



Detecting abnormal cell behaviors from dry mass time series

Romain Bailly, Marielle Malfante, Cédric Allier, Chiara Paviolo, Lamya Ghenim, Kiran Padmanabhan, Sabine Bardin, Jerome Mars

► To cite this version:

Romain Bailly, Marielle Malfante, Cédric Allier, Chiara Paviolo, Lamya Ghenim, et al.. Detecting abnormal cell behaviors from dry mass time series. 2023. hal-04275745

HAL Id: hal-04275745

<https://cnrs.hal.science/hal-04275745>

Preprint submitted on 8 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Detecting abnormal cell behaviors from dry mass time series

Romain Bailly^{1,7}, Marielle Malfante^{1,*}, Cédric Allier^{2,3}, Chiara Paviolo², Lamya Ghenim⁴, Kiran Padmanabhan⁵, Sabine Bardin⁶, and Jérôme Mars⁷

¹Univ. Grenoble Alpes, CEA, List, Grenoble, F-38000, France

²Univ. Grenoble Alpes, CEA, Leti, Grenoble, F-38000, France

³Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

⁴Univ. Grenoble Alpes, INSERM, CEA-IRIG, BGE, Biomix, Grenoble, F-38000, France

⁵Institut de Génomique Fonctionnelle de Lyon, UMR 5242, Ecole Normale Supérieure de Lyon, CNRS, Université Claude Bernard Lyon 1, Lyon, France

⁶CNRS, UMR 144, Molecular Mechanisms of Intracellular Transport, PSL Research University, Institut Curie

⁷Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, Grenoble, 38000, France

*marielle.malfante@cea.fr

ABSTRACT

The prediction of pathological changes on single cell behaviour is a challenging task for deep learning models. Indeed, in self-supervised learning methods, no prior labels are used for the training and all of the information for event predictions are extracted from the data themselves. We present here a novel self-supervised learning model for the detection of anomalies in a given cell population, StArDusTS. Cells are monitored over time, and analysed to extract time-series of dry mass values. We assessed its performances on different cell lines, showing a precision of 94% in the automatic detection of anomalies. Additionally, anomaly detection was also associated with cell measurement errors inherent to the acquisition or analysis pipelines, leading to an improvement of the upstream methods for feature extraction. Our results pave the way to novel architectures for the continuous monitoring of cell cultures in applied research or bioproduction applications, and for the prediction of pathological cellular changes.

1 Introduction

Predicting the future behavior of single cells within a large population is an interesting task for deep learning models. To this end, different approaches have already been developed to predict cell trajectories from transcriptomics dataset¹, detect cells that deviate from normal behavior in time-lapse imaging², and learning cell competition rules³. Once a complete behavioral pattern is known within a temporal interval, any deviation from “normality” can be classified as “abnormal”. This is of crucial importance when monitoring a cell population, as identifying cells that harbor cancerous mutations based on deviation from a “typical” trajectory would allow for predictions about pathological changes.

Recently, novel techniques in bioimaging coupled to advanced deep learning models have allowed the visualization, quantification, and monitoring of important cellular features over time on large populations and in label-free conditions⁴. Lensfree phase microscopy allows indeed the extraction of important cellular features over time, such as the dry mass (i.e. the weight of the cell’s content other than water). This parameter is a critical biological feature, as it remains invariant over different cell generations and reflects the growth and division stages of a cell. Dry mass is therefore key to the understanding of cellular behavior, including cell size, cycle, state, and homeostasis⁵⁻⁷. Monitoring the dry mass of a cell over time is thus a proxy of the overall cellular status, and could allow the prediction about possible abnormal deviations.

We propose a method using self-supervised learning to detect abnormalities or anomalous cell behavior on temporal trajectories of dry mass. Self-supervised learning is a recent training paradigm that does not need any label to train a machine learning model. It uses information extracted from the data themselves, called pseudo-labels. These pseudo-labels are used to train a neural network to perform a pretext task, that can then be used in the downstream pipeline, in our case anomaly detection. Forecasting the future dry mass of a cell is the pretext task used in this paper.

Unsupervised (i.e. no labels are manually provided to the model) anomaly detection has been used lately in a wide variety of domains⁸, including but not limited to, astronomy⁹, earth science¹⁰, neuroscience¹¹, oceanography¹² or physics^{13,14}. We focused on anomaly detection on time series without any prior labels as presented by Gupta *et al.*. Here, the detection is achieved with a two stage anomaly detector, relying on the comparison of the measured and predicted nominal trajectory.

We call the proposed model StArDusTS, for Self-supervised Anomaly Detection on Time Series, and use it on distinct

datasets of cellular dry mass time series. To assess its performances, we designed an experimental validation using different cell lines, *i.e.* a human cancer HeLa cell line and murine fibroblasts cells, cultured and imaged in different laboratories. Overall, we report a precision of 94% in the automatic detection of anomalies present in these different datasets. Several types of biological anomalies from the measurement of cell dry mass alone and without any human priors were detected in the different time-lapses, *e.g.* cells dividing to three cells, very large cells and cell fusion. Additionally, anomaly detection was associated not only with abnormal cell behavior but also with cell measurement errors inherent to the acquisition or analysis pipelines, such as segmentation and tracking. This could lead to an improvement of the upstream methods for cell imaging and analysis. Our results pave the way to novel architectures for the continuous monitoring of cell cultures in applied research or bioproduction applications.

2 Method

Here we describe the cell imaging and analysis pipeline section 2.1, the architecture of the StArDusTS model 2.2 and the different experiments conducted to validate the algorithms. Importantly, we designed an experimental plan featuring three sets of live cell acquisitions conducted in different laboratories.

2.1 Lensfree microscopy for cell dry mass measurement

Lensfree microscopy, a quantitative phase imaging technique, allows for acquisition of phase images of cells over a large field of view of $\sim 30\text{mm}^2$ directly in a cell incubator¹⁵. Using this method, thousands of cells are simultaneously imaged over several days. The obtained dataset allow the tracking of thousands of individual cells over several tens of hours. Furthermore, it allows the computation of the cell dry mass through the measurement of the optical path difference (OPD) introduced by the sample^{15–17}. OPD is measured by the integral of the sample's refractive index along the optical path.

$$\varphi(x, y)_{\text{shift}} = (\varphi(x, y)) - (\varphi_{\text{medium}}) \quad (1) \quad \text{OPD}(x, y) = \lambda \frac{\varphi_{\text{shift}}(x, y)}{2\pi} = \int_0^h [n(x, y, z) - n_{\text{medium}}] dz \quad (2)$$

where n is the local sample refractive index and n_{medium} the surrounding medium refractive index, z the position along the optical axis, h the thickness of the sample and λ the illumination wavelength¹⁵. The optical volume difference (OVD) is obtained by integrating the OPD over the total projected area. The OVD can then be converted into cell dry mass according to eq. (3). In our notation, it is a function of α , the specific refractive increment which relates the refractive index change to the increase in mass density¹⁸. The specific refractive index of the different intracellular substances falls with a narrow range, allowing the definition of a constant α of $0.18 \mu\text{m}^3 \cdot \text{pg}^{-1}$ for most eukaryotic cells¹⁸.

$$\text{OVD} = \int_S \text{OPD}(x, y) dx dy \quad (3) \quad \text{CDM} = \frac{\text{OVD}}{\alpha} \quad (4)$$

The dry mass measurements have been obtained using a previously described cell imaging analysis pipeline⁴. The latter includes the acquisitions of raw images with a lensfree microscope at a frame rate of one acquisition every 10 minutes, the reconstruction of OPD images with the algorithm described in¹⁹, the detection of the cell with a dedicated 2D-CNN and tracking of each individual cell by means of Fiji plugin Trackmate²⁰ and a cell segmentation performed by a watershed-algorithm. For the cell dry mass measurement, several sources of noise are present, *i.e.* in the acquisition, the reconstruction of the OPD, the cell detection, and segmentation. The measured cell dry mass values are in the order of a few hundreds of picograms (pg), while the precision of our measurements was estimated to be about 35 pg¹⁷. Figure 1 shows an example of the cell imaging analysis pipeline in terms of segmentation (a) and cell dry mass time series (b). Figure A in supplementary materials shows an example of lensfree microscopy acquisition before the segmentation and tracking of cells.

2.2 StArDusTS model

To eliminate any potential human bias when identifying abnormal cells, we introduce the StArDusTS model, which leverages artificial intelligence to autonomously acquire insights from cellular data and identify abnormal patterns within it. StArDusTS, an acronym denoting “Self-supervised Anomaly Detection on Time Series”, comprises two independent algorithmic components. The first component, focuses on learning a data representation that surpasses the utility of raw time series data for the purpose of abnormal cell detection. This learned representation is subsequently channeled into the model's second component, the anomaly detection module, detailed after.

Representation learning block

Self-supervised learning: Traditionally, supervised learning^{21,22} has been the dominant paradigm in machine learning, where models are trained on meticulously labeled datasets. However, this labeling process is often labor-intensive, time-consuming,

expensive, and can induce human biases. Self-supervised learning, on the other hand, seeks to alleviate these limitations by enabling models to learn directly from the data itself.

At its core, self-supervised learning operates on the principle of leveraging inherent structures and relationships within the data^{23,24}. It does so by creating surrogate tasks, also called pretext tasks, that generate pseudo-labels or objectives from the input data. These pretext tasks require the model to predict missing parts of the data, reorder sequences, or otherwise make sense of the information it encounters. By solving these tasks, the model learns to capture meaningful features and representations that are useful for downstream tasks, such as image classification, language understanding, recommendation systems, or here, anomaly detection.

Examples of pretext tasks in computer vision include the reconstruction of an image^{25,26}, namely auto-encoders, the prediction of rotation of an image²⁷, the coloring of black and white images²⁸, image inpainting²⁹ or even artifact detection³⁰.

Pretext task adapted to time series : In the context of StArDusTS, we propose the utilization of time series prediction as a pretext task. Specifically, we partition 30-hours (180-points) windows of cell dry mass data into two segments: an input segment and a corresponding label. The initial 20 hours (120 points) of data serve as the input for the proposed neural network. It is tasked with forecasting the subsequent 10 hours (60 points) of cell dry mass, knowing those first 20 hours.

The length of the window has been chosen in order to always include a cell division inside the input window. If a smaller input window had been chosen, the cell division would have been a rare event in an input and would therefore not have been learned well, since it would have been subject to catastrophic forgetting³¹. It should also be noted that the decision to exclude mother-daughter cell acquisitions lasting less than 30 hours may introduce a potential bias into our algorithm. This bias arises from the restriction to studying only cells with sufficiently extended lifespans. Nevertheless, this selection was imperative to enable the model to acquire a meaningful representation of the dataset.

Evaluation metric for prediction: To evaluate the predictive performance, we use the Mean Squared Error (\mathcal{MSE}) metric, as expressed in eq. (5). In this equation, y is the actual value of the time series, \hat{y} denotes the predicted values, y_t and \hat{y}_t are the actual and predicted values at each time step t , and $N = 60$ corresponds to the number of time steps in a prediction. A lower value of \mathcal{MSE} indicates superior predictive performances.

$$\mathcal{MSE}(y, \hat{y}) = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2 \quad (5)$$

Neural network architecture: In conjunction with our choice of a pretext task, it is imperative that we meticulously define the architecture of the AI model that will be assigned with this task. The architecture dictates the network's structure, layer configurations, and parameter settings, all of which play a pivotal role in the model's ability to extract meaningful patterns and make accurate predictions.

A one-dimensional convolutional neural network (1D-CNN) is a deep learning architecture designed specifically for processing one-dimensional data, such as time series. Unlike traditional convolutional neural networks (2D-CNN) that operate on two-dimensional grids, 1D-CNNs convolve filters over a single axis, typically time or sequence steps. In a 1D-CNN, convolutional layers are responsible for sliding a small set of learnable filters over the input data, capturing local patterns and features. These filters can detect characteristics like edges, gradients, or more complex temporal patterns in the data. Subsequent layers, such as pooling and fully connected layers, help consolidate these features and enable the network to learn high-level representations³².

The choice of the 1D-CNN was motivated by its use in a wide variety of time series applications such as ECG classification^{33,34}, fault detection³⁵⁻⁴⁰, or speech recognition^{41,42}. 1D-CNN for time series processing is beneficial because it can capture local patterns and dependencies in the data through its convolutional filters, making it effective for tasks like feature extraction and anomaly detection.

The 1D-CNN architecture used in this study is presented figure 2.a) and contains 3 blocks of 3 Conv1D layers with 64 kernels of size 3, paired with `tanh` activation functions. The blocks of 3 1D-CNN are separated with `maxpooling` layers of size 2. The features are then fed in `dense` layers of size 64 and 32 with `relu` activation functions and finally an output layer of size 60. This architecture was selected after a manual optimization of hyperparameters of 1D-CNN⁴³.

Anomaly detection block

The representation learned by the 1D-CNN is then fed to the anomaly detection block. For this application of detection of abnormal cells, we propose two complementary detectors, the second one working on top of the first one.

Window level anomaly detection: A first threshold detector is used to detect the anomalies within a single predicted window. It is based on the value of the prediction metric. The \mathcal{MSE} computes the l_2 distance between the prediction and the actual value

of the future cell dry mass. The larger the MSE , the larger the error of the predictor. Selecting the windows with the larger values of metric is therefore selecting the most anomalous windows.

The threshold τ_w determining which prediction windows have to be considered anomalous is computed such that windows with metric value outside the 95% confidence interval are anomalous. The threshold τ_w is computed eq. (6) with μ_{training} and σ_{training} respectively the mean and standard deviation of the MSE of the training set.

$$\tau_w = \mu_{\text{training}} + 2 \cdot \sigma_{\text{training}} \quad (6)$$

Cellular level anomaly detection: The detection of abnormal cells consists in the aggregation of multiple window-wise anomalies. For instance, the dry masses of the mother cell and its daughter cells are observed for 50 hours. This full length time series is seen independently through 121 overlapping windows of 30 hours by the threshold detector. In order to aggregate all the window level detections into a full cellular level, we build a second detector on top of the first one.

The anomaly score \mathbb{A} , eq. (7), uses all the results of detection of all the windows extracted of a cell dry mass and computes a single score from them. This score is the ratio of abnormal windows to the total number of windows of this full-lengthed time series. Cells with a higher score are expected to be *more abnormal* than those with a lower score.

$$\mathbb{A} = \frac{\text{\# of abnormal windows}}{\text{total \# of windows}} \quad (7)$$

Figure 2.a) shows the window-wise anomaly detection based on the representation extracted with a 1D-CNN and Figure 2.b) shows the StArDusTS model for a whole time series, including the 1D-CNN representation learning block, the window-wise anomaly detection with the threshold detector and the aggregation of window-wise results with the anomaly score.

2.3 Experimental plan and constructed datasets

Cell culture acquisition

We designed a set of three live cell time-lapses to validate the models. To obtain a first assessment, a set of acquisitions a were performed with HeLa cells cultured and imaged in the laboratory. To test generalization, the model trained based on acquisitions a have been applied to another set of acquisitions b . The latter was obtained with HeLa cells cultured and imaged in a second laboratory. This is a strong generalization test since there are differences between the HeLa cell lines but also between the cell culture protocols used in the two different laboratories.

Finally, with a last set of acquisitions c , we could train a deep learning model with a wild type murine fibroblast cell line and test it on abnormal mutated fibroblasts d . Acquisitions c and d were conducted in the same laboratory.

- HeLa cell culture (dataset a) coming from the ATCC catalog. HeLa cells were cultured in high glucose Dulbecco's Modified Eagle Medium (DMEM) supplemented with GlutaMAX, pyruvate, and 10% (v/v) calf serum (Gibco). Cells were grown onto 35 mm glass bottom (0.17 mm) dishes and imaged every 10 minutes on a Cytonote 1W (Iprasense) for 24-48 hours at 37 °C and 5% CO_2 .
- HeLa cell culture (dataset b) donated by Dr Dimitrios Skoufias of the *Institut de biologie structurale* in Grenoble. These cells were grown in DMEM supplemented with GlutaMAX, 10% (v/v) heat-inactivated fetal calf serum (FCS), and 1% penicillin and streptomycin. For imaging, 6-well glass bottom culture plates were coated with fibronectin (25 µg/mL) for 1 h. Cells were seeded at a concentration of $2 \cdot 10^4$ cells per well and imaged on a Cytonote 6W (Iprasense) every 10 minutes at 37 °C and 5% CO_2 .
- For acquisitions c and d , wild type mouse fibroblasts were isolated from C57BL/6 mice (acquisition c) while Per0 fibroblasts were isolated from *Period1* ($mPer1^{ldc/-}$); *Period2* ($Per2^{ldc/-}$); *Period3* ($mPer3^{-/-}$) triple knock-out mice (PMID:35606517, (acquisition d))⁴⁴. All cell lines were cultured in standard DMEM (high glucose) supplemented with 10% FCS (Thermo Fisher), penicillin (25 units/mL, Thermo Fisher) and streptomycin (25 units/mL, Thermo Fisher). Cells were passaged following trypsinization at low density ($2 - 5 \cdot 10^4$) onto 35mm glass bottom (0.17mm) dishes (FD-35, Fluoro-dish WPI) and imaged every 10 minutes for 24-96 hours after attachment on a Cytonote 1W (Iprasense) housed inside standard cell culture incubator at a controlled temperature and humidity.

Construction of the datasets

The dry mass time series obtained on the basis of the different acquisitions mentioned above have been post-processed to train and validate the models. The time series are split into training, validation and test datasets containing respectively 80%, 10% and 10% of the data points²². These *sub-dataset* have been independently normalized such that their mean value is 0 and their standard deviation is 1. Mother-daughter tracks shorter than 30 hours are discarded. We used a sliding window, 30 hours wide,

moved along the full length time series to generate numerous slightly modified versions of the same time series at a 10-minute interval difference. This augmentation technique⁴⁵ is used to maximize the data fed to neural networks in order to better learn the phenomenon. Finally, each 30-hour window is split in two segments: the first 20 hours that will be given as **inputs** to the neural network and the last 10 hours to be predicted and therefore unknown to the neural network. The algorithms must predict the value of the dry mass during this next 10 hours. A 1-hour sliding window is used to smooth the 10-hours signals to be predicted. A summary of the respective number of time series in the splits of each dataset is available in Table 1.

Table 1. Train/Validation/test distribution for all four datasets. For each one, the total number of cells, the numbers of cells with a life-span over 30 hours and the number of extracted cell dry mass 30 hours time series are reported.

Acq	Dataset	Cell line	Location	Number of	Total	Train (80%)	Validation (10%)	Test (10%)
<i>a</i>	\mathcal{A}	HeLa	Grenoble CEA	Cell trajectories	36 396	29 118	3 639	3 639
				Cells trajectories > 30 h	6 700	5 520	621	559
				30 hours time series	559 141	440 930	60 179	58 032
<i>b</i>	\mathcal{B}	HeLa	Paris Curie institute	Cell trajectories	27 528	22 024	2 752	2 752
				Cells trajectories > 30 h	7 283	5 859	706	718
				30 hours time series	834 918	665 280	86 363	83 275
<i>c</i>	\mathcal{C}	Fibroblasts Wild Type	Lyon École Normale Supérieure	Cell trajectories	28 243	22 596	2 823	2 824
				Cells trajectories > 30 h	3 986	3 169	404	413
				30 hours time series	156 381	123 950	16 259	16 172
<i>d</i>	\mathcal{D}	Fibroblasts Knock Out	Lyon École Normale Supérieure	Cell trajectories	5 035			5 035
				Cells trajectories > 30 h	315			315
				30 hours time series	26 688			26 688

3 Results and discussion

3.1 Experiment 1 : Anomaly detection

The purpose of this first experiment is to assess the anomaly detection performances of the StarDusTS model. Therefore, the algorithm is trained, validated and tested on the **same** datasets. The aim is to check whether the StArDusTS model is capable to learn a representation and detect anomalies among the **same** cell culture. This experiment is run on both dataset \mathcal{A} and \mathcal{B} . In order to assess the performances of the detection, some of the videos were manually analyzed to detected abnormal cells. These labels allowed the assesement of the precision (i.e. the ratio between the anomalies and the detections) of the model.

After training, the StArDusTS models raised respectively 104 and 198 abnormal cells from datasets \mathcal{A} and \mathcal{B} . These anomalies are manually annotated thanks to the original videos from which the dry mass time series were extracted. All cells detected in dataset \mathcal{A} are labeled. On dataset \mathcal{B} , 104 cells are randomly drawn from the 198 detected as abnormal by the model. Only those cells are annotated to better compare the results on both datasets. We identified three main causes for the model to raise an anomaly:

- (i) The cell, flagged as abnormal by StArDusTS has an abnormal behavior. Such behavior are discussed in detail after and are referred to as **cellular anomalies** or **biological anomalies**.
- (ii) The acquisition system, including the lensfree microscope, the reconstruction algorithm, the segmentation algorithm and the tracking one, resulted in an error on the input fed into StArDusTS. Such anomalies are referred to as **acquisition anomalies**.
- (iii) The cell seems to have a normal behavior, from the video and available time series perspectives. Such anomalies are **False Positives** (FP) in the detection.

Two more classes emerged from the manual annotation of anomalies. On the one hand, some abnormal cell behaviors mislead the acquisition system, therefore leading to erroneous times series. Such detected anomalies are both **acquisition AND biologic anomalies**. On the other hand, some anomalies are impossible for us to classify. As it impossible for us to say whether they are acquisition anomaly or biologic ones, they are called **acquisition OR biologic anomalies**.

Table 2 gives the details of the manual annotation of this experiment on both datasets \mathcal{A} and \mathcal{B} . The precision is the ratio of the number of good detections to the total number of detections. StArDusTS algorithm was able to detect anomalies with a **precision** up to 94.2% on cells from Grenoble CEA and 83.5% on cells from Curie's Institute.

Table 2. Distribution of annotated anomalies for experiment 1 on datasets \mathcal{A} and \mathcal{B} . Underlined anomalies are biologic ones which are detailed in table 3

	\mathcal{A}	\mathcal{B}
Acquisition	49% (51)	31.1% (32)
Biologic	36.5% (<u>38</u>)	38.8% (<u>40</u>)
Acq and Bio	2.9% (3)	3.9% (4)
Acq or Bio	7.7% (8)	9.7% (10)
False positives	3.8 % (4)	16.5% (17)
Precision	94.2%	83.5%

Table 3. Manual annotation of biologic anomalies for experiment 1 on datasets \mathcal{A} and \mathcal{B}

	\mathcal{A} (38)	\mathcal{B} (40)
(a) Abnormal division	8 % (3)	15% (6)
(b) Abnormal Growth	50% (19)	40% (16)
(c) Merging cells	26% (10)	5% (2)
(d) Too big cells	5 % (2)	20% (8)
(e) Dead cells	8 % (3)	20% (8)
(f) Others	3 % (1)	0% (0)

Most biological anomalies are different one from another. However, we propose here to group them into 5 major classes. The latter identified during the manual labeling of abnormal cells:

- (a) Cells with an **abnormal division**, including cells that do not divide and stagnate on a plateau of dry mass, cells that divide asymmetrically or cells that divide in more than two daughter cells,
- (b) Cells with an **abnormal growth**, which may include growth that is too long, too short, non-linear or with unexplained dry mass loss,
- (c) Cells that **merge** with each other,
- (d) Cells that are **too big**,
- (e) **Dead** cells
- (f) All **other** cellular anomalies.

Table 3 shows the class distribution of these biological anomalies, both on cells from Grenoble CEA and Curie’s Institute. Figures 3a to 3f show the original acquisitions time-lapses (*i.e.* series of snapshots) of one example for each cellular anomaly class. For each time-lapse, we display the dry mass time series from which the anomalies were detected. Each instant corresponding to the images shown above are framed in red. Figure 3g focuses on an acquisition anomaly. More examples of all anomalies are available in the supplementary materials figure B. The videos from which these snapshots were extracted are also available as supplementary material.

3.2 Experiment 2 : Generalization

We showed in the previous experiment that the StArDusTS model is able to detect anomalies in the dataset on which it has been train. To test its generalization capabilities, the model trained on \mathcal{A} has been applied to a different set of HeLa cell acquisitions \mathcal{B} . HeLa cells were considered as a good model for detecting anomalies, as they often present one extra version of most chromosomes with up to five copies detected in a single cell⁴⁶. Moreover, cells in a and in b were cultured and imaged under different protocols and laboratories, therefore increasing the differences between the phenotypes.

When conducting this experiment, the StArDusTS model identified 939 abnormal cells from the 7283 cells of dataset \mathcal{B} . Among those cells, 152 were already detected as abnormal in the previous experiment. This shows a strong consistency of the models in the detection of abnormal cells, since both experiments detected the same cells as abnormal. Tables 4 and 5 respectively show the distribution of the 104 manually annotated cells and the classification of biologic anomalies.

When the model was trained and tested on the same cells from dataset (\mathcal{A}), it detected 49% of acquisition anomalies and 36.5% of biological anomalies. During the generalization tests, StArDusTS detected 42.3% of acquisition anomalies and 43.3% of biological anomalies, proving the good generalization of the representation.

StArDusTS model obtained a better performance on the anomaly detection for dataset \mathcal{B} when trained on dataset \mathcal{A} (precision of 94.2%) than when it was trained on dataset \mathcal{B} itself (83.5% precision). A plausible explanation is the difference in image quality between the 2 datasets. This experiment also shows that the model trained on dataset \mathcal{A} can be transferred to other datasets. Indeed, it kept the exact same precision of 94.2% regardless of whether it is tested on dataset \mathcal{A} or \mathcal{B} .

Two conclusions can be drawn from this experiment : First, we showed that StArDusTS model is general enough to be able to detect anomalies in dataset it has not been trained on. Second, the performances of the model, and especially its precision, is highly impacted by the quality of the dataset used for training.

Table 4. Distribution of annotated anomalies for experiment 2

Acquisition	42.3 % (44)
Biologic	43.3 % (45)
Acq and Bio	6.7 % (7)
Acq or Bio	3.8 % (4)
False positives	3.8 % (4)
Precision	94.2%

Table 5. Manual classification of the **biologic anomalies** of experiment 2

(a) Abnormal division	20 % (9)
(b) Abnormal Growth	49% (22)
(c) Merging cells	9% (4)
(d) Too big cells	7 % (3)
(e) Dead cells	11 % (5)
(f) Others	4 % (2)

3.3 Experiment 3: Controlled experiment

The previous experiments have shown the capability of StArdusTS for anomaly detection. However, dataset \mathcal{A} and \mathcal{B} could not be used for the recall evaluation. At this purpose, wild type and genetically modified fibroblasts were analyzed to quantify the number of undetected anomalies and completely characterize the StArdusTS model. For the analysis, cells from dataset \mathcal{C} were labeled as normal, while cells from acquisition \mathcal{D} as abnormal.

To address this problem, we propose to **artificially create** a “labeled” dataset. Indeed, cells from dataset \mathcal{D} are almost identical to cells from \mathcal{C} with the difference that they are genetically modified. Because of this modification, cells from \mathcal{D} are expected to behave differently. Thus, we can label cells from \mathcal{C} as “normal” and cells from \mathcal{D} as “abnormal”. Now, both precision and recall of the StArdusTS model can be measured.

This experiment allows the computation of the recall to 0.68 at the threshold τ_w .

Figure 4 shows the ROC (Receiving curve for the detection of abnormal windows on datasets $\mathcal{C}_{\text{test}}$ and \mathcal{D}). The orange reference is the ROC curve of a random classifier. Each point is a threshold value of τ_w . The red cross is the τ_w value computed from the training dataset \mathcal{C} such that the time series outside the 95% interval of confidence are abnormal. It is however important to underline that the absolute values of precision and recall for the detections cannot be taken into account due to the strong hypothesis used for this experiment. Indeed, the assumption that all cells from dataset \mathcal{C} are normal and \mathcal{D} are abnormal might be biologically unrealistic.

The ROC curves show that choice of the τ_w value depending on the 95% confidence interval of the training dataset was made to have a good balance between the true and false positive rate. This is one of the best choice of threshold that could have been done since it is one of the points closest to the top left-hand corner of the graph.

4 Conclusion

In this paper, we propose a model for Self-supervised Anomaly Detection on Time Series called StArdusTS, which we applied to the detection of abnormal cells from their dry mass over time series.

StArdusTS relies on the learning of a representation of normal cells with 1D convolutional neural network trained to predict the future cell dry mass. Thanks to self-supervised learning, the detection is processed without any human induced biases during training.

In a first experiment, we validate the anomaly detection abilities of the StArdusTS model by successfully detecting abnormal time series on two datasets with a precision up to 94.2%. We were able to manually identify 2 causes of anomalies, either being cellular anomalies or acquisition anomalies. Biological anomalies were then classified into 6 sub-classes. The acquisition anomalies that we report can be used to compare and improve acquisition pipelines if needed. A second experiment validated that the representation learned from one dataset is general enough to be able to detect anomalies from cells grown in another lab. Moreover, it shows that the results are even better than those obtained with a model trained on the same data. Finally, a third experiment with dummy labels of known anomalies was set up to validate the choice of the anomaly detector.

While the representation is learned only from the dry mass time series, the StArdusTS model could be extended for the prediction of multiple features such a cell area, thickness or speed. Adding more modalities such as cell area or thickness could bring more information about the cell population being analyzed and thus allow the predictions of pathological changes.

Declarations

Competing interests

The authors declare no competing interests.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

Acknowledgments

This work was supported by the French ANR via Carnot funding. This work has partially received funding from the European Unions Horizon 2020 research program under grant agreement no. 101016726.

Authors' contributions

- RB: contributed to the conception and the developement of StArDusTS. It also contributed in the writing of this scientific paper,
- MM: contributed to the conception of StArDusTS and the revision of this paper,
- CA: contributed in the acquisition of the data used in this paper and the manual annotation of the raised anomalies. He also contributed in the writing of this paper,
- CP: contributed to the generation of cellular figures and the writing of this paper,
- LG: contributed in the acquisition of cellular data and the revision of this paper,
- KP: contributed in the acquisition of cellular data and the revision of this paper,
- SB: contributed in the acquisition of cellular data and the revision of this paper,
- JM: contributed to the conception of StArDusTS and the revision of this paper.

References

1. Yeo, G. H. T., Saksena, S. D. & Gifford, D. K. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nat. Commun.* **12**, 3222, DOI: [10.1038/s41467-021-23518-w](https://doi.org/10.1038/s41467-021-23518-w) (2021).
2. Soelistyo, C. J., Vallardi, G., Charras, G. & Lowe, A. R. Learning biophysical determinants of cell fate with deep neural networks. *Nat. Mach. Intell.* **4**, 636–644, DOI: [10.1038/s42256-022-00503-6](https://doi.org/10.1038/s42256-022-00503-6) (2022).
3. Soelistyo, C. J., Vallardi, G., Charras, G. & Lowe, A. R. Learning the rules of cell competition without prior scientific knowledge, DOI: [10.1101/2021.11.24.469554](https://doi.org/10.1101/2021.11.24.469554) (2021).
4. Allier, C. *et al.* CNN-Based Cell Analysis: From Image to Quantitative Representation. *Front. Phys.* **9** (2022).
5. Tzur, A., Kafri, R., LeBleu, V. S., Lahav, G. & Kirschner, M. W. Cell Growth and Size Homeostasis in Proliferating Animal Cells. *Science* **325**, 167–171, DOI: [10.1126/science.1174294](https://doi.org/10.1126/science.1174294) (2009).
6. Liu, X., Yan, J. & Kirschner, M. W. Beyond G1/S regulation: How cell size homeostasis is tightly controlled throughout the cell cycle?, DOI: [10.1101/2022.02.03.478996](https://doi.org/10.1101/2022.02.03.478996) (2022).
7. Ghenim, L. *et al.* A new ultradian rhythm in mammalian cell dry mass observed by holography. *Sci. Reports* **11**, 1290, DOI: [10.1038/s41598-020-79661-9](https://doi.org/10.1038/s41598-020-79661-9) (2021).
8. Wang, H. *et al.* Scientific discovery in the age of artificial intelligence. *Nature* **620**, 47–60, DOI: [10.1038/s41586-023-06221-2](https://doi.org/10.1038/s41586-023-06221-2) (2023).
9. Naul, B., Bloom, J. S., Pérez, F. & van der Walt, S. A recurrent neural network for classification of unevenly sampled variable stars. *Nat. Astron.* **2**, 151–155, DOI: [10.1038/s41550-017-0321-z](https://doi.org/10.1038/s41550-017-0321-z) (2018).
10. Rafique, M. *et al.* Delegated Regressor, A Robust Approach for Automated Anomaly Detection in the Soil Radon Time Series Data. *Sci. Reports* **10**, 3004, DOI: [10.1038/s41598-020-59881-9](https://doi.org/10.1038/s41598-020-59881-9) (2020).
11. Chamberland, M. *et al.* Detecting microstructural deviations in individuals with deep diffusion MRI tractometry. *Nat. Comput. Sci.* **1**, 598–606, DOI: [10.1038/s43588-021-00126-8](https://doi.org/10.1038/s43588-021-00126-8) (2021).
12. Pastore, V. P., Zimmerman, T. G., Biswas, S. K. & Bianco, S. Annotation-free learning of plankton for classification and anomaly detection. *Sci. Reports* **10**, 12142, DOI: [10.1038/s41598-020-68662-3](https://doi.org/10.1038/s41598-020-68662-3) (2020).
13. Kasieczka, G. *et al.* The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics. *Reports on Prog. Phys.* **84**, 124201, DOI: [10.1088/1361-6633/ac36b9](https://doi.org/10.1088/1361-6633/ac36b9) (2021).
14. Govorkova, E. *et al.* Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider. *Nat. Mach. Intell.* **4**, 154–161, DOI: [10.1038/s42256-022-00441-3](https://doi.org/10.1038/s42256-022-00441-3) (2022).

15. Allier, C. *et al.* Imaging of dense cell cultures by multiwavelength lens-free video microscopy. *Cytom. Part A* **91**, 433–442, DOI: [10.1002/cyto.a.23079](https://doi.org/10.1002/cyto.a.23079) (2017).
16. Allier, C. *et al.* Lens-free Video Microscopy for the Dynamic and Quantitative Analysis of Adherent Cell Culture. *J. Vis. Exp. JoVE* 56580, DOI: [10.3791/56580](https://doi.org/10.3791/56580) (2018).
17. Allier, C. *et al.* Quantitative phase imaging of adherent mammalian cells: A comparative study. *Biomed. Opt. Express* **10**, 2768–2783, DOI: [10.1364/BOE.10.002768](https://doi.org/10.1364/BOE.10.002768) (2019).
18. Barer, R. Interference Microscopy and Mass Determination. *Nature* **169**, 366–367, DOI: [10.1038/169366b0](https://doi.org/10.1038/169366b0) (1952).
19. Hervé, L. *et al.* Alternation of inverse problem approach and deep learning for lens-free microscopy image reconstruction. *Sci. Reports* **10**, 20207, DOI: [10.1038/s41598-020-76411-9](https://doi.org/10.1038/s41598-020-76411-9) (2020).
20. Tinevez, J.-Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods (San Diego, Calif.)* **115**, 80–90, DOI: [10.1016/j.ymeth.2016.09.016](https://doi.org/10.1016/j.ymeth.2016.09.016) (2017).
21. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).
22. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning*. Springer Series in Statistics (Springer, New York, NY, 2009).
23. Kolesnikov, A., Zhai, X. & Beyer, L. Revisiting Self-Supervised Visual Representation Learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1920–1929, DOI: [10.1109/CVPR.2019.00202](https://doi.org/10.1109/CVPR.2019.00202) (IEEE, Long Beach, CA, USA, 2019).
24. Chen, T., Kornblith, S., Swersky, K., Norouzi, M. & Hinton, G. E. Big Self-Supervised Models are Strong Semi-Supervised Learners. *Adv. Neural Inf. Process. Syst.* **33**, 22243–22255 (2020).
25. Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning, ICML '08*, 1096–1103, DOI: [10.1145/1390156.1390294](https://doi.org/10.1145/1390156.1390294) (Association for Computing Machinery, New York, NY, USA, 2008).
26. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. & Manzagol, P.-A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *The J. Mach. Learn. Res.* **11**, 3371–3408 (2010).
27. Gidaris, S., Singh, P. & Komodakis, N. Unsupervised representation learning by predicting image rotations. In *ICLR 2018* (Vancouver, Canada, 2018).
28. Larsson, G., Maire, M. & Shakhnarovich, G. Learning Representations for Automatic Colorization. *arXiv:1603.06668 [cs]* (2017). [1603.06668](https://arxiv.org/abs/1603.06668).
29. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. & Efros, A. A. Context Encoders: Feature Learning by Inpainting. *arXiv:1604.07379 [cs]* (2016). [1604.07379](https://arxiv.org/abs/1604.07379).
30. Jenni, S. & Favaro, P. Self-Supervised Feature Learning by Learning to Spot Artifacts. *arXiv:1806.05024 [cs]* (2018). [1806.05024](https://arxiv.org/abs/1806.05024).
31. McCloskey, M. & Cohen, N. J. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In Bower, G. H. (ed.) *Psychology of Learning and Motivation*, vol. 24, 109–165, DOI: [10.1016/S0079-7421\(08\)60536-8](https://doi.org/10.1016/S0079-7421(08)60536-8) (Academic Press, 1989).
32. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q. (eds.) *Advances in Neural Information Processing Systems 25*, 1097–1105 (Curran Associates, Inc., 2012).
33. Kiranyaz, S., Ince, T. & Gabbouj, M. Personalized Monitoring and Advance Warning System for Cardiac Arrhythmias. *Sci. Reports* **7**, 1–8, DOI: [10.1038/s41598-017-09544-z](https://doi.org/10.1038/s41598-017-09544-z) (2017).
34. Li, D., Zhang, J., Zhang, Q. & Wei, X. Classification of ECG signals based on 1D convolution neural network. In *2017 IEEE 19th International Conference on E-Health Networking, Applications and Services (Healthcom)*, 1–6, DOI: [10.1109/HealthCom.2017.8210784](https://doi.org/10.1109/HealthCom.2017.8210784) (2017).
35. Abdeljaber, O. *et al.* 1-D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data. *Neurocomputing* **275**, 1308–1317, DOI: [10.1016/j.neucom.2017.09.069](https://doi.org/10.1016/j.neucom.2017.09.069) (2018).
36. Avci, O., Abdeljaber, O., Kiranyaz, S. & Inman, D. Structural Damage Detection in Real Time: Implementation of 1D Convolutional Neural Networks for SHM Applications. In Niezrecki, C. (ed.) *Structural Health Monitoring & Damage Detection, Volume 7*, Conference Proceedings of the Society for Experimental Mechanics Series, 49–54, DOI: [10.1007/978-3-319-54109-9_6](https://doi.org/10.1007/978-3-319-54109-9_6) (Springer International Publishing, Cham, 2017).

37. Eren, L., Ince, T. & Kiranyaz, S. A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier. *J. Signal Process. Syst.* **91**, 179–189, DOI: [10.1007/s11265-018-1378-3](https://doi.org/10.1007/s11265-018-1378-3) (2019).
38. Ince, T., Kiranyaz, S., Eren, L., Askar, M. & Gabbouj, M. Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks. *IEEE Transactions on Ind. Electron.* **63**, 7067–7075, DOI: [10.1109/TIE.2016.2582729](https://doi.org/10.1109/TIE.2016.2582729) (2016).
39. Khan, A., Ko, D.-K., Lim, S. C. & Kim, H. S. Structural vibration-based classification and prediction of delamination in smart composite laminates using deep learning neural network. *Compos. Part B: Eng.* **161**, 586–594, DOI: [10.1016/j.compositesb.2018.12.118](https://doi.org/10.1016/j.compositesb.2018.12.118) (2019).
40. Zhang, W., Li, C., Peng, G., Chen, Y. & Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **100**, 439–453, DOI: [10.1016/j.ymssp.2017.06.022](https://doi.org/10.1016/j.ymssp.2017.06.022) (2018).
41. van den Oord, A. *et al.* WaveNet: A generative model for raw audio. *CoRR* **abs/1609.03499** (2016). [1609.03499](https://arxiv.org/abs/1609.03499).
42. van den Oord, A. *et al.* Conditional Image Generation with PixelCNN Decoders. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I. & Garnett, R. (eds.) *Advances in Neural Information Processing Systems* 29, 4790–4798 (Curran Associates, Inc., 2016).
43. Bailly, R., Malfante, M., Allier, C., Ghenim, L. & Mars, J. Deep anomaly detection using self-supervised learning: Application to time series of cellular data. In *ASPAI 2021 - 3rd International Conference on Advances in Signal Processing and Artificial Intelligence* (2021).
44. Tartour, K. *et al.* Mammalian PERIOD2 regulates H2A.Z incorporation in chromatin to orchestrate circadian negative feedback. *Nat. Struct. & Mol. Biol.* **29**, 549–562, DOI: [10.1038/s41594-022-00777-9](https://doi.org/10.1038/s41594-022-00777-9) (2022).
45. Cui, Z., Chen, W. & Chen, Y. Multi-Scale Convolutional Neural Networks for Time Series Classification. *arXiv:1603.06995 [cs]* (2016). [1603.06995](https://arxiv.org/abs/1603.06995).
46. Landry, J. J. M. *et al.* The Genomic and Transcriptomic Landscape of a HeLa Cell Line. *G3 Genes|Genomes|Genetics* **3**, 1213–1224, DOI: [10.1534/g3.113.005777](https://doi.org/10.1534/g3.113.005777) (2013).

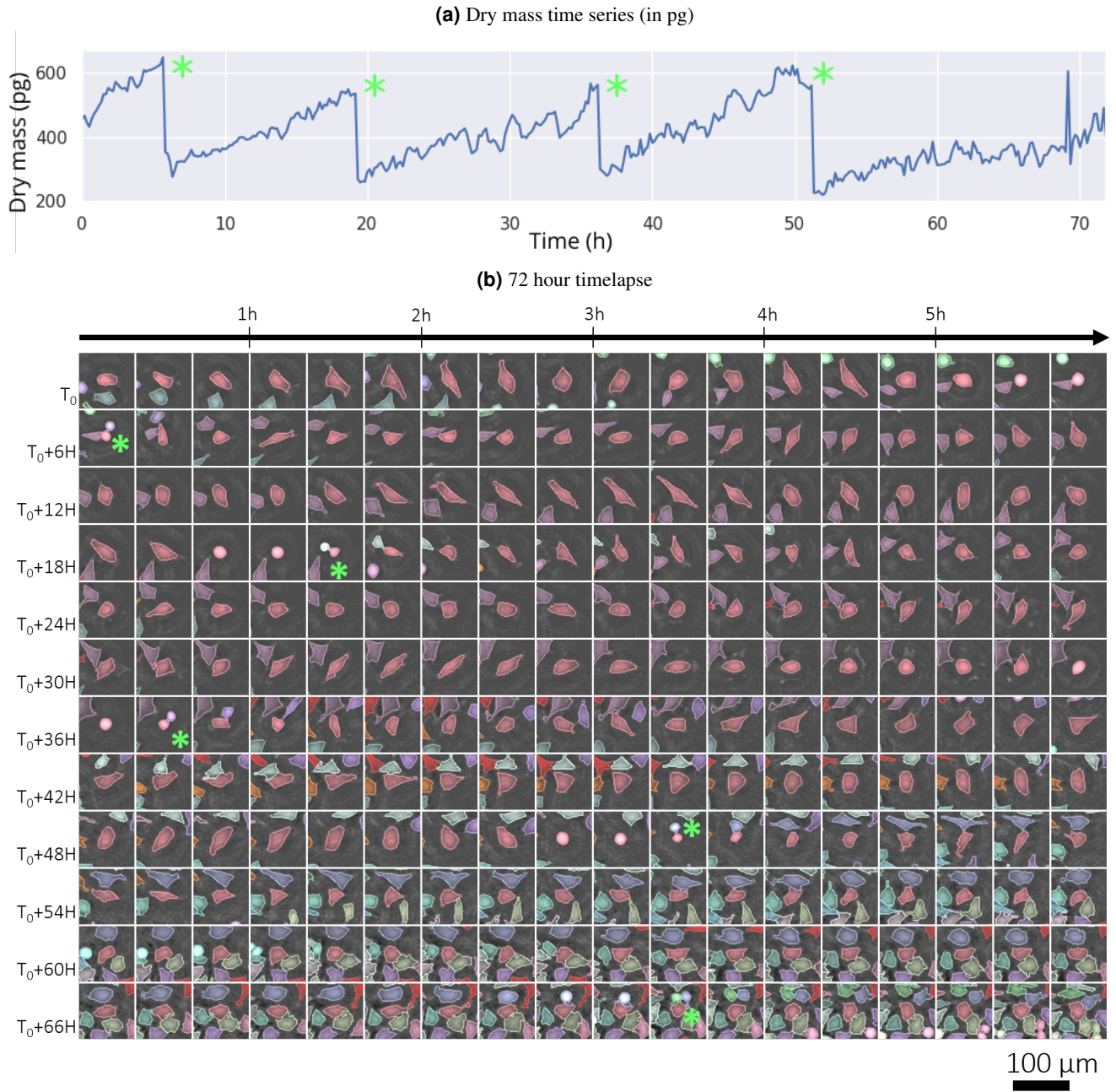


Figure 1. (a) Dry mass measurements (in pg) for the cell of interest as a function of time. (b) 72 hours time-lapse acquisition of a HeLa cell. Cell tracking and cell segmentation are computed together to obtain this time-lapse series. Each cell successfully tracked is depicted with a different color. Every cropped image is $100 \times 100 \mu m^2$. The cell of interest is centered in the cropped image (light red). Four green asterisks point cell divisions. For readability purposes, the time between two images is 20 minutes, even if the acquisition rate is one image every 10 minute. The four cell divisions are associated with a decrease in cell dry mass by a factor of two.

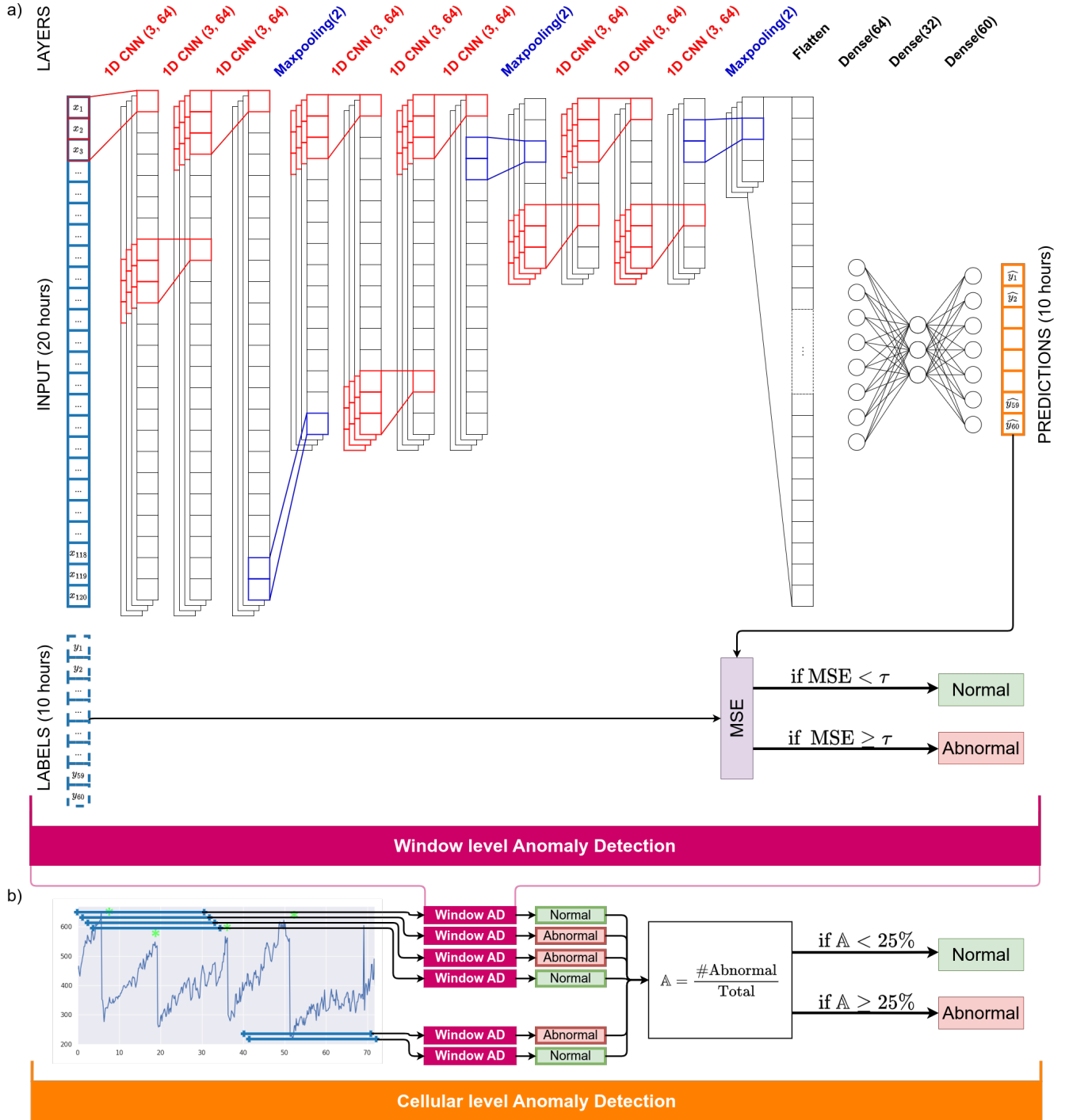


Figure 2. Full scheme of the StArDusTS model. a) shows the window-wise anomaly detection based on the representation extracted with a 1D-CNN. b) shows the whole StArDusTS model for a cell, including the 1D-CNN representation learning block, the window-wise anomaly detection and the aggregation of window-wise results with the anomaly score.

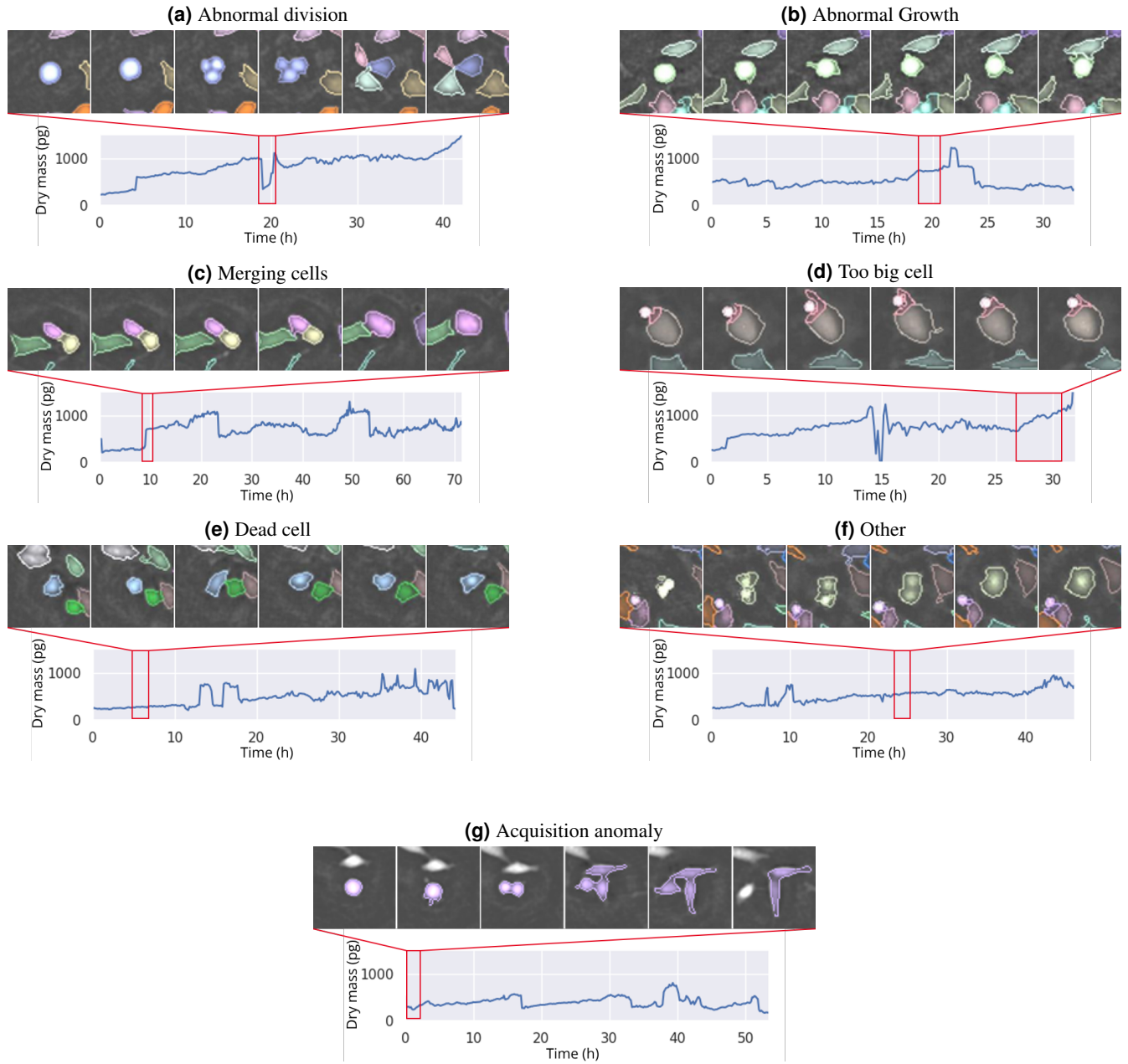


Figure 3. Time-lapse acquisition of adherent cells (Hela) that were detected as abnormal by StArDusTS. Each subplot depict an example of a subclass of anomalies. Every cropped image is $100 \times 100 \mu m^2$. The time between two images is 20 minutes. Cell tracking and cell segmentation are computed together to obtain time-lapse series. Each cell successfully tracked is depicted with a different color. The cell of interest is centered in the cropped image. Under each cell, the dry mass (in pg) time series from which the anomalies have been detected is printed. The time steps corresponding to the pictures are framed in red.

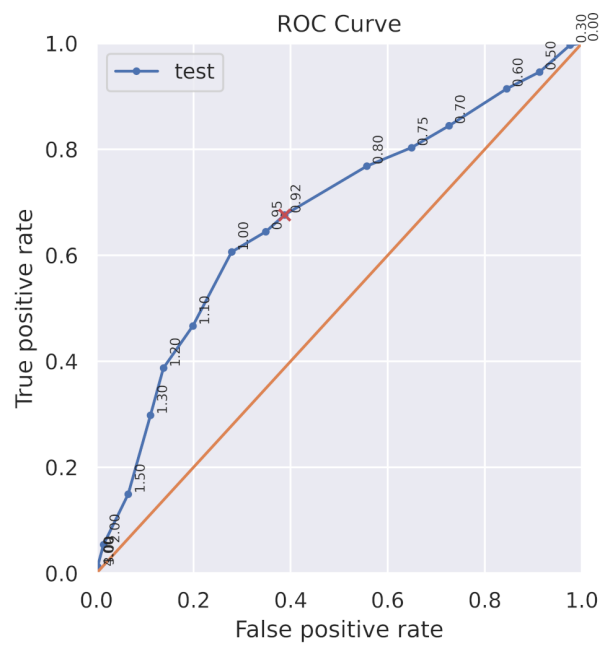


Figure 4. ROC curve of the anomaly detection on datasets \mathcal{C} and \mathcal{D} . Each point is a threshold value of τ_w for the consideration of a window as abnormal. The red cross is the τ_w value computed from the training dataset \mathcal{C} such that the time series outside the 95% interval of confidence are abnormal.