Dear editors,

Thank you for sending us the referee comments on our manuscript titled “Machine-learning skyrmions”. We have made considerable revisions to the content while trying to address each of the referee comments. Please see below our responses to comments, and the summary of changes made in the updated manuscript.

Report of Referee A -- LT16311/Singh

The manuscript comprises four parts. Here are the questions and comments for each of them.

1) The first part provides a detailed description of PCA method. The PCA method is well known and one can find its description across the literature. The problem is that from the text it is not clear how this method improves the operation of ML program or facilitates phase classification and observables evaluation. The authors just stated that «In fact, very similar leading eigen-images are obtained for all three phases: Sp, SpSk, and Sk», which means that PCA cannot be utilized for phases classification problem. This point should be clarified.

Reply: We have decided to completely remove the PCA part from the renewed manuscript. Instead we now sharply focus on the very issue we set out to address at the beginning of this project: proving the utility of machine learning in the interpretation of experimental data. To establish strong connection to the experiment, several new pieces of calculations have been performed, and some of the older results were struck out including the PCA part. The ones we removed are of more marginal nature in light of our overall agenda. The ones we do keep and add in the revision are of much practical importance to those engaged in the experimental science. With this new format, we believe the new manuscript greatly enhances the accessibility to a large audience and finally deserves publication.

2) The next part is devoted to application of a specific ML scheme to phase classification problem. The authors found that the chosen ML method is not capable to characterize mixed phases. They said that “… one must seek other means of characterizing mixed phases.” In this concern it is desirable to discuss results of other works connected with mixed phases classification. For instance, the authors of work [arXiv: 1803.06682] have performed a qualitative analysis of mixed SkFm and SkSp phases simulated with a similar Hamiltonian containing DMI interactions by means of simple neural network.

Reply: In the earlier version of the paper we did not detail the differences between our

work and arXiv:1803.06682. In part this was due to the fact that their work did not address the phase co-existence region carefully. In the new manuscript we decided to confront their work head-on, making comparisons in both the ML architectures used by the respective group and the results obtained. Since they do not mention the mixed-phase issue in detail (actually not at all), it was not possible to make quantitative comparisons between their conclusions and ours on this particular issue, even after revision. Our ML architecture differs significantly from theirs in the inclusion of several CNN layers and hidden layers. Only one hidden layer was used in their work, and no CNN layer was ever invoked. These differences are mentioned in the main text now. Instead of delegating the description of the architecture we used to the supplementary material, we decided to move it to the main text so that readers interested in the ML implementation aspect of the problem can reproduce our data if necessary.

The authors of arXiv:1803.06682 used the z-component of the spin configuration data to train their model. In the new version, we show results from training the program on the basis of all xyz components, xy-components only, and z-component only. The picture we provide is therefore far more comprehensive and addresses the versatility of the ML training, since all training procedures are shown to lead to the same prediction regardless of the data sets (xyz, xy, or z) used in the training. We do not find other works in the literature with this level of demonstration for the universality of ML training. Most importantly, our new demonstration is able to establish key connections between experimental data from LTEM (which measures xy components of spin) and MFM (which measures z component of spin). We propose that feeding such kind of experimentally harvested data to the machine learning software we developed could lead to predictions for spin components not measurable by a given apparatus. This is a very interesting new bridge connecting experimental data and machine learning.

One of the drawbacks concerning this part of the work is that the machine learning procedure is not reproducible. For example, the authors write that they used 16 (6x6) filters in the first layer and 32 (3x3) ones in the second layer, but structure of filters wasn't presented and there is no rule how to construct them.

Reply: The 6x6 filter size in the first CNN layer is motivated by the inherent size of the spiral and skyrmion structures in the model. The second CNN layer is chosen to have 3x3 size from trial-and-error. A certain amount of trial-and-error decisions are very common in the machine learning architecture. Our choice is motivated by what gives out the best outcome. The rationale for the 6x6 filter size is explained in detail in the main text. A 6x6 convolution filter consists of 36 unknown parameters which are learned during the machine learning. A convolution filter can learn a wide variety of local feature like local magnetization, spin chirality etc. The convolution operation involves inner product of all the neighborhood elements of a considered lattice point with the convolution filter. So the image forwarded to the next layers has the data of local behaviour at every lattice point making it easier for the prediction of net global

features. The reason for using 16 (or multiple) filters is that not every filter reads the same local feature. Every filter works independently to understands different local features and modify themselves in order to match the exact result.

It is assumed that the D/J ratio was always kept equal to the square root of 6 in this work, which set a specific spin spiral period and radius of skyrmions. Keeping in mind the application of the proposed method to experimental data usually having larger sizes of skyrmions (~700 Å for FeGe, [Nature Mater. 10, 106]), the natural question arises: how one can scale the method in order to use it with different skyrmion sizes? The lack of ML architecture implementation details does not allow other researchers to apply this approach even for different model parameters. Thus, it is impossible to estimate the universality of the proposed method.

Reply: The lattice spacing in the model does not need to have a one-to-one correspondence with the actual crystal lattice spacing. One should think of our lattice spacing as the artificial spacing, introduced to discretize what was originally a continuum effective model of spin dynamics. In turn, the experimental data can be coarse-grained in a way that the effective new lattice spacing becomes the same as the one used in the ML training. This kind of procedure is familiar from the real-space renormalization, and one can easily implement it to the raw data. For instance, let's say we have the raw data from some imaging tool that gives spin values over 100x100 grid points for a single skyrmion. Using a sensible averaging procedure one can reduce this information to a set of unit vectors defined on a coarse-grained 6x6 grid points. After such coarse-graining treatment, the new data set can be fed to the ML algorithm we have developed to make predictions on quantities that could not be measured in an experiment. All of these schemes are outlined carefully in the new version.

3) The third part is aimed at the extraction of mechanical (net magnetization, chirality) and thermodynamic (temperature, magnetic field) properties of the system from ground state magnetic configuration. One can see, that the ML program gives qualitatively correct estimations of the above properties once they were averaged over a number of Monte Carlo calculations (Ref. 25) and over obtained magnetic configurations, respectively. On the other hand, considered method gives "a greater degree of fluctuation" for an individual magnetic configuration. Such a feature of the method should be discussed and analyzed in the main text rather than in bibliography.

Reply: In the new figure we present unaveraged data taken from one input image. Since the statistical fluctuations were quite small, even a small prediction was already very close to the actual value. We removed the comment and any reference to the statistical error.

Another limitation of the approach is that it requires the information about all three components of local magnetization, whereas experimental techniques could extract either z- or xy-components during a single measurement and evaluation of approximate vector magnetization density requires careful post-processing of a number of measurements [Nature Physics 7, 713; Nature Mater. 10, 106]. Despite all these limitations, the authors state in conclusion, that the proposed method could be easily applied to experimentally produced data. Can the authors elaborate on details of the application of the proposed method to experimental data?

Reply: Motivated by the referee comment we carried out a set of new numerical experiments, as reported in the revised manuscript. Rather than training the ML algorithm on the full xyz components of the Monte Carlo-generated magnetization profile, we created new training procedures where only the xy-components, or even just the z-component of the data was fed. Despite such limited input of information, all three ML programs trained on the xyz, xy, and z-component data, respectively, gave equally good performance in predicting values of spin chirality, magnetization, magnetic field, and the temperature. This is a testament to the powerful universality of the machine learning algorithm, and a strong suggestion that ML can be an extremely useful "recovery tool" to reconstruct pieces of data missing in an experimental measurement. By following the referee's suggestion seriously, we believe not only that the content of the article is significantly improved, but the outcome starts to have genuine application to the experiments. We are grateful to the referee for this key suggestion.

The motivation of this part of work is also not clear. Indeed, if one has information about all components of local magnetic moments in a configuration then it is possible to simply calculate exact values of chirality and total magnetization without complicated ML approach. Similarly, temperature and magnetic field are usually known from experimental conditions. In my opinion, this study does not pave the way for notable progress in an existing field, which is important for Physical Review Letters standards.

Reply: After the new round of calculations based on xyz, xy, and z component data, we are confident that the motivation and the thrust of the paper is extremely clear. We made very careful statements as to the utility of our suggested ML program in the hope that readers can see the virtue of our idea clearly. Ultimately, we believe some experimentalists will come round to our way of thinking and start implementing our suggestion into their experiments.

Concerning the details of the ML process, one could notice that the training set was obtained at D/J equal to the square root of 6 in the temperature and field ranges [0.1, 2.0] and [0.1, 4.0], respectively. Figure 3 and similar figures in SM show that feature extraction has

been performed in the same ranges of parameters. Was the ML program tested on truly unseen data, using the above training set? (at higher fields and temperatures or different D/J ratio).

Reply: At higher fields we reach the spin-polarized ferromagnetic state, at which the ML prediction becomes trivially correct. We do not believe that using one value of D/J for training and applying it to the test set generated at another value of D/J is a good idea. The CNN filter size is geared toward the inherent spacings of the spiral and the skyrmion. A different D/J value gives rise to an entirely new spiral and skyrmion sizes. We believe the most effective application of the ML idea is to apply it to the images having the same inherent length scale. To this end, we are suggesting that experimental data be re-adjusted and coarse-grained to fit the scale we used in the training. The suggested procedure is laid out carefully in the final paragraph of the revision.

The last but not least, the authors avoid the phases labelling during ML program operation, however, they return back to this problem later in different formalism. I have doubts that the proposed phases classification scheme, which consider only total magnetization and chirality, is reliable. For instance, the determination of mixed phases boundaries using only net magnetization and chirality is highly subjective. For a more robust classification it is necessary to analyze structure factors and/or specific heat of the considered system. The presented ML program does not evaluate these quantities and one have to calculate them separately. It is desirable to discuss the criteria of mixed phases boundaries determination within this approach.

Reply: We agree that the previous manuscript left a strong impression that our work is yet another piece of attempt at the phase labeling problem. No, this work is not focused on that, and we try quite hard to make that clear in the revision. The earlier work arXiv:1803.06682 already addressed this problem partially. We were not fully content with the level of thoroughness in this earlier work, which is why we decided to construct a more careful ML architecture, use a far greater set of training data, and so on, to make sure that the problem of phase identification has been addressed adequately. To our surprise, it was not addressed adequately at all! So although the phase identification is no longer the main thrust of the paper, we find it in our scientific obligation to report our findings as part of the manuscript. Any remarks hinting that our paper is on the mixed-phase identification have been either removed or modified. We are now simply stating our observation in regard to the phase labeling as a novelty, worth bearing in mind to future practitioners who might be interested in the same issue in their own line of investigation.

4) And in the final section the authors tested the trained ML program on magnetic configurations produced by different Hamiltonian. The authors stated that Hamiltonian term H_K (Eq (6)) reflects “realistic

situation of material". Could the authors explain why in real materials the easy magnetization axes of impurities should be collinear to the z- axis? For a realistic modeling it is probably better to consider the case with randomly oriented easy axes of impurities.

Reply: Inspired by the referee comment, we introduce another model of disorder and make tests on both kinds of impurity Hamiltonians. The results are equally good with both models of disorder. Please see our revised manuscript for the reports of new findings.

In the fourth paragraph of page 4 it was stated that H(K,p) stays adiabatically connected to H_{HDMZ} as long as K remains sufficiently small compared to other energy scales. However, below authors have demonstrated that the error in evaluation is still relatively small even for values of K=2 (see Table I) which is the same order of energy scale in system (D is equal the square root of 6 ≈ 2.45). There is a natural question: is the adiabatic connection of Hamiltonians crucial for ML program operation?

Reply: We have removed most of the discussion on H(K,p) and the adiabatic connectivity. When ML works, it works, and whether it works only for the adiabatically connected set of Hamiltonians is an issue that we nor anyone else at this moment can address adequately. So we are focusing the reports of new investigations and leaving out grand remarks on the adiabatic connectivity.

Minor corrections:

1. Please, replace subscript index k of [I_k]_{iz} to 1 in Fig. 2 for clarity.

Reply: We do not find such symbols in Fig. 2. Perhaps the referee is referring to Fig. 1? That figure on PCA has been removed as we decided to delete the entire discussion on PCA.

2. In the introduction authors mentioned that "very little attention has been paid to the utility of ML in advancing the frontier of this field", however they cited about 15 works devoted to solving of physics problems by means of ML. Thus, the application of ML as itself to problems of physics is not a breakthrough as it was few years ago.

Reply: We fully agree. That's why our paper has been entirely revised to bring forth a new agenda of ML. The new agenda, as made clear in the revised manuscript, is the application of ML as a information recovery tool. I do hope the new agenda we are putting forth is made sufficiently clear in the writing.

3. SkSp phase mentioned in the first paragraph of page 3 is actually SkFm phase (box 4 in Fig. 2).

Reply: Corrected. Thank you.

4. Figure 4 does not have a blue curve denoted as “Ref.” in corresponding legend. It has black line instead which is not labeled in the legend.

Reply: We have removed fig. 4 in the revision process.

5. Missed Table number in the 4th paragraph of page 4.

Reply: The table has been removed.

6. The main text discusses SM material instead of referencing to it (the second paragraph of page 5). The Supplementary material in its turn has a lot of references to the main text and figures and contain important data for understanding of article. I suggest including this material in appendix or main text of the revised version of the manuscript.

Reply: With the removal of PCA part, we decided that SM is no longer necessary.

7. Supplementary material is full of broken and unclear references to figures from SM and from main text. The last paragraph of page 3 contains the sentence “The schematic diagram of the architecture can be found in the Supplementary Material (SM)” probably originated from the full-text version.

Reply: With the removal of PCA part, we decided that SM is no longer necessary.

8. The bibliography contains unused references ([24], [26]). The reference [15] is not correct.

Reply: All references are now cited. The reference number for [15] has been corrected.

Report of Referee B -- LT16311/Singh

The current manuscript presents a study on magnetic properties of thin film magnetic systems studied by machine learning tool. The topic is interesting for a broad readership but, in my opinion, the presented results do not provide a large advance in the field. Therefore, I

cannot recommend publication in Physical Review Letters.

Here are my questions and comments:

1. The title is in my opinion confusing, since the ML tools describe magnetic states in ultrathin films and not necessarily skyrmion properties.

Reply: We have adopted the new title "Machine learning application to two-dimensional Dzyaloshinskii-Moriya ferromagnets" in better reflection of the actual material covered in the paper.

2. In my opinion, many terms are not correctly introduced and the paper is not understandable for a broader audience. For example, I don't understand what PCA and the idea behind is and the connection to machine learning. What is a neural network and how is it used here? In my opinion, the authors should include more details or references to the paper.

Reply: We have decided to drop all PCA-related discussions for lack of relevance under the thrust of the new manuscript. Machine learning is a popular theme both in condensed matter physics and computer science community at large. Some new references are added in the bibliography for readers who wish to catch up on the basic vocabulary and ideas of machine learning.

3. The paper focused on the machine learning tools, but nonetheless, they should introduce the magnetic system they study with more details. What are skyrmion phases, spin spirals and so one? What determines in which phase the system is? This has been studied before and should be described here.

Reply: A more explicit and detailed definition of, and introduction to various phases of the HDMZ model is given in the new main text before delving into the ML application now.

4. Do I understand it right that the machine learning is used to predict the phase diagram for two-dimensional systems? This has been studied in many details by numerical simulations. How can the ML tools compete with the existing tools? How good can they predict the phase transitions? The Hamiltonian is rather simple, what happens if interactions beyond nearest neighbors are included? Are the tools able to include these states?

Reply: The ML program can be applied to models that are more complex than the simple model Hamiltonian we work with. At the end of the paper we do work with two variations on the model and find successful application of the machine learning

concept. We are confident that the general applicability of the machine learning is not limited to a specific model Hamiltonian. The two explicit examples we work with in the revised manuscript delivers the point, in our opinion. As the new manuscript makes clear, being able to predict spin chirality and magnetization based on faulty data is not a task that can be accomplished by any other apparatus we have available now. We hope the new manuscript makes it clear that ML offers a unique new opportunity to reconstruct some pieces of information one cannot acquire by existing toolbox.

5. At the end of the conclusions the authors discuss the possibility
to evaluate data from experiments. But to me, this seems a bit vague.
Temperature and magnetic field should be known for the experiments.
What is the advantage in using the presented tools?

Reply: Please see our reply to Referee A's comment on a related issue. With additional calculations on xy and z-component trained machine learning algorithm, we are very confident that the scheme we propose should have direct relevance to the experiment. Our reasoning and supporting evidences can be found in the revised manuscript.

With best regards,
Vinit Kumar Singh and Jung Hoon Han

Summary of changes:

1. Title of the paper has been renamed.
2. Discussion of PCA has been removed, first for lack of relevance to the main theme of the revised article, secondly to make room for additional calculations and arguments.
3. Figure 4 and the table has been removed. They are deemed less relevant than the additional materials we include as the new figure 4.
4. Figure 1 describing the main architecture of machine learning is moved from the supplementary section to the main text.
5. The overall tone of the article has been revamped to emphasize the new perspective and the close connection of our study to the experiments.
6. Supplementary Material is no longer deemed necessary, and removed.