

RESEARCH

AMES: Automated evaluation of sarcomere structures in cardiomyocytes

Maximilian Hillemanns^{1*}, Heiko Lemcke^{2,3}, Robert David^{2,3}, Thomas Martinetz⁴, Markus Wolfien¹ and Olaf Wolkenhauer^{1,5,6}

*Correspondence:

maximilian.hillemanns@uni-rostock.de

¹Department of Systems Biology and Bioinformatics, Faculty of Computer Science and Electrical Engineering, University of Rostock 18051 Rostock Germany
Full list of author information is available at the end of the article

Abstract

Background:: Arrhythmias are severe cardiac diseases and lethal if untreated. To serve as an in vitro drug testing option for anti-arrhythmic agents, cardiomyocytes are being generated in vitro from induced pluripotent stem cells (iPSCs). Unfortunately, these generated cardiomyocytes resemble fetal cardiac tissue rather than adult cardiomyocytes. An automated tool for an unbiased evaluation of cardiomyocytes would highly facilitate the establishment of new differentiation protocols to increase cellular maturity.

Results:: In this work, a novel deep learning-based approach for this task is presented and evaluated. Different convolutional neural networks (CNNs) including 2D and 3D models were trained on fluorescence images of human iPSC-derived cardiomyocytes, which were rated based on their sarcomere content (sarcomerisation) and the orientation of sarcomere filaments (directionality) beforehand by a domain expert. The CNNs were trained to perform classifications on sarcomerisation, directionality ratings, and cell source, including primary adult and differentiated cardiomyocytes. The best accuracies are reached by a 3D model with a classification accuracy of about 90 % for sarcomerisation classification, 63 % for directionality classification, and 80 % for cell source classification. The trained models were additionally evaluated using two explanatory algorithms, IGrad and Grad-CAM. The heatmaps computed by those explainability algorithms show that the important regions in the image occur inside the cell and at the cellular borders for the classifier, and, therefore, validate the calculated regions.

Conclusion:: In summary, we showed that cellular fluorescence images can be analyzed with CNNs and subsequently used to predict different states of sarcomere maturation. Our developed prediction tool AMES (<https://github.com/maxhillemanns/AMES>) can be used to make trustworthy predictions on the quality of a cardiomyocyte, which ultimately facilitates the optimized generation of cardiomyocytes from iPSCs and improves the quality control in an automated, unbiased manner. The applied workflow of testing different CNN models, adjusting parameters, and using a variety of explanatory algorithms can be easily transferred to further image based quality control, stratification, or analysis setups.

Keywords: Deep Learning; Fluorescence Image Analysis; Cardiomyocytes; Sarcomere Structures; Reasonable AI

¹Background

²Differentiation, generation, and maturation of cardiomyocytes and pacemaker cells ²
³Cardiomyocytes, the muscle cells of the heart, can be generated by various methods³
⁴in wet lab conditions: i) the differentiation of adipose tissue-derived mesenchymal⁴
⁵stem cells [1], ii) the differentiation of murine or human embryonic stem cells [2, 3],⁵
⁶and iii) the reprogramming of somatic cells [4], especially induced PSCs (iPSCs).⁶
⁷It is also possible to extract adult cardiomyocytes from murine hearts for compar-⁷
⁸ative analyses. These programming methods produce a cardiomyocyte aggregate,⁸
⁹where some cells possess so-called pacemaker abilities and some do not. In recent⁹
¹⁰approaches, the amount of non-pacemaker cells in this aggregate is still at around¹⁰
¹¹20 % [5, 6]. Moreover, the electrophysiological properties of these cells resemble¹¹
¹²fetal cardiac tissue instead of adult cardiomyocytes [7]. The maturation level of¹²
¹³cardiomyocytes may be critical for drug development, as immature cardiomyocytes¹³
¹⁴are far more sensitive to potassium-channel blockers [8].¹⁴

¹⁵In general, fully developed cardiomyocytes possess a well aligned and highly orga-¹⁵
¹⁶nized sarcomere network [9]. Longer sarcomere structures correlate to an improved¹⁶
¹⁷cardiac mechanical function [10]. Likewise, the mechanical, as well as the electrical¹⁷
¹⁸function, are also dependent on the orientation of myofibrils and subsequently the¹⁸
¹⁹sarcomere structures in a cell [11]. This orientation depends on the cell shape and¹⁹
²⁰the principal stress directions in the cell [12].²⁰

²¹In order to validate and evaluate cardiomyocyte generation protocols, scientists²¹
²²need to examine the maturity of these cell aggregates. Hence, the need for an easily²²
²³applicable method to distinguish between different maturation states is apparent²³
²⁴[13].²⁴

²⁶Cellular image analysis with deep learning

²⁷One possible approach to distinguish between different maturation states of car-²⁷
²⁸diomyocytes can be the analysis of cellular images with a variety of deep learning²⁸
²⁹(DL) applications. The main approaches commonly used refer to image segmen-²⁹
³⁰tation (partitioning an image into meaningful parts or objects), object tracking³⁰
³¹(identifying and following an object through a time series), augmented microscopy³¹
³²(extraction of latent information from biological images), and, finally, image classi-³²
³³fication. Deep learning image classification has been used on a variety of different³³
³⁴cells and tasks, like identifying changes in cell state [14], sorting cells into different³⁴
³⁵phenotypes [15, 16, 17], and distinguishing between differentiated and undifferen-³⁵
³⁶tiated cells on bright-field images [18]. By using DL, it is also possible to extract³⁶
³⁷feature vectors from cellular images in order to cluster these vectors and gain insight³⁷
³⁸on morphological patterns [19]. In comparison to classical machine learning (ML)³⁸
³⁹approaches like support vector machines (SVMs) [20] or logistic regression [21], DL³⁹
⁴⁰has shown to be more efficient at cell analysis tasks. In many cases, ML or DL is⁴⁰
⁴¹only applied after features were extracted from the images [20, 22]. In this study,⁴¹
⁴²DL was used directly on the images.⁴²

⁴³Transfer learning is also commonly used in biological image analysis due to the⁴³
⁴⁴lack of available training data. In transfer learning, a neural network pretrained on⁴⁴
⁴⁵another data set is applied to a new data set and the weights are retrained. Con-⁴⁵
⁴⁶volution Neural Networks (CNNs) are neural networks, which are able to extract⁴⁶

¹features or patterns from images themselves, without any need for sophisticated¹
²preprocessing. CNNs combine this extraction part with the ability to make classifi-²
³cation based upon these extractions. They were introduced by Bengio and Lecun in³
⁴1997 [23]. In cellular image analysis, CNNs are mainly used for image segmentation⁴
⁵and not image classification [24, 25]. For image classification, fully connected layers⁵
⁶are transferred after the convolutional layers to translate these features into a la-⁶
⁷bel. The labels that will be used here are different maturation degrees of sarcomere⁷
⁸structures that have been introduced by a biological domain expert. 8

⁹9

¹⁰Explainability analysis of image analysis models 10

¹¹As CNNs modulate highly nonlinear functions, they are too complex to allow for¹¹
¹²straightforward interpretability. This is referred to as the black box problem. A clas-¹²
¹³sifier may produce good classification results, but its inner workings and reasoning¹³
¹⁴are unattainable [26]. The question "Why does this model decide the way it does?"¹⁴
¹⁵plays an increasingly important role, especially in the life sciences. In the last few¹⁵
¹⁶years, a lot of algorithms have been developed to lift the lid of the black box. They¹⁶
¹⁷can be sorted by two main criteria: local vs. global explanation and model-specific¹⁷
¹⁸vs. model-independent. Local explanations are computed with the model, a data¹⁸
¹⁹point and an output. In most cases, this is the predicted output of the model, al-¹⁹
²⁰though a different label can be used to find weaknesses in the model (e.g., finding²⁰
²¹outputs the model might confuse for each other) [27]. Global explaining approaches²¹
²²take the model itself into account. An example would be the calculation of fea-²²
²³ture importances in Random Forests [28]. Model-independent approaches can be²³
²⁴applied on every classifier, while model-specific explaining algorithms are designed²⁴
²⁵for one type of classifiers, e.g., SVMs or CNNs. Two examples for local explaining²⁵
²⁶algorithms are Sensitivity Analysis (SA) and Layer-Wise Relevance Propagation²⁶
²⁷(LRP). They both produce a heatmap with pixels relevant to the classification [29].²⁷

²⁸This study compares and evaluates different CNN architectures upon their ability²⁸
²⁹to correctly identify cardiomyocyte's differentiation status/quality. Furthermore, a²⁹
³⁰comparison between a 2D and 3D analysis of fluorescence images is made and ex-³⁰
³¹plainability methods will be applied onto the classifiers to investigate the reasoning³¹
³²for a certain cellular stratification. In the end, it will be examined whether the³²
³³predictions match the biological criteria for differentiated cardiomyocytes. 33

³⁴Results 34

³⁵Individual model development and classification 35

³⁶Figure 1 shows the categorical accuracies, validation accuracies, and confusion ma-³⁶
³⁷trices on the test set for the singular 2D model on all classification tasks. All ac-³⁷
³⁸curacies have a rather logarithmic progression over time, a typical training curve.³⁸
³⁹For sarcomerisation classification, the training accuracy reached a plateau at almost³⁹
⁴⁰100 % after around 100 epochs. It took the validation accuracy around 80 epochs to⁴⁰
⁴¹reach a value of around 70 %, where it remained for the rest of the training epochs.⁴¹
⁴²The confusion matrix for sarcomerisation classification has its highest values along⁴²
⁴³the diagonal, with the next highest values in row 2, column 1 and 3, which means⁴³
⁴⁴that most of the data set is correctly classified. The accuracy of sarcomerisation⁴⁴
⁴⁵classification on the test data set is 68.68 %. The training accuracy for direction-⁴⁵
⁴⁶ality classification settled at almost 100 % after around 150 epochs. The validation⁴⁶

1 Images/Figure1.png 1
2
3 **Figure 1 Accuracy plots and confusion matrices for the singular model.** Shown are the
4 categorical training (train, blue) and validation (val, orange) accuracies over time as well as the
5 confusion matrices computed on the test set. (a-b): Sarcomerisation classification; (c-d):
6 Directionality classification; (e-f): Cell source classification

6 Images/Figure2.png 6
7
8 **Figure 2 Confusion matrices for the one-vs-all ensemble model.** Shown are the confusion
9 matrices computed on the test set. (a): Sarcomerisation classification; (b): Directionality
10 classification; (c): Cell source classification

11
12
13 accuracy reached a value of around 60 % after approximately 120 epochs. Again,
14 the highest value per row lies along the diagonal of the confusion matrix. Direc-
15 tionality classification has an accuracy of 57.95 % on the test data set. Cell source
16 classification's training accuracy reached a plateau of almost 100 % after about 110
17 epochs. The validation accuracy settled at around 75 % after approximately 130
18 epochs. In the confusion matrix computed on the test data set, the highest values
19 are located on the diagonal. The accuracy of cell source classification on the test
20 data set is 74.85 %.

21 In Fig. 2, the confusion matrices computed on the test set for the 2D ensemble
22 model are shown. For all three classification tasks, the highest value per row lies
23 on the diagonal. Sarcomerisation and directionality classification each have one cell
24 classified as none of the ratings. The accuracy of sarcomerisation classification on
25 the test data set is 70.93 %. It is 64.16 % for directionality classification and 82.06 %
26 percent for cell source classification. As the individual classifiers were trained succes-
27 sively, no single training/validation accuracy for the whole model can be computed.
28 They are, however, computed for each individual classifier in the model. The first
29 sarcomerisation classifier, a one-vs-all model for rating "1", has a final training
30 accuracy of 99.78 % and a final validation accuracy of 91.10 % after 1000 epochs
31 of training. The second classifier, for rating "2" respectively, settled at a training
32 accuracy of 99.53 % and a validation accuracy of 85.23 %. For classifiers 3 and 4,
33 the final training accuracies are 97.12 % and 99.56 %, while the final validation
34 accuracies are 78.44 % and 92.90 %, respectively.

35 The first directionality classifier has a final training accuracy of 99.66 % and a final
36 validation accuracy of 84.44 %, each after 1000 epochs of training. For classifiers 2
37 and 3, final accuracies settled at 99.75 % and 99.16 % (training) and 81.91 % and
38 81.33 % (validation). The fourth directionality classifier reached a final training ac-
39 curacy of 96.89 % and a final validation accuracy of 84.79 %.

40 The first cell source classifier is an one-vs-all model for the cell lineage "adult". It
41 has a final training accuracy of 99.96 % and a final validation accuracy of 97.10 %
42 after 1000 epochs of training. For the "iPSC" classifier, the final training accuracy
43 is 99.89 % and the final validation accuracy 97.53 %. The training accuracy for
44 the "neonatal" classifier reached 99.67 %, while the validation accuracy reached
45 92.84 %, each after 1000 epochs.

46 The accuracy plots for the 3D model based upon the singular model, also follow
47 a typical training trend (see Fig. 3). For sarcomerisation classification, the training

1 Images/Figure3.png 1
 2
 3 **Figure 3 Accuracy plots and confusion matrices for the 3D model.** Shown are the categorical 3
 4 training (train, blue) and validation (val, orange) accuracies over time as well as the confusion 4
 5 matrices computed on the test set. (a-b): Sarcomerisation classification; (c-d): Directionality 5
 6 classification; (e-f): Cell source classification

6 Images/Figure4.PNG 6
 7
 8 **Figure 4 Comparison of the weighted f1-scores between all evaluated models.** 8
 9

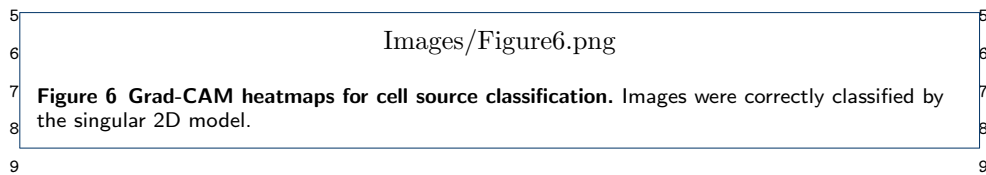
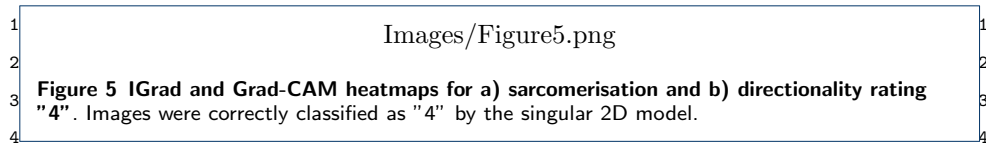
10
 11 accuracy reached 100 % after 50 epochs, while the validation accuracy remained at 11
 12 around 78 % after 100 epochs. The confusion matrix has its highest values along the 12
 13 diagonal, which means that the vast majority of images (90.91 %) were correctly 13
 14 classified. For directionality classification, the training accuracy is almost 100 % 14
 15 after 500 epochs. The validation accuracy reaches a peak at about 68 % after around 15
 16 150 epochs and then slowly declines to around 62 % after 500 epochs. The accuracy 16
 17 of directionality classification on the test data set is 62.60 % and the confusion 17
 18 matrix has its highest values along the diagonal with a notable peak in the last 18
 19 element of the second row. The training accuracy for cell source classification is 19
 20 100 % after 500 epochs of training, while the validation accuracy settles at around 20
 21 80 % after 70 epochs. In the confusion matrix computed on the test set, the highest 21
 22 values are located along the diagonal and in the second column of rows 4 and 5. 22
 23 The accuracy for cell source classification on the test set is 79.04 %. 23
 24

25 **Performance Comparison** 25

26 The weighted f1-scores for the classifiers evaluated in this study are summarized in 26
 27 Fig. 4. It is different from the normal multi-class f1-score, as the per-class f1-scores 27
 28 are not only averaged, but also weighted according to the occurrence of their re- 28
 29 spective class. For the singular 2D model, directionality classification performs with 29
 30 a weighted f1-score of 57.96 %, while sarcomerisation classification reaches 67.33 %, 30
 31 and cell source classification reaches 74.81 %. Sarcomerisation classification yields 31
 32 the second best results for the 2D ensemble model (70.16 %), while directionality 32
 33 classification reaches 64.19 %. This model performs best on cell source classification 33
 34 by achieving a weighted f1-score of 82.06 %. Directionality classification for the sin- 34
 35 gular 3D model has a weighted f1-score of 62.63 %, while the weighted f1-score for 35
 36 cell source classification is 79.14 %. The weighted f1-score for sarcomerisation clas- 36
 37 sification lies at 90.87 %. Table 1 shows the test accuracies and weighted f1-scores 37
 38 computed on the test data set for all evaluated models. 38
 39

40 **Table 1 Test accuracies and weighted f1-scores for the models evaluated in this study.** All values 40
 41 are presented in percent [%] and were computed on the respective test set. For the two transfer 41
 42 models, only the final results after retraining all layers are presented. In each column, the lowest and 42
 43 highest values are highlighted. (sarc: sarcomerisation, dir: directionality, cs: cell source) 43

	Test accuracy			Weighted f1-score		
	sarc	dir	cs	sarc	dir	cs
2D singular model	68.68	57.95	74.85	67.33	57.96	74.81
2D ensemble model	70.93	64.16	82.06	70.16	64.19	81.94
3D singular model	90.91	62.60	79.04	90.87	62.63	79.14



10 Explainability Heatmaps 10

11 In order to understand the reasoning behind the singular 2D models decisions, the 11

12 explainability algorithms iGrad and Grad-CAM were applied. In Fig. 5a, heatmaps 12

13 produced by IGrad as well as Grad-CAM for the singular 2D model can be seen. It 13

14 correctly classified these images with a rating of "4" for sarcomerisation, meaning 14

15 this particular cell has a high degree of sarcomerisation. The heatmaps produced 15

16 by IGrad have generally lower values than those produced by Grad-CAM. They lack 16

17 areas of high importance, as the cells themselves have no impact on classification. 17

18 There are however, slight accumulations around the edges of cells. When look- 18

19 ing at the Grad-CAM heatmaps, the highest values lie inside the cell, with peaks 19

20 around the edges. Almost the whole cell has a positive impact on classification. The 20

21 heatmaps for the other classes can be seen in the supplemental material. 21

22 The heatmaps Grad-CAM produced for the directionality have their peaks located 22

23 inside the cardiomyocytes, while the one produced by IGrad barely have peaks at 23

24 all (Fig. 5b). The IGrad heatmaps have very few values greater than zero. 24

25 The heatmaps produced by IGrad for the cell source have only very few values 25

26 greater than zero. Inside the cells, there are none visible. The Grad-CAM heatmaps 26

27 have their highest values inside the cells for classes "adult" and "iPSC" and on the 27

28 cellular border for class "neonatal" (see Fig. 6). Note that for "adult" and "iPSC", 28

29 Grad-CAM also highlights the borders of a cell. 29

30 **Discussion** 30

31 **Cardiomyocyte quality is determinable via CNNs** 31

32 When looking at the accuracies and the f1-scores of the models, it is clear that in 32

33 principle, the evaluation of cardiomyocytes using DL is possible. This is a promising 33

34 result, as in past approaches, neural networks mainly delivered well-suited results 34

35 on bright field microscopy images, as opposed to fluorescence images [20, 25]. The 35

36 classification regarding sarcomerisation, directionality, and even cell source yields 36

37 encouraging findings in terms of accuracy. Because of the balanced data set created 37

38 by data augmentation, the baseline classification accuracy (the model predicts every 38

39 data point to belong to the majority class) lies at around 20 % for sarcomerisation 39

40 and directionality classification and at around 30 % for cell source classification. 40

41 For all tasks, these baseline accuracies can be considerably outperformed. It also 41

42 appears that a shallow model is sufficient to tackle all classification tasks. The sin- 42

43 gular 2D model has far less parameters than other, established image classification 43

44 networks like the VGG-16 or MobileNet [30]. This suggests that the features ex- 44

45 tracted by a CNN do not need to be overly specific or detailed to ensure an accurate 45

46 46

¹classification of either sarcomerisation, directionality, or cell source. From a human ¹
²domain expert perspective, the rating of sarcomerisation and directionality are in-²
³tuitively easy, as humans can easily spot parallel patterns and fractions of objects.³
⁴Both are relatively easy features and so it is quite surprising that directionality⁴
⁵classification consistently performs worse than sarcomerisation and cell source clas-⁵
⁶sification. However, directionality classification, when looked at in detail, are two⁶
⁷separate tasks. First, the main cellular axis has to be found, which may be easy⁷
⁸for adult cardiomyocytes, but has its challenges for the other analyzed cell sources.⁸
⁹Second, the sarcomere structures must be evaluated according to their axis, poten-⁹
¹⁰tially leading to consequential errors, if the main axis is not determined properly.¹⁰
¹¹This may be an explanation to why it is more difficult for neural networks to evalu-¹¹
¹²ate on directionality than on sarcomerisation or cell source. It has been shown that¹²
¹³cardiomyocytes generated from iPSCs resemble fetal cardiomyocytes rather than¹³
¹⁴adult ones [31]. Our analysis also confirms this finding, as all three analyzed cell¹⁴
¹⁵types can be distinguished from each other, meaning that there is still an observable¹⁵
¹⁶difference between primary and generated cardiomyocytes(see Figs. 1-3). ¹⁶

¹⁷Stochastic Gradient Descent, although being a widespread optimizer, has its prob-¹⁷
¹⁸lems with large data sets and/or high dimensional feature space [32, 33]. The latter,¹⁸
¹⁹high dimensional feature space, holds true for the data used in this work, especially¹⁹
²⁰for 3D classification. The Adam optimizer tackles these problems and was conse-²⁰
²¹quently used for all models. The learning rate was set to 10^{-5} and made adaptive.²¹
²²Adaptive learning rates have been shown to boost classification results [34], which²²
²³was also the case here. Another approach, increasing the batch size over training in-²³
²⁴stead of decreasing the learning rate, was not applied, although it showed promising²⁴
²⁵results in previous works [35]. For the applications in this study, it was not appli-²⁵
²⁶cable, as the batch sizes cannot be increased arbitrarily because the images are too²⁶
²⁷large in file size, especially for the 3D classification. Although transfer models have²⁷
²⁸not been evaluated in this work, they have been applied onto cellular image analysis²⁸
²⁹tasks in the past. For example, Dong et al. were able to correctly distinguish malaria²⁹
³⁰infected cells from healthy ones using bright field microscopy with an accuracy of³⁰
³¹up to 98.13 % [20]. Cascio et al. built a transfer model, which was able to classify³¹
³²indirect immunofluorescence images of Human Epithelial type 2 (HEp-2) into fluo-³²
³³rescence intensity classes with an accuracy of 93.80 % [36]. HEp-2 is a marker for³³
³⁴antinuclear antibodies and, therefore, autoimmune diseases. Both approaches made³⁴
³⁵use of the AlexNet [37], which is rather simple in terms of architecture and number³⁵
³⁶of parameters. ³⁶

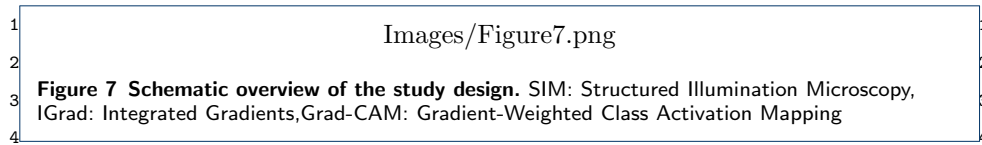
³⁷3D classification outperforms 2D classification in terms of accuracy ³⁷

³⁸The comparison between the 2D singular model and the 3D model, two similar ³⁸
³⁹models with different dimensionalities, shows a slight improvement in the classifica-³⁹
⁴⁰tion accuracies when 3D images were analysed. An increase in model performance ⁴⁰
⁴¹comes to no surprise, as the images are bigger and able to store more information in ⁴¹
⁴²the additional dimension. However, they do not seem to become more complicated, ⁴²
⁴³as the same shallow architecture is able to extract features, which suffice for classifi-⁴³
⁴⁴cation. The addition of a third dimension improves the classification marginally for ⁴⁴
⁴⁵directionality and cell source classification (about 5 % each) and drastically for sar-⁴⁵
⁴⁶comerisation classification (about 20 %). The computation time however increases ⁴⁶

¹drastically. Training a classifier on 2D data for 500 epochs is a matter of hours, on
²3D data, it takes days. Interestingly, both 2D and 3D models take up about the
³same storage space with around 6 megabytes. Still, the increase in computation
⁴time, not only for the training of the classifier, but also for image preprocessing,
⁵outweighs the improvement of classification accuracy at this point. There are two
⁶options for the application of this evaluation: Use the 2D classifier, and get quick,
⁷but slightly imprecise results, or use the 3D classifier, and get more sensitive and
⁸accurate results at the expense of computation time. As computational hardware
⁹is constantly improving, the difference in computation time might be significantly
¹⁰reduced in the future. 3D analysis has been long proven to outperform 2D analysis
¹¹on a wide range of tasks[38]. In the life sciences, 3D image analysis with DL is
¹²often used for segmentation [39, 40, 41]. In this context, it has been shown that
¹³3D segmentation outperforms the segmentation of all individual 2D slices [42]. 3D
¹⁴image classification approaches for biological images are rare. One example is the
¹⁵classification of functional connectomes by Khosla et al. They were able to correctly
¹⁶classify 73.30 % of functional magnetic resonance images (fMRI) using a custom
¹⁷build model trained on the Autism Brain Imaging Data Exchange (ABIDE) data
¹⁸set [43]. This is similar to the accuracies reached by 3D classification in this study
¹⁹(62.60 % to 90.91 %).

²¹Cellular borders are of high interest

²²The implementation of IGrad leads to no interpretable results. This could be due
²³to an error in the implementation of IGrad in the keras-explain package. Other pos-
²⁴sible error sources could be the shallowness of the evaluated model, the proximity
²⁵of a monochrome cellular image to the black image baseline or a combination of
²⁶these two. In contrast, the heatmaps produced by Grad-CAM are very informative,
²⁷although they lack detailed spatial information of pixel-wise impact on the classifi-
²⁸cation. This is due to the shallow architecture of the evaluated model. The deeper
²⁹a CNN is, the more specific the features in the last feature map are. The singular
³⁰model simply seems not deep enough to extract detailed features from the images.
³¹Still, it is apparent that the most important regions for all classification tasks are
³²within the cell. This is not surprising, as the sarcomere network lies within the cell
³³and thus, all valid decisions based upon this sarcomere network must be traceable
³⁴to the cell itself. Interestingly, there are outlier images, though, whose Grad-CAM
³⁵heatmaps locate the important regions to be outside of the cell. This could be due
³⁶to "zero-filters", convolutional filters that learn to find background or, as in this
³⁷case, black regions in the images. Especially for sarcomerisation, these filters can
³⁸be useful, as they allow for reverse explanations. If a large part of the image and/or
³⁹the cell is black, the sarcomerisation rating will probably be low. Cellular borders
⁴⁰seem to be of high importance according to the heatmaps produced by Grad-CAM.
⁴¹For all three classification tasks one can find examples of this (e.g., Fig. 5, Fig. 6).
⁴²This could hint at edge detection being learned by the networks. For sarcomerisa-
⁴³tion classification, this makes sense, as the network may recognize where the cells
⁴⁴are on the image and, consequently, which areas in the image should have a dense
⁴⁵sarcomere network. Adult cardiomyocytes tend to be elongated and thin, while car-
⁴⁶diomyocytes generated from iPSCs resemble fetal cardiac muscle cells, which are



more likely to be round or have irregular shapes [44]. Therefore, edge detection could benefit the distinction between these cell sources. In general, Grad-CAM produced heatmaps of the singular model that seemed to highlight regions in the image that would also be deemed important by human curators.

Conclusion

In this work, a novel unbiased tool to evaluate the sarcomerisation, directionality, and cell origin of a cardiomyocyte is presented. Several different CNNs were trained on 2D and 3D fluorescence images of cardiomyocytes with the sarcomere network stained. The cardiomyocytes used in this study were rated based on their sarcomerisation and the orientation of sarcomere structures (directionality) beforehand (CCRS scheme). The trained models were subsequently evaluated by feeding them into two explainability algorithms, IGrad and Grad-CAM, which highlight the areas in an image that are most important for the respective classification.

IGrad and Grad-CAM both produce heatmaps, where IGrad did not provide interpretable results in our data. The heatmaps produced with Grad-CAM have their highest values inside the cell and at cellular borders for all three classification tasks, meaning that these regions are important for classification. However, these heatmaps do not contribute to novel findings, but highlight that cells are being recognized by the classifier.

In general, it is shown that cellular fluorescence images can be analysed with CNNs. A classifier was built that is capable of predicting 82 % of cardiomyocyte origins, 71 % of sarcomerisation ratings, and 64 % of directionality ratings correctly. This classifier can be used to make independent and trustworthy predictions on the quality of generated cardiomyocytes based on the sarcomere network. This underlying work will significantly benefit the unbiased evaluation of cardiomyocytes, as a fast and reliable tool for cardiomyocyte aggregates is now available.

Methods

In this study, 2D and 3D fluorescence images of cardiomyocytes were analyzed with CNNs. These CNNs were trained to distinguish between different ratings for cardiac muscle states, which were assigned to the cells beforehand. The CNNs were evaluated according to their accuracy and f1-score. Additionally, the best performing model was analyzed with explainability algorithms in order to visualize the network's behaviour and criteria for classification. The workflow is shown schematically in Fig. 7.

Cell origin and image acquisition

A comparison of murine-derived cardiomyocytes (adult and neonatal) and human-derived induced pluripotent stem cells (iPSCs), which both differentiated into cardiomyocytes, serve as the basic sarcomere models in this work. This allows for a

Table 2 Antibodies used for immunostaining of cardiomyocytes.

Type	Antibody
Primary	Sarcomeric <i>alpha</i> -actinin antibody ea53 Abcam plc, USA
Secondary	F(ab') ₂ -Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody Thermo Fisher Scientific, USA

Images/Figure8.png

Figure 8 Example images for all different sarcomerisation and directionality ratings.

validation of the iPSCs, which were differentiated and harvested by the RTC. The cellular images used in this study were acquired using 3D fluorescence structured illumination microscopy (SIM), which provides for resolutions of about 100 nm [45, 46]. Between 36 and 83 images have been taken and arranged into a z-stack per cell. The x- and y-sizes of the images range from 564 to 2002 pixels and from 392 to 2027, respectively, depending on the size of the cell captured. In order to obtain the 3D fluorescence images, the cells were stained as follows. At first, the cells were fixed by adding pre-warmed 4 % paraformaldehyde (PFA) directly into the culture medium with a ratio of 1:1 for five minutes at 37 °C. The cells were washed two times with phosphate buffered saline (PBS) for 5 min each. After this, they were permeabilized with 0.2 % Triton for five minutes and again washed twice with PBS for five minutes each. Next, the immunostaining took place. Therefore, the unspecific binding sites of the cells were blocked with 1 % bovine serum albumin (BSA) at room temperature for 60 min. The cells were then stained with the primary antibody against *alpha*-actinin (see Table 2), which is diluted in 1 % BSA, for 60 min at room temperature. *alpha*-actinin binds to the actin filaments of the sarcomeres and stabilizes the muscle contractile apparatus [47]. Two washes with 0.2 % BSA for five minutes each followed. Then, the cells were incubated with the secondary antibody (see Table 2), which was diluted in 1 % BSA, for 45 minutes at room temperature. Again, the cells were washed, twice with 0.2 % BSA and twice with PBS, each for five minutes. Coverslips were rinsed with distilled water and the cells were embedded on slides using mounting medium containing DAPI.

Introduction of a cardiomyocyte cell rating system (CCRS)

The cells in the fluorescence images have been rated regarding the orientation of sarcomere structures relative to the longitudinal axis of the cell (directionality) and their sarcomerisation, which both correlate to the maturity of a cardiomyocyte. "1" marks the lowest rating and "4" the highest, with "2" and "3" as intermediate steps. A cell with sarcomere structures parallel to the cell's longitudinal axis would be marked as "1" for directionality, whereas a cell with perpendicular orientated sarcomere structures would be marked as "4". Similarly, cells with a high degree of sarcomerisation were marked as "4", as opposed to ones with a low level, which were marked as "1". Example images for all ratings can be seen in Fig. 8.

Cardiomyocyte image processing, classification, and explanation

As the original images are three-dimensional, 2D images had to be created with ImageJ (version 1.52a) [48]. The common practice of reducing the dimension by a maximum illumination projection along the z-axis was used to obtain two-dimensional

Table 3 Comparison of the number of parameters between the different models used. For the ensemble model, the sum of the individual models has been calculated.

Model	Number of parameters			
	Total	Trainable		
		Output neurons		
	5	5	5	6
2D singular model	1.517.489	1.517.510	1.517.105	1.517.126
2D ensemble model	3.004.020	3.775.025	3.002.996	3.753.745
3D singular model	1.406.193	1.406.214	1.405.809	1.405.830

images of the cardiomyocytes fluorescence stacks [49]. Image preprocessing consisted of resizing the images to a models respective input size and scaling the pixels between 0 and 255. Data augmentation was implemented to increase the number of data points. The augmentation was performed depending on the relative occurrence of a class to simultaneously guarantee for a balanced data set. The class with the fewest data points was multiplied by a factor of 15 for 2D classification and 9 for 3D classification. This is because more augmented 3D images exceed the available storage space. Images were either flipped horizontally, vertically, rotated by a random degree or a combination of these three methods. This does not distort the image and allows validation of the results [50].

Two different 2D classification approaches will be presented in the following: one singular model with varying architectures and one ensemble model. The first model is shallow, and consists of only three convolutional blocks, and, thus, has only a few parameters. The second model makes use of an ensemble of neural networks. For each class, a binary classifier was trained to distinguish between this class and all other classes.

3D image classification is often accompanied by a task like depth perception (e.g., human pose estimation [51]) or shape reconstruction (mesh/point cloud classification [52]). There are few cases, where 3D images are classified as a whole, mainly for medical purposes, but these approaches lack a common model used. Therefore, a native 3D classification model has been evaluated. The 3D model closely resembles the singular model. It is also made up from three blocks of 3D convolution, batch normalization, max pooling, and dropout layers. The detailed architecture can be seen in the appendix. The network was trained with the Adam optimizer with a learning rate of 10^{-5} , which decays by 10^{-8} each epoch.

Table 3 shows a comparison between the parameters of each model. The number of parameters strongly influences the storage space of a model, as well as the time needed to train and test it. Models with fewer parameters are easier to implement and take less time to make predictions.

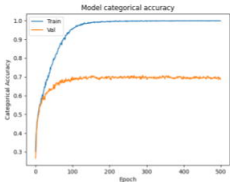
Two different methods for explainability analysis were applied in this thesis: Integrated Gradients (IGrad, [53]) and Gradient-weighted Class Activation Mapping (Grad-CAM, [54]). Both methods allow for a pixel-wise decomposition of CNNs. Here, we made use of the keras-explain 0.0.1 package for Python 3.7 to implement both methods and compute the heatmaps [55]. The heatmaps were scaled between 0 and 255.

All models were initialized with random weights and biases following a uniform distribution around zero. Native models were each trained for 500 epochs. The code was implemented in Python 3.7.2 using Keras 2.1.6 with Tensorflow backend

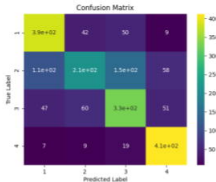
¹ and scikit-learn (version 0.20.3) packages [56, 57, 58]. Tests were run on an nVidia ¹	1
² GeForce GTX 2080 GPU with 8 GB RAM.	2
³	3
⁴	4
Ethics approval and consent to participate	
⁵ Not applicable.	5
⁶	6
Consent for publication	
⁷ Not applicable.	7
⁸	8
Availability of data and materials	
⁹ The datasets used and analysed during the current study are available from the corresponding author on reasonable request.	9
¹⁰ The python code used for analysis is available at github [59].	10
¹¹	11
Competing interests	
¹² The authors declare that they have no competing interests.	12
¹³	13
Funding	
¹⁴ This work was supported by the EU Social Fund (ESF/14-BM-A55-0024/18, ESF/14-BM-A55-0027/18), the DFG (DA1296/6-1), the German Heart Foundation (F/01/12), the FORUN Program of Rostock Medical University (889001/889003), the Josef and Käthe Klinz Foundation (T319/29737/2017), the DAMP Foundation (2016-11), and the BMBF (VIP+00240, 031L0106C).	15
¹⁶	16
Author's contributions	
¹⁷ Conceptualization, MH, TM, MW and OW; Formal analysis, MH; Funding acquisition, OW and RD; Investigation, MH and HL; Methodology, MH, HL, RD and MW; Supervision, RD, TM and OW; Visualization, MH; Writing — original draft, MH, HL and MW	17
¹⁸	18
¹⁹ All authors have read and agreed to the published version of the manuscript.	19
²⁰	20
Acknowledgements	
²¹ Not applicable.	21
²²	22
Author details	
²³ Department of Systems Biology and Bioinformatics, Faculty of Computer Science and Electrical Engineering, University of Rostock 18051 Rostock Germany. ² Reference and Translation Center for Cardiac Stem Cell Therapy (RTC), Department of Cardiac Surgery, Rostock University Medical Center Schillingallee 69 18057 Rostock Germany. ³ Department of Life, Light, and Matter of the Interdisciplinary Faculty at Rostock University Albert-Einstein-Straße 2 18059 Rostock Germany. ⁴ Institute for Neuro- and Bioinformatics, University of Lübeck Ratzeburger Allee 160 23562 Lübeck Germany. ⁵ Stellenbosch Institute of Advanced Study, Wallenberg Research Centre, Stellenbosch University 7602 Stellenbosch South Africa. ⁶ Leibniz-Institute for Food Systems Biology, Technical University of Munich Lise-Meitner-Straße 34 85354 Freising Germany.	23
²⁴	24
²⁵	25
²⁶	26
²⁷	27
²⁸	28
References	
²⁹ 1. Feng, Y., Yang, P., Luo, S., Zhang, Z., Li, H., Zhu, P., Song, Z.: Shox2 influences mesenchymal stem cell fate in a co-culture model in vitro. <i>Molecular medicine reports</i> 14 (1), 637–642 (2016)	29
³⁰	30
³¹ 2. Dorn, T., Goedel, A., Lam, J.T., Haas, J., Tian, Q., Herrmann, F., Bundschu, K., Dobrova, G., Schiemann, M., Dirschinger, R., <i>et al.</i> : Direct nkx2-5 transcriptional repression of isl1 controls cardiomyocyte subtype identity. <i>Stem cells</i> 33 (4), 1113–1129 (2015)	31
³²	32
³³ 3. Protze, S.I., Liu, J., Nussinovitch, U., Ohana, L., Backx, P.H., Gepstein, L., Keller, G.M.: Sinoatrial node cardiomyocytes derived from human pluripotent cells function as a biological pacemaker. <i>Nature biotechnology</i> 35 (1), 56–68 (2017)	33
³⁴	34
³⁵ 4. Hu, Y.-F., Dawkins, J.F., Cho, H.C., Marbán, E., Cingolani, E.: Biological pacemaker created by minimally invasive somatic reprogramming in pigs with complete heart block. <i>Science translational medicine</i> 6 (245), 245–9424594 (2014)	35
³⁶	36
³⁷ 5. Jung, J.J., Husse, B., Rimbach, C., Krebs, S., Stieber, J., Steinhoff, G., Dendorfer, A., Franz, W.-M., David, R.: Programming and isolation of highly pure physiologically and pharmacologically functional sinus-nodal bodies from pluripotent stem cells. <i>Stem cell reports</i> 2 (5), 592–605 (2014)	37
³⁸	38
³⁹ 6. Rimbach, C., Jung, J.J., David, R.: Generation of murine cardiac pacemaker cell aggregates based on es-cell-programming in combination with myh6-promoter-selection. <i>Journal of visualized experiments: JoVE</i> (96) (2015)	39
⁴⁰	40
⁴¹ 7. Blazeski, A., Zhu, R., Hunter, D.W., Weinberg, S.H., Zambidis, E.T., Tung, L.: Cardiomyocytes derived from human induced pluripotent stem cells as models for normal and diseased cardiac electrophysiology and contractility. <i>Progress in biophysics and molecular biology</i> 110 (2-3), 166–177 (2012)	41
⁴²	42
⁴³ 8. Tan, S.H., Ye, L.: Maturation of pluripotent stem cell-derived cardiomyocytes: a critical step for drug development and cell therapy. <i>Journal of cardiovascular translational research</i> 11 (5), 375–392 (2018)	43
⁴⁴	44
⁴⁵ 9. Lemcke, H., Skorska, A., Lang, C.I., Johann, L., David, R.: Quantitative evaluation of the sarcomere network of human hipsc-derived cardiomyocytes using single-molecule localization microscopy. <i>International journal of molecular sciences</i> 21 (8), 2819 (2020)	45
⁴⁶	46
⁴⁷ 10. Botcherby, E.J., Corbett, A., Burton, R.A., Smith, C.W., Bollensdorff, C., Booth, M.J., Kohl, P., Wilson, T., Bub, G.: Fast measurement of sarcomere length and cell orientation in langendorff-perfused hearts using remote focusing microscopy. <i>Circulation research</i> 113 (7), 863–870 (2013)	47

11. Wong, J., Kuhl, E.: Generating fibre orientation maps in human heart models using poisson interpolation. *Computer methods in biomechanics and biomedical engineering* **17**(11), 1217–1226 (2014)
12. Yuan, H., Marzban, B., Kit Parker, K.: Myofibrils in cardiomyocytes tend to assemble along the maximal principle stress directions. *Journal of biomechanical engineering* **139**(12) (2017)
13. Morris, T.A., Naik, J., Fibben, K.S., Kong, X., Kiyono, T., Yokomori, K., Grosberg, A.: Striated myocyte structural integrity: Automated analysis of sarcomeric z-discs. *PLoS computational biology* **16**(3), 1007676 (2020)
14. Sommer, C., Hoefler, R., Samwer, M., Gerlich, D.W.: A deep learning and novelty detection framework for rapid phenotyping in high-content screening. *Molecular biology of the cell* **28**(23), 3428–3436 (2017)
15. Doan, M., Case, M., Masic, D., Hennig, H., McQuin, C., Caicedo, J., Singh, S., Goodman, A., Wolkenhauer, O., Summers, H.D., *et al.*: Label-free leukemia monitoring by computer vision. *Cytometry Part A* **97**(4), 407–414 (2020)
16. Doan, M., Sebastian, J.A., Caicedo, J.C., Siegert, S., Roch, A., Turner, T.R., Mykhailova, O., Pinto, R.N., McQuin, C., Goodman, A., *et al.*: Objective assessment of stored blood quality by deep learning. *Proceedings of the National Academy of Sciences* **117**(35), 21381–21390 (2020)
17. Godinez, W.J., Hossain, I., Lazic, S.E., Davies, J.W., Zhang, X.: A multi-scale convolutional neural network for phenotyping high-content cellular images. *Bioinformatics* **33**(13), 2010–2019 (2017)
18. Buggenthin, F., Buettner, F., Hoppe, P.S., Ende, M., Kroiss, M., Strasser, M., Schwarzfischer, M., Loeffler, D., Kokkalis, K.D., Hilsenbeck, O., *et al.*: Prospective identification of hematopoietic lineage choice by deep learning. *Nature methods* **14**(4), 403–406 (2017)
19. Pawlowski, N., Caicedo, J.C., Singh, S., Carpenter, A.E., Storkey, A.: Automating morphological profiling with generic deep convolutional networks. *BioRxiv*, 085118 (2016)
20. Dong, Y., Jiang, Z., Shen, H., Pan, W.D., Williams, L.A., Reddy, V.V., Benjamin, W.H., Bryan, A.W.: Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 101–104 (2017). IEEE
21. Chen, C.L., Mahjoubfar, A., Tai, L.-C., Blaby, I.K., Huang, A., Niazi, K.R., Jalali, B.: Deep learning in label-free cell classification. *Scientific reports* **6**(1), 1–16 (2016)
22. Zhang, H., Shao, X., Peng, Y., Teng, Y., Saravanan, K.M., Zhang, H., Li, H., Wei, Y.: A novel machine learning based approach for ips progenitor cell identification. *PLoS computational biology* **15**(12), 1007351 (2019)
23. LeCun, Y., Bengio, Y., *et al.*: Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* **3361**(10), 1995 (1995)
24. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241 (2015). Springer
25. Van Valen, D.A., Kudo, T., Lane, K.M., Macklin, D.N., Quach, N.T., DeFelicce, M.M., Maayan, I., Tanouchi, Y., Ashley, E.A., Covert, M.W.: Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS computational biology* **12**(11), 1005177 (2016)
26. Binder, A., Bach, S., Montavon, G., Müller, K.-R., Samek, W.: Layer-wise relevance propagation for deep neural network architectures. In: Information Science and Applications (ICISA) 2016, pp. 913–922. Springer, ??? (2016)
27. Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., Samek, W.: On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one* **10**(7), 0130140 (2015)
28. Altmann, A., Tolosi, L., Sander, O., Lengauer, T.: Permutation importance: a corrected feature importance measure. *Bioinformatics* **26**(10), 1340–1347 (2010)
29. Samek, W., Wiegand, T., Müller, K.-R.: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296* (2017)
30. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017)
31. Karbassi, E., Fenix, A., Marchiano, S., Muraoka, N., Nakamura, K., Yang, X., Murry, C.E.: Cardiomyocyte maturation: advances in knowledge and implications for regenerative medicine. *Nature Reviews Cardiology* **17**(6), 341–359 (2020)
32. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
33. Le, Q.V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., Ng, A.Y.: On optimization methods for deep learning. In: ICML (2011)
34. Senior, A., Heigold, G., Ranzato, M., Yang, K.: An empirical study of learning rates in deep neural networks for speech recognition. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 6724–6728 (2013). IEEE
35. Smith, S.L., Kindermans, P.-J., Ying, C., Le, Q.V.: Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489* (2017)
36. Cascio, D., Taormina, V., Cipolla, M., Bruno, S., Fauci, F., Raso, G.: A multi-process system for hep-2 cells classification based on svm. *Pattern Recognition Letters* **82**, 56–63 (2016)
37. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25**, 1097–1105 (2012)
38. Chang, K.I., Bowyer, K.W., Flynn, P.J.: Face recognition using 2d and 3d facial data. In: Workshop in Multidimensional User Authentication Pp25-32 (2003). Citeseer
39. Castilla, C., Maška, M., Sorokin, D.V., Meijering, E., Ortiz-de-Solorzano, C.: Segmentation of actin-stained 3d fluorescent cells with filopodial protrusions using convolutional neural networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 413–417 (2018). IEEE
40. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 424–432 (2016). Springer

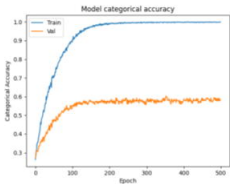
1	41. Weigert, M., Royer, L., Jug, F., Myers, G.: Isotropic reconstruction of 3d fluorescence microscopy images using	1
2	convolutional neural networks. In: International Conference on Medical Image Computing and	2
3	Computer-Assisted Intervention, pp. 126–134 (2017). Springer	
4	42. Patravali, J., Jain, S., Chilamkurthy, S.: 2d-3d fully convolutional neural networks for cardiac mr segmentation.	3
5	In: International Workshop on Statistical Atlases and Computational Models of the Heart, pp. 130–139 (2017).	4
6	Springer	
7	43. Khosla, M., Jamison, K., Kuceyeski, A., Sabuncu, M.R.: 3d convolutional neural networks for classification of	5
8	functional connectomes. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical	6
9	Decision Support, pp. 137–145. Springer, ??? (2018)	
10	44. Baumgartner, S., Halbach, M., Krausgrill, B., Maass, M., Srinivasan, S.P., Sahito, R.G.A., Peinkofer, G.,	7
11	Nguemo, F., Müller-Ehmsen, J., Hescheler, J.: Electrophysiological and morphological maturation of murine	8
12	fetal cardiomyocytes during electrical stimulation in vitro. <i>Journal of cardiovascular pharmacology and</i>	9
13	<i>therapeutics</i> 20 (1), 104–112 (2015)	
14	45. Lukosz, v.W., Marchand, M.: Optischen abbildung unter überschreitung der beugungsbedingten	10
15	auflösungsgrenze. <i>Optica Acta: International Journal of Optics</i> 10 (3), 241–255 (1963)	10
16	46. Gustafsson, M.G., Shao, L., Carlton, P.M., Wang, C.R., Golubovskaya, I.N., Cande, W.Z., Agard, D.A., Sedat,	11
17	J.W.: Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination.	12
18	<i>Biophysical journal</i> 94 (12), 4957–4970 (2008)	
19	47. Sjöblom, B., Salmazo, A., Djinović-Carugo, K.: α -actinin structure and regulation. <i>Cellular and molecular life</i>	13
20	<i>sciences</i> 65 (17), 2688–2701 (2008)	
21	48. Schneider, C.A., Rasband, W.S., Eliceiri, K.W.: Nih image to imagej: 25 years of image analysis. <i>Nature</i>	14
22	<i>methods</i> 9 (7), 671–675 (2012)	
23	49. Medalia, O., Weber, I., Frangakis, A.S., Nicastro, D., Gerisch, G., Baumeister, W.: Macromolecular architecture	15
24	in eukaryotic cells visualized by cryoelectron tomography. <i>Science</i> 298 (5596), 1209–1213 (2002)	16
25	50. Bloice, M.D., Roth, P.M., Holzinger, A.: Biomedical image augmentation using augmentor. <i>Bioinformatics</i>	17
26	35 (21), 4522–4524 (2019)	
27	51. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time	18
28	human pose recognition in parts from single depth images. In: CVPR 2011, pp. 1297–1304 (2011). Ieee	19
29	52. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and	20
30	segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.	
31	652–660 (2017)	20
32	53. Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: International Conference on	21
33	Machine Learning, pp. 3319–3328 (2017). PMLR	
34	54. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations	22
35	from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on	23
36	Computer Vision, pp. 618–626 (2017)	
37	55. Godec, P.: keras-explain: Explanation toolbox for Keras models.	24
38	https://github.com/PrimoZGodec/keras-explain (2018)	25
39	56. Chollet, F., et al.: Keras. https://keras.io (2015)	
40	57. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin,	26
41	M., et al.: Tensorflow: Large-scale machine learning on heterogeneous distributed systems. <i>arXiv preprint</i>	27
42	<i>arXiv:1603.04467</i> (2016)	
43	58. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P.,	28
44	Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. <i>the Journal of machine Learning</i>	29
45	<i>research</i> 12 , 2825–2830 (2011)	
46	59. Hillemanns, M.: AMES. https://github.com/maxhillemanns/AMES (2021)	30
31	Additional Files	31
32	Additional file 1 — Singular2D.png	32
33	Architecture of the singular 2D model in Keras notation.	33
34	Additional file 2 — Ensemble2D.png	34
35	Architecture of the ensemble 2D model in Keras notation.	35
36	Additional file 3 — Singular3D.png	36
37	Architecture of the singular 3D model in Keras notation.	37
38	Additional file 4 — CellSource-IGrad.png	38
39	Grad heatmaps for cell source classification. Images were correctly classified by the singular 2D model.	39
40		40
41		41
42		42
43		43
44		44
45		45
46		46



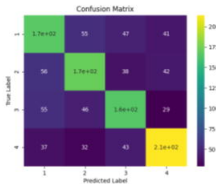
(a)



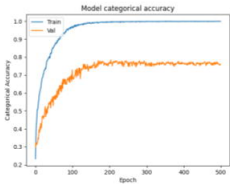
(b)



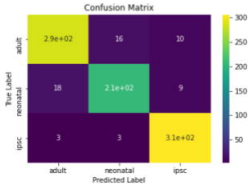
(c)



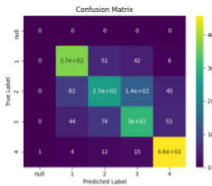
(d)



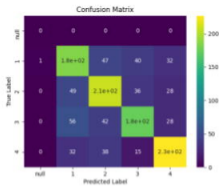
(e)



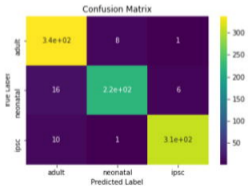
(f)



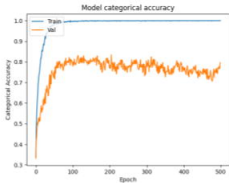
(a)



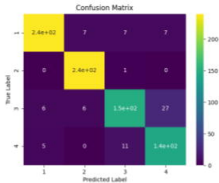
(b)



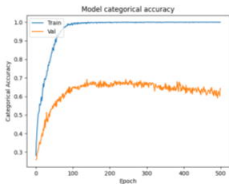
(c)



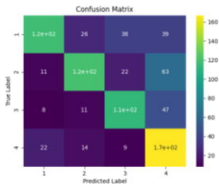
(a)



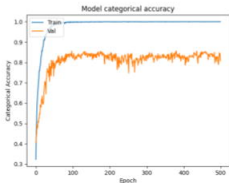
(b)



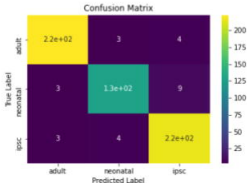
(c)



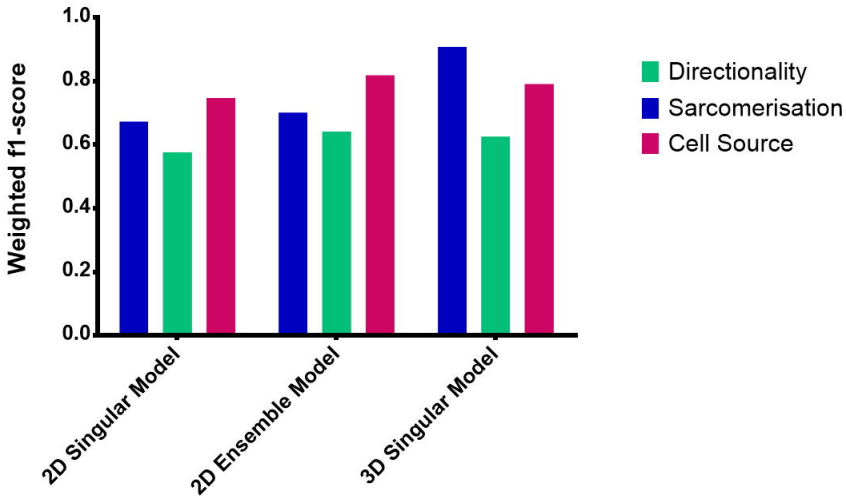
(d)



(e)



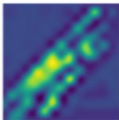
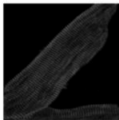
(f)



original image:

IGrad for class 4

Grad-CAM for class 4

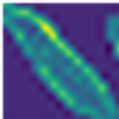
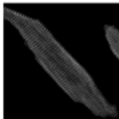


(a)

original image:

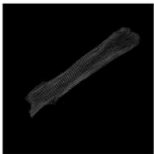
IGrad for class 4

Grad-CAM for class 4

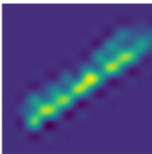


(b)

original image:

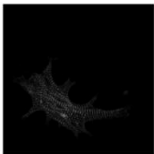


Grad-CAM for class adult

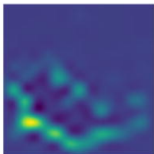


(a)

original image:

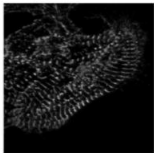


Grad-CAM for class neonatal

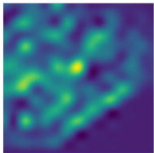


(b)

original image:



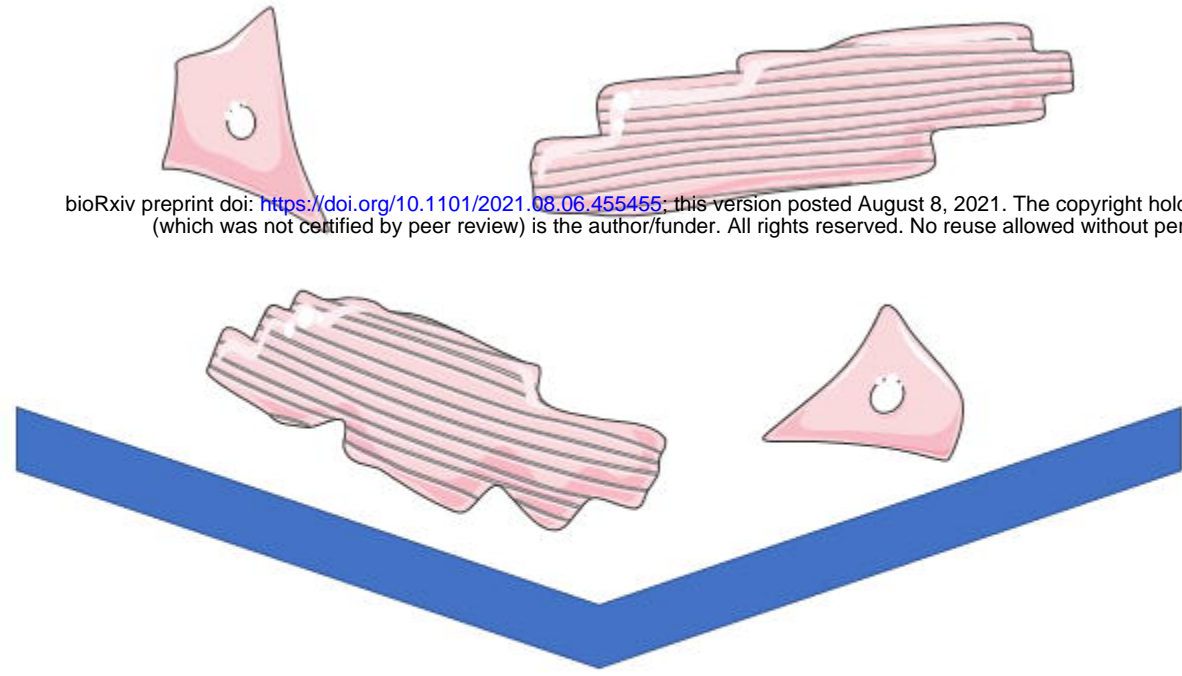
Grad-CAM for class ipsc



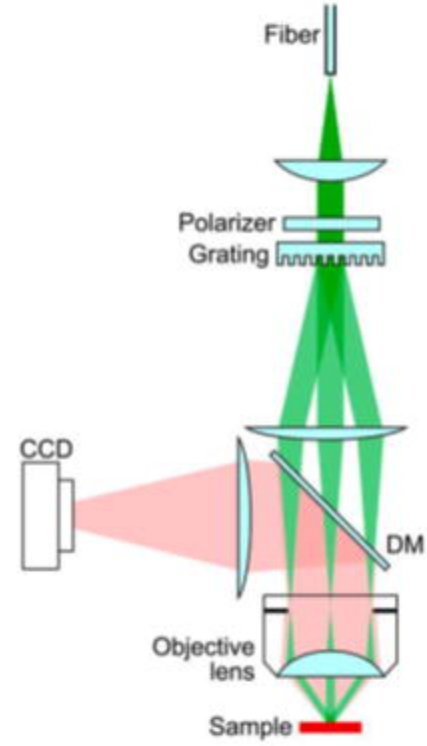
(c)

Cardiomyocytes

bioRxiv preprint doi: <https://doi.org/10.1101/2021.08.06.455455>; this version posted August 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



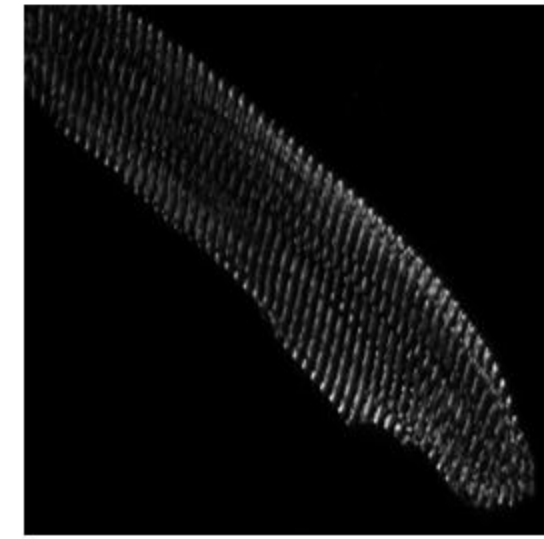
3D-SIM



Rated 3D Stacks

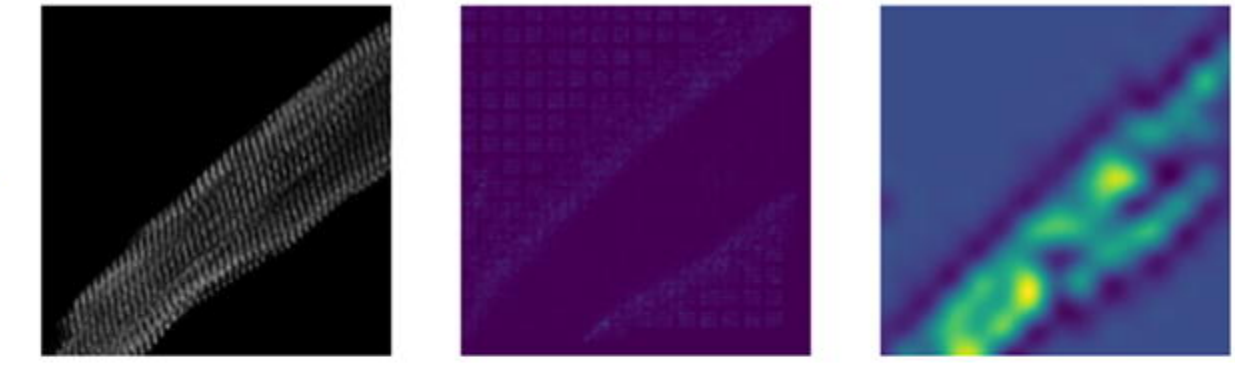


Rated 2D Images



Maximum Illumination Projection

Explainability Analysis (IGrad, Grad-CAM)

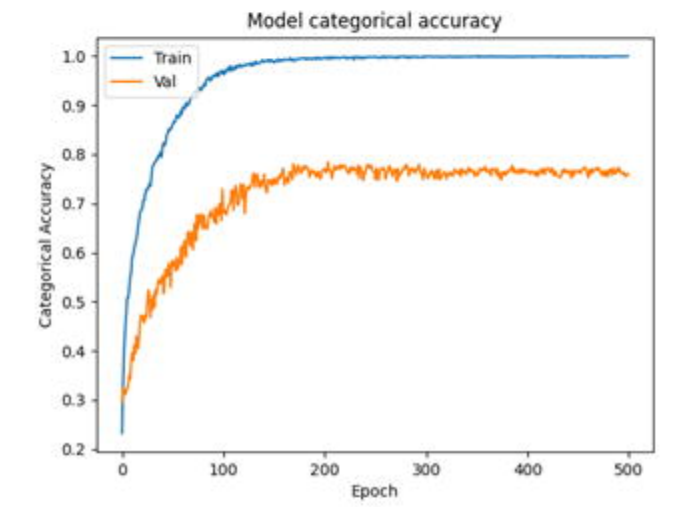


Explainability Heatmaps

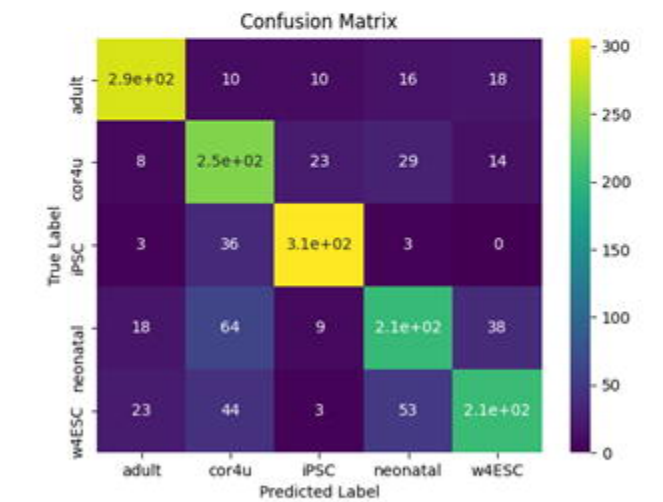
Convolutional Neural Networks

2D

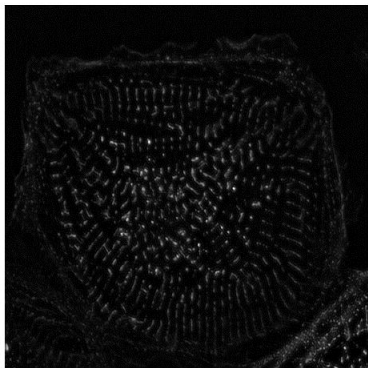
3D



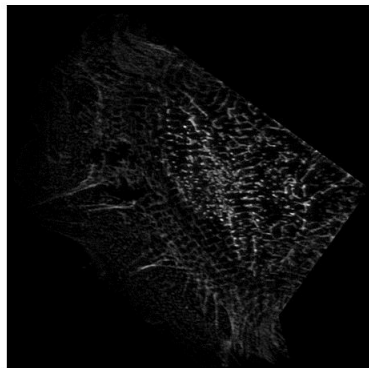
Accuracies



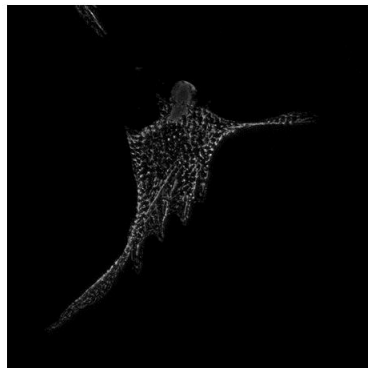
f1-Scores



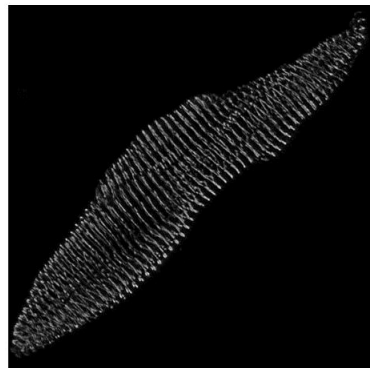
(a)



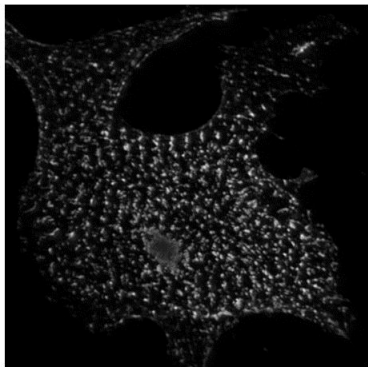
(b)



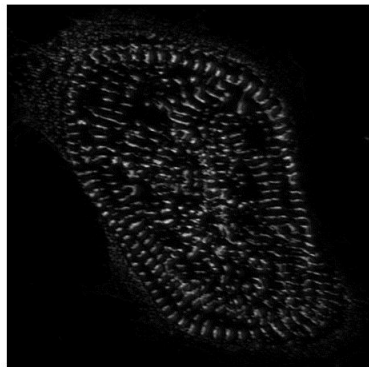
(c)



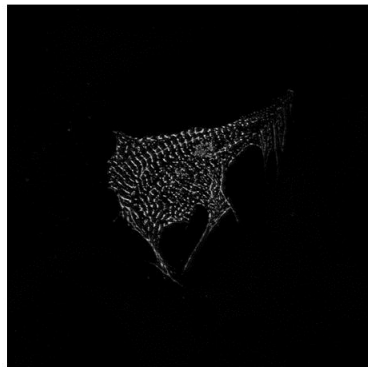
(d)



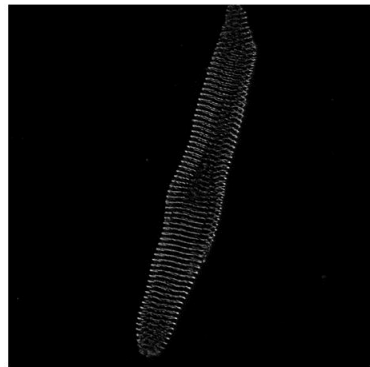
(e)



(f)

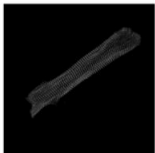


(g)



(h)

original image:

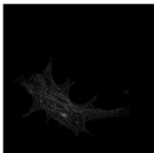


IGrad for class adult

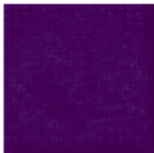


(a)

original image:

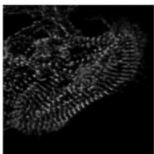


IGrad for class neonatal

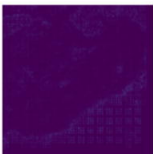


(b)

original image:



IGrad for class ipsc



(c)