Peer Review File

Article information: http://dx.doi.org/10.21037/atm-20-1806

Comment 1: ***A major disadvantage of the manuscript is a lack of essential information crucial for implementation of the proposed methods (DenseNets and feature engineering in combination with random forest and Xgboost). In the case of the DenseNet, the following must be included into the manuscript: used optimizer and its setting (learning rate, stopping criteria, minibatch size, …), and used augmentation methods and their settings. DenseNet topology allows control of a model compactness via a compression factor. Please provide also information about its setting. In the case of the random forest and Xgboost, add information about settings of all hyperparameters. Specify a loss function used within training processes.***

Reply 1: Thank you for your constructive suggestion. We felt so sorry that the description of the model parameters was obscure that some essential information lacked. As for DenseNet, the optimizer is Adam and the compression factor of DenseNet is 0.5. The training of DenseNet is controlled by the epoch without stopping criteria. Other hyperparameters of DenseNet, such as learning rate, minibatch, are shown in Table 1 in the current letter.

Table 1 The hyperParameters of DenseNet.

HyperParameters	Values	
Learning rate	0.005	Epoch<50
	0.0005	50≤Epoch＜100
	0.00005	100≤Epoch
Epoch	150	
Training BatchSize	20	
Testing BatchSize	8	

Details about Augmentation methods are as following: In the training process, an online augmentation method was used to increase the number of the samples, which employed horizontal flip, vertical flip and horizontal vertical flip to generate new samples. The number of samples was quadrupled by this way. At each batch, the augmentation method was applied on the samples before training. We have added descriptions about augmentation methods at **line 266-270** in the manuscript.

The MSE loss function was used within training processes of DenseNet. n is the number of the samples, ${y_i}^h, {y_i}^b$ represent the labels of EHL and EBL respectively, ${\hat{y}_i}^h, {\hat{y}_i}^b$ represent the predictive values of EHL and EBL respectively.

$$Loss = \frac{1}{2n}\sum_{i=1}^{n}(({y_i}^b - {\hat{y}_i}^b)^2 + ({y_i}^h - {\hat{y}_i}^h)^2)$$

The hyperparameters of Xgboost and Random Forest are showed in Table 2 and Table 3 (hyperparameters of EBL and EHL are the same in Random Forest). Thanks again.

Table 2 HyperParameters of Xgboost

HyperParameters	EBL	EHL
Learning rate	0.1	0.05
Number of estimators	600	300
Max depth	3	4
Min child weight	4	6

Subsample	0.6	0.6
Gama	0.1	0.1
Reg_alpha	0.1	0.1
Reg_lambda	0.1	0.1
Eval_metric	RMSE	RMSE

Table 3 HyperParameters of Random Forest

HyperParameters	**Values**
Number of estimators	100
Min_samples_split	2
Criterion	MSE
Min_samples_leaf	1
Min_impurity_decrease	$1e^{-07}$

Change in the text: We have rewritten the part of "***Establish models based on Dense Network method***" in the method part and have revised the supplementary materials to make the description clearer and more precise for readers. As for Comment 1, we added description about the parameters of DenseNet and augmentation methods in the manuscript and hyperparameters of Xgboost and RF in the supplementary materials. Main changes are as following:

(1) We **added** "*The MSE loss function was used within training processes of DenseNet. n is the number of the samples,* y_i^h, y_i^b *represent the labels of EHL and EBL respectively,* $\hat{y}_i^h, \hat{y}_i^b$ *represent the predictive values of EHL and EBL respectively.* $Loss = \frac{1}{2n}\sum_{i=1}^{n}((y_i^b - \hat{y}_i^b)^2 + (y_i^h - \hat{y}_i^h)^2)$" at **line 306-309** in the **Method** part of the manuscript.

(2) We **added** "*The growth rate of DenseNet was 4. The optimizer is Adam and the compression factor of DenseNet is 0.5. The training of DenseNet is controlled by the epoch without stopping criteria. The maximum epoch is 150. When the epoch is below 50, the learning rate is 0.005; when the epoch is between 50 to 100, the learning rate is 0.0005; When the epoch is between 100 and 150, the learning rate is 0.00005. The training Batchsize is 20, while the testing batchsize 8*" at **line 310-315** in the **Method** part of the manuscript.

(3) We **added** "*Since more images would be required for deep learning, we did image augmentation. In the training process, an online augmentation method was used to increase the number of the samples, which employed horizontal flip, vertical flip and horizontal vertical flip to generate new samples. The number of samples was quadrupled by this way. At each batch, the augmentation method was applied on the samples before training*" at **line 266-270** in the **Method** part of the manuscript.

(4) We **added** "T*he hyperparameters are fine-tuned through training dataset. And the detailed information about hyperparameters of Xgboost and Random Forest for EBL and EHL were presented in the supplementary materials(shown in supplementary Table 2 and Table 3)*" at **line 260-263** in the **Method** part of manuscript. We **added** Table 2 and Table 3 in the response letter as Supplementary Table 2 and Table 3 in the **Supplementary materials** at **line 55-57.** We also **added** "*Supplementary Table.2 and supplementary Table.3 show the hyperparameters of Random Forest and Xgboost used in feature engineering method*" at **line 48-49** in the **Supplementary materials.**

Comment 2: ***The topology of the presented DenseNet remarkably differs from topologies proposed by Huang et al. (If this is your intention, please state it in the text). Input of DenseNets proposed by Huang et al. consists of a convolution followed by a batch normalization (BN), ReLU and max pooling (https://github.com/flyyufelix/DenseNet-Keras/blob/master/densenet121.py). In your topology, only convolution is used (line 199).*** **According to Huang et al., a transition layer consists of BN-1x1 Conv followed by 2x2 average pooling (11). In your topology, only convolution and average pooling are used (lines 200 and 201). Further, modify the sentence at the lines 201-202 as follows: "At the end of the network, feature layers produced by the last dense block were concatenated, bath normalized, and then, …". Specify the output of the network (as far as I understand, the output is a 1D vector that consists of two values).**

Reply 2: Thank you for your kindly reminder very much. We felt sorry that our description was obscure and imprecise to make reviewers and readers confuse. In fact, the DenseNet in this paper contains four dense blocks and three transition layers. The topology of the presented DenseNet is modified from topologies proposed by Huang et al. The transition layers consist of *batch normalization (BN), ReLU,* 1×1 convolution and 4×4 average pooling. There were three modifications in our DensNet model. First of all, the input in the presented DenseNet consisted one 3×3 convolutional layer. Second, the kernel size of average pooling layers in the presented DenseNet was 4×4. The two modifications would reduce the number of parameters in the DenseNet. Third, as with your comments, the linear layer at the end of the presented DenseNet outputted 1D vector that consists of two values, which is different with DenseNet by Huang et al. As with your comments, the linear layer at the end of the presented DenseNet outputted 1D vector that consists of two values. The purpose of this modification is to simultaneously output the predictive values of EHL and EBL. The experimental results show that these modifications can get a satisfactory predictive performance. So we revised the description as "*At the end of the last dense block, the architecture consisted of a bath normalization and rectified linear unit followed by a 4×4 average pooling layer, and then, a linear layer was attached to obtain the output. The output is 1D vector that consists of two values which are the predictive values of EBL and EHL*" as you suggested at **line 290-294** in the **Method** part. Thank you.

Change in the text: We have rewritten the part of "*Establish models based on Dense Network method"* in the Method part and have revised the supplementary materials to make the description clearer and more precise for readers. As for Comment 2, we revised descriptions about architecture of the DenseNet. The main changes are as following:

(1) We **deleted** the text "*and used data augmentation methods to increase the number of samples* (**at line 286-287**)", "*that each had two layers in DN* (**at line 289-290**)", "*The growth rate of DN was 4* (**at line 292**)", "*At the beginning, a 3×3 convolution with 8 output channels was performed on the input images*(**at line 294-295**)", "*Between two contiguous dense blocks, a transition layer with a 1×1 convolution and 2×2 average pooling was set. At the end of the last dense block, bath normalization was performed, and then, a linear layer was attached to obtain the estimated value of the image. At the end of the last dense block, bath normalization was performed, and then, a linear layer was attached to obtain the results of EBL and EHL* (**at line 300-305**)" in the **Method** part of the manuscript.

(2) We **added** the text "*To reduce the number of parameters in the DenseNet, we did some modifications in DenseNet in this paper. In detail, the input in the presented DenseNet consisted*

one 3×3 convolutional layer. Second, the kernel size of average pooling layers in the presented DenseNet was 4×4 (**at line 282-285**)", "*At the beginning, a 3×3 convolution with 8 output channels was performed on the input images* (**at line 287-288**)", "*The dense block contains two layers* (**at line 290**)", "*The transition layers consist of batch normalization (BN), ReLU, 1×1 convolution and 4×4 average pooling. At the end of the last dense block, the architecture consisted of a bath normalization and rectified linear unit followed by a 4×4 average pooling layer, and then, a linear layer was attached to obtain the output. The output is 1D vector that consists of two values which are the predictive values of EBL and EHL (shown in Table 1)* (**at line 295-300**)" in the **Method** part of the manuscript.

(3) We **revised** "*There were four dense blocks* (**at line 288-289**)" as "*Then, there were four dense blocks and three transitional layers in DenseNet* (**at line 288-289**)".

Comment 3: ***Training and evaluation of all presented solutions must be performed on different (independent) datasets. Provide details related to the training and evaluation processes.***

Reply 3: Thank you very much for the good comments. We quite agree with your comments and we trained and tested our data on independent datasets. In brief, the dataset was randomly divided into training dataset and testing dataset and the hyperparameters are fine-tuned through training dataset. 10-flod cross validation was used in the experiment. First, we arranged the samples in ascending order according to the gradients of blood loss. Second, we randomly shuffled the order of the samples and divided them into 10 equal subsets. The labels of divided subsets showed that each subset contains all the gradients blood loss. While the proportion among the number of samples corresponding to the gradients of blood loss was slightly different in each subset. Each subset is then used once as a testing set while the 9 remaining subsets form the training set. As a consequence, the ratio of the number of samples between the training dataset and the testing dataset was about 9:1. The feature engineering method and deep learning method used the same training dataset and testing dataset. The hyperparameters of Xgboost and Random Forest used in the training process are shown in Table 2 and Table 3 in the response letter. The optimizer used in DenseNet was Adam and the compression factor of DenseNet is 0.5. The training of DenseNet was controlled by the epoch without stopping criteria. Other hyperparameters of DenseNet, such as learning rate, minibatch, are shown in Table 1 in the current letter. Thanks again.

Change in the text: We **added** descriptions about datasets and training process in the **Method** part of the manuscript. In detail, we **added** "*After image preprocessing, the dataset was randomly divided into training dataset and testing dataset. In the current study, 10-flod cross validation was used to form the training and testing datasets for methods based on the feature engineering and deep learning. First of all, we arranged the samples in ascending order according to the gradients of blood loss. Then, we randomly shuffled the order of the samples and divided them into 10 equal subsets, in which contained all the gradients blood loss. While the proportion among the number of samples corresponding to the gradients of blood loss was slightly different in each subset. Afterwards, each subset was then used once as a testing dataset while the 9 remaining subsets formed the training dataset. As a consequence, the ratio of the number of samples between the training dataset and the testing dataset was about 9:1* (at **line 241-252** in the ***Establish models based on Feature Engineering method***)" and "*Images were* *divided into training and testing datasets as previously described* (at **line 265** in the ***Establish models based on Dense Network method***)".

Comment 4: ***The paragraph at the lines 224-226 should be split-up and moved on appropriate places in section "Methods". At the line 152, add information about lighting conditions. At the line 108, add reference on an image recognition system based on DenseNets (e.g. 10.1371/journal.pone.0216720). Further, I recommend use of the conventional shortcut DenseNet instead of DN. Optionally, add a table or figure clarifying the topology of the DenseNet (including spatial dimensions of feature maps). At the line 276, check the small r.***

Reply 4: Greatly appreciate your valuable comments. We **deleted** the paragraph at the lines 224-226 (**at line 353-356 in the revised vision**) and revised the description in the section "**Methods**" with trace of modifications (detailed information was shown in the **changes in the text**). In the image acquisition step, we took picture in the normal illumination for the operation room. The corresponding description was revised and marked red **at line 213**. We quite agree with you that this article (***Skrabanek P, Zahradnikova A. Automatic assessment of the cardiomyocyte development stages from confocal microscopy images using deep convolutional networks. PLOS ONE, 2019, 14(5). DOI: 10.1371/journal.pone.0216720***) should be cited to indicate the promising role of DenseNets in image recognition. Meanwhile, we use the conventional shortcut DenseNet instead of DN as you suggested. Moreover, we added a table to make the topology of the DenseNet more clear and friendly to readers (shown in Table 4 in this response letter and in Table 1 in the revised manuscript). The r and the related description at the **line 276 (line 429-432 in the revised vision)** may be not so proper here, so we deleted it. Thank you for your kindness again.

Table 4 The detailed architecture of DenseNet.

Layers	Composition	Output Size
Convolution	3×3 conv, stride 1,padding 1	8×256×256
Dense Block	[BatchNormalization ReLU 3×3 conv, stride 1,padding 1] ×2	16×256×256
Trainsition Layers	BatchNormalization ReLU 1×1 conv, stride 1	8×256×256
	4×4 average pool, stride 4	8×64×64
Dense Block	[BatchNormalization ReLU 3×3 conv, stride 1,padding 1] ×2	16×64×64
Trainsition Layers	BatchNormalization ReLU 1×1 conv, stride 1	8×64×64
	4×4 average pool, stride 4	8×16×16
Dense Block	[BatchNormalization ReLU 3×3 conv, stride 1,padding 1] ×2	16×16×16
Trainsition Layers	BatchNormalization	8×16×16

	ReLU 1×1 conv, stride 1	
	4×4 average pool, stride 4	8×4×4
Dense Block	[BatchNormalization ReLU 3×3 conv, stride 1, padding 1] ×2	16×4×4
Regression Layer	BatchNormalization ReLU 4×4 average pool, stride 4	16×1
	Linear	1×2

ReLU, Rectified Linear Unit.

Change in the text: As for Comment 4, we did some main changes as following:

(1) We **deleted** "*Finally, we collected non-recycled blood form 34 surgical patients. Therefore, 569 portions of blood-soaked images were employed for feature engineering. For deep learning, the images were augmented so that 2274 blood-soaked images were analysed by DN methods*" at **line 353-356** in the **Results** part. And we **revised** the description as "*Finally, we collected non-recycled blood form 34 surgical patients. Therefore, 569 portions of blood-soaked images were employed for feature engineering* (at **line 218-220**)", "*Since more images would be required for deep learning, we did image augmentation. In the training process, an online augmentation method was used to increase the number of the samples, which employed horizontal flip, vertical flip and horizontal vertical flip to generate new samples. The number of samples was quadrupled by this way. At each batch, the augmentation method was applied on the samples before training* (at **line 266-270**)" in the **Method** part of the manuscript.

(2) We **added** the text "*in the normal illumination for the operation room*" at **line 213** in the **Method** part of the manuscript.

(3) We **added** a table about architecture of DenseNet as Table 1 at **line 572** in the manuscript.

(4) All the shortcut DN was replaced by DenseNet as reviewer suggested.

(5) We deleted "*According to the published work, r is used to describe the correlation between the estimated and actual EBL, which strongly indicates to us that the Triton System model is based on linear regression*" at **line 429-432** in the **Discussion** part.