# Deep Convolutional Neural Network for Classification and Segmentation of Wireless Capsule Endoscopy Images

Sreevidya C[1], Vysak Valsan[2]

*[1]PG Scholar, [2]Asst. Profeesor*

*Dept of Electronics and Communication Engineering, Kerala Technological University*

***Abstract—The objective of this paper is to develop convolutional neural network (CNN)-based deep learning to identify bleeding and non-bleeding CE images, where a Res-Net is used to train a CNN that carries out the identification. Moreover, bleeding zones in a bleeding image are also identified using deep learning-based semantic segmentation.***

***Keywords— Capsule endoscopy Convolutional neural network Deep learning classification Segmentation***

## I. INTRODUCTION

Bleeding is a very common symptom of many GI tract diseases such as vascular lesions, small bowel tumors, coeliac disease, and Crohn's disease. It captures images while moving along the GI tract to detect abnormalities and bleeding in colon, esophagus, small intestinal and stomach in a non-invasive way in which flexible endoscope (e.g., colonoscopy and gastroscopy) will not be able to access.

### *A. Wireless capsule endoscopy*

The patient swallows a capsule like a pill comprises of then the capsule travels through the GI track and captures the images by the camera and sent out wirelessly to a special recorder attached to the patients waist. WCE consists of light source, lens, camera, radio transmitter and batteries This process will continue depending on the battery's life. Finally, all recorded images will be uploaded to computer work station for physicians' analysis. Capsule travels about eight hours in the GI track during which it captures 50,000 images.

### *B. Challenging issues on working with WCE image analysis*

- First problem identified are due to the long period of time spent for the inspection of huge number of images produced by WCE device.
- Second problem identified is that different physicians would provide different diagnosis or interpretation of the same image
- Third problem faced when dealing with WCE images is that these images are rather dark and vague as such physicians would find difficulties in analysing and giving diagnosis
- WCE has some drawbacks in terms of its low transmission power and bandwidth constraints, which leads to poor clarity of the images

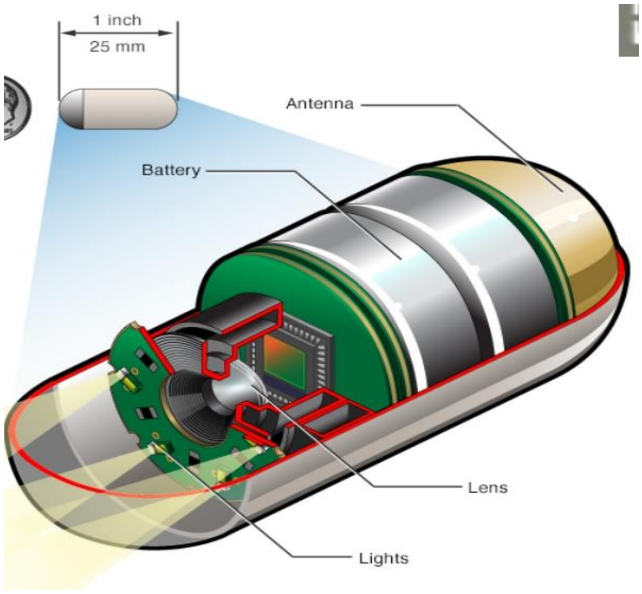


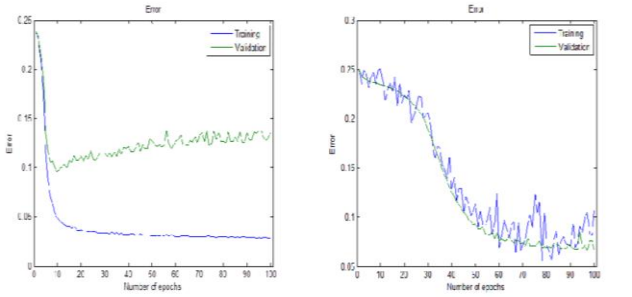

Fig. 1 Wireless Capsule Endoscopy

## II. SYSTEM MODEL

### *A. Enchancement of WCE images*

In classifying the bleeding area of WCE images, we start the experiments on raw images to train the neural network using DCNN. WCE images are divided into three parts: training, testing and validation set. Here, pre-processing technique is not applied on input of WCE images on training data of the neural network. Validation data set is used to know how well the pre trained model has been trained. The validation and training errors are monitored to avoid over-fitting occurring in the training network using early stopping as its regularization methods. Both these errors should decrease when number of epoch increases. However, these experiments of training neural network without any pre-processing technique shows that training error decreases with validation error increases after

some point even though decrease initially. This problem is due to the difference in the colour of bleeding and normal areas on both training images and unknown WCE images are very prominent. The colour of bleeding of training images is totally different from the unknown WCE images.




(a) Early stopping before colour normalization applied. (b) Early stopping after colour normalization applied.

Fig. 2 Early stopping during training neural network

the colour of bleeding of training images is totally different from the testing WCE images.The objective of the learning algorithm for neural network is to minimize the training and validation errors of each iteration process. After each iteration process, both the training and validation errors were evaluated. When minimizing the training and validation errors using preprocessing the performance of neural network using DCNN is maximized and improve accuracy in classification. Color normalisation reduces the variation in color in WCE image and it increases the performance of Neural network

### *B. Convolutional neural network ResNet*

CNN is composed of an input layer, output layer, and many hidden convolutional layers as intermediate layers. These layers perform operations that alter the data with the intent of learning features specific to the data.

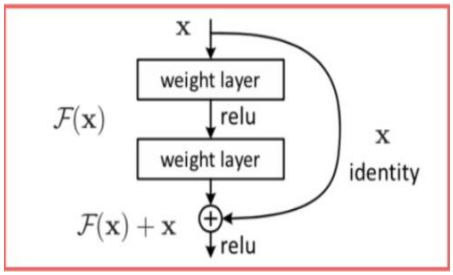



Fig3 Building block of Res-Net

The network depth is defined as the largest number of sequential convolutional or fully connected layers on a from the input layer to the output layer. In total, ResNet-50 has 177 layers. Res-Net consists of: convolution, activation function such as Rectified Linear Unit (ReLU), batch normalization layer and fully connected layer. CNN has varied filters which is used for pattern recognition ranging from simple to complex for automatic image feature extraction. Images features may be colors, texture, boundaries or shapes which are detected by using various filters. The convolutional layer extracts the feature maps out of input images which is also the main layer in the CNN model. ResNet-50 is a convolutional neural network that is 50 layers deep. The pretrained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories.

This uses the rectified linear unit (ReLU) as the activation function with the convolutional layer because natural images are non-linear so ReLU can be helpful to increase the non-linearity in the input image. ReLu when used for a non-linear function. It provides faster convergence than a tanh function. Residual connections that bypass the convolutional units of the main branch. The outputs of the residual connections and convolutional units are added element-wise. When the size of the activations changes, the residual connections must also contain 1-by-1 convolutional layers. Residual connections enable the parameter gradients to flow more easily from the output layer to the earlier layers of the network, which makes it possible to train deeper networks. Stochastic gradient descent with momentum optimizer is used with a reasonable minimum batch size for each training epoch. Moreover, to reduce further over-fitting, we use data augmentation. For training, horizontal reflections and image translations in both horizontal and vertical directions are considered.

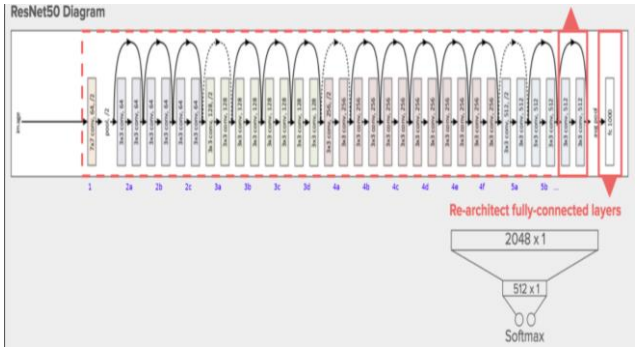



Fig. 4 Res-Net architecture

### *C. SegNet Architecture*

Semantic image segmentation is the task of classifying each pixel in an image from a predefined set of classes. Since bleeding areas have arbitrary shape and size, recognizing a image in pixel level, which means the trained network can assign each pixel in the image to an object class. In here classifying the image in to two classes: bleeding and non-bleeding. Fully connected layers will be discarded to keep the high-resolution feature maps at the deepest encoder output. Finally, fully connected layer can be replaced by the convolution layer. The decoder output will be fed into the classifier to classify each pixel separately. Segmentation contains encoder and decoder that enables segmentation

maps from different size. Each such layer is followed by batch normalisation and Relu. The encoder generates the transferred pool indices then the input is up-sampled by the decoder by recalling the corresponding max pooling indices in the up sample to produce a sparse feature map. Five max-pooling and five up-sampling layers are used. The trainable filter bank then helps to perform the convolution to density the feature map. Finally, a softmax classifier is fed by the decoder output feature maps for pixel-wise classification. For the encoder network, max-pooling and sub-sampling are used to achieve translation in-variance for the tiny spatial shifts of the input image and then produce a large spatial window for all pixels.

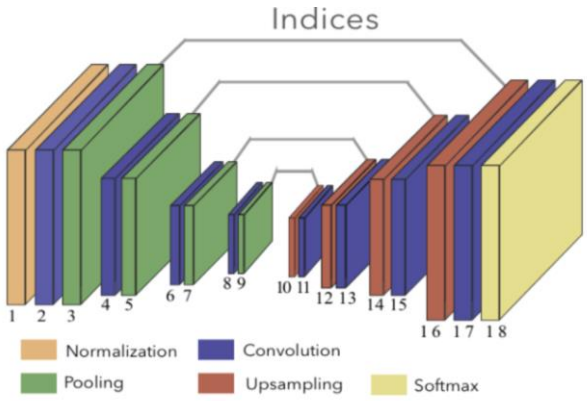



Fig. 5 Seg-Net architecture

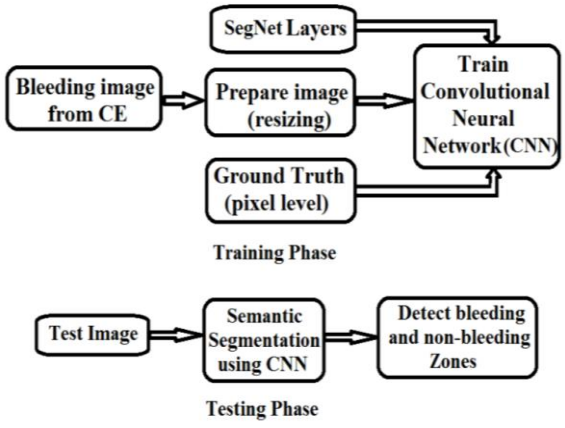



Fig. 6 Block diagram of segmentation

These processes will increase the lossy image representation at the boundaries, which is unacceptable for segmentation. Thus, the boundary information should be fully stored before the sub-sampling. However, the encoder feature maps cannot be fully stored in practice because of memory limits. Thus, for each encoder feature map, only the max-pooling indices, which include the maximum features in each pooling window, are stored. The layers which down sample the input are the part of the encoder and layers which up sample are the part of decoder. The feature maps are up sampled by using the max pooling indices then convolved with a trainable filter.

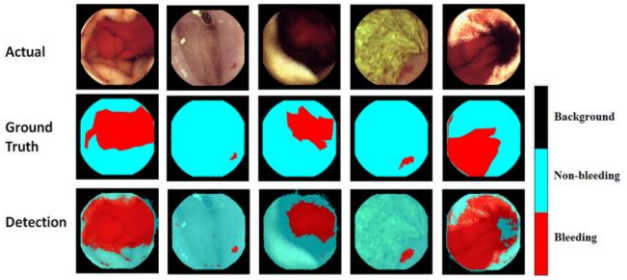



Fig. 7 Bleeding segmentation output

## III. RELATED WORKS

A hybrid CNN with Extreme Learning Machine (ELM) was proposed by [11]. The CNN was constructed as a data-driven feature extractor and cascaded ELM acts as strong classifier. In [12], the authors fine-tuned the layers in CNN on the ImageNet database to diagnose the celiac disease. Binary classification was performed using softmax classifier and Support Vector Machines (SVM). DCNN was used to classify the digestive organs using WCE images[13]. [16] proposed an automatic bleeding detection strategy based on a DCNN to learn high-level features where rectified linear units (ReLUs)were used as the activation function. They used a softmax function to minimize the cross-entropy loss for prediction. However, the performance in detecting abnormalities of the other classes was poorer than the state-of-the-art technique. Then, the authors improved the complexity of the DCNN [14] by combining hand crafted (HC) features and DCNN with fewer training samples. [15] focused on small-size imbalanced endoscopy images for bleeding detection thus CNN could learn with very limited training data. Data augmentation and image resampling were employed to increase the size of training database.[17] proposed two combined CNNs to avoid the edge feature caching and speed up the hook worm classification in WCE images. [5] investigated the colour spaces (RGB, HSB, and YUV) that has the ability to disclose lesion structure while geometry, colour and texture were combined in their analysis. The features were extracted from inside the masked region or also known as Region of Interest (ROI) where the abnormality category then classified to be either ulcer or bleeding by using Support Vector Machines(SVM) and Vector Supported Convex Hull Method (VSCH). [18] proposed colour histogram based on index image to extract the colour texture of bleeding. SVM was used to detect bleeding and normal regions from WCE videos. However, this method has weakness since it is relying on intensity range of MSB (Most Significant Bit) and LSB (Least Significant Bit) in RGB colour spaces. [19] proposed to extract bleeding area using R to G pixel intensity ratio and different statistical parameters. K-nearest neighbour (KNN) was used to classify between bleeding or normal region. A was utilized to characterize the feature vector from region of interest (ROI). Then, classification based on SVM and

K-nearest neighbour (KNN) was employed to detect bleeding. The authors also investigated on which colour spaces (RGB, HSV, YCbCr and LAB) are the most appropriate in describing bleeding characteristic. However, their proposed method focused on easy WCE images, in which bleeding colour is relatively uniform in training and test images. super-pixel segmentation was proposed by [10] by grouping the pixel to reduce computational complexity. Although these three algorithms are relatively insensitive to colour variation, the computational costs are rather high, which makes it unsuitable for processing a huge number of WCE images. [9] used K-means clustering to make the most of the colour information of the bleeding. Then, colour histogram was utilized to characterize the feature vector from region of interest (ROI). Then, classification based on SVM and K-nearest neighbour (KNN) was employed to detect bleeding. The authors also investigated on which colour spaces (RGB, HSV, YCbCr and LAB) are the most appropriate in describing bleeding characteristic. However, their proposed method focused on easy WCE images, in which bleeding colour is relatively uniform in training and test images. In order to detect bleeding frames and zones, a probabilistic neural network-based method is investigated in [5], demonstrating a satisfactory level of sensitivity, which is a measure of the correctness of bleeding frame detection. Statistical features and a region growing method is proposed in [6], where a user needs to select the initial seed of the bleeding zone. Since the user selects a starting point of a region growing method, rather than an automatic annotation, it is a semi-automatic system. A word-based histogram method is presented in [7], where K-means clustering is used to fnd the word dictionary that will be used to detect bleeding frames and a two-stage salient map is proposed to detect bleeding regions. This method used a large number of features (80) and the reported sensitivity is 92%, which is not considered satisfactory. The study in [8] introduces super-pixel-based bleeding segmentation, which is computationally complex. A cluster-based statistical feature is proposed in [9], which used unsupervised K-means clustering to detect bleeding zones. The intensity variation profile among normalized RGB colour planes is introduced in [10].

## IV. PROPOSED METHODOLOGY

There are a variety to deep learning models available for CNN architecture. The architecture of each CNN in the proposed system is ResNet architecture. The optimizing algorithm used is s Stochastic Gradient Descent with Momentum (SGDM). Segmentation results of bleeding detection are shown in Fig. 7, where five sample images are presented. Among the five bleeding images, active bleeding is shown in column 1, 2 and 3 and inactive bleeding is shown in column 2 and 4. In the case of active bleeding, the proposed method detects the bleeding regions with some false positive while on the other case of inactive bleeding, bleeding regions are detected more precisely. It is to be noted that the objective of this work is to assist physician in the reviewing process. The detection bleeding regions reduce the effort of the physician to find out the area and they can concentrate only on those regions. Therefore, we find that bleeding region segmentation can be achieved with SegNet network. This network can definitely help the physician to diagnose diseases. So in this proposed system, Res Net architecture is used for to classify the WCE images in to bleeding and non-bleeding region without prior knowledge.

There are four possible outcomes when classifying bleeding and non-bleeding images:

1. True positive (TB): A bleeding image is correctly detected as bleeding,

2. True negative (TNB): A non-bleeding image is correctly detected as non-bleeding,

3. False negative (FNB): A bleeding image is wrongly detected as non-bleeding image, and

4. False positive (FB): A non-bleeding image is wrongly detected as bleeding.

Semantic segmentation metrics aggregated over the data set, specified by this parameter

Global Accuracy — Ratio of correctly classified pixels to total pixels, regardless of class.

Mean Accuracy — Ratio of correctly classified pixels in each class to total pixels, averaged over all classes. The value is equal to the mean of class metrics accuracy.

Mean IoU — Average intersection over union (IoU) of all classes. The value is equal to the mean of class metrics IoU

Weighted IoU — Average IoU of all classes, weighted by the number of pixels in the class.

IoU is the most widely used metric, also known as the Jaccard similarity coefficient. The proportion of correctly classified pixels to the total number of ground truth and predicted pixels in that class is measured by IoU. Since bleeding and non-bleeding zones are disproportionally sized classes, in order to decrease the impact of errors in the smaller class, weighted IoU is introduced. This is the average IoU of each class, weighted by the number of pixels in that class. The percentage of correctly recognized pixels for each class is defined by accuracy. Mean accuracy is the average accuracy of all classes. Moreover, global accuracy indicates the proportion of correctly classified pixels, despite of class, relative to the total number of pixels.

## V. RESULT AND SIMULATION

During training, all images are resized to 224 × 224× 3. Randomly selected 80% images are used to train the

ResNet and the remaining 20% images are selected for testing. From this experiment we are concluding the Res-Net has high classification accuracy and best efficiency.


Fig. 8 Training Accuracy during training phase

The loss function vs. training iteration number is shown in Fig. 8, showing that the loss value is considerably low after eight iterations. By using just 20 images, accuracy is very high. The WCE images are classified in to bleeding and non-bleeding images. The training of the training phase is done with an accuracy of 83.3% with 20 images. So that by increasing the iteration number, training accuracy can be increased further.

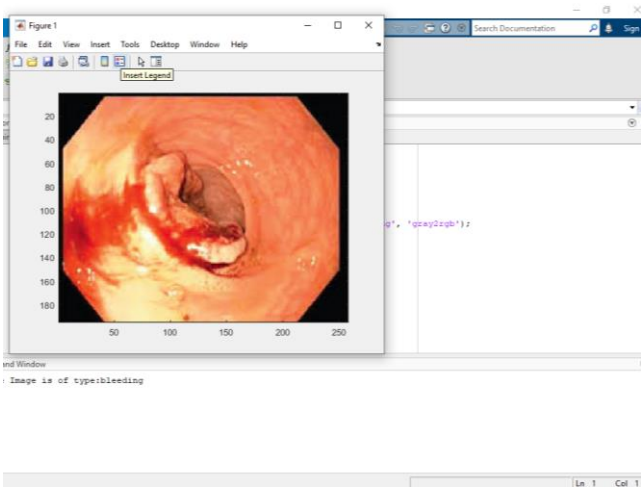



Fig. 9 Single image testing for classification

## VI. CONCLUSION

Deep learning is used to classify the bleeding and non-bleeding images. The exact region of bleeding is identified using semantic segmentation.

## REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] I.N. Figueiredo, S. Kumar, C. Leal, P.N. Figueiredo, Computerassisted bleeding detection in wireless capsule endoscopy images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 1(4), 198 (2013)

[3] M. Liedlgruber and A. Uhl, “Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review,” IEEE Reviews in Biomedical Engineering, Vol. 4, 2011, pp. 73-88.

[4] R. Shahril, S. Baharun, A. K. M. M. Islam, and S. Komaki, “Anisotropic contrast diffusion enhancement using variance for wireless capsule endoscopy images,” in Proceedings of International Conference on Informatics, Electronics Vision, 2014, pp. 1-6.

[5] T. Ghosh, S.A. Fattah, S. Bashar, C. Shahnaz, K. Wahid, W.P. Zhu, M.O. Ahmad, An automatic bleeding detection technique in wireless capsule endoscopy from region of interest. in 2015 IEEE International Conference on Digital Signal Processing (DSP), pp. 1293–1297 (2015)

[6] S. Suman, F.A.B. Hussin, N. Walter, A.S. Malik, S.H. Ho, K.L. Goh, Detection and classifcation of bleeding using statistical color features for wireless capsule endoscopy images. In: International Conference on Signal and Information Processing (IConSIP) , pp. 1–5 (2016)

[7] A. Novozámskỳ, J. Flusser, I. Tachecí, L. Sulík, J. Bureš, O. Krejcar, Automatic blood detection in capsule endoscopy video. Journal of Biomedical Optics 21(12), 126007 (2016)

[8] T. Ghosh, S. Fattah, C. Shahnaz, A. Kundu, M. Rizve, Block based histogram feature extraction method for bleeding detection in wireless capsule endoscopy. in 2015 IEEE Region 10 Conference (TENCON) (2015), pp. 1–4

[9] E. Hu, H. Sakanashi, H. Nosato, E. Takahashi, Y. Suzuki, K. Takeuchi, H. Aoki, M. Murakawa, Bleeding and tumor detection for capsule endoscopy images using improved geometric feature. Journal of Medical and Biological Engineering 36(3), 344 (2016)

[10] T. Ghosh, S.A. Fattah, K.A. Wahid, Automatic computer aided bleeding detection scheme for wireless capsule endoscopy (wce) video based on higher and lower order statistical features in a composite color. Journal of Medical and Biological Engineering 38(3), 482 (2018

[11] J. Yu, J. Chen, Z. Q. Xiang, and Y. Zou, “A hybrid convolutional neural networks with extreme learning machine for wce image classification,” in Proceedings of IEEE International Conference on Robotics and Biomimetics, 2015, pp. 1822-1827.

[12] G.Wimmer, A. Vcsei, and A. Uhl, “Cnn transfer learning for the automated diagnosis of celiac disease,” in Proceedings of the 6th International Conference on Image Processing Theory, Tools and Applications, 2016, pp. 1-6.

[13] Y. Zou, L. Li, Y. Wang, J. Yu, Y. Li, and W. J. Deng, “Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network,” in Proceedings of IEEE International Conference on Digital Signal Processing, 2015, pp. 1274-1278.

[14] X. Jia and M. Q. H. Meng, “Gastrointestinal bleeding detection in wireless capsule endoscopy images using

handcrafted and cnn features," in Proceedings of the 39th Annual International Conference of IEEE Engineering in Medicine and Biology Society, 2017, pp. 3154-3157

[15] X. Li, H. Zhang, X. Zhang, H. Liu, and G. Xie, "Exploring transfer learning for gastrointestinal bleeding detection on small-size imbalanced endoscopy images," in Proceedings of the 39th Annual International Conference of IEEE Engineering in Medicine and Biology Society, 2017, pp. 1994-1997.

[16] A. K. Sekuboyina, S. T. Devarakonda, and C. S. Seelamantula, "A convolutional neural network approach for abnormality detection in wireless capsule endoscopy," in Proceedings of IEEE 14th International Symposium on Biomedical Imaging, 2017, pp. 1057-1060.

[17] J. He, X. Wu, Y. Jiang, Q. Peng, and R. Jain, "Hookworm detection in wireless capsule endoscopy images with deep learning," IEEE Transactions on Image Processing, Vol. 27, 2018, pp. 2379-2392.

[18] T. Ghosh, S. A. Fattah, C. Shahnaz, and K. A. Wahid, "An automatic bleeding detection scheme in wireless capsule endoscopy based on histogram of an rgb-indexed image," in Proceedings of the 36th Annual International Conference of IEEE Engineering in Medicine and Biology Society, 2014, pp. 4683-4686.

[19] T. Ghosh, S. K. Bashar, M. S. Alam, K. Wahid, and S. A. Fattah, "A statistical feature based novel method to detect bleeding in wireless capsule endoscopy images," in Proceedings of International Conference on Informatics, Electronics Vision, 2014, pp. 1-4.