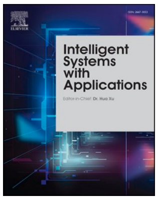
Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications



Attention-effective multiple instance learning on weakly stem cell colony segmentation

Novanto Yudistira [a], Muthu Subash Kavitha [b,*], Jeny Rajan [c], Takio Kurita [d]

[a] Informatics Department, Faculty of Computer Science, Brawijaya University, Jalan Veteran 8, Malang, 65145, Malang, Indonesia
[b] School of Information and Data Sciences, Nagasaki University, Nagasaki city 852-8521, Nagasaki, Japan
[c] Department of Computer Science and Engineering, National Institute of Technology, Karnataka, Surathkal, India
[d] Graduate School of Advanced Science and Engineering, Hiroshima University, Hiroshima, Japan

ARTICLE INFO

Keywords:
Multiple instance
Weakly supervised segmentation
Colony
Annotation
Inference

ABSTRACT

The detection of induced pluripotent stem cell (iPSC) colonies often needs the precise extraction of the colony features. However, existing computerized systems relied on segmentation of contours by preprocessing for classifying the colony conditions were task-extensive. To maximize the efficiency in categorizing colony conditions, we propose a multiple instance learning (MIL) in weakly supervised settings. It is designed in a single model to produce weak segmentation and classification of colonies without using finely labeled samples. As a single model, we employ a U-net-like convolution neural network (CNN) to train on binary image-level labels for MIL colonies classification. Furthermore, to specify the object of interest we used a simple post-processing method. The proposed approach is compared over conventional methods using five-fold cross-validation and receiver operating characteristic (ROC) curve. The maximum accuracy of the MIL-net is 95%, which is 15% higher than the conventional methods. Furthermore, the ability to interpret the location of the iPSC colonies based on the image level label without using a pixel-wise ground truth image is more appealing and cost-effective in colony condition recognition.

1. Introduction

Induced pluripotent stem cells (iPSC) can self-renew infinitely and generate into every human body's cell type. The iPSCs are helpful to substitute deteriorated tissue of the human body and thus it is highly demanded clini- cal drug development (Takahashi et al., 2007). To realize reliable and secured tissue regeneration, it is essential to determine the conditions of the cells during their culture. Identifying the good quality cells and colonies (cluster of identical cells) for subsequent treatment therapy is generally observed by the eye in terms of the morphological features, such as colonies with a densely packed cell appearance and almost a well-defined edge. On the contrary to the morphology of the excellent quality cells, harmful quality colonies are detected. However, manual evaluations of cell conditions highly rely on human experts and cost errors (Fan et al., 2017). Furthermore, the assessment of a massive amount of cell conditions in culturing is tedious and laborious. Hence non-invasive automatic classification technique would benefit from tracing large numbers automatically interestedly without any classification errors.

On the other hand, MIL (Multiple Instance Learning), is a machine learning technique that involves learning from a dataset of labeled bags, and each bag containing multiple instances (examples). In the biological and medical fields, MIL has been applied in a variety of contexts, including drug discovery, protein function prediction, and disease diagnosis. Several applications of MIL in medical images have shown outperforming classical training approach. Melanoma detection using color and texture features (Fuduli et al., 2019), by means of MIL to discriminate melanomas and common nevi (Astorino et al., 2020), viral pneumonia images classification by MIL (Zumpano et al., 2021), and diabetic retinopathy images classification via MIL (Vocaturo and Zumpano, 2021) are some of proposed MIL in this field that shows outperforming classical learning.

Motivated by the studies mentioned above, we intended to utilize the effectsof MIL through a weakly supervised approach, where the binary image- level label (colony with dense cells as good /sparse cells as bad) is given to the group of instances. However, the aforementioned MIL-based CNN archi- tectures extracted local to global features from the multiple non-linear layers limiting the performance by insisting on the

* Corresponding author.
E-mail address: kavitha@nagasaki-u.ac.jp (M.S. Kavitha).

https://doi.org/10.1016/j.iswa.2023.200187
Received 14 April 2022; Received in revised form 27 December 2022; Accepted 14 January 2023
Available online 24 January 2023

intensity profile with shape features. In order to improve that, we intend to push the local colony structure information in the MIL-net by highlighting the essential features for colony conditions. MIL is used to determine a sample when all of the instances from a sample must be taken into account without any specific pixel label to each of the instances. In our case, MIL attempts to discover the target variable from the instances of sparse and dense patterns of stem cells by extracting the feature maps through several convolution layers and transformed into a low dimensional space. There it can generate a single bag level representation using average weighted pooling and classifies the bag into good or bad colony image. Furthermore, supervised learning demanding a large amount of annotated images, which are tedious and time-consuming. Alternatively, the proposed approach based on a U-net-like structure to predict the pixel-level segmentation with the boundaries of the colonies without a finely-labeled sample is promising in cell detection.

The contribution of this study can be summarized as follows:

1) Proposes a multiple instance approach in form of weakly supervised for iPSC colony segmentation and colony conditions classification based on Unet-like architecture in an end-to-end manner without using finely-labeled samples.
2) Involves simple post-processing in the learning output to automatically specify region of interests and removes the unwanted pixel localizations as false positives.
3) Compares the performance of the proposed framework over architectures that includes U-net with fully connected layer (hereafter termed as baseline), patch-based shallow U-net, ResNet-50, deep V-CNN and SVM.
4) Investigates the performance of the proposed model using five-fold cross-validation.
5) Evaluates the performance of all architectures by using mean accuracy, precision, recall, F-score and receiver operating characteristic curve (ROC) measures.

2. Related works

2.1. State-of-the-art method in iPSC classification

Several automated techniques have been developed to classify various conditions of iPSCs (Yuan-Hsiang et al., 2017; Kavitha et al., 2017; Joutsijoki et al., 2016). Several research works using digital image processing techniques exploiting preliminary filtering and thresholding to detect the shape of the objects of the colonies (Kavitha et al., 2017; Chen and Zhang, 2009). However, the feature assessment using image analysis techniques depends on prior parameters and manual interactions, prone to large-scale assessment errors (Kavitha, et al., 2020; Kavitha, et al., 2016). Furthermore, the morphology of colonies is dynamically changed in subsequent reprogramming stages. Thus prior parameter setting approaches were not appropriate for evaluating the colonies (Nagasaka et al., 2017; Kato et al., 2016). In order to alleviate manual interaction, few approaches used machine learning techniques. However, machine learning methods relied on hand-crafted microscopic morphology-based and texture-based features of colonies to classify cell conditions (Joutsijoki et al., 2016; Stumpf and MacArthur, 2019; Zhang et al., 2019). Specifically, hand-crafted features-based support vector machine (SVM) models were commonly applied and produced satisfactory results for the classification of conditions of colonies (Raytchev et al., 2016; Kavitha et al., 2018).

Recently deep learning methods are extensively used in detecting cell images because of the ability to recognize the changes and development of stem cells without manual interventions (Waisman et al., 2019; Moen et al., 2019). The open-source package and Xception network were effective in differentiating the types of neural stem cells (Zhu et al., 2021). A vector-based convolutional neural network (V-CNN) was developed and matrix transformation is added as a pre-processing layer in the two- dimensional CNN (Kavitha et al., 2017). The authors in Kavitha et al. (2017) claimed that the V-CNN produced better performance than SVM for the classification of colonies. A simple LeNet architecture with an image processing algorithm efficiently derived cell types from iPSCs with high performance (Kusumoto et al., 2018). However, the methods mentioned above highly relied on the number of pre-processing ways to locate most related features for iPSC colony classification. The pre-processing steps are often problem-specific and required prior parameter settings, which is not always appropriate for evaluating the variations in iPSCs heterogeneity.

Thus we intend to develop a single model without pre-processes for reducing the risk of biased results and inconsistencies for colony conditions evaluation. We used a customized version of the popular encoder-decoder based U-net architecture (Ronneberger et al., 2015). Previously, several biological imaging tasks have been utilized U-net or attention mechanism for segmentation due to its ability to capture coarse-to-fine structures (Yudistira et al., 2020; Du et al., 2020; Oktay et al., 2018). Differently in this study, we proposed to use an end-to-end MIL-net without pre-processing that implements local connectivity patterns between the neurons of the adjacent layers and average pooling for the attention features at the end of the architecture. MIL is a specific type of supervised learning, where instances are grouped into sets, termed as bags, and labels are only given at the bag level and not for each individual instance level. In our case, MIL attempts to discover the target variable from the instances of sparse and dense patterns of stem cells. The instances of stem cells are extracted through several convolution layers and transformed into a low dimensional space. There it can generate a single bag level representation using average weighted pooling with highest attention to show landmarks of stem cell region as well as classifies the bag into good or bad colony image. The classical global pooling methods can only detect approximate pixel location, and thus, global weighted average pooling was used to evaluate the pixel level localization (Qiu, 2018). A fully convolutional neural network trained with fewer ground truth bounding boxes and many image-level labels was found to be effective in locating the pixel-level objects on the benchmark datasets (Qiu, 2018). Attention gating as Sononet was used in VGG or U-net to detect salient regions on the medical images (Schlemper et al., 2019). Attention mechanism using GRADCam in U-net with logistic regression classifier enhanced Alzheimer's disease classification (Kavitha et al., 2019). A modified 25-layers of U-net was effectively used to diagnose cardiac arrhythmia based on the electrocardiographic signals (Oh et al., 2019). A weakly-supervised approach using feedback CNN and global average pooling with binary labels was used to locate the satellite images (Liu et al., 2016). The attention mask generated from the attention U-net improved the iris region detection (Lian et al., 2018).

2.2. Applications of MIL in medical field

One of the applications of MIL in the medical field is disease diagnosis (Fuduli et al., 2019; Astorino et al., 2020; Zumpano et al., 2021; Vocaturo and Zumpano, 2021). For example, a bag might represent a patient's medical record, and the instances within the bag represent different symptoms or test results. The label for the bag indicate whether the patient has a particular disease. By training a MIL model on a dataset of labeled bags, it can learn to predict the presence of a disease based on the patient's symptoms and test results.

Some examples of recent works in which the MIL approach has been shown to perform well in the medical field have been proposed. In traditional machine learning, MIL has been incorporated to improve classification performance. In the study of Melanoma detection using color and texture features (Fuduli et al., 2019; Astorino et al., 2020), the authors used MIL to classify images of skin lesions as benign or malignant (melanoma). They found that MIL outperformed traditional machine learning approaches regarding accuracy and sensitivity, such as support vector machines and decision trees. In the study of viral pneumonia images classification (Zumpano et al., 2021), the authors used MIL to classify chest X-ray images as normal or abnormal (pneumonia). They found that MIL achieved higher accuracy compared to a traditional machine learning approach, such as support vector machines. In the study of Diabetic Retinopathy classification, the authors (Vocaturo and Zumpano, 2021) used MIL to classify retinal images as healthy or containing diabetic retinopathy. They found that MIL outperformed other methods, such as support vector machines and decision trees, in terms of accuracy. There are many other studies that have used MIL for tasks such as cancer diagnosis, disease prediction, and image classification, and have demonstrated the potential of this approach for solving complex, real-world problems.

3. Materials and methods

3.1. Dataset

This study included a set of 94 images of iPSC colonies. Out of 94 images, 60 were maintained as described elsewhere (Okita et al., 2007) and 34 were received from American Type Culture Collection. The details of gathering the iPSC colonies and phase contrast microscopic image collection settings are explained in our previous study (Kavitha et al., 2017). We also used another dataset namely Raabin-WBC (Kouzehkanan et al., 2021), which is the large-scale white blood images data. The dataset contains granulocytes (neutrophils, basophils, and eosinophils), lymphocyte, and monocytes classes. 1145 cropped images including 242 lymphocytes, 242 monocytes, 242 neutrophils, 201 eosinophils, and 218 basophils were randomly selected, and their ground truths were extracted by experts. Each of the image from 1145 selected cells has ground truth of the nucleus and cytoplasm. In order to extract the nucleus ground truth, image processing tricks were used. The ground truth of the whole cell is prepared for basophils, without the basophils' cytoplasm and nucleus. This makes it challenging for MIL to extract cytoplasm and nucleus level ground truth.

3.2. Data augmentation

The gray scale images of iPSC input images are fed into the network. Thetraining data are augmented thus there is increase in number and variation. We used augmentation images involved with vertical or horizontal flip and rotation of 90, 180 and 270° degrees. To resolve imbalanced classes within training data, the minority class is oversamped, eventhough the level of imbalance is not severe.

3.3. Proposed MIL for discriminative cell patterns for colony condition

MIL is a type of machine learning approach in which the learning algorithm is presented with a set of labeled bags, each of which contains a set of instances (e.g., images, text documents, etc.). The label for a bag is determined by the labels of the instances it contains, rather than the label being assigned directly to each instance. The learning algorithm would need to identify the relevant features of the object in each image in the bag in order to correctly classify the bag. MIL is often used in situations where it is difficult to label individual instances accurately, but it is still possible to label groups of instances (e.g., bags) based on the majority label of the instances they contain. This can be useful for tasks such as object recognition, where the object of interest may appear in different positions, scales, or orientations within an image, making it challenging to label each instance individually. The specific approach used will depend on the characteristics of the data and the specific task being addressed. MIL can be formulated as a supervised learning problem, where the goal is to learn a function that can map a set of instances to a label. The instances are grouped into bags, and the label for a bag is determined by the labels of the instances it contains. One common formulation of MIL uses a function *f* that maps a bag *X* to a label *y*:

$$f : X- > y \tag{1}$$

where *X* is a bag containing a set of instances $\{x_1, x_2, \ldots, x_n\}$, and *y* is the label for the bag. The bags can be termed as relation between instances. Hence the positive bag included positive instances and negative bag included negative instances with label $Y = \{y_1, y_2, \ldots y_n\}$ of $+1$ and -1, respectively. Then MIL label follows the equation as,

$$X(\omega) = \begin{cases} +1 & \text{if exists } y_i : y_i = +1 \\ -1 & \quad \text{if Otherwise} \end{cases} \tag{2}$$

In this formulation, the goal of the learning algorithm is to find the function *f* that best approximates the true underlying relationship between the bags and their labels. This can be done using a variety of techniques, such as supervised learning algorithms that optimize a loss function using gradient descent. One way to formulate MIL using a supervised learning algorithm is to use a neural network with multiple hidden layers and an output layer. The input to the network would be the set of instances in a bag, and the output would be the predicted label for the bag. The weights and biases of the network could then be adjusted based on the errors made during training, using an optimization algorithm such as stochastic gradient descent.

To learn the function *f*, we can use a supervised learning algorithm that is trained on a labeled dataset of bags. The algorithm adjusts the parameters of the function *f* (e.g., the weights and biases of a neural network) based on the errors made during training. One way to express the learning objective for MIL is to use a loss function *L* that measures the difference between the predicted label y' and the true label *y* for a given bag:

$$L = L(y', y) \tag{3}$$

The goal of the learning algorithm is to find the parameters of f that minimize the average loss over the training dataset. This can be done by using an optimization algorithm such as gradient descent to adjust the parameters of f based on the gradient of the loss function with respect to the parameters. Colony condition recognition is a typical binary image classification problem for a learning algorithm. Consider *X* is the input image, and *N* represents the total number of classes. Hence, $\mathrm{L} \in \{1, \ldots, N\}$ is the corresponding class label of X. The training algorithm proposes to find a function $f : \mathrm{X} \rightarrow \mathrm{L}$. In conventional image classification frameworks for colony condition detection, *f* is often defined as $G(E(X))$, where $E(X)$ and $G(.)$ indicate the feature extractors and classifiers, respectively. We used the criterion of bags and instances in the MIL settings. In this study, a single channel 250X250 pixels size of good and bad quality iPSC colony images are used with their categorical labels to train the model.

3.4. Multiple instance classification

Image level labels or bag labels of input data is used to generate pixel locations or instance labels of the cell area. The bag labels consist of two classes of images in this study. Furthermore, in this study we did not label the exact location or each pixel in the stem cell regions, instead image-wise good and bad labels are used to train the network. However, each pixel locations were generated from the network weights that trained after the backpropagation by using average activation. There, it can automatically find the region of interest by visualizing the stem cell colonies and it is indicated as a weak segmentation in this study. Furthermore, the highest attention that contributing the cell region enhance the performance in classifying the conditions of the good and bad colonies. The steps involved in the proposed MIL-net are; classification of the classes of good or bad iPSCs until convergence and weakly supervised segmentation of the cell colonies retrieve from the last convolutional layer of the network.

3.5. Proposed architecture of U-Net based MIL

In this study, we propose a U-net-like CNN architecture that automatically finds regions of interest and differentiates between different cell conditions of iPSC, such as good and bad, in an end-to-end fashion (Fig. 1). U-Net has been shown to perform well on a variety of images. Specifically, it has a symmetrical architecture that effectively localizes objects or features in an image. Moreover, U-net-like architecture is relatively simple and can be easily trained using standard CNN training techniques. This is useful for tasks such as identifying the boundaries of objects or cells in biomedical im- ages. The feature maps of decoder are concatenated with the feature maps skipped from the encoder through the skip connections. Thus, it can able to retrieve the full spatial resolution at the network output. Furthermore, it added an average pooling layer at the end of the network layer to enhance the classification capacity of the CNN. The proposed architecture is designed by large receptive fields of the output neurons, essential for multiple instance classifications. The new addition of average pooling and fully connected layer at the final layer complement their plain counterparts in the classical U-net architecture. Therefore, it is trained by leveraging and back-propagating the network to give classification results.

The network implemented $m \times m$ convolution filters for features extraction and dimension reduction. The softmax is used to find the probability of the colony conditions followed by the average pooling and fully connected layer. The encoder and decoder structure combines the feature maps. The encoder consists of eight layers of convolutions and leaky ReLU. Each convolutional layer has 4 × 4 kernels with stride 2. The encoder part starts with 64-dimensional features or channels and it increases until to reach 512 feature maps. In the encoder, the input x convolved with filter of w^c and residual bias b_{cl}, where c is channel number and l is layer number, before fed into non-linear activation function of f (leaky ReLU). It is defined as

$$y^l_{ij} = f\left(\sum_{p=0}^{m}\sum_{q=0}^{m} w^c_{p,q} x^{l-1}_{(p+i)(q+j)} + b_{cl}\right) \tag{4}$$

The convolutions are done spatially in 2-dimensional space, where p and q are width and height of the input, respectively. The previous output layer x^{l-1} is convoluted and activated to produce the activation output of y^l.

The decoder path consists of eight up-sampling convolutions called de-convolution layers (Yang et al., 2021) and ReLU activation functions. Drop out operation is added on first three up-sampling convolutions that is after ReLU. It maximizes feature maps by 4, and minimizes the number of features dimension by half. If there is any negative activations, ReLu returns zero and hence the gradient become zero for all the inputs to the following layers. However, Leaky ReLu returns very small value for any negative inputs. Hence, we used Leaky ReLu in the encoder and ReLu in the decoder. The output of the last up-sampling layer is passed into the average pooling layer followed by fully connected layer. The high-level feature vectors of the last convolution layer are derived from low to high layers are passed into the average pooling. *A*verage pooling helped to reduce the number of parameters as well as make the features invariant to varying locations, rotations, and scales that beneficial for generalization. Hence, the precise attention and compact features derived from the average pooling flow into the softmax cross entropy for classification can be capable to maintain the most relevant features for stem cell region and that enhance the network efficiency in categorizing the classes of colony conditions. The attention of the average pooling is used to localize the region of interest for visual interpretation and termed as weakly localization of the stem cell regions in this study.

The filters in the decoder are also trainable parameters. The output x of the previous layer is transposed and convoluted with filter of w^c with the bias value of b_j before nonlinear activation function of f_1. The total number of trainable parameters obtained from eight convolution layers of encoder and seven deconvolution layers of decoder are 54,653,008. The last convolution in the decoder is used for the reconstruction to the original size.

Furthermore, in the concepts of MIL-based CNN architectures, f consisted of multiple non-linear layers with convolutional layers, each followed by pooling and one fully connected and softmax classifier to extract local to global features. However, in the process of colony condition recognition, the local definite structure information in the cell colony is often asymmetrically distributed. Hence, learning global features from the conventional multiple non-linear layers-based CNN limited the performance in distinguishing the colony condition. Furthermore, it demands the intensity profile along with the shape features. To resolve this issue, we constructed an end-to-end U-net like deeper CNN architecture adopted to MIL framework by pushing the local colony structure information for colony condition detection. More of interest, automatically partition the informative local colony contour according to the conditions of the colonies without finely labeled samples are promising in cell detection.

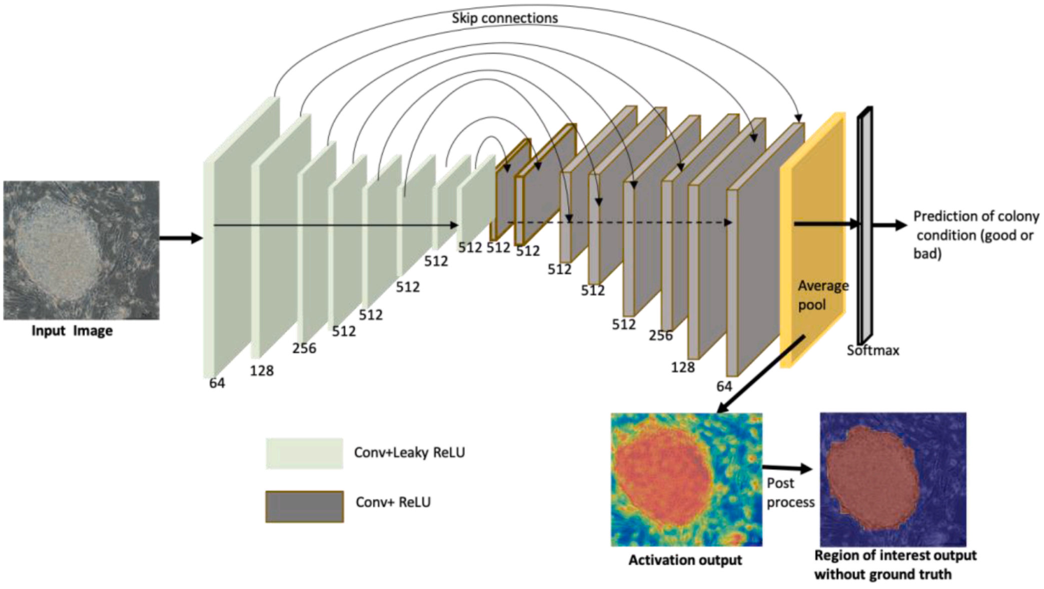


Fig. 1. Proposed attention-effective learning network for induced pluripotent stem cell colony conditions classification.

The proposed framework fits into MIL criterion and consists of train encoder and decoder stages (Fig. 1). We trained our proposed architecture independently from scratch. No pretraining or transfer learning was used in any of our experiments in this study. It fits into the weakly supervised learning in which input data is labeled as good/bad called bags (Zhi-Hua et al., 2012; Foulds and Frank, 2010).

Furthermore, the direct localization via classification using the MIL-net based CNN method is evaluated by removing the average pooling with three fully connected layers and patch-based shallow number of layers for the conditions of the iPSC colony. The input and the number of neurons in the three fully connected layers are 5000, 1000 and two, respectively, corresponding to the good and bad conditions of colony. For comparison, additionally we built a MIL-net using patch based input. It is evaluated to reduce the computational burden. It included shallower layers of three convolutions and deconvolutions with average pooling at the final layer before the softmax.

3.6. *Weakly supervised visualization*

We used the softmax cross-entropy loss function by learning from image-level labels to train the proposed network. The proposed MIL-net directly learns the input and output relationship of different colony conditions. The classification of the colonies is learned from the network weights that learn after the convergence. The intermediate activation output of the last convolutional layer is used to visualize the localization of the region of interest. The weakly supervised learning of the proposed network visualizes the texture features of the colonies by minimizing the irrelevant neuron activation. The output investigates distinctive textures that identify the condition of the colonies.

3.7. *Post-processing*

The colony region obtained from the learning is not clearly delineated and still included some outliers which are not important for the detection. Particularly the neural network is highly uncertain with less the number of training data and thus this condition often occurs. Hence a simple post-processing step is needed to remove the unwanted pixel localizations. In order to achieve this, we used the morphological operation such as an opening that keeps the largest localized object visible. The remaining objects are removed as false positives. The opening operation consists of erosion and then dilation with structuring element or kernel. We used rectangle-based kernel of z with 20 × 20 in size by experiment. The formulation of opening operation is shown as follows:

$$S \circ z = (S \ominus z) \oplus z \tag{5}$$

Where S and z are input and kernel, respectively. The parameter z performs erosion morphological operation ($\ominus$) on input S. And then dilation morphological operation ($\oplus$) is performed on S. These consecutive operations removed the small noisy artifacts and keeps iPSC colony region as the region of interest.

4. Experimental setup

4.1. *Experimental evaluation settings*

The proposed MIL-net classification method is applied to iPSC of good and bad colonies to evaluate its utility and effectiveness. Out of 94 images, the number of good and bad colonies used in this study are 54 and 40, respectively. The dataset is randomly partitioned into 74 training and 20 testing images without using any same instances in both train and test set. The good and bad colonies are 44 and 30, respectively in the training and 10 and 10, respectively in the testing set. To avoid overfitting we applied several regularizer methods throughout the network during training phase. In our architecture, we used dropout at the first three deconvolutional layers, batch normalization at Conv1 to DeConv7, ReLU for all layers, and weight decay. Dropout is important to prevent co-dependent neuron units and thus only the key properties are selected within thinned networks. To prevent covariate shift which occurs when the distribution between training and testing data are different while the conditional label distributions are the same, batch normalization is utilized. Finally, Leaky ReLU and ReLU activations are applied across layers to guarantee sparseness by removing unnecessary negative values which is beneficial for generalization. The proposed approach is compared over baseline U-net, patch-based shallow Unet, ResNet50, deep V-CNN and SVM methods. In patch-based shallow U-net we used patches of input as (48X48) to train the network. Different from U-net, ResNet-50 only considers using encoder as end to end learning of image with skip connections and blocks (He et al., 2016; Jifara et al., 2019). And it can overcome vanishing gradient problems of deep network and thus it can allow to train with deeper layer without severe over- fitting. The performance of all architectures used in this study is compared using accuracy, precision, recall and F1-score. Additionally, the performance of the proposed approach is evaluated using five-fold cross validation by randomly splitting train and test using five times without repeating the same instances in each fold. Furthermore, we used receiver operating characteristic (ROC) curve to evaluate the performance of the architectures in classifying the colony qualities. We used the same training and testing splits for all the methods compared in this study. All the methods used in the 5-fold cross validation experiment are used the same train and test set splits for all the folds. The best hyper parameters are selected heuristically using the proposed model. The networks are trained using Adam optimizer with starting alpha (learning rate of Adam), beta, and weight decay of 0.0001, 0.5, and 0.000001, respectively. To avoid local minima, the alpha value of Adam is decreased by multiplying it with 0.9 for every 20,000 iterations out of 300 epochs. The learning rate is dynamically and gradually reduced from 0.0001 to 0.00001 which make the loss smoothly decreased overtime. Thus, the final learning rate of 0.00001 is achieved and that lead smooth convergence. By experiment we set the learning rate of 0.001 to the baseline U-net. All the architectures were implemented in Python using the Chainer framework.

5. Results and discussions

As shown in Table 1, the proposed MIL-net learning architecture outperforms all other architectures experimented in this study. The accuracy of the MIL-net is higher than the baseline, patch-based shallow-net and ResNet-50 by 5.0%, 35.0%, and 15.0%, respectively. Patch based segmentation or classification becomes alternative to network to learn via data augmentation. It increases variation thus network can learn more. However, in this study patch-based shallow network can not perform well because patch based is failed to learn global texture.

Though, the experimental results of the ResNet-50 is high with reasonable accuracy, the spatial pooling nature of the encoder make the network difficult to visualize. While compared to the SVM and V-CNN, the MIL-net outperformed by 12.0% and 2.0%, respectively. and does not require pre-training and pre-processing techniques to extend the localization of cell regions through the classification. Furthermore, the proposed approach has lower number of parameters than the baseline

Table 1
Performance comparison of the MIL-net with other architectures.

Architectures	Accuracy	Precision	Recall	F-Score
MIL-net	**0.95**	**0.99**	0.89	**0.95**
Baseline	0.90	**1.0**	0.8	0.89
Shallow-net	0.60	0.57	0.80	0.70
ResNet-50	0.80	0.75	**0.90**	0.82
V-CNN	0.93	0.90	**0.90**	0.90
SVM	0.83	0.84	0.82	0.82

and yields high F-score with the value of 95.0% compared to 89.0%. of baseline.

As shown in Table 2, using an average of five-fold cross validation, the MIL-net is still the best in terms of recall with 96.0%. The baseline shows good performance in terms of F-score. However, it has higher number of parameters than the MIL-net. Thus the proposed structure is still considered to be beneficial.

Fig. 2 shows the ROC graph of the proposed network based on five-fold validation. It shows that the probabilities of different thresholds produce almost similar accuracy based on the network output. Furthermore, our data set used both separate as well as combined not more than two colonies in the dataset. The stem cell colony condition detection is not intended to separate the boundary of the combined colonies and hence, the performance of the MIL-net is not affected with the combined colonies. Compared to MIL-net as in Fig. 2, the baseline as in Fig. 3 is better in terms of mean ROC of five-folds with AUC of 97.0%. However, in terms of the ROC graph, MIL-net outperforms as revealed from figure.

Fig. 4 shows the ROC of the ResNet-50 in classifying the colony conditions using five-fold cross validation. It describes lower performance than the proposed and baseline with the mean AUC of 93.0%. The proposed MIL-net classify the colony quality followed by automatic localization of cell areas which is different from the traditional cell classification that performed classification after feeding the localized cell regions from several pre-processing steps. The approach used large series of mouse embryonic stem cells based on various architectures of deep CNN with annotations revealed 99.0% accu- racy in differentiating two different types of cells (Stumpf and MacArthur, 2019). The detection rate of neuron in neural stem cells using Xception network was 92.0% (Raytchev et al., 2016). The stem cell differentiation using simple and shallow CNN networks produced 75–90% accuracy (Kavitha et al., 2018). The performance of our proposed approach in detecting iPSC colony conditions is almost similar with those of the above studies. However the above mentioned stem cell detection studies used large series of training data. Whereas in this study we used limited number of dataset and different nature of stem cells than those of the above studies.

5.1. *Learning for localization*

Though using limited number of training data, the decoder of the trained MIL-net can able to generate accurate activation of the region of interest. Figs. 5 and 6 demonstrate the representative examples of the inferred colony regions of the proposed network. The weekly localization of the cell colony can be clearly visualized from the texture features gathered from the average pooling of the MIL-net. The effect of MIL justified in this study is the extension of localization through the classification, though it is not directed to do so.

One potential drawback of using MIL and DL approaches is that they can be complex and may not be as transparent as other methods. This lack of transparency can make it difficult for researchers and practitioners to understand how the AI system arrived at a particular conclusion or recommendation. In addition, these approaches may not be as interpretable as other methods, which can make it difficult to explain the results to others or to validate the results.

Therefore, it is important for researchers to carefully consider the trade-offs between the performance and explanatory power of different AI approaches, and to clearly communicate the limitations and potential drawbacks of their proposed methods in their research. This will help ensure that the results of AI research are transparent, understandable, and reliable, which is essential for the responsible and ethical use of AI in medical informatics and other fields.

Table 2
Performance comparison of the MIL-net using five-fold cross validation.

Architectures	Accuracy	Precision	Recall	F-score
MIL-net	**0.92**	0.84	0.96	0.88
Baseline	0.87	0.83	1.0	0.90
Shallow-net	0.50	0.58	1.0	0.73
ResNet-50	0.83	0.80	0.90	0.85
V-CNN	0.92	0.87	0.86	0.87
SVM	0.77	0.87	0.86	0.77

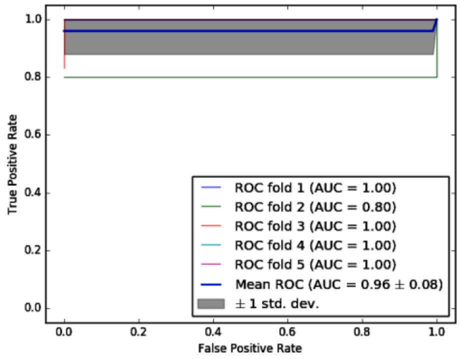


Fig. 2. Evaluation of the receiver operating characteristic curve of the proposed approach based on five-fold cross validation.

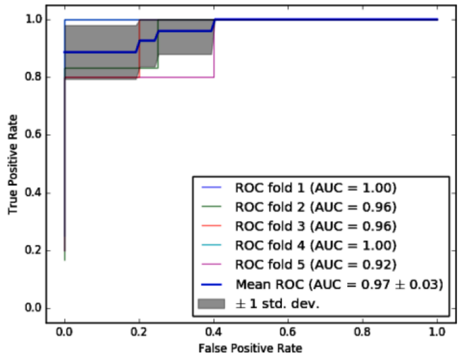


Fig. 3. Evaluation of the receiver operating characteristic curve of the baseline based on five-fold cross validation.

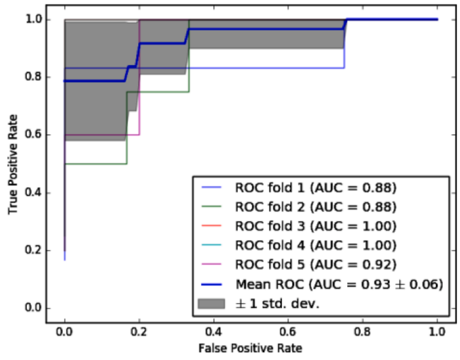


Fig. 4. Evaluation of the receiver operating characteristic curve of the ResNet-50 based on five-fold cross validation.

. To answer the explanatory power and reliability of the proposed study we compare the region of interest results of our method with the ResNet- 50. Figs. 7 demonstrates the representative examples of the region of interests of the inferred colony regions of the proposed network and ResNet- 50 using post-processing method. Compared to the ResNet50, MIL-net localization is more concentrated into the main cell (true positives) than ResNet50's first layer output localization. It proves that autoencoder like architecture like U-Net can neglect noise and occlusion occurred in images.

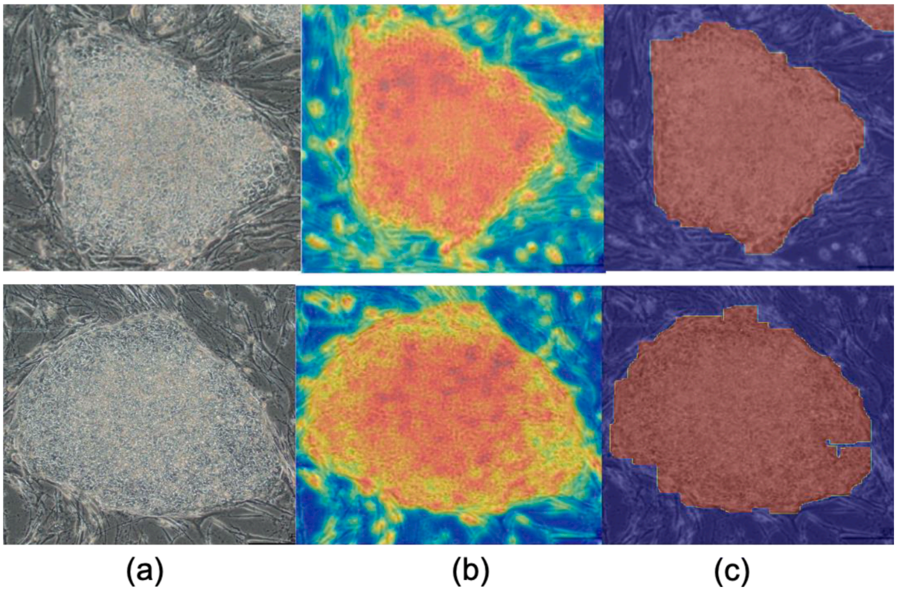


Fig. 5. Visualization of weak segmentation of good condition of colonies showing dense cells.(a) input image, (b) dense activation output, and (c) region of interest output.

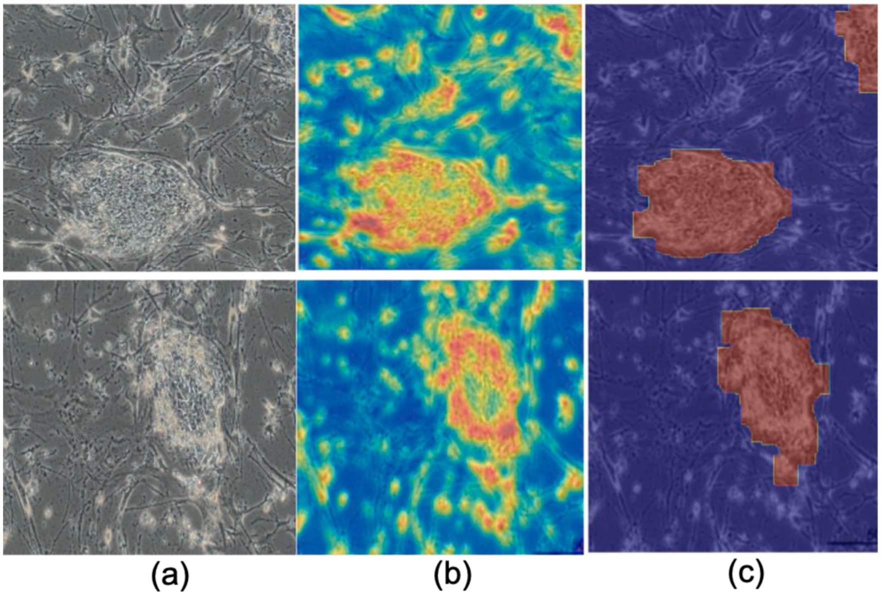


Fig. 6. Visualization of weak segmentation of bad condition of colonies showing sparse cells.(a) input image, (b) sparse activation output, and (c) region of interest output.

Fig. 8 shows the results of the white blood cell localization using the Raabin-WBC dataset. Each row of the figure represents basophils, eosinophils, lym-phocytes, monocytes, and neutrophils respectively. In each column it shows the original cell image, manual labeling, activation map on the final MIL-net layer, and localization results sequentially. The classification results of the five classes using five-fold cross-validation resulting the average accuracy of 70%. However, if we see the results of the localization from the MIL-net are more detailed than the manual labeling. With a note that the manual labeling is not done at the cytoplasm and nucleus level. Whereas in MIL-net visual localization looks close to the level of cytoplasm and nucleus, so that it is more detailed and better than manual labeling by experts.

6. Limitations

To extend the proposed work, argumentation approaches for explainable AI that involve the use of game theory and argumentation frameworks can be utilized to provide insights into the reasoning behind the decisions made by AI systems. These approaches seek to make the decision-making process more transparent and interpretable by human, and constructing, evaluating arguments and counter arguments based on the data and knowledge used by the AI system. In general, argumentation approaches can be seen as a way to provide a more human-like explanation of the decision-making process of AI systems. They can be used to identify and evaluate the relative strengths and weaknesses of different arguments and to determine which argument is the most reasonable or justified with given the available evidence.

Argumentation approaches can be applied to a wide range of AI

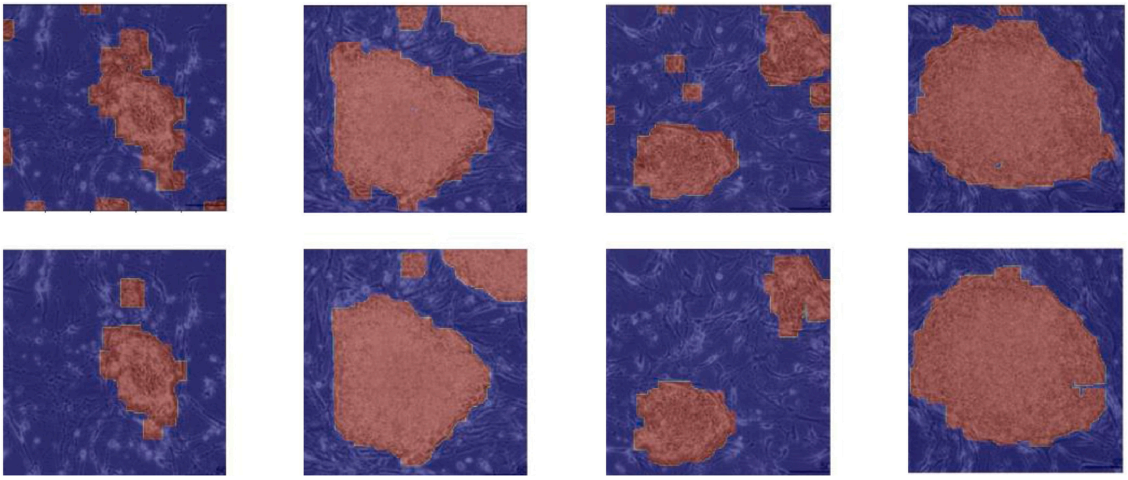


Fig. 7. Comparison between the localization results of the first layer output of ResNet- 50 (first row) and the last layer activation output of MIL-net (second row).

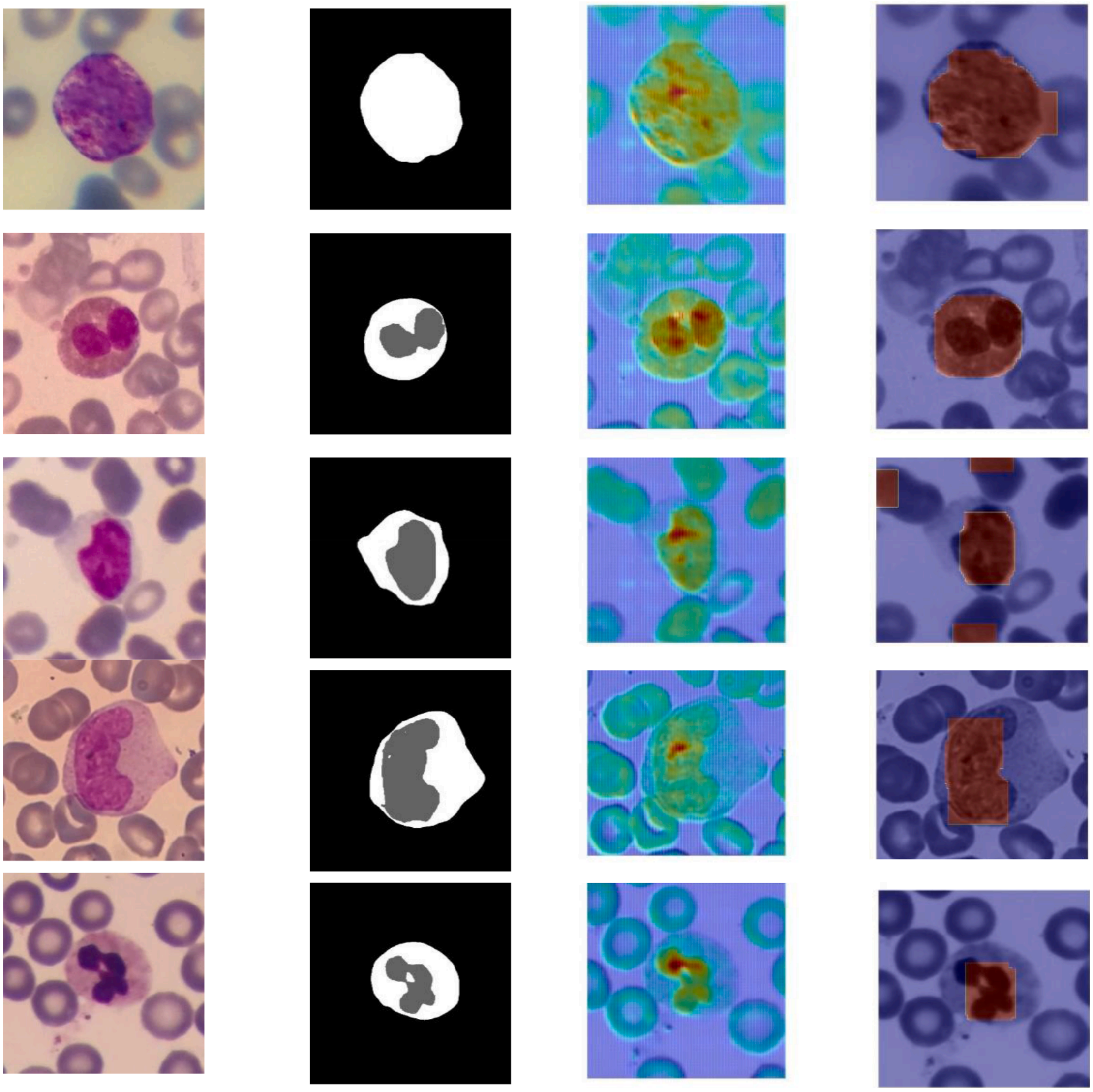


Fig. 8. Each row of the figure represents basophils, eosinophils, lymphocytes, monocytes, and neutrophils respectively. Each column shows the original cell image, manual labeling, activation map on the final MIL-net layer, and localization results, respectively.

systems and domains, including medical informatics (Caroprese et al., 2022), where they can be used to provide explanations for the diagnosis and treatment recommendations made by AI-powered medical decision support systems. However, there are several potential drawbacks to using argumentation approaches for explainable AI in medical informatics. Argumentation approaches can be quite complex, as they involve constructing and evaluating arguments and counterarguments. This complexity makes it difficult for non-experts to understand and interpret the results of these approaches. Argumentation approaches may not be suitable for all types of AI systems or for all types of medical data. They can be more effective for more straightforward or well-defined problems, but may be less effective for more complex or open-ended problems. Moreover, constructing and evaluating arguments and counter-arguments can be time-consuming and resource-intensive, which may make it difficult to apply these approaches in practice. While argumentation approaches may provide some level of transparency, they may not fully reveal the inner workings of the AI system or the reasoning behind its decisions. Like any other approach, argumentation approaches can be subject to biases, either in the construction of the arguments or in the evaluation of them. This can potentially lead to biased or unfair results. It is important to note that these drawbacks are not necessarily unique to argumentation approaches, and similar challenges may also be present in other explainable AI approaches. It is necessary to carefully consider the trade-offs and limitations of any approach when applying it in practice.

7. Conclusion

This study proposed a single network multiple instance learning in a weakly supervised settings based on U-net like architecture for annotating colonies and classifying the colony conditions. Most appealing in this study is the automatic visualization of the segmentation output of the cell regions without using the pixel-wise ground truth. Thus it reduced the annotation cost and maximized the classification accuracy in detecting the colony conditions. Experimentally we proved the robustness of our proposed approach by comparing the performance with state-of-the-art methods. Furthermore, through the experiments, we observed that the proposed approach has fewer number of parameters and high detection ability when compared over CNN-based and SVM methods. Hence it indicated its simplicity and the reliability. Thus the proposed approach for extending the localization through classification is highly useful to explain the reasons for decision making in identifying the colony conditions. Though our approach reveals high performance and produced attention-effective learning with weakly labels using iPSC dataset, the approach is needed to evaluate on different cell types data and different medical image dataset. It helps to understand the generalization ability of the proposed approach. In addition, the MIL-net with self-supervised setting is needed to test the optimal procedure of the architecture.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Data availability

Data will be made available on request.

References

Astorino, A., Fuduli, A., Veltri, P., & Vocaturo, E. (2020). Melanoma detection by means of multiple instance learning. *Interdisciplinary Sciences: Computational Life Sciences, 12*(1), 24–31.

Caroprese, L., Vocaturo, E., & Zumpano, E. (2022). Argumentation approaches for explanaible ai in medical informatics. *Intelligent Systems with Applications, 16*, Article 200109.

Chen, W.-. B., & Zhang, C. (2009). An automated bacterial colony counting and classification system. *Information Systems Frontiers, 11*(4), 349–368.

Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical image segmentation based on u-net: A review. *Journal of Imaging Science and Technology, 64*(2), 20508–1.

Fan, K., Zhang, S., & Zhang, Y.e.a. (2017). A machine learning assisted, label-free, non-invasive approach for somatic reprogramming in induced pluripo- tent stem cell colony formation detection and prediction. *Scientific Reports, 7*.

Foulds, J., & Frank, E. (2010). A review of multi-instance learning assumptions. *Knowledge Engineering Review, 25*(1), 1–25.

Fuduli, A., Veltri, P., Vocaturo, E., & Zumpano, E. (2019). Melanoma detection using color and texture features in computer vision systems. *Advances in Science, Technology and Engineering Systems Journal, 4*(5), 16–22.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Jifara, W., Jiang, F., Rho, S., et al. (2019). Medical image denoising using con- volutional neural network: A residual learning approach. *Journal of Supercomputing, 75*, 704–718.

Joutsijoki, H., Haponen, M., Rasku, J., Aalto-Setälä, K., & Juhola, M. (2016). Machine learning approach to automated quality identification of human induced pluripotent stem cell colony images. *Computational and Mathematical Methods in Medicine*, 1–15.

Kato, R., Matsumoto, M., Sasaki, H., Joto, R., Okada, M., Ikeda, Y., et al. (2016). Parametric analysis of colony morphology of non-labelled live human pluripotent stem cells for cell quality control. *Scientific Reports, 6*.

Kavitha, M. S., Kurita, T., Park, S.-. Y., Chien, S.-. I., Bae, J.-. S., & Ahn, B.-. C. (2017). Deep vector-based convolutional neural network approach for automatic recognition of colonies of induced pluripotent stem cells. *PloS One, 12*(12).

Kavitha, M. S., Kurita, T., & Ahn, B.-. C. (2018). Critical texture pattern fea- ture assessment for characterizing colonies of induced pluripotent stem cells through machine learning techniques. *Computers in Biology and Medicine, 94*, 55–64.

Kavitha, M.s., Yudistira, N., & Kurita, T. (2019). Multi instance learning via deep cnn for multi-class recognition of alzheimer's disease. In *2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA)* (pp. 89–94). IEEE.

Kavitha, M. S., Lee, C-H, Shibudas, K., et al. (2020). Deep learning enables automated localization of the metastatic lymph node for thyroid cancer on ^{131}I post-ablation whole-body planar scans. *Sci Rep, 10*, 7738.

Kavitha, M.S., Park, S-Y., Heo, M-S., Chien, S-I. (2016). Distributional Variations in the Quantitative Cortical and Trabecular Bone Radiographic Measurements of Mandible, between Male and Female Populations of Korea, and its Utilization, PLoS One 21;11 (12).

Kouzehkanan, Z.M., Saghari, S., Tavakoli, E., Rostami, P., Abaszadeh, M., Mirzadeh, F., et al., et al., Raabin-wbc: A large free access dataset of white blood cells from normal peripheral blood, bioRxiv (2021).

Kusumoto, D., Lachmann, M., Kunihiro, T., Yuasa, S., Kishino, Y., Kimura, M., et al. (2018). Automated deep learning-based system to identify endothelial cells derived from induced pluripotent stem cells. *Stem Cell Reports, 10*(6), 1687–1695.

Lian, S., Luo, Z., Zhong, Z., Lin, X., Su, S., & Li, S. (2018). Attention guided u-net for accurate iris segmentation. *Journal of Visual Communication and Image Representation, 56*, 296–304.

Liu, X., Zhang, A., Tiecke, T., Gros, A., Huang, T.S., Feedback neu- ral network for weakly supervised geo-semantic segmentation, ArXiv abs/1612.02766 (2016).

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., & Van Valen, D. (2019). Deep learning for cellular image analysis. *Nature Methods, 16*(12), 1233–1246.

Nagasaka, R., Matsumoto, M., Okada, M., Sasaki, H., Kanie, K., Kii, H., et al. (2017). Visualization of morphological categories of colonies for monitor- ing of effect on induced pluripotent stem cell culture status. *Regenerative Therapy, 6*, 41–51.

Oh, S. L., Ng, E., Tan, R., & Acharya, U. (2019). Automated beat-wise arrhythmia diagnosis using modified u-net on extended electrocardiographic record- ings with heterogeneous arrhythmia types. *Computers in Biology and Medicine, 105*, 92–101.

Okita, K., Ichisaka, T., & Yamanaka, S. (2007). Generation of germline competent induced pluripotent stem cells. *Nature, 448*, 313–317.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Mis- awa, K., et al., et al., Attention u-net: Learning where to look for the pancreas, arXiv preprint arXiv:1804.03999 (2018).

Qiu, S., Global weighted average pooling bridges pixel-level localization and image-level classification, ArXiv abs/1809.08264 (2018).

Raytchev, B., Masuda, A., Minakawa, M., Tanaka, K., Kurita, T., Ima- mura, T., et al. (2016). Detection of differentiated vs. undifferentiated colonies of ips cells using random forests modeled with the multivariate polya distribution. In *Conference proceedings: MICCAI* (pp. 667–675).

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241). Springer.

Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., et al. (2019). Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis, 53*, 197–207.

Stumpf, P. S., & MacArthur, B. D. (2019). Machine learning of stem cell identities from single-cell expression data via regulatory network archetypes. *Frontiers in Genetics, 10*(2).

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., et al. (2007). Induction of pluripotent stem cells from adult human fi- broblasts by defined factors. *Cell, 131*(5), 861–872.

Vocaturo, E., & Zumpano, E. (2021). Diabetic retinopathy images classification via multiple instance learning. In *2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)* (pp. 143–148). IEEE.

Waisman, A., La Greca, A., Mobbs, A. M., Scarafia, M. A., Santín Velazque, N. L., Neiman, G., et al. (2019). Deep learning neural networks highly pre- dict very early onset of pluripotent stem cell differentiation. *Stem Cell Reports, 12*(4), 845–859.

Yang, Y., Zhang, W., Wu, J., Zhao, W., Chen, A., Deconvolution-and- convolution networks, arXiv preprint arXiv:2103.11887 (2021).

Yuan-Hsiang, C., Abe, K., Yokota, H., Sudo, K., Nakamura, Y., & e. a. Cheng-Yu, L. (2017). Human induced pluripotent stem cell region recognition in microscopy images using convolutional neural networks. In *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference* (pp. 4058–4061).

Yudistira, N., Kavitha, M.s., Itabashi, T., Iwane, A. H., & Kurita, T. (2020). Pre- diction of sequential organelles localization under imbalance using a bal- anced deep u-net. *Scientific Reports, 10*(1), 1–11.

Zhang, H., Shao, X., Peng, Y., Teng, Y., Saravanan, K., Zhang, H., et al. (2019). A novel machine learning based approach for ips progenitor cell identi- fication. *Plos Computational Biology, 15*(12).

Zhi-Hua, Z., Min-Ling, Z., Sheng-Jun, Y.-. F., & Huang, L. (2012). Multi-instance multi-label learning. *Artificial Intelligence, 176*(1), 2291–2320.

Zhu, Y., Huang, R., Wu, Z., Song, S., Cheng, L., & Zhu, R. (2021). Deep learning- based predictive identification of neural stem cell differentiation. *Nature Communications, 12*(1).

Zumpano, E., Fuduli, A., Vocaturo, E., & Avolio, M. (2021). Viral pneumonia im- ages classification by multiple instance learning: Preliminary results. In *25th International Database Engineering and Applications Symposium* (pp. 292–296).