CIU Journal | 4(1) 17–35

Evaluation of Deep Neural Networks for Predicting Optical Properties of Silicon-rich Silicon Nitride Waveguide

M R Karim[1], Abrar Hussain[2], Al Kayed[3], B M A Rahman[4]

Abstract

Deep learning (DL) has recently emerged as a potential platform for estimating linear and nonlinear optical phenomena of waveguides due to its high computational power, high-level structures and flexible usages. In this work, we performed a comparative analysis of four DL based Deep Neural Network (DNN) configurations for predicting and analyzing the effective mode area of a planar Silicon-rich Silicon Nitride (SRN) waveguide, its nonlinear coefficient, effective index and dispersion in the wavelength range of 0.65 μm – 3.05 μm, waveguide core width of 1 μm – 5 μm and waveguide height of 0.3 μm – 0.4 μm. We found that out of four DNN structures analyzed, ELU-ELU-ReLU-70-9000 structure showed superior performance in terms of mean squared error values. The computational time required with deep neural network (for training) and finite-element method (FEM) solutions is also compared and found that the training time of DNN structures increased with a number of epochs and due to the ReLU activation function. This simple and fast-training DNN employed here predict the output for unfamiliar parameter setting of the optical waveguide faster than traditional numerical simulation techniques.

Keywords

Deep neural network, deep learning, silicon-rich nitride, planar waveguide, dispersion, nonlinearity, integrated photonics, nonlinear optics, ultrafast optics

Introduction

Silicon-On-Insulator (SOI) technology facilitates the CMOS compatible fabrication effectively to build low cost and scalable on-chip components similar to silicon photonics. Due to having four times higher Kerr nonlinearity, high refractive index and large energy bandgap in comparison with silica, Silicon-rich Silicon Nitride (SRN) chemically known as Si2N material pene-

[1] Associate Professor, Department of EEE, Chittagong Independent University, Chattogram, Bangladesh.
[2] Doctoral Student, Concordia University, Canada.
[3] Student, CUET, Chattogram, Bangladesh.
[4] Professor, City University of London, United Kingdom

Corresponding author:
M R Karim, Associate Professor, Department of EEE, Chittagong Independent University, Chattogram, Bangladesh.
Email: mrkarim@ciu.edu.bd

trated into the competitive market in the silicon chip industry. This material consists of 65% Silicon (Si) and 35% Nitrogen (N) contains a huge bandgap of 2.05 eV which eliminates the two-photon absorption at 1.55 μm (Yang et al., 2018). The Kerr nonlinearity of 2.8×10^{-13} cm^2/W and refractive index of about 3.1 to tailor dispersion. The loss is taken for merely material by ignoring substrate linkage is 6 dB/cm since of Si-H bonds formation for chemical vapor deposition (Wang et al., 2015).

In ultrafast nonlinear optics, designing an integrated photonics compatible efficient broadband supercontinuum (SC) source has been an increasing demand owing to the frequent usage of this light source in optical coherence tomography, spectroscopy, biomedical imaging, optical frequency metrology and various industrial sensing applications (Jones et al., 2000; Mirnov et al., 2006; Hartl et al., 2001). SRN material-based optical waveguide has the ability to produce SC spectrum up to mid-infrared (MIR) region (Karim et al., 2020a; Karim et al., 2020b). Additionally, due to having a low Raman response, SRN is highly coherent spread over the whole SC spectrum (Wang et al., 2015).

For efficient waveguide design to be used for producing ultra-broadband SC generation up to the MIR region, specific waveguide modal properties, including effective index (n_{eff}), effective mode area (A_{eff}), dispersion (D) and non-linearity (γ) can be easily controlled by changing width and thickness (height) of a planar waveguide in the wavelength region of interest (Karim et al., 2020a; Karim et al., 2020b). This is because the broadening of SC is dependent on the group-velocity dispersion within the waveguide provided with the position of pump wavelength and peak power. This broadening can further be extended far into the MIR spectrum by adopting a tapering approach during waveguide design. Selective enhancement of spectral flatness across the spectral coverage can also be observed by applying the tapering approach (Zia et al., 2020; Singh et al., 2019; Ciret & Gorza, 2017) on silicon made core material-based waveguides.

To design and optimize the device parameters of the optical waveguide mentioned in the above paragraph, sophisticated numerical methods such as the finite difference method (Yu & Chang, 2004) and finite element method (FEM) (Cucinotta et al., 2002) are usually used. However, these methods necessitate a substantial amount of power of a personal computer's Central Processing Unit (CPU) when dealing with intricate optical waveguide layouts which require be simulated several time periods to settle with an optimized design. These complex waveguide layouts have their own input design parameters to be optimized which makes the whole iterative process even more complex and time-consuming. One way to solve this problem is to estimate the parameters using prediction algorithms and a deep learning-based deep neural network (DNN) can readily give a way out to optimize the parameters quickly with very good accuracies. Deep learning has become a vital approach to solving a big-data-driven problem. It has found tremendous applications in computer vision and natural language processing (Huang et al., 2020). Deep learning has received much attention worldwide because it can efficiently

process and analyze a vast number of datasets based on the multilevel abstraction of data using hierarchically structured layers.

Over the last couple of years, in photonics, Deep Neural Network (DNN) has been employed to improve the state-of-the-art photonics application in optical fiber sensing, laser characterization, quantum communications, imaging and device design (Zibar et al., 2017; Rivenson et al., 2017; Sinha et al., 2017; Turduev et al., 2018; Malkiel et al., 20017; Ma et al., 2018; Liu et al., 2018; Kojima et al., 2017). Very recently, two of the research works have been published (Chugh et al., 2019a; Chugh et al., 2019b) on optimizing the PCF's optical parameters using DNN. However, they have dealt with the limited size of the dataset due to the parameter settings, and also their data is not highly nonlinear and not well-distributed over the wavelength region considered. Moreover, these articles were limited in tuning the training and optimization hyperparameters in a wide range to produce the optimized design.

To overcome the above-mentioned limitations, we have made the contributions as pointed out below.

1) We considered a completely different material (SRN waveguide) for this article and optimized its parameters using DNN.

2) We generated a well-balanced and comparatively larger dataset having highly nonlinear input datapoints distributed over the wavelength region considered.

3) We considered a variety of structures of simple DNN, just varying their structural and also the optimization hyperparameters, to reach to an optimized design of SRN waveguide.

In this article, four distinct configurations of DNN are applied on highly nonlinear optical parametric simulated training data of SRN waveguide and predicted on separate set of test data where the DNNs never seen those test data before during training. We predicted effective mode index (n_{eff}), effective mode area (A_{eff}), nonlinear coefficient (γ) and group-velocity dispersion (D) parameters using those DNNs, evaluated and compared the performance of the DNNs in terms of mean squared error. The next section describes about the modeling and method, where SRN waveguide model and also the DNN structures are described. The subsequent sections explain the results and discussion. This article concludes with a conclusion section included with future works.

Modeling and Method

In this section, we first describe our 3-mm-long SRN waveguide and the waveguide structure is shown in Fig. 1.

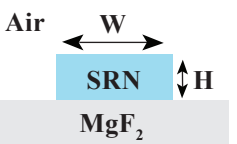


Figure 1. SRN Waveguide
Source: The authors.

The waveguide's core, upper and lower cladding are made of Si_2N, air and MgF_2, respectively. To compute the refractive index of proposed SRN and MgF_2 materials, the Sellmeier equation (Karim et al., 2020b) is considered. In this study, a set of SRN waveguides geometric data of width of core (W), height of core (H) and wavelength (λ) have taken as input variables and n_{eff}, A_{eff}, γ and D as output variables to make a labelled training dataset. We have considered a short wavelength ranging 0.65 µm – 3.05 µm (0.5 µm in steps), core width finely distributed within 1 µm – 5 µm (0.25 µm in steps) and core thickness of 0.3 µm – 0.4 µm (0.10 µm in steps). Subsequently, n_{eff}, A_{eff}, γ and D values are calculated using FEM solution based on COMSOL Multiphysics® for 8789 samples. We used extremely fine mesh in COMSOL and mesh elements of 17782. The above-mentioned parameters are calculated using the equations (Obayya et al., 2005) below.

$$A_{eff} = \frac{(\iint |E^2| \, dx \, dy)^2}{(\iint |E|^4 \, dx \, dy)^2} \tag{1}$$

$$D(\lambda) = -\frac{\lambda}{c} \frac{d^2 Re(n_{eff})}{d\lambda^2} \tag{2}$$

$$\gamma = \frac{n_2 \omega_o}{C A_{eff}} \tag{3}$$

where, E is the transverse electric field vector, ∬ is the area bounded by the computing domain. Re signifies the real part and c is the speed of light in free space.

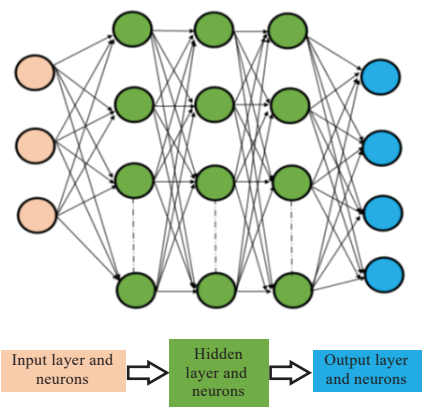


Figure 2. DNN Structure
Source: The authors.

Secondly, we have designed and varied the structure of a DNN having three hidden layers and each layer contains either 50 or 70 neurons. We found that increasing the number hidden layers and their associated nodes contributed

very insignificantly in the error value and therefore we made a good starting value of 50 for number of nodes and restricted our models trying a different number of 70 hidden nodes in the three hidden layers. The general network structure is shown in Fig. 2. These hidden layers are interlocked as a whole, which means layers' mimic human brain where neurons maintain an interconnectivity between each and every layers. We used pytorch machine learning framework to design the DNNs. We have a total data point of 8789 which is generated through COMSOL Multiphysics® numerical simulation. We have divided the whole dataset into training data (90% of total data), validation data (10% of total data). We settled with 7910 training data points and 879 validation data point. The validation data is randomly selected from the training data to validate the performance of the neural network. Also, for the test dataset, we have simulated a new range of data for the width of the core material 2.3 μm corresponding to core height of 0.335 μm and new wavelength spanning between 0.675 μm and 3.075 μm. We got 47 datapoints as test data. All the training, validation and testing data points are normalized between 0 and 1 without distorting differences in the ranges of values and to avoid losing information.

Our target is to choose the best combination of DNN structural hyperparameters to optimize the waveguide structure. Another vital reason is to see the effect of changing each hyperparameters on the output. The list hyperparameters and how it evolves in each epoch of 5000, 7000 and 9000 are given in Table 1. The epochs are selected by the user targeting to reduce the overfitting and underfitting of data on the model.

Table 1. List of Hyperparameters

No of epochs	No of layers	No of nodes	Activation function	Dropout	Optimizer	Batch normalization
5000	3	50	ELU-ELU-ReLU	0.1	Adam	After hidden layer 1
5000	3	50	ELU-ELU-PReLU	0.1	Adam	After hidden layer 1
7000	3	50	ELU-ELU-ReLU	0.1	Adam	After hidden layer 1
9000	3	70	ELU-ELU-ReLU	0.1	Adam	After hidden layer 1

Source: The authors.

It is evident from Table 1 that, for training data, exponential linear unit (ELU) activation function is used in the first two hidden layers as it can take negative inputs and rectified linear unit (ReLU) and parametric rectified linear unit (PReLU) are used in the last hidden layer interchangeably. An Adam optimizer with a learning rate of 0.0001 has been used to adjust the weights during the training process. We also have used batch normalization after the first hidden layer as it makes faster and more stable training through the normalization of

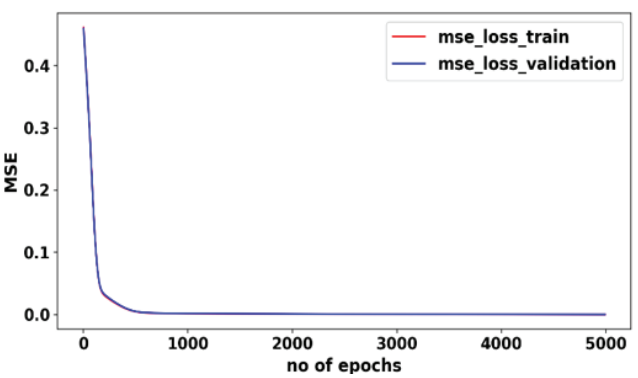
the input layer. Once trained, the DNN model estimates outputs after each epoch. The difference between the predicted and original output is calculated using mean squared error (MSE). The back-propagation algorithm (LeCun et al., 1988) is used on the difference iteratively to create new weights of the interconnected hidden layers for each and every iteration. When our DNN model gets optimized with training data through having a steady mean squared error surface, we produce designated outputs for the brand new input optical waveguide parameters (as test data) which is not given to the model during the training term.

Results and Discussion

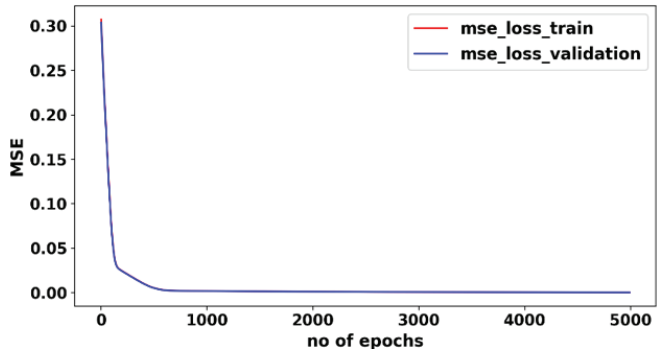
In this section, we evaluated DNN structures using validation data and test data. Also, computational time is compared between DNN models and complex numerical simulations. Initially, we have run the DNN models for 50 neurons in each of three hidden layers and 5000 epochs. We used combination of ELU-ELU-ReLU activation functions in first, second and third hidden layer respectively. For the sake of explanation, we named this ELU-ELU-ReLU-50-5000 as "DNN architecture 1." Additionally, for the same number of layers and hidden nodes, we used ELU-ELU-PReLU combination to see the effect of these functions in the last hidden layer on accuracy and computational time required. We named this ELU-ELU-PReLU-50-5000 as "DNN architecture 2." We then finally ran ELU-ELU-ReLU-50-7000 (DNN architecture 3) and ELU-ELU-ReLU-70-9000 (DNN architecture 4). For each of the above DNN architectures, we computed and predicted n_{eff}, A_{eff}, γ and D and it is described in the following subsections.

Prediction of A_{eff}

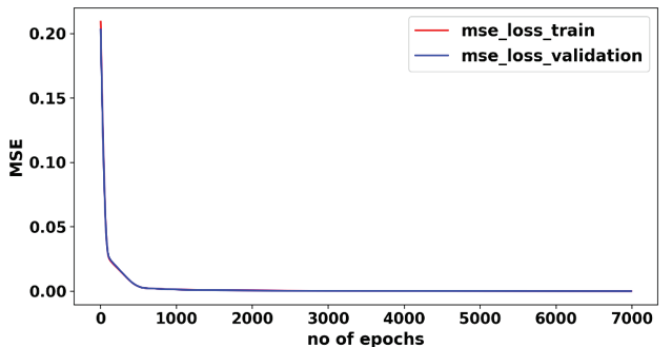
The prediction of effective mode area first comes with the analysis of training loss and validation loss for architecture 1, architecture 2, architecture 3 and architecture 4. These are shown in Fig. 3.



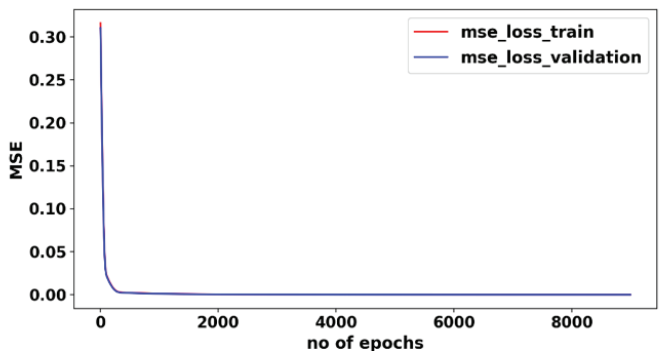
(a)



(b)



(c)

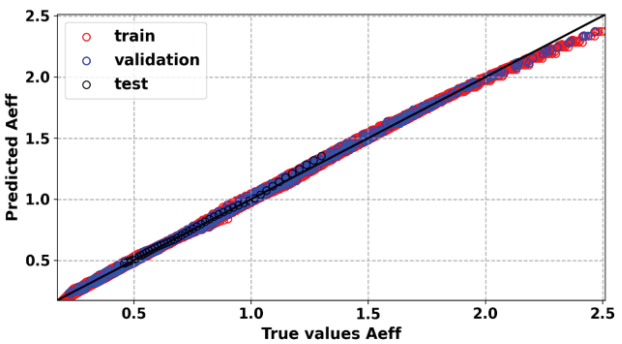


(d)

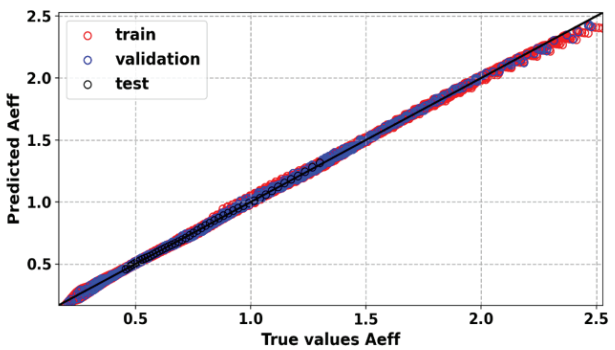
Figure 3. MSE Curves for a) Architecture 1; b) Architecture 2; c) Architecture 3; and d) Architecture 4.

Source: The authors.

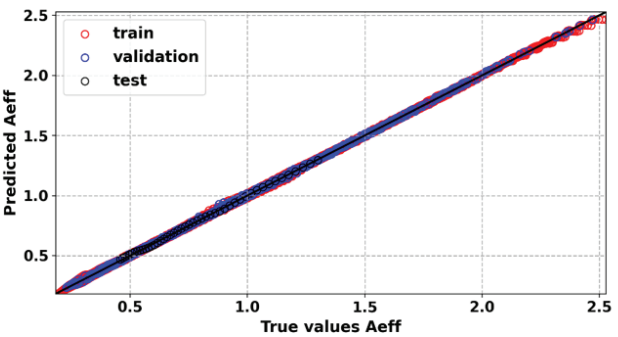
It is depicted from Fig. 3 that DNN models have well trained and validated with training and validation data respectively. Here, MSE values change through epochs and give the lowest values possible. As expected, MSE curves converge faster for 7000 and 9000 epochs compared to 5000 epochs and it is seen in Figs. 3 (c) and 3(d). There is no effect seen on MSE visually, by other parameters, considered in the architectures. Figure 4 shows a scatter plot of the A_{eff} values of the dataset used for training, validation and testing of the DNN models.



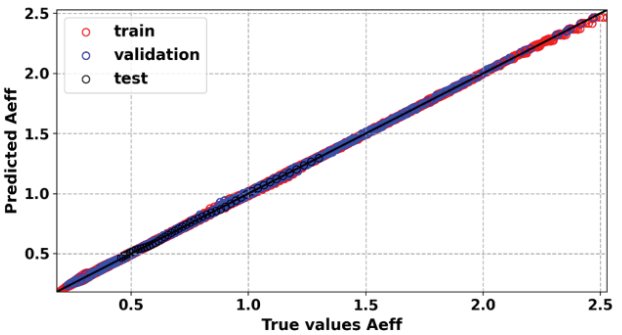
(a)



(b)



(c)



(d)

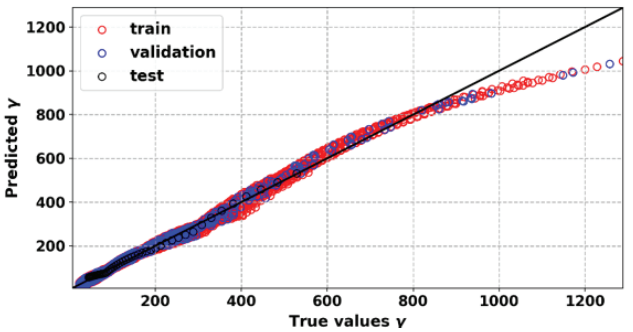
Figure 4. True and Predicted Values of A_{eff} for a) Architecture 1; b) Architecture 2; c) Architecture 3; and d) Architecture 4.

Source: The authors.

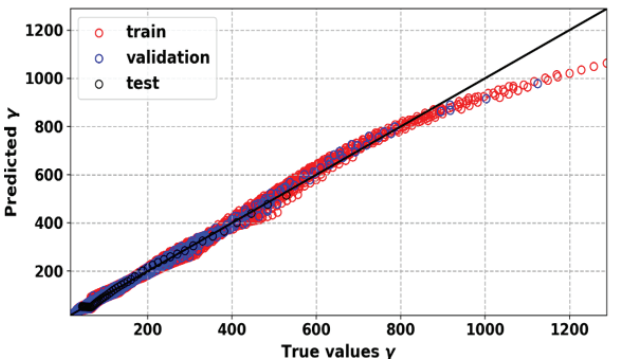
These figures explain that how well the DNN predicts the A_{eff} by indicating how close or far are the circled data points to the ideal black solid line. As we initially experimented, it is found that training data points are evenly distributed throughout the wavelength range considered and closer to ideal line when we applied ELU-ELU-ReLU than we applied ELU-ELU-PReLU, and this is visible in Figs. 4(a) and 4(b). This is because ReLU has more data generalization capability compared to PReLU. On the other hand, ELU avoided the dead neurons in the deep structured DNN networks. However, training data started deviating from the ideal line both for ELU-ELU-ReLU and ELU-ELU-PReLU after 2 μm^2. Later we understood that increasing the number of epochs to 7000 and 9000, made the training data test data perfectly fit to the ideal line which means architecture 3 and 4 predict A_{eff} quite accurately. This is depicted in Figs. 4(c) and 4(d). This is supported by the fact that MSE drops from 0.000346 (ELU-ELU-ReLU-50-5000) to 0.000023 (ELU-ELU-ReLU-70-9000)

Prediction of γ

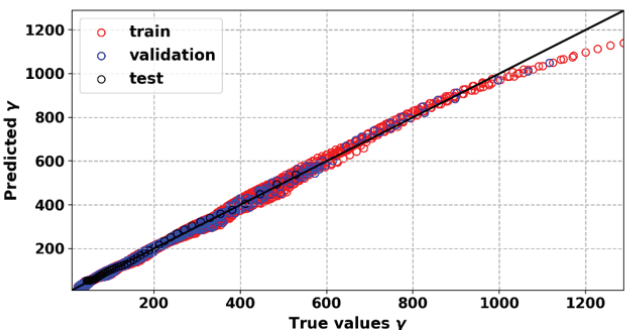
It is described in Fig. 5 that for the epochs 5000, γ values are poorly trained on architecture 1 and 2, especially after 800 W-1m-1 and it is verified in Figs. 5(a) and 5(b). However, it is observed in Figs. 5(c) and 5(d) that training data points after 800 W-1m-1 got closer to ideal values of for epochs 7000 and 9000. This is also validated that validation data do not deviate much from the y=x line for data values up to 400 W-1m-1. Also, all the architectures predict test γ values, more or less, well as most of the black circles aligned well to the y=x line.


train
validation
test
Predicted y
True values y
200
400
600
800
1000
1200

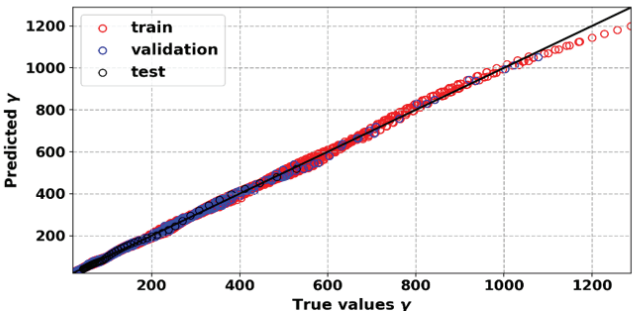
(a)


train
validation
test
Predicted y
True values y
200
400
600
800
1000
1200

(b)


train
validation
test
Predicted y
True values y
200
400
600
800
1000
1200

(c)



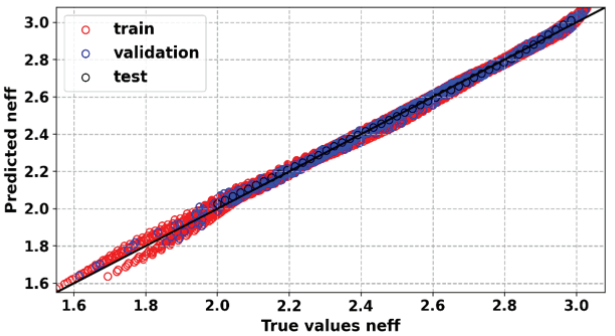
(d)

Figure 5. True and Predicted Values of γ for a) Architecture 1; b) Architecture 2; c) Architecture 3; and d) Architecture 4.
Source: The authors.

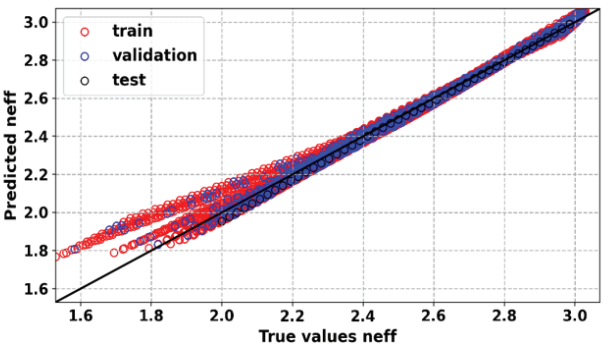
These figures explain that how well the DNN predicts the A_{eff} by indicating how close or far are the circled data points to the ideal black solid line. As we initially experimented, it is found that training data points are evenly distributed throughout the wavelength range considered and closer to ideal line when we applied ELU-ELU-ReLU than we applied ELU-ELU-PReLU, and this is visible in Figs. 4(a) and 4(b). This is because ReLU has more data generalization capability compared to PReLU. On the other hand, ELU avoided the dead neurons in the deep structured DNN networks. However, training data started deviating from the ideal line both for ELU-ELU-ReLU and ELU-ELU-PReLU after 2 μm^2. Later we understood that increasing the number of epochs to 7000 and 9000, made the training data test data perfectly fit to the ideal line which means architecture 3 and 4 predict A_{eff} quite accurately. This is depicted in Figs. 4(c) and 4(d). This is supported by the fact that MSE drops from 0.000346 (ELU-ELU-ReLU-50-5000) to 0.000023 (ELU-ELU-ReLU-70-9000)

Prediction of n_{eff}

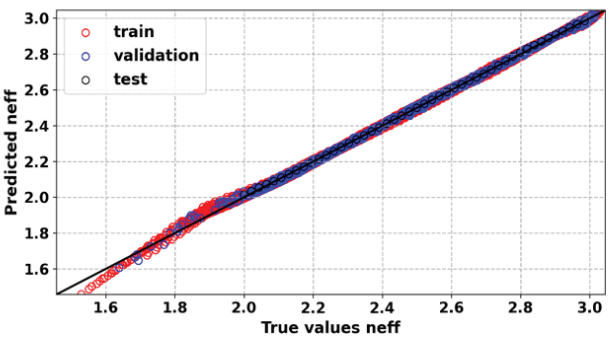
The prediction of effective mode area first comes with the analysis of training loss and validation loss for architecture 1, architecture 2, architecture 3 and architecture 4. These are shown in Fig. 3.


train
validation
test
Predicted neff
True values neff
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0

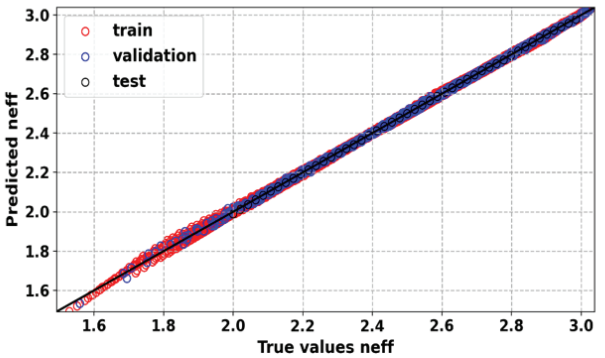
(a)


train
validation
test
Predicted neff
True values neff
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0

(b)


train
validation
test
Predicted neff
True values neff
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0

(c)



(d)

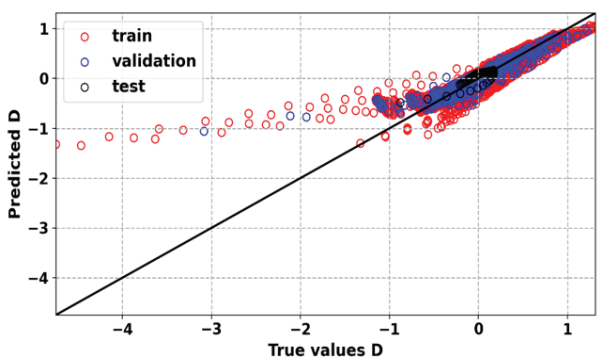
Figure 6. True and Predicted Values of n_{eff} for a) Architecture 1; b) Architecture 2; c) Architecture 3; and d) Architecture 4.

Source: The authors.

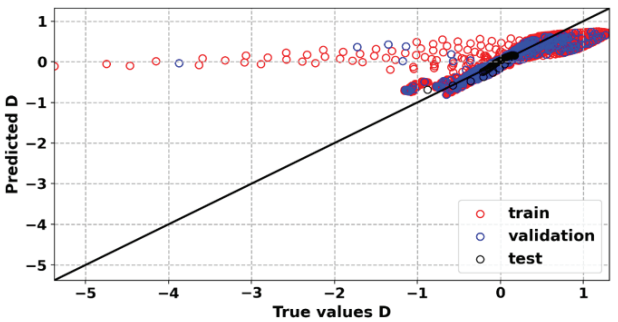
When PReLU comes into play in the last hidden layer, data gets overfitted. According to Figs. 6(c) and 6(d), this situation is not seen for architecture 3 and 4 as number of iterations of 7000 and 9000 along with number of hidden nodes have made the training data and lot closer to the ideal line although few of the training points in the beginning still did not fit to the straight line. For the same reason, test data was well predicted by the model which was never given to the model before during training and this is shown in Figs. 6(c) and 6(d).

Prediction of D

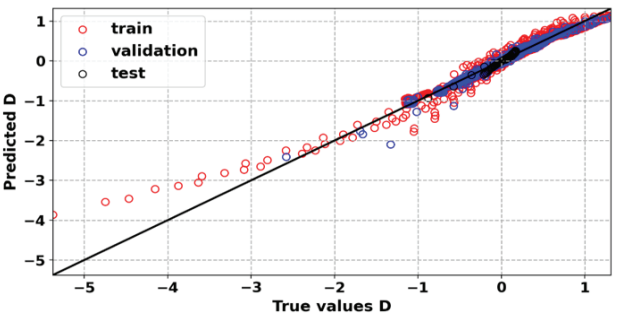
The dispersion (D) of waveguide characterization is a vital nonlinear optical phenomenon for many nonlinear optics applications, such as SC generation and frequency comb generation. The true and predicted values for training, test and validation are portrayed in Fig. 7.



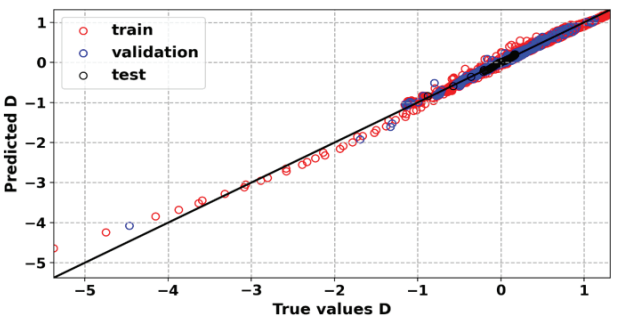
(a)



(b)



(c)



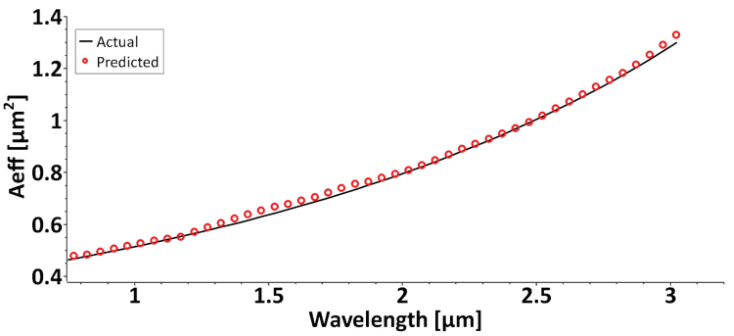
(d)

Figure 7. True and Predicted Values of D a) Architecture 1; b) Architecture 2; c) Architecture 3; and d) Architecture 4.
Source: The authors.

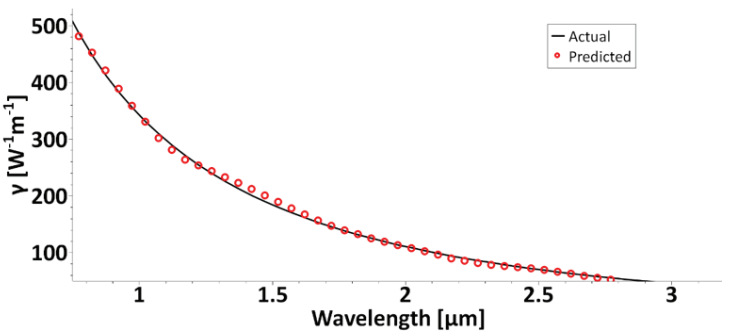
It is very clear from Figs. 7(a) and 7(b) that dispersion values are trained poorly on architecture 1 and this situation accelerated in architecture 2 due to data generalization problem created by PReLU activation function. The training data points which are especially in negative region, are located far away from the ideal solid black line as the network did not learn the inherent alpha value. On the contrary, as the architecture 3 went through 7000 iterations, most of the initial red circled data (up to -3 ps/nm/m) came closer to y=x line. However, not that much of them to count in and this observed in Fig. 7(c). Moreover, due to having higher number of hidden layers and being run for 9000 epochs, architecture 4 fit the training data well for the whole wavelength considered. Our designed architecture 4 fit well to the test data given that it runs for 9000 epochs. This is well seen in Fig. 7(d).

Comparison of Actual and Predicted Values

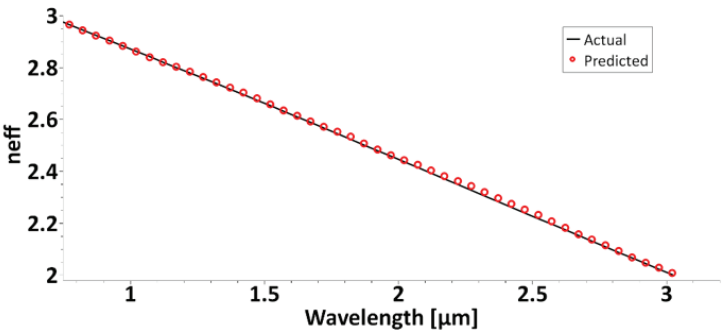
In this section, we compare the actual and predicted values of all the optical properties mentioned in the previous sections in terms of the wavelength considered for testing architecture 4. This is represented in Fig. 8.



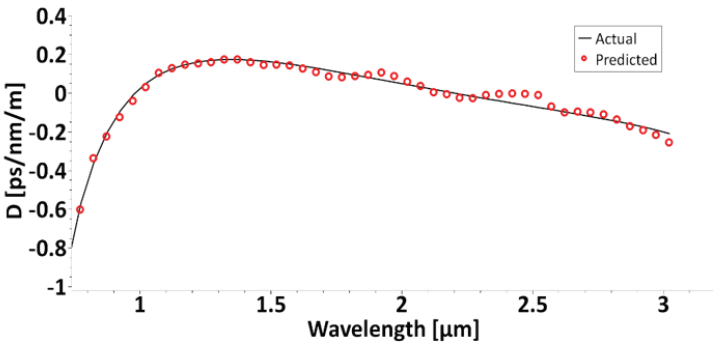
(a)



(b)



(c)



(d)

Figure 8. Optical Properties Vs Wavelength for a) A_{eff} ; b) γ; c) n_{eff} ; and d) D.

Source: The authors.

It is clear from Fig. 8(a) that predicted values of A_{eff} follow the actual line (black solid line) and it increases with wavelength. It is also observed that architecture 4 predicted γ and n_{eff} quite well in whole wavelength region considered. However, D values did follow the actual values for wavelength only until 1.5 µm.

Table 2. Training Time for Different DNNs

Layers	Nodes	Activation functions	Epochs	Training time (sec)
3	50	ELU-ELU-ReLU	5000	66.86
3	50	ELU-ELU-PReLU	5000	58.51
3	50	ELU-ELU-ReLU	7000	90.58
3	70	ELU-ELU-ReLU	9000	164.81

Source: The authors.

Computational Performance

The training and testing of the DNN model have run on a personal computer (PC) having the configuration of Intel Core i7 CPU @ 2.60 GHz, 16 GB RAM, NVIDIA GeForce RTX 2060 1 core GPU. The PC runs on Windows 10 operating system. The training actually runs on the GPU. The training time for our four architectures is given in Table 2. The time to train the DNN model relies on the dataset size and different training hyper-parameters. It is obvious that training time increases with increasing the number of layers and hidden nodes. In our designed models of DNN, as we fix the number of hidden layers, increasing the number of epochs from 5000 to 7000 and 9000 have increased the training time on large scales. We noticed from Table 2 that applying PReLU in the last hidden layer requires less training time than applying ReLU. This is because it considers the negative values of input. On the other hand, numerical simulation requires 2222 seconds for one particular structural parameter (W and H) of waveguide for the desired wavelength region, which can even take longer for finer values of such parameters during the optimization. Therefore, it can be said that training and optimizing the waveguide parameters are way faster than computing and optimizing the parameters using numerical simulation techniques.

Conclusion and Future Works

In this paper, we have designed and tuned four different deep neural network structures based on the deep learning concept. We predicted four general but vital optical parameters of the SRN planar waveguide using our designed deep learning (DL) models. We saw that the MSE value of training and validation accuracies decreases very sharply with the increase of the number of epochs. It is found that our designed DL architectures predicted the four characteristic nonlinear optical parameters well with test data generated for a specific wavelength range, width and height of waveguide which were never exposed to models during the training phase. We found when neural networks go deep into the structures with an increased number of nodes, their prediction accuracies increase tremendously. It was also evident that being one of the important hyperparameters, activation function also contributed to the prediction power of DNN structures. We can conclude that ELU-ELU-ReLU-70-9000 structure works best among four of the structures analyzed. With this best structure, we further verified the predicted and actual data of all parameters against the wavelength and we found that this architecture is somewhat shallow to predict dispersion in longer wavelength regions. In another part, we described that deeper structures need more time to be trained. This research opened a pathway to deliver more research works in the future which can consider more deep networks with proper tuning to predict the nonlinear optical parameters especially n_{eff}, as it should be predicted with precision and accuracy to generate supercontinuum spectra. Also, in future, we will include more DNN models such as LSTM/Bi-LSTM or combinations, conventional machine learning models to make a rigorous comparison among these models, so that the machine learning researchers in photonics will get a clear idea on which models are best for different applications available.

References

Chugh, S., Ghosh, S., Gulistan, A., & Rahman, B. M. A. (2019a). Machine learning regression approach to the nanophotonic waveguide analyses. *Journal of Lightwave Technology, 37*(24), 6080-6089.

Chugh, S., Gulistan, A., Ghosh, S., & Rahman, B. M. A. (2019b). Machine learning approach for computing optical properties of a photonic crystal fiber. *Optics Express, 27*(25), 36414-36425.

Ciret, C., & Gorza, S. P. (2017). Generation of ultra-broadband coherent supercontinua in tapered and dispersion-managed silicon nanophotonic waveguides. *JOSA B, 34*(6), 1156-1162.

Cucinotta, A., Selleri, S., Vincetti, L., & Zoboli, M. (2002). Holey fiber analysis through the finite-element method. *IEEE Photonics Technology Letters, 14*(11), 1530-1532.

Hartl, I., Li, X. D., Chudoba, C., Ghanta, R. K., Ko, T. H., Fujimoto, J. G., ... & Windeler, R. S. (2001). Ultrahigh-resolution optical coherence tomography using continuum generation in an air–silica microstructure optical fiber. *Optics Letters, 26*(9), 608-610.

Huang, L., Xu, L., & Miroshnichenko, A. E. (2020). Deep learning enabled nanophotonics. In *Advances and applications in deep learning.* IntechOpen.

Jones, D. J., Diddams, S. A., Ranka, J. K., Stentz, A., Windeler, R. S., Hall, J. L., & Cundiff, S. T. (2000). Carrier-envelope phase control of femtosecond mode-locked lasers and direct optical frequency synthesis. *Science, 288*(5466), 635-639.

Karim, M. R., Al Kayed, N., & Rahman, B. M. A. (2020a, June). Broadband supercontinuum generation in the mid-infrared range (0.8 μm–4.5 μm) using low power dispersion engineered LiNbO3 cladded silicon-rich Nitride waveguide. In *2020 IEEE region 10 symposium (TENSYMP)* (pp. 246-249). New Jersey, USA: IEEE.

Karim, M. R., Al Kayed, N., Hossain, M. R., & Rahman, B. M. A. (2020b). Study of low peak-power highly coherent broadband supercontinuum generation through a dispersion-engineered Si-rich silicon nitride waveguide. *Applied Optics, 59*(20), 5948-5956.

Kojima, K., Wang, B., Kamilov, U., Koike-Akino, T., & Parsons, K. (2017, July). Acceleration of FDTD-based inverse design using a neural network approach. In *Integrated photonics research, silicon and Nano photonics* (pp. ITu1A-4). Washington D. C.: Optical Society of America.

LeCun, Y., Touresky, D., Hinton, G., & Sejnowski, T. (1988). A theoretical framework for back-propagation. *Proceedings of the 1988 Connectionist Models Summer School, 1*,21-28.

Liu, D., Tan, Y., Khoram, E., & Yu, Z. (2018). Training deep neural networks for the inverse design of nanophotonic structures. *Acs Photonics, 5*(4), 1365-1369.

Ma, W., Cheng, F., & Liu, Y. (2018). Deep-learning-enabled on-demand design of chiral metamaterials. *ACS Nano, 12*(6), 6326-6334.

Malkiel, I., Nagler, A., Mrejen, M., Arieli, U., Wolf, L., & Suchowski, H. (2017). Deep learning for design and retrieval of nano-photonic structures. *arXiv Preprint arXiv:1702.07949.*

Obayya, S. S. A., Rahman, B. M. A., & Grattan, K. T. V. (2005). Accurate finite element modal solution of photonic crystal fibers. *IEE Proceedings-Optoelectronics, 152*(5), 241-246.

Rivenson, Y., Göröcs, Z., Günaydin, H., Zhang, Y., Wang, H., & Ozcan, A. (2017). Deep learning microscopy. *Optica, 4*(11), 1437-1443.

Singh, N., Vermulen, D., Ruocco, A., Li, N., Ippen, E., Kärtner, F. X., & Watts, M. R. (2019). Supercontinuum generation in varying dispersion and birefringent silicon waveguide. *Optics Express, 27*(22), 31698-31712.

Sinha, A., Lee, J., Li, S., & Barbastathis, G. (2017). Lensless computational imaging through deep learning. *Optica, 4*(9), 1117-1125.

Smirnov, S. V., Ania-Castanon, J. D., Ellingham, T. J., Kobtsev, S. M., Kukarin, S., & Turitsyn, S. K. (2006). Optical spectral broadening and supercontinuum generation intelecom applications. *Optical Fiber Technology, 12*(2), 122-147.

Turduev, M., Bor, E., Latifoglu, C., Giden, I. H., Hanay, Y. S., & Kurt, H. (2018). Ultracompact photonic structure design for strong light confinement and coupling into nanowaveguide. *Journal of Lightwave Technology, 36*(14), 2812-2819.

Wang, T., Ng, D. K., Ng, S. K., Toh, Y. T., Chee, A. K., Chen, G. F., ... & Tan, D. T. (2015). Supercontinuum generation in bandgap engineered, back-end CMOS compatible silicon rich nitride waveguides. *Laser & Photonics Reviews, 9*(5), 498-506.

Yang, M., Xu, L., Wang, J., Liu, H., Zhou, X., Li, G., & Zhang, L. (2018). An octave spanning optical parametric amplifier based on a low-dispersion silicon-rich nitride waveguide. *IEEE Journal of Selected Topics in Quantum Electronics, 24*(6), 1-7.

Yu, C. P., & Chang, H. C. (2004). Applications of the finite difference mode solution method to photonic crystal structures. *Optical and Quantum Electronics, 36*(1), 145-163.

Zia, H., Lüpken, N. M., Hellwig, T., Fallnich, C., & Boller, K. J. (2020). Supercontinuum generation in media with sign-alternated dispersion. *Laser & Photonics Reviews, 14*(7), 2000031.

Zibar, D., Wymeersch, H., & Lyubomirsky, I. (2017). Machine learning under the spotlight. *Nature Photonics, 11*(12), 749-751.