

Redes Neurais Artificiais Aplicadas à Previsão Antecipada de Precipitações na Região Central de Manaus

Rafaela dos Santos Sousa, Liliana Oliveira da Silva,
Elloá B. Guedes, Maria Betânia Leal de Oliveira

¹Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Av. Darcy Vargas, 1200 – Manaus – Amazonas

{rass.eng, lods.eng, ebgcosta, mloliveira}@uea.edu.br

Abstract. *This work aims at showing the results of artificial neural networks for forecasting rainfall 30 min ahead in the central area of Manaus from data of meteorological stations located around the city. A total of 523 neural networks were proposed and trained for the problem under consideration where two of them showed better accuracy and precision. A second training with a balanced proportion of the existing classes resulted in an increase in the precision of the tests. The results obtained can be used by several agencies for strategic decision making aiming at minimizing the negative effects of rainfall in this city.*

Resumo. *Este trabalho apresenta os resultados da utilização de redes neurais artificiais para previsão antecipada em 30 min da ocorrência de precipitações na região central da cidade de Manaus a partir de dados oriundos de estações meteorológicas localizadas no entorno da cidade. Foram propostas e treinadas 523 redes neurais adequadas para o problema considerado, das quais duas foram elencadas por possuírem maiores acurácia e precisão. Um segundo treinamento com balanceamento das classes existentes permitiu verificar aumento na precisão dos testes. Os resultados obtidos podem ser utilizados por diversos órgãos para tomada de decisões estratégicas para minimizar os efeitos negativos da precipitação nesta cidade.*

1. Introdução

A previsão de precipitações pode ser endereçada segundo duas abordagens diferentes. A primeira delas envolve o estudo dos fenômenos de precipitação com o intuito de modelar as leis físicas que os regem. Porém, esta abordagem pode não ser viável, pois: (i) a precipitação é o produto final de processos atmosféricos complexos, que variam tanto no espaço quanto no tempo; (ii) mesmo que estes processos sejam descritos de maneira concisa e completa, o número de cálculos envolvidos pode ser proibitivo; e (iii) os dados que estão disponíveis para auxiliar nesta tarefa são limitados tanto nas dimensões espaciais quanto temporais. A segunda abordagem de previsão, por sua vez, baseia-se na metodologia de reconhecimento de padrões, segundo a qual características relevantes tanto espaciais quanto temporais são extraídas de padrões históricos de precipitação, podendo ser utilizadas na previsão de novos eventos [Luk et al. 2001].

Diferentes técnicas da Inteligência Computacional já foram consideradas na previsão de precipitações. As *redes neurais artificiais*, em particular, foram utilizadas em diversos cenários de previsão de precipitações em diferentes lugares do mundo [Darji et al. 2015, Bushara and Abraham 2013, Nayak et al. 2013, Gardner and Dorling 1998]. As redes neurais artificiais são formalmente definidas como sendo sistemas distribuídos

massivamente paralelos, formadas a partir de unidades de processamento simples e com uma propensão natural para o armazenamento e posterior disponibilização de conhecimento obtido a partir da experiência [Haykin 2009]. Este modelo de aprendizagem de máquina é bastante utilizado neste tipo de problema por várias razões, a citar: (i) RNAs são dirigidas por dados (*data driven*) e não demandam pré-requisitos restritivos sobre o que está sendo modelado; (ii) RNAs podem prever padrões que não são fornecidos durante o treinamento, isto é, são capazes de generalizar; (iii) RNAs são eficientes no treinamento de grandes amostras de dados graças a sua capacidade de processar em paralelo; (iv) RNAs possuem a habilidade de detectar relações complexas e não-lineares entre as variáveis dependentes e independentes [Darji et al. 2015].

Considerando a previsão de precipitações em Manaus, as chuvas se caracterizam por serem abundantes e não uniformes, sendo um dos principais fatores de influência no clima da região [Sioli 1991]. Esta cidade, em particular, possui característica equatorial quente e úmida, resultando em desconforto térmico. Este desconforto é amenizado principalmente diante da ocorrência de precipitações [da Silva 2009]. Descrever o regime de precipitações em Manaus, entretanto, não é uma tarefa fácil, pois vários sistemas atmosféricos atuam nesta cidade, modificando as características do tempo meteorológico e ocasionando ou suprimindo a ocorrência de precipitações. Alguns desses sistemas atuam em maior escala, como a Zona de Convergência Intertropical, e outros em escalas inferiores, como os sistemas convectivos de Mesoescala [da Silva 2012].

Trabalhos anteriores [Guedes et al. 2016, de Lima and Guedes 2015b, de Lima and Guedes 2015a, de Lima 2016] investigaram a possibilidade de prever a ocorrência de precipitações em Manaus considerando a adoção de redes neurais artificiais. Para tanto, utilizaram 40 anos de dados diários (1970 a 2010, totalizando 12 mil exemplos) advindos de uma estação meteorológica automática localizada na cidade para treinar e testar 38 redes neurais diferentes. Os resultados obtidos por estes autores demonstraram que as redes neurais se mostram boas predictoras para este cenário, obtendo uma elevada acurácia. Prever a ocorrência de precipitações em uma cidade como Manaus, embora seja uma tarefa complexa, pode impactar positivamente em diversos aspectos, tais como na prevenção de alagamentos, por exemplo. Porém, embora a previsão da ocorrência da precipitação um dia à frente colabore na tomada de decisões estratégicas, é essencial estimar mais precisamente quando esta precipitação ocorrerá ao longo de um dia, permitindo o deslocamento de efetivo ou a geração de alertas junto às populações de risco. Na Austrália, por exemplo, este cenário já foi considerado. Os dados de diferentes estações meteorológicas auxiliaram na previsão da ocorrência antecipada de precipitações em uma região propensa à alagamentos [Luk et al. 2001], permitindo a geração de alertas com 15 minutos de antecedência.

Levando em consideração as preocupações mencionadas, este trabalho teve como objetivo tirar proveito de uma rede de estações meteorológicas localizadas no entorno da zona urbana de Manaus para treinar redes neurais artificiais com o intuito de realizar previsões de precipitação na região central da cidade, com 30 minutos de antecedência. Dados meteorológicos coletados durante 4 anos por 3 estações meteorológicas auxiliaram na identificação de redes neurais adequadas para este cenário, resultando em 97% de acurácia para a previsão de precipitações meia hora à frente.

Para apresentar os resultados obtidos, o trabalho em questão está organizado como segue. Uma breve descrição da rede de estações meteorológicas utilizada como fonte de dados para este trabalho é apresentada na Seção 2. Uma visão geral do conjunto de dados

utilizado neste trabalho é mostrada na Seção 3. Os materiais e métodos encontram-se descritos na Seção 4. Os resultados obtidos são detalhados e discutidos na Seção 5. Por fim, as considerações finais são listadas na Seção 6.

2. Rede de Estações Meteorológicas de Manaus

A Rede de Estações Meteorológicas de Manaus é administrada pelo Laboratório de Instrumentação Meteorológica (LabInstru) da Escola Superior de Tecnologia da Universidade do Estado do Amazonas. Esta rede é composta de dez estações meteorológicas automáticas distribuídas pela cidade de Manaus e região metropolitana, conforme ilustrado na Figura 1, adquiridas com recursos do Projeto Rede de Mudanças Climáticas da Amazônia [de Oliveira 2017].

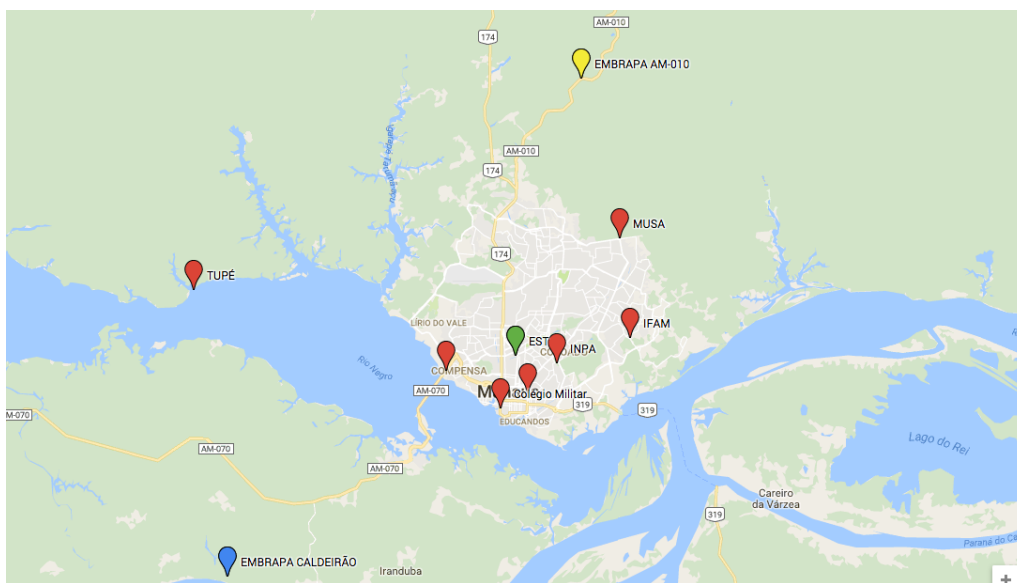


Figura 1: Mapa da Rede de Estações Meteorológicas de Manaus. Em particular, destacam-se as estações Embrapa-AM010 (cor amarela), Embrapa-Caldeirão (cor azul) e EST (cor verde), as quais serão utilizadas no escopo deste trabalho.

Dados como temperatura, pressão, velocidade do vento, volume de precipitação, dentre outros, são coletados pelas estações meteorológicas em intervalos de tempo fixos de 5 minutos e enviados, via rede móvel de dados, para um servidor. Estas informações são manipuladas e analisadas pelos pesquisadores e professores do LabInstru para a realização de pesquisas no contexto científico-tecnológico, para a divulgação junto à comunidade em geral e para alimentar centros operacionais de previsão do tempo. Os dados gerados pela rede de estações também alimentam órgãos dos setores ambientais, produtivos, de infraestrutura e de Defesa Civil [de Oliveira 2017].

3. Visão Geral do Conjunto de Dados

Para consolidar um conjunto de dados para o problema considerado, primeiramente foi necessário recorrer à base de dados das estações meteorológicas para checar a disponibilidade das medições. Embora a rede de estações meteorológicas tenha iniciado o seu funcionamento em 2012, as estações foram sendo implantadas progressivamente, algumas das quais começaram a coletar dados apenas 5 anos depois. Além disso, problemas na transmissão de dados para a rede móvel podem ter comprometido algumas medições

mesmo após a implantação das estações. Estes fatores culminaram na indisponibilidade de algumas medições.

Para contornar a indisponibilidade mencionada, foi necessário contabilizar as estações viáveis para o contexto considerado neste trabalho. Como resultado, mostrado detalhadamente na Tabela 1, duas estações meteorológicas localizadas no entorno de Manaus, a citar, Estações Embrapa-AM010 e Embrapa-Caldeirão, e uma estação localizada na região central da cidade, Estação EST, apresentaram uma quantidade de dados adequada para observação dos fenômenos de precipitação.

Tabela 1: Disponibilidade das medições das diferentes estações ao longo dos meses. O símbolo ■ indica que todas as medições estão presentes, o símbolo □ indica que há dados faltantes e a ausência de símbolos indica a ausência de medições.

2013												
	01	02	03	04	05	06	07	08	09	10	11	12
EST				□	□	■	□	■	□	■	■	■
EMBRAPA									□	■	□	■
CALDEIRÃO						□	■	■	■	■	■	■

2014												
	01	02	03	04	05	06	07	08	09	10	11	12
EST	■	■	■	■	■	■	■	□	■	■	■	■
EMBRAPA	■	■	■	□	■	■	□	□	■	■	■	■
CALDEIRÃO	■	■	■	■	■	■	■				□	

2015												
	01	02	03	04	05	06	07	08	09	10	11	12
EST			■	■	■						■	■
EMBRAPA	■	■	■	■	■	■	■	□	□	■	■	■
CALDEIRÃO	■	■	■	■	□	□	■	■	■	□	□	■

2016												
	01	02	03	04	05	06	07	08	09	10	11	12
EST	■			■	■			■	■	■	■	■
EMBRAPA	■	■	■	■	■	■	■	■	■	■	■	■
CALDEIRÃO	■	■	■	■	■	■	■	□	■	□		

A partir dos dados destas estações, foi constituído um *dataset* com 146.985 exemplos coletados durante os anos de 2013 a 2016, os quais contêm dados da pressão atmosférica, temperatura, umidade relativa do ar e precipitação das estações Embrapa-AM010 e Embrapa-Caldeirão. Estes atributos foram utilizados como preditores da precipitação 30 min *a posteriori* observada na Estação EST. O *dataset* é composto, portanto, de atributos numéricos intervalares (temperatura, pressão, umidade) e atributos categóricos (ocorrência ou não de precipitação, por cada estação).

Considerando o *dataset* consolidado, partiu-se para a realização de uma estatística descritiva das variáveis do conjunto de dados, cujos resultados encontram-se sintetizados nas Tabelas 2 e 3. O gráfico *boxplot* da Figura 2 ilustra a distribuição dos atributos de temperatura, pressão e umidade coletados das estações. Calculou-se também o coeficiente de correlação de Pearson dos atributos preditores com o atributo alvo (precipitação na Estação EST), ilustrado na última coluna da Tabela 2, com o intuito de analisar alguma associação de natureza linear. Porém, os valores baixos obtidos para este coeficiente não evidenciam a existência dessa associação.

Tabela 2: Estatística descritiva dos atributos intervalares do conjunto de dados.

Estação	Atributo	Média	Mediana	Desvio Padrão	Correlação
Embrapa Caldeirão	Pressão	1004.261	1004.30	2.221	0.048848
	Temperatura	26.744	25.74	2.970	-0.065570
	Umidade Relativa	90.026	95.00	11.537	0.064536
Embrapa AM010	Pressão	999.030	999.00	2.194	0.048752
	Temperatura	26.044	24.75	3.324	-0.060573
	Umidade Relativa	87.251	96.60	20.484	0.058460

De acordo com a Tabela 3, é possível verificar que há um desbalanceamento no número de exemplos por classe no conjunto de dados, no qual o número de não ocorrência

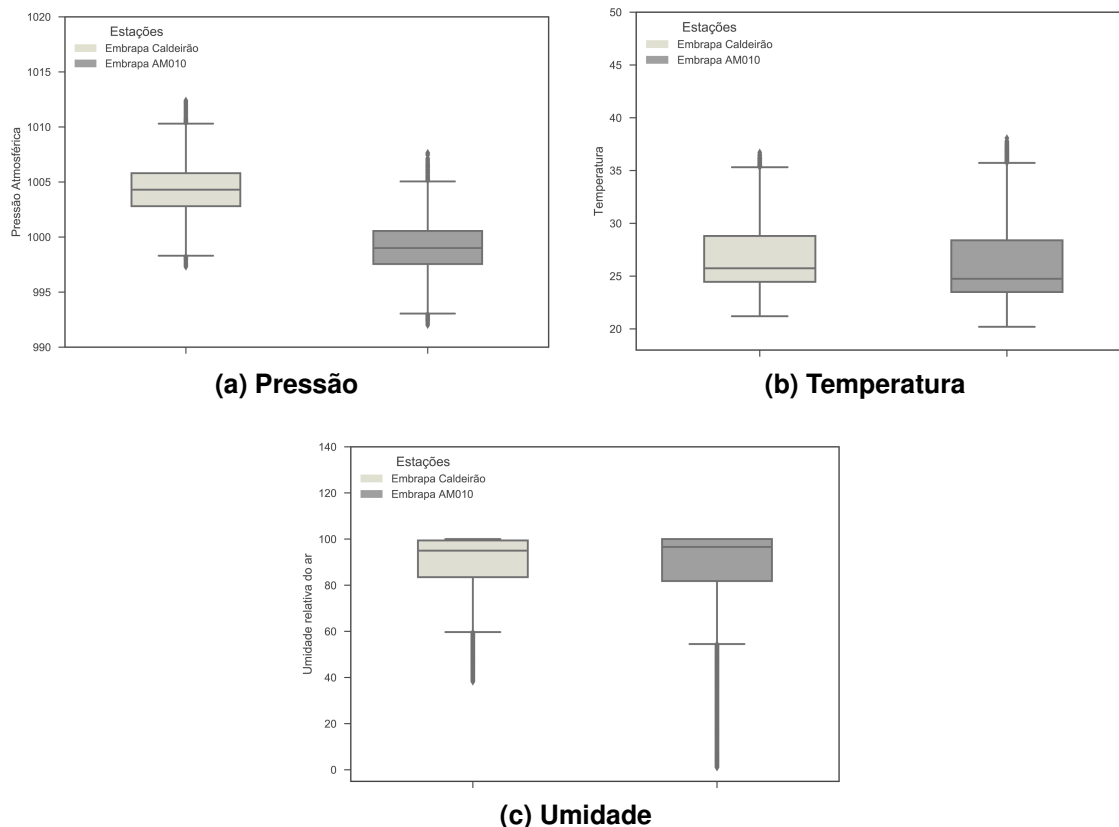


Figura 2: Gráficos de *boxplot* para os atributos preditores do conjunto de dados.

Tabela 3: Frequência de eventos de precipitação por estação.

Estação	Ocorrências	Não Ocorrências
Embrapa Caldeirão	3.0%	97%
Embrapa AM010	3.3%	96.7%
EST	2.7%	97.3%

de precipitações é significativamente maior do que o número de ocorrências. Este fator precisa ser levado em consideração ao analisar os resultados obtidos pelas redes neurais e será retomado posteriormente.

4. Materiais e Métodos

Considerando o *dataset* apresentado na seção anterior, considerou-se uma partição dos 146985 exemplos em 70% para treinamento e 30% para testes, escolhidos aleatoriamente. Da fração reservada para treinamento, 10% será utilizada para fins de validação com vistas a prevenir *overfitting*.

No cenário considerado, há 8 atributos preditores, são eles: pressão, temperatura, umidade relativa e ocorrência de precipitação nas Estações Embrapa-AM010 e Embrapa-Caldeirão; e há apenas um atributo alvo, a precipitação nos 30 min posteriores na Estação EST. Para abordar este problema, considerou-se a utilização de redes neurais *multilayer perceptron* a serem treinadas com o algoritmo *backpropagation* associado ao método de otimização *Levenberg-Marquadt* [Haykin 2009].

O dimensionamento das redes neurais para abordar o problema em questão fez uso de algumas sugestões da literatura. Apenas redes neurais com 1 a 2 camadas ocultas serão adotadas, em virtude de serem aproximadoras universais de qualquer função [Haykin 2009]. Quanto ao número de neurônios nessas camadas, uma das sugestões disponíveis envolve escolher valores no intervalo $0.5 \cdot N_i \leq N_h \leq 4 \cdot N_i$, em que N_i é o número de neurônios na camada de entrada ($N_i = 8$) e N_h é o número de neurônios na camada oculta [Palit and Popovic 2005]. Como resultado deste dimensionamento, foram identificadas 523 diferentes redes neurais *multilayer perceptron* com até duas camadas ocultas contendo de 4 até 32 neurônios por camada. Considerou-se a utilização da função de ativação tangente hiperbólica e diferentes valores para a taxa de aprendizado, a qual tem comportamento adaptativo, ou seja, decresce sempre que possível.

Para permitir a comparação entre as diferentes redes neurais propostas, serão utilizadas as seguintes métricas de desempenho: acurácia e a precisão. A acurácia é definida como sendo a quantidade de acertos da rede em relação aos valores esperados, expressa em termos percentuais. A precisão, por sua vez, é obtida pela razão o número de positivos verdadeiros pela soma dos erros positivos verdadeiros com os falsos positivos, podendo ser compreendida intuitivamente como a capacidade da rede em não rotular como positiva uma amostra que é negativa. Duas práticas serão adotadas na obtenção dessas métricas, a citar: (1) as métricas serão derivadas apenas utilizando resultados obtidos do conjunto de testes, pois fornecem uma estimativa da capacidade de generalização da rede; (2) para minimizar o viés introduzido pelas escolhas aleatórias dos pesos iniciais e da randomização na partição do conjunto de dados, serão considerados os resultados médios após 10 execuções de cada rede.

Levando em conta as redes identificadas e as métricas de desempenho estabelecidas, a seção a seguir apresenta os resultados do treino e testes na previsão antecipada de precipitações na zona urbana de Manaus.

5. Resultados e Discussão

As 523 redes neurais identificadas para o cenário em questão foram implementadas, treinadas e testadas com ferramental da linguagem de programação Python, nos quais o *framework* `sci-kit learn`¹ e a biblioteca `pandas`² possuíram papel central.

Os resultados obtidos revelam uma alta acurácia das redes neurais para o problema em questão, cujos resultados podem ser vistos na Tabela 4 e na Figura 3. Analisando o desvio padrão e o *boxplot* da acurácia para os resultados de todas as redes, é possível também identificar que todas as redes testadas tiveram um desempenho similar, sem muita dispersão. A maior acurácia obtida foi observada na rede neural com arquitetura (8, 5, 18, 1), isto é, com duas camadas ocultas com 5 e 18 neurônios, respectivamente.

Uma vez que os valores da acurácia foram satisfatórios, partiu-se para examinar a outra métrica de performance previamente selecionada, a precisão, cujos resultados são sintetizados na Tabela 5. A maior precisão foi observada na rede com arquitetura (8, 8, 3, 1), a qual teve acurácia igual a 97.26%.

Embora a acurácia tenha se apresentado alta para todas as redes, os resultados para a precisão precisam ser examinados mais detalhadamente, pois não se mostram tão satisfatórios. Ao examinar a matriz de confusão obtida a partir dos dados testados na rede

¹<http://scikit-learn.org/>

²<http://pandas.pydata.org/>

Tabela 4: Sumário dos resultados obtidos para a acurácia das redes neurais testadas.

Estatística	Resultado
Média	97.26%
Mínimo	97.14%
Máximo	97.32%
Desvio Padrão	0.05

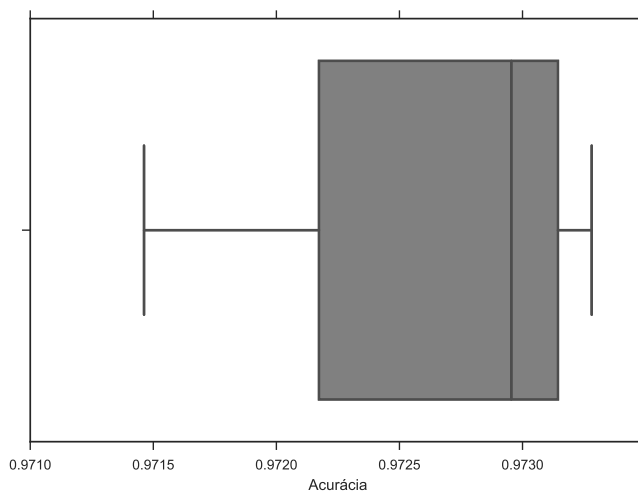


Figura 3: Gráfico de *boxplot* para as acurácias obtidas pelas 523 redes neurais.

Tabela 5: Sumário dos resultados obtidos para a precisão das redes neurais testadas.

Estatística	Resultado
Média	5.53%
Mínimo	0.08%
Máximo	12.07%
Desvio Padrão	2.86

(8, 8, 3, 1), apresentada na Tabela 6, é possível perceber que esta rede acerta muitos testes em que não há precipitação (42693 acertos) ao passo que comete equívocos prevendo que não iria chover quando, na verdade, a precipitação ocorre (1041 erros desse tipo).

Tabela 6: Matriz de confusão para as previsões do conjunto de testes da rede neural (8, 8, 3, 1). No sentido vertical encontram-se os valores desejados e no sentido horizontal encontram-se os valores obtidos pela rede.

	Não Ocorrência	Ocorrência
Não Ocorrência	42693	219
Ocorrência	1041	143

Uma das hipóteses que pode justificar a baixa precisão dessas redes para a precipitação em Manaus é o baixo percentual de exemplos de eventos de precipitação disponível

no conjunto de dados, conforme apontado anteriormente na Seção 3. Como consequência, as redes foram expostas a um número maior de exemplos onde não houve ocorrência de precipitação, aprendendo melhor sobre esta classe do que sobre a outra.

Para testar esta hipótese, todos os eventos de precipitação disponíveis no conjunto de dados foram considerados, juntamente com uma quantidade igual ao dobro de eventos de não ocorrência de precipitações, escolhidos aleatoriamente. Esta proporção respeita a existência de mais eventos de não precipitação do que de precipitação no conjunto de dados original. Estes exemplos passaram a compor um segundo conjunto de dados. As redes com melhor acurácia e precisão foram novamente treinadas, considerando uma partição dos dados similar à descrita na Seção 3, e para a rede (8, 5, 18, 1) a acurácia resultante foi de 66.35% com precisão de 47.22%, e para a rede (8, 8, 3, 1) obteve-se acurácia 66.14% de com precisão de 21.32%. É importante salientar que a obtenção do novo conjunto de dados com uma proporção mais balanceada dos exemplos não foi feita de maneira exaustiva. Outras metodologias ainda podem ser aplicadas neste aspecto para derivar resultados mais satisfatórios.

6. Considerações Finais

Este trabalho utilizou redes neurais artificiais e dados de uma rede de estações meteorológicas para prever, com 30 minutos de antecedência, a ocorrência de precipitações na região central de Manaus a partir de informações oriundas de estações localizadas no entorno da cidade. Foram propostas e treinadas 523 redes neurais para o cenário em questão, das quais foi possível elencar duas delas, com arquiteturas (8, 5, 18, 1) e (8, 8, 3, 1), como preditoras com melhor acurácia e precisão nos testes realizados. Um novo treino com um subconjunto melhor balanceado dos dados permitiu o aumento na precisão, reflexo de uma melhor captura nos padrões de precipitação disponíveis.

Os resultados deste trabalho podem ser utilizados de maneira prática por órgãos de previsão do tempo e para geração de alertas junto à Defesa Civil, Bombeiros, Polícia de Trânsito, etc. para um melhor posicionamento de efetivo e tomada de decisão para prevenção de desastres e de alagamentos.

Em trabalhos futuros, será analisada a viabilidade de integrar as redes neurais identificadas à plataforma de gerenciamento de dados da rede de estações meteorológicas de Manaus. Além disso, novas investigações serão feitas com o intuito de estimar melhor um balanceamento dos exemplos do conjunto de treinamento e também para prever o volume das precipitações, aprimorando os resultados obtidos neste trabalho.

Agradecimentos

As autoras agradecem ao LabInstru pelos cessão dos dados utilizados neste trabalho. As autoras Rafaela dos Santos Souza e Liliana Oliveira da Silva agradecem o apoio financeiro provido pela Universidade do Estado do Amazonas e pela Fundação de Amparo à Pesquisa do Estado do Amazonas por meio do Programa de Apoio à Iniciação Científica.

Referências

- Bushara, N. O. and Abraham, A. (2013). Computational intelligence in weather forecasting: A review. *Journal of Network and Innovative Computing*, 1:320–331.
- da Silva, D. A. (2009). *A influência das áreas verdes no clima da cidade de Manaus*. PhD thesis, Universidade Federal do Amazonas.

- da Silva, D. A. (2012). Função da precipitação no conforto do clima urbano da cidade de manaus. *Revista Geonorte*.
- Darji, M. P., Dabhi, V. K., and Prajapati, H. B. (2015). Rainfall forecasting using neural network: A survey. In *International Conference on Advances in Computer Engineering and Applications*, India. IMS Engineering College.
- de Lima, P. M. (2016). Redes neurais para predição de séries temporais de precipitação em Manaus, Amazonas. Monografia de Engenharia de Computação, Universidade do Estado do Amazonas.
- de Lima, P. M. and Guedes, E. B. (2015a). Rainfall prediction for manaus, amazonas with artificial neural networks. In *Latin America Congress on Computational Intelligence*, Curitiba, Paraná.
- de Lima, P. M. and Guedes, E. B. (2015b). Uma abordagem baseada em redes neurais para a predição de chuva em manaus, amazonas. In *Congresso Amazônico de Computação e Sistemas Inteligentes*, pages 149–158, Manaus, Amazonas.
- de Oliveira, M. B. L. (2017). Laboratório de instrumentação meteorológica. Acessado em 11 de setembro de 2017.
- Gardner, M. W. and Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14/15):2627–2636.
- Guedes, E. B., de Lima, P. M., and de Oliveira, M. B. L. (2016). Neural networks for time series rainfall forecasting: A case study in manaus, amazonas. In *Encontro Nacional de Inteligência Artificial e Computacional*, pages 337–348, Recife, Pernambuco.
- Haykin, S. (2009). *Neural Networks and Learning Machines*. Pearson, Nova Jersey, 3 edition.
- Luk, K. C., Ball, J. E., and Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. *Mathematical and Computing Modelling*, 33:683–693.
- Nayak, D. R., Mahapatra, A., and Mishra, P. (2013). A survey on rainfall prediction using artificial neural network. *International Journal of Computer Applications*, 72(16):32–40.
- Palit, A. K. and Popovic, D. (2005). *Computational Intelligence in Time Series Forecasting - Theory and Engineering Applications*. Springer, Londres, 1 edition.
- Sioli, H. (1991). *Amazônia: Fundamentos da ecologia da maior região de florestas tropicais*. Vozes, Rio de Janeiro.