

Article

An Automatic Analysis System for High-Throughput *Clostridium Difficile* Toxin Activity Screening

Megan Garland ^{1,*}, Joanna Jaworek-Korjakowska ^{2,*}, Urszula Libal ³, Matthew Bogyo ^{4,5} and Marcin Sięńczyk ⁶

¹ Cancer Biology Program, Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA

² Department of Automatic Control and Robotics, AGH University of Science and Technology, 30-059 Krakow, Poland

³ Signal Processing Systems Department, Faculty of Electronics, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; urszula.libal@pwr.edu.pl

⁴ Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305, USA; mbogyo@stanford.edu

⁵ Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA

⁶ Faculty of Chemistry, Division of Medicinal Chemistry and Microbiology, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; marcin.sienczynk@pwr.edu.pl

* Correspondence: garlandm@stanford.edu (M.G.); jaworek@agh.edu.pl (J.J.-K.); Tel.: +48-12-617-50-65 (M.G. & J.J.-K.)

† These authors contributed equally to this work.

Received: 31 July 2018; Accepted: 22 August 2018; Published: 1 September 2018

Abstract: *Clostridium difficile* infection (CDI) is an increasing global health threat and major worldwide cause of hospital-acquired diarrhea. The development of novel therapies to effectively treat this bacterial pathogen is an unmet clinical need. Here, we describe an image processing and classification algorithm that automatically identifies toxin-induced cytotoxicity to host cells based on characteristic morphological changes. This efficient and automatic algorithm can be incorporated into a screening platform to identify novel anti-toxin inhibitors of the *C. difficile* major virulence factors TcdA and TcdB, and contains the following steps: image enhancement, cell segmentation, and classification. We tested the algorithm on 504 images (containing 5096 cells) and achieved 93% sensitivity and 91% specificity, indicating that the proposed computational approach correctly classified most of the cells and provided reliable information for an effective screening platform. This algorithm achieved higher classification results compared to existing cell counter and analysis programs, scoring 92.6% accuracy. Compared to visual examination by a researcher, the algorithm significantly decreased classification time and identified toxin-induced cytotoxicity in an unbiased manner. *Availability:* Examples are available at home.agh.edu.pl/jaworek/CDI.

Keywords: *Clostridium difficile*; image analysis; pattern recognition; classification; screening system; anti-toxin inhibitors

1. Introduction

Clostridium difficile infection (CDI) is an increasing global health threat and major cause of hospital-acquired diarrhea [1–3]. The development of novel therapies to effectively treat this bacterial pathogen is an urgent unmet clinical need evidenced by increasing infection rates, high incidence of recurrence, and rising antibiotic resistance [4,5]. An alternative therapeutic strategy to antibiotic treatment is the development of anti-toxin compounds that neutralize the toxin mediators of disease—toxins A and B (TcdA and TcdB) of *C. difficile* [6]. Here, we describe a computer application

based on image processing and classification algorithms to rapidly screen and identify anti-toxin inhibitors of TcdA and TcdB activity based on characteristic morphological changes to host cells upon toxin-mediated cytotoxicity. Representative images of cells before (a) and after (b) toxin treatment are shown in Figure 1. This technology can be applied to the screening of novel anti-toxin compounds for use as monotherapy or in combination with antibiotics to treat CDI.

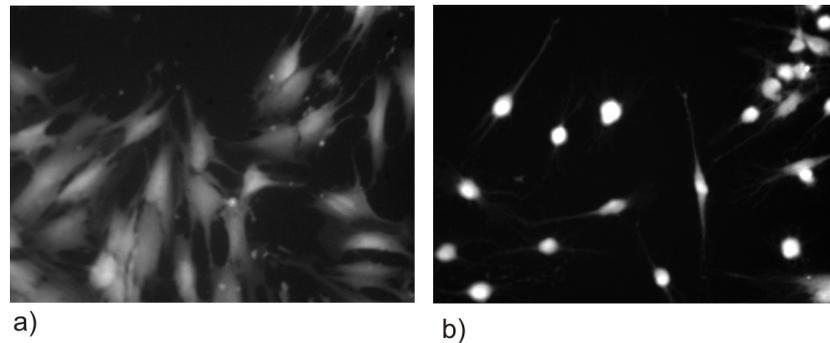


Figure 1. Cell-rounding assay to identify anti-toxin inhibitors of TcdB. (a) Fluorescence image of human foreskin fibroblast (HFF) cells before toxin challenge. Adherent HFF cells are stained with cytosolic cell tracker dye. (b) Fluorescence image of HFFs after challenge with TcdB. Adherent HFFs round upon the addition of toxin, indicating actin cytoskeletal dysfunction following intracellular toxin activity against host Rho/Rac GTPases.

1.1. Clinical Definition and Motivation

C. difficile is a Gram-positive anaerobic bacterium that is a major cause of hospital-acquired infections [1–3]. The incidence of CDI has been rising in hospitals and long-term health care facilities during the past several years (see Figure 2) [7]. Additionally, infections are becoming more severe with more complications, deaths, and higher healthcare-associated costs [2]. This increased morbidity and mortality requires a change not only in management and treatment, but also in early diagnosis.

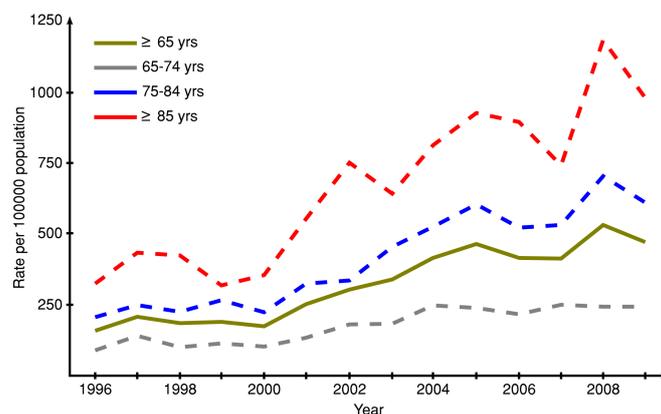


Figure 2. Increases in *C. difficile* infection over time. Graph includes patients aged 65 years and older who were either hospitalized with CDI or who acquired *C. difficile* during a hospital stay. Rates for 1996–1999 were based on U.S. Census Bureau civilian population estimates adjusted for the net underenumeration in the 1990 census. Rates for 2000–2009 were calculated using U.S. Census Bureau 2000-based postcensal civilian population estimates [7].

Infection by *C. difficile* usually occurs when patients are placed on broad-spectrum antibiotics for an unrelated condition, altering the microbiome in the gastrointestinal tract and allowing this opportunistic pathogen to thrive. CDI causes a range of symptoms from diarrhea and abdominal pain to pseudomembranous colitis, toxic megacolon, and death [8,9]. Even though antibiotics are a main

risk factor for the development of CDI, the antibiotics vancomycin, metronidazole, and fidaxomicin are the gold standard and first-line treatments for CDI. Initial cure rates with these antibiotics range from 72.7–81.1%, and 15–25% of patients will have recurrent infection [10,11]. Further, after the first recurrence, patients have approximately a 50% chance of subsequent recurrences [11]. In addition to rising rates of infection and recurrence, antibiotic use contributes to increasing antibiotic resistance and precludes the recovery of a healthy microbiome. An alternative or complementary treatment strategy for CDI is to target the exotoxin mediators of clinical disease with anti-toxin strategies [12,13].

CDI is mediated by two large multi-subunit clostridial toxins: TcdA and TcdB. Epidemiologic studies show that only *C. difficile* expressing toxin are able to cause clinical disease [14,15]. Each toxin contains a glucosyltransferase domain (GTD), a cysteine protease domain (CPD), a translocation domain, and a receptor-binding domain [1,16]. The toxins are taken up into host enterocytes by endocytosis, where acidification of the endosome mediates conformational changes that allow the translocation domain to extrude the GTD and CPD into the host cell cytosol [16]. The cytosolic, mammalian-specific sugar 1D-myo-inositol hexakisphosphate (IP_6) activates the CPD to cleave off the GTD, where it exerts its toxigenic effects via monoglucosylation of members of the Rho/Rac family of small GTPases [16]. Glucosylation of Rho/Rac GTPases causes actin cytoskeletal disruption in the host cell that can be seen by a characteristic cellular shrinking and rounding morphology [1] (see Figure 1).

Due to the essential function of the toxins in mediating disease, strategies that neutralize toxin activity could treat CDI as a monotherapy or in combination with antibiotics. Efforts to identify small-molecule inhibitors of the CPD and GTD domains have demonstrated the feasibility of this strategy to effectively treat CDI *in situ* and with *in vivo* animal models [12,17–21]. However, these proof-of-concept inhibitors have yet to advance to clinical trials. To identify compounds that are non-toxic to host cells and neutralize toxin function, a medium-throughput screen that can rapidly read out the characteristic rounded cell morphology compared to healthy cells would significantly increase the number of compounds that could be screened (see Figure 3).

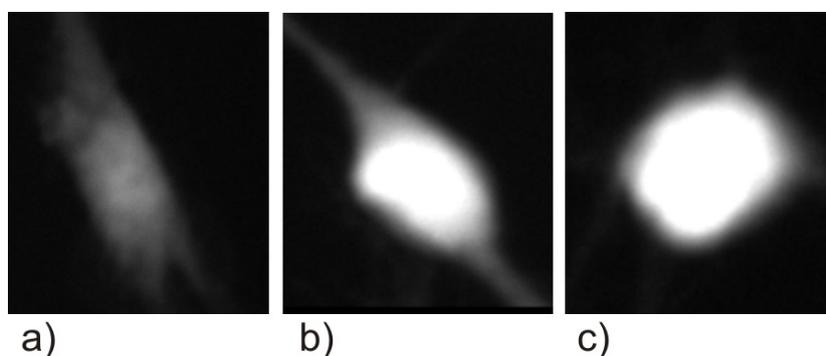


Figure 3. Characteristic morphology changes with TcdB challenge. (a) Adherent HFFs without toxin treatment exhibit diffuse cytosolic staining with a cell tracker dye. Upon TcdB treatment, actin cytoskeletal disruption leads to a progression from (b) partial to (c) complete cell rounding.

1.2. Related Works

The cell-rounding assay was adapted from an approved clinical test used to diagnose *C. difficile* in human patients. In this clinical test, evaluation and differentiation between adherent and rounded cells is performed by visual examination. Therefore, adaptation of this assay into a screen to identify potential anti-toxin compounds in a library composed of thousands of molecules is limited in its throughput due to the necessity of manually analyzing each image. Additionally, when including the replicates needed to achieve statistically reliable data, the resulting set of images cannot be easily analyzed manually within a reasonable time frame. To improve the high-throughput potential of this assay, we developed an algorithm to automatically count and classify cells according to the morphological changes seen upon toxin challenge. To the best of our knowledge, there is no commercial

software to analyze and classify such images in an unsupervised way, but some existing software can be adapted for counting or classification functions. For cell counting, ImageJ—an open source image processing program designed for scientific multidimensional images—contains a derivate (*Fiji*) released in 2003 to manually count cells [22–25]. *Icy*, a JAVA plugin, is another interactive counter [26]. For cell classification, the shape parameter circularity in ImageJ classifies detected cells based on a function of the perimeter P and the area A [27]. However, classification of adherent cells with irregular borders remains a challenge. If circularity is not a sufficient exclusion criterion, the *BioVoxxel Toolbox* [28] is an advanced particle analyzer. However, solutions are based on a permanent interaction with the user, and the detection and classification parameters are chosen manually [28]. Therefore, use of this technology does not significantly decrease the time required to analyze images.

Additionally, without image processing and machine learning knowledge, cell counting, analysis, and classification steps can contain high error rates for large datasets. To overcome the current limitations in the automatic image analysis of phenotypic cell-based screens, we created an automated approach for the differentiation of rounded versus adherent cells exposed to *C. difficile* toxins and potential inhibitors. Moreover, this method can be easily adapted and applied to many diagnostics and antitoxin screens for a variety of cytotoxic agents. An overall summary and comparison of the proposed and described methods are presented in Table 1.

Table 1. Comparison of classification results for different algorithms. ACC: accuracy; SN: sensitivity; SP: specificity.

Software	Classification Algorithm	SN (%)	SP (%)	ACC (%)
ImageJ	Circularity	67	71	69
BioVoxxel Toolbox	Shape analysis	82	84	82
Described algorithm	Preprocessing, image enhancement, shape analysis, pixel brightness	93	91	92.6

2. Methods

2.1. Automated Classification System

We created a computer-aided analysis system to automatically classify rounded versus adherent cells in a phenotypic screen of toxin-induced cytotoxicity. This system is designed to reproduce manual visual classification outcomes using fluorescent images of cell nuclei (DAPI) and cellular morphology (GFP). An outline of the methodology used to discriminate between adherent and rounded cells is presented in Figure 4. The proposed application was implemented in *MATLAB* using the *Image Processing and Statistics* and *Machine Learning Toolboxes*. The system is divided into four main stages including: preprocessing of GFP and DAPI images, nuclei detection on DAPI images, cell segmentation on GFP images, and classification. The automatic system based on the identification of cells on both images (DAPI and GFP) classifies them into two classes: adherent or rounded.

2.2. Image Preprocessing

The preprocessing step improves the image quality by reducing or removing unrelated and surplus parts in the images. GFP and DAPI images are fairly low-contrast, and therefore we applied the contrast-limited adaptive histogram equalization method (contrast adjustment). The results are presented in Figure 5.

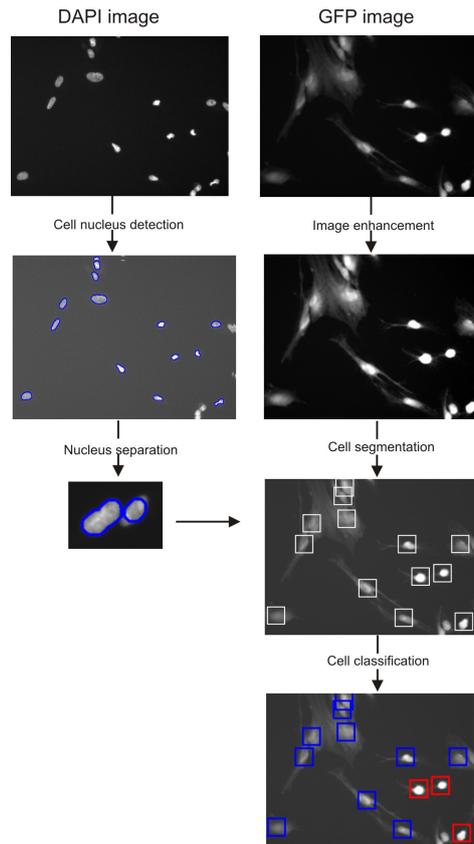


Figure 4. Fundamental steps of the analysis algorithm. For each set of images, a DAPI and a GFP image is preprocessed and analyzed. (1) Original DAPI and GFP images are enhanced. (2) Cell nuclei are detected and separated from each other. (3) Based on the position of nuclei, the cells are segmented in the GFP images. (4) Based on the segmented area, each cell is analyzed and classified as adherent (blue) or rounded (red).

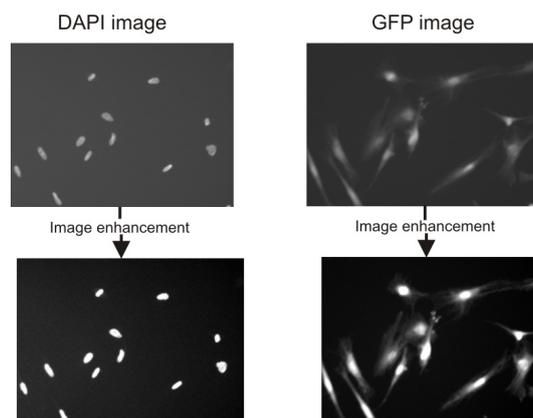


Figure 5. Outcome of the preprocessing step. The image quality is improved, and nuclei and cell areas become more visible.

2.3. Cell Detection

The *cell detection* step locates cells using images of Hoescht fluorescence (DAPI channel). DAPI images contain only the cell nuclei and are used to locate each cell and, more importantly, to quantify the number of cells. This step is fundamental for the correct separation of cells, because the

cell nuclei and/or cytosol of adjacent cells may overlap. First, we identified cell nuclei in the DAPI images and then applied the watershed transform to separate the nuclei.

The cell nuclei were automatically detected by implementing the Otsu thresholding algorithm [29]. As presented in Figure 6, some of the nuclei overlap or border with each other, and therefore need more calculation to be separated properly. First, the watershed transform segments the nuclei. This method consists of placing a water source in each regional minimum in the relief in order to flood the entire relief from sources, and build barriers when different water sources meet. The watershed transform returns a label matrix L that identifies the watershed regions of the input matrix A , which can have any dimension. The watershed transform finds “catchment basins” or “watershed ridge lines” in an image by treating it as a surface where light pixels represent high elevations and dark pixels represent low elevations. The elements of L are integer values greater than or equal to 0. The elements labeled 0 do not belong to a unique watershed region. The elements labeled 1 belong to the first watershed region, the elements labeled 2 belong to the second watershed region, and so on. By default, watershed uses 8-connected neighborhoods for 2-D inputs. The implemented method uses the Fernand Meyer algorithm [30]. A watershed definition for the continuous case can be based on distance functions. Depending on the distance function used, one may arrive at different definitions. In this work, we used the distance function described in [30]. Assume that the function f is an element of the space $C(D)$ of real twice continuously differentiable functions on a connected domain D with only isolated critical points (the class of Morse function on D forms).

Then, the topographical distance between p and q is the minimum of the topographical distances along all paths between p and q :

$$T_f(p, q) = \text{MIN}_{\pi \in [p \rightsquigarrow q]} T_f^\pi(p, q), \quad (1)$$

where the set of all paths from p to q is denoted by $[p \rightsquigarrow q]$. The topographical distance between a point $p \in D$ and a set $A \subseteq D$ is defined as $T_f(p, A) = \text{MIN}_{a \in A} T_f(p, a)$.

The path with shortest T_f distance between p and q is the path of steepest slope [31]. Several definitions of the watershed transform have been proposed: an excellent review can be found in [31], where additional details can be found.

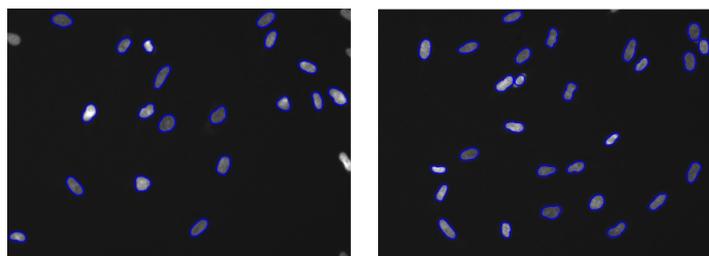


Figure 6. Nucleus detection and segmentation of DAPI images. The watershed transform separates nuclei from each other.

2.4. Cell Segmentation

Based on the results of the *cell detection* step, we calculated centroids of the nucleus that fit the shape of nuclei in the DAPI image. We calculated centroids using the assumptions that every cell has one nuclei and that it lies inside the cell, mostly in the center. Our method takes the coordinates of a nucleus in the digital image and outlines a squared segment around it for the classification step.

The DAPI image with superimposed centroid locations was transformed into a set of 50×50 pixel images. The number of small segments depends on the number of cells in the examined DAPI image. Each detected centroid determines the 50×50 pixel segment location. Although all nuclei are detected in the previous step, only cells that do not lie on the border of the image are processed. We rejected cells on the borders of the image because the image only captures part of the cell, preventing accurate

classification. Squares corresponding to the 50×50 pixel squares identified in the DAPI image are then identified in the GFP image as presented in Figure 7. The GFP segments are statistically analyzed and have a crucial impact on the final classification rule.

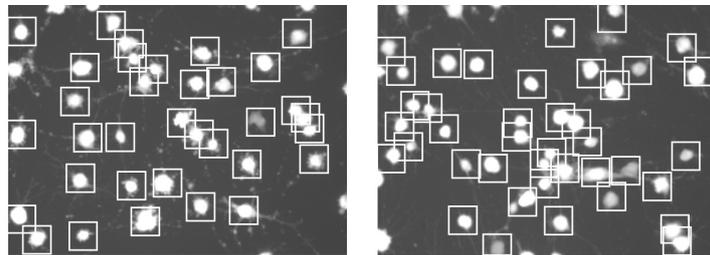


Figure 7. Outcome of the cell segmentation step. Detected cells are outlined in square segments of 50×50 pixels.

2.5. Classification

The classifier generates a classification rule using information gathered based on 50×50 pixel segments cut from GFP images (corresponding to 50×50 pixel DAPI segments), divided into two groups: adherent and rounded cells. This procedure required the preparation of learning data. In particular, many cells were first manually defined as adherent or rounded by an expert. The dataset was divided into training and test sets based on the k -fold cross-validation method [32,33] with $k = 10$. In k -fold cross-validation, the original sample is randomly partitioned into k equal-sized subsamples. The k -fold cross-validation method uses all observations for both training and validation, with each of the k subsamples used exactly once as the validation data. Then, the classifier learns (parameter adjustment) based on the positive and negative control images and runs using new images, without any external help. Each square GFP image segment produced in the previous step, with one cell in the center, is classified. The distinction between adherent and rounded cells is made based on probability density functions (PDFs) of gray-scale levels. The main difference in PDF shapes leads to a compact thresholding method. The classifier indicates whether the cell contained in a segment is adherent or rounded.

The proposed method is based on a naïve Bayes classifier, a classical well-known algorithm [34–38] that is still widely applied, with satisfactory results in various applications [39–44]. The naïve Bayes classifier assumes that boundary distributions are independent and in general multidimensional cases (for D dimensions, feature vector is $x = (x_1, \dots, x_D)$) can be defined as the following classification rule:

$$\Psi(x) = \arg \max_{c \in \{1,2\}} p(c) \prod_{d=1}^D p_c(x_d), \tag{2}$$

where $p(c)$ is the prior probability of occurrence of class c , and $p_c(x)$ is a probability density function (PDF) in class c . The notation $c \in \{1, 2\}$ results from the fact that we consider a two-class problem. In our case, the classification problem was transformed to a one-dimensional task ($D = 1$), because of the use of grayscale distributions for the horizontal axis. The distributions for horizontal and vertical axes were very similar, and this kind of symmetry can result from the huge freedom of cell rotation, shown in the images (see Figure 6). The simplification leads to the following classifier form:

$$\Psi(x) = \begin{cases} 1, & \text{if } p_1(x_1) > p_2(x_1), \\ 2, & \text{if } p_1(x_1) < p_2(x_1), \end{cases} \tag{3}$$

where we assumed that rounded and adherent cell classes are equiprobable (i.e., $p(1) = p(2) = 0.5$). The one-dimensional naïve Bayes classifier is in fact the optimal Bayes classifier, in terms of the lowest classification risk [35]. The probability distributions of gray-scale illumination in both classes were estimated in previously cut squares containing cells (for details of cell segmentation, see Section 2.4).

The probability density functions in classes were estimated non-parametrically by histograms (see Figure 8). In the presented case, the separation point between classes in the horizontal axis was achieved by a naïve Bayes classifier for 69.4% of brightness.

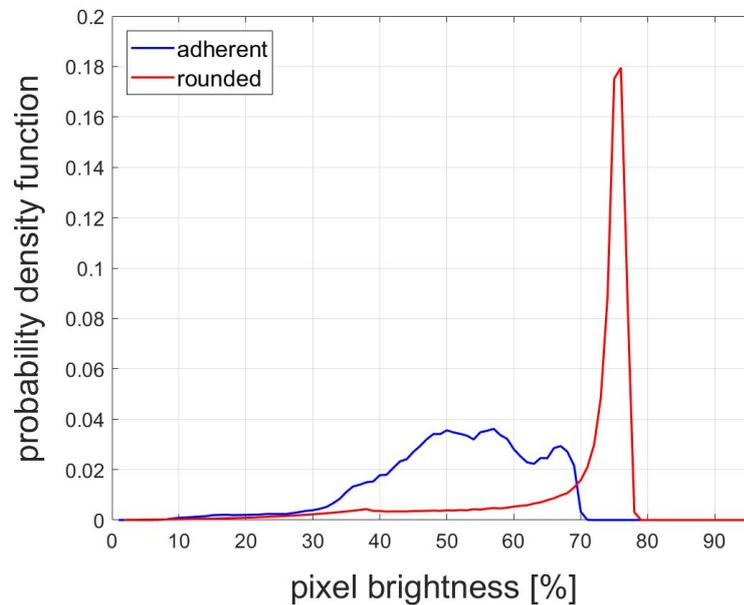


Figure 8. Probability density functions of pixel brightness in the horizontal axis (brightness scale from 1 to 96) for adherent cells (blue) and rounded cells (red).

The final classification result is shown in Figure 9. The rounded cells are marked with red while adherent cells are outlined in blue.

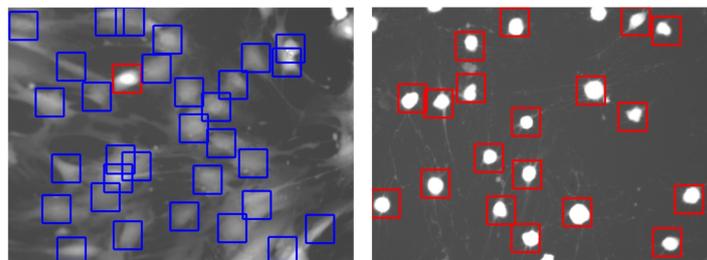


Figure 9. Classification results. Adherent cells are marked with blue squares, while rounded cells are indicated by red squares.

2.6. Cell-Rounding Assay

Human foreskin fibroblasts (HFFs) were cultured in Dulbecco's modified medium (DMEM) supplemented with 10% (*v/v*) fetal calf serum, 10 U/mL penicillin G, and 100 µg/mL streptomycin, and were plated in 96-well dishes at a density of 12,500 cells/well and allowed to adhere overnight. Cells were washed with PBS, then treated with 1 µM CellTracker™ green 5-chloromethylfluorescein diacetate (CMFDA) (1 mM stock in DMSO) and Hoechst 33342 nuclear stain in a final volume of 100 µL/well in serum-free DMEM for 30 min at 37 °C. TcdB (125 ng) or buffer control was pretreated with 1 µL DMSO or 200× concentration of inhibitor in 14 µL buffer (20 mM TRIS, 60 mM NaCl, pH 8.0) for 1 h at 37 °C. Following toxin pretreatment, 185 µL DMEM was added to the toxin/inhibitor reaction, mixed, and applied to cells. Cells were incubated with toxin for 1 h at 37 °C, then imaged on a BioTek Cytation 3 plate reader in bright field, DAPI, and GFP channels. Four sets of images per well were taken.

3. Results

3.1. Image Database

The complete dataset includes a DAPI and GFP image for each of 504 outcomes of the cell-rounding assay to identify anti-toxin inhibitors of TcdB (5096 cells in total). All images were assessed manually. The preprocessing and cell segmentation steps achieved a detection error less than 2% due to high contrast and did not affect subsequent classification analysis.

3.2. Classification Results

To compare our algorithm with popular methods for cell segmentation, we evaluated the segmentation performance by quantifying errors in the number of cells counted. Accurate segmentation of rounded cells can be measured with morphological cell features such as cell area. However, the adherent cells had irregular borders, which complicated analysis. For the control of performance, we used different types of GFP and DAPI images with various levels of noise and contrast. While building our solution, we compared the algorithm with other segmentation methods, such as active contour models for DAPI images or the region-growing algorithm for GFP images. Segmentation errors of the algorithm led to mis-segmentation rates of 3% of the cells, with errors mostly caused by rounded cells undergoing nuclear fragmentation that were therefore counted as two rounded cells.

Diagnostic exactness is often used to mean the fraction of cases on which the automated image analysis system gave a correct answer, where correctness is determined by comparing the diagnostic decision to some definition of “truth” [45]. There are different ways to define “truth”. In our case, we depended on the separate gold standard, which is determination by the results of a pathology expert. We then assessed the accuracy of the designed classification system by comparing its results against the gold standard. The analyzed task is a binary problem, where the answer is “1” when the cell is adherent or “0” when it is rounded. Table 2 shows the classification results obtained with the implemented algorithm in a confusion matrix.

Table 2. List of parameters evaluated by the proposed method. FN: false negative; FP: false positive; TN: true negative; TP: true positive.

	Adherent (Predicted)	Rounded (Predicted)	Total
Adherent (Actual)	2552 (TP)	213 (FP)	2765
Rounded (Actual)	164 (FN)	2167 (TN)	2331
	Sensitivity: 93%	Specificity: 91%	Total: 5096

To measure the diagnostic performance, we calculated accuracy, sensitivity, and specificity. The fundamental definitions of these performance measures can be illustrated as follows.

The classification accuracy of a technique depends upon the number of correctly classified samples (i.e., true negative and true positive), and is calculated as follows:

$$Accuracy = \frac{\#TP + \#TN}{N}, \quad (4)$$

where TP is the true positive answer, TN is the true negative answer, and N is the total number of samples present in the images.

The sensitivity (also called the true positive rate—TPR) is the probability that the cell is said to be adherent given that it is. This can be estimated by relative frequency [45,46]:

$$\text{Sensitivity} = \frac{\#TP}{\#TP + \#FN} \quad (5)$$

where TP is the true positive answer and FN is the false negative answer. The value of sensitivity ranges between 0 and 1, where 0 and 1 mean worst and best classification, respectively.

On the other hand, we calculated specificity (also called false positive rate—FPR), which is defined as the proportion of the true negatives against all negative results. Specificity is defined by the following equation:

$$\text{Specificity} = \frac{\#TN}{\#TN + \#FP} \quad (6)$$

where TN is the true negative answer and FP is the false positive answer. The value of specificity ranges between 0 and 1, where 0 and 1, respectively, meaning worst and best classification, respectively.

Another dominant technique for assessing computer classification performance is the receiver operating characteristic (ROC) analysis [32,46–48]. A ROC curve is a plot of sensitivity versus (1—specificity) (i.e., TPR versus FPR; see Figure 10). Area under the ROC curve is the most commonly used accuracy index. The area under a perfect-accuracy ROC curve, when sensitivity and specificity reach 1.0, is also 1.0.

To create the ROC curve shown in Figure 10, we used 10-fold cross-validation and a method based on the Mann–Whitney statistic [47,49–51]. The ROC curve was not computed point-by-point, but for those random folds in every step. Because of the use of Mann–Whitney statistic and 10-fold cross-validation, the presented ROC curve has only 10 points connected by a staircase function. For each data fold, we first performed the classifier training using 9/10 of the data to find the gray-scale illumination threshold (see Figure 8) optimally discriminating classes, and then the classifier testing was done using the remaining 1/10 of the data set.

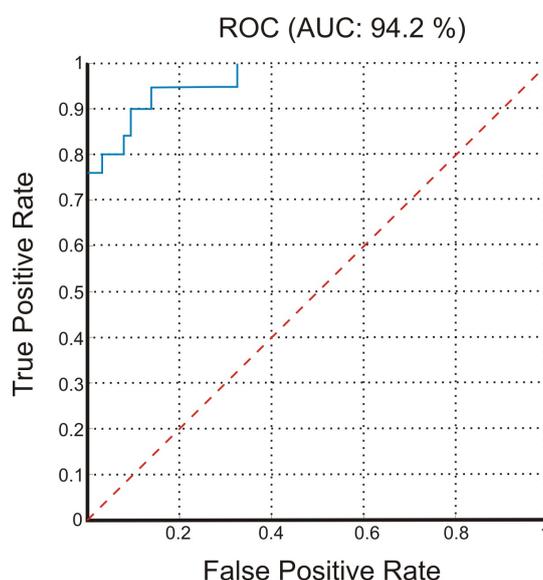


Figure 10. Receiver operating characteristic (ROC) curve analysis (blue line) for the implemented automatic analysis system for high-throughput screening of *C. difficile* anti-toxin compounds. AUC: area under the curve.

Table 1 summarizes the classification results of the works reported in Section 1.2. The comparison between the described applications was not easy to perform due to differences in the algorithms. For example, the proposed algorithm automatically selects the image enhancement threshold, the contrast adjustment parameters, and the nucleus separation values, whereas the described toolboxes need to have an exact value chosen manually. However, we chose 200 images and compared the results using the most relevant set of parameters.

The algorithm we developed was comparable to the gold-standard of manual scoring (accuracy = 92.6%, sensitivity = 93%, specificity = 91%, and area under ROC curve = 94.2%). In addition to its high level of accuracy, this automated algorithm is much less time consuming than manual scoring, and the results are objective and unbiased. Table 1 shows a significant improvement of the algorithm over alternate programs such as ImageJ and the BioVoxel Toolbox in terms of sensitivity, specificity, and overall accuracy. Overall, these results show that cell classification can be supported by a powerful system based on the proposed algorithms to improve phenotypic cell-based high-throughput assays (see Figure 11).



Figure 11. MATLAB graphical user interface (GUI) of the attached software for image selection, data processing, and statistical analysis.

4. Discussion and Conclusions

Our work presents an algorithm for the automatic discrimination of phenotypic cellular changes to improve the high-throughput screening of structurally diverse compound libraries to identify inhibitors of *C. difficile* toxins. In our system, human fibroblasts were treated with TcdB. Computer-aided analysis of images showing morphological changes to cells was a time-efficient method to identify toxin-mediated cell damage. Our system achieved 93% sensitivity and 91% specificity towards adherent and rounded cells.

Although the system was designed to search for potential agents neutralizing *C. difficile* toxins using a human fibroblast cell line, it could easily be applied to different cell-based assays. The algorithm could be adapted for use in diagnostics aimed at detecting many different types of cytotoxic compounds or toxins, or for other applications, such as screening for novel antitumor agents. Our future plans will concentrate on expanding our database and implementing a deep neural network architecture like convolutional neural network (CNN) or U-Net to solve the remaining problems and achieve even better results.

Author Contributions: Conceptualization, M.G., J.J.-K. and M.B.; Data curation, M.G.; Formal analysis, U.L.; Funding acquisition, M.G.; Methodology, J.J.-K. and U.L.; Project administration, M.G., J.J.-K. and U.L.; Resources, M.G. and M.B.; Software, J.J.-K. and U.L.; Supervision, M.B. and M.S.; Validation, M.G.; Writing—original draft, M.G., J.J.-K., U.L. and M.S.

Funding: This work was funded in part by an National Institutes of Health grant R56 AI127455 to Matthew Bogoy, and the Howard Hughes Medical Institute Medical Research Fellows program (to Megan Garland), and by the Stanford Medical Scientist Training Program (to Megan Garland), and by the AGH University of Science and Technology Status Funds 11.11.120.714 (to Joanna Jaworek-Korjakowska), and Wroclaw University of Science and Technology grant 0402/0088/17 (to Urszula Libal), and Marcin Sienczyk is grateful to Wroclaw University of Science and Technology for support (Statute Funds 0401/0195/17).

Acknowledgments: J.J.-K., U.L. and M.S. are laureates of the “TOP 500 Innovators” program developed by the Polish Ministry of Science and Higher Education and focused on science management, technology transfer, commercialization and university collaboration. We would like to thank the Polish Ministry of Science and Higher Education, Stanford University, Piotr Moncarz, and our friends for a life-changing experience. You stretched our thoughts and provided a road map for even greater achievements.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Carroll, K.C.; Bartlett, J.G. Biology of *Clostridium difficile*: Implications for epidemiology and diagnosis. *Annu. Rev. Microbiol.* **2011**, *65*, 501–521. [PubMed]
2. Goudarzi, M.; Seyedjavadi, S.S.; Goudarzi, H.; Mehdizadeh Aghdam, E.; Nazeri, S. *Clostridium difficile* Infection: Epidemiology, Pathogenesis, Risk Factors, and Therapeutic Options. *Scientifica* **2014**, *2014*. [CrossRef] [PubMed]
3. Magill, S.S.; Edwards, J.R.; Bamberg, W.; Beldavs, Z.G.; Dumyati, G.; Kainer, M.A.; Lynfield, R.; Maloney, M.; McAllister-Hollod, L.; Nadle, J. Multistate point-prevalence survey of health care-associated infections. *N. Engl. J. Med.* **2014**, *370*, 1198–208. [CrossRef] [PubMed]
4. Hughes, D.; Karlen, A. Discovery and preclinical development of new antibiotics. *Ups. J. Med. Sci.* **2014**, *119*, 162–169. [CrossRef] [PubMed]
5. O'Brien, S. Meeting the societal need for new antibiotics: The challenges for the pharmaceutical industry. *Br. J. Clin. Pharmacol.* **2015**, *79*, 168–172. [CrossRef] [PubMed]
6. Pruitt, R.N.; Lacy, D.B. Toward a structural understanding of *Clostridium difficile* toxins A and B. *Front. Cell. Infect. Microbiol.* **2012**, *2*, 28. [CrossRef] [PubMed]
7. Centers for Disease Control and Prevention. QuickStats: Rates of *Clostridium Difficile* Infection among Hospitalized Patients Aged ≥ 65 Years,* by Age Group—National Hospital Discharge Survey, United States, 1996–2009. Available online: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6034a7.htm> (accessed on 22 August 2018).
8. Abt, M.C.; McKenney, P.T.; Pamer, E.G. *Clostridium difficile* colitis: pathogenesis and host defence. *Nat. Rev. Microbiol.* **2016**, *14*, 609–620. [CrossRef] [PubMed]
9. Khanna, S.; Pardi, D.S. *Clostridium difficile* Infection: New Insights Into Management. *Mayo Clin. Proc.* **2012**, *87*, 1106–1117. [CrossRef] [PubMed]
10. Johnson, S.; Louie, T.J.; Gerding, D.N.; Cornely, O.A.; Chasan-Taber, S.; Fitts, D.; Gelone, S.P.; Broom, C.; Davidson, D.M. For the Polymer Alternative for CDI Treatment (PACT) Investigators. Vancomycin, metronidazole, or tolevamer for *Clostridium difficile* infection: Results from two multinational, randomized, controlled trials. *Clin. Infect. Dis.* **2014**, *59*, 345–354. [CrossRef] [PubMed]
11. Kelly, C.P. Can we identify patients at high risk of recurrent *Clostridium difficile* infection? *Clin. Microbiol. Infect.* **2012**, *18*, 21–27. [CrossRef] [PubMed]
12. Garland, M.; Loscher, S.; Bogyo, M. Chemical Strategies To Target Bacterial Virulence. *Chem. Rev.* **2017**, *117*, 4422–4461. [CrossRef] [PubMed]
13. Upadhyay, A.; Mooyottu, S.; Yin, H.; Surendran Nair, M.; Bhattaram, V.; Venkitanarayanan, K. Inhibiting Microbial Toxins Using Plant-Derived Compounds and Plant Extracts. *Medicines* **2015**, *2*, 186–211. [CrossRef] [PubMed]
14. Kyne, L.; Warny, M.; Qamar, A.; Kelly, C.P. Asymptomatic Carriage of *Clostridium difficile* and Serum Levels of IgG Antibody against Toxin A. *N. Engl. J. Med.* **2000**, *342*, 390–397. [CrossRef] [PubMed]
15. Shim, J.K.; Johnson, S.; Samore, M.H.; Bliss, D.Z.; Gerding, D.N. Primary symptomless colonisation by *Clostridium difficile* and decreased risk of subsequent diarrhoea. *Lancet* **1998**, *351*, 633–636. [CrossRef]
16. Shen, A. *Clostridium difficile* toxins: Mediators of inflammation. *J. Inmate Immun.* **2012**, *4*, 149–158. [CrossRef] [PubMed]
17. Bender, K.O.; Garland, M.; Ferreyra, J.A.; Hryckowian, A.J.; Child, M.A.; Puri, A.W.; Solow-Cordero, D.E.; Higginbottom, S.K.; Segal, E.; Banaei, N.; et al. A small-molecule antivirulence agent for treating *Clostridium difficile* infection. *Sci. Transl. Med.* **2015**, *7*. [CrossRef] [PubMed]
18. Lacy, D.B. Pre-Clinical Evaluation of *Clostridium Difficile* Toxin Inhibitors. Available online: <http://grantome.com/grant/NIH/I01-BX002943-01> (accessed on 22 August 2018).
19. Larabee, J.L.; Bland, S.J.; Hunt, J.J.; Ballard, J.D. Intrinsic toxin-derived peptides destabilize and inactivate *Clostridium difficile* TcdB. *MBio* **2017**, *8*. [CrossRef] [PubMed]
20. Letourneau, J.J.; Stroke, I.L.; Hilbert, D.W.; Sturzenbecker, L.J.; Marinelli, B.A.; Quintero, J.G.; Sabalski, J.; Ma, L.; Diller, D.J.; Stein, P.D.; et al. Identification and initial optimization of inhibitors of *Clostridium difficile* (*C. difficile*) toxin B (TcdB). *Bioorg. Med. Chem. Lett.* **2018**, *28*, 756–761. [CrossRef] [PubMed]

21. Tam, J.; Beilhardt, G.L.; Auger, A.; Gupta, P.; Therien, A.G.; Melnyk, R.A. Small Molecule Inhibitors of Clostridium difficile Toxin B-Induced Cellular Damage. *Chem. Biol.* **2015**, *22*, 175–185. [[CrossRef](#)] [[PubMed](#)]
22. Dordea, A.C.; Bray, M.-A.; Allen, K.; Logan, D.J.; Fei, F.; Malhotra, R.; Gregory, M.S.; Carpenter, A.E.; Buys, E.S. An open-source computational tool to automatically quantify immunolabeled retinal ganglion cells. *Exp. Eye Res.* **2016**, *147*, 50–56. [[CrossRef](#)] [[PubMed](#)]
23. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*, 676–682. [[CrossRef](#)] [[PubMed](#)]
24. Schindelin, J.; Rueden, C.T.; Hiner, M.C.; Eliceiri, K.W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol. Reprod. Dev.* **2015**, *82*, 518–529. [[CrossRef](#)] [[PubMed](#)]
25. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [[CrossRef](#)] [[PubMed](#)]
26. Chaumont, F. Manual Counting: Documentation. Available online: www.icy.bioimageanalysis.org/plugin/Manual_Counting. (accessed on 22 August 2018).
27. Helmy, I.M.; Abdel Azim, A.M. Efficacy of ImageJ in the assessment of apoptosis. *Diagn. Pathol.* **2012**, *7*, 15. [[CrossRef](#)] [[PubMed](#)]
28. Brocher, J.; Wagner, T. BioVoxxel Toolbox (ImageJ/Fiji). Available online: www.biovoxxel.de (accessed on 22 August 2018).
29. Feng, Y.; Zhao, H.; Li, X.; Zhang, X.; Li, H. A multi-scale 3D Otsu thresholding algorithm for medical image segmentation. *Digit. Signal Process.* **2017**, *60*, 186–199. [[CrossRef](#)]
30. Meyer, F. Topographic distance and watershed lines. *Signal Process.* **1994**, *38*, 113–125. [[CrossRef](#)]
31. Zanaty, E.A.; Afifi, A. A watershed approach for improving medical image segmentation. *Comput. Meth. Biomech. Biomed. Eng.* **2013**, *16*, 1262–1272. [[CrossRef](#)] [[PubMed](#)]
32. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: Berlin, Germany, 2009.
33. Japkowicz, N.; Shah, M. *Evaluating Learning Algorithms: A Classification Perspective*; Cambridge University Press: Cambridge, MA, USA, 2011.
34. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin, Germany, 2006.
35. Devroye, L.; Györfi, L.; Lugosi, G. *A Probabilistic Theory of Pattern Recognition*; Springer: Berlin, Germany, 1996.
36. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed.; Wiley: New York, NY, USA, 2000.
37. Mitchell, T. *Machine Learning*; McGraw Hill: New York, NY, USA, 1997.
38. Webb, A.R.; Copsey, K.D. *Statistical Pattern Recognition*, 3rd ed.; Wiley: New York, NY, USA, 2011.
39. Al-Aidaros, K.M.; Bakar, A.A.; Othman, Z. Medical data classification with Naïve Bayes approach. *Inform. Technol. J.* **2012**, *11*, 1166–1174. [[CrossRef](#)]
40. John, G.H.; Langley, P. Estimating continuous distributions in bayesian classifiers. In Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, 18–20 August 1995.
41. Khan, A.U.R.; Khan, M.; Khan, M.B. Naïve Multi-label Classification of YouTube Comments Using Comparative Opinion Mining. *Procedia Comput. Sci.* **2016**, *82*, 57–64. [[CrossRef](#)]
42. Ontivero-Ortega, M.; Lage-Castellanos, A.; Valente, G.; Goebel, R.; Valdes-Sosa, M. Fast Gaussian Naïve Bayes for searchlight classification analysis. *Neuroimage* **2017**, *163*, 471–479. [[CrossRef](#)] [[PubMed](#)]
43. Sheikhan, H.; Delavar, M.R.; Stein, A. Predictive Modelling of Seismic Hazard Applying Naïve Bayes and Granular Computing Classifiers. *Procedia Environ. Sci.* **2015**, *26*, 49–52. [[CrossRef](#)]
44. Vangelis, M.; Ion, A.; Geogios, P. Spam Filtering with Naïve Bayes—Which Naïve Bayes? In Proceedings of the 3rd Conference on Email and Anti-Spam, Mountain View, CA, USA, 27–28 July 2006.
45. Hajian-Tilaki, K. Sample size estimation in diagnostic test studies of biomedical informatics. *J. Biomed. Inform.* **2014**, *48*, 193–204. [[CrossRef](#)] [[PubMed](#)]
46. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
47. Feng, D.; Cortese, G.; Baumgartner, R. A comparison of confidence/credible interval methods for the area under the ROC curve for continuous diagnostic tests with small sample size. *Stat. Methods Med. Res.* **2015**, *26*, 2603–2621. [[CrossRef](#)] [[PubMed](#)]
48. Kim, T.; Chung, B.D.; Lee, J.S. Incorporating receiver operating characteristics into naive Bayes for unbalanced data classification. *Computing* **2017**, *99*, 203–218. [[CrossRef](#)]

49. Airola, A.; Pahikkala, T.; Waegeman, W.; De Baets, B.; Salakoski, T. An experimental comparison of cross-validation techniques for estimating the area under the ROC curve. *Comput. Stat. Data Anal.* **2011**, *55*, 1828–1844. [[CrossRef](#)]
50. Bamber, D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* **1975**, *12*, 387–415. [[CrossRef](#)]
51. Cortes, C.; Mohri, M. AUC optimization vs. error rate minimization. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2004.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).