Identification of light sources using machine learning (F)

Chenglong You,[1,a] (iD) Mario A. Quiroz-Juárez,[2] (iD) Aidan Lambert,[1] Narayan Bhusal,[1] (iD) Chao Dong,[1]
Armando Perez-Leija,[3] Amir Javaid,[1] Roberto de J. León-Montiel,[2] (iD) and Omar S. Magaña-Loaiza[1]

AFFILIATIONS

[1] Quantum Photonics Laboratory, Department of Physics and Astronomy, Louisiana State University, Baton Rouge,
Louisiana 70803, USA
[2] Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Apartado Postal 70–543,
04510 Ciudad de México, Mexico
[3] Max-Born-Institut, Max-Born-Straße 2A, 12489 Berlin, Germany

[a] Author to whom correspondence should be addressed: cyou2@lsu.edu

ABSTRACT

The identification of light sources represents a task of utmost importance for the development of multiple photonic technologies. Over the last decades, the identification of light sources as diverse as sunlight, laser radiation, and molecule fluorescence has relied on the collection of photon statistics or the implementation of quantum state tomography. In general, this task requires an extensive number of measurements to unveil the characteristic statistical fluctuations and correlation properties of light, particularly in the low-photon flux regime. In this article, we exploit the self-learning features of artificial neural networks and the naive Bayes classifier to dramatically reduce the number of measurements required to discriminate thermal light from coherent light at the single-photon level. We demonstrate robust light identification with tens of measurements at mean photon numbers below one. In terms of accuracy and number of measurements, the methods described here dramatically outperform conventional schemes for characterization of light sources. Our work has important implications for multiple photonic technologies such as light detection and ranging, and microscopy.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5133846

I. INTRODUCTION

The underlying statistical fluctuations of the electromagnetic field have been widely utilized to identify diverse sources of light.[1,2] In this regard, the Mandel parameter constitutes an important metric to characterize the excitation mode of the electromagnetic field and consequently to classify light sources.[3] Similarly, the degree of optical coherence has also been extensively utilized to identify light sources.[3–6] Despite the fundamental importance of these quantities, they require large amounts of data, which impose practical limitations.[6–10] This problem has been partially alleviated by incorporating statistical methods, such as bootstrapping, to predict unlikely events that are hard to measure experimentally.[6,8–11] Unfortunately, the constraints of these methods severely impact the realistic implementation of photonic technologies for metrology, imaging, remote sensing, and microscopy.[10,12–17]

The potential of machine learning has motivated novel families of technologies that exploit self-learning and self-evolving features of artificial neural networks to solve a large variety of problems in different branches of science.[18,19] Conversely, quantum mechanical systems have provided new mechanisms to achieve quantum speedup in machine learning.[20,21] In the context of quantum optics, there has been an enormous interest in utilizing machine learning to optimize quantum resources in optical systems.[22–26] As a tool to characterize quantum systems, machine learning has been successfully employed to reduce the number of measurements required to perform quantum state discrimination, quantum separability, and quantum state tomography.[25–32]

In this article, we demonstrate the potential of machine learning to perform discrimination of light sources at extremely low light levels. This is achieved by training single artificial neurons with the statistical fluctuations that characterize coherent and thermal states of light. The self-learning features of artificial neurons enable the dramatic reduction in the number of measurements and the number of photons required to perform identification of light sources. For the first time, our experimental results demonstrate the possibility of using tens of measurements to identify light sources with mean photon numbers below one. In addition, we demonstrate similar experimental results

using the naive Bayes classifier, which are outperformed by our single neuron approach. Finally, we present a discussion on how a single artificial neuron based on an ADAptive LINear Element (ADALINE) model can dramatically reduce the number of measurements required to discriminate signal photons from ambient photons. These results are validated through the Helstrom and Chernoff bounds. Our work has strong implications for realistic implementation of light detection and ranging (LiDAR), remote sensing, and microscopy.

II. EXPERIMENTAL SETUP AND MODEL

As shown in Fig. 1(a), we utilize a continuous-wave (CW) laser beam that is divided by a 50:50 beam splitter. The transmitted beam is focused onto a rotating ground glass, which is used to generate pseudo-thermal light with super-Poissonian statistics. The beam emerging from the ground glass is collimated using a lens and attenuated by neutral-density (ND) filters to mean photon numbers below one. The attenuated beam is then coupled into a single-mode fiber (SMF). The fiber directs photons to a superconducting nanowire single-photon detector (SNSPD). Furthermore, the beam reflected by the beam splitter is used as a source of coherent light. This beam, characterized by Poissonian statistics, is also attenuated, coupled into a SMF and detected by another SNSPD. The SNSPDs' bias voltages are set to achieve high-efficiency photon counting with less than five dark counts per second. The mean photon number of the coherent beam is matched to that of the pseudo-thermal beam of light.

In order to perform photon counting with our SNSPDs, we use the surjective photon counting method described in Ref. [33]. In this case, the transistor -transistor logic (TTL) pulses produced by our SNSPDs were detected and recorded by an oscilloscope. The data were divided in time bins of 1 μs, which corresponds to the coherence time of our CW laser. Moreover, the 20 ns recovery time of our SNSPDs ensured that we perform measurements on a single-temporal-mode field. Voltage peaks above $\sim$0.5 V were considered as one photon event. The number of photons (voltage peaks) in each time bin was counted to retrieve photon statistics. The effect of dark counts was omitted, since the average number for dark counts per measurement is less than 5×10^{-6}. These events were then used for training and testing our ADALINE neuron and naive Bayes classifier.

The probability of finding n photons in coherent light is given by $P_{coh}(n) = e^{-\bar{n}}(\bar{n}^n/n!)$, where $\bar{n}$ denotes the mean photon number of the beam. Furthermore, the photon statistics of thermal light is given by $P_{th}(n) = \bar{n}^n/(\bar{n}+1)^{n+1}$. It is worth noting that the photon statistics of thermal light is characterized by random intensity fluctuations with a variance greater than the mean number of photons in the mode. For coherent light, the maximum photon-number probability sits around $\bar{n}$. For thermal light, the maximum is always at vacuum. However, when the mean photon number is low, the photon number distribution for both kinds of light becomes similar. Consequently, it becomes extremely difficult to discriminate one source from the other. The conventional approaches to discriminate light sources make use of millions of measurements.[7,9,34,35] Unfortunately, these methods are not only time consuming, but also impose practical limitations.

In order to dramatically reduce the number of measurements required to identify light sources, we make use of an ADALINE neuron. ADALINE is a single neural network model based on a linear processing element, proposed by Bernard Widrow,[36] for binary classification. In general, the neural networks undergo two stages: training and test. In the training stage, ADALINE is capable of learning the correct outputs (named as output labels or classes) from a set of inputs, so-called features, by using a supervised learning algorithm. In the test stage, this neuron produces the outputs of a set of inputs that were not in the training data, taking as reference the acquired experience in the training stage. Although we tested architectures far more complex than a single neuron for the identification of light sources, we found that a simple ADALINE offers a perfect balance between accuracy and simplicity (for more details, see Appendix A). The structure of the ADALINE model is shown in Fig. 1(b). The neuron input features are denoted by $P(n)$, which corresponds to the probability of detecting n photons, in a single measurement event, for a given light source. Furthermore, the parameters ω_i are the synaptic weights and
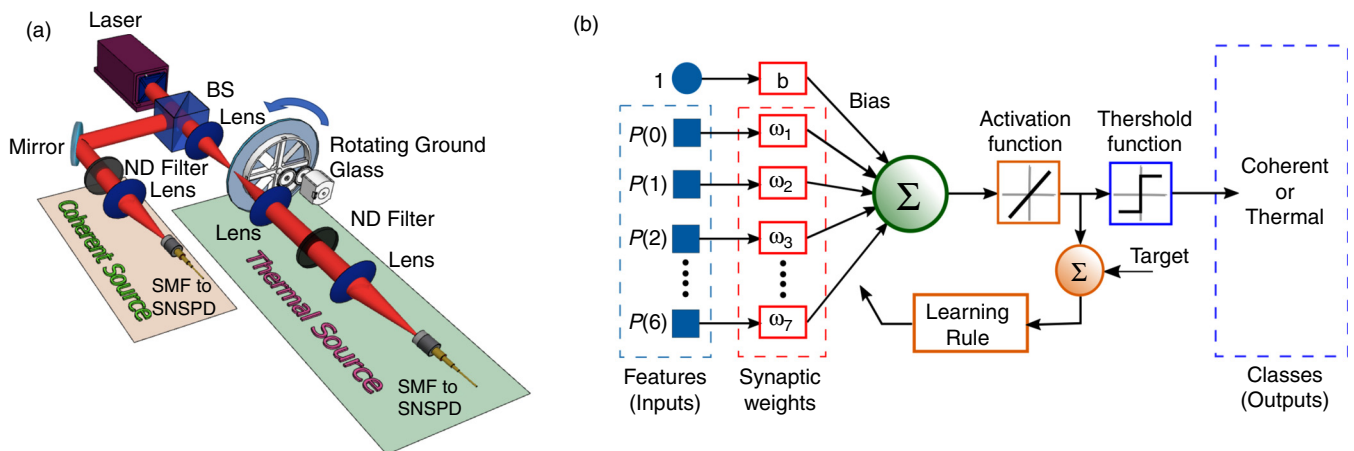

FIG. 1. (a) Schematic representation of the experimental setup. A laser beam is divided by a beam splitter (BS); the two replicas of the beam are used to generate light with Poissonian (coherent) and super-Poissonian (thermal) statistics. The thermal beam of light is generated by a rotating ground glass. Neutral density (ND) filters are utilized to attenuate light to the single-photon level. Coherent and thermal light beams are measured by superconducting nanowire single-photon detectors (SNSPDs). (b) Flow diagram of the ADALINE neuron used for demonstration of light source identification. Additional details are discussed in the body of the article.

b represents a bias term. During the training period, these parameters are optimized through the learning rule by using the error between the target output and neuron's output as reference. For binary classification (coherent or thermal), the neuron's output is fed into the identity activation function and, subsequently, into the threshold function.

To train the ADALINE, we make use of the so-called delta learning rule,[37] in combination with a database of experimentally measured photon-number distributions, considering different mean photon numbers: $\bar{n} = 0.44, 0.53, 0.67, 0.77$. The database for each mean photon number was divided into subsets comprising $10, 20, \ldots, 150, 160$ data points. The ADALINE neurons are thus prepared by using one hundred thousands of those subsets, where 70% are devoted to training and 30% to testing. In all cases, the training was stopped after 50 epochs.

We have established the baseline performance for our ADALINE neuron by using a naive Bayes classifier. This is a simple classifier based on Bayes's theorem.[38] Throughout this article, we assume that each measurement is independent. Moreover, we represent the measurement of the photon number sequence as a vector $\mathbf{x} = (x_1, \ldots, x_k)$. Then, the probability of this sequence generated from coherent or thermal light is given by $p(C_j|x_1, \ldots, x_k)$, where C_j could denote either coherent or thermal light. Using Bayes's theorem, the conditional probability can be decomposed as $p(C_j|\mathbf{x}) = p(C_j)p(\mathbf{x}|C_j)/p(\mathbf{x})$. By using the chain rule for conditional probability, we have $p(C_k|x_1, \ldots, x_k) = p(C_j) \prod_{i=1}^{k} p(x_i|C_j)$. Since our light source is either coherent or thermal, we assume $p(C_j) = 0.5$. Thus, it is easy to construct a naive Bayes classifier, where one picks the hypothesis with the highest conditional probability $p(C_j|\mathbf{x})$. We used theoretically generated photon-number probability distributions as the prior probability $p(x_i|C_j)$ and used the experimental data as the test data.

III. RESULTS

In Fig. 2, we compare the histograms for the theoretical and experimental photon number distributions for different mean photon numbers $\bar{n} = 0.40, 0.53, 0.67$, and 0.77. The bar plots are generated by experimental data with one million measurements for each source; the curves in each of the panels represent the expected theoretical photon number distributions for the corresponding mean photon numbers. Figure 2 shows excellent agreement between theory and experiment, which demonstrates the accuracy of our surjective photon counting method. Furthermore, from Fig. 2(a)–2(d), we can also observe the effect of the mean photon number on the photon number probability distributions. As shown in Fig. 2(a), it is evident that millions of measurements enable one to discriminate light sources. On the other hand, Fig. 2(d) shows a situation in which the source mean-photon number is low. In this case, the discrimination of light sources becomes cumbersome, even with millions of measurements. In Fig. 3, we illustrate the difficulty of using limited sets of data to discriminate light sources at a mean photon number of $\bar{n} = 0.77$. The histograms in this figure are generated with data points that represent 10, 20, 50, 100, and 100 000 realizations of the experiment. Thus, each data point is equivalent to a photon number-resolving measurement of the light source. As shown in Fig. 3, the photon number distributions obtained with a limited number of measurements do not resemble those in the histograms shown in Fig. 2(a), for neither coherent nor thermal light beams.

To evaluate the performance of the ADALINE and naive Bayes classifiers, we calculate the accuracy P_{acc}, which is defined as the ratio of number of correct predictions C_P to the total number of input
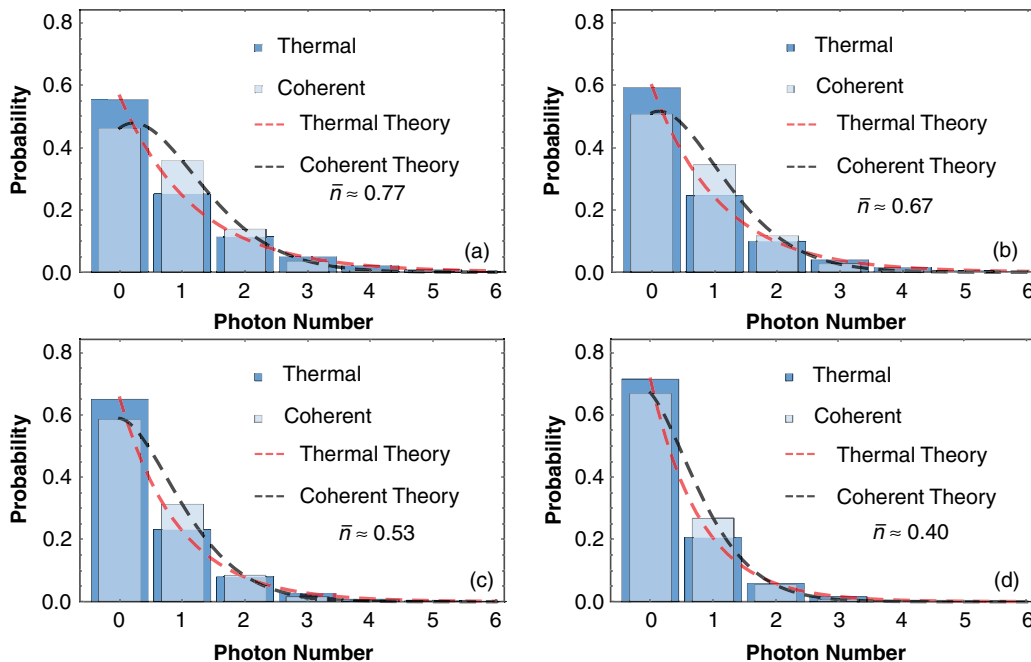

FIG. 2. A set of histograms displaying theoretical and experimental photon number probability distributions for coherent and thermal light beams with different mean photon numbers. Our experimental results are in excellent agreement with theory. The photon number distributions illustrate the difficulty in discriminating light sources at low-light levels even when large sets of data are available.
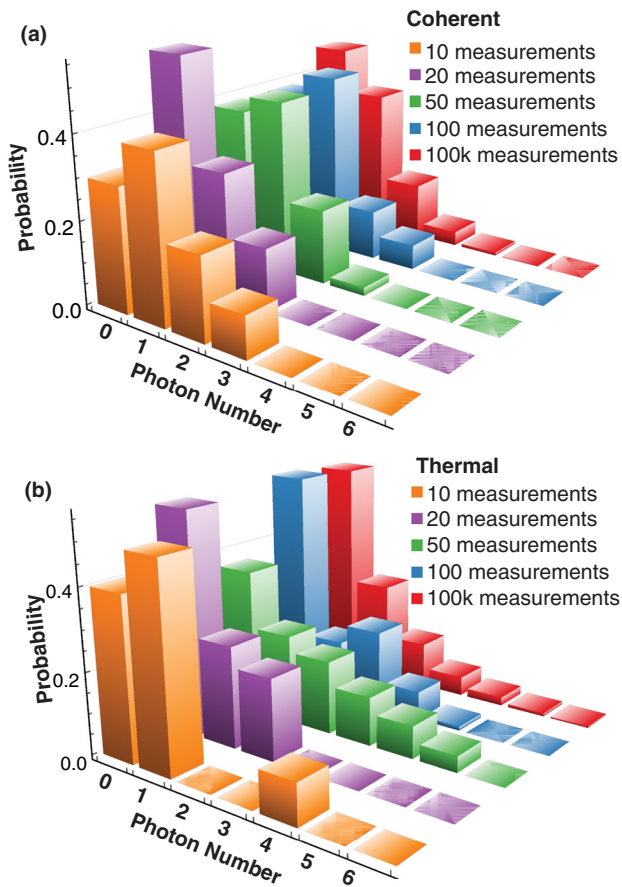
FIG. 3. Probability distributions of coherent and thermal light, for varying dataset sizes (10, 20, 50, 100, 100k). Data used here is randomly selected from the measurement presented in Fig. 2(a).

FIG. 4. Overall accuracy of light discrimination vs the number of data points used in a naive Bayes classifier. The curves represent the accuracy of light discrimination for $\bar{n} = 0.40$ (red line), $\bar{n} = 0.53$ (blue line), $\bar{n} = 0.67$ (green line), and $\bar{n} = 0.77$ (orange line). The error bars are generated by dividing the test dataset into ten subsets.

samples T_N, that is, $P_{\mathrm{acc}}(\%) = C_P/T_N \times 100$. Note that the accuracy is computed from the test dataset. These datasets are unseen by the algorithm in the training stage. Here, the number of input samples belonging to each class is equal. Since the accuracy only quantifies the successful events from a balanced test dataset, this is equivalent to the probability of discrimination. Therefore, we can also measure the probability of misclassification defining the discrimination probability error P_e as $P_e = 1 - P_{\mathrm{acc}}$.

In Fig. 4, we show the overall accuracy for light discrimination using a naive Bayes classifier. The accuracy increases with the number of data points. For example, when $\bar{n} = 0.40$, the accuracy of discrimination increases from approximately 61% to 90% as we increase the number of data points from 10 to 160. It is worth noting that even with a small increase in the number of measurements, the naive Bayes classifier starts capturing the characteristic features of different light sources, given by distinct sequences of photon number events. This is obvious since larger sets of data contain more information pertaining to the probability distribution. Furthermore, the mean photon number of the light field significantly changes the discrimination accuracy profile. As the mean photon number increases, the overall accuracy converges faster toward 100%, as expected. This is due to the fact that the
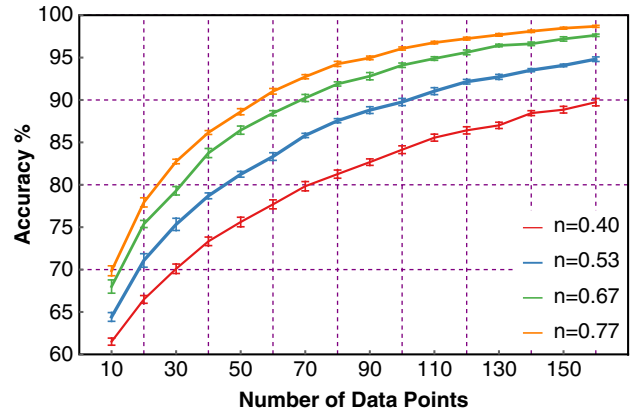
photon number probability distributions become more distinct at a higher mean photon number.

The overall accuracy of light-source discrimination with respect to the number of data points using ADALINE is shown in Fig. 5. In this case, the information provided by only 10 data points leads to an average accuracy of around 63% for $\bar{n} = 0.40$; whereas for 160 data points, the accuracy is greater than 90%. The comparison between Figs. 4 and 5 reveals that ADALINE and the naive Bayes classifier exhibit similar accuracy levels. As one might expect, in both cases, the accuracy increases with the number of data points and mean photon numbers. However, ADALINE requires far less computational resources than a naive Bayes classifier. Indeed, the execution time of ADALINE is one order of magnitude smaller than that of the naive Bayes classifier. The identification was performed using a computer with an Intel Core i7–4710MQ CPU (@2.50 GHz) and 32 GB of RAM
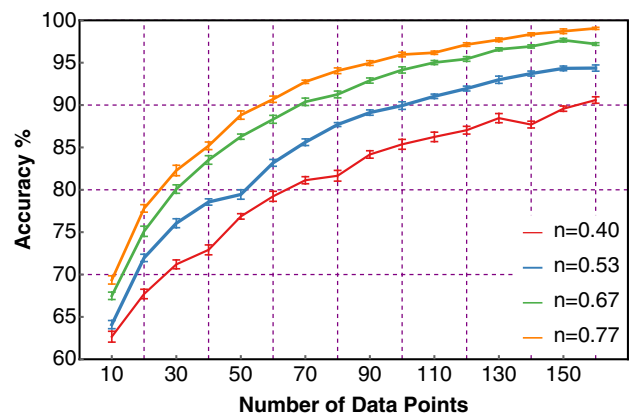

FIG. 5. Overall accuracy of light discrimination vs the number of data points used in ADALINE. The curves represent the accuracy of light discrimination for $\bar{n} = 0.40$ (red line), $\bar{n} = 0.53$ (blue line), $\bar{n} = 0.67$ (green line), and $\bar{n} = 0.77$ (orange line). The error bars are generated by dividing the test dataset into ten subsets.
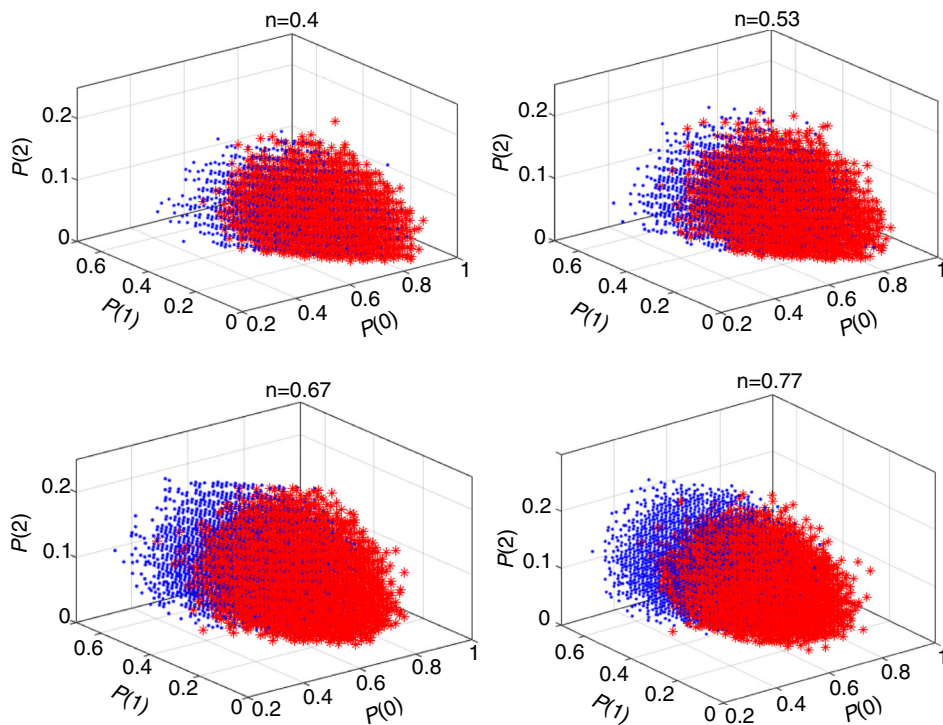
FIG. 6. Projection of the feature space on the plane $[P(0), P(1), P(2)]$ for different mean photon numbers: (a) $\bar{n} = 0.4$, (b) $\bar{n} = 0.53$, (c) $\bar{n} = 0.67$, and (d) $\bar{n} = 0.77$. The blue points correspond to photon statistics of coherent light, whereas the red stars describe photon statistics of thermal light. The number of points and stars is 50 000 for all cases. Moreover, each point or star is obtained from a probability distribution generated with 60 data points.
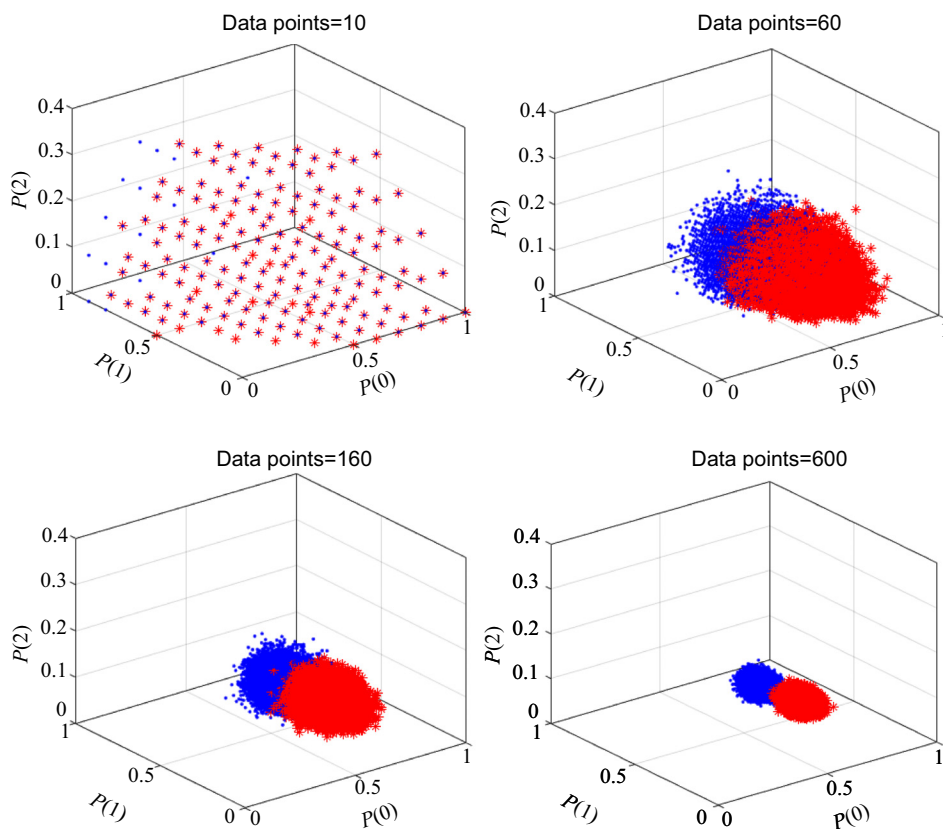

FIG. 7. Projection of the feature space on the plane $(P(0), P(1), P(2))$ for different number of data points used to generate a probability distribution: (a) 10, (b) 60, (c) 160, and (d) 600. The blue points correspond to photon statistics of coherent light, whereas the red stars describe photon statistics of thermal light. In both cases, the size of the dataset is 50 000. In addition, the mean photon number is set to $\bar{n} = 0.77$.

with MATLAB 2019a. However, the convergence rate for the naive Bayes is slightly higher than that observed for the ADALINE classifier. This implies that at low mean photon numbers ADALINE outperforms the naive Bayes classifier in the sense that the former requires less computational resources than the latter.

To understand why a single ADALINE neuron is enough for light discrimination, we first realize that ADALINE is a linear classifier. Therefore, the decision surface is expressed by a seven-dimensional hyperplane. This is defined by the seven features described by $P(n)$ (with $n = 0, 1, \ldots, 6$). In our case, the range of $[0,6]$ is selected based on the extremely low probability of observing a seven-photon event when the number of data points is 100. Interestingly, one can find that the datasets at the space of probability-distribution values are linearly separable. This can be seen from Fig. 6, where we plot the projection of the feature space on a three-dimensional sub-space defined by $[P(0), P(1), P(2)]$ for different mean photon numbers. In all cases, each point or star is obtained from a probability distribution generated with 60 data points. Within this subspace, the photon statistics for thermal (red stars) and coherent (blue points) light sources show a separation from each other that increases with $\bar{n}$. This effect is more evident when the number of data points used to generate one photon probability distribution is increased, and the mean photon number remains fixed at $\bar{n} = 0.77$ (see Fig. 7). Evidently, the fact that both thermal and coherent light form two linearly separated classes makes ADALINE a good classifier for identification of coherent and thermal light sources.

IV. CONCLUSION

For more than 20 years, there has been an enormous interest in reducing the number of photons and measurements required to perform imaging, remote sensing, and metrology at extremely low light levels.[12,17] In this regard, photonic technologies operating at low-photon levels utilize weak photon signals that make them vulnerable against detection of environmental photons emitted from natural sources of light. Indeed, this limitation has made unfeasible the realistic implementation of this family of technologies.[6,10,13] So far, this vulnerability has been tackled through conventional approaches that rely

on the measurement of coherence functions, p-values, the implementation of thresholding, and quantum state tomography.[6,10,13,44] Unfortunately, these approaches to characterizing photon-fluctuations rely on the acquisition of large number of measurements that impose constraints on the identification of light sources. Here, for the first time, we have demonstrated a smart protocol for discrimination of light sources at mean photon numbers below one. Our work demonstrates a dramatic improvement in both the number of photons and measurements required to identify light sources.[6,10,13,44] Furthermore, our results indicate that a single artificial neuron outperforms the naive Bayes classifier at low light levels. Interestingly, this neuron has simple analytical and computational properties that enable low-complexity and low-cost implementations of our technique. We are certain that our work has important implications for multiple photonic technologies, such as LiDAR and microscopy of biological materials.

APPENDIX A: IMPLEMENTATION DETAILS

For sake of completeness, we provide additional details of our calculations as well as comparisons among different methods for light identification.
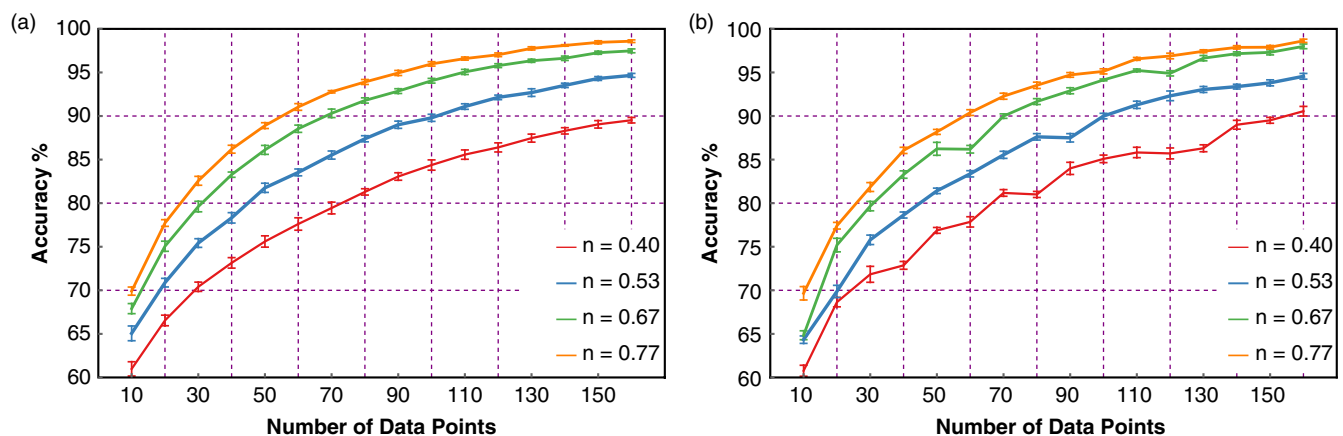

FIG. 8. Overall accuracy of light discrimination vs the number of data points used in (a) naive Bayes classifier and (b) ADALINE. The training stages are performed with ideal theoretical data. Furthermore, the test stages are carried out using experimental acquired data. The curves represent the accuracy of light discrimination for $\bar{n} = 0.40$ (red line), $\bar{n} = 0.53$ (blue line), $\bar{n} = 0.67$ (green line), and $\bar{n} = 0.77$ (orange line). The error bars are generated by dividing the test dataset into ten subsets.
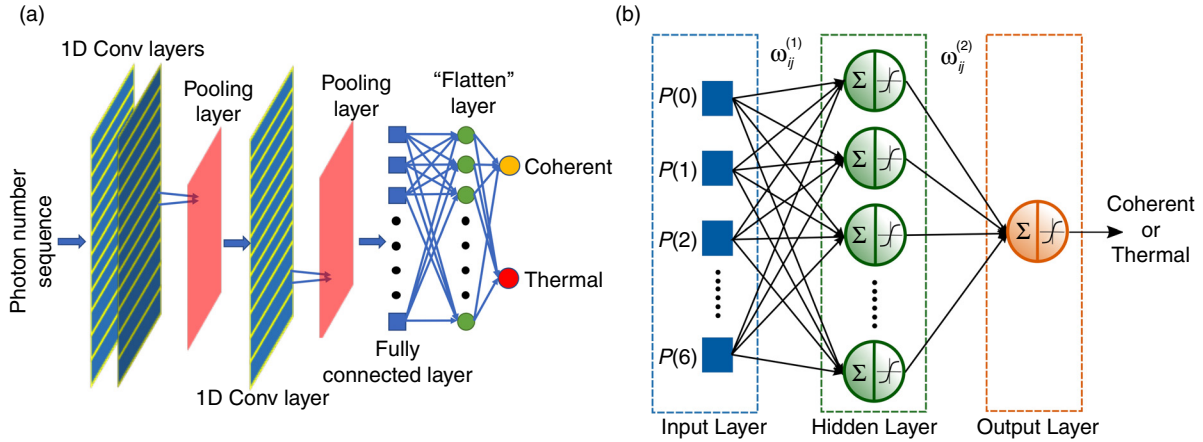
FIG. 9. Schematic representations: (a) one-dimensional convolutional neural network and (b) multilayer neural network used for demonstration of light source identification.

Since the naive Bayes classifier and the ADALINE methods presented in Sec. II were trained on experimental data, it is important to study the role of imperfections and their impact on the performance of our scheme for light identification. In Fig. 8, we report the overall accuracy of our naive Bayes classifier and our single neuron trained with numerically generated data and tested with experimental data. Naturally, the training is performed with ideal coherent and thermal statistics. A comparison among Figs. 4, 5 and 8 indicates a good agreement between theory and experiment. These results demonstrate the robustness of our method to identify light sources.

In addition to naive Bayes and ADALINE, we evaluate two additional machine-learning algorithms, namely a one-dimensional convolutional neural network (1D CNN) and a multilayer neural network (MNN). Despite the fact that both algorithms are effective to identify light sources, they are analytically and computationally more sophisticated than the simple ADALINE model. Nevertheless,

their recognition rates do not present substantial differences. Figures 9(a) and 9(b) show the structure of the 1D-CNN and MNN, respectively.

A convolutional neural network is a deep learning algorithm that extracts automatically relevant features of the input.[45] Here, our one-dimensional convolutional neural network is composed of two 1D-convolutional layers that extract the low- and high-level features of the input. Outcomes from these two layers are subsequently fed into a convolutional layer sandwiched between two max-pooling layers. The pooling layers downsample the input representation, and therefore its dimensionality, leading to a computational simplification by removing redundant and unnecessary information. The activation function, implemented in all layers, is the rectified linear unit function (ReLU). Finally, a fully connected and a flattening layer precedes the output layer consisting of two softmax functions, whose outputs are the probability distributions over labels.
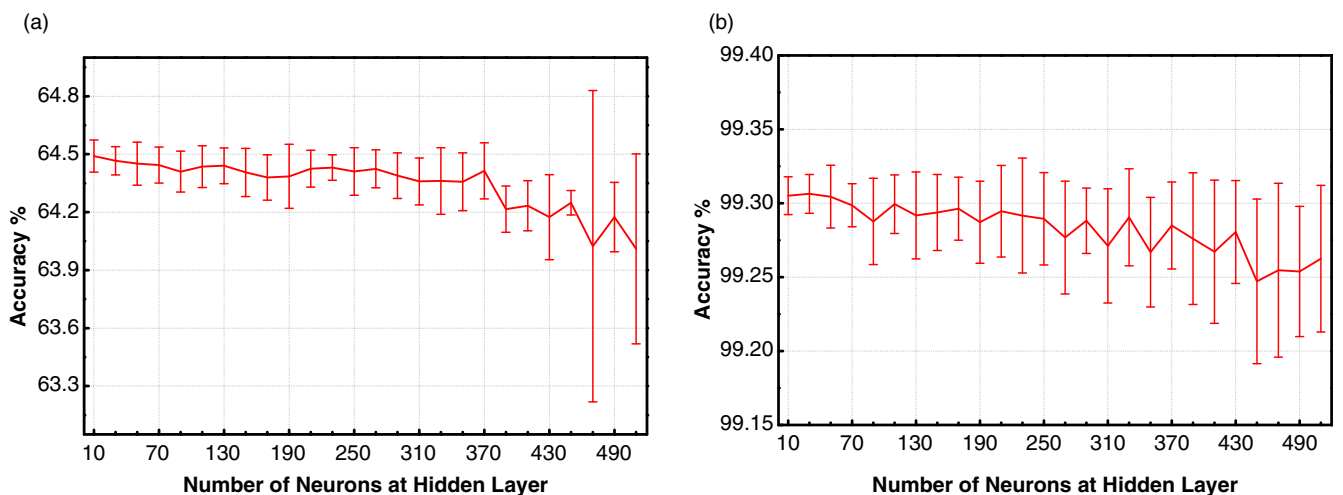

FIG. 10. Overall accuracy of light discrimination vs the number of neurons in the hidden layer of the MNN for two different mean photon numbers: (a) $\bar{n} = 0.4$ and (b) $\bar{n} = 0.77$. The error bars represent the standard deviation of the training stages.
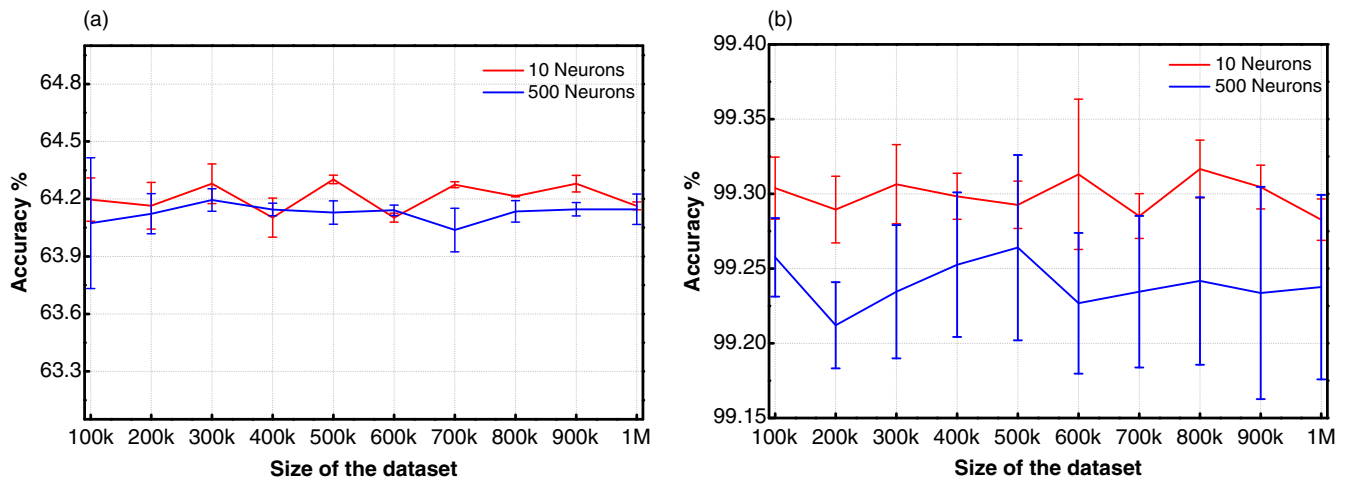
FIG. 11. Overall accuracy of light discrimination vs the size of the dataset of the MNN for two different mean photon numbers: (a) $\bar{n} = 0.40$ and (b) $\bar{n} = 0.77$. Red line and blue line correspond to neural networks with 10 and 500 neurons, respectively. The error bars represent the standard deviation of the training stages.

On the other hand, the multilayer neural network belongs to a classical machine learning algorithm, where the feature vector should be manually determined.[46] In our case, this vector is given by the probabilities of the photon number distribution, $P(n)$. As depicted in Fig. 9(b), the model corresponds to a two-layer feed-forward network: the hidden layer contains ten sigmoid neurons and the output layer consists of a softmax function. To determine a suitable neuron number in the hidden layer of the MNN, we trained different MNNs by changing the neuron number in the hidden layer and followed the accuracy values for each net. Figures 10(a) and 10(b) show the overall accuracy for light discrimination vs the number of neurons in the hidden layer for different mean photon numbers, $\bar{n} = 0.4$ and 0.77, respectively. Note that in both cases, the accuracy becomes lower as the number of neurons increases. This is because many neurons lead to over-parameterization,

causing poor generalization of the test-stage data. Additionally, as the number of neurons increases, the training becomes computationally more intensive. Figures 11(a) and 11(b) show the overall accuracy for light discrimination vs the size of the dataset for different mean photon numbers, $\bar{n} = 0.40$ and $\bar{n} = 0.77$, respectively. Note that in both cases, the accuracy exhibits similar performances as the size of the dataset increases, irrespective of the number of neurons in the hidden layer. Importantly, for large size datasets, additional computational time is required. Our results indicate that comparable accuracy can be achieved with smaller sized datasets. All the MNNs were trained by using the scaled conjugate gradient backpropagation method, where the cross-entropy was employed as the cost function. Since the output of sigmoid neurons is ranged in the interval [0,1], the cross-entropy function is ideal for the classification task. The network training was stopped after 200 epochs.

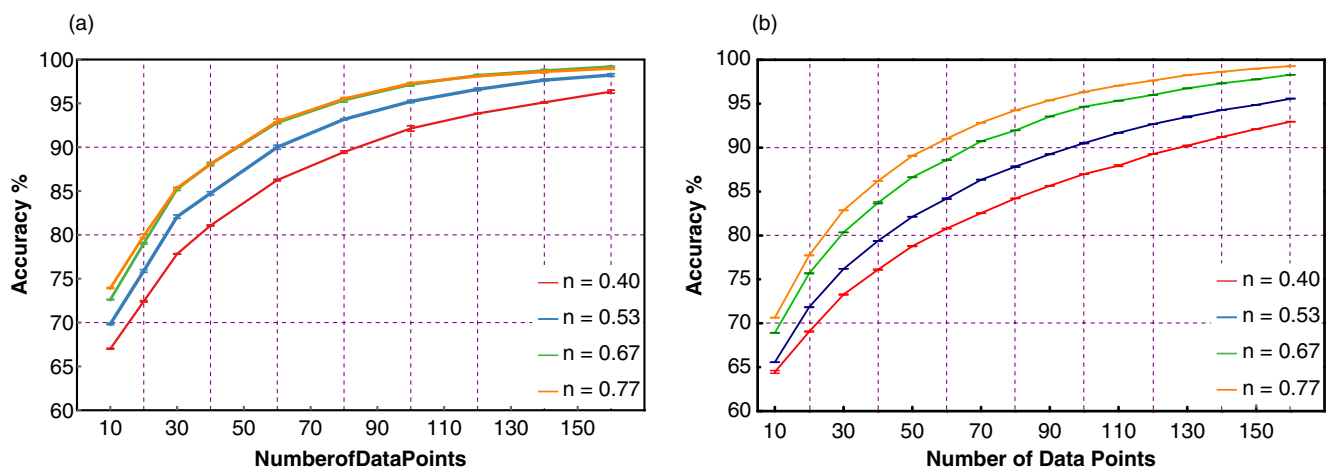**FIG. 12.** Overall accuracy of light discrimination vs the number of data points used in (a) 1D-CNN and (b) MNN. The curves represent the accuracy of light discrimination for $\bar{n} = 0.40$ (red line), $\bar{n} = 0.53$ (blue line), $\bar{n} = 0.67$ (green line), and $\bar{n} = 0.77$ (orange line). The error bars represent the standard deviation of the training epochs for 1D-CNN and training stages for MNN.

1D-CNNs and MNNs were trained with the same training set described in the main manuscript. Despite the fact that deep neural networks should be trained with a larger amount of data, we use 70% of the dataset for the training and the rest for testing both networks. Note that the same procedure was used for the ADALINE model. Figures 12(a) and 12(b) show the overall light-discrimination accuracy for increasingly larger number of data points for (a) 1D-CNNs and (b) MNNs. In both cases, the accuracy increases with the number of data points, because larger sets of data contain more information about the probability distribution. Interestingly, the accuracy of 1D-CNNs for $\bar{n} = 0.67$ and $\bar{n} = 0.77$ are almost the same; this indicates that in the low mean photon-number regime, the peak performance for 1D-CNN saturates much faster than the MNN classifier. As one might expect, this fast accuracy convergence carries the cost of a much more complex computation as compared to the one needed for the MNN classifier.

APPENDIX B: COMPARISON WITH $g^{(2)}$ AND P-VALUE CLASSIFICATION

Additionally, we have calculated the degree of second-order correlation functions for both sources. In this case, the correlation function $g^{(2)}(\tau)$ was measured using the definition $g^{(2)}(\tau) = 1 + (\langle (\Delta \hat{n})^2 \rangle - \langle \hat{n} \rangle) / \langle \hat{n} \rangle^2$, where $\langle \cdots \rangle$ denotes statistical average of the input dataset. The second-order correlation functions at $\tau = 0$ for different sizes of datasets and mean photon numbers are presented in Figs. 13(a) and 13(b). In both cases, the $g^{(2)}(0)$ calculation presents large standard deviations due to the limited data points. These large variations impose important difficulties in the identification of light sources. In addition, these large variations make it hard to justify a good $g^{(2)}(0)$ to identify coherent and thermal sources. To provide visual evidence of the last statement, we compare the $g^{(2)}(0)$ results with the statistical fluctuations that characterize coherent and thermal sources. Inspired by the level of significance commonly used in statistical hypothesis testing, we established tolerance bands of 5% around the theoretical values of $g^{(2)}(0)$. The overall accuracy obtained for both sources is reported in Fig. 13(c). Notably, the accuracy is not greater

than 25%, even when the mean photon number is $\bar{n} = 0.77$. Remarkably, the ADALINE neuron reaches an accuracy of about 95% for the same mean photon number.

Moreover, the so-called p-value is a measure of the probability of an observed value, assuming that the null hypothesis is true. The null hypothesis is rejected, if the p-value is less than or equal to the level of significance α, which by convention is set to $\alpha = 0.05$.[39] Then, inspired by the previous results, we take our estimator for the p-value to be the variance of the photon probability distribution. Note that the variance for a coherent state is given by the mean photon number $\bar{n}$; whereas the variance for a thermal distribution is given by $\bar{n} + \bar{n}^2$. This difference in the variance makes it a perfect candidate for an estimator of the p-value. For this purpose, we set the null hypothesis H_0 to be a Poissonian distribution and the alternative hypothesis H_1 to be a thermal distribution. Thus, the p-value is defined by $p_{\text{value}} = \Pr(s^2(X) \le p_\alpha | H_0)$. Here, p_α is the critical bound which limits the decision region; furthermore, $s^2(X)$ corresponds to the sample variance with X representing a random variable for the observed data. It is worth mentioning that the variances $s^2(X)$ follow a normal distribution. Therefore, p_α can be calculated as $\Pr(s^2(X) \ge p_\alpha | H_0) = \alpha$. Assuming that $\alpha = 0.05$, then $p_\alpha = 1.64\sigma + \bar{n}$, the standard deviation of the variance distribution is described by σ.

Figures 14(a) and (b) show the p-value for different dataset sizes for sources with various mean photon numbers. Note that p-values shown in Fig. 14(a) for a coherent source are greater than the level of significance for any mean photon number. Consequently, the null hypothesis H_0 must be accepted, and therefore, the data are produced by a source with a Poissonian distribution. For the case of a thermal source, [Fig. 14(b)], we expected to find p-values less than $\alpha = 0.05$ to reject the null hypothesis and to accept an alternative hypothesis. However, it is worth noticing that for any mean photon number, the p-values are beyond the level of significance for datasets with sizes smaller than 150. Thus, we cannot reject H_0, and one concludes that the data were produced by a source with a Poissonian distribution, which is clearly wrong. Finally, Fig. 14(c) shows the accuracy for different mean photon numbers using the p-value as a measurement to discriminate light sources. For the lowest mean photon number, the accuracy shows high performance for data sizes beyond 500 whereas, for $\bar{n} = 0.77$, the accuracy reaches
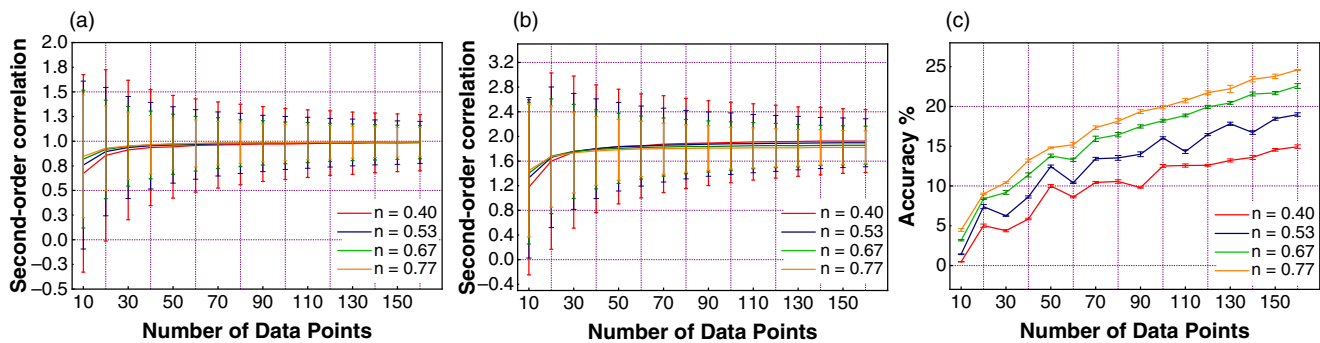

FIG. 13. (a) Second-order correlation function $g^{(2)}$ for coherent source. (b) Second-order correlation function $g^{(2)}$ for thermal source. (c) Overall accuracy of light discrimination vs the number of data points using $g^{(2)}$ estimator. The curves represent the accuracy of light discrimination for $\bar{n} = 0.40$ (red line), $\bar{n} = 0.53$ (blue line), $\bar{n} = 0.67$ (green line), and $\bar{n} = 0.77$ (orange line).

FIG. 14. (a) *p*-value for coherent source assuming that H_0 is true. (b) *p*-value for thermal source assuming that H_0 is true. (c) Overall accuracy of light discrimination vs the number of data points using the *p*-value and a level of significance of 5%. The curves represent the accuracy of light discrimination for $\bar{n} = 0.40$ (red line), $\bar{n} = 0.53$ (blue line), $\bar{n} = 0.67$ (green line), and $\bar{n} = 0.77$ (orange line).
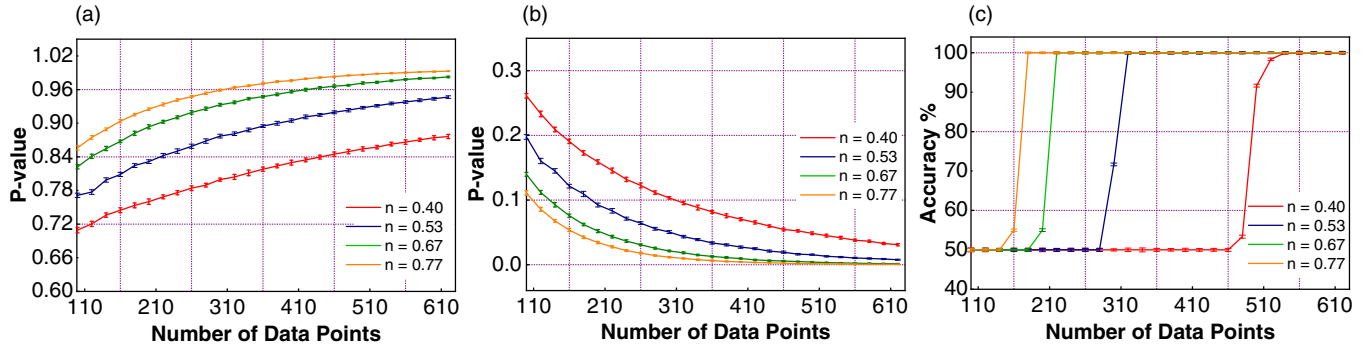
the best performance with 180 data points. In any case, the ADALINE and naive Bayes classifiers offer a better performance when compared to the calculated *p*-value with fewer data points.

APPENDIX C: COMPARISON WITH SIGNAL-DISCRIMINATION THEORETICAL BOUNDS

Finally, since our scheme for light identification is inherently related to signal discrimination, it is important to compare our results with the theoretical bounds predicted by the Helstrom[40,41]

and the Chernoff bounds.[42,43] If two quantum states are described by the density matrices ρ and σ, then Helstrom bound is given by $p_{\mathrm{H}} = \frac{1}{2}[1 + D_{\mathrm{Tr}}(\rho, \sigma)]$, where $D_{\mathrm{Tr}}(\rho, \sigma) = \frac{1}{2}\mathrm{Tr}[\rho - \sigma]$ is the trace distance. Thus, the Helstrom bound indicates the lower bound on the error probability $P_e = 1 - P_{\mathrm{H}}$ in a single realization of the experiment. In the limit of repeating the experiment n times, the error probability P_e will decreases exponentially given by $P_{e,n} \sim \exp(-n\xi_{\mathrm{QCB}})$, where $\xi_{\mathrm{QCB}} = -\log\min_{0 \le s \le 1}\mathrm{Tr}(\rho^s\sigma^{1-s})$ is the quantum Chernoff bound. As shown in Fig. 15, indeed, in the limit of a single data point, our accuracy is lower than that dictated
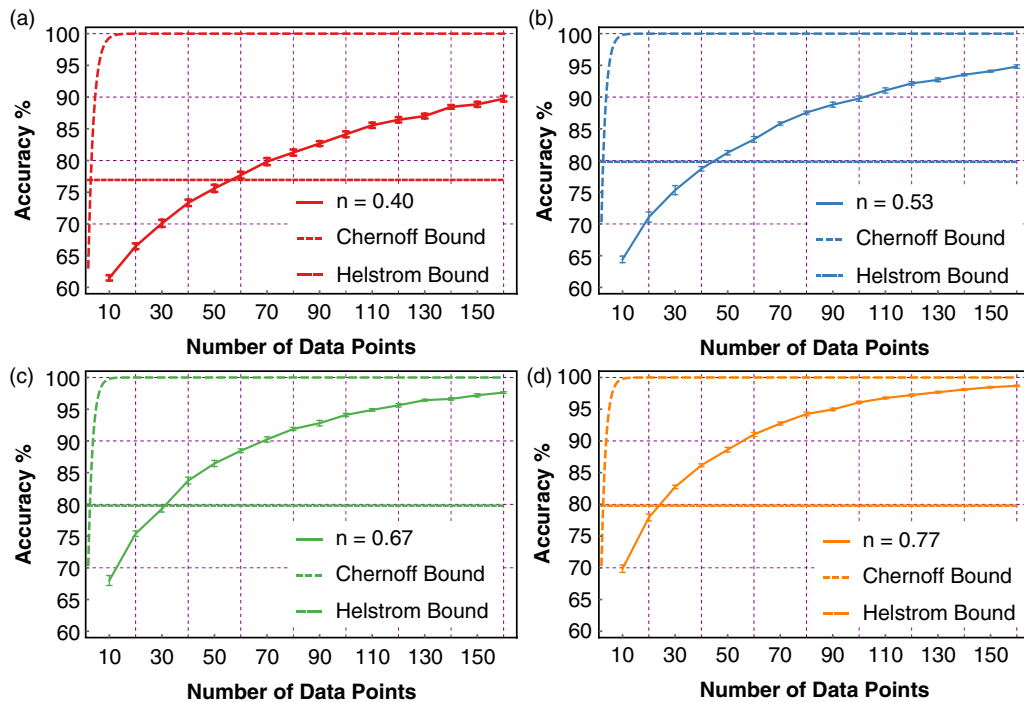

FIG. 15. Overall accuracy of a naive Bayes classifier plotted with the accuracy predicted by the Helstrom and Chernoff bounds. The plots represent the accuracy of light discrimination for (a) $\bar{n} = 0.40$, (b) $\bar{n} = 0.53$, (c) $\bar{n} = 0.67$, and (d) $\bar{n} = 0.77$. The dashed lines represent the Helstrom bound and the dotted lines represent the Chernoff bound.

by the Helstrom bound. However, in the limit of increasing the number of data points, our accuracy approaches the Chernoff bound quickly. These results validate the performance and accuracy of our method for light identification.

REFERENCES

[1] R. J. Glauber, "The quantum theory of optical coherence," Phys. Rev. **130**, 2529 (1963).

[2] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*, Cambridge University Press (1995).

[3] L. Mandel, "Sub-Poissonian photon statistics in resonance fluorescence," Opt. Lett. **4**, 205–207 (1979).

[4] L. Mandel and E. Wolf, "Coherence properties of optical fields," Rev. Mod. Phys. **37**, 231 (1965).

[5] J. Liu and Y. Shih, "Nth-order coherence of thermal light," Phys. Rev. A **79**, 023819 (2009).

[6] J. Hloušek, M. Dudka, I. Straka, and M. Ježek, "Accurate detection of arbitrary photon statistics," Phys. Rev. Lett. **123**, 153604 (2019).

[7] L. Dovrat, M. Bakstein, D. Istrati, A. Shaham, and H. S. Eisenberg, "Measurements of the dependence of the photon-number distribution on the number of modes in parametric down-conversion," Opt. Express **20**, 2266–2276 (2012).

[8] L. Dovrat, M. Bakstein, D. Istrati, E. Megidish, A. Halevy, L. Cohen, and H. S. Eisenberg, "Direct observation of the degree of correlations using photon-number-resolving detectors," Phys. Rev. A **87**, 053813 (2013).

[9] G. Zambra, A. Andreoni, M. Bondani, M. Gramegna, M. Genovese, G. Brida, A. Rossi, and M. G. A. Paris, "Experimental reconstruction of photon statistics without photon counting," Phys. Rev. Lett. **95**, 063602 (2005).

[10] L. A. Howard, G. G. Gillett, M. E. Pearce, R. A. Abrahao, T. J. Weinhold, P. Kok, and A. G. White, "Optimal imaging of remote bodies using quantum detectors," Phys. Rev. Lett. **123**, 143604 (2019).

[11] A. Ling, A. Lamas-Linares, and C. Kurtsiefer, "Accuracy of minimal and optimal qubit tomography for finite-length experiments," arXiv preprint arXiv:0807.0991 (2008).

[12] J. P. Dowling and K. P. Seshadreesan, "Quantum optical technologies for metrology, sensing, and imaging," J. Light. Technol. **33**, 2359 (2015).

[13] Y. Sher, L. Cohen, D. Istrati, and H. S. Eisenberg, "Low intensity LiDAR using compressed sensing and a photon number resolving detector," Emerging Digital Micromirror Device Based Syst. Appl. X **10546**, 105460J (2018).

[14] Q. Wang, L. Hao, Y. Zhang, C. Yang, X. Yang, L. Xu, and Y. Zhao, "Optimal detection strategy for super-resolving quantum LiDAR," J. Appl. Phys. **119**, 023109 (2016).

[15] J. P. Dowling, "Quantum optical metrology – the lowdown on High-N00N states," Contemp. Phys **49**, 125–143 (2008).

[16] O. S. Magaña-Loaiza, R. J. León-Montiel, A. Perez-Leija, A. B. U'Ren, C. You, K. Busch, A. E. Lita, S. W. Nam, R. P. Mirin, and T. Gerrits, "Multiphoton quantum-state engineering using conditional measurements," Npj Quantum Inf. **5**, 80 (2019).

[17] O. S. Magaña-Loaiza and R. W. Boyd, "Quantum imaging and information," Rep. Prog. Phys. **82**, 124401 (2019).

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).

[19] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, "Machine learning and the physical sciences," Rev. Mod. Phys. **91**, 045002 (2019).

[20] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," Nature **549**, 195–202 (2017).

[21] V. Dunjko, J. M. Taylor, and H. J. Briegel, "Quantum-enhanced machine learning," Phys. Rev. Lett. **117**, 130501 (2016).

[22] A. Hentschel and B. C. Sanders, "Machine learning for precise quantum measurement," Phys. Rev. Lett. **104**, 063603 (2010).

[23] A. Lumino, E. Polino, A. S. Rab, G. Milani, N. Spagnolo, N. Wiebe, and F. Sciarrino, "Experimental phase estimation enhanced by machine learning," Phys. Rev. Appl. **10**, 044033 (2018).

[24] A. A. Melnikov, H. P. Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, "Active learning machine learns to create new quantum experiments," PNAS **115**, 1221–1226 (2018).

[25] C. L. Cortes, S. Adhikari, X. Ma, and S. K. Gray, "Accelerating quantum optics experiments with statistical learning," arXiv preprint arXiv:1911.05935 (2019).

[26] Z. A. Kudyshev, S. Bogdanov, T. Isacsson, A. V. Kildishev, A. Boltasseva, and V. M. Shalaev, "Rapid classification of quantum sources enabled by machine learning," arXiv preprint arXiv:1908.08577 (2019).

[27] S. Lohani and R. T. Glasser, "Turbulence correction with artificial neural networks," Opt. Lett. **43**, 2611–2614 (2018).

[28] J. Gao, L.-F. Qiao, Z.-Q. Jiao, Y.-C. Ma, C.-Q. Hu, R.-J. Ren, A.-L. Yang, H. Tang, M.-H. Yung, and X.-M. Jin, "Experimental machine learning of quantum states," Phys. Rev. Lett. **120**, 240501 (2018).

[29] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, "Neural-network quantum state tomography," Nat. Phys. **14**, 447–450 (2018).

[30] F. Flamini, N. Spagnolo, and F. Sciarrino, "Visual assessment of multi-photon interference," Quantum Sci. Technol. **4**, 024008 (2019).

[31] I. Agresti, N. Viggianiello, F. Flamini, N. Spagnolo, A. Crespi, R. Osellame, N. Wiebe, and F. Sciarrino, "Pattern recognition techniques for Boson sampling validation," Phys. Rev. X **9**, 011013 (2017).

[32] M. Bentivegna, N. Spagnolo, C. Vitelli, D. Brod, A. Crespi, F. Flamini, R. Ramponi, P. Mataloni, R. Osellame, E. Galvão, and F. Sciarrino, "Bayesian approach to Boson sampling validation," Int. J. Quantum Inf. **12**, 1560028 (2014).

[33] S. M. H. Rafsanjani, M. Mirhosseini, O. S. Magaña-Loaiza, B. T. Gard, R. Birrittella, B. E. Koltenbah, C. G. Parazzoli, B. A. Capron, C. C. Gerry, J. P. Dowling, and R. W. Boyd, "Quantum-enhanced interferometry with weak thermal light," Optica **4**, 487–491 (2017).

[34] I. A. Burenkov, A. K. Sharma, T. Gerrits, G. Harder, T. J. Bartley, C. Silberhorn, E. A. Goldschmidt, and S. V. Polyakov, "Full statistical mode reconstruction of a light field via a photon-number-resolved measurement," Phys. Rev. A. **95**, 053806 (2017).

[35] N. Montaut, O. S. Magaña-Loaiza, T. J. Bartley, V. B. Verma, S. W. Nam, R. P. Mirin, C. Silberhorn, and T. Gerrits, "Compressive characterization of telecom photon pairs in the spatial and spectral degrees of freedom," Optica **5**, 1418 (2018).

[36] B. Windrow and M. E. Hoff, "Adaptive switching circuits," "Technical report no. 1553–1," Stanford University, Stanford-California, Stanford Electronics Laboratories (1960).

[37] S. I. Gallant, *Neural Network Learning and Expert Systems*, MIT Press (1993).

[38] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, Chapman and Hall/CRC (2013).

[39] J. B. Ramsey, *The Elements of Statistics: With Applications to Economics and the Social Sciences*, Cengage Learning (2001).

[40] C. H. Helstrom, *Quantum Detection and Estimation Theory*, Academic Press (1976).

[41] Z. Puchała, Ł. Pawela, and K. Życzkowski, "Distinguishability of generic quantum states," Phys. Rev. A **93**, 062112 (2016).

[42] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," Ann. Math. Stat. **23**, 493–507 (1952).

[43] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, "Discriminating states: The quantum Chernoff bound," Phys. Rev. Lett. **98**, 160501 (2007).

[44] L. Cohen, E. S. Matekole, Y. Sher, D. Istrati, H. S. Eisenberg, and J. P. Dowling, "Thresholded quantum LiDAR: Exploiting photon-number-resolving detection," Phys. Rev. Lett **123**, 203601 (2019).

[45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press (2016).

[46] C. Bishop, *Pattern Recognition and Machine Learning*, Springer (2006).