Combinatorial Methods of Feature Selection for Cell Image Classification

Sergiy Vilkomir, Jiabin Wang, Nam Le Thai, Junhua Ding
Department of Computer Science
East Carolina University
Greenville, NC, USA
vilkomirs@ecu.edu, {wangj15, thain11}@students.ecu.edu, dingj@ecu.edu

***Abstract*—Feature selection is a major task in machine learning for selecting the most impactful features that will result in better accuracy and/or performance. One of the main approaches to this task is experiment-based feature selection. However, this approach is infeasible in many cases due to the large number of experiments needed in order to effectively select the optimized feature set. In this paper, we propose the use of an experiment-based feature selection guided by combinatorial methods such as t-way coverage. This approach can significantly reduce the required number of experiments in a study while selecting the best feature set to achieve a high level of accuracy.**

We conducted a case study of the feature selection process for a Support Vector Machine (SVM) classification of biomedical images for cell typing. Three thousand labeled images were used in the experimental SVM classification, and 32 features were evaluated for each image. We considered feature sets of different sizes from 6 to 32, and for each size, we generated t-way combinations of features for t from 2 to 6. The accuracy of each combinatorial combination was evaluated, which allowed us to select the optimal set of features for use in the classification process. The experiment results show that the proposed approach is an effective way to conduct feature selection and that it can be easily adapted for feature selection in other machine learning algorithms.

***Keywords*-combinatorial method; machine learning; support vector machine; diffraction image; GLCM; feature selection**

I. Introduction

Noise data such as invalid data and incorrectly labeled data can decrease the accuracy of data analytics [1]. One solution is to filter out the noise data using machine learning approaches to reduce their impact [2]. In machine learning, attributes, associated to items or instances of data, are known as *features* [3]. Feature selection is a major task in machine learning for selecting the most impactful features that will result in better accuracy and/or performance. One of the main approaches to this task is experiment-based feature selection. However, this approach is infeasible in many cases due to the large number of experiments needed in order to effectively select the optimized feature set. In this paper, we propose the use of an experiment-based feature selection guided by combinatorial methods [4] such as t-way coverage. This approach can significantly reduce the required number of experiments in a study while selecting the optimal feature set to achieve a high level of accuracy.

The feature selection task is important, in particular, for image classification. We investigate an approach for automated separation of noise data from massive-scale biomedical image data using a Support Vector Machine (SVM) [5]. An SVM classifier was designed for automatically classifying the regular image data and noise data into different categories. The experiment was conducted on diffraction images of biological cells of different types. The diffraction images of a cell were acquired using a polarization diffraction imaging flow cytometer (p-DIFC) for quantifying and profiling 3D morphology of cells [6]. The 3D morphological features of a cell captured in the diffraction images were used for classifying cell types.

Three thousand labeled images were used in the experimental SVM classification, and 32 features were evaluated for each image. We considered feature sets of different sizes from 6 to 32, and for each size, we generated t-way combinations of features for t from 2 to 6. The accuracy of each combinatorial combination was evaluated, which allowed us to select the optimal set of features for use in the classification process. The experiment results show that the proposed approach is an effective way to conduct feature selection and that it can be easily adapted for feature selection in other machine learning algorithms.

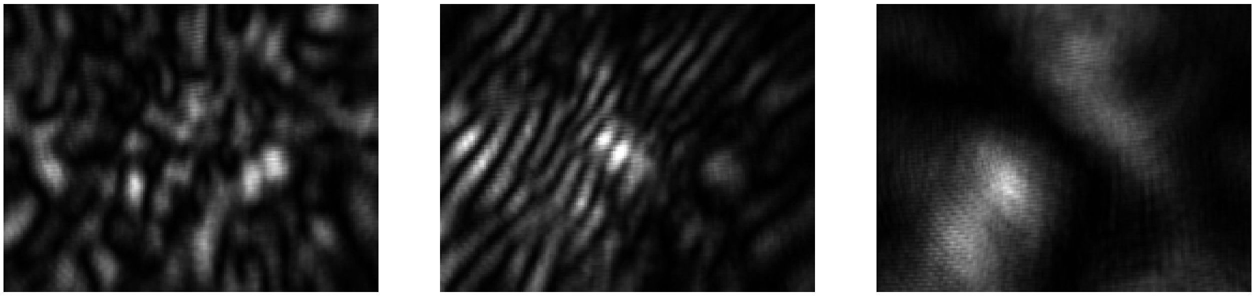
The rest of this paper is organized as follows: Section 2 introduces the background of this research including cell imaging, features of diffraction images, and SVM based automated image classification. Section 3 considers an application of combinatorial methods for feature selection. We provide a brief explanation of the combinatorial approach and describe used data sets and the organization of our experimental evaluation. The analysis of the numerical experimental results is given in Section 4. Section 5 discusses the related work and Section 6 concludes the paper.

II. Image Classification

A. *Types of cell images*

Cells have intensely varied and convoluted 3-dimensional (3D) structures by intracellular organelles. The morphology information can be captured in the diffraction images for cell classification. In our approach, diffraction images are classified into three categories. The sample diffraction image of each category is shown in Fig. 1.

These classification categories are diffraction images of viable cells with intact structures (*cells*), diffraction images of ghost cell bodies or aggregated spherical particles (*fractured*



(a) (b) (c)

Fig. 1. Sample p-DIFC acquired diffraction image of (a) cell, (b) fractured cell, and (c) debris.

cells), and diffraction images of cell debris or small particles (*debris*). The two latter two categories of the images are considered as noise data. In most cases, cell samples for p-DIFC imaging include such data, which should be recognized and filtered out.

B. Image Features and SVM based Image Classification

Image features are calculated from the Gray-Level Co-Occurrence Matrix (GLCM) [7], which reflects the frequency of different combinations of gray level pixels occur in an image. In our experimental study, the set {CON, VAR, IDM, SVA, DVA, DIS, ASM, SEN, ENT, DEN, CLS, CLP, MAP, COR, SAV, MEA} of 16 different features was used. The definitions and explanations of these features can be found in [8]. Since cytometer p-DIFC takes a pair of diffraction images in s-polarization and p-polarization and the same features are used for s- and p-images, the total number of features for each image pair becomes 32.

GLCM is calculated based on a given pixel distance d to its neighbor at a particular angle θ. In our study, images are normalized to an 8-bit gray-level range, d =1, and calculations are in average from 4 different angles $\{0^o, 45^o, 90^o, 135^o\}$. A parallel program using NVIDIAs CUDA [9] on GPUs was developed for calculating GLCM and the features. The paired images and their features are used together for the SVM based classication.

SVM builds a set of hyperplanes in a high-dimensional space through analyzing data for classification or regression analysis [5]. LIBSVM [10], a widely used open source toolkit for SVM is used to support SVM classification in this project.

III. Application of Combinatorial Methods for Feature Selection

A. Combinatorial t-way Approach

Combinatorial methods are based on the mathematical theory of the orthogonal arrays and are widely used in such different areas as software testing [4], design of experiments, statistics, cryptography, and many others. The most general t-way methods requires that for any t from total k parameters (variables, components, options) all possible t-tuples (t-way combinations) of the values of these parameters should be covered (tested, considered, used). To use the same terminology in different areas, these parameters are often called *factors* and their values *levels*. Combinatorial t-way method is usually used for t from 2 to 6 where the most common method is 2-way or *pairwise*.

In this paper we apply combinatorial methods for feature selection during a cell image classification process. We divided all 32 features into several groups and consider each group as a factor and each specific feature in the group as a level. The number of factors varies from 6 to 32 and, for each factor t-way (t=2...6), combinations of features are generated. The classification process is repeated for each combination and the accuracy is evaluated in order to find a combination of features which provides the highest level of accuracy.

To generate t-way combinations of features, we used Automated Combinatorial Testing for Software (ACTS) tool [11], developed by the National Institute of Standards and Technology (NIST) and the University of Texas at Arlington. The tool has a convenient easy-to-use interface and can generate combinations of values for 2-way through 6-way.

B. Organization of Experimental Evaluation

Our experimental investigation includes five major steps (Fig. 2).

The first step is preliminary and is conducted ones at the beginning of the investigation. Its aim is to prepare and provide necessary information for feature matrix building. The next four steps are the main part of the experiment and are repeated as a cycle for different combinations of features. Their aim is to detect the best combination of features to obtain the highest level of the classification accuracy.

More precisely, these five steps are:

- Step 1. Labelling of 3000 paired images (6000 images total) according to the three different types mentioned in Section 2. Calculation of the feature values for each image based on GLCM.
- Step 2. Generation of t-way feature combinations of different sizes and for different t using the ACTS tool.
- Step 3. Building a feature matrix to allow classification via the SVM tool.
- Step 4. Creation of the SVM classifier using training and testing data.

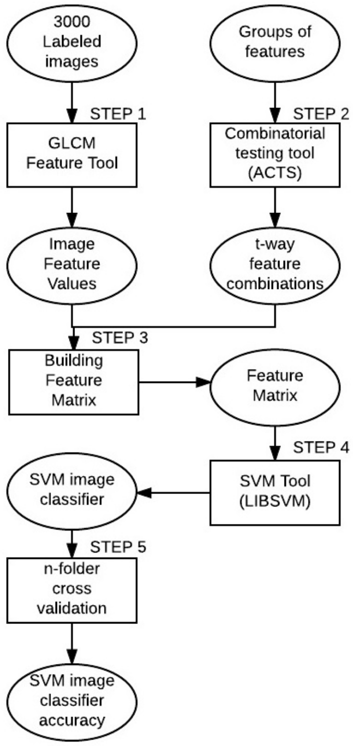


Fig. 2. Experiment Organization

- Step 5. Calculation of the classifier accuracy with 10 folder cross validation.

The obtained SVM classifier with the highest accuracy can be used in the classification process for any other cell diffraction images.

IV. Experimental Results

To generate combinatorial combinations of features of different sizes, we divided step by step the set of all 32 features into different subsets. For each combinatorial combination, one feature from each subset was taken. It means that for each step the size of combinatorial combinations coincides with the number of feature subsets.

As an initial step, both s- and p-polarized features were divided into three subgroups: contrast measures {CON, VAR, IDM, SVA, DVA, DIS}, uniformity or orderliness {ASM, SEN, ENT, DEN, CLS, CLP, MAP}, and correlation and other descriptive statistics {COR, SAV, MEA}. Totally, six subgroups were created at this initial step that allowed us to generate combinatorial combinations of size 6.

At the next step, the uniformity s- and p-subgroups were divided into two subgroups each, {ASM, CLS, CLP, MAP} and {SEN, ENT, DEN}, and combinatorial combinations of size 10 were generated. At each next step the next subgroup was divided and the size of combinations was increased by 2. The aim was to keep the correlated features in the same subgroup and separate independent features. At the last step, all features were divided in 32 subgroups with one feature in each subgroup. It gives us one combination, which includes all features.

TABLE I
NUMBER OF t-WAY COMBINATIONS OF FEATURES.

	Number of t-way combinations				
Number of Features	2-way	3-way	4-way	5-way	6-way
6	52	334	1783	5736	15876
8	36	161	708	2575	8008
10	22	93	368	1264	3802
12	17	67	225	567	1754
14	14	39	119	328	808
16	12	31	91	205	476
18	10	25	54	127	272
20	10	22	47	115	230
22	10	20	44	93	178
24	8	18	38	73	120
26	8	12	28	32	64
28	6	8	16	-	-
30	4	-	-	-	-
32	1	-	-	-	-

For each size of feature combinations, five combinatorial t-way sets (t=2...6) were generated. The number of combinations in each combinatorial se for all sizes and all values of t is presented in Table I.

As it is shown in Table I, 60 combinatorial sets of feature combinations were generated for different numbers of features and different values of t. In each such set, accuracy was calculated for each combination in the set. The maximum, minimum, and average values of accuracy were also detected for each of 60 sets. To evaluate stability of accuracy among different combinations, the standard deviation and the coefficient of variation (the ratio of standard deviation to mean) were calculated. The example of obtained data for one set of size 10 and t=2 (pairwise) is presented in Table II.

The values of the standard deviation and the coefficient of variation in this example confirms stability (low variability relative to the mean) of accuracy values. The difference between maximum and minimum values (range) of accuracy is less than 5% that also confirm stability of accuracy for different pairwise combinations of size 10. This situation is also typical for combinatorial sets for other sizes and values of t.

The results of the accuracy for all sizes of feature combinations and all values of t are presented in Table III. The maximum value of accuracy 72.23% was reached for the combination of 18 features. This result was obtained when 5- and 6-way sets of features were evaluated.

The analysis of data from Table III allows us making the following conclusions:

- For each size of feature combinations, the maximum level of the accuracy was found while evaluating 5- and 6-way sets of features. This result was expected because sets of combinations for t=5...6 have significantly more combinations than for t=2...4.

TABLE II

EXAMPLE OF EXPERIMENTAL DATA FOR PAIRWISE COMBINATIONS OF 10 FEATURES.

Index of Combination	S-Polarized Features					P-Polarized Features					Accuracy, %
	1-ASM 13-CLS 14-CLP 16-MAP	7-SEN 9-ENT 10-DEN	3-COR 6-SAV 17-MEA	2-CON 5-IDM 11-DVA 12-DIS	4-VAR 8-SVA	21-ASM 33-CLS 34-CLP 36-MAP	27-SEN 29-ENT 30-DEN	23-COR 26-SAV 37-MEA	22-CON 25-IDM 31-DVA 32-DIS	24-VAR 28-SVA	
1	1	9	6	5	8	21	29	26	25	28	67.36
2	1	10	17	11	4	33	30	37	31	24	68.06
3	1	7	3	12	8	34	27	23	32	24	65.26
4	1	9	17	2	4	36	27	26	22	28	65.73
5	13	10	3	11	4	21	29	37	32	28	65.90
6	13	7	6	12	8	33	30	23	22	28	66.53
7	13	10	17	2	8	34	27	37	25	24	68.00
8	13	7	3	5	4	36	29	23	31	24	67.90
9	14	9	17	12	8	21	30	26	31	24	67.26
10	14	7	6	2	4	33	27	37	32	28	66.06
11	14	10	3	5	4	34	29	26	22	24	66.26
12	14	7	6	11	8	36	30	23	25	24	68.00
13	16	7	17	2	4	21	27	23	31	28	67.73
14	16	9	3	5	8	33	30	26	32	24	69.03
15	16	9	6	11	8	34	29	37	22	28	65.20
16	16	10	3	12	4	36	29	37	25	28	68.36
17	13	9	3	2	4	21	30	23	22	28	66.70
18	14	10	6	2	4	33	29	23	25	28	68.53
19	13	7	6	2	8	34	30	26	31	28	65.96
20	13	10	17	5	4	36	27	37	32	28	67.73
21	16	10	17	12	4	36	29	37	22	24	67.60
22	1	9	3	11	4	33	27	26	22	28	64.23
Maximum of accuracy											69.03
Minimum of accuracy											64.23
Range of accuracy											4.80
Average accuracy											66.97
Standard Deviation											1.23
Coefficient of variation											0.018

TABLE III

ACCURACY OF t-WAY COMBINATIONS OF FEATURES

Number of Features	Accuracy of t-way combinations of features, %					Max Accuracy, %
	2-way	3-way	4-way	5-way	6-way	
6	67.16	67.13	67.23	67.47	67.50	67.50
8	67.73	68.53	69.10	69.10	69.06	69.10
10	69.03	69.90	70.33	70.63	70.60	70.63
12	69.40	70.60	70.60	70.83	70.96	70.96
14	70.60	70.96	71.06	71.13	71.53	71.53
16	70.83	71.23	71.50	71.53	71.60	71.60
18	71.63	71.66	71.90	72.23	72.23	72.23
20	71.76	71.96	71.86	72.03	71.96	72.03
22	71.73	71.83	71.96	72.00	72.00	72.00
24	71.90	71.83	71.90	71.86	71.90	71.90
26	71.90	72.00	72.00	72.00	72.00	72.00
28	71.93	71.93	71.93	-	-	71.93
30	71.10	-	-	-	-	71.10
32	71.06	-	-	-	-	71.06

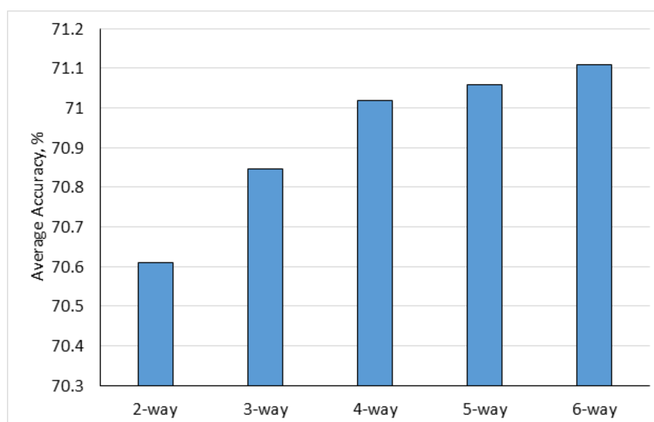


Fig. 3. Average Accuracy of T-way Combinations of Features

- The difference between the accuracy for the same size but different t is not significant. Indeed, the average values of the accuracy are presented in Fig. 3. While the level of the accuracy increases when t rises, the difference between values for t=2 and t=6 is only around 0.5%. It means that if it is sufficient to obtain the level of the accuracy close to optimal in the range of 0.5-1% (but not optimal), only pairwise combinations can be used. This can significantly decrease the number of experimental evaluations.
- For small values of used features, the accuracy increases when the size rises. Thus, the accuracy for 8 features is 1.6% higher than for 6 features, etc.
- However, it is not true for large values of used features. It is not necessary to use all feature to obtain the optimal accuracy. Opposite, at one moment adding a new feature just decreases the accuracy. This fact is illustrated at Fig. 4, where the maximum levels are reached for size 28

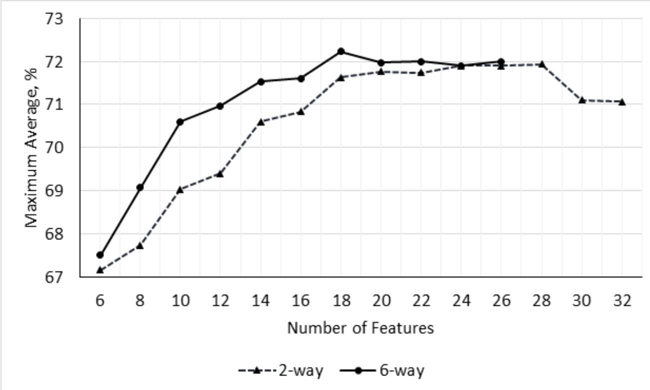


Fig. 4. Accuracy of 2-way and 6-way combinations with differnt sizes of features

(t=2) and size 18 (t=6). The behavior for t=3...5 is similar with maximum values of accuracy reached for sizes 26 (t=3 and t=4) and 18 (t=5).

It is necessary to mention that it is not necessary to evaluate all t-way combinations from Table I to find the optimal combination. Thus, we first evaluated pairwise combinations for all sizes to approximately determine the interval where the optimum can be reached (sizes 18-28 in our case). Then 5- and 6-way combinations were evaluated only for this interval of sizes that allowed us to select the optimal combination of 18 features with the maximum value of accuracy of 72.23%.

Later we evaluated all other combinations of features including 5- and 6-way combinations for other sizes. However, we conducted these experiments only for research goals to completely investigate the approach (reflected in Table III). Evaluation of the large numbers of combinations (for example, 5- and 6-way combinations for sizes 6-16) was not necessary and only a small number of feature combinations is required to be evaluated to find the optimal combination. This is the main benefit of our approach that makes its practical application very promising.

V. Related Work

The related to our investigation works belong to two different areas: combinatorial methods and feature selection in image classification and, more generally, in machine learning.

Combinatorial methods (t-way coverage), which we used in this investigation, are well-known in software testing. Detailed information can be found in book [4], which considers the theoretical basis of these approaches and their applications. The list of references contains 232 publications that make [4] a comprehensive survey in this area. Other surveys on combinatorial testing methods are [12]–[15]. One of the main methods for generation of t-way combinations is the In-Parameter-Order-General (IPOG) algorithm [16]. This algorithm is implemented in the ACTS tool [11], which we used in our work.

Feature selection approaches can be classified into two groups: wrapper approaches and filter approaches. Wrappers use the machine learning models to rank feature sets, and filters use some criteria such as the correlation with the value to be learned, which is independent to the machine learning algorithms, to evaluate feature sets [17]. The two approaches are used together in many cases. Thati *et al.* [8] reported their work on feature selection for classifying diffraction images, where an experiment approach called Extensive Feature Correction Study (EFCs) was used to select optimized GLCM features for classifying diffraction images. The selected feature set was cross checked with an algorithm based feature selection approach called Correlation based Feature Selection Algorithm (CFS). The selected feature sets were evaluated based on the classification accuracy. Wang [18] has recently conducted a feature selection experiment study for an SVM based diffraction image classification using a greedy method. In the experiment study, feature subsets were evaluated in forward and in backward through adding or removing one feature each time for the highest classification accuracy. However, it is not necessary true that a feature subset with higher classification accuracy combined with a feature will result in better classification accuracy than the other feature subset combined with the same feature. Greedy method only evaluates a small portion of the all possible feature combinations.

Feature selection guided by combinatorial method has been reported in the several research works. Dreisetitl *et al.* [17] proposed and experimented a feature selection based on the classification performance of pairwise feature sets. Silva and Fred [19], Li and Oh [20] have done the similar work. In these works, only pairwise features were fully evaluated and ranked. The pairwise feature ranks serve the basis for further selections, which were conducted using traditional experiment and algorithm based approaches. In our paper, feature selection is conducted purely on t-way combinations. Each feature set is evaluated on its classification performance and the one with the highest performance is selected.

VI. Conclusions and Future Work

Machine learning researchers often need to make the trade-off between using better learning models or using better training data when they look for a machine learning solution [21]. Feature selection is a major way for improving machine learning models. In our paper, we proposed a combinatorial method guided feature selection. Comparing to other wrapper approaches, the proposed approach can find the feature set that achieves the same or better classification accuracy as other approaches but requires less experiments for this.

To obtain the highest level of the accuracy, we considered combinations with different numbers of features (different sizes) from 6 to 32. For each size we generated t-way sets of combinations of features for t=2...6 and used each combinations in the classification process. Totally, 60 combinatorial sets were created.

The interesting research result is on dependency of accuracy on the number of features. For small number of features, each additional feature for classification process increases the accuracy. However, starting from one point, adding new features is not useful more and even decreases the accuracy. It means that using all features is not effective and that there is a specific number of features that provides the maximum value of the accuracy. In our case, this number is 18 (features) and the maximum accuracy is 72.23%.

Another observation is that the difference of the accuracy for 2-way and 6-way combinations is only around 0.5% that is not significant. So, if we accept detecting features providing not optimal accuracy but still very close (0.5-1.0 % difference) to optimal, we can use only pairwise combinations of features. This significantly reduces the number of required experimental evaluations.

To generate combinatorial combination, we divided all features into several groups and selected one feature from each group according to t-way coverage. The important direction for future research is formalizing this dividing process and investigation how dividing impacts the accuracy. In the future, we also will conduct more experimental studies of the proposed approach on machine learning models that have the much higher number of features.

ACKNOWLEDGMENT

This research is supported in part by grants #1262933 and #1560037 from the National Science Foundation. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

REFERENCES

[1] E. Giannoulatou, S.-H. Park, D. Humphreys, and J. Ho, "Verification and validation of bioinformatics software without a gold standard: a case study of bwa and bowtie," *BMC Bioinformatics*, vol. 15(Suppl 16):S15, 2014.

[2] J. A. Saez, B. Krawczyk, and M. Wozniak, "On the influence of class noise in medical data classification: Treatment using noise filtering methods," *Applied Artificial Intelligence*, vol. 30, no. 6, pp. 590–609, Jul. 2016.

[3] M. Mohri, A. Rostamizadeh, and A. Talwalkar.(2012), *Foundations of machine learning*. MIT press.

[4] D. R. Kuhn, R. Kacker, and Y. Lei, *Introduction to Combinatorial Testing*. Chapman and Hall/CRC, 2013, p. 341.

[5] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, Jan. 1998.

[6] K. Jacobs, J. Lu, and X. Hu, "Development of a diffraction imaging flow cytometer," *Opt. Lett.*, vol. 34, no. 19, p. 29852987, 2009.

[7] R. M. Haralick, K. Shanmugan, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 610–621, 1973.

[8] S. K. Thati, J. Ding, D. Zhang, and X. Hu, "Feature selection and analysis of diffraction images," in *4th IEEE Intl. Workshop on Information Assurance*, Vancouver, Canada, August 2015.

[9] NVIDIA, "Cuda parallel computing platform," April 2017. [Online]. Available: http://www.nvidia.com/object/cuda_home_new.html

[10] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[11] *Automated Combinatorial Testing for Software (ACTS)*, National Institute of Standards and Technology (NIST) Std., April 2017. [Online]. Available: http://csrc.nist.gov/groups/SNS/acts/

[12] M. Grindal, J. Offutt, and S. Andler, "Combination testing strategies: a survey," *Software Testing, Verification and Reliability*, vol. 15, no. 3, pp. 167–199, Mar. 2005.

[13] C. Nie and H. Leung, "A survey of combinatorial testing," *ACM Computing Surveys (CSUR)*, vol. 43, no. 2, 2011.

[14] J. Lawrence, R. Kacker, Y. Lei, D. R. Kuhn, and M. Forbes, "A survey of binary covering arrays," *The Electronic Journal of Combinatorics*, vol. 18, no. 1, 2011.

[15] S. Vilkomir, "Combinatorial testing of software with binary inputs: A state-of-the-art review," in *Proceedings of the 2016 IEEE International Conference on Software Quality, Reliability and Security (QRS 2016) Companion*, Vienna, Austria, 1-3 August 2016, pp. 55–59.

[16] Y. Lei, R. Kacker, D. Kuhn, V. Okun, and J. Lawrence, "IPOG/IPOD: efficient test generation for multi-way combinatorial testing," *Software Testing, Verification, and Reliability*, vol. 18, no. 3, pp. 125–148, Sept. 2008.

[17] S. Dreiseitl and M. Osl, "Feature selection based on pairwise classification performance," in *R. Moreno-Daz R., F. Pichler, A. Quesada-Arencibia (eds) Computer Aided Systems Theory - EUROCAST 2009, LNCS*, vol. 5717, 2009.

[18] J. Wang, "Automated classification of massive scale image data," Master Thesis, Dept. of Computer Science at East Carolina University, 2016.

[19] H. Silva and A. Fred, "Pairwise vs global multi-class wrapper feature selection," in *Proceedings of the 6th Conference on 6th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases*, vol. 6, 2007, pp. 1–6.

[20] S. Li and S. Oh, "Improving feature selection performance using pairwise pre-evaluation," *BMC Bioinformatics*, vol. 17, no. 1, p. 312, 2016.

[21] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.