Machine Learning Algorithms for Classification of Microcirculation Images from Septic and Non-Septic Patients

Perikumar Javia[1], Aman Rana[1], Nathan Shapiro[2], and Pratik Shah[1†]

[1] MIT Media Lab
Massachusetts Institute of Technology
Cambridge, MA, USA
{*pjavia, arana, pratiks*}*@media.mit.edu*

[2] Emergency Medicine
Beth Israel Deaconess Medical Center
Boston, MA, USA
nshapiro@bidmc.harvard.edu

Abstract—Sepsis is a life-threatening disease and one of the major causes of death in hospitals. Imaging of microcirculatory dysfunction is a promising approach for automated diagnosis of sepsis. We report a machine learning classifier capable of distinguishing non-septic and septic images from dark field microcirculation videos of patients. The classifier achieves an accuracy of 89.45%. The area under the receiver operating characteristics of the classifier was 0.92, the precision was 0.92 and the recall was 0.84. Codes representing the learned feature space of trained classifier were visualized using t-SNE embedding and were separable and distinguished between images from critically ill and non-septic patients. Using an unsupervised convolutional autoencoder, independent of the clinical diagnosis, we also report clustering of learned features from a compressed representation associated with healthy images and those with microcirculatory dysfunction. The feature space used by our trained classifier to distinguish between images from septic and non-septic patients has potential diagnostic application.

Index Terms—machine learning, convolutional neural networks, sepsis, microcirculation, t-SNE embeddings, feature saliency

I. Introduction

Sepsis is a life-threatening disease where the host response to an infection leads to inflammation that may result in multiple organ failure and death of the patient [1]. Sepsis is associated with very high mortality rates that make early detection crucial for treatment. Microcirculation is the flow of blood in the smallest elements of the cardiovascular system, the capillary network, where the exchange of oxygen takes place and is pivotal in the maintenance of homeostasis [2]. Dysfunctions in microcirculation system, such as changes in flow of blood in the capillary network, constriction of capillary vessel etc. manifest early in the pathophysiology of sepsis. Previous studies demonstrate as association between microcirculatory derangements and sepsis severity and outcomes [3] [14].

Improvements in microcirculation function after early resuscitation were associated with reduction in subsequent organ failure [3] [16] [13]. This evaluation of microcirculation function holds potential for treating patients with sepsis. Invasive

methods such as serum lactate physiology are often used as indirect measures for the evaluation of the microcirculatory function [3]. Direct clinical evaluation of microcirculatory dysfunction using dark field imaging, a relatively fast and non-invasive approach, has been investigated by researchers for diagnosis and prognosis of sepsis. Measurements relating to capillary density and flow in side-stream dark field (SDF) videos have been used to evaluate microcirculatory dysfunction either by human raters, simple software systems or computer vision methods [3] [10] [11] [17].

Machine learning has been used successfully to automate diverse tasks such as recognition of objects and semantic segmentation in self-driving cars. Deep learning architectures such as convolutional neural networks (CNNs) can learn to segment distinguishing features in images with minimal human supervision [4]. CNNs are comprises of layers of stacked neurons where input features are transformed into higher dimensional representation using convolution. The convolution is often wrapped with non-linear activation function. Max pooling is usually used for downsampling and the final compressed high dimensional features are fed to softmax classifier for classification. CNNs have achieved high accuracy in skin cancer detection from RGB images, lung nodule classification using CT images and segmentation of periodontal diseases from oral images [5] [6] [7] [8]. Previous work has used deep learning and computer vision to investigate microcirculatory dysfunction in dark field imaging videos of pigs and other animals [9] [12]. To our knowledge no previous study has successfully used deep learning to detect microcirculation dysfunction from critically ill human subjects.

In this work, we implement a CNN to analyze microcirculatory dysfunction captured by dark field imaging in human patients and distinguish between septic and non-septic images with high accuracies. We also investigate outputs from the last convolutional layer in our architecture, and using a t-SNE embedding show that the representation learned by the classifier successfully differentiates the frames. Additionally, an unsupervised learning approach, independent of clinical labels, was used to investigate the feature space of the microcirculation image validation dataset and showed clustering
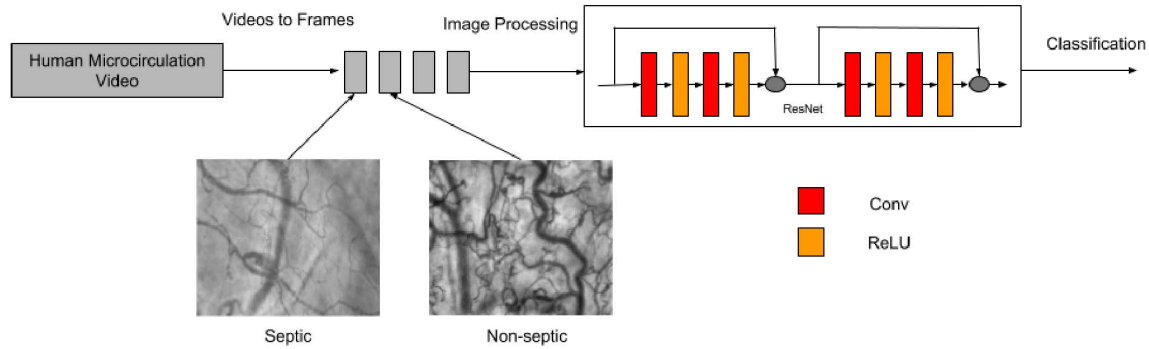
---

†Corresponding author: pratiks@media.mit.edu

Fig. 1. Frames were sampled from microcirculation videos, resized to 224 x 224 x 3 pixels and fed into CNN to distinguish images from non-septic and septic patients. A truncated version of ResNet18 was used as architecture.

of images from non-septic and septic patients. To our knowledge, this is the first study that successfully classifies human microcirculation image data using a deep neural network. We reason the salient feature space used by our trained classifier may have diagnostic applications.

II. RELATED WORK

In previous studies, microcirculation dysfunction was experimentally induced in animal models with drugs and other agents followed by SDF imaging [12]. SDF imaging is a non-invasive imaging modality and has been used to track changes in the microcirculation on mucosal surfaces. SDF uses green polarized light with wavelength of 550 nm which is absorbed by hemoglobin and makes red blood cells visible [22]. The distinguishing parameters and measurements associated to microcirculation include MFI (Microcirculation Flow Index), PVD (Perfused Vessel Density), TVD (Total Vessel Density) and PPV (Portion of Perfused Vessels) [3]. These measurements require identification of vessels in microcirculation video frames. Several software systems have been developed to analyze the microcirculation images and videos for these measurements, but often fail to provide desired results [18]. Marking of hand engineered features on each frame of the video is required at some stage of analyses and is time-consuming and not accurate [17].

Computer vision techniques have previously been used to automate microcirculation parameters and associated measurements to detect dysfunction. Liu et. al report a framework to estimate the flow rate statistics from microcirculation SDF imaging after the video is stabilized, enhanced and micro-vessels are extracted from each frame [9]. Similarly, Bezemer et. al report an automated assessment of the TVD measurement using contrast score thresholding to validate the vessel detection on microcirculation videos collected from patients [11]. Demir et. al automate the estimation of functional capillary density (FCD) by preprocessing multi-thresholding segmentation on four blocks of five consecutive frames; followed by morphological operations on the mean image of each block and finally region growth models to extract the vessels to calculate the FCD [10]. Graph-based approaches have also been leveraged to extract features from flow and capillary structure to assess microcirculation dysfunction caused due to sepsis in mice [19]. McIlroy et. al reported use of machine learning to classify microcirculation videos from animals into discrete intensities of drug induced microcirculation inflammation. They report no significant differences in prediction accuracy when comparing the performance of training on videos vs. frames [12].

III. METHODS

A. Data Acquisition and Processing

Deidentfied microcirculation videos were acquired from BIDMC (Beth Israel Deaconess Medical Centre, Boston, MA) collected using protocol 2008P-000089. De-identified videos were transferred to Massachusetts Institute of Technology and analyzed according to Committee on the Use of Humans as Experimental Subjects protocol 1705964591. Videos were captured using a CytoCam Incidence Dark Field imaging camera system (Braedius Medical B.V., The Netherlands). Videos were acquired from consenting patients in the Emergency Department of Intensive care unit (ICU) who were diagnosed with septic shock (low blood pressure from sepsis). Non-septic consenting patients were also imaged as controls. The videos were acquired by placing the imaging device under the tongue of the patients and were recorded at two different intervals for septic patients i.e. at zero hours and 12 hours after study enrollment. Control videos were taken only at zero hours. The diagnosis of patients as non-septic or septic following clinical evaluation was considered as the ground truth for training and validation of machine learning classifiers but not for unsupervised learning using autoencoders.

Raw videos were reviewed for quality control and videos frames that did not meet expected standards of illumination, focus, duration, content, stability and pressure were discarded [18]. In total 133 videos acquired from 53 patients were selected for machine learning. Videos were stabilized and cropped to remove unnecessary image background. Frames

608

Label	Number of training frames	Number of validation frames
Septic	8191	2410
Non-septic	14026	2613

were extracted from the videos and split into 80% for training and 20% for validation (Table 1). All the frames from a single patient were used either for training or the validation dataset to prevent over fitting on patient-specific signatures.

B. Classifier

ResNet18 was modified by removing layers, resulting in a 10-layer architecture used in this study, while retaining the ResNet structure [15]. Due to variability in size of videos frames were cropped and were resized to 224 x 224 x 3 pixels to create a uniform dataset. A batch size of four was used during training and Adam optimizer was used as an optimizer [20]. Cross Entropy was used as the loss metric. The training process is shown in Fig. 1. NVIDIA GeForce GTX Titan X GPU was utilized to speed up the training time and the code was written using PyTorch library.

C. Visualization of CNN codes

t-SNE was used to visualize the convolutional feature vectors for the validation dataset in two dimension [23]. After the architecture was trained layers were frozen and the frames in the validation set were fed to the architecture. The 128-dimensional output of the last convolutional layer (convolutional feature vector) was used for t-SNE embedding and visualization (Fig. 2).
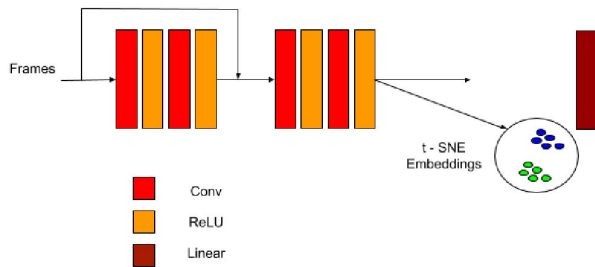

Fig. 2. The last convolutional layer used for visualization of t-SNE embedding.

D. Autoencoders

For investigating the feature space of the extracted frames, an unsupervised convolutional autoencoder was trained with reconstruction loss (Fig. 3). The autoencoder consisted of convolutional layers in the encoder followed by ReLU and BatchNorm [24]. The encoder and decoder comprised of three convolutional layers each. No skip connections were used. The autoencoder was trained using Adam optimizer and mean squared error as the loss metric. The training and test dataset remained unchanged but clinical labels were not used. A few selected images from the validation dataset were randomly sampled after every epoch to monitor the similarity between input and output frames. Once autoencoder was trained, the feature vector (1152 dimensional bottleneck layer) was extracted and k-means algorithm was used for clustering and visualization [21]. For designing bottleneck layer the number of neurons were kept as few as possible in order to avoid identity mapping and still achieve better reconstruction loss.
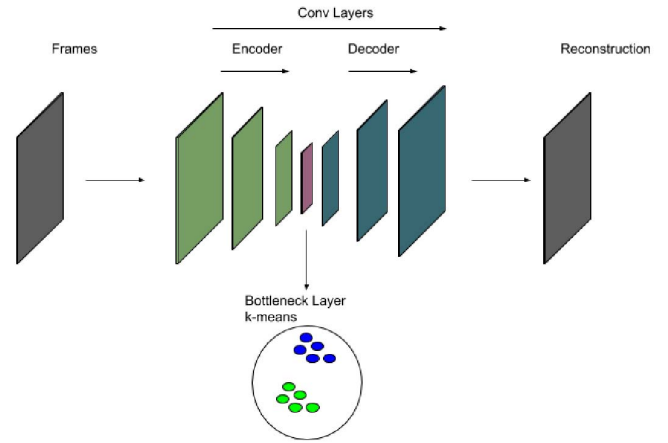

Fig. 3. Feature vector from the trained autoencoder used for k-means clustering and visualization.

IV. RESULTS

We report a machine learning classifier that successfully distinguishes between non-septic patients and septic frames extracted from SDF microcirculation videos of septic and non-septic patients. The classifier achieves an accuracy of 89.45% on a validation dataset comprising of 5,023 frames. The area under the receiver operating characteristic curve (AUROC) was 0.92 (Fig. 4). Supplementary table S1 shows mapping of correctly classified frames to individual patients in the validation dataset. Although the total number of septic and non-septic patients were not equal, the total number of frames were balanced as shown in Table I.
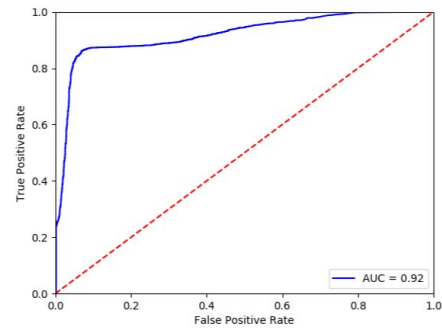

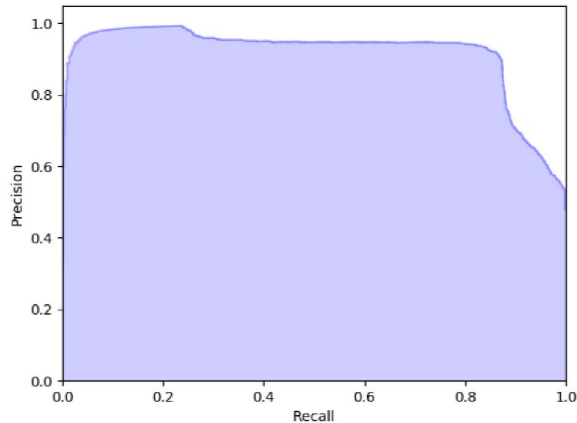Fig. 4. Receiver operating characteristic curve.

Fig. 5. Precision-recall curve for the validation data.

Fig. 6. Visualization of t-SNE embedding using the last convolutional layer of the trained CNN model.

The precision and recall were 0.92 and 0.84 respectively (Fig. 5). Figure 6 shows the t-SNE embedding of the convolutional features on the validation set, with visually separable clusters. Thus, showing that CNN does indeed separate images from septic and non-septic patients into distinct clusters for classification. The results of autoencoder were measured based on reconstruction loss in terms of mean squared error, with a value of 0.005362; thus, showing that the sampled decoded images visually matched the input images. The accuracy on validation set using k-means clustering was 73.32%.

V. DISCUSSION

The AUROC of the CNN classifier is 0.92 (Fig. 4), which indicates that the accuracy of the classifier is above chance. This also shows that there are distinct learnable representations, independent of temporal information, which can differentiate between frames from non-septic and septic patients. The classifier achieves a precision value of 0.92, indicating that 92% of its predictions were accurately classified. The classifier was able to recall frames from septic and non-septic patients correctly 84% of the time. Our patient-level data in supplementary Table S1 shows that seven patients in the validation dataset had more 90% of their images classified accurately, with 100% accuracy for four individuals. Images from two patients were classified with low accuracies. Variability in image capture and background have been reported as potential confounding factors precluding clinical evaluation. Inherent heterogeneity in patient anatomical features could also play a role in reducing classification accuracies. Images from a single patient were used either for training or in the validation dataset as one potential approach to prevent over fitting on patient-specific signatures.

Stuart McIlroy et. al report an accuracy of 83% when detecting inflammation in microcirculation videos of animals; and 80% accuracy when classifying the degree of inflammation [12]. They train a CNN model on microcirculation videos of animals with drug induced inflammation. We use microcirculation videos from human subjects where the inflammation was induced as the body's natural response to infection. We
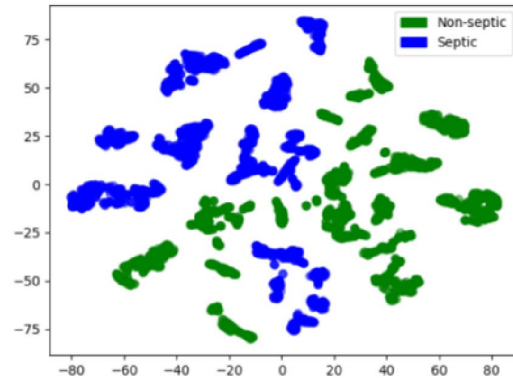
report an accuracy of 89.45% when detecting presence of sepsis in the images and also show t-SNE embedding of the features learned by the convolutional layers. We further extend our work by showing that an unsupervised autoencoder can distinguish between images from septic and non-septic patients without using clinical labels. Thus, our work is significantly different compared to this study and novel.

Previous studies use temporal information to detect blood flow in the capillaries [9] [10] [11] [12]. We on the other hand use static frames extracted from the microcirculation videos for classification. The first step for the previously described studies in the field required detection of vessels in the videos, followed by calculation of measurements such as MFI, PVD, TVD etc, which were used to evaluate dysfunction in microcirculation. Hand engineered features, image segmentation and computer vision algorithms have also been used to extract vessels in the videos [3] [10] [11] [17]. While we, use a CNN to automatically learn features in the raw frames to classify septic vs. non-septic patients. We hypothesize that our neural network may be learning features, that are independent of temporal information, such as variation in numbers, length and density of vessels between images from septic and non-septic patients. Previous work by Stuart McIlroy et. al has also alluded about the dispensbility of temporal learning for successfully classification in animals. The overlap, if any, between diagnostic features used by human experts and machine learning algorithms is an active area of investigation in our research group.

Visualization and examination of the convolutional features learned by the CNN has been shown (Fig. 6). The cluster separation indicate that the classifier had learned to effectively distinguish between images of septic and non-septic patients. The slight differences in the clustering could be attributed to variability in illumination, stability, contrast between images. The features learned by bottleneck layer of the unsupervised autoencoder were extracted using reconstruction loss, and thereby are not influenced by the sepsis labels provided by clinicians. The accuracy of 73.32% using the k-means

algorithm on these features show that the images indeed had general features that can be leveraged or refined to distinguish between septic and non-septic frames.

VI. Conclusion

We report the first CNN classifier that successfully classifies human SDF microcirculation images with high accuracy. Representation and visualization of the learned embeddings of the trained classifier supports the conclusion that it may use unique features to distinguish between immages from septic or non-septic patients. Additionally, we use an unsupervised autoencoder to show that indeed there may be differential features in human microcirculation. We propose that the salient feature space used by our trained classifier has diagnostic applications for evaluating microcirculation dysfunction in humans.

References

[1] Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801–810. doi:10.1001/jama.2016.0287

[2] Ince, C. (2005). The microcirculation is the motor of sepsis. Critical Care, 9(Suppl 4), S13–S19. http://doi.org/10.1186/cc3753

[3] Charlton, M., Sims, M., Coats, T., & Thompson, J. P. (2017). The microcirculation and its measurement in sepsis. Journal of the Intensive Care Society, 18(3), 221–227. http://doi.org/10.1177/1751143716678638

[4] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge MA: MIT Press, 2016.

[5] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., ... & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis, 36, 61-78.

[6] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115.

[7] Shen, W., Zhou, M., Yang, F., Yang, C., & Tian, J. (2015, June). Multi-scale convolutional neural networks for lung nodule classification. In International Conference on Information Processing in Medical Imaging (pp. 588-599). Springer, Cham.

[8] Rana, A., Yauney, G., Wong, L. C., Gupta, O., Muftu, A., & Shah, P. (2017, November). Automated segmentation of gingival diseases from oral images. In Healthcare Innovations and Point of Care Technologies (HI-POCT), 2017 IEEE (pp. 144-147). IEEE.

[9] Liu, C., Gomez, H., Narasimhan, S.G., Dubrawski, A., Pinsky, M.R., & Zuckerbraun, B.S. (2015). Real-time visual analysis of microvascular blood flow for critical care. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2217-2225.

[10] Demir, S. U., Hakimzadeh, R., Hargraves, R. H., Ward, K. R., Myer, E. V., & Najarian, K. (2012). An automated method for analysis of microcirculation videos for accurate assessment of tissue perfusion. BMC Medical Imaging, 12, 37. http://doi.org/10.1186/1471-2342-12-37

[11] Bezemer, R., Dobbe, J. G., Bartels, S. A., Christiaan Boerma, E., Elbers, P. W. G., Heger, M., & Ince, C. (2011). Rapid automatic assessment of microvascular density in sidestream dark field images. Medical & Biological Engineering & Computing, 49(11), 1269–1278. http://doi.org/10.1007/s11517-011-0824-1

[12] McIlroy, S., Kubo, Y., Trappenberg, T., Toguri, J., & Lehmann, C. (2017, May). In vivo classification of inflammation in blood vessels with convolutional neural networks. In Neural Networks (IJCNN), 2017 International Joint Conference on (pp. 3022-3027). IEEE.

[13] Trzeciak, S, McCoy, JV, Phillip Dellinger, R. Early increases in microcirculatory perfusion during protocol-directed resuscitation are associated with reduced multi-organ failure at 24 h in patients with sepsis. Intensive Care Med 2008; 34: 2210–2217.

[14] Sakr, Y, Dubois, MJ, De Backer, D. Persistent microcirculatory alterations are associated with organ failure and death in patients with septic shock. Crit Care Med 2004; 32: 1825–1831.

[15] He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. CoRR, abs/1512.03385.

[16] De Backer D, Creteur J, Preiser JC, Dubois MJ, Vincent JL. Microvascular blood flow is altered in patients with sepsis. Am J Respir Crit Care Med. 2002;166:98–104.

[17] De Backer, D., Hollenberg, S., Boerma, C., Goedhart, P., Büchele, G., Ospina-Tascon, G., ... Ince, C. (2007). How to evaluate the microcirculation: report of a round table conference. Critical Care, 11(5), R101. http://doi.org/10.1186/cc6118

[18] Massey, M. J., & Shapiro, N. I. (2016). A guide to human in vivo microcirculatory flow image analysis. Critical Care, 20, 35. http://doi.org/10.1186/s13054-016-1213-9

[19] Zoghbi, J. M., De La Cruz, L. T., Galarreta, M. A. G. V., Jackowski, M. P., Vieira, J. C. F., Liberatore, A. M. A., & HJ, K. I. (2014, August). Graph based characterization of microcirculation in sepsis using sidestream dark field imaging. In 2014 27th Sibgrapi Conference On Graphics, Patterns and Images (sibgrapi) (pp. 312-318). Ieee.

[20] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[21] David Arthur and Sergei Vassilvitskii. 2007. k-means++: the advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '07). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1027-1035.

[22] Goedhart, P. T., Khalilzada, M., Bezemer, R., Merza, J., & Ince, C. (2007). Sidestream Dark Field (SDF) imaging: a novel stroboscopic LED ring-based imaging modality for clinical assessment of the microcirculation. Optics express, 15(23), 15101-15114.

[23] Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(Nov), 2579-2605.

[24] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167.

Supplementary Material

TABLE S1
Patient Level accuracy distribution

Patient	Total frames	Correct pred.	Incorrect pred.	Percent accuracy	Clinical label
P1	377	377	0	100 %	Septic
P2	961	957	4	99.58 %	Septic
P3	1072	714	358	66.60 %	Septic
P4	958	954	4	99.58 %	Non-septic
P5	135	135	0	100 %	Non-septic
P6	519	463	56	89.21 %	Non-septic
P7	194	86	108	44.32 %	Non-septic
P8	300	300	0	100 %	Non-septic
P9	507	507	0	100 %	Non-septic