Contents lists available at ScienceDirect

SLAS Technology

journal homepage: www.elsevier.com/locate/slast

Technical Brief

Towards an automated approach for smart sterility test examination

Clemens Dierks [a,c], Robert Söldner [a], Kevin Prühl [b], Niels Wagner [b], Noushin Delmdahl [b], Andreas Dominik [c], Michael W. Olszowy [d], Jonas Austerjost [a,*]

[a] Corporate Research, Sartorius Stedim Biotech GmbH, Germany
[b] Product Development Lab Consumables, Sartorius Lab Instruments GmbH & Co. KG, Germany
[c] Technische Hochschule Mittelhessen - University of Applied Sciences, Germany
[d] Corporate Research, Sartorius Stedim North America Inc., USA

ABSTRACT

As new technologies emerge, deep learning applications are often integral parts of new products as features and often as differentiating benefits. This is especially notable in commercial consumer products in everyday applications, such as voice assistants or streaming content recommendation systems. Due to the power and applicability of these deep learning technologies significant efforts are being directed to the development and integration of appropriate models into science and engineering applications to supplant analogue systems that may be highly prone to human error. Here we present an innovative, low-cost approach to advance sterility assessment workflows that are required and regulated within drug release/manufacturing processes. The model system leverages off-the-shelf hardware as well as deep learning models to detect and classify different microbial contaminations in test containers. The paired hardware and software tools were evaluated in experiments using common model organisms (*C. sporogenes*, *P. aeruginosa*, *S. aureus*). With this approach we were able to detect all three test organisms across 40 experiments, furthermore we were capable of classifying the present organisms with an average classification accuracy of over 87%.

Introduction

Good Manufacturing Practices (GMP) are applied across industries to protect consumers. For products directed to supporting human health, sterility testing is critical to ensure the products are free from contaminating microorganisms [1]. Sterility testing methods are used widely in the food and beverage industry, as well as the pharmaceutical and medical industries. The regulatory groundwork that covers product sterility testing is articulated within the United States Pharmacopeia (USP) <71>. This compendium has been harmonized mainly with the European and Japanese pharmacopeias [2–4]. According to the USP <71>, membrane filtration is one method that is prescribed for filterable pharmaceutical products as it allows screening of large sample volumes. The underlying concept of the membrane filtration method relies in part on the product batch being filtered through microorganism-retentive filters that are then incubated in Fluid Thioglycollate Medium (FTM) as well as Soybean–Casein Digest Medium (SCDM) at 30-35˚C and 20-25˚C, respectively. This is done to accommodate different types of organisms and their unique growth requirements. The respective samples are in-

cubated for at least 14 days and visually inspected at prescribed intervals by highly qualified and trained personnel for signs of growth by looking for turbidity. When microbial growth is detected, the entire lot is maintained in quarantine until a detailed root cause analysis is completed.

Due to the long incubation times of the conventional sterility test, it is not always suitable for sterility testing of drugs with a short shelf life, such as gene and cell therapy products. To address the requirements of these new products, the USP informational chapter <1071> was released in 2018. This chapter proposes several methods for Rapid Microbial Testing (RMT) such as Adenosine Triphosphate (ATP) bioluminescence, flow cytometry, calorimetric change and nucleic acid amplification based on Real-Time Polymerase Chain Reaction (RT-PCR) [5–10].

Although these methods provide testing outcomes in a comparably shorter amount of time than the compendial methods outlined in USP <71>, culture-based methods remain the gold standard to examine product sterility in the biopharmaceutical industry [11]. One advantage over RMT methods is that they do not require expensive instrumentation. Nevertheless, culture interpretation is subjective and the manual
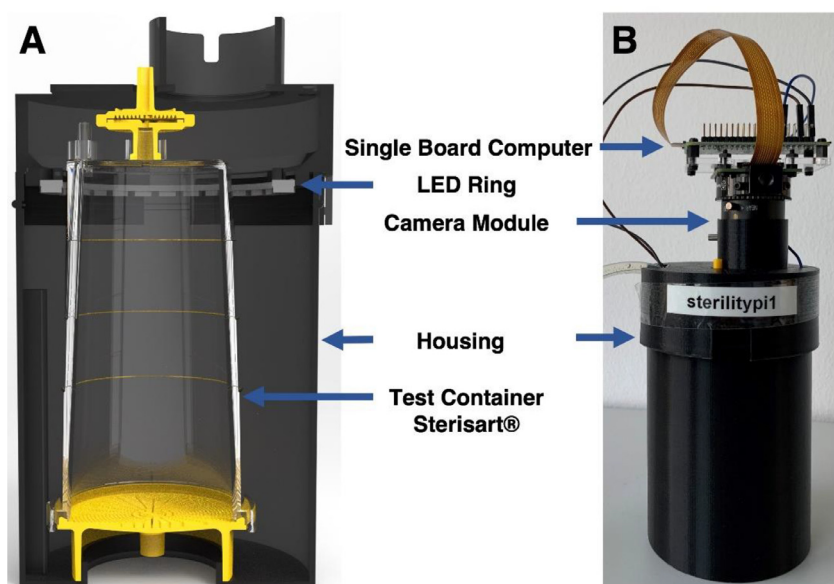
Fig. 1. Hardware components of the custom image acquisition prototypes: A and B depict the interior and exterior design, respectively, as well as the applied components.

visual inspection cycles are in contradiction with the general trend towards automated solutions in the biopharmaceutical industry [12].

Deep Learning (DL) components are notable parts of our everyday lives as they enable applications such as face recognition or natural language processing. DL approaches have become more accessible over the past decade and are now found in almost every science and engineering domain [13]. Applications of DL in biotechnology range from cell segmentation models, that support microscopy image analytics over image-based foam sensing in cultivation processes, to the control of biomanufacturing processes [14–16].

In this study, we present a hardware and software concept in support of the USP <71> membrane filtration method by providing an automated approach for image capture as well as deep learning methods to detect and classify contaminations in sterility testing units. The model system was implemented using off-the-shelf hardware for image material acquisition, as well as open-source software for model development and image data processing.

Image data of 3 different organisms (*Clostridium sporogenes, Pseudomonas aeruginosa, Staphylococcus aureus*) growing in sterility test containers, as well as negative controls, have been collected to develop a model for growth detection and a model for subsequent organism classification. These models were then applied on test data sets acquired by the same approach.

Materials & methods

Experimental setup

All sterility tests were performed in accordance with the regulatory requirements for sterility testing of pharmaceutical liquids [2,17]. In this study, sterile and contaminated pharmaceutical formulation samples were modeled with either sterile sodium phosphate buffer (Becton & Dickinson (BD), Franklin Lakes, NJ) or sodium phosphate buffer inoculated with one of the three validation organisms (*Clostridium sporogenes* ATCC 19404, *Pseudomonas aeruginosa* ATCC 9027, *Staphylococcus aureus* ATCC 6538; all acquired from BD), respectively. Here, the inoculation concentration ranged from 7 to 58 colony forming units (CFU). The sterile or inoculated buffer solutions were pumped out of each test container (Sterisart®, Sartorius Stedim Biotech, Göttingen, Germany) through integral 0.45 μm membrane filter discs. Through this initial step, pre-existing contaminants were retained by the membrane on the luminal side of each container. Each container was then refilled with fluid thio-

glycolate medium (FTM; BD). The Sterisart® systems were then placed into specially designed hardware fixtures and finally placed in an incubator (Thermo Fisher Scientific, Waltham, MA) at 32°C for up to 7 days. All medium/buffer pumping procedures were performed with the Sterisart® Universal Pump (Sartorius Stedim Biotech, Göttingen, Germany) to ensure a uniform fill level of 200 ml in the interest of cross culture reproducibility.

Image acquisition

To collect training and evaluation data for the development of the DL models, images of sterility tests were acquired by 6 matched prototypes (see Fig. 1). For this purpose, single-board computers (Raspberry Pi Zero W) were equipped with 12-MP HQ cameras (both Raspberry Pi Foundation, Cambridge, United Kingdom) for the acquisition of top-view RGB images (1640×1232 pixels) of the test vessels.

Prior to initial acquisition, the focal points of all camera systems were manually adjusted to the membrane surface within the Sterisart® test containers. Because all prototypes were deployed simultaneously in a dark incubator environment, separate illumination units (NeoPixel LED Ring 24, Adafruit, New York, NY) and a custom 3D-printed housing were introduced for each image acquisition unit. This was done to ensure consistent lighting conditions across each test unit. The prototypes were controlled via an embedded Node-RED (v1.0.6) workflow with a graphical user interface that allowed live monitoring of the experiment in progress as well as the adjustment of various acquisition parameters (exposure, acquisition interval, labeling). Images of the experiments were taken at 5-minute intervals (see Supplemental Material for time-lapse videos; Vid. A1-A3) and sent directly from the prototypes to a dedicated cloud storage service (Microsoft Azure Blob Storage, Microsoft Corporation, Redmond, WA) for further processing. In total, 46070 images were acquired in 40 experiments, of which 10,444 images were obtained from 10 sterile controls.

Model training

The acquired image data sets were used to train DL models for microbial growth detection and classification. For this purpose, two convolutional neural networks (CNNs) were developed. Initially, a convolutional autoencoder was trained in an unsupervised anomaly detection approach. Autoencoders are hourglass-shaped neural networks that consist of at least one input layer, a bottleneck layer with lower dimension-

ality, and usually an output layer with the same dimensions as the input layer [18]. During model training, the network was taught to reconstruct images from the compressed feature representation stored in the bottleneck layer. In the following, a second CNN was trained to classify the microbial growth present in the captured images.

Software for image processing and model development was implemented using the libraries OpenCV (v. 4.5.3) and PyTorch (v.1.10; Facebook Research, Menlo Park, CA) in the Python programming language (v. 3.8) [19,20]. Both models were trained on a virtual machine in the Azure Machine Learning (Microsoft Corporation, Redmond, WA) environment using a NVIDIA Tesla K80 graphics processing unit with an Intel Xeon E5-2690 v3 processor.

Before model training, all images were rescaled to 492×369 pixels and cropped to fit the area of the filter membrane, resulting in a resolution of 180×180 pixels. While a market-ready application would need to ensure a full view into the test container, the centric position of the venting filter prevents full sight. Therefore, the filter membrane was identified as the region of interest because it serves as the best model for simulating the background of a full view application. Furthermore, its circular shape allows for random rotations (360°) without inducing major artefacts, which were combined with slight shearing (5° in x and y direction) and structural changes like random changes in brightness and saturation (10%) to increase the variance of the data set. These augmentations were applied on-the-fly during the training of both proposed models.

The autoencoder was trained to minimize the reconstruction loss of images from negative control runs. Every 10th image of 5 control runs was selected for training, resulting in a total of 1000 images over several days. This temporal variance was crucial because the reduction-oxidation indicator resazurin, which is present in the cell culture media used, triggers a color change to pink. The implemented model uses 3 convolutional layers with max pooling to compress the input information and forces the model to find a latent space representation to store its most notable features. Three transposed convolutional layers were used to unfold this representation into the original image size (180×180 pixels). To calculate the distance, also called reconstruction error (RE), between the original image and the reconstruction, the mean-squared error (MSE) was used as a loss function (see Eq. 1).

$$MSE = \frac{1}{3 * h * w} \sum_{c=1}^{3} \sum_{y}^{h} \sum_{x}^{w} (p(x, y) - \hat{p}(x, y))^2 \tag{1}$$

Here, h and w represent the height and the width of the image, respectively. p and $\hat{p}$ are the input and reconstructed output, that are presented as a 3-dimensional matrix with x and y coordinates and 3 color channels (c). The model was trained with a batch size of 50 and a learning rate of $1e^{-3}$, which was constantly updated by the Adam optimizer [21]. Additionally, an early stopping mechanism was implemented to prevent the model from overfitting on the training data. After 112 epochs, the lowest reconstruction error of $3.2e^{-3}$ was achieved and the corresponding model was used for evaluation.

Microbial growth classification was performed using a ResNet34 connected to a classifier consisting of two fully connected layers to assign the extracted features to the 3 investigated microorganisms [22]. Because the classification is to be performed immediately after detection, the training and test dataset was composed of images that had a RE greater than $5e^{-3}$, but were acquired no later than 4 hours after this detection limit. However, this selection resulted in a balanced training data set of only 5325 training images, making a transfer learning approach by using the pretrained weights on the ImageNet data set necessary [23]. The model was trained using cross-entropy loss and a batch size of 100 images. The initial learning rate of $1e^{-3}$ was optimized by AdamW and continuously reduced every 7 epochs by a factor of 0.1 [24]. Before training, one third of the training data set was excluded class-wise at random and used for on-the-fly validation. The best model performance on the validation data was achieved after 33 epochs as indicated by the
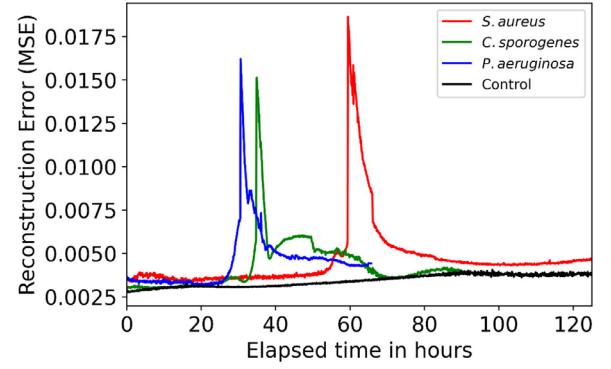

Fig. 2. Reconstruction Error/time series plot of trials with all validation organisms and a control run show the general applicability of the proposed autoencoder model to detect contaminations in sterility testing. The initial inoculated concentrations of *C. sporogenes*, *P. aeruginosa* and *S. aureus* for the experiments shown were 7, 48 and 34 CFU, respectively.

loss plot. Loss plots of model trainings are displayed in the appendix (see supplemental material Fig. A9 and A10).

Model validation

The predictive power of the detection model was evaluated using time-series plots and heat maps showing the temporal progression of the reconstruction error during a sterility test and the image regions that contribute most to its increase, respectively. The heat maps were generated using the method adapted from Chow *et al.* [25]. In this way, the pixel-wise, squared difference of the original image and the reconstructed output was visualized. The image data sets of 30 different sterility tests performed were used for model evaluation. Subsequently, the model was validated with image data sets of the control runs excluded from the training data.

The classification model was evaluated on a test data set containing 1617 images selected the same way as the training data, but from different experiments. The 3 classes *C. sporogenes*, *P. aeruginosa* and *S. aureus* contained 700, 500 and 417 images, respectively. Due to the imbalanced class distribution, weighted accuracy and weighted F1-scores were used to assess the model performance, which are defined as follows:

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$F1 = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{4}$$

$$Weighted\ Accuracy = \sum_{c=1}^{3} w_c * \left(\frac{(TP + TN)}{N_c} \right) \tag{5}$$

$$Weighted\ F1 - Score = \sum_{c=1}^{3} w_c * F1_c \tag{6}$$

Here, TP, TN, FP, and FN correspond to the true positive/negative and false positive/negative values, respectively. To compensate for the bias due to unbalanced class sets c, the weighted accuracy and F1 score were used. Accuracy and F1 score were multiplied by the proportion of the class w_c in the total number of samples per N_c.

Results & discussion

This study proposes a concept for automated and digitized support for compendial sterility testing. First, a hardware prototype was developed to collect reproducible imagery of the experiments performed. Second, a convolutional autoencoder was trained to detect morphological
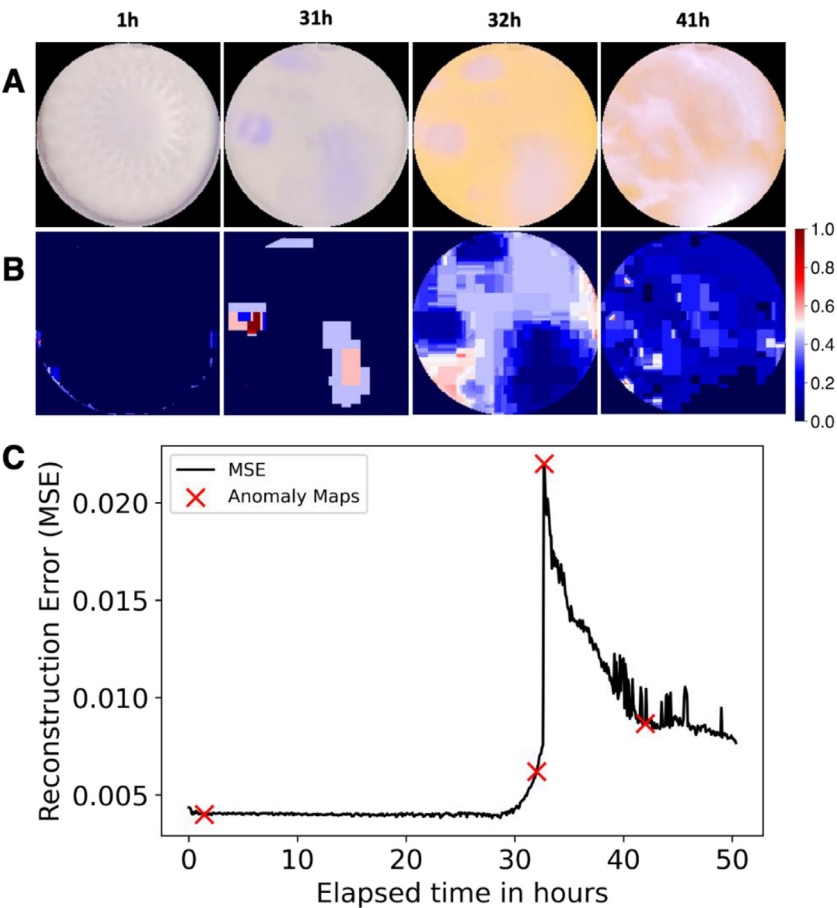
Fig. 3. A: Region of interest acquired for exemplary P. aeruginosa batch over specific time points (1h, 31h, 32h and 41h after inoculation). B: Resulting anomaly maps of the same batch. The matrix underlying the anomaly maps was normalized [0, 1], with red and blue representing high and low reconstruction error, respectively. C: Reconstruction Error/time plot with marked time points that were used for anomaly map generation.
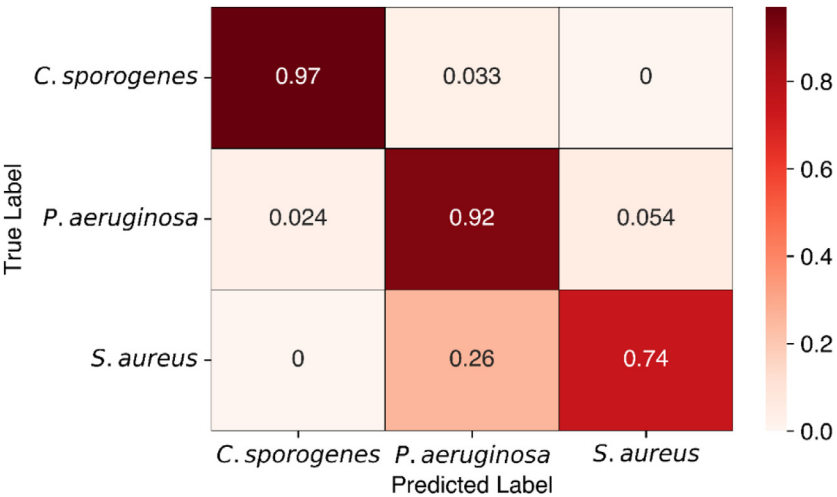

Fig. 4. Confusion matrix of the proposed classification model.

changes within the test container. Finally, a second model was developed to classify the microbial growth visible in the acquired images using three exemplary model organisms.

Microbial growth of all organisms investigated was detected in all 30 inoculated experiments within the first 72h. Fig. 2 shows the time course of the reconstruction error (RE) for all three validation organisms during exemplary experiments. It can be observed that the RE of the sterile control run (black) stays relatively constant between 2.5 e^{-3} and 5e^{-3}, while the REs of the inoculated experiments (blue: *P. aeruginosa*, green: *C. sporogenes,* and red: *S. aureus*) peak at approximately 31, 38 and 61 hours, respectively, with an error up to 1.8e^{-2}. The decrease in RE cor-

relates with the spread of whitish growth patterns in later stages of the experiments, as these are more similar to the training data showing the white membrane of the test vessel than the intermediate growth stages (see Fig. 3A).

The temporal difference of the error plots between the organisms can be explained by the different growth rates and inoculation concentrations. The reconstruction error of different time points is visualized in Fig. 3C, that shows a time-series plot of an exemplary *P. aeruginosa* batch and corresponding heat maps. The displayed heat or anomaly maps are based on the squared pixel-based difference between the original input image and the reconstructed output. The calculated difference was

subsequently normalized between [0,1] [25]. Comparison of the original images with the corresponding heatmaps show that the image regions with growth patterns contribute most to the increase of the RE (see Fig. 3B). The supplementary material includes videos with the graphs shown in Fig. 3 of experiments with all three validation organisms. Additionally, a video and a figure of a sterile sample are included (see supplemental material Vid. A4-A7 and Fig. A8).

Furthermore, the developed classification model showed strong performance on an independent test set, yielding a weighted accuracy and weighted F1-score of 88% and 90%, respectively. This indicates reliable determination of the detected contamination by the developed MV application.

The confusion matrix (see Fig. 4) shows accuracies of beyond 90% for *C. sporogenes* and *P. aeruginosa*, however, only 74% for *S. aureus*. Due to the high nonlinearity of the model, the lower predictive performance for *S. aureus* cannot be fully explained; but it might be based on the color and structural similarity of the image material collected of *P. aeruginosa* and *S. aureus* (see supplemental material Vid. A2 & A3).

Concluding remarks & outlook

Sterility testing is a GMP requirement in the production process of pharmaceutical products. The compendial method for sterility testing of pharmaceuticals is a highly manual approach, requiring trained and skilled personnel. In addition, the manual visual examination of sterility testing units is a subjective and time-consuming task. In this study, we conceptualized a cost-efficient hard- and software setup that appears to overcome these disadvantages by providing an objective and automated supporting tool for sterility testing. The hardware prototype has proven to be a useful system to acquire image material in incubator environments. For the three evaluation organisms (*C. sporogenes, P. aeruginosa* and *S. aureus)*, the models demonstrated high sensitivity for the detection of microbial growth as well as robust performance for the classification of the respective organism present in each test unit.

Due to the agnostic nature of the presented detection approach, its general applicability to different organisms can be assumed. However, although our approach shows promising results in terms of the general classification ability of CNNs for macroscopic images of growth in a liquid environment; further experiments with different organisms, including organisms with similar morphologies and respective strains, need to be performed to show if the approach is applicable in real test environments. Therfore, future work includes data acquisition and model building for additional common microbial organisms that might be present, as well as data acquisition and model implementation for cross-contamination. In addition, based upon the flexible software architecture, we are planning to integrate and connect the hardware and respective models to an alert system that would notify the user as soon as a contamination advances.

In conclusion, this proof of concept represents the foundational work for the implementation of a machine vision workflow in sterility testing using state-of-the-art deep learning techniques. The ultimate aim is to streamline a subjective and manual process by transforming it into an objective and automated workflow that allows for ultimate integration into smart laboratory environments.

Funding

Data availability statement

The models generated during the current study are available from the corresponding author on reasonable request.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jonas Austerjost has patent pending to Sartorius Stedim Biotech GmbH. Robert Söldner has patent pending to Sartorius Stedim Biotech GmbH.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.slast.2022.09.005.

References

[1] Gouveia BG, Rijo P, Gonçalo TS, Reis CP. Good manufacturing practices for medicinal products for human use. J Pharmacy Bioallied Sci 2015;7(2):87–96. doi:10.4103/0975-7406.154424.
[2] United States Pharmacopeia Sterility Tests (USP 71). Bethesda, MD; 2020.
[3] European Pharmacopeia Biological Tests (EP 2.6.1); 2008.
[4] Japanese Pharmacopeia Sterility Tests (JP 4.06); 2016.
[5] United States Pharmacopeia Rapid Sterility Testing (USP 1071); 2019.
[6] Bugno A, Almodovar AAB, Saes DPS, et al. Evaluation of an amplified ATP bioluminescence method for rapid sterility testing of large volume parenteral. J Pharma Innov 2019;14(2):152–8. doi:10.1007/s12247-018-9344-y.
[7] Chollet R, Kukuczka M, Halter N, et al. Rapid detection and enumeration of contaminants by ATP bioluminescence using the milliflex® rapid microbiology detection and enumeration system. J Rapid Methods Autom Microbiol 2008;16(3):256–72. doi:10.1111/j.1745-4581.2008.00132.x.
[8] Mohr H, Lambrecht B, Bayer A, et al. Basics of flow cytometry-based sterility testing of platelet concentrates. Transfusion 2006;46(1):41–9. doi:10.1111/j.1537-2995.2005.00668.x.
[9] Fricke C, Harms H, Maskow T. How to speed up the detection of aerobic microbial contaminations by using isothermal microcalorimetry. J Therm Anal Calorim 2020;142(5):1933–49. doi:10.1007/s10973-020-09986-0.
[10] De Boer E, Beumer RR. Methodology for detection and typing of foodborne microorganisms. Int J Food Microbiol 1999;50(1–2):119–30. doi:10.1016/S0168-1605(99)00081-1.
[11] England MR, Stock F, Gebo JET, Frank KM, Lau AF. Comprehensive evaluation of compendial USP<71>, BACT/ALERT DUAl-T, and BACTEC FX for detection of product sterility testing contaminants. J Clin Microbiol 2019;57(2). doi:10.1128/JCM.01548-18.
[12] Bremme L, Darino L, Parry B, Teo K. Automation and the future of work in the US biopharma industry | McKinsey 2020. https://www.mckinsey.com/industries/life-sciences/our-insights/automation-and-the-future-of-work-in-the-us-biopharma-industry (accessed January 10, 2022).
[13] Emmert-Streib F, Yang Z, Feng H, Tripathi S, Dehmer M. An introductory review of deep learning for prediction models with big data. Front Artif Intell 2020;3. doi:10.3389/frai.2020.00004.
[14] von Chamier L, Laine RF, Jukkala J, et al. Democratising deep learning for microscopy with ZeroCostDL4Mic. Nat Commun 2021;12(1):1–18. doi:10.1038/s41467-021-22518-0.
[15] Austerjost J, Söldner R, Edlund C, Trygg J, Pollard D, Sjögren R. A machine vision approach for bioreactor foam sensing. SLAS Technology 2021;26(4):408–14. doi:10.1177/24726303211008861.
[16] Treloar NJ, Fedorec AJH, Ingalls B, Barnes CP. Deep reinforcement learning for the control of microbial co-cultures in bioreactors. PLoS Comput Biol 2020;16(4):e1007783. doi:10.1371/JOURNAL.PCBI.1007783.
[17] European Pharmacopeia. Biological Tests (EP 2.6.1). 2008.
[18] Zheng J, Du J, Liang Y, et al. Deeppipe: A semi-supervised learning for operating condition recognition of multi-product pipelines. Process Saf Environ Prot 2021;150:510–21. doi:10.1016/j.psep.2021.04.031.
[19] Gary B. The OpenCV Library. Dr Dobb's J Softw Tools 2008;25(2236121):120–3.
[20] Paszke A, Gross S, Massa F, et al. PyTorch: An imperative style, high-performance deep learning library Advances in Neural Information Processing Systems, 32; 2019.
[21] Kingma DP, Ba JL. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings; 2015.
[22] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society; 2016. p. 770–8. 2016- Decem. doi:10.1109/CVPR.2016.90.
[23] Deng Jia, Dong Wei, Socher Richard, Li Li-Jia, Li Kai, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition; 2009. p. 248–55.
[24] Loshchilov I, Hutter F. Decoupled weight decay regularization. 7th International Conference on Learning Representations, ICLR 2019; 2019.
[25] Chow JK, Su Z, Wu J, Tan PS, Mao X, Wang YH. Anomaly detection of defects on concrete structures with the convolutional autoencoder. Adv Eng Inf 2020;45:101105. doi:10.1016/j.aei.2020.101105.