warm up do
8     atualiza os estados dos neurônios do RC;
9 end
10 while até que o critério de parada seja atingido do
11     for cada valor do conjunto de entrada do
12         atualiza os estados dos neurônios do RC;
13     end
14     Calcula a pseudo-inversa de Moore-Penrose para encontrar a matriz de pesos
que conecta o RC à camada de saída;
15     for cada valor do conjunto de validação cruzada do
16         atualiza os estados dos neurônios do RC;
17     end
18     Calcula os valores de saída do RC;
19     Calcula o EMQ;
20     Verifica se o critério de parada já foi atingido;
21 end
22 for cada valor do conjunto de testes do
23     atualiza os estados dos neurônios do RC;
24 end
25 Calcula os valores de saída do RC;
26 Calcula o EPMA // para o caso de previsão;
27 Calcula a taxa de acerto //para classificação;
```