Explainable Classification Techniques for Quantum Dot Device Measurements

Daniel Schug[1,2], Tyler J. Kovach[3], M. A. Wolfe[3], Jared Benson[3], Sanghyeok Park[3], J. P. Dodson[3*], J. Corrigan[3†], M. A. Eriksson[3], Justyna P. Zwolak[2]

[1] Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, USA
[2] National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA
[3] Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA
dschug1@umd.edu, tkovach@wisc.edu, mawolfe4@wisc.edu, jcbenson2@wisc.edu, sanghyeok.park@wisc.edu, jpdodson42@gmail.com, joelle.corrigan@intel.com, maeriksson@wisc.edu, jpzwolak@nist.gov

Abstract

In the physical sciences, there is an increased need for robust feature representations of image data: image acquisition, in the generalized sense of two-dimensional data, is now widespread across a large number of fields, including quantum information science, which we consider here. While traditional image features are widely utilized in such cases, their use is rapidly being supplanted by Neural Network-based techniques that often sacrifice explainability in exchange for high accuracy. To ameliorate this trade-off, we propose a variant of a Gabor Wavelet Transform-based technique that results in explainable features. We show, using Explainable Boosting Machines (EBMs), that this method offers superior explainability without sacrificing accuracy. Specifically, we show that there is a meaningful benefit to this technique in the context of quantum dot tuning, where human intervention is necessary at the current stage of development.

Introduction

There has been a longstanding trade-off between the accuracy of a candidate machine learning (ML) model and its interpretability. This is evident in the extreme example of deep neural networks (DNNs), which can offer excellent accuracy for many problems but are limited in their interpretability due to the number of inaccessible layers. Alternatively, there are simple techniques, such as linear models or decision trees, that offer the user full comprehension of the internal weights. However, these are often unable to model the complex relationships seen in modern datasets. For tabular data, there has been considerable progress toward finding a middle ground, typically through explaining complex models with surrogates such as LIME (Ribeiro, Singh, and Guestrin 2016) and Shapley (Lundberg and Lee 2017).

In contrast with many black-box ML models, explainable boosting machines (EBMs) are a glass-box method that enables the model to be directly interpretable as opposed to relying on surrogate explanations (Lou et al. 2013). Specifically, EBMs extend generalized additive models to include pairwise interactions, which allows one to observe the relationship between features. EBMs often provide accuracy on par with many black-box models with the additional advantage of enhanced intelligibility, which makes them an appealing replacement for other models.

For spatial data, such as images, interpretability is more challenging. This is in part due to their composition by structures which are highly correlated at multiple scales in the two-dimensional (2D) space. This feature is one of the main reasons why convolutional neural networks (CNNs) have quickly become the dominant ML approach for many computer vision tasks. However, the black-box nature of CNNs makes their use prohibitive in applications of consequence such as healthcare (Caruana et al. 2015) or finances (Israel, Kelly, and Moskowitz 2020). While making CNNs more interpretable has been a very active area of research, techniques proposed to date vary in utility in many applications, especially as the depth of the neural network increases (Adebayo et al. 2018).

Our previous work developed a methodology that addresses some of these concerns by combining vectorization methods to image data with EBMs. The possibility of using EBMs as models for image data poses numerous challenges, the principal of which is the mapping from images to a vector representation that could then be used directly with EBMs. In our previous work, we used the Gabor Wavelet transform in conjunction with a constrained optimization procedure to extract key image features from the data (Schug et al. 2024). We also applied a highly custom feature engineering to tailor this process to the particular dataset (Schug et al. 2023). In both cases, we relied on domain knowledge and understanding of the physical systems under investigation to inform the feature extraction process. This was essential in ensuring that the resulting model produces human-agreeable interpretations. Here, we demonstrate that the same approach can be successfully applied to assist in the tune-up of accumulation mode Si/SiGe quantum dot (QD) devices. We also propose an alternative image vectorization method involving the generation of synthetic data to approximate the experimentally acquired scans. We then show that both methods result in a comparable performance, but the latter produces features that are easier to interpret.

*Present Address: HRL Laboratories, LLC, 3011 Malibu Canyon Road, Malibu, CA 90265, USA
†Present Address: Intel Corp., Hillsboro, OR 97124, USA
XAI4Sci: Explainable machine learning for sciences, AAAI-24 (xai4sci.github.io)

Background: Quantum Dot Tuning Problem and Triangle Plots

Arrays of QDs – electrons confined in a semiconductor heterostructure with unique properties that allow them to act as artificial atoms – are a leading candidate for use as qubits, the fundamental information carriers in quantum computers. We use a type of gate-defined QD, as shown in Figure 1, a three-layer overlapping gate structure fabricated on top of a Si/SiGe heterostructure device is used to form and control QDs (Zajac et al. 2016; Dodson et al. 2020). As the number of QDs and thus the number of gates needed to control them grows, the manual tuning process becomes intractable. An autotuning framework incorporating ML tools was originally proposed and validated off-line using premeasured experimental scans capturing a large range of gate voltages in (Kalantre et al. 2019) and then deployed online (i.e., *in situ*) to tune a double QD in real-time in (Zwolak et al. 2020). A detailed description of the tuning process is available in (Zwolak and Taylor 2023).

One sub-problem within the tuning procedure is to correctly determine the voltage placement of gates to allow the formation of isolated current channels inside the 2D electron gas formed at the intersection of the Si and SiGe layers in the heterostructure, see Figure 1(a). One way to achieve proper gate voltages is by sweeping certain gates [S1, S2, S3, and S4 in Figure 1(a)] until the current begins flowing in the channel. The bright bars at the edges of the images in Figure 1(b) and (c), called *walls*, are due to the current flowing strictly under those gates. The region with the ridged pattern Figure 1(b) is the area of interest – it indicates that the current is not flowing under any screening gates, but instead, it is flowing between them. For the purpose of our study, such images are considered *good*. If there is no current in this region, as shown in Figure 1(c), it is necessary to accumulate all of the gates over the current channel and try again. Images that do not contain the triangle region are considered *bad*.

In gate-defined QD devices, the area of interest usually takes the form of a triangle – thus the name *triangle plots* – and indicates the formation of an isolated current channel. The patterned texture inside of this region is due to charge defects and variability in gate uniformity near the current channel (Ye et al. 2021). While researchers can easily visually identify these regions, finding them automatically using simple threshold analysis is challenging. In moving towards automating the tuning process, it is desirable to have an algorithm that can (1) predict whether the triangle plot indicates a well-behaving current channel and (2) provide an explanation for this prediction.

Methods

The triangle plots form a dataset with image data that has a greater variety of scales and orientations. In addition to the approach used in our previous work (Schug et al. 2023), we consider a radically different approach to producing a vector representation from images that is better adapted to the complexity of the triangle plots dataset. While the original vectorization method produces a model that performs well

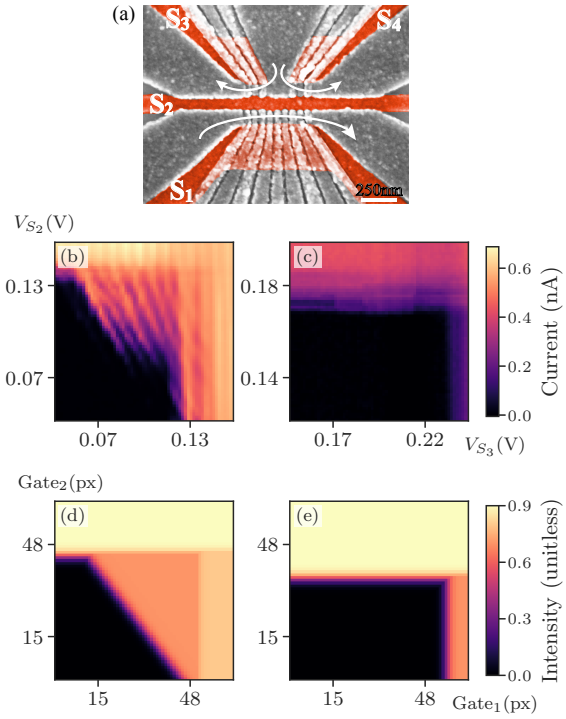

Figure 1: (a) False-colored SEM micrograph of a typical Si/SiGe heterostructure device. The orange highlighted gates are the screening gates that are swept for the triangle plot. This device has 3 distinct current channels between the screening gates marked with white arrows. (b) An example of a good triangle plot with high visibility in the triangle region above the background. (c) An example of a bad triangle plot with little to no current in the triangle region. An example of of (d) good and (e) bad synthetic triangle plot.

with the classification task, the interpretability and usefulness of this model are limited. The new approach produces qualitatively superior explanations.

Gabor Filterbank Approach

To vectorize the image data, we utilize the 2D Gabor wavelet transform, an oriented multi-scale representation that is shown in (Daugman 1988) to be a model for complex neurons in mammalian vision. As such, this representation seems ideal for extracting oriented features, as well as textures, from image data.

Definition 1 *The 2D Gabor Kernel for parameters* $p = \{\sigma_x, \sigma_y, \lambda, \theta\}$ *for* $(x, y) \in \mathbb{R}^2$ *is defined as*

$$G_p(x,y) = \frac{e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)}}{\sqrt{2\pi}\sigma_x\sigma_y} e^{i\lambda(x\sin(\theta)+y\cos(\theta))}, \qquad (1)$$

where σ_x *and* σ_y *represent scale in* x *and* y*, and* θ *and* λ *are the wave direction and wavelength. We further denote the convolutional application of kernel* G_p *to an image* $u(x, y)$ *to be* $G_p(x, y) * u(x, y)$.

Specifically, we consider small filterbanks of Gabor wavelets which capture particular scales and orientations directly relevant to our image data, which leads to a more compact representation.

Definition 2 *The 2D Gabor Filterbank for discrete set of N parameters $P = \{p_0, ..., p_N\}$*

$$G_P(x, y) = \{G_{p_i}(x, y)\}_{i=0}^{N}. \quad (2)$$

We denote the application of the filterbank G_P to an image u to be $G_P * u = \{G_{p_i}(x, y) * u(x, y)\}_{i=0}^{N}$

In practice, we attempt to construct P such that the Fourier transform of G_P has support covering prominent frequencies observed in the Fourier transform of u. This can be accomplished with optimization or other filterbank construction techniques. For simplicity, we perform the construction manually.

To produce the vector representation of the image, we use the physical understanding that the dominant discriminant factor in discriminating *good* and *bad* triangles is the presence of line-like features at approximately 45° orientation corresponding to the interaction between the two gates, see Figure 1(b). The image is vectorized by taking the L^2-norm and TV^2-norm of Gabor filters with this orientation at different scales. We also perform additional feature engineering in the form of extracting the estimated location of this feature and the location of the bright bars, which we accomplish with narrow edge-like Gabor filters.

Synthetic Triangle Plots

To enable the creation of interpretable features, we generate a set of crude synthetic triangle plots. The synthetic data share certain salient features visible in real triangle plots, namely the presence of walls of varying width and a possible diagonal region. It is important to note that, while the resulting images are visually similar to experimental data, as depicted in Figure 1(d) and (e), this approach does not produce physically realizable data.

For generating the synthetic triangles we utilize the widely known 2D sigmoid function $s_{m,r,b}(x) = \frac{m}{1+\exp(r(x-b))}$, where m, r, and b control the magnitude, rate, and shift of the sigmoid, respectively. The motivation for this choice is the resemblance of the idealized walls and triangle region to the Heaviside step function. In practice, the transition from no signal to wall and triangle should be smoother. Triangles are assumed to have a horizontal and vertical wall corresponding to individual gates, as well as a possible third diagonal region linking the walls. To construct the 2D triangle plot we compose three sigmoids:

$$\Delta_{(x,y)} = \max\left(s_h(x), s_v(y), s_d(x\sin\theta + y\cos\theta)\right) \quad (3)$$

where s_h, s_v, and s_d are the parameters for the horizontal, vertical, and diagonal sigmoid, respectively, and θ is the orientation of the diagonal sigmoid. In practice, we evaluate this function on an evenly spaced grid of (x, y) points.

The criteria used to determine the presence of a triangle region are derived trivially from the definition of the triangle plot: we require that $s_d(x\sin(\theta) + y\cos(\theta))$ is greater than s_h and s_v. We also impose a constraint that the orientation of the diagonal sigmoid is sufficiently diagonal, with the cutoff defined as $0 \leq \theta \leq \pi/2$.

It is important to restrict the parameters used to define the synthetic triangles to physically reasonable ranges. The simple constraints we chose include a requirement that the sigmoids cannot be infinitely dilated and that the walls must be in the image and have positive heights.

Synthetic Triangle Fitting Approach

Since our images have an assumed form of a triangle plot, we can use the synthetic triangles to model real triangles. In other words, for a given experimentally acquired triangle plot, the parameters defining a synthetic triangle most similar to it visually and sharing all of the important visual features can be used to parameterize it. The resulting vector provides a powerful and compact representation for ML. Moreover, this vectorization approach has a key advantage in that the vector features correspond directly to our original image, and that fewer assumptions about the frequency content of the data are made.

We start by defining an appropriate measure of image similarity. The experimentally acquired triangle plots contain considerable noise, texture, and other high-frequency features that are not essential in characterizing the overall structure. The synthetic triangles can serve as a representation of the boundaries of the vertical, horizontal, and diagonal walls defined by a shifted, scaled, and dilated sigmoid function. Since, by design, the synthetic and experimentally acquired plots are assumed to be structurally similar, both can be transformed into a region of the frequency domain that they are expected to share.

In practice, this involves applying an ideal low-pass filter, such as the Gaussian. To focus on the details of the structure, we use the magnitude gradient of the low-pass filtered image. While this can be thought of as a simple edge detection, we do not have *a priori* knowledge about the appropriate scale of the Gaussian for any particular image. To quantify the similarity between an experimentally acquired and synthetic triangle plot, we use the L_2-norm applied to both thus transformed images.

Definition 3 *Let U and V be two images subject to some transform F with parameters p. The similarity measure between U and V is defined as*

$$\mathcal{S}_p^F(U, V) = ||F_p(U) - F_p(V)||_2 \quad (4)$$

The desired vector representation of an experimentally acquired image $\mathcal{I}_{\text{exp}}$ is obtained by optimizing the parameters defining the corresponding synthetic image $\mathcal{I}_{\text{sim}}$ such that $\mathcal{S}_p^F(\mathcal{I}_{\text{exp}}, \mathcal{I}_{\text{sim}})$ is minimized subject to boundary constraints. We set up the optimization problem as:

$$p^*, \sigma^* = \min_{p,\sigma} \big\{ \lambda\, \mathcal{S}_\sigma^{DG}((\mathcal{I}_{\text{exp}}, \mathcal{I}_{\text{sim}}) + (1-\lambda)\, \mathcal{S}^I((\mathcal{I}_{\text{exp}}, \mathcal{I}_{\text{sim}}) + \epsilon(p) \big\}, \quad (5)$$

where $DG_\sigma(U) = \nabla(G_\sigma(U))$ is the gradient of the Gaussian transform; $I(U) = U$ is the identity transform; λ is a hyperparameter giving us control over the balance

between trusting DG_σ and I similarity measures; and $\epsilon(p) = \{C : \text{p violates boundary constraints}, 0 : \text{otherwise}\}$ for some arbitrarily large penalty constant C.

We use the Nelder-Mead optimizer (Nelder and Mead 1965) implemented in SciPy (Jones et al. 2001). For the initial guess, we use values in the middle of the boundary constraints, although we notice some sensitivity to this choice (which is reasonable for local optimizers). We assume that the low-pass filtering in the transform leads to more consistent convergence evidenced by a heuristic determination that $\lambda \geq 0.5$ yields qualitatively better fits.

The optimal parameters, as well as the value of the objective function at those parameters, define the final vector representation of image $\mathcal{I}_{exp}$:

$$\mathbf{v}(\mathcal{I}_{exp}) = [\sigma^*, \mathcal{H}_{(B,M,R)}, \mathcal{V}_{(B,M,R)}, \mathcal{D}_{(B,M,R,\theta)}, \mathcal{F}] \quad (6)$$

where B, M, and R denote the boundary, magnitude, and rate for the optimal horizontal $\mathcal{H}$, vertical $\mathcal{V}$, and diagonal $\mathcal{D}$ boundary, respectively; θ indicates the orientation of the diagonal boundary; σ^* is the optimal scale [see. Eq. (5)]; and $\mathcal{F}$ is the value of the objective function for the optimized parameters.

Results

For the experiments, we use 902 experimentally acquired triangle plots resized to 64 by 64-pixels. Reshaping the images is done for convenience so that we can use fixed parameter filterbanks, and so that synthetic triangles have comparable ranges of parameters. Each image is then vectorized following the procedure described in the *Synthetic Triangle Fitting Approach* section. We present the results of several experiments, starting with a qualitative assessment of the fitting procedure.

In the transformed space, the synthetic images look qualitatively strikingly similar to the experimentally acquired images, which is a strong indicator that the fits are capturing the critical structural information. The optimal fit plot captures the location and average intensity of the walls and diagonal component, despite the presence of the ridged pattern observed in the experimental data.

Performance and Interpretability

The vectorized features $\mathbf{v}(\mathcal{I}_{exp})$ are used to train an EBM. We utilize the smoothing and greedy rounds parameters to improve the accuracy and interpretability at a modest training time cost (Nori et al. 2019). For a more thorough discussion of this, see (Nori et al. 2021). We carry out five 6-fold stratified cross-validations and report averaged results, with precision and recall analyzed at the class level. The results from all experiments are presented in Table 1. The relatively close performance of the more intuitive synthetic triangle fitting approach compared to the Gabor filterbank approach confirms the validity of the new vectorization method.

To understand how the model makes decisions globally we rely on the EBM's feature importance, as well as the individual plots of the important features. The feature importance, shown in Figure 2(a), reveals that the model is indeed picking up the correct features, with $\mathcal{D}_B$, $\mathcal{D}_\theta$, $\mathcal{D}_M$, and

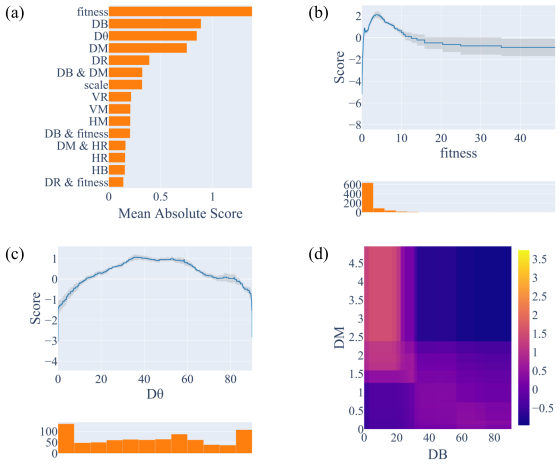

Figure 2: Global EBM results for the synthetic triangle fitting approach. (a) The feature importance plot. (b) Feature curve for the fitness feature $\mathcal{F}$. (c) Feature curve for the hypotenuse orientation feature $\mathcal{D}_\theta$. (d) Pairwise interaction plot for features $\mathcal{D}_B$ and $\mathcal{D}_M$.

$\mathcal{D}_R$ being among top five most important features. Somewhat unexpected is the emphasis on the overall fitness $\mathcal{F}$ as the most important feature. Looking at the feature curve for $\mathcal{F}$ depicted in Figure 2(b), we see that bad images have a lower cost function in fitting, whereas good images tend to fit more poorly. This correlation is reasonable considering that the triangle regions in good images often have a significantly more complicated structure than in bad images whereas walls are practically very similar to the vertical and horizontal sigmoid.

The angle of the diagonal boundary, shown in Figure 2(c), aligns well with the intuition that the hypotenuse in good triangles is ideally oriented at $(45 \pm 15)^\circ$, which is where we see peak scores for $\mathcal{D}_\theta$. We also see negative scores as $\mathcal{D}_\theta$ approaches 0° and 90° for diagonal boundaries, as expected.

Figure 2(d) shows the pairwise interaction between $\mathcal{D}_B$ (length of the hypotenuse) and $\mathcal{D}_M$ (intensity of response within the triangle region). The interaction between these features reveals the ability of the EBM model to recognize that even when the diagonal boundary is not complete, if the diagonal magnitude is large, the scan is likely good. This

Metric	Class	Model	
		Gabor Filt.	Synth Tri.
Accuracy		92.2(7)	87.4(1.4)
Precision	Good	87.1(1.7)	78.6(3.0)
	Bad	93.6(1.2)	89.4(1.5)
Recall	Good	77.8(4.3)	62.2(6.2)
	Bad	96.5(6)	94.9(9)

Table 1: The results of five 6-fold cross-validation for the experimentally acquired data. The value(uncertainty) notation is used to express uncertainties.

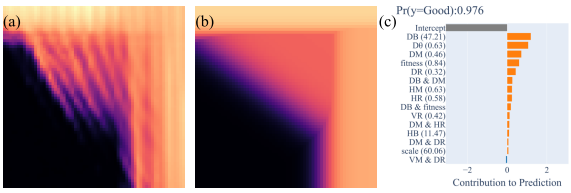

Figure 3: (a) A sample experimentally acquired scan, the same as in Fig. 1(b). (b) Synthetic fit to scan shown in (a). (c) Local EBM results for the fitted synthetic triangle shown in (b).

corresponds to the case where the triangle region is towards the top right corner of the image, as in Figure 1(b), and is in line with our intuition.

EBMs also gives us the ability to examine predictions for individual images, which in our case also has a clear visual meaning as a consequence of our approach. Figure 3(a) and (b) show an example of a good image and its synthetic fit, respectively. The fit effectively captures the structural content of the observed triangle region. In the relative feature importance plot, shown in Figure 3(c), the dominant features – $\mathcal{D}_B$, $\mathcal{D}_\theta$, and $\mathcal{D}_M$ – are largely associated with the strong diagonal component of the observed triangle, and, consequentially, the model is extremely confident that this is a good plot.

Conclusion

In this work, we demonstrate an alternative approach to vectorizing image data for use with EBMs that relies on generating synthetic data that best fit the experimentally acquires scans. While the Gabor filterbank features are somewhat more performant on this dataset, they may be less robust in the sense that they do not comprehend the underlying structure of the data. The proposed method retains comparable accuracy, performing only slightly worse than the Gabor filterbank method, while producing features that are in strong alignment with the human intuition about the data and can be tied directly back to the scientific problem.

Finally, despite the ability of the synthetic triangle fitting approach to fit the majority of scans in the dataset, it fails in about $5\ \%$ more cases than Gabor filterbank approach. It is reasonable to assume that the Gabor technique excels due to its ability to represent the ridged pattern in the triangle region which is considered an important characteristic of a good scan. Refining the penalty function and incorporating some of the filters from the filterbank may help remedy this issue. Other measures of similarity may improve the methods ability to fit in the presence of artifacts, and improve the overall visual correspondence between our synthetic and observed triangle plots.

Acknowledgments

We acknowledge Patrick Walsh for experimental assistance. We acknowledge HRL Laboratories, LLC for support and L.F. Edge for providing one of the Si/SiGe heterostructures used in this work. This research was sponsored in part by the Army Research Office (ARO) under Grant Nos. W911NF-23-1-0110 and W911NF-17-1-0274. Development and maintenance of the growth facilities used for fabricating one of the heterostructure samples were supported by DOE (DE-FG02-03ER46028). We acknowledge the use of facilities supported by NSF through the UW-Madison MRSEC (DMR-2309000). The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government or the ARO. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright noted herein. Any mention of equipment, instruments, software, or materials; it does not imply recommendation or endorsement by the National Institute of Standards and Technology.

References

Adebayo, J.; Gilmer, J.; Muelly, M.; Goodfellow, I.; Hardt, M.; and Kim, B. 2018. Sanity Checks for Saliency Maps. In *Advances in Neural Information Processing Systems*, volume 31.

Caruana, R.; Lou, Y.; Gehrke, J.; Koch, P.; Sturm, M.; and Elhadad, N. 2015. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1721–1730.

Daugman, J. 1988. Complete discrete 2-d gabor transforms by Neural Networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(7): 1169–1179.

Dodson, J. P.; Holman, N.; Thorgrimsson, B.; Neyens, S. F.; MacQuarrie, E. F.; McJunkin, T.; Foote, R. H.; Edge, L. F.; Coppersmith, S. N.; and Eriksson, M. A. 2020. Fabrication process and failure analysis for robust quantum dots in silicon. *Nanotechnology*, 31: 505001.

Israel, R.; Kelly, B.; and Moskowitz, T. 2020. Can Machines 'Learn' Finance? *J. Invest. Manag.*, 18(2): 23–36.

Jones, E.; Oliphant, T.; Peterson, P.; et al. 2001. SciPy: Open source scientific tools for Python. [Online; accessed February 15, 2024].

Kalantre, S. S.; Zwolak, J. P.; Ragole, S.; Wu, X.; Zimmerman, N. M.; Stewart, M. D.; and Taylor, J. M. 2019. Machine learning techniques for state recognition and auto-tuning in quantum dots. *npj Quantum Inf.*, 5(6): 1–10.

Lou, Y.; Caruana, R.; Gehrke, J.; and Hooker, G. 2013. Accurate intelligible models with pairwise interactions. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 623–631.

Lundberg, S. M.; and Lee, S.-I. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, volume 30.

Nelder, J. A.; and Mead, R. 1965. A Simplex Method for Function Minimization. *Comput. J.*, 7(4): 308–313.

Nori, H.; Caruana, R.; Bu, Z.; Shen, J. H.; and Kulkarni, J. 2021. Accuracy, Interpretability, and Differential Privacy via Explainable Boosting. *arXiv:2106.09680*.

Nori, H.; Jenkins, S.; Koch, P.; and Caruana, R. 2019. InterpretML: A Unified Framework for Machine Learning Interpretability. *arXiv:1909.09223*.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Schug, D.; Yerramreddy, S.; Caruana, R.; Greenberg, C.; and Zwolak, J. P. 2023. Extending Explainable Boosting Machines to Scientific Image Data. *arXiv:2305.16526*.

Schug, D.; Yerramreddy, S.; Egleston, B.; Caruana, R.; Spielman, I. B.; Greenberg, C.; and Zwolak, J. P. 2024. Explainable Machine Learning for Ultracold Atoms Image Data. *(in preparation)*.

Ye, F.; Ellaboudy, A.; Albrecht, D.; Vudatha, R.; Jacobson, N. T.; and Nichol, J. M. 2021. Characterization of individual charge fluctuators in Si/SiGe quantum dots. *arXiv:2401.14541*.

Zajac, D. M.; Hazard, T. M.; Mi, X.; Nielsen, E.; and Petta, J. R. 2016. Scalable Gate Architecture for a One-Dimensional Array of Semiconductor Spin Qubits. *Phys. Rev. Appl.*, 6(5): 054013.

Zwolak, J. P.; McJunkin, T.; Kalantre, S. S.; Dodson, J.; MacQuarrie, E. R.; Savage, D.; Lagally, M.; Coppersmith, S.; Eriksson, M. A.; and Taylor, J. M. 2020. Autotuning of Double-Dot Devices In Situ with Machine Learning. *Phys. Rev. Appl.*, 13(3): 034075.

Zwolak, J. P.; and Taylor, J. M. 2023. *Colloquium*: Advances in automation of quantum dot devices control. *Rev. Mod. Phys.*, 95(1): 011006.