

# Learning Deep Features for Stain-free Live-dead Human Breast Cancer Cell Classification

**Gisela Pattarone**

Universidad de Buenos Aires, Facultad de Farmacia y Bioquímica, Buenos Aires

**Laura Acion**

Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

**Marina Simian**

Instituto de Nanosistemas, Universidad Nacional de San Martín

**Emmanuel Iarussi** (✉ [emmanueliarussi@gmail.com](mailto:emmanueliarussi@gmail.com))

National Technological University

---

## Research Article

**Keywords:** Automated cell classification, cancer biology, classifiers clusterize bright-field images

**Posted Date:** November 25th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-109542/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Learning deep features for stain-free live-dead human breast cancer cell classification

Gisela Pattarone<sup>1,2</sup>, Laura Acion<sup>3,6</sup>, Marina Simian<sup>4,6</sup>, and Emmanuel Iarussi<sup>5,6,\*</sup>

<sup>1</sup>Universidad de Buenos Aires, Facultad de Farmacia y Bioquímica, Buenos Aires, Argentina

<sup>2</sup>Albert Ludwigs University of Freiburg, Faculty of Medicine, Freiburg, Germany

<sup>3</sup>Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina

<sup>4</sup>Instituto de Nanosistemas, Universidad Nacional de San Martín, Argentina

<sup>5</sup>Universidad Tecnológica Nacional, Buenos Aires, Argentina

<sup>6</sup>Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

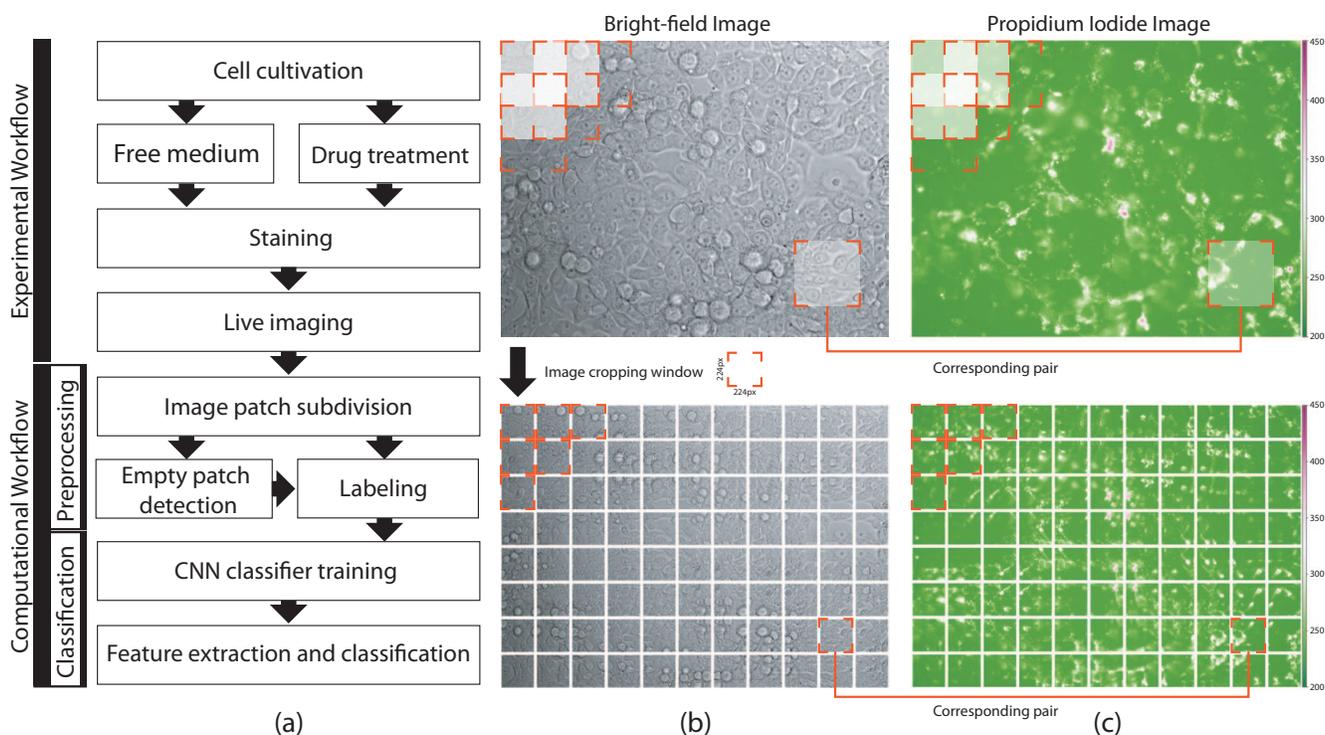
\*eiarussi@conicet.gov.ar

## ABSTRACT

Automated cell classification in cancer biology is a challenging topic in computer vision and machine learning research. Breast cancer is the most common malignancy in women that usually involves phenotypically diverse populations of breast cancer cells and an heterogeneous stroma. In recent years, automated microscopy technologies are allowing the study of live cells over extended periods of time, simplifying the task of compiling large image databases. For instance, there have been several studies oriented towards building machine learning systems capable of automatically classifying images of different cell types (i.e. motor neurons, stem cells). In this work we were interested in classifying breast cancer cells as live or dead, based on a set of automatically retrieved morphological characteristics using image processing techniques. Our hypothesis is that live-dead classification can be performed without any staining and using only bright-field images as input. To our knowledge, there is no previous work attempting this task on in vitro studies of breast cancer cells, nor is there a dataset available to explore solutions related to this issue. We tackled this problem using the JIMT-1 breast cancer cell line that grows as an adherent monolayer. First, a vast image set composed by JIMT-1 human breast cancer cells that had been exposed to a chemotherapeutic drug treatment (doxorubicin and paclitaxel) or vehicle control was compiled. Next, several state-of-the-art classifiers were trained based on convolutional neural networks (CNN) to perform supervised classification using labels obtained from fluorescence microscopy images associated with each bright-field image. Model performances were evaluated and compared on a large number of bright-field images. The best model reached an AUC = 0.941 for classifying breast cancer cells without treatment. Furthermore, it reached AUC = 0.978 when classifying breast cancer cells under drug treatment. Our results highlight the potential of machine learning and computational image analysis to build new diagnosis tools that benefit the biomedical field by reducing cost, time, and stimulating work reproducibility. More importantly, we analyzed the way our classifiers clusterize bright-field images in the learned high-dimensional embedding and linked these groups to salient visual characteristics in live-dead cell biology observed by trained experts.

## Introduction

Breast cancer is the most frequently diagnosed malignancy in women worldwide; one out of eight women are expected to develop breast cancer at some point in their lifetime<sup>1</sup>. As a disease, it involves biologically diverse subtypes with high intratumor heterogeneity that determine different pathological characteristics and have different clinical implications. Understanding the intricacy of the molecular cross-talk within the cell death pathway highlights the need for developing methods to characterize the morphological cell response to therapy with anticancer drugs. The emergence of automatic microscopes made it possible to develop large datasets of live fluorescence images and single cell analysis, and more recently, these data started to be massively studied by means of computational tools. Some efforts are focused on developing image processing programs able to identify cells and separate them from the extracellular matrix, performing segmentation and tracking cells using contrast fluorescence<sup>2</sup>. More recent efforts are based on automatic classification of images using deep learning techniques<sup>3,4</sup>, a form of automatic learning<sup>5,6</sup> enabling improved data analysis for high-throughput microscopy<sup>7,8</sup>. For example, deep convolutional neural networks have been trained<sup>9</sup> with labeled images from different cell types like motor neurons, stem cells, and Jurkat cells<sup>10</sup>. In order to label each cell, Hoechst and DAPI have been used to identify nuclear areas, CellMask to highlight plasma membranes and Propidium Iodide to spot cells with compromised membranes. These automatic methods were able to make accurate pixel predictions of the location and intensity of the different structures represented by the fluorescence. More recently, machine learning classifiers were trained to perform stain-free hierarchical classification of human white blood cells in flow

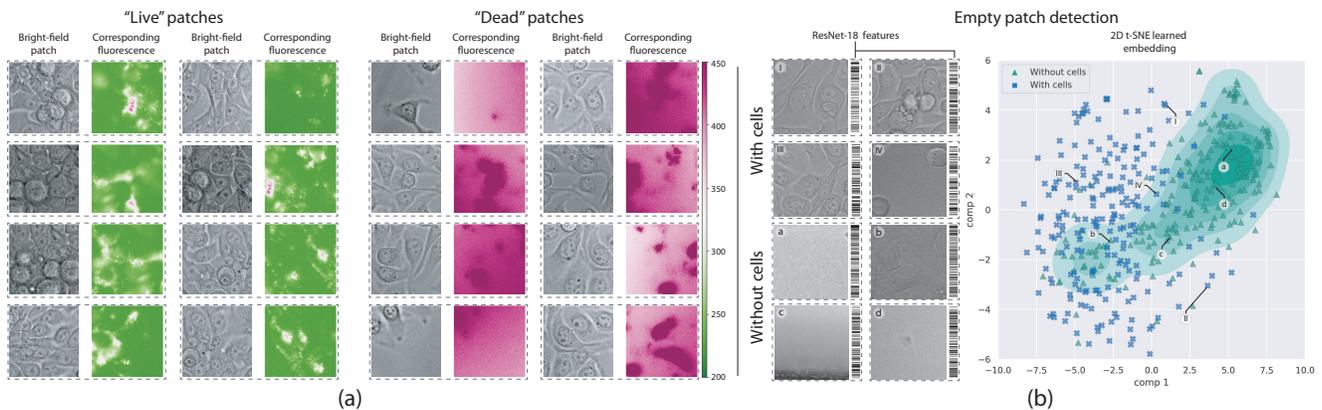


**Figure 1.** (a) Experimental and computational steps in our automated cell classification pipeline. (b-c) Top: bright-field and corresponding fluorescence images resulting from the imaging step (experimental workflow). High fluorescence values (white and red areas) indicate cell death. Bottom: as a post-process, images are cropped into 224x224px patches and paired with their corresponding fluorescence patch. Notice cropping overlaps contiguous patches (horizontal and vertical) in order to augment the number of images

cytometry images<sup>11</sup>. Similar methods have been used to distinguish dead from living microalgae *Chlorella vulgaris* with features extracted from individual cells<sup>12</sup>. In both cases, the acquisition technique isolates cells, simplifying segmentation and labeling tasks in the image preprocessing step. In the context of cancer cell growth, this type of isolation is difficult to achieve, making it necessary to use techniques which can aggregate image information and automatically extract features for classification. Other innovative biological applications related to automated image processing methods are morphological classification of hematopoietic cells, pluripotent stem cells<sup>13</sup> and 3D cell boundary and nuclear segmentation<sup>14</sup>.

Empowered by recent advances in image processing and deep learning, in this work we were interested in the study of morphological characteristics showing death signs in breast cancer cells. Particularly, in the context of live cell fluorescence, the live-dead labeling method has many limitations like low contrast or differences in pixel intensities, resulting in heterogeneous staining for individual cells and requiring a final human-assisted cell segmentation. Additionally, fluorescent stains are expensive and usually several stains are required to precisely identify a cell<sup>11</sup>. Fluorescence-free cell classification could potentially offer substantial improvements in detection specificity, sensitivity, and accuracy for physiological and pathological cell condition diagnosis. Furthermore, the cells could remain in their regular culture conditions without any intervention. Our purpose is to evaluate the potential of automatically classifying cancer cells as live or dead without staining, using only bright-field images as input.

First, we present a new massive dataset of breast cancer cell images of the JIMT-1 breast cancer cell line<sup>15</sup>. We studied cellular growth before and after the introduction of in vitro drugs treatments with Doxorubicin and Paclitaxel. After characterizing the biological behavior within chambered coverslips, each image was split into smaller patches containing a very limited amount of cells and properly tagged as live or dead using the information available in the form of fluorescence images from calcein and propidium iodide. To our knowledge, this is the biggest dataset of labeled JIMT-1 cell images ever compiled. We then used this dataset to train deep CNN models for cell image classification. These trained classifiers learned to label cancer cells as live or dead without staining and using only bright-field images as input. A diagram of the presented workflow is shown in Fig. 1 (a). We additionally studied the learned embeddings and identified clusters of images with similar visual cues which are often associated with living and dead cells. We believe our results could be helpful as a diagnostic and



**Figure 2.** (a) Bright-field and corresponding propidium iodide fluorescence images. The columns under Live patches (green) show images with mostly live cells. The columns under Dead patches (red) present images with mostly dead cells in our dataset. (b) Left, samples of empty and non-empty patches with the associated barcode visualization of the 512-dimensional feature vector from ResNet-18 last convolutional output per image. Right, dimensionality reduction visualization of patch features and the high-dimensional decision function (green level-sets) learned by the SVM. Notice how empty patches mostly lie inside the highlighted region.

complementary tool for cancer and normal cell biology, allowing a better understanding of the capabilities of image-based automatic classification. Furthermore, we foresee potential applications in the pharmaceutical field, as automatic live/dead cell classification in preclinical trials for drug tests is of high interest, complementing the information related to pharmacokinetics and pharmacodynamics characteristics of new anti-cancer drugs development.

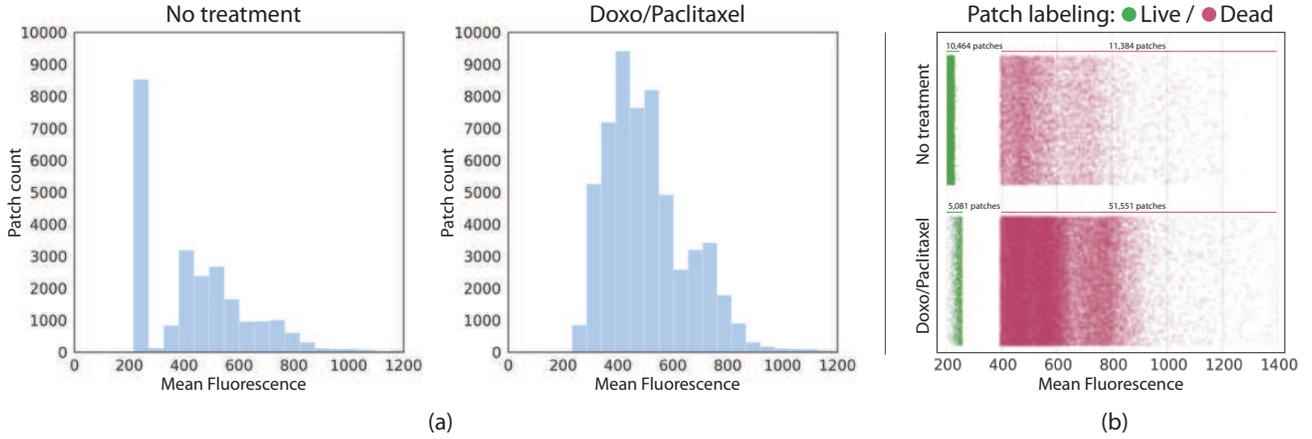
## Results

### Cell preparation and image acquisition

To ensure a biologically representative set of breast cancer cell images in our dataset, we first analyzed and characterized the development of JIMT-1 within the Ibidi chamber slides. JIMT-1 cells are positive for cytokeratins 5/14 and 8/18, are estrogen and progesterone receptor negative, and overexpress HER2 as a consequence of HER2 amplification. JIMT-1 cells are classified as basal-like and represent the subgroup that occasionally carry HER2 amplifications. JIMT-1 cells act like a triple negative subtype breast cancer given their lack of response to trastuzumab<sup>15</sup>. To induce JIMT-1 cell death we designed a treatment scheme consisting of a 4 hr exposure to doxorubicin followed by 24 hr of paclitaxel. In order to capture the images, we performed live fluorescence imaging using a live-dead cell imaging kit of cells cultured in chambered coverslips with 8 independent wells and a non-removable polymer coverslip-bottom, over extended time periods. This setup has high optical quality, with a total growth area per well of 1.0 cm<sup>2</sup>, tolerates live fluorescence, and allows the tracking of breast cancer cells during a maximum of five days. We constructed a biologically representative dataset of breast cancer cells grown in culture medium supplemented with the sequential treatment of doxorubicin followed by paclitaxel. After cultivation and drug treatment, we measured the effect of the therapeutic agents on the percentage area shown by calcein and propidium fluorescence. Both were studied in comparison with a control sample. The area of activity of the PI fluorescence was higher in comparison to control. Simultaneously, the calcein percentage area was lower at the end of the treatment. Both facts combined showed that the treatment with drugs was effective in inducing cell death and ensured that our image dataset contained both cells states, live and dead. We compiled 964 raw images into a dataset we named Doxo/Paclitaxel. We additionally collected 339 raw images from the cell growth and death process occurring spontaneously (without therapy) during the same time period and named it No treatment. In both datasets, each bright-field image has a corresponding fluorescence PI image indicating cell death (Fig. 1 (b-c), top).

### Image pre-processing

We curated the raw images to be suitable for training automatic classifiers. We identified several problems with the raw images that we solved individually in order to prepare the final image set. The first issue relates closely to the image size. Raw images cover large areas of the Ibidi device with a resolution of 1344x1024 pixels, and often the associated PI fluorescence strongly varies across it. This represents a problem in our setup, since a single label indicating live or dead must be assigned to each image to train the classifier. Therefore, we decided to partition raw images into smaller patches (Fig. 1 (b-c), bottom). By



**Figure 3.** (a) Histograms of mean fluorescence per patch on each dataset. (b) Scatter diagram showing the samples distribution after the automatic patch labeling based on their mean fluorescence. Due to treatment, Doxo/Paclitaxel has significantly less live patches.

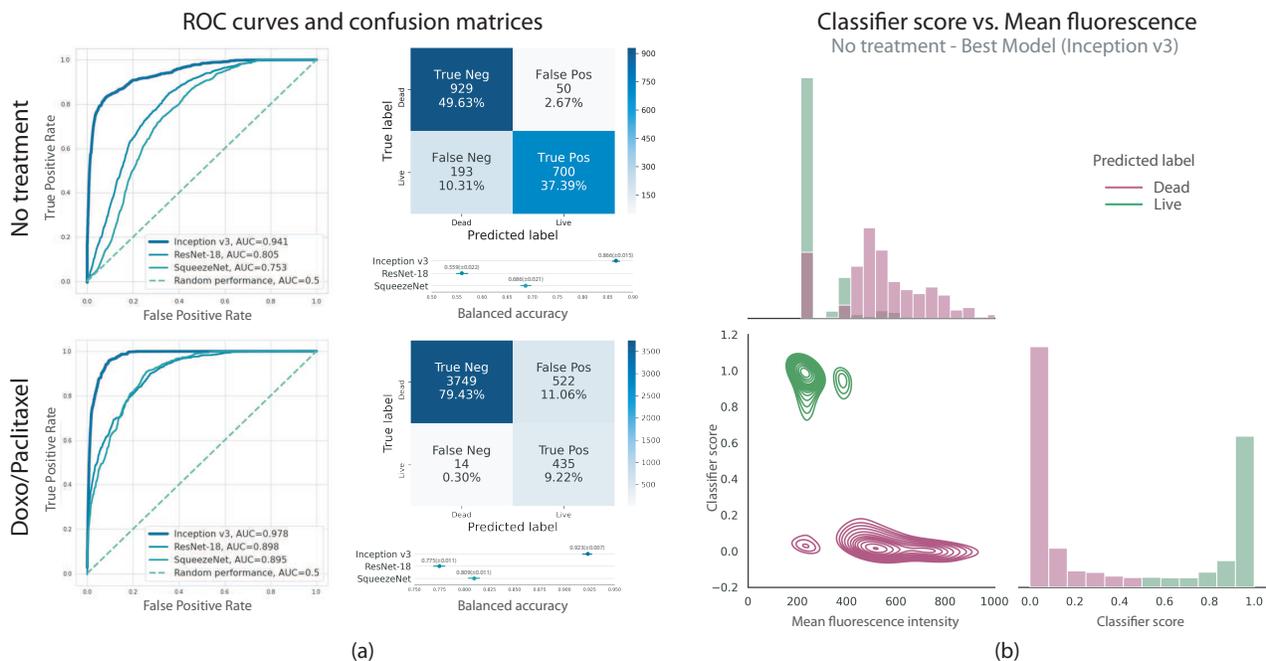
cropping smaller areas, we increased the reliability of the labels for each patch, since neighboring cells often have the same state. However, setting a proper granularity for this operation is not trivial. On the one hand, individually labeling each cell could lead to very accurate labels, but the topology of cell growth in the device makes it difficult to automatically isolate cells. On the other hand, cropping large areas could lead to overlapping labels, with interfering residual fluorescence from neighboring patches. Despite the fact that PI has the characteristic of only entering the cell when its membrane is compromised, we noticed the fluorescence spectrum emissions are not uniform and may overlap or even occupy more than one cell diameter. We therefore found a compromise between these two options by using a fixed size sliding cropping window. Conveniently, we cropped 224-pixel wide square patches, a standard size that facilitates the use of state-of-the-art CNN architectures (see Classifiers Training). In our datasets, each bright-field cropped patch has a corresponding cropped fluorescence image (Fig. 2 (a)).

After cropping, we noticed many image patches did not capture any cells. This is especially common in data coming from the first culture days, where a uniform distribution is not yet achieved. When training automatic classifiers, empty images can decrease network performance because no real feature extraction process occurs without cells in the image. We therefore implemented a mechanism to easily detect and discard empty patches. First, we manually labeled a subset of 226 bright-field patches that didn't contain any cells or unsuitable data, such as out of focus images, and 226 patches containing properly captured cells (Fig. 2 (b), left). For each of these images, we computed a 512-feature vector by taking the output of the last convolutional layer from a pretrained ResNet-18 on ImageNet<sup>16</sup>. The dataset contained patches from both: No treatment and Doxo/Paclitaxel data partitions. We then trained a support vector machine (SVM)<sup>17</sup> to perform outlier detection. The trained model learned to detect most of the empty images (f1-score = 0.833). Fig. 2 (b) presents a 2D t-distributed stochastic neighbor embedding (t-SNE)<sup>18</sup> visualization of the learned high-dimensional decision function when classifying image patches as with or without cells. After cropping and filtering empty patches, the No treatment set contains 21,848 images and the Doxo/Paclitaxel set contains 56,632 images.

Once most empty images were removed from the datasets, we prepared them for supervised training, that requires a single binary label indicating whether the image represents live or dead cells. We therefore averaged the fluorescence values and set up a threshold splitting the image in two non-overlapping sets: a set labelled as containing live cells and another containing dead cells. We found the threshold for defining each set by computing histograms of the mean fluorescence intensities for the No treatment and Doxo/Paclitaxel datasets (Fig. 3 (a)). Choosing a very high threshold (high fluorescence values) would assure more certainty for image patches labeled as dead, but it would end up labeling as alive many images that are far from the low fluorescence values indicating live signs. Conversely, the opposite effect would be observed if choosing a very low threshold. We solved this issue by fitting a Gaussian mixture model<sup>19</sup> to the No treatment distribution of means which contained the most balanced number of live and dead cells. We identified two main clusters in data: one with very low fluorescence values containing mostly live cells ( $\bar{x}_{live} = 224.51$ ,  $s_{live} = 34.46$ ), and another one with high fluorescence values containing mostly dead cells ( $\bar{x}_{dead} = 550.44$ ,  $s_{dead} = 153.55$ ). We used this model to label as live all images with mean fluorescence lower than  $\bar{x}_{live} + s_{live} = 258.97$ , and as dead all the images with mean fluorescence above  $\bar{x}_{dead} - s_{dead} = 396.89$ . Patches in the range  $(\bar{x}_{live} + s_{live}, \bar{x}_{dead} - s_{dead})$  were discarded. We applied the same threshold to both, No treatment and Doxo/Paclitaxel datasets (Fig. 3 (b)). Only very low fluorescence values are considered as containing

**Table 1.** Dataset summary. The first row corresponds to raw, full resolution images (1344x1024 pixels). The number of cropped images are reported after empty patch filtering. The classification classes are unbalanced, particularly for the Doxo/Paclitaxel data.

	Raw images	Cropped images	Total Live	Total Dead	Train Live	Train Dead	Train total	Valid. Live	Valid. Dead	Valid. total	Test Live	Test Dead	Test total
No treatment	339	21,848	10,464	11,384	8,680	9,480	18,160	891	925	1,816	979	893	1,872
Doxo/Paclitaxel	964	56,632	5,081	51,551	4,195	42,966	47,161	437	4,314	4,751	449	4,271	4,720



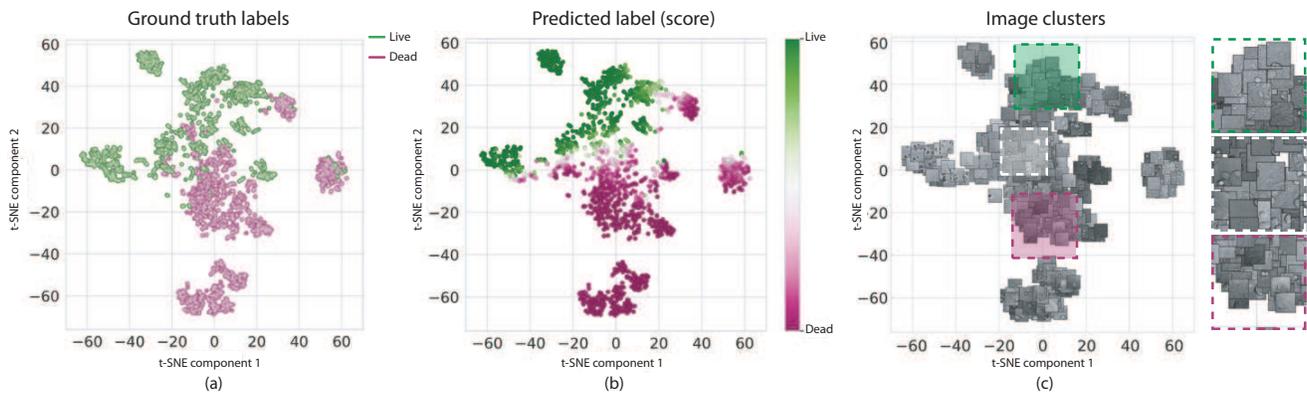
**Figure 4.** (a) ROC curves showing the classification performance over the testing datasets for each CNN architecture. The Inception-v3 model outperforms ResNET and SqueezeNET. (b) Mean fluorescence vs. classifier-score analysis for the model that performed best over No treatment data (Inception-v3). Higher mean fluorescence intensities tend to cluster together for lower classification scores used to label dead cells. Simultaneously, lower mean fluorescence intensities are grouped near higher classification scores that signal live cells.

live cells. Table 1 summarizes the number of images included in each pre-processing step and available in this repository: <https://github.com/emmanueliarussi/live-dead-JIMT-1>.

### Classifiers Training

We trained three different state-of-the-art CNN architectures to perform binary live-dead classification using the curated cell image dataset: ResNET-18<sup>20</sup>, SqueezeNET<sup>21</sup>, and Inception-v3<sup>22</sup>. Each network architecture was trained twice using a cross entropy loss function and the No treatment and Doxo/Paclitaxel dataset partitions. Three splits for each dataset were constructed to allow training and subsequent evaluation tasks (Table 1). Approximately 80% of the images were used for training, 10% for validation and 10% for testing, as suggested in the literature<sup>23</sup>. Since each cropped image patch was tagged with an identifier corresponding to the ID of the raw image from which it came, we were able to avoid patches from the same raw image to belong to more than one partition simultaneously. In other words, there are no overlapping images among training, validation, and testing partitions since we carefully selected patches from different raw images for each set.

A common problem when training classifiers is their sensitivity to class imbalance<sup>24</sup>. Therefore, to compensate for the strong imbalance in our dataset, we sampled images by means of a weighted random sampler with replacement. Weights were computed as the inverse of the sample count for each class. Additionally, data were augmented by random 90-degree rotations and vertical/horizontal flipping of each image. This type of data augmentation leads to better generalization performance<sup>11</sup>. We empirically found that fine-tuning network weights pre trained on Image-NET<sup>25</sup> performed significantly better than training from a random initialization. Therefore, we adopted a transfer-learning approach for all the reported results. Training hyperparameters were adjusted based on the network performances over the validation set. More training details can be found in the Methods section.



**Figure 5.** Visualization of the Inception-v3 learned feature space for our No treatment testing dataset. The 2048-dimensional features were projected to a 2D space using t-SNE, and colored according to ground truth labels (a), and predicted labels (b). Cells with the same state tend to cluster together. Visual inspection of the images in each cluster further reveals the shared characteristics within each group.

After training, each model was validated using non-augmented instances from the validation set. In order to evaluate and compare the performances of the trained classifiers, we relied on several metrics. In particular, we computed the balanced accuracy, which is defined as the average of the recall obtained on each class<sup>26,27</sup>. This metric is well-suited for our setup, since it does not favor a classifier that exploits class imbalance by biasing toward the majority class. Together with the balanced accuracy, we computed confusion matrices and pairwise relationships between mean fluorescence and the classifier score. Fig. 4 (a) summarizes the performance of the trained classifiers. Overall, the three models outperformed random performance for both datasets and were able to automatically extract relevant image features in order to classify JIMT-1 cell images as living or dead. Inception-v3 was the best performing model, with over 85% accuracy over both testing datasets, No treatment balanced accuracy = 0.866 (95% CI=[0.851, 0.881]), AUC = 0.941; Doxo/Paclitaxel balanced accuracy = 0.923 (95% CI=[0.916, 0.930]), AUC = 0.978. Confusion matrices and ROC curves in Fig. 4 further illustrate the classifiers' performance. Furthermore, we computed the correlations between the mean values of PI and the classification score obtained for each image in the testing set to explore the association between classification and fluorescence images. A significant inverse Pearson correlation was found in both training scenarios, No treatment:  $r = -0.705$  ( $p = 0.024$ ) and Doxo/Paclitaxel:  $r = -0.281$  ( $p = 0.025$ ), indicating the scores are correlated to the fluorescence levels, a relationship that could be explored in future work in order to predict fluorescence images from bright-fields (Fig. 4 (b)).

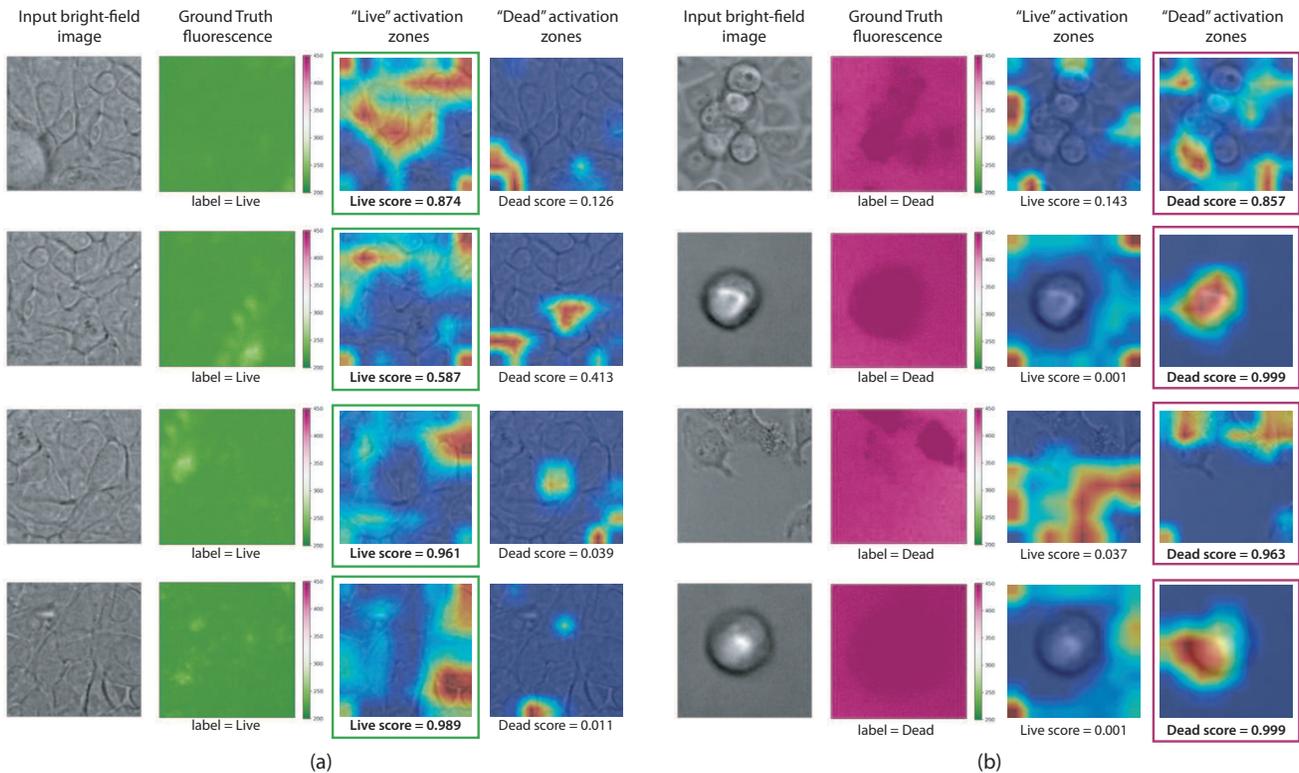
### Visualizing Learned Features

In line with previous work<sup>10,11</sup>, we took advantage of well-known visualization techniques in order to gain further insight into the classifiers' automatically learned space to uncover their biological meaning. In particular, we applied a nonlinear dimensionality reduction technique suited for embedding high-dimensional data into a low-dimensional space, namely t-SNE<sup>18</sup>, which preserves local structures of the high-dimensional input space.

The learned features of a CNN are encoded by the intermediate activation after the last convolutional layer of the network. Therefore, given an input image which is fed to the CNN to perform classification, we extract the activation pattern of the last layer before classification. This high-dimensional vector becomes a signature of the input image. Scatter plots in Fig. 5 (a-b) illustrate the emerging clusters after projecting the 2048-dimensional features of Inception-v3 into two components for all testing samples. To further understand each cluster, we also show a version of the scatter plot where each dot is replaced by the corresponding bright-field image thumbnail (Fig. 5 (c)). This enhanced visualization reveals that groups of cell images with similar visual characteristics tend to cluster together in the learned feature space. This visualization of the feature space learned by the classifiers also provided a visual validation of the classification confusion occurring between live and dead cells. We found that the boundary between main live and dead clusters (white dots in Fig. 5 (b)) correspond to images in which a mixture of live and dead cells appear.

We further investigated the relation between input bright-field images and the produced outcomes by means of the gradient-weighted class activation mapping (Grad-CAM)<sup>28</sup>. This visualization technique uses the class-specific gradient information flowing into the final convolutional layer of a CNN to produce a coarse localization map of the important regions in the image which triggered the classifier output. These regions can be visualized by means of a heatmap (Fig. 6).

Overall, in the case of living and untreated cells, morphology looks as expected, with the presence of an uncompromised



**Figure 6.** Bright-field testing patches paired with gradient-weighted class activation mapping (Grad-CAM) visualizations for the Inception-v3 model. Live patches are shown on the left (a), and dead on the right (b). The activation maps show which zones of the input bright-field are triggering the classifier response. These maps can be computed for both labels and help to identify zones in the input images activating a live or a dead response from the convolutional neural network. The maps also have an associated score, indicating the probability of each label, which determines the final classifier response.

membrane, organelles, nuclei, and nucleolus (Fig. 6 (a)). The membrane can be often seen clearly without any special enhancement (green fluorescence rows in Fig. 6). This integrity of the cell membrane is necessary to keep the position of its organelles, mainly rough endoplasmic reticulum and golgi apparatus. Cells in this group have a mostly uniform gray color, scattered by very tiny dark circles, possibly corresponding to the cell nuclei. These are expected morphological characteristics of a cell that remains active and where its chromatin remains partially in the form of a nucleolus and that is decomposed and used according to the needs of the biological machinery. Biological aspects of dead cells are different. It can be seen in patches containing stained and classified as dead cells (red fluorescence rows in Fig. 6 (b)), that the compromised membrane appears more as a blurred dark halo. This is expected since the PI staining enters the cell only when the cell membrane has been compromised and binds to DNA by intercalating between the bases with little or no sequence preference. It can be generally observed that the harmonic disposition evidenced as a smooth gray of the organelles is lost, probably due to a contraction of the cytoplasm that occurs in the processes of cell death. The cell death process leads to a series of intracellular events, regulated, and coordinated by the activation of different enzymes that perform proteolysis cascades and controlled destruction of organelles and genetic material. The final phase of this process is evident inside the cells classified as dead. The circular genetic material known as nucleolus is not evident, but rather there is a deletion of it as can be clearly noted in cells identified as dead. On the contrary, cells identified as live maintain the central dark gray nucleoli. Differences in cell death images in the two groups, No treatment and Doxo/Paclitaxel datasets, can be seen in the process of contraction of the cytoplasm and DNA degradation. The pharmacological effect of Doxorubicin on the cancer cells is induced by intercalation into DNA and disruption of topoisomerase-II-mediated DNA repair and generation of free radicals that damage cellular membranes, DNA, and proteins<sup>29</sup>. This is supplemented by the effect of Paclitaxel on tubulin that polymerizes into small tubes called microtubules, which are responsible for mitosis, cell movements, preservation of cell shape, as well as the intracellular trafficking of organelles and macromolecules. Paclitaxel stabilizes microtubules and reduces their dynamicity, promoting mitotic halt and cell death<sup>30</sup>. Both pharmacological effects can be visualized in the cytoplasm that present a kind of effacement and bright spot in the brightfield image, without evidence of destruction of organelles and genetic material.

## Discussion

All evaluated state-of-the-art network architectures were able to autonomously extract relevant information from bright-field imagery in order to perform live-dead classification. This automatic feature extraction can be improved in future work, by combining it with cell characteristics i.e. cell diameter, area, and radius, similar to the work of Reimann et al.<sup>12</sup>. The mixture of learned and engineered features can improve performance as well as interpretability of the classifier behaviour. In order to push further in this hybrid direction there is a need for more robust methods able to individualize and segment cells growing as an adherent monolayer. At the beginning of this project, we explored the alternative of segmenting and labeling each cell individually before classification, but the extremely irregular cellular contours and the occasional overlap among them made this approach inapplicable. We believe the work of Lugagne et al.<sup>31</sup> highlights the next steps to overcome these issues.

The curated image data was of paramount importance for the achieved performances of the classifiers. In general, the lack of large image datasets greatly hampers the applicability of deep learning techniques. Even if our dataset was big enough to learn and generalize to unseen samples, we believe a larger effort in building bigger and more diverse datasets is still necessary. For example, all our images come from a single capture device, which could limit the applicability of the trained models to images from a different acquisition setup. We also worked on a single cell line and stain. More data will definitely contribute to make these tools widely available across the scientific and medical community. Future work should consider compiling images in a variety of capture scenarios.

Automatic cell classification is a very challenging and interdisciplinary problem, involving simultaneous efforts from computer vision, machine learning and biomedical research. In the context of human breast cancer, machine learning can bring new tools to support diagnosis that benefit the biomedical field by reducing cost and time. In this work we investigated the applicability of deep learning techniques to stain-free live-dead breast cancer cell classification from bright-field images. Since our aim was that others may reuse our findings and data, we used open-source Python packages and we made freely available our image dataset online.

## Methods

### Experimental methods

*Cell culture.* JIMT-1 cells ATCC 589 (DSMZ) were cultured in complete DMEM medium (Gibco), supplement with fetal calf serum heat-inactivated (FBS) 10% (w/v) (Gibco), L-glutamine 2 mmol L<sup>-1</sup> (Gibco), penicillin 100 units mL<sup>-1</sup>, streptomycin 100 g mL<sup>-1</sup> (Gibco) at 37°C in an incubator with 5% CO<sub>2</sub>. Cells were resuspended with trypsin 0.50 mg mL<sup>-1</sup> and EDTA-4Na 0.2 mgmL<sup>-1</sup> (Gibco), and incubated at 37°C for 3 min. Trypsin was inactivated with FBS and cells were washed with phosphate buffer solution (PBS) (NaH<sub>2</sub>PO<sub>4</sub> 50 mmol L<sup>-1</sup>, NaCl 300 mmol L<sup>-1</sup>, (pH = 7.6) and centrifuged at 1200 rpm for 5 min. Finally, the cells were resuspended in the same complete DMEM medium. We use the 8-well slide (Ibidi GmbH) and 12000 cells per well were used to perform culture assays according to the manufacturer's protocol.

*Cell viability staining.* We used the live-dead cell imaging kit (Sigma) to evaluate cell viability in the Ibidi chip. The cells were loaded into the Ibidi devices and cell viability was evaluated at third, fourth, and fifth days; we PBS to wash the culture chambers in the models for 1–3 min. Then, the cells were incubated with the live-dead cell imaging kit for 15–30 min at 37°C. Next, we used PBS again to wash out the reagent for 3–5 min and observed the culture chambers under a fluorescent microscope.

*Autophagy and Apoptosis activity staining.* We used the autophagy cell imaging kit (CYTO-ID) and caspase-3 and-7 cell imaging kit (Invitrogen). In both assays performed separately, the cells are stained green. The procedure with negative and positive controls were performed as recommended by the manufacturers' instructions (Enzo ENZ-51031-K200) 32.

*Doxorubicin and Paclitaxel schematic treatment.* For the drug schematic tests, the effects of paclitaxel (Sigma Aldrich) and doxorubicin (Sigma Aldrich) combined were studied (Holmes, F., 1996). First, JIMT-1, were loaded into the Ibidi chips, as described previously, and 24 h later when the cells were adherent, the medium was replaced with fresh culture medium supplemented with 0.01 M doxorubicin (DOX). Then, after 4 hours it was subsequently replaced with a fresh medium containing 0.001 M paclitaxel (PAX) for 24 hours. Live imaging and biological characterization with different staining as described before was performed for the whole experiment.

*Microscopy.* Cell images were captured using the Olympus ScanR microscope. The images collected for the dataset were taken in each biological step related to cellular growth and the use of different chemotherapeutic agents and drug schemes. A 20x magnification was used, according to this each image has the dimension of 433x330 m, with a conversion factor 0.32250 m/pixel, and a final pixel per image 16 bit of 1346x1024 pixels. Each brightfield image taken by the microscope was triplicated in the same position by different filters chosen to show the biological structure labeled with the correspondent fluorescence.

For the H $\ddot{o}$ chst filter we used an excitation filter of 377/50 with an emission filter of 437–475 nm, for the propidium iodide filter we used an excitation filter of 575/25 with an emission filter of 600–662 nm, and for autophagy and caspase we used an excitation filter of 494/20 with an emission filter of 510–552 nm.

## Computational methods

*Dataset construction.* We converted the raw 16 bit microscope images to an unsigned 8 bit type (both bright-field and fluorescence images). Pre-computations were implemented in Python using OpenCV (Open Source Computer Vision Library) framework, an open source computer vision and machine learning software library.

*Neural Networks.* The network architectures and training were implemented in Python using PyTorch Framework<sup>32</sup> and the aforementioned pre-trained models. We used the same hyperparameters for all network architectures and training scenarios: learning rate lr = 1e-5, batch size bz = 4, epochs e = 30. We optimized our objective function by means of the Adam, a state of the art adaptive learning rate optimizer implemented in PyTorch (b0 = 0.5; b0 = 0.999), with weight decay wd = 1e-5.

*Equipment.* A notebook was used for the creation of the dataset. Training of the CNN was performed on an Intel Xeon server equipped with two Graphics Processing Unit (GPU) Nvidia Titan Xp and 32Gb of RAM.

## Data Availability

The image dataset and further resources are available in the public github repository:

<https://github.com/emmanueliarussi/live-dead-JIMT-1>

## References

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *Int. journal cancer* **136**, E359–E386 (2015).
2. Deter, H. S., Dies, M., Cameron, C. C., Butzin, N. C. & Buceta, J. A cell segmentation/tracking tool based on machine learning. In *Computer Optimized Microscopy*, 399–422 (Springer, 2019).
3. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436–444 (2015).
4. Gupta, A. *et al.* Deep learning in image cytometry: a review. *Cytom. Part A* **95**, 366–380 (2019).
5. Asri, H., Mousannif, H., Al Moatassime, H. & Noel, T. Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Comput. Sci.* **83**, 1064–1069 (2016).
6. Moen, E. *et al.* Deep learning for cellular image analysis. *Nat. methods* 1–14 (2019).
7. Blasi, T. *et al.* Label-free cell cycle analysis for high-throughput imaging flow cytometry nat (2016).
8. Dao, D. *et al.* Cellprofiler analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinformatics* **32**, 3210–3212 (2016).
9. Christiansen, E. M. *et al.* In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792–803 (2018).
10. Eulenberg, P. *et al.* Reconstructing cell cycle and disease progression using deep learning. *Nat. communications* **8**, 1–6 (2017).
11. Lippeveld, M. *et al.* Classification of human white blood cells using machine learning for stain-free imaging flow cytometry. *Cytom. Part A* **97**, 308–319 (2020).
12. Reimann, R. *et al.* Classification of dead and living microalgae *Chlorella vulgaris* by bioimage informatics and machine learning. *Algal Res.* **48**, 101908 (2020).
13. Kusumoto, D. *et al.* Automated deep learning-based system to identify endothelial cells derived from induced pluripotent stem cells. *Stem cell reports* **10**, 1687–1695 (2018).
14. Kesler, B., Li, G., Thiemicke, A., Venkat, R. & Neuert, G. Automated cell boundary and 3d nuclear segmentation of cells in suspension. *Sci. reports* **9**, 1–9 (2019).
15. Tanner, M. *et al.* Characterization of a novel cell line established from a patient with herceptin-resistant breast cancer. *Mol. cancer therapeutics* **3**, 1585–1592 (2004).

16. He, K., Zhang, X., Ren, S. & Sun, J. Identity mappings in deep residual networks. In *European conference on computer vision*, 630–645 (Springer, 2016).
17. Bishop, C. M. *et al.* *Neural networks for pattern recognition* (Oxford university press, 1995).
18. Maaten, L. v. d. & Hinton, G. Visualizing data using t-sne. *J. machine learning research* **9**, 2579–2605 (2008).
19. Marin, J.-M., Mengersen, K. & Robert, C. P. Bayesian modelling and inference on mixtures of distributions. *Handb. statistics* **25**, 459–507 (2005).
20. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
21. Iandola, F. N. *et al.* Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. *arXiv preprint arXiv:1602.07360* (2016).
22. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).
23. Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning* (MIT press, 2016).
24. Japkowicz, N. & Stephen, S. The class imbalance problem: A systematic study. *Intell. data analysis* **6**, 429–449 (2002).
25. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255 (Ieee, 2009).
26. Brodersen, K. H., Ong, C. S., Stephan, K. E. & Buhmann, J. M. The balanced accuracy and its posterior distribution. In *2010 20th International Conference on Pattern Recognition*, 3121–3124 (IEEE, 2010).
27. Kelleher, J. D., Mac Namee, B. & D’arcy, A. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies* (MIT press, 2020).
28. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626 (2017).
29. Patel, A. G. & Kaufmann, S. H. Cancer: How does doxorubicin work? *Elife* **1**, e00387 (2012).
30. Weaver, B. A. How taxol/paclitaxel kills cancer cells. *Mol. biology cell* **25**, 2677–2681 (2014).
31. Lugagne, J.-B. *et al.* Identification of individual cells from z-stacks of bright-field microscopy images. *Sci. reports* **8**, 1–5 (2018).
32. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, 8026–8037 (2019).

## Acknowledgements

This study was supported by Agencia Nacional de Promoción Científica y Tecnológica, Argentina, PICT 2018-04517, Préstamo BID-PICT 2016-0222 and BID-PICT 2018-01582, donations from the Federico Deutsch Jack Yael Foundation, the Banchemo Family and Grupo Día to M.S, PID UTN 2018 (SIUTNBA0005139), PID UTN 2019 (SIUTNBA0005534), and NVIDIA GPU hardware grant that supported this research with the donation of two Titan Xp graphic cards. We also want to deeply acknowledge Ing. Gustavo Rosero, PhD. Marie Follo and Prof. Dr. Roland Mertelsmann of the Zentrum für Translationale Zellforschung (ZTZ) Universitätsklinikum Freiburg, Lighthouse Core Facility, who provided us with access to the broad spectrum of fluorescence-based technologies used for this research.

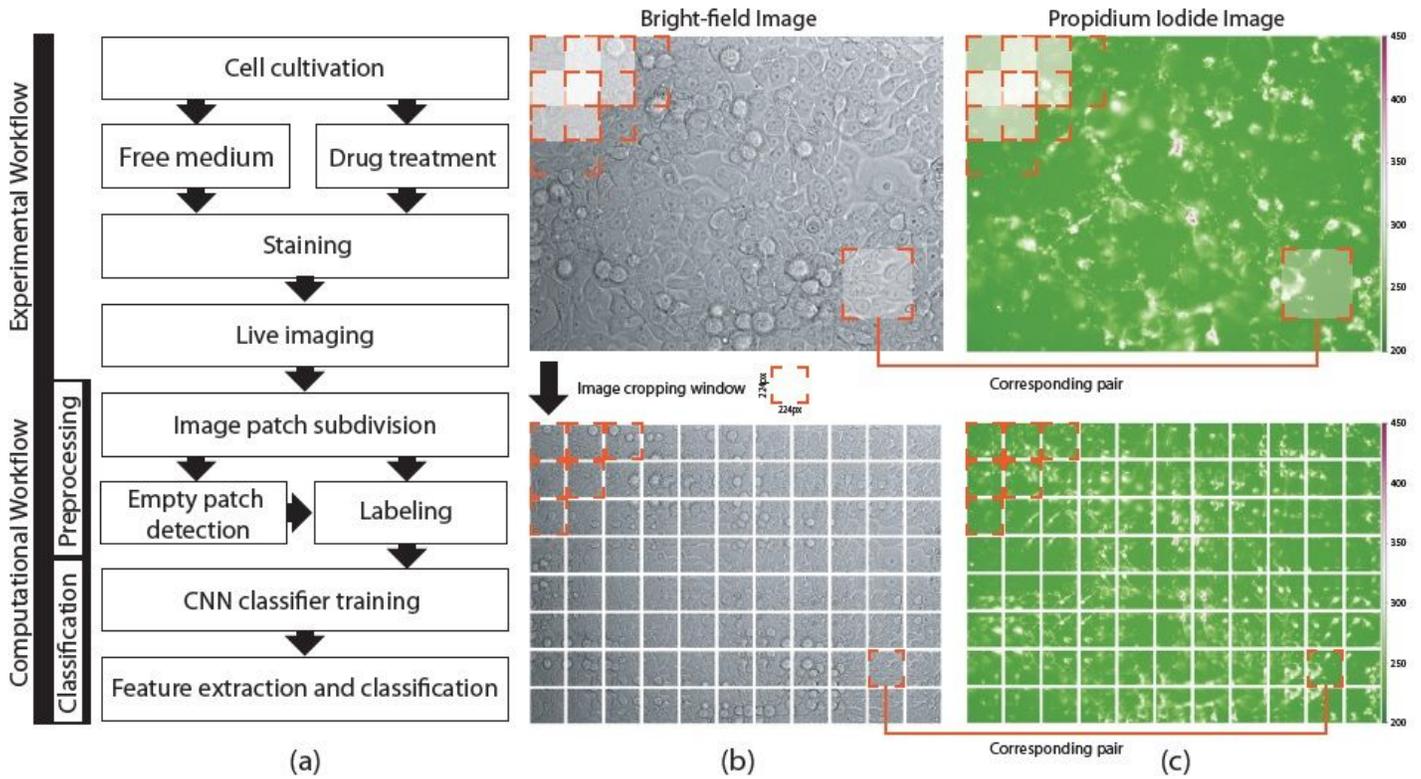
## Author contributions statement

G.P. conceived the lab experiments and dataset, G.P. and E.I. conducted the computational pipeline. All authors analysed the results, wrote and reviewed the main manuscript text.

## Additional information

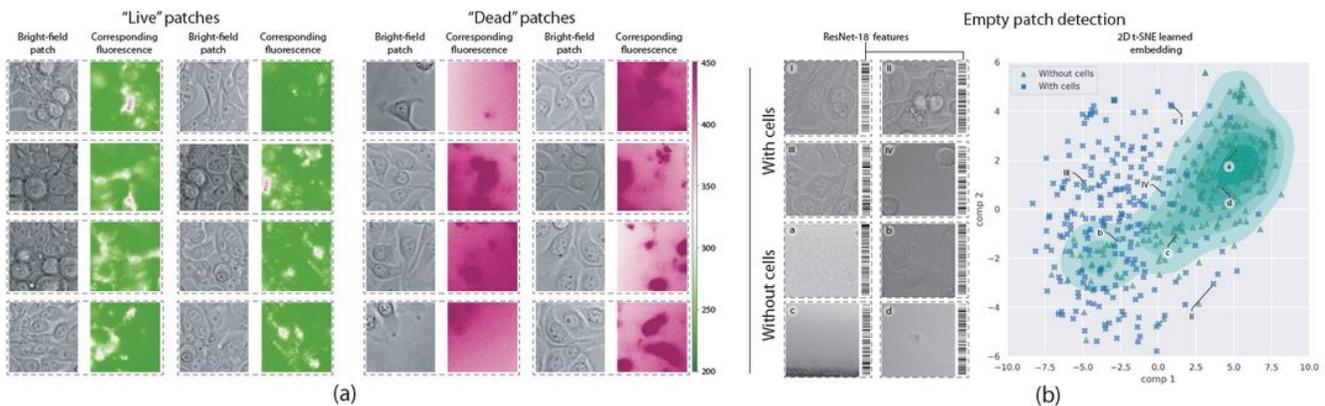
**Competing interests.** The authors declare no competing interests.

# Figures



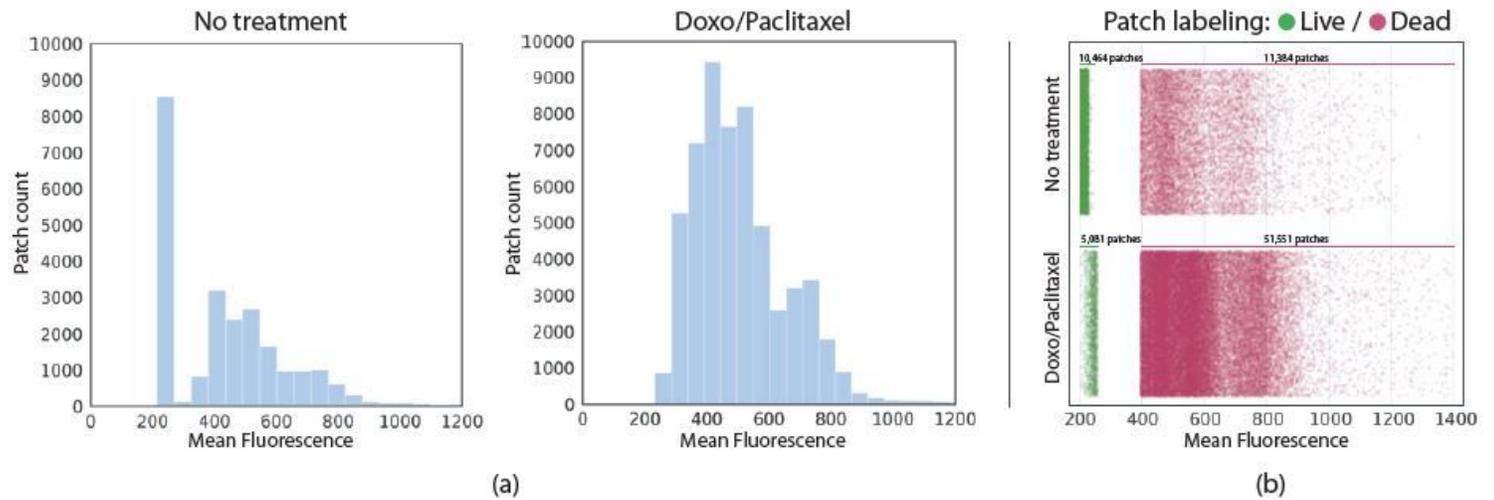
**Figure 1**

(a) Experimental and computational steps in our automated cell classification pipeline. (b-c) Top: bright-field and corresponding fluorescence images resulting from the imaging step (experimental workflow). High fluorescence values (white and red areas) indicate cell death. Bottom: as a post-process, images are cropped into 224x224px patches and paired with their corresponding fluorescence patch. Notice cropping overlaps contiguous patches (horizontal and vertical) in order to augment the number of images



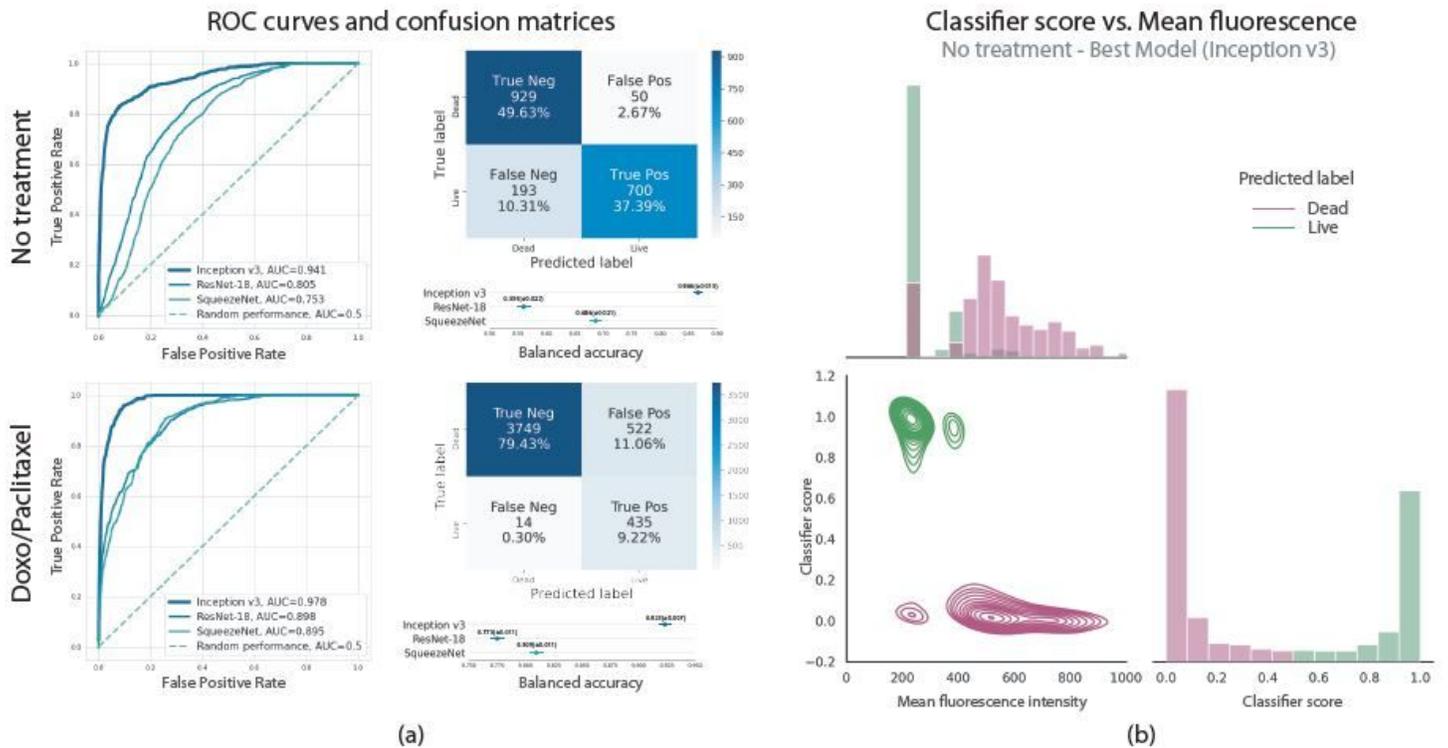
**Figure 2**

(a) Bright-field and corresponding propidium iodide fluorescence images. The columns under Live patches (green) show images with mostly live cells. The columns under Dead patches (red) present images with mostly dead cells in our dataset. (b) Left, samples of empty and non-empty patches with the associated barcode visualization of the 512-dimensional feature vector from ResNet-18 last convolutional output per image. Right, dimensionality reduction visualization of patch features and the high-dimensional decision function (green level-sets) learned by the SVM. Notice how empty patches mostly lie inside the highlighted region.



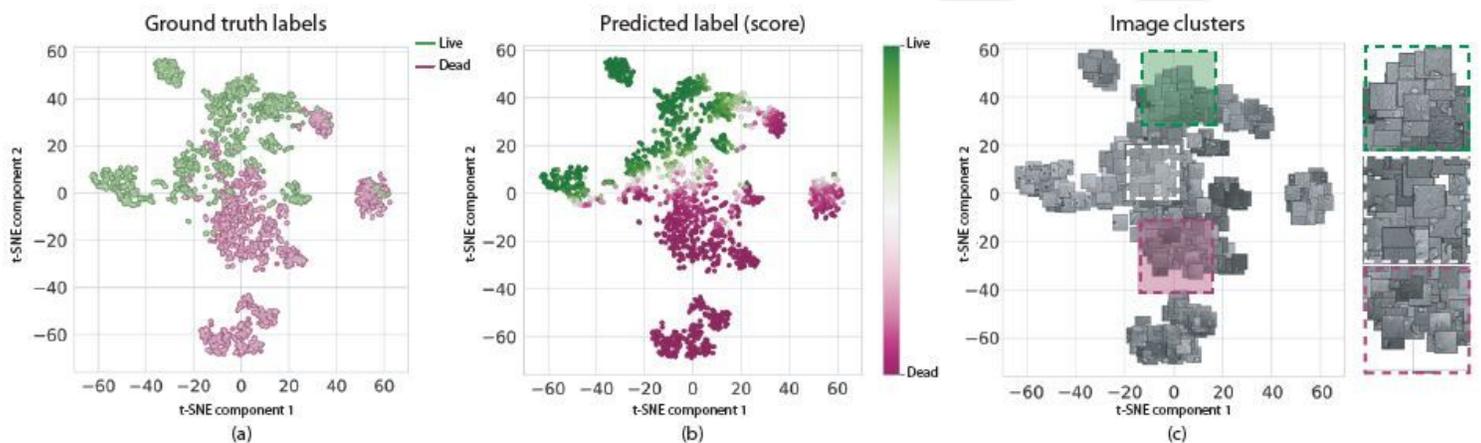
**Figure 3**

(a) Histograms of mean fluorescence per patch on each dataset. (b) Scatter diagram showing the samples distribution after the automatic patch labeling based on their mean fluorescence. Due to treatment, Doxo/Paclitaxel has significantly less live patches.



**Figure 4**

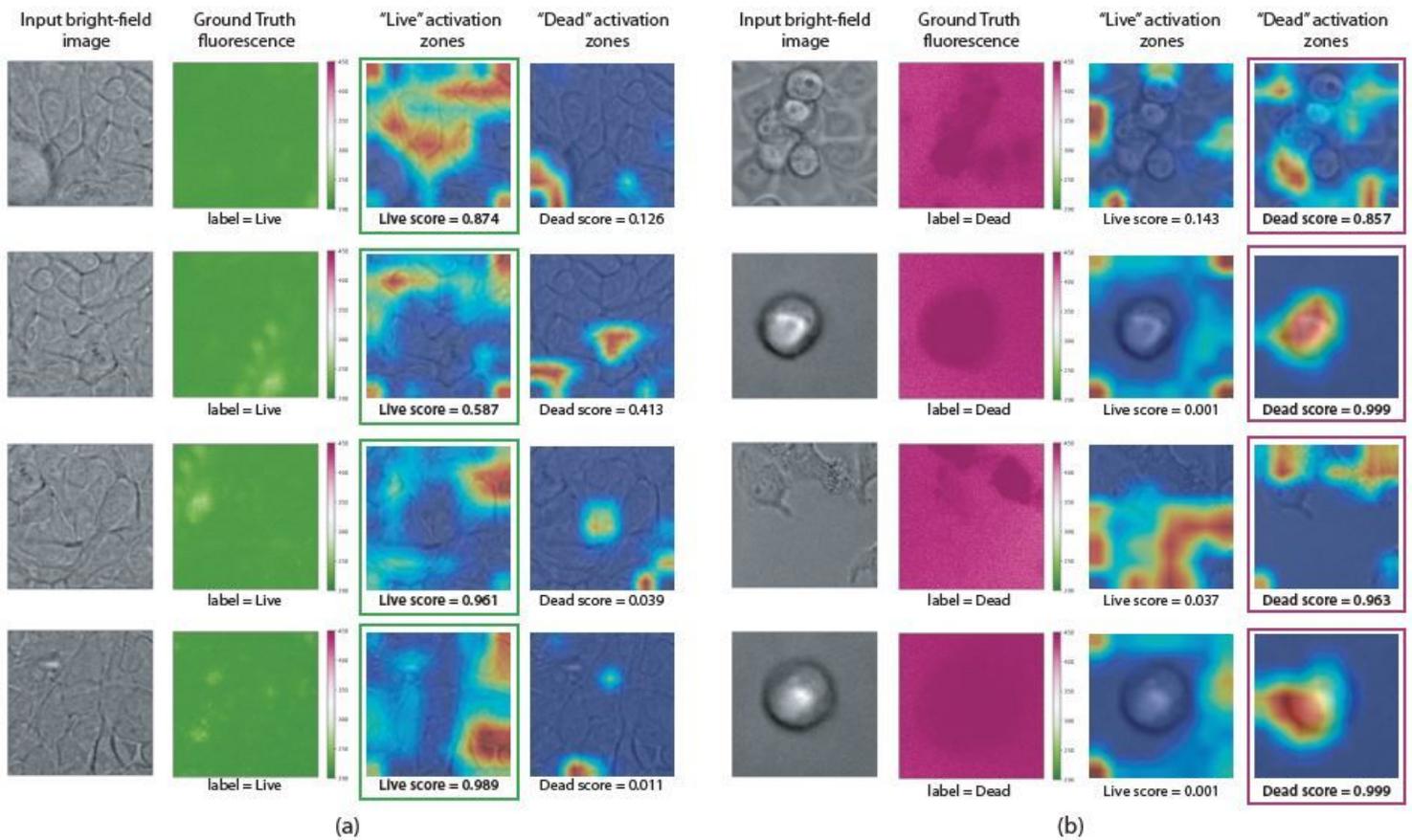
(a) ROC curves showing the classification performance over the testing datasets for each CNN architecture. The Inception-v3 model outperforms ResNET and SqueezeNET. (b) Mean fluorescence vs. classifier-score analysis for the model that performed best over No treatment data (Inception-v3). Higher mean fluorescence intensities tend to cluster together for lower classification scores used to label dead cells. Simultaneously, lower mean fluorescence intensities are grouped near higher classification scores that signal live cells.



**Figure 5**

Visualization of the Inception-v3 learned feature space for our No treatment testing dataset. The 2048-dimension features were projected to a 2D space using t-SNE, and colored according to ground truth

labels (a), and predicted labels (b). Cells with the same state tend to cluster together. Visual inspection of the images in each cluster further reveals the shared characteristics within each group.



**Figure 6**

Bright-field testing patches paired with gradient-weighted class activation mapping (Grad-CAM) visualizations for the Inception-v3 model. Live patches are shown on the left (a), and dead on the right (b). The activation maps show which zones of the input bright-field are triggering the classifier response. These maps can be computed for both labels and help to identify zones in the input images activating a live or a dead response from the convolutional neural network. The maps also have an associated score, indicating the probability of each label, which determines the final classifier response.