Journal of the Korean Physical Society (2021) 79:173–184
https://doi.org/10.1007/s40042-021-00180-5

Online ISSN 1976-8524
Print ISSN 0374-4884

ORIGINAL PAPER

Machine learning study of the deformed one-dimensional topological superconductor

Jae Hyuck Lee[1,2] · **Hyun Cheol Lee**[1]

Received: 8 March 2021 / Accepted: 29 March 2021 / Published online: 17 May 2021

Abstract
A one-dimensional p-wave topological superconductor deformed by a sine-square-deformation is studied in the framework of machine learning. A supervised learning algorithm is applied with a convolutional neural network to discern the existence of a Majorana zero mode, which is the hallmark of topological superconductivity. The machine learning algorithm learns features of the Majorana zero mode, and the neural network trained with the dataset from the link deformed case turns out to be the most effective.

Keywords Machine learning · Topological superconductors · Majorana fermions · Sine-square deformation

1 Introduction

Machine learning (ML) is playing an increasingly important role in various fields of physics [1–3]. ML itself is a subfield of artificial intelligence and aims at developing algorithms that can learn from data automatically.

The methods of ML can be broadly divided into three main categories: supervised learning, unsupervised learning, and reinforcement learning. A hybrid of the above three approaches is also possible, e.g., semi-supervised learning falls between supervised and unsupervised learning. In supervised learning, the input training data are paired to the corresponding target output (the label), and the trained algorithm makes prediction to certain test dataset or a new input. Classification and regression are typical problems addressed in supervised learning. In unsupervised learning, the training input data are not paired with the target output (unlabelled). As such, it aims at finding hidden structures (or patterns) of a given data. A typical problem treated in unsupervised learning is the clustering problem. In reinforcement learning, the algorithm seeks to find a set of actions that maximize (numerical) reward by essentially trial and error. Famous examples are the Deep Mind Go and its important applications to quantum computation.

Ass there are numerous types of ML algorithms, the applications are extremely diverse, and here, we mention just a few which are the most relevant to condensed matter physics: (1) the ML phases of matter [4], (2) the relation ML to the renormalization group [5], (3) the representation of quantum states using a neural network [6, 7] and (4) the relations among quantum entanglement, tensor network, and ML [8].

In this paper, we employ a supervised learning algorithm in the context of the ML phases of matter as studied in Ref. [4] and apply it to the deformed one-dimensional topological superconductor [9–12]. We focus on understanding how ML discerns the topological character of a physical system [13].

Topological materials possess properties that are robust against certain continuous deformations. These properties of topological materials can be characterized by topological invariants, such as the Chern number for the topological insulator with broken time reversal symmetry or the Z_2 index for the topological insulator with time reversal symmetry. These topological invariants defined above manifest themselves as gapless excitations living on boundaries. For example, the Chern number is identical to the number of edge states, and the Z_2 index can be identified with the Z_2 parity of the number of edge/surface states [9–11].

For the one-dimensional p-wave superconductor, such topological boundary states associated with particle-hole symmetry are the Majorana fermions of zero energy with

✉ Hyun Cheol Lee
hyunlee@sogang.ac.kr

[1] Department of Physics, Sogang University, Seoul 04107, South Korea

[2] Department of Physics and Astronomy, Seoul National University, Seoul 08826, South Korea

1/2 fractional fermion number [14] localized at the ends of the superconductor [12]. In other words, the superconductor is topologically nontrivial if the Majorana zero modes exist at the ends, and vice versa.

The authors have studied the influences of the sine-square deformation (SSD) [15–18] on a one-dimensional p-wave superconductor [19]. SSD was designed to suppress the finite size and the boundary effects so that it can simulate an infinite size or a periodic boundary condition for finite systems with a boundary. As such investigating the effects of SSD on the boundary states of topological materials is very interesting.

The goal of this paper is to understand how ML delineates the effects of SSD on the Majorana zero modes of one-dimensional topological superconductors, using a classification model, to analyze the associated topological phase. We emphasize that our focus is on understanding the operating mechanism of ML applied to a physical system rather than building a highly accurate prediction model. The main result of this paper is presented in Fig. 8.

We convert particular combinations of eigenvectors of the deformed Hamiltonian of the topological superconductor into image formats and implement a supervised learning algorithm via the convolutional neural network (CNN). The CNN was specifically designed with translational invariance and locality highlighted [1] so that it can classify images very efficiently. This feature has also been exploited in the classification of the phases of the matter [4, 20–22].

We have built a CNN structure which perceived the features of the Majorana zero mode (the spinor structure and the spatial profile), and the CNN trained with a dataset extracted from the link-deformed Hamiltonian (see below) turns out to be most effective in prediction.

This paper is organized as follows: we review the basic properties of the deformed one-dimensional p-wave superconductors in Sect. 2. Data preparation and the CNN structure are described in Sects. 3 and 4, respectively. The results are presented in Sect. 5, and we conclude this paper with discussions and summary in Sect. 6.

MATLAB is used as a computing platform [23].

2 Deformed one-dimensional p-wave topological superconductor

To set the stage, we review our work on the deformed one-dimensional, p-wave, topological superconductor in Ref. [19]. The second quantized lattice model of the spin-polarized, one-dimensional, p-wave superconductor in the Bogoliubov-de-Gennes formalism (BdG) is given by [10, 12]

$$\begin{aligned}\mathsf{H}_{\mathrm{BdG}} =& -t\sum_j \left(c_{j+1}^\dagger c_j + c_j^\dagger c_{j+1}\right)\\ &-\mu\sum_j c_j^\dagger c_j + \sum_j \left(\Delta\, c_{j+1}^\dagger c_j^\dagger + \Delta^*\, c_j c_{j+1}\right),\end{aligned} \tag{1}$$

where $t(t>0)$, μ, and Δ are the hopping amplitude, the chemical potential, and the pairing amplitude, respectively. For the sake of physical relevance, we assume $t > |\Delta|$. The p-wave pairing symmetry is reflected in the nearest-neighbor coupling of the pairing term. By introducing the Nambu spinor on lattice site j

$$\Psi_j = \begin{pmatrix} c_j \\ c_j^\dagger \end{pmatrix} \tag{2}$$

one can recast the BdG Hamiltonian in Eq.(1) as

$$\begin{aligned}\mathsf{H}_{\mathrm{BdG}} =& \sum_{j=1}^{N-1} \Psi_j^\dagger \begin{bmatrix} -t/2 & -\Delta/2 \\ \Delta^*/2 & +t/2 \end{bmatrix} \Psi_{j+1}\\ &+ \sum_{j=1}^{N-1} \Psi_{j+1}^\dagger \begin{bmatrix} -t/2 & \Delta/2 \\ -\Delta^*/2 & +t/2 \end{bmatrix} \Psi_j\\ &- \frac{\mu}{2}\sum_{j=1}^{N} \Psi_j^\dagger \sigma_3 \Psi_j,\end{aligned} \tag{3}$$

where N is the number of sites, and $\sigma_{1,2,3}$ are the Pauli matrices. An irrelevant constant term is dropped. If a periodic boundary condition is imposed on the lattice, the Hamiltonian $\mathsf{H}_{\mathrm{BdG}}$ can be diagonalized in momentum space:

$$\mathsf{H}_{\mathrm{BdG}} = \frac{1}{2}\sum_{k\in[-\pi,\pi]} \Psi_k^\dagger \begin{bmatrix} \xi_k & -2i\Delta\sin k \\ +2i\Delta^*\sin k & -\xi_k \end{bmatrix} \Psi_k, \tag{4}$$

where $\xi_k = -2t\cos k - \mu$. The energy spectrum is given by

$$E_\pm(k) = \pm\sqrt{\xi_k^2 + 4|\Delta|^2\sin^2 k}. \tag{5}$$

The energy spectrum is particle-hole symmetric, and the energy gap exists for $\mu \neq \pm 2t$. If $\mu = \pm 2t$, the energy gap closes at $k = 0, \pi$. For the gapful cases with a periodic boundary condition, there are no midgap states such as zero energy states (zero modes).

The Hamiltonian in Eq. (3) can be expressed in terms of the single-particle Hamiltonian as

$$\begin{aligned}\mathsf{h}_{\mathrm{BdG}} =& \sum_{j=1}^{N-1} |j\rangle\langle j+1| \otimes \begin{bmatrix} -t/2 & -\Delta/2 \\ \Delta^*/2 & +t/2 \end{bmatrix}\\ &+ \sum_{j=1}^{N-1} |j+1\rangle\langle j| \otimes \begin{bmatrix} -t/2 & \Delta/2 \\ -\Delta^*/2 & +t/2 \end{bmatrix} - \frac{\mu}{2}\sum_{j=1}^{N} |j\rangle\langle j| \otimes \sigma_3.\end{aligned} \tag{6}$$

From now on, we impose the open boundary condition and assume that Δ is real and positive.

Now, we introduce the SSD into our system. The SSD spatially modulates the energy scales of the Hamiltonian for one-dimensional systems with the following profile function:

$$f_j = \sin^2\left[\frac{\pi}{N}(j-\frac{1}{2})\right]. \tag{7}$$

The SSD can be applied to both the interaction defined on the lattice link (the first two terms of Eq. (6)) and the local site interaction (the last term of Eq. (6)). Because the characters of these two interactions are very different, we will consider two types of SSDs: (I) an SSD applied only to the link interaction and (II) an SSD applied to both the link and the local site interaction. The case of a SSD applied only to local site interaction is not considered, because it is similar to an ordinary system with a spatially varying potential energy.

Then, the single particle BdG Hamiltonian of type (I) is given by

$$\begin{aligned}\mathsf{h}_{\mathrm{SSD,link}} &= \sum_{j=1}^{N-1} f_{j+\frac{1}{2}}|j\rangle\langle j+1|\otimes\begin{bmatrix}-t/2 & -\Delta/2\\ \Delta/2 & +t/2\end{bmatrix}\\ &+\sum_{j=1}^{N-1} f_{j+\frac{1}{2}}|j+1\rangle\langle j|\otimes\begin{bmatrix}-t/2 & \Delta/2\\ -\Delta/2 & +t/2\end{bmatrix}-\frac{\mu}{2}\sum_{j=1}^{N}|j\rangle\langle j|\otimes\sigma_3,\end{aligned} \tag{8}$$

and the single-particle BdG Hamiltonian of type (II) is given by

$$\begin{aligned}\mathsf{h}_{\mathrm{SSD,\ all}} &= \sum_{j=1}^{N-1} f_{j+\frac{1}{2}}|j\rangle\langle j+1|\otimes\begin{bmatrix}-t/2 & -\Delta/2\\ \Delta/2 & +t/2\end{bmatrix}\\ &+\sum_{j=1}^{N-1} f_{j+\frac{1}{2}}|j+1\rangle\langle j|\otimes\begin{bmatrix}-t/2 & \Delta/2\\ -\Delta/2 & +t/2\end{bmatrix}-\frac{\mu}{2}\sum_{j=1}^{N} f_j|j\rangle\langle j|\otimes\sigma_3.\end{aligned} \tag{9}$$

The energy eigenvalues and the eigenvectors of the Hamiltonians Eqs. (6), (8), and (9) comprise the dataset of the ML study of this paper.

To understand the physics of the above Hamiltonians, it is instructive to consider the continuum limit of Eq. (4) near $k = 0$ (with negative μ) in the low-energy limit, where it reduces to the one-dimensional Dirac Hamiltonian:

$$\mathsf{h}_{\mathrm{continuum}} = \Delta\sigma_2(-i\frac{\partial}{\partial x}) - \mu_c\sigma_3, \tag{10}$$

where $\mu_c = \mu + 2t$. The topological nature of this Dirac Hamiltonian can be revealed by considering a spatially varying $\mu_c(x)$ of the kink (or domain wall) profile [14, 19]. Then, it can be readily shown that Eq. (10) has only one normalizable zero mode solution:

$$\psi_{\mathrm{zero}}(x) \sim \frac{1}{[\cosh(x-x_0)]^{\mu_{c0}/\Delta}}\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \tag{11}$$

which demonstrates that the zero mode is localized at the location of the kink. Furthermore, this zero mode is a Majorana fermion satisfying the Majorana condition

$$\psi^c_{\mathrm{zero}}(x) = \psi_{\mathrm{zero}}(x), \tag{12}$$

where ψ^c is the charge conjugated spinor given by $\psi^c = \sigma_1\psi^*$ in our convention of Dirac matrices. Also, this zero mode is well-known to carry 1/2 (fractional) fermion number [14]. We quickly add that a similar analysis can be done for the anti-kink, and in that case, the spinor part becomes $(1,-1)^{\mathrm{t}}$ (t denotes the matrix transpose).

For an open boundary condition, the exterior of the superconductor plays the role of topologically trivial state so that the boundaries (two ends of the wire) effectively become a kink and an anti-kink. In view of the discussions of the continuum model, the Majorana zero mode localized at the ends of the wire should exist if the superconductor is topologically nontrivial. Kitaev [12] showed, by the representing electron operator c_j as a (complex) sum of two Majorana fermions, which manifests the fractional nature of a Majorana zero mode, that a Majorana zero mode localized at the end exists for $-2 < \mu/t < 2$. (Hence, the superconductor is topological.).

3 Data preparation

The energy spectrum and the eigenvectors of the Hamiltonians in Eqs. (6), (8), and (9) have been obtained by direct diagonalization using MATHEMATICA, and these data are exported to MATLAB for the actual running of ML.

In the absence of a SSD, we find that the Majorana zero mode is localized at one end for the range $-2 < \mu/t < 2$ (see panel (a) of Figs. 1 and 2), confirming the result obtained by Kitaev [12]. However, for smaller values of Δ and N, the criterion for the zero mode becomes ambiguous (see Fig. 1). An SSD makes the demarcation between the zero and the non-zero modes even more challenging (see Fig. 3), which implies that the detection of topological superconductivity becomes very nontrivial. This is where ML comes into play. We want to understand how ML responds to the topological nontriviality of topological superconductors if it is given input data with certain ambiguities.

We employ supervised learning based on the CNN, which means that we need labelled data, namely, zero mode or non-zero mode, for our binary classification problem. Because we are interested in Majorana fermions (which are charge neutral), presenting the data in a form that exploits the particle–hole symmetry of the Hamiltonians is advantageous. For this purpose, let us re-express the eigenvalue problem in the following form (temporarily assuming complex Δ for generality):

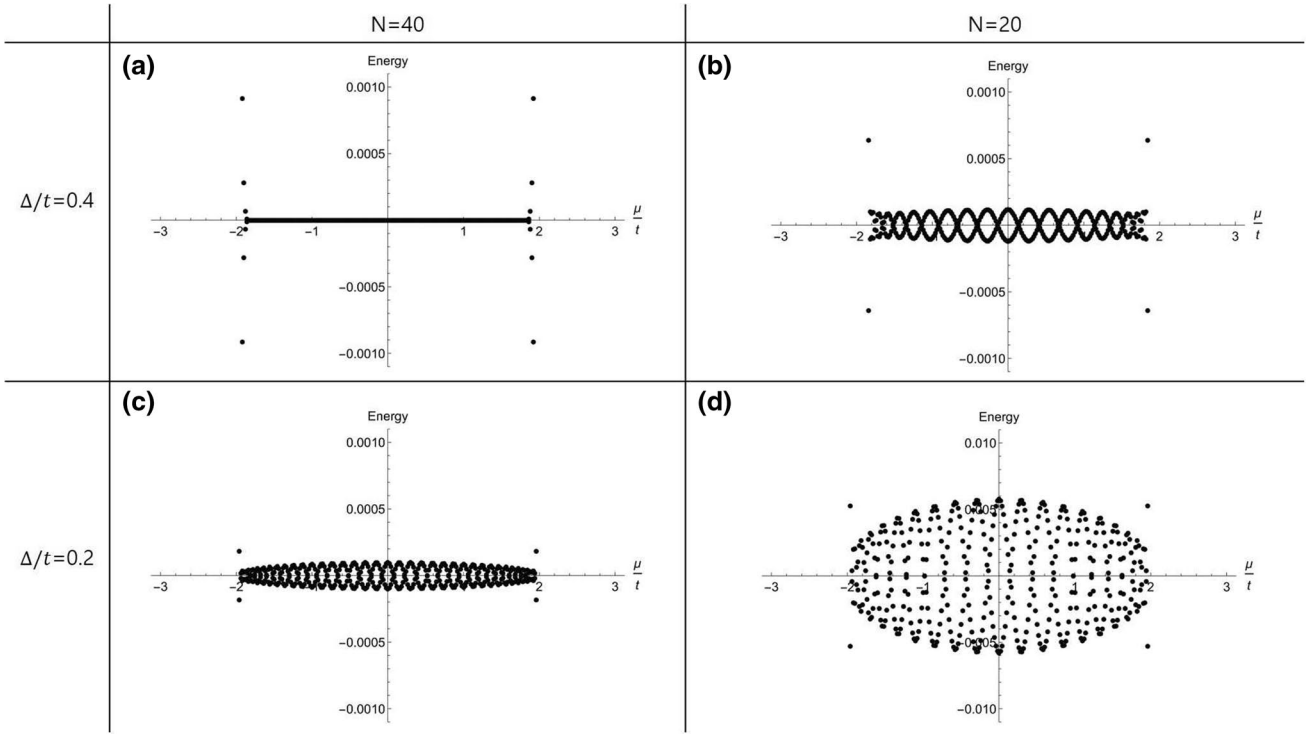

Fig. 1 Energy spectrum of the NoSSD case as a function of μ/t in the vicinity of zero energy. The number of lattice sites N is 40, 20 (left, right), and $\Delta/t = 0.4, 0.2$ (top to bottom). Note the different energy scale in (d)

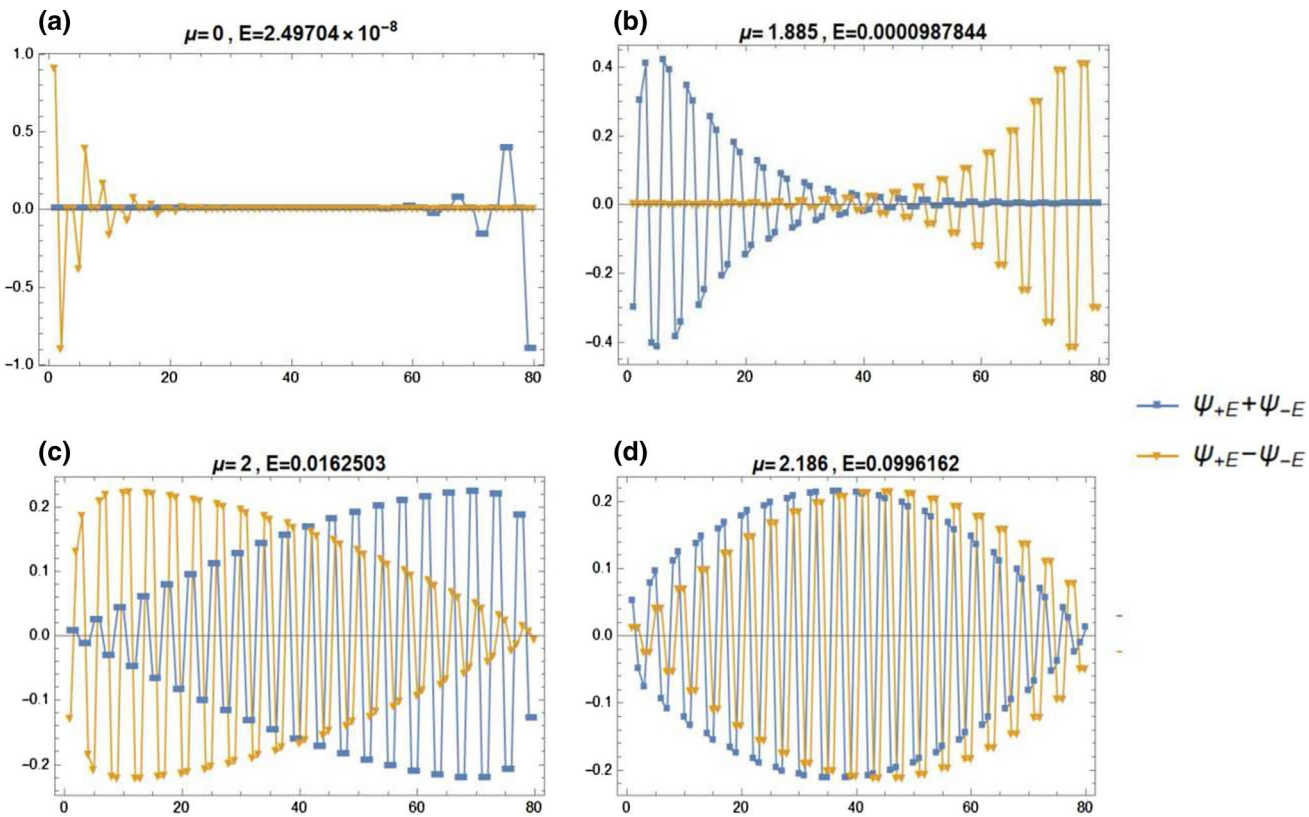

Fig. 2 Typical examples of $\psi_E \pm \psi_{-E}$ for the NoSSD case. (a) and (b) are the Majorana zero modes, while (c) and (d) are the Majorana non-zero modes. $N = 40$ and $\Delta/t = 0.4$. Each case is labeled by the energy and by μ/t. $E_{\text{upper}} = 0.1t$, and $E_{\text{threshold}} = 0.001t$ (cf. Fig. 1a)

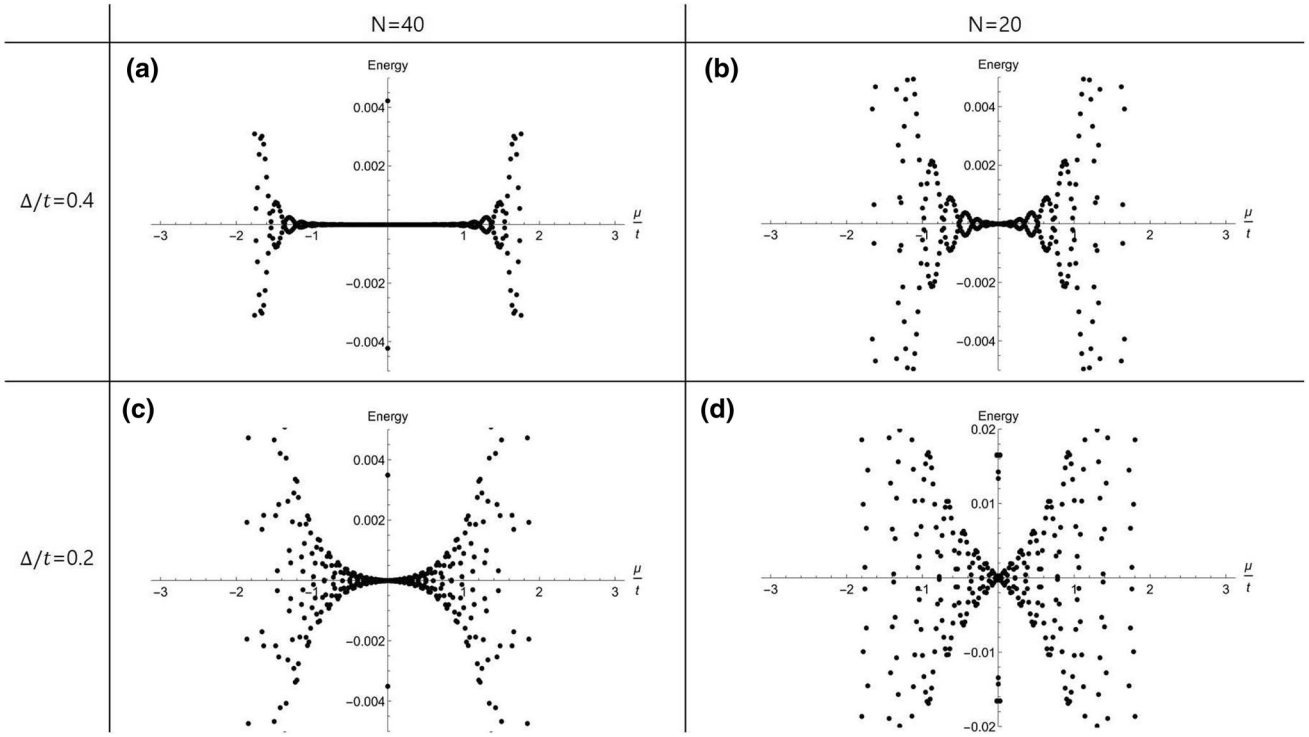

Fig. 3 Energy spectrum of the LinkSSD case as a function of μ/t in the vicinity of near zero energy. Note the different energy scale in (d)

$$\frac{1}{2}\begin{pmatrix} \mathsf{H}_0 & \Delta \\ -\Delta^* & -\mathsf{H}_0^* \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = E\begin{pmatrix} u \\ v \end{pmatrix}, \tag{13}$$

where u and v denote the spatial parts of the spinor (with N components). Then, the spinor $(+v^*, u^*)^{\mathrm{t}}$ is seen to have $-E$ as an eigenvalue. Let us choose the phases of the eigenvectors to be real and take the following combinations of the eigenvectors ψ_E, ψ_{-E} with the energy eigenvalues $+E$ and $-E$:

$$\psi_E + \psi_{-E} = (u+v, u+v) = (u+v)(1,1), \quad \psi_E - \psi_{-E} = (u-v, v-u) = (u-v)(1,-1). \tag{14}$$

Thus, the spinor parts become identical with those of the Majorana zero modes discussed in Section 2. In fact, the above shows that the zero modes should be doubly degenerate. Because we are looking for the zero energy states, that the most relevant data would be a set of eigenvectors with energy lower than a certain upper bound (E_{upper}) is evident: then, Eq. (14) is the proper combination to compare with the fractionalized Majorana zero-mode state. The combinations of Eq. (14) turn out to actually be equivalent to taking combinations of even- and odd-parity eigenvectors [24] owing to the parity symmetry of the Hamiltonians.

The data we use to train and test our deep learning algorithm are made up of the combination of the eigenvectors Eq. (14), where the eigenvectors with (the absolute value of) energy below E_{upper} are chosen for our dataset. Also, practically the energy of the zero mode is not strictly zero, so we have to introduce a certain threshold energy $E_{\mathrm{threshold}}$ such that the states with energies lower than $E_{\mathrm{threshold}}$ are regarded to be zero modes. E_{upper} and $E_{\mathrm{threshold}}$ are adjusted according to both the type of SSD and the parameters Δ/t and N. We designate the types of SSD as NoSSD (Eq. (6)), LinkSSD (Eq. (8)), and AllSSD (Eq. (9)). The data are selected by picking out 2000 random values of μ/t in the range of [–3, 3] for three different values of Δ/t (0.2, 0.3, and 0.4), and all the corresponding eigenvector combinations are labelled as either a zero mode or a non-zero mode. The combinations are labeled as Majorana zero mode states only when μ/t is within the topologically non-trivial range *and* when they have the lowest eigenvalues for the μ/t in consideration. We note that because only one zero mode exists for a given value of μ/t (for $|\mu/t| > 2$, none exist), the non-zero mode data tend to be more than zero mode data for the LinkSSD and the AllSSD cases. The numbers of extracted data are presented in Table 1.

Now, let us examine the data for the NoSSD case. The spectrum of the eigenvalues with energy less than E_{upper} and

Table 1 Number of data used in the ML training and the test of the zero mode

Δ/t	NoSSD			LinkSSD			AllSSD		
	0.2	0.3	0.4	0.2	0.3	0.4	0.2	0.3	0.4
Zero mode	2676	2530	2514	2526	2332	2386	2702	2572	2694
non-Zero mode	1978	1598	1242	4200	10938	12348	41546	40188	39336

the combinations of the eigenvectors are presented in Figs. 1 and 2, respectively.

The Majorana zero mode is clearly visible in the range $-2 < \mu/t < 2$. The eigenvectors of the Majorana zero modes are depicted in Fig. 2. For the case of $\mu = 0$, the combinations of the zero-mode eigenvectors are well localized at *one end only* while as μ/t approaches the critical value ± 2, the localized nature weakens, for the exponent μ_{c0}/Δ of Eq. (11) becomes smaller near the critical point. As is evident from Fig. 2, away from the critical points $\mu/t = \pm 2$, the Majorana zero modes can be clearly distinguished from the non-zero modes.

Because we are focusing on the zero energy states, the influence coming from the energy fluctuation caused by the finite size effect should be taken into account. The input data taken from the simulated data are subject to the finite-size effect, as is demonstrated in Fig. 1 by comparing the left and the right panels. Clearly, for smaller lattice size, the criterion for the zero mode becomes blurry, and this affects the performance of ML. Interestingly, the finite-size effect becomes weaker as Δ increases. This is because at smaller Δ, the topological stability of the topological superconductor becomes weaker.

Next, we examine the case of LinkSSD. The eigenvalues with energies below E_{upper} and the combinations of the eigenvectors are presented in Fig. 3 and Fig. 4, respectively. Comparing Fig. 3 with Fig. 1, we conclude that LinkSSD diminished the stability domain for the zero mode. In particular, the demarcation between non-zero modes and zero modes becomes ambiguous around the critical points $\mu/t = \pm 2$.

As for $\psi_E \pm \psi_{-E}$, we have to note that the combinations in panels (a) and (b) (the zero mode) of Fig. 4 are localized at only one end while in the panels (c) and (d) (the non-zero

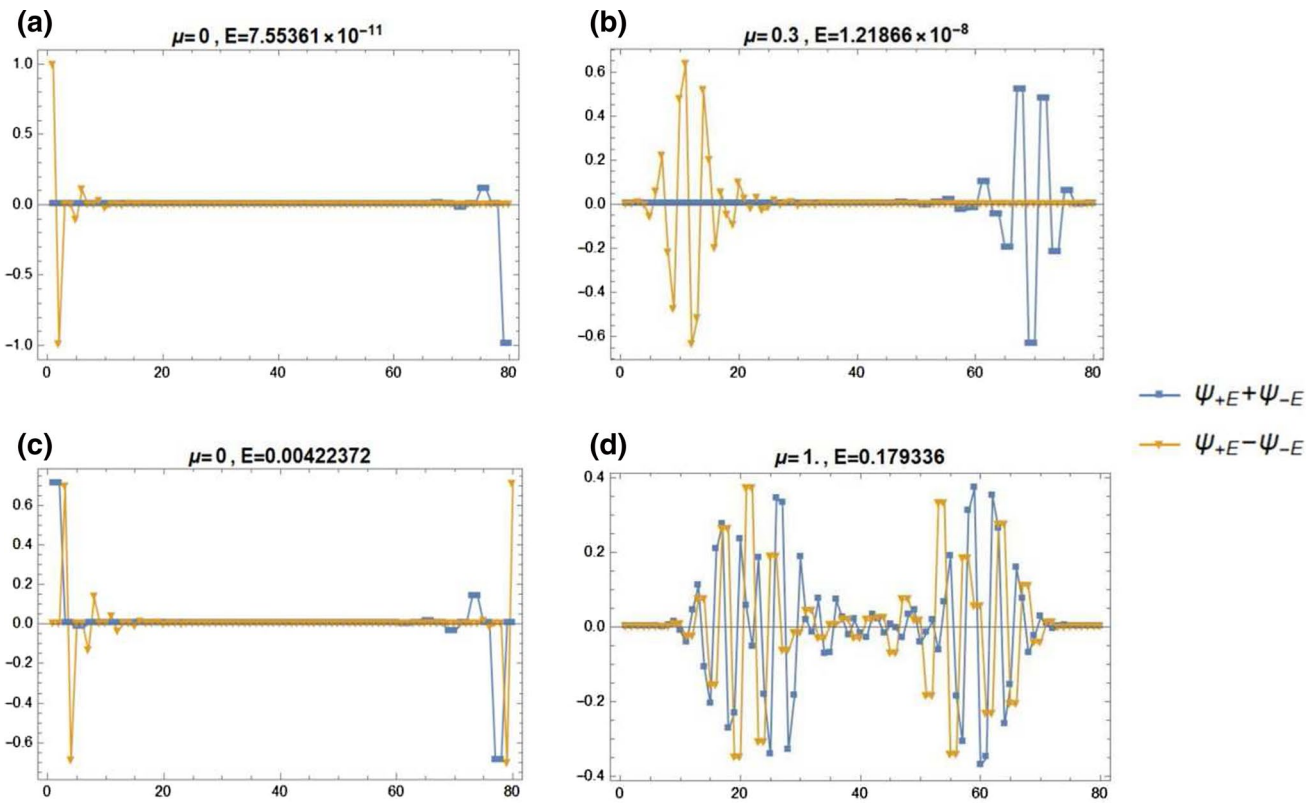

Fig. 4 Typical examples of $\psi_E \pm \psi_{-E}$ for the LinkSSD case. **a** and **b** are the Majorana zero modes while **c** and **d** are the Majorana non-zero modes. $\Delta/t = 0.4$, and $N = 40$. $E_{\text{upper}} = 0.2t$, and $E_{\text{threshold}} = 0.004t$ (cf. Fig. 3a)

mode), they appear to be localized at both ends. These features will be learned by the ML algorithm. A qualitative explanation for the eigenvectors in panel (b) of Fig. 4 can be found in Ref. [19].

Finally, we examine the case of AllSSD. The eigenvalue spectrum and the combinations of the eigenvectors are presented in Figs. 5 and 6, respectively. For the AllSSD case, the difference between the Majorana zero modes and non-zero modes is significantly robust, compared to the LinkSSD and the NoSSD cases. The Majorana zero mode is clearly visible in the range $-2 < \mu/t < 2$ and is well localized at one end shown in panel a of Fig. 6. A qualitative explanation for this behavior is given in Ref. [19]. Also, finite size effects are seen to be strongly suppressed compared to the LinkSSD case. In particular, we must note that the wavefunction of panel c of Fig. 6 (at $\mu = 2$) is labelled as a non-zero mode even though it is very similar to that of panel b.

4 Neural network structure

A neural network is one of the most powerful techniques used by supervised learning algorithms. From diverse neural networks, we choose to deploy the CNN, which was specifically designed for image processing. The idea is to convert the raw labelled data of the combinations of the eigenvectors into image formats and to let CNN classify the image dataset into two categories: zero mode (topological superconductor) or non-zero mode (non-topological superconductor). A schematic of the data processing and CNN structure is presented in Fig. 7.

The $\psi_E \pm \psi_{-E}$ combinations of eigenvectors, which have 80 components ($N = 40$), are divided into 8 equal components and stacked from top to bottom and left to right, resulting in a dataset of 8×10 matrices. Next, these matrices are converted into gray images and augmented by a factor of five in order to make training smoother and more flexible. The CNN consists of two convolution layers: the first layer having 16 5×5 sized filters, and the second having 32 3×3 sized filters. A maximum pooling layer with a pooling size of 4×5 is placed between the two convolution layers. Although not shown in the figure, a batch normalization layer exists between the convolution layer and the activation function, which is chosen to be the leaky rectified linear unit (leaky ReLU). The last convolution layer is followed by a fully connected layer and a softmax classifier with two neurons each representing a zero or a non-zero mode.

We to implement the above CNN use MATLAB's inbuilt algorithm provided by the Deep Learning Toolbox. The Stochastic Gradient Descent method using momentum optimization is used in the training process [25]. Our

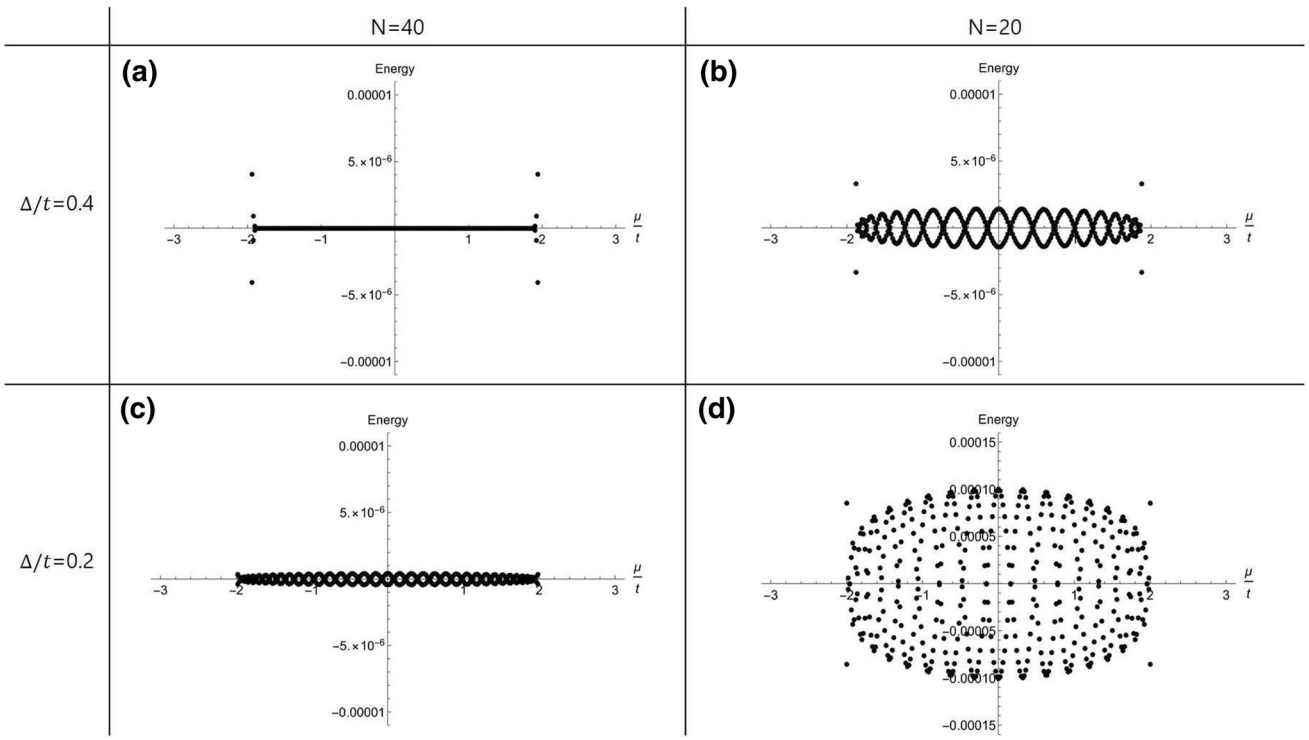

Fig. 5 Spectrum of the energy eigenvalues for the AllSSD case as a function of μ/t in the vicinity of zero energy. Note the different energy scale in (**d**)

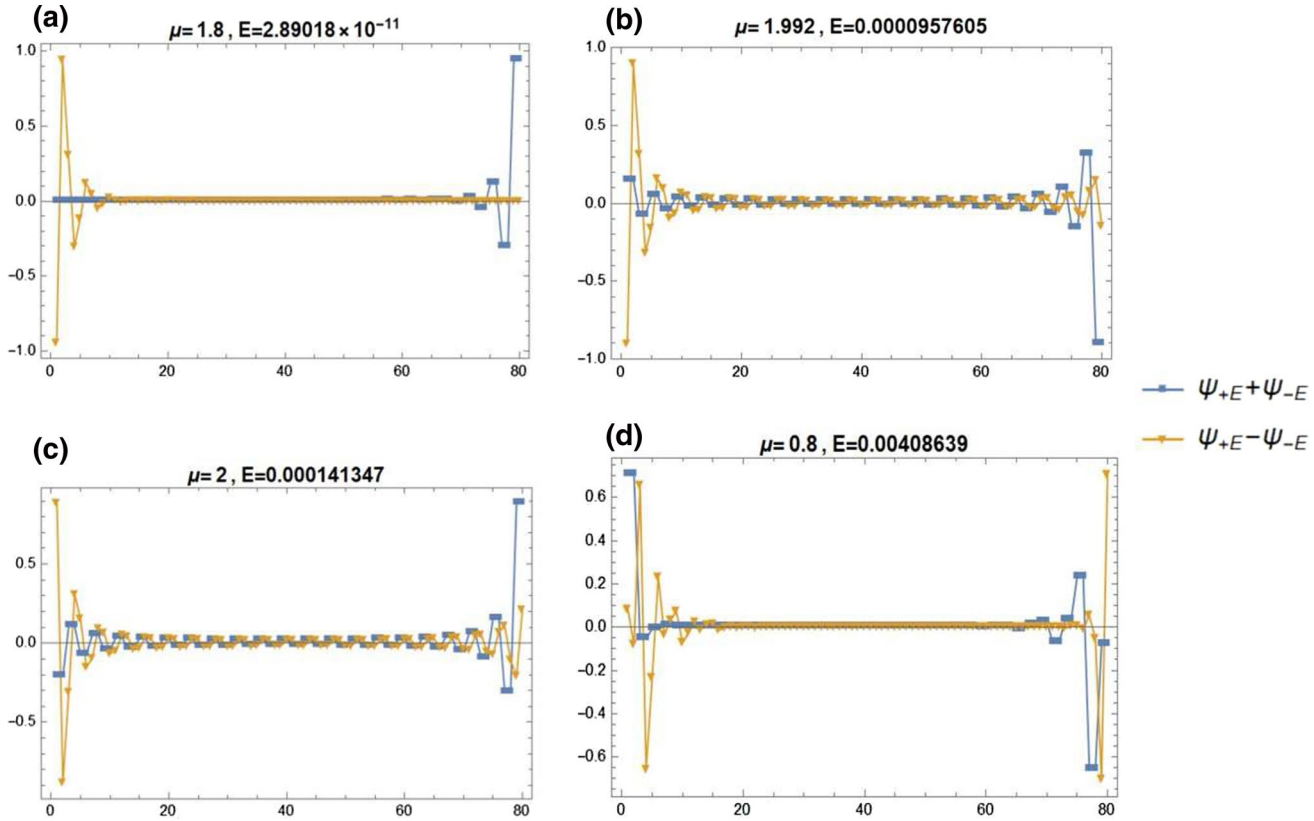

Fig. 6 Typical examples of $\psi_E \pm \psi_{-E}$ for the AllSSD case. **a** and **b** are for Majorana zero modes, while **c** and **d** are the Majorana non-zero modes. $N = 40$, and $\Delta/t = 0.4$. $E_{\text{upper}} = 0.1t$, and $E_{\text{threshold}} = 0.0001t$ (cf. Fig. 5a)

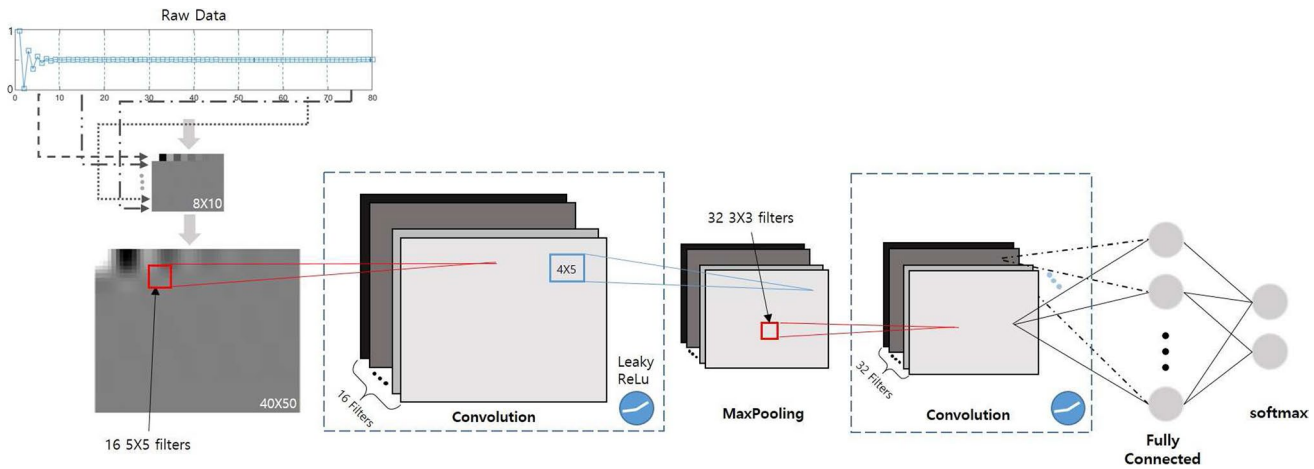

Fig. 7 Schematic of the data processing and the main structure of the CNN employed in this paper

image data are relatively simple and intuitive (even though the data for the deformed cases can be quite ambiguous), so a small number of features (such as the numbers of filters and epochs) will suffice for our purposes. This allows, and actually forces, us to set the network parameters lower than those that would be used to prevent over-fitting. A network consisting of two convolution layers performs well while preventing the network from become too deep and complex. In fact, the CNN with three layers does not yield a significant improvement of the classification accuracy. We also comment that for prototypical networks with a single convolution layer, adding a pooling layer

reduces the prediction accuracy for other SSD case data substantially.

L2 regularization has been employed throughout the whole training process, and the maximum number of training epoch is kept at 10, with the Validation Patience parameter set at 10 [25]. Such measures were necessary, because the training accuracy reaches almost 100%, while the loss function also rapidly converges to almost 0% within the first epoch. Normally, such behaviors imply either overfitting or an *easy* dataset compared to the complexity of the CNN model, which indicates that most of the important features are learned during the first few iterations.

We also tried a simpler network consisting of one convolution layer with a limited number of filters (1, 2, 4). The rest of the network structure were held fixed. The performance of the simpler network was compared with that of the main two-layered network model. Even though the loss function died out substantially within the first epoch, a complete plateau toward zero value was reached only after several more runs, in contrast with the main two-layered network. Despite such improvements, the critical downside is that compared to the two-layer network, the simpler networks yield worse classification accuracy for datasets obtained from different SSD cases.

In view of these observations, retaining the slightly more complicated overall network structure of two convolution layers with limited iterations and epochs appears to be the optimal choice. The input data are extracted with labels from three cases, NoSSD, LinkSSD, and AllSSD, independently, as discussed in Sect. 3. These data are input into our neural network *separately* without mixture.

5 Results

The labeled dataset obtained in Sect. 3 is the input to the CNN developed in Sect. 4. We randomly split each dataset into train, validation, and test datasets in the approximate ratio of 70%, 15%, and 15%, respectively. The training dataset are selected only from the eigenvectors with $\Delta/t = 0.4$ to reduce the finite-size effects.

Once the training procedure has been completed using the train and the validation datasets, evaluation on the model is provided by the remaining test dataset. Once this cycle is finished, the final CNN is applied to various datasets of the Majorana combination of eigenvectors for topological phase classifications.

The result of the CNN and a schematic of the training and the classification processes are presented in Fig. 8.

The accuracy of the results is defined as the percentage of eigenvectors classified as a corresponding mode that belong to the actual labeled state. Note that the total average accuracy plotted on the main graphs of Fig. 8 may be biased due to the fact that the initial datasets contain many more non-zero modes than zero modes (see Table 1), which, in fact, may reflect the experimental difficulties in obtaining the true zero-mode Majorana fermions.

To address this point, we provide the inset plots of Fig. 8 to assess the performance of the three trained networks in classifying the zero modes as zero modes and the non-zero modes as non-zero modes. The dotted lines in the inset data are obtained from classifying test data made up of zero modes only, and similarly, the semi-dotted line data are obtained for the non-zero modes. These results demonstrate that the trained CNN accurately predicts zero mode data as zero mode and non-zero mode data as non-zero mode, so that *no bias* is caused by the overwhelming abundance of non-zero mode data.

Comparing panels a–c of Fig.8, we find that, evidently, the CNN trained with LinkSSD data performs the best in distinguishing between zero and non-zero modes of the test datasets from the NoSSD, LinkSSD, and AllSSD cases. Because the structure of the CNN is fixed as of Fig. 7, such a difference in performance implies that the LinkSSD training data set is a most diverse and flexible dataset, which is in line with what we discussed in Sect. 2. In terms of bias-variance tradeoff, given a finite amount of training data, we can restate the above result by saying that the LinkSSD-trained CNN is less dependent on the particular realization of the training data, i.e., has a lower variance.

On the contrary, the two other models display poorer performances in classifying other types of datasets. For the NoSSD case, this may stem from the fact that it is provided with a smaller amount of data compared to the other cases (see Table 1) and is, thus, more liable to finite-size sampling noise. The NoSSD data with wide error bars support this observation in general. Additionally, all the zero modes in the NoSSD dataset have the highest peak at one end of the wire. From the inset of Fig. 8c, we confirm that the NoSSd-trained model cannot perceive the zero modes of the LinkSSD (see the dotted lines with the star marker) while successfully picking out those of the AllSSD with almost 100% accuracy (see the dotted lines with the square markers). Comparing Figs. 2a, 4b, and 6a, we can infer that such an inflexibility in the training data creates a model with high variance.

This behavior is also seen in the AllSSD case, despite the fact that it is provided with the largest training dataset. The rigidity of the zero-mode eigenvector combinations is the strongest in AllSSD due to the suppression of the finite size effect by the SSD; i.e., all the zero modes (except for the ones in the vicinity of the boundaries of the critical range of μ/t) of the AllSSD have a single sharply localized peak at one end of the wire. From the inset of Fig. 8b, we confirm that AllSSD trained models have low performance rates for

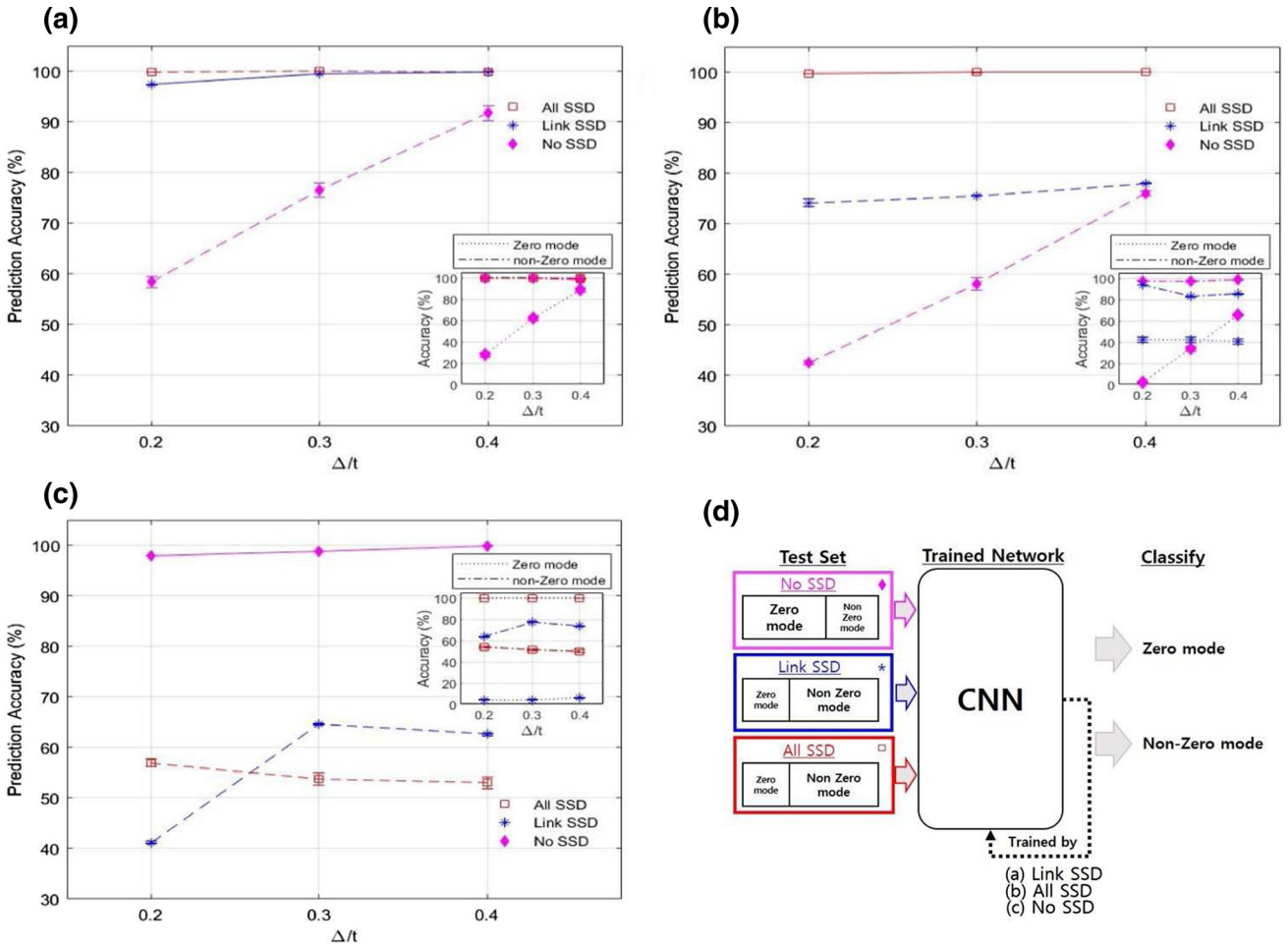

Fig. 8 Average accuracy for the classification of zero and non-zero Majorana modes at each value of Δ/t. The vertical error bars represent the standard errors of the mean over twelve separate runs of the training process. From **a**–**c**, each figure shows the performance of three separate CNNs that are trained with data extracted from three different cases; LinkSSD, AllSSD, and NoSSD, all with $\Delta/t = 0.4$. **d** Schematic of the training and the classification processes

classifying zero modes of other SSD data (see the two dotted lines).

Also, returning to the inset of Fig. 8c, the non-zero mode prediction for the AllSSD case is also quite low at about 50% no matter the value of Δ/t (see the semi-dotted lines with the square markers), because the data with *two* narrow peaks positioned at both ends of the wire, such as in Fig. 6d, were inaccurately predicted as zero modes.

Thus, we conclude that when images consisting of a localized peak at one end are provided as the only zero modes, the algorithm *learns* the following: zero mode data have peaks that must be localized at the *ends* of the wire, and the number of peaks does not matter as long as the highest peaks are located at the ends. On the other hand, when the CNN is trained with data that has a wavefunction localized away from the ends, as in the zero-mode LinkSSD data (Fig. 4b), then the network finally learns to seek out the number of peaks while being flexible to the actual location and width of the distribution, thus achieving generalization, i.e., a lower variance.

Additionally, in Fig. 8a and b, the prediction accuracy for the NoSSD test set is higher for datasets corresponding to larger value of Δ/t (see the diamond shaped markers). Such behavior is manifest in the total accuracy, as well as the accuracy for only zero modes, in the inset graph. Because the accuracy for the test data consisting only of non-zero modes always reaches near 100%, that the strong correlation between Δ/t and the accuracy of the NoSSD data classification originates from the properties of the zero-mode dataset is evident. To put it in simpler terms, the trained networks in Fig. 8a and b are less effective in predicting the zero modes with smaller values of Δ/t while being perfectly capable for non-zero modes.

As explained in Sect. 2, at lower Δ/t values, the eigenvalues and the eigenvectors become more susceptible to the finite size effects and perturbations. In particular, our test dataset becomes more rapidly ambiguous near the critical boundaries of $\mu/t = \pm 2$ (namely, the localization behavior weakens). Considering that Fig. 8a and b are trained by using the LinkSSD and the AllSSD train datasets, where the zero-mode data display a single sharp peak at the edges or a single bell-shaped localization, their inability to discern the zero-mode states with a broad profile, such as in Fig. 2b, is understandable.

6 Discussion and summary

As an extension of the result in Sect. 5, the performance of our model can be further improved if the network is trained using mixed types of data from different SSD cases rather than just using the dataset from one case. However, our focus does not lie in building a network that operates at high accuracy, but rather in understanding what attributes the network learns when provided with specific types of data.

For example, when we train the CNN by using entire AllSSD dataset together with zero-mode images from the LinkSSD, the classification accuracy for NoSSD zero mode improved substantially to about 91%. This implies that the addition of diverse zero-mode data of LinkSSD enables the algorithm to learn about more generalized attributes; thus, it becomes a model with lower variance.

Recall that for one convolution layer network having 1, 2, or 4 filters (discussed in Sect. 3), the overall classification accuracy for other SSD cases was low. However, when we increase the number of filters up to 16, the network trained with the LinkSSD dataset only, exhibits a better prediction accuracy (nearly 95% on average) with smaller standard errors for the NoSSD zero-mode images while it performs poorly on AllSSD zero modes, 58% on average. In contrast, our main CNN in Fig. 7 classifies the AllSSD test dataset with almost 100% accuracy while its performance is relatively low (nearly 89%) for NoSSD zero-mode data, as can be seen in Fig. 8a. The comparison of these reversed behaviors causes us to speculate that the addition of an extra convolution layer, i.e., increasing the depth of the CNN, enables the network to learn smaller-range localized features. This is in line with the general fact that CNN with multiple layers can probe more detailed and localized features of input images.

Although in our paper ML processed the data from theoretical models, it can truly excel is in the analysis of complex experimental data subject to noise and other (undesirable or interesting) perturbations. In our study, such perturbations are simulated through the SSD and the finite size of our system. From the perspective of our theoretical work on topological superconductors, the tunneling microscopy experiments [26] were particularly relevant, because they give direct information on the local density of states.

In summary, we have applied a supervised machine learning algorithm in the neural network architecture to the problem of classifying the topological phase of a SSD one-dimensional topological superconductor. Our study clearly illuminates the aspects of the operating mechanism of the CNN when applied to physical systems.

Acknowledgements This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2019R1F1A1058671).

References

1. P. Mehta, M. Bukov, C.-H. Wang, A.G.R. Day, C. Richardson, C.K. Fisher, D.J. Schwab, A high-bias, low-variance introduction to machine learning for physicists. Phys. Rep. **810**, 1–124 (2019)
2. J. Carrasquilla, Machine learning for quantum matter. Adv. Phys. X **5**, 1797528 (2020)
3. G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová, Machine learning and the physical sciences. Rev. Mod. Phys. **91**, 045002 (2019)
4. J. Carrasquilla, R.G. Melko, Machine learning phases of matter. Nat. Phys. **13**, 431 (2017)
5. P. Mehta and D. J. Schwab, An exact mapping between the variational renormalization group and deep learning, arXiv:14103831 (2014)
6. R.G. Melko, G. Carleo, J. Carrasquilla, J.I. Cirac, Restricted Boltzmann machines in quantum physics. Nat. Phys. **15**, 887 (2019)
7. G. Carleo, M. Troyer, Solving the quantum many-body problem with artificial neural networks. Science **355**, 602 (2017)
8. Y.-Z. You, Z. Yang, X.-L. Qi, Machine learning spatial geometry from entanglement features. Phys. Rev. B **97**, 045153 (2018)
9. M.Z. Hasan, C.L. Kane, Topological insulators. Rev. Mod. Phys. **82**, 3045 (2010)
10. X.-L. Qi, S.-C. Zhang, Topological insulators and superconductors. Rev. Mod. Phys. **83**, 1057 (2011)
11. B.A. Bernevig, T.L. Hughes, *Topological Insulators and Topological Superconductors* (Princeton University Press, Princeton, 2013)
12. A.Y. Kitaev, Unpaired Majorana fermions in quantum wires. Phys. Usp. **44**, 131 (2001). ((**Number 10S**))
13. P. Zhang, H. Shen, H. Zhai, Machine learning topological invariants with neural networks. Phys. Rev. Lett. **120**, 066401 (2018)
14. R. Jackiw, C. Rebbi, Solitons with fermion number 1/2. Phys. Rev. D **13**, 3398 (1976)
15. A. Gendiar, R. Krcmar, and T. Nishino, Spherical deformation for one-dimensional quantum systems, Prog. Theor. Phys. **122**, 953 (2009); **123**, 393 (2010)
16. T. Hikihara, T. Nishino, Connecting distant ends of one-dimensional critical systems by a sine-square deformation. Phys. Rev. B **83**, 060414(R) (2011)
17. H. Katsura, Exact ground state of the sine-square deformed XY spin chain. J. Phys. A **44**, 252001 (2011)
18. H. Katsura, Sine-square deformation of solvable spin chains and conformal field theories. J. Phys. A **45**, 115003 (2012)
19. J.H. Lee, H.C. Lee, The sine-square deformation of the one-dimensional p-wave topological superconductor. J. Kor. Phys. Soc. **75**, 997 (2019)

20. A. Bohrdt, C.S. Chiu, G. Ji, M. Xu, D. Greif, M. Greiner, E. Demler, F. Grusdt, M. Knap, Classifying snapshots of the doped Hubbard model with machine learning. Nat. Phys. **15**, 921–924 (2019)
21. C. Miles, A. Bohrdt, R. Wu, C. Chiu, M. Xu, G. Ji, M. Greiner, K. Q. Weinberger, E. Demler, and E.-A. Kim, Correlator convolutional neural networks: an interpretable architecture for image-like quantum matter data.arXiv:2011.03474 (2020)
22. E. Khatami, E. Guardado-Sanchez, B.M. Spar, J.F. Carrasquilla, W.S. Bakr, R.T. Scalettar, Visualizing strange metallic correlations in the two-dimensional Fermi-Hubbard model with artificial intelligence. Phys. Rev. A **102**, 033326 (2020)
23. M. Paluszek, S. Thomas, *Practical Matlab Deep Learning* (Apress, New York, 2020).
24. S.-R. Eric Yang, Soliton fractional charges in graphene nanoribbon and polyacetylene: similarities and differences. Nanomaterials **9**, 885 (2019)
25. http://www.mathworks.com/help/deeplearning/ref/trainingoptions.html
26. M. Ziatdinov, A. Maksov, L. Li, A.S. Sefat, P. Maksymovych, S.V. Kalinin, Deep data mining in a real space:separation of interwined electronic responses in a lightly doped $BaFE_2$ As_2. Nanotechnology **27**, 475706 (2016)