Self-attention based BiLSTM-CNN classifier for the prediction of ischemic and non-ischemic cardiomyopathy

Kavita Dubey[1], Anant Agarwal[2], Astitwa Sarthak Lathe[1], Ranjeet Kumar[3] and Vishal Srivastava[1]*

1 Electrical and Instrumentation Engineering Department, Thapar Institute of Engineering and Technology, Patiala, India
2 Electronics and Communication Engineering Department, Thapar Institute of Engineering and Technology, Patiala, India
3 University of St Andrews, Scotland

***Corresponding author:** vishalsrivastava17@gmail.com

Abstract: Approximately 26 million individuals are suffering from heart failure, according to the global annual report. Despite higher inter-rater variability, endomyocardial biopsy (EMB) is still regarded the gold standard for assessing heart failure. Therefore, we proposed and implemented a new unified architecture consist of convolutional layers, bidirectional LSTM (BiLSTM), and self-attention mechanism to predict the ischemic and non- ischemic cardiomyopathy using histopathological images. The proposed model is based on self-attention that implicitly focus to the information outputted from the hidden layers of BiLSTM. Through our results we demonstrate that this framework carries high learning capacity and is able to improve the classification performance.

***Keywords*:** Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Self-Attention, Histopathological Images and Hear disease.

Introduction

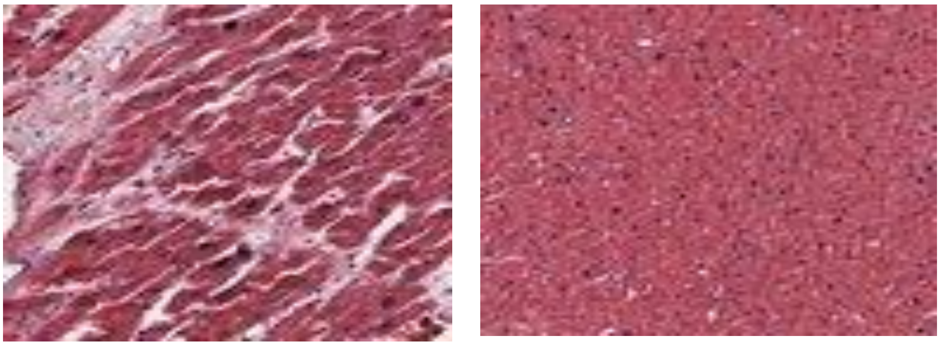
Cardiovascular disease (CD) is major cause of heart failures and related causalities worldwide. According to the survey done by Centres for Disease Control (CDC) in 2011, CD is the foremost cause of deaths in the United States, Australia, United Kingdom and Canada [1]. Heart failure is a severe, progressive clinical syndrome that results in inadequate systemic perfusion. It is evident through other common symptoms like arrhythmia (irregular heartbeats), myocardial infarction (commonly known as heart-attack), and coronary heart disease [3]. Cardiomyopathy is a disease related to hardening of the heart muscle and is considered the root cause of heart failure [2]. Depending upon causing condition, cardiomyopathy has been further divided into two categories ischemic and non-ischemic cardiomyopathy [4]. Ischemic cardiomyopathy is a coronary heart disease caused by left ventricle dysfunction due to chronic absence of oxygen to the heart muscle. On other hand non- ischemic cardiomyopathy is not caused by coronary artery disease and is often associated with the organ illnesses and exhibits common symptoms such as; breathlessness, sweating, fast breathing, high level of fatigue etc. [5]. Cardiomyopathy leads to approximately one third of all clinical heart failure cases. Therefore, timely diagnosis with high accuracy and inter-rater acceptability is pertinent to prevent the CD related casualties. Advanced imaging modalities have allowed to diagnose the cause associated with heart-diseases and are detrimental to decide the course of appropriate treatment. For example, myocardial perfusion imaging and stress echocardiography are commonly used non-invasive imaging modalities to assess for the prognosis of diversified ischemic heart disease [6]. However, movement of coronary arteries during the cardiac cycle causes motion-induced drift which is confined to a very small spatial dimension and must be captured using high spatio-temporal resolution imaging technique [7, 8]. In this context, non-invasive imaging modalities like compute tomography (CT), magnetic resonance imaging (MRI) are preferred choices because of improved image quality and diagnostic accuracy [9, 10]. In addition to CT and MRI, an endomyocardial biopsy (EMB) is considered in some instances complementary; or is the only procedure for diagnosing both ischemic and non-ischemic conditions in unexplained cardiomyopathy cases [11-13]. The current pathological examination depends upon the conviction of pathologists and hence ample room for inter-rater variability [14]. The quantitative interpretation of pathology images is very important for an accurate diagnosis [15]. At present, quantitative analysis in standard clinical environment is a time-consuming affair and carries high inter-rater variability, and so an automated quantification algorithm fostering diagnostic outcomes with universal acceptability is urgently required.

Recent, burgeoning and impeccable applications of deep learning in the field of medicine and digital pathology provides an exclusive opportunity to learn about hidden patterns that may not be visible by human examination [14]. In fact, deep learning architectures have been shown to be successful in the automated classification and segmentation of disease in digital histopathology [15]. The benefit of using deep learning models is that they get familiar with the most appropriate representation in a progressive way as part of the training process. However, the traditional CNN's output layer is fully connected to the hidden layer but it includes inefficient or multiple kernel which extract the same information repeatedly or trivial information. To improve the accuracy further, network can be enlarged by adding more convolutional kernels, convolutional layers, and pooling layers but it will increase the computational cost and also there is a chance of overfitting [16, 17]. LSTM is more appropriate to handle time-dependence problems [18]. It can filter and fuse the empty input, similar information, and irrelevant information extracted from the convolutional kernels, so that the effective and relevant information can be stored for a long time in the state cell. Benefit of a memory cell and controlling gates is that it prevents gradient from vanishing too quickly thereby propagating information without loss [19-21]. But LSTM only exploits the preceding or past information. BiLSTM is an advancement of LSTM in which forward hidden layer is combined with backward hidden layer, that can access both the preceding and subsequent information. The vector representation of high intra-variability and sparsity causes histopathological images of high-dimensional vector [22]. The high-dimensional vector acting as the BiLSTM input will raise the number of network parameters and thus making it harder to optimize the network. The convolutional operation reduces the dimensionality of the data [23-25]. Although, BiLSTM with CNN can obtain the contextual information, but it can't focus on the relevant information in the contextual information acquired [26]. Attention mechanism by setting the distinct weights can highlight the important information from the acquired information. The mechanism of attention is an efficient method which enables a model to pay more attention to significant information. The integration of BiLSTM-CNN with self-attention mechanism can further enhance the performance of the model.

In this manuscript, we reported a novel framework based on CNN (Inception V3) and BiLSTM with self-attention for the prediction of ischemic and non-ischemic cardiomyopathy. The proposed framework has an intrinsic self-attention ability i.e., using the feature maps acquired from high-layers as the attention mask of a low layers, rather than learning the attention mask with additional layer. Results reveals that the proposed model attain high performance with histopathology images than the conventional CNN approaches.

Methodology

Stage 1. Preparation of Input Image: Images of the size of 96 x 96 pixels are considered as an input image to the network. High-resolution histopathological images which can easily annotated by the pathologist are used to generate a binary mask. The approach is associated with binary image categorization, where for each case features like textural properties, colour or spatial features. In our binary mask, '1' is assigned to ischemic cardiomyopathy and '0' is assigned to non- ischemic cardiomyopathy. Dataset was split randomly, i.e., 100 images were used for training, 80 images were used for validation and 29 images were held out as an independent test set.



(a) (b)

Fig. 1 (a) Ischemic and (b) Non-ischemic histopathology images of heart.

Model Description: Structure of the model presented in this manuscript was implemented in KERAS. The architecture of the model consists of Inception-V3, BiLSTM cells and self-attention mechanism and FCNN as shown in the Fig. 2. Input to the inception-V3 is an image of the size of 96 x 96 pixels. Inception-V3 consists of convolutional neural network (CNN) layer is used to extract the features. The goal of the inception module is to attain a “multi-level feature extractor” by computing 1×1, 3×3, and 5×5 convolutions within the same module of the network. By reducing the number of features involved without the loss of information, CNN can achieve an appropriate fitting to the image dataset. Along the channel dimensions, the output of these filters is stacked before being fed into the next layer of the network. To avoid the problem of overfitting, layers are followed by a dropout layers during training with dropping probability of 0.5 and Kernel regularization is done by using L2 regularization with λ= 0.01[22]. Further, data augmentation is used to enlarge the size of datasets as well as to ameliorate the overfitting issue [23]. We perform augmentation by flipping, rotation and zoom scaling without altering the actual features of images. Padding is done to ensure the same size of input as well as output. *Tanh* function is used as activation functions in all CNN the layers of the network. Adam optimizer with a learning rate of 0.0001 and decay-rate of 0.000001 is used as optimizing function for the proposed model. The model is trained until convergence is achieved. Reshaping process is required before giving the output of CNN as an input to the BiLSTM. CNN have 2-D representation at its output which is reshaped to 1D array before giving as an input to BiLSTM which accepts 1D at its input. The memory cell of BiLSTM used to store the information and is accessed by different controlling gates (like input gate, output gate, forget gate) [24-26]. In the proposed model the input to the BiLSTM is 1D array of size 1 x 2048. By knowing the input, past, future states of its local neighbours, Bi-LSTM can predict the present state. Dropout of 0.5 is applied on the layer of Bi-LSTM (input, output as well as on the image representation vector).

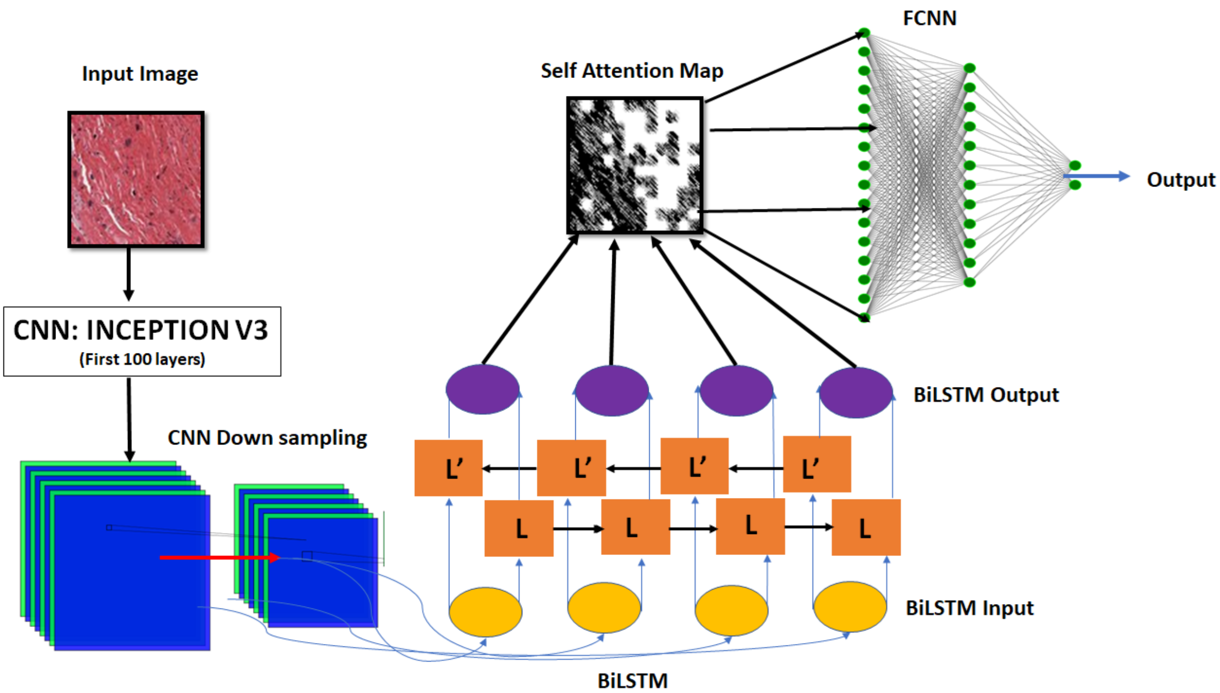


Fig.2. Architecture of the proposed model, L & L' : Layers.

In the next step, to compute a context vector for the decoder containing most useful information from all hidden states of the encoder, averaging of weights is done and the process is called the attention process. When process is applied to every position of the source sequence, it is called a self-attention mechanism. A self-attention model with a 1-layer stacked configuration is used in the network structure. It works with the aim to find the effectiveness during classification in terms of accuracy, sensitivity and specificity. Query, key and value are the three vectors on which self-attention model works. By applying learned linear projections these vectors can be created. The output of the BiLSTM is taken as an input to the self-attention model. To obtain the query, key, and value, three feedforward layers were used. For 1-layer stacked configuration the process needs no repetition. Attention width of 256 is used and Sigmoid is used as an activation function in the self-attention mechanism. By averaging out the last layer (output) of self-attention and then applied to feed-forward layer a self-attention map is created, which will be used for future classification. In the end, the image representative vector is passed through the dense layer. *Tanh* is used by all feed-forward layers as an activation function and the final prediction is achieved using the SoftMax function.

All the experiments were executed on single GTX1080 Ti, with a 3.3 GHz and 32 Gb RAM and i7 processor. The model was implemented in KERAS 2.0.8 with Tensorflow 1.7 backend, using cuDNN 5.1.5 and CUDA 9.1.

Results & Discussion

The dataset is downloaded from https://idr.openmicroscopy.org/webclient/?experimenter=-1 [27], which consist two type of patients left ventricular heart tissue, the one with end-stage heart failure who are clinically diagnosed either ischemic cardiomyopathy or other non-ischemic cardiomyopathy. The dataset consists both male and female. The average age of patients suffering from end-stage heart failure and is 53.4 ± 15.3 years and without heart failure is 55.5 ± 11.6 years. During the collection stage, all patients with tissue sectioned, scanned and scanned were included for investigation. The proposed self-attention based BiLSTM-CNN was used for an automated classification heart tissue. In order to achieve so, we first differentiate the ischemic/ non- ischemic histopathology images from normal one. Further, to evaluate the performance of proposed model, parameters like accuracy, sensitivity, and specificity were calculated. On a training set of 100 images (both ischemic and non-ischemic) inception-V3 is trained as a binary classifier. The trained model is then validated and tested on 80 and 29 new images of ischemic and non-ischemic heart disease to evaluate the performance of the developed model. Two models were tested on the datasets. First an Inception-V3 model is tested on the present dataset and after that proposed based on Inception-V3, BiLSTM and self-attention were also tested. To evaluate the performance of the model kernel size varies and to capture spatiotemporal correlations larger state-to-state kernels are appropriate. Whereas, for larger kernels, later states have larger receptive fields and are related to a wide range of input. Finally, in order to test the generalization ability of the model we tested it by feeding some other images as well. Although, self- attention is used to generate attention aware features by using attention modules consisting of feed-forward structures. Attention aware features are directly dependent on depth (i.e. depth of the sample used) which changes as depth increases. Based on the training, validation and testing, the model based on Invception-V3, BiLSTM and self-attention achieved excellent performance in distinguishing the images (ischemic and non-ischemic). At each stage, the mean absolute response of the output layer is calculated for a better understanding of self-attention model. In the proposed model we used attention one time for evaluation of images. As the self-attention mechanism is used to store relevant information by suppressing the noise, thus making optimization much easier and also helps in classifying the represented features. This technique is applied to train the network required for assessment. Thus, a new classifier model was developed using a pre-trained network which worked more effectively in differentiating histopathological images based on deep learning algorithm self-attention mechanism. Figure 3 (a) and (b) shows the performance curve of the proposed model during training and validation of the system, respectively. Figure 3 (c) shows the loss curve for the training and for the validation sets during training of the proposed model. By using adequate number of classified images during training the accuracy of the proposed model can reach 95 % for the validation datasets while discriminating ischemic and non-ischemic images based on histopathology images. Table 1 shows the comparative performance metrics of the different networks on the validation datasets. The proposed model has a better performance as compare to other conventional CNN model in terms of accuracy, sensitivity and specificity to predict the cardiac outcome.

Table 1. Performance metrics of different models on the validation datasets of H&E stained images.

Model	Accuracy	Sensitivity	Specificity
Inception-V3	91.25 %	94.59 %	88.3 %
Inception-V3 + LSTM	95 %	97.5 %	92.5 %
Proposed Model (Inception-V3 + BiLSTM + Self-Attention)	96.25 %	97.5 %	92.68 %

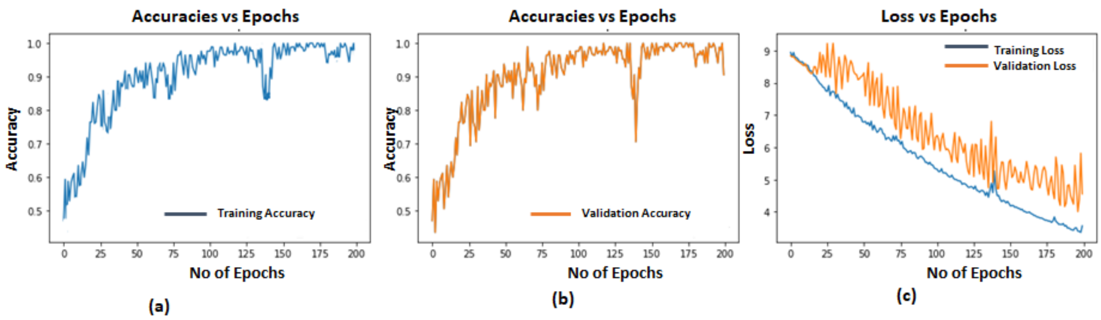


Fig. 3. Curves of (a, b) accuracy vs epochs (c) loss vs epochs for the training and validation dataset.

The proposed model achieved 93.10 % accuracy, 93.33% sensitivity and 86.66 % specificity for testing datasets. Although, our results reveal that we achieved much better performance but the use of CNN is limited and doesn't perform well when the size of available training dataset is very less and sometimes when there is very complex mix response. Figure 4 shows the histopathological images that were misclassified from all the models including the proposed one. Figure 4(a, b) is misclassified as ischemic and non-ischemic cardiomyopathy. The primary reason for the network failure is sometimes heart tissues have both ischemic and non-ischemic characteristic of cardiomyopathy, as a significant region of heart tissues is influence by both coronary artery disease as well as some organs dysfunction. This makes a classification between ischemic and non-ischemic cardiomyopathy hard for the network as well as pathologist.

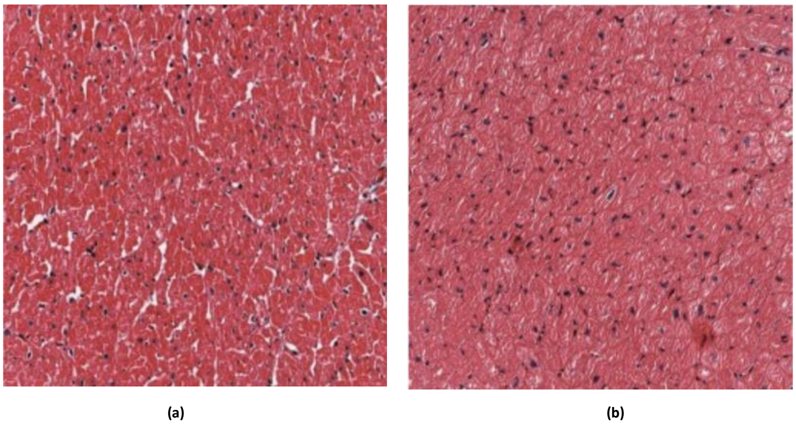


Fig. 4. Evaluation of misclassified histopathological images.

Conclusion

The proposed model gave promising results and outperforms the state of the art for an automatic classification of ischemic and non-ischemic cardiomyopathy using histopathological images. The model is an integration of inception-V3, BiLSTM and self-attention map to improve the accuracy of classification even for a small dataset. Data analysis using proposed model revealed that it has attained a high accuracy of 93.10 %, sensitivity 93.33 % and specificity of 86.66 % in classifying the ischemic and non-ischemic images for testing. Some misclassification has also been noted. In the assessment of cardiac failure, more study is needed to fully assess the potential of deep learning in clinical practice and its implications.

Disclosure: The author declares that there are no conflicts of interest.

Acknowledgement: Authors gratefully acknowledge the financial assistance from Science Engineering and Research Board, Government of India for the Project No. EMR/2016/000677.

REFERENCES

[[1] Kochanek KD, Xu J, Murphy SL, Minino AM, Kung HC. Deaths: final data for 2009. Natl Vital Stat Rep. 2011; 60(3):1±116. Epub 2011/12/29.
[2] Geleijnse ML, Elhendy A, Van Domburg RT et al., "*Cardiac imaging for risk stratification with dobutamine atropine stress testing in patients with chest pain*", *Circulation*, 137–47, 96, 1997.
[3] Falk RH., "Diagnosis and management of the cardiac amyloidosis", *Circulation*, 2047–60, 112, 2005.
[4] Beltrami CA, Finato N, Rocco M, Feruglio GA, Puricelli C, Cigola E, et al., "Structural basis of end-stage failure in ischemic cardiomyopathy in humans", *Circulation*, 151–63, 89(1), 1994.
[5] Dubrey SW, Burke MM, Hawkins PN, Banner NR., "Cardiac transplantation for amyloid heart disease: The United Kingdom experience", *J Heart Lung Transplant*, 1142–53, 23, 2004.
[6] Gu¨nalp B, Dokumaci B, Uyan C et al., "Value of dobutamine technetium-99m-sestamibi SPECT and echocardiography in the detection of coronary artery disease compared with coronary angiography", *J Nucl Med*, 889–94, 34, 1993.
[7] Bax JJ, Poldermans D, Elhendy A et al., "Sensitivity, specificity, and predictive accuracies of various non-invasive techniques for detecting hibernating myocardium", *Curr Probl Cardiol*., 141–86, 26, 2001.
[8] Berry C, Kellman P, Mancini C, Chen MY, Bandettini WP, Lowrey T, et al., "Magnetic resonance imaging delineates the ischemic area at risk and myocardial salvage in patients with acute myocardial infarction", *Circ Cardiovasc Imaging*, 527–35, 3(5), 2010.
[9] Fieno DS, Kim RJ, Chen EL, Lomasney JW, Klocke FJ, Judd RM. "Contrast-enhanced magnetic resonance imaging of myocardium at risk: distinction between reversible and irreversible injury throughout infarct healing", *J Am Coll Cardiol*., 1985–91, 36(6), 2003.
[10] Carlsson M, Ubachs JF, Hedstrom E, Heiberg E, Jovinge S, Arheden H. "Myocardium at risk after acute infarction in humans on cardiac magnetic resonance: quantitative assessment during follow-up and validation with single-photon emission computed tomography",
[11] Ardehali H, Kasper EK, Baughman KL., "Diagnostic approach to the patient with cardiomyopathy: whom to biopsy", *Am Heart J*., 149, 7(12), 2012.
[12] Sloan KP, Bruce CJ, Oh JK, Rihal CS., "Complications of echocardiography-guided endomyocardial biopsy", *J Am Soc Echocardiogr*., 2009; 321-324, 22(3), 2009.
[13] Ghaznavi, F., Evans, A., Madabhushi, A., and Feldman, M., "Digital imaging in pathology: Whole-slide imaging and beyond," *Annual Review of Pathology: Mechanisms of Disease*, 331–359, 81(1), 2013.
[14] Bengio, Y., Courville, A., and Vincent, P., "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1798–1828, 35(8), 2013.
[15] Cruz-Roa, A., Arevalo, J., Madabhushi, A., and Gonzalez, F., "A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection," *in Medical Image Computing and Computer-Assisted Intervention*, 403–410, 2013.
[16] Jurgen Schmidhuber, "Learning complex, extended sequences using the principle of history compression", *Neural Computation*, 234–242, 4(2), 1992.
[17] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh., "A fast learning algorithm for deep belief nets", *Neural computation*, 1527–1554, 18(7) 2006.
[18] Sepp Hochreiter and Jurgen Schmidhuber, "Long short-term memory", *Neural Computation*, 1735– 1780, 9(8), 1997.
[19] S. Hochreiter and J. Schmidhuber. "Long short-term memory", *Neural Computation*, 1735–1780, 9(8), 1997.
[20] A. Graves., "Generating sequences with recurrent neural networks", *arXiv preprint arXiv*:1308, 8(50), 2013.
[21] Y. Bengio, I. Goodfellow, and A. Courville. Deep Learning. Book in preparation for MIT Press, 2015.
[22] Felix A. Gers, Jurgen Schmidhuber, and Fred Cummins, "Learning to forget: Continual prediction with ¨ LSTM", *In ICANN*, 850–855, 2, 1999.
[23] LeCun Y, Bengio Y, Hinton G., "Deep learning", *Nature*, 436-44, 521(7553) 2015.
[24] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, 1929–1958, 15(1), 2014.
[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *NIPS*, 1097–1105, 2012.
[26] Wang, Xiaolong, Kaiming He, and Abhinav Gupta. "Transitive invariance for self-supervised visual representation learning." *Proceedings of the IEEE international conference on computer vision*. 2017.
[27] Nirschl, Jeffrey J., et al. "A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of H&E tissue." *PloS one* 13(4) 2018.