Deep Learning Methods for Optimizing Terahertz Leaky Wave Antenna Design

J. Neronha, H. Guerboukha, and D. Mittleman

School of Engineering, Brown University, Providence, RI 02912

Introduction

The predictive power of neural networks and machine learning techniques more generally have been applied to a vast array problems ever since the neural network was proposed as a computational technique loosely modeling the brain in the mid-twentieth century. [1] This, of course, includes communications technology – neural networks have been used extensively in the field because of their unique ability to approximate accurate solutions to nonlinear problems that are commonplace in antenna design. Machine learning models are particularly useful in situations where an analytical solution cannot be obtained and/or numerical simulations are expensive [2, 3]. These types of problems are usually split into two cases, forward and inverse problems. The former refers to situations where models attempt to generate the electromagnetic response given some parametric input, while the latter attempts the reverse: predicting those parameters given some signal [4].

Examples of direct problems include approximating an antenna's output given a geometry [5] or modeling a metasurface [6], which could take a finite-element model hours or longer to solve, limiting real-time simulation and design. Inverse problem considers the opposite direction, for example reconstructing an image from scattered light [7]. There have been many applications of neural networks to communications-related problems, for example, in the optimization of massive multiple-input multiple-output (MIMO) systems to improve signal reconstruction quality [8] or physical layer design [9]. *MORE EXAMPLES HERE?*

We are particularly interested in the inverse problem of terahertz leaky-wave antenna (LWA) design because of its high applicability to the field of communications – given some arbitrarily desired far-field pattern, how can we quickly design, fabricate, and test an antenna that meets our needs? Periodic leaky wave-antennas are especially interesting for signal design because of their characteristic emissions at various angles, both forward and backwards scattering, that depend on the periodicity, allowing for a wide variety of potential peak angles [10]. And while traditional communications systems demand broad angular emission, the terahertz range (i.e. > 100 GHz) requires focusing power along narrow directional beams due to substantial power loss as a result of free path losses at higher frequencies [11, 12], meaning that highly specific peak profile design is crucial.

LWAs are simple metallic waveguides that have proven very effective in the terahertz range and are particularly interesting because they emit radiation at a frequency-dependent angle with a one-to-one relationship between frequency and angle, which is quite valuable given the narrow character of beams in the terahertz range. [13] They have been used in a number of diverse applications including link discovery [12], multiplexing and demultiplexing [14, 15], and for radar and object

detection purposes [16, 17]. Leaky-wave antennas also stand out for our purposes of rapid design and experimentation because of the ease of fabricating them using hot-stamping techniques [18]. This inverse problem has been explored in the terahertz range using deep neural networks in the context of designing structures to obtain an optimized geometry that produces the desired signal, particularly in relation to metasurface design [19, 20]. The inverse Leaky-wave antenna problem has also been explored, but using a genetic algorithm and outside of the terahertz range at much higher frequencies [21].

As a result, in this paper, we propose a model to predict the ideal LWA geometry that will generate desired far-field radiation in the terahertz range. Specifically, we suggest an antenna design where a LWA is broken into sub-slots, each of which can either be metallic, transparent, or somewhere in between (i.e. partially transparent). Thus, we find that any slot composed of an array of these sub-slots forms a linear superpositioning of periodic modes that can be used to generate highly specific peak profiles.

Methodology

We train a model to predict the slot geometry of the LWA for a desired far-field pattern. We consider a LWA based on a parallel plate waveguide, the fundamental transverse electric (TE_1) mode, and a frequency of 200 GHz. Instead of a uniform 1D slot [13], we discretize the slot into 36 0.5-mm sub-slots that can be transparent, metallic, or somewhere in between (i.e. semi-transparent). This approach is similar to the approach taken in past explorations of metasurface design [22, 21]. According to Floquet theory [10], a periodic array of slots excites an infinite number of space harmonics each with its distinct dispersion constant β_p:

$$\beta = \sqrt{k_0^2 - (\frac{\pi}{h})^2} + \frac{2\pi p}{\Lambda} \tag{1}$$

where k_0 is the wave number, h is the height of the waveguide, and Λ is the slot periodicity. These different modes peak at different angles defined by $\cos(\theta) = \beta_p/k_0$, given that the excited modes can be consider fast-wave (i.e., $|\beta_p| > k_0$). Our primary objective is to generate multiple beams with specific magnitudes at specific locations. Therefore, instead of a periodic geometry, we consider a non-periodic array of slots. By Fourier decomposition, we note that any non-periodic slot design can be viewed as a linear combination of periodic designs β_p, each peaking at an angle $\cos(\theta) = \beta_p/k_0$. The strength (amplitude) of individual peaks can be related to the Fourier coefficients of this decomposition.

Our discretized geometry imposes limits on the possible excited Floquet modes, by constraining the values of Λ. Indeed, due to the discretization, the possible Λ are given by $\Lambda =$. Furthermore, due to the fast-wave requirement (i.e., $|\beta_p| > k_0$), only a few select modes will be able to leak out of the waveguide.

A schematic is shown in Figure 1 below.

see, I can keep typing things.

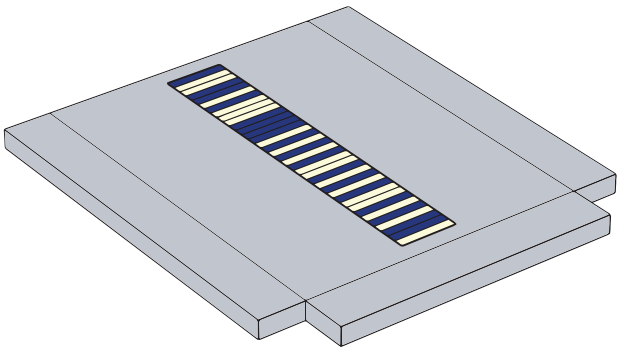


Figure 1: The leaky wave antenna geometry. Dark-colored sub-slots represent transparent areas while light colors indicate metallic boundaries; in order to maintain constant power, the number of each type of slot is kept equal.

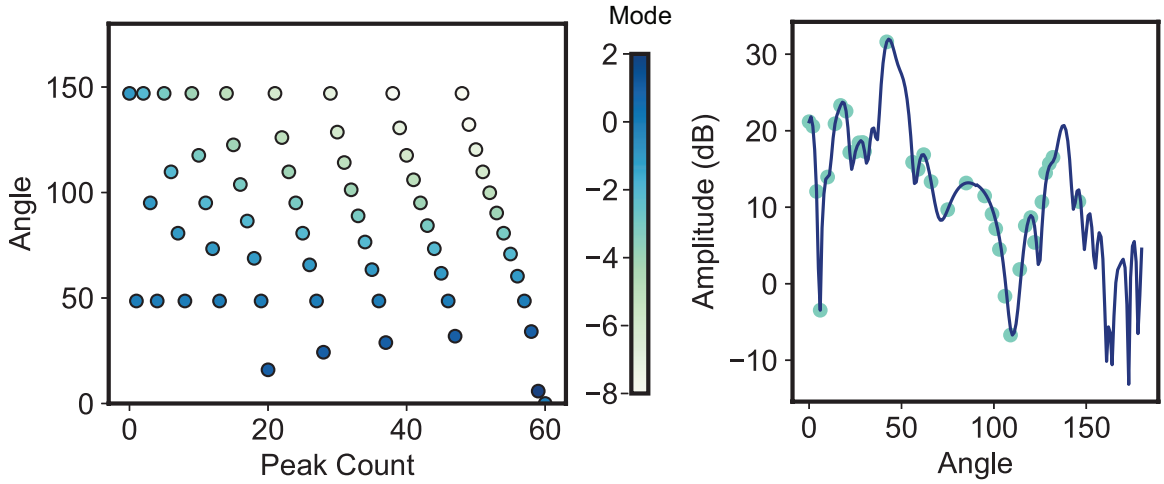
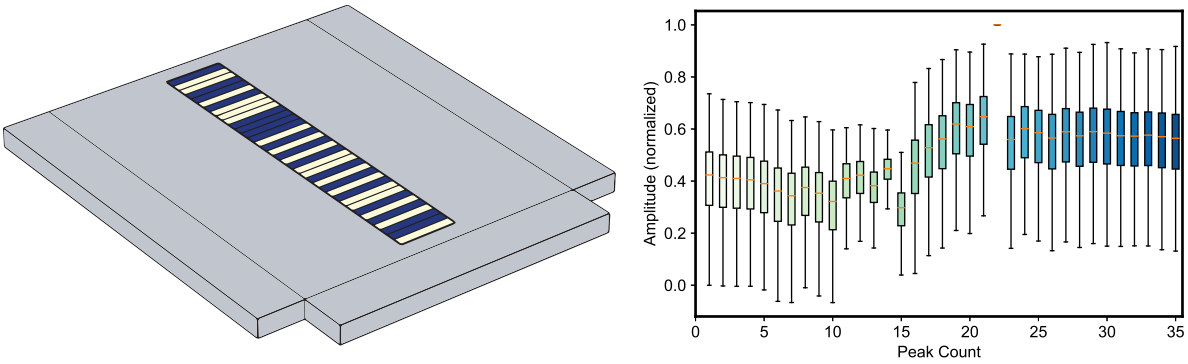


Figure 2: The left panel shows the possible Floquet modes from any linear combination of modes predicted by Equation 1; as expected, the positive modes occur in forward scattering (¡ 90deg) and the negative modes in back-scattering. The right panel shows these predicted Floquet modes overlaid on a sample peak profile generated in COMSOL.

To generate our training data, we build a model of a leaky-wave antenna in COMSOL Multiphysics with a slot of length 18 mm divided into 36 sub-slots, each of which is 500 microns in length. 18 sub-slots (exactly half) of the slots are randomly chosen as to be transparent (i.e. a scattering boundary layer) whereas the balance remain metallic and thus do not leak radiation. We automate this random geometry generation and simulation using MPh [23], an open-source Python package that enables controlling COMSOL via its API. Basic combinatorics theory tells us that there are over 9 billion possible slot designs that can be generated, indicating the usefulness of deep neural networks in this prediction task given the sheer number of possible outputs; using only a few tens of thousands sets of training data, we can predict an optimal slot design for an arbitrary peak profile. The simulation geometry is shown in Figure 2a below

Figure 3: Simulation geometry (left) and possible peak profiles (right)



(a) The leaky wave antenna geometry used to generate training data. Dark-colored sub-slots represent transparent areas while light colors indicate metallic boundaries; both are randomly generated so that the slots are half and half.

(b) A box plot showing the amplitudes at each peak; peak #23 is always the highest amplitude because it is the primary emission angle of all LWAs, periodic and not, at 200 GHz. A wide array of values are evident.

25,000 instances of COMSOL simulation data is used as the data source for our deep neural network, with a test-train split of 80-20 [24]. Figure 2b above shows the possible values for each peak, giving an approximate range over which we can design peak profiles for this particular geometry. The network, which is built with TensorFlow [25], a popular deep-learning library, has one convolutional layer for feature extraction connected to six sequential fully-connected layers, each of which have 2000 neurons and use a LeakyReLU activation function with $\alpha = 0.1$. The model uses the Adam optimizer and has a learning rate of 0.0001 and predicts the output of each individual sub-slot with a sigmoid activation function indicating the probability that a given sub-slot should be transparent. The model architecture is visualized in Figure 3.

The tricky part of the inverse problem, as we consider here, is finding an appropriate loss function – the primary objective is to generate a slot that minimizes the difference in signal compared to what is desired, but in order to use this idea as our loss function, we would need to compute the far-field signal for each slot we generate as we train the model, which would be extraordinary (and unfeasibly) computationally expensive. Thus, instead, we turn to the idea of making the generated

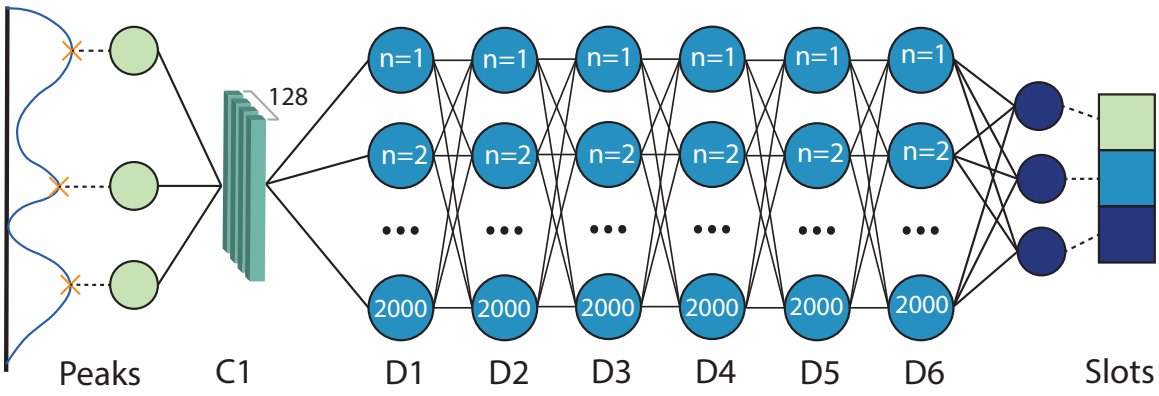


Figure 4: Neural network architecture schematic; peaks are used as the input data and passed through six hidden layers to predict the ideal slot design that will yield the closest possible signal.

slot look as similar to the given training slot for a given peak profile as possible. Given that this is essentially a binary classification problem for each sub-slot, we use three weighted binary cross-entropy (BCE) loss functions, which are summed together to form our overall loss function with the following weightings:

1. BCE between the true and predicted slot (2x)
2. BCE between the first derivative of the true and predicted slot using a Prewitt filter (1x)
3. BCE between the second derivative of the true and predicted slot using a Laplacian filter (1x)

Results and Discussion

Binary Predictions

For some given peak geometry, the model returns each individual sub-slot's probability of being transparent (i.e. a value close to 0 indicates the sub-slot should be metal, whereas a value close to 1 indicates a transparent sub-slot) using a sigmoid activation function. In Figure 4 below, for two sample peak profiles, the true slots (left), predicted slot (center) and rounded ML slot (right, i.e. making the 16 slots with the highest probabilities transparent and vice versa). Both of these predicted slots share many similarities with the true slot but there are also some differences – the testing accuracy was about 68% averaged over all of the training samples.

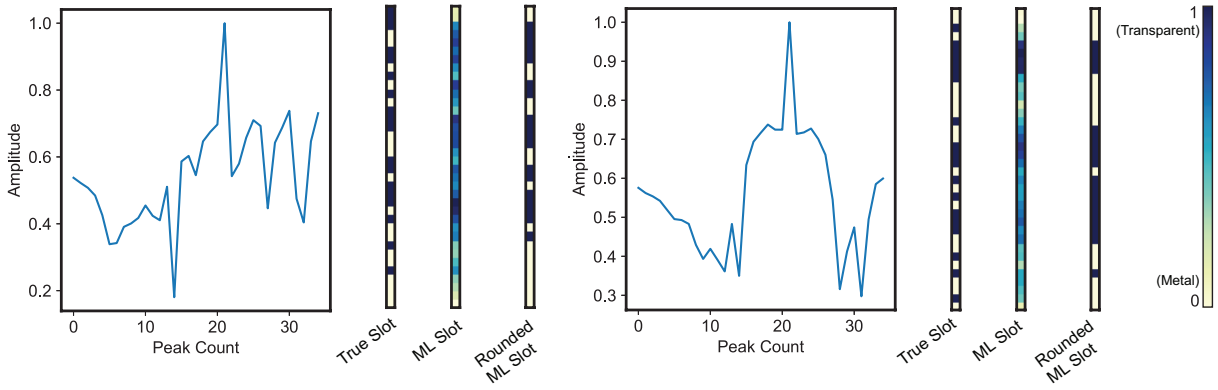


Figure 5: Model predictions for two different peak profiles. The left-most slot design is the randomly generated slot used to generate this profile, the center plot shows the sigmoid probabilities of a transparent sub-slot, and the right-most plot shows the top 16 sub-slots from the probabilities.

However, what determines the usefulness of our model is not the similarity of the slot but the similarity of the signal it generates to the objective. In order to consider the accuracy and usefulness of our model, we extract the predicted slot geometries, choose the top 16 slots based on the probabilities, and pass 500 of these new generated slots into our COMSOL model in order to compare the signal from the ML-generated slot to the originally desired peak profile. We then compare the peak profiles to the objective signal and calculate using the mean square error (MSE), as shown in the top left panel of Figure 5 below; the shape of the distribution matches that of similar work [6]. In the other panels of Figure 2, we plot the objective peak profile in black against the peak profile of an ML-generated slot, where the color corresponds to the four colored buckets depicted in the histogram. Overall, the agreement is quite excellent, particularly for the first two buckets (green and light blue), which compose about 60% of testing samples with trends and relative

amplitudes of the peaks agreeing very well. Even in the lower-tiered buckets, like in the bottom right panel of Figure 2, many of the peak profile trends are well matched – however, there are some areas where the model does not perform as well, resulting in a higher MSE. Overall, though, the agreement is excellent and shows that the model can be used to design an acceptable slot for many potentially desired peak profiles.

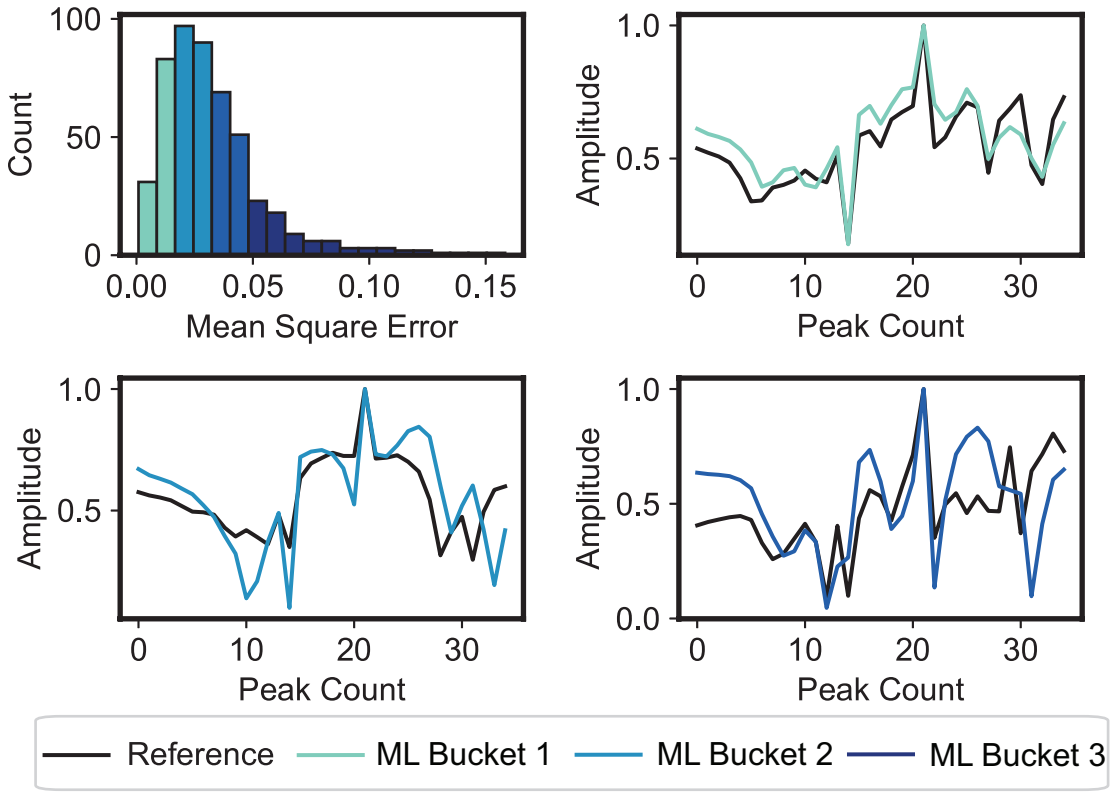


Figure 6: Results from COMSOL simulations of ML-generated slots. The top left panel shows an MSE histogram while the others compare the ML-slot generated peak profile to the objective.

Conclusion

References

[1] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[2] Y. Kim, "Application of machine learning to antenna design and radar signal processing: A review," in *2018 International Symposium on Antennas and Propagation (ISAP)*, pp. 1–2, 2018.

[3] A. Massa, D. Marcantonio, X. Chen, M. Li, and M. Salucci, "Dnns as applied to electromagnetics, antennas, and propagation—a review," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 11, pp. 2225–2229, 2019.

[4] L.-Y. Xiao, W. Shao, F.-L. Jin, B.-Z. Wang, and Q. H. Liu, "Inverse artificial neural network for multiobjective antenna design," *IEEE Transactions on Antennas and Propagation*, vol. 69, no. 10, pp. 6651–6659, 2021.

[5] H. M. Yao and L. J. Jiang, "Machine learning based neural network solving methods for the fdtd method," in *2018 IEEE International Symposium on Antennas and Propagation USNC/URSI National Radio Science Meeting*, pp. 2321–2322, 2018.

[6] C. C. Nadell, B. Huang, J. M. Malof, and W. J. Padilla, "Deep learning for accelerated all-dielectric metasurface design," *Opt. Express*, vol. 27, pp. 27523–27535, Sep 2019.

[7] Y. Sun, Z. Xia, and U. S. Kamilov, "Efficient and accurate inversion of multiple scattering with deep learning," *Opt. Express*, vol. 26, pp. 14678–14688, May 2018.

[8] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive mimo csi feedback," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, 2018.

[9] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Deep learning based MIMO communications," *CoRR*, vol. abs/1707.07980, 2017.

[10] F. Xu and K. Wu, "Understanding leaky-wave structures: A special form of guided-wave structure," *IEEE Microwave Magazine*, vol. 14, no. 5, pp. 87–96, 2013.

[11] J. Ma, R. Shrestha, L. Moeller, and D. M. Mittleman, "Invited article: Channel performance for indoor and outdoor terahertz wireless links," *APL Photonics*, vol. 3, no. 5, p. 051601, 2018.

[12] Y. Ghasempour, R. Shrestha, A. Charous, E. Knightly, and D. M. Mittleman, "Single-shot link discovery for terahertz wireless networks," *Nature Communications*, vol. 11, no. 1, p. 2017, 2020.

[13] H. Guerboukha, R. Shrestha, J. Neronha, O. Ryan, M. Hornbuckle, Z. Fang, and D. M. Mittleman, "Efficient leaky-wave antennas at terahertz frequencies generating highly directional beams," *Applied Physics Letters*, vol. 117, no. 26, p. 261103, 2020.

[14] N. J. Karl, R. W. McKinney, Y. Monnai, R. Mendis, and D. M. Mittleman, "Frequency-division multiplexing in the terahertz range using a leaky-wave antenna," *Nature Photonics*, vol. 9, no. 11, pp. 717–720, 2015.

[15] J. Ma, N. J. Karl, S. Bretin, G. Ducournau, and D. M. Mittleman, "Frequency-division multiplexer and demultiplexer for terahertz wireless links," *Nature Communications*, vol. 8, no. 1, p. 729, 2017.

[16] Y. Amarasinghe, R. Mendis, and D. M. Mittleman, "Real-time object tracking using a leaky thz waveguide," *Opt. Express*, vol. 28, pp. 17997–18005, Jun 2020.

[17] Y. Amarasinghe, H. Guerboukha, Y. Shiri, and D. M. Mittleman, "Bar code reader for the thz region," *Opt. Express*, vol. 29, pp. 20240–20249, Jun 2021.

[18] H. Guerboukha, Y. Amarasinghe, R. Shrestha, A. Pizzuto, and D. M. Mittleman, "High-volume rapid prototyping technique for terahertz metallic metasurfaces," *Opt. Express*, vol. 29, pp. 13806–13814, Apr 2021.

[19] Y. Deng, S. Ren, K. Fan, J. M. Malof, and W. J. Padilla, "Neural-adjoint method for the inverse design of all-dielectric metasurfaces," *Opt. Express*, vol. 29, pp. 7526–7534, Mar 2021.

[20] E. Dejband, J.-W. Li, P.-C. Peng, and T.-H. Tan, "Prediction of thz absorption and inverse design of graphene-based metasurface structure using deep learning," in *2021 30th Wireless and Optical Communications Conference (WOCC)*, pp. 21–23, 2021.

[21] S. Jafar-Zanjani, S. Inampudi, and H. Mosallaei, "Adaptive genetic algorithm for optical metasurfaces design," *Scientific Reports*, vol. 8, no. 1, p. 11040, 2018.

[22] C. Liu, Q. Ma, Z. J. Luo, Q. R. Hong, Q. Xiao, H. C. Zhang, L. Miao, W. M. Yu, Q. Cheng, L. Li, and T. J. Cui, "A programmable diffractive deep neural network based on a digital-coding metasurface array," *Nature Electronics*, vol. 5, no. 2, pp. 113–122, 2022.

[23] J. Hennig, M. Elfner, and J. Feder, "Mph-py/mph: Mph 1.1.5," Feb. 2022.

[24] A. Rácz, D. Bajusz, and K. Héberger, "Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification," *Molecules*, vol. 26, no. 4, 2021.

[25] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.