Deep Learning-Based In Vitro Detection Method for Cellular Impurities in Human Cell-Processed Therapeutic Products

Yasunari Matsuzaka 1,2, Shinji Kusakawa 3, Yoshihiro Uesawa 1, Yoji Sato 3 and Mitsutoshi Satoh 4,*

1 Department of Medical Molecular Informatics, Meiji Pharmaceutical University, Kiyose 204-8588, Tokyo, Japan; yasunari80808@ims.u-tokyo.ac.jp (Y.M.); uesawa@my-pharm.ac.jp (Y.U.)
2 Division of Molecular and Medical Genetics, Center for Gene and Cell Therapy, The Institute of Medical Science, University of Tokyo, Minato-ku 108-8639, Tokyo, Japan
3 Division of Cell-Based Therapeutic Producers, National Institute of Health Sciences, Kawasaki 201-9501, Kanagawa, Japan; kusakawa@nihs.go.jp (S.K.); yoji@nihs.go.jp (Y.S.)
4 Department of Toxicology and Pharmacology, Faculty of Pharmaceutical Sciences, Meiji Pharmaceutical University, Kiyose 204-8588, Tokyo, Japan
* Correspondence: satoh@my-pharm.ac.jp; Tel.: +81-42-495-8949

Abstract: Automated detection of impurities is in demand for evaluating the quality and safety of human cell-processed therapeutic products in regenerative medicine. Deep learning (DL) is a powerful method for classifying and recognizing images in cell biology, diagnostic medicine, and other fields because it automatically extracts the features from complex cell morphologies. In the present study, we construct prediction models that recognize cancer-cell contamination in continuous long-term (four-day) cell cultures. After dividing the whole dataset into Early- and Late-stage cell images, we found that Late-stage images improved the DL performance. The performance was further improved by optimizing the DL hyperparameters (batch size and learning rate). These findings are first report for the implement of DL-based systems in disease cell-type classification of human cell-processed therapeutic products (hCTPs), that are expected to enable the rapid, automatic classification of induced pluripotent stem cells and other cell treatments for life-threatening or chronic diseases.

Keywords: human cell-processed therapeutic products; deep learning; convolutional neural networks; hyperparameters; image classification

1. Introduction

In regenerative medicine, human cell-processed therapeutic products (hCTPs) derived from human pluripotent stem cells (hPSCs), such as induced pluripotent stem cells (iPSCs) and embryonic stem cells, are promising treatment modalities for life-threatening or incurable diseases owing to their pleiotropic effects. Specifically, living cells have infinite self-renewal capacity and can differentiate into various cell types, providing new resources with high efficiency [1–4]. However, the difficulty of managing the quality of the new cells is problematic because quality is highly correlated with the efficacy and safety of the hCTPs, which is demanded for the stable supply of clinical needs. The generation protocol of hPSCs must be down-scaled for monolayer cell culture conditions, and tumor formation has been reported in patients receiving transplantation of human somatic cells [1,2,5–14]. By integrating control systems, we could realize the automatic management of raw materials, intermediates, final products, and process parameters for good manufacturing practice on a simple and cost-effective platform [15–20]. In particular, hCTPs can become contaminated with tumor cells and other abnormal cells during the manufacturing process, which hinders their application in regenerative medicine. During long-term suboptimal culturing, contaminants are generated by two main mechanisms: (1) spontaneous transformation of the original component cells induced by accumulation of abnormal karyotypes and (2) cross-contamination with other cell...
types such as fibrosarcoma, osteosarcoma, and glioma cell lines. Therefore, establishing a highly accurate method that detects abnormal cells at each step of the hCTP production process is essential [10,11,21–31]. For quality and safety reasons, hCTPs are comprehensively evaluated in multiple tumorigenicity-related tests to remove the risk of tumorigenic cells in transplants [1,9–11,32–34]. However, the throughput, accuracy, and cost of these time-consuming tests must be improved before administering hCTPs to humans. This problem might be resolved by deep learning (DL) approaches that automatically and rapidly classify different cell types by detecting their small differences among a large number of phase-contrast cell images. DL algorithms can morphologically classify cells into fibroblast-like (elongated), epithelial-like (polygonal), and lymphoblast-like (spherical). The main component in DL algorithms is a convolutional neural network (CNN), which has proven successful in the classification and semantic segmentation of live cell images [31,35–43]. The other classification approach is conventional machine learning (ML), which differentiates cellular morphologies, predicts cell differentiation potential, and provides non-invasive evaluations of human iPSCs. When surveying cellular conditions, ML requires trial-by-error feature extraction, feature selection, and classification steps. DL generally outperforms ML because it leverages the powerful capability of the CNN to automatically extract and learn the image data representing cells with complicated and uniformed morphologies [44–49]. CNNs have become useful tools for quality control of cultured cells because they successfully discriminate between abnormal and normal cells. After training, a CNN automatically classifies and recognizes cells in different types of images [31,50–59]. For example, Toratani et al. [53] recently reported applying a CNN to microscopic images of patient-derived cancer and radiation-resistant cells to determine the effectiveness of radiation therapy. CNNs make it possible to discriminate small differences in cell images that cannot be discriminated by the human eye. Novel technologies based on CNNs are expected that can use images to predict not only the cell type but also its responsiveness to treatment. Isozaki et al. [60] reported a DL-based technology that uses image analysis to identify individual cells at high speed. This technology can be used to image large cell populations at high speed and to discriminate specific cells in real time based on information processing technology. In an experiment to demonstrate the versatility of their technology, they imaged cells of different sizes (3–30 µm) at high speed and identified potential cancer cells circulating in the blood of cancer patients. However, a manufacturing process has not yet been developed that can ensure high-quality and safe hCTP production.

In this study, we constructed model systems for distinguishing abnormal cervical cancer cells (HeLa-GFP) when mixed together with human embryonic lung fibroblasts (MRC-5). Images of the cells at different stages of a continuous culture period were input to a DL algorithm. The prediction performance was improved in the late stage of the culture period. The prediction performances of the constructed models were further improved by optimizing the DL hyperparameters, namely, the batch size (BS) and learning rate (LR). These results suggest that DL-based prediction models can effectively classify different cell types in cell images; in particular, DL can automatically extract the meaningful features from complex morphological cell information.

2. Materials and Methods
2.1. Cell Culture System

Human cervical cancer (HeLa-GFP) cells (1 × 10^2 cells/200 µL/well, Cell Biolabs Inc, San Diego, CA, USA) and human embryonic lung fibroblasts (MRC-5 cells; 1 × 10^4 cells/200 µL/well, ATCC, Manassas, VA, USA) were co-cultured in Dulbecco’s modified Eagle medium supplemented with 10% fetal bovine serum, 1% non-essential amino acid, and 0.03% SphereMax™ containing LA717 natural polymer (Nissan Chemical Co., Tokyo, Japan). HeLa-GFP cells (1 × 10^2 cells/200 µL/well) were cultured alone in the same medium. All cultures were performed in 96-well ultra-low attachment microplates (Corning, Corning, NY, USA) for 4 days at 37 °C in a 5% CO₂ atmosphere.
2.2. Preparation of the Image Data of Culture Cells

The images of the MRC-5 cells alone and co-cultured MRC-5 cells and HeLa-GFP cells were taken on a Confocal Quantitative Image Cytometer, CQ1 (Yokogawa Electric Corporation, Tokyo, Japan) every 20 min for 4 days. Thirty positive areas (pArea) in which normal cells and cancer cells coexist in co-cultured wells and 60 negative areas (nArea) in which normal cells alone exist in monoculture wells were excised from whole-well bright-field images over time acquired with CQ1, as inputs for a DL algorithm (Figure 1b). The dataset was split into training (Tra) and validation (Val) datasets at a ratio of 1:1. The Total dataset of the cell culture period (Days 0–4) comprised 4320 images in each of the Tra and Val datasets of the positive area and 8640 images in each of the Tra and Val datasets of the negative area (Table S1). The Total dataset was then divided into two groups. The Early_stage dataset (Days 0–2) comprised 2016 images in the Tra dataset and 2304 images in the Val dataset of the positive area and 4320 images in each of the Tra and Val datasets of the negative area. The Late_stage dataset (Days 2–4) comprised 2304 images in the Tra dataset and 1616 images in the Val dataset of the positive area and 4320 images in each of the Tra and Val datasets of the negative area (Table 1).

![Figure 1](image)

**Figure 1.** The culture model for assessing the presence of abnormal cells in a product. In this suspension culture system, MRC-5, normal cells as product model, and HeLa-GFP, abnormal cells as cellular impurity model, were co-cultured (a). Positive areas (pArea) in which normal cells and cancer cells coexist in co-cultured wells and negative areas (nArea) in which normal cells alone exist in monoculture wells were excised from whole-well bright-field images over time in each well, as inputs for a DL algorithm (b).

| Table 1. Image datasets used in Deep Learning. |
|-----------------------------------------------|
|                  | Total | Pre_Stage | Late_Stage |
|                  | Positive Area | Negative Area | Positive Area | Negative Area | Positive Area | Negative Area |
| Tra              | 4320               | 8640              | 2016           | 4320           | 2304           | 4320           |
| Val              | 4320               | 8640              | 2304           | 4320           | 2016           | 4320           |

2.3. Deep Learning

All image files produced using fluorescence microscopy were resized using NVIDIA DL GPU Training System (DIGITS) version 4.0.0 software (NVIDIA, Santa Clara, CA, USA). The final image files, which had a fixed resolution of 256 × 256 pixels, were used as input for DL. The DL process included data management, design, and training of the neural networks on four-GPU systems (Tesla-V100-PCIE; 31.7 GB), real-time monitoring neural
networks on four-GPU systems (Tesla-V100-PCIE; 31.7 GB), real-time monitoring of the model performance, and selection of the best performance model from the results.

The Total, Early\_stage, and Late\_stage prediction models were constructed using the corresponding Tra datasets and were evaluated using the corresponding Val datasets. The evaluation indices were the loss and accuracy:

\[
\text{Loss} = -1/N \sum_{i=1}^{N} [y_i \log p_i + (1 - y_i) \log(1 - p_i)] = -1/N \sum_{i=1}^{N} \log p_i' \quad (1)
\]

\[
\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \quad (2)
\]

For the loss, \(y_i\) has a value of 1 or 0 for positive and negative samples, respectively. \(p_i\) represents the prediction probability of each record being a positive example. The loss function computes the prediction probability of the true value, so a lower loss value indicated a better performance [61]. \(p_i'\) denotes the prediction probability of a true value. For a positive true value, \(p_i' = p_i\); for a negative example, \(p_i' = 1 - p_i\). A positive example is given a large penalty when the probability of being a positive example is low; similarly, a negative example is given a large penalty when the probability of being a positive example is high. Differentiating the score \(L_i = -[y_i \log p_i + (1 - y_i) \log(1 - p_i)]\) of record \(i\) with respect to the predicted probability \(p_i\) gets \(\frac{\partial L_i}{\partial p_i} = \frac{p_i - y_i}{p_i(1 - p_i)}\). Clearly, \(L_i\) is minimized when \(p_i = y_i\).

For the accuracy, TP, FN, TN, and FP denote the numbers of true positives, false negatives, true negatives, and false positives, respectively. To test the independence of the dataset, McNemar’s test was performed using a free algorithm on a website: https://www2.ccrb.cuhk.edu.hk/stat/confidence\%20interval/McNemar\%20Test.htm (accessed on 8 September 2021) [62].

3. Results

In the present study, we employed the recently reported novel 3D cell culture system with a low molecular weight agar, LA717 that would promote uniform suspension of cells on low-attachment culture plates [63]. In this culture environment, normal cells are unable to proliferate and die or aggregate with neighboring cells over time, on the other hand, abnormal cells, such as cancer cells, form colonies through anchorage-independent growth (Figure 1a). We used MRC-5, normal cells as a hCTP model, and HeLa-GFP, abnormal cells as a cellular impurity model in the culture system. Positive areas (pArea) in which normal MRC-5 cells and cancer HeLa-GFP cells coexist in co-cultured wells and negative areas (nArea) in which MRC-5 cells alone exist in monoculture wells were excised from whole-well bright-field images over time in each well, as inputs for a DL algorithm (Figure 1b).

To construct the prediction models of contamination by other cell types in hCTP-cell culture systems, we trained the CNN on images of single and mixed cell types during continuous cell culture. We first analyzed the effects of the hyperparameters on the prediction performance of the constructed models. In DL with the stochastic gradient descent method, the dataset is divided into several subsets during training to reduce the effects of outliers. The BS specifies the number of samples in each training iteration [62,64–66]. The gradients were averaged over each population of the Tra dataset, which was training dataset and divided into \(N\) samples. After the weight was updated according to the gradient, the next group was processed. When all groups were finally processed (i.e., all epochs), the order of the Tra dataset was randomly changed for learning. However, when a large BS is set for training, the DL shows a well-known generalization gap that remarkably degrades its generalization performance [67,68]. To optimize the hyperparameters of the DL for detecting HeLa-GFP contaminants co-cultured with normal human MRC-5 cells, all images containing pArea and nArea during the cell culture period were divided into three datasets: Total (Days 1–4), Early\_stage (Days 1–2), and Late\_stage (Days 3–4). The Total, Early\_stage, and Late\_stage models had BSs of 26 (5–350), 11 (100–200), and 24 (5–200), respectively (see Figure 2 and Tables S1–S3). The mean loss(Val) and Acc(Val) denote the loss and accuracy values, respectively, with the Val datasets; they were 0.718 ± 0.068 and 53.7% ±
11.8%, respectively, for Total; 1.268 ± 0.220 and 35.3% ± 3.1%, respectively, for Early_stage; and 0.660 ± 0.067 and 66.3% ± 3.9%, respectively, for Late_stage (Figure 2, Tables S1–S3). The best loss(Val) and Acc(Val) performances were 0.649 at BS = 30 and 64.3% at BS = 7, respectively, for Total; 0.986 at BS = 180 and 40.0% at BS = 5, respectively, for Early_stage; and 0.623 at BS = 189 and 70.2% at BS = 189 or 190, respectively, for Late_stage (Figure 2, Table 3 and Tables S1–S3). Figure 3 plots the prediction performances as a function of the other hyperparameter (LR). For fine-tuning, we set the LR to 16 for Total (−log10[L]R = 1.40–7.00), 7 for Early_stage (−log10[L]R = 2.00–9.00), and 22 for Late_stage (−log10[L]R = 2.00–5.00). The mean loss(Val) and Acc(Val) were 0.690 ± 0.049 and 58.2% ± 11.2%, respectively, for Total; 1.099 ± 0.879 and 51.7% ± 15.8%, respectively, for Early_stage; and 0.634 ± 0.020 and 67.2% ± 4.4%, respectively, for Late_stage (Figure 3, Tables S1–S3). The best loss(Val) and Acc(Val) performances were 0.648 at −log10[L]R = 1.699 and 64.4% at −log10[L]R = 2.0, respectively, for Total; 0.682 at −log10[L]R = 9.000 and 65.9% at −log10[L]R = 7.0, respectively, for Early_stage; and 0.612 at −log10[L]R = 2.585 and 70.6% at −log10[L]R = 5.0, respectively, for Late_stage (Figure 3, Table 5 and Tables S1–S3).

Figure 2. Performance measures of the DL prediction models as functions of the BS. Plotted are the loss values (left) and accuracies (right) with the Val datasets over the total period (upper), Days 1 and 2 (middle), and Days 3 and 4 (bottom) of continuous cell culture.

Table 2. Mean loss(Val) and mean Acc(Val) in total, early_stage, and late_stage datasets with BS.

|                      | Total ± SD | Early_Stage ± SD | Late_Stage ± SD |
|----------------------|------------|------------------|-----------------|
| mean loss (Val)      | 0.718      | 1.268            | 0.660           |
| mean Acc (Val)       | 53.7       | 35.3             | 66.3            |

Table 3. best loss(Val) and best Acc(Val) in total, early_stage, and late_stage datasets with BS.

|                      | Total | Early_Stage | Late_Stage |
|----------------------|-------|-------------|------------|
| best loss (Val)/BS   | 0.649 | 0.986       | 0.623      |
| best Acc (Val)/BS    | 64.3  | 40.0        | 70.2       |

Table 4. mean loss(Val) and mean Acc(Val) in total, early_stage, and late_stage datasets with LR.

|                      | Total ± SD | Early_Stage ± SD | Late_Stage ± SD |
|----------------------|------------|------------------|-----------------|
| mean loss (Val)      | 0.718      | 1.268            | 0.660           |
| mean Acc (Val)       | 53.7       | 35.3             | 66.3            |

Table 5. best loss(Val) and best Acc(Val) in total, early_stage, and late_stage with LR.

|                      | Total | Early_Stage | Late_Stage |
|----------------------|-------|-------------|------------|
| best loss (Val)/LR   | 0.648 | 0.986       | 0.623      |
| best Acc (Val)/LR    | 64.4  | 40.0        | 70.2       |
| Total (BS) | Early_Stage (BS) | Late_Stage (BS) |
|-----------|-----------------|-----------------|
| best loss (Val)/BS | 0.649 | 0.986 | 0.623 |
| best Acc (Val)/BS | 64.3 | 40.0 | 70.2 |

**Table 4.** Mean loss(Val) and mean Acc(Val) in total, early_stage, and late_stage datasets with LR.

| Total ± SD | Early_Stage ± SD | Late_Stage ± SD |
|------------|-----------------|-----------------|
| mean loss (Val) | 0.690 ± 0.049 | 1.099 ± 0.879 | 0.634 ± 0.020 |
| mean Acc (Val) | 58.2 ± 11.2 | 51.7 ± 15.8 | 67.2 ± 4.4 |

**Table 5.** Best loss(Val) and best Acc(Val) in total, early_stage, and late_stage with LR.

| Total (LR) | Early_Stage (LR) | Late_Stage (LR) |
|------------|-----------------|-----------------|
| best loss (Val)/LR | 0.648 | 0.682 | 0.612 |
| best Acc (Val)/LR | 64.4 | 65.9 | 70.6 |
The prediction performances of five DL solver types—Adaptive Delta (AdaDelta), Adaptive gradient (AdaGrad), Adaptive Moment Estimation (Adam), Nesterv’s accelerated gradient, and root mean square propagation (RMSprop)—were then evaluated at the optimal hyperparameter values. With the Early-stage dataset, RMSprop achieved a slightly lower loss(Val) value than the other solvers, but all five solvers obtained very similar Acc(Val) values (Figure 4a,b, Tables S2 and S3). The prediction performances were also evaluated for five policies: step down, exponential decay, inverse decay, polynomial decay, and sigmoidal decay in the DL hyperparameters. With the Late-stage dataset, the loss(Val) value was slightly lower under the sigmoidal decay policy than under the other policies, but all five policies yielded very similar Acc(Val) values (Figure 4c,d), Tables S2 and S3). To test the independence of the Val dataset, McNemar’s chi-square test was applied to compare the accuracies among the different algorithm solvers and policies. The calculation results did not show significant differences in accuracy among the solvers or policies ($p = 0.074$). These results suggest that the accuracy and policies used for the prediction model may not affect the accuracy. On the contrary, an association study of Acc(Val) and loss(val) with BSs of 100–200 between Early-stage and Late-stage data indicated significant differences ($P_c < 3.77 \times 10^{-20}$ for Acc(Val) and $P_c < 1.02 \times 10^{-5}$ for loss(Val) according to Student’s t-test with the Bonferroni correction). According to these results, the prediction performance was improved by evaluating images from the Late-stage dataset. This suggests that easily identifiable contaminant cells appear later in the culture period and that the prediction performance can be improved by acquiring a dataset containing cells of clearly different morphologies.

Figure 4. Performances of the DL prediction models with different solver types (a,b) and policies (c,d). Plotted are the loss values (left) and accuracies (right) with the validation dataset.
4. Discussion

Generalizing the learning results for time series prediction with test data is an important issue for DL. However, conventional methods have problems such as a poor generalization performance and suboptimal solutions. The generalization ability, which is the ability to predict future results of input data that have never been seen before, is determined by the generalization error, which is the difference between the prediction errors with the training and test datasets. This demonstrates the importance of developing a prediction method with a small error. Although various predictive methods have been reported for DL, current mainstream approaches have some limitations. DL can be used to extract features automatically, but overfitting is likely to occur with fluctuations in training data, which increases the generalization error. Therefore, the dimensionality needs to be reduced to obtain information with good representativeness. The most widely used approach to dimensionality reduction and feature extraction is principal component analysis, which extracts a large amount of noise that is not related to prediction. Therefore, it is not always compatible with maximizing the generalization ability. A state-space model that infers the hidden state variables and parameters of the nonlinear generation process for the input signal is currently widely used for prediction. However, this model has drawbacks such as not being able to infer the generation process correctly because it may fall into a non-optimal solution (i.e., local minimum) when the underlying dynamics are unknown. Thus, traditional approaches are known to have large generalization errors or local minimum values, both of which increase the prediction error with new test data. Lin et al. [69] recently reported that, with sufficient training samples and high-dimensional inputs, the state variables, parameters, and dimensions of standard nonlinear generation processes can be identified with high accuracy. Their results suggest that their approach has high robustness to observed noise because generalization and feature extraction can be maintained with only limited training data. Because this system is easy to configure with a neural network and has a low computational cost, it is expected to be applicable to highly reliable and explainable AI with a mathematically guaranteed optimal generalization strategy when implemented in parallel with our constructed prediction model.

The application of the DL algorithms in our study to various fields of regenerative medicine requires addressing various problems such as classification accuracy reproduction, implementation difficulty, and learning times. Classification accuracy reproduction refers to the difficulty of knowing whether a performance level can be achieved until actual implementation. Implementation difficulty refers to the higher level of programming skills and mathematical knowledge required for DL than for other types of ML. DL is based on mathematical principles with varied backgrounds, and it takes a long time to understand. Finally, DL models take a long time to train. To reduce the learning time, GPUs have been applied to speeding up calculations [70–73].

In addition to implementation problems, some fundamental DL challenges remain. One goal for DL is to minimize the generalization error. However, this is not an easy task, because the tissues handled by DL are considered ill-posed problems (i.e., some of the required information to obtain the solution is lacking). By selecting the appropriate heuristics, information can be added to transform an ill-posed problem into a well-posed problem. Domain-specific heuristics are used for efficient solving, but if the algorithm is not clear, the correct answer cannot be obtained. For ill-posed problems where not all data can be collected, the criteria are unclear, and a formal and rigorous verification cannot be performed. Therefore, remembering that the results contain some uncertainties is important for DL implementation. Another issue with DL is the “no free lunch” theorem, which postulates that a good DL model is theoretically impossible for any problem, and if one model outperforms another, it is specialized for a particular problem. In particular, if the problem changes, the algorithm should change, and it is important to devise a solution that is specific to each problem based on preconditions and prerequisite knowledge as much as possible. The performance of ML methods including DL can be improved for
specific problems by applying an inductive bias, which is prior knowledge or a hypothesis besides the training data.

Then, in what kind of problems does DL specialize? Simple and realistically sized neural networks can sufficiently approximate the following problems:

1. Low-degree polynomial models: any polynomial can be approximated by a neural network consisting of about four times as many neurons as the number of multiplications required for calculation.
2. Locality: a local Markov network can be approximated using a neural network consisting of a number of neurons proportional to the number of nodes.
3. Symmetry: CNNs that explicitly incorporate mobile and temporal universality can significantly reduce the number of parameters required for learning, which also greatly reduces the apparent complexity [74].

According to the “ugly duckling” theorem, it is theoretically impossible to classify or judge similarity without some prior knowledge or induction bias. This is conceptually similar to the no free lunch theorem, which argues that there is no universal one-size-fits-all ML model or search or optimization algorithm that can solve any problem efficiently. Meanwhile, the ugly duckling theorem argues that, without assumptions or prior knowledge, there is no optimal feature representation and feature set that will result in better classification performance. This raises the issue that objective or general-purpose classification is difficult. Therefore, this theorem also indicates that algorithms need to be formulated according to the classification problem to be solved.

Also, in general, DL has some problems, such as “overfitting or overtraining”, that if the training period is too long or the training data is not typical, it will adapt to certain random features of the training data, that are unrelated and unwanted features. The state in which unknown test data cannot be adapted and be generalized will lead to a lock of generalization ability [75–82]. One of the causes is that the model is complicated and has too many degrees of freedom compared to the number of training data, such as too many parameters to fitting to the statistical model. The main strategies for suppressing the overfitting are the following:

1. Increase the number of training data

To make sure that training is a better generalization, the variation of training data is increased. For the purpose, there is also a method of extending the existing training data by generating a variety of image data with changing the saturation, brightness, and direction of one image. Thus, these approaches make it possible to acquire abundant data and build a flat learning model that does not cause the overfitting by making good use of limited data. However, in the process of the increase of training data, if only biased data is given, it will adversely affect the construction of the model. If a bias, which is a representation of the difference between the predicted results and measured value, is small, the prediction, which represents variability in prediction results, accuracy is high, but if the variance is too large, the versatility is low, and overfitting may occur. In order to build an appropriate prediction model that does not cause the overfitting, it is necessary to keep the balance between bias and variance constant.

2. Model simplification

Introduction complex and sophisticated algorithms allows for advanced analysis with the addition of various parameters, but at the same time increases the risk of the overfitting. Therefore, the simplest way to prevent the overfitting is to reduce the size of the model, i.e., the number of learnable parameters in the model, which is determined by the number of layers and units per layer. By probabilistically selecting the units of each layer of the multi-layer network and invalidating the units other than the selected ones, it is possible to create a “temporary network with a small degree of freedom”. By training with this process, the effect of simplifying the model can be obtained.
(3) Regularization

There is a method of adding a regularization term to the error function to suppress the complexity and degree of freedom of the model and prevent overfitting. The most common regularizations are L1 regularization and L2 regularization. With L1 regularization, some parameters can be set to zero, which is similar in feature selection, resulting in a sparse model. Furthermore, if there are many zero, it can be expressed as a sparse matrix and can be calculated at high speed. On the other hand, L2 regularization is a method of approaching zero according to the size of the data and is characterized by being able to build a smooth model. Thus, while DL has grown into a program that can produce amazing output through long-term operation, proper tuning and verification by human must be continued to avoid the overfitting, because the overfitting can be prevented depending on the approach, and even after it occurs, DL can be updated to a normal model by troubleshooting in the correct way. Further, another of the challenges facing machine learning (ML) is the selection of appropriate ML models for classifying domain-specific data. Some of the main ML techniques are:

(1) Support vector machine (SVM)

The SVM is a model of unsupervised learning that can handle both analysis and regression, and mainly solves classification tasks, and can make predictions with high accuracy even for unknown data by maximizing the margin that draws a boundary so that the distance of data near the boundary is maximized.

(2) K-nearest neighbor (kNN) algorithm

The kNN method plots the given training data on the vector space, and if unknown data is obtained, acquire any k pieces in order of distance from them, and estimate the class to which the data belongs by the majority vote.

(3) Naïve Bayes classifier

This method is classifier to determine which category a certain data belongs to and makes predictions according to Bayesian discrimination rules based on a simple stochastic model.

In addition, it was reported that comparison of the DL and traditional ML models for classification of normal and abnormal cells by using seven different ML algorithms such as logistic regression, kNN, SVM, naïve Bayes, random forest, XGBoost, and deep neural network. Of these algorithms, deep neural network indicated outperformance compared with other algorithms [83].

One of the major concerns with the manufacture of transplanted cells used in regenerative medicine is the risk of cross-contamination, which causes accidental mixing of cell samples. Control of the manufacturing process requires eliminating the existence of harmful impurities such as cancer cells and ensuring the quality of the transplanted cells. A soft agar colony formation (SACF) test is commonly used to evaluate the anchorage-independent growth, which is one of characteristics of malignant transformed cells such as cancer cells. This test can detect the presence of abnormal cells in normal cells in a short time. Our group [9] developed a digital assay screening system that detects cancer cells among normal cells with better sensitivity and efficiency than the conventional SACF test. However, this approach requires a skillful technique to make multiple layers of viscous agar medium with different concentrations in many wells. Therefore, in the present study, we applied the novel 3D cell culture system with 0.03% LA717 containing medium and low-attachment culture plates, as a culture method to evaluate the anchorage-independent growth of cells more easily and effectively than the conventional soft agar culture. The co-culture of MRC-5 and HeLa-GFP in this method showed efficient growth of HeLa-GFP only, suggesting that this culture system is useful as a model for evaluating contamination of normal cells by cellular impurities such as cancer cells equivalent to HeLa cells. Through this culture system, a large number of image data over time were acquired by
time-lapse imaging, which is necessary for developing a DL-based prediction model of cellular impurity contamination.

5. Conclusions

We constructed DL-based prediction models that detect the contamination of human embryonic lung fibroblasts (MRC-5 cells) with human cervical cancer (HeLa-GFP) cells. The cultures were incubated for 4 days. The image data collected from Days 1 to 4 were divided into two groups: Early stage (Days 1 and 2) and Late stage (Days 3 and 4). The prediction performance of the DL algorithm was higher after training on the Late stage dataset than on the Early stage dataset. The performance was further improved by optimizing the BS and LR hyperparameters in the DL. This method is first reports that the implement of the DL for the classification of the contamination of cancer cells in hCTPs, that expected to provide high-throughput and low-cost alternatives to traditional methods; especially, the culture cell conditions can be adjusted during the generation process of hCTPs.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/app11209755/s1, Table S1: loss(Val) and Acc(Val) in total dataset with BS and LR, Table S2: loss(Val) and Acc(Val) in early stage dataset with BS and LR, Table S3: loss(Val) and Acc(Val) in late stage dataset with BS and LR.

Author Contributions: Conceptualization, initiation and supervision of the study, S.K.; Y.U.; Y.S.; and M.S.; Experiments and analysis of cell culture, S.K.; Y.S.; and M.S.; Computer analysis, Y.M.; and Y.U.; drafting and editing of the manuscript, Y.M.; S.K.; Y.U.; and M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded in part by research grants from JSPS KAKENHI Grant numbers 19K07177 to M.S. and 18K12137 to S.K. and the Japan Agency for Medical Research and Development, AMED (20km0104118j0003 and 21mk0101195j0001 to Y.S.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We are compliant with all data requirements.

Acknowledgments: The environmental setting was supported by Shunichi Sasaki and Kota Kurosaki.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yasuda, S.; Sato, Y. Tumorigenicity assessment of human cell-processed therapeutic products. *Biologicals* 2015, 43, 416–421. [CrossRef] [PubMed]
2. Kono, K.; Sawada, R.; Kuroda, T.; Yasuda, S.; Matsuyama, S.; Matsuyama, A.; Mizuguchi, H.; Sato, Y. Development of selective cytotoxic viral vectors for concentration of undifferentiated cells in cardiomyocytes derived from human induced pluripotent stem cells. *Sci. Rep.* 2019, 9, 3630. [CrossRef]
3. Gao, Y.; Pu, J. Differentiation and application of human pluripotent stem cells derived cardiovascular cells for treatment of heart diseases: Promises and challenges. *Front. Cell Dev. Biol.* 2021, 9, 658088. [CrossRef]
4. McGarvey, S.S.; Ferreyros, M.; Kogut, I.; Bilousova, G. Differentiating induced pluripotent stem cells toward mesenchymal stem/stromal cells. *Methods Mol. Biol.* 2021, in press. [CrossRef]
5. Theopold, C.; Hoeller, D.; Velander, P.; Demling, R.; Eriksson, E. Graft site malignancy following treatment of full-thickness burn with cultured epidermal autograft. *Plast. Reconstr. Surg.* 2004, 114, 1215–1219. [CrossRef]
6. Amariglio, N.; Hirshberg, A.; Scheithauer, B.W.; Cohen, Y.; Loewenthal, R.; Trakhtenbrot, L.; Paz, N.; Koren-Michowitz, M.; Waldman, D.; Leider-Trejo, L.; et al. Donor-derived brain tumor following neural stem cell transplantation in an ataxia telangiectasia patient. *PLoS Med.* 2009, 6, e1000029. [CrossRef]
7. Ben-David, U.; Benvenisty, N. The tumorigenicity of human embryonic and induced pluripotent stem cells. *Nat. Rev. Cancer* 2011, 11, 268–277. [CrossRef]
8. Dlouhy, B.J.; Awe, O.; Rao, R.C.; Kirby, P.A.; Hitchon, P.W. Autograft-derived spinal cord mass following olfactory mucosal cell transplantation in a spinal cord injury patient: Case report. *J. Neurosurg. Spine* 2014, 21, 618–622. [CrossRef] [PubMed]
33. Luijten, M.; Corvi, R.; Mehta, J.; Corvao, M.; Delrue, N.; Felter, S.; Haas, B.; Hewitt, N.J.; Hilton, G.; Holmes, T.; et al. A comprehensive view on mechanistic approaches for cancer risk assessment of non-genotoxic agrochemicals. *Regul. Toxcol. Pharmacol.* **2020**, *118*, 104789. [CrossRef]

34. Chour, T.; Tian, L.; Lau, E.; Thomas, D.; Itzhaki, I.; Malak, O.; Zhang, J.Z.; Qin, X.; Wardak, M.; Liu, Y.; et al. Method for selective ablation of undifferentiated human pluripotent stem cell populations for cell-based therapies. *JCI Insight* **2021**, *6*, e142000. [CrossRef]

35. Held, M.; Schmitz, M.H.; Fischer, B.; Walter, T.; Neumann, B.; Olma, M.H.; Peter, M.; Ellenberg, J.; Gerlich, D.W. CellCognition: Time-resolved phenotype annotation in high-throughput live cell imaging. *Nat. Methods* **2010**, *7*, 747–754. [CrossRef]

36. Thillaikkarasi, R.; Saravanan, S. An enhancement of deep learning algorithm for brain tumor segmentation using Kernel based unsupervised learning. *J. Med. Syst.* **2014**, *38*, 403–406. [CrossRef]

37. Niioka, H.; Asatani, S.; Yoshimura, A.; Ohigashi, H.; Tagawa, S.; Miyake, J. Classification of C2C12 cells at differentiation by convolutional neural network of deep learning using phase contrast images. *Hum. Cell* **2018**, *31*, 87–93. [CrossRef]

38. Yao, K.; Rochman, N.D.; Sun, S.X. Cell type classification and unsupervised morphological phenotyping from low-resolution images using deep learning. *Sci. Rep.* **2019**, *9*, 13467. [CrossRef] [PubMed]

39. Maruthamuthu, M.K.; Raffee, A.H.; De Oliveira, D.M.; Ardekani, A.M.; Verma, M.S. Raman spectra-based deep learning: A tool to identify microbial contamination. *Microbiologyopen* **2020**, *9*, e1122. [CrossRef] [PubMed]

40. Mzurikwao, D.; Khan, M.U.; Samuel, O.W.; Cinatl, J., Jr.; Wass, M.; Michaelis, M.; Marcelli, G.; Ang, C.S. Towards image-based cancer cell lines authentication using deep neural networks. *Sci. Rep.* **2020**, *10*, 19857. [CrossRef] [PubMed]

41. Chen, X.; Li, Y.; Wyman, N.; Zhang, Z.; Fan, H.; Le, M.; Gannon, S.; Rose, C.; Zhang, Z.; Mercuri, J.; et al. Deep learning provides high accuracy in automated chondrocyte viability assessment in articular cartilage using nonlinear optical microscopy. *Biomed. Opt. Express* **2021**, *12*, 2759–2772. [CrossRef] [PubMed]

42. Hu, H.; Liu, A.; Zhou, Q.; Guan, Q.; Li, X.; Chen, Q. An adaptive learning method of anchor shape priors for biological cells detection and segmentation. *Comput. Methods Programs Biomed.* **2021**, *208*, 106260. [CrossRef]

43. Iseoka, H.; Sasai, M.; Miyagawa, S.; Takekita, K.; Date, S.; Ayame, H.; Nishida, A.; Sanami, S.; Hayakawa, T.; Tawa, Y. Rapid and sensitive mycoplasma detection system using image-based deep learning. *J. Artif. Organs* **2021**, in press. [CrossRef]

44. Theriault, D.H.; Walker, M.L.; Kwan, J.Y.; Betke, M. Cell morphology classification and clutter mitigation in phase-contrast microscopy images using machine learning. *Mach. Vis. Appl.* **2012**, *23*, 659–673. [CrossRef]

45. Matsuoka, F.; Takeuchi, I.; Agata, H.; Kagami, H.; Shiono, H.; Kiyota, Y.; Honda, H.; Kato, R. Morphology-based prediction of osteogenic differentiation potential of human mesenchymal stem cells. *PLoS ONE* **2013**, *8*, e55082. [CrossRef]

46. Tokunaga, K.; Saitoh, N.; Goldberg, I.G.; Sakamoto, C.; Yasuda, Y.; Yoshida, Y.; Yamanaka, S.; Nakao, M. Computational image analysis of colony and nuclear morphology to evaluate human induced pluripotent stem cells. *Sci. Rep.* **2014**, *4*, 6996. [CrossRef]

47. Sasaki, K.; Sasai, H.; Takahashi, A.; Kang, S.; Yuasa, T.; Kato, R. Non-invasive quality evaluation of confluent cells by image-based orientation heterogeneity analysis. *J. Biosci. Bioeng.* **2016**, *121*, 227–234. [CrossRef] [PubMed]

48. Boumaraf, S.; Liu, X.; Wan, Y.; Zheng, Z.; Ferkous, C.; Ma, X.; Li, Z.; Bardou, D. Conventional machine learning versus deep learning for magnification Dependent histopathological breast cancer image classification: A comparative study with visual explanation. *Diagn. Histopathol.* **2021**, *11*, 528. [CrossRef]

49. Shoebibi, A.; Khodatars, M.; Jafari, M.; Moridian, P.; Rezaei, M.; Alizadehsani, R.; Hesamzadeh, F.; Gorriz, J.M.; Hilsenbeck, O.; et al. Prospective identification of hematopoietic lineage choice by deep learning. *Nat. Methods* **2019**, *16*, 403–406. [CrossRef] [PubMed]

50. Van Valen, D.A.; Kudo, T.; Lane, K.M.; Macklin, D.N.; Quach, N.T.; DeFelice, M.M.; Maayan, I.; Tanouchi, Y.; Ashley, E.A.; Covert, M.W. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS Comput. Biol.* **2021**, *17*, e1009545. [CrossRef] [PubMed]

51. Chang, Y.-H.; Abe, K.; Yokota, H.; Sudo, K.; Nakamura, Y.; Lin, C.-T.; Tsai, M.-D. Human induced pluripotent stem cell recognition in microscopy images using Convolutional Neural Networks. In Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea, 11–15 July 2017; pp. 4058–4061. [CrossRef]

52. Hong, J.; Park, B.-Y.; Park, H. Convolutional neural network classifier for distinguishing Barrett’s esophagus and neoplasia endomicroscopy images. In Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea, 11–15 July 2017; pp. 2892–2895. [CrossRef]

53. Toratani, M.; Konno, M.; Asai, A.; Koseki, J.; Kawamoto, K.; Tamari, K.; Li, Z.; Sakai, D.; Kudo, T.; Satoh, T.; et al. A convolutional neural network uses microscopic images to differentiate between mouse and human cell line nuclei and their radioreistant clones. *Cancer Res.* **2018**, *78*, 6703–6707. [CrossRef]

54. Guo, X.; Wang, F.; Teodoro, G.; Farris, A.B.; Kong, J. Liver steatosis segmentation with deep learning methods. In Proceedings of the 16th IEEE International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 24–27. [CrossRef]

55. Orita, K.; Sawada, K.; Koyama, R.; Schwarzfischer, M.; Loeffler, D.; Kokkaliaris, K.D.; Hilsenbeck, O.; et al. Prospective identification of hematopoietic lineage choice by deep learning. *Nat. Methods* **2017**, *14*, 403–406. [CrossRef] [PubMed]

56. Thillaikkarasi, R.; Saravanan, S. An enhancement of deep learning algorithm for brain tumor segmentation using Kernel based CNN with M-SVM. *J. Med. Syst.* **2019**, *43*, 84. [CrossRef]
57. Lin, Y.H.; Liao, K.Y.; Sung, K.B. Automatic detection and characterization of quantitative phase images of thalassemic red blood cells using a mask region-based convolutional neural network. *J. Biomed. Opt.* 2020, 25, 116502. [CrossRef]

58. Ma, L.; Shuai, R.; Ran, X.; Liu, W.; Ye, C. Combining DC-GAN with ResNet for blood cell image classification. *Med. Biol. Eng. Comput.* 2020, 58, 1251–1264. [CrossRef]

59. Yu, S.; Zhang, S.; Wang, B.; Dun, H.; Xu, L.; Huang, X.; Shi, E.; Feng, X. Generative adversarial network based data augmentation to improve cervical cell classification model. *Math. Biosci. Eng.* 2021, 18, 1740–1752. [CrossRef] [PubMed]

60. Isozaki, A.; Mikami, H.; Tezuka, H.; Matsumura, H.; Huang, K.; Akamine, M.; Hiramatsu, K.; Iino, T.; Ito, T.; Karakawa, H.; et al. Intelligent image-activated cell sorting 2.0. *Lab Chip.* 2020, 20, 2263–2273. [CrossRef] [PubMed]

61. Janocha, K.; Czarnecki, W.M. On loss functions for deep neural networks in classification. *arXiv 2017*, arXiv:1702.05659v1. Available online: https://arxiv.org/abs/1702.05659 (accessed on 15 October 2021). [CrossRef]

62. Bradley, J. *Distribution-Free Statistical Tests*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1968; p. 388.

63. Abe-Furukawa, N.; Otsuka, K.; Aihara, A.; Itasaki, N.; Nishino, T. Novel 3D Liquid Cell Culture Method for Anchorage-independent Cell Growth. *Cell Imaging and Automated Drug Screening. Sci. Rep.* 2018, 8, 3627. [CrossRef]

64. Yao, Z.; Gholami, A.; Arfeen, D.; Liaw, R.; Gonzalez, J.; Keutzer, K.; Mahoney, M. Large batch size training of neural networks with adversarial training and second-order information. *arXiv 2018*, arXiv:1810.01021v3. Available online: https://arxiv.org/abs/1810.01021 (accessed on 15 October 2021).

65. Perrone, M.P.; Khan, H.; Kim, C.; Kyrillidis, A.; Quinn, J.; Salapura, V. Optimal mini-batch size selection for fast gradient descent. *arXiv 2019*, arXiv:1911.06459v1. Available online: https://arxiv.org/abs/1911.06459 (accessed on 15 October 2021).

66. Alfarra, M.; Hanzely, S.; Albasyoni, A.; Ghanem, B.; Richtarik, P. Adaptive learning of the optimal mini-batch size of SGD. *arXiv 2005*, arXiv:2005.01097v1. Available online: https://arxiv.org/abs/2005.01097 (accessed on 15 October 2021).

67. Qian, X.; Klabjan, D. The impact of the mini-batch size on the variance of gradients in stochastic gradient descent. *arXiv 2004*, arXiv:2004.13146v1. Available online: https://arxiv.org/abs/2004.13146 (accessed on 15 October 2021).

68. Gao, F.; Zhong, H. Study on the large batch size training of neural networks based on the second order gradient. *arXiv 2012*, arXiv:2012.08795v1. Available online: https://arxiv.org/abs/2012.08795 (accessed on 15 October 2021).

69. Lin, T.; Kong, L.; Stich, S.U.; Jaggi, M. Extrapolation for large batch training in deep learning. *arXiv 2006*, arXiv:2006.05720v1. Available online: https://arxiv.org/abs/2006.05720 (accessed on 15 October 2021).

70. Isomura, T.; Toyozumi, T. Dimensionality reduction to maximize prediction generalization capability. *Nat. Mach. Intell.* 2021, 3, 434–446. [CrossRef]

71. Kontaxis, C.; Bol, G.H.; Lagendijk, J.J.W.; Raaymakers, B.W. DeepDose: Towards a fast dose calculation engine for radiation therapy using deep learning. *Phys. Med. Biol.* 2020, 65, 075013. [CrossRef] [PubMed]

72. Torkey, H.; Atlam, M.; El-Fishawy, N.; Salem, H. A novel deep autoencoder based survival analysis approach for microarray dataset. *Peer J. Comput. Sci.* 2021, 7, e492. [CrossRef]

73. Zhao, M.; Jha, A.; Liu, Q.; Mills, B.A.; Mahadevan-Jansen, A.; Lu, L.; Landman, B.A.; Tyska, M.J.; Hsu, Y. Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking. *Med. Image Anal.* 2021, 71, 102048. [CrossRef]

74. Zhang, C.; Weingärtner, S.; Moeller, S.; Uğurbil, K.; Akçağay, M. Fast GPU implementation of a scan-specific deep learning reconstruction for accelerated magnetic resonance imaging. *IEEE Int. Conf. Electro. Inf. Technol.* 2018, 2018, 399–403. [CrossRef]

75. Lee, J.G.; Jun, S.; Cho, Y.W.; Lee, H.; Kim, G.B.; Seo, J.B.; Kim, N. Deep Learning in Medical Imaging: General Overview. *Korean J. Radiol.* 2017, 18, 570–584. [CrossRef]

76. Salman, S.; Liu, X. Oversfitting Mechanism and Avoidance in Deep Neural Networks. *arXiv 2019*, arXiv:1901.06566v1. Available online: https://arxiv.org/abs/1901.06566 (accessed on 15 October 2021).

77. Rice, L.; Wong, E.; Kolter, J.Z. Oversfitting in adversarially robust deep learning. *arXiv 2002*, arXiv:2002.11569v2. Available online: https://arxiv.org/abs/2002.11569 (accessed on 15 October 2021).

78. Zhang, C.; Vinyals, O.; Munos, R.; Bengio, S. A Study on Oversfitting in Deep Reinforcement Learning. *arXiv 2018*, arXiv:1804.06893v2. Available online: https://arxiv.org/abs/1804.06893 (accessed on 15 October 2021).

79. Arief, H.A.; Indahl, U.G.; Strand, G.-H.; Tveite, H. Addressing Oversfitting on Pointcloud Classification using Atrous XCRF. *arXiv 2019*, arXiv:1902.03088v1. Available online: https://arxiv.org/abs/1902.03088 (accessed on 15 October 2021).

80. Daee, P.; Peitola, T.; Vehtari, A.; Kaski, S. User Modelling for Avoiding Oversfitting in Interactive Knowledge Elicitation for Prediction. *arXiv 2018*, arXiv:1710.04881v2. Available online: https://arxiv.org/abs/1710.04881 (accessed on 15 October 2021).

81. Cogswell, M.; Ahmed, F.; Girshick, R.; Zitnick, L.; Batra, D. Reducing Oversfitting in Deep Networks by Decorrelating Representations. *arXiv 2016*, arXiv:1511.06086v4. Available online: https://arxiv.org/abs/1511.06086 (accessed on 15 October 2021).

82. Ghojogh, B.; Crowley, M. The Theory Behind Oversfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial. *arXiv 2019*, arXiv:1905.12787v1. Available online: https://arxiv.org/abs/1905.12787 (accessed on 15 October 2021).

83. Yilmaz, A.; Demircali, A.A.; Kocaman, S.; Uvet, H. Comparison of Deep Learning and Traditional Machine Learning Techniques for Classification of Pap Smear Images. *arXiv 2009*, arXiv:2009.06366. Available online: https://arxiv.org/abs/2009.06366 (accessed on 15 October 2021).