Use of Artificial Neural Network to calculate satellite link attenuation caused by rainfall

Heitor Eugênio Gonçalves, Alexandre Coutinho Mateus, Luciano Xavier Medeiros e André Luiz Aguiar da Costa

Resumo—A chuva é capaz de atenuar significativamente os sinais em links de satélite com frequências acima de 10 GHz. Nesse contexto, o setor de Radiocomunicações da International Telecommunication Union (ITU-R) desenvolveu um modelo para prever atenuações causadas por precipitação. No entanto, este método contém inúmeras operações complexas, incluindo potenciação, raiz quadrada, entre outras funções com alto custo computacional.

Este artigo propõe o uso de uma Rede Neural Artificial com baixo tempo de execução para calcular a atenuação. Esta rede foi desenvolvida após testes com diferentes quantidades de neurônios, funções de ativação, épocas, entre outros parâmetros. Os resultados obtidos comprovam que a Rede Neural é uma boa aproximação para o modelo ITU-R com um custo computacional muito menor.

Palavras-Chave—Atenuação por chuva, Rede Neural Artificial, Modelo ITU-R.

Abstract—Rain is able to significantly attenuate signals on satellite links with frequencies above 10 GHz. In this context, the Radiocommunication sector of the International Telecommunication Union (ITU-R) developed a model to predict attenuations caused by precipitation. However, this method contains numerous complex operations, including potentiation, square root, among other functions with high computational cost.

This article proposes the use of an Artificial Neural Network with low execution time to calculate attenuation. This network was developed after tests with different amounts of neurons, activation functions, epochs, among other parameters. The results obtained prove that the Neural Network is a good approximation for the ITU-R model with a much lower computational cost.

Keywords—Rain attenuation, Artificial Neural Network, ITU-R model.

I. Introdução

O uso de um satélite de comunicação permite o envio de informações a receptores que se encontram sobre a superfície terrestre, sem a necessidade de uma infraestrutura em solo para isso. Por esse motivo, satélites são utilizados em sistemas de rádio, de televisão, em redes móveis, GPS, entre diversas outras áreas.

Em enlaces via satélite com frequências acima de 10 GHz, a chuva é um fator extremamente prejudicial, capaz de causar

Heitor Eugênio Gonçalves, Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, Uberlândia, MG, 38400-902, Brasil, e-mail: heitor.goncalves@ufu.br; Alexandre Coutinho Mateus, Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, Uberlândia, MG, 38400-902, Brasil, e-mail: acmateus@ufu.br; Luciano Xavier Medeiros, Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, Uberlândia, MG, 38400-902, Brasil, e-mail: lucianox@ufu.br; André Luiz Aguiar da Costa, Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, Uberlândia, Uberlândia, MG, 38400-902, Brasil, e-mail: alacosta@ufu.br. Este trabalho foi possível graças a bolsa de Iniciação Científica fornecida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

atenuações significativas no sinal [1], [2]. Desse modo, muitos métodos foram criados com o intuito de prever esse tipo de atenuação. Entre eles, um dos mais utilizados é o modelo desenvolvido pelo *International Telecommunication Union - Radiocommunication Sector* (ITU-R). Este modelo é descrito na recomendação mais atual P. 618-13 [3], publicada em dezembro de 2017.

Destaca-se que o método do ITU-R foi desenvolvido a partir de extensa análise de dados de precipitação e atenuação em diversas regiões. Esses locais de coleta concentraram-se, principalmente, em países do continente Europeu.

As equações usadas pelo método do ITU-R, conforme será mostrado na Seção II, são extensas e trabalhosas, tendo um custo computacional elevado devido ao grande número e complexidade de operações. Uma forma de simplificar e reduzir os cálculos em relação aos modelos anteriormente citados é utilizando Rede Neural Artificial (RNA).

As Redes Neurais Artificiais são fundamentadas no funcionamento do sistema nervoso humano. Por meio das RNA é possível otimizar resultados, modelar dados, encontrar objetos em imagens, entre várias outras funcionalidades. Cita-se, por exemplo, a elaboração de redes para prever o preço das ações [4] e reconhecimento de padrões em imagens [5].

No trabalho de Bai e outros [6], uma rede neural é desenvolvida para calcular a atenuação de um canal de satélite na Banda Q. Para isso, são utilizados dados meteorológicos e de receptores instalados em Hampshire na Inglaterra. A RNA elaborada contém seis camadas intermediárias com 16 neurônios na primeira, 32 na segunda e quarta e 64 na terceira. Esta rede utilizou a função a média quadrática como função de erro.

Em Alencar et al. [7], foi construída uma Rede Neural Artificial para calcular as atenuações de enlace via satélite de frequência entre 11 e 20 GHz. Esses valores foram comparados com os obtidos pelo modelo do ITU-R e outras 4 variações desse modelo. A rede apresentou um erro quadrático médio que variou de 15% a 20%, enquanto, para as demais técnicas testadas, este erro esteve entre 30% e 40%. A camada intermediária da RNA foi desenvolvida com 15 neurônios, sendo adotada a função de ativação tangente hiperbólica.

O artigo desenvolvido por Yang et al. [8] compara os valores de atenuação previstos por RNA e pelo modelo do CCIR, que foi publicado pela ITU-R em 1990 [9]. Os resultados foram satisfatórios, apresentando um erro quadrático médio de 2,01 dB para a rede e 2,70 dB para o modelo de CCIR.

A desvantagem dos três últimos trabalhos citados está no fato de não variarem as funções de ativação e de erro com o objetivo de chegar ao melhor modelo de rede. Ademais,

estes artigos não especificam o tempo computacional gasto no cálculo das atenuações, o que é extramente importante quando o número de dados a ser utilizado é grande.

Neste trabalho é construído uma Rede Neural Artificial que pode ser utilizada quando é desejado obter rapidamente os valores de atenuações causadas pela chuva em enlaces via satélite. A RNA é treinada a partir de atenuações calculadas pelo modelo criado pela ITU-R. Esta rede é testada com diferentes funções de ativação e de erro, quantidades de épocas, neurônios. Além disso, este artigo utiliza uma grande variação de entradas, 1000 dados.

II. Fundamentação Teórica

A. Modelo do ITU-R

A Recomendação ITU-R-618-13 permite calcular a atenuação excedida em 0.01% ao ano a partir das etapas a seguir.

- 1) Determinar a altura média da camada de formação da chuva h_R em km.
- 2) Para um ângulo de elevação da antena, θ , maior ou igual a 5°, tem-se que a distância que o sinal percorreu com interferência da precipitação, L_S , em km é dada por:

$$L_S = \frac{h_R - h_S}{\operatorname{sen}\theta}. (1)$$

3) Para θ menor que 5°, L_S é, então, calculado por:

$$L_S = \frac{2(h_R - h_S)}{\sqrt{\sin^2 \theta + 2(h_R - h_S)/R_e} + \sin \theta}.$$
 (2)

onde $h_R - h_S \ge 0$.

- 4) Obter a projeção horizontal de do percusso, L_G , em que $L_G = L_S \cos \theta$.
- 5) Encontrar a taxa de atenuação específica de acordo com a Eq. (3), em que k e α são parâmetros que dependem de fatores como polarização, ângulo de elevação da antena e frequência do sinal, conforme pode ser observado na Recomendação ITU-R P.838-3 [10].

$$\gamma_R = k(R_{0,01})^{\alpha},\tag{3}$$

onde $R_{0,01}$ é a precipitação excedida em 0.01% ao ano em mm/h

6) Determinar o fator de redução horizontal, $r_{0,01}$, para 0,01% do ano por meio da seguinte equação:

$$r_{0,01} = \frac{1}{1 + 0.78\sqrt{\frac{L_G \gamma_R}{f}} - 0.38 \left(1 - e^{-2L_G}\right)}.$$
 (4)

7) Obter o fator de ajuste vertical, $v_{0,01}$, para 0,01% do ano a partir da Eq. (5)

$$\nu_{0,01} = \frac{1}{1 + \sqrt{\operatorname{sen}\theta} \left[31 \left(1 - e^{-\frac{\theta}{1+\chi}} \right) \frac{\sqrt{L_R \gamma_R}}{f^2} - 0, 45 \right]}.$$
 (5)

Para determinar os parâmetros da Eq. (5), calcula-se primeiramente o ângulo ζ , dado por:

$$\zeta = \tan^{-1} \left(\frac{h_R - h_S}{L_G r_{0.01}} \right). \tag{6}$$

Desse modo, se ζ é maior que zero, tem-se que

$$L_R = \frac{L_G r_{0,01}}{\cos \theta},\tag{7}$$

caso contrário,

$$L_R = \frac{h_R r_{0,01}}{\cos \theta}. (8)$$

Ademais, se a latitude da estação terrena, φ tem módulo menor que 36°, então,

$$\chi = 36 - |\varphi|,\tag{9}$$

senão, $\chi = 0^{\circ}$.

8) Calcular o comprimento efetivo do enlace, L_E , em que

$$L_E = L_R \nu_{0.01}. \tag{10}$$

9) Encontrar o valor da atenuação média excedida para 0.01% de um ano, $A_{0.01}$ como

$$A_{0.01} = \gamma_R L_E. {(11)}$$

B. Redes Neurais Artificiais

As Redes Neurais Artificiais são compostas, de forma geral, por um conjunto de camadas interligadas por meio de pesos. Essas camadas são formadas, com exceção da entrada, por neurônios cuja as saídas são determinadas por funções de ativação aplicadas sobre as combinações lineares dos dados ponderados que se associam a estas estruturas [11], [12].

Existem diversos tipos de Redes Neurais Artificiais e estas diferenciam-se por fatores como o modo em que é feito a correção dos pesos, o número e as conexões das camadas, entre outras condições. Nesse sentido, salienta-se que este trabalho utiliza uma rede neural conhecida como *Percéptron de Múltiplas Camadas* (PMC) [13].

O PMC é caracterizado por uma conjunto de entradas que conectam-se a camadas ocultas ou intermediárias, e estas interligam-se a camada de saída. Desse modo, os erros entre os valores de saída pela rede e os que eram esperados permitem realizar o ajuste dos pesos com o intuito de ensinar o PMC.

A correção dos pesos é feita pelo método de *backpropagation*, isto é, ocorre uma propagação do erro da camada de saída para as camadas anteriores. Outrossim, o negativo do gradiente da função custo - totalização dos erros - é utilizado para minimizar esta de forma rápida.

III. METODOLOGIA

A criação das Redes Neurais Artificiais pode ser divida em três etapas: *Obtenção de Dados para Teste e Treino, Escolha dos Parâmetros* e *Treinamento*.

A. Obtenção de Dados para Teste e Treino

Conforme foi apresentado na Seção II-B, o método proposto neste trabalho utiliza rede PMC com uma única saída, a atenuação causada pela chuva, que varia de acordo com as 7 entradas da rede: latitude, longitude, frequência, ângulo de elevação, altura da antena, precipitação excedida em 0,01% do ano e polarização do sinal.

Para treinar e testar a RNA, foram utilizados dados de 1000 variações de valores de entrada e, consequentemente, de atenuação calculada pelo modelo do ITU-R. Desta base de dados, 670 foram destinados ao treinamento, enquanto 330 para os testes.

B. Escolha dos Parâmetros

Com o propósito de criar redes neurais com bons resultados, foram escolhidos diferentes parâmetros para as RNA de modo a observar quais dentre estes fornecem os menores erros na saída. Desta forma, variou-se o número de neurônios nas camadas intermediárias, as funções de ativação, os otimizadores.

A RNA foi construída com 3 camadas ocultas. Destas, uma foi fixada com 6 neurônios e 2 tiveram a quantidade de neurônios variando em 3, 4 e 5. Paralelamente, as funções de ativação alteraram-se em função linear, *ReLU* e *Softplus*. Cada uma destas são detalhadas a seguir.

1) Função Linear Definida por

$$f(x) = a \cdot x,\tag{12}$$

em que *a*, chamado de coeficiente angular, determina a inclinação da reta. Assim, o seu valor é igual a tangente do ângulo formado pelo eixo das abcissas e a reta no sentido anti-horário.

O gráfico da Fig. 1 mostra a característica de uma função linear. Nota-se que a reta cruza necessariamente a origem do plano de coordenadas.

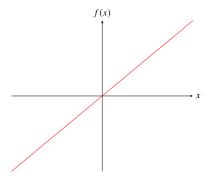


Fig. 1: Função Linear $f(x) = a \cdot x$ para a > 0

2) Função ReLU (Rectified Linear Unit) É não linear e tem 0 como limite inferior, isto é,

$$f(x) = \max(0, x). \tag{13}$$

A Fig. 2 representa a forma de uma função ReLU. Percebese que entradas negativas têm os seus valores zerados pela função de ativação ReLU. Dessa maneira, apenas neurônios com valores positivos são ativados.

Outrossim, os otimizadores, isto é, as formas usadas para corrigir os pesos, são baseados no método do *Gradiente de Descida Estocástico* (GDE). Esta técnica, proposta por Bottou [13], consiste em calcular o gradiente da função custo com apenas parte dos dados disponíveis por iteração [14]. Desse modo, amostras redundantes são evitadas e o algoritmo tornase mais rápido.

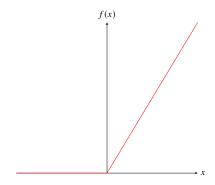


Fig. 2: Função ReLU

As variações do GDE empregadas nas PCM são o GDE com Momento e o Adam. Estes otimizadores são descritos a seguir.

1) GDE com Momento

Uma forma aumentar a rapidez da descida do gradiente em direção à minimização do custo, é utilizando um coeficiente conhecido como momento [15]. Assim, seja $J(\theta)$ a função custo, tem-se

$$v_{t+1} = \alpha v_t - \eta \nabla J(\theta) \tag{14a}$$

$$\theta_{t+1} = \theta_t + v_{t+1} \tag{14b}$$

em que $\alpha \in [0,1]$ é o momento e $\eta > 0$ é a taxa de aprendizado.

2) Adam (Adaptive Moment Estimation)

O algoritmo *Adam* aplica os momentos m_t e v_t , que são, respectivamente, a média e a variância não centrada dos gradientes [16]. Logo, tem-se

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{15a}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$
 (15b)

onde β_1 e β_2 são coeficientes de decaimento e g_t é o gradiente da função custo em t.

Em geral, os momentos são inicializados com zero, o que faz com que eles tendam a zero nos primeiros valores de t. Por esse motivo, usa-se os coeficientes \hat{m}_t e \hat{v}_t , de modo que

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t},\tag{16a}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t},\tag{16b}$$

para realizar a atualização de Adam por meio da Eq. 17

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t, \tag{17}$$

em que $\epsilon \ll 1$ evita que o denominador seja zero.

Nos algoritmos desenvolvidos neste trabalho são utilizados os valores $\beta_1=0,9,\ \beta_2=0,999$ e $\epsilon=10^{-7},$ recomendados por Kingma et al. em [16].

As funções de erro foram a média quadrática e absoluta, dadas, respectivamente, por

$$E_q = \frac{\sum_{i=1}^{n} (y_i - \dot{y}_i)^2}{n}$$
 (18)

$$E_a = \frac{\sum_{i=1}^{n} |y_i - \dot{y}_i|}{n} \tag{19}$$

onde n é o número de amostras, y_i é a saída da rede e \dot{y}_i é o valor esperado.

As RNA foram testadas com 2500 e 5000 épocas. Acentuase que o número de épocas é a quantidade de vezes conjunto de dados passou pela rede

C. Treinamento e Teste

Dentre os dados disponíveis, 66% são destinados para o treinamento das redes neurais e 33% para a realização de testes. As variáveis de entrada são a latitude, longitude, frequência, ângulo de elevação, altura da antena, precipitação excedida em 0,01% do ano e polarização do sinal.

D. Implementação do algoritmo

Para a escrita do algoritmo utilizado no trabalho, foi usada a linguagem *Python* na versão 3.7. Ademais, as bibliotecas *TensorFlow* e *Keras* foram empregadas na construção e treinamento das RNA.

O código foi implementado no IDE (*Integrated Development Environment*) Spyder 5.0.0, que foi executado no sistema operacional Linux Mint Debbie 4, em um computador com 4 GB de memória RAM e processador Intel Core i3-6006U.

IV. RESULTADOS

Ao combinar os diferentes parâmetros apresentados na Seção III, obteve-se 48 modelos de Redes Neurais. Dentre estas RNAs testadas, a que apresentou o menor erro absoluto médio é representada na Figura 3 e os seus parâmetros são descritos a seguir. Portanto, os próximos testes a serem descritos referemse a esta rede.

- Função de Ativação nas camadas intermediárias e de entrada = Relu.
- Função de Ativação na camada de saída = Linear.
- Otimizador = Adam.
- Número de Épocas = 5000.
- Quantidade de Neurônios = 4.

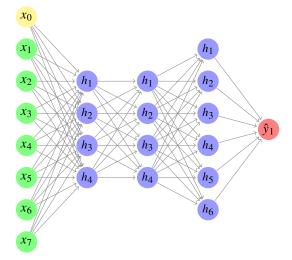


Fig. 3: Estrutura da RNA com os melhores parâmetros.

Para os 330 dados destinados para teste, os cálculos das atenuações por meio da RNA atingem um erro absoluto médio de 2,1254 dB. Além disso, a média dos valores de atenuação pelo modelo do ITU-R é 19,6281 dB, enquanto, pela RNA, é 19,1596 dB. Nota-se, portanto, que a rede aparenta realizar uma boa aproximação para o método.

A Tabela I mostra as porcentagens de valores determinados pela RNA em diferentes intervalos de erros absolutos. A partir da análise desta tabela, identifica-se que aproximadamente 33% das atenuações determinadas pela RNA menor que 1 dB de diferença dos valores calculados pelo modelo do ITU-R. Nota-se também que 93,9394% dos dados previstos pela rede são menores do que 5 dB. Assim, pode-se afirmar que a rede pode ser utilizada para o cálculo de atenuação por apresentar resultados próximos aos do modelo do ITU-R.

TABELA I: Intervalos de Erro.

Erro Absoluto (dB)	Porcentagem de valores (%)	
[0,1[31,8182	
[1,2[25,1515	
[2,3[19,0909	
[3,4[13,3333	
[4,5[4,5455	
[5,+∞)	6,0606	

A Figura 4 mostra de forma visual os valores de atenuação determinados pela rede e pelo método do ITU-R.

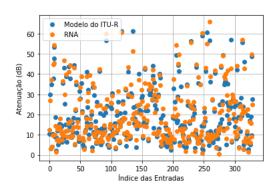


Fig. 4: Atenuações determinadas pelo modelo de ITU-R em comparação com a RNA

Ao observar a Figura 4, percebe-se que valores de atenuação muito elevados, acima de 50 dB, tendem a não serem determinados com precisão pela RNA. Isso ocorre porque poucos dados acima de 50 dB foram usados no treinamento da rede devido ao fato desses valores serem raros de ocorrerem na realidade. Não obstante, é notável que a RNA tem uma boa precisão para dados abaixo de 50 dB.

A Figura 5 mostra as curvas linearizadas das atenuações calculadas pelo modelo do ITU-R e pela RNA proposta neste trabalho. Para que seja possível analisar com uma maior nitidez a relação entre os gráficos, foram usados apenas 50 índices de entrada nessa imagem.

Com o intuito de comparar o custo computacional, aumentou-se o número de dados e comparou-se o tempo gasto

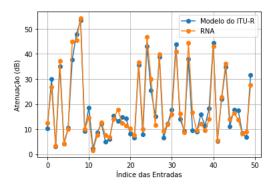


Fig. 5: Linearização das atenuações determinadas pelo modelo de ITU-R em comparação com a RNA

pelo modelo de ITU-R e pela RNA na determinação das atenuações. Estes valores de tempo estão representados na Tabela II.

TABELA II: Custo computacional do modelo do ITU-R e da RNA.

Algoritmo	Número de Dados	Tempo (s)
Modelo do ITU-R	660	1,2662
	1320	3,2991
	2640	9,9651
	5280	32,3829
RNA	660	0,1110
	1320	0,1405
	2640	0,1656
	5280	0,2883

Ao observar a Tabela II, identifica-se que para um total de 660 dados, o tempo empregado pelo método de ITU-R é de 1,2662 segundos. Este tempo é 11,4072 vezes maior que o gasto pela RNA.

Para 1320 dados houve um aumento significativo no tempo gasto pelo modelo da ITU-R em comparação com o processamento pela RNA. Isso torna-se mais evidente ao colocar 5280 dados, em que o tempo de processamento pelo modelo do ITU-R é de 32,3829 segundos, enquanto a rede gasta apenas 0,2883 segundos.

Por meio dos resultados mostrados na Tabela II, verifica-se que a RNA desenvolvida neste artigo contém, à medida que aumenta-se o número de dados, um custo computacional muito menor do que o modelo do ITU-R.

O menor custo computacional da RNA deve-se ao fato dela utilizar apenas somas e multiplicações. Estas operações possuem baixo tempo de execução, ao contrário, da potenciação, raízes quadradas, funções trigonométricas e logarítmicas que são usadas no modelo de ITU-R.

V. Conclusão

Para enlaces com frequências maiores que 10 GHz, a chuva é capaz de atenuar substancialmente o sinal. Para calcular esta atenuação, o modelo do ITU-R é bastante utilizado. Porém, esta técnica tem um elevado custo computacional devido a

complexidade de suas operações, que incluem potenciação, raízes quadradas, funções trigonométricas e logarítmicas. Pensando nisso, este artigo propõe a utilização de uma RNA com 7 entradas para o cálculo de atenuações.

Os 330 valores de atenuação determinados pela Rede Neural Artificial proposta neste trabalho apresentaram um erro absoluto médio de 2,1254 dB. Isso, indica que o método desenvolvido é uma boa aproximação para o modelo do ITU-R.

A vantagem de utilizar a RNA foi comprovada a partir da comparação do seu custo computacional com o modelo do ITU-R. Para um número de 5280 dados, por exemplo, a rede foi mais de 112 vezes mais rápida que o método do ITU-R.

Desse modo, é possível constatar que a RNA pode ser usada para o cálculo de atenuação por apresentar um custo computacional muito menor atenuações próximas as calculadas por meio do modelo do ITU-R.

Apêndice

Banco de dados para treinamento e teste https://drive.google.com/drive/folders/ 1aJLu-M0njdzwZsZwj4qCGOs1nAa84f7c

Referências

- [1] L. J. Ippolito Jr, Satellite communications systems engineering: atmospheric effects, satellite link design and system performance. John Wiley & Sons, 2017.
- [2] B. R. Elbert, The satellite communication applications handbook. Artech House. 2004.
- [3] Recommendation ITU-R P.618-13, "Propagation data and prediction methods required for the designof earth-space telecommunication systems," in *International Telecomunication Union*, December 2017.
- [4] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, "An innovative neural network approach for stock market prediction," *The Journal of Supercomputing*, vol. 76, no. 3, pp. 2098–2118, 2020.
- [5] M. Liang and X. Hu, "Recurrent convolutional neural network for object recognition," in *Proceedings of the IEEE conference on computer vision* and pattern recognition, pp. 3367–3375, 2015.
- [6] L. Bai, C.-X. Wang, Q. Xu, S. Ventouras, and G. Goussetis, "Prediction of channel excess attenuation for satellite communication systems at q-band using artificial neural network," *IEEE antennas and wireless* propagation letters, vol. 18, no. 11, pp. 2235–2239, 2019.
- [7] G. A. Alencar, L. P. Caloba, and M. Assis, "Artificial neural networks as rain attenuation predictors in earth-space paths," in *Proceedings of* the 2003 International Symposium on Circuits and Systems, ISCAS'03., vol. 5, IEEE, 2003.
- [8] H. Yang, C. He, W. Song, and H. Zhu, "Using artificial neural network approach to predict rain attenuation on earth-space path," in IEEE Antennas and Propagation Society International Symposium. Transmitting Waves of Progress to the Next Millennium. 2000 Digest. Held in conjunction with: USNC/URSI National Radio Science Meeting, vol. 2, pp. 1058–1061, IEEE, 2000.
- [9] P. Data, "Prediction methods required for earth-space telecommunication systems," *Proceedings of the CCIR Plenary Assembly, Dusseldorf, Germany*, pp. 447–505, 1990.
- [10] Recommendation ITU-R P.838-3, "Specific attenuation model for rain for use in prediction methods," in *International Telecomunication Union*, 2005
- [11] J. E. Dayhoff, Neural network architectures: an introduction. Van Nostrand Reinhold Co., 1990.
- [12] S. Haykin, Redes neurais: princípios e prática. Bookman Editora, 2007.
- [13] M. A. Nielsen, *Neural networks and deep learning*, vol. 25. Determination press San Francisco, CA, 2015.
- [14] L. Bottou, "Large-scale machine learning with stochastic gradient descent," pp. 177–186, 2010.
- [15] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv preprint arXiv:1609.04747, 2016.
- [16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.