

UNIVERSIDADE FEDERAL DO PARANÁ

JAQUELINE VALLE

CLASSIFICAÇÃO DE HIDROMETEOROS EM RADARES METEOROLÓGICOS
UTILIZANDO REDES NEURAS ARTIFICIAIS

CURITIBA

2021

JAQUELINE VALLE

CLASSIFICAÇÃO DE HIDROMETEOROS EM RADARES METEOROLÓGICOS
UTILIZANDO REDES NEURAS ARTIFICIAIS

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Métodos Numéricos em Engenharia, área de concentração em Programação Matemática, no curso de Pós Graduação em Métodos Numéricos em Engenharia, Setores de Tecnologia e Ciências Exatas, Universidade Federal do Paraná..

Orientador: Prof. Dr. Paulo Henrique Siqueira

CURITIBA

2021

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

V181c Valle, Jaqueline
Classificação de hidrometeoros em radares meteorológicos
utilizando redes neurais artificiais [recurso eletrônico] / Jaqueline
Valle – Curitiba, 2021.

Dissertação - Universidade Federal do Paraná, Setor de Ciências
Exatas, Programa de Pós-graduação em Métodos Numéricos em
Engenharia.

Orientador: Prof. Dr. Paulo Henrique Siqueira

1. Radar Meteorológico. 2. Redes Neurais (Computação). I.
Universidade Federal do Paraná. II. Siqueira, Paulo Henrique. III.
Título.

CDD: 621.3848

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585

ATA DE SESSÃO PÚBLICA DE DEFESA DE MESTRADO PARA A OBTENÇÃO DO GRAU DE MESTRE EM MÉTODOS NUMÉRICOS EM ENGENHARIA

No dia vinte de abril de dois mil e vinte e um às 14:00 horas, na sala de videoconferência, Google Meet, foram instaladas as atividades pertinentes ao rito de defesa de dissertação da mestranda **JAQUELINE VALLE**, intitulada: **CLASSIFICAÇÃO DE HIDROMETEOROS EM RADARES METEOROLÓGICOS UTILIZANDO REDES NEURAIS ARTIFICIAIS**, sob orientação do Prof. Dr. PAULO HENRIQUE SIQUEIRA. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: PAULO HENRIQUE SIQUEIRA (UNIVERSIDADE FEDERAL DO PARANÁ), LEONARDO CALVETTI (UNIVERSIDADE FEDERAL DE PELOTAS), LUIZ CARLOS MATIOLI (UNIVERSIDADE FEDERAL DO PARANÁ), SÉRGIO SCHEER (UNIVERSIDADE FEDERAL DO PARANÁ). A presidência iniciou os ritos definidos pelo Colegiado do Programa e, após exarados os pareceres dos membros do comitê examinador e da respectiva contra argumentação, ocorreu a leitura do parecer final da banca examinadora, que decidiu pela APROVAÇÃO. Este resultado deverá ser homologado pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais definidos pelo programa. A outorga de título de mestre está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, PAULO HENRIQUE SIQUEIRA, lavrei a presente ata, que vai assinada por mim e pelos demais membros da Comissão Examinadora.

CURITIBA, 20 de Abril de 2021.

Assinatura Eletrônica

20/04/2021 16:32:50.0

PAULO HENRIQUE SIQUEIRA

Presidente da Banca Examinadora

Assinatura Eletrônica

12/05/2021 10:48:36.0

LEONARDO CALVETTI

Avaliador Externo (UNIVERSIDADE FEDERAL DE PELOTAS)

Assinatura Eletrônica

20/04/2021 17:38:22.0

LUIZ CARLOS MATIOLI

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

20/04/2021 16:18:58.0

SÉRGIO SCHEER

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **JAQUELINE VALLE** intitulada: **CLASSIFICAÇÃO DE HIDROMETEOROS EM RADARES METEOROLÓGICOS UTILIZANDO REDES NEURAIS ARTIFICIAIS**, sob orientação do Prof. Dr. PAULO HENRIQUE SIQUEIRA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 20 de Abril de 2021.

Assinatura Eletrônica
20/04/2021 16:32:50.0
PAULO HENRIQUE SIQUEIRA
Presidente da Banca Examinadora

Assinatura Eletrônica
12/05/2021 10:48:36.0
LEONARDO CALVETTI
Avaliador Externo (UNIVERSIDADE FEDERAL DE PELOTAS)

Assinatura Eletrônica
20/04/2021 17:38:22.0
LUIZ CARLOS MATIOLI
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
20/04/2021 16:18:58.0
SÉRGIO SCHEER
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

AGRADECIMENTOS

Primeiramente a Deus.

Aos meus pais, Keli e Gilmar, por serem minha base, meu exemplo e minha inspiração. Por acreditarem em mim mesmo nos momentos mais difíceis e estarem sempre ao meu lado. Por me pegarem no colo no momento em que minhas pernas fraquejaram, por segurarem minha mão quando tive medo e por festejarem comigo cada pequena vitória. Obrigada por todo amor, apoio e paciência.

À minha irmã, Maria Rita, por ter me apoiado e me acompanhado nesta jornada, mesmo que sentando ao meu lado em silêncio para me incentivar nos momentos de exaustão.

Ao meu namorado, Gabriel, por escutar meus anseios e me encorajar, por ter estado ao meu lado nos bons e maus momentos. Obrigada por todo amor, acolhimento e compreensão.

Aos meus avós, Claudionor, Rita, Antônio, Isaura e Maria. Avós são como segundos pais, obrigada por todo carinho, zelo e amor dedicado a mim, por torcerem pelo meu sucesso e se preocuparem comigo.

À todos os meus familiares, por torcerem por mim, enviarem suas boas energias e compartilharem bons momentos.

Ao professor Paulo, por aceitar me orientar nesta caminhada, por todos os ensinamentos e direcionamentos.

Aos professores Sérgio, Luiz e Leonardo por aceitarem avaliar e contribuir com este trabalho.

À todos os professores do PPGMNE, o conhecimento é o bem mais precioso que podemos compartilhar com alguém, obrigada por compartilharem os vossos conhecimentos conosco.

Ao SIMEPAR pelos dados cedidos para a realização deste estudo.

Aos amigos conquistados ao longo desta caminhada, em especial Adriana, Wagner, Kally, Alyne e Lilian, vocês tornaram a caminhada muito mais leve e possível. Obrigada por segurarem minha mão nos maus momentos, por compartilharem as preocupações e principalmente pelas boas conversas e risadas.

Aos meus amigos de longa data, Franciara, Mariana, Matheus, Rafael e Maiara, por todos os momentos de descontração e os "encontros mensais" não tão mensais... Fran, você esteve comigo na maior parte do meu processo formativo, desde a educação

básica, fico muito feliz e grata por compartilhar mais este acontecimento com você.

Ao meu amigo, agora anjo, Filipe, obrigada pelo tempo que passamos juntos. Sei que está olhando por mim de onde está.

À todos minha imensa gratidão, esta trajetória foi intensa, uma montanha russa de altos e baixos, onde aprendi muito mais que conteúdos acadêmicos. Obrigada por embarcarem comigo.

RESUMO

As condições meteorológicas afetam diretamente diversos setores das atividades humanas, tornando a previsão da ocorrência de hidrometeoros um importante recurso. Uma das ferramentas de sensoriamento remoto que fornecem informações sobre as condições meteorológicas são os radares meteorológicos, porém somente os radares de dupla polarização fazem a classificação de hidrometeoros, de modo que devem ser usadas técnicas alternativas para fazer essa classificação nas áreas monitoradas por radares de polarização simples. O objetivo deste trabalho é aplicar técnicas de redes neurais artificiais para fazer a classificação de hidrometeoros a partir de dados de radares de polarização simples utilizando como base a classificação fornecida por radares de dupla polarização com área de cobertura comum. Foram aplicadas a rede Perceptron Multicamadas, MLP, a rede Long Short-Term Memory, LSTM, bem como a técnica de rede neural híbrida, que trata-se da junção de técnicas de rede auto-organizadas com rede de aprendizagem supervisionada. Como métrica para avaliar o resultado utilizou-se a acurácia, ou seja, a porcentagem de alvos que a rede classificou corretamente tendo como parâmetro a classificação fornecida pelos radares de dupla polarização utilizados. A segunda técnica gerou melhores resultados, sendo que a acurácia média da execução da rede híbrida no conjunto de teste foi de até 76,99%, enquanto a melhor acurácia obtida sem o agrupamento foi de 74,21%. Isto porque a rede k-means, utilizada como a rede auto-organizada da técnica híbrida, identificou as semelhanças entre os dados formando agrupamentos que potencializaram o desempenho das redes MLP e LSTM. A inclusão das variáveis polarimétricas como estímulos para as redes também melhorou seus resultados, gerando uma acurácia média de até 95,69% para o conjunto de teste. Esta inclusão, na execução da rede, só foi possível a partir da aproximação dessas variáveis por uma rede MLP, uma vez que os dados da região monitorada somente por radar de polarização simples não incluem tais valores.

Palavras-chaves: Classificação de hidrometeoros, Radar Meteorológico, Redes Neurais Artificiais, Rede Neural Híbrida.

ABSTRACT

Meteorological conditions directly affect several sectors of human activities, making the prediction of hydrometeors an important resource. One of the remote sensing tools that provide information on weather conditions is weather radar, but only dual polarization radars classify hydrometeors, so alternative techniques should be used to make this classification in areas monitored by polarization radars. simple. The objective of this work is to apply artificial neural network techniques to classify hydrometeors from simple polarization radar data using the classification provided by dual polarization radars with common coverage area as a basis. The Multilayer Perceptron network, MLP, the Long Short-Term Memory network, LSTM, as well as the hybrid neural network technique, which is the combination of self-organized network techniques with supervised learning network, were applied. As a metric to evaluate the result, accuracy was used, that is, the percentage of targets that the network correctly classified using as parameter the classification provided by the dual polarization radars used. The second technique generated better results, with the average accuracy of the hybrid network execution in the test set being up to 76.99%, while the best accuracy obtained without the cluster was 74.21%. This is because the k-means network, used as the self-organized network of the hybrid technique, identified the similarities between the data forming clusters that enhanced the performance of the MLP and LSTM networks. The inclusion of polarimetric variables as stimuli for the networks also improved their results, generating an average accuracy of up to 95.69 % for the test set. This inclusion, in the execution of the network, was only possible from the approximation of these variables by an MLP network, since the data of the region monitored only by simple polarization radar do not include such values.

Key-words: Hydrometeor classification, Weather Radar, Artificial Neural Networks, Hybrid Neural Network.

LISTA DE ILUSTRAÇÕES

FIGURA 1 – COMPONENTES DO RADAR	20
FIGURA 2 – POLARIZAÇÃO DE ONDAS ELETROMAGNÉTICAS	22
FIGURA 3 – DIFERENÇA NA DETECÇÃO DE ALVOS	23
FIGURA 4 – DISTRIBUIÇÃO DE GOTAS EM UMA REGIÃO CONTRIBUINTE	24
FIGURA 5 – VALORES DE REFLETIVIDADE PARA O RADAR DE CASCAVEL	25
FIGURA 6 – VARIAÇÃO DE Z_{DR} PARA DIFERENTES FORMATOS DE ALVOS	27
FIGURA 7 – VARIAÇÃO DE Z_{DR} PARA DIFERENTES TIPOS DE HIDROME- TEOROS	28
FIGURA 8 – VALORES DE Z_{DR} PARA O RADAR DE CASCAVEL	28
FIGURA 9 – VARIAÇÃO DE ρ_{HV} PARA DIFERENTES TIPOS DE ALVOS . . .	29
FIGURA 10 – VALORES DE ρ_{HV} PARA O RADAR DE CASCAVEL	30
FIGURA 11 – DIFERENÇA DE FASE CAUSADA POR HIDROMETEOROS . . .	31
FIGURA 12 – VALORES DE Φ_{DP} PARA O RADAR DE CASCAVEL	31
FIGURA 13 – VALORES DE K_{DP} PARA O RADAR DE CASCAVEL	32
FIGURA 14 – CLASSIFICAÇÃO DE HIDROMETEOROS PARA O RADAR DE CASCAVEL	37
FIGURA 15 – NEURÔNIO BIOLÓGICO	41
FIGURA 16 – VARIAÇÃO DO POTENCIAL DE AÇÃO DO NEURÔNIO	42
FIGURA 17 – MODELO DE UM NEURÔNIO ARTIFICIAL	43
FIGURA 18 – FUNÇÕES DE ATIVAÇÃO	46
FIGURA 19 – FUNCIONAMENTO DO ALGORITMO K-MEANS	52
FIGURA 20 – ARQUITETURA DA REDE MLP	54
FIGURA 21 – SINAIS PROCESSADOS AO LONGO DE UMA REDE MLP . . .	54
FIGURA 22 – REDE DE ELMAN E REDE DE JORDAN	60
FIGURA 23 – DESDOBRAMENTO DO LOOP NO TEMPO DE UM BLOCO DE MEMÓRIA	61
FIGURA 24 – BLOCO DE MEMÓRIA LSTM	62
FIGURA 25 – EXEMPLO DAS CONEXÕES DE UMA REDE COM CAMADA LSTM	63
FIGURA 26 – COBERTURA E ELEVAÇÕES DO RADAR DE TEIXEIRAS SOARES	66
FIGURA 27 – COBERTURA E ELEVAÇÕES DO RADAR DE CASCAVEL	66
FIGURA 28 – COBERTURA E ELEVAÇÕES DO RADAR DE LONTRAS	67
FIGURA 29 – MOSAICO DOS DADOS DOS RADARES CAS, TXS E LON	69
FIGURA 30 – CURVA MÉTODO ELBOW	71

FIGURA 31 – ERRO E ACURÁCIA PARA ARQUITETURA (100,100,100) DA REDE MLP	75
FIGURA 32 – QUANTIDADE DE DADOS POR GRUPO PARA 5 AGRUPAMENTOS	79
FIGURA 33 – QUANTIDADE DE DADOS POR GRUPO PARA 7 AGRUPAMENTOS	79
FIGURA 34 – QUANTIDADE DE DADOS POR GRUPO PARA 10 AGRUPAMENTOS	80
FIGURA 35 – DISTRIBUIÇÃO DOS DADOS DENTRO DOS 5 AGRUPAMENTOS EM RELAÇÃO A CLASSE DE HIDROMETEOROS QUE PERTENCEM	80

LISTA DE TABELAS

TABELA 1 – REVISÃO BIBLIOGRÁFICA: CLASSIFICAÇÃO DE HIDROMETEOROS	33
TABELA 2 – DATA E HORÁRIO DOS DADOS UTILIZADOS	67
TABELA 3 – RESULTADOS PARA DIFERENTES ARQUITETURAS DA REDE MLP PARA TODOS OS DADOS	75
TABELA 4 – RESULTADOS PARA DIFERENTES TAMANHOS DE LOTES DE TREINAMENTO DA REDE MLP	76
TABELA 5 – RESULTADOS PARA DIFERENTES TAXAS DE APRENDIZADO DA REDE MLP	77
TABELA 6 – RESULTADOS PARA DIFERENTES FUNÇÕES DE ATIVAÇÃO DA REDE MLP	77
TABELA 7 – RESULTADOS PARA DIFERENTES ARQUITETURAS DA REDE LSTM PARA TODOS OS DADOS	78
TABELA 8 – RESULTADOS DAS REDES MLP PARA 5 AGRUPAMENTOS COM K-MEANS++	81
TABELA 9 – RESULTADOS DAS REDES MLP PARA 5 AGRUPAMENTOS COM INICIALIZAÇÃO ALEATÓRIA DE CENTROIDES	81
TABELA 10 – DESEMPENHO MÉDIO DAS REDES MLP PARA 5 AGRUPAMENTOS	82
TABELA 11 – RESULTADOS DAS REDES MLP PARA 7 AGRUPAMENTOS COM K-MEANS++	82
TABELA 12 – DESEMPENHO MÉDIO DAS REDES MLP PARA 7 AGRUPAMENTOS	82
TABELA 13 – RESULTADOS DAS REDES MLP PARA 10 AGRUPAMENTOS COM K-MEANS++	83
TABELA 14 – DESEMPENHO MÉDIO DAS REDES MLP PARA 10 AGRUPAMENTOS	83
TABELA 15 – RESULTADOS DAS REDES LSTM PARA 5 AGRUPAMENTOS COM K-MEANS++	84
TABELA 16 – DESEMPENHO MÉDIO DAS REDES LSTM PARA 5 AGRUPAMENTOS	84
TABELA 17 – RESULTADOS DAS REDES LSTM PARA 7 AGRUPAMENTOS COM K-MEANS++	84
TABELA 18 – DESEMPENHO MÉDIO DAS REDES LSTM PARA 7 AGRUPAMENTOS	85

TABELA 19 – RESULTADOS DAS REDES LSTM PARA 10 AGRUPAMENTOS COM K-MEANS++	85
TABELA 20 – DESEMPENHO MÉDIO DAS REDES LSTM PARA 10 AGRUPA- MENTOS	85
TABELA 21 – DESEMPENHO DA REDE MLP PARA APROXIMAÇÃO DAS VA- RIÁVEIS POLARIMÉTRICAS	86

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVO	16
1.1.1	Objetivo geral	16
1.1.2	Objetivos específicos	16
1.2	MÉTODO	16
1.3	LIMITAÇÕES DO TRABALHO	17
1.4	ESTRUTURA DO TRABALHO	17
2	CONCEITOS METEOROLÓGICOS	18
2.1	RADAR METEOROLÓGICO	19
2.1.1	Funcionamento	19
2.1.2	Tipos de Radares	21
2.1.3	Polarização	22
2.2	VARIÁVEIS MEDIDAS	23
2.2.1	Refletividade (Z)	24
2.2.2	Refletividade Diferencial (Z_{DR})	26
2.2.3	Coeficiente de Correlação Copolar (ρ_{HV})	28
2.2.4	Diferencial de Fase de Propagação Φ_{DP}	30
2.2.5	Diferencial de Fase Específica (K_{DP})	32
2.3	CLASSIFICAÇÃO DE HIDROMETEOROS	33
3	REDES NEURAIS	39
3.1	NEURÔNIOS ARTIFICIAIS	40
3.1.1	Funções de Ativação	43
3.1.2	Arquitetura	45
3.2	TREINAMENTO E PROCESSO DE APRENDIZAGEM	47
3.2.1	Regras de Aprendizagem	48
3.3	REDE K-MEANS	49
3.4	REDE PERCEPTRON MULTICAMADAS (MLP)	53
3.5	REDE LONG SHORT-TERM MEMORY (LSTM)	59
4	MÉTODO	65
4.1	DADOS	65
4.1.1	Py-ART	68
4.1.2	NumPy	68
4.2	SELEÇÃO DOS DADOS PONTUAIS	69

		14
4.3	APLICAÇÃO DAS REDES NEURAS ARTIFICIAIS	70
4.3.1	TensorFlow e Keras	72
4.4	VARIÁVEIS UTILIZADAS	72
5	RESULTADOS	74
5.1	REDE PERCEPTRON MULTICAMADAS ÚNICA	75
5.2	REDE LSTM ÚNICA	77
5.3	REDE HÍBRIDA: K-MEANS + MLP	79
5.4	REDE HÍBRIDA: K-MEANS + LSTM	83
5.5	UTILIZANDO AS VARIÁVEIS POLARIMÉTRICAS	85
6	CONCLUSÃO	88
6.1	SUGESTÃO PARA TRABALHOS FUTUROS:	89
	REFERÊNCIAS	90

1 INTRODUÇÃO

As condições meteorológicas afetam diretamente diversos setores das atividades humanas, como agrícola, aviação, geração de energia, turismo e de prevenção de desastres naturais. Assim, um dos pontos importantes é a previsão da ocorrência de hidrometeoros, que são os meteoros formados por um conjunto de partículas de água, líquidas ou sólidas, como chuva, granizo, neve, nevoeiro, entre outros.

A classificação correta de hidrometeoros permite a análise microfísica de sistemas precipitantes e conseqüentemente melhora a estimativa da taxa de precipitação. A importância da estimativa de precipitação vem crescendo significativamente nos últimos anos, pois é a principal forçante do ciclo hidrológico. O resultado dessa estimativa pode ser utilizado nos alertas de temporais, enchentes e inundações, com aplicações diretas em diversos setores da sociedade (OLIVEIRA, 2014).

O reconhecimento de hidrometeoros potencialmente perigosos, tais como os granizos de tamanho grande, é útil na previsão imediata de tempestades severas. Estas previsões são particularmente úteis na região sul do Brasil, onde frequentemente ocorre o desenvolvimento de sistemas convectivos intensos (RAMIREZ, 2018).

Uma das ferramentas de sensoriamento remoto que fornecem informações sobre as condições meteorológicas são os radares meteorológicos, porém somente os radares de dupla polarização fazem a classificação de hidrometeoros. Desta forma, técnicas alternativas devem ser usadas para fazer essa classificação nas áreas monitoradas por radares de polarização simples.

Neste trabalho utilizou-se uma rede neural artificial, RNA, para fazer a classificação de hidrometeoro a partir dos dados gerados por radar de polarização simples. O treinamento da rede foi feito a partir de dados obtidos na região de interseção do monitoramento do radar de polarização simples com radares de dupla polarização.

Para realizar este estudo foram utilizados dados cedidos pelo Sistema Meteorológico do Paraná, SIMEPAR, de modo que o objetivo é fazer a classificação de hidrometeoros para a área monitorada pelo radar de Teixeira Soares, que é de polarização simples, através dos dados de radares de dupla polarização, o radar de Cascavel e de Lontras, que possuem área comum ao radar de Teixeiras Soares.

O foco deste trabalho é a classificação de hidrometeoros em: *Ground Clutter* (GC) - eco de terreno, *Sea Clutter* (SC) - eco do mar, *Biological* (BI) - dispersores biológicos, *Dry Snow* (DS) - neve seca, *Wet Snow* (WS) - neve molhada, *Ice Crystals* (IC) - cristais de gelo, *Graupel* (GR) - graupel, *Big Drops* (BD) - grandes gotas, *Rain* (RA) - chuva leve e moderada, *Heavy Rain* (HR) - chuva forte e *Hail* (HA) - granizo.

1.1 OBJETIVO

1.1.1 Objetivo geral

Aplicar técnicas de redes neurais artificiais para fazer a classificação de hidrometeoros a partir de dados de radares de polarização simples utilizando como base a classificação fornecida por radares de dupla polarização com área de cobertura comum.

1.1.2 Objetivos específicos

- Selecionar dados de radares meteorológicos de eventos com ocorrência de hidrometeoros abordados neste trabalho;
- Identificar pontos comuns entre os radares de dupla polarização e o radar de Teixeira Soares;
- Enumerar as variáveis significativas para a classificação de hidrometeoros;
- Verificar a influência e possível relação entre as variáveis selecionadas;
- Avaliar o desempenho das redes neurais artificiais para a classificação de hidrometeoros na região do radar de Teixeira Soares.

1.2 MÉTODO

Para classificação de hidrometeoros na região do radar de Teixeira Soares foi utilizada a técnica de Redes Neurais Artificiais. A proposta é utilizar uma rede neural híbrida, que trata-se da junção de técnicas de rede auto-organizadas com rede de aprendizagem supervisionada.

Nesta abordagem os dados foram inicialmente agrupados através da rede de k-means, para que a rede identifique as particularidades dos dados e agrupe-os de modo que os dados semelhantes estejam no mesmo grupo.

Na sequência, para cada grupo de dados formado é treinada uma rede de aprendizagem supervisionada, no caso deste trabalho foram aplicadas as rede Perceptron Multicamadas, MLP, e a rede Long Short-Term Memory, LSTM. Assim, cada vez que um dado novo for apresentado para rede, inicialmente será identificado a qual grupo o mesmo pertence e então aplicada a rede MLP ou LSTM correspondente.

Uma técnica semelhante foi aplicada por Coadou e Benabdeslem (2006), porém para a previsão de séries temporais, nesse caso os autores treinaram para cada grupo uma rede MLP. A rede MLP é uma rede de arquitetura *feedforward* com

múltiplas camadas, com aprendizagem supervisionada e algoritmo de aprendizagem usual *backpropagation*.

Também foi testada a rede LSMT por ser uma rede neural recorrente que tem como benefício o uso de uma das capacidades do cérebro humano, a memória. Assim a cada iteração do treinamento a rede carrega uma memória de curto e de longo prazo das iterações anteriores, o que pode facilitar sua convergência para a classificação correta.

1.3 LIMITAÇÕES DO TRABALHO

Este trabalho limita-se a utilizar apenas dados de três radares meteorológicos da região sul do Brasil e da estação de radiossondagem de Curitiba. Por isso, não há garantia de precisão na aproximação da temperatura na posição do alvo em questão.

Neste estudo foram aplicadas apenas técnicas de aprendizagem de máquina através de redes neurais artificiais.

1.4 ESTRUTURA DO TRABALHO

Este trabalho se inicia com a introdução, na qual são abordados a motivação e os objetivos do trabalho. O segundo capítulo apresenta uma revisão dos conceitos meteorológicos relevantes para o desenvolvimento deste trabalho, como informações sobre o radar meteorológico e as variáveis meteorológicas medidas pelo mesmo. No Capítulo 3 é abordada a teoria de redes neurais artificiais com foco na rede auto-organizável com regra de aprendizagem competitiva, k-means, a rede de arquitetura *feedforward* com aprendizado supervisionado, MLP, e por fim a rede recorrente LSTM. Os métodos utilizados para a classificação de hidrometeoros são descritos no Capítulo 4. Já o Capítulo 5 apresenta os resultados obtidos, com as comparações entre o desempenho de cada técnica. Por fim, no Capítulo 6 são apresentadas as conclusões do trabalho e sugestões para trabalhos futuros.

2 CONCEITOS METEOROLÓGICOS

A curiosidade natural nos impulsiona a buscar entender o funcionamento e as características do meio ambiente em que vivemos, entre eles as condições e transformações climáticas. Foi nesse contexto que por volta de 350 a.C Aristóteles escreveu o primeiro texto sobre meteorologia, onde descreveu o tempo e clima da época com base em suas observações em sua obra intitulada *Meteorologica* (PETTERSEN, 1941).

O primeiro registro do termo meteorologia está nessa obra de Aristóteles, que deriva da palavra grega *metéōros* (meteoros), uma vez que na época tudo que se observava na atmosfera, como as nuvens, chuva, vento, relâmpagos, etc, era considerado um meteoro. Atualmente o termo meteorologia se refere a ciência ambiental que estuda os processos físicos, químicos e dinâmicos da atmosfera, bem como sua relação com os sistemas litosfera, hidrosfera, criosfera e biosfera (YNOUE et al., 2017).

Desde os primeiros indícios de estudos na área que hoje é considerada meteorologia, que datam de aproximadamente 500 a.C, até que houvesse algum progresso significativo passaram-se aproximadamente quatorze séculos. De acordo com Pettersen (1941), a meteorologia é essencialmente física, matemática, mecânica e química aplicadas à atmosfera. E devido a isso, o progresso real da mesma só foi possível após o desenvolvimento dessas ciências.

Foi então que em 30 de abril de 1904 o engenheiro Christian Hülsmeier de Düsseldorf, registrava a patente número DE165546 intitulada "*Telemobiloskop*", que se refere a um objeto que transmite e recebe ondas Hertzianas (GALATI, 2016). O objeto em questão é a primeira versão de um radar, que mais tarde seria considerado por alguns como "A invenção que mudou o mundo"(BUDERI, 1998). O sistema desenvolvido pelo engenheiro tinha como objetivo detectar objetos metálicos (GALATI, 2016).

Durante a Segunda Guerra Mundial, mais precisamente no ano de 1940 o aprimoramento do radar foi impulsionado pela necessidade de detectar a aproximação de aviões, navios e submarinos inimigos. Um dos aprimoramentos se deu através do magnetron, uma válvula termiônica que gera micro-ondas eletromagnéticas (SANTOS, 2011). O magnetron permitiu a redução significativa no tamanho dos radares, facilitando sua locomoção e permitindo que os mesmos fossem instalados em aeronaves (FABRY, 2015).

E é nesse contexto que a história dos radares se encontra com a da meteorologia, os radares que utilizavam magnétron detectaram grandes manchas de eco, que foram então identificadas como resultantes de precipitação. Como na época haviam

muitos meteorologistas a serviço do exército, uma vez que a previsão do tempo era importante para as estratégias de guerra, essas imagens rapidamente chegaram a eles, e assim se deu o início dos estudos para o uso de radares na meteorologia.

2.1 RADAR METEOROLÓGICO

Os radares meteorológicos se tornaram referência para o monitoramento e previsão a curto prazo (CALVETTI; BENETI; MEDEIROS, 2017), unindo o termômetro, a radiossonda e imagens de satélite, tornou-se também um instrumento amplamente utilizado para pesquisas com foco nos fenômenos meteorológicos, sendo uma das melhores ferramentas para detectar ocorrências meteorológicas severas como fortes tempestades, granizo e tornados (FABRY, 2015).

O termo radar é um acrônimo para *Radio Detection and Ranging*, e se refere a um mecanismo de sensoriamento, assim como o tato, paladar, audição e visão, que utilizamos para obter informações dos objetos que nos cercam. O tato e paladar são exemplos de dispositivos de sensoriamento *in situ*, onde há o contato direto entre o sensor e o objeto. A audição e a visão, assim como o radar, são instrumentos de sensoriamento remoto, não entram em contato direto com o objeto, a informação viaja entre o objeto e o sensor, através de ondas acústicas ou eletromagnéticas como luz, calor e rádio (FABRY, 2015).

No caso do radar são utilizadas ondas eletromagnéticas, as microondas emitidas pelo radar se propagam na velocidade da luz até encontrarem um objeto, o alvo, o qual reflete parte dessa energia em direção ao radar. O radar, após receber esse eco extrai o máximo de informações possíveis do mesmo, como a potência e tempo de propagação, processando esses dados para gerar informações meteorológicas.

Os alvos nos quais a meteorologia está interessada são os hidrometeoros, porém eles não são os únicos que o radar detecta, por vezes acontece de o mesmo detectar insetos, aviões, entre outros (CALVETTI; BENETI; MEDEIROS, 2017). Por isso a necessidade de obter informações suficientes para identificar corretamente o alvo.

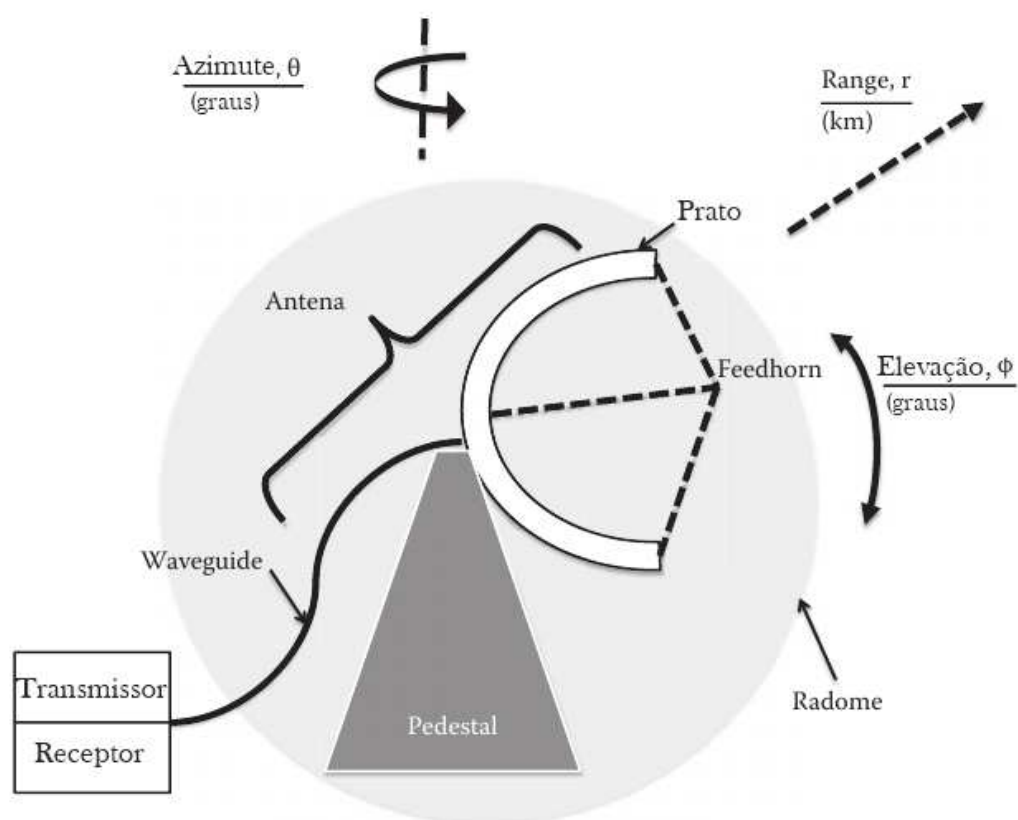
2.1.1 Funcionamento

Como citado anteriormente, o radar identifica a presença de hidrometeoros emitindo ondas eletromagnéticas. Sendo que o transmissor gera a radiação eletromagnética, esse pulso é transmitido através do *waveguide* até a antena, que consiste geralmente de um prato parabólico, girado a partir de um pedestal. Assim o sinal chega até o *feedhorn*, um dispositivo que transmite o pulso como uma onda de rádio até o centro do prato, onde a mesma é refletida formando o feixe de radar que se propaga

através da atmosfera até o alvo (HONG; GOURLEY, 2014).

Ao encontrar o alvo a energia refletida em direção ao radar é chamada de retroespalhamento, que chega até a antena do radar onde é concentrada no *feedhorn*, o qual converte essa energia novamente em voltagem, direcionando o sinal ao receptor que amplifica o mesmo e o envia aos processadores (HONG; GOURLEY, 2014). Os processadores extraem os dados desse sinal, e então os dados são enviados aos hardware e software de geração de produtos meteorológicos (FABRY, 2015). Tais componentes do radar podem ser observados na FIGURA 1.

FIGURA 1 – COMPONENTES DO RADAR



FONTE: Adaptado de Hong e Gourley (2014)

Cada dado fornecido pelo radar está associado a um ponto dado em coordenadas esféricas (r, θ, ϕ) que indica sua posição em relação ao radar, sendo que r representa o range, ou seja, a distância do ponto ao radar, θ o ângulo de azimute e ϕ o ângulo de elevação da antena (FIGURA 1).

Os radares abordados nesse trabalho fazem a varredura volumétrica de uma área fixando inicialmente um ângulo de elevação ϕ , que é medido em relação a posição vertical. Uma vez fixado o ângulo de elevação a antena direciona o sinal ao longo de

cada ângulo de azimute θ , o ângulo azimute é medido tendo como referência o norte geográfico e representa o ângulo de rotação da antena em torno do eixo do pedestal, fazendo assim uma varredura de 360° a cada elevação.

Para determinar a distância r da tripla, utiliza-se o tempo de propagação do sinal, que é um dos dados que o radar extrai do eco. De fato a velocidade média de um objeto se relaciona com o espaço percorrido (Δs) e o tempo decorrido (Δt) pela equação $v = \frac{\Delta s}{\Delta t}$, considerando que o sinal do radar se propaga na velocidade da luz ($v = c$) e que o deslocamento do mesmo, partindo do radar até o alvo e depois sendo refletido ao radar, é de $2r$. Temos então que a distância r de um ponto ao radar é dada pela equação (2.1) (RAUBER; NESBITT, 2018).

$$r = \frac{c\Delta t}{2} \quad (2.1)$$

2.1.2 Tipos de Radares

Dois dos fatores que caracterizam um radar são o comprimento de onda (λ) e a frequência de rádio (f) com os quais o radar opera. Esses atributos podem variar de um radar para outro, porém estão relacionados com a velocidade de propagação pela relação $c = f\lambda$, onde c é a velocidade da luz, logo é a mesma para todos os radares (HONG; GOURLEY, 2014).

De acordo com Rauber e Nesbitt (2018), em geral o foco dos meteorologistas é sobre o comprimento de onda, uma vez que o mesmo pode ser comparado com o tamanho dos hidrometeoros, além disso o comprimento da onda influencia diretamente na atenuação do sinal. Alvos com seção transversal, na orientação de polarização da onda, menores que $\frac{\lambda}{16}$ apresentam assinaturas de radar previsíveis para pingos de chuva de tamanhos diferentes (HONG; GOURLEY, 2014).

Assim, radares com comprimento de ondas menores só conseguem diferenciar tamanhos menores de alvos. Porém quanto maior o comprimento da onda (λ), maior deve ser o diâmetro (d) da antena do radar para que o ângulo do feixe (α) tenha um tamanho viável, uma vez que a relação entre eles é dada por $\alpha = \frac{73\lambda}{d}$ (HONG; GOURLEY, 2014). Portanto, apesar de radares com comprimentos maiores de onda oferecerem vantagens na detecção de pingos de chuvas com tamanhos diferentes, possui desvantagens ao exigir um tamanho de antena maior, consumir mais energia e em geral serem mais caros (RAUBER; NESBITT, 2018).

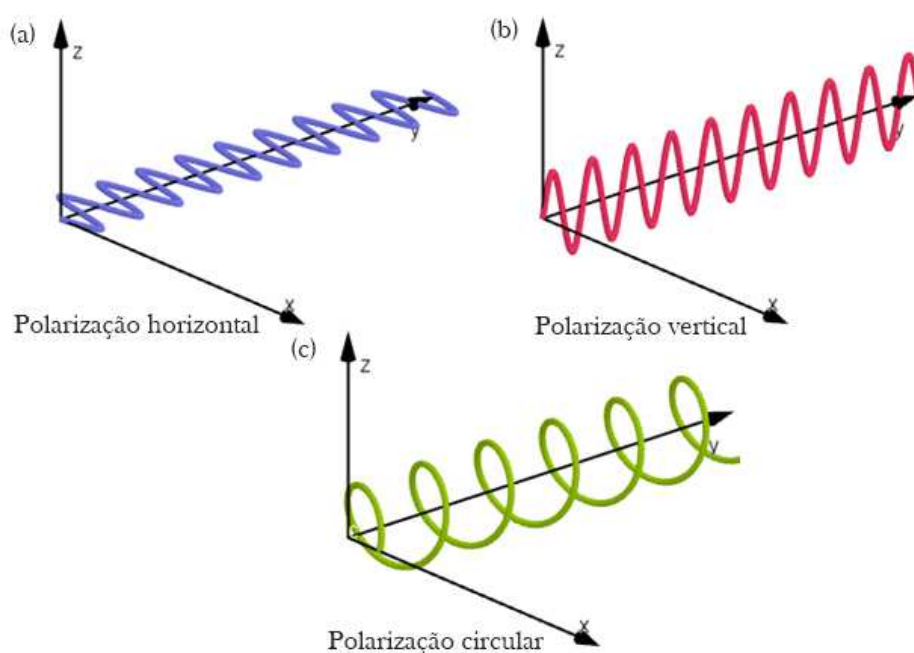
Desta forma, aspectos como a missão do radar, o local onde o radar será montado e a energia disponível para seu funcionamento são determinantes na escolha do comprimento de onda a ser utilizado (RAUBER; NESBITT, 2018). Os radares empregados para coletar os dados utilizados nesse trabalho são radares de banda S

que tem uma frequência f de 2 a 4 GHz, um comprimento de onda λ de 7,5 a 15 cm e sua principal missão é estimar precipitação de intensidade moderada e pesada.

2.1.3 Polarização

Segundo Fabry (2015) a maioria das ondas de rádio e microondas emitidas são polarizadas, o que também é o caso das ondas eletromagnéticas emitidas pelos radares. O plano de polarização de uma onda eletromagnética é o plano no qual o campo elétrico oscila, sendo que a orientação da onda é dada pela orientação do vetor do campo elétrico (RAUBER; NESBITT, 2018). Na FIGURA 2 é possível observar três modelos de propagação relevantes para o estudo de radares meteorológicos.

FIGURA 2 – POLARIZAÇÃO DE ONDAS ELETROMAGNÉTICAS



FONTE: Adaptado de R. M. Rauber e Nesbitt (2018)

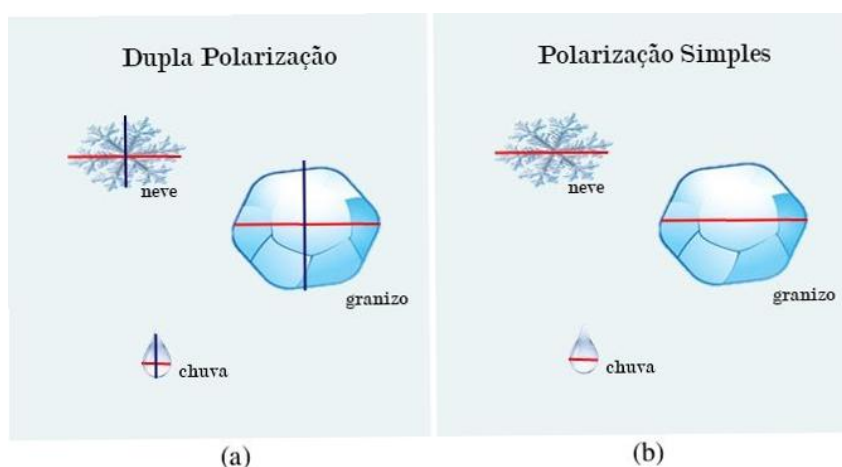
LEGENDA: Propagação do vetor campo elétrico em uma onda (a) polarizada horizontalmente, (b) polarizada verticalmente e (c) polarizada circularmente à direita

Os radares convencionais de polarização simples são os que possuem apenas uma polarização, em sua maioria, emitem ondas polarizadas horizontalmente (RAUBER; NESBITT, 2018). Já os radares de dupla polarização transmitem ondas em duas polarizações, que em geral são na polarização vertical e polarização horizontal, essa emissão pode acontecer tanto alternadamente como simultaneamente (FABRY, 2015).

As ondas polarizadas horizontalmente permitem estimar as dimensões horizontais dos alvos, assim como as ondas polarizadas verticalmente estão relacionadas às dimensões verticais. Um dos principais benefícios do radar de dupla polarização é poder comparar os dados de ambas as polarizações e assim gerar informações sobre a forma e o tamanho do alvo (FABRY, 2015).

Quando o foco é a classificação de hidrometeoros os radares de dupla polarização são ainda mais relevantes, uma vez que os alvos não são esféricos nem simétricos, ou seja, sem a comparação dos dados nas dimensões horizontais e verticais, não é possível classificar o alvo. Por exemplo, as gotas de chuva têm formato oblato, os flocos de neve possuem formatos com diversas orientações, os dispersores biológicos como insetos e pássaros são ainda menos simétricos em relação ao eixo de propagação da onda polarizada horizontal. Assim esses alvos diferentes possuem assinaturas diferentes somente quando se trata de radares de dupla polarização (FABRY, 2015). Na FIGURA 3 é possível observar a diferença na mensuração de partículas por um radar de polarização simples e um radar de dupla polarização.

FIGURA 3 – DIFERENÇA NA DETECÇÃO DE ALVOS



FONTE: Adaptado de METED (2017)

LEGENDA: Diferença na detecção de alvos por radares de (a) dupla polarização, vertical e horizontal, (b) polarização única horizontal.

2.2 VARIÁVEIS MEDIDAS

O dado fornecido pelo radar é gerado a partir da potência da energia do sinal de retorno, que é convertida em refletividade (CALVETTI; BENETI; MEDEIROS, 2017). A refletividade é um parâmetro básico fornecido pelos diversos tipos de radares.

Os radares de dupla polarização fornecem também como variáveis de medida a refletividade diferencial, o coeficiente de correlação copolar, diferencial de fase de propagação e o diferencial de fase específica, essas variáveis são chamadas de variáveis polarimétricas.

2.2.1 Refletividade (Z)

A refletividade é a variável, ou parâmetro, mais importante gerada pelo radar meteorológico, sendo uma medida amplamente utilizada pelos meteorologistas (RAUBER; NESBITT, 2018). A refletividade mede o tamanho das gotas em determinado volume a partir da quantidade de gotas no mesmo, sendo assim ela pode ser determinada a partir da equação (2.2):

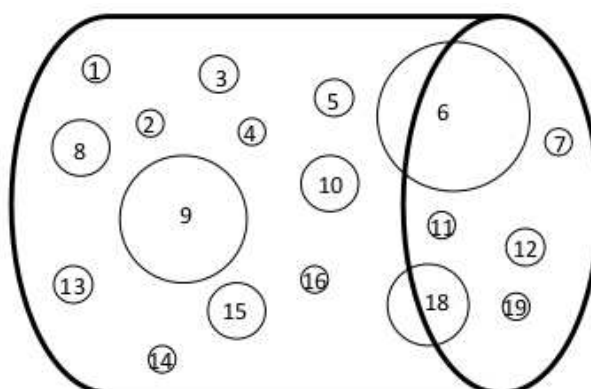
$$z = \frac{\sum_j D_j^6}{V_c} \quad (2.2)$$

onde D_j representa o diâmetro da j -ésima gota dentro da região contribuinte, V_c é o volume da região contribuinte e o somatório é sobre o número total de gotas dentro da região (MULLER, 2013).

Por exemplo, considerando uma região contribuinte de 1 metro cúbico do feixe de radar que contém 19 gotas de diversos tamanhos distribuídas dentro de seu volume, como na FIGURA 4. Nesse caso a refletividade relativa a essa seção seria dada pela equação (2.3) (MULLER, 2013).

$$z = \frac{\sum_{j=1}^{19} D_j^6}{1} = D_1^6 + D_2^6 + D_3^6 + \dots + D_{18}^6 + D_{19}^6 \quad (2.3)$$

FIGURA 4 – DISTRIBUIÇÃO DE GOTAS EM UMA REGIÃO CONTRIBUINTE



FONTE: Muller (2013)

LEGENDA: Exemplo de distribuição de gotas de diversos tamanhos dentro de uma região contribuinte de 1 metro cúbico.

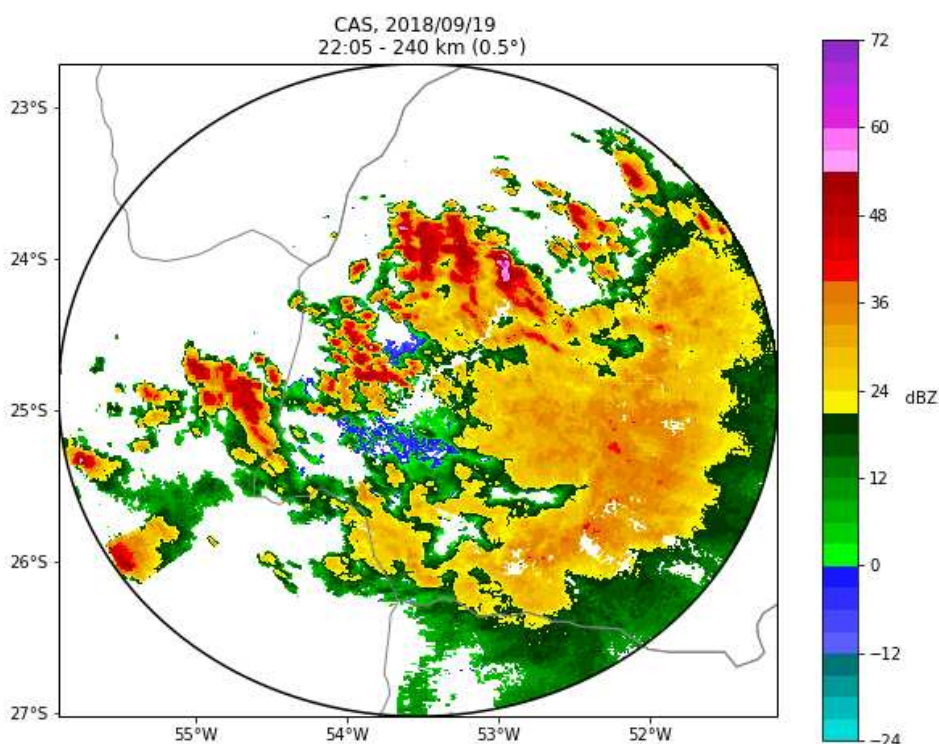
A unidade de medida original da refletividade é dada por mm^6m^{-3} , uma vez que a dimensão das gotas é dada em milímetros e o volume da região contribuinte em metros cúbicos. Utilizando essa unidade de medida a refletividade apresenta uma grande variação na ordem de magnitude (RAUBER; NESBITT, 2018), assumindo valores desde aproximadamente $0,001 mm^6m^{-3}$, para regiões com névoa úmida, até $36.000.000 mm^6m^{-3}$ para regiões com granizo grande (QUEIROZ, 2008).

Para amenizar essa variação costuma-se utilizar a refletividade em escala logarítmica, de modo que seu valor é dado pela equação (2.4):

$$Z = 10 \log \left(\frac{z}{1mm^6m^{-3}} \right) \quad (2.4)$$

onde z é a refletividade medida em mm^6m^{-3} e Z é a refletividade na escala logarítmica com unidade de medida dBZ (FABRY, 2015). Dessa forma os valores de refletividade costumam variar entre -24 dBZ e 72 dBZ como mostra a FIGURA 5.

FIGURA 5 – VALORES DE REFLETIVIDADE PARA O RADAR DE CASCAVEL



FONTE: a autora

LEGENDA: Valores Z para a região do radar de Cascavel no dia 19/09/2018 às 22:05

Na prática a distribuição do tamanho das gotas dentro da região de contribuição não é conhecida, então o software do radar calcula o fator refletividade a partir da

equação da potência retornada p_r (MULLER, 2013) que é dada, para cada radar, pela equação (2.5):

$$p_r = \frac{\pi^3 p_t g^2 \theta_0 \varphi_0 l |K|^2 z}{1024 \ln(2) \lambda^2 r^2} \quad (2.5)$$

onde p_t é a energia transmitida pelo radar, g representa o ganho da antena, θ_0 a largura do feixe horizontal, φ_0 a largura do feixe vertical, l a largura do pulso, $|K|^2$ é a constante dielétrica para hidrometeoros (geralmente 0,93 para água líquida e 0,197 para gelo), z o fator de refletividade do radar, λ o comprimento de onda do pulso do radar e r a distância do alvo (RINEHART, 2004).

Dessa forma o valor da refletividade z é calculado pela equação (2.6):

$$z = \frac{p_r 1024 \ln(2) \lambda^2 r^2}{\pi^3 p_t g^2 \theta_0 \varphi_0 l |K|^2} \quad (2.6)$$

Para simplificar a escrita da equação (2.6) considera-se uma única constante para representar as constantes do radar, definindo assim c_r através da equação (2.7):

$$c_r = \frac{1024 \ln(2) \lambda^2}{\pi^3 p_t g^2 \theta_0 \varphi_0 l |K|^2} \quad (2.7)$$

onde o valor da constante dielétrica $|K|^2$ é considerado fixo como 0,93 (OLIVEIRA, 2014).

Portanto a refletividade z , medida em $mm^6 m^{-3}$ é dada pela equação (2.8).

$$z = p_r c_r r^2 \quad (2.8)$$

2.2.2 Refletividade Diferencial (Z_{DR})

A refletividade diferencial expressa a relação entre a refletividade referente ao sinal polarizado horizontalmente, denotada por Z_H e a refletividade resultante do sinal polarizado verticalmente, denotada por Z_V (CALVETTI; BENETI; MEDEIROS, 2017). Mais especificamente a refletividade diferencial é a razão, ou proporção, entre Z_H e Z_V , sendo obtida pela equação (2.9):

$$Z_{DR} = 10 \log \left(\frac{Z_H}{Z_V} \right) = dBZ_H - dBZ_V \quad (2.9)$$

onde dBZ_H e dBZ_V são as refletividades horizontal e vertical, respectivamente, em escala logarítmica (KNIGHT; VIVEKANANDAN; LASHER-TRAPP, 2002).

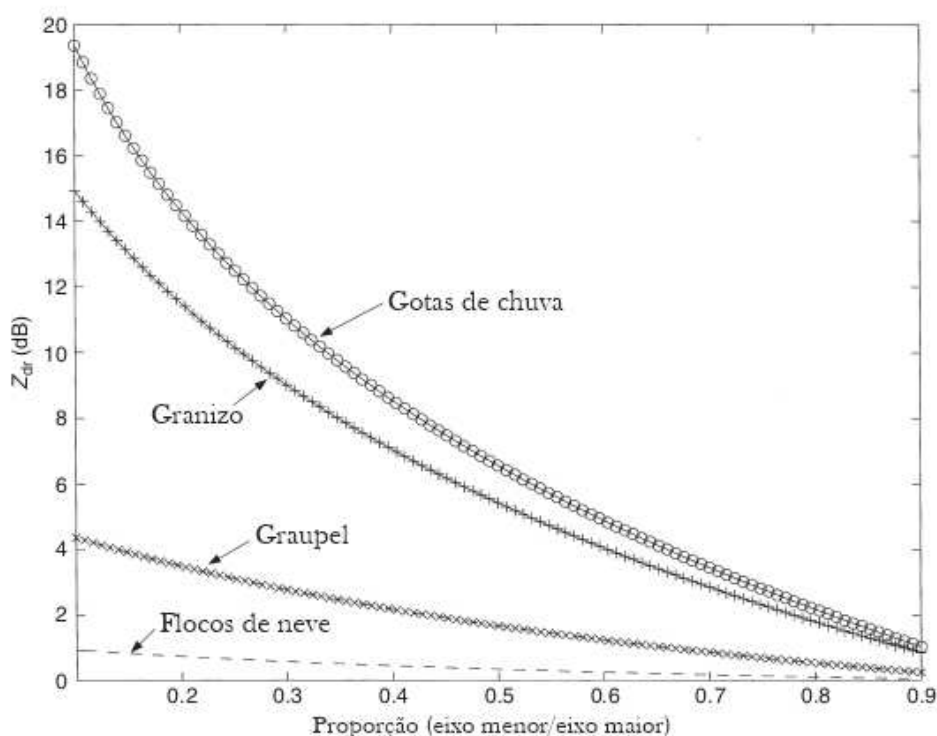
Através da refletividade diferencial é possível quantificar a excentricidade do hidrometeoro (CALVETTI; BENETI; MEDEIROS, 2017), de modo que se $Z_{DR} = 0$ então $Z_H = Z_V$, o que significa que o alvo é esférico. O formato esférico de partículas costuma estar associado a gotas pequenas ou pequenas partículas de granizo, logo

valores de Z_{DR} próximos de zero indicam pequenas gotas ou granizo (RYZHKOV et al., 2005).

Grandes gotas, por sua vez, estão associadas à valores maiores de Z_{DR} . Já pedras maiores de granizo costumam gerar um valor negativo de Z_{DR} , isso porque o granizo costuma cair com sua maior dimensão na direção vertical. Portanto em tempestades com alta refletividade, valores negativos ou próximos de zero para Z_{DR} indicam a presença de granizo. No caso da neve, os valores de Z_{DR} variam com menos intensidade, assim, apenas cristais de gelo alongados geram valores altos para Z_{DR} . Por fim, o Z_{DR} permite diferenciar ainda insetos de garoa, uma vez que a refletividade de ambos são semelhantes, porém o valor de Z_{DR} para os insetos é alto, por possuírem um formato alongado (FABRY, 2015).

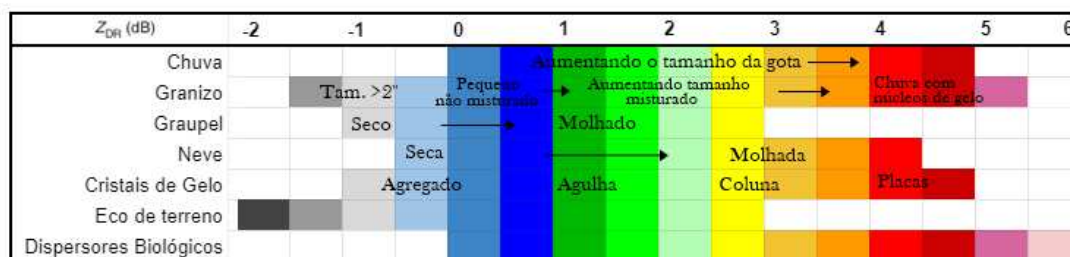
A FIGURA 6 apresenta a variação dos valores de Z_{DR} para cada tipo de hidrometeoro de acordo com a variação do seu formato. Na FIGURA 7 por sua vez, é possível observar a variação dos valores da refletividade diferencial para algumas características de cada hidrometeoro. Já a FIGURA 8 apresenta os valores de Z_{DR} para a região do radar de Cascavel no dia 19 de setembro de 2018 às 22:05, na mesma imagem estão indicados os pontos com a ocorrência de determinados hidrometeoros.

FIGURA 6 – VARIÇÃO DE Z_{DR} PARA DIFERENTES FORMATOS DE ALVOS



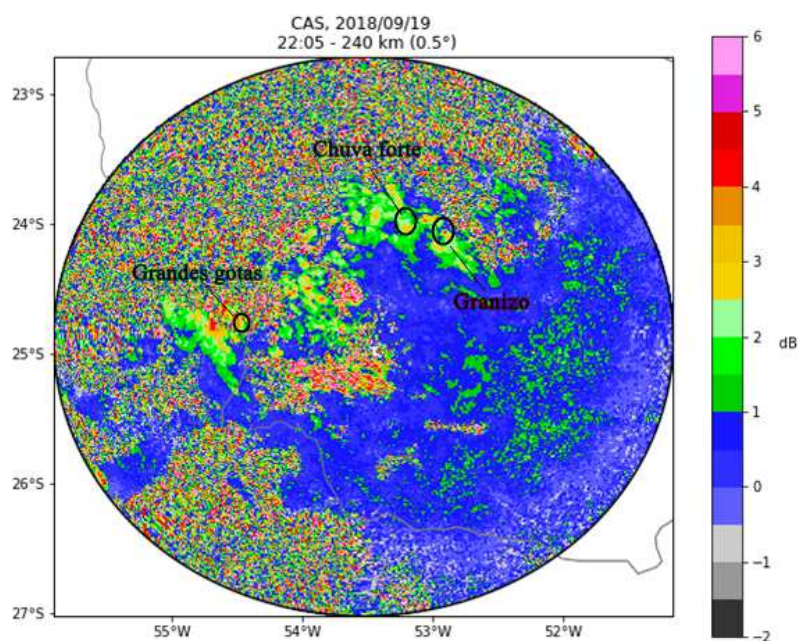
FONTE: adaptado de Bringi e Chandrasekar (2001)

LEGENDA: Variação dos valores da refletividade diferencial em função da razão entre as dimensões vertical e horizontal das partículas.

FIGURA 7 – VARIAÇÃO DE Z_{DR} PARA DIFERENTES TIPOS DE HIDROMETEOROS

FONTE: adaptado de Fabry (2015)

LEGENDA: Valores típicos da refletividade diferencial em relação ao tipo de hidrometeoro.

FIGURA 8 – VALORES DE Z_{DR} PARA O RADAR DE CASCAVEL

FONTE: a autora

LEGENDA: Valores Z_{DR} para a região do radar de Cascavel no dia 19/09/2018 às 22:05, com a indicação de hidrometeoros.

2.2.3 Coeficiente de Correlação Copolar (ρ_{HV})

O coeficiente de correlação copolar, ρ_{HV} , é a correlação entre os sinais polarizados vertical e horizontalmente. Sendo assim seus valores variam entre 0 e 1, onde 0

significa que não há correlação entre os sinais e 1 representa a correlação perfeita dos sinais (FABRY, 2015).

O valor de ρ_{HV} pode ser calculado através da equação (2.10):

$$\rho_{HV} = \frac{\langle S_{VV} S_{HH}^* \rangle}{\sqrt{\langle |S_{HH}|^2 \rangle} \sqrt{\langle |S_{VV}|^2 \rangle}} \quad (2.10)$$

onde S é a intensidade do sinal, S^* representa o conjugado complexo da intensidade do sinal e os índices representam a forma de emissão e de captação do sinal, vertical e horizontal (SELUZNIAK, 2016). Considerando uma região contribuinte com n alvos, o coeficiente de correlação copolar associado a esta região é dado pela equação (2.11):

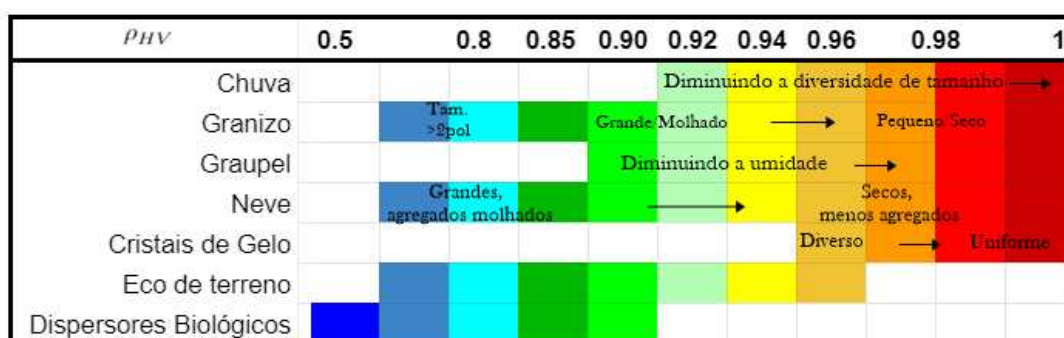
$$\rho_{HV} = \frac{\sum_{i=1}^n (S_i^{VV} - \overline{S^{VV}}) (S_i^{*HH} - \overline{S^{*HH}})}{\sqrt{\sum_{i=1}^n (S_i^{VV} - \overline{S^{VV}})^2 \sum_{i=1}^n (S_i^{*HH} - \overline{S^{*HH}})^2}} \quad (2.11)$$

onde \overline{S} representa a média dos sinais (RAUBER; NESBITT, 2018).

Assim o coeficiente de correlação copolar é uma variável polarimétrica que mede a uniformidade de forma dentro de uma região contribuinte, sendo que volumes com alvos uniformes estão associados a valores mais altos de ρ_{HV} e volumes com alvos de diversas formas geram valores mais baixos de ρ_{HV} (FABRY, 2015).

De acordo com Calvetti, Beneti e Medeiros (2017) valores de ρ_{HV} acima de 0,96 indicam um volume constituído por gotículas e cristais de gelo, uma vez que esse valor representa pouca diversidade de hidrometeoros. Já valores de ρ_{HV} entre 0,85 e 0,95 estão associados a volumes com cristais de gelo de grandes diâmetros e formas diferentes, podendo conter partículas que não são hidrometeoros. Por fim valores de ρ_{HV} menores que 0,85 em geral estão associados a alvos não meteorológicos.

FIGURA 9 – VARIACÃO DE ρ_{HV} PARA DIFERENTES TIPOS DE ALVOS

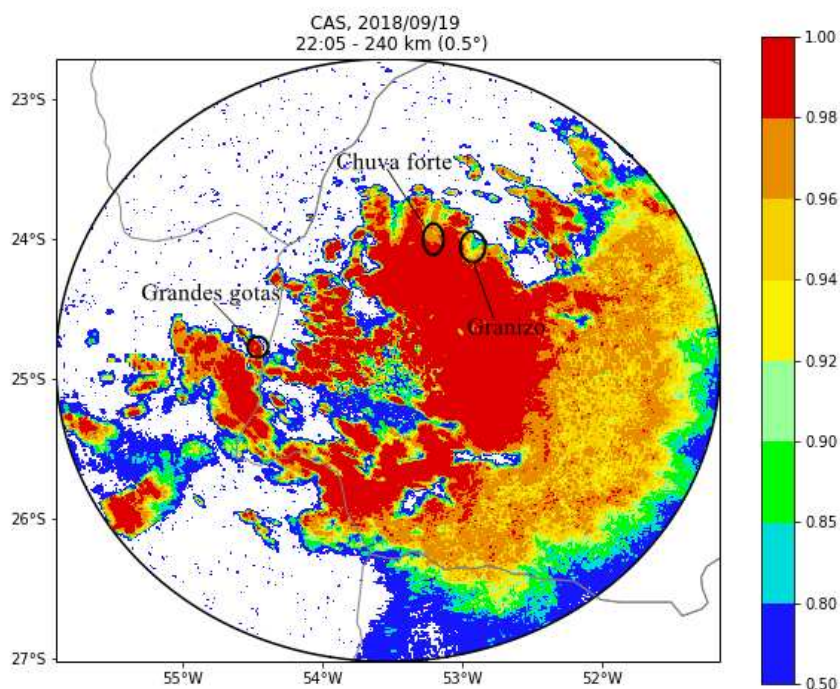


FONTE: adaptado de Fabry (2015)

LEGENDA: Valores típicos do coeficiente de correlação copolar em relação aos diferentes alvos.

Na FIGURA 9 é possível observar os valores típicos de ρ_{HV} associados a diferentes tipos de alvos. A FIGURA 10 apresenta os valores de ρ_{HV} para a região do radar de Cascavel no dia 19 de setembro de 2018 às 22:05.

FIGURA 10 – VALORES DE ρ_{HV} PARA O RADAR DE CASCAVEL



FONTE: a autora

LEGENDA: Valores ρ_{HV} para a região do radar de Cascavel no dia 19/09/2018 às 22:05, com a indicação de hidrometeoros.

2.2.4 Diferencial de Fase de Propagação Φ_{DP}

O diferencial de fase de propagação, Φ_{DP} , é a diferença entre a fase do pulso polarizado na horizontal e a fase do pulso polarizado na vertical, sendo determinado pela equação (2.12):

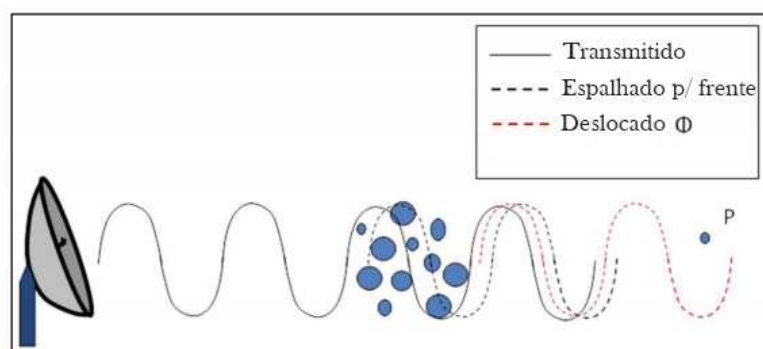
$$\Phi_{DP} = \Phi_H - \Phi_V \quad (2.12)$$

onde Φ_H é a fase da onda polarizada horizontalmente e Φ_V representa a fase da onda polarizada verticalmente, logo Φ_{DP} é medido em graus (CALVETTI; BENETI; MEDEIROS, 2017).

O valor de Φ_{DP} tende a aumentar monotonicamente, porém irregularmente, com o aumento do range, ou seja, conforme o sinal se distancia do radar. Isso acontece porque esse valor reflete as características de todo caminho entre o radar e o alvo

observado, sendo que sua taxa de aumento é uma função combinada do número, tamanho e alongamento dos alvos (FIGURA 11) (FABRY, 2015). A FIGURA 12 apresenta os valores de Φ_{DP} para a região do radar de Cascavel no dia 19 de setembro de 2018 às 22:05.

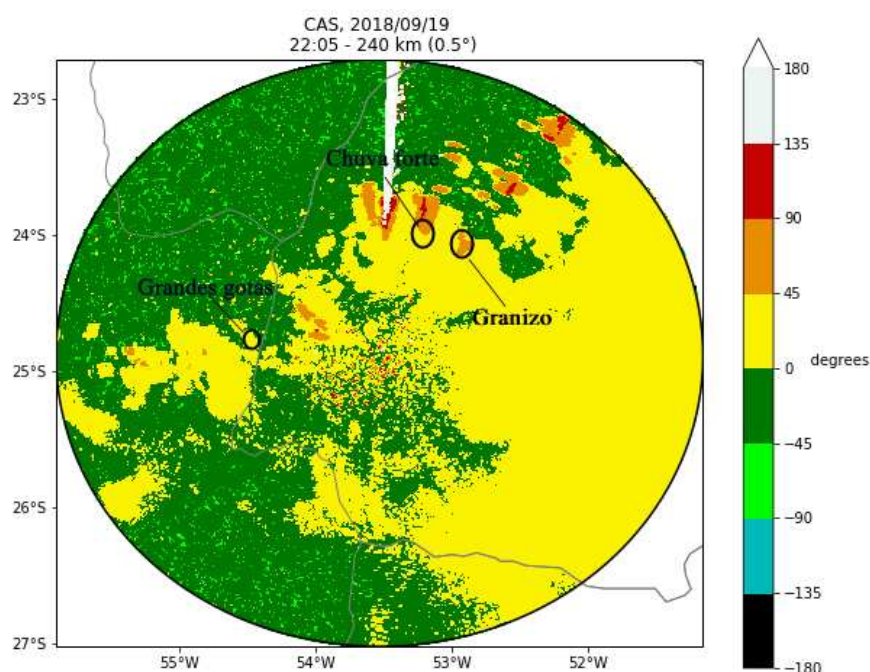
FIGURA 11 – DIFERENÇA DE FASE CAUSADA POR HIDROMETEOROS



FONTE: adaptado de (MELO, 2016)

LEGENDA: Exemplo de diferença de fase no sinal causada pela presença de hidrometeoros.

FIGURA 12 – VALORES DE Φ_{DP} PARA O RADAR DE CASCAVEL



FONTE: a autora

LEGENDA: Valores Φ_{DP} para a região do radar de Cascavel no dia 19/09/2018 às 22:05, com a indicação de hidrometeoros.

2.2.5 Diferencial de Fase Específica (K_{DP})

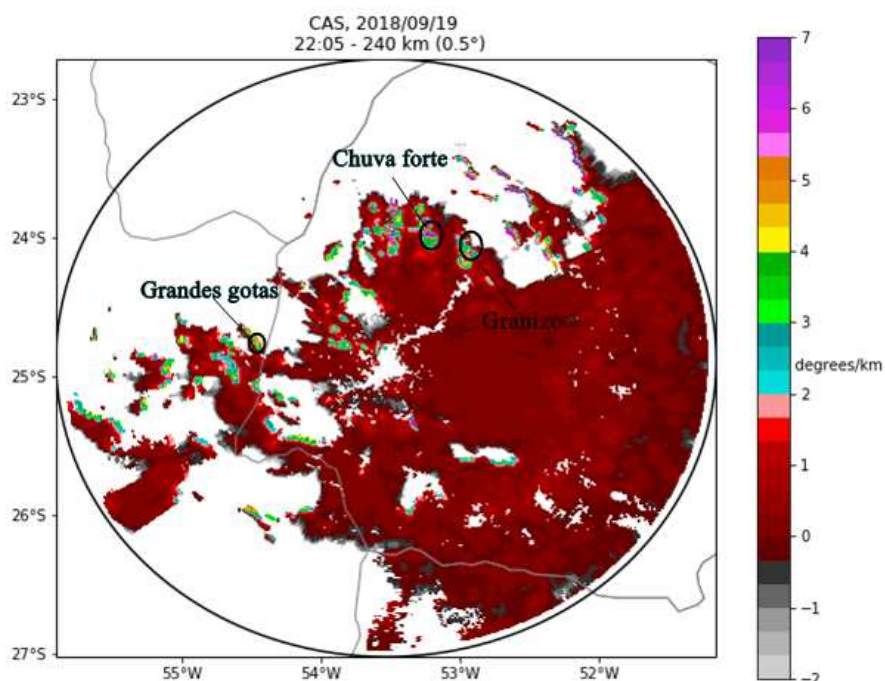
O diferencial de fase específica, K_{DP} , é a derivada espacial de Φ_{DP} em relação a distância do alvo ao radar (SELUZNIAK, 2016). Sendo assim K_{DP} pode ser determinado a partir da equação (2.13):

$$K_{DP} = \frac{\Phi_{DP}(r_2) - \Phi_{DP}(r_1)}{2(r_2 - r_1)} \quad (2.13)$$

onde r_1 e r_2 são as distâncias entre os pontos 1 e 2 e a antena do radar, medidas em quilômetros, com $r_1 < r_2$. E Φ_{DP} é o diferencial de fase de propagação no ponto em questão (CALVETTI; BENETI; MEDEIROS, 2017).

Uma das vantagens do K_{DP} em relação as demais variáveis polarimétricas é que ele não é afetado pela atenuação (FABRY, 2015). Valores de K_{DP} próximos de zero indicam que o pulso polarizado horizontalmente se propagou com velocidade menor que o pulso polarizado verticalmente, esta é uma situação típica de regiões com pequenas gotículas de chuva ou granizo esféricos. Já valores mais altos de K_{DP} costumam indicar volumes com grandes gotas e cristais de gelo não uniformes (FIGURA 13) (CALVETTI; BENETI; MEDEIROS, 2017).

FIGURA 13 – VALORES DE K_{DP} PARA O RADAR DE CASCAVEL



FONTE: a autora

LEGENDA: Valores K_{DP} para a região do radar de Cascavel no dia 19/09/2018 às 22:05, com a indicação de hidrometeoros.

2.3 CLASSIFICAÇÃO DE HIDROMETEOROS

Os hidrometeoros são meteoros constituídos por água, ou seja, consistem em um conjunto de partículas de água que podem estar em estado líquido ou sólido. Os hidrometeoros podem ser de queda ou precipitação, como chuva, granizo e neve, de suspensão, como nevoeiro, de deposição, como geada e orvalho, ou de levantamento, como espuma.

As principais técnicas e algoritmos utilizados para fazer a classificação de hidrometeoros a partir de dados de radares com dupla polarização, costumam utilizar os valores de Z_H , Z_{DR} , ρ_{HV} e K_{DP} , como é o caso dos trabalhos citados na TABELA 1. Além disso as técnicas mais utilizadas para esse estudo são lógica fuzzy, neuro-fuzzy, estatística bayesiana e simulações teóricas (RAMIREZ, 2018).

TABELA 1 – REVISÃO BIBLIOGRÁFICA: CLASSIFICAÇÃO DE HIDROMETEOROS

Autores	Banda	Algoritmo	Variáveis Polarimétricas
Höller et al. (1994)	C	árvore de decisão	Z_H , Z_{DR} , LDR
Vivekanandan et al. (1999)	S	lógica fuzzy	Z_H , Z_{DR} , LDR
Straka, Zrnić e Ryzhkov (2000)	S	lógica fuzzy	Z_H , Z_{DR} , Φ_{DP} , K_{DP} , ρ_{HV} , LDR
Liu e Chandrasekar (2000)	S	neuro-fuzzy	Z_H , Z_{DR} , K_{DP} , ρ_{HV} , LDR
Keenan (2003)	C	lógica fuzzy	Z_H , Z_{DR} , K_{DP} , ρ_{HV}
Marzano et al. (2006)	C	lógica fuzzy	Z_H , Z_{DR} , K_{DP}
Marzano et al. (2008)	C	estatística bayesiana	Z_H , Z_{DR} , K_{DP}
Dolan e Rutledge (2009)	X	lógica fuzzy	Z_H , Z_{DR} , K_{DP} , ρ_{HV}
Park et al. (2009)	S	lógica fuzzy	Z_H , Z_{DR} , K_{DP} , ρ_{HV} , $SD(Z)$, $SD(\Phi_{DP})$
Snyder et al. (2010)	X	lógica fuzzy	Z_H , Z_{DR} , LK_{DP} , ρ_{HV}
Kouketsu e Uyeda (2010)	X	lógica fuzzy	Z_H , Z_{DR} , K_{DP} , ρ_{HV}
Schuur et al. (2012)	S	árvore de decisão	Z_H , Z_{DR} , LK_{DP} , ρ_{HV} , $SD(Z)$, $SD(\Phi_{DP})$
Al-Sakka et al. (2013)	X, C, S	lógica fuzzy	Z_H , Z_{DR} , K_{DP} , ρ_{HV}

FONTE: adaptada e atualizada de Al-Sakka et al. (2013)

LEGENDA: Revisão bibliográfica das técnicas e variáveis utilizadas para a classificação de hidrometeoros.

Os radares de Cascavel e Lontras, que são os radares de polarização dupla utilizados neste trabalho, utilizam lógica fuzzy para, a partir da refletividade e das variáveis polarimétricas, fazer a classificação do alvo em relação aos hidrometeoros (SELUZNIAK, 2016). A abordagem de lógica fuzzy é um método não linear considerado uma boa abordagem para a classificação de hidrometeoros (AL-SAKKA et al., 2013).

A primeira publicação sobre lógica fuzzy foi feita pelo professor de Ciências da Computação da Universidade da Califórnia, Lofti Asker Zadeh, nos anos 60. Este fato só foi possível pelo desenvolvimento dos conjuntos com grau de pertinência feito por Jan Lukasiewicz em 1920. Isso porque a lógica fuzzy é uma classificação, que diferentemente da lógica booleana, assume infinitos valores entre 0 e 1. Com isto, é possível encontrar valores não inteiros, como 0.5 que representa a meia verdade. Esses valores são úteis para melhorar a identificação do hidrometeoro independente da sua forma, sua inclinação ou até mistura com outros, algo que não é possível através da lógica booleana (SOUSA RIGNEL; CHENCI; LUCAS, 2011).

Portanto a lógica fuzzy é um sistema não linear de mapeamento de um vetor de entrada em uma saída escalar através do estudo dos princípios formais do raciocínio aproximado. Nos métodos de lógica fuzzy, a interseção entre um conjunto e seu complemento pode não ser um conjunto nulo. Por exemplo, regiões com chuva e sem chuva podem conter partículas de fase mista, como partículas de gelo derretidas ou misturas de chuva e granizo. Assim, existe imprecisão ou um conjunto não vazio entre um objeto e seu oposto. Objetos ou eventos que exibem imprecisão são classificados como conjuntos fuzzy (VIVEKANANDAN et al., 1999).

Em uma abordagem baseada em lógica fuzzy, há uma transição suave nas fronteiras polarimétricas observáveis entre os tipos de precipitação. O processo de classificação utilizando lógica fuzzy começa pela "fuzzificação", que é a conversão de números reais para o domínio fuzzy, isto é, associando a cada número real um grau de pertinência apropriado para o conjunto que dado número pertence. Para a abordagem proposta por Vivekanandan et al. (1999), por exemplo, para cada um dos dados de entrada, os valores particulares de Z_H , Z_{DR} e razão de despolarização linear, LDR, as funções de pertinência determinam o grau em que a observação pertence a cada um dos conjuntos fuzzy apropriados, no caso, chuva e granizo.

Os resultados da "fuzzificação" para cada variável são ponderados por um peso predeterminado w e somados para produzir um único valor para cada um dos conjuntos fuzzy. Essa unificação de saídas fuzzy para todas as regras de conjuntos fuzzy é chamada de agregação. O banco de regras, ou base de conhecimento, possui todas as informações dos universos de referência, agindo como o cérebro do sistema fuzzy.

As saídas fuzzy do processo de agregação passam então pela "defuzzificação" por meio de uma função que determina o máximo entre os conjuntos fuzzy. O conjunto difuso com o valor máximo é identificado como o tipo de partícula para os valores Z_H , Z_{DR} e LDR fornecidos (VIVEKANANDAN et al., 1999).

Höller et al. (1994) utilizaram as variáveis polarimétricas Z_H , Z_{DR} e LDR, juntamente com a informação da altura do alvo em relação ao nível de derretimento, para através da técnica de árvore de decisão classificar os hidrometeoros em granizo grande e úmido, granizo úmido, granizo seco, chuva com granizo grande, chuva com granizo pequeno, grandes gotas, chuva leve, graupel e neve. A comparação de ondas de granizo derivadas deste estudo e observações terrestres de danos de granizo verificou a assinatura de granizo LDR - Z_{DR} definida neste trabalho.

Vivekanandan et al. (1999) desenvolveram um classificador de chuva e granizo base-

ado na teoria de lógica fuzzy, utilizando as variáveis polarimétricas Z_H , Z_{DR} e LDR. A partir da implementação da lógica fuzzy entre tipos de precipitação (chuva e granizo) fizeram uso de operações matemáticas simples, para implementar em tempo real. Um ponto positivo deste método é que o mesmo não é afetado de forma negativa pelos erros típicos de medição de algumas variáveis de entrada.

Straka, Zrnić e Ryzhkov (2000) utilizaram um algoritmo de classificação baseado na lógica fuzzy para, a partir das variáveis polarimétricas Z_H , Z_{DR} , Φ_{DP} , K_{DP} , ρ_{HV} e LDR fazer um resumo de informações baseadas nas observações e modelagem de assinatura dos hidrometeoros para deduzir seus tipos e quantidades. Fizeram também a identificação, a estimação e a quantificação dos hidrometeoros baseados na espécie, tamanho, concentrações e quantidades.

Liu e Chandrasekar (2000) aplicaram em seus estudos a técnica de neuro-fuzzy, que é uma lógica fuzzy acoplada a classificação de redes neurais, onde a lógica fuzzy foi usada para inferir o tipo de hidrometeoro, e o algoritmo de aprendizagem da rede neural foi usado para ajuste automático dos parâmetros dos conjuntos fuzzy no sistema de lógica fuzzy de acordo com ao conhecimento prévio. Como entrada para o algoritmo foram utilizadas as variáveis polarimétricas Z_H , Z_{DR} , K_{DP} , ρ_{HV} e LDR juntamente com a altitude correspondente ao alvo. Os hidrometeoros foram então classificados em garoa, chuva, neve seca de baixa densidade, cristais secos de alta densidade, neve úmida e derretida, graupel seco, graupel molhado, granizo pequeno, granizo grande e chuva com granizo. A validação do estudo foi feita através de observações in situ.

Keenan (2003) mostrou empiricamente um esquema baseado na lógica fuzzy. Neste método, primeiramente são consideradas as medições das variáveis polarimétricas Z_H , Z_{DR} , K_{DP} , ρ_{HV} para definir uma função de associação em espaços bidimensionais e, em seguida, são acopladas a uma função de pertinência dependente da temperatura para classificar os hidrometeoros em dez espécies. Os resultados alcançados foram bons, mas tiveram dificuldades nas regiões.

Marzano, Scaranari, Celano et al. (2006) aplicaram um método de classificação de lógica fuzzy que usa um conjunto reduzido de observáveis polarimétricos, ou seja, Z_H , Z_{DR} , K_{DP} , de dados provenientes de sites de radar localizados em Gattatico e S. Pietro Capofiume no norte da Itália. O objetivo final foi mostrar melhorias qualitativas na precisão com relação ao uso de um conjunto de dez funções de pertinência bidimensionais.

Marzano, Scaranari, Vulpiani et al. (2008) desenvolveram um algoritmo baseado na estatística Bayesiana supervisionado por um modelo microfísico de retroespalhamento para classificação de hidrometeoros. O objetivo deste trabalho foi representar dez diferentes classes de hidrometeoros em água, gelo e fase mista. A precisão obtida neste estudo foi melhor do que a obtida a partir do algoritmo de classificação de lógica fuzzy desenvolvido anteriormente (Marzano, Scaranari, Celano et al. (2006)).

Dolan e Rutledge (2009) desenvolveram um algoritmo de identificação de hidrometeoros para radares de banda X baseado em simulações teóricas usando o modelo de es-

palhamento de matriz T de sete tipos diferentes de hidrometeoros: chuva, garoa, agregados, cristais de gelo, graupel de baixa densidade, graupel de alta densidade e gelo vertical. Granizo e hidrometeoros de fase mista não foram considerados neste estudo. Os intervalos de valores das variáveis Z_H , Z_{DR} , K_{DP} e ρ_{HV} com base nas simulações teóricas foram utilizados para criar funções de pertinência beta unidimensionais de associação de lógica fuzzy que formam a base do classificador de hidrometeoros.

Park et al. (2009) desenvolveram modificações no método aplicado anteriormente em Straka, Zrnić e Ryzhkov (2000). Os refinamentos do algoritmo de classificação de hidrometeoros anterior com base nos princípios da lógica fuzzy, incluem a estimativa de fatores de confiança que caracterizam os possíveis impactos de todas as fontes de erro nas medições de radar, a atribuição da matriz de pesos que caracteriza o poder de classificação de cada variável em relação a cada classe de alvo de radar, e a implementação de um sistema de designação de classe com base na distância do radar e nos parâmetros da camada de fusão. Essas adições forneceram flexibilidade e melhoraram a discriminação entre hidrometeoros líquidos e congelados. Este esquema de classificação utiliza todas as variáveis polarimétricas disponíveis e distingue 10 classes diferentes de hidrometeoros de radar.

Snyder et al. (2010) aplicaram um algoritmo de classificação de hidrometeoros de lógica fuzzy para estimar a distribuição tridimensional de hidrometeoros usando dados de radar de banda X. Este estudo também testou vários procedimentos de correção de atenuação usando medições de polarização dupla para fazer a correção prévia dos dados que foram utilizados na aplicação do algoritmo.

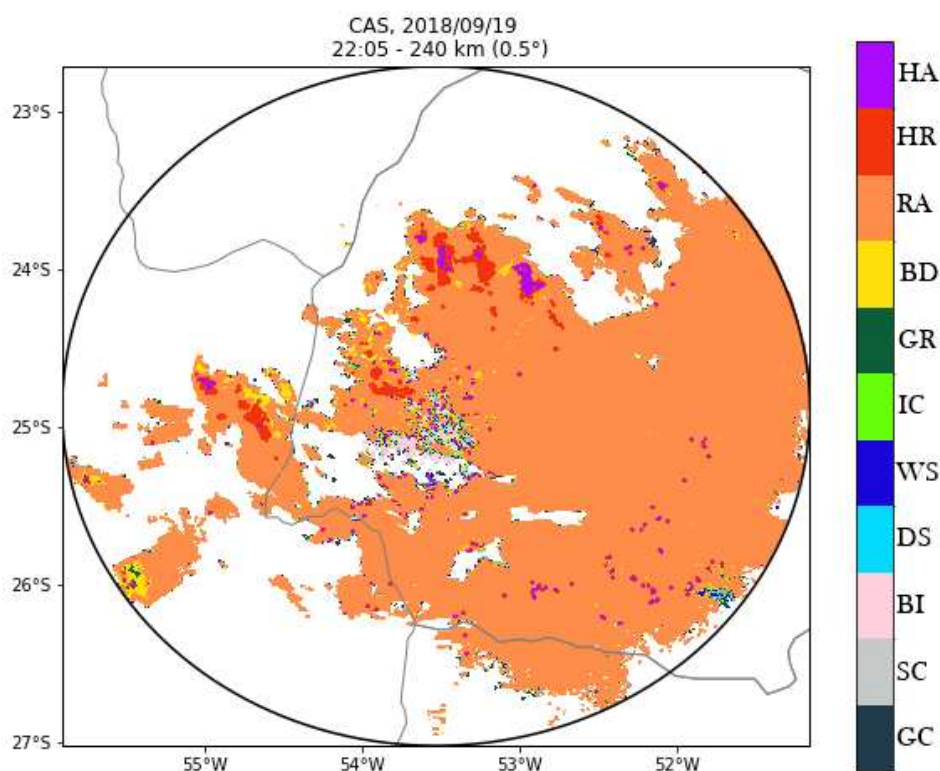
Kouketsu e Uyeda (2010) aplicaram um método de classificação de hidrometeoros baseado em lógica fuzzy, que foi adequado para observação de hidrometeoros sólidos em ambientes úmidos que produzem pouco ou nenhum granizo. Este método identifica o hidrometeoro mais provável em cada volume de amostragem de radar dentre oito categorias: garoa, chuva, neve úmida agregada, neve seca agregada, cristais de gelo, graupel seco, graupel úmido e chuva com granizo. Os dados utilizados pelo método foram os valores das variáveis polarimétricas Z_H , Z_{DR} , K_{DP} e ρ_{HV} juntamente com a temperatura e informações da umidade relativa. A validação da técnica foi feita a partir de observações terrestres e in situ de hidrometeoros sólidos.

Schuur et al. (2012) apresentam um algoritmo de classificação de hidrometeoros que combina a saída termodinâmica do modelo *Rapid Update Cycle* (RUC) com observações polarimétricas de radar. O algoritmo aprimorou as técnicas de classificação existentes até então, que dependiam exclusivamente de observações polarimétricas de radar, usando informações termodinâmicas para ajudar a diagnosticar processos microfísicos. Primeiramente o algoritmo usa perfis verticais de temperatura de bulbo úmido derivados da saída do modelo RUC para fornecer um tipo de classificação de precipitação de fundo. A partir de um conjunto de regras empíricas, os dados do radar polarimétrico são então usados para refinar as categorias do tipo de precipitação quando as observações são consideradas inconsistentes com a classificação de fundo. O algoritmo foi testado em um evento transicional de tempestade de inverno que

produziu uma combinação de chuva, chuva congelada, pelotas de gelo e neve e passou sobre o centro de Oklahoma em 30 de novembro de 2006.

Al-Sakka et al. (2013) propõem um algoritmo de classificação de hidrometeoros baseado em lógica fuzzy que leva em consideração funções de associação baseadas em dados, condições de medição e informações de temperatura tridimensionais fornecidas pelo modelo de Aplicação de Pesquisa para Operações em Mesoescala, que é um modelo de previsão de tempo numérico não hidrostático de alta resolução. Este algoritmo emprega funções de associação bivariadas que foram estabelecidas usando dados reais coletados pelos radares polarimétricos franceses e simulações de matriz T para as variáveis Z_H , Z_{DR} , K_{DP} e ρ_{HV} .

FIGURA 14 – CLASSIFICAÇÃO DE HIDROMETEOROS PARA O RADAR DE CASCAVEL



FONTE: a autora

LEGENDA: Classificação de hidrometeoros para a região do radar de Cascavel no dia 19/09/2018 às 22:05.

Os radares que geraram os dados utilizados neste trabalho fazem a classificação dos seus alvos em 11 classes: eco de terreno - GC (Ground Clutter), eco do mar - SC (Sea Clutter), dispersores biológicos - BI (Biological), neve seca - DS (Dry Snow), neve molhada - WS (Wet Snow), cristais de gelo - IC (Ice Crystals), graupel - GR (Graupel), grandes gotas - BD (Big Drops), chuva leve e moderada - RA (Rain), chuva forte - HR (Heavy Rain) e granizo - HA (Hail) (PARK et al., 2009). É possível observar o resultado dessa classificação para o radar de Cascavel na FIGURA 14.

Como fica evidente pelos estudos desenvolvidos para a classificação de hidrometeoros, os valores das variáveis polarimétricas permitem estimar algumas características dos mesmo, assim é possível determinar qual hidrometeoro é o alvo utilizando estas medidas. Portanto, o objetivo da classificação de hidrometeoros é identificar padrões e similaridades nos valores das variáveis polarimétricas (RAMIREZ, 2018). Para identificar estas similaridades neste trabalho foram aplicadas técnicas de redes neurais artificiais, as quais são descritas no próximo capítulo.

3 REDES NEURAIS

Redes neurais artificiais, RNA, são modelos computacionais inspirados no sistema nervoso de seres vivos. Os primeiros trabalhos sobre o tema foram publicados há mais de 60 anos, derivados do sonho de construir uma máquina inteligente e autônoma (SILVA; SPATTI; FLAUZINO, 2010). Uma vez que é reconhecido que o cérebro humano computa de uma maneira diferente do computador convencional, possuindo uma alta complexidade, pois seu processamento é não linear e paralelo, assumiu-se então a missão de reproduzir computacionalmente o funcionamento do mesmo. O cérebro humano é capaz de organizar seus neurônios de modo a realizar reconhecimento de padrões, percepção e controle motor muitas vezes mais rápido do que um computador digital (HAYKIN, 2008).

As RNA's começaram a ser amplamente estudadas em 1990, sendo que suas principais aplicações incluem avaliação de imagens captadas por satélites, classificação de padrões de escrita e fala, reconhecimento de faces em visão computacional, previsão de ações no mercado financeiro, identificação de anomalias em imagens médicas, controle de aparelhos eletrônicos e eletrodomésticos, entre outras (SILVA; SPATTI; FLAUZINO, 2010).

Uma rede neural é um processador constituído por um conjunto de unidades de processamento simples, os neurônios artificiais, os quais armazenam conhecimento experiencial e são interligados por um grande número de interconexões, as chamadas sinapses artificiais (SILVA; SPATTI; FLAUZINO, 2010; HAYKIN, 2008). Uma RNA se assemelha ao cérebro humano em dois pontos: a rede adquire seu conhecimento através de um processo de aprendizagem, e a intensidade das conexões entre os neurônicos, chamada de pesos sinápticos, são usados para armazenar o conhecimento adquirido (HAYKIN, 2008).

De acordo com Silva, Spatti e Flauzino (2010) e (HAYKIN, 2008) as principais características e recursos das redes neurais que se tornam úteis em suas aplicações são:

- **Adaptação por experiência:** os parâmetros internos da rede, como seus pesos sinápticos, são adaptados e ajustados a partir da apresentação sucessiva de exemplos, padrões, amostras e medidas. Além disso, quando ocorrem pequenas mudanças no ambiente em que a rede neural foi treinada para atuar, a mesma pode ser facilmente retreinada para se adaptar a essas mudanças;
- **Capacidade de aprendizado:** a partir da execução de um método de treinamento a rede é capaz de extrair a relação existente entre as variáveis que compõem a aplicação;
- **Habilidade de generalização:** após seu treinamento, a rede é capaz de generalizar seu conhecimento estimando soluções para conjuntos de dados até então desconhecidos;
- **Organização de dados:** a rede é capaz de se organizar internamente possibilitando o agrupamento de padrões semelhantes, a partir das características intrínsecas do conjunto de informações a respeito do processo;

- **Tolerância a falhas:** graças ao elevado nível de interconexões entre os neurônios artificiais, a rede neural constitui um sistema tolerante a falhas quando parte de sua estrutura interna é corrompida;
- **Armazenamento distribuído:** como o conhecimento sobre um determinado processo dentro de uma arquitetura neural é distribuído entre as diversas sinapses dos neurônios artificiais, a rede é capaz de desenvolver um incremento na robustez da arquitetura quando alguns neurônios ficam inativos;
- **Facilidade de prototipagem:** a maioria das redes neurais possuem uma fácil prototipagem em hardware ou software, uma vez que, após treinadas, seus resultados geralmente são obtidos por algumas operações matemáticas elementares;
- **Não linearidade:** as funções de ativação dos neurônios artificiais podem ser lineares ou não lineares, logo a rede neural em si, por ser um conjunto de neurônios artificiais interligados, pode ser linear ou não linear. A não linearidade de uma rede é uma propriedade importante, principalmente se o objeto de estudo da rede possui características não lineares, como por exemplo, o sinal de voz;
- **Uniformidade:** no universo das redes neurais artificiais a mesma notação é usada em todas suas aplicações, isso porque os neurônios são comuns a todas as redes neurais. Graças a essa uniformidade é possível compartilhar teorias e algoritmos de aprendizagem em diferentes aplicações de redes neurais.

De acordo com Fausett (1994) as RNA's possuem três características que as definem, que são sua arquitetura, ou seja, o modo como os neurônios estão conectados uns aos outros, o tipo de treinamento, ou seja, o modo como os pesos das ligações entre os neurônios são definidos e sua(s) função(ões) de ativação.

3.1 NEURÔNIOS ARTIFICIAIS

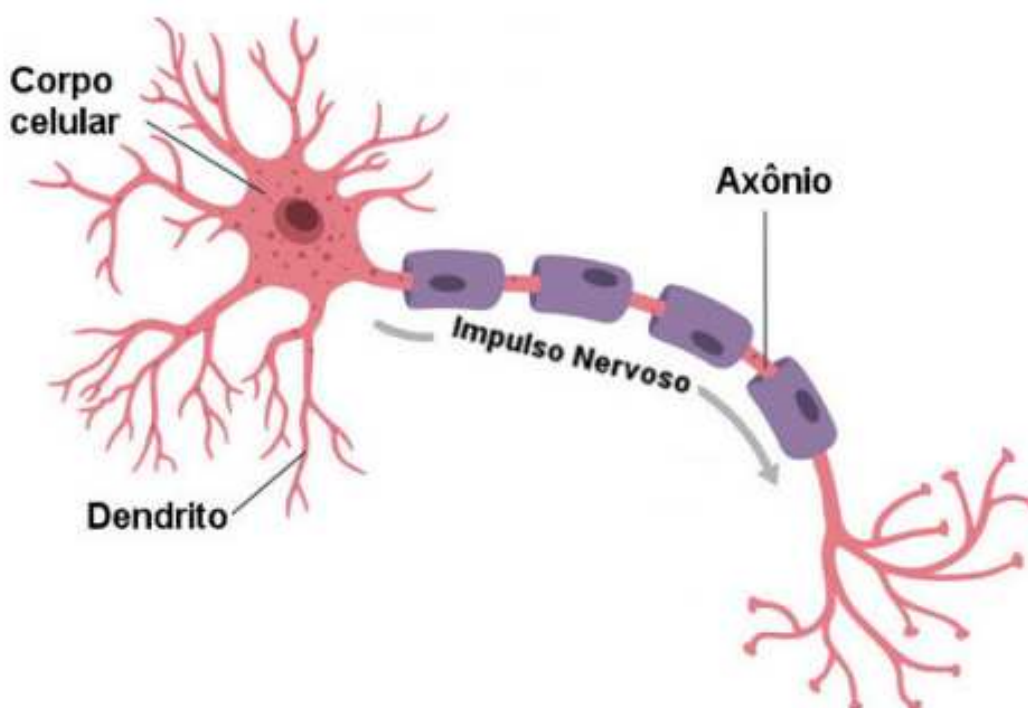
A primeira ideia e modelo de um neurônio artificial foi dada por McCulloch e Pitts (1943), onde os autores realizaram a primeira modelagem matemática de um neurônio biológico (SILVA; SPATTI; FLAUZINO, 2010). Isso só foi possível após o trabalho pioneiro de Cajal (1911) em descrever os neurônios como constituintes estruturais do cérebro (HAYKIN, 2008).

No caso dos neurônios biológicos, eles funcionam em paralelo, com o objetivo de cumprir suas funcionalidades de pensar e memorizar. O papel do neurônio biológico é basicamente conduzir impulsos, que são estímulos elétricos advindos de reações físico-químicas (SILVA; SPATTI; FLAUZINO, 2010).

Os neurônios biológicos podem ser divididos em três partes principais (FIGURA 15), os dendritos, corpo celular ou soma e axônio. A função dos dendritos é captar continuamente os estímulos vindos dos demais neurônios ou do meio externo, caso seja um neurônio sensitivo. O corpo celular, ou soma, por sua vez, é responsável por processar os sinais ou informações

recebidas para produzir o potencial de ativação e assim determinar se o neurônio deve disparar um impulso elétrico ao longo de seu axônio. Sendo assim a missão do axônio é conduzir os impulsos nervosos para os demais neurônios, podendo esses serem os neurônios efetadores que estão conectados diretamente com o tecido muscular (SILVA; SPATTI; FLAUZINO, 2010).

FIGURA 15 – NEURÔNIO BIOLÓGICO



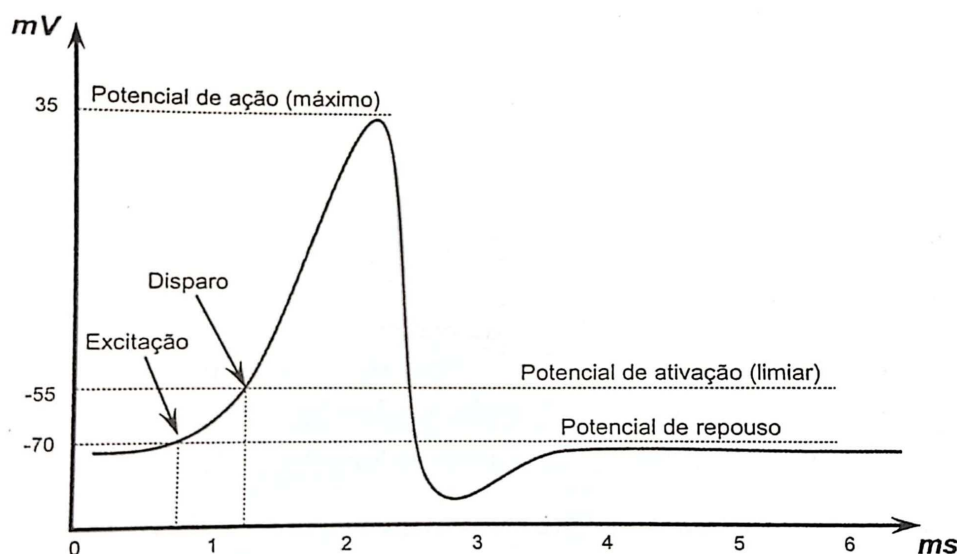
FONTE: Adaptado de Siqueira (2018)

LEGENDA: Principais componentes de um neurônio biológico.

As sinapses são as responsáveis em promover a interação entre os neurônios, ou seja, fazer com que seja possível transferir o impulso elétrico de um neurônio para outro. Como não existe contato físico entre os neurônios, um processo pré-sináptico libera uma substância transmissora, os neurotransmissores, que se difunde através da junção sináptica entre os neurônios e é responsável por ponderar a transmissão do impulso elétrico entre os mesmos (HAYKIN, 2008; SILVA; SPATTI; FLAUZINO, 2010).

Os neurônios possuem o que é chamado de potencial de ação da membrana neural (FIGURA 16). Quando o neurônio está em repouso, esse potencial é negativo, em geral $-70mV$, quando essa célula nervosa é estimulada e ultrapassa o limiar de ativação, que é $-55mV$, a mesma sai do estado de repouso e atinge um potencial de ação, que faz com que o neurônio dispare seu impulso elétrico, podendo alcançar a amplitude máxima de $35mV$ (SILVA; SPATTI; FLAUZINO, 2010; KOVÁCS, 2006).

FIGURA 16 – VARIAÇÃO DO POTENCIAL DE AÇÃO DO NEURÔNIO



FONTE: Silva, Spatti e Flauzino (2010)

Apesar de as atividades desempenhadas por cada neurônio individualmente serem relativamente simples, o conjunto dessas células nervosas é o principal responsável pelos processamentos executados e gerenciados pelo cérebro humano. Estima-se que a rede neural biológica seja constituída por 100 bilhões de neurônios, cada um interligado, por conexões sinápticas, com cerca de 6 mil outros neurônios, resultando assim em 600 trilhões de sinapses (SILVA; SPATTI; FLAUZINO, 2010; HAYKIN, 2008).

Os neurônios artificiais (FIGURA 17), por sua vez, são unidades de processamento de informação constituindo um modelo simplificado dos neurônios biológicos. Essas unidades de processamento não-lineares fornecem saídas tipicamente contínuas e realizam funções simples, entretanto são fundamentais para operação de uma rede neural (HAYKIN, 2008; SILVA; SPATTI; FLAUZINO, 2010).

Os sinais de entrada advindos do meio externo, ou seja, o conjunto de valores das variáveis da aplicação, são análogos aos impulsos elétricos externos captados pelos dendritos dos neurônios biológicos. Esse conjunto de entrada é denotado por $\{x_1, x_2, x_3, \dots, x_n\}$ (SILVA; SPATTI; FLAUZINO, 2010).

Nos neurônios biológicos, o sinal é recebido pela célula nervosa através da sinapse, no caso dos neurônios artificiais, cada um deles está associado a um conjunto de pesos sinápticos, denotado por $\{w_1, w_2, w_3, \dots, w_n\}$ que fazem o papel dos neurotransmissores na ponderação do sinal recebido (SILVA; SPATTI; FLAUZINO, 2010).

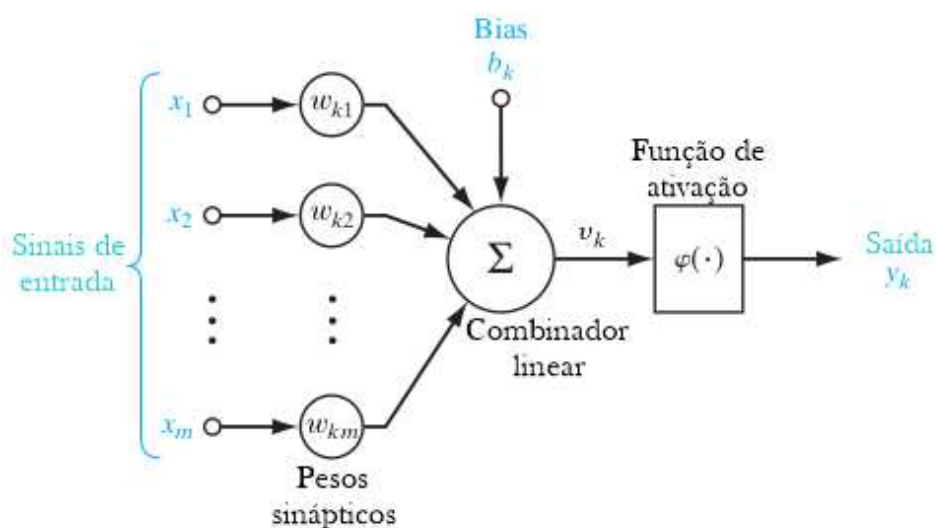
O potencial de ação ou potencial de ativação de um neurônio artificial é produzido através do combinador linear (equação 3.1):

$$u = \sum_{i=1}^n x_i \cdot w_i \quad (3.1)$$

que soma as ponderações dos sinais de entrada, o resultado dessa soma é o potencial de

ativação u do neurônio. Em geral os modelos de neurônios artificiais também possuem um valor associado chamado bias (b), esse valor pode aumentar ou diminuir o potencial de ativação (HAYKIN, 2008). Nos casos em que o bias assume um valor negativo ele pode ser interpretado como um limiar de ativação, ou seja, valor que especifica qual será o patamar para que o resultado produzido pelo combinador linear seja disparado para saída do neurônio (SILVA; SPATTI; FLAUZINO, 2010).

FIGURA 17 – MODELO DE UM NEURÔNIO ARTIFICIAL



FONTE: adaptado de Haykin (2008)

Sendo assim o potencial de ativação é o resultado final da junção do combinador linear com o bias (equação 3.2). Esse valor é a entrada da função de ativação, cujo objetivo é limitar a saída do neurônio dentro de um intervalo de valores a serem assumidos pela sua imagem, por isso ela também recebe o nome de função restritiva. A função de ativação após processar o potencial de ativação do neurônio gera o sinal de saída que é o valor final produzido pelo neurônio, o qual será enviado para os demais neurônios conectados.

$$u = \sum_{i=1}^n x_i \cdot w_i + b \quad (3.2)$$

3.1.1 Funções de Ativação

Segundo Fausett (1994) as funções de ativação são uma das três principais características que definem uma RNA. Pode-se dividir as funções de ativação em dois grupos, as parcialmente diferenciáveis, ou seja, possuem pontos onde a derivada de primeira ordem não existe, e as totalmente diferenciáveis, ou seja, possuem derivada de primeira ordem para todos os pontos de seu domínio (SILVA; SPATTI; FLAUZINO, 2010).

As principais funções de ativação parcialmente diferenciáveis são:

- **Função degrau:** a imagem da função degrau é positiva igual a 1 para potenciais de ativação maiores ou iguais a zero, e nula para potenciais negativos (equação 3.3).

$$f(u) = \begin{cases} 1, & \text{se } u \geq 0 \\ 0, & \text{se } u < 0 \end{cases} \quad (3.3)$$

- **Função degrau bipolar:** a imagem dessa função assumirá valor 1 quando o potencial de ativação for maior que zero, valor nulo quando o potencial for zero e valor -1 quando o potencial for menor que zero (equação 3.4).

$$f(u) = \begin{cases} 1, & \text{se } u > 0 \\ 0, & \text{se } u = 0 \\ -1, & \text{se } u < 0 \end{cases} \quad (3.4)$$

- **Função rampa simétrica:** a imagem dessa função é o próprio potencial para valores dentro de um intervalo $[-a, a]$ e assumirá os valores extremos do intervalo para potenciais fora do mesmo (equação 3.5).

$$f(u) = \begin{cases} a, & \text{se } u > a \\ u, & \text{se } -a \leq u \leq a \\ -a, & \text{se } u < -a \end{cases} \quad (3.5)$$

- **Função reLU:** a imagem dessa função é o próprio potencial para valores potenciais positivos e nulo para os demais valores (equação 3.6).

$$f(u) = \max\{0, u\} = \begin{cases} u, & \text{se } u > 0 \\ 0, & \text{se } u \leq 0 \end{cases} \quad (3.6)$$

As principais funções de ativação totalmente diferenciáveis são:

- **Função logística:** a imagem dessa função são números reais dentro do intervalo de $(0, 1)$, sua expressão é dada pela equação (3.7) onde β é uma constante real que está associada a inclinação da função em relação ao seu ponto de inflexão (FIGURA 18). Note que conforme os valores de β aumentam a função logística se aproxima da função degrau. Quando $\beta = 1$, alguns autores chamam essa função de função sigmoide.

$$f(u) = \frac{1}{1 + e^{-\beta u}} \quad (3.7)$$

- **Função tangente hiperbólica:** a imagem dessa função é o intervalo real $(-1, 1)$, sua expressão é dada pela equação (3.8), onde β é uma constante real, que assim como no caso da função logística, está associada a inclinação da função. Conforme o valor de β aumenta na função tangente hiperbólica, a mesma se aproxima da função degrau bipolar.

$$f(u) = \frac{1 - e^{-\beta u}}{1 + e^{-\beta u}} \quad (3.8)$$

- **Função gaussiana:** a imagem dessa função é o intervalo real $(0, 1]$, sua expressão é dada pela equação (3.9), de modo que os pontos que estão a uma mesma distância da sua média c terão a mesma imagem (FIGURA 18).

$$f(u) = e^{-\frac{(u-c)^2}{2\sigma^2}} \quad (3.9)$$

- **Função linear:** a imagem dessa função é idêntica aos valores do potencial de ativação, sua expressão é dada pela equação(3.10).

$$f(u) = u \quad (3.10)$$

- **Função softmax:** a função softmax é utilizada na camada de saída de uma RNA para a classificação de padrões, sendo que ela retorna, para cada classe i , a probabilidade dos dados analisados pela rede pertencerem a classe, ou seja, sua imagem está contida no intervalo real $(0, 1]$. No caso da função de ativação softmax, sua entrada não é somente o potencial de ativação do neurônio em questão, mas sim os potenciais de ativação dos neurônios que estão na camada onde a função está atuando, sendo assim a expressão dessa função para cada classe é dada pela expressão da equação (3.11), onde \mathbf{u} é o vetor dos potenciais de ativações dos neurônios da camada, u_i é o potencial do neurônio em que a função está sendo aplicada e n é o número de neurônios na camada.

$$f_i(\mathbf{u}) = \frac{e^{u_i}}{\sum_{j=1}^n e^{u_j}} \quad (3.11)$$

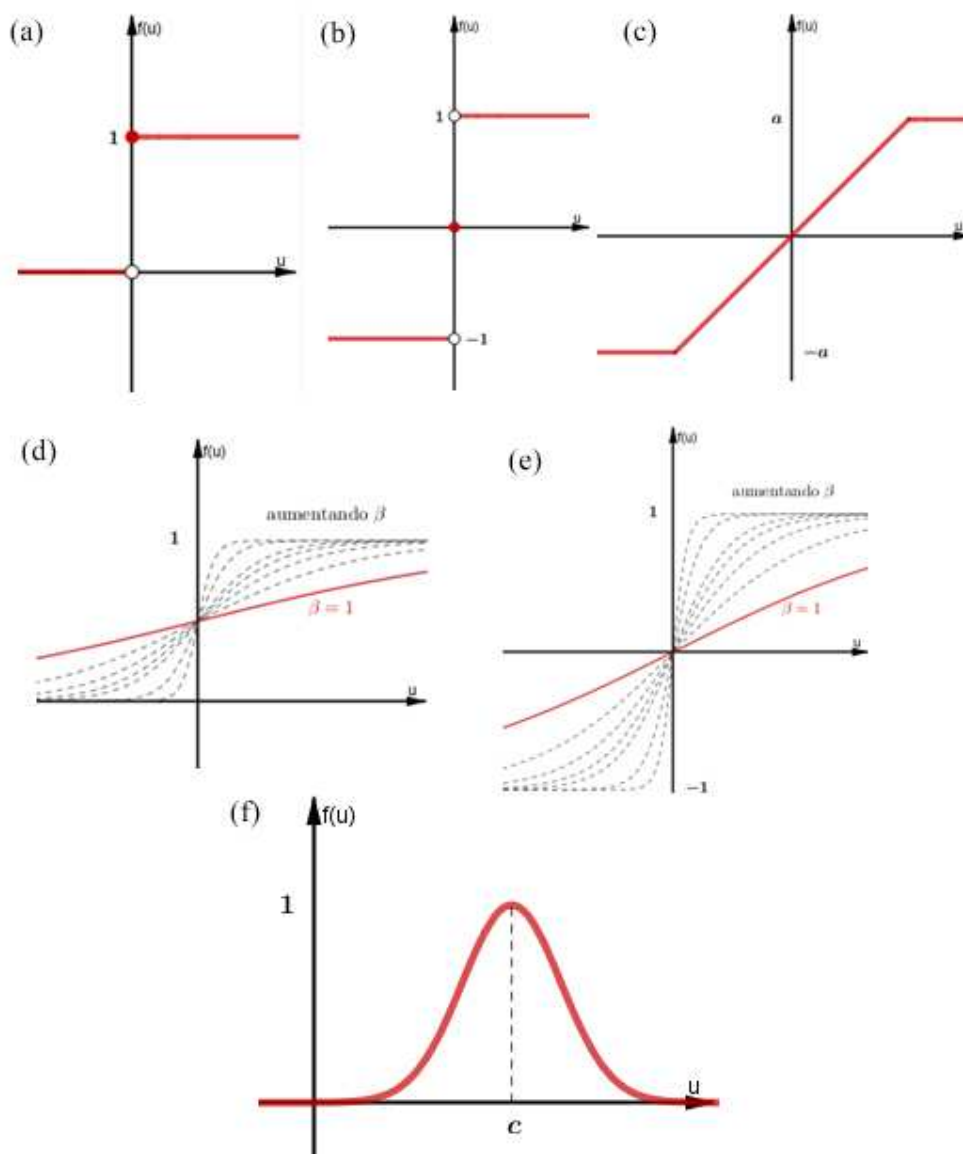
3.1.2 Arquitetura

A arquitetura de uma rede neural artificial é definida pela forma como os neurônios estão estruturados dentro da mesma, ou seja, a forma como os neurônios estão conectados uns aos outros, configurando uma das principais características que definem uma rede neural (FAUSETT, 1994). Em geral os neurônios artificiais são organizados na forma de camadas, essas camadas podem ser camada de entrada, camadas ocultas ou intermediárias e camada de saída (SILVA; SPATTI; FLAUZINO, 2010; HAYKIN, 2008).

A camada de entrada é a responsável pelo recebimento das informações ou dados externos, sendo assim é através da camada de entrada que a rede recebe os valores das variáveis da aplicação estudada. Os neurônios que compõem essa camada são análogos aos neurônios biológicos sensitivos. Os valores recebidos pela camada de entrada costumam ser normalizados em relação a faixa de variação dos valores produzidos pela função de ativação aplicada na rede, uma vez que essa normalização melhora a precisão nas operações matemáticas da rede (SILVA; SPATTI; FLAUZINO, 2010).

Camadas ocultas ou intermediárias são compostas pelos neurônios que são análogos aos neurônios biológicos conectores, esses neurônios são responsáveis por extrair as características associadas a aplicação estudada, sendo assim quase todo o processamento interno

FIGURA 18 – FUNÇÕES DE ATIVAÇÃO



FONTE: a autora

LEGENDA: Exemplos de funções de ativação: (a) função degrau, (b) função degrau bipolar (c) função rampa simétrica, (d) função logística, (e) função tangente hiperbólica e (f) função gaussiana

da rede é realizado nessas camadas. O número de camadas ocultas utilizadas em uma rede bem como o número de neurônios em cada uma dessas camadas é variável, sendo definido de acordo com o objetivo da rede e seu melhor desempenho (SILVA; SPATTI; FLAUZINO, 2010).

A camada de saída por sua vez é a responsável pela produção e apresentação dos resultados finais da rede, sendo assim os neurônios que a constituem são análogos aos neurônios biológicos efetadores (HAYKIN, 2008).

Em geral podemos identificar quatro classes de arquiteturas de rede, são elas redes *feedforward* de camada simples, redes *feedforward* de camadas múltiplas, redes recorrentes e

redes reticuladas (SILVA; SPATTI; FLAUZINO, 2010).

As redes *feedforward* recebem esse nome pela orientação dos sinais enviados entre as camadas, o fluxo de informação nessas redes são unidirecionais, sendo que os sinais de uma camada de neurônios são enviados somente para a camada adjacente seguinte (HAYKIN, 2008).

No caso das redes *feedforward* de camada simples a camada de entrada está diretamente conectada à camada de saída. As principais aplicações dessas redes são problemas envolvendo classificação de padrões e filtragem linear. Exemplos de redes com essa característica são o Perceptron e Adaline, que utilizam como algoritmo de aprendizado principal a regra de Hebb e a regra Delta (ambas abordadas na sequência) (SILVA; SPATTI; FLAUZINO, 2010).

Redes *feedforward* de camadas múltiplas se diferenciam das de camadas simples pela presença de camadas ocultas. Entre suas principais aplicações estão problemas de aproximação de funções, classificação de padrões, identificação de sistemas, otimização, robótica e controle de processos. As principais redes representantes desse grupo são o Perceptron Multicamadas ou *Multilayer Perceptron* (MLP) e as redes de base Radial (RBF) (SILVA; SPATTI; FLAUZINO, 2010; HAYKIN, 2008).

No caso das redes recorrentes ou realimentadas a característica que as diferencia é a realimentação do sinal de saída de uma camada para camadas anteriores quando processado o próximo sinal, ou seja, o sinal de saída da camada n para amostra t é dado como entrada para camada $n - 1$ para amostra $t + 1$. Entre as principais representantes desse grupo estão a rede Perceptron multicamadas com realimentação entre neurônios, a rede de Elman e a rede LSTM, cujas principais aplicações são previsão de séries temporais, otimização e identificação de sistemas e controle de processos (HAYKIN, 2008).

As redes com arquitetura em estrutura reticulada são caracterizadas pela relevância da disposição espacial dos neurônios, sendo assim a localização espacial dos neurônios está diretamente ligada com o processo de atualização dos pesos sinápticos. As principais representantes desse grupo são as redes de Kohonen e as redes de k-means, que possuem aplicações principalmente em problemas de agrupamento, reconhecimento de padrões, otimização de sistemas e grafos (SILVA; SPATTI; FLAUZINO, 2010).

3.2 TREINAMENTO E PROCESSO DE APRENDIZAGEM

Uma das características primordiais de uma rede neural é a sua habilidade de aprender e melhorar seu desempenho através da aprendizagem. De acordo com Mendel e McLaren (1970) a aprendizagem é o processo pelo qual os parâmetros, pesos, de uma rede neural são adaptados. Isso acontece através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela forma como ocorre a modificação dos parâmetros (HAYKIN, 2008).

A rede neural aprende sobre seu ambiente através do processo de ajuste de seus pesos sinápticos e bias, sendo que idealmente, a rede se torna mais instruída sobre seu

contexto a cada iteração do processo. É chamada de iteração a apresentação e processamento completo, dentro da rede, de uma amostra (HAYKIN, 2008).

Algoritmos de aprendizagem são o conjunto de regras ou passos definidos para solução de um problema de aprendizagem, a execução desse conjunto é o que consideramos o treinamento de uma rede neural (HAYKIN, 2008; SILVA; SPATTI; FLAUZINO, 2010). Assim a forma de treinamento de uma RNA é a forma como os pesos sinápticos são atualizados após a rede processar um sinal de entrada. De acordo com Fausett (1994) a forma de treinamento é uma das principais características que definem uma rede neural.

Esse processo de aprendizagem ou treinamento acontece dentro da rede seguindo a sequência: a rede é estimulada pelo ambiente, ou seja, recebe os dados de determinada amostra, acontecem modificação em seus parâmetros, e por fim a rede dá uma nova resposta ao ambiente, resposta esta resultante das modificações ocorridas em sua estrutura (HAYKIN, 2008).

Os processos de treinamento de uma rede neural podem ser divididos em dois grandes grupos: treinamento supervisionado e treinamento não supervisionado. No caso do treinamento supervisionado é como se houvesse um "professor" ensinando a resposta correta para rede, na prática o que acontece é que cada amostra apresentada para a rede é associada a uma saída (alvo) esperada que também é apresentada para a rede. Assim a rede ajusta seus pesos sinápticos com o objetivo de minimizar a diferença entre a resposta desejada e a resposta real da rede produzida pelo sinal de entrada (SILVA; SPATTI; FLAUZINO, 2010; HAYKIN, 2008).

Para os processos de treinamento não supervisionado a saída desejada não é pré-estabelecida ou apresentada para a rede. Assim, a própria RNA é responsável por se auto-organizar de modo a identificar os subconjuntos dos dados cujos elementos possuem características ou particularidades comuns (SILVA; SPATTI; FLAUZINO, 2010).

Neste trabalho foram aplicadas redes neurais com as duas formas de treinamento, a rede neural com treinamento não supervisionado foi utilizada para fazer o agrupamento dos dados, enquanto a rede neural de treinamento supervisionado foi utilizada para fazer a classificação de hidrometeoros. Optou-se por uma rede com treinamento supervisionado pois as classes desejadas são pré-estabelecidas, ou seja, tem-se a resposta desejada.

3.2.1 Regras de Aprendizagem

As regras de aprendizagem indicam a forma com que os pesos sinápticos serão atualizados ao longo das iterações, considerando como w_{ij} o peso sináptico da ligação do neurônio i com o neurônio j , a regra base para atualização dos pesos para a iteração $n + 1$ é dada pela equação (3.12). Entre as principais regras de aprendizagem estão a regra de Hebb, regra delta e a competitiva (HAYKIN, 2008).

$$w_{ij}(n + 1) = w_{ij}(n) + \Delta w_{ij}(n) \quad (3.12)$$

A regra de Hebb é a mais famosa e antiga regra de aprendizagem, idealizada por Hebb (1949) que se pautou na hipótese de que os pesos sinápticos das ligações entre dois neurônios

que estão simultaneamente ativos deve ser reforçada, assim o incremento da atualização dos pesos sinápticos é dado pela equação (3.13):

$$\Delta w_{ij}(n) = \alpha \cdot y_i(n) \cdot y_j(n) \quad (3.13)$$

onde y_i é o sinal de saída do neurônio i , y_j é o sinal de saída do neurônio j e α é uma constante positiva que simboliza a taxa de aprendizagem, ou seja, a velocidade de aprendizagem da rede (HAYKIN, 2008; T. RAUBER, 2005).

Já a regra delta, que é conhecida também como regra de aprendizagem por correção de erro, foi idealizada por Widrow e Hoff (1960) pensando na atualização dos pesos sinápticos baseada no erro produzido entre a saída alvo da rede e a saída real gerada pela mesma. Sendo assim a regra delta é aplicada somente para algoritmos de aprendizagem supervisionada, e o incremento de adaptação dos pesos sinápticos é dado pela equação (3.14):

$$\Delta w_{ij}(n) = \alpha \cdot e_j(n) \cdot y_i(n) \quad (3.14)$$

onde y_i é o sinal de saída do neurônio i , α é a taxa de aprendizagem e e_j é o erro dado pela diferença entre a saída real, y_j , do neurônio j e a saída alvo d_j , ou seja, $e_j = y_j - d_j$. Assim o incremento da adaptação dos pesos sinápticos que segue a regra delta também pode ser escrito como na equação (3.15) (HAYKIN, 2008; T. RAUBER, 2005).

$$\Delta w_{ij}(n) = \alpha(d_j(n) - y_j(n))y_i(n) \quad (3.15)$$

Na regra de aprendizagem competitiva, como o próprio nome já diz, os neurônios de saída da rede neural competem entre si para serem ativados. Nesta aprendizagem somente um único neurônio de saída pode estar ativo em determinada iteração, sendo assim essa regra de aprendizagem se mostra adequada para determinar características salientes que podem ser utilizadas para classificar um conjunto de padrões de entrada. Cada neurônio individualmente aprende a se especializar em agrupamentos de padrões semelhantes tornando-se assim detectores de características para classes diferentes de padrões de entrada. Nesta regra só há aprendizado, ou seja, adaptação de peso, para o neurônio ativo assim se um neurônio não responde a um padrão de entrada então não ocorrerá aprendizado naquele neurônio. O incremento do peso sináptico de acordo com a regra de aprendizagem competitiva é dado pela equação (3.16).

$$\Delta w_{ij}(n) = \begin{cases} \alpha(x_i(n) - w_{ij}(n)), & \text{se o neurônio } j \text{ vencer a competição} \\ 0, & \text{se o neurônio } j \text{ perder a competição} \end{cases} \quad (3.16)$$

onde α é o parâmetro taxa de aprendizagem e x_i é a entrada do neurônio i . Esta regra tem o efeito de mover o vetor peso \mathbf{w} do neurônio vencedor j em direção ao parâmetro de entrada \mathbf{x} (HAYKIN, 2008).

3.3 REDE K-MEANS

A rede k-means é um método de aprendizagem não supervisionada utilizada para análise de agrupamentos através da regra de aprendizagem competitiva. O algoritmo k-means

faz parte da família de análise de agrupamentos chamada de *hard clustering* que baseia-se no fato de que cada amostra, ou seja, vetor de entrada está associado a um único grupo (*cluster*). Assim dado o conjunto de vetores $X = \{x_1, x_2, \dots, x_n\}$ o problema de agrupamento *hard clustering* consiste na determinação do conjunto de grupos $G = \{G_1, G_2, \dots, G_k\}$ que satisfaçam as restrições da equação (3.17) (THEODORIDIS; KOUTROUMBAS, 2009).

$$\begin{cases} G_i \neq \emptyset & \forall i = 1, 2, \dots, k \\ \bigcup_{i=1}^k G_i = X \\ G_i \cap G_j = \emptyset & \forall i, j = 1, 2, \dots, k \text{ com } i \neq j \end{cases} \quad (3.17)$$

Os grupos são formados objetivando obter uma grande homogeneidade interna, entre elementos do mesmo grupo, e grande heterogeneidade externa, entre os grupos (DAMIAN, 2011). Para atingir esse objetivo é necessário dar atenção a três importantes pontos, a determinação do número de agrupamentos na solução final, pois a técnica de k-means exige a definição prévia do número de grupos e do posicionamento inicial de seus centros. Outro ponto importante é a medição de similaridade, ou seja, definir a medida que será utilizada para determinar o grau de correspondência entre os objetos ao longo do processo. Por fim temos a formação de agrupamentos como terceiro ponto de atenção, pois uma vez definida a similaridade entre os elementos os mesmos devem ser agrupados (JUNIOR et al., 2009).

Utiliza-se o conceito de centroides como protótipos representativos dos grupos, esses centroides são os centros dos grupos. Assim, a técnica inicia com a escolha do número de grupos e do posicionamento inicial dos centroides (FONSECA; BELTRAME, 2010). A medida de similaridade adotada para as redes k-means é a distância euclidiana (equação 3.18):

$$d(x_i, C_j) = \|x_i - c_j\| = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + \dots + (x_{im} - c_{jm})^2} \quad (3.18)$$

onde $(x_{i1}, x_{i2}, \dots, x_{im})$ são os valores associados a entrada x_i , que na abordagem de k-means representam as coordenadas do vetor x_i . Da mesma forma $(c_{j1}, c_{j2}, \dots, c_{jm})$ representa as coordenadas do centroide C_j do grupo j . Para efetuar o cálculo da distância todas as variáveis devem ter a mesma ordem de grandeza, por isso é necessário efetuar a normalização dos dados (DAMIAN, 2011).

Assim o objetivo da rede é minimizar a função objetivo:

$$\sum_{j=1}^k \sum_{x_i \in G_j} \|x_i^{(j)} - c_j\|^2 \quad (3.19)$$

onde C_j é o centroide do grupo j e $x_i^{(j)}$ é o dado i do grupo j , ou seja, o objetivo é minimizar a soma das distâncias entre os dados de um grupo e seu respectivo centro (OLIVEIRA, 2014).

Como citado anteriormente, a rede de k-means utiliza a regra de aprendizagem competitiva, assim o centroide, que representa o neurônio de saída, é deslocado na direção dos vetores de entrada que o ativaram, ou seja, é deslocado na direção das entradas que foram classificadas como sendo de seu grupo. Assim, o centroide se desloca em direção ao centro

de massa de seu grupo, que é o ponto minimizador da soma das distâncias de si aos demais pontos do grupo, dessa forma a atualização do centroide é dada pela equação (3.20):

$$C_j = \frac{1}{|G_j|} \sum_{x_i \in G_j} x_i^{(j)} \quad (3.20)$$

onde $|G_j|$ é a cardinalidade do grupo j , ou seja, a quantidade de elementos que pertencem a esse grupo. C_j é o centroide do grupo j e cada $x_i^{(j)}$ representa um dado pertencente ao grupo j (PALMA, 2018)

A seqüência de funcionamento do algoritmo k-means se dá pela seguinte forma:

- **passo 1:** inicializar a quantidade de centroides e suas coordenadas;
- **passo 2:** apresentar as entradas de dados x para a rede e executar os passos de 3 a 4 enquanto houver alterações significativas da posição dos centroides;
- **passo 3:** calcular a distância euclidiana (equação 3.18) da entrada em relação a cada um dos centroides e classificá-la como pertencente ao grupo do centroide mais próximo;
- **passo 4:** deslocar o centroide de cada grupo em direção ao centro de massa do mesmo (equação 3.20).

A FIGURA 19 ilustra esse processo de aprendizagem da rede k-means. Em (a) os centroides dos grupos, representados por \diamond , são inicializados aleatoriamente, em (b) é calculada a distância euclidiana entre os centroides e cada uma das amostras, e feito o agrupamento das mesmas no grupo com centroide mais próximo. Em (c) é realizada a atualização do centroide na direção do centro de massa do grupo formado. Esse processo se repete em (d) e (e) até chegar em uma configuração como em (f) onde não há mais alteração das amostras nos grupos e conseqüentemente seus centroides permanecem inalterados.

Uma das grandes vantagens da rede k-means é sua baixa complexidade computacional que é da ordem de $O(Nkq)$, onde N é o número de dados a serem processados, k é o número de grupos a serem formados e q é o número de iterações necessárias para a convergência (THEODORIDIS; KOUTROUMBAS, 2009). Entretanto, um ponto não muito positivo para a técnica é sua suscetibilidade a inicialização dos centroides, de modo que inicializações diferentes de centroides convergem para resultados diferentes, uma vez que o algoritmo pode convergir para um mínimo local do valor da função de critério dependendo da partição inicial escolhida (OLIVEIRA, 2014; FONSECA; BELTRAME, 2010).

Para contornar esse problemas Arthur e Vassilvitskii (2007) desenvolveram uma forma de escolher a inicialização dos centroides da rede k-means que promete encontrar um agrupamento melhor que o algoritmo padrão, esse método recebeu o nome de k-means++ e baseia a escolha dos centros dos grupos na probabilidade:

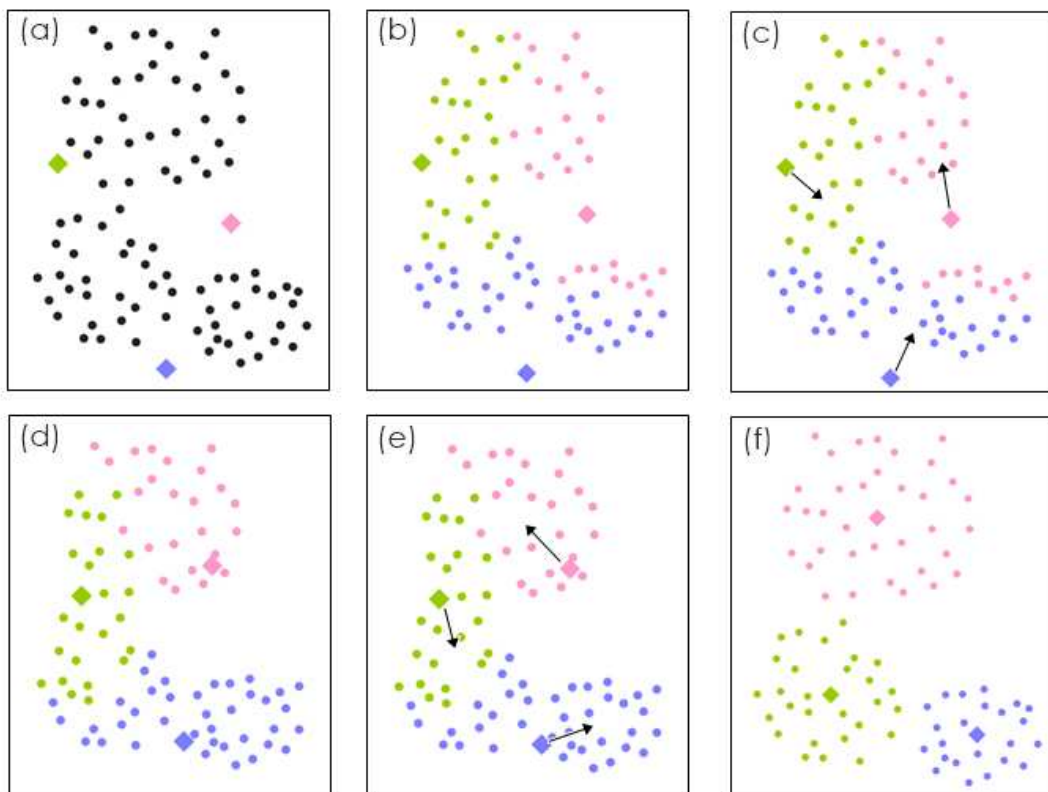
$$\frac{D(x')^2}{\sum_{x_i \in X} D(x)^2} \quad (3.21)$$

onde $D(x)$ é a distância do dado x ao centroide mais próximo já determinado e x' representa a amostra que está sendo considerada.

O algoritmo de inicialização dos centros por k-means++ é (ARTHUR; VASSILVITSKII, 2007):

- **passo 1:** inicialização aleatória do centroide C_1 tomado dentro do conjunto de dados X ;
- **passo 2:** determinação de um novo centroide $C_i = x' \in X$ de modo que x' é o dado com maior probabilidade (equação 3.21);
- **passo 3:** repetição do passo 2 até determinar os k centroides;
- **passo 4:** execução do algoritmo k-means padrão

FIGURA 19 – FUNCIONAMENTO DO ALGORITMO K-MEANS



FONTE: a autora.

De acordo com (THEODORIDIS; KOUTROUMBAS, 2009) algumas das principais aplicações de análise de agrupamentos como k-means são: redução de dados, onde a rede analisa uma grande quantidade de dados e realiza o agrupamento com o objetivo de processar cada grupo como uma única entidade. Outra aplicação é a geração de hipóteses, onde aplica-se a análise de agrupamentos para o conjunto de dados afim de inferir algumas hipóteses sobre a

natureza dos dados. Há também a aplicação para a previsão baseada em grupos, nesse caso os grupos resultantes são caracterizados pelos padrões formados, de modo que é possível caracterizar um padrão desconhecido ao determinar a qual grupo é mais provável que pertença esse novo dado.

Neste trabalho a rede k-means foi aplicada como pré-processamento de dados, com o objetivo de agrupar dados semelhantes de modo a potencializar o desempenho posterior da rede neural com aprendizagem supervisionada.

3.4 REDE PERCEPTRON MULTICAMADAS (MLP)

A rede perceptron multicamadas (*multilayer perceptron*), MLP, é uma generalização da rede perceptron proposta por Rosenblatt (1958) inspirada na retina humana. O que diferencia a rede MLP de sua precursora é a presença de pelo menos uma camada oculta em sua arquitetura.

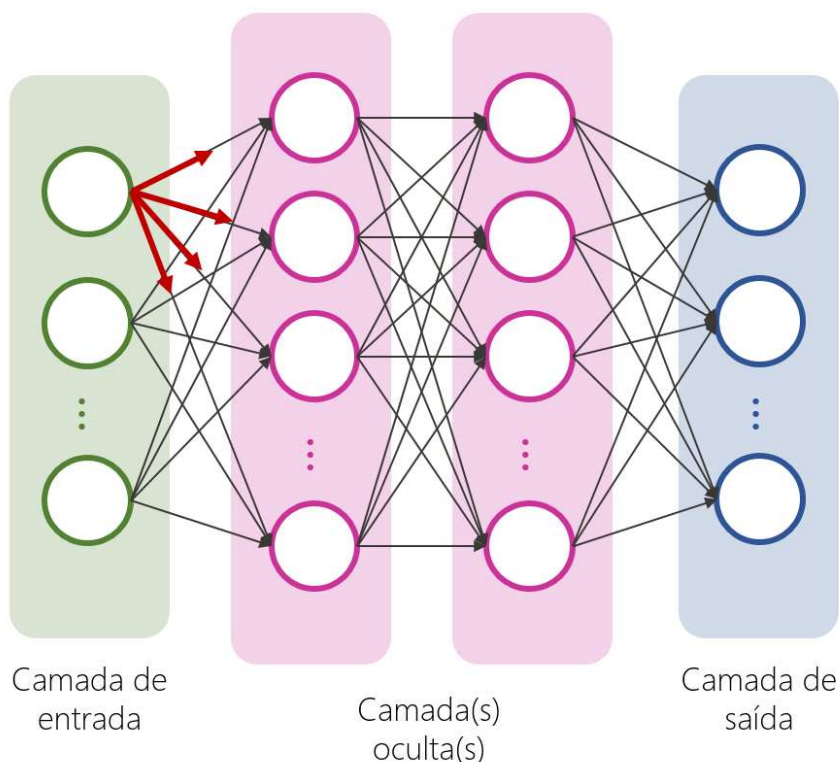
Trata-se assim de uma rede com arquitetura *feedforward* de camadas múltiplas, ou seja, o sinal é propagado dentro da rede no sentido unidirecional, da camada de entrada em direção a camada de saída (FIGURA 20). Com aprendizagem supervisionada, sua regra de aprendizagem é chamada de *backpropagation*, também conhecida como regra delta generalizada, que consiste no ajuste dos pesos sinápticos através da retropropagação do erro ao longo da rede (HAYKIN, 2008; SILVA; SPATTI; FLAUZINO, 2010).

O algoritmo de aprendizagem *backpropagation* consiste na aplicação sucessiva de duas fases, a fase de propagação do sinal a frente (*forward*) e a fase de retropropagação do erro (*backward*). Na primeira fase a rede recebe o vetor de entrada \mathbf{x} e propaga os sinais gerados pelos neurônios ao longo de suas camadas até produzir um sinal de saída. A partir daí começa a segunda fase, onde são comparados os sinais de saída produzidos pela rede com a saída desejada associada ao sinal de entrada em questão. O erro produzido pela diferença desses valores é retropropagado pela rede através de suas camadas e os pesos sinápticos são atualizados de modo que a resposta real da rede se mova para mais perto da resposta desejada (HAYKIN, 2008; SILVA; SPATTI; FLAUZINO, 2010).

Sendo assim a rede MLP tem dois tipos de sinais, que são os sinais funcionais, ou seja, os sinais de estímulo; sinais de entrada da rede, que propagam-se para a frente ao longo da rede na fase *forward*; e os sinais de erro que se originam nos neurônios de saída e se propagam para trás na fase *backward* (FIGURA 21) (HAYKIN, 2008).

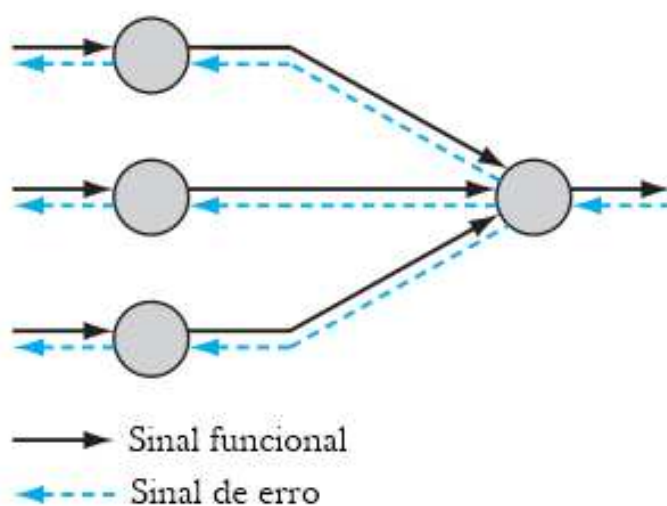
A especificação da configuração topológica de uma rede perceptron multicamadas, ou seja, a quantidade de camadas ocultas e a quantidade de neurônios em cada uma dessas camadas, é particular a cada aplicação, podendo assumir diferentes configurações, que

FIGURA 20 – ARQUITETURA DA REDE MLP



FONTE: a autora.

FIGURA 21 – SINAIS PROCESSADOS AO LONGO DE UMA REDE MLP



FONTE: adaptado de Haykin (2008)

dependem de diversos fatores. A quantidade de vetores na camada de saída da rede, por exemplo, depende da quantidade de saídas do processo que está sendo mapeado pela rede, de modo que se o processo possui n saídas, então a camada de saída da rede é composta por n neurônios (SILVA; SPATTI; FLAUZINO, 2010).

Para abordar os processos matemáticos realizados ao longo do algoritmo de aprendizagem da rede MLP considera-se a seguinte notação:

- $\mathbf{x}(n)$ vetor de entrada da n -ésima amostra de treinamento;
- $\mathbf{d}(n)$ vetor da saída esperada para a n -ésima amostra de treinamento;
- $\tau = \{\mathbf{x}(n), \mathbf{d}(n)\}_{n=1}^N$ conjunto de amostras utilizado para o treinamento da rede;
- j um neurônio da camada de saída da rede;
- $d_j(n)$ j -ésima componente do vetor $\mathbf{d}(n)$, ou seja, a saída desejada para o neurônio j ;
- $y_k(n)$ sinal funcional da saída do neurônio k gerado pelo sinal de estimulação da amostra n ;
- $e_j(n)$ sinal de erro na saída do neurônio j , para a amostra n ;
- $\xi(n)$ energia total do sinal de erro produzido pelo processamento da amostra n ;
- $\xi_{md}(N)$ energia média do erro produzido pelos sinais de erro do conjunto de amostras do treinamento;
- m_l o número de neurônios da camada l , com $l = 0, 1, \dots, L$. Assim m_0 representa o número de neurônios na camada de entrada e m_L , que também é denotado por M , representa o número de neurônios na camada de saída;
- φ_k função de ativação do neurônio k ;
- $u_k(n)$ potencial de ativação do neurônio k gerado pelo processamento do sinal da amostra n ;
- $w_{ik}(n)$ peso sináptico da conexão da saída do neurônio i com a entrada no neurônio k , para a amostra n ;
- $\Delta w_{ik}(n)$ incremento que representa a adaptação do peso sináptico $w_{ik}(n)$ associada ao processamento da amostra n ;
- b_k bias associado ao neurônio k , também referido como w_{0k} ;
- α constante positiva que representa a taxa de aprendizagem;
- $\delta_k(n)$ gradiente local associado ao neurônio k para a amostra n .

Para aplicação da regra de aprendizagem *backpropagation* é preciso definir a função representativa do erro de aproximação, que é responsável por comparar a resposta produzida pela rede com a resposta alvo. Em geral opta-se pelo método de Gauss-Newton (HAYKIN, 2008), então o algoritmo de aprendizagem da rede MLP que será abordado aqui segue tal método.

O sinal de erro gerado na saída do neurônio j para a amostra n é dado por:

$$e_j(n) = d_j(n) - y_j(n) \quad (3.22)$$

Pela abordagem do método Gauss-Newton o desempenho local associado aos sinais de erro produzidos a partir da amostra n é medido pela função erro quadrático, sendo assim, a energia total do sinal de erro associado a amostra n é dada pela equação (3.23):

$$\xi(n) = \frac{1}{2} \sum_{j=1}^M e_j^2(n) \quad (3.23)$$

Já o desempenho global da rede é medido pela função erro quadrático médio, sendo assim, é dado pela equação (3.24) (HAYKIN, 2008; SILVA; SPATTI; FLAUZINO, 2010).

$$\xi_{md}(N) = \frac{1}{2N} \sum_{n=1}^N \xi(n) \quad (3.24)$$

Ambas as energias de erro, $\xi(n)$ e $\xi_{md}(N)$, são essencialmente funções dos pesos sinápticos, uma vez que o sinal de saída do neurônio j é dado por:

$$y_j(n) = \varphi_j(u_j(n)) \quad (3.25)$$

sendo que $u_j(n)$ é o potencial de ativação do neurônio j gerado pela combinação linear dos seus sinais de entrada ponderados pelos pesos sinápticos. Como os sinais de entrada do neurônio j são os sinais de saída dos neurônios da sua camada anterior, o potencial de ativação do neurônio j é dado pela equação (3.26):

$$u_j(n) = \sum_{i=0}^m w_{ij}(n)y_i(n) \quad (3.26)$$

onde m é o número de neurônios da camada do neurônio i e o peso sináptico w_{0j} é o bias b_j com sinal de entrada associado fixo $y_0 = 1$ (HAYKIN, 2008).

A atualização dos pesos sinápticos é feita com o objetivo de minimizar o erro da rede, que pode ser feito seguindo dois modos: o modo sequencial, ou on-line, onde os pesos sinápticos são atualizados ao final de cada iteração, buscando minimizar o valor da função $\xi(n)$ (equação 3.23). O segundo modo é o aprendizado por lote, ou off-line, onde a adaptação dos pesos sinápticos acontece a cada época, ou seja, a cada vez que todas as amostras do conjunto τ são apresentadas a rede, assim a função a ser minimizada nesse caso é a $\xi_{md}(N)$ (equação 3.24) (HAYKIN, 2008). Aqui será apresentado o desenvolvimento da adaptação dos pesos sinápticos para o modo sequencial, porém o mesmo é análogo ao caso off-line, basta alterar a função erro considerada.

Como citado anteriormente a regra de aprendizagem utilizada é a regra delta generalizada, que também é conhecida como gradiente descendente, isso porque a atualização dos pesos sinápticos não é feita somente na intenção de minimizar a função erro, mas também de fazer isso no sentido mais eficiente possível. Sendo assim, como o gradiente da função indica sua direção de maior crescimento, os pesos são atualizados na direção oposta do gradiente

da função, ou seja, na direção de $-\nabla\xi(n)$ que indica a direção de maior decréscimo do erro (SILVA; SPATTI; FLAUZINO, 2010; HAYKIN, 2008).

Assim considerando um neurônio j da camada de saída da rede, o incremento de atualização do peso sináptico $w_{ij}(n)$ é dado por:

$$\Delta w_{ij}(n) = -\alpha \nabla \xi(n) \quad (3.27)$$

sendo que o gradiente da função erro $\xi(n)$ é obtido pela equação:

$$\nabla \xi(n) = \frac{\partial \xi(n)}{\partial w_{ij}(n)} \quad (3.28)$$

aplicando a regra da cadeia para a derivada parcial tem-se:

$$\frac{\partial \xi(n)}{\partial w_{ij}(n)} = \frac{\partial \xi(n)}{\partial e_j(n)} \cdot \frac{\partial e_j(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial u_j(n)} \cdot \frac{\partial u_j(n)}{\partial w_{ij}(n)} \quad (3.29)$$

Desenvolvendo cada uma das derivadas parciais individualmente temos: derivando a equação (3.23) obtemos a equação (3.30):

$$\frac{\partial \xi(n)}{\partial e_j(n)} = e_j(n) \quad (3.30)$$

Derivando a equação (3.22) temos:

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \quad (3.31)$$

Da equação (3.25) obtemos:

$$\frac{\partial y_j(n)}{\partial u_j(n)} = \varphi'_j(u_j(n)) \quad (3.32)$$

Por fim, derivando a equação (3.26) obtemos a equação (3.33):

$$\frac{\partial u_j(n)}{\partial w_{ij}(n)} = y_i(n) \quad (3.33)$$

substituindo os resultados das derivadas na equação (3.29) concluímos:

$$\frac{\partial \xi(n)}{\partial w_{ij}(n)} = -e_j(n) \cdot \varphi'_j(u_j(n)) \cdot y_i(n) \quad (3.34)$$

Logo o incremento de adaptação do peso sináptico que conecta a saída no neurônio i com a entrada do neurônio j , após o processamento da amostra n é dado pela equação (3.35):

$$\Delta w_{ij}(n) = \alpha e_j(n) \cdot \varphi'_j(u_j(n)) \cdot y_i(n) \quad (3.35)$$

Para simplificar a notação e generalização da adaptação dos pesos sinápticos, vários autores costumam definir o termo $\delta_k(n)$ como sendo o gradiente local dado pela equação:

$$\delta_k(n) = -\frac{\partial \xi(n)}{\partial u_k(n)} \quad (3.36)$$

assim para o neurônio da camada de saída temos

$$\delta_j(n) = -\frac{\partial \xi(n)}{\partial u_j(n)} = -\frac{\partial \xi(n)}{\partial e_j(n)} \cdot \frac{\partial e_j(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial u_j(n)} = e_j(n) \varphi'_j(u_j(n)) \quad (3.37)$$

portando a adaptação do peso sináptico $w_{ij}(n)$ em função do gradiente local é dado pela equação (3.38) (HAYKIN, 2008; T. RAUBER, 2005).

$$\Delta w_{ij}(n) = \alpha \delta_j(n) \cdot y_i(n) \quad (3.38)$$

Para um neurônio de uma camada oculta a equação (3.38) continua sendo válida, porém a aplicação da regra da cadeia para a derivada não resulta na equação (3.37), pois como não é um neurônio da camada de saída, não possui uma saída alvo d , ou seja, não está diretamente associado a um sinal de erro. Assim o sinal do erro para um neurônio de uma camada oculta deve ser determinado recursivamente, em termos dos sinais de erro de todos os neurônios aos quais o neurônio oculto está diretamente conectado.

Considerando então um neurônio i da última camada oculta, o sinal que cada um dos neurônios de saída recebem, em algum momento passou pelo neurônio i , assim cada um dos sinais de erro $e_j(n)$ resultantes do processamento da amostra n são dependentes do sinal de saída de y_i . O gradiente para atualização do peso w_{ki} será dado então por:

$$\frac{\partial \xi(n)}{\partial w_{ki}(n)} = \sum_{j=1}^M \left(e_j(n) \cdot \frac{\partial e_j(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial u_j(n)} \cdot \frac{\partial u_j(n)}{\partial y_i(n)} \cdot \frac{\partial y_i(n)}{\partial u_i(n)} \cdot \frac{\partial u_i(n)}{\partial w_{ik}(n)} \right) \quad (3.39)$$

substituindo as equações (3.31) e (3.32) na equação (3.39) temos:

$$\frac{\partial \xi(n)}{\partial w_{ki}(n)} = - \sum_{j=1}^M \left(e_j(n) \cdot \varphi'_j(u_j(n)) \frac{\partial u_j(n)}{\partial y_i(n)} \cdot \frac{\partial y_i(n)}{\partial u_i(n)} \cdot \frac{\partial u_i(n)}{\partial w_{ik}(n)} \right) \quad (3.40)$$

Além disso temos que:

$$\frac{\partial u_j(n)}{\partial y_i(n)} = w_{ij}(n) \quad (3.41)$$

$$\frac{\partial y_i(n)}{\partial u_i(n)} = \varphi'_i(u_i(n)) \quad (3.42)$$

$$\frac{\partial u_i(n)}{\partial w_{ik}(n)} = y_k(n) \quad (3.43)$$

substituindo as equações (3.37), (3.41), (3.42) e (3.43) na equação (3.40) temos:

$$\frac{\partial \xi(n)}{\partial w_{ki}(n)} = -y_k(n) \cdot \varphi'_i(u_i(n)) \sum_{j=1}^M \delta_j(n) \cdot w_{ij}(n) \quad (3.44)$$

Portando a adaptação do peso sináptico $w_{ki}(n)$ pode ser escrito como:

$$\Delta w_{ki}(n) = \alpha \delta_i(n) \cdot y_k(n) \quad (3.45)$$

onde $\delta_i(n)$ é o gradiente local, que a partir da equação (3.44) é dado pela equação (3.46).

$$\delta_i(n) = \varphi'_i(u_i(n)) \sum_{j=1}^M \delta_j(n) \cdot w_{ij}(n) \quad (3.46)$$

Podemos então descrever o algoritmo de aprendizagem da rede MLP de acordo com os passos:

- **passo 1:** inicialização do vetor de pesos sinápticos \mathbf{w} com valores aleatórios, e da taxa de aprendizagem α ;
- **passo 2:** realizar os passos de 3 a 7 para $n = 1, 2, \dots, N$;
- **passo 3:** apresentar a amostra n para a rede, determinando os sinais de saída de cada camada (equação 3.25) e as saídas reais produzida pela rede;
- **passo 4:** cálculo dos sinais de erro (equação 3.22) e da energia total do sinal de erro $\xi(n)$ (equação 3.23);
- **passo 5:** adaptação dos pesos sinápticos conectados a camada de saída (equações 3.38 e 3.37);
- **passo 6:** para cada camada oculta l , $l = L - 1, L - 2, \dots, 1$ fazer a adaptação de seus pesos sinápticos (equações 3.45 e 3.46);
- **passo 7:** acrescentar os resultados da amostra n ao cálculo da energia média do erro $\xi_{md}(N)$ (equação 3.24);
- **passo 8:** verificar o critério de parada. Se a diferença entre a energia média do erro das últimas duas épocas, ou seja, o erro global dos últimos dois processamento de todas as amostras, for menor que um valor ϵ parar, caso contrário voltar ao passo 2.

De acordo com Silva, Spatti e Flauzino (2010) as redes MLP's são caracterizadas pelas diversas possibilidades de aplicações, sendo considerada uma das arquiteturas mais versáteis quanto à aplicabilidade. Entre suas aplicações é possível destacar a aproximação universal de funções, reconhecimento de padrões, identificação e controle de processos, previsão de séries temporais e otimização de sistemas. Neste trabalho a rede perceptron multicamadas foi utilizada para realizar a classificação de hidrometeoros a partir do reconhecimento de padrões nos dados de entrada.

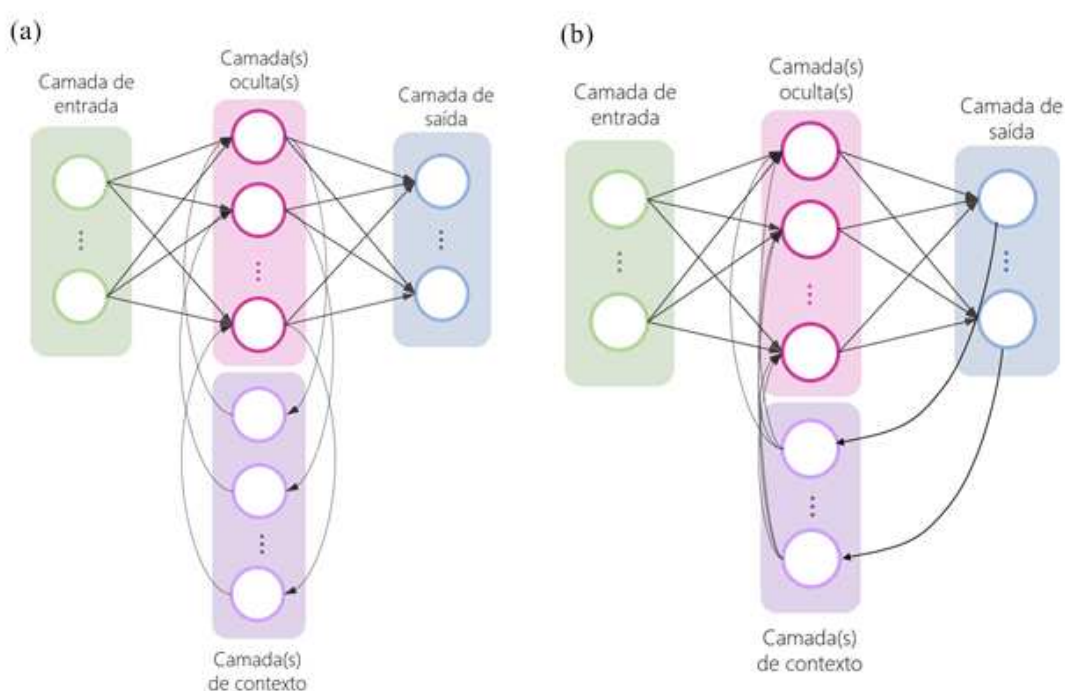
3.5 REDE LONG SHORT-TERM MEMORY (LSTM)

A rede Long Short-Term Memory, LSTM, é uma rede neural recorrente (RNN) proposta por Hochreiter e Jürgen Schmidhuber (1997). As redes neurais recorrentes são caracterizadas pela presença de camadas de contexto, que permitem conexões cíclicas, assim a rede é alimentada pela memória de saídas produzidas por estímulos anteriores recebidos pela rede. Uma das características que classifica uma RNN é a forma como a camada oculta está inserida na arquitetura da rede, alguns exemplos são: a rede neural de Jordan, proposta por Jordan (1986) em que a camada de contexto é alimentada pela camada de saída da rede (FIGURA 22). Outra famosa rede recorrente é a rede de Elman, proposta por Elman (1990) em que a camada de contexto é alimentada pela saída da(s) camada(s) oculta(s) (FIGURA 22) (GRAVES, 2012).

Assim as RNN's tem como principal benefício o uso de uma das capacidades do cérebro humano que é a memória. Porém arquiteturas de redes recorrentes como a rede de

Elman e a rede de Jordan possuem um problema, a influência de uma determinada entrada na camada oculta e, portanto, na saída da rede, pode decair ou explodir exponencialmente à medida que circula pelas conexões recorrentes da rede. Este efeito é frequentemente referido na literatura como o *"vanishing gradient problem"* ou problema do gradiente de fuga. A rede LSTM foi projetada justamente para contornar esse problema, utilizando os conceitos de memória de curto prazo e memória de longo prazo (HOCHREITER; Jürgen SCHMIDHUBER, 1997).

FIGURA 22 – REDE DE ELMAN E REDE DE JORDAN



FONTE: a autora.

LEGENDA: Arquitetura de redes neurais recorrentes: (a) rede de Elman e (b) rede de Jordan.

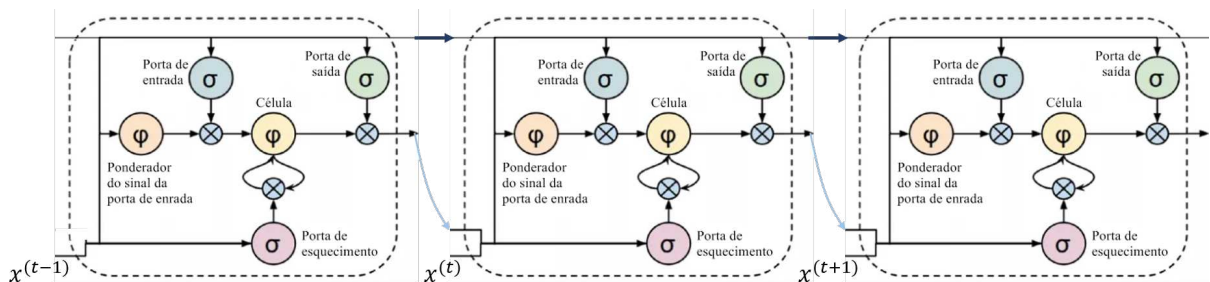
A arquitetura LSTM consiste em um conjunto de sub-redes conectadas recorrentemente (FIGURA 23), conhecidas como blocos de memória (FIGURA 24), esses blocos de memória podem ser considerados uma versão dos chips de memória de um computador digital. Cada bloco contém pelo menos uma célula de memória autoconectada e três unidades multiplicativas: as portas de entrada, saída e esquecimento. Tais unidades multiplicativas realizam o análogo de operações de gravação, leitura e redefinição (GRAVES, 2012).

Para entender o funcionamento de um bloco de memória da rede LSTM serão utilizadas as seguintes notações:

- w_{ij} peso entre as unidades i e j ;

- $u_j^{(t)}$ potencial de ativação da unidade j no tempo t ;
- $b_j^{(t)}$ saída da unidade j no tempo t ;
- l indica a porta de entrada;
- ϕ indica a porta de esquecimento;
- ω indica a porta de saída;
- c indica a célula;
- $s_c^{(t)}$ estado da célula c no tempo t ;
- x_i entradas do bloco de memória advindas do processamento do estímulo da amostra n ;
- h indica valores de contexto;
- N total de entradas recebidas pelo bloco de memória referentes a camada anterior;
- H total de neurônios na camada de contexto;
- C total de blocos de memórias na camada;
- $u_j^{(t)}$ potencial de ativação da unidade j no tempo t .

FIGURA 23 – DESDOBRAMENTO DO LOOP NO TEMPO DE UM BLOCO DE MEMÓRIA



FONTE: a autora.

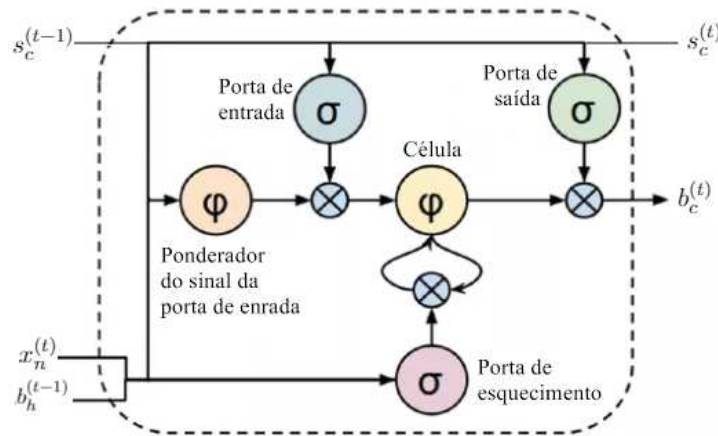
Quando um bloco de memória recebe um sinal de entrada referente a amostra n , ele também recebe o estado das células dos blocos de memória da sua camada no tempo $t - 1$, bem como as saídas das camadas de contexto (FIGURA 25). Além disso, o bloco processa esses sinais primeiramente na porta de esquecimento que determina o quanto esse sinal é relevante e deve ser lembrado ou esquecido pela célula, assim é determinado o potencial de ativação dessa unidade pela equação:

$$u_{\phi}^{(t)} = \sum_{i=1}^N w_{n\phi} x_n^{(t)} + \sum_{h=1}^H w_{h\phi} b_h^{(t-1)} + \sum_{c=1}^C w_{c\phi} s_c^{(t-1)} \quad (3.47)$$

esse potencial de ativação é o argumento da função de ativação da porta de esquecimento, que em geral é a função logística com $\beta = 1$ e assim produz a saída da porta de esquecimento que é dada pela equação (3.48) (GRAVES, 2012).

$$b_{\phi}^{(t)} = f(u_{\phi}^{(t)}) \quad (3.48)$$

FIGURA 24 – BLOCO DE MEMÓRIA LSTM



FONTE: a autora.

Esses mesmos sinais passam também pela porta de entrada, produzindo um potencial de ativação para a porta de entrada que é dado por:

$$u_l^{(t)} = \sum_{i=1}^N w_{ni} x_n^{(t)} + \sum_{h=1}^H w_{hl} b_h^{(t-1)} + \sum_{c=1}^C w_{cl} s_c^{(t-1)} \quad (3.49)$$

a função de ativação da porta de entrada mais usual também é função sigmoide, ou seja, a função logística com $\beta = 1$. Logo a saída da porta de entrada é dada pela equação (3.51) (GRAVES, 2012).

$$b_l^{(t)} = f(u_l^{(t)}) \quad (3.50)$$

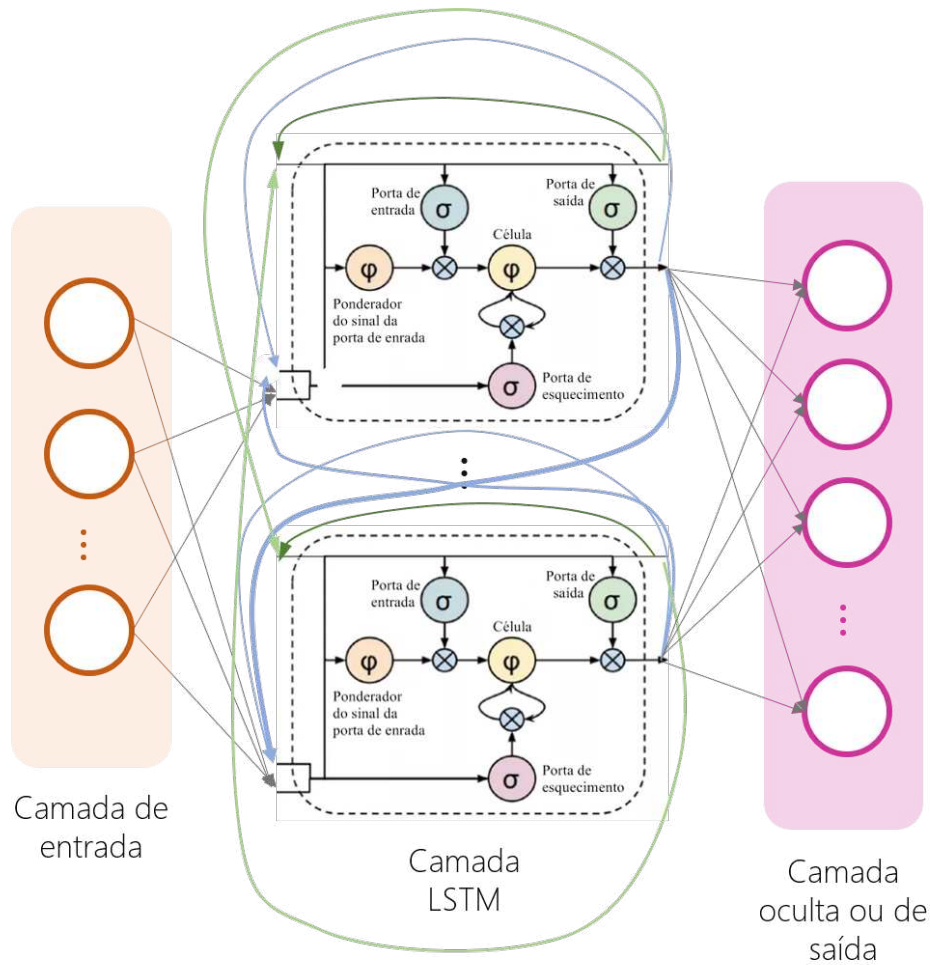
Ambos os sinais, o sinal de saída da porta de esquecimento e o sinal da porta de entrada são enviados para a célula do bloco de memória. Porém o sinal da porta de entrada é ponderado por uma variável que indica o quanto ele é importante, assim o estado da célula c no tempo t é dado por:

$$s_c^{(t)} = b_{\phi}^{(t)} \cdot s_c^{(t-1)} + b_l^{(t)} \cdot g(u_c^{(t)}) \quad (3.51)$$

onde g é a função de ativação que atua sobre o potencial da célula u_c , nesse caso a função mais usual é a tangente hiperbólica. O potencial de ativação da célula do bloco de memória, por sua vez, é dado pela equação (3.52) (GRAVES, 2012).

$$u_c^{(t)} = \sum_{i=1}^N w_{nc} x_n^{(t)} + \sum_{h=1}^H w_{hc} b_h^{(t-1)} \quad (3.52)$$

FIGURA 25 – EXEMPLO DAS CONEXÕES DE UMA REDE COM CAMADA LSTM



FONTE: a autora.

Por fim os sinais passam pelo portão de saída, cujo potencial de ativação é dado por:

$$u_{\omega}^{(t)} = \sum_{i=1}^N w_{n\omega} x_n^{(t)} + \sum_{h=1}^H w_{h\omega} b_h^{(t-1)} + \sum_{c=1}^C w_{c\omega} s_c^{(t)} \quad (3.53)$$

que é o argumento da função de ativação da porta da saída, que geralmente é também a função sigmoide. O valor da porta de saída é dado pela equação (3.54):

$$b_{\omega}^{(t)} = f(u_{\omega}^{(t)}) \quad (3.54)$$

porém essa saída ainda não é saída final do bloco de memória, a saída do bloco de memória é dada pela equação (3.55), onde φ é a função de ativação do bloco (GRAVES, 2012).

$$b_c^{(t)} = b_{\omega}^{(t)} \cdot \varphi(s_c^{(t)}) \quad (3.55)$$

O estado de ativação da célula do bloco s_c é associado a memória de longo prazo, já a saída do bloco b_c é associada a memória de curto prazo (HOCHREITER; Jürgen SCHMIDHUBER, 1997). As adaptações dos pesos w_{ij} são feitas pela regra de aprendizagem

backpropagation, sendo assim segue a mesma ideia abordada na seção anterior para a rede MLP.

Gers, J. Schmidhuber e Cummins (1999) aplicaram a rede LSTM para um problema de *benchmark*, ou seja, um problema de comparação. De acordo com os autores, a rede LSTM padrão superou outros algoritmos RNN na resolução do problema abordado. Eles ainda afirmam que uma rede LSTM pode resolver muitas tarefas não resolvidas por algoritmos de aprendizagem anteriores para redes neurais recorrentes.

Sundermeyer, Schluter e Ney (2012) utilizaram a rede LSTM para a modelagem de linguagem, no caso, modelagem do inglês e francês. Os experimentos realizados com a rede mostram melhorias de cerca de 8% em relação à perplexidade se comparado aos algoritmos de redes neurais recorrentes padrão. Além disso, os autores afirmam terem conquistado um sistema de reconhecimento de voz de última geração. Eles ainda apontam o principal fator que faz das redes recorrentes candidatas melhores que as redes *feedforward* para tarefas de modelagem de linguagem. De acordo com os autores, enquanto as redes *feedforward* exploram apenas um comprimento de contexto fixo para prever a próxima palavra de uma sequência, conceitualmente, as redes neurais recorrentes podem levar em consideração todas as palavras predecessoras.

Huang, Xu e Yu (2015) propuseram uma variedade de modelos baseados no modelo da rede LSTM para marcação de sequência, mais especificamente marcação de sequência de *benchmark* de programação neurolinguística. As técnicas aplicadas por eles foram, além da própria rede LSTM, redes LSTM bidirecionais, LSTM com uma camada de campo aleatório condicional e LSTM bidirecional com uma camada de campo aleatório condicional.

Greff et al. (2017) afirmam que nos últimos anos, as redes LSTM tornaram-se modelos de última geração para uma variedade de problemas de aprendizado de máquina. Sendo assim os autores desenvolveram uma análise em grande escala de oito variantes da rede LSTM em três tarefas: reconhecimento de fala, reconhecimento de caligrafia e modelagem polifônica de música. De acordo com os resultados obtidos os autores concluíram que nenhuma das variantes pode melhorar significativamente a arquitetura LSTM padrão e que a porta de esquecimento e a função de ativação de saída são seus componentes mais críticos.

A rede LSTM, assim como as redes recorrentes costumam ser bastante aplicadas para análise de séries temporais, previsão de palavras, composição musical, entre outras (GRAVES, 2012). Conforme citado anteriormente, esta rede vem se destacando nos contextos aplicados gerando resultados satisfatórios, sendo assim, neste trabalho ela foi aplicada para a classificação de hidrometeoros, assim como a rede MLP.

4 MÉTODO

Neste capítulo serão descritos os processos utilizados no desenvolvimento do trabalho, desde a seleção e tratamento dos dados até a execução das redes neurais.

Este trabalho foi desenvolvido utilizando a linguagem de programação Python, uma linguagem de altíssimo nível, orientada a objeto, interpretada e interativa (BORGES, 2010), sendo conhecida por ser versátil, possuir uma vasta biblioteca padrão e módulos desenvolvidos por terceiros, além de apresentar a vantagem de fácil aprendizagem (SELUZNIAK, 2016).

A linguagem Python foi desenvolvida nos anos 90 por Guido van Rossum, no Instituto Nacional de Pesquisa para Matemática e Ciência da Computação da Holanda (CWI) com foco em usuários como físicos e engenheiros. Por possuir uma sintaxe clara e concisa, que favorece a legibilidade do código fonte, tornando a linguagem mais produtiva, além de ser uma linguagem de código aberto, podendo ser utilizada em várias plataformas, tornou-se aceita na indústria por diversas empresas de alta tecnologia, tais como Google, Nokia, Nasa, Philips e Yahoo (BORGES, 2010).

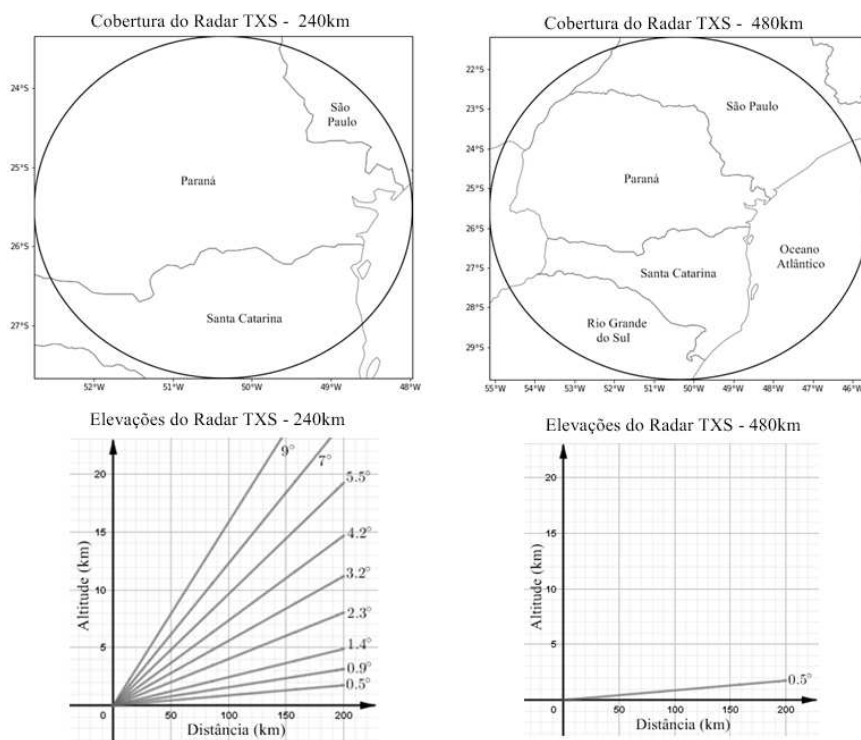
Outro ponto positivo desta linguagem é a possibilidade de integração com outras linguagens, como C, Fortran, Java. Com isso, é possível fazer o desenvolvimento de aplicações híbridas que unem código em Python e outra linguagem (SELUZNIAK, 2016).

4.1 DADOS

Os dados utilizados neste trabalho foram cedidos pelo SIMEPAR - Sistema Meteorológico do Paraná e são provenientes de três radares meteorológicos:

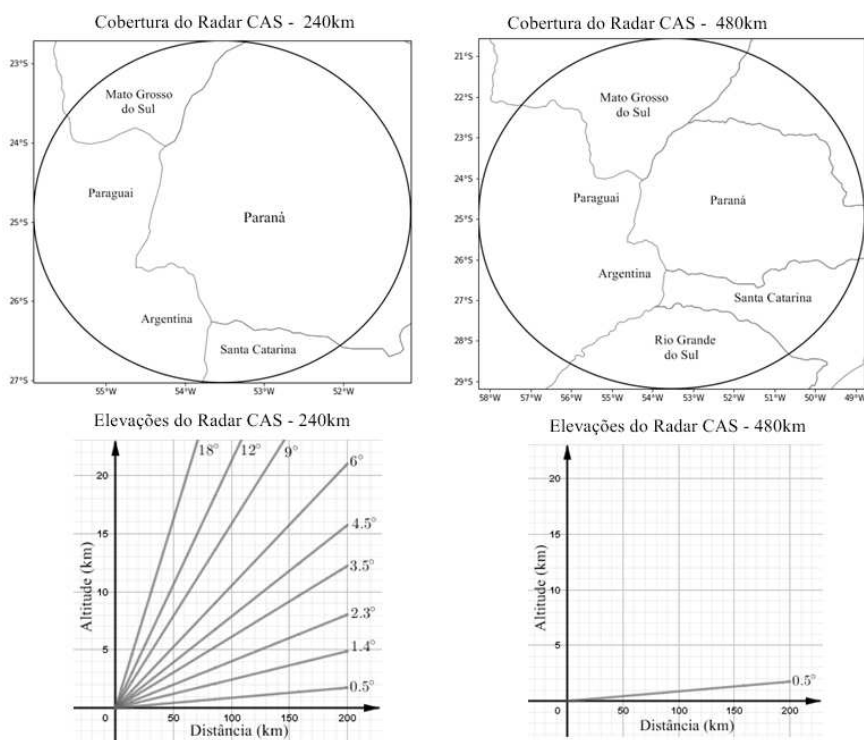
- **Radar de Teixeira Soares (TXS):** é um radar Banda-S de polarização simples que está instalado no município de Teixeira Soares no centro-sul do estado do Paraná. A cada 10 minutos este radar faz uma varredura com raio de 480 quilômetros com uma elevação e duas varreduras com raio de 240 quilômetros e 9 elevações (FIGURA 26).
- **Radar de Cascavel (CAS):** radar Banda-S de polarização dupla que está instalado no município de Cascavel no oeste do estado do Paraná. A cada 10 minutos este radar faz uma varredura com raio de 480 quilômetros com uma elevação e duas varreduras com raio de 240 quilômetros e 9 elevações (FIGURA 27).
- **Radar de Lontras (LON):** radar Banda-S de polarização dupla que está instalado no município de Lontras no alto vale do Itajaí no estado de Santa Catarina. A cada 4 minutos este radar faz uma varredura com raio de 240 quilômetros e 11 elevações (FIGURA 28).

FIGURA 26 – COBERTURA E ELEVAÇÕES DO RADAR DE TEIXEIRAS SOARES



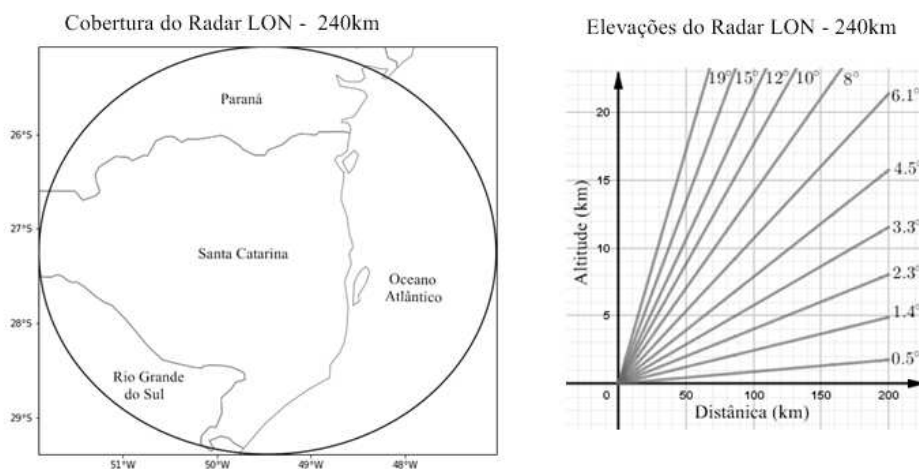
FONTE: a autora.

FIGURA 27 – COBERTURA E ELEVAÇÕES DO RADAR DE CASCAVEL



FONTE: a autora.

FIGURA 28 – COBERTURA E ELEVAÇÕES DO RADAR DE LONTRAS



FONTE: a autora.

Foram selecionadas sete datas para o estudo dos dados, esta seleção foi feita a partir da observação da ocorrência de hidrometeoros mais raros como granizo, sendo que o período considerado foi de 2018, onde este estudo começou a ser pensado, e março de 2019 quando começou-se a desenvolver este trabalho. Dentre os dados dessas datas foram selecionados aqueles com maior ocorrência de hidrometeoros dentro da área de cobertura dos radares, assim as datas e os horários dos dados utilizados estão na TABELA 2.

TABELA 2 – DATA E HORÁRIO DOS DADOS UTILIZADOS

Datas	Horários
07/fev/2018	18:05, 20:05 e 21:00
24/ago/2018	09:56, 10:50, 12:35, 14:00, 15:30, 16:21, 18:45, 19:50, 20:25, 21:39, 22:16, 23:05
04/set/2018	00:03
19/set/2018	16:45, 18:10, 21:45
04/out/2018	02:33, 19:05
24/out/2018	00:05, 02:45, 03:33, 21:52
02/fev/2019	05:25, 16:45, 17:33, 19:30, 20:15

FONTE: a autora

Para ler e tratar os dados fornecidos pelos radares foram utilizados pacotes como Py-ART e Numpy. Sendo o primeiro um pacote desenvolvido especificamente para dados de radares e o segundo um pacote matemático do Python.

4.1.1 Py-ART

O Python ARM Radar Toolkit, Py-ART, é um pacote para ler, visualizar, corrigir e analisar dados de radares meteorológicos que foi inicialmente desenvolvido para atender as necessidades do Atmospheric Radiation Measurement Climate Research Facility (ARM). Entretanto em 2013 o pacote foi lançado como software de código aberto e o escopo do projeto foi expandido para que possa ser utilizado por outras comunidades de pesquisas para examinar processos e analisar dados de diversos tipos de radares meteorológicos (HELMUS; COLLIS, 2016).

Desde o surgimento da tecnologia de radar meteorológico na década de 1940, as pesquisas têm buscado explorar todo o potencial das observações de radar meteorológico. As melhorias na tecnologia de radar estiveram intimamente ligadas aos avanços na ciência da computação e engenharia de software, de modo que a utilização de radares modernos não é possível sem software (HEISTERMANN et al., 2015).

O Py-ART está disponível no GitHub como uma base de código da comunidade e contém contribuições de desenvolvedores em universidades, programas governamentais e entusiastas de radar de todo o mundo. Além de várias bibliotecas Python científicas de código aberto como NumPy, SciPy e matplotlib, o Py-ART incorpora bibliotecas de código aberto específicas para radar, como RSL e o esquema de distribuição Doppler quadridimensional (HEISTERMANN et al., 2015; HELMUS; COLLIS, 2016).

Neste trabalho o Py-ART foi aplicado utilizando várias de suas funcionalidades, tais como a leitura dos dados de radar a partir dos arquivos no formato NetCDF e HDF5 através dos subpacotes *io* e *aux_io*. A visualização de dados foi feita no *Plan Position Indicator* (PPI), que é a forma básica de visualização das variáveis medidas pelo radar, uma vez que fornece a projeção no plano horizontal, dos dados obtidos através de uma varredura em azimute para um ângulo de elevação constante. Os PPI's foram gerados a partir da classe **RadarDisplay** do Py-ART.

Foi utilizado também o submódulo de mapa, **map**, para a criação das grades cartesianas regulares a partir dos dados dos três radares, pois as observações brutas dos radares estão em coordenadas esféricas. Porém é uma pré-condição para a integração das observações dos radares que as variáveis em coordenadas polares sejam projetadas para um sistema de referência cartesiano. Os pontos dessa grade são encontrados por interpolação de todas as portas de radar dentro de um raio de influência especificado usando a ponderação proposta por Cressman (1959) ou a proposta por Barnes (1964). Aqui foi aplicada a ponderação de Barnes por produzir melhores resultados e maior eficiência no contexto testado (HELMUS; COLLIS, 2016; HEISTERMANN et al., 2015).

4.1.2 NumPy

O NumPy traz o poder computacional de linguagens como C e Fortran para Python, de modo que quase todos os cientistas que trabalham com essa linguagem contam com o

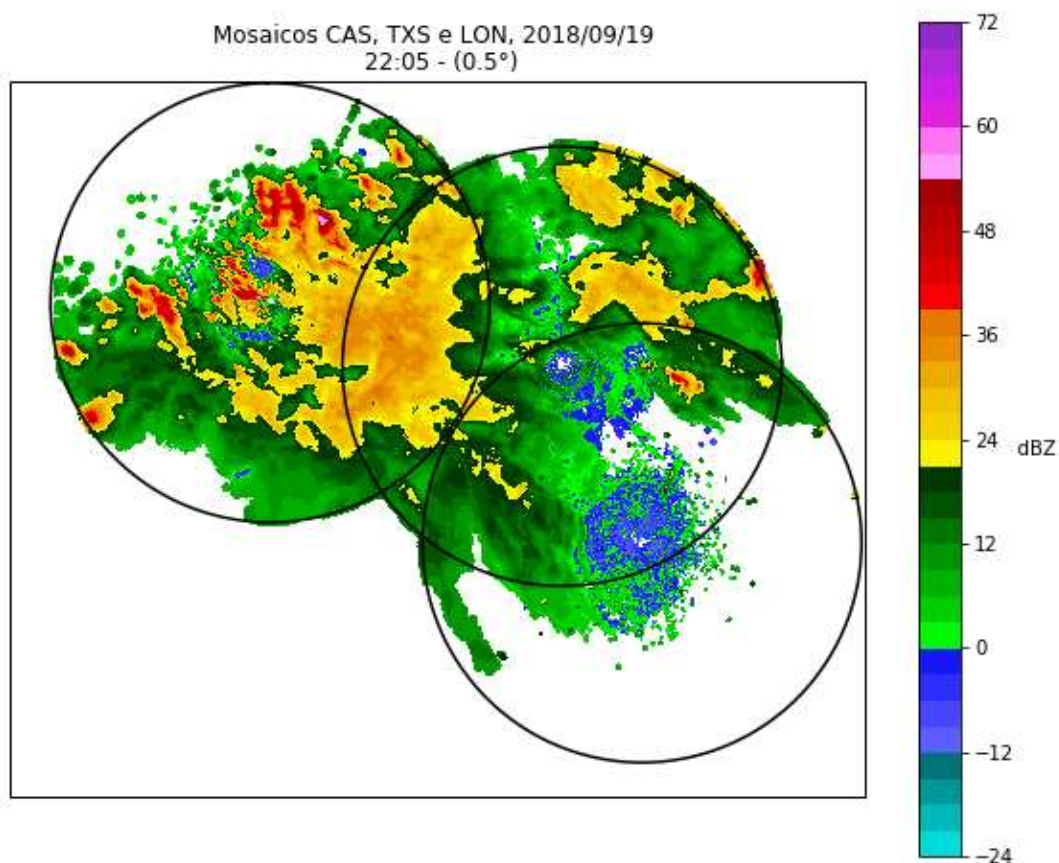
poder do NumPy. Além disso esse pacote contém uma grande variedade de módulos voltados para resolução de problemas numéricos de forma eficiente, sendo assim o NumPy se configura como o pacote fundamental para computação científica em Python (COELHO, 2007).

Numpy tornou-se o padrão para representação de dados numéricos em Python, pois possibilita a implementação eficiente de algoritmos de processamento numérico em uma linguagem de alto nível. Além de dispor de inúmeras funções para manipulação de vetores multidimensionais e rotinas de álgebra linear, o que proporciona algumas conveniências em relação à produtividade, reduzindo consideravelmente o esforço de programação (SELUZNIAK, 2016).

4.2 SELEÇÃO DOS DADOS PONTUAIS

O objetivo principal deste trabalho é fazer a classificação de hidrometeoros em regiões monitoradas por radares de polarização simples, sendo assim a área de nosso maior interesse é a área monitorada pelo radar de Teixeira Soares. Para a aplicação da técnica de RNA com aprendizagem supervisionada é necessário que os dados de entrada estejam associados a seu alvo, ou seja, a sua saída desejada, para isso foi feita a associação pontual dos dados da região comum do radar de Teixeira Soares com os radares de Cascavel e Lontras (FIGURA 29).

FIGURA 29 – MOSAICO DOS DADOS DOS RADARES CAS, TXS E LON



FONTE: a autora

Para tornar essa associação possível foi necessário passar os dados para uma grade de projeção cartesiana comum, conforme citado anteriormente, assim os dados passaram a ser associados a uma localização (x, y, z) no R^3 , ao invés da tupla original em coordenadas polares (r, θ, ϕ) .

Para cada dado pontual do radar de TXS, dentro das áreas comuns, foi associado um ponto dos radares de CAS ou LON, a associação foi feita considerando o ponto mais próximo disponível dentro de uma distância máxima de 500 metros. A partir da associação, a dupla de pontos passou a ser tratada como um único ponto. Dessa forma cada ponto considerado da área de TXS passou a ter uma classificação de hidrometeoros correspondente.

A seleção de dados foi finalizada eliminando os dados sem valores de variáveis polarimétricas associados, ou seja, pontos onde não há a presença de alvos meteorológicos. Após esse pré-processamento resultaram 10.003.865 dados para serem utilizados na aplicação das técnicas de RNA selecionados.

4.3 APLICAÇÃO DAS REDES NEURAIIS ARTIFICIAIS

Para realizar a classificação de hidrometeoros foi utilizada uma proposta de rede neural artificial híbrida, onde primeiramente os dados passaram por uma rede neural auto-organizável de aprendizagem competitiva não supervisionada, fazendo o agrupamento dos dados.

O objetivo é que esta primeira rede identifique semelhanças relevantes entre os elementos de cada grupo, de modo que o treinamento da segunda parte da rede, a rede com aprendizagem supervisionada seja mais eficiente.

A primeira rede foi utilizada a rede k-means, sua implementação foi realizada através da biblioteca Scikit-learn do Python, que oferece uma ferramenta simples e eficiente para a mineração e análise de dados. O módulo utilizado foi o `sklearn.cluster.KMeans()`, que agrupa os dados utilizando o método k-means.

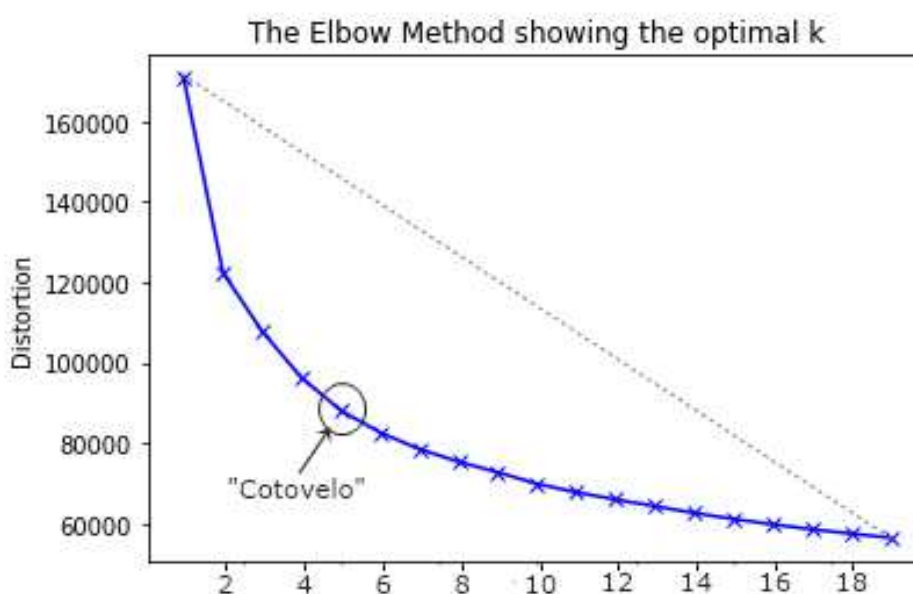
Os principais parâmetros de entrada para o módulo aplicado são, a matriz de dados, a quantidade de grupos desejados e a forma de inicialização dos centroides. A quantidade de grupos foi determinada através de variados testes e a construção da chamada curva do cotovelo, derivada do método *Elbow*.

O método *Elbow* testa a variância dos dados em relação ao número de grupos. A quantidade ideal de grupos é aquela que tem uma menor função objetivo, dada pela equação (3.19) e ao mesmo tempo o menor número de grupos (HAN; KAMBER; PEI, 2011). Essa configuração gera a situação ideal de maior homogeneidade dentro dos grupos e heterogeneidade entre os grupos. A identificação do número ideal de grupos acontece através do ponto do "cotovelo" na curva (FIGURA 30), sendo este o ponto da curva que está mais distante da reta que passa pelas extremidades do gráfico. Assim a quantidade ideal de grupos pelo método *Elbow* é 5, entretanto foram também feitos testes com 7 e 10 grupos para comparar os resultados.

Já no que diz respeito a inicialização dos centroides dos grupos, a mesma foi feita de forma aleatória e também através do algoritmo k-means++.

Para cada agrupamento formado foi treinada uma rede com aprendizagem supervisionada, as redes utilizadas nesta etapa foram a perceptron multicamadas e a rede LSTM. Modelos assim já foram propostos em trabalhos como Coadou e Benabdeslem (2006) para predição de séries temporais, por Alves, Araújo e Librantz (2006) para a classificação de imagens digitais e por Martins, Guimarães e Fonseca (2002) para a classificação de imagens geradas por sensoriamento remoto.

FIGURA 30 – CURVA MÉTODO ELBOW



FONTE: a autora

De fato um projeto envolvendo uma RNA necessita da definição de diversos parâmetros, tais como arquitetura da rede, algoritmo de treinamento, função de ativação, parâmetros de treinamento e critério de parada. Por isso foram realizados testes em diferentes cenários em relação a arquiteturas, parâmetros e funções de ativação.

Foram comparados cenários com o processamento dos dados por uma única rede de aprendizagem supervisionada, sem o agrupamento prévio, bem como o desempenho das redes MLP e LSTM.

Para aplicar as redes neurais artificiais foi utilizada a biblioteca Keras que utiliza o TensorFlow em seu backend.

4.3.1 TensorFlow e Keras

O TensorFlow é um framework desenvolvido originalmente pelos pesquisadores e engenheiros do projeto *Google Brain* dentro do grupo de Inteligência de Máquina do Google. Sendo utilizado para escrever e executar algoritmos de aprendizado profundo de máquina.

Seus cálculos são realizados através de grafo de fluxo de dados, onde as matrizes de dados multidimensionais são as arestas dos grafos e os nós do mesmo são as operações matemáticas, de modo que as matrizes são trocas entre os nós, que também executam os algoritmos escrito pelos usuários.

O TensorFlow é escrito como uma API em Python e interface em C++, assim é fortalecido sobre um motor C/C++, o que otimiza seu desempenho. Um dos seus benefícios é o suporte à computação distribuída com suporte a múltiplas placas de vídeo.

Já o Keras é uma biblioteca minimalista utilizada para construir redes neurais profundas, cuja principal motivação é oferecer uma prototipação rápida para chegar aos resultados o mais rápido possível. O resultado dessa motivação é uma biblioteca que oferece uma interface de mais alto nível se comparada a outras bibliotecas de aprendizado profundo.

4.4 VARIÁVEIS UTILIZADAS

Conforme abordado anteriormente, as variáveis polarimétricas geradas pelos radares de polarização dupla oferecem informações cruciais para a classificação de hidrometeoros, sendo assim elas seriam de grande contribuição para o treinamento de nossa rede neural. Porém nosso foco de estudo é justamente a região do radar de Teixeira Soares, que é um radar de polarização simples. Assim, esse radar não fornece os valores das variáveis polarimétricas.

Foram então testados dois cenários na aplicação das redes, o primeiro deles não aplicando as variáveis Z_{HV} , ρ_{HV} e K_{DP} . E um segundo cenário inferindo o valor dessas variáveis através de uma rede neural MLP e utilizando esse resultado como entrada para realização do agrupamento e do treinamento da rede de classificação dos hidrometeoros.

As variáveis aplicadas em todos os cenários foram a localização cartesiana do ponto, a refletividade horizontal e a temperatura do ponto em função de sua altitude. Para associar a temperatura ao dado foi considerado somente sua altitude, uma vez que a temperatura foi obtida através da interpolação dos dados da estação de radiossonda.

Como a distribuição de estações de radiossondagem no Brasil é muito esparsa, optou-se por selecionar uma única estação para avaliar o perfil de temperatura na área de cobertura do radar, sendo o radar referencial o de TXS, por ser o foco do nosso estudo. Assim foi selecionada a estação de radiossondagem de Curitiba (código 83840 da Organização Meteorológica Mundial - OMM). Foram utilizados os dados das 00 UTC e 12 UTC disponíveis virtualmente, na página do Departamento de Ciências Atmosféricas da Universidade de Wyoming, nos Estados Unidos (<http://weather.uwyo.edu/upperair/sounding.html>).

As variáveis que compõem os dados são de grandezas diferentes e algumas assumem

uma variação de valores de grande amplitude, assim os dados foram normalizados para contribuir com a convergência da rede. A normalização foi feita considerando a função de ativação a ser utilizada, assim para aplicação da função de ativação sigmoide os dados foram normalizados para o intervalo $[0, 1]$ através da equação (4.1):

$$\hat{x}_{ij} = \frac{x_{ij}}{\max\{x_i\}} \quad (4.1)$$

onde \hat{x}_{ij} é o valor normalizado da entrada j para a variável i , x_{ij} é o valor original da entrada j da variável i e $\max\{x_i\}$ é o valor máximo assumido pela variável i .

Para utilizar a função tangente hiperbólica como função de ativação os dados foram normalizados para o intervalo $[-1, 1]$ utilizando a equação (4.2):

$$\hat{x}_{ij} = 2 \left(\frac{x_{ij} - \min\{x_i\}}{\max\{x_i\} - \min\{x_i\}} \right) - 1 \quad (4.2)$$

onde \hat{x}_{ij} é o valor normalizado da entrada j para a variável i , x_{ij} é o valor original da entrada j da variável i , $\max\{x_i\}$ é o valor máximo assumido pela variável i e $\min\{x_i\}$ é o valor mínimo da variável i .

5 RESULTADOS

O computador utilizado para o processamento de dados, o treinamento e execução das redes neurais possui as seguintes características:

- Sistema operacional: Ubuntu 20.04.2 LTS x86z_64.
- Versão do núcleo do sistema: 5.4.0-70-generic.
- Shell: bash 5.0.17.
- CPU: Intel i5-8400 (6) @ 4.000GHz.
- GPU: Intel UHD Graphics 630.
- Memória RAM: 16GB.

Foram testados diversos cenários para a classificação de hidrometeoros através das redes neurais, como por exemplo, treinamento supervisionado por agrupamento, treinamento supervisionado único, treinamento com a rede MLP e com a rede LSTM, alterando a arquitetura das redes, o algoritmo de treinamento e os parâmetros de treinamento.

Aqui serão apresentados os resultados obtidos nos testes, sendo que as métricas utilizadas para verificação do desempenho das técnicas, no conjunto de treinamento e de teste, são o erro quadrado médio e a acurácia. A acurácia representa a porcentagem dos dados que a rede está classificando corretamente, é considerada uma classificação correta quando a rede retorna a mesma classificação de hidrometeoro indicada pelo software dos radares de Cascavel ou Lontras.

Um dos parâmetros alterados para testes foi o *batch_size*, que é aplicado para treinamento de redes neurais utilizando lote, ou seja, os pesos sinápticos não são adaptados após cada sinal recebido, mas sim após um lote de sinais de entrada processados. O parâmetro *batch_size* indica o tamanho desses lotes, e influencia significativamente no tempo de processamento do treinamento, principalmente quando se tem um conjunto de dezenas de milhares de dados.

A proporção de dados considerada para o treinamento e teste da rede foram respectivamente 70% para o treinamento e 30% para teste, sendo que nos casos em que foi realizado o agrupamento dos dados, todo o conjunto de dados foi apresentado e agrupado, e então foi considerada a proporção 80% e 20% para o treinamento e teste da rede de cada agrupamento.

Como padrão em todos os testes foi utilizada a camada de saída com 11 neurônios, um para cada classe de hidrometeoros. E como função de ativação dessa camada a função *softmax* que retorna para cada neurônio a probabilidade da amostra analisada pertencer a classe representada pelo mesmo, escolhendo-se a classe com maior probabilidade.

5.1 REDE PERCEPTRON MULTICAMADAS ÚNICA

Para o cenário de aplicação direta de apenas uma rede MLP, sem o agrupamento prévio dos dados, foi fixado o número de épocas de treinamento em 100, pois mais épocas resultaram em um tempo de treinamento elevado porém sem ganhos significativos em melhora de erro ou acurácia. Foram testadas diversas arquiteturas, as mais relevantes com os respectivos resultados estão descritas na TABELA 3.

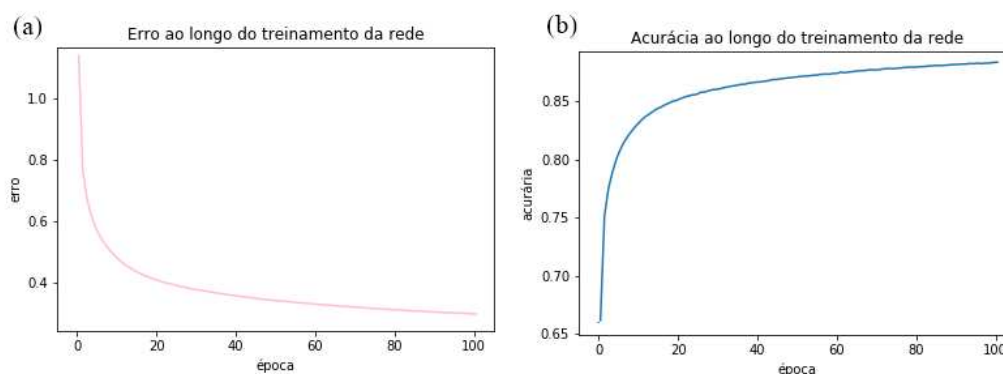
TABELA 3 – RESULTADOS PARA DIFERENTES ARQUITETURAS DA REDE MLP PARA TODOS OS DADOS

Camadas ocultas	Neurônios nas camadas	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
2	100, 100	0.6768	73.26%	1.8807	65.93%
2	10, 10	1.0227	58.2%	1.5551	36.94%
3	100, 100, 100	0.285	86.25%	1.0764	76.39%
3	300, 100, 200	0.3798	77.96%	1.2814	73.12%
3	30, 10, 20	0.9439	61.67%	1.3981	43.77%
3	10, 10, 10	0.9866	60.01%	1.4269	40.92%
4	100, 100, 100, 100	0.344	85.61%	1.1885	70.13%

FONTE: a autora

Assim a arquitetura que apresentou melhores resultados foi a arquitetura com 3 camadas ocultas, cada uma com 100 neurônios. As curvas do erro e da acurácia ao longo do treinamento para essa configuração são apresentados na FIGURA 31. Como essa foi a arquitetura que apresentou melhores resultados, os demais cenários foram testados fixando essa arquitetura.

FIGURA 31 – ERRO E ACURÁCIA PARA ARQUITETURA (100,100,100) DA REDE MLP



FONTE: a autora

LEGENDA: (a) gráfico da evolução do erro ao longo do treinamento (b) gráfico da evolução da acurácia ao longo do treinamento

Para o *batch_size*, que indica o tamanho dos lotes de treinamento foram realizados também alguns testes, conforme mostra a TABELA 4, sendo consideradas 100 épocas em todos os casos. A maior diferença está no tempo necessário para o treinamento da rede. Para *batch_size* = 100 o tempo de treinamento da rede foi de aproximadamente 3 horas e 28 minutos, já para *batch_size* = 500, por exemplo, o tempo de treinamento para as mesmas configurações do anterior, foi de aproximadamente 40 minutos. Vale lembrar que o total de dados estudados é aproximadamente de 10 milhões, sendo assim um treinamento por lote de 500 dados indica que os pesos sinápticos serão adaptados cerca de 20 mil vezes por época, e para lotes de 100 dados terão aproximadamente 100 mil atualizações de pesos sinápticos por época.

Neste caso para decidir qual a configuração mais adequada é preciso observar não somente os resultados de erro e acurácia mas também o tempo necessário para isso. Como observa-se na TABELA 4 o melhor resultado em termos de erro e acurácia foi obtido para um *batch_size* de 200, entretanto seu tempo de processamento foi 23 minutos mais longo que a configuração de *batch_size* = 500. Assim, considerando que a diferença de erro e acurácia entre essas configurações não é grande, considerou-se para os demais testes lotes de treinamento de 500 dados cada.

TABELA 4 – RESULTADOS PARA DIFERENTES TAMANHOS DE LOTES DE TREINAMENTO DA REDE MLP

<i>batch_size</i>	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste	Tempo
100	0.2716	86.4%	1.018	77.36%	3h28min
200	0.2596	86.62%	1.1917	76.59%	1h3min
500	0.2752	86.25%	1.0764	76.64%	0h40min
1000	0.284	86.08%	1.0525	77.04%	0h29min

FONTE: a autora

A taxa de aprendizado α de uma rede neural indica a velocidade com que a rede aprende, neste caso indica a velocidade com que os pesos sinápticos se deslocam na direção de maior decrescimento da função erro. Foram realizados testes para diferentes valores de α , cujos resultados mais relevantes estão descritos na TABELA 5

Valores extremos, ou seja, muito pequenos ou muito grandes, para a taxa de aprendizagem da rede não geram bons resultados no contexto aplicado. Sendo que os valores que geraram melhores resultados foram obtidos para $0.1 \leq \alpha \leq 0.8$, decidiu-se então adotar $\alpha = 0.5$ para o restante dos testes.

Por fim foram testadas diferentes funções de ativação para os neurônios, as funções de ativação foram alteradas para as camadas, porém todos os neurônios de uma mesma camada aplicam a mesma função de ativação. Conforme citado anteriormente a função de ativação na camada de saída foi fixada como sendo a função softmax, uma vez que a mesma gerou acurácia cerca de 35% maior que a função sigmoide ou tangente hiperbólica.

TABELA 5 – RESULTADOS PARA DIFERENTES TAXAS DE APRENDIZADO DA REDE MLP

α	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
0.01	0.3668	84.72%	1.146	73.96%
0.1	0.3246	86.02%	1.1269	74.16%
0.5	0.3166	86.30%	1.1300	74.56%
0.8	0.3192	86.13%	1.1580	74.94%
1.5	0.3527	84.79%	1.1862	73.70%
5	1.2806	44.66%	1.9210	23.39%

FONTE: a autora

Os melhores resultados obtidos, e a configuração correspondente, estão descritos na TABELA 6, para arquitetura (100,100,100), ou seja, 3 camadas ocultas com 100 neurônios cada.

TABELA 6 – RESULTADOS PARA DIFERENTES FUNÇÕES DE ATIVAÇÃO DA REDE MLP

φ	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
reLU, reLU, reLU	0.3290	85.83%	1.1744	72.56%
tanh, tanh, tanh	0.3140	86.59%	1.0924	74.21%
reLU, tanh, tanh	0.3149	86.49%	1.0784	71.46%
tanh, reLU, reLU	0.3410	85.27%	1.0546	72.59%
sigmoid, sigmoid, sigmoid	0.3728	84.04%	1.1316	73.11%
reLU, tanh, sigmoid	0.3202	86.29%	1.0507	74.46%

FONTE: a autora

NOTA: tanh representa a função de ativação tangente hiperbólica. E sigmoid representa a função logística com $\beta = 1$

As funções de ativação com suas devidas normalizações produzem resultados semelhantes, sendo que os dois cenários com melhores resultados foram (tanh, tanh, tanh) e (reLU, tanh, sigmoid). Assim optou-se por utilizar a função de ativação tangente hiperbólica para os demais testes realizados.

5.2 REDE LSTM ÚNICA

A rede LSTM tem como característica própria a aplicação de determinadas funções de ativação para suas portas, as quais foram mantidas nos testes realizados. Assim a função de ativação para a porta de esquecimento, a porta de entrada e a porta de saída foi a função sigmoide, ou seja, a função logística com $\beta = 1$. Para ponderar o sinal de entrada enviado a célula do bloco foi aplicada a função de ativação tangente hiperbólica.

Esta rede exige mais poder computacional assim como tempo de processamento, de modo que para serem executadas 100 épocas de treinamento da rede o tempo necessário foi de pelo menos 5 horas e 16 minutos para a configuração mais rápida testada.

Os testes de arquitetura que geraram os melhores resultados estão descritos na TABELA 7, onde as camadas com blocos de memória LSTM foram conectadas a camadas de neurônios da rede perceptron multicamadas. A camada de saída foi mantida com 11 neurônios com função de ativação softmax. Já a função de ativação da camada de neurônios da rede MLP aplicada foi a tangente hiperbólica, levando em consideração os resultados obtidos para a própria rede MLP.

TABELA 7 – RESULTADOS PARA DIFERENTES ARQUITETURAS DA REDE LSTM PARA TODOS OS DADOS

Camadas ocultas	Tipo de camada	Neurônios nas camadas	Tempo treinamento
2	LSTM, Dense	50, 50	5h16min
2	LSTM, Dense	100, 50	9h45min
2	LSTM, LSTM	50, 25	7h16min
3	LSTM, LSTM, Dense	50, 25, 50	7h47min
Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
0.9681	51.42%	1.5460	37.15%
0.9697	51.38%	1.5540	37.46%
0.9691	51.41%	1.5585	37.15%
0.9680	51.51%	1.5497	37.47%

FONTE: a autora

NOTA: LSTM indica uma camada de blocos LSTM e dense indica uma camada de neurônios densamente conectados, ou seja, camadas como as que formar uma rede MLP

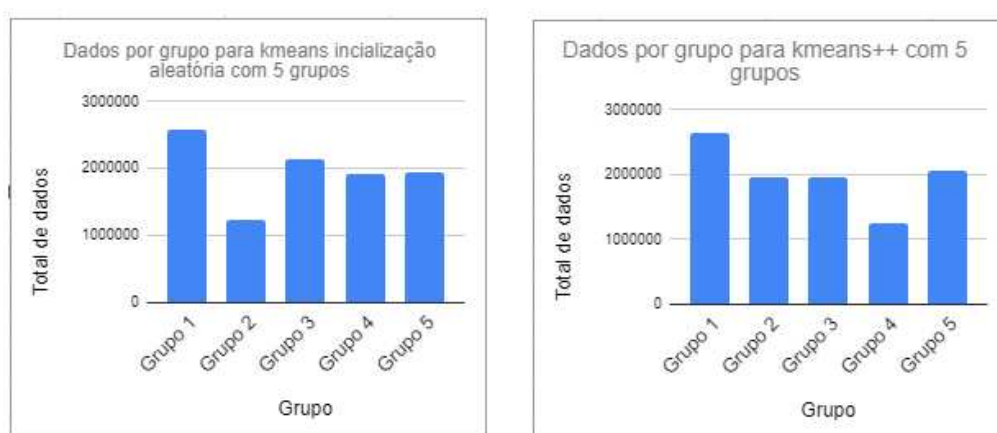
As alterações na arquitetura da rede LSTM não geraram uma variação grande em seus resultados, assim a rede LSTM se mostrou menos sensível a alteração de arquitetura do que a rede MLP (TABELA 3). Em todos os cenários testados para a rede LSTM única, mesmo as melhores resultados ficaram distantes do desempenho da rede MLP, que se mostrou mais eficiente para a classificação de hidrometeoros dos dados não agrupados.

Uma justificativa para esse fato foi apresentada pelos próprios Hochreiter e Jürgen Schmidhuber (1997) que afirmaram existir tarefas que podem ser resolvidas rapidamente por adaptação de pesos sinápticos sem recorrência de sinal, mas não por algoritmo LSTM truncado com inicializações de peso pequeno.

5.3 REDE HÍBRIDA: K-MEANS + MLP

Foram realizados testes aplicando a rede k-means para gerar 5, 7 e 10 agrupamentos de dados, utilizando para isso a inicialização dos centroides através do algoritmo k-means++ e também de forma aleatória. As FIGURA 32, FIGURA 33 e FIGURA 34 apresentam a quantidade resultante de dados em cada agrupamento. Já a FIGURA 35 apresenta a distribuição da classificação de hidrometeoros dentro de cada agrupamento, considerando o cenário com 5 agrupamentos.

FIGURA 32 – QUANTIDADE DE DADOS POR GRUPO PARA 5 AGRUPAMENTOS



FONTE: a autora

FIGURA 33 – QUANTIDADE DE DADOS POR GRUPO PARA 7 AGRUPAMENTOS



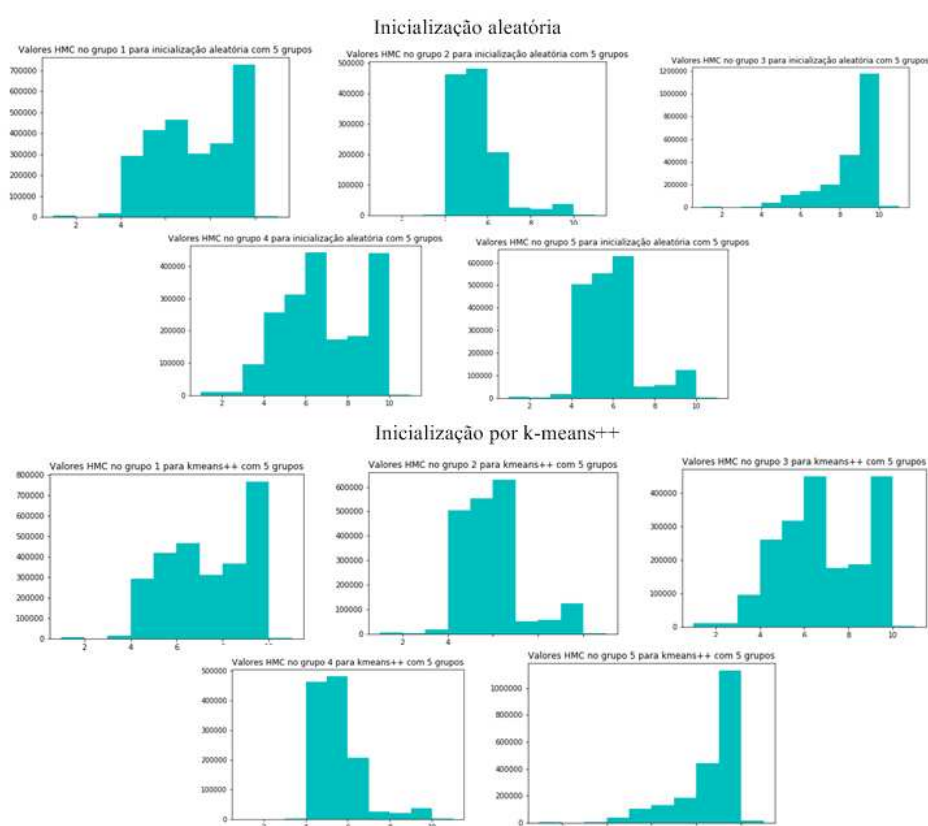
FONTE: a autora

FIGURA 34 – QUANTIDADE DE DADOS POR GRUPO PARA 10 AGRUPAMENTOS



FONTE: a autora

FIGURA 35 – DISTRIBUIÇÃO DOS DADOS DENTRO DOS 5 AGRUPAMENTOS EM RELAÇÃO A CLASSE DE HIDROMETEOROS QUE PERTENCEM



FONTE: a autora

Há dois pontos a serem ressaltados, primeiramente, independente da forma de inicialização dos centroides, os agrupamentos resultantes se assemelham em quantidade e distribuição de dados, considerando a característica foco do estudo. Segundo a rede k-means por si só não faz a classificação de hidrometeoros, uma vez que os agrupamentos formados apresentam dados heterogêneos no que diz respeito a classificação de hidrometeoros.

Os testes realizados aqui consideraram o cenário de melhor desempenho da rede

MLP quando aplicada independentemente, assim considerou-se a arquitetura com 3 camadas escondidas com 100 neurônios cada e função de ativação tangente hiperbólica, para a camada de saída com 11 neurônios com função de ativação softmax e a taxa de aprendizagem aplicada foi $\alpha = 0.5$. Como cada rede treinada nesse cenário recebeu uma quantidade menor de dados do que no caso em que não houve agrupamento prévio, foi aplicado um treinamento por lote de 100 dados, ao invés de 500, e a proporção de dados para treinamento e dados para teste foi de 80% para 20%.

Considerando o cenário com 5 agrupamentos, pela distribuição dos dados nos grupos nota-se a correspondência entre os grupos gerados pelo algoritmo k-means++ e os grupos gerados pela inicialização aleatória dos centros, de modo que o grupo 1 gerado pela primeira técnica corresponde ao grupo 1 gerado pela segunda técnica, o grupo 2 ao grupo 5, o grupo 3 ao grupo 4, o grupo 4 ao grupo 2 e o grupo 5 ao grupo 3.

Os resultados obtidos pelas redes MLP para os agrupamentos gerados pelo algoritmo k-means++ estão descritos na TABELA 8. Os resultados das redes MLP treinadas a partir dos agrupamentos resultantes da rede k-means com inicialização aleatória dos centroides estão apresentados na TABELA 9.

TABELA 8 – RESULTADOS DAS REDES MLP PARA 5 AGRUPAMENTOS COM K-MEANS++

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
grupo 1	0.3085	95.55%	0.9909	84.95%
grupo 2	0.3337	81.28%	1.2098	69.76%
grupo 3	0.3821	75.48%	1.1623	62.30%
grupo 4	0.3444	91.99%	0.9676	70.28%
grupo 5	0.3079	98.06%	0.8137	83.05%

FONTE: a autora

TABELA 9 – RESULTADOS DAS REDES MLP PARA 5 AGRUPAMENTOS COM INICIALIZAÇÃO ALEATÓRIA DE CENTROIDES

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
grupo 1	0.309	95.42%	0.977	84.94%
grupo 2	0.3141	93.53%	0.9892	71.06%
grupo 3	0.3041	97.06%	0.8578	82.48%
grupo 4	0.388	75.53%	1.1739	62.50%
grupo 5	0.3218	81.84%	1.1895	68.14%

FONTE: a autora

Portanto os agrupamentos correspondentes (correspondência indicada anteriormente) não apenas são semelhantes como também resultam em redes MLP com desempenho idêntico.

Em decorrência disso, para os demais casos de agrupamento serão apresentados somente os resultados das redes MLP para os agrupamentos originados do algoritmo k-means++.

Algumas das redes treinadas para os 5 agrupamentos produziram resultados melhores do que a rede treinada para todos os dados, porém outras tiveram um desempenho pior. O desempenho médio das cinco redes treinadas uma para cada agrupamento é apresentado na TABELA 10.

TABELA 10 – DESEMPENHO MÉDIO DAS REDES MLP PARA 5 AGRUPAMENTOS

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
média	0.3353	88.47%	1.0288	74.06%

FONTE: a autora

Para o cenário com 7 agrupamentos, o resultado do desempenho das 7 redes MLP treinadas cada uma para os dados de um agrupamento são descritos na TABELA 11

TABELA 11 – RESULTADOS DAS REDES MLP PARA 7 AGRUPAMENTOS COM K-MEANS++

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
grupo 1	0.2559	97.56%	0.8338	82.11%
grupo 2	0.3065	81.02%	0.8793	70.77%
grupo 3	0.2441	98.29%	0.7751	82.35%
grupo 4	0.2607	96.37%	0.8595	81.63%
grupo 5	0.4212	72.82%	1.0481	66.51%
grupo 6	0.3111	88.88%	0.8568	76.18%
grupo 7	0.2939	82.90%	1.0921	69.28%

FONTE: a autora

O desempenho médio das 7 redes treinadas uma para cada agrupamento é apresentado na TABELA 12.

TABELA 12 – DESEMPENHO MÉDIO DAS REDES MLP PARA 7 AGRUPAMENTOS

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
média	0.2990	88.26%	0.9063	75.54%

FONTE: a autora

Para o cenário com 10 agrupamentos, o resultado do desempenho das 10 redes MLP treinadas cada uma para os dados de um agrupamento são descritos na TABELA 13

TABELA 13 – RESULTADOS DAS REDES MLP PARA 10 AGRUPAMENTOS COM K-MEANS++

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
grupo 1	0.2486	96.74%	0.7507	86.78%
grupo 2	0.3188	90.27%	0.8827	81.24%
grupo 3	0.2104	98.82%	0.8136	85.14%
grupo 4	0.2902	90.17%	0.8898	79.81%
grupo 5	0.3009	81.45%	1.046	68.37%
grupo 6	0.2844	85.08%	1.3214	68.18%
grupo 7	0.2327	90.71%	0.8121	82.23%
grupo 8	0.4675	70.95%	1.4997	62.99%
grupo 9	0.3031	83.41%	1.2051	67.97%
grupo 10	0.2487	98.35%	0.6984	87.20%

FONTE: a autora

O desempenho médio das 10 redes treinadas uma para cada agrupamento é apresentado na TABELA 14.

TABELA 14 – DESEMPENHO MÉDIO DAS REDES MLP PARA 10 AGRUPAMENTOS

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
média	0.2905	88.59%	0.9919	76.99%

FONTE: a autora

Segundo os resultados a configuração com 10 agrupamentos gerou maior acurácia média no treinamento e teste. Sendo que sua acurácia no teste foi 1,45% maior que a acurácia média para 7 agrupamentos e 2,93% maior que a acurácia média para 5 agrupamentos.

5.4 REDE HÍBRIDA: K-MEANS + LSTM

Da mesma forma que foi realizado para as redes MLP, foram treinadas redes LSMT considerando os mesmos 5, 7 e 10 agrupamentos. Utilizou-se para esses treinamentos a rede com duas camadas LSTM conectadas a uma camada densamente conectada que por sua vez está conectada a saída. A quantidade de neurônios ou blocos de memória em cada uma dessas camadas foi respectivamente 50, 25 e 50, com 11 neurônios na camada de saída. Esta arquitetura foi escolhida tendo como base o desempenho da rede quando treinada com a totalidade de dados.

Para o cenário com 5 agrupamentos, o resultado do desempenho das 5 redes LSTM treinadas cada uma para os dados de um agrupamento são descritos na TABELA 15

TABELA 15 – RESULTADOS DAS REDES LSTM PARA 5 AGRUPAMENTOS COM K-MEANS++

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
grupo 1	1.0745	48.89%	1.2942	39.18%
grupo 2	0.8161	58.23%	1.5481	35.09%
grupo 3	1,1496	46.00%	1.6548	31.71%
grupo 4	0.641	60.77%	1.3919	50.92%
grupo 5	0.6436	73.44%	1.132	61.63%

FONTE: a autora

O desempenho médio das 5 redes treinadas uma para cada agrupamento é apresentado na TABELA 16.

TABELA 16 – DESEMPENHO MÉDIO DAS REDES LSTM PARA 5 AGRUPAMENTOS

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
média	0.8649	57.46%	1.4042	43,70%

FONTE: a autora

Para o cenário com 7 agrupamentos, o resultado do desempenho das 7 redes LSTM treinadas cada uma para os dados de um agrupamento são descritos na TABELA 17

TABELA 17 – RESULTADOS DAS REDES LSTM PARA 7 AGRUPAMENTOS COM K-MEANS++

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
grupo 1	0.873	64.38%	1.5006	50.49%
grupo 2	0.7261	53.28%	1.5597	38.83%
grupo 3	0.6873	70.81%	1.1339	60.83%
grupo 4	0.8825	52.96%	1.5978	36.02%
grupo 5	1.1848	47.87%	1.5496	35.02%
grupo 6	0.6389	61.42%	1.4846	48.92%
grupo 7	0.9708	66.62%	1.5732	41.76%

FONTE: a autora

O desempenho médio das 7 redes treinadas uma para cada agrupamento é apresentado na TABELA 18.

TABELA 18 – DESEMPENHO MÉDIO DAS REDES LSTM PARA 7 AGRUPAMENTOS

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
média	0.8519	59.62%	1.4856	44.55%

FONTE: a autora

Para o cenário com 10 agrupamentos, o resultado do desempenho das 10 redes LSTM treinadas cada uma para os dados de um agrupamento são descritos na TABELA 19

TABELA 19 – RESULTADOS DAS REDES LSTM PARA 10 AGRUPAMENTOS COM K-MEANS++

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
grupo 1	0.885	53.14%	1.6349	37.86%
grupo 2	0.959	57.70%	1.6834	39.08%
grupo 3	0.7928	65.94%	1.303	54.66%
grupo 4	0.6115	50.25%	1.5541	43.06%
grupo 5	0.7247	56.26%	1.6192	39.78%
grupo 6	1.0242	59.29%	1.6066	47.57%
grupo 7	0.5438	61.45%	0.2684	52.48%
grupo 8	1.1883	44.90%	1.5358	41.88%
grupo 9	0.9582	68.60%	1.5517	42.81%
grupo 10	0.649	72.60%	0.0338	66.94%

FONTE: a autora

O desempenho médio das 10 redes treinadas uma para cada agrupamento é apresentado na TABELA 20.

TABELA 20 – DESEMPENHO MÉDIO DAS REDES LSTM PARA 10 AGRUPAMENTOS

Grupo	Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
média	0.8336	59.01%	1.479	46.61%

FONTE: a autora

5.5 UTILIZANDO AS VARIÁVEIS POLARIMÉTRICAS

Conforme abordado anteriormente, as variáveis polarimétricas estão intimamente ligadas a classificação de hidrometeoros, assim é esperado que apresentar as mesmas como estímulo para a rede neural artificial gere um desempenho melhor da mesma. De fato, enquanto os cenários mais promissores das redes treinadas utilizando a localização, a temperatura e a

refletividade horizontal produziram uma acurácia média de aproximadamente 86% no treinamento e 75% no teste. Nos cenários testados para redes MLP com as variáveis polarimétricas, o melhor desempenho foi da rede híbrida com 7 agrupamentos que resultou em uma acurácia média de 98.45% no treinamento e 95.69% no teste.

As redes LSTM, mesmo tendo as variáveis polarimétricas como entrada, continuaram tendo um desempenho inferior ao das redes MLP, atingindo uma acurácia média de 71.97% no treinamento e 69.02% no teste. Assim, apesar das novas variáveis melhorarem o desempenho dessas redes, o mesmo ainda não foi superior ao das redes MLP.

Porém, para utilizar as variáveis polarimétricas para a classificação de hidrometeoros na região do radar de Teixeira Soares é necessário estimar esses valores, uma vez que o próprio radar não fornece tais variáveis. Para aproximar os valores de Z_{DR} , K_{DP} e ρ_{HV} utilizou-se uma rede MLP com 5 camadas ocultas e 300 neurônios em cada uma delas e uma camada de saída com 3 neurônios, um para aproximar cada uma das variáveis.

A função de ativação aplicada para as camadas ocultas foi a tangente hiperbólica e para a camada de saída a função sigmoide. A taxa de aprendizagem utilizada foi de 0.5 com a regra de aprendizagem *backpropagation*. Utilizou-se também do treinamento em lote, considerando lotes com 500 dados cada e uma divisão dos dados de 70% para o treinamento e 30% para o teste. Os resultados obtidos a partir dessa configuração estão descritos na TABELA 21.

TABELA 21 – DESEMPENHO DA REDE MLP PARA APROXIMAÇÃO DAS VARIÁVEIS POLARIMÉTRICAS

Erro treinamento	Acurácia treinamento	Erro teste	Acurácia teste
0.0051	96.49%	0.0076	93.83%

FONTE: a autora

Se considerarmos que a acurácia do teste realizado aplicando a rede MLP com as variáveis polarimétricas é de 95.69% e essas variáveis são aproximadas com uma precisão de 93.83%, chegamos a um cenário com acurácia de aproximadamente 89.78%.

De um modo geral, no contexto aplicado aqui, as redes MLP tiveram um melhor desempenho que as redes LSTM. Os resultados da rede MLP se mostraram sensíveis a alterações na arquitetura da rede, já para a rede LSTM esta sensibilidade não foi observada, entretanto a alteração em sua arquitetura influenciou no tempo de processamento do treinamento da rede.

Quando combinado o método da rede k-means com a rede MLP ou LSTM, o desempenho médio das redes foi superior ao desempenho da rede MLP ou LSTM única. Sendo que, em alguns grupos o desempenho da rede treinada especificamente para o grupo foi consideravelmente melhor que a mesma rede sem agrupamento. Na maioria dos agrupamentos a rede aplicada ao mesmo gerou um resultado semelhante ao cenário de rede única, e ainda,

em alguns grupos a rede do mesmo teve desempenho inferior a rede individual.

Outro fator que gerou uma melhora nos resultados de todos os cenários testados foi a incorporação das variáveis polarimétricas no conjunto de entradas. Este ato necessitou de um processamento prévio dos dados e o treinamento de uma rede neural para fazer a aproximação das variáveis polarimétricas, as quais não são originais dos dados de radares de polarização simples. A rede MLP treinada para a estimar estas variáveis se mostrou eficiente, de modo que, considerando a acurácia da aproximação das variáveis com a acurácia do método de classificação de hidrometeoros por rede híbrida, o resultado alcançado foi superior ao resultado no mesmo cenário sem as variáveis polarimétricas.

6 CONCLUSÃO

Os radares meteorológicos se tornaram referência para o monitoramento e previsão a curto prazo. Unindo o termômetro, a radiossonda e imagens de satélite, tornaram-se também um instrumento amplamente utilizado para pesquisas com foco nos fenômenos meteorológicos, sendo uma das melhores ferramentas para detectar ocorrências meteorológicas perigosas como fortes tempestades, granizo e tornados.

Porém, somente os radares de dupla polarização fazem a classificação de hidrometeoros. De modo que devem ser usadas técnicas alternativas para fazer essa classificação nas áreas monitoradas por radares de polarização simples. Assim, assumiu-se neste estudo o objetivo principal de aplicar técnicas de redes neurais artificiais para fazer a classificação de hidrometeoros a partir de dados de radares de polarização simples utilizando como base a classificação fornecida por radares de dupla polarização com área de cobertura comum.

Os hidrometeoros de interesse principal são neve seca, neve molhada, cristais de gelo, graupel, grandes gotas, chuva leve e moderada, chuva forte e granizo, além de diferenciar os mesmos de alvos não meteorológicos como dispersores biológicos e ecos de terreno.

Para o estudo foram selecionadas 7 datas com ocorrência dos hidrometeoros de interesse, de modo especial, o granizo. Dentre essas datas foram selecionados 30 momentos com maior ocorrência de hidrometeoros, a partir desses momentos foram extraídos os dados pontuais das áreas em comum entre o radar de Teixeira Soares e os radares de Cascavel e Lontras.

A classificação dos hidrometeoros foi feita através da aplicação de técnicas de redes neurais artificiais, especificamente a rede de k-means, a rede Perceptron Multicamadas, MLP, e a rede Long Short-Term Memory, LSTM. Sendo que a rede k-means foi utilizada para fazer o agrupamento dos dados, com o objetivo de formar grupos com dados homogêneos potencializando o desempenho das redes MLP e LSTM que foram treinadas cada uma para um agrupamento.

A rede MLP demonstrou um desempenho melhor que a rede LSTM para este ambiente, sendo que o resultado médio melhorou quando estas técnicas foram aplicadas em conjunto com a rede de k-means, neste cenário, para dez agrupamentos atingiu-se uma acurácia média de 88.59% no treinamento e 76.99% na execução da rede no conjunto de teste.

Outro fator que possibilitou melhores resultados foi a incorporação das variáveis polarimétricas no conjunto de entradas da rede, fato esse que foi possível pela aproximação dessas variáveis através de uma rede MLP. Neste cenário as redes MLP aplicadas para 7 agrupamentos geraram uma acurácia média de 98.45% no treinamento e 95.69% no teste para a classificação de hidrometeoros, já a rede MLP utilizada para inferir os valores das variáveis polarimétricas Z_{DR} , ρ_{HV} e K_{DP} apresentou uma acurácia de 96.49% no treinamento e 93.83% no teste.

Os resultados obtidos para cada um dos agrupamentos gerados são discrepantes, sendo que alguns grupos proporcionam uma acurácia consideravelmente melhor que outros. Por isso, fazer um estudo específico nos grupos com menor acurácia pode gerar uma melhora no resultado do mesmo e conseqüentemente na acurácia média geral.

6.1 SUGESTÃO PARA TRABALHOS FUTUROS:

Como sugestão para trabalhos futuros propõe-se:

- Aprofundar os estudos nas características dos grupos onde a rede treinada obteve menor acurácia, a fim de determinar processos que melhorem a mesma;
- Aplicação de outras técnicas de agrupamentos para potencializar o desempenho das redes neurais aplicadas em cada agrupamento;
- Teste de diferentes técnicas ou arquiteturas de rede para cada agrupamento;
- Utilizar técnicas para seleção direcionada de dados de modo a obter uma representatividade maior de dados associados a hidrometeoros com menor ocorrência.

REFERÊNCIAS

- ALVES, Wonder Alexandre Luz; ARAÚJO, Sidnei Alves de; LIBRANTZ, André Felipe Henriques. Reconhecimento de Padrões de Textura em Imagens Digitais Usando uma Rede Neural Artificial Híbrida. **Exacta – Engenharia de Produção**, v. 4, p. 325–332, 2006.
- ARTHUR, David; VASSILVITSKII, Sergei. k-means++: the advantages of careful seeding. **Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms**, v. 7, p. 1027–1035, 2007.
- BARNES, Stanley L. A Technique for Maximizing Details in Numerical Weather Map Analysis. **Journal of Applied Meteorology and Climatology**, v. 3, p. 396–409, 1964.
- BORGES, Luiz Eduardo. **Python para Desenvolvedores**. 2. ed. [S.n.], 2010.
- BRINGI, V.N.; CHANDRASEKAR, V. **Polarimetric Doppler Weather Radar: Principles and Applications**. [S.l.: s.n.], ago. 2001. ISBN 9780521623841. DOI: 10.1017/CB09780521623841.
- BUDERI, Robert. **The Invention That Changed the World: How a Small Group of Radar Pioneers Won the Second World War and Launched a Technical Revolution**. 1. ed. Touchstone, 1998.
- CAJAL, Santiago Ramón y. **Histologie du Systéms Nerveux de l’homme et des vertébrés**. 1. ed. Hachette Livre, 1911.
- CALVETTI, Leonardo; BENETI, Cesar; MEDEIROS, Otávio. In: **Meteorologia em Tópicos**. Edição: André Becker Nunes e Glauber Lopes Mariano. 1. ed. Pelotas: 2017. v. 4 Identificação de fenômenos meteorológicos por radar meteorológico.
- COADOU, Yves Le; BENABDESLEM, Khalid. Optimizing local modeling for times series prediction. **International Journal of Computational Intelligence Research**, v. 2, p. 86–90, 2006.
- COELHO, Flávio Codeço. **Computação Científica com Python: Uma introdução à programação para cientistas**. 1. ed. Petrópolis, Rio de Janeiro: [s.n.], 2007.
- CRESSMAN, George P. An Operational Objective Analysis System. **Monthly Weather Review**, v. 87, p. 367–374, 1959.
- DAMIAN, Eloá Alano. **Dois Metodologias Aplicadas a Classificação de Precipitação Convectiva e Estratiforme com Radar Meteorológico: SVM e K-means**. 2011. Diss. (Mestrado) – Universidade Federal do Paraná.

DOLAN, Brenda; RUTLEDGE, Steven A. A Theory-Based Hydrometeor Identification Algorithm for X-Band Polarimetric Radars. **Journal of Atmospheric and Oceanic Technology**, v. 26, p. 2071–2088, 2009.

ELMAN, Jeffrey L. Finding Structure in Time. **Cognitive Science**, v. 24, p. 179–211, 1990.

FABRY, F. **Radar meteorology: Principles and practice**. Cambridge University Press, 2015.

FAUSETT, Laurene. **Fundamentals of Neural Networks: architectures, algorithms and applications**. 1. ed. New Jersey: Prentice-Hall: Pearson, 1994.

FONSECA, Felipe Cesar Stanzani; BELTRAME, Walber Antonio Ramos. Aplicações Práticas dos Algoritmos de Clusterização Kmeans e Bisecting K-means. **SIBGRAPI - Congresso de Gráficos, Padrões e Imagens**, v. 32, p. 1124–1131, 2010.

GALATI, Gaspare. **100 Years of Radar**. 1. ed. Springer International Publishing, 2016.

GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. Learning to forget: continual prediction with LSTM. **Ninth International Conference on Artificial Neural Networks ICANN 99**, v. 2, p. 850–855, 1999.

GRAVES, Alex. **Supervised Sequence Labelling with Recurrent Neural Networks**. 1. ed. [S.l.]: Springer, 2012.

GREFF, K. et al. LSTM: A Search Space Odyssey. **IEEE Transactions on Neural Networks and Learning Systems**, v. 28, p. 2222–2232, 2017.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining: Concepts and Techniques**. 3. ed. [S.l.]: Morgan Kaufmann, 2011.

HAYKIN, Simon. **Neural Networks and Learning Machines**. 3. ed. Prentice Hall, 2008.

HEBB, Donald Olding. **The Organization of Behavior: A Neuropsychological Theory**. 1. ed. Wiley, New York: Psychology Press, 1949.

HEISTERMANN, M. et al. The Emergence of Open-Source Software for the Weather Radar Community. **Bulletin of the American Meteorological Society**, v. 96, p. 117–128, 2015.

HELMUS, Jonathan J; COLLIS, Scott M. The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language. **Journal of Open Research Software**, v. 4, p. 25, 2016.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long Short-Term Memory. **Neural computation**, v. 9, p. 1735–1780, dez. 1997.

HÖLLER, H. et al. Life Cycle and Precipitation Formation in a Hybrid-Type Hailstorm Revealed by Polarimetric and Doppler Radar Measurements. **Journal of the Atmospheric Sciences**, v. 51, p. 2500–2522, 1994.

- HONG, Yang; GOURLEY, Jonathan J. **Radar Hydrology: Principles, models, and applications**. 1. ed. CRC Press, 2014.
- HUANG, Zhiheng; XU, Wei; YU, Kai. Bidirectional LSTM-CRF Models for Sequence Tagging. **CoRR**, abs/1508.01991, ago. 2015.
- JORDAN, Michael I. Attractor Dynamics and Parallelism in a Connectionist Sequential Machine. **Proc. of the Eighth Annual Conf. of the Cognitive Science Society**, p. 531–546, 1986.
- JUNIOR, Joseph F. Hair et al. **Análise Multivariada de Dados**. 6. ed. Bookman, 2009.
- KEENAN, T. Hydrometeor classification with a C-band polarimetric radar. **Australian Meteorological Magazine**, v. 52, p. 23–31, 2003.
- KNIGHT, CHARLES A.; VIVEKANANDAN, JOTHIRAM; LASHER-TRAPP, SONIA G. First Radar Echoes and the Early ZDR History of Florida Cumulus. **Journal of the Atmospheric Sciences**, v. 59, p. 1454–1472, 2002.
- KOUKETSU, Takeharu; UYEDA, Hiroshi. Validation of Hydrometeor Classification Method for X-band Polarimetric Radar - Comparison with ground observation of solid hydrometeor. **Sixth European Conf. on Radar in Meteorology and Hydrology (ERAD2010)**, p. 195–201, 2010.
- KOVÁCS, Zsolt László. **Redes Neurais Artificiais: fundamentos e aplicações**. 4. ed. Livraria da Física, 2006.
- LIU, H.; CHANDRASEKAR, V. Classification of hydrometeors based on polarimetric radar measurements: Development of fuzzy logic and neuro-fuzzy systems, and in situ verification. **J. Atmos. Oceanic Technol**, v. 17, p. 140–164, 2000.
- MARZANO, F. S.; SCARANARI, D.; CELANO, M. et al. Hydrometeor classification from dual-polarized weather radar: extending fuzzy logic from S-band to C-band data. **Advances in Geosciences**, v. 7, p. 109–114, 2006.
- MARZANO, F. S.; SCARANARI, D.; VULPIANI, G. et al. Supervised Classification and Estimation of Hydrometeors From C-Band Dual-Polarized Radars: A Bayesian Approach. **IEEE Transactions on Geoscience and Remote Sensing**, v. 46, p. 85–98, 2008.
- MCCULLOCH, Warren Sturgis; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 32, p. 1124–1131, 1943.
- MELO, Jeferson Aparecido Arestides de. **Caracterização de uma Linha de Instabilidade Amazônica Utilizando Radar Polarimétrico Durante o Projeto Chuva – Belém**. 2016. Diss. (Mestrado) – Universidade Federal de Campina Grande.
- MENDEL, J.M.; MCLAREN, R.W. In: **Adaptive, Learning and Pattern Recognition Systems**. Edição: J.M. Mendel e K.S. Fu. Elsevier, 1970. v. 66 8 Reinforcement-Learning Control and Pattern Recognition Systems.

- METED. **Weather radar fundamentals**. [S.l.: s.n.], 2017. Disponível em: <http://www.meted.ucar.edu/radar/basic_wxradar>. Acesso em: 5 dez. 2020.
- MULLER, Brad. **Radar equation**. [S.l.: s.n.], 2013. Disponível em: <https://wx.erau.edu/faculty/mullerb/Wx365/Radar_equation/radar_equation.pdf>. Acesso em: 10 dez. 2020.
- OLIVEIRA, Camila de. **Identificação e Correção da Banda Brilhante em Dados de Radar Meteorológico**. 2014. Diss. (Mestrado) – Universidade Federal do Paraná.
- PALMA, Luann Farias. **Agrupamento de Dados: K-médias**. 2018. Monografia (Graduação) – Universidade Federal do Recôncavo da Bahia.
- PARK, Hyangsuk et al. The Hydrometeor Classification Algorithm for the Polarimetric WSR-88D: Description and Application to an MCS. **Weather and Forecasting**, v. 24, p. 730–748, 2009.
- PETTERSEN, S. **Introduction to Meteorology**. 1. ed. New York e London: McGRAW-HILL BOOK COMPANY, 1941.
- QUEIROZ, Antônio Paulo de. **Monitoramento e Previsão Imediata de Tempestades Severas Usando Dados de Radar**. 2008. Diss. (Mestrado) – Instituto Nacional de Pesquisas Espaciais.
- RAMIREZ, Yusvelis Maribel Barzaga. **Inferência de hidrometeoros a partir de um radar meteorológico de dupla polarização banda X**. 2018. Diss. (Mestrado) – Universidade de São Paulo.
- RAUBER, Robert M.; NESBITT, Stephen W. **Radar Meteorology: A First Course**. 1. ed. Wiley-Blackwell, 2018. v. 1.
- RAUBER, Thomas. Redes neurais artificiais, mar. 2005.
- RINEHART, Ronald E. **Radar for meteorologists**. 4. ed. Nevada, United States of America: Rinehart Publications, 2004.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, p. 386–408, 1958.
- RYZHKOV, Alexander V. et al. Calibration Issues of Dual-Polarization Radar Measurements. **Journal of Atmospheric and Oceanic Technology**, v. 22, p. 1138–1155, 2005.
- AL-SAKKA, Hassan et al. A New Fuzzy Logic Hydrometeor Classification Scheme Applied to the French X-, C-, and S-Band Polarimetric Radars. **Journal of Applied Meteorology and Climatology**, v. 52, p. 2328–2344, 2013.
- SANTOS, Karla Susane Borges dos. **Magnetron: do Radar ao Forno de Micro-Ondas**. 2011. f. 22. Monografia (Graduação) – Universidade Católica de Brasília, Brasília.

- SCHUUR, Terry J. et al. Classification of Precipitation Types during Transitional Winter Weather Using the RUC Model and Polarimetric Radar Retrievals. **Journal of Applied Meteorology and Climatology**, v. 51, p. 763–779, 2012.
- SELUZNIAK, Renan Henrique Luz. **Controle de Qualidade em Dados de Radar Polarimétrico com Apoio de Visualizador Interativo de Dados**. 2016. Diss. (Mestrado) – Universidade Federal do Paraná.
- SILVA, Ivan Nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. **Redes Neurais Artificiais**: para engenharia e ciências aplicadas. 1. ed. São Paulo: Artliber, 2010.
- SIQUEIRA, Paulo Henrique. **Apostila de Metaheurísticas e Aplicações**. [S.l.: s.n.], 2018. Disponível em: <http://www.exatas.ufpr.br/portal/deggraf_paulo/wp-content/uploads/sites/4/2021/02/apostila_2020.pdf>. Acesso em: 16 jan. 2021.
- SNYDER, Jeffrey C. et al. Attenuation Correction and Hydrometeor Classification of High-Resolution, X-band, Dual-Polarized Mobile Radar Measurements in Severe Convective Storms. **Journal of Atmospheric and Oceanic Technology**, v. 27, p. 1979–2001, 2010.
- SOUSA RIGNEL, Diego Gabriel de; CHENCI, Gabriel Pupin; LUCAS, Carlos Alberto. Uma Introdução a Lógica Fuzzy. **Revista Eletrônica de Sistemas de Informação e Gestão Tecnológica**, v. 1, p. 17–28, 2011.
- STRAKA, Jerry M.; ZRNIĆ, Dusan S.; RYZHKOV, Alexander V. Bulk Hydrometeor Classification and Quantification Using Polarimetric Radar Data: Synthesis of Relations. **Journal of Applied Meteorology and Climatology**, v. 39, p. 1341–1372, 2000.
- SUNDERMEYER, Martin; SCHLUTER, Ralf; NEY, Hermann. LSTM Neural Networks for Language Modeling. **INTERSPEECH-2012**, p. 194–197, 2012.
- THEODORIDIS, Sergios; KOUTROUMBAS, Konstantinos. **Pattern recognition**. 4. ed. [S.l.]: Academic Press, 2009.
- VIVEKANANDAN, J. et al. Cloud Microphysics Retrieval Using S-Band Dual-Polarization Radar Measurements. **Bulletin of the American Meteorological Society**, v. 80, p. 381–388, 1999.
- WIDROW, B.; HOFF, M. Adaptive Switching Circuits. **WESCON Convention Record**, v. 4, p. 96–104, 1960.
- YNOUE, Rita Yuri et al. **Meteorologia**: noções básicas. 1. ed. Oficina de Textos, 2017.