Multi-label Classification of Stem Cell Microscopy Images Using Deep Learning

Adam Witmer
Department of Bioengineering
University of California, Riverside
Riverside, CA, 92521
Email: awitm001@ucr.edu

Bir Bhanu
Department of Electrical and Computer cEngineering
University of California, Riverside
Riverside, CA, 92521
Email: bhanu@ee.ucr.edu

Abstract— This paper develops a pattern recognition and machine learning system to localize cell colony subtypes in multi-label, phase-contrast microscopy images. A convolutional neural network is trained to recognize homogeneous cell colonies, and is used in a sliding-window patch based testing method to localize these homogeneous cell types within heterogeneous, multi-label images. The method is used to determine the effects of nicotine on induced pluripotent stem cells expressing the Huntington's disease phenotype. The results of the network are compared to those of an ECOC classifier trained on texture features. The ability of the network to localize cell phenotypes within heterogeneous colonies is visualized and the temporal behavior of stem cells is analyzed.
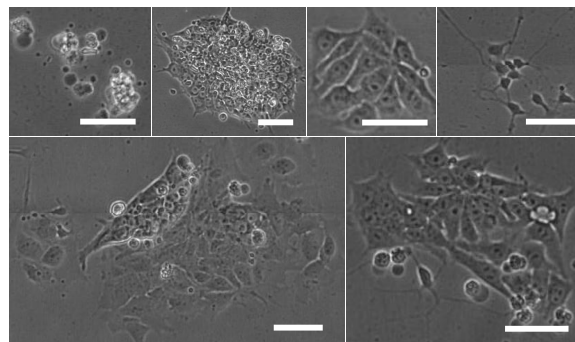
Fig. 1. Representative examples of cell colony classes (with scale bars to indicate relative size). From left to right: Top (homogeneous classes): Debris, Dense, Spread, Differentiated. Bottom (heterogeneous classes): Partially Spread, Partially Differentiated. Morphological differences can be exploited to distinguish between classes using texture pattern features. Without molecular biomarkers, it is difficult to localize cell colony subtypes within the heterogeneous classes.

I. INTRODUCTION

Huntington's disease is a rare neurodegenerative disorder caused by a mutation of the Huntingtin protein. This mutation causes a functional gain of protein activity, resulting in the degradation of neurons in the striatum of the brain. Patients typically display involuntary movements (chorea), and decline in fine-motor function, memory, and behavioral stability for many years after onset, until death [1]. There is a need to understand the underlying molecular mechanisms of the disease in order to determine effective treatment methods. One potential therapeutic agent is nicotine, which has been used to treat Parkinson's disease patients by slowing the progression of the disease and reducing associated symptoms via activation of nicotinic acetyl choline receptors in the brain [2]. Additionally, nicotine has been shown to have a neuroprotective effect in Huntington's disease patients and Huntington's rat models [3] [4]. Developmental models employing Human Embryonic Stem Cells (HESC's) are also crucial to improving our understanding of disease formation and progression.

HESC's are a reliable developmental model for early embryonic growth because of their ability to divide indefinitely (pluripotency), and differentiate, or functionally change, into any adult cell type. Their morphological behavior is indicative of developmental status and health, and can be used as a metric for classification during cellular experimentation. These cell colonies can either be homogeneous, displaying only one morphological class, or heterogeneous, displaying more than one morphological class within the colony. Time-lapse, phase-contrast microscopy is a common mode of data collection for stem cell experimentation, as it allows for non-invasive observation of temporal dynamics.

Normally, data analysis requires hours of by-hand quantification including area measurements and sorting [5]. Many advances have been made in computer aided analysis including the implementation of heuristic image processing algorithms, and classification using image features [6] [7] [10]. Additionally pixel-wise classification has been used to segment colonies in biological datasets of various forms including images and videos [11] [9] [12]. Many times, these computer vision and pattern recognition methods are combined with molecular biomarker expression in order to validate the location of cell colony subtypes and structures. Fluorescent biomarkers require that cells be sacrificed and stained before being imaged, which prevents the collection of time-lapse videos. However, using morphological colony patterns, it is possible to localize homogeneous portions within contiguous heterogenous colonies.

In this paper a patch-based deep learning classification method, trained on homogeneous images can is used to localize cellular subtypes within multi-label, heterogeneous images. Colony regions-of-interest (ROI) are first detected in high-resolution microscope images using a segmentation algorithm and ground-truth the resulting dataset into six morphological classes. A convolutional neural network (CNN) is
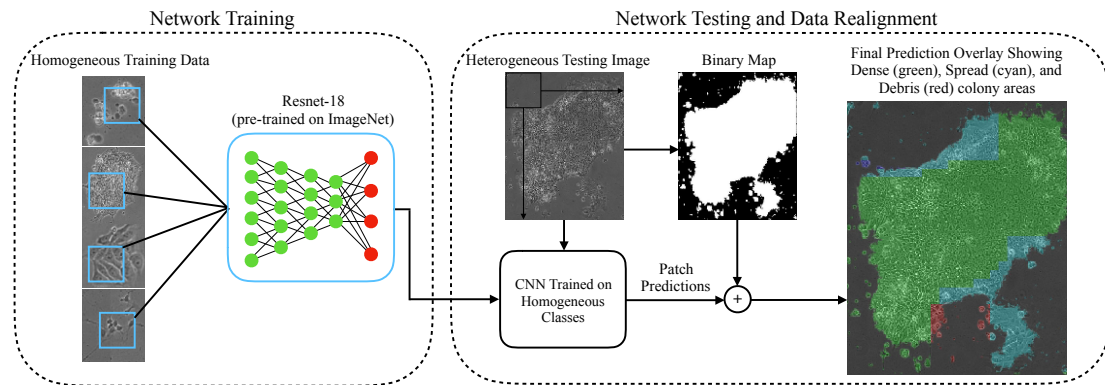
Fig. 2. Diagram of network training, testing and data realignment. A network is first trained on random crops of homogeneous cell colonies. This network is then used to classify overlapping patches of our heterogeneous testing image. Those patch predictions are combined with the corresponding binary map in order to localize colony subclasses within the original image. The results are visualized as a colored prediction map where green area represents dense portions, cyan area corresponds to spread morphology, red is debris colonies, and blue is differentiated portions.

trained on homogenous images, and used in a sliding-window testing method to recognize these cell types in heterogeneous colonies.

Classification results of two deep CNN's (Resnet18 [13], and VGG [14]) are compared to those of an SVM classifier trained on features from various state-of-the-art texture descriptors including Local Binary Pattern (LBP) [16], Gray Level Co-occurance Matrix (GLCM) [17], and Segmentation-based Fractal Texture Analysis (SFTA) [18]. Classification performance is quantified using accuracy, true positive rate, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC). The applied method addresses the issue of performing heterogeneous colony classification without the use of invasive biomarker data collection. Experimental findings regarding stem cell behavior are discussed.

II. RELATED WORK AND CONTRIBUTIONS

A. Related Work

Computer vision and pattern recognition have been used extensively for biological applications to improve the accuracy and repeatability of experimental analysis. Zahedi, et al. [7] combine morphological and dynamic features extracted from time-lapse, phase-contrast, stem cell microscopy images, to perform classification. They classify human embryonic stem cell colonies as healthy, unhealthy, or dying during exposure to varying levels of cigarette smoke. They track cell behavior over time and use a combination of morphological and temporal features to train an SVM classifier to achieve an accuracy of 97%. Similarly, Perestrelo, et. al. [10] tackle automated monitoring of induced pluripotent stem cell reprogramming using morphological segmentation and random forest machine learning classification. They achieve high accuracy (90-97%) classification of fluorescently labeled colonies from high-resolution images. More recently, deep neural networks have been implemented to automatically learn features for classification of large, biological datasets. Tasks such as segmentation and classification can be accurately combined into one network in order to avoid hand-crafting

algorithms for each dataset.

Buggenthin et. al. [8] employ deep CNNs to predict the downstream differentiation result of hematopoietic stem cells using bright field microscope images. They use a recurrent neural network to incorporate temporal information and achieve a ROC AUC metric of 87% for their two class problem. Chen et al. [9] perform cell counting using convolutional regression networks trained on synthetic fluorescent data. Van Valen et al. [11] apply deep learning to perform region detection of cell colonies using a patch based classification method. While, all of these methods incorporate automated segmentation, feature extraction, and deep learning classification, none of them perform multi-label classification of contiguous, heterogeneous cell colony images without the use of molecular biomarkers. Therefore, the following contributions for this paper are proposed:

B. Contributions

1) A novel biological dataset of grayscale, time-lapse images, with ground-truth labels is established. It will be made publicly available in the future.
2) Patch based, multi-label classification of heterogeneous colonies using features learned from homogenous samples is performed.
3) Biologically relevant conclusions are automatically drawn from stem cell experiments using pattern recognition and deep learning.

III. TECHNICAL APPROACH

A. Data & Classes

Data comes from the laboratory of Dr. Prue Talbot in the Department of Cellular and Developmental Biology at the University of California, Riverside. The experimental null hypothesis is as follows: *If there is no neuronal response of cells expressing Huntington's disease after exposure to nicotine, a significant change in morphological behavior corresponding to phenotypic changes towards the differentiated class should not*

TABLE I
CLASS DESCRIPTIONS

Class	Morphological Description	Implication
Debris	Individual cells or aggregates of circular cells with high intensity white area	Distressed, dead (apoptotic/necrotic) cell colonies or individual cells
Dense	Homogeneous aggregates of small cells with indiscernible cell boundaries, no clear nucleus	Induced pluripotent stem cell colonies
Spread	Homogeneous aggregates of large cells with discernible cell boundaries, clear nuclei, large protrusions	Down stream lineage, intermediate or progenitor cells
Differentiated	Individual cells or homogeneous aggregates of cells with distinct, dark cell body and axon like protrusion(s)	Differentiated neurons
Partially Spread	Heterogeneous aggregates of dense and spread colonies, no differentiated cells	Interacting cell colonies or pluripotent colonies changing towards down stream intermediates
Partially Differentiated	Heterogeneous aggregates of colonies that include at least one differentiated cell or colony	Interacting cell colonies or cell colonies changing to downstream neuronal lineages

be observe. To test this hypothesis, human induced pluripotent stem cells (HiPSC's) reprogrammed from a patient expressing the Huntington's disease phenotype were cultured in standard culture conditions (37°C, 90% humidity, 5% CO_2). Cultures were exposed, via culture medium, to various levels of nicotine solution (Control, 10^{-5} M, 10^{-4} M) and allowed to grow for 48 hours in the Biostation CT incubator/microscope unit [Nikon]. Five time-lapse videos were collected for each experimental condition in order to observe colony dynamics. The Biostation was programmed to automatically collect one phase-contrast microscope image of each culture well every hour for the entirety of the growth period of 48 hours, resulting in a total of 15 videos * 48 images/video = 720 images. Each image is a composite of stitched images making up a 2908 x 2908 pixel resolution field of view.

B. Ground-Truth

From these videos it is observed that various homogenous and heterogeneous cell colonies grow and change towards downstream lineages over time. While assumptions can be made about developmental status for more pronounced morphologies, such as neuron-like formations, the dataset does not contain molecular biomarker validation, and therefore phenotypes of each cell colony can not be accurately localized. Instead, colonies are classified based on morphological appearance, and inferences are made about the nature of cellular changes based on experimental findings and previous knowledge of stem cell behavior. With the assistance of experimental collaborators, the following classes for our data were determined: *Debris, Dense, Partially Spread, Spread, Partially Differentiated, and Differentiated*. Figure 1 displays example images for each morphological class, and Table I details the criteria for classification. Additionally, a hierarchical decision tree that is based on the downstream differentiation process of neurons is used to standardize image labeling. The hierarchy is as follows: *Differentiated, Partially Differentiated, Spread, Partially Spread, Dense, Debris*. Ground-truth labels for each image were provided by-hand over many months, highlighting the need for automated analysis tools. Images with more than one cell type are labeled as Partially differentiated or partially spread based on the highest ranking cell type visible in the

image. The details of our approach are provided in the next section.

TABLE II
DATA BREAKDOWN

Class Name	Number of Images
Debris	3679
Dense	4600
Differentiated	662
Spread	10458
Partially Differentiated	1435
Partially Spread	1225
Total	**22059**

C. Approach

The general analysis pipeline is outlined in Figure 2. Individual cell colonies are first segmented using morphological operations and labeled into six classes, saving the binary maps for patch based testing. Image samples from the four homogeneous classes are used to train a deep CNN and the trained network is used to gather predictions for multi-label, heterogeneous testing image patches using an overlapping, sliding-window method. The resulting image predictions are combined with the ROI data from our binary segmentation maps to localize homogeneous classes within the heterogeneous images. The final prediction map is used to measure cell colony area by class, and visualized by overlaying the prediction map on our input image (Figure 3).

1) Morphological Segmentation: A morphological segmentation algorithm is implemented to provide colony maps for every image. This approach is employed because of the lack of biomarker validation such as fluorescent labeling. The algorithm is as follows : 1. Gray scale images are smoothed using a Gaussian filter. 2. An entropy filter is used to localize textured colony areas by exploiting the difference between background and foreground intensity distributions. 3. Thresholding is used to binarize the image, segmenting high entropy areas. 4. Finally, morphological opening, hole filling, and small object removal are used to produce the final binary map. Colony area bounding boxes are cropped out and saved along with corresponding binary maps that are combined later with patch

predictions to localize colony areas. In total, 22,059 individual cell colony images are detected and labeled each image into six morphological classes, a breakdown of images by class is provided in Table II. The four homogenous class images are then used to extract features for the classifiers.

2) Feature Extraction & Classification: Different colony classes display distinct morphological patterns, as described in Table I. Of the six classes, four represent homogeneous cell colonies (Dense, Spread, Differentiated, Debris), being composed of only one cell type, and two represent heterogenous colonies (Partially Differentiated, Partially Spread), being composed of more than one cell type. A CNN is used to extract features from homogeneous colony images, and a patch based classification approach is implemented to localize colony types in the heterogeneous, multi-label, images.

Classification results of Resnet18 and VGG are networks are compared to those of multi-class error correcting output codes, support vector machine classifiers (ECOC) [15] trained on various texture descriptors including Local Binary Patterns (LBP) [16], Gray level cooccurence matrix (GLCM) [17], and Segmentation-based Fractal Texture Analysis (SFTA) [18].

3) Network Training & Data Augmentation: The neural networks and feature descriptors take fixed size, 3-channel, gray-scale images as input. For the texture feature classifier, images smaller than 224×224 are scaled to have a smallest size of 224 (while maintaining aspect ratio) and then train and test on normalized, center crops from the datasets. During neural network training, images are resized in the same manner, and random 224×224 crops are randomly flipped horizontally and vertically with independent probability of 0.5. For the ECOC classifier, data is split with an 80:20 train:test split, and for Resnet18 and VGG11 networks, 80:10:10, train:validate:test split is used, all with 10-fold cross validation. These networks are sufficiently complex to model the experimental data and allow for faster training times than extremely deep networks such as Resnet50. ECOC classifiers are trained to convergence, and early stopping is implemented during neural network training. The neural network accounts for data imbalance using a weighted cross-entropy loss function, Equation 1. The softmax multinomial distribution outputs a network prediction probability, x, of class, c, for which the negative log likelihood loss is multiplied by the weight proportion, w, of class c in the training dataset, magnifying the effect that the sample has on network learning. The network is trained using Stochastic Gradient Descent on mini-batches of 128 images with a learning rate of 0.001, weight decay regularization of 0.0001, momentum of 0.9, reducing the learning rate by a factor of ten every 50 epochs (values determined via parameter search). The results of our classifiers are discussed in the following section.

$$Loss(x, c) = w(c) * (-x(c) + \log(\sum_j e^{x_j}))$$ (1)

IV. RESULTS

A. Classification Results

The results of the CNN classifiers are compared to those of an ECOC classifier trained on features from three state-of-the-art texture descriptors (Table III). The CNN's out perform the texture-descriptor trained ECOC classifiers by more than 25%. Resnet and VGG both perform very similarly, achieving an overall accuracy of 89%. The texture/classifiers show much lower accuracy, and a lower true positive rate. In contrast, the CNN's maintain a high true positive rate, and also serve as more robust classifiers, having ROC-AUC's of approximately 86% (Figure 5). Table IV details associated confusion matrices for homogeneous classification. The trained network is then used to perform patch testing over our whole multi-label images in order to localize sub classes in heterogenous colonies.

B. Patch Testing

After the whole image is processed, pixel areas for each subclass are measured and incremented based on the time stamp of the image. Figure 3 shows the resulting prediction maps for our heterogenous classes at various time points. Temporal behavior changes of the control and experimental colonies are observed. For the partially differentiated class, the colony at the 23 hr mark shows a small differentiated area (blue) that spreads and detaches over time, being connected only by a single axon at the end of experimentation. For the Partially spread class, cells on the outer edge of the colony begin to change before cells in the center, which is a common trait of differentiating colonies.

Table IV details the classification accuracy of the patch based method for the heterogeneous, multi-label images. The network localizes cell subtypes in heterogeneous images with a true positive rate of 67% for the partially spread class, and 27% for partially differentiated. The majority of negative classifications fall into the differentiated class, indicating that the network is able to localize the differentiated area in these multi-label images with low proportions of spread of dense area. There is also a tradeoff between stride parameter and computation time, as decreasing the stride parameter increases accuracy but also results in prohibitively expensive computations.

C. Experimental Results

Figure 4 displays the change in colony area over time by class, for each experiment, as determined during patch testing. The experimental null hypothesis is: *If there is no neuronal response of cells expressing Huntington's disease phenotype after exposure to nicotine a significant change in morphological behavior corresponding to phenotypic changes towards differentiated lineages should not be observed.* Experimental observations confirm the null hypothesis, as a large change in total differentiated colony area is not observed. The growth rates of spread colony areas increase in a similar manner for all experimental conditions, and the growth rate of dense areas is greater under exposure to nicotine, than in control conditions. Coupled with the stable levels of debris colony

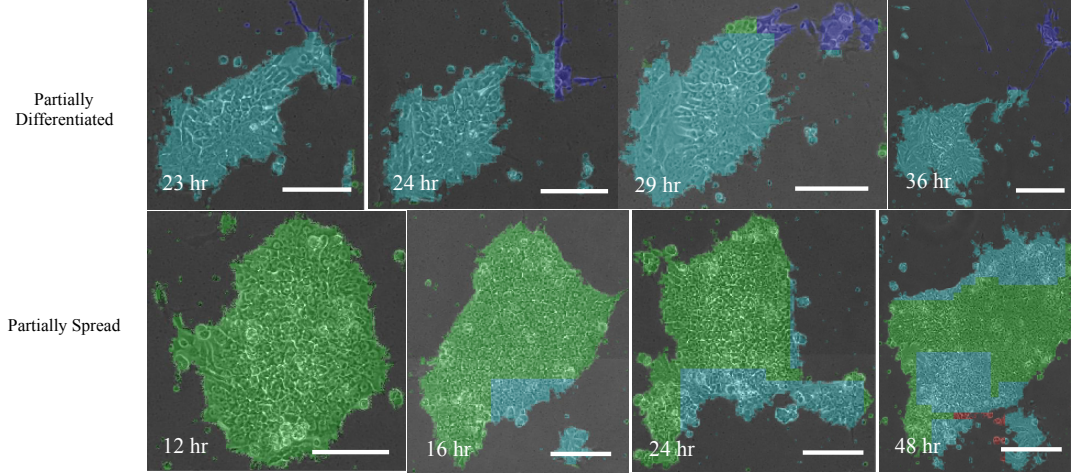| Method	Accuracy ± std. dev. (%)	True Positive Rate ± std. dev. (%)	ROC AUC ± std. dev.	Train+Test Time (min)
VGG	0.8935 ± 0.0111	0.8788 ± 0.0141	0.8635 ± 0.0123	74.000
Resnet	0.8916 ± 0.0064	0.8995 ± 0.0078	0.8518 ± 0.0144	84.618
GLCM + ECOC	0.7414 ± 0.0057	0.5553 ± 0.0041	0.6941 ± 0.0115	9.829
LBP + ECOC	0.7160 ± 0.0072	0.4976 ± 0.0467	0.6695 ± 0.0097	9.211
SFTA + ECOC	0.4203 ± 0.0981	0.3858 ± 0.0915	0.5847 ± 0.1269	191.691

Fig. 3. Final image class prediction overlaid on original image, for both classes, over time. Changes are observed from spread (cyan) to differentiated (blue) in the partially differentiated colonies (top) and from dense (green) to spread in the partially spread colonies (bottom) with areas of debris (red) located in the bottom right image.
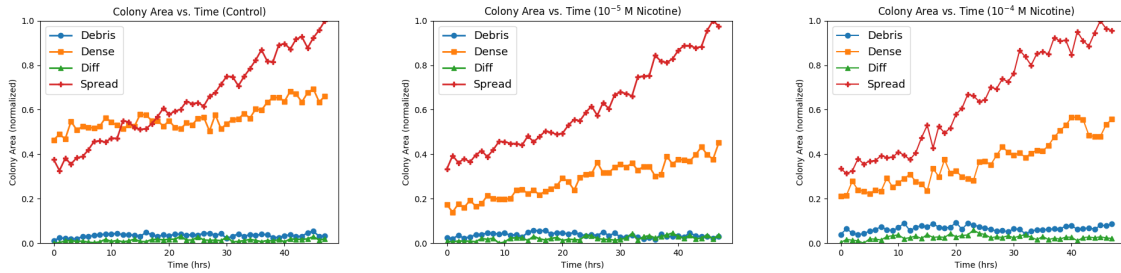

Fig. 4. Graphs of normalized colony pixel area over a 48 hour period. A linear increase in spread and dense cell colony areas is magnified under exposure to nicotine. However, significant changes in the total area of differentiated colonies over time are not observed, indicating that nicotine does not have a neuronal effect on Huntington's disease stem cells.
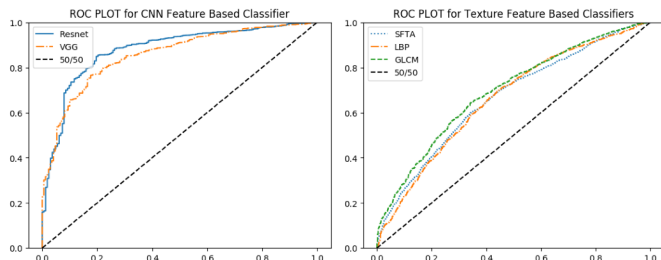

Fig. 5. Graphs of Receiver Operating Characteristic Curves for Classifiers

measurements, these findings indicate that nicotine does not have toxic effect on overall colony health or development, and is directing colonies away from pluripotency, towards intermediate progenitor phenotypes associated with the spread morphology.

V. CONCLUSION & DISCUSSION

In this paper, deep learning and pattern recognition are used to classify stem cell colony ROI crops into four homogeneous, morphological classes that have implications on cellular behavior and phenotype. The deep CNN method outperforms multi-support vector machine error correcting output codes classifiers trained on texture feature vectors by over 25%, achieving a ROC AUC of 87%. The use of a CNN model allows circumvents the need to hand-craft feature maps for each data type and is robust against changes in illumination

TABLE IV
CONFUSION MATRIX FOR CLASSIFICATIONS OF PARTIALLY DIFFERENTIATED AND PARTIALLY SPREAD TESTING IMAGES WITH PERCENT OF TOTAL

Class	Stride Size	Partially Differentiated (%)	Partially Spread (%)	Dense (%)	Differentiated (%)	Spread (%)	Debris (%)
Part Differentiated	64	343 (23.90)	106 (7.38)	296 (20.62)	491 (34.21)	158 (11.01)	41 (2.85)
Part Differentiated	32	396 (27.59)	96 (6.68)	281 (19.58)	465 (32.40)	161 (11.21)	36 (2.50)
Part Spread	64	233 (19.03)	766 (62.58)	0(0)	195 (15.93)	39 (3.18)	1 (0.08)
Part Spread	32	187 (15.27)	827 (67.56)	0 (0)	150 (12.25)	59 (4.82)	1 (0.08)

TABLE V
CONFUSION MATRICES ASSOCIATED WITH RESNET18 ROC CURVE AT
VARIOUS CLASSIFICATION THRESHOLDS

Threshold = 0.2	Debris	Dense	Differentiated	Spread
Debris	367	6	1	3
Dense	4	378	0	32
Differentiated	0	0	70	6
Spread	5	17	4	814
Threshold = 0.7	Debris	Dense	Differentiated	Spread
Debris	355	12	1	13
Dense	7	365	0	65
Differentiated	0	0	66	10
Spread	14	33	8	767

and contrast. the A patch based classification method is implemented to accurately detect cell colony subtypes within multi-label, time-lapse, gray scale images. Prediction maps of multi-label images are visualized and the localized areas of each class are measured over time across the entire dataset. Experimental findings indicate that nicotine has a minimal neuronal effect on Huntington's disease induced pluripotent cells, as a large increase in differentiated colony area is not observed in either under control or experimental conditions. A steady increase in dense and spread colony areas implies that nicotine does not have a toxic effect on the cells at the experimental concentrations, and stimulates colony growth at low levels. Without the use of molecular biomarkers, it is difficult to localize cell area within heterogeneous colonies. Experimental analysis for which ground-truth is not provided via molecular biomarkers is often tedious, biased and inaccurate, taking many months to perform by-hand. While this work allows for the automated detection of cellular behavior changes using pattern recognition and deep learning, the results can be improved in the following ways: 1. incorporating temporal information; 2. testing on other datasets to validate model robustness; 3. testing multiple custom network configurations and comparing to standard methods (comprehensively); 4. performing statistical analysis of experimental findings. Along with the completed work, these improvements lend to increasing the accuracy and reliability of biological data analysis involving cellular microscopy.

ACKNOWLEDGMENT

REFERENCES

[1] F. Walker, M.D. *Huntingtons disease.* The Lancet, 2007
[2] M. Quik. *Smoking, nicotine and parkinson's disease.* TRENDS in Neurosciences, 2004.
[3] A. McGregor, J. Dysart, M. Tingle, B. Russell, R. Kydd, G.Finucane. *Varenicline improves motor and cognitive symptoms in early Huntington's disease.* Neuropsychiatric Disease and Treatment, 2016.
[4] M. Tariq, H. Khan, I. Elfaki, S. Al Deeb, H. Al Moutaery. *Neuroprotective effect of nicotine against 3-nitroproionic acid(3-NP)-induced experimental Huntington?s disease in rats.* Brain Research Bulletin, 2005.
[5] P. Talbot, N. zur Nieden, S. Lin, I. Martinez, B. Guan, B. Bhanu. *Use of Video Bioinformatics Tools in Stem Cell Toxicology.* Handbook of Nanotoxicology, Nanomedicine and Stem Cell Use in Toxicology, 2014.
[6] B.Guan, B. Bhanu, P. Talbot, S. Lin. *Bio-Driven Cell Region Detection in Human Embryonic Stem Cell Assay.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2014.
[7] A. Zahedi, V. On, S. Lin, B. Bays, E. Omaiye, B. Bhanu, P. Talbot. *Evaluating Cell Processes, Quality, and Biomarkers in Pluripotent Stem Celss Using Video Bioinformatics.* PLOS one, 2016.
[8] F. Buggenthin, F. Buettner, P. Hoppe, M. Endele, M. Kroiss, M. Strasser, M. Schwarzfischer, D. Loeffler, K. Kokkaliaris, O. Hilsenbeck, T. Schroeder, F. Theis, C. Marr. *Prospective identification of hematopoietic lineage choice by deep learning.* Nature Methods, 2017.
[9] O. Ronneberger, P. Fischer, T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation.* MICCAI, 2015.
[10] T. Perestrelo, W. Chen, M. Correia, C. Le, S. Pereira, A. Rodrigues, M. Sousa, J. Ramalho-Santos, D. Wirtz. *Pluri-IQ: Quantification of Embryonic Stem Cell Pluripotency through an Image Based Analysis Software.* Stem Cell Reports, 2017.
[11] D. Van Valen, T. Kudo, K. Lane, D. Macklin, N. Quach, M. DeFelice, I. Maayan, Y. Tanouchi, E. Ashley, M. Covert. *Deep Learning Automates the Quantitative Analysis of Individual Celss in Live-Cell Imaging Experiments.* PLOS Computational Biology, 2016.
[12] W. Xie, A. Nobel, A. Zisserman. *Microscopy cell counting and detection with fully convolutional regression networks.* Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2016.
[13] K. He, X. Zhang, S. Ren, J. Sun. *Deep Residual Learning for Image Recognition.* arXiv:1512.03385v1, 2015
[14] K. Simonyan, A. Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition.* International Conference on Learning Representations, 2015.
[15] J. Furnkranz. *Round Robin Classification.* Journal of Machine Learning Research, Vol. 2, 2002, pp. 721?747.
[16] T. Ojala, M. Pietikainen, T. Maenpaa *Multiresolution Gray Scale and Rotation Invariant Texture Classification With Local Binary Patterns.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.
[17] R. Haralick, K. Shanmugan, I. Dinstein, *Textural Features for Image Classification.* IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-3, 1973, pp. 610-621.
[18] A. Ferraz Costa, G. Humpire-Mamani, A. Juci Machado Traina. *An Efficient Algorithm for Fractal Analysis of Textures.* IEEE Conference on Graphics, Patterns and Images, 2012.