Quantifying Data Difficulty with Polarized K-Entropy for Assessing Machine Learning Models

Ayomide Afolabi
School of Data Science and Analytics
Kennesaw State University
Kennesaw, GA USA
aafolab4@students.kennesaw.edu

Ramazan Aygun
Computer Science Department
Kennesaw State University
Kennesaw, GA USA
raygun@kennesaw.edu

Truong X. Tran
Penn State Harrisburg
The Pennsylvania State University
Middletown, PA USA
truong.tran@psu.edu

***Abstract*—Data difficulty level measurement is a critical aspect of machine learning performance evaluation. Several measures have been used to assess the difficulty level of classifying data points in binary classification. However, these measures typically involve building a machine learning model first, which is then used to assess the data difficulty level. In this paper, we propose a novel model agnostic measure named as *polarized K-entropy* to evaluate the difficulty of classifying a data instance. Our measure leverages the computation of entropy based on the nearest neighbors of a data point. We conducted experiments to evaluate the effectiveness of our proposed method by analyzing how the accuracy of machine learning models change with respect to data difficulty. We used Spearman's rank correlation coefficient to analyze this relationship for neural network, support vector machine, and random forest. Our results show that our measure outperformed the non-conformity measure in all the experiments conducted for six datasets using the selected machine learning models.**

***Index Terms*—Data difficulty, polarized K-entropy, non-conformity**

I. INTRODUCTION

Data utilized in training, validating, and evaluating machine learning models has garnered significant research interest in recent years due to its profound influence on model performance. When a machine learning model may perform as expected in a controlled environment while evaluating on numerous scenarios, it is a big question what causes the model to fail in a real-life deployment. The investigation and analysis of these data to determine why certain test data may not be correctly classified has become a crucial aspect of classification tasks [1].

Trustworthy machine learning has several vectors, such as explainability, fairness, avoiding biases, and safety, to gain the trust of users. It is important to determine the vulnerability or weaknesses of machine learning models to minimize errors in classification. In other words, understanding the knowledge limits of a machine learning model may help to improve overall performance rather than just blindly providing a prediction. To manage this, quantifiable measures need to be developed to know how some data are related to each other and its impact on the likelihood of correct or incorrect classification. Therefore, we delve into *data difficulty* to understand a specific data instance's treatment by machine learning models with respect to a training set. To improve the model, it is helpful to understand the difficulty level of the data and identify situations where the model is likely to perform poorly. Developing methods to enhance the model by refining the training set may prove beneficial in addressing these challenges.

There are methods, such as WisdomNet, that aim to minimize misclassification. WisdomNet [2] adds a rejection node to a pre-existing deep learning model to reject decisions on instances that are likely to be misclassified. WisdomNet also aims to maintain a low rejection rate while achieving high accuracy on unseen data. Based on posterior analysis, it has been observed that data points that are difficult to classify correctly can increase the rejection rate of these models. This is because the data to be classified may originate from a different distribution than its training set or contain outliers and noise that were not present during training. Therefore, determining the level of difficulty associated with classifying data will provide insight into the capacity of trustworthy machine learning models and play a vital role in ranking the performance of other machine learning classifiers [3].

While a variety of methods have been employed in previous research to assess the level of difficulty associated with data classification. Data difficulty could be measured with respect to a machine learning model or could be measured independently. Many of these techniques are machine learning model-based. These methods, such as computing entropy loss and employing the number of hidden layers in a deep learning model [4], utilizing margin-based approaches, and calculating silhouette scores, are contingent upon the construction of a machine learning model. While these methods could be helpful in assessing a specific dataset with respect to a pre-determined machine learning model, the data difficulty and model's ability to learn are interweaved together. A misclassification may result from data difficulty as well as the model's inability to learn. Conversely, a few procedures that are model agnostic have also been developed to evaluate data classification difficulty [5]. These techniques involve the application of non-conformity measures. However, it has been

2835-5776/24/$31.00 ©2024 IEEE
DOI 10.1109/IRI62200.2024.00015

observed that existing model agnostic measures are not highly effective in gauging data classification difficulty, necessitating the development of better model independent measures.

In this paper, we propose a measure called the *polarized K-entropy*, which computes the entropy based on the K-nearest neighbors. The purpose of this measure is to determine the level of difficulty associated with our test data in a classification task using either a tabular or image dataset. It is called polarized since the class of data to be analyzed should be considered for this entropy measure. The polarized K-entropy measure computes the entropy of K-nearest neighbors, and low entropy for a data instance in a different class indicates high difficulty, whereas a data instance in the same class indicates low difficulty. Additionally, comparing this measure with existing measures, such as the non-conformity of nearest neighbors, reveals its level of performance in assessing machine learning models.

In summary, our main contributions are provided as follows:

- We propose a novel polarized to measure the data points difficulty level and evaluate its performance on five datasets.
- We show that the proposed measure can easily assess how well our different machine learning classifiers will perform.

The remainder of this paper is organized as follows: The next section presents an in-depth analysis of the related work done in developing different data classification difficulty measures. Section III describes our polarized K-entropy measure and non-conformity measure. Section IV demonstrates the results of our experiments on our proposed measure. Finally, Section V concludes the work and provides relevant information for future work.

II. RELATED WORK

A variety of measures have been developed to assess data difficulty while training machine learning models and evaluating them. Some of these measures are based on the machine learning model's capability to discern different classes, whereas other measures are model agnostic. These measures can provide insight into individual instances as well as datasets.

In various studies, researchers have utilized model-based data difficulty measures at both instance and dataset levels. For example, Ho et al. [6], and Komornicza et al. [7] have utilized measures such as the error rate of a linear classifier, the non-linearity of a linear classifier, and the error rate of the nearest neighbor classifier. Rahane et al. [8] introduced model-based entropy measures, such as pixelwise Shannon entropy, GLCM feature Shannon entropy, and delentropy, to measure image classification difficulty. Baldock et al. [4] proposed neural network prediction depth as a measure to gauge data difficulty level, while Zhang et al. [9] proposed DIME, an information-theoretic difficulty measure for datasets based on Fano's inequality and a neural network estimation of the conditional entropy of the sample-label distribution. However, these model-based measures have some drawbacks. These measures are typically specific to a particular model and may not be applicable to assessing the performance of other machine learning models. Additionally, they may not be consistent across different datasets, and they may not be able to accurately distinguish the difficulty of datasets since the model's capability is part of the process. Furthermore, model-based dependent measures are less robust to changes in the underlying data or model assumptions, which means that they are more affected by variations in model complexity.

To address these limitations, some researchers have suggested model agnostic measures. Ho et al. [6] and Komornicza et al. [7]. introduced feature-based methods such as the maximum Fisher's discriminant ratio, volume of overlapping region, and maximum individual feature efficiency. Torquette et al. [10] introduced several instance hardness measures, such as tree depth, which provides the decision tree depth of a leaf node at which an instance is classified. Other measures include class likelihood difference, the fraction of nearby instances of different classes, the ratio of intra-class and extra-class distances, and k-Disagreeing Neighbors (kDN(x_i)). Shafer et al. [5] presented non-conformity measures for nearest neighbors, which measure how well a data point in a dataset deviates from the rest of the dataset. However, these model agnostic measures may not always correlate well in assessing model performance. This paper will focus on using polarized K-entropy on addressing limitations arising from both model based and existing model agnostic measures.

III. METHOD

In machine learning binary classification tasks, data points from a particular class that are close to or overlapping with the opposite class present a level of difficulty in accurately classifying those points. In this section, we introduce and describe our *polarized K-entropy* measure that quantifies data difficulty. While we provide our examples on binary classes, this measure could easily be extended for multiple classes. This measure helps assess the machine learning behavior with respect to data difficulty. Additionally, we briefly explain the non-conformity score [5] to compare and evaluate the effectiveness of our proposed method using Spearman's rank correlation coefficient.

There are two properties expected from a data difficulty measure.

Property 1. Correlation with Model Performance. The performance of models should decrease as the data difficulty level increases.

Property 2. Diversity of Difficulty. The measure should be able to distinguish data difficulty of different data by generating diverse scores.

A. *Polarized K-Entropy Measure*

Polarized means to assign either a positive or negative value to each test data entropy value based on its class. Our polarized K-entropy measure utilizes the nearest neighbors surrounding our data based on chosen k. The entropy of these nearest

Algorithm 1 Polarized K-Entropy Based Algorithm

Input: Test data R, Train data T, Number of neighbors K
Output: Polarized K-entropy list

Polarized K-entropy list $\leftarrow \emptyset$
for $d_r \in R$ **do**
  distances $\leftarrow \emptyset$
  for $d_t \in T$ **do**
    distance $\leftarrow$ compute_distance(d_r, d_t)
    distances.append(distance)
  end for
  K_nearest_indices $\leftarrow$ find_K_nearest_indices(distances, K)
  $Z_r \leftarrow$ extract_K_nearest_labels(T, k_nearest_indices)
  $E(Z_r) \leftarrow$ compute_entropy(Z_r)
  if $d_r \in C^+$ **then**
    if $p(C^+|Z_r) < 0.5$ **then**
      $P_r \leftarrow 1 - \frac{E(Z_r)}{2}$
    else
      $P_r \leftarrow \frac{E(Z_r)}{2}$
    end if
  else if $d_r \in C^-$ **then**
    if $p(C^-|Z_r) < 0.5$ **then**
      $P_r \leftarrow \frac{E(Z_r)}{2} - 1$
    else
      $P_r \leftarrow -\frac{E(Z_r)}{2}$
    end if
  end if
  Polarized K-entropy list.$append(P_r)$
end for
return Polarized K-entropy list

neighbors, which can be either from the positive or negative class, is measured as follows:

$$E(Z_r) = -\sum_{i=1}^{2} p_i \log_2 p_i \tag{1}$$

The notation $E(Z_r)$ denotes the entropy of K-nearest neighbor set Z_r for data instance d_r. The notation p_i denotes the proportion of class i in the K-nearest neighbors respectively. For simplicity, assume that two classes are positive (C^+) and negative (C^-). The polarization of the resulting entropy of a given test instance in Equation 1 is shown in Equation 2, and the polarized K-entropy is denoted as P_r for a data instance d_r.

$$P_r = \begin{cases} 1 - \dfrac{E(Z_r)}{2}, & \text{if } d_r \in C^+ \text{ and } p(C^+|Z_r) < 0.5 \\ \dfrac{E(Z_r)}{2}, & \text{if } d_r \in C^+ \text{ and } p(C^+|Z_r) \geq 0.5 \\ \dfrac{E(Z_r)}{2} - 1, & \text{if } d_r \in C^- \text{ and } p(C^-|Z_r) < 0.5 \\ -\dfrac{E(Z_r)}{2}, & \text{if } d_r \in C^- \text{ and } p(C^-|Z_r) \geq 0.5 \end{cases} \tag{2}$$

The value derived from this measure spans between -1 to 1. For positive class data, the values fall within the range of 0 to 1, while negative class data values fall within the range of -1 to 0. As P_r gets farther from 0, it is harder to classify data. As an extreme point, $P_r = 1$ indicates the data of interest is in the positive class and very hard to classify. Similarly, $P_r = -1$ indicates that the data of interest is in the negative class and very hard to classify. Hence, the data difficulty can be computed as $|P_r|$.

Algorithm 1 provides a pseudo-code of our methodology. Initially, the distance of the test instance from our training data instances is calculated and arranged in descending order based on the top K-nearest distances. The corresponding top K-nearest neighbors are then returned. The entropy of the nearest neighbors associated with each test instance is determined, and this entropy is utilized to calculate polarized K-entropy, taking into account the class of the data and the probability of that class in its nearest neighbors.

B. Non-conformity Score Measure

The efficacy of our proposed method is evaluated by comparing it with a non-conformity measure [5], which is the ratio of the distance to the nearest instance of the same class to the distance to the nearest instance of the different class. The non-conformity values are measured as follows:

$$\alpha_r = \frac{min\{|d_r - T| : 1 \leq r \leq n-1, C_r = C\}}{min\{|d_r - T| : 1 \leq r \leq n-1, C_r \neq C\}} \tag{3}$$

The notations d_r, T, C_r, and C denote data instance, train data, data instance class, and nearest instance class, respectively. Also, the non-conformity measure, denoted by α_r, gauges the extent to which an instance in a dataset deviates from the rest of the dataset. Specifically, a larger α_r value indicates a greater degree of dissimilarity between the test instance and the rest of the data points. To ensure the non-conformity values fall within the range of 0 to 1, the fraction of instances in the dataset, p_r, which have their non-conconformity scores less than that of our data instance, d_r, is computed as follows:

$$p_r = \frac{\text{number of j such that } 1 \leq j \leq n \text{ and } \alpha_j \leq \alpha_r}{n} \tag{4}$$

C. Spearman's Rank Correlation Coefficient

Spearman's rank correlation coefficient [11] is a nonparametric measure utilized to evaluate the degree to which the relationship between two variables can be represented by a monotonic function. This is used to summarise the direction and strength of their relationship. We assume that accuracy and the data difficulty level, which is an ordinal data exhibit a monotonic relationship, and thus, Spearman's rank correlation coefficient can be employed to assess the correlation between the two variables. For a sample size of n, the ranks of X and Y, denoted as x_i and y_i, respectively, the Spearman's coefficient ρ is computed as follows:

$$\rho = 1 - \frac{6 * \sum (x_i - y_i)^2}{n(n^2 - 1)} \tag{5}$$

The Spearman's rank correlation coefficient ranges from -1 to 1. A coefficient of +1 and -1 signifies a perfect positive and negative monotonic relationship, respectively, while a coefficient of 0 indicates the absence of a monotonic relationship.

IV. EXPERIMENTS

In this section, we provide a brief description of the dataset used in our experiments and categorize our data into different data difficulty levels based on their polarized K-entropy and non-conformity values. To evaluate the performance of our measure, we use neural networks, support vector machines, and random forest classifiers. We then assess the data difficulty level from our measures with corresponding test accuracy using Spearman's rank correlation coefficient. Our experiments primarily utilized Python 3, and all machine learning models were trained using TensorFlow and Scikit Learn.

A. Datasets

The following binary datasets, including Dry beans [12], MNIST [13], CIFAR10 [14] (excluding Banana quality, Raisin [15], and Phoneme), were created by selecting two classes of interest. A summary of these datasets showing the number of attributes, train, validation, and test count is provided in Table I.

TABLE I: Data Description

Data	Attributes	Train	Val	Test
Drybeans(Sira-Dermason)	17	4945	618	619
Banana quality	8	6400	800	800
Raisin	8	720	90	90
Phoneme	6	4323	540	541
MNIST(2-7)	[28-28-1]	10991	1232	2060
CIFAR10(Cats-Dogs)	[32-32-3]	8962	1038	2000

B. Model Agnostic Measure and Classifier Accuracy

The quantification of the data difficulty level for each test instance is assessed using polarized K-entropy and non-conformity measures, as comprehensively detailed in the method section. For our experiment, the absolute values of the polarized K-entropy values is taken to ensure they fall within the range of 0 to 1. These test data's absolute polarized K-entropy and non-conformity values are sorted in ascending order and categorized into specific polarized K-entropy or non-conformity ranges.

Table II displays that 10.95% of the test data has an absolute polarized K-entropy value ranging from [0 - 0.40], which corresponds to the polarized K-entropy data difficulty level 1. Level 1 consists of the easiest data to classify. We used a range for setting levels since as the difficulty level increases, less and less data belongs to difficult levels. This yields fluctuations in accuracy computation since even one correct classification in a small set (at the corresponding difficulty level) has a big impact on accuracy. In this case, the range is set to 0.4 to contain a good number of samples per level. The range of each polarized K-entropy level is shifted forward on both the lower and upper bound by parameter value of 0.01, till it reaches the range of [0.6 - 1.00]. This forward shift is intended to gradually increase with the data difficulty level, as we expect that test data with very high polarized K-entropy values should be difficult to classify. Furthermore, the accuracy for 10.95% of the test data in data difficulty level 1 is 99.85%.

TABLE II: MNIST(2-7) Difficulty Level Quantification with Polarized K-Entropy Measure

Level	Polarized K-Entropy Range	Test data(%)	NN Accuracy
1	[0.00 - 0.40]	10.95	0.9985
2	[0.01 - 0.41]	0.26	0.9388
...	...	...	...
61	[0.60 - 1.00]	0.05	0

In a similar manner, a starting range of [0.0 - 0.1] and a parameter value of 0.01 for the forward shift is selected for the non-conformity measure. Table III shows that 1.04% for the test data falls within that range, and its test accuracy by the neural network is 100%.

TABLE III: MNIST(2-7) Difficulty Level Quantification with Non-conformity Measure

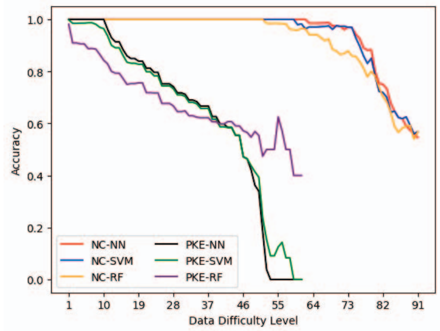
Level	Non-conformity Range	Test data(%)	NN Accuracy
1	[0.00 - 0.10]	1.04	1.0000
2	[0.01 - 0.11]	1.00	1.0000
...	...	...	...
91	[0.90 - 1.00]	1.17	0.9358

We should note that since we use a sliding range, the difficulty levels are overlapping. We made sure that both non-conformity and polarized K-entropy had enough samples in each level. To visualize how data difficulty levels for both the polarized K-entropy values and non-conformity values vary with respect to different machine learning algorithms, a plot of accuracy with the data difficulty levels for both polarized K-entropy value and non-conformity values is depicted for all the datasets in Fig. 1 (a)-(d) and Fig. 2 (a)-(b).

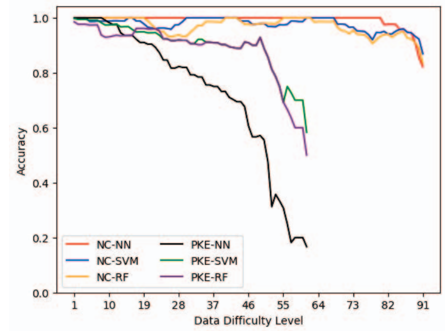
Neural networks (NN), support vector machines (SVMs), and random forest(RF) algorithms are utilized to gauge the accuracy of test points within a specific data difficulty level. For fairness, feature sets obtained from neural networks are used in our experiments for all the models. The trained and fine-tuned models used train and validation data are depicted in Table I. Finally, the tuned model is employed to evaluate the accuracy.

C. Further Analysis and Discussion

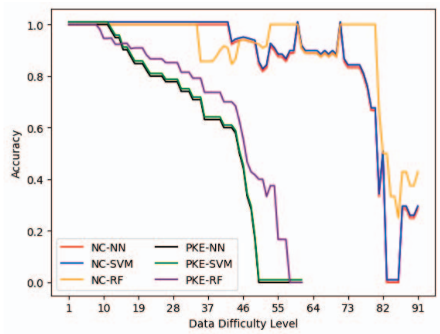
Spearman's rank correlation coefficient is able to assess model performance accuracy as the data difficulty levels increase, as our data difficulty level is ordinal. Table IV presents Spearman's rank correlation coefficient results for both difficulty measures and their respective accuracy based on neural network, support vector machine, and random forest classifier for various datasets. For the MNIST dataset, the correlation between the accuracy and ranked data difficulty level as measured by our polarized K-entropy measure and the accuracy for the neural network classifier is -0.9934. This measure is almost a perfect negative value and indicates that it is highly effective in assessing the decrease in performance of the neural network classifier as the difficulty data level increases. Therefore, it is capable of discerning differences



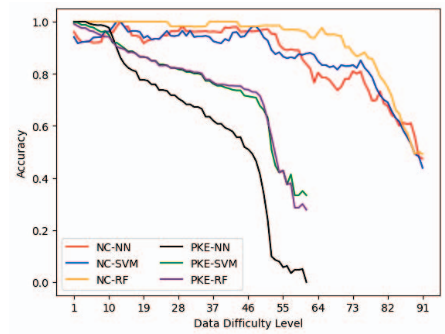
(a) Drybeans (Sira-Dermason)



(b) Banana quality

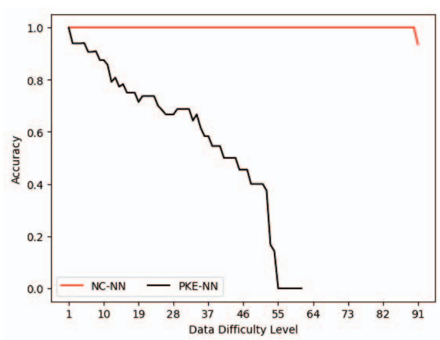


(c) Raisin

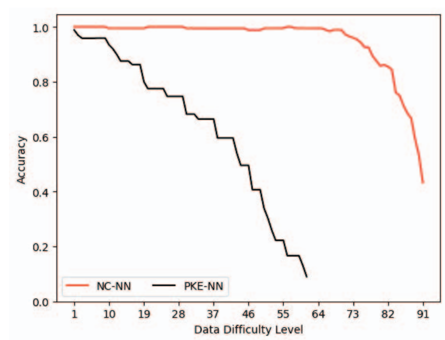


(d) Phoneme

Fig. 1: Comparison of data difficulty level assessment of tabular dataset accuracy between polarized K-entropy(PKE) and Non-conformity(NC) measure for Neural Network(NN), Support Vector Machine(SVM) and Random Forest(RF) Models



(a) MNIST (2-7)



(b) CIFAR10 (Cats-Dogs)

Fig. 2: Comparison of data difficulty level assessment of image dataset accuracy between polarized K-entropy(PKE) and Non-conformity(NC) measure for Neural Network(NN) Model

TABLE IV: Spearman's Rank Correlation Coefficient Results

	Spearman's Rank Correlation Coefficient					
	Polarized K-Entropy Measure			Non-conformity Score Measure		
Data sets	NN	SVM	RF	NN	SVM	RF
Drybeans(Sira-Dermason)	-0.9960	-0.9975	-0.9805	-0.8262	-0.8489	-0.9068
Banana quality	-0.9970	-0.9641	-0.9334	-0.5660	-0.5229	-0.5934
Raisin	-0.9925	-0.9925	-0.9968	-0.8965	-0.8965	-0.5664
Phoneme	-0.9992	-0.9993	-0.9993	-0.7099	-0.7099	-0.9174
MNIST (2-7)	-0.9934	-	-	-0.1805	-	-
CIFAR10 (Cats-Dogs)	-0.9961	-	-	-0.7888	-	-

in data difficulty levels. In contrast, non-conformity measure has a value of -0.1805, which suggests that it is not effective in assessing the decrease in performance of our neural network classifier as the data difficulty level increases.

In addition, Table IV shows that our polarized K-entropy measure correlation coefficient value for the banana quality datasets are -0.9970, -0.9641, and -0.9334 for the neural network, support vector machine, and random forest classifiers, respectively. Meanwhile, the correlation coefficient values for the non-conformity measure are -0.5660, -0.5229, and -0.5934 for the same classifiers. Also, polarized K-entropy measures have correlation coefficient values of -0.9992, -0.9993, and -0.9993 for neural networks, support vector machines, and random forest classifiers, respectively. These values represent a better negative correlation coefficient than -0.7099, -0.7099, and -0.9174 of non-conformity measures for the same classifiers. These findings indicate that regardless of the classifier or dataset used, the polarized K-entropy measure is more effective than non-conformity measures in assessing the decrease in model performance as the data difficulty level increases. Therefore, it distinguishes data difficulty better.

Our graph of accuracy and data difficulty levels for both difficulty measures from our experiments also gives a representation of the better measure. In Fig. 2 (a) and (b), the accuracy of MNIST(2-7) and CIFAR10(Cats-Dogs) is very close to 1 up to data difficulty levels of 90 and 73 for the non-conformity measure, respectively. So it is not able to data distinguish difficulty levels. On the other hand, accuracy starts dropping proportionally with the polarized K-entropy measure in Fig. 2 (a) and (b). For different classifiers, Fig. 1(a) shows that accuracy only starts dropping below 1 around 64, 63, and 55 for neural networks, support vector machines, and random forests, respectively, for the non-conformity measure. On the other hand, for our polarized K-entropy measure, the accuracy starts dropping almost proportionally up to data difficulty level 45 before it experiences a steep drop for the neural network and support vector machine classifier with the polarized K-entropy measure. A similar pattern also occurs in Fig. 1 (b), (c), and (d). Therefore, these graphic illustrations clearly show that the performance of machine learning models decreases as the difficulty level increases for our measure. Moreover, the polarized K-entropy measure is able to distinguish the difficulty of different data irrespective of the classifier.

The polarized K-entropy measure may be subject to certain limitations. The distance of the nearest neighbors of opposite classes to a data instance of interest could potentially influence the polarized K-entropy, and the algorithm may be susceptible to outliers if the hyperparameter K is selected as a small number.

V. CONCLUSION

In this paper, we introduce a novel polarized K-entropy measure for determining the difficulty level of data when evaluating machine learning models. This methodology involves computing the entropy of the nearest neighbors and polarizing the entropy. We further compared this approach with the non-conformity score by employing Spearman's rank correlation coefficient to gauge the extent to which the relationship between our classifier accuracy and data difficulty level, as described by the polarized K-entropy level, and non-conformity level. Our experiments demonstrate that the polarized K-entropy method is more effective than the non-conformity score in quantifying data difficulty for evaluating machine learning models. It is important to recognize that the performance of machine learning models in terms of data difficulty can serve as an indicator of their limitations and provide a way to improve them. We intend to expand this approach to multi-class classification in the future, which will enable us to encompass all datasets with multiple classes. Furthermore, this extension will prove beneficial in addressing the challenges associated with multi-class classification.

REFERENCES

[1] F. Östlund and E. Fahlman, "Data complexity and its effect on classification accuracy in multi class classification problems: A study using synthetic datasets," 2022.

[2] T. X. Tran and R. S. Aygun, "Wisdomnet: trustable machine learning toward error-free classification," *Neural Computing and Applications*, vol. 33, no. 7, pp. 2719–2734, 2021.

[3] V. V. de Melo and A. C. Lorena, "Using complexity measures to evolve synthetic classification datasets," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2018.

[4] R. Baldock, H. Maennel, and B. Neyshabur, "Deep learning through the lens of example difficulty," *Advances in Neural Information Processing Systems*, vol. 34, pp. 10876–10889, 2021.

[5] G. Shafer and V. Vovk, "A tutorial on conformal prediction.," *Journal of Machine Learning Research*, vol. 9, no. 3, 2008.

[6] T. K. Ho and M. Basu, "Complexity measures of supervised classification problems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 3, pp. 289–300, 2002.

[7] J. Komorniczak, P. Ksieniewicz, and M. Woźniak, "Data complexity and classification accuracy correlation in oversampling algorithms," in *Fourth International Workshop on Learning with Imbalanced Domains: Theory and Applications*, pp. 175–186, PMLR, 2022.

[8] A. A. Rahane and A. Subramanian, "Measures of complexity for large scale image datasets," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 282–287, IEEE, 2020.

[9] P. Zhang, H. Wang, N. Naik, C. Xiong, *et al.*, "Dime: An information-theoretic difficulty measure for ai datasets," in *NeurIPS 2020 Workshop: Deep Learning through Information Geometry*, 2020.

[10] G. P. Torquette, V. S. Nunes, P. Y. Paiva, and A. C. Lorena, "Instance hardness measures for classification and regression problems," *Journal of Information and Data Management*, vol. 15, no. 1, pp. 152–174, 2024.

[11] T. D. Gauthier, "Detecting trends using spearman's rank correlation coefficient," *Environmental forensics*, vol. 2, no. 4, pp. 359–362, 2001.

[12] "Dry Bean." UCI Machine Learning Repository, 2020. DOI: https://doi.org/10.24432/C50S4B.

[13] L. Deng, "The mnist database of handwritten digit images for machine learning research [best of the web]," *IEEE signal processing magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[14] K. Alex, "Learning multiple layers of features from tiny images," *https://www. cs. toronto. edu/kriz/learning-features-2009-TR. pdf*, 2009.

[15] "Raisin." UCI Machine Learning Repository, 2023. DOI: https://doi.org/10.24432/C5660T.