Deep Learning Algorithm for the Confirmation of Mucosal Healing in Crohn's Disease, Based on Confocal Laser Endomicroscopy Images

Anca Loredana Udriştoiu[1], Daniela Ştefănescu[2], Gabriel Gruionu[3,4], Lucian Gheorghe Gruionu[3], Andreea Valentina Iacob[1], John Gásdal Karstensen[5,6,7], Peter Vilmann[6,7], Adrian Săftoiu[2]

1) Faculty of Automation, Computers and Electronics, University of Craiova, Craiova, Romania;

2) Research Center of Gastroenterology and Hepatology, University of Medicine and Pharmacy Craiova, Romania;

3) Faculty of Mechanics, University of Craiova, Craiova, Romania; 4) Krannert Cardiovascular Institute, Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA; 5) Gastro Unit, Hospital Hvidovre, Copenhagen, Denmark; 6) Gastro Unit, Division of Endoscopy, Copenhagen University Hospital Herlev, Copenhagen, Denmark; 7) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

Address for correspondence:
Gabriel Gruionu
Faculty of Mechanics,
University of Craiova, Craiova,
Romania;
ggruionu@iu.edu

ABSTRACT

Background & Aims: Mucosal healing (MH) is associated with a stable course of Crohn's disease (CD) which can be assessed by confocal laser endomicroscopy (CLE). To minimize the operator's errors and automate assessment of CLE images, we used a deep learning (DL) model for image analysis. We hypothesized that DL combined with convolutional neural networks (CNNs) and long short-term memory (LSTM) can distinguish between normal and inflamed colonic mucosa from CLE images.

Methods: The study included 54 patients, 32 with known active CD, and 22 control patients (18 CD patients with MH and four normal mucosa patients with no history of inflammatory bowel diseases). We designed and trained a deep convolutional neural network to detect active CD using 6,205 endomicroscopy images classified as active CD inflammation (3,672 images) and control mucosal healing or no inflammation (2,533 images). CLE imaging was performed on four colorectal areas and the terminal ileum. Gold standard was represented by the histopathological evaluation. The dataset was randomly split in two distinct training and testing datasets: 80% data from each patient were used for training and the remaining 20% for testing. The training dataset consists of 2,892 images with inflammation and 2,189 control images. The testing dataset consists of 780 images with inflammation and 344 control images of the colon. We used a CNN-LSTM model with four convolution layers and one LSTM layer for automatic detection of MH and CD diagnosis from CLE images.

Results: CLE investigation reveals normal colonic mucosa with round crypts and inflamed mucosa with irregular crypts and tortuous and dilated blood vessels. Our method obtained a 95.3% test accuracy with a specificity of 92.78% and a sensitivity of 94.6%, with an area under each receiver operating characteristic curves of 0.98.

Conclusions: Using machine learning algorithms on CLE images can successfully differentiate between inflammation and normal ileocolonic mucosa and can be used as a computer aided diagnosis for CD. Future clinical studies with a larger patient spectrum will validate our results and improve the CNN-SSTM model.

Key words: confocal laser endomicroscopy – inflammatory bowel disease – Crohn's disease – convolutional neural network – deep learning.

Abbreviations: AUC: the area under the ROC curve; CD: Crohn's disease; CLE: confocal laser endomicroscopy; eCLE: endoscope-based CLE system; CNN: convolution neural network; CRC: colorectal carcinoma; DL: deep learning; IBD: inflammatory bowel disease; LSTM: long short-term memory; MH: mucosal healing; ML: machine learning; pCLE: probe-based CLE; ROC: Receiver Operating Characteristics; Se: sensitivity; Sp: specificity.

INTRODUCTION

Crohn's disease (CD) is an inflammatory bowel disease (IBD) characterized by the presence of multifactorial inflammation in the intestinal wall caused by a dysregulation of the immune system [1, 2]. Intermittent disease flares and chronic inflammation lead to irreversible mucosal damage, intestinal complications, disability and even colitis-associated neoplasia [1].

Due to the risk of relapse, a long-term therapeutic strategy is necessary to achieve a good quality of life without surgery. The newest guidelines suggest that histological inflammation is predictive of future flares, with consequent decreased rates of sustained remission, and increased need for corticosteroids and colectomy [3]. Recent evidence suggests that mucosal healing (MH) is associated with a stable course of the disease,

less hospitalization, lower surgery rates and also lower risk for colorectal carcinoma (CCR) [4]. In CD, MH can be defined as the absence of erosions and ulcerations. It is characterized by complete healing of all ulcerative and inflammatory lesions as diagnosed by endoscopic examination.

In order to assess inflammatory disease activity, MH is considered a more objective endpoint than clinical remission [5]. For MH assessment, regular endoscopic procedures, biopsy and histopathological evaluation are needed, but these procedures are associated with post biopsy bleeding, increased need for careful samples manipulation and increased time to diagnosis [6]. Novel endoscopic techniques such as confocal laser endomicroscopy (CLE) have recently been developed, providing the possibility of obtaining in vivo high-magnification images of the gut epithelium [7]. This technique allows real-time examination of the gastrointestinal mucosa at the cellular and subcellular level [8]. Because CLE is an operator-dependent technique, it requires a trained endoscopist to interpret images and give a diagnosis. Currently, the CLE diagnosis can have large variations between different experts [9]. Diagnostic errors can occur during examinations, because of subjectivity regarding interpretation, many seemingly similar features in the image, training or physical factors of the operator.

Machine learning (ML) algorithms which analyze a very large number of existing medical images in order to identify specific common features and differences between normal and pathologic samples, improve the precision and efficiency of diagnosis [10] and create actionable prediction models [11]. Recent articles reported similar results in medical applications using the deep learning (DL) technique of Convolution Neural Networks (CNNs) [12]. For example, CNNs were applied to primary breast cancer grading [13, 14] and detection [15], and also to glioma grading [16]. Other authors used the DL technique, called stacked denoising auto-encoders to perform cell detection and segmentation in histopathological images [17]. With CLE images, we have previously used DL for automated CCR diagnosis [18]. Furthermore, the combination between CNN and long short-term memory (LSTM) was used in the literature to discover the time dependencies in image sequences [19, 20].

In the present study, we used a CNN-LSTM model with four convolution layers and one LSTM layer for automatic detection of MH and CD diagnosis from CLE images. We have used our large existing CLE image database to train the CNN algorithm to discriminate between normal and inflamed CLE ileocolonic images and then confirm MH.

MATERIAL AND METHODS

Human Subjects and the CLE Procedure

We included 6,205 endomicroscopy grayscale images from 54 patients who were referred for ileocolonoscopy to the Gastro Unit of the Copenhagen University Hospital Herlev, Denmark. Fifty patients had a CD diagnosis and either active inflammation (32 patients) or mucosal healing (18 patients). Because mucosal healing represents a clinical goal for CD treatment, we considered the 18 mucosal healing patients as part of the control group. Four additional patients which were examined for adenoma surveillance but had no history of IBD were also included in the control group. Every person included in the study signed a written informed consent form. Prior to the procedure, the age, gender, current medication and surgery were recorded for each patient. None of the patients included in the study were pregnant or breast-feeding, had allergy to fluorescein or an impaired renal function or under 18 years of age. The patients' ages ranged from 19 to 77 years, with a slight predominance of male gender. The patient population was described in more detail before [21].

An endoscope-based CLE system (eCLE) was used to perform ileocolonoscopy (EC-3837CILK; Pentax, Tokyo, Japan). The images were obtained from five areas of the gastrointestinal tract: rectum, left colon, transverse colon, right colon and terminal ileum. The final diagnosis was established by histopathology. All images were classified in two categories: normal and pathologic (characterized by inflammation) (Fig. 1).

The eCLE system has the CLE technology integrated into the distal tip of the endoscope. The CLE probe contains a 488 nm-wavelength low-power blue laser to illuminate the tissue and capture a microscopic grey-scale image. To obtain confocal images, 5 ml of fluorescein sodium dye was administered intravenously before imaging. CLE generated serial images
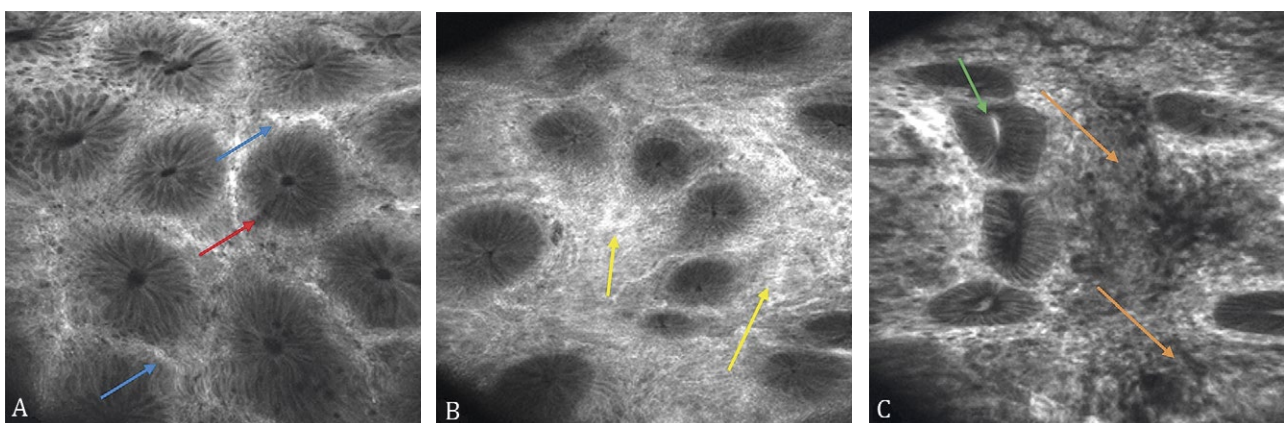

Fig. 1. Confocal laser endomicroscopy of the colon using intravenous fluorescein: A) normal colon with round shaped crypts (red arrow) and regular architecture of blood vessels (blue arrows); B), C) Crohn's disease with tortuous blood vessels (yellow arrows), distorted crypts with fluorescein leakage (green arrow) and inflammatory cells infiltration (orange arrows).

down to 250 μm into the colon mucosa at an image resolution of 1024 x 1024 pixels and lateral resolution of 0.7 μm. The acquisition frequency was 0.8 images/second and the field of view was 475 x 475 μm [22]. The dataset consists of two types of CLE images (Fig. 1): normal (2,533) and pathologic, i.e. characterized by inflammation (3,672).

The CNN-LSTM model

To classify the medical images, we used a fusion method between CNN and LSTM in order to better utilize spatial and configuration information from 2D images than a CNN model alone. A CNN is a type of artificial neural network used to recognize and classify images. It is composed of convolutional layers alternating with pooling, batch normalization and fully connected layers (Fig. 2). The convolutional layers filter the input image by detecting patterns at different positions in the image. A pooling layer follows a convolution layer to down-sample the feature map from the preceding convolution layer. To normalize the input layers, the batch normalization layers are used to adjust and scale the activations. Dropout is the method used to reduce overfitting and forces the model to learn multiple independent representations of the same data by randomly disabling neurons in the learning phase.

On the other hand, we used a LSTM network model as a type of recurrent neural network that is able to learn and remember over long sequences of input data. Our LSTM model consists of a series of 100 units. Each unit uses an input gate, a forget gate, an output gate and a memory cell. The input gate controls how much new information enters the unit and alters the state of the memory cell. The forget gate controls what to be remembered and what to be forgotten, the cell memory gate is a summation of the incoming information, and the output gate allows the state of the memory cell to have an effect on the current hidden state or other units [20]. Our LSTM model captures long-term temporal dynamics inside the feature matrix extracted by CNNs: each row of the feature matrix is a basic unit while the number of rows is the length of a sequence. So, the CLE images are converted into sequential data. Also, the long-term dependencies property of the LSTM is used to filter and fuse the rows of the feature matrix to improve the classification performance.

Model implementation and hyper parameter selection

Keras [23] with Google TensorFlow backend [24] was used to implement the CNN in this study, together with other scientific computing libraries as numpy [25] and scikit-learn [26]. In order to implement the CNN-LSTM algorithm proposed in this study, the relevant hyper-parameters under the DL framework [23, 24] are shown in Table I.

Our CNN-LSTM architecture contains 4 convolutional layers, 2 max-pooling layers, 2 batch normalization layers and 4 dropout layers (Fig. 2). The rows of the feature matrix extracted by the CNN are fed into the LSTM network for feature fusion. The LSTM layer with 100 memory blocks is followed by 2 fully connected (dense) layers. The final dense layer with a sigmoid non-linear activation function outputs the inflammation or normal diagnosis. In order to reduce the overfitting, we applied three techniques: 1) rectified linear unit (ReLU) for non-linear activation function [27], 2) dropout for randomly deactivating a fraction of the units or connections in the network on each training iteration to generalize and avoid overfitting of training data [28], and 3) data augmentation to generate more training data from existing images by augmenting the samples via a number of random transformations so that the CNN-LSTM algorithm does not analyze the same image twice. The random transformations used on images were randomly rotation of images, randomly translations, shearing transformations, and randomly zooms in images [23].

Our method was trained using the Adam optimization algorithm [29] for 200 epochs on an NVIDIA Quadro K4200. The architecture on which we run the experiments is: Intel(R) Xeon(R) CPU E5-1620 v3 @ 3.50GHz, 32 GB RAM.

Table I. Parameters of the proposed CNN-LSTM architecture

Layers	Conv1	Conv2	Pool1	Conv3	Conv4	Pool2	Fully connected1	LSTM	Fully connected2	Fully connected3
Kernel	3*3	3*3	2*2	3*3	3*3	2*2	-	-	-	-
Channel	32	32	32	64	128	128	128	100	50	1

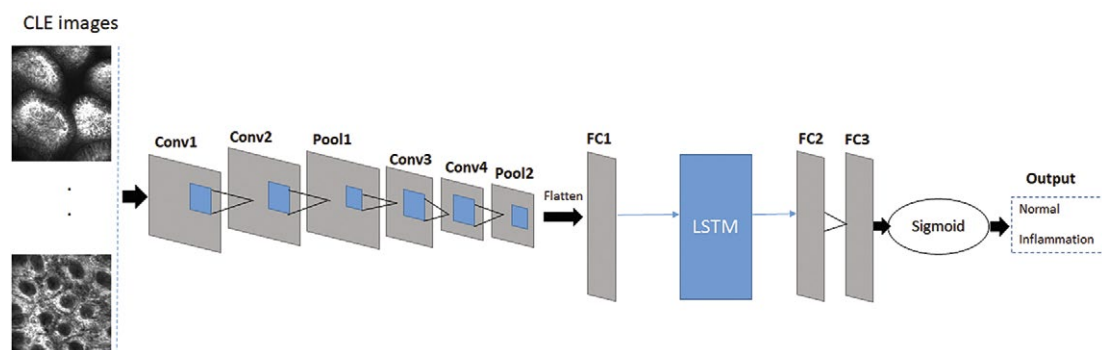CNN: convolutional neural networks; LSTM: long short-term memory

Fig. 2. The convolutional neural networks (CNN)- long short-term memory (LSTM) architecture

Image dataset

All images were classified accordingly to histopathological results. The dataset contains 6,205 grayscale CLE images as described in Table II. Of those, 5,081 images were used for training and 1,124 images for testing. For the training dataset, we used 2,892 images with inflammation and 2,189 control images. For the test dataset we used 780 images with inflammation and 344 control images. The initial resolution was 1024 x 1024 pixels and we reduced it to 300x300px by manual cropping of images, in order to delete the information of equipment, patients and borders.

Table II. The distribution of the images and patients for training and testing datasets.

Image Datasets	Inflammation	Control	Total Images/Patients
Training	2,892 /32	2,189 /22	5,081/54
Testing	780/32	344/22	1,124/54
Total	3,672/32	2,533/22	6,205/54

Visual analysis

For a better understanding of the mechanism of feature extraction and evolution of the algorithm, the results at each convolution and pool layer were visualized. Visualizing intermediate activations (outputs of intermediate layers) of our CNN-LSTM model consists of displaying the feature maps that are generated by the convolution and pooling layers in our CNN model, from input images of normal colonic mucosa and CD (Fig. 1). The feature maps show how an input image is decomposed into different filters learned by the network. Each channel encodes relatively independent features for normal and CD diagnosis (Fig. 3).

Evaluation metrics

To analyze the performance of our CNN-LSTM model, we used the following metrics that are widely adopted in the medical diagnosis field: specificity (Sp), sensitivity (Se), the test accuracy. The "true positive" are the images with inflammation diagnosis and a CNN-LSTM model prediction of inflammation. The "true negative" are images with normal diagnosis and prediction. The "false positive" are images

with normal diagnosis but predicted as inflammation. The "false negative" are the patients' images with an inflammation diagnosis that are predicted as normal.

We used two diagnostic tools for the interpretation of probabilistic prediction for binary classification: the area under the Receiver Operating Characteristics (ROC) curve (AUC) and Precision-Recall curve [30]. The ROC curves show the trade-off between the true positive rate (4) and false positive rate (5) for a predictive model using different probability thresholds. The top left corner of the ROC curve plot is the ideal point (Fig. 4A). The Precision-Recall curves show the balance between the true positive rate and the positive predictive value for a predictive model using different probability thresholds (Fig. 4B) [31].

RESULTS

Confocal laser endomicroscopy imaging reveals the appearance of normal colonic mucosa as round crypts, with dark goblet cells due to the presence of mucin and a uniform caliber of blood vessels (Fig. 1). The crypts are situated at a relative equal distance from one another. In CD patients with mucosal inflammation, the architecture of the mucosa becomes abnormal due to the presence of inflammation cellularity that produces an irregular arrangement of crypts while the blood vessels are dilated and tortuous. The crypts can be distorted with an irregular and bright crypt lumen due to fluorescein leakage (Fig. 1B and C).

On a normal ileum, intestinal villi have a regulated aspect, without any discontinuation of the epithelial layer which is uniform. Goblet cells appear dark due to the presence of mucin. Blood vessels have a uniform caliber, becoming visible after fluorescein administration. In CD, when inflammation is localized in the ileum, villi have micro-erosions which appear as a discontinuation of the epithelial layer. This layer has an irregular thickness alternating thick and narrow areas. Fluorescein leakage is present, and blood vessels have a tortuous and inequal caliber.

We used a CNN-LSTM model to detect the characteristic features in each CLE image. Our results show a final diagnosis accuracy of 95.30% and the AUC of 0.98. By comparison, when using only the CNN model the final diagnosis accuracy was 92.79% and the AUC of 0.97. The diagnosis Se - the percentage of the true positives - was 92.78%, and the specificity (Sp)
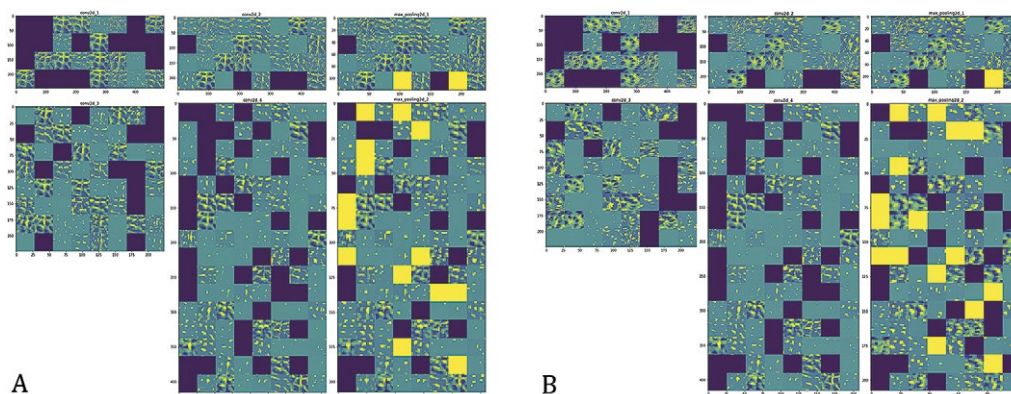

Fig. 3. The visualization of outputs of the intermediate layers of our CNN model for an image diagnosed as normal (A) and Crohn's disease (B).

Table III. Evaluation metrics for classification.

	Accuracy	ROC AUC	PR AUC	Sp	Se
CNN	92.79%	0.97	0.87	91.37%	89.46%
CNN-LSTM	95.30%	0.98	0.93	94.6%	92.78%
	Accuracy	ROC AUC	PR AUC	Sp	Se
CNN	92.79%	0.97	0.87	91.37%	89.46%
CNN-LSTM	95.30%	0.98	0.93	94.6%	92.78%

CNN: convolutional neural networks; LSTM: long short-term memory; ROC: receiver operating characteristics; AUC: the area under the ROC curve; PR: Precision-Recall; Se: sensitivity; Sp: specificity.

– the percentage of true negatives - was 94.6%. All metrics computed for CNN-LSTM vs. CNN are summarized in Table III and showed the improvements achieved by using the CNNs together to LSTM models.

The comparison of the ROC curves for the two models (CNN-LSTM vs. CNN alone) reveals slightly better results for CNN-LSTM method (Fig. 4A). The same superior performance of the CNN-LSTM is shown in the display of the precision-recall curves for the two models (Fig. 4B).
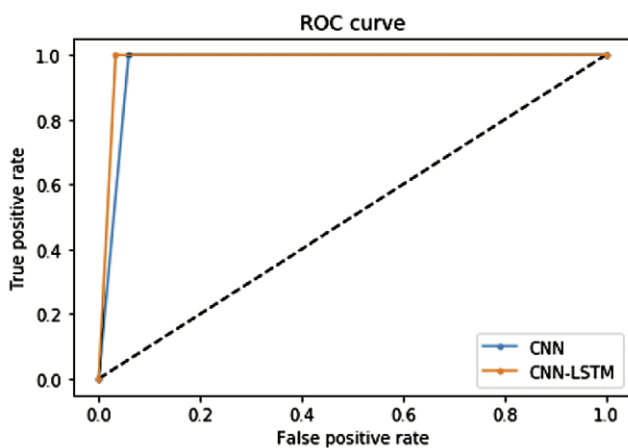

Fig. 4A. Comparison between CNN-LSTM vs. CNN: the ROC curve. For abbreviations see Table III.
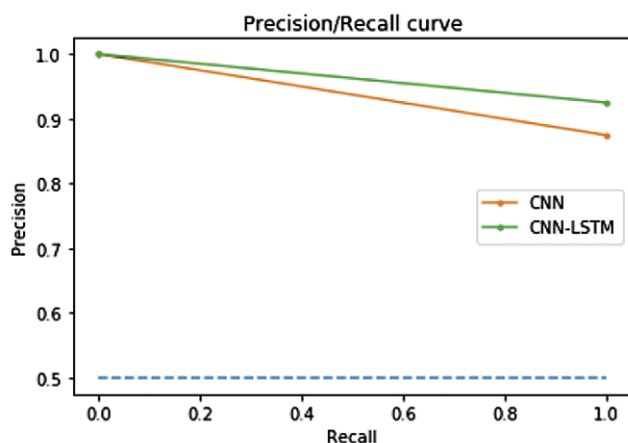

Fig. 4B. Comparison between CNN-LSTM vs. CNN: the precision/recall curve. For abbreviations see Table III.

DISCUSSION

Recently, CLE has emerged as an accurate method to identify defects of the mucosal barrier in patients with IBD. It can detect the loss of intestinal barrier function in the areas of cell shedding and identify mucosal leakage in macroscopically normal IBD [32].

Several studies concluded that CLE can reveal different types of focal lesions, similarly to histopathology [33, 34]. In addition, CLE has a distinct advantage over histopathology: several dynamic changes can be visualised in vivo, at the microscopic level (i.e. colonic crypt tortuosity, enlarged crypt lumen, hyper-vascularization, micro-erosions, mononuclear cell infiltrates and fluorescein leakage), suggesting an impaired mucosal integrity and clinical inflammation relapse [21, 22]. Despite these advantages, CLE has its limitations; being an operator-dependent imaging technique, it needs an accurate image interpretation.

Our results demonstrate that the DL techniques could be successfully applied to computer-aided disease diagnosis for detecting inflammation on endomicroscopy images. Our DL method improved the performance over traditional ML models. When applied for the classification of pediatric IBD, the traditional ML model achieved a diagnostic accuracy of 82.7% and an AUC of 0.87 [35]. When we used a DL CNN-LSTM method trained with 5,081 images and tested on 1,124 images, the final test accuracy was 95.3% with AUC of 0.98.

Our results are comparable or better than similar digital diagnosis studies. In the last years, several studies addressed the potential use of DL techniques for automated detection of pathologies from CLE images. A recent study presented an approach for the automatic detection of motion artifacts in probe-based confocal laser endomicroscopy (pCLE) images using a pre-trained Inception v3 network [9]. The authors obtained an AUC value of 0.92 and an overall accuracy of 94.8%. In another recent study, the authors reported an accuracy of 80.77% when a CNN was used to classify CLE images with intestinal metaplasia, gastric metaplasia and esophageal neoplasia [36]. When comparing CNN capacity of grading severity of ulcerative colitis with human reviewers on endoscopic images, it has been demonstrated that the performance of DL was similar with that of experienced human reviewers. The CNN could distinguish endoscopic remission from moderate-to-severe disease with a Sn of 83%, Sp of 96.0%, negative predictive value of 0.94 and positive predictive value of 0.87 [37]. Maeda et al. [38] obtained the following results: 91% accuracy, 74% sensitivity and 97% specificity in a study based on testing a CAD system potential to identify

persistent histologic inflammation on images obtained during endocytoscopy in 187 ulcerative colitis patients [38].

By considering several normal and pathologic images from the same patient, our model differentiates between normal and pathological samples from the same patient as well as between pathological and normal images from different patients. By training the CNN-LSTM model on normal images from the same patient as well as other patients we took into account both intra- and inter- patient variances. Nonetheless the model can be further improved by increasing the number of patients.

Our patient spectrum included the inflammation group of patients with active inflammation and CD diagnosis on one hand, and the control group composed of patients with mucosal healing and normal mucosa on the other hand. Most of the control patients (82%) were mucosal healing patients; therefore, our test is mostly focused on differentiating between actively inflamed and healing mucosa samples rather than between mucosal inflammation and healthy patients. Although, in the control group we combined mucosal healing and some healthy patients (18% of control patients) which could decrease the overall accuracy of the test, the application of our CNN-LSTM model resulted in over 95% diagnosis accuracy. Future studies will further stratify the patients in three (CD, mucosal healing and healthy patients) or more categories to delineate more subtle differences between each group.

Further algorithm development will include prospective data. As required by DL technique, the image data set was analyzed retrospectively to be able to train the model on known diagnosis. Future development will include new images which will be digitally diagnosed with our algorithm and confirmed with traditional clinical diagnosis. Also, our DL analysis was not designed to differentiate between active and inactive inflammation. More stratified intestinal mucosa images are necessary to train the ML algorithm on more detailed features. Furthermore, in clinical practice, the eCLE technology was replaced by pCLE which is used through the biopsy channel of a conventional scope. Thus, these results should be validated on images obtained with the pCLE technology, as well.

In clinical practice, our DL algorithm could be used as a digital diagnostic tool for MH assessment in IBD patients from real-time virtual biopsies during colonoscopy. The patient's CLE image is run through our DL algorithm in real time and compared with our database of thousands of normal and pathologic cases. A diagnostic of normal or inflammatory condition is automatically generated by the computer algorithm independently of the operator's experience and subsequently confirmed by a pathologist. Following the diagnosis, the gastroenterologists decide whether the current treatment is appropriate, or any modifications are needed. When MH is achieved, therapy can be de-escalated or even stopped. In our future work, we plan to collect more clinical data to increase our database and use different variations of CNN-LSTM architecture with transfer learning to further improve the performance of IBD classification.

CONCLUSION

In this study, we focused on finding a way to differentiate inflammation from normal colonic mucosa, in order to assess MH from CLE images. For this purpose, we developed a novel DL algorithm that is superior to the traditional ML methods as it does not require an initial description of visual features of medical images. The proposed CNN-LSTM model had a good performance in terms of the Sn (92.78%), Sp (94.6%), test accuracy (95.3%) and AUC (0.98).

Conflicts of interest: None to declare.

REFERENCES

1. Atreya R, Neurath M. Current and Future Targets for Mucosal Healing in Inflammatory Bowel Disease. Visc Med 2017;33:82-88. doi:10.1159/000458006

2. Flamant M, Roblin X. Inflammatory bowel disease: towards a personalized medicine. Therap Adv Gastroenterol 2018;11:1756283X1774502. doi:10.1177/1756283X17745029

3. Lamb CA, Kennedy NA, Raine T, et al. British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. Gut 2019;68(Suppl 3):s1-s106. doi:10.1136/gutjnl-2019-318484

4. Cintolo M, Costantino G, Pallio S, Fries W. Mucosal healing in inflammatory bowel disease: Maintain or de-escalate therapy. World J Gastrointest Pathophysiol 2016;7:1-16. doi:10.4291/wjgp.v7.i1.1

5. Pineton de Chambrun G, Peyrin-Biroulet L, Lémann M, Colombel J. Clinical implications of mucosal healing for the management of IBD. Nat Rev Gastroenterol Hepatol 2010;7:15-29. doi:10.1038/nrgastro.2009.203

6. Dave M, Loftus EV Jr. Mucosal healing in inflammatory bowel disease-a true paradigm of success? Gastroenterol Hepatol 2012;8:29-38.

7. Fugazza A, Gaiani F, Carra M, et al. Confocal Laser Endomicroscopy in Gastrointestinal and Pancreatobiliary Diseases: A Systematic Review and Meta-Analysis. Biomed Res Int 2016;2016:4638683. doi:10.1155/2016/4638683

8. Iacucci M, Ghosh S. Looking beyond symptom relief: evolution of mucosal healing in inflammatory bowel disease. Therap Adv Gastroenterol 2011;4:129-143. doi:10.1177/1756283X11398930

9. Aubreville M, Stoeve M, Oetter N, et al. Deep learning-based detection of motion artifacts in probe-based confocal laser endomicroscopy images. Int J Comput Assist Radiol Surg 2018;14:31-42. doi:10.1007/s11548-018-1836-1

10. Imran Razzak M, Naz S, Zaib A. Deep Learning for Medical Image Processing: Overview, Challenges and the Future. In: Dey N, Ashour AS, Borra S. (eds.) Classification in BioApps. Cap. 14. Springer International Publishing 2017:323-350. doi:10.1007/978-3-319-65981-7

11. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagn Progn Res 2018;2:11. doi:10.1186/s41512-018-0033-6

12. Shen D, Wu G, Suk H. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng 2017;19:221-248. doi:10.1146/annurev-bioeng-071516-044442

13. Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. Med Image Comput Comput Assist Interv 2013;16:411-418. doi:10.1007/978-3-642-40763-5_51

14. Wang H, Cruz-Roa A, Basavanhally A, et al. Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection. Conference Paper. SPIE Medical Imaging 2014. doi:10.1117/12.2043902

15. Xu J, Xiang L, Liu Q, et al. Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. IEEE Trans Med Imaging 2016;35:119-130. doi:10.1109/TMI.2015.2458702

16. Ye F, Pu J, Wang J, Li Y, Zha H. Glioma grading based on 3D multimodal convolutional neural network and privileged learning. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Kansas City, MO,2017. doi:10.1109/BIBM.2017.8217751

17. Su H, Xing F, Kong X, Xie Y, Zhang S, Yang L. Robust Cell Detection and Segmentation in Histopathological Images Using Sparse Reconstruction and Stacked Denoising Autoencoders. Med Image Comput Comput Assist Interv 2015;9351:383-390. doi:10.1007/978-3-319-24574-4_46

18. Ştefănescu D, Streba C, Cârţână ET, Săftoiu A, Gruionu G, Gruionu LG. Computer Aided Diagnosis for Confocal Laser Endomicroscopy in Advanced Colorectal Adenocarcinoma. PLoS One 2016;11:e0154863. doi:10.1371/journal.pone.0154863

19. Lu N, Wu Y, Feng L, Song J. Deep Learning for Fall Detection: Three-Dimensional CNN Combined With LSTM on Video Kinematic Data. IEEE J Biomed Health Inform 2019;23:314-323. doi:10.1109/JBHI.2018.2808281

20. Donahue J, Hendricks LA, Rohrbach M, et al. Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. IEEE Trans Pattern Anal Mach Intell 2017;39:677-691. doi:10.1109/TPAMI.2016.2599174

21. Karstensen JG, Săftoiu A, Brynskov J, et al. Confocal laser endomicroscopy: a novel method for prediction of relapse in Crohn's disease. Endoscopy 2016;48:364-372. doi:10.1055/s-0034-1393314

22. Kim ES. Role of Advanced Endoscopic Imaging Techniques in the Management of Inflammatory Bowel Disease. Clin Endosc 2017;50:424-428. doi:10.5946/ce.2017.143

23. keras-team/keras [Internet]. GitHub. 2020. Cited 1 September 2020. Available from: https://github.com/keras-team/keras

24. Abadi M. TensorFlow: learning functions at scale. ICFP 2016: Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming September 2016;51:1. doi:10.1145/2951913.2976746

25. van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. Comput Sci Eng 2011;13:22-30. doi:10.1109/MCSE.2011.37

26. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825-2830.

27. Nair V, Hinton G. Rectified Linear Units Improve Restricted Boltzmann Machines. Proceedings of the 27th International Conference on Machine Learning, June 21-24, Haifa, Israel, 2010.

28. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J Mach Learn Res 2014;15:1929–1958.

29. Kingma DP, Ba JL. Adam: A Method for Stochastic Optimization. Conference paper at ICLR 2015. Cited 1 September 2020]. Available from: https://arxiv.org/pdf/1412.6980.pdf

30. Deep Learning with Python. Machine Learning Mastery. 2018. Available at: https://machinelearningmastery.com/machine-learning-with-python

31. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLoS One 2015;10:e0118432. doi:10.1371/journal.pone.0118432

32. Chapman CG, Konda VJ. Confocal laser endomicroscopy in inflammatory bowel disease: achieving new depths in mucosal healing. Gastrointest Endosc 2016;83:792-794. doi:10.1016/j.gie.2015.11.002

33. Neumann H, Vieth M, Atreya R, et al. Assessment of Crohn's disease activity by confocal laser endomicroscopy. Inflamm Bowel Dis 2012;18:2261-2269. doi:10.1002/ibd.22907

34. De Palma GD, Maione F, Esposito D, et al. In vivo assessment of tumour angiogenesis in colorectal cancer: the role of confocal laser endomicroscopy. Colorectal Dis 2016;18:O66-O73. doi:10.1111/codi.13222

35. Mossotto E, Ashton JJ, Coelho T, Beattie RM, MacArthur BD, Ennis S. Classification of Paediatric Inflammatory Bowel Disease using Machine Learning. Sci Rep 2017;7:2427. doi:10.1038/s41598-017-02606-2

36. Hong J, Park B, Park H. Convolutional neural network classifier for distinguishing Barrett's esophagus and neoplasia endomicroscopy images. Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference. 2017:2892-2895.

37. Stidham RW, Liu W, Bishu S, et al. Performance of a Deep Learning Model vs Human Reviewers in Grading Endoscopic Disease Severity of Patients with Ulcerative Colitis. JAMA Netw Open 2019;2:e193963. doi:10.1001/jamanetworkopen.2019.3963

38. Maeda Y, Kudo SE, Mori Y, et al. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). Gastrointest Endosc 2019;89:408-415. doi:10.1016/j.gie.2018.09.024