

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA
PROJETO DE GRADUAÇÃO



PEDRO HENRIQUE DE OLIVEIRA SOARES

PROJETO DE REDES NEURAIS PARA
PROPAGAÇÃO NÃO LINEAR DE PULSOS EM FIBRA

VITÓRIA-ES

DEZEMBRO/2023

Pedro Henrique De Oliveira Soares

PROJETO DE REDES NEURAIIS PARA PROPAGAÇÃO NÃO LINEAR DE PULSOS EM FIBRA

Parte manuscrita do Projeto de Graduação do aluno Pedro Henrique De Oliveira Soares, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Vitória-ES

Dezembro/2023

Pedro Henrique De Oliveira Soares

PROJETO DE REDES NEURAIS PARA PROPAGAÇÃO NÃO LINEAR DE PULSOS EM FIBRA

Parte manuscrita do Projeto de Graduação do aluno Pedro Henrique De Oliveira Soares, apresentado ao Departamento de Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Engenheiro Eletricista.

Aprovado em 15 de dezembro de 2023.

COMISSÃO EXAMINADORA:



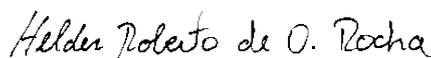
**Prof. Dr. Carlos Eduardo Schmidt
Castellani**

Universidade Federal do Espírito Santo
Orientador



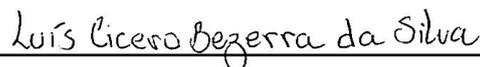
Eng. Gustavo Rossi Martins

Universidade Federal do Espírito Santo
Coorientador



**Prof. Dr. Helder Roberto de Oliveira
Rocha**

Universidade Federal do Espírito Santo
Examinador



Dr. Luís Cícero Bezerra da Silva
Universidade Federal do Espírito Santo
Examinador

Vitória-ES

Dezembro/2023

Aos meus avós, Nicácio Soares de Oliveira e Pedro Almerindo Soares.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, que me deu energia e benefícios para concluir todo esse trabalho.

Aos meus pais, Marcelo e Marcell, que me apoiaram durante toda a minha graduação, me incentivando a estudar e me dedicar, dando sempre o melhor de mim.

À minha irmã Marcella, que sempre demonstrou seu carinho e dedicação comigo.

À minha namorada Livia, que esteve sempre comigo durante todo o processo, mesmo em momentos difíceis.

Aos meus professores da UFES por todo o conhecimento, em especial ao meu orientador Prof. Carlos Eduardo, que me acompanha há anos em projetos de Iniciação Científica e esteve sempre disposto a me ajudar e tirar dúvidas.

Agradeço também ao Gustavo Martins, meu coorientador, que tirou diversas dúvidas sobre o projeto e colaborou fortemente com todas as etapas.

À banca examinadora pela aceitação do convite e pelo tempo investido para leitura e avaliação desse trabalho.

Agradeço à Universidade Federal do Espírito Santo pela minha formação.

RESUMO

Fibras óticas se tornaram a base dos sistemas de telecomunicações modernos, garantindo a interligação de continentes via cabos submarinos. Intensas pesquisas no campo foram capazes de reduzir as perdas deste meio para apenas 0,2 dB/km. Entretanto descobriu-se, também, uma gama de efeitos não lineares que podem ser destrutivos para os sistemas de comunicações. A simulação destes efeitos são baseados na equação não linear de Schrödinger (NLSE), que torna as simulações complexas. O método estado-da-arte para simulações atualmente é o *split-step* Fourier (SSF), que produz ótimos resultados, mas possui um tempo de execução muito grande, podendo se tornar um gargalo para simulações em tempo real ou para grandes comprimentos de fibra. O presente projeto faz uso de uma rede neural convolucional para tentar solucionar o problema de lentidão gerado pelo SSF. Para realizar a tarefa foram utilizadas as técnicas de pré-processamento de dados e foi feito um estudo aprofundado sobre redes convolucionais, detalhando seu funcionamento interno. Foi possível alcançar um coeficiente de determinação sobre o *dataset* de teste de 91,98%, demonstrando um bom ajuste do modelo à curva real.

Palavras-chave: Fibra óptica; Pulsos ópticos; *Deep-learning*; Redes neurais convolucionais.

ABSTRACT

Optical fibers have become the foundation of modern telecommunication systems, ensuring the interconnection of continents through submarine cables. Intensive research in the field has been able to reduce the losses of this medium to only 0.2 dB/km. However, it has also been discovered that a range of nonlinear effects can be destructive to communication systems. The simulation of these effects is based on the nonlinear Schrödinger equation (NLSE), which makes the simulations complex. Currently, the state-of-the-art method for simulations is the split-step Fourier (SSF), which produces excellent results but has a very long runtime, potentially becoming a bottleneck for real-time simulations or long fiber lengths. This project uses a convolutional neural network to solve the slowness problem generated by SSF. Data preprocessing techniques were applied to perform this task, and an in-depth study of convolutional networks was conducted, detailing their internal functioning. It was possible to achieve a determination coefficient on the test dataset of 91.98%, demonstrating a good fit of the model to the real curve.

Keywords: Optical fiber; Optical pulses; Deep-learning; Convolutional neural network.

LISTA DE FIGURAS

Figura 1 – Fibra Óptica.	15
Figura 2 – Neurônio <i>feed-forward</i>	19
Figura 3 – Funções de ativação sigmoideal, tangente hiperbólica e ReLU, respectivamente.	20
Figura 4 – Exemplo de rede <i>feed-forward</i> com duas camadas ocultas.	21
Figura 5 – Exemplo de convolução.	22
Figura 6 – Arquitetura da rede utilizada.	26
Figura 7 – Intensidade temporal do sóliton em distâncias selecionadas $P = 20, 5$ W e $T_{FWHM} = 1, 13$ ps.	28
Figura 8 – Intensidade temporal do sóliton $P = 20, 5$ W e $T_{FWHM} = 1, 13$ ps.	29
Figura 9 – Intensidade temporal do sóliton em distâncias selecionadas $P = 29, 3$ W e $T_{FWHM} = 1, 13$ ps.	29
Figura 10 – Intensidade temporal do sóliton $P = 29, 3$ W e $T_{FWHM} = 1, 13$ ps.	30
Figura 11 – Propagação do erro da rede neural.	30

LISTA DE TABELAS

Tabela 1 – Descrição dos dados gerados	24
Tabela 2 – Descrição dos dados normalizados	25
Tabela 3 – Performance da rede neural	28
Tabela 4 – Comparação de tempo de execução do método SSF e tempo de predição da rede em segundos	31

LISTA DE ABREVIATURAS E SIGLAS

CNN	<i>Convolutional Neural Network</i>
FFT	<i>Fast Fourier Transform</i>
FWHM	<i>Full Width at Half Maximum</i>
GVD	<i>Group-Velocity Dispersion</i>
MSE	Erro Quadrático Médio
NLSE	<i>Nonlinear Schrödinger Equation</i>
PINN	<i>Physical informed neural network</i>
R^2	Coefficiente de Determinação
RMSE	Raiz do Erro Quadrático Médio
RNA	Redes neurais artificiais
RNN	<i>Recurrent neural network</i>
SSF	<i>Split-step Fourier</i>
TOD	<i>Third-Order Dispersion</i>
UFES	Universidade Federal do Espírito Santo

LISTA DE SÍMBOLOS

a	Raio do núcleo
b	Raio da casca
n_1	Índice de refração do núcleo
n_2	Índice de refração da casca
Φ	Ângulo de incidência
Φ_c	Ângulo crítico
$A(z, T)$	Amplitude do pulso (W)
α	Atenuação total da fibra ótica
β_2	Parâmetro de dispersão GVD (s^2m^{-1})
β_3	Parâmetro de dispersão TOD (s^3m^{-1})
γ	Parâmetro não linear ($W^{-1}m^{-1}$)
$U(z, T)$	Amplitude normalizada do pulso (W)
α_{dB}	Atenuação total da fibra ótica (dB/km)
P_0	Potência de pico do pulso incidente (W)
T_0	Largura do pulso incidente (s)
T_{FWHM}	Largura à meia altura do pulso (s)
N	Ordem do sóliton

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Apresentação	12
1.2	Objetivos	13
2	REFERENCIAL TEÓRICO	15
2.1	Fibra Óptica	15
2.1.1	Perdas na fibra	15
2.1.2	Dispersão cromática	16
2.1.3	Não linearidades da fibra	16
2.1.4	Propagação do Pulso	17
2.2	Método <i>Split-Step</i> Fourier	18
2.3	Redes Neurais Artificiais	19
2.3.1	Funções de ativação	19
2.3.2	Redes Neurais <i>Feed-Forward</i>	20
2.3.3	Redes Neurais Convolucionais	21
3	PROPOSTA	23
3.1	Modelagem da Fibra	23
3.2	Descrição dos dados	23
3.2.1	Normalização	24
3.3	Estrutura da Rede Neural	25
4	RESULTADOS	27
4.1	Introdução	27
4.2	Métricas de Desempenho	27
4.3	Desempenho da rede	28
4.3.1	Ajuste do modelo	28
4.3.2	Propagação de erro	30
4.3.3	Tempo de execução	31
5	CONCLUSÃO E TRABALHOS FUTUROS	32
5.1	Conclusão	32
5.2	Temas a serem pesquisados	32
	REFERÊNCIAS	34

1 INTRODUÇÃO

1.1 Apresentação

Os sistemas de comunicação modernos estão cada vez mais dependentes de velocidades mais altas, confiabilidade aprimorada e maior largura de banda na transmissão de dados. No entanto, essas demandas têm impulsionado a necessidade de explorar novos meios de transmissão de informações. Na década de 1950, surgiu o campo de pesquisa das fibras ópticas, que possibilitou o desenvolvimento contínuo de fibras cada vez melhores e mais viáveis no mercado (AGRAWAL, 2019).

Inicialmente, as primeiras fibras ópticas apresentavam alta perda (mais de 1000 dB/km), o que limitava sua aplicabilidade, pois em poucos quilômetros todo o sinal era perdido. No entanto, ao longo dos anos, as perdas foram sendo reduzidas progressivamente. Um marco significativo foi alcançado em 1979, quando a perda foi reduzida para 0,2 dB/km para comprimentos de onda de 1,55 μm , o que permitiu uma ampla utilização das fibras ópticas no campo das telecomunicações (CASTELLANI, 2013).

Além dos grandes avanços na diminuição das perdas na fibra, foram descobertos efeitos não lineares gerados a partir da interação entre a luz e o meio. Tais efeitos são um grande campo de pesquisa na área e já são aplicados para a implementação de sistemas de chaveamento e sensoriamento (AGRAWAL, 2019). Para simular esses efeitos e estudar novas aplicações é preciso realizar simulações computacionais do modelo, que se tornam computacionalmente exaustivas, uma vez que é preciso basear-se na equação de Schrödinger não linear (NLSE, do inglês *nonlinear Schrödinger equation*), que pode se tornar bastante complexa conforme o número de ordens simuladas aumenta.

Atualmente, a melhor solução para esse problema é o algoritmo *split-step* Fourier (SSF), que utiliza transformadas de Fourier para resolver a equação, dividindo a fibra em pequenos segmentos e solucionando-a para cada um (ALESHAMS; ZARIFKAR; SHEIKHI, 2005). No entanto, esse método pode ser demorado para sistemas de longa distância, representando um desafio quando o objetivo é simular múltiplas voltas da luz em uma cavidade. Neste projeto de pesquisa, propõe-se a utilização de redes neurais convolucionais (CNN, do inglês *convolutional neural network*) para a simulação da propagação de pulsos ópticos, oferecendo uma solução mais rápida e adequada para o problema.

O *machine learning* tem sido amplamente utilizado para a solução de diversos problemas complexos, como processamento de linguagem natural e visão computacional, devido a sua capacidade de abstrair as informações e fornecer soluções muito otimizadas. As redes

neurais já têm se mostrado presentes na área da fotônica há alguns anos, sendo utilizadas diversas arquiteturas diferentes para solucionar o problema.

Os projetos de Martins et al. (2022) e Salmela et al. (2021), por exemplo, utilizam redes recorrentes (RNN, do inglês *recurrent neural network*), tratando o problema de forma similar ao SSF, gerando a saída de cada passo baseado na saída do passo anterior. Já os trabalhos de Zhou e Yan (2021) e Jiang et al. (2022) tratam o problema com a utilização de redes neurais fisicamente informadas (PINN, do inglês *physical informed neural network*), que utilizam as equações diferenciais que resolvem a propagação como parte da função de perda da rede.

Este projeto faz o uso de CNN para resolver a propagação do pulso, com o objetivo de obter, além de um baixo erro, uma maior velocidade em comparação com o SSF, visto que esta arquitetura é capaz de reduzir drasticamente a quantidade de pesos utilizados pela rede, reduzindo a quantidade de memória e processamento utilizados. O algoritmo SSF será utilizado para gerar o banco de dados para o treinamento da rede. Esse banco de dados será gerado a partir da simulação de diversas amostras de pulso sendo propagados separadamente na fibra.

Serão utilizadas métricas como erro quadrático médio (MSE, do inglês *mean squared error*), sua raiz quadrada (RMSE, do inglês *root mean squared error*) e o coeficiente de determinação R^2 . As métricas serão avaliadas em função dos valores obtidos do SSF, permitindo avaliar a eficácia da rede quando comparada a modelos atuais de cálculo.

1.2 Objetivos

Objetivo Geral

- O desenvolvimento de uma rede neural capaz de modelar a propagação de pulsos em fibra óptica. Para isso, pretende-se utilizar redes convolucionais, para um melhor desempenho do modelo.

Objetivos Específicos

- Estudar efeitos não lineares em fibra;

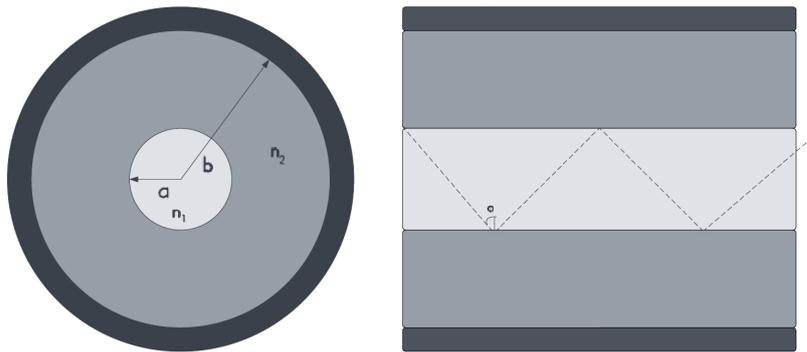
- Implementar um modelo com boa acurácia e que funcione para uma ampla variedade de dados de entrada, como potência de entrada e largura do pulso;
- Obter um modelo capaz de gerar resultados de forma mais rápida que o método SSF.

2 REFERENCIAL TEÓRICO

2.1 Fibra Óptica

A fibra óptica consiste basicamente de um núcleo de vidro (*core*) com raio a e uma casca (*cladding*) com raio b , onde o índice de refração n_2 da casca é ligeiramente menor que o índice de refração n_1 do núcleo, conforme ilustrado na Figura 1. Para fibras de baixas perdas, tanto o núcleo quanto a casca são feitos de sílica (SiO_2), sendo utilizados dopantes para controlar o índice de refração em cada região. Entretanto, para a geração de efeitos não lineares é comum se utilizar elementos calcogênicos na produção da fibra. O germânio (Ge) e o fósforo (P) são dopantes comuns no núcleo, pois aumentam o índice de refração da sílica pura. Na casca, dopantes como boro e flúor são utilizados, pois são capazes de reduzir o índice de refração (AGRAWAL, 2019).

Figura 1 – Fibra Óptica.



Fonte: Produção do próprio autor.

2.1.1 Perdas na fibra

Um parâmetro importante nas fibras ópticas são as perdas durante a transmissão de sinais. A equação (2.1) descreve o cálculo da potência transmitida P_T na fibra, considerando o comprimento L , a potência inicial P_0 , e o coeficiente de perda α ,

$$P_T = P_0 \cdot \exp(-\alpha L). \quad (2.1)$$

O coeficiente α é determinado pela soma das perdas de todas as fontes que ocorrem dentro da fibra, dado o comprimento de onda da luz propagante. Cada tipo de fibra tem um comprimento de onda ótimo para a propagação da luz, onde são alcançadas as menores perdas. Um exemplo comum é a fibra de modo único de sílica (*single-mode silica fiber*), que possui um coeficiente de perda mínimo (aproximadamente 0,2dB/km) próximo ao

comprimento de $1,55\mu\text{m}$. Isso significa que após a propagação em uma fibra de 50km, ainda restaria 10% da potência inicial do pulso (CASTELLANI, 2013).

2.1.2 Dispersão cromática

A resposta da interação de uma onda eletromagnética com um meio dielétrico depende da frequência da onda. Essa propriedade, conhecida como dispersão cromática, é baseada na frequência de ressonância em que o meio absorve a radiação eletromagnética devido à oscilação dos elétrons. A equação de Sellmeier (2.2) descreve a relação entre o índice de refração da onda n e a frequência angular ω , onde β_j é a intensidade da ressonância de número j e m representa o número de componentes que serão calculadas na série (AGRAWAL, 2019),

$$n^2(\omega) = 1 + \sum_{j=1}^m \frac{\beta_j \omega_j^2}{\omega_j^2 - \omega^2}. \quad (2.2)$$

Podemos expandir a equação (2.2) usando a série de Taylor para obter as componentes de propagação β , como mostrado nas equações (2.3) e (2.4),

$$\beta(\omega) = n(\omega) \frac{\omega}{c} = \beta_0 + \beta_1(\omega - \omega_0) + \frac{1}{2}\beta_2(\omega - \omega_0)^2 + \dots, \quad (2.3)$$

onde

$$\beta_m = \left(\frac{d^m \beta}{d\omega^m} \right)_{\omega=\omega_0}, \quad (m = 0, 1, 2, \dots). \quad (2.4)$$

Para este projeto, em que será utilizada a propagação de pulsos em fibras ópticas, as componentes β_2 e β_3 serão utilizadas no equacionamento. A componente β_2 está diretamente relacionada ao alargamento dos pulsos, também conhecida como dispersão de velocidade de grupo (GVD, *group-velocity dispersion*). Já a componente β_3 é utilizada para a propagação de pulsos ultracurtos ($T_0 < 1\text{ps}$) e é conhecida como dispersão de terceira ordem (TOD, *third-order dispersion*) (AGRAWAL, 2019).

2.1.3 Não linearidades da fibra

Fibras ópticas, assim como qualquer meio dielétrico, responde de forma não linear a campos magnéticos intensos. Essa resposta ocorre devido ao movimento gerado nos elétrons do meio pelo campo elétrico aplicado. É possível calcular a polarização P induzida no campo elétrico E a partir da equação (2.5),

$$P = \epsilon_0 (x^{(1)} \cdot E + x^{(2)} \cdot E^2 + x^{(3)} \cdot E^3 + \dots), \quad (2.5)$$

onde ϵ_0 é a permissividade do vácuo e $x^{(j)}$ é a susceptibilidade de ordem j . O índice de susceptibilidade linear $x^{(1)}$ é incluído por meio do índice de refração n e o coeficiente de atenuação α . O índice de segunda ordem $x^{(2)}$ é o responsável por efeitos não lineares como a geração de segundo harmônico e *sum-frequency generation*. Em geral, os efeitos de segunda ordem não ocorre em fibras comuns de sílica, pois o SiO_2 é uma molécula simétrica, anulando tais efeitos. Efeitos de ordens maiores se tornam cada vez mais fracos e difíceis de serem observados, exigindo ondas de alta potência e condutores específicos (AGRAWAL, 2019).

2.1.4 Propagação do Pulso

A propagação de pulsos ópticos em fibra é regida pela equação de Schrödinger não linear (NLSE), que é representada por (AGRAWAL, 2019)

$$i \frac{\partial A}{\partial z} = \frac{\beta_2}{2} \frac{\partial^2 A}{\partial T^2} + i \frac{\beta_3}{6} \frac{\partial^3 A}{\partial T^3} - i \frac{\alpha}{2} A - \gamma |A|^2 A, \quad (2.6)$$

onde β_2 e β_3 estão ligados aos efeitos de dispersão, α representa as perdas na propagação, γ é o coeficiente das não linearidades da interação do pulso com o meio, A é o envelope do pulso e T é a medida de tempo em relação ao *frame* de referência do envelope óptico (AGRAWAL, 2019).

Por questões de simplificação, serão ser consideradas apenas duas soluções: pulso gaussiano e sóliton. Ambos os pulsos podem ser descritos pela largura inicial T_0 e sua potência de pico incidente P_0 . O pulso gaussiano normalizado pode ser representado pela equação (2.7), onde $U(0, T)$ representa o pulso normalizado,

$$U(0, T) = \exp\left(-\frac{T^2}{2T_0^2}\right). \quad (2.7)$$

Outra forma de se caracterizar um pulso é a partir do T_{FWHM} (FWHM, de *full width at half maximum*), que representa a largura temporal do pulso na altura da metade de sua potência. O T_{FWHM} se relaciona com o T_0 , em pulsos gaussianos, a partir da equação (2.8),

$$T_{FWHM} = 2\sqrt{\ln(2)}T_0 \approx 1,665T_0. \quad (2.8)$$

A equação que descreve a secante hiperbólica está representado em (2.9),

$$U(0, T) = N \operatorname{sech}\left(\frac{T}{T_0}\right), \quad (2.9)$$

onde N , no contexto de sólitons, é sua ordem, que é representada por (2.10),

$$N^2 = \frac{\gamma P_0 T_0^2}{|\beta_2|}. \quad (2.10)$$

Para a secante hiperbólica, a relação entre T_0 e T_{FWHM} pode ser descrita pela Equação (2.11).

$$T_{\text{FWHM}} = 2 \ln(1 + \sqrt{2})T_0 \approx 1,763T_0. \quad (2.11)$$

O sóliton fundamental, ou seja, uma onda solitária que se propaga sem sofrer alterações, possui um formato natural de secante hiperbólica e é de ordem $N = 1$. A propagação sem alterações ocorrerá somente em situações ideais, isto é, γ , P_0 e β_2 são constantes. Para sólitons de ordem maior, haverá mudanças periódicas no pulso (CASTELLANI, 2013).

2.2 Método *Split-Step Fourier*

A NLSE, por se tratar de um problema de equação diferencial parcial não linear, o sistema não possui solução analítica trivial, exigindo, portanto, uma solução numérica para resolvê-lo. O método estado da arte utilizado para resolver este problema é o *split-step Fourier* (SSF), que utiliza da transformada rápida de Fourier (FFT, de *fast Fourier transform*), para calcular os efeitos de propagação não linear de forma rápida.

Para facilitar o entendimento, podemos reescrever a Equação (2.6) como

$$\frac{\partial A}{\partial z} = (\hat{D} + \hat{N})A, \quad (2.12)$$

onde,

$$\hat{D} = -\frac{i\beta_2}{2} \frac{\partial^2}{\partial T^2} + \frac{\beta_3}{6} \frac{\partial^3}{\partial T^3} - \frac{\alpha}{2}, \quad (2.13)$$

e

$$\hat{N} = i\gamma|A|^2. \quad (2.14)$$

Nesse caso, \hat{D} representa o operador de dispersões e perdas lineares e \hat{N} é o operador não linear de propagação. A partir disso, o método SSF supõe que os efeitos ocorrem de forma independente, de modo que a fibra possa ser dividida em pequenos trechos de comprimento $2h$ e, para cada um deles, é possível supor que na primeira metade $\hat{N} = 0$ e na segunda metade $\hat{D} = 0$. Dessa forma, pode-se simplificar as equações e solucioná-las individualmente.

A propagação do pulso para cada trecho é dada por (AGRAWAL, 2019)

$$A(z + h, T) = \exp[h(\hat{D}\hat{N})]A(z, T). \quad (2.15)$$

Utilizando a fórmula de Baker-Campbell-Hausdorff e aplicando transformadas rápidas de Fourier a cada passo obtemos (AGRAWAL, 2019)

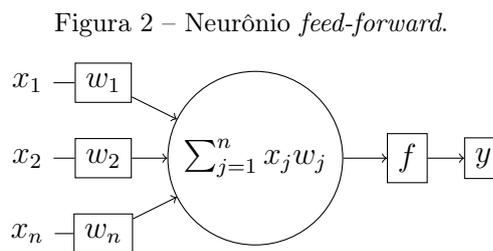
$$A(L, T) \approx \left(\prod_{m=1}^M e^{h\hat{D}} e^{h\hat{N}} \right) A(0, T), \quad (2.16)$$

onde L é o comprimento da fibra, M é o número total de passos e $A(0, T)$ e $A(L, T)$ representam o pulso óptico na entrada e saída da fibra, respectivamente.

2.3 Redes Neurais Artificiais

As redes neurais artificiais (RNA) são modelos inspirados no cérebro humano, capazes de reconhecer padrões em conjuntos de dados. Esses modelos artificiais consistem na estruturação de uma rede a partir de pequenas unidades conhecidas como neurônios, que geram valores reais de ativação.

O modelo mais simples de neurônio, o perceptron de Rosenblatt (ROSENBLATT, 1958), introduz o conceito de múltiplas entradas x_n , onde cada uma é multiplicada por um peso ω_n e somada, resultando em uma saída de $f(\sum_n \omega_n x_n)$, onde f é a função de ativação do perceptron.



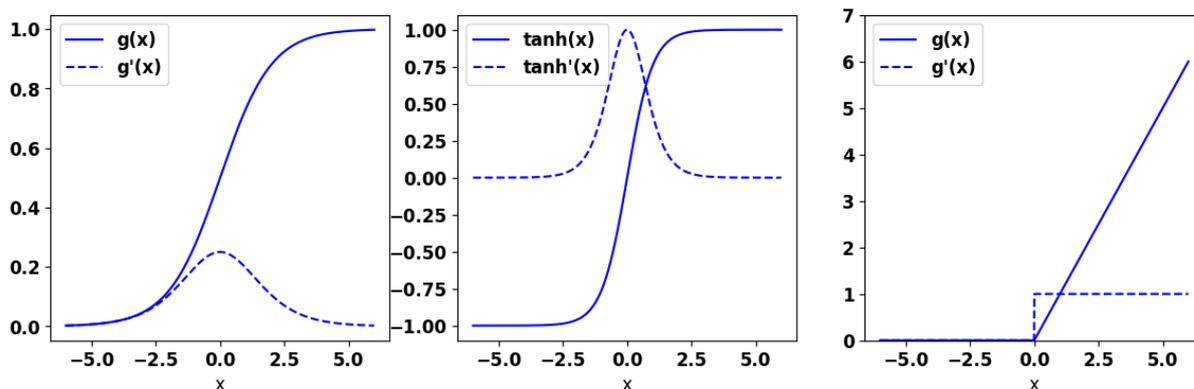
Fonte: Produção do próprio autor.

2.3.1 Funções de ativação

Uma rede neural natural funciona a partir da geração de impulsos elétricos que podem ou não, de acordo com sua intensidade, estimular o neurônio. No caso da RNA, esse *threshold* é simulado a partir de funções de ativação, que definem se o próximo neurônio deve ser ativado (DING; QIAN; ZHOU, 2018). Antes da utilização de funções de ativação as RNA eram somente capazes de solucionar problemas lineares, dessa forma, essas funções foram cruciais para criação de tecnologias modernas baseadas em redes neurais, além de serem grandes colaboradoras para a popularização das RNA.

Entre as funções de ativação mais comuns estão a sigmoide, tangente hiperbólica e ReLU, que estão representadas por suas funções e derivadas na Figura 3.

Figura 3 – Funções de ativação sigmoidal, tagente hiperbólica e ReLU, respectivamente.



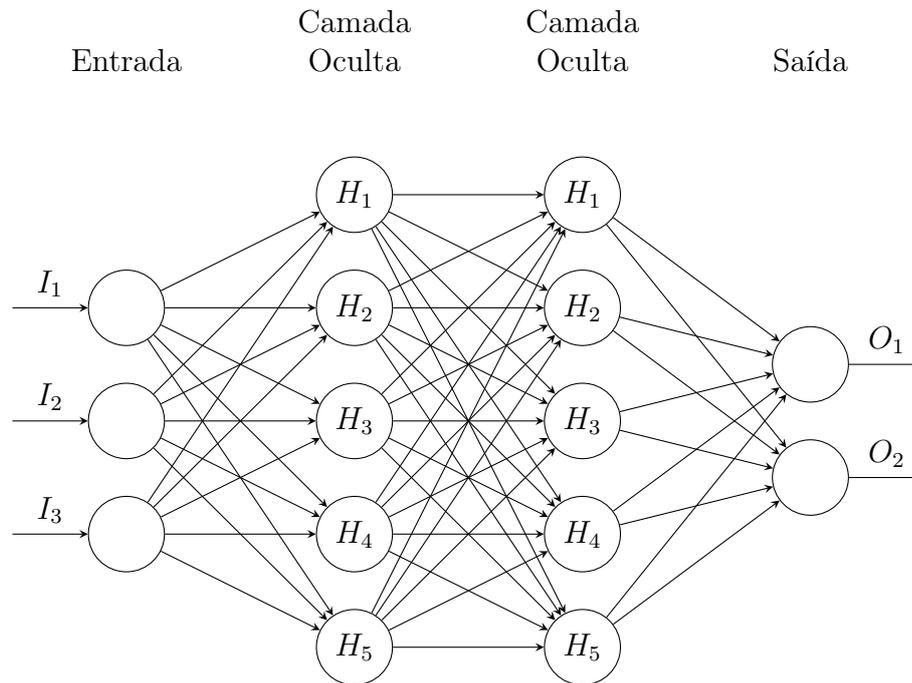
Fonte: Produção do próprio autor.

2.3.2 Redes Neurais *Feed-Forward*

As redes *feed-forward* são modelos que consistem em uma ou mais camadas intermediárias de perceptrons, onde os dados são processados por várias camadas, cada uma com n neurônios, permitindo uma maior variedade na complexidade dos problemas que podem ser solucionados (KROGH, 2008).

A rede neural profunda, uma rede com duas ou mais camadas intermediárias exemplificada na Figura 4, é capaz de combinar uma estrutura simples, em que cada perceptron em uma camada está conectado a todos os perceptrons da camada anterior e da camada seguinte, com uma capacidade de abstração maior, permitindo solucionar problemas de alta complexidade. No entanto, redes profundas com muitas camadas podem enfrentar problemas durante o treinamento, devido ao aumento do número de pesos a serem ajustados, podendo gerar a necessidade de uma grande quantidade de memória e processamento disponível para treinar a rede.

Esse tipo de rede é muito utilizada ainda para solucionar problemas de classificação ou regressão, por exemplo, pois costuma ser mais simples de ser implementada, entretanto, é deixada de lado quando o problema trata de soluções matriciais, pois começa a aumentar drasticamente o número de pesos conforme o tamanho da entrada ou número de camadas ocultas aumenta.

Figura 4 – Exemplo de rede *feed-forward* com duas camadas ocultas.

Fonte: Produção do próprio autor.

2.3.3 Redes Neurais Convolucionais

As redes neurais convolucionais se tornaram um dos modelos mais representativos de rede no campo do *deep learning*. Amplamente utilizada na área de visão computacional, esse modelo permitiu a construção de tecnologias como reconhecimento facial e veículos autônomos, que eram considerados impossíveis há poucas décadas (LI et al., 2022).

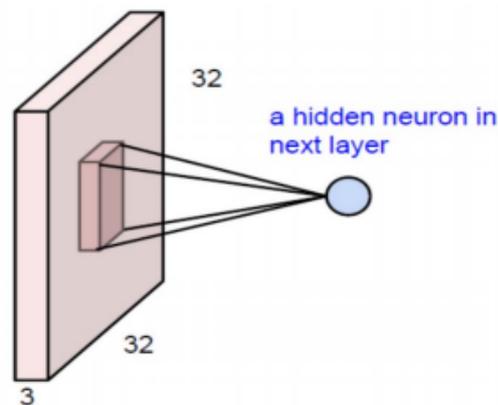
A grande vantagem apresentada pela CNN está na sua capacidade de reduzir drasticamente a quantidade de pesos da rede, permitindo trabalhar com modelos maiores e utilizando menor quantidade de memória e processamento. Por exemplo, supondo uma rede *feed-forward* tradicional com entrada de imagem colorida de tamanho $32 \times 32 \times 3$ pixels, para conectar a camada de entrada com um único neurônio serão necessários $32 \times 32 \times 3$ pesos. Para cada novo neurônio conectado na próxima camada, serão adicionados novos $32 \times 32 \times 3$ pesos, ou seja, para apenas dois neurônios, são utilizados mais de 6000 pesos (ALBAWI; MOHAMMED; AL-ZAWI, 2017).

Sabe-se que uma rede com somente dois neurônios é, geralmente, insuficiente para processar imagens de 32×32 pixels, considerando, então, o número de conexões para uma nova

camada de 32×32 neurônios, serão necessários $(32 \times 32 \times 3) \times (32 \times 32)$, totalizando um total de 3.145.728 parâmetros (ALBAWI; MOHAMMED; AL-ZAWI, 2017).

A fim de otimizar esse modelo, a CNN foi projetada para utilizar, ao invés de neurônios totalmente conectados, um *kernel*, ou seja, cada neurônio da próxima camada se conecta à $N \times N$ neurônios da camada anterior, onde N é o tamanho do *kernel* utilizado, como ilustrado na Figura 5. Pensando no mesmo problema citado, mas agora utilizando uma CNN com *kernel* de tamanho 5×5 , a quantidade de parâmetros para ligar as camadas seria de 76.800, ou seja, um valor 98% menor.

Figura 5 – Exemplo de convolução.



Fonte: Albawi, Mohammed e Al-Zawi (2017)

3 PROPOSTA

3.1 Modelagem da Fibra

O algoritmo SSF utilizado foi o mesmo desenvolvido por Martins et al. (2022). Os parâmetros de simulação foram ajustados para um problema de propagação em uma fibra de 12 metros, altamente não linear. Os coeficientes utilizados foram de $\alpha_{dB} = 0,05 dBm^{-1}$, $\beta_2 = -5,23 \times 10^{-27} s^2 m^{-1}$ (regime de dispersão anômalo), $\beta_3 = 4,27 \times 10^{-41} s^3 m^{-1}$, e $\gamma = 18,4 \times 10^{-3} W^{-1} m^{-1}$, similares ao usado por Salmela et al. (2021) para realizar uma compressão de soliton superior. O perfil do pulso utilizado é de uma secante hiperbólica, com potência P_0 e largura de pulso T_0 variáveis.

Esse modelo simula um caso extremo de propagação de lasers pulsados, que possui uma dinâmica não linear complexa. Este caso foi escolhido pois consiste de um modelo numérico complexo, dificultando a modelagem da solução.

3.2 Descrição dos dados

Para a geração da base de dados, a potência de pico na entrada da fibra P_0 foi variada no intervalo de 18,41 a 34,19 W, e a duração T_{FWHM} varia de 0,77 a 1,43 ps (obtendo T_0 no intervalo de 0,44 a 0,81 ps), caracterizando um soliton de ordem N entre 3,4 e 8,9 (SALMELA et al., 2021).

A base de dados gerada simula 3000 pulsos com diferentes condições iniciais e armazena os valores simulados em diversos pontos da fibra. A fim de obter um baixo erro gerado pelo SSF, a fibra foi dividida em 1000 passos e o pulso foi amostrado com 1024 pontos. Utilizando essas dimensões citadas, será obtida uma base de dados com mais de $3,072 \times 10^9$ de elementos ($3000 \times 1000 \times 1024$). Essa quantidade de dados, apesar de gerar bons resultados para o treinamento da rede, torna o processo muito lento, além de exigir sistemas com grande quantidade de memória para o treinamento.

A fim de facilitar o processo de treino da rede, os dados foram subamostrados, de modo que o pulso fosse amostrado com somente 128 pontos e com 101 passos realizados na fibra. Feito isso, teremos uma base de dados com $3,88 \times 10^7$ elementos, diminuindo drasticamente o tempo e recursos necessários para o treinamento da rede neural.

Tabela 1 – Descrição dos dados gerados

	0	1	2	3	...	124	125	126	127
0	0,1437	0,1588	0,1754	0,1938	...	0,2140	0,1937	0,1754	0,1588
1	0,1437	0,1588	0,1754	0,1938	...	0,2140	0,1937	0,1754	0,1588
2	0,1437	0,1588	0,1754	0,1938	...	0,2140	0,1937	0,1754	0,1587
3	0,1437	0,1588	0,1754	0,1937	...	0,2140	0,1937	0,1754	0,1587
4	0,1436	0,1588	0,1754	0,1937	...	0,2140	0,1937	0,1754	0,1587
...
96	0,1400	0,1555	0,1701	0,2064	...	0,2068	0,1877	0,1877	0,1545
97	0,1400	0,1555	0,1700	0,2063	...	0,2049	0,1893	0,1689	0,1560
98	0,1401	0,1552	0,1701	0,2064	...	0,2037	0,1903	0,1678	0,1570
99	0,1402	0,1550	0,1703	0,2065	...	0,2037	0,1900	0,1677	0,1569
100	0,1405	0,1545	0,1705	0,2067	...	0,2049	0,1887	0,1687	0,1558

Fonte: Produção do próprio autor.

A base de dados foi dividida em 2900 simulações para o treinamento e 100 simulações para testes. Cada uma das simulações é representada por um vetor de entrada da rede de tamanho 128 e uma matriz de saída de tamanho 101×128 .

A Tabela 1 retrata a estrutura de uma das simulações que serão utilizadas para o treinamento da rede.

3.2.1 Normalização

A normalização é um processo fundamental para um bom resultado do treinamento de uma rede, acelerando o treinamento e aumentando as chances de conversão de um modelo (ACHILLES; BONFIGLIOLI, 2018).

O processo de normalização transforma toda a base de dados em um conjunto de valores entre 0 e 1, e pode ser descrita por (3.1)

$$X_{std} = \frac{X - X_{min}}{X_{max} - X_{min}}, \quad (3.1)$$

onde X se refere à base de dados e X_{min} e X_{max} se referem ao menor e maior valor encontrados na base, respectivamente. A Tabela 2 ilustra os dados normalizados obtidos.

Tabela 2 – Descrição dos dados normalizados

	0	1	2	3	...	124	125	126	127
0	0,00046	0,00050	0,00061	0,00068	...	0,00068	0,00061	0,00056	0,00050
1	0,00046	0,00050	0,00061	0,00068	...	0,00068	0,00061	0,00056	0,00050
2	0,00046	0,00050	0,00061	0,00068	...	0,00068	0,00061	0,00056	0,00050
3	0,00046	0,00050	0,00061	0,00068	...	0,00068	0,00061	0,00056	0,00050
4	0,00046	0,00050	0,00061	0,00068	...	0,00068	0,00061	0,00056	0,00050
...
96	0,00044	0,00049	0,00054	0,00065	...	0,00066	0,00059	0,00054	0,00049
97	0,00044	0,00049	0,00054	0,00065	...	0,00065	0,00060	0,00054	0,00049
98	0,00044	0,00049	0,00054	0,00065	...	0,00065	0,00060	0,00053	0,00050
99	0,00044	0,00049	0,00054	0,00065	...	0,00065	0,00060	0,00053	0,00050
100	0,00045	0,00049	0,00054	0,00065	...	0,00065	0,00060	0,00053	0,00049

Fonte: Produção do próprio autor.

3.3 Estrutura da Rede Neural

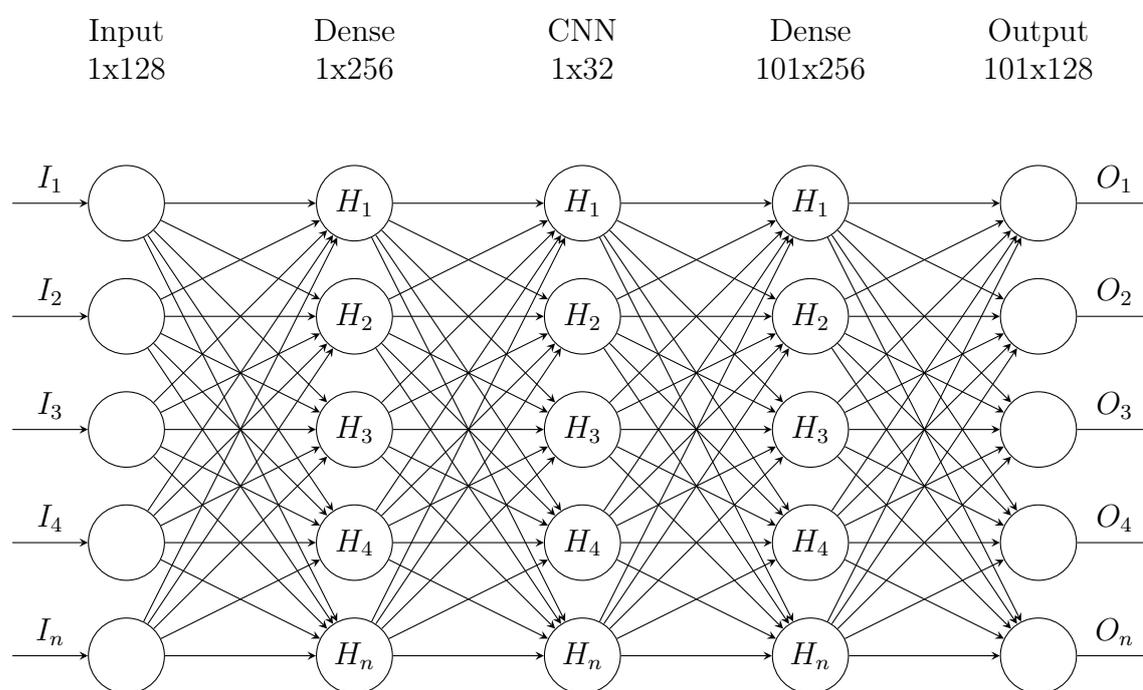
A estrutura de rede utilizada consiste em duas camadas intermediárias totalmente conectadas e uma camada convolucional, além das camadas de entrada e saída. As camadas densas e de entrada possuem função de ativação do tipo ReLU, já a camada convolucional e de saída têm como função de ativação a tangente hiperbólica e sigmoideal, respectivamente.

Para a entrada e saída da rede foram utilizados 128 neurônios, e para as camadas ocultas foram 256 neurônios. A Figura 6 ilustra a estrutura da rede utilizada.

A estrutura da rede foi inspirada na utilizada por Martins et al. (2022), porém adaptada para a arquitetura de uma rede convolucional *feed-forward*, visto que no projeto citado foi utilizado uma arquitetura de rede recorrente. Além disso, a quantidade de neurônios presentes nas camadas escondidas foi aumentado para 256 e a ordem das camadas foi, também, alterada.

Para o processo de treinamento da rede, foi utilizado um *batch* de 800 amostras. Além disso, o processo de treino da rede foi realizado em 240 épocas, utilizando como otimizador o RMSProp (de *root mean square propagation*) com taxa de aprendizagem de 1×10^{-4} e função de perda MSE (de *mean squared error*).

Figura 6 – Arquitetura da rede utilizada.



Fonte: Produção do próprio autor.

4 RESULTADOS

4.1 Introdução

Neste capítulo serão discutidas as métricas utilizadas para avaliar o desempenho da rede e, logo em seguida, serão apresentados os resultados que foram obtidos utilizando as predições da rede nos dados de teste. A análise sobre a base de teste é importante para descartar os possíveis efeitos de *overfitting* sobre os dados de treino.

4.2 Métricas de Desempenho

Após a finalização do treinamento da rede, é realizada uma análise do desempenho da rede ao prever as saídas para os dados de teste.

Duas métricas utilizadas para a análise serão o erro quadrático médio (MSE) e sua raiz (RMSE), descritos pelas Equações (4.1) e (4.2),

$$MSE = \frac{1}{N} \sum_{m=1}^N (x_m - \hat{x}_m)^2, \quad (4.1)$$

$$RMSE = \sqrt{MSE}, \quad (4.2)$$

onde x_m é o valor obtido pela SSF e \hat{x}_m o valor de saída da rede neural. Os valores de MSE e RMSE indicam o erro do modelo em relação ao método SSF, sendo assim, quanto menor o valor obtido, mais precisa é a predição. Em contrapartida, o coeficiente de determinação R^2 , descrito pela Equação (4.3), mede o índice de ajuste entre as curvas de referência e preditas, sendo, portanto, buscado um valor próximo de 1.

$$R^2 = 1 - \frac{\sum_{m=1}^N (x_m - \hat{x}_m)^2}{\sum_{m=1}^N (x_m - \bar{x})^2}, \quad (4.3)$$

onde \bar{x} representa a média dos valores de x_m .

4.3 Desempenho da rede

4.3.1 Ajuste do modelo

O modelo treinado apresentou boa capacidade de ajuste à curva do pulso calculada utilizando o SSF. As métricas de desempenho geral da rede obtidas estão apresentadas na Tabela 3 e foram obtidas a partir dos valores preditos pela rede comparados com os gerados pelo SSF para a base de teste.

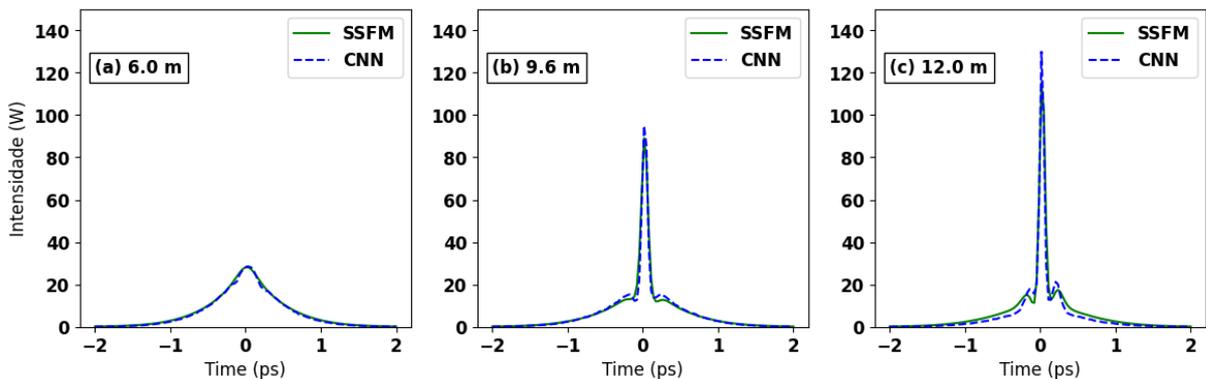
Tabela 3 – Performance da rede neural

RMSE	MSE	R^2
$11,05 \times 10^{-3}$	$12,22 \times 10^{-5}$	91,98%

Fonte: Produção do próprio autor.

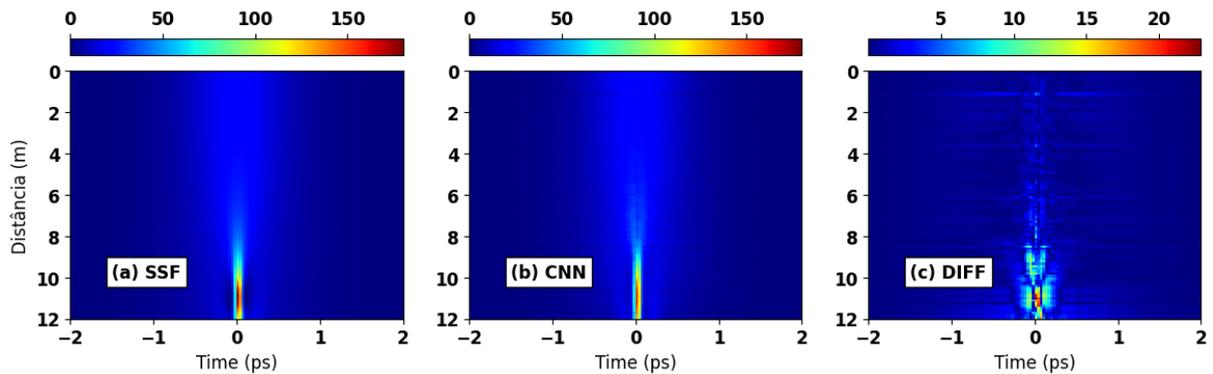
A primeira análise gráfica a ser realizada para avaliar o desempenho da rede será sobre as previsões para condições iniciais de $P = 20,5 \text{ W}$ e $T_{\text{FWHM}} = 1,13 \text{ ps}$, onde foi capaz de alcançar um valor de $R^2 = 98,37\%$. A Figura 7 mostra a saída da rede para estas condições, deixando claro que o modelo foi eficaz em prever o formato e a potência de pico do pulso nos pontos selecionados da fibra, sendo capaz de exibir com clareza o efeito das não linearidades sofridas pelo sóliton. Analisando a saída para as mesmas condições, porém para todo o comprimento da fibra, obtém-se os gráficos observados na Figura 8. Percebe-se que a diferença entre as imagens (Figura 8(c)) se posiciona próximo de zero na maior parte dos valores, além de apresentar sua maior variação em cerca de 25 W , que é, também, um valor aceitável dada a escala das previsões.

Figura 7 – Intensidade temporal do sóliton em distâncias selecionadas $P = 20,5 \text{ W}$ e $T_{\text{FWHM}} = 1,13 \text{ ps}$.



Fonte: Produção do próprio autor.

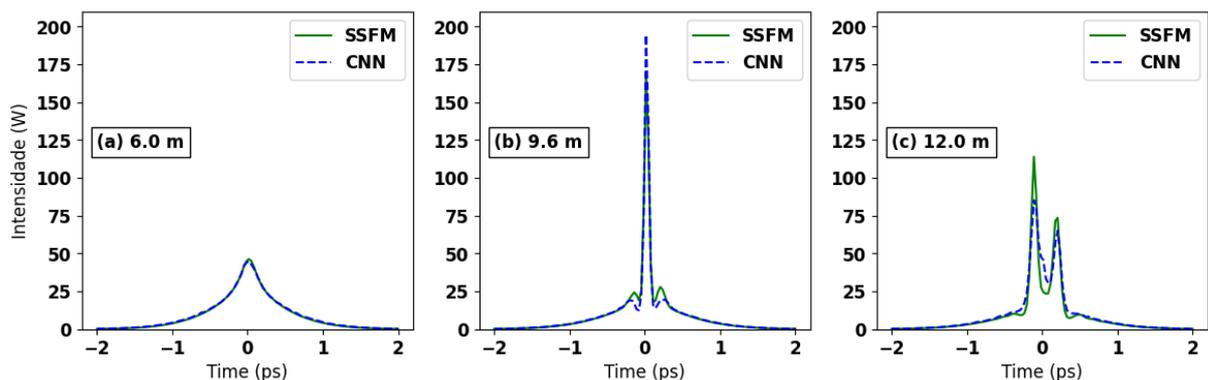
O próximo teste foi realizado com um pulso com o mesmo valor de T_{FWHM} , porém com potência de $29,3 \text{ W}$. Para estas condições foi possível alcançar um valor de $R^2 = 95,05\%$.

Figura 8 – Intensidade temporal do sóliton $P = 20,5$ W e $T_{FWHM} = 1,13$ ps.

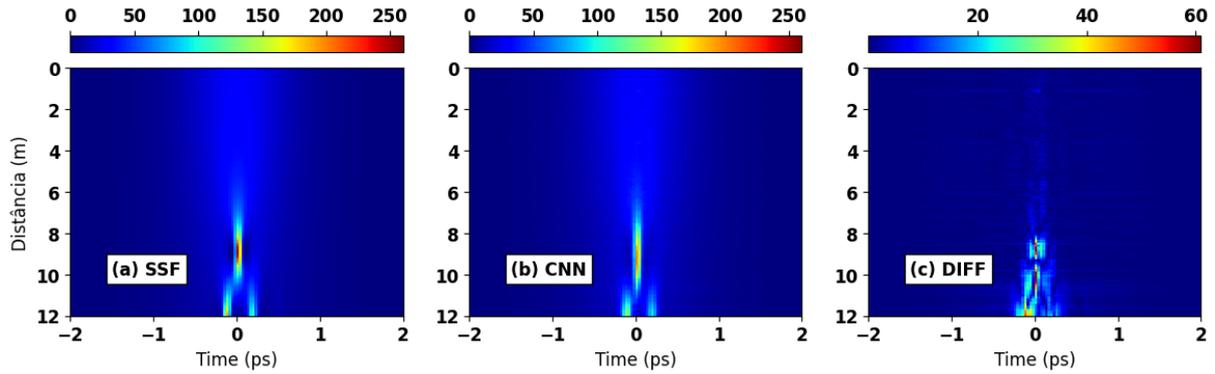
Fonte: Produção do próprio autor.

O valor de potência quase 50% maior tende a intensificar drasticamente os efeitos não lineares presentes na fibra, permitindo uma análise do desempenho da fibra em condições ainda mais intensas de não linearidade.

A Figura 9 deixa claro que a rede apresentou um ajuste um pouco pior, quando comparado ao caso anterior, cuja potência era menor, entretanto, ainda pode-se observar o formato do pulso bem próximo do gerado pelo SSF. A Figura 10 mostrou um aumento considerável na diferença entre os gráficos, alcançando seu pico em cerca de 60 W, que se limita a alguns pontos localizados.

Figura 9 – Intensidade temporal do sóliton em distâncias selecionadas $P = 29,3$ W e $T_{FWHM} = 1,13$ ps.

Fonte: Produção do próprio autor.

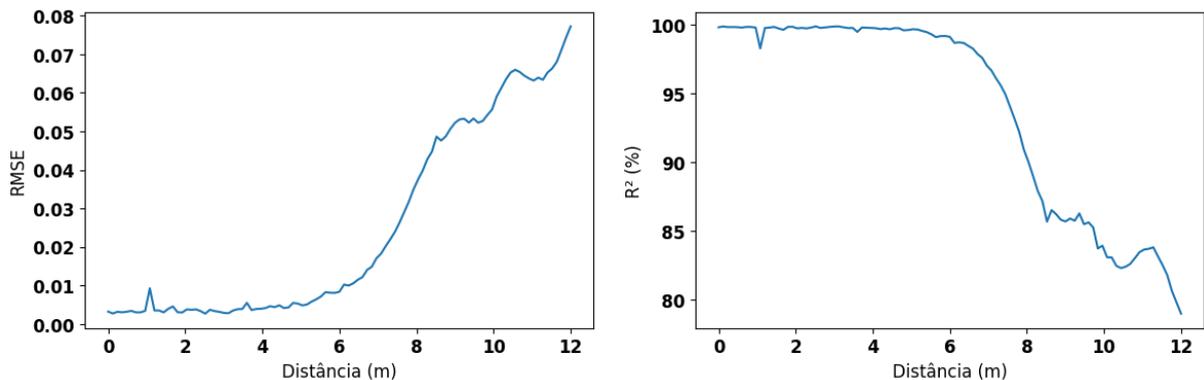
Figura 10 – Intensidade temporal do sóliton $P = 29,3$ W e $T_{FWHM} = 1,13$ ps.

Fonte: Produção do próprio autor.

4.3.2 Propagação de erro

Outra análise importante a ser realizada para avaliar o desempenho do modelo e encontrar pontos de melhoria é a propagação do RMSE e o coeficiente de determinação durante os passos na fibra, que está representado pela Figura 11, onde cada valor representa o RMSE e o R^2 médio dos dados de teste em cada distância da fibra. Percebe-se, a partir dos gráficos apresentados, o aumento gradual do erro obtido conforme a distância percorrida aumenta, alcançando seu ponto máximo nos 12 m. Considerando a análise feita na seção anterior, entende-se que o erro gerado em distâncias maiores realmente diminui o desempenho da rede para distâncias maiores, mas não o suficiente para comprometê-lo. A partir de 12 m de fibra a rede tende a não apresentar boas previsões (SALMELA et al., 2021), por isso não foram realizados treinamentos para distâncias maiores.

Figura 11 – Propagação do erro da rede neural.



Fonte: Produção do próprio autor.

4.3.3 Tempo de execução

Como última análise, foi comparado o tempo de execução do algoritmo SSF e de predição da rede neural. Este teste foi repetido 100 vezes e foi obtido a média dos resultados. Ambos os modelos foram testados em situações semelhantes, onde cada um deve gerar 800 resultados de mesmo tamanho e sendo executados na mesma máquina. Os resultados obtidos deste teste se encontram na Tabela 4.

Tabela 4 – Comparação de tempo de execução do método SSF e tempo de predição da rede em segundos

SSF	Rede Neural
65,9 ms	14,5 ms

Fonte: Produção do próprio autor.

É possível observar que a rede neural foi capaz de gerar o mesmo número de resultados que o SSF, porém em um tempo quase 5 vezes menor. Essa diferença pode se tornar um fator chave para a criação de simulações em tempo real, onde grandes tempos de execução tornam o sistema inviável. Vale ressaltar que para obter um baixo erro o SSF seria realizado com pelo menos 10.000 passos, que aumentaria drasticamente o tempo de execução do algoritmo, mas que não seria um problema enfrentado pela rede neural, gerando uma diferença de velocidades ainda maior.

5 CONCLUSÃO E TRABALHOS FUTUROS

5.1 Conclusão

O objetivo desse trabalho foi propor um modelo neural capaz de resolver o problema de propagação não linear de pulsos ópticos, garantindo um desempenho próximo ao algoritmo *split-step* Fourier, entregando uma alternativa mais flexível e de menor custo computacional.

Em um primeiro momento, foi feita uma análise geral sobre o desempenho da rede em gerar previsões sobre as curvas do pulso. A partir das métricas analisadas, foi obtido um R^2 de aproximadamente 92%, que se apresenta como um bom valor de aproximação das curvas, dado que a rede deverá prever não somente um pulso de saída, mas sim o estado dele em cada momento da fibra, sendo, portanto, uma tarefa complexa de ser realizada.

Em seguida, a análise se voltou para o erro de previsão da rede em cada passo da fibra, que mostrou um erro consideravelmente maior a partir de 6 m de propagação. Entretanto, foi possível observar que, mesmo com o aumento do erro, os gráficos obtidos para a parte final da fibra foram muito bem preditos pela rede, onde pode-se observar de forma clara os efeitos não lineares operando sobre o pulso.

Por fim, foi realizado uma análise a respeito do tempo de previsão da rede em comparação com a execução do SSF, e a CNN foi capaz de gerar resultados quase 5x mais rápido. Além disso, ainda há processos de otimização que podem ser realizados na rede para que se obtenha desempenho ainda melhor, permitindo pesquisas mais aprofundadas sobre desempenho desse modelo e sua viabilidade em simulações de larga escala e em tempo real.

5.2 Temas a serem pesquisados

Como trabalhos futuros: serão pesquisadas novas arquiteturas de rede capazes de modelar os efeitos apresentados de forma mais robusta e estável, sendo capaz de manter um baixo erro mesmo em distâncias de propagação maiores. Além disso, o modelo neural criado é capaz de ser adaptado à problemas de telecomunicações, onde a propagação é quilométrica, porém com efeitos não lineares mais amenos. Por fim, propõe-se a aplicação de *machine learning* à análise do espectro óptico de sistemas não lineares, sendo capaz de simular

supercontínuos, efeito onde há espalhamento da banda espectral em grande escala.

REFERÊNCIAS

- ACHILLES, R.; BONFIGLIOLI, A. The early proofs of the theorem of campbell, baker, hausdorff, and dynkin. Archive for History of Exact Sciences, 2018. Citado na página 24.
- AGRAWAL, G. P. Nonlinear Fiber Optics. 6. ed. [S.l.]: Academic Press, 2019. Citado 5 vezes nas páginas 12, 15, 16, 17 e 18.
- ALBAWI, S.; MOHAMMED, T. A.; AL-ZAWI, S. Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET). [S.l.: s.n.], 2017. p. 1–6. Citado 2 vezes nas páginas 21 e 22.
- ALESHAMS, M.; ZARIFKAR, A.; SHEIKHI, M. Split-step fourier transform method in modeling of pulse propagation in dispersive nonlinear optical fibers. In: Proceedings of CAOL 2005. Second International Conference on Advanced Optoelectronics and Lasers, 2005. [S.l.: s.n.], 2005. v. 2, p. 124–126 vol. 2. Citado na página 12.
- CASTELLANI, C. E. S. All-fibre wavelength versatile short pulsed laser sources. Tese ((Doutorado)) — Imperial College London, UK, 2013. Disponível em: <<https://doi.org/10.25560/24135>>. Citado 3 vezes nas páginas 12, 16 e 18.
- DING, B.; QIAN, H.; ZHOU, J. Activation functions and their characteristics in deep neural networks. In: 2018 Chinese Control And Decision Conference (CCDC). [S.l.: s.n.], 2018. p. 1836–1841. Citado na página 19.
- JIANG, X.; WANG, D.; FAN, Q.; ZHANG, M.; LU, C.; LAU, A. P. T. Physics-informed neural network for nonlinear dynamics in fiber optics. Laser & Photonics Reviews, v. 16, n. 9, p. 2100483, 2022. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/lpor.202100483>>. Citado na página 13.
- KROGH, A. What are artificial neural networks? Nature Biotechnology, v. 26, p. 195–197, 2008. Disponível em: <<https://doi.org/10.1038/nbt1386>>. Citado na página 20.
- LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Transactions on Neural Networks and Learning Systems, v. 33, n. 12, p. 6999–7019, 2022. Citado na página 21.
- MARTINS, G. R.; SILVA, L. C. B.; SEGATTO, M. E. V.; ROCHA, H. R. O.; CASTELLANI, C. E. S. Design and analysis of recurrent neural networks for ultrafast optical pulse nonlinear propagation. Opt. Lett., Optica Publishing Group, v. 47, n. 21, p. 5489–5492, Nov 2022. Disponível em: <<https://opg.optica.org/ol/abstract.cfm?URI=ol-47-21-5489>>. Citado 3 vezes nas páginas 13, 23 e 25.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, v. 65(6), p. 386–408, 1958. Disponível em: <<https://psycnet.apa.org/doiLanding?doi=10.1037%2Fh0042519>>. Citado na página 19.
- SALMELA, L.; TSIPINAKIS, N.; FOI, A.; BILLET, C.; DUDLEY, J. M.; GENTY, G. Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network. Nature Machine Intelligence, 2021. Citado 3 vezes nas páginas 13, 23 e 30.

ZHOU, Z.; YAN, Z. Deep learning neural networks for the third-order nonlinear schrödinger equation: bright solitons, breathers, and rogue waves. Communications in Theoretical Physics, IOP Publishing, v. 73, n. 10, p. 105006, sep 2021. Disponível em: <<https://dx.doi.org/10.1088/1572-9494/ac1cd9>>. Citado na página 13.