Expert Systems With Applications 299 (2026) 130302

Contents lists available at ScienceDirect

Expert Systems With Applications

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

VAE-MOTION: A deep generative model for cardiomyocyte contractility analysis for improving drug efficacy evaluation

Giorgia Curci [a,b,*], Paola Casti [a,b], Luca Sala [c,d], Marcella Brescia [e], Pasquale Cascarano [f], Michele D'Orazio [a,b], Joanna Filippi [a,b], Gianni Antonelli [a,b], Arianna Mencattini [a,b], Massimo Mastrangeli [g], Berend J. van Meer [e,h], Eugenio Martinelli [a,b]

[a] Department of Electronic Engineering, University of Rome Tor Vergata, Rome, Italy
[b] Interdisciplinary Center for Advanced Studies on Lab-on-Chip and Organ-on-Chip Applications (ICLOC), Via del Politecnico 1, 00133 Rome, Italy
[c] Department of Biotechnology and Biosciences, University of Milano – Bicocca, Milan, Italy
[d] Istituto Auxologico Italiano IRCCS, Center for Cardiac Arrhythmias of Genetic Origin and Laboratory of Cardiovascular Genetics, Milan, Italy
[e] Department of Anatomy & Embryology, Leiden University Medical Center, The Netherlands
[f] Department of the Arts, University of Bologna, Bologna, Italy
[g] Microelectronics Deparment, Delft University of Technology, Delft, The Netherlands
[h] Sync Biosystems, Leiden, The Netherlands

ARTICLE INFO

Keywords:
Variational autoencoders
Time-lapse microscopy
Contraction analysis
Data restoration

ABSTRACT

Deep learning has proven to be one of the most effective methods in analyzing biological images to extract parameters fundamental for studying physiological functions and pathological conditions. In particular, when coupled with time-lapse microscopy (TLM), deep learning proves particularly effective in studying behaviors involving temporal dynamics. However, TLM videos are often affected by experimental noise and setup limitations, which can lead to inaccurate and poorly reproducible results. Taking advantage of the variational and generative capabilities of Variational Autoencoders (VAEs), we propose VAE-MOTION, a deep learning-based model for the analysis of cardiac contractile dynamics. By incorporating a temporal encoder into its architecture, our model allows the restoration of video quality by removing noise or increasing resolution, while simultaneously extracting accurate contraction-related signals from the latent space. The generation of synthetic videos allowed extensive training of VAE-MOTION, which subsequently validated on real videos from two different cardiac tissue models: 2D monolayers and 3D microtissues. VAE-MOTION was compared to two gold-standard methods in extracting contraction parameters relevant to drug efficacy or toxicity studies, demonstrating its potential for analyzing temporal dynamics in a given phenomenon or process.

1. Introduction

Time-lapse microscopy (TLM) videos of in-vitro cell culture experiments hold great potential for providing breakthrough insights into disease mechanisms and yielding advances in the development of pharmacological compounds (Ascione et al., 2014; Antonelli et al., 2024). TLM enables the observation and analysis of cellular responses to stimuli and processes over time and it can provide critical insights into dynamic biological processes when combined with well-suited image analysis algorithms (Strbkova et al., 2020; Mencattini et al., 2021). Machine learning, in particular, is increasingly being used in video analysis for its unbiased contribution to evaluate the content of such videos, resulting in more sensitive, consistent, and accurate methods (D'Orazio et al., 2022). However, the quality of the collected videos may be influenced by several constraints related to the experimental platform, such as microscope type, illumination, and camera specifications. The video acquisition process can undermine the quality of obtained data, often compromised by artifacts and other kinds of degradation

* Corresponding author at: Department of Electronic Engineering, University of Rome Tor Vergata, Via del Politecnico 1, 00133 Roma, Italy.
E-mail addresses: giorgia.curci@uniroma2.it (G. Curci), casti@ing.uniroma2.it (P. Casti), luca.sala@auxologico.it (L. Sala), M.Dias_Brescia@lumc.nl (M. Brescia), pasquale.cascarano2@unibo.it (P. Cascarano), michele.d.orazio@uniroma2.it (M. D'Orazio), filippi@ing.uniroma2.it (J. Filippi), g.antonelli@ing.uniroma2.it (G. Antonelli), mencattini@ing.uniroma2.it (A. Mencattini), M.Mastrangeli@tudelft.nl (M. Mastrangeli), berend.van.meer@demcon.com (B.J. van Meer), martinelli@ing.uniroma2.it (E. Martinelli).

https://doi.org/10.1016/j.eswa.2025.130302
Received 7 January 2025; Received in revised form 21 October 2025; Accepted 2 November 2025
Available online 4 November 2025

(Mencattini, D'Orazio, et al., 2023). Moreover, high-resolution videos require significant storage and processing power for effective handling and analysis. These resources are not always available, resulting in a compromise between spatial and temporal resolutions. Deficits in spatial resolution might miss out on capturing detailed morphological features; conversely, deficits in temporal resolution might filter out fast dynamic changes. These aspects can result in incomplete or biased data, affecting the reliability of derived descriptors and quantitative measurements (Comes et al., 2019).

In this context, deep learning stands out as a powerful tool for TLM video analysis because of its ability to extract deep features and recognize high-complexity patterns that traditional algorithms cannot elicit (W. J. Zhang et al., 2018). Deep learning techniques automatically learn hierarchical feature representations from raw data, a capability that traditional machine learning approaches generally lack (LeCun et al., 2015). While classical machine learning methods rely heavily on handcrafted features, deep learning models can discover complex patterns and temporal dependencies directly from the input. Furthermore, when appropriately trained, deep learning models can effectively overcome disturbances and compensate for low spatio-temporal resolution, addressing many of the challenges posed by limitations in video quality and experimental setups (Yun et al., 2024). Numerous deep models have been successfully implemented to recognize, learn, and reduce noise tailored to the specific type and application (Tian et al., 2020). Many existing models are based on convolutional neural networks, CNNs (LeCun et al., 2015), which, with their flexible architecture, are very popular for image analysis and restoration (Cascarano et al., 2021; Qian et al., 2021). Zhang and coworkers (2017) proposed denoising CNN (dnCNN) with residual learning (He et al., 2016; Jifara et al., 2019) and batch normalization techniques (Ioffe & Szegedy, 2015) for image restoration. CNNs, originally developed as discriminative models, are fundamental in the recognition of specific functions for segmentation purposes, object recognition in images, and classification (Tian et al., 2020). Due to the challenges of obtaining appropriate experimental data mentioned above, defining methodologies to generate synthetic training data can be advantageous. For the purpose of providing new realistic images and data, generative models have been designed (Celard et al., 2023). One of the earliest examples were Generative Adversarial Networks (GANs), introduced in (Goodfellow et al., 2020). GANs consist of a generator, trained to produce images that closely resemble real ones, and a discriminator, tasked with distinguishing between real images and generated ones. This framework allows GANs to generate new data and, moreover, has proven to be effective in several applications such denoising (Chen et al., 2024; Yu & Ma, 2018) and super-resolution imaging (Comes et al., 2020, 2021). These capabilities are highly functional for biological investigation, where the accuracy of observations is crucial to avoid artifact-related misinterpretation. Controlling the impact of these limitations on cell imaging turned out to play a crucial role in understanding complex biological processes (Harrison & Baker, 2018; Mencattini, D'Orazio, et al., 2023).

In this context, when combined with accurate information extraction methods, TLM can be used as a pipeline for investigating biological functions (Filippi et al., 2024). Extracting significant features relies not only on a precise design of experimental setups but also on the strength of the algorithms used for the analysis, which process and codify the data.

From this perspective, Variational Autoencoders (VAEs) introduced by (Kingma & Welling, 2013), possess the capability of encoding and modeling cell properties in a latent space, which can be interpreted in light of the generated images. Indeed, compared to other generative models, VAEs learn to encode data distributions in a probabilistic structure, which enhances training stability, interpretability, and computational efficiency (Bond-Taylor et al., 2022; Sami & Mobin, 2019). These models can be the best compromise for computational costs and explainability when reconstructing data with specific characteristics. Furthermore, VAEs learn to encode high-dimensional input data, e.g. microscopy video frames, into a latent space and cluster similar patterns and behaviors together (Ehrhardt & Wilms, 2022). This helps to reduce the dimensionality and to focus on essential features from TLM data, thus enabling more precise and robust analysis (Casti et al., 2023; Mencattini, Casti et al., 2023). These characteristics are particularly relevant when addressing cell and tissue dynamics, such as migration tests or muscle contraction. Monitoring these phenomena is fundamental in assessing tissue function, identifying disease mechanisms, as well as evaluating safety and efficacy of pharmacological treatment. VAEs have already found applications in TLM studies, such as extracting features for image classification (Casti et al., 2023) and encoding cell shape while incorporating local tissue organization and temporal features to predict future states (Soelistyo et al., 2022). Additionally, VAEs have also been applied as image restoration tools (Zheng et al., 2022) or combined with long short-time memory (LSTM) networks for anomaly detection in mono-dimensional time series (Lin et al., 2020; Rong et al., 2022). However, to the best of our knowledge, no VAE-based model has been reported to date to codify dynamics information directly from TLM videos, extracting dynamic features while simultaneously addressing and counteracting eventual disturbances. In fact, while video restoration and denoising are classical problems, standard approaches typically require separate steps for noise reduction and feature extraction, which may fail when dealing with high-dimensional, temporally structured, and noisy TLM data. Therefore, there is a need for models capable of simultaneously handling noise and extracting dynamic information from the videos, ensuring both robustness and accuracy in the analysis of cellular behavior and other complex biological processes. In this work, we propose a novel VAE-based approach designed to improve dynamics evaluation in TLM videos tailored for cardiac contraction studies, providing a robust framework that can eventually be adapted also to other dynamics investigations.

1.1. Analysis of cardiomyocyte contractility

Studies on dynamics have become crucial in the characterization of contractile tissues, such as cardiomyocytes (CM) (H. Li et al., 2023). The heart's ability to adapt its contractile rhythm in response to different stimuli, known as inotropism, is an indispensable aspect of physiological regulation. Several in vitro study methodologies allow the replication of complex cardiac tissue physiology (Campostrini et al., 2021; Stiefbold et al., 2024). Among these, organoids and organ-on-chip technology are well-suited due to their ability to enable real-time observation of the evolution of fundamental features, miniaturize experimental setups, standardize and precisely reproduce complex models, such as cardiac developmental paths (Yang et al., 2024; Y. S. Zhang et al., 2015).

Accurate quantification of cyclic CM contraction and relaxation involves measuring specific parameters of the contraction dynamics, such as contraction amplitude, contraction velocity, relaxation velocity, and beat frequency. These measurements can allow the estimation of dose–response curves for drug testing, predicting the eventual cardiotoxic effects of specific drugs. Cell contraction can be monitored at multiple levels. At single-cell level, the patch clamp technique is often employed for electrophysiological recording of action potentials (Bers, 2002) and fluorescence assays to track key molecular concentrations involved in contraction, such as intracellular Ca^{2+} (Wang et al., 2002). At tissue level, contraction models involve measuring force during isometric or isotonic contraction to study contraction velocity and viscoelastic properties, e.g. through 3D pillar structures or in organ-on-chip devices with the integration of electrodes for stimulation or signal detection (Agrawal et al., 2017; T. Li et al., 2024). Combining these methods provides comprehensive insights into tissue contractile properties, as no gold standard exists. Numerous methodologies to analyze dynamic patterns from TLM videos have been implemented. Some require sophisticated hardware or software systems, whose the absence could undermine reproducibility across models and laboratories.

Alternatively, some existing software tools enable quantitative analysis of cardiac contraction dynamics, disease patterns, and drug responses by estimating specific parameters (Kim et al., 2024; Sala et al., 2017; Scalzo et al., 2021). Among the physical quantities to be evaluated in TLM videos to extract and track information about movement, one of the most commonly used is the gradient in light intensity between consecutive frames, analyzed pixel by pixel. Indeed, detecting how brightness varies over time makes it possible to use it as a tracking tool. Exploiting this methodology, Kim and collaborators (2024) proposed BeatProfiler, a suite of cardiac analysis tools designed to quantify contractile function, calcium handling, and force generation for multiple in vitro cardiac models and apply downstream machine learning methods for deep phenotyping and classification. The software tool CONTRACTIONWAVE, presented by Scalzo and his collaborators (2021), uses dense optical flow (Barron et al., 1994; Farnebäck, 2003) to obtain contractility parameters from images obtained with video microscopy. Some software tools exploit particle image velocimetry (PIV) algorithms (Thielicke & Sonntag, 2021) to calculate the displacement vector field between consecutive images within an interrogation window, e.g. PIVMyoMonitor, proposed by Lee and colleagues (2024). Sala and coworkers (2017) proposed MUSCLEMOTION (MM), a software tool that enables rapid and easy extraction of contraction signals from high-speed velocity videos. These methods are truly valuable for obtaining pharmacological data and modeling cardiac disease phenotypes. Still, evaluating only the variance in luminosity as an indicator of displacement can lead to results influenced by the experimental setup, the user performing the analyses, and the context in which they are conducted. All these factors can affect the baseline from which the analysis is made, influencing the results' reliability.

Estimating these parameters and avoiding errors caused by artifacts introduced in the experimental setup is crucial to evaluate therapeutic efficacy and safety and, eventually, to customize the treatment for individual patients in the context of personalized medicine.

For this reason, we propose a novel VAE-based model, hereafter referred to as VAE-MOTION (Fig. 1). By incorporating a temporal encoder composed of time-convolutional filters, our model can process video data to extract dynamic descriptors within its latent space, accurately capturing temporal behaviors in response to various external stimuli. The temporal module is directly embedded within the network architecture in series, positioned as the initial block, making it an intrinsic part of the model itself. At the same time, VAE-MOTION is specifically designed to identify and learn degradation artifacts, enabling it to restore the quality of degraded videos. This dual functionality, which combines robust analysis of dynamic descriptors with artifact-free video reconstruction, is unique among existing methods. In this work, VAE-MOTION is tailored to analyze cardiomyocyte contractility. It is trained to reconstruct high-quality data from degraded inputs, effectively addressing the limitations associated with experimental noise and the trade-offs between high spatial and temporal resolution in microscopy data. In particular, our model defines contraction and relaxation cycles within the latent space of the VAE architecture. This allows for encoding and modeling contraction, enhancing the estimation of dose–response curves even in the presence of video degradations that affect measurement results. The proposed method improves the quality and applicability of the derived contraction descriptors, even when working with data constrained by experimental setups or storage limitations. A classical approach could have been to first train a denoising model specifically tailored for this application and then apply standard analysis methods. In contrast, our VAE-based model is designed to simultaneously perform denoising and extract the relevant contractility signal directly from raw TLM videos. This simultaneous approach is particularly important because noise is very common in TLM videos, and classical methods often fail when data are noisy. By integrating these two functions within a single framework, VAE-MOTION ensures both

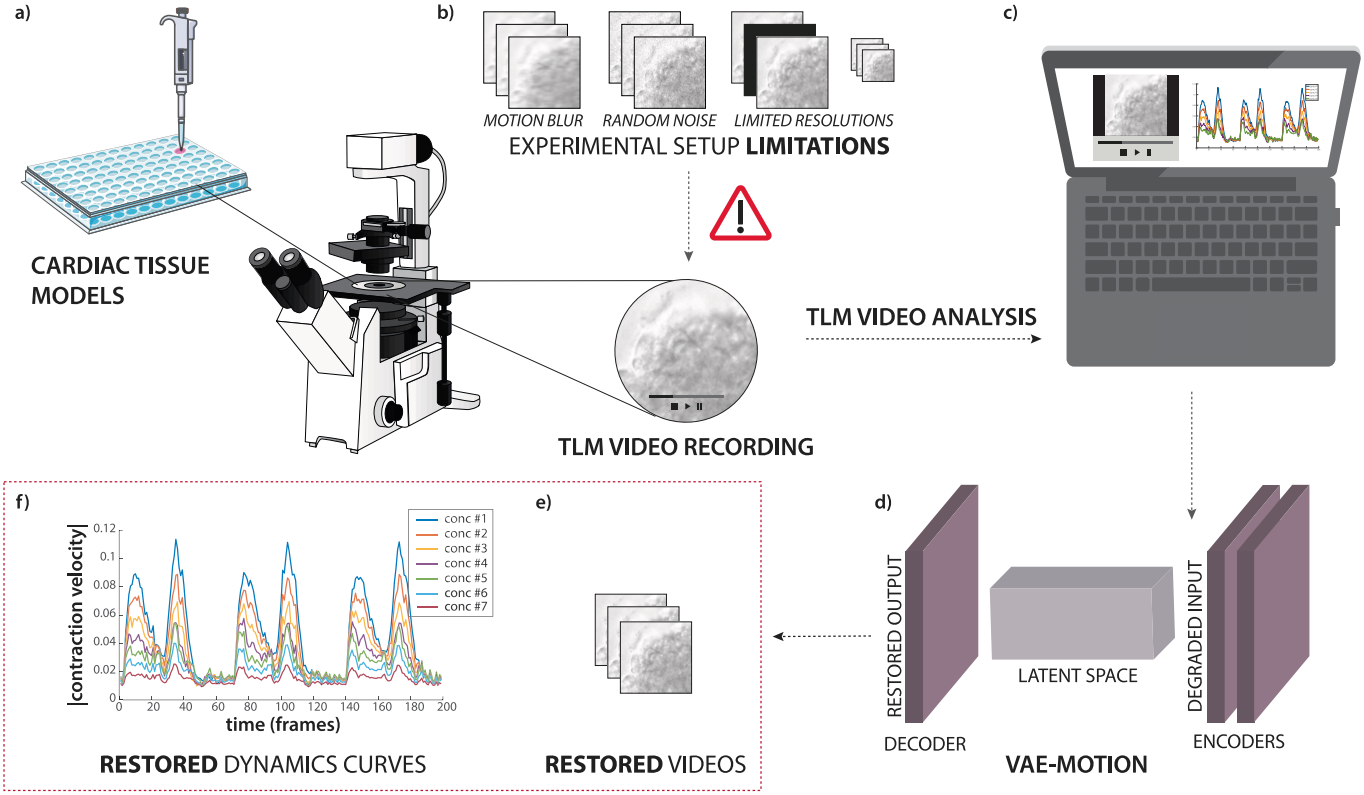

Fig. 1. Workflow of the proposed VAE-MOTION methodology. a) Experimental set up for TLM-based study of contractile cardiac models. b) Degradations often introduced experimentally during TLM recordings. c) Video analysis and extraction of dynamic curves. d) VAE-MOTION schematic representation. e,f) Restored outputs: restored videos obtained from the decoder output (e) and restored dynamics curves obtained from the latent space (f).

robustness to artifacts and accurate extraction of dynamic features, making it novel and well-suited for the analysis of cardiomyocyte contractility. To validate the proposed approach, we tested VAE-MOTION on two case studies: 2-dimensional monolayers of human-induced pluripotent stem cell (hiPSC)-derived cardiomyocytes and 3-dimensional cardiac microtissues. Moreover, we developed an algorithm that generates synthetic videos by modifying the displacement field of a single baseline video recorded under control conditions (i.e., without pharmacological treatment or pacing). Through this algorithm, we simulated the effects of various drug conditions, altering extracted contraction parameters, thereby producing phantom videos that served as training dataset. These videos were also generated in degraded form to mimic noise and experimental imperfections. The VAE-MOTION model was then trained on these datasets and subsequently validated on real (i.e., experimentally recorded) videos, both in their clean and degraded versions. From the latter, latent-space signals were extracted and systematically compared against those obtained using state-of-the-art methods, enabling the evaluation of both accuracy and robustness. Such a procedure, based on generating controlled synthetic data, training on diverse conditions, and validating on real experimental signals, can be further optimized and extended to other contractile tissues or, more generally, to the study of dynamic biological processes.

2. Methods

2.1. Variational autoencoders

VAEs (Kingma & Welling, 2013) are deep neural network architectures that combine generative and probabilistic modeling, exploiting the principles of variational Bayesian inference. As autoencoders (AEs), they compress data input, thereby reducing dimensionality, and then decompress it back into an approximation of the input itself. The input image $x \in \mathbb{R}^n$ is firstly compressed into a low-dimensional data point $z \in \mathbb{R}^d$ ($d < n$) through the encoder $E(x)$, then decompressed back through the decoder $D(z)$. Encoder and decoder functions are parameterized by deep neural network blocks, in which parameters are represented by the weights and biases of those networks in terms of code. The main difference compared to traditional AEs is that the compressed representation of the input sample x, i.e. z, is determined in a probabilistic way and randomly sampled from a Gaussian distribution. Thanks to its variational nature, the latent space is represented as a smooth, continuous distribution; hence VAEs can generate a wider variety of novel data samples, leading to better generalization. Input data x is compressed in z, chosen from the prior distribution $p_\theta(z)$. Then, the generative block recreates an approximation of x from the likelihood $p_\theta(x|z)$. Variational inference is used to form an approximation for $p_\theta(z|x)$ due to its analytical intractability. In this way, the posterior distribution over a set of unobserved variables $z = \{z_1, z_2, \cdots, z_d\}$ given some input data x is approximated by a so-called variational distribution $q_\phi(z|x) \approx p_\theta(z|x)$. The distribution $q_\phi(z|x)$ can be defined as:

$$q_\phi(z_1, \ldots, z_d|x) = \prod_{j=1}^{d} q_\phi(z_j|g(z_j), x) \tag{1}$$

with $g(z_j)$ being a set of variables derived from variables z_j.

Defining a differentiable loss function is necessary to update the network weights through backpropagation (Raúl Rojas, 1996). In VAEs, the idea is to optimize jointly the generative parameters θ and ϕ, to maximize the similarity between data input and data output, thus reducing the reconstruction error and making $q_\phi(z|x)$ the best approximation for $p_\theta(z|x)$, respectively. The objective function of this model is the so-called evidence lower bound (ELBO):

$$ELBO = \log p_\theta(x) - D_{KL}(q_\phi(z|x)||p_\theta(z|x)) \tag{2}$$

Maximizing ELBO means maximizing simultaneously the log-likelihood of the observed data and minimizing the divergence of the approximate posterior $q_\phi(z|x)$ from the exact posterior $p_\theta(z|x)$.

Indeed, the second term in Eq. 3 represents the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951) between $q_\phi(z|x)$ and $p_\theta(z|x)$, which is non-negative, so that $ELBO \leq \log p_\theta(x)$. The reparameterization trick makes the sampling process differentiable, allowing the model to be trained using backpropagation. Instead of directly sampling a latent variable z from a distribution $N(\mu, \Sigma)$, it is reparametrized as $z = \mu + \sigma \bullet \epsilon$, where $\epsilon \sim N(0, I)$ is a standard normal random variable, and σ is derived from the covariance matrix Σ (e.g., using the Cholesky decomposition $\Sigma = \sigma\sigma^T$). This allows gradients to flow through μ and Σ, enabling efficient optimization via backpropagation. Prior $p_\theta(x)$ is considered as a standard normal distribution, i.e., $p_\theta(x) \sim N(z; 0, I)$, while the variational distribution $q_\phi(z|x)$ is assumed to be derived from the generative process as isotropic multivariate Gaussian form $N(z; \mu, \sigma^2 I)$ and $\mu := h_\phi^{(1)}(x), \log\sigma^2 := h_\phi^{(2)}(x)$, where $h_\phi^{(1)}$ and $h_\phi^{(1)}$ are two outputs of the encoding network. Under this assumption, D_{KL} can be expressed analytically as:

$$D_{KL}(q_\phi(z|x)||p_\theta(z)) = -\frac{1}{2}\sum_{j=1}^{d}(1 + \log\sigma_j^2 - \mu_j^2 - \sigma_j^2) \tag{4}$$

where d is the dimensionality of z. Averaging the results of random sampling, simplifying integrals and expectations through Monte Carlo approximation, the exact loss VAE (as negative of ELBO) results:

$$loss_{VAE}(\phi, \theta; x) = -\sum_{i=1}^{n} E_{q_\phi(z|x_i)}[\log p_\theta(x_i|z)] - D_{KL}(q_\phi(z|x_i)||p_\theta(z)) \tag{5}$$

The loss function can be minimized with respect to both generative and variational parameters through stochastic gradient descent learning (Robbins & Monro, 1951). The model aims to reconstruct each x_i accurately from its compressed representation, z_i, thanks to the first term of Eq. (4), which therefore can be seen as a Reconstruction Loss. It also enforces that the latent variables z_i follow a standard normal distribution with the second term, the regularization one. This dual objective ensures that while the model captures the essential features of the data, the latent space remains well-structured and regularized, making it easier to generate new, realistic samples.

2.2. VAE-motion

We exploited generative capabilities and structured organization of latent spaces in VAEs to investigate temporal dynamic phenomena. VAEs learn a probabilistic mapping from input data to a lower-dimensional latent space, capturing essential features and enabling the generation of future states. Here, we introduce VAE-MOTION, an architectural variant of the VAE that explicitly emphasizes temporal continuity within the latent space. In addition, we propose a training approach that presents input data in a way that enhances the similarity of sequences between adjacent frames, thereby improving the model's ability to extrapolate signals and predict future states more accurately.

VAE-MOTION, unlike standard VAEs, introduces a pre-encoder that preprocesses input data, producing a representation with the same input frame dimensionality while accounting for temporal convolution in each input sequence through the use of 3D convolutional layers. The "time encoder" utilizes convolution filters that operate in both space and time. This allows it to capture intricate spatiotemporal patterns and preserve dynamic relationships between frames. As a result, the latent space benefits from a more regular structure. Whereas standard VAEs treat each frame as an independent entity and process it separately, this architecture considers dynamics as a fundamental feature for representing frames in the latent space. VAE-MOTION utilizes these capabilities to encode temporal dynamics, facilitating the creation of a latent space useful for studying contraction dynamics and generating accurately

reconstructed output frames. This design guarantees that each reconstructed frame is generated by processing a sequence of consecutive frames, enabling more accurate predictions and improved temporal coherence in video reconstruction (see Fig. 2).

2.2.1. VAE-motion architecture

The model's architecture is shown in Fig. 2 and detailed in Table 1. For each iteration, the first block receives a temporal series of data as input, with a total of m sequential frames of [128x128] pixels and performs 3D convolution on them, with m depending on the case study. Then, a max pooling layer ensures that the dimension of the obtained Time Encoder output is the same as the original data (128x128), concentrating relevant information. After this temporal pre-encoding, it enters the architecture of a standard VAE and, in the latent space, it results in a single data point $z \in \mathbb{R}^d$, with d fixed to 20. Latent signals are extracted as a sequence of data points corresponding to a series of temporally ordered batches (with stride equal to one), each one of dimension 128x128x m. ReLU functions follow convolutional 2D layers of the main encoder and decoder as activation layers. With this approach, the input processed by the main encoder is not an individual video frame but the result of spatiotemporal encoding, represented as an image that encapsulates dynamic information. The minimization of the cost function in Eq. (4), specifically the difference between reconstructed images and original images, occurs through the generation of an intermediate representation. This intermediate data not only ensures optimal reconstruction but also aids in building a more robust and temporally meaningful latent space. The idea above is to capture variation and movements not only in the spatial but also in the temporal domain, helping the model to follow dynamic changes, crucial features in the study of tissue contraction behaviors.

Table 1
Layers of the proposed VAE-MOTION architecture.

Time Encoder	Main Encoder	Decoder
m x m x m conv3D, ReLU	3 x 3 x 32 conv2D, ReLU	Fully connected
1 x 1 x m maxPooling3D	3 x 3 x 32 conv2D, ReLU	8 x 8 x 64 transposedConv2D, ReLU
	3 x 3 x 32 conv2D, ReLU	3 x 3 x 32 transposedConv2D, ReLU
	3 x 3 x 64 conv2D, ReLU	3 x 3 x 16 transposedConv2D, ReLU
	Fully connected	3 x 3 x 16 transposedConv2D, ReLU
		3 x 3 x 8 transposedConv2D, ReLU
		3 x 3 x 1 transposedConv2D, ReLU

2.2.2. Time series extraction

The main concept of the proposed approach is to utilize the latent space to encode input contractile videos and subsequently extract dynamic information, such as time series, from the latent space itself. In particular, two temporal series, named *latent distance* and *latent cumulative distance,* are extracted from the latent variables $z(t)$ with $z(t) \in \mathbb{R}^d$, as VAE representation of CMs dynamics (Fig. 2), respectively of tissue contraction velocity, $v_{VAE}(t)$, and of contraction amplitude (e.g. cumulative displacement), $A_{VAE}(t)$, as follows:

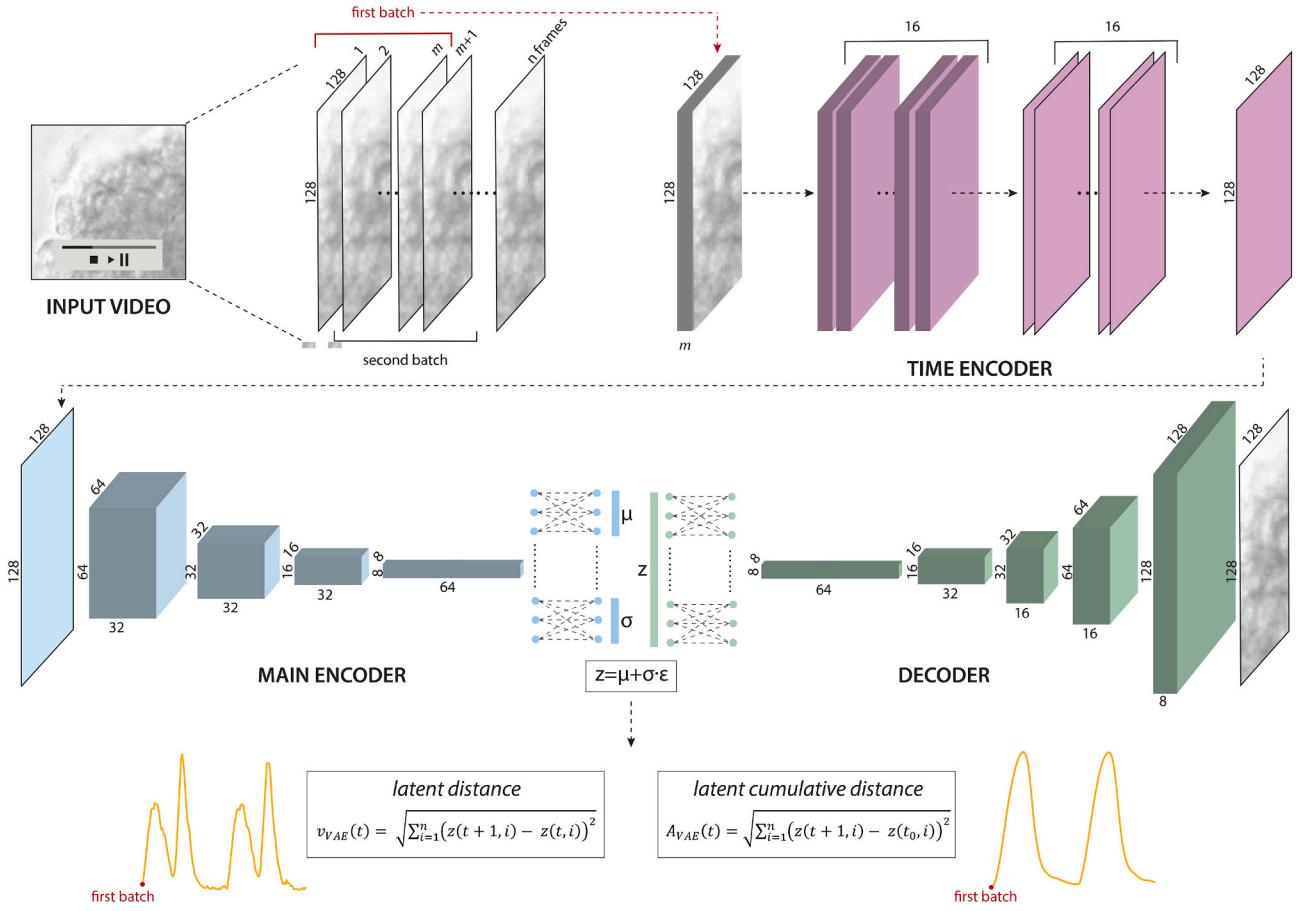

Fig. 2. VAE-MOTION exploded architecture. Input videos are organized into batches of m frames. For each iteration, a single batch is processed through the Time Encoder, Main Encoder, and Decoder, resulting in the reconstruction of a single frame. In the latent space, each batch is represented by a single point. Sequences of temporal adjacent batches are codified in dynamic time series.

$$v_{VAE}(t) = \sqrt{\sum_{i=1}^{d} (z(t+1,i) - z(t,i))^2} \quad (6)$$

$$A_{VAE}(t) = \sqrt{\sum_{i=1}^{d} (z(t+1,i) - z(t_0,i))^2} \quad (7)$$

In essence, these signals quantify how the latent representation of each frame evolves over time. The latent distance measures the frame-to-frame changes in the latent space (relative evolution with respect to the previous frame), capturing the instantaneous contraction velocity $v_{VAE}(t)$. The latent cumulative distance measures the deviation of the latent code from a reference frame t_0, capturing the overall contraction amplitude $A_{VAE}(t)$. These metrics thus provide a way to translate the latent dynamics of the tissue into interpretable temporal signals reflecting contraction behavior.

2.3. Phantom video generation

Training of deep learning models requires a substantial amount of data. To address shortages of data, synthetic data are often generated to augment the dataset (Benezeth et al., 2024; Prakosa et al., 2013). We generated a set of phantom videos simulating realistic scenarios to train the model. Then, the effectiveness of these videos was tested using real data. By isolating and manipulating specific parameters, we were able to generate phantom videos, creating a dataset with controlled variability that replicates various biological scenarios.

We generated a series of simulated videos using the displacement fields extrapolated from TLM videos of cardiac tissues beating under normal conditions (baseline), without any pharmacological treatment and without pacing. This approach enhances the model's capacity to discern features pertinent to the application context, e.g., in our case study, modifying the displacement vector field to simulate the effect of a specific drug on cardiomyocyte contractility.

To initially obtain displacement fields of baseline videos, we analyzed them by exploiting PIV (Harmand et al., 2013). PIV measures vector fields of displacement, evaluating the cross-correlation between pairs of adjacent frames in each sequence. In particular, for a given observation window, it produces the most probable position that the same window occupies in the next frame. This work uses tailored modifications of MATLAB's PIVlab algorithms (Thielicke & Stamhuis, 2014; Thielicke & Sonntag, 2021).

Before applying the PIV algorithm, each frame undergoes a pre-processing step to enhance image quality and facilitate the analysis. The first step increases the contrast of the input image by mapping intensity values linearly; the second step transforms the values using contrast-limited adaptive histogram equalization, CLAHE (Zuiderveld, 1994). Small sub-images (i.e., interrogation windows) of an image pair are cross-correlated, giving back the most probable particle displacement in the interrogation areas in uniform grids with fixed resolution.

The cross-correlation matrix (implemented as in Eq. (7)) is used to derive pattern matching between interrogation area A and interrogation area B of a different frame (Huang et al., 1997):

$$C(m,n) = \sum_i \sum_j A(i,j)B(i-m, j-n) \quad (8)$$

The correlation function is calculated in the frequency domain with the Fast Fourier Transform (FFT) function to reduce computational costs. Background noise is avoided by running several passes on the same grid. The result of the first pass is used as an offset for the following ones and so forth.

Subpixel precision in terms of displacement is achieved by the principle of the Gaussian 2x3 point fitting (Thielicke & Stamhuis, 2014): a more precise location can be derived from the location of the intensity peaks in the correlation matrix by fitting a Gaussian function from three points for each of the two dimensions separately.

In summary, the video generation process follows the following steps:

1. Extraction of the displacement field via PIV from a baseline video.
2. Bicubic interpolation to derive estimates of pixel-by-pixel displacements (Thielicke & Stamhuis, 2014), obtaining the components of the displacement field $u(x,y,t)$ and $v(x,y,t)$.
3. Modification of the displacement field $\tilde{u}(x,y,t) = f(u(x,y,t))$, $\tilde{v}(x,y,t) = g(v(x,y,t))$, with f and g independent functions, chosen by the operator to simulate specific conditions.
4. Selection of a reference frame $I(x,y,t_0)$ to obtain $t = 1$ corresponding to a frame of stillness before the onset of contraction.
5. The iterative process of generation of successive frames via application of the modified displacement field for $t_n = 2,3,\ldots t_N$, with t_N corresponding to the end of the relaxation cycle, as $I(x,y,t_n) = I(x - \tilde{u}(x,y,t_{n-1}), y - \tilde{v}(x,y,t_{n-1}), t_{n-1})$.

When working with non-integer displacement values, finding the exact new pixel intensity value at specific coordinates is not always possible. Therefore, each frame generation involves considering the four neighboring pixels at the resulting position and computing a weighted average based on their luminous intensity values. The weight assigned to each pixel is determined by its distance from the actual position identified by the displacement field. This process assumes rigid and bijective pixel-by-pixel displacements, which simplifies the computations but introduces an approximation. As a result, the reconstruction error tends to increase with successive iterations. To reduce this error, the final video is obtained as a weighted average between two versions: one obtained by applying the displacement field forward in time from frame $I(x,y,t_0)$ to $I(x,y,t_N)$ and one obtained by applying the displacement field backward from frame $I(x,y,t_N)$ to $I(x,y,t_0)$. This approach reduces the impact of systematic errors that accumulate iteration after iteration. Combining the two versions using a weighted average, the complementary information from the two temporal flows is leveraged.

3. Experimental analysis

This section presents a detailed experimental analysis to validate the proposed VAE-MOTION for analyzing contractile dynamics in human pluripotent stem cell-derived cardiomyocytes (hPSC-CMs). The following subsections describe the generation of synthetic datasets, the simulation of degradation effects, and the training and testing procedures for the model, ultimately demonstrating its efficacy in reconstructing both visual and physiological data under challenging conditions.

3.1. Experimental dataset

To demonstrate the efficacy of this new method, we applied VAE-MOTION to two physiologically and pharmacologically relevant experimental models that involve hPSC-CMs: (1) 2-D monolayer of hPSC-CMs cultures and (2) cardiac microtissues composed of hPSC-CMs and hPSC-derived endothelial cells (hPSC-ECs) (Giacomelli et al., 2017). All videos were recorded under bright-field conditions with an optical microscope with 10x objective and 100 s^{-1} frame rate. The videos have already been analyzed using the MM tool (Sala et al., 2017).

To train the model, we considered three different experiments for each of the two case studies. Each experiment included one baseline case video, showing tissue contraction without any treatment or pacing as a control, and six videos monitoring the tissue's response to pharmacological treatment with six concentration levels of nifedipine, an L-type calcium channel blocker. Nifedipine alters the heart's contractile force with a negative inotropic effect. Thus, with the increase in drug concentration, contractile force progressively decreases (Schwinger et al.,

1991).

3.2. Generation of synthetic datasets at different drug concentrations

We extracted the displacement fields of each video and, starting from them, the corresponding contraction velocity signals were derived as:

$$sig(t) = \sqrt{mean(|u(t)|)^2 + mean(|v(t)|)^2} \quad (9)$$

Analyzing these signals, we extrapolated the temporal indices that mark each beginning of contraction and end of relaxation phases. These indices allowed us to select reference frames from original videos and use them to generate synthetic videos that simulate the effect of increasing nifedipine concentrations.

In the case of monolayers, the tissue contraction occupied the entire field of vision, with the tissue tightly adhering to the entire substrate. However, in 3-dimensional cardiac microtissues, the contraction occurred exclusively in the tissue portion within the microtissue itself. In addition, for the 3D scenario at different drug concentrations for the same experiment, a remarkable shift in the x-y plane was observed in the videos because of changes in the experimental conditions caused by the perfusion of the drug on the microtissues, which led to significant morphological differences in the frames.

To obtain morphologically comparable videos of the different concentrations, we extracted correlation maps between the reference frames of all the videos belonging to the same experiment. This way, it was possible to identify regions of [512 x 512] pixels that maintained the maximum correlation between different drug concentrations, ensuring that the region's center corresponded as much as possible to the cultured microtissue centroid.

Thus, we used the displacement fields of only the baseline videos (untreated) and modified them by progressively reducing, for each pixel, their module (in Eq.8) with a factor of reduction of $r(t) = \frac{M\bullet(t-t_0)}{t_{peak}-t_0}$ $with t \in (t_0, t_{peak})$; $r(t) = \frac{M\bullet(t-t_{peak})}{t_{end}-t_{peak}}$ $with t \in (t_{peak}, t_{end})$, considering M varying from 0 to 0.12 in increments of 0.02. The obtained module was $s(t) = sig(t) - r(t)$. For each pixel, the corresponding displacements in both x and y coordinates were derived as $x(t) = s(t)*\cos\alpha$ and $y(t) = s(t)*\sin\alpha$, by maintaining the same direction of the movement, α, to simulate the decrease in contractile force induced by nifedipine. With the approach described above, we obtained from the displacement fields, for each experiment, a set of seven synthetic videos, respectively six simulated drug concentrations and one simulated control, each responding to a decrease in the peak of contraction velocity consistent with the intensification of nifedipine (See Supplementary Video 1 and 2 for one of the control videos, respectively for monolayer and microtissue). The choice of M depends on the maximum value of the baseline signal and is fundamental for generating curves that align with the real scenarios, particularly to represent potential drug effects. Specifically, in this case, the highest level of M, $(M = 0.12)$ corresponds to a video where the beating stops, resulting in a curve where the beating becomes nearly undetectable, mimicking the real effect of a high drug concentration. We opted to produce a linear reduction in contraction amplitude to streamline the algorithm's implementation. In fact, this phantom dataset was not designed to perfectly replicate real drug-treated videos at different concentrations, but rather to provide a controlled and reproducible framework where the contraction dynamics could be modulated in a predictable manner.

This approach maintains the efficacy of simulating changes in contraction dynamics while ensuring computational efficiency and robustness without constraining the training process to specific real concentrations. Velocity signals from CMs are composed of two peaks for each beat, corresponding respectively to the contraction and to the relaxation phase. Evaluating the trend of the peak in contraction velocity is one of the key parameters in the study of drug response. We resized each frame of phantom videos with a scale factor of 0.5 and extracted 4 crops of [128 x 128] pixels to obtain 28 video crops per experiment (7 concentration levels, 4 crops each), as shown in Fig. 3.

3.3. Simulation of degradations effect on videos

Several limitations can occur in experimental setups. Some relate to the necessity of collecting large volumes of data to achieve statistically significant results, which may require compromising resolutions to enable rapid acquisitions or extended recording time, while others relate to unpredictable environmental events, such as temperature fluctuations, vibrations, and illumination changes, beyond operators' control.

The core of our study was to assess the ability of VAE-MOTION to reconstruct original video data correctly. This approach involved reconstructing input frames to restore them to their original condition, regardless of whether data contained degradations or not. We aimed to restore not only the visual content of input images but also significative physiological metrics, such as dynamic responses, allowing for the extraction of accurate dose–response curves, critical when derived from degraded conditions.

In this work, we considered two kinds of degradation to validate the efficacy of our VAE-based methodology: (1) unwanted noise introduced during data acquisition and (2) known resolution limitations.

In (1), we considered two scenarios:

- *Random Gaussian noise*, typically introduced by various sources in the acquisition system, like cables, cameras, or sensors.
- *Motion blur*, introduced by unwanted shifts in the chip/well position relative to the microscope or vibrations affecting the microscope table, often caused by external factors such as machinery, building oscillations, or environmental disturbances.

In (2), we considered the two possible scenarios:

- *Limited spatial resolution*, leading to the inability of the acquisition system to capture fine spatial details in the x-y plane. This may arise while aiming for a high temporal resolution.
- *Limited temporal resolution*, resulting in the loss of rapid dynamic changes in time.

Thus, we generated five datasets for each experiment: one control dataset without degradation and four degraded datasets, each corresponding to one of the scenarios outlined above.

Random Gaussian noise was added to the phantom training dataset with mean $\mu = 0$ and standard deviation $\sigma = 0.1$ and $\sigma = 0.07$ for monolayers (see Supplementary video 3, left) and microtissues (see Supplementary Video 7, left), respectively.

Motion blur noise was added to phantom videos through a bidimensional filter that approximates, once convolved with an image, the linear motion of a camera with a length that varies randomly from 0 to 15 pixels and an angle from 0 to 30 degrees clockwise for both monolayers and microtissues case (respectively Supplementary Video 4 and 8, left).

Spatial resolution was degraded by downsampling. In the case of monolayers, resolution was decreased by a factor of 16 for each axis (Supplementary Video 5, left) and in microtissues by a factor of 8 for each axis (Supplementary Video 9, left). To preserve the network architecture, the frame sequences were then returned to their initial dimensionality of [128 x 128] pixels through an upsampling with bilinear interpolation.

Time resolution was reduced. One frame every 11 was maintained in the case of monolayers (Supplementary Video 6, left) and one every 9 in the case of microtissues (Supplementary Video 10, left), reducing the overall number of frames. As for spatial resolution degradation, frames underwent linear interpolation over time to maintain the architecture and restore the initial number of frames. In the microtissue case study, the noise level inducing problems in the evaluation of contractile dynamics was lower than in the case study of monolayers. Hence, the

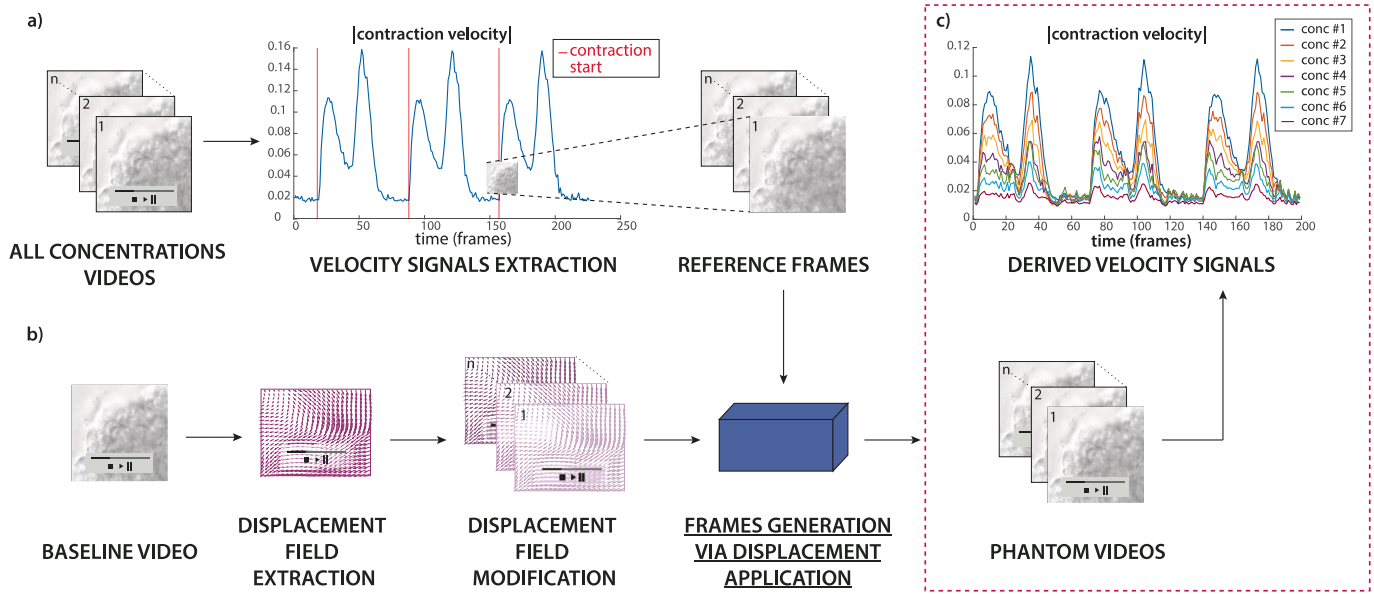

Fig. 3. Steps for phantom video generation. a) Extraction of the reference frames corresponding to the beginning of the contraction beat. b) Extraction and modification of the displacement field to simulate the effect of different drug concentrations. c) Generated synthetic videos and related contraction signals.

above choice of the levels of degradation was introduced.

3.4. Model training and test

To handle the data effectively, we trained the VAE-based model separately for each experiment and each specific degradation condition, completing the training after 400 epochs per model. We organized the data to obtain a batch in each iteration (total of 400) containing m of the 400 frames for all 28 crops used in the training. Input data were organized as random sequences of m frames so that the input dimension of 3-dimensional convolutional layer was 128x128x m, with $m = 5$. Only in the case of decreased temporal resolution, $m = 11$ (monolayer), and $m = 9$ (microtissue). Indeed, each training batch received just one original frame. The learning rate was fixed to 0.0005. We used the adaptive moment estimation (Adam) optimizer algorithm (Kingma & Ba, 2014) to update the network learnable parameters. All the simulations were performed in MATLAB R2023a on a system with NVIDIA GeForce RTX 4060, 32 GB RAM, i7-12700F CPU, and Windows 11 Enterprise 23H2. As described above, the training phase was conducted using phantom-generated videos. Each model was trained to extract $d = 20$ latent variables from the latent space. This value was chosen heuristically from the experiments. We trained the model to reconstruct the data in their original aspect without degradation. The effectiveness of this method was then tested on real original videos.

4. Results

This section aims to showcase the results of applying the proposed VAE-MOTION framework, emphasizing the role of the time-encoder on both unaltered and degraded video datasets. The evaluation focuses on its ability to accurately reconstruct dynamic contraction signals and extract essential physiological parameters. Additionally, the proposed method is compared to conventional approaches, such as PIV and MM.

4.1. Unaltered-videos condition

Firstly, the model has been rigorously optimized and validated for its capability to capture dynamic responses in tissue models. Specifically, we demonstrated that the system is able to detect the decrease in CM's contractile performance as nifedipine's concentration increases. This ability is crucial for assessing drug effects on cardiac tissue functionality, serving as an excellent readout for pharmacological testing and research.

Signals extracted from our model closely resemble those produced using other established methods, specifically MM and PIV. Both MM and PIV can be used to capture contractile behavior. However, our approach delivers equivalent high-fidelity data even when video degradation or noise is present, restoring video quality and yielding significant contraction parameters. Fig. 4 illustrates the resulting signals obtained from videos without degradation, considering two beats for each video, consisting of $n = 138$ total frames. In Fig. 4a, we compared the normalized velocity of contraction ($v_{VAE}(t)$) computed from phantom videos during the training phase and from real videos during the test phase with that obtained with MM and PIV tools on both 2D monolayers and 3D microtissue case studies.

By covering both model types, we can validate our system's general applicability and reliability across different tissue architectures. The model performed accurately during both training and testing phases, maintaining consistency when applied to real time lapse videos. The same accuracy is obtained by extracting normalized signals of cumulative displacement (contraction amplitude $A_{VAE}(t)$), as shown in Fig. 4b. In Fig. 4c, we highlight one of the most critical parameters for studying contractile trends: the peak in contraction velocity. This peak provides a clear indicator of the maximum rate at which the tissue contracts, offering a quantitative metric that can be used to compare the effects of various drug concentrations on tissue contractility.

4.2. Contribution of the time-encoder in VAE-motion

The use of a VAE is functional in encoding dynamics and extracting contraction signals from latent space, as described in Sect. 2.3. Indeed, a standard VAE system, with proper training, is already capable of reconstructing the video provided as input, producing from the latent space representations of the data from which we could derive indicators of the contraction dynamics.

In this work, we show that the implementation of a double encoder layer, as done in the case of VAE-MOTION, is necessary to overcome limitations due to the presence of noise. By encoding temporal variations in the deep architecture, the system is able to restore videos in the presence of degradation effects, contemporaneously extracting clean

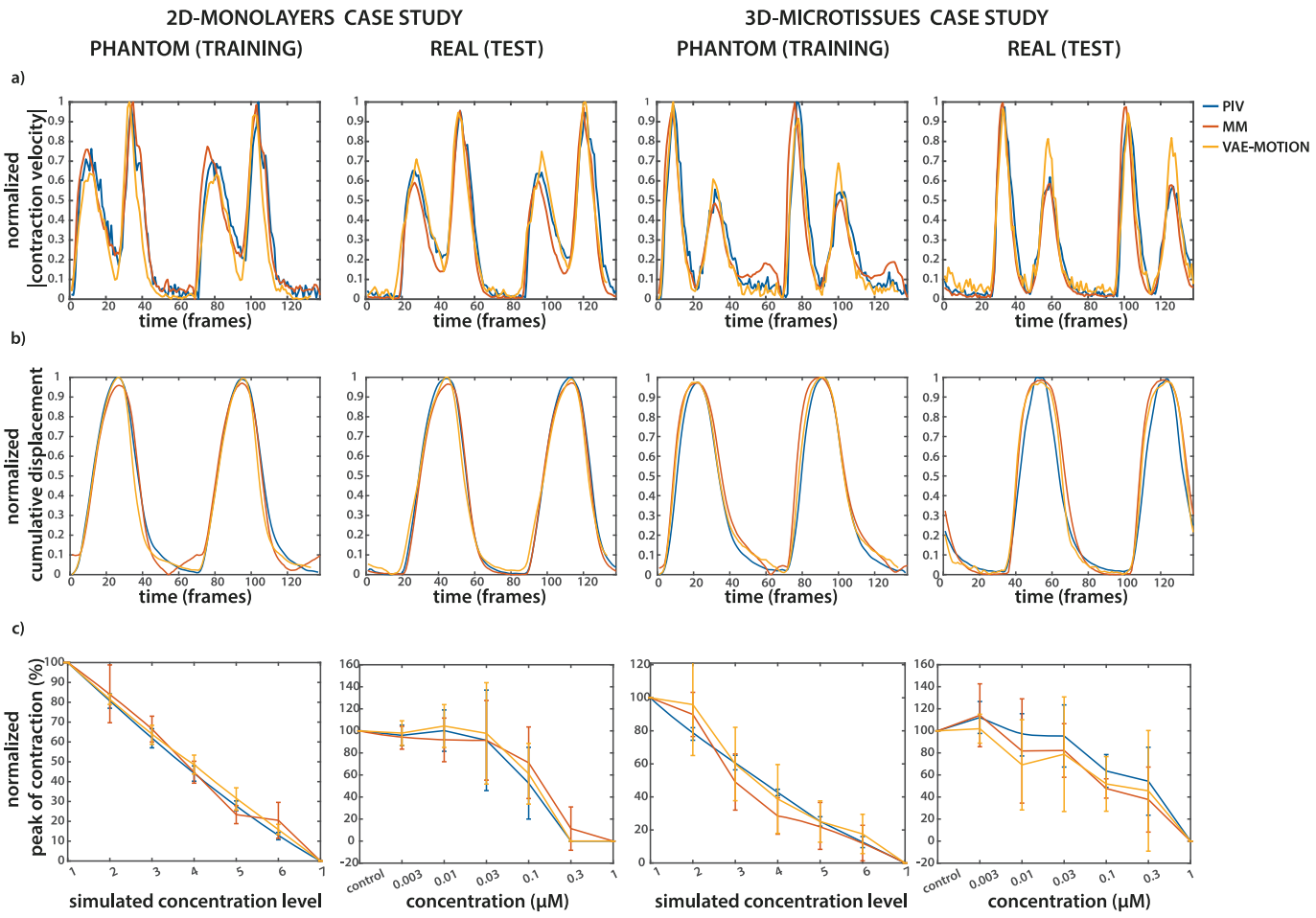

Fig. 4. Signals extracted from VAE-MOTION's latent space obtained during the training and test phases for both 2D monolayers of hPSC-CMs cultures and cardiac 3D microtissues. The results were compared with those extracted with MM and PIVlab tools. a) Normalized absolute value of velocity, b) normalized cumulative displacement, c) normalized trends in peak of contraction as function of varying drug concentrations, simulated in training, real in test.

latent signals indicative of the velocity of contraction and contraction amplitude. VAE-MOTION can not only extract useful signals but also disregard situations in which noise or limitations on resolutions appear, conditions that frequently occur when performing experiments.

Fig. 5 compares the results obtained using classical VAE architecture and those obtained using VAE-MOTION after applying Gaussian noise to videos. The standard VAE was implemented with a single encoder and decoder that share the same architecture as VAE-MOTION's main encoder and decoder, utilizing identical layers and training parameters. The comparison shown in Fig. 5 involves phantom videos generated for the case of 2D monolayers. It demonstrates the efficacy of a standard VAE model in processing unaltered videos, extracting curves coherent with the ones used to generate them, but also the necessity of an additional step for counteracting disturbances. The first column shows the results obtained using a standard VAE on videos without any degradation. The second column presents the results achieved with the same standard VAE when tasked with recovering videos affected by Gaussian noise. The third column displays the results obtained with VAE-MOTION under the same Gaussian noise recovery task. Fig. 5a shows the time-variant reconstruction loss, calculated as the average mean squared error (MSE) between the original image and the one pulled out by the VAE-MOTION decoder. The MSE has a similar trend in all three cases, but it is noisier in presence of noise processed by a standard VAE. Its temporal behavior has the same periodicity as the contraction velocity. It shows a peak in correspondence with the max value in the velocity signal related to the frames with the highest level of deformation relative to the reference frames. The distributions present similar mean values, with a slightly higher mean (and median) in the case of VAE-MOTION, as shown in Fig. 5b. Despite that, the use of VAE-MOTION proved to be necessary for the extraction of meaningful signals from the latent space (Fig. 5c). Indeed, excluding the presence of degradations the latent distances obtained with a standard VAE approach, i.e. without the double encoder block, are accurate and manage to estimate the dynamic trend over time at various concentrations. However, when the degradation is introduced, the obtained curves are noisy and do not allow further evaluation to extract the fundamental contraction parameters. VAE-MOTION, by proposing more attention to dynamics, allows the same curves to be restored. Fig. 5d shows the calculated error between the estimated curves and those obtained with the original displacement field, derived via PIV and used to generate the phantom videos. The reduction shown in the mean error distributions on VAE dynamic curves highlights the effectiveness of the VAE-MOTION architecture in improving dynamics estimates.

4.3. Real videos after degradation addition

VAE-MOTION's proposed architecture offers an effective solution for estimating critical contraction parameters even in noise-affected videos.

Signals extracted with traditional methods, such as PIV and MM, can be advanced tools for studying contraction. However, they become difficult to handle once noise or low-resolution conditions are introduced (Fig. 6a). As a result, none of these approaches allows the extraction of meaningful parameters in the study of contractile functions, thus preventing the evaluation of the effect of drug concentration

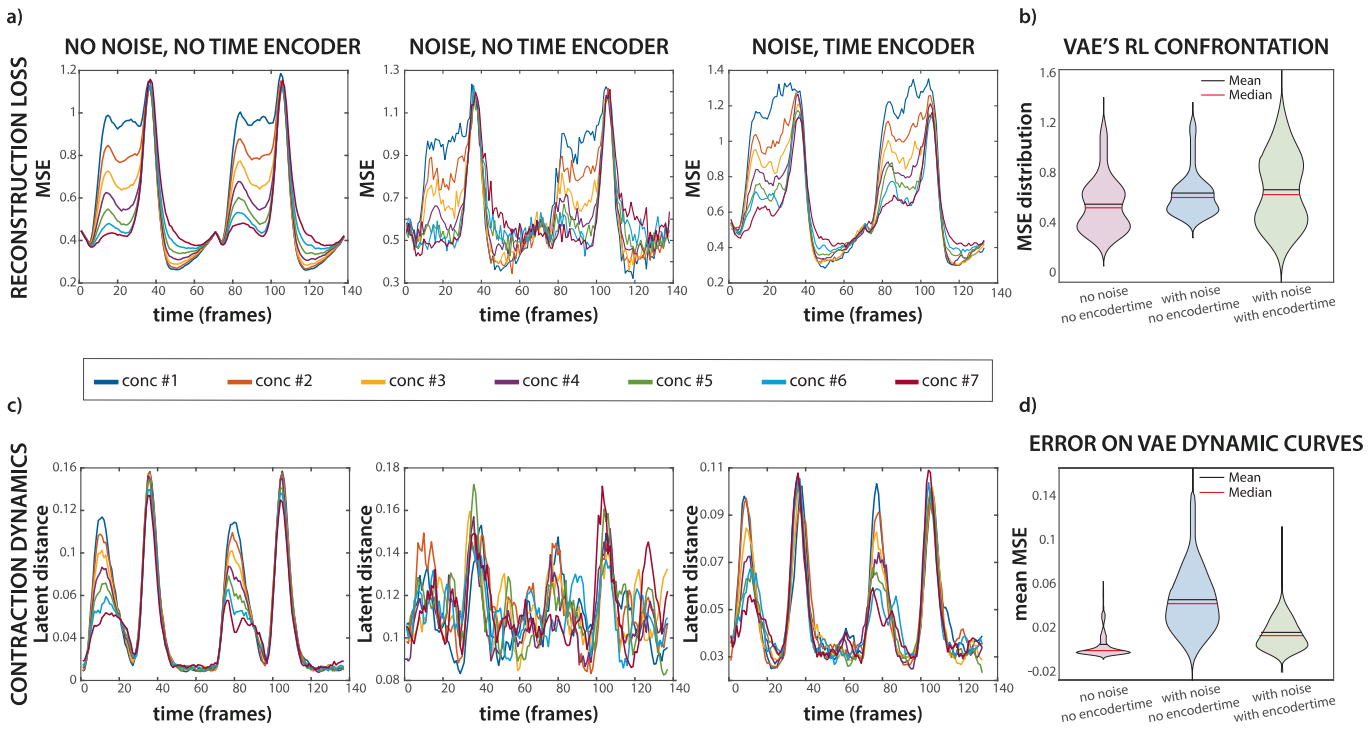

Fig. 5. Comparison between results obtained with a standard VAE receiving unaltered input videos, a standard VAE receiving degraded input videos, and VAE-MOTION receiving degraded input videos. a) Temporal trend of the reconstruction error. b) Violin plot of RL distributions. c) Contraction-dynamic curves extracted from the latent space. d) Error between obtained curves and those obtained with the original displacement field, derived via PIV and used to generate the phantom videos (used as reference).

on cardiac models. Fig. 6b shows the concentration–response trend of the contraction peak, one of the key parameters, comparing results obtained from training the VAE-MOTION network to restore videos in the presence of Gaussian noise, motion blur, and low spatial and temporal resolutions.

For both the 2D monolayer model and the 3D microtissues model, the results obtained (Fig. 6b) remain consistent with the literature (Sala et al., 2017) and similar to those observed in Fig. 4c, generated from unaltered videos, demonstrating the effectiveness of VAE-MOTION even on perturbed data. Similar to the control case, the deviation from the mean values is associated with the differences between videos, while still showing the correct trend regarding the effect of the investigated drug.

See the Right part of Supplementary Videos 1–8 for a visual demonstration of the real videos recovered with VAE-MOTION.

4.4. Statistical comparison between extracted signals

To further validate the accuracy and robustness of VAE-MOTION compared to traditional methodologies, we performed a statistical analysis on synthetic phantom videos where the ground truth was clearly defined. As described in Section 3.2, seven videos were generated by linearly modifying the baseline PIV-derived vector fields to mimic the effect of drug-induced changes in contraction velocity. Each video was divided into four ROIs, resulting in 28 signals per method. The peak values extracted from PIV, MM, and VAE-MOTION signals were compared against the ground-truth PIV reference using paired two-sided *t-tests*. The resulting p-values (Table 2) were all greater than 0.05, indicating no statistically significant differences between VAE-MOTION and the existing methods in noise-free conditions. This demonstrates that VAE-MOTION provides contraction signals fully consistent with established approaches under baseline scenarios.

To complement this analysis, the behavior of the three methods under different degradation conditions (Gaussian noise, motion blur, spatial and temporal resolution reduction) was evaluated. Performance decay curves (Fig. 7) were computed by progressively applying intermediate noise levels to the synthetic videos and measuring the agreement between reconstructed and ground-truth signals using the Concordance Correlation Coefficient (CCC):

$$CCC = \frac{2\rho\,\sigma_x\sigma_y}{\sigma_x{}^2 + \sigma_y{}^2 + \left(\mu_x - \mu_y\right)^2}$$

where $\rho = corr(x, y)$ is the Pearson correlation, μ_x and μ_y are the means, and σ_x, σ_y are the standard deviations of the two compared signals. Importantly, no additional VAEs were trained; instead, the videos with intermediate level of noise were directly provided as input to the VAEs trained on high-noise data (detailed in Section 3.3). Results show that the performance of PIV and MM drops sharply as noise or resolution degradation increases, leading to signals that are either unquantifiable (Gaussian/blur noise, Fig. 7c–d) or that lose critical spectral information required for parameter extraction (resolution degradations, Fig. 7e–f).

In contrast, VAE-MOTION preserves signal quality across conditions, maintaining robust concordance with the ground truth and enabling reliable estimation of contraction parameters.

Together, these results indicate that while VAE-MOTION yields comparable outcomes to standard methods in ideal scenarios, its key advantage emerges under low signal-to-noise ratio or resolution-limited conditions, delineating its advantageous operating range for functional cardiac analysis.

5. Discussion

The proposed model has been shown to allow in-depth analysis of contractile dynamics in cardiac tissues, particularly for the study of drug toxicity, while overcoming common laboratory challenges such as noise or low image resolution. To explore and analyze the dynamics, our system exploits the generative variational behavior of a VAE. By

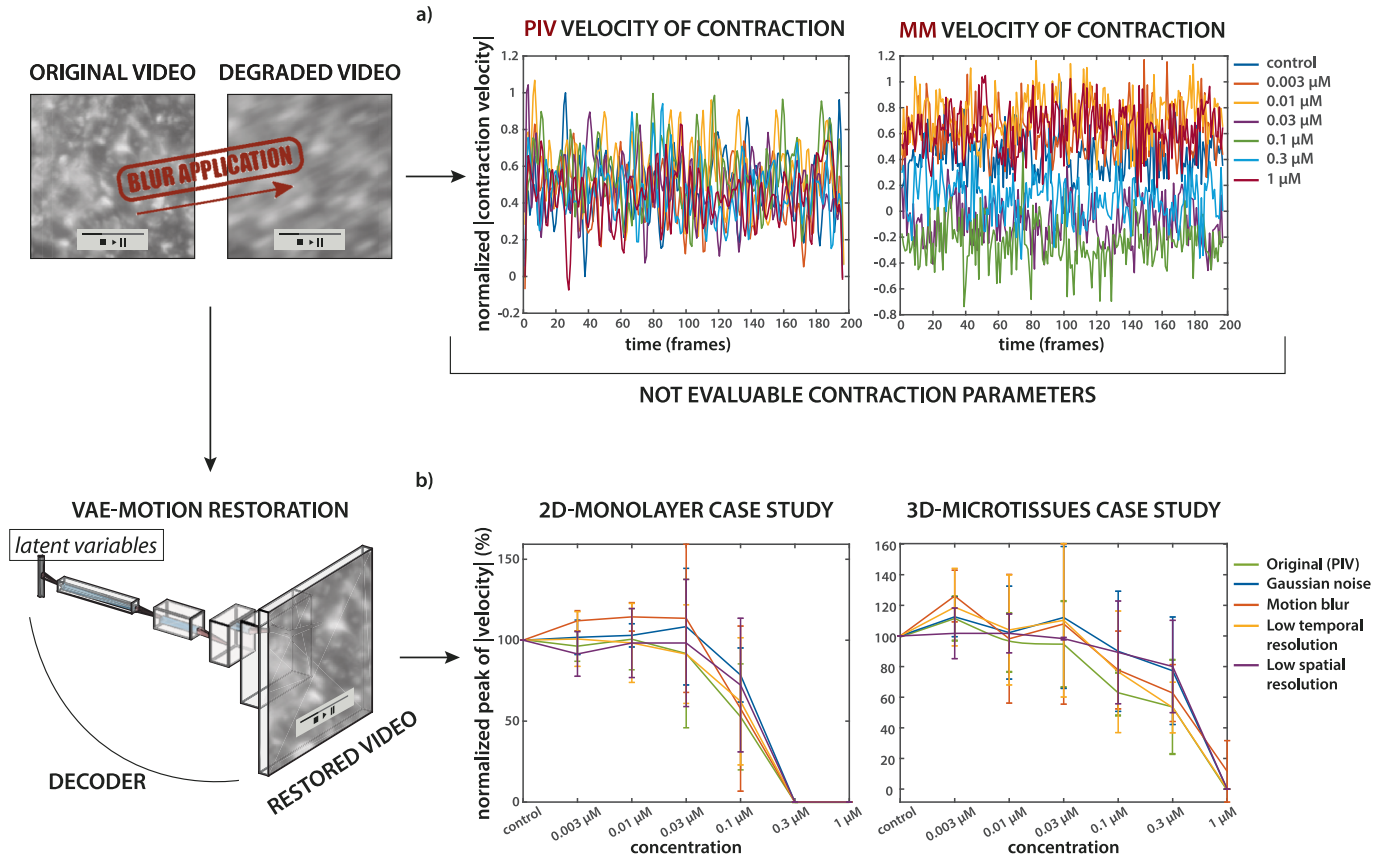

Fig. 6. Results obtained from degraded videos in the real case scenario with 2D monolayers of hPSC-CMs cultures and 3D cardiac microtissues. a) Curves obtained with PIV or MM, wherefrom contraction parameters cannot be extracted. b) Concentration-response curves obtained for the contraction's peak during the test phase for four degradation scenarios compared to curves obtained from PIV analysis on original videos.

Table 2
P-values resulting from *t*-test comparison in peak values extraction through different methodologies with respect to the ground truth.

Comparison	p-value
Original Signal vs PIV	0.6827
Original Signal vs MM	0.1591
Original Signal vs VAE	0.1192

regularizing the latent space, a VAE is able to encode samples that are similar to each other at close positions within the latent space. The basic idea is to exploit this latent space property to represent the contraction dynamics in TLM videos. In particular, our tests show that temporally close frames, characterized by minimal deformation and morphological changes, are encoded at nearby positions in the latent space. In contrast, frames that morphologically differ, such as those related to the onset and the end of contraction, are encoded in more distant positions.

The$d = 20$ latent variables extracted individually show periodic behavior aligned with the frequency of contraction of the video under analysis. However, in relation to what has just been described, the signal of greatest interest is the Euclidean distance between encoding points in latent space. For each temporal instant, the Euclidean distance between the point $z(t) \in \mathbb{R}^d$, encoding of the frame at time t, and $z(t-1) \in \mathbb{R}^d$, encoding of the previous frame is calculated. This distance provides an indicator of the morphological difference perceived by the network between two consecutive frames (Eq. (5). The greater the distance, the more significant the difference between adjacent frames. In particular, the maximum distance corresponds to the point of maximum contraction. The signal thus obtained can be interpreted as an indicator of the velocity of contraction. Similarly, calculating the Euclidean distance between each point in the latent space and a fixed point corresponding to a reference frame yields a signal that can be assimilated to the amplitude of contraction, i.e., the cumulative displacement (Eq. (6). Thus described, a VAE in its original form is able to replicate the same results produced by tools such as PIVlab or MM, providing the ability to perform the same analyses on contraction indicators useful for quantifying drug toxicity or efficacy. However, variations in the latent distances might occur not only to contraction events or morphological variations but also to variations related to the presence of noise, artifacts, or degraded experimental conditions.

The innovation of VAE-MOTION stems from the desire to combine the variational and generative capabilities of VAEs and also exploit the input restoration for the benefit of contraction analysis.

Experimental conditions often impose limitations in TLM video recording with noisy results and/or low resolution. This leads to uncertain and poorly reproducible results. The proposed model is able to overcome the noise-related limitations, producing more accurate signals and thus analysis. Indeed, VAE-MOTION was trained to recognize noise or low resolutions in low-quality images and then reconstruct clean, unaltered versions of them. VAE-MOTION training exploited the generation of phantom videos that allowed simulation of all possible conditions from a single control video. During training, the implemented architecture was designed to minimize the difference between the reconstructed image and the original unaltered version (Reconstruction Loss). However, the input provided to the encoder was the degraded video. In this way, VAE-MOTION can represent the salient features of the original image despite imperfections in the input, enabling more accurate reconstructions and signals for dynamic analysis. The original version of VAE alone cannot meet the demand (Fig. 5). This led to the

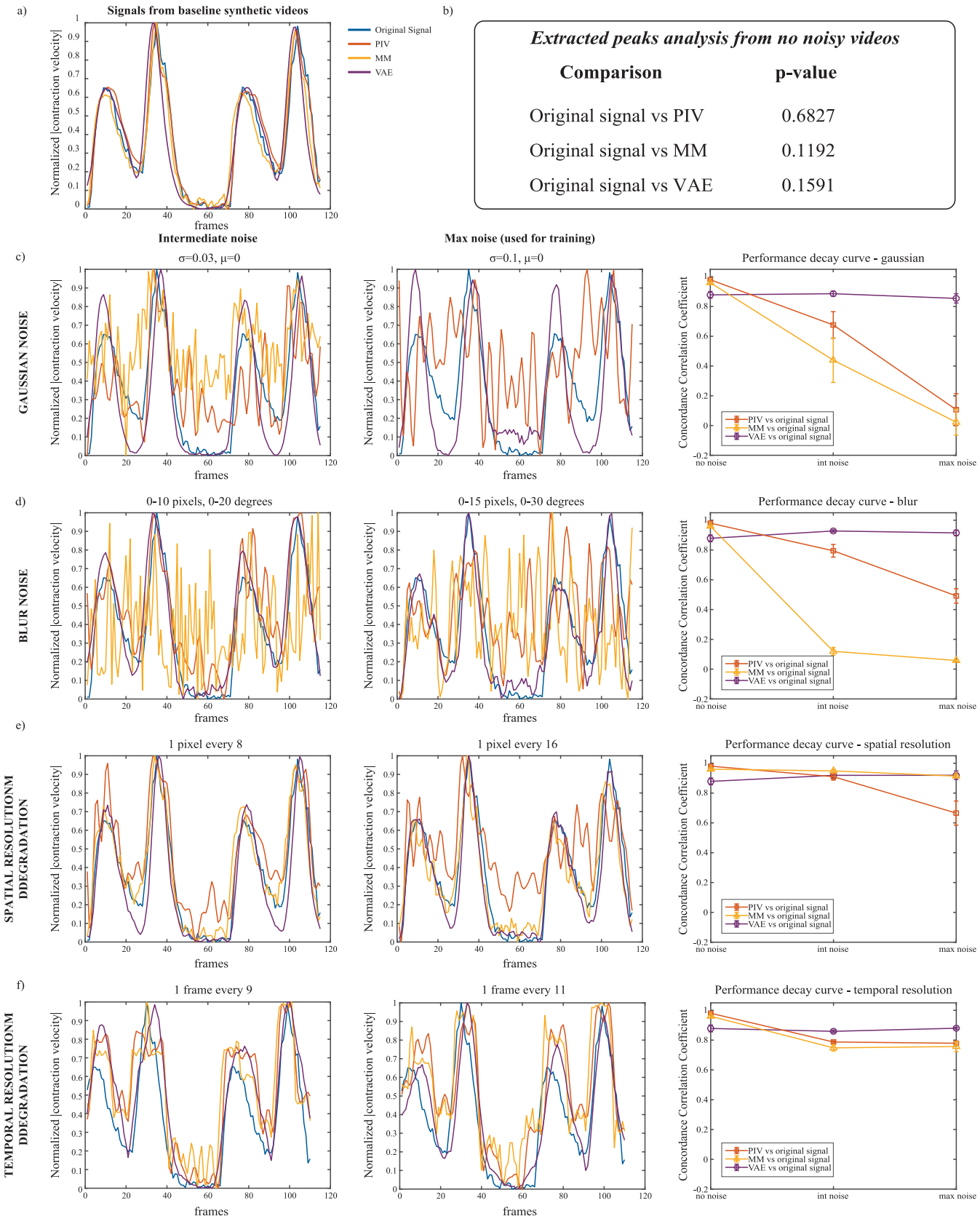

Fig. 7. Statistical comparison between extracted signals and performance decay analysis. a) Signals extracted from the synthetic baseline video with PIV, MM, VAE-MOTION compared to the Original Signal. b) Paired t-test results comparing contraction peak values obtained from PIV, MM, and VAE-MOTION signals against the ground-truth PIV reference in synthetic phantom videos. c–f) Performance decay curves of PIV, MM, and VAE-MOTION under different degradation conditions.

implementation of an additional block to the main Encoder, the "Time Encoder" block, which provided information on tissue dynamics. The block receives, as input, time sequences of adjacent frames, degraded and not, and applies three-dimensional convolutional filters (with third dimension being time) and returns output of spatial dimensions identical to the input but with the third dimension equal to one. The output of this block enters the Main Encoder and VAE-MOTION makes sure that the first frame of the input sequence is reconstructed in its unaltered version. The addition of Gaussian noise is taken as an example in Fig. 5, and it shows how the improvements obtained do not specifically reside in the morphological reconstruction of each frame. Instead, the reconstruction error (RL, computed as MSE) shows, in all three comparative cases, a trend correlated to the speed of contraction, i.e., the reconstruction is more difficult in the presence of higher contraction velocities. The improvement obtained with VAE-MOTION is, however, clearly visible in the extraction of signals from the latent space. The introduced noise, if not recognized and further modeled with a temporal block, is also reflected in the extracted signals, which are not evaluable to extract fundamental kinematic parameters. The signals extracted with VAE-MOTION in the case of added noise are shown to be comparable to those extracted in the case of no added noise. Thus, our model allows the evaluation of contraction parameters, even where other known methods fail, as demonstrated in Fig. 6. The concentration–response curves, for example, of the peak of the contraction velocity are kept the same as in Fig. 4c, which were extracted from the unaltered videos.

The model's ability to accurately track and extract this peak velocity further solidifies its utility in pharmacological testing, enabling precise monitoring of contractile dynamics and facilitating detailed drug effect analysis. The system has been tested on both 2D and 3D models, demonstrating its versatility of application for temporal analysis in different contexts even though, morphologically, the contraction develops in a completely different way.

6. Conclusion

This paper presents a new deep learning model based on a VAE for contraction analysis in TLM videos, named VAE-MOTION. The method provides a new way to extract information about the contractile activity of cardiac tissues under optimal and sub-optimal experimental conditions. By exploiting an additional temporal encoder in the VAE architecture, the model is able to account for the input dynamics and produce accurate analysis, even in the presence of degradations, such as noise or low-resolution conditions, due to very common limitations in experimental setups.

The proposed network was trained using synthetic videos of contractile CMs in two biological case studies: 2D monolayer of hPSC-CMs cultures and 3D cardiac microtissues of hPSC-CMs and hPSC-derived endothelial cells. Phantom videos were produced by extracting the displacement field of the baseline video and modifying it to simulate the effect of the drug nifedipine, a negative inotrope, and applying it to a reference frame. Multiple synthetic conditions were simulated and applied to reference frames of original videos to ensure reliability of the morphological structures in the simulated videos. The tests were conducted on real videos with the addition of noise, complicating the evaluation of dynamics using already existing methods. The proposed model proved effective in recovering the high quality of the videos, repairing degradations, and extracting useful and consistent information for the assessment of possible drug toxicity.

VAE-MOTION proved to be a powerful deep learning tool for analyzing cardiac contraction in TLM videos, demonstrating adaptability to different biological contexts. By virtue of its features, the model could be optimized and used in further studies of dynamics, both in the context of cardiac tissue and other in-vitro models, and in completely different applications, even outside the TLM framework.

CRediT authorship contribution statement

Giorgia Curci: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Paola Casti:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Luca Sala:** Data curation, Formal analysis, Writing – review & editing. **Marcella Brescia:** Data curation, Formal analysis, Writing – review & editing. **Pasquale Cascarano:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing. **Michele D'Orazio:** Writing – review & editing. **Joanna Filippi:** Methodology, Writing – review & editing. **Gianni Antonelli:** Methodology, Writing – review & editing. **Arianna Mencattini:** Conceptualization, Methodology, Writing – review & editing. **Massimo Mastrangeli:** Data curation, Formal analysis, Methodology, Funding acquisition, Writing – review & editing. **Berend J. van Meer:** Data curation, Formal analysis, Writing – review & editing. **Eugenio Martinelli:** Conceptualization, Data curation, Formal analysis, Methodology, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Italian Ministry of University and Research (MUR) through the PRIN-PNRR-2022 project "AI-HEART: AI-guided generation of beating and sensing heart-on-chip for drug screening" under grant number P20229Z3CM_002 and PRIN-2022 project "Sarcopenia-on-chip: an integrated platform based on chemical sensors, microfluidic devices, and machine learning algorithms for the development and testing of personalized treatment for sarcopenia disease (SELENE)", grant number 2022RWYH2K_001. This work was supported by European Union - Next Generation EU project No. H45E22001210006. This work was supported by the Netherlands Organ-on-Chip Initiative (an NWO Gravitation project (024.003.001) funded by the Ministry of Education, Culture and Science of the Government of the Netherlands)

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2025.130302.

Data availability

Data will be made available on request.

References

Agrawal, G., Aung, A., & Varghese, S. (2017). Skeletal muscle-on-a-chip: An in vitro model to evaluate tissue formation and injury. *Lab on a Chip, 17*(20), 3447–3461. https://doi.org/10.1039/C7LC00512A

Antonelli, G., Camera, F., Mencattini, A., Casciati, A., Tanori, M., Zambotti, A., Casti, P., Curci, G., Filippi, J., D'Orazio, M., Merla, C., & Martinelli, E. (2024). Lab-on-Chip Label-Free Sensing System for Electroporation Based on Time-Lapse Microscopy. *IEEE Sensors Journal, 24*(22), 36306–36315. https://doi.org/10.1109/JSEN.2024.3463959

Ascione, F., Caserta, S., Perris, R., & Guido, S. (2014). Investigation of cell dynamics in vitro by time lapse microscopy and image analysis. *Chemical Engineering Transactions, 38*, 517–522. https://doi.org/10.3303/CET1438087

Barron, J. L., Fleet, D. J., & Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision, 12*(1), 43–77. https://doi.org/10.1007/BF01420984

Benezeth, Y., Krishnamoorthy, D., Botina Monsalve, D. J., Nakamura, K., Gomez, R., & Mitéran, J. (2024). Video-based heart rate estimation from challenging scenarios

using synthetic video generation. *Biomedical Signal Processing and Control, 96*, Article 106598. https://doi.org/10.1016/j.bspc.2024.106598

Bers, D. M. (2002). Cardiac excitation–contraction coupling. *Nature, 415*(6868), 198–205. https://doi.org/10.1038/415198a

Bond-Taylor, S., Leach, A., Long, Y., & Willcocks, C. G. (2022). Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(11), 7327–7347. https://doi.org/10.1109/TPAMI.2021.3116668

Campostrini, G., Windt, L. M., van Meer, B. J., Bellin, M., & Mummery, C. L. (2021). Cardiac Tissues From Stem Cells: New Routes to Maturation and Cardiac Regeneration. *Circulation Research, 128*(6), 775–801. https://doi.org/10.1161/CIRCRESAHA.121.318183

Cascarano, P., Comes, M. C., Mencattini, A., Parrini, M. C., Piccolomini, E. L., & Martinelli, E. (2021). Recursive Deep Prior Video: A super resolution algorithm for time-lapse microscopy of organ-on-chip experiments. *Medical Image Analysis, 72*, Article 102124. https://doi.org/10.1016/j.media.2021.102124

Casti, P., Cardarelli, S., Comes, M. C., D'Orazio, M., Filippi, J., Antonelli, G., Mencattini, A., Di Natale, C., & Martinelli, E. (2023). S3-VAE: A novel Supervised-Source-Separation Variational AutoEncoder algorithm to discriminate tumor cell lines in time-lapse microscopy images. *Expert Systems with Applications, 232*, Article 120861. https://doi.org/10.1016/j.eswa.2023.120861

Celard, P., Iglesias, E. L., Sorribes-Fdez, J. M., Romero, R., Vieira, A. S., & Borrajo, L. (2023). A survey on deep learning applied to medical images: From simple artificial neural networks to generative models. *Neural Computing & Applications, 35*(3), 2291–2323. https://doi.org/10.1007/s00521-022-07953-4

Chen, Y., Zhou, G., Li, L., Chen, A., Wang, Y., & Li, L. (2024). A high-quality self-supervised image denoising method based on SDDW-GAN and CHRNet. *Expert Systems with Applications, 258*, Article 125157. https://doi.org/10.1016/j.eswa.2024.125157

Comes, M. C., Casti, P., Mencattini, A., Di Giuseppe, D., Mermet-Meillon, F., De Ninno, A., Parrini, M. C., Businaro, L., Di Natale, C., & Martinelli, E. (2019). The influence of spatial and temporal resolutions on the analysis of cell-cell interaction: A systematic study for time-lapse microscopy applications. *Scientific Reports, 9*(1), 6789. https://doi.org/10.1038/s41598-019-42475-5

Comes, M. C., Filippi, J., Mencattini, A., Casti, P., Cerrato, G., Sauvat, A., Vacchelli, E., De Ninno, A., Di Giuseppe, D., D'Orazio, M., Mattei, F., Schiavoni, G., Businaro, L., Di Natale, C., Kroemer, G., & Martinelli, E. (2021). Multi-scale generative adversarial network for improved evaluation of cell–cell interactions observed in organ-on-chip experiments. *Neural Computing and Applications, 33*(8), 3671–3689. https://doi.org/10.1007/s00521-020-05226-6

Comes, M. C., Filippi, J., Mencattini, A., Corsi, F., Casti, P., De Ninno, A., Di Giuseppe, D., D'Orazio, M., Ghibelli, L., Mattei, F., Schiavoni, G., Businaro, L., Di Natale, C., & Martinelli, E. (2020). Accelerating the experimental responses on cell behaviors: A long-term prediction of cell trajectories using Social Generative Adversarial Network. *Scientific Reports, 10*(1), 15635. https://doi.org/10.1038/s41598-020-72605-3

D'Orazio, M., Murdocca, M., Mencattini, A., Casti, P., Filippi, J., Antonelli, G., Di Giuseppe, D., Comes, M. C., Di Natale, C., Sangiuolo, F., & Martinelli, E. (2022). Machine learning phenomics (MLP) combining deep learning with time-lapse-microscopy for monitoring colorectal adenocarcinoma cells gene expression and drug-response. *Scientific Reports, 12*(1), 8545. https://doi.org/10.1038/s41598-022-12364-5

Ehrhardt, J., & Wilms, M. (2022). Autoencoders and variational autoencoders in medical image analysis. In *Biomedical Image Synthesis and Simulation* (pp. 129–162). Elsevier. https://doi.org/10.1016/B978-0-12-824349-7.00015-3.

Farnebäck, G. (2003). *Two-Frame Motion Estimation Based on Polynomial Expansion* (pp. 363–370). https://doi.org/10.1007/3-540-45103-X_50.

Filippi, J., Corsi, F., Casti, P., Antonelli, G., D'Orazio, M., Capradossi, F., Capuano, R., Curci, G., Ghibelli, L., Mencattini, A., & Martinelli, E. (2024). Optically Induced Dielectrophoresis and Machine Learning Algorithms for the Identification of the Circulating Tumor Cells. *Eurosensors, 2023*, 71. https://doi.org/10.3390/proceedings2024097071

Giacomelli, E., Bellin, M., Sala, L., van Meer, B. J., Tertoolen, L. G. J., Orlova, V. V., & Mummery, C. L. (2017). Three-dimensional cardiac microtissues composed of cardiomyocytes and endothelial cells co-differentiated from human pluripotent stem cells. *Development*. https://doi.org/10.1242/dev.143438

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139–144. https://doi.org/10.1145/3422622

Harmand, S., Pellé, J., Poncet, S., & Shevchuk, I. V. (2013). Review of fluid flow and convective heat transfer within rotating disk cavities with impinging jet. *International Journal of Thermal Sciences, 67*, 1–30. https://doi.org/10.1016/j.ijthermalsci.2012.11.009

Harrison, J. U., & Baker, R. E. (2018). The impact of temporal sampling resolution on parameter inference for biological transport models. *PLOS Computational Biology, 14*(6), Article e1006235. https://doi.org/10.1371/journal.pcbi.1006235

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Huang, H., Dabiri, D., & Gharib, M. (1997). On errors of digital particle image velocimetry. *Measurement Science and Technology, 8*(12), 1427–1440. https://doi.org/10.1088/0957-0233/8/12/007

Ioffe, S., & Szegedy, C. (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. http://arxiv.org/abs/1502.03167.

Jifara, W., Jiang, F., Rho, S., Cheng, M., & Liu, S. (2019). Medical image denoising using convolutional neural network: A residual learning approach. *The Journal of Supercomputing, 75*(2), 704–718. https://doi.org/10.1007/s11227-017-2080-0

Kim, Y., Wang, K., Lock, R. I., Nash, T. R., Fleischer, S., Wang, B. Z., Fine, B. M., & Vunjak-Novakovic, G. (2024). BeatProfiler: Multimodal In Vitro Analysis of Cardiac Function Enables Machine Learning Classification of Diseases and Drugs. *IEEE Open Journal of Engineering in Medicine and Biology, 5*, 238–249. https://doi.org/10.1109/OJEMB.2024.3377461

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. http://arxiv.org/abs/1412.6980.

Kingma, D. P., & Welling, M. (2013). *Auto-Encoding Variational Bayes*.

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, H., Kim, B., Yun, J., Bae, J., Park, S., Jeon, J., Jang, H. R., Lee, J., & Lee, S. (2024). PIV-MyoMonitor: An accessible particle image velocimetry-based software tool for advanced contractility assessment of cardiac organoids. *Frontiers in Bioengineering and Biotechnology, 12*. https://doi.org/10.3389/fbioe.2024.1367141

Li, H., Sundaram, S., Hu, R., Lou, L., Sanchez, F., McDonald, W., Agarwal, A., Chen, C. S., & Bifano, T. G. (2023). Dynamic Control of Contractile Force in Engineered Heart Tissue. *IEEE Transactions on Biomedical Engineering, 70*(7), 2237–2245. https://doi.org/10.1109/TBME.2023.3239594

Li, T., Nie, M., Morimoto, Y., & Takeuchi, S. (2024). Pillar electrodes embedded in the skeletal muscle tissue for selective stimulation of biohybrid actuators with increased contractile distance. *Biofabrication, 16*(3), Article 035022. https://doi.org/10.1088/1758-5090/ad4ba1

Lin, S., Clark, R., Birke, R., Schonborn, S., Trigoni, N., & Roberts, S. (2020). Anomaly Detection for Time Series Using VAE-LSTM Hybrid Model. In *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4322–4326). https://doi.org/10.1109/ICASSP40776.2020.9053558

Mencattini, A., Casti, P., D'Orazio, M., Antonelli, G., Filippi, J., & Martinelli, E. (2023). Uncertainty-Based Feature Selection for Improved Adequacy of Dermoscopic Image Classification. *IEEE Transactions on Instrumentation and Measurement, 72*, 1–9. https://doi.org/10.1109/TIM.2023.3303498

Mencattini, A., D'Orazio, M., Casti, P., Comes, M. C., Di Giuseppe, D., Antonelli, G., Filippi, J., Corsi, F., Ghibelli, L., Veith, I., Di Natale, C., Parrini, M. C., & Martinelli, E. (2023). Deep-Manager: A versatile tool for optimal feature selection in live-cell imaging analysis. *Communications Biology, 6*(1), 241. https://doi.org/10.1038/s42003-023-04585-9

Mencattini, A., Spalloni, A., Casti, P., Comes, M. C., Di Giuseppe, D., Antonelli, G., D'Orazio, M., Filippi, J., Corsi, F., Isambert, H., Di Natale, C., Longone, P., & Martinelli, E. (2021). NeuriTES. Monitoring neurite changes through transfer entropy and semantic segmentation in bright-field time-lapse microscopy. *Patterns, 2*(6), Article 100261. https://doi.org/10.1016/j.patter.2021.100261

Prakosa, A., Sermesant, M., Delingette, H., Marchesseau, S., Saloux, E., Allain, P., Villain, N., & Ayache, N. (2013). Generation of Synthetic but Visually Realistic Time Series of Cardiac Images Combining a Biophysical Model and Clinical Images. *IEEE Transactions on Medical Imaging, 32*(1), 99–109. https://doi.org/10.1109/TMI.2012.2220375

Qian, Z., Lv, Y., Lv, D., Gu, H., Wang, K., Zhang, W., & Gupta, M. M. (2021). A New Approach to Polyp Detection by Pre-Processing of Images and Enhanced Faster R-CNN. *IEEE Sensors Journal, 21*(10), 11374–11381. https://doi.org/10.1109/JSEN.2020.3036005

Rojas, R. (1996). *Neural networks: A systematic introduction*. Springer.

Robbins, H., & Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics, 22*(3), 400–407. https://doi.org/10.1214/aoms/1177729586

Rong, C., OuYang, S., & Sun, H. (2022). Anomaly Detection in QAR Data Using VAE-LSTM with Multihead Self-Attention Mechanism. *Mobile Information Systems, 2022*, 1–14. https://doi.org/10.1155/2022/8378187

Sala, L., van Meer, B. J., Tertoolen, L. G. J., Bakkers, J., Bellin, M., Davis, R. P., Denning, C., Dieben, M. A. E., Eschenhagen, T., Giacomelli, E., Grandela, C., Hansen, A., Holman, E. R., Jongbloed, M. R. M., Kamel, S. M., Koopman, C. D., Lachaud, Q., Mannhardt, I., Mol, M. P. H., & Burton, F. L. (2017). *Versatile open software to quantify cardiomyocyte and cardiac muscle contraction in vitro and in vivo.*. https://doi.org/10.1101/160754

Sami, M., & Mobin, I. (2019). A Comparative Study on Variational Autoencoders and Generative Adversarial Networks. *International Conference of Artificial Intelligence and Information Technology (ICAIIT), 2019*, 1–5. https://doi.org/10.1109/ICAIIT.2019.8834544

Scalzo, S., Afonso, M. Q. L., da Fonseca, N. J., Jesus, I. C. G., Alves, A. P., Mendonça, C. A. T. F., Teixeira, V. P., Biagi, D., Cruvinel, E., Santos, A. K., Miranda, K., Marques, F. A. M., Mesquita, O. N., Kushmerick, C., Campagnole-Santos, M. J., Agero, U., & Guatimosim, S. (2021). Dense optical flow software to quantify cellular contractility. *Cell Reports Methods, 1*(4), Article 100044. https://doi.org/10.1016/j.crmeth.2021.100044

Schwinger, R. H. G., Böhm, M., & Erdmann, E. (1991). Negative Inotropic Activity of the Calcium Antagonists Isradipine, Nifedipine, Diltiazem, and Verapamil in Diseased Human Myocardium. *American Journal of Hypertension, 4*(2_Pt_2), 185S–187S. https://doi.org/10.1093/ajh/4.2.185S

Soelistyo, C. J., Vallardi, G., Charras, G., & Lowe, A. R. (2022). Learning biophysical determinants of cell fate with deep neural networks. *Nature Machine Intelligence, 4*(7), 636–644. https://doi.org/10.1038/s42256-022-00503-6

Stiefbold, M., Zhang, H., & Wan, L. Q. (2024). Engineered platforms for mimicking cardiac development and drug screening. *Cellular and Molecular Life Sciences, 81*(1), 197. https://doi.org/10.1007/s00018-024-05231-1

Strbkova, L., Carson, B. B., Vincent, T., Vesely, P., & Chmelik, R. (2020). Automated interpretation of time-lapse quantitative phase image by machine learning to study cellular dynamics during epithelial–mesenchymal transition. *Journal of Biomedical Optics, 25*(08). https://doi.org/10.1117/1.JBO.25.8.086502

Thielicke, W., & Sonntag, R. (2021). Particle Image Velocimetry for MATLAB: Accuracy and enhanced algorithms in PIVlab. *Journal of Open Research Software, 9*(1), 12. https://doi.org/10.5334/jors.334

Thielicke, W., & Stamhuis, E. J. (2014). PIVlab – Towards User-friendly, Affordable and Accurate Digital Particle Image Velocimetry in MATLAB. *Journal of Open Research Software, 2*. https://doi.org/10.5334/jors.bl

Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W., & Lin, C.-W. (2020). Deep learning on image denoising: An overview. *Neural Networks, 131*, 251–275. https://doi.org/10.1016/j.neunet.2020.07.025

Wang, Z.-M., Messi, M. L., & Delbono, O. (2002). Sustained Overexpression of IGF-1 Prevents Age-Dependent Decrease in Charge Movement and Intracellular Ca2+ in Mouse Skeletal Muscle. *Biophysical Journal, 82*(3), 1338–1344. https://doi.org/10.1016/S0006-3495(02)75489-1

Yang, Q., Xiao, Z., Lv, X., Zhang, T., & Liu, H. (2024). Fabrication and Biomedical Applications of Heart-on-a-chip. *International Journal of Bioprinting, 7*(3), 370. https://doi.org/10.18063/ijb.v7i3.370.

Yu, S., & Ma, J. (2018). Deep learning for denoising. *International Geophysical Conference, Beijing, China, 24-27 April 2018*, 461–464. https://doi.org/10.1190/IGC2018-113.

Yun, J.-S., Kim, M. H., Kim, H.-I., & Yoo, S. B. (2024). Kernel adaptive memory network for blind video super-resolution. *Expert Systems with Applications, 238*, Article 122252. https://doi.org/10.1016/j.eswa.2023.122252

Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing, 26*(7), 3142–3155. https://doi.org/10.1109/TIP.2017.2662206

Zhang, W. J., Yang, G., Lin, Y., Ji, C., & Gupta, M. M. (2018). On Definition of Deep Learning. In *2018 World Automation Congress (WAC)* (pp. 1–5).

Zhang, Y. S., Aleman, J., Arneri, A., Bersini, S., Piraino, F., Shin, S. R., Dokmeci, M. R., & Khademhosseini, A. (2015). From cardiac tissue engineering to heart-on-a-chip: Beating challenges. *Biomedical Materials, 10*(3), Article 034006. https://doi.org/10.1088/1748-6041/10/3/034006

Zheng, D., Zhang, X., Ma, K., & Bao, C. (2022). *Learn from Unpaired Data for Image Restoration: A Variational Bayes Approach.*

Zuiderveld, K. (1994). Contrast Limited Adaptive Histogram Equalization. In *Graphics Gems* (pp. 474–485). Elsevier. https://doi.org/10.1016/B978-0-12-336156-1.50061-6.