ACCEPTED MANUSCRIPT

Predicting the number of printed cells during inkjet-based bioprinting process based on droplet velocity profile using machine learning approaches

Cite this Accepted Manuscript (AM) as: Accepted Manuscript (AM) version of Xi Huang, Wei Ng and Wai Yeong, Predicting the number of printed cells during inkjet-based bioprinting process based on droplet velocity profile using machine learning approaches, Journal of Intelligent Manufacturing https://doi.org/10.1007/s10845-023-02167-4

This Accepted Manuscript (AM) is a PDF file of the manuscript accepted for publication after peer review, when applicable, but does not reflect post-acceptance improvements, or any corrections. Use of this AM is subject to the publisher's embargo period and AM terms of use. Under no circumstances may this AM be shared or distributed under a Creative Commons or other form of open access license, nor may it be reformatted or enhanced, whether by the Author or third parties. By using this AM (for example, by accessing or downloading) you agree to abide by Springer Nature's terms of use for AM versions of subscription articles:
https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms

The Version of Record (VOR) of this article, as published and maintained by the publisher, is available online at: https://doi.org/10.1007/s10845-023-02167-4. The Version of Record is the version of the article after copy-editing and typesetting, and connected to open research data, open protocols, and open code where available. Any supplementary information can be found on the journal website, connected to the Version of Record.

For research integrity purposes it is best practice to cite the published Version of Record (VOR), where available (for example, see ICMJE's guidelines on overlapping publications). Where users do not have access to the VOR, any citation must clearly indicate that the reference is to an Accepted Manuscript (AM) version.

Accepted manuscript

Predicting the number of printed cells during inkjet-based bioprinting process based on droplet velocity profile using machine learning approaches

Xi Huang[1+], Wei Long Ng[1,2+*], Wai Yee Yeong[1,2*]

[1] HP-NTU Digital Manufacturing Corporate Lab, Nanyang Technological University (NTU), 65 Nanyang Drive, Singapore 637460

[2] Singapore Centre for 3D Printing (SC3DP), School of Mechanical and Aerospace Engineering, Nanyang Technological University (NTU), 50 Nanyang Avenue, Singapore 639798

[+] Both authors have contributed equally towards this paper.

Email: ng.wl@ntu.edu.sg; wyyeong@ntu.edu.sg

Abstract:

In this work, our proof-of-concept study can be used to predict the number of cells within printed droplets based on droplet velocity at two different points along the nozzle-substrate distance using machine learning approaches. A novel high-throughput contactless method that combines the use of an optical system and machine learning algorithms was utilized for various applications such as cell detection within single droplets (presence/absence of cells) and prediction of the total number of printed cells within multiple droplets by measuring the droplet deceleration between two positions along the nozzle-substrate distance. The proposed method in this work has demonstrated good accuracy in cell prediction within single droplet (presence/absence of cells) and low prediction error in determining number of cells within multiple droplets by reducing the error by a factor of $\sqrt{N}$ for N droplets measured in a batch. The performance of five different machine learning algorithms such as linear regression, support vector regression, decision tree regressor, random forest regression, and extra tree regression were compared to determine the best algorithm for each type of application. The random forest regressor algorithm demonstrated the highest accuracy at 80% in cell prediction (presence/absence of cells) within single droplets, while the extra tree regressor demonstrated the lowest mean error of 12% in predicting the number of printed cells within multiple droplets (e.g., 20 droplets on same spot). By incorporating these models in a droplet monitoring system, live assessment of the number of printed cells during an inkjet-based bioprinting process can be achieved.

Keywords: 3D bioprinting; 3D printing; Biofabrication; Machine learning; Drop-on-demand bioprinting; living cells

1. Introduction

Over the years, there is an increased interest in the use of 3D bioprinting technologies for tissue engineering, regenerative medicine, and biomedical applications due to its exceptional ability to precisely deposit highly viable cells during the bioprinting process to fabricate biomimetic 3D tissue-engineered constructs (He et al., 2021; Levato et al., 2020; Ng, Chan, et al., 2020; Ng et al., 2019; W. L. Ng et al., 2016; Sun et al., 2020). These 3D bioprinting technologies can be categorized into 3 distinct groups according to the ASTM standards – jetting-based (Choe & Kim, 2020; X. Li et al., 2020; Wei Long Ng et al., 2016), extrusion-based (Meng et al., 2020; Ng, 2014; Ozbolat & Hospodiuk, 2016) and vat photopolymerization-based (W. Li et al., 2020; Ng, Lee, et al., 2020). Each bioprinting technology has its distinctive advantages and limitations and the selection of an appropriate bioprinting technology should be dependent on its intended application (Lee et al., 2019). The key advantages of each bioprinting technology are as follows: the jetting-based bioprinting approach facilitates contactless on-demand bioprinting of different living cells and biomaterials to enhance cell-cell and cell-matrix interactions (Ng, Lee, et al., 2017; W. L. Ng et al., 2018), the extrusion-based bioprinting approach facilitates printing highly viscous bio-inks with fast fabrication speed (Yang et al., 2022; Zhuang et al., 2019), and lastly the vat polymerization-based bioprinting approach can achieve high printing resolution and printable cell concentration (Nieto et al., 2020).

For bioprinting applications, the living cells are usually encapsulated within hydrogel-based bio-inks and printed using different bioprinting technologies which have their own specific bio-ink requirements. The printed cells in jetting-based bioprinting approach are usually encapsulated within un-crosslinked bio-inks of low viscosity (Ng et al., 2022; Ng et al., 2023), whereas the printed cells in extrusion-based and vat polymerization-based bioprinting approaches are usually encapsulated within highly-viscous crosslinked bio-inks (W.L. Ng et al., 2016; Suntornnond et al., 2022). As compared to the other 3D bioprinting approaches, the jetting-based bioprinting approach facilitates higher degree of control over printed cell density by manipulating the number of printed cell-laden droplets at specific positions (Ng et al., 2021; Takagi et al., 2019; Xu et al., 2013). Hence, the jetting-based bioprinting approach facilitates precise deposition of multiple types of living cells and biomaterials by manipulating the spatial positions of these functional living cells and biomaterials for fundamental studies that aim to elucidate the cell-cell and cell-biomaterial interactions (Gudupati et al., 2016; Wei Long Ng et al., 2018; Ng & Yeong, 2019).

Traditional methods for cell detection include the use of microfluidics, image-based analysis and laser induced fluorescence (LIF). The main limitations of these traditional methods include challenges to automate the image processing process using advanced image analysis algorithms which lead to low throughput monitoring for real-time applications (Köster et al., 2008; Mary et al., 2011; Schoeman et al., 2014), and the need to stain the printed cells with fluorescence dyes (Chen et al., 2022). For microfluidics, the cell count can be quantified through microflow cytometry which utilizes multiple techniques to count the cells such as optical (Piyasena & Graves, 2014), acoustic (Chen et al., 2014), or impedance measurement (Tang et al., 2017). Although the microflow cytometer can measure cell count in microfluidic flow channel with high throughput, the cell quantification usually occurs before the droplet is ejected from the nozzle. Hence, it may not be representative of the number of printed cells within the droplets. For image-based analysis, a high-speed camera is utilized to capture the frames with cells within the droplet and provide detailed information on the shape of the cells (Köster et al., 2008; Mary et al., 2011; Schoeman et al., 2014). However, it is challenging to process the images with automated image analysis algorithms and hence manual cell counting of images results in low throughput rates. Laser induced fluorescence (LIF) provides higher throughput by reading fluorescence signal from a droplet excited by lasers (Sciambi & Abate, 2015). With this method, since at least 95% of droplets containing cells would contain a single cell at low cell occupancy (< 0.1 cell per droplet) (Klein et al., 2015; Macosko et al., 2015), therefore the number of cells can be calculated by assuming that all ejected droplets with cells were detected as a single cell droplet. These traditional methods, despite their efficacy, cannot be implemented for real-time cell counting and no existing method can monitor the number of printed cells at high throughput rates. There is existing work that utilizes high speed camera to capture the images of fluorescence-labelled cells excited by multiple class 3b and class 4 lasers at various incidence angles and perform the calculation of average cell count over multiple droplets in real time at a high throughput (Chen et al., 2022). This method calculates the average cell count from the area with fluorescence over many droplets. However, the high intensity from the lasers damaged the encapsulated cells within the droplets (~69% cell viability after laser excitation) and the use of fluorescence-labelled cells may not be desirable for some bioprinting applications. The ability to accurately monitor the number of printed cells is critical for fabrication of *in-vitro* 3D cell-laden tissue constructs in a scalable and reproducible manner.

Hence, the goal of this study is to utilize machine learning approaches on high-speed images captured to predict the number of encapsulated cells during the droplet bioprinting process. Machine learning is a subset of artificial intelligence which can identify the relationship between large datasets efficiently, resulting in a model that is useful in the prediction of new inputs (Xames et al., 2022). Machine learning is useful when statistical relationships exist in the dataset but the model for the dataset cannot be determined analytically, such as optimizing printing parameters (Bonatti et al., 2022; Fu et al., 2021; Law et al., 2023), predicting the cell viability from the process parameter and identifying the significant process parameters for high cell viability (Xu et al., 2022). There are three basic approaches to apply machine learning for different situations; supervised learning (Caruana & Niculescu-Mizil, 2006), unsupervised learning (Ghahramani, 2004; Hastie et al., 2009), and reinforcement learning (Arulkumaran et al., 2017; Kaelbling et al., 1996; Wiering & Van Otterlo, 2012). Supervised learning creates a model from labeled data to predict the output from input, unsupervised learning extracts features out of unlabeled data, while reinforcement learning deduces the best step to maximize a reward function. Supervised learning is the most common form of machine learning, and it requires a large dataset of accurately labelled images. This model would undergo training iteration to compute the error between the predicted probabilities and ground-truth labels while internal weights were adjusted to minimize the error value. On the other hand, unsupervised learning averts the laborious step of data annotation such that the model can extract features out of unlabelled data and learn useful representations from the data. Lastly, reinforcement learning involves a neural network (agent) that is trained to maximize the value of a reward through a trial-and-error process.

In this work, supervised learning is the approach used to predict the cell count in droplets from various inputs. The relationship between the change in the velocity profile and the cell count of the printed droplet will be studied with machine learning (ML). The 2 main methods in ML to predict a number from input are regression and decision tree. In regression, the algorithm tries to fit each coefficient in a formula to maximize the accuracy or minimize the loss. In decision tree, the algorithm create a tree of decision that decide the output based on the input (Quinlan, 1987). Regression-based ML algorithm used in this work are linear regression and support vector regression (SVR) (Smola & Schölkopf, 2004), while the decision tree-based ML algorithm used are decision tree regressor, and ensemble method (Bauer & Kohavi, 1999) such as random forest regressor (Breiman, 2001) and extra tree regressor (Geurts et al., 2006). The mentioned ML technique is graphically depicted in **Figure 1**. Linear regression is the simplest regressor where the coefficient is fitted to minimize the loss for a best fit line, this method is fast and easy to implement, but is not accurate when the statistical relationship between data is not linear. Support vector regressor maximize the number of data point in its hyperplane area, SVR perform better than linear regression when there is outlier in data. Ensemble method combines multiple decision trees and predict the output by averaging the output from the multiple decision tree, which increase the efficiency of decision tree and prevent overfitting. Random forest regressor and extra tree regressor are both tree ensemble method with similar approach, but with a slight difference in which extra tree regressor use the whole original sample for training to reduce bias and choosing the split node randomly to reduce variance. The performance of these ML techniques on our samples are compared in this work to determine the optimal algorithm for cell prediction in both single and multiple droplets during jetting-based bioprinting process.

Figure 1: Graphic representation of various machine learning techniques.

2. Materials and Methods

2.1. Preparation of Cell-laden Bio-inks

The primary human dermal fibroblasts (HDFs) used in this work were purchased from CellnTec Advanced Cell Systems and routinely passaged in tissue culture flasks (Passages 3 – 5). Fibroblasts were cultured in CnT-Prime Fibroblast Proliferation Medium (CnT-PR-F, 1% serum medium supplemented with fully defined growth factors and co-factors) at a temperature of 37^{o}C, 5% CO_2 and the culture medium were changed once every three days. The adherent fibroblasts were harvested using CnT Accutase cell detachment solution (CnT-Accutase 100) at 90% confluency. Based on our previous work, the use of 2% PVP-based bio-inks improved the droplet printing quality and cell viability during the inkjet bioprinting process (Ng, Yeong, et al., 2017). Hence, the detached fibroblast cells were suspended in 2% w/v PVP-based bio-inks to obtain different concentrations of cell-laden bio-inks (1 – 3 million cells/ml).

2.2. Bioprinting process

A thermal inkjet printer (HP D300e Digital Dispenser) and non-commercial prototype cassettes with 8 embedded thermal inkjet print-heads with large nozzle orifice diameter to enable cell printing were used for cell bioprinting. The thermal inkjet print-head dispensed discrete droplets of constant volume of ~0.345 nL at a printing frequency of 1 kHz. The desired volume can be controlled by manipulating the number of printed droplets at the same spot.

2.3. Experimental Setup

Cell-laden bio-inks (2% w/v PVP of varying cell concentrations from 1 – 3 million cells/ml) were ejected onto glass slides using a thermal inkjet printer (HP D300e Digital Dispenser). Multiple single, discrete cell-laden droplets were dispensed and captured using a Photron FASTCAM Nova S12 (high-speed camera) at 120,000 frames per second (fps) as shown in **Figure 2a**. The image of the experimental setup is presented in **Figure A** of Supplementary Information. During the printing process, the underlying substrate translated along the x-axis to obtain single, discrete deposited cell-laden droplets. These deposited droplets were then counted using a light microscope to determine the number of printed cells within each deposited droplet. The droplet velocity profile of the cell-laden bio-inks at different positions (0.7 mm, 1.0 mm, 1.5 mm, 2.0 mm, 2.5 mm, and 3.0 mm) along the nozzle-substrate distance was analyzed using the high-speed camera images. The number of printed cells within each deposited droplet and its corresponding droplet velocity profile were used as training data in the machine learning model.

2.4. Image Processing

As depicted in **Figure 2b**, the captured high-speed image is first filtered with the seeded threshold method to isolate the main droplet and remove the noise from the image for further processing (Xi, 2021). In the seeded threshold method, a seed threshold is applied to the image on top of the usual threshold. The pixels filtered out by the first threshold will be grouped into their connected area respectively. A connected area that is below the threshold and contain pixel below the seed threshold will be identified as the droplet, whereas the connected area that is below the threshold but without a pixel below the seed threshold will be removed as noise. This method removes noises by assuming the real droplets will always have a pixel that is darker than all noises or negligible satellite droplets.

By using a multiple object centroid tracker, the displacement of each droplet at each frame is identified. The centroid tracker calculates the centroid of each droplet and adds the droplet’s velocity to the coordinate of the centroid to predict the droplet’s position at the next subsequent frame (assuming an initial droplet velocity of 20 m/s). Each droplet is compared to the predicted droplet position from the previous frame, and the droplet with the smallest displacement from the predicted droplet position is matched as the same droplet. The droplet velocity is then tagged to every matched droplet, and the displacement data of all discrete droplets at every frame are recorded.

Displacement data is converted to velocity with the formula $v_{i+1} = (d_{i+1} - d_i)/T$, where v_{i+1} and d_{i+1} is the velocity and displacement of current frame respectively, and T is the time interval between each frame (1/120000 second). The velocity at the displacement of 0.7 mm, 1.0 mm, 1.5 mm, 2.0 mm, 2.5 mm, and 3.0 mm is interpolated from the previous velocity-displacement data.

2.5. Sample Selection

The distribution of the samples acquired are as shown in **Figure 2c**. More samples at concentration of 3 million/ml were included to increase the sample with higher cells count. We had selected 67 droplets with 0 cell, 61 droplets with 1 cell, and 28 droplets with 2 cells. The sample with 3 cells or above is not included in the machine learning sample set as the sample count is too low and can be seen as outlier. The sample set has a train-test split of 60 to 40. The training set contains 91 samples, and the testing set contains 65 samples. The split is done completely random from the sample set.

2.6. Training Algorithm

The cell count of the sample was trained against the velocity of the sample at 2 different positions with 5 different algorithms to determine the efficiency of each algorithm in determining the availability of cells in single droplet and cell count in multiple droplets. The architecture for the training of model and the usage of the model for prediction is detailed in **Figure 2d**.

Figure 2: (a) Schematics of experimental setup. (b) Seeded threshold method and centroid tracker to extract displacement from image. M1 is the main droplet to be detected while the S1 is the satellite droplet to be ignored. (c) Distribution of samples by cell counts and cell concentration of the solution. (d) Flowchart of the training and usage of the machine learning model for cell prediction.

The 5 algorithms being used are linear regression (LR), support vector regression (SVR), decision tree regressor (DTR), random forest regression (RFR), and extra tree regression (ETR). LR and SVR are both regression-based model, where LR assumes a linear model while SVR assumes a non-linear model by using the radial basis function (RBF) kernel. DTR, RFR, and ETR are decision tree-based model, in which their suitability for predicting the cell count will be compared against regression-based model. DTR is the basic for the decision tree-based technique, while the RFR and ETR are both decision tree-based method that utilized the ensemble method to improve the accuracy of decision tree with the trade-off of having low readability and high computational cost. The machine learning models are implemented with Scikit-learn (Pedregosa et al., 2011). A brief introduction for the models and the tuning parameters used will be detailed in **Section 2.6.1** to **Section 2.6.5**.

2.6.1. Linear regression

The LR model was chosen as the control to compare with other machine learning models, as it is easy to understand and has the highest readability out of all machine learning method. This model assume that the dataset is a linear model. LR fit the model with ordinary least square method, $y = \omega_0 + \omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_p x_p$, where y is the predicted cell count, x_i is the independent variables (velocity of droplet at different positions), and ω_i is the coefficient to be fitted. The model is fitted by minimizing the residual sum of squared error between the observed and predicted cell count in the dataset, $J(\omega_0,\omega_1,\ldots,\omega_n) = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2$, where $J(\omega_0,\omega_1,\ldots,\omega_n)$ is the residual sum of the squared error, and ε_i is the residual error for the prediction for the i^{th} data point.

2.6.2. Support vector regression

The SVR model is a commonly used ML regression model and is used to compare with the accuracy of LR model as they are both regression-based method. SVR maximize the number of data points in the hyperplane area, thus performing better than LR when there is outlier in data (Smola & Schölkopf, 2004). The kernel function used is RBF for a non-linear regression, $K(x_1,x_2) = \exp(- \gamma\|x_1 - x_2\|^2)$, where x_1 and x_2 are 2 points in the dataset, and γ is the tuneable parameter gamma. The tuneable gamma value used is 'scale', $\gamma = 1/[\text{number of features} * X.\text{var}()]$. The regularization parameter C is set to 1.0, epsilon as 0.1, and the tolerance for stopping criterion is 0.1. The solver is iterated with no limit until the stopping criterion is achieved.

2.6.3. Decision tree regressor

The DTR learns to approximate the value of target variable with a set of if-else-then functions (Quinlan, 1987). The advantage of DTR is that it is easy to visualize and interpret a dataset with good fit, but it came with the disadvantage where it is easy to overfit. The DTR is chosen as it is the most basic technique for decision tree based regressor. The criterion for splitting used in this work is mean squared error, where the strategy used by the splitter is best split approach. The maximum feature for best split is the square root of the total feature. The hyperparameters used are determined by grid search method, with a maximum depth of 5, minimum sample splits of 10, and minimum sample in leaf node of 4.

2.6.4. Random forest regression

RFR creates multiple DTR using subset of the samples, followed by averaging the output of the multiple trees for the prediction of the target (Breiman, 2001). RFR is better than DTR in term of accuracy and avoiding overfitting. However, it is very difficult to visualize or interpret a RFR model. The criterion for splitting used is mean squared error. The maximum feature for best split is the square root of the total feature. The number of trees in the forest is set as 300 with a maximum depth of 5, minimum sample splits of 10, and minimum sample in leaf node of 4.

2.6.5. Extra tree regression

ETR is similar to RFR in term of approach, with the difference in usage of entire original sample set instead of bootstrap replica to reduce bias, and choosing the sample cut point randomly instead of optimally to reduce variance (Geurts et al., 2006). The criterion for splitting used is mean squared error. The maximum feature for best split is the square root of the total feature. The number of trees in the forest is set as 300 with a maximum depth of 5, minimum sample splits of 10, and minimum sample in leaf node of 4.

3. Results and discussions

The velocity profile of ejected ink with 3 million cells/ml in **Figure 3a** shows some obvious variation between different droplets, with some of droplets sharing similar velocity profiles to ink with no cells. We have attributed this variation to the number of cells in a droplet. The presence of cells in a droplet changes the mass distribution of the droplet, affecting droplet shape, which alters drag force on the droplet as the shape and area of the droplet with respect to the velocity direction change. As shown in **Figure 3b**, a change in droplet shape is observed with increasing cell concentration. It is observed that most droplets at higher cell concentration (3 million cells/ml) exhibited slimmer droplet head as compared to those droplets without cells (0 million cell/ml). The slimmer droplet head at higher cell concentration lowers the drag force, which reduces the droplet deceleration. There is a huge variability in the average velocity profiles of different bio-inks (0 to 3 million cells/ml) as shown in **Figure 3c**; whereby the R-squared value of the best-fit line decreases with increasing cell concentration (from R^2 = 0.9191 in 0 million cell/ml to R^2 = 0.5009 in 3 million cells/ml). A higher variance in the number of printed cells (0 – 0 million cell/ml, 0.13 – 1 million cells/ml, 0.41 – 2 million cells/ml and 1.16 cells – 3 million cells/ml) is observed with increasing cell concentration. Furthermore, an increase in the cell concentration also leads to a higher variance in the droplet velocity profile. Hence, there is a strong correlation between the number of printed cells and droplet velocity profile.

Spearman correlation ranks the monotonic relation between variables, with the correlation being 1 when the two variables are a perfect monotonic relation of each other. The closer the correlation is to 0, the less the two variables are monotonic related. The Spearman correlation between variables in the dataset, as shown in **Figure 3d**, reveals that a low monotonic relationship exists between number of cells and other variables in the dataset, with the highest rank being cell concentration, with the value of 0.53, and 0.12 to 0.23 for velocity at various positions. This is understandable as more cells will be ejected with higher cell concentration in the ink solution.

This can be confirmed from **Figure 3e**, where the droplet is plotted on a graph of velocity at 2 mm against the velocity at 1mm. When the deceleration of the droplet is higher, there are more droplets without cells compared to droplets with cells. In **Figure 3f**, deceleration of droplets from a displacement of 1 mm to 3 mm are plotted against the cell count in the droplet. Although the average droplet deceleration is lower with higher cell counts, there is no statistical distinction between both type of droplets as the datasets still overlaps with each other. This means that a more complex relationship exists between these variables. Machine learning approaches can be utilized to understand if a connection exists between velocity at different positions and hence cell count in the droplet when the relationship between the variables is difficult to be interpreted manually. In the following sections, we present two machine learning methods for cell count prediction in single droplet and cell count prediction in multiple droplets respectively.

Figure 3: (a) Droplet velocity profile of different droplets with cells (black dot – 3 million cells/ml) and no cell (red dot – 0 million cell/ml) plotted against the displacement from the nozzle. (b) Influence of cell concentration on the shape of droplet head – wider droplet head at 0 million cell/ml and slimmer droplet head at 3 million cells/ml. (c) Average velocity profile of cell-laden bio-inks (best fitted from velocity profile of 6 droplets) with cell concentration of 0 to 3 million cells/ml. (d) Spearman correlation of all the variables in the dataset. (e) Droplet with cell (black dot) and no cell (red dot) plotted with their velocity at 1 mm and 3mm from the nozzle. (f) Deceleration of droplet from 1 mm to 3 mm plotted against cell count.

3.1. Cell availability prediction in single droplet

The cell count in the training set is fitted against the velocity of the droplet at 2 positions chosen from 0.7 mm, 1.0 mm, 1.5 mm, 2.0 mm, 2.5 mm, and 3.0 mm. The optimized model is used to predict the number of cells in the testing set. The value acquired is rounded and compared against the actual value in the testing set. The accuracy in predicting the availability of cell in a single droplet of different model is shown in **Figure 4**. The random forest regression model provides the best accuracy, up to 80% accuracy in predicting the availability of cells in a single droplet, when velocity of the droplet at 1.0 mm and 2.0 mm is used as the input for the model.

Figure 4: Accuracy in determining if the droplets contain cells. The value in bracket is the best accuracy among the matrices of different detection position. (a) Linear regression (0.77) (b) SVR (0.77) (c) Decision Tree (0.77) (d) Random Forest Regression (0.80) (e) Extra Tree Regression (0.79).

LR has the highest accuracy at 77% when the position used are 0.7 mm and 2.0 mm. The other 4 methods, SVR, DTR, RFR, and ETR, have the highest accuracy when the position used are 1.0 mm and 2.0 mm. Both SVR and DTR has the highest accuracy at 77%, while RFR and ETR has the highest accuracy of 82% and 80% respectively. Although the accuracy of the models might be improved with droplet velocity taken from more than 2 points, it is more desirable to have droplet velocity taken at lesser points to reduce the complexities in optimizing the desired position for velocity detection. **Figure 5** demonstrates the confusion matrix prediction for the best 2 positions. The accuracy of the confusion matrix is the ratio of the correct prediction to all other predictions. Precision is the ratio of the true positive of the category to both the true positive and false positive of that category. This metric is indicative of how good the model is at predicting a particular category. Recall is the ratio of the true positive of the category to the sum of the true positive and false negative of that category, this is the percentage in which the model can detect a particular category. Accuracy is the ratio of the sum of true positive and true negative of the category to the sum of all results, it is indicative of the overall prediction accuracy.

Figure 5: Confusion matrix for single droplet detection with the best 2 features as input. (a) Linear regression (b) SVR (c) Decision Tree (d) Random Forest Regression (e) Extra Tree Regression.

The velocity at the position of 1 mm and 2 mm gives the highest accuracy in determining the availability of the cell in droplets for 4 out of 5 machine learning algorithms. The deviation in accuracies of five different type of algorithm is relatively small (77% to 82%). This indicates that there exists a linear relationship between the number of printed cells and the velocity of the droplet at 2 positions. From the linear regression, the cell count (number of printed cells) can be estimated as

$$\text{Cell Count} = -0.0862V_{0.7} + 0.1430V_{2.0} - 0.5873$$

where $V_{0.7}$ and $V_{2.0}$ are the velocity of the droplet at 0.7 mm and 2.0 mm respectively in m/s. By defining acceleration, $a = (V_{2.0} - V_{0.7})/t$, and the time taken for the droplet to travel from 0.7 mm to 2.0 mm, $t = 0.0013/V_a$, where V_a is the average velocity between 0.7 mm and 2.0mm, the model can be rearranged to

$$\text{Cell Count} = \left(\frac{1.1206 \times 10^{-4}}{V_a}\right)a + 0.0568V_{2.0} - 0.5873$$

The linear model shows that the previous assumption on lower cell count will result in higher deceleration (lower acceleration) is indeed true.

Among all the models, LR has the highest precision at 100% when predicting droplet with no cells, while ETR has the highest precision of 91% for droplet with cells. This implies that ETR gives the best prediction if the user wished to ensure that the droplet contain cells, and LR gave the best prediction for droplet without cells. For a balanced precision between droplet containing cells and no cells, RFR has the highest accuracy of 82% as compared to other 4 models.

Predicting the availability of a cell in a single droplet is useful in determining the number of cells in droplets with low cell occupancy (<0.1 cell per droplet). At low cell occupancy, 95% of the non-empty droplet contain only a single cell (Klein et al., 2015; Macosko et al., 2015). Therefore, a binary output of yes or no can be approximated to 1 cell or 0 cell respectively. This provides a contactless high throughput method to determine the cell count in individual printed droplet for ink with low cell concentration.

3.2. Cell number prediction in multiple droplets

While the method from last section works well in predicting cell count in single droplet at low cell occupancy, it has a very low accuracy for the sample used in this work as seen in **Figure 6a**, with a large 79% mean error for ETR method. This is likely due to the sample used for this work which has a high cell occupancy at 0.76 cell per droplet (117 cells in 154 droplets). At such a high error rate, this method is not feasible to predict the cell count in cell-laden ink with high cell concentration. To improve the accuracy of the predictor, a multiple droplet approach was taken. The cell count in the training set was fitted against the velocity in 2 positions (Position 1 = 1.0 mm and Position 2 = 2.0 mm) with the best accuracy from **Figure 4**. The optimized model was then used to predict the cell count in the testing set. 1 to 40 random droplets were chosen from the testing set, in which the actual cell counts in the chosen droplets were summed up and compared with the summed predicted cell count (**Figure 6a**). This process is replicated 10 times for each droplet count. The mean error and the standard deviation of the prediction at 5, 10, 20, and 40 droplets for each model is shown in **Figure 6b**.

Figure 6: (a) Mean error in determining the total cell count in multiple droplets of 1 to 40 droplets. (b) Mean error and standard deviation in determining the total cell count in multiple droplets of 5, 10, 20, 40 droplets.

At 20 droplets, ETR has the best accuracy with a mean error of 17%, followed by SVR with a mean error of 19%, DTR with a mean error of 20%, RFR with a mean error of 22%, and LR with the worst accuracy at a mean error of 31%. The confusion matrix for the cell count prediction of 20 droplets is shown in **Figure 7**. 20 droplets were selected instead of other droplet numbers for comparison as it represents the optimized number of droplets to balance accuracy and low number of droplets. A higher droplet count per prediction provides more accuracy but will increases the cell count deposited per batch of prediction and therefore reduce the resolution for cell count control. At 20 droplets, mean error drops below 20% for some of the algorithms. For further reduction in the mean error, the droplet counts required will be growing exponentially and is not realistic for real application.

Figure 7: Confusion matrix for cell prediction in 20 droplets. (a) Linear regression (b) SVR (c) Decision Tree (d) Random Forest Regression (e) Extra Tree Regression

From the confusion matrix (**Figure 7**), the regression-based method (LR and SVR) generally tries to predict the average of the cell count in the entire dataset. As example, the range of LR prediction is from 13 to 16 while the actual range of cell count is from 9 to 18. Meanwhile, the decision tree-based method (DTR, RFR and ETR) avoided this tendency. This shows that for the cell number prediction, a decision tree-based method should be used instead of a regression-based method to prevent the model reliance on average values.

The reduction in error by increasing number of droplets can be explained by the summation of normal distributed variables. The error in the prediction of the cell count in a single droplet of extra tree regressor is shown in **Figure 9a**. Assuming that the error of the prediction is normally distributed with a mean of 0 and a variance of σ^2, the error of a single prediction will be $\pm 2\sigma$ with a 95% confidence. With N droplet count, the variance of the error distribution, σ_N^2 is

$$\sigma_N^2 = N\sigma^2$$

The standard deviation of the prediction of N droplets is

$$\sigma_N = \sqrt{N\sigma^2}$$

$$\sigma_N = \sqrt{N}\sigma$$

Thus, the error of the prediction with N droplets will be $\pm 2\sqrt{N}\sigma$ with 95% confidence. The distribution of error is reduced by a factor of $\sqrt{N}$. From this calculation, the mean error can be halved by increasing the droplet count with a factor of 4 from 5 to 20 for ETR. The mean error for ETR at 5 droplets is 29%, while the mean error at 20 droplets is 17%, which is a 41% reduction in mean error. This value is close to the predicted error reduction, which confirms that increasing the droplet count reduces the prediction error. The deviation in the error reduction is a result of the mean prediction error being > 0 as assumed, and the distribution of the error is skewed.

The source of prediction error is mainly due to the lack of the data for the cell diameter and the spatial positioning of the printed cells in each individual droplet. The velocity profile of a droplet is affected by the bio-ink properties, number of printed cells in the droplet, cell diameter, and the spatial positioning of the printed cells in the droplet. Although the rheological properties of the bio-inks are assumed to be uniformly constant, it is important to note that there is a variation in the cell diameter (18.9 ± 1.3 μm for human fibroblast cells used in this work) and the spatial positioning of cells within each individual droplet varies even for the same cell count (e.g., the printed cells could be closer to the core in one droplet, whereas the printed cells in another droplet could be closer to the edge). The measurement of these data in real-time is challenging and hence an increased sampling count is used to reduce the random errors.

3.3. Model validation for cell number prediction

The effectiveness of the optimized models in real application is validated on a completely new set of samples. In the validation set, 10 droplets are printed consecutively on the same location with the cell-laden bio-inks (3 million cells/ml), in which the total number of cells in that location is measured with a microscope. The velocity of each droplet at 1.0 mm and 2.0 mm displacement from the nozzle is analyzed with a high-speed camera at 50,000 fps. This lower fps will be helpful in testing the robustness of the model when the velocity measured has lower accuracy. The total number of cells in 2 consecutive locations are combined, such that each data point consists of total cell count and velocities of 20 droplets. 8 data points (160 droplets) are used to validate the random forest regressor and the extra tree regressor models.

The confusion matrix for the cell count prediction of the validation set is given in **Figure 8**. The RFR has a mean error of 13% while the ETR has a mean error of 12%. From the previous sample set, the cell occupancy of the bio-ink with 3 million cells/ml is 1.29, therefore the average cell count in 20 droplets are 25.8 cells. Using the average cell count as prediction will give an error of 23% for the validation set. Therefore, these results show that the model is an improvement over using the average cell count as prediction for the total cell count in 20 consecutive droplets, even when the velocity measured has lower accuracy.

Figure 8: Confusion matrix for the validation set. (a) Random Forest Regression (b) Extra Tree Regression

3.4. Model optimization by removing unusable data

It may seem logical to include the cell concentration of the solution as an input variable given it shows the highest Spearman correlation value with respect to cell count in droplet. However, the performance of the model is reduced when the cell concentration is included alongside the velocity at 2 positions from the nozzle. Using the ETR model as comparison in **Figure 9d**, the mean error at 20 droplets is increased to 27% as compared to 17% of the original model. This could be due to the overfitting of data points when the concentration is included. This same problem occurs when velocity at more than 2 points is included, and performance is degraded as compared to just 2 points. Therefore, for a robust machine learning model, only the velocity at 1 mm and 2 mm should be used as the input variables.

The droplet velocity for bio-ink containing 0 million cell/ml is expected to be similar to those droplets containing 0 cell and the inclusion of these droplet velocity data for 0 million cell/ml should improve the accuracy of ML algorithm. However, the generated result as shown in **Figure 9d** and **Figure 9e** suggested otherwise when 25 samples of bio-ink with 0 million cell/ml is included in the training. The mean error of the model is significantly higher than the original model, with a mean error of 49% at 20 droplets. This large deterioration in the accuracy could not be explained by overfitting alone unlike the case when including concentration data as part of the input variable. This indicates that the velocity profile at 1 mm and 2 mm for bio-ink with 0 million cell/ml is very different from droplet containing no cells ejected from cell-laden bio-ink. It can be concluded that the bio-ink with no cells is not compatible as the sample for the training of ML models to detect cells from velocity at different points.

Figure 9: (a) Histogram of prediction error for the cell count in a single droplet through extra tree regressor, fitted with gaussian distribution. (b) Confusion matrix for extra tree regression model when concentration is included. (c) Confusion matrix for extra tree regression model when ink with no cells is included. (d) Mean error in determining the cell count when concentration or ink with no cells are included as a variable for extra tree regression model 1 to 40 droplets. (e) Mean error and standard deviation in determining the cell count when concentration or ink with no cells are included as a variable for extra tree regression model at 5, 10, 20, 40 droplets.

3.5. Laser detection system for velocity measurement

These results can be utilized to build an *in-situ* droplet monitoring system where the deposited cell count is predicted from the velocity of the droplets at 2 positions, as depicted in **Figure 2d**. The velocity of droplets can be detected using a high-speed camera or a laser. High speed cameras will provide a better accuracy for the measurement of the droplet velocity but come with several disadvantages including expensive equipment, bulky incorporation in a 3D bioprinting process, requirement of a strong light source that may degrade the bio-ink, and intensive computational requirements for the image processing of droplet’s image. By contrast, a laser-based measurement system is cheap and is easy to incorporate in an existing 3D bioprinter as shown in **Error! Reference source not found.**. The D300e Digital Dispenser used in this schematic is a thermal inkjet printer and the proposed setup is applicable for most droplet-based printers. A weaker laser, such as a Class 2 laser at 1mW with a wavelength of 532 nm should be used to minimize the laser-induced damage to the printed cells within droplets. The laser would be directed to 1 mm and 2 mm below the nozzle to be detected by the photodiode sensor across the printer. When a droplet passes the laser line, the light will be scattered or absorbed by the droplet, which lowers the intensity count in the sensor. By utilizing both the firing timing and the time where the droplet is detected by the laser, the velocity of the droplets at 2 different displacements from the nozzle can be calculated. This value can either be used to train a model or be fed into the trained model to predict the cell counts per batches of droplets (such as 20 droplets for the model in this work). Once the system determines that the total number of cells in a particular position is sufficient, the cell printing could be halted.

Figure 10: Design schematic for laser detection system in a D300e Digital Dispenser to measure the velocity of the droplets.

Although the difference in the bio-ink properties (viscosity, surface tension and density), cell diameter, nozzle design, and printing parameters (e.g., pulse voltage) may affect the initial velocity and the deceleration rate of the droplets, the effect of these parameters on the accuracy of the models is not in the scope of this work. Furthermore, multiple fold of samples will be required to obtain a more robust model with increased inputs using the aforementioned parameters. Some of the experimental constraints include hardware limitations such as the need for huge data storage space to store the images captured using the high-speed camera (~16 GB of image data per 27 samples), lengthy processing time and potential human error while handling the massive amount of data. This work has proved that it is possible to predict the cell count by using droplet velocity at 2 different points along the nozzle-substrate distance, hence the laser detection system is suggested to eliminate the need for image processing in future work.

The main advantage of the laser detection system is that it will significantly reduce the processing complexity needed for a set of experiment. Capturing image with high-speed camera involve calibration of camera focus, backlight intensity, and parameters for image processing. These processes are time-consuming and will induce human errors. With this laser detection system, it is possible to avoid these processes to cut down the overhead time from hours to minutes per run. The sampling rate can be increased significantly, and the data quality will be improved as there is lesser error. This will allow us to train a deep neural network (DNN) model with the samples. While it is possible for DNN to perform better than the machine learning algorithm that is used in this work, the number of samples required is massive. We attempted to perform training a deep neural network with our samples, but the sample count was not sufficient for the network to achieve a good fit. This work demonstrates the relationship between the velocity of a droplet at multiple points and the cell count in the droplet. This data lends confidence to further exploration of predicting cell count with DNN by including more samples. Using a DNN for cell prediction is more attractive than the conventional machine learning due to the ability for transfer learning (Tan et al., 2018). Once a DNN model is trained for a specific cell laden solution, this model can be transferred to other solutions by retraining with a much smaller number of samples. Furthermore, the robustness of the DNN means that it is possible to come up with a universal model to predict the cell count by using the material properties of the fluid and the velocity of the droplet.

4. Conclusion

The main goal of this work is to demonstrate the possibility of using droplet velocity at 2 different points along the nozzle-substrate distance to predict the number of printed cells within each ejected droplet during the inkjet-based bioprinting process. This work demonstrates two different applications such as cell detection within single droplets (presence/absence of cells) and prediction of the total number of printed cells within multiple droplets by comparing 5 machine learning algorithms to determine the performance of each model. The first method predicts the presence of a cell in a single droplet with an accuracy of 82% when the velocity of the droplet at 1.0 mm and 2.0 mm are taken as input through random forest regression. This method is suitable for droplets with low cell occupancy but is inaccurate for droplets with high cell occupancy (more than 0.1 cell per droplet). The second method reduced the error in the cell count prediction by a factor of $\sqrt{N}$ through cell count prediction in N droplets instead of single droplet. In the validation set, a mean error of 12% for cell counts in 20 droplets with high cell occupancy (1.29 cell per droplet) is achieved with the extra tree regression model, which is an improvement over the 23% error by predicting the cell count using the average cell per droplet. Cell count prediction with the decision tree-based algorithm performed better than the regression-based algorithm as the regression-based algorithm tends to underfit and generalize all prediction as a similar value. This work also demonstrates that cell concentration of a bio-ink should not be included as the input for the ML model given that bio-ink containing no cells significantly worsens accuracy of the models.

This work provides the basis for a method to control the localized cell concentration within printed droplets by live cell count monitoring. It would be interesting to study the relationship of other parameters such as the rheological properties of the bio-inks, average cell diameter, nozzle properties, and printing parameter for future work. The setup in **Figure 10** can help to reduce the data processing time needed for each set of samples from hours to minutes while improving the sample quality. The massively increased sample count can be fed into a DNN for a more robust model to elucidate their complex interactions.

Author Contribution

Xi Huang: Methodology, Investigation, Writing. **Wei Long Ng**: Conceptualization, Methodology, Investigation, Writing, Review & Editing **Wai Yee Yeong**: Funding acquisition, Review.

Data Availability

The raw data for the training and the validation of the models, and the trained models for validation is available at https://github.com/huangxi90/Cell-prediction-by-velocity-raw-data/.

Acknowledgements

This study is supported under the RIE2020 Industry Alignment Fund – Industry Collaboration Projects (IAF-ICP) Funding Initiative, as well as cash and in-kind contribution from the industry partner, HP Inc., through the HP-NTU Digital Manufacturing Corporate Lab. We would also like to acknowledge and thank the D300e HP team for supplying the C8 celldispensing cassettes for the experiments. Wei Long Ng would like to acknowledge support from NTU Presidential Postdoctoral Fellowship.

Conflict of interest

The authors declare no potential conflicts of interest.

5. Reference

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep Reinforcement Learning: A Brief Survey. *IEEE Signal Processing Magazine*, *34*(6), 26-38. https://doi.org/10.1109/MSP.2017.2743240

Bauer, E., & Kohavi, R. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning*, *36*(1), 105-139. https://doi.org/10.1023/A:1007515423169

Bonatti, A. F., Vozzi, G., Chua, C. K., & De Maria, C. (2022). A Deep Learning Quality Control Loop of the Extrusion-Based Bioprinting Process. *International Journal of Bioprinting*, *8*(4), 620. https://doi.org/10.18063/ijb.v8i4.620

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Caruana, R., & Niculescu-Mizil, A. (2006). An Empirical Comparison of Supervised Learning Algorithms. *Icml '06*, 161–168. https://doi.org/10.1145/1143844.1143865

Chen, X., O'Mahony, A. P., & Barber, T. (2022). The assessment of average cell number inside in-flight 3D printed droplets in microvalve-based bioprinting. *Journal of Applied Physics*, *131*(22), 224701. https://doi.org/10.1063/5.0096468

Chen, Y., Nawaz, A. A., Zhao, Y., Huang, P. H., McCoy, J. P., Levine, S. J., . . . Huang, T. J. (2014). Standing surface acoustic wave (SSAW)-based microfluidic cytometer. *Lab on a Chip*, *14*(5), 916-923. https://doi.org/10.1039/c3lc51139a

Choe, Y. E., & Kim, G. H. (2020). A PCL/cellulose coil-shaped scaffold via a modified electrohydrodynamic jetting process. *Virtual and Physical*

Prototyping, *15*(4), 403-416. https://doi.org/10.1080/17452759.2020.1808269

Fu, Z., Angeline, V., & Sun, W. (2021). Evaluation of printing parameters on 3D extrusion printing of pluronic hydrogels and machine learning guided parameter recommendation. *International Journal of Bioprinting*, *7*(4), 434. https://doi.org/10.18063/ijb.v7i4.434

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3-42. https://doi.org/10.1007/s10994-006-6226-1

Ghahramani, Z. (2004). Unsupervised Learning. In O. Bousquet, U. von Luxburg, & G. Rätsch (Eds.), *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, Tübingen, Germany, August 4 - 16, 2003, Revised Lectures* (pp. 72-112). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-28650-9_5

Gudupati, H., Dey, M., & Ozbolat, I. (2016). A Comprehensive Review on Droplet-based Bioprinting: Past, Present and Future. *Biomaterials*, *102*, 20-42. https://doi.org/10.1016/j.biomaterials.2016.06.012

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised Learning. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (pp. 485-585). Springer New York. https://doi.org/10.1007/978-0-387-84858-7_14

He, J., Mao, M., Li, X., & Chua, C. K. (2021). Bioprinting of 3D Functional Tissue Constructs. *International Journal of Bioprinting*, *7*(3), 1-2. https://doi.org/10.18063/ijb.v7i3.395

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, *4*, 237-285. https://doi.org/10.1613/jair.301

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., . . . Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, *161*(5), 1187-1201. https://doi.org/10.1016/j.cell.2015.04.044

Köster, S., Angilè, F. E., Duan, H., Agresti, J. J., Wintner, A., Schmitz, C., . . . Weitz, D. A. (2008). Drop-based microfluidic devices for encapsulation of single cells [10.1039/B802941E]. *Lab on a Chip*, *8*(7), 1110-1115. https://doi.org/10.1039/B802941E

Law, A. C. C., Wang, R., Chung, J., Kucukdeger, E., Liu, Y., Barron, T., . . . Kong, Z. (2023). Process parameter optimization for reproducible fabrication of layer porosity quality of 3D-printed tissue scaffold. *Journal of Intelligent Manufacturing*, 1-20. https://doi.org/10.1007/s10845-023-02141-0

Lee, J. M., Ng, W. L., & Yeong, W. Y. (2019). Resolution and shape in bioprinting: Strategizing towards complex tissue and organ printing. *Applied Physics Reviews*, *6*(1), 011307, 011301-011316. https://doi.org/10.1063/1.5053909

Levato, R., Jungst, T., Scheuring, R. G., Blunk, T., Groll, J., & Malda, J. (2020). From shape to function: the next step in bioprinting. *Advanced Materials*, *32*(12), 1906423. https://doi.org/10.1002/adma.201906423

Li, W., Mille, L. S., Robledo, J. A., Uribe, T., Huerta, V., & Zhang, Y. S. (2020). Recent Advances in Formulating and Processing Biomaterial Inks for Vat Polymerization-Based 3D Printing. *Advanced Healthcare Materials*, *9*(15), 2000156. https://doi.org/10.1002/adhm.202000156

Li, X., Liu, B., Pei, B., Chen, J., Zhou, D., Peng, J., . . . Xu, T. (2020). Inkjet Bioprinting of Biomaterials. *Chemical Reviews*, *120*(19), 10793-10833. https://doi.org/10.1021/acs.chemrev.0c00008

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., . . . Martersteck, E. M. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, *161*(5), 1202-1214. https://doi.org/10.1016/j.cell.2015.05.002

Mary, P., Chen, A., Chen, I., Abate, A. R., & Weitz, D. A. (2011). On-chip background noise reduction for cell-based assays in droplets [10.1039/C1LC20159J]. *Lab on a Chip*, *11*(12), 2066-2070. https://doi.org/10.1039/C1LC20159J

Meng, Z., He, J., Li, J., Su, Y., & Li, D. (2020). Melt-based, solvent-free additive manufacturing of biodegradable polymeric scaffolds with designer microstructures for tailored mechanical/biological properties and clinical applications. *Virtual and Physical Prototyping*, *15*(4), 417-444. https://doi.org/10.1080/17452759.2020.1808937

Ng, W. L., Ayi, T. C., Liu, Y.-C., Sing, S. L., Yeong, W. Y., & Tan, B.-H. (2021). Fabrication and Characterization of 3D Bioprinted Triple-layered Human Alveolar Lung Models. *International Journal of Bioprinting*, *7*(2), 332. https://doi.org/10.18063/ijb.v7i2.332

Ng, W. L., Chan, A., Ong, Y. S., & Chua, C. K. (2020). Deep learning for fabrication and maturation of 3D bioprinted tissues and organs. *Virtual and Physical Prototyping*, *15*(3), 340-358. https://doi.org/10.1080/17452759.2020.1771741

Ng, W. L., Chua, C. K., & Shen, Y.-F. (2019). Print Me An Organ! Why We Are Not There Yet. *Progress in Polymer Science*, *97*, 101145. https://doi.org/10.1016/j.progpolymsci.2019.101145

Ng, W. L., Goh, M. H., Yeong, W. Y., & Naing, M. W. (2018). Applying Macromolecular Crowding to 3D Bioprinting: Fabrication of 3D Hierarchical Porous Collagen-based Hydrogel Constructs. *Biomaterials Science*, *6*(3), 562-574. https://doi.org/10.1039/C7BM01015J

Ng, W. L., Huang, X., Shkolnikov, V., Goh, G. L., Suntornnond, R., & Yeong, W. Y. (2022). Controlling Droplet Impact Velocity and Droplet Volume: Key Factors to Achieving High Cell Viability in Sub-Nanoliter Droplet-based Bioprinting. *International Journal of Bioprinting*, *8*(1), 424. https://doi.org/10.18063/ijb.v8i1.424

Ng, W. L., Huang, X., Shkolnikov, V., Suntornnond, R., & Yeong, W. Y. (2023). Polyvinylpyrrolidone-based bioink: Influence of bioink properties on printing performance and cell proliferation during inkjet-based bioprinting. *Bio-Design and Manufacturing*. https://doi.org/10.1007/s42242-023-00245-3

Ng, W. L., Lee, J. M., Yeong, W. Y., & Win Naing, M. (2017). Microvalve-based bioprinting – process, bio-inks and applications. *Biomaterials Science*, *5*(4), 632-647. https://doi.org/10.1039/C6BM00861E

Ng, W. L., Lee, J. M., Zhou, M., Chen, Y.-W., Lee, K.-X. A., Yeong, W. Y., & Shen, Y.-F. (2020). Vat polymerization-based bioprinting–process, materials, applications and regulatory challenges. *Biofabrication*, *12*(2), 022001. https://doi.org/https://doi.org/10.1088/1758-5090/ab6034

Ng, W. L., Tan, Z. Q., Yeong, W. Y., & Naing, M. W. (2018). Proof-of-concept: 3D bioprinting of pigmented human skin constructs. *Biofabrication*, *10*(2), 025005, 025001-025013. https://doi.org/10.1088/1758-5090/aa9e1e

Ng, W. L., Wang, S., Yeong, W. Y., & Naing, M. W. (2016). Skin Bioprinting: Impending Reality or Fantasy? . *Trends in Biotechnology*, *34* (9), 689 - 699. https://doi.org/10.1016/j.tibtech.2016.04.006

Ng, W. L., & Yeong, W. Y. (2019). The Future of Skin Toxicology Testing - 3D Bioprinting Meets Microfluidics. *International Journal of Bioprinting*, *5*(2.1), 237. https://doi.org/10.18063/ijb.v5i2.1.237

Ng, W. L., Yeong, W. Y., & Naing, M. W. (2016). Microvalve bioprinting of cellular droplets with high resolution and consistency. *Proceedings of the International Conference on Progress in Additive Manufacturing*, 397-402. https://doi.org/10.3850/2424-8967_V02-236

Ng, W. L., Yeong, W. Y., & Naing, M. W. (2016). Polyelectrolyte gelatin-chitosan hydrogel optimized for 3D bioprinting in skin tissue engineering. *International Journal of Bioprinting*, *2* (1), 53-62. https://doi.org/10.18063/IJB.2016.01.009

Ng, W. L., Yeong, W. Y., & Naing, M. W. (2017). Polyvinylpyrrolidone-Based Bio-Ink Improves Cell Viability and Homogeneity during Drop-On-Demand Printing. *Materials*, *10*(2), 190, 191-112. https://doi.org/10.3390/ma10020190

Ng, W. L., Yeong, W. Y. & Naing, M. W. (2014). Potential of Bioprinted Films for Skin Tissue Engineering. *Proceedings of the 1st International Conference on Progress in Additive Manufacturing*, 441-446. https://doi.org/10.3850/978-981-09-0446-3_065

Nieto, D., Marchal Corrales, J. A., Jorge de Mora, A., & Moroni, L. (2020). Fundamentals of light-cell–polymer interactions in photo-cross-linking based bioprinting. *APL bioengineering*, *4*(4), 041502. https://doi.org/10.1063/5.0022693

Ozbolat, I. T., & Hospodiuk, M. (2016). Current advances and future perspectives in extrusion-based bioprinting. *Biomaterials*, *76*, 321-343. https://doi.org/10.1016/j.biomaterials.2015.10.076

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Piyasena, M. E., & Graves, S. W. (2014). The intersection of flow cytometry with microfluidics and microfabrication. *Lab on a Chip*, *14*(6), 1044-1059. https://doi.org/10.1039/C3LC51152A

Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, *27*(3), 221-234. https://doi.org/10.1016/S0020-7373(87)80053-6

Schoeman, R. M., Kemna, E. W. M., Wolbers, F., & van den Berg, A. (2014). High-throughput deterministic single-cell encapsulation and droplet pairing, fusion, and shrinkage in a single microfluidic device. *Electrophoresis*, *35*(2-3), 385-392. https://doi.org/10.1002/elps.201300179

Sciambi, A., & Abate, A. R. (2015). Accurate microfluidic sorting of droplets at 30 kHz [10.1039/C4LC01194E]. *Lab on a Chip*, *15*(1), 47-51. https://doi.org/10.1039/C4LC01194E

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Sun, W., Starly, B., Daly, A. C., Burdick, J. A., Groll, J., Skeldon, G., . . . Ozbolat, I. T. (2020). The bioprinting roadmap. *Biofabrication*, *12*(2), 022002. https://doi.org/10.1088/1758-5090/ab5158

Suntornnond, R., Ng, W. L., Huang, X., Yeow, E. C. H., & Yeong, W. Y. (2022). Improving printability of hydrogel-based bio-inks for thermal inkjet bioprinting applications via saponification and heat treatment process. *Journal of Materials Chemistry B*, *10*, 5989-6000. https://doi.org/10.1039/D2TB00442A

Takagi, D., Lin, W., Matsumoto, T., Yaginuma, H., Hemmi, N., Hatada, S., & Seo, M. (2019). High-precision three-dimensional inkjet technology for live cell bioprinting. *International Journal of Bioprinting*, *5*(2), 208. https://doi.org/10.18063/ijb.v5i2.208

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. *Artificial Neual Networks and Machine Learning - ICANN 2018*, 270-279. https://doi.org/10.1007/978-3-030-01424-7_27

Tang, W., Tang, D., Ni, Z., Xiang, N., & Yi, H. (2017). Microfluidic Impedance Cytometer with Inertial Focusing and Liquid Electrodes for High-Throughput Cell Counting and Discrimination. *Analytical Chemistry*, *89*(5), 3154-3161. https://doi.org/10.1021/acs.analchem.6b04959

Wiering, M. A., & Van Otterlo, M. (2012). Reinforcement learning. *Adaptation, learning, and optimization*, *12*(3), 729. https://doi.org/10.1007/978-3-642-27645-3

Xames, M. D., Torsha, F. K., & Sarwar, F. (2022). A systematic literature review on recent trends of machine learning applications in additive manufacturing. *Journal of Intelligent Manufacturing*, 1-27. https://doi.org/10.1007/s10845-022-01957-6

Xi, H. (2021). Synthesis and Characterization of Crystalline Embedded with Metallic Silicide Nanoparticles. https://scholarbank.nus.edu.sg/handle/10635/199993

Xu, H., Liu, Q., Casillas, J., McAnally, M., Mubtasim, N., Gollahon, L. S., . . . Xu, C. (2022). Prediction of cell viability in dynamic optical projection stereolithography-based bioprinting using machine learning. *Journal of Intelligent Manufacturing*, *33*(4), 995-1005. https://doi.org/10.1007/s10845-020-01708-5

Xu, T., Zhao, W., Zhu, J. M., Albanna, M. Z., Yoo, J. J., & Atala, A. (2013). Complex heterogeneous tissue constructs containing multiple cell types prepared by inkjet printing technology. *Biomaterials*, *34*(1), 130-139. https://doi.org/10.1016/j.biomaterials.2012.09.035

Yang, Y., Xu, R., Wang, C., Guo, Y., Sun, W., & Ouyang, L. (2022). Recombinant human collagen-based bioinks for the 3D bioprinting of full-thickness human skin equivalent. *International Journal of Bioprinting*, *8*(4), 611. https://doi.org/10.18063/ijb.v8i4.611

Zhuang, P., Ng, W. L., An, J., Chua, C. K., & Tan, L. P. (2019). Layer-by-layer ultraviolet assisted extrusion-based (UAE) bioprinting of hydrogel constructs with high aspect ratio for soft tissue engineering applications. *PLoS One*, *14*(6), e0216776, 0216771-0216721. https://doi.org/10.1371/journal.pone.0216776