Detection and Classification of Human Stool Using Deep Convolutional Neural Networks

YIN PUI CHOY¹, GUOQING HU¹ and JIA CHEN²

¹School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: 201920155701@mail.scut.edu.cn, gqhu@scut.edu.cn)
²School of Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou 510006, China (e-mail: tchenjia@gdpu.edu.cn)

Corresponding author: Jia Chen (e-mail: tchenjia@gdpu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 81703956.

ABSTRACT The diagnosis of functional gastrointestinal disorders and chronic digestive system diseases such as irritable bowel syndrome relies heavily on macroscopic examination of human stool specimens. However, traditional manual stool analysis processes are time-consuming and prone to human subjectivity errors that may lead to incorrect judgments. In this study, we employed deep convolutional neural networks (CNN) to automatically recognize and classify stools in macroscopic images. This approach is advantageous because it reduces the amount of direct interaction required by patients or medical staff, lowers the risk of cross-infection, and removes subjectivity in stool analysis. The U-Net segmentation model is applied to define the region of stools by generating a mask image. After combining the mask and the corresponding original image, the ResNeXt-50 classifier classifies the pre-processed stool image according to the modified Bristol Stool Form Scale (BSFS). Overall, the U-Net model yielded a mean intersection-over-union (mIoU) of 93.75% and an F-score of 0.9570, as the ResNeXt-50 classifier had a classification accuracy of 94.35% and showed decent performance in terms of predictive power. Our study may help to improve the quality of diagnosis and monitoring of diseases associated with bowel movement habits by providing reliable measurements in terms of stool form and consistency.

INDEX TERMS bristol stool form scale, convolutional neural network, deep learning, functional bowel disease, stool classification, stool segmentation.

I. INTRODUCTION

The human stool is a heterogeneous complex waste product naturally excreted from the body that involves a mixture of bacteria and food residues. The novel coronavirus “COVID-19”, which has caused a pandemic and has spread globally since the year 2020, was discovered in stool specimens from the infected patients, implying that it could be a potential source of virus transmission [1]. Besides, the global prevalence of functional bowel diseases is estimated to be around 11.7% [2], and irritable bowel syndrome (IBS) is the most common among all gastrointestinal (GI) disorder problems, accounting for 9.2% of all morbidity [3]. Unfortunately, only about 30% of people who have suffered IBS symptoms will seek medical attention to address their health problems [4], and approximately 10% of IBS patients may experience inappropriate medication interactions causing severe health problems and disrupts the quality of life [5], [6].

Unlike ordinary pathological assessments applied for cancer treatment, stool analysis depends primarily on visual inspection. Stool analysis is essential to diagnose various diseases involving functional bowel disorders and GI symptoms (including IBS [7], [8], pancreatitis [9], colorectal cancer [10], digestive system infections [11], chronic constipation [12], and chronic diarrhea [13]). But most comprehensive stool analyses can only be performed in medical laboratories [14], which would invariably necessitate the use of more expensive analytical equipment. Instead of inspecting the biochemical composition of stools under a microscope, stool samples can be examined macroscopically in terms of consistency, form, size, and color [15]. Analyzing the stool forms and consistency provides a more clinically feasible and straightforward method for gaining clinical insights into a wide range of diseases affecting the human GI tract [16]. The Bristol stool form scale (BSFS) is the most popular...
A significant amount of effort and time. Furthermore, few owing to a scarcity of stool images labeled by experts. The detection and classification systems have been developed, work is advantageous, only a few automatic human stool evaluation of treatment options [26].

According to a study, equipment and are routinely used in diagnostic procedures, cal images are typically generated by specialized diagnosis gradually evolved to facilitate accurate detection, recognition, and classification tasks of medical images [25]. Medical images are typically generated by specialized diagnosis equipment and are routinely used in diagnostic procedures, which benefits in the analysis of disease symptoms and the evaluation of treatment options [26].

Even though using CNN to replace manual inspection work is advantageous, only a few automatic human stool detection and classification systems have been developed, owing to a scarcity of stool images labeled by experts. The annotation and labeling of the stool images often require a significant amount of effort and time. Furthermore, few patients were willing to provide their stool samples due to shameful and embarrassing reasons [27], [28]. Hence, the shortage of labeled stool image datasets is currently an impediment to innovation in this field. To the extent of our knowledge, a few works [20], [29] obtained their fully labeled stool image dataset according to the BSFS, but they were unavailable at the time of writing. The other publicly available stool image datasets compiled by Leng et al. [28] and Yang et al. [30] were incompatible with our study because the sample images were not classified using the BSFS. In this manner, we attempted to collect and compile our dataset for research purposes that adhered to the BSFS classification scheme.

We present a deep learning-based approach to detect and classify human stools in macroscopic images in this paper. This study aims to provide an automatic assessment of bowel movements by analyzing stool forms and consistency. Our contributions are summarized below:

1) We attempt to develop an automated method for stool detection and type recognition using deep learning via CNN. The encoder-decoder architecture U-Net is used to remove background noise from the stool images, and the 50-layer CNN ResNet-50 is then used to classify the type of stool based on BSFS. On our own collected stool image dataset, the proposed method achieves higher segmentation and classification accuracy results when compared to related state-of-the-art methods.

2) Our method is applicable in a variety of medical and homecare settings because it can provide valuable and accurate information in assessing bowel movement habits. It also helps to reduce the burden and risk of cross-infection on medical staff while also assisting doctors in improving disease diagnosis and the efficacy of future treatments for patients.

The rest of this paper is structured as follows: Section II introduces the related works. Section III goes into detail about the proposed method. Section IV illustrates the experiments and results. Section V provides the discussion of our study, and Section VI draws the overall conclusions.

II. RELATED WORKS

A. MEDICAL IMAGE SEGMENTATION

The primary goal of image segmentation is to detect, localize, and predict the dense pixel-wise output of the object of interest. Deep learning has recently demonstrated great promise for general image semantic segmentation tasks in the healthcare sector [31]. For example, recognizing brain tumors in Magnetic Resonance Imaging (MRI) images [32], cancer detection [33], [34] and lesions identification [35], [36]. These works have all produced excellent results when neither the sizes, shapes, nor positions are fixed and will be useful to doctors in assessing and managing medical treatments and pathogenesis [37].

Nevertheless, object detection of human stools has been strategically challenging due to a lack of annotated stool
image datasets, which have prevented the widespread use of the automatic stool detection and tracking system. Some methods used threshold-based segmentation techniques to localize the stool region in the original images [27], [28], [30]. However, because of the inability to remove all unrelated background information, these methods frequently produced unsatisfactory predictions. Unlike the threshold-based segmentation scheme, Hachuel et al. [29] used a deep learning CNN segmentation algorithm named SegNet [38] and produced moderate results.

B. STOOL SPECIMEN CLASSIFICATION
Stool form refers to the visible shape and texture of the stool, which can be examined visually and used as a proxy measurement. In the 7-level BSFS, the type of stools near the clinical decision points (such as the boundary between type-2 and type-3 stools; and the boundary between type-5 and type-6 stools) were found extremely difficult to distinguish [18], [21]. Several studies have revealed that modifying the BSFS is required [18], [39] to improve the classification accuracy, and a common approach is merging the 7-level BSFS into three main classes (see Table 1) [39].

Hachuel et al. [29] introduced a deep structured CNN residual neural network (ResNet) [40] to classify macroscopic synthetic stool images. Because of its faster execution time and smaller parameter size, ResNet-18, the shallowest architecture (with 18 denoting the number of layers contained in the model), was chosen over other variations. Park et al. [20] utilized the Inception v.3 [41] CNN to create a human stool image analysis unit and integrated it into a mountable toilet system. Both works have used the modified BSFS as their classification standard and proved that modified BSFS may improve the recognition performance.

Some other researchers discovered shallow structured CNN are useful to classify stool medical images in hospital settings. Leng et al. [28] developed a lightweight practical CNN classifier for recognizing feces traits. Yang et al. [30] proposed a similar shallow framework CNN dubbed “StoolNet” to classify the colors of fecal images. But the goals of these works were not to investigate the stool forms and consistency. Hence, the classification schemes did not follow the well-known BSFS.

III. MATERIALS AND METHODS
In this section, we will first introduce the stool dataset that we collected for this study. Then we present the encoder-decoder U-Net CNN architecture to detect the location of stools through semantic segmentation. Next, after localizing the stool region, the deep structured CNN named ResNeXt-50 classifies the preprocessed image based on the modified BSFS taxonomy. Fig. 2 depicts the overall framework of our proposed method.

A. OVERVIEW OF THE STOOL IMAGE DATASET
To the best of our knowledge, no publicly available stool image datasets followed the BSFS as the classification standard. Therefore, we collected 1103 stool image samples from online search engines (keywords included ‘human stool’ and ‘feces’) and with the addition of 15 stool images captured by ourselves using the back camera of an Honor Magic 2 smartphone. Due to a large portion of the stool specimen images in our dataset being either provided or anonymously uploaded to the internet, the age and gender of patients, location, and time range of data collection were all unknown. In general, the dataset mainly contains images of stools in toilets, and Table 2 shows a breakdown of our dataset.

| Stool image collection strategy | Total |
|--------------------------------|-------|
| Acquired from the internet     | 1,103 |
| Captured by ourselves          | 15    |
| **Total**                      | **1,118** |

We noticed a range of factors that affect the image quality, such as complicated background illumination, containing irrelevant items causing severe occlusions, and low-resolution imaging. This is primarily because each image was taken from different resolutions, various toilet constructions, distinct background illumination levels, and altered camera configurations. Furthermore, when the stools were immersed or soaked in toilet water, their texture could soften and their shape may alter, making the stool form recognition relatively difficult with low data quality issues.

To ensure reliable predictions for the classification task, we removed the stool images that were either out of focus or displayed severe occlusions. We are aware that any incorrect labels or descriptions provided by specialists may significantly impair the training efficiency of the CNN for classification. As a result, only 724 clear and identifiable stool image samples were evaluated by expert physicians.
using the modified BSFS shown in Table 1. After the expert physicians categorized the stool images according to the modified BSFS, we experienced an unequal distribution of classes in the dataset as shown in Fig. 3.

The encoder module is also known as the feature extraction backbone. As recommended by the authors in [42], the encoder backbone is based on the pre-trained 16-layer Visual Geometry Group (VGG-16) framework [44]. In this manner, we employed transfer learning to shorten training time and conserve computational resources during training. Transfer learning implies that the deep learning network is first trained to solve one initial specific task, and then the trained model is used as a new starting point for solving another related task [45].

The decoder module includes four feature layer blocks and an output layer. The main objective is to combine and semantically project the learned features from the encoder onto the pixel space. A series of operations such as up-sampling and concatenation was used to obtain a final feature layer that contains all contextual information. Each layer block starts with two 3 × 3 convolutional layers, followed by a batch-normalization (BN) layer and Rectified Linear Unit (ReLU) activation. The convolutional layers have filter numbers 64, 128, 256, and 512, which correspond to the VGG-16 layers in the encoder. At the end of the network, the classification output layer is a 1 × 1 convolutional layer with sigmoid activation to generate the prediction map, which estimates the probability of each pixel in the original input image belonging to the target region.

Selecting the suitable loss function for a CNN is essential as it delivers a way to quantify the discrepancy between the predicted segmentation results and the ground truth manual annotations. Class imbalance issues are common in stool image segmentation because fecal matter areas can take shape in small or thin segments of the entire image. Therefore, we adopted the “Dice and cross-entropy (CE)” hybrid loss function, which is effective in addressing class imbalanced issues [46]. The “Dice and CE” loss function improves segmentation quality by gradually learning better model parameters and preventing model parameters from remaining at bad local minima, and is computed by the following equations:

\[
Loss_{Total} = Loss_{CE} + Loss_{Dice} \\
Loss_{CE} = -\sum_{i}^{N} y_i \log p_i \\
Loss_{Dice} = 1 - \frac{\sum_{i}^{N} y_i \cdot p_i}{\sum_{i}^{N} y_i + \sum_{i}^{N} p_i}
\]
where \( y_i \) and \( p_i \) are the ground-truth value and the predicted value of the \( i \)-th pixel of an image, respectively, and \( N \) represents the total number of pixels in the stool image.

### C. IMAGE PREPROCESSING STAGE

One of the preliminary steps in solving general classification problems is to remove the background from the image and identify the region of interest (ROI). The ROI has a direct impact on the accuracy predicted by the latter CNN classifier. We obtain the ROI accurately by combining the mask image generated by the U-Net model from the previous stage with the corresponding original image, as displayed in Fig. 5.

![Figure 5. Localizing the stool region through image processing.](image)

The previous deep structured CNN stool image classification methods by Hachuel et al. [29] and Park et al. [20] were selected as the basis of our approach. Following this manner, we decided to utilize ResNeXt [47], as the invention of the ResNeXt architecture draws on the advantages of both the ResNet [40] and Inception [41] networks.

The ResNeXt CNN is constructed by stacking a series of identical topology blocks together with aggregate residual transformations. It incorporates the main architectural characteristics of ResNet and Inception, such as multiscale pathway block, split-transform-merge paradigm, and skip connections between layers to boost the accuracy while maintaining low network complexity. The authors in [47] introduced the cardinality parameter, which is defined as the number of aggregate transformations. Fig. 6 portrays the residual block structure of ResNeXt that consists of 32 blocks with the same topology, meaning that the cardinality value is 32.

The ResNeXt-50 (32 × 4d) was chosen for the stool classification task because it has the shallowest architecture among all other ResNeXt variant models. As implied by its name, it contains 50 layers and the parameters inside the bracket represent cardinality and width of the bottleneck layer, respectively. For the sake of simplicity, this network is referred to as ResNeXt-50 throughout the rest of this paper.

Moreover, we establish a trichotomous decision classifier, which corresponds to the three different types of stool in the modified BSFS classes shown in Table 1 (i.e. “constipated,” “normal,” or “loose”). We adopted the CE loss function (equation 2) because it is well-suited to multiclass classification problems. Softmax function is then applied at the

![Figure 4. Structure of the U-Net segmentation model](image)
probability distribution. Softmax is defined as follows:

$$softmax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$  \hspace{1cm} (4)$$

where i represents the class index, $z_i$ is the probability of i-th class, and k is the total number of classes.

IV. EXPERIMENTS AND RESULTS

In this section, we use experimental comparisons to assess the performance of our proposed methodology. Each module is described in detail, including data processing and augmentation, evaluation metrics, and implementation details. We also demonstrate the results of the CNN by comparing our method to other methods in the same field.

Furthermore, all experiments in this paper were carried out with our own collected stool image dataset. The CNN models were trained and configured with an Intel Core i7-7500U CPU @2.70GHz, 16GB internal storage, NVIDIA GeForce 940MX GPU, and a Windows 10 64-bit operating system. The image preprocessing and training codes were written in Python, with Pycharm serving as the integrated development environment (IDE) compiler.

A. STOOL IMAGE SEGMENTATION EXPERIMENTS AND RESULTS

1) Data augmentation

Data augmentation is useful to avoid the problem of overfitting and algorithm bias. During the segmentation experiment, the training data are artificially expanded to boost the expressive power. We utilized the following augmentation transformation techniques: random horizontal and vertical flips, random rotations and skews, and random crops or zooms with scales ranging from 0.5 to 1.5.

2) Evaluation metrics

The mean intersection-over-union (mIoU) and F-score (also known as Dice score) are commonly used to evaluate the segmentation performance of a CNN. The mIoU is used to compare the similarity between the stool region recognized by the model and the ground truth, and it can be calculated in the following equations:

$$IoU = \frac{TP}{TP + FP + FN}$$  \hspace{1cm} (5)$$

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} IoU_i$$  \hspace{1cm} (6)$$

where TP represents the number of true positives; FP is the number of false positives; FN means false negatives estimated by the model.

F-score is the weighted harmonic mean of precision and recall and is defined as:

$$F-score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$  \hspace{1cm} (7)$$

where the mean precision is the fraction of accurately predicted stool pixels to the number of pixels predicted to be stool; the mean recall is the fraction of accurately predicted stool pixels to the number of ground truth stool pixels in the image.

In addition, inference time is used to assess the computational efficiency of the SegNet and our U-Net model, which denotes the milliseconds (ms) required for segmentation on a single stool image.

3) Implementation details

We loaded the VGG-16 pre-trained weights which have been previously trained on the ImageNet dataset (around 1.4 million labeled images and 1000 object categories) [48]. Moreover, we employ the reduced learning rate on the plateau strategy to let the learning rate automatically be reduced by a constant factor during training. The base learning rate is set to 0.001. We also adopt the stochastic gradient descent (SGD) with a momentum value of 0.9 and a batch size of 2 to optimize the U-Net for a duration of 100 iterations. Before being fed into the CNN, each original stool image in our dataset was resized using cubic spline interpolation, whereas each ground-truth label was resized using nearest-neighbor interpolation to a resolution of 512 $\times$ 512. Following the standard CNN training procedure, we randomly divided the stool image dataset into 75% for training and the remaining 25% for validation testing.

4) Comparison with state-of-the-art methods

Since there are very few developed deep CNN-based stool image segmentation methods, we could only compare the test performance of our U-Net with the state-of-the-art model named SegNet in [29]. However, the approach in [29] takes the binary CE loss function into account, and the model was trained using their proprietary dataset. Therefore, we separately trained and tested a SegNet model using the same...
setup as our experiments. Additionally, to evaluate the effectiveness of the “Dice and CE” loss function, we trained and tested a comparison experiment with the binary CE loss function on both U-Net and SegNet. In Table 3, we present the numerical test results of comparisons, which include training and validation loss values as well as evaluation metrics such as the inference time, F-score, and mIoU. We also compared our F-score and mIoU ratings to Otsu’s threshold-based edge detection method [49], which had previously been used by researchers in the same field [27], [28], [30].

From Table 3, we can see that our U-Net with “Dice and CE” loss function increases the mIoU rating from the most of 50.69% to 93.75% by 43.06%, and we achieve an F-score of 0.9570. SegNet and U-Net have comparable average inference times because both models have a similar structure and number of total trainable parameters. In addition, the “Dice and CE” loss function tends to yield lower training loss and validation loss values than the “Binary CE” loss function in both SegNet and U-Net models. In particular, U-Net has the lowest training and validation loss values of 0.0564 and 0.0453, respectively, indicating greater robustness. The parameters of the loss function become smaller as the mIoU between the predicted map and the ground-truth mask increases.

Fig. 7 shows the training and validation loss curves of U-Net with “Dice and CE” loss. At the start of the training process, both loss curves drop rapidly. After around 10 epochs, the training loss curve continues to decline while the validation loss curve fluctuates, signaling that the model can be further optimized. Later, both loss curves become smooth, with only minor fluctuations remaining, demonstrating that the U-Net has appropriately converged.

Fig. 8 depicts a visual qualitative analysis of the image segmentation method outputs. Fig. 8(a) shows the original stool images taken in different flush toilets, Fig. 8(b) shows the ground-truth mask images, and Fig. 8(c), (d), (e) displays the predicted mask images generated by Otsu’s method, SegNet and our proposed method, respectively.

We observe that Otsu’s method cannot exclude the edges of the toilet construction, resulting in unsatisfactory and noisy results in Fig. 8(c). On the other hand, although Table 3 listed that the F-score and mIoU ratings for both U-Net and SegNet are similar, the predicted stool mask images are quite different according to Fig. 8(d) and (e). U-Net predicted masks are much more similar in shape to the ground truth with appropriate stool contours, whereas SegNet tends to segment the target with irregularities along the ROI boundaries. Even though the position, sizes, and shapes of the stool, and the level of background illumination vary in the original images, U-Net can still accurately segment the ROI as a continuous region, demonstrating the robustness of the model. Hence, we can conclude that our proposed method achieves accurate and clearer results in the stool detection task.

**B. STOOL IMAGE CLASSIFICATION EXPERIMENTS AND RESULTS**

1) **Image processing and data augmentation**

We used the simple binary bitwise operations from the OpenCV library [50] in Python to suppress background noise and localize the stool region for each image sample in our dataset.

As previously stated, we removed the images with occlusions caused by reflections or irrelevant objects, leaving only 724 clear stool images remaining. But we experienced an unequal distribution of classes in the dataset as shown in Fig. 3, the classifier estimations could then be one-sided or biased towards the group that appeared more frequently in the dataset.

As we observed, the classifier method was unable to handle the different types of stools. The classifier consistently predicted the “Constipated” class, even though the stool images were quite different. To address this issue, we decided to augment the stool images in specific classes. Each image in the “Constipated” class was rotated seven times,

---

**TABLE 3.** Experimental results for stool segmentation (best results marked bold)

| Model            | Inference Time (ms) | Loss Function   | Training Loss | Validation Loss | F-score | mIoU(%) |
|------------------|---------------------|-----------------|---------------|-----------------|---------|---------|
| SegNet [29]      | 475                 | Binary CE       | 0.0978        | 0.0864          | 0.9285  | 74.82   |
|                  |                     | Dice and CE     | 0.0730        | 0.0680          | 0.9403  | 90.14   |
| U-Net (Proposed) | 462                 | Binary CE       | 0.1150        | 0.1052          | 0.9042  | 76.93   |
|                  |                     | Dice and CE     | 0.0564        | 0.0453          | 0.9570  | 93.75   |
| Otsu’s           | -                   | -               | -             | -               | 0.2654  | 50.69   |

FIGURE 7. The loss trends of U-Net.
while each image in the “Loose” class was rotated three times. The augmented images were then horizontally flipped. After that, we added Gaussian noise to each original and augmented image to increase the total number of training samples. At last, the dataset contains 3540 stool images, with 1218, 1236, and 1088 images in the “Constipated”, “Normal”, and “Loose” classes, respectively.

Some of the augmented images with horizontal and vertical flips and rotations produced different shapes. However, the other augmented images salted with Gaussian noise were assumed to have the same stool reference shape and orientation as the original image. Thus, instead of randomly dividing the expanded dataset, we try to place images with identical reference stool shapes in either the training or validation sets whenever possible. We were driven to accomplish this procedure for two reasons. Firstly, on most practical occasions, finding two duplicate stool specimens is very uncommon. Secondly, we purposefully enhanced the complexity by limiting the number of shape styles that may be learned through training, which is reasonable for verifying the effect and robustness of our proposed CNN classifier. After all, we split the dataset into 75% (2605 samples) for training and the remainder 25% (885 samples) for validation. The images were resized using cubic spline interpolation to a resolution of $224 \times 224$ pixels while maintaining the average aspect ratio to suit the input size of the ResNeXt-50 network.

2) Evaluation metrics
To effectively evaluate the performance of the classification CNN, we combined the experience from previous studies. Leng et al. [28], Hachuel et al. [29], and Yang et al. [30] selected the accuracy as a measurement indicator, while Park et al. [20] used the area under the receiver operating characteristic (ROC) curve (AUC) as a performance metric. Moreover, we included the inference time metric to examine the average time required for each model to classify a single stool image. The accuracy is a popular assessment criterion to describe the percentage of accurately predicted images made by the classifier and is defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

where the variables TP, TN, FP, FN denotes True Positive, True Negative, False Positive, and False Negative, respectively.

We also adopt the AUC-ROC analysis to evaluate the predictive power of the classifier. But because the AUC is typically used for diagnostic tests with dichotomous outcomes, we employ the one-versus-all approach to reduce the multi-class classification into multiple binary classification problems for each of the modified BSFS classes. This approach compares how well the classifier can predict positive samples from the corresponding class against the negative samples from all other classes. Hence, in this experiment, we generate ROC curves and AUC for each of the “Constipated,” “Normal,” and “Loose” classes, and AUC is defined as follows:

$$AUC = \frac{1}{2} \cdot \sum_{i=1}^{m-1} (x_{i+1} - x_i)(y_i + y_{i+1}) \quad (9)$$

where $x$ represents the value of true positive rate (TPR) and $y$ is the false positive rate (FPR) from the ROC curve. In this experiment, we obtained the AUC by using the algorithm provided in the Scikit-learn Python library. Additionally, the classification accuracy, the inference time, ROC curves, and AUC scores are computed based on the validation dataset, which contains 304, 309, and 272 stool images, respectively, for the “Constipated,” “Normal,” and “Loose” classes.
3) Implementation details

The adaptive momentum (Adam) [51] optimizer is adopted with a learning rate of 0.001, a batch size of 8, and the maximum epoch is set to 100. Following the standard practice in transfer learning, we loaded the ResNeXt-50 pre-trained weights which have been previously trained on the ImageNet dataset [48].

4) Comparison with state-of-the-art methods

Previous research used either deep or shallow structured CNN to classify stool types automatically. However, no comparisons were made between the various methods for recognizing stools with the BSFS. For this reason, we compared our proposed method to state-of-the-art models on our stool image dataset using the same experimental setup, including both deep structured CNN (ResNet-18 [29] and GoogLeNet Inception v.3 [20]) and lightweight shallow CNN (the model proposed by Leng et al. [28] and StoolNet [30]).

As shown in Table 4, the deep CNN structured models have a significantly longer inference time because of the complexity of the structure. However, the ResNeXt-50 had a classification accuracy of at least 2.94% higher than the other state-of-the-art methods. We also demonstrated the confusion matrix in Fig. 9 to present a more intuitive understanding of the results. We expect most of the values to be on the diagonal cells of the confusion matrix, indicating that the samples are correctly classified. The darker the blue color in the confusion matrix square, the greater the value, while a light blue color cell represents lower values.

Fig. 9 shows that most of the values are along with the dark blue colored diagonal cells, and the other cells have remained in light blue. It indicates the classification accuracy of the model was excellent. Furthermore, the predictions made by the ResNeXt-50 were unbiased, meaning that we have addressed the issue of uneven stool specimen ratings through data augmentation.

As for the AUC-ROC analysis, we display the AUC scores for the corresponding modified BSFS classes of various CNN methods in Table 5. We then present the ROC curves for our ResNeXt-50 classifier and the related state-of-the-art methods in Fig. 10. The black diagonal dotted line represents a random guessing classifier with an AUC of 0.5, which means it can only predict whether an image belongs to the appropriate category at random.

As can be seen in Fig. 10, all ROC curves for different methods have a convex shape and are significantly higher above the random guessing black dotted line. The ROC curves of the ResNeXt-50 model in Fig. 10(c) showed that slopes are the steepest, and Table 5 shows that the AUC scores for our method are the highest in each of the modified BSFS classes when compared to other various methods. It means that our proposed method has better performance in terms of predictive power. Hence, the ResNeXt-50 outperforms all other state-of-the-art methods by a considerable margin and can accurately follow the labeling decisions made by the expert physicians in our collected dataset.

V. DISCUSSION

Although our approach is similar to the Hachuel et al. [29] study on the recognition and characterization of human stool in macroscopic images, the two are not directly comparable due to several significant differences, which are summarized below:

1) There are differences between the stool image datasets. The method described in [29] trained their deep-CNN classifier on a stool image dataset containing only Type-1 to Type-5 synthetic stool samples. In contrast, our dataset contains real human stool images, including loose stools (Type-6 and Type-7) from all modified BSFS categories;

2) The method described in [29] did not preprocess the stool images to remove background noise before feeding them into the classifier for training and validation. On the other hand, we preprocessed each stool image sample in our dataset by combining the generated mask image and the corresponding original image;

3) The experimental results show that the method described in [29] is incapable of maintaining the level of performance demonstrated in this paper. Our proposed method performs well in both the automatic detection/segmentation and classification tasks.

One current limitation of our approach is that the type-7 stools have a similar phase and color appearance of urine in deep flush toilets. The annotation of type-7 stool images is challenging because it lacks a distinct shape or size, and we only have a limited number of type-7 stool image samples in our collected dataset. Consequently, the segmentation of the watery stool region becomes challenging in these extreme cases.
TABLE 4. The classification accuracy and inference time comparison of different methods (best results marked bold)

| CNN structure | Model | Accuracy(%) | Inference Time (ms) |
|---------------|-------|-------------|---------------------|
| Deep          | Inception v.3 [20] | 55.37 | 408                  |
|               | ResNet-18 [29]    | 91.41 | 167                  |
|               | ResNeXt-50 (Proposed) | **94.35** | **240**             |
| Shallow       | Leng et al. [28]  | 62.49 | 33                   |
|               | StoolNet [30]     | 79.77 | 89                   |

FIGURE 10. Multiclass receiver operating characteristic (ROC) curve and area under the curve (AUC) for our proposed model and the various state-of-the-art methods. (a) Inception v.3 classifier. (b) ResNet-18 classifier. (c) Our proposed ResNeXt-50 classifier. (d) Light-weight classifier proposed by Leng et al. [28]. (e) StoolNet classifier.

TABLE 5. Values of AUC between various methods under different modified BSFS classes (best results marked bold)

| Model             | Constipated | Normal | Loose  |
|-------------------|-------------|--------|--------|
| Inception v.3 [20]| 0.943       | 0.881  | 0.940  |
| ResNet-18 [29]    | 0.976       | 0.959  | 0.982  |
| ResNeXt-50 (Proposed) | **0.988** | **0.979** | **0.994** |
| Leng et al. [28]  | 0.881       | 0.784  | 0.835  |
| StoolNet [30]     | 0.870       | 0.900  | 0.788  |

Hence, in the future, we intend to enlarge our stool image dataset. We seek to obtain more stool image samples from an actual hospital setting, especially the watery stool images.

To avoid watery stools mixing with urine or dissolving in toilet water, we could recommend the patients dispose of their stools properly in a transparent excrement container, if possible. The dataset scale expands when more stool images are collected, and the performance of both segmentation and classification CNN will be further improved [20].

The CNN models used in our study have a deep depth network, which requires longer average inference time to generate results than shallow structures. So, we will try to optimize our approach further using Multi-Task Joint Learning (MTL) to improve the computational efficiency of our algorithms, ultimately leading to the development of a quick, precise, and robust stool detection and examination method.

In addition, we can use a similar approach described in this paper to evaluate other aspects of stool, such as its color and size. This will be useful to provide a detailed description...
of the hydrodynamics of human defecation [52]. At last, we look forward to applying our method in conjunction with additional clinically relevant assays to a “Smart Toilet” system for prognostic and continuous health monitoring purposes.

VI. CONCLUSION

In this paper, we describe an automated clinically relevant stool analysis using a deep-CNN approach. We acquired actual human stool images to establish a dataset for appearance detection and classification of stool form. The technical contribution of our paper is to combine the U-Net model with the “Dice and CE” hybrid loss function and the ResNeXt-50 model to extract and characterize the stool object in macroscopic images. To that end, the VGG-16 network is used in the U-Net encoder module; both the VGG-16 and the ResNeXt-50 networks are off-the-shelf networks that allow us to take advantage of transfer learning. Comprehensive experimental results revealed that U-Net could precisely segment the boundaries of the stool target region, and the ResNeXt-50 classifier accurately classifies the images based on the modified BSFS with clinical-level decision-making performance. Overall, our proposed method satisfies the requirements for practical application, such as serving as a treatment evaluation or a daily bowel movement tracking system to facilitate accurate communication between patients and doctors. The findings of our study are potentially an important contribution to improving the diagnosis and monitoring of disorders connected with bowel movement behaviors.

REFERENCES

[1] J. Hindson, “COVID-19: faecal-oral transmission?”, Nat Rev Gastroen- terol Hepatol, vol. 17, no. 5, p. 259, 2020.
[2] A. D. Sperber et al., “Worldwide Prevalence and Burden of Functional Gastrointestinal Disorders, Results of Rome Foundation Global Study,” Gastroenterology, vol. 160, no. 1, pp. 99–114, 2021.
[3] P. Oka, H. Parr, B. Barberio, C. J. Black, E. V. Savarino, and A. C. Ford, “Global prevalence of irritable bowel syndrome according to Rome III or IV criteria: a systematic review and meta-analysis,” Lancet Gastroenterol Hepatol, vol. 3, no. 5, pp. 908–917, 2020.
[4] C. Canavan, J. West, and T. Card. “The epidemiology of irritable bowel syndrome,” Clin Epidemiol, vol. 6, pp. 71–80, 2014.
[5] F. Mearin et al., “Irritable bowel syndrome with constipation and func- tional constipation in adults: Treatment (Part 2 of 2),” Semergen, vol. 43, no. 2, pp. 123–140, 2017.
[6] T. R. Card, J. Siffledeen, and K. M. Fleming. “Are IBD patients more likely to have a prior diagnosis of irritable bowel syndrome? Report of a case-control study in the General Practice Research Database,” United European Gastroenterol J, vol. 2, no. 6, pp. 505–512, 2014.
[7] O. S. Palsson, J. Baggish, and W. E. Whitehead, “Episodic nature of symptoms in irritable bowel syndrome,” Am J Gastroenterol, vol. 109, no. 9, pp. 1450–1460, 2014.
[8] S. E. Fehnel et al., “Development of the Diary for Irritable Bowel Syn- drome Symptoms to Assess Treatment Benefit in Clinical Trials: Founda- tional Qualitative Research,” Value Health, vol. 20, no. 4, pp. 618–626, 2017.
[9] B. M. Shandro, R. Nagarajah, and A. Poullis, “Challenges in the manage- ment of pancreatic exocrine insufficiency,” World J Gastroenterol, vol. 9, no. 5, pp. 39–46, 2018.
[10] T. F. Imperiale, D. F. Ransohoff, and S. H. Izikowitz, “Multitarget stool DNA testing for colorectal-cancer screening,” N Engl J Med, vol. 371, no. 2, pp. 187–188, 2014.
[11] A. Nakarii, J. Kato, S. Hiraoka, S. Takashima, D. Takei, T. Inokuchi, Y. Sugihara, M. Takahara, K. Harada, and H. Okada, “Ulcerative colitis patients in clinical remission demonstrate correlations between fecal immunochemical test results, mucosal healing, and risk of relapse,” World J Gastroenterol, vol. 22, no. 21, pp. 5079–5087, 2016.
[12] A. D. Markland, O. Palsson, P. S. Goode, K. L. Burgio, J. Busby-Whitehead, and W. E. Whitehead, “Association of low dietary intake of fiber and liquids with constipation evidence from the national health and nutrition examination survey,” Am J Gastroenterol, vol. 108, no. 5, pp. 796–803, 2013.
[13] C. M. Theriot and J. R. Fletcher. “Human fecal metabolomics profiling could inform Clostridiodes difficile infection diagnosis and treatment,” J Clin Invest, vol. 129, no. 9, pp. 3539–3541, 2019.
[14] O. T. Nkamgang, D. Khotsiotop, H. B. Fotsin, P. K. Talla, V. Louis Dorr, and D. Wolf, “Automating the clinical stools exam using image processing integrated in an expert system,” Informatics in Medicine Unlocked, vol. 15, p. 100165, 2019.
[15] M. M. Lane, D. I. Czyzewski, B. P. Champitazzi, and R. J. Shulman, “Reliability and validity of a modified Bristol Stool Form Scale for children,” J Pediatr, vol. 159, no. 3, pp. 437–441, 2011.
[16] S. J. Lewis and K. W. Heaton, “Stool form scale as a useful guide to intestinal transit time,” Scand J Gastroenterol, vol. 32, no. 9, pp. 920–924, 1997.
[17] L. Vork, E. Wilms, J. Penders, and D. M. A. E. Jonkers, “Stool Consis- tency: Looking Beyond the Bristol Stool Form Scale,” J Nutr Gastroenterol, vol. 25, no. 4, p. 625, 2019.
[18] M. R. Blake, J. M. Raker, and K. Whelan, “Validity and reliability of the Bristol Stool Form Scale in healthy adults and patients with diarrhoea- predominant irritable bowel syndrome,” Aliment Pharmacol Ther, vol. 44, no. 7, pp. 693–703, 2016.
[19] A. Malhotra, N. Shah, J. Depasquale, W. Baddoura, R. Spira, and T. Rec- tor, “Use of Bristol Stool Form Scale to predict the adequacy of bowel preparation - a prospective study,” Colorectal Disease, vol. 18, no. 2, pp. 200–204, 2016.
[20] S. M. Park et al., “A mountable toilet system for personalized health monitoring via the analysis of excreta,” Nat Biomed Eng, vol. 4, no. 6, pp. 624–635, 2020.
[21] E. P. Halmos, J. R. Biesiekierski, E. D. Newnham, R. E. Burgell, J. G. Muir, and P. R. Gibson, “Inaccuracy of patient-reported descriptions and satisfaction with bowel actions in irritable bowel syndrome,” Neuro- gastroenterol Motil, vol. 30, no. 2, p. e13187, 2018.
[22] A. Vial, D. Stirling, M. Field, M. Ros, C. Ritz, M. Carolan, L. Holloway, and A. A. Miller, “The role of deep learning and radiomic feature ex- traction in cancer-specific predictive modelling: a review,” Translational Cancer Research, vol. 7, no. 3, pp. 803–816, 2018.
[23] M. G. R. Alam, S. F. Abedin, S. I. Moon, A. Talukder, and C. S. Hong, “Healthcare IoT-Based Affective State Mining Using a Deep Convolu- tional Neural Network,” IEEE Access, vol. 7, pp. 75189–75202, 2019.
[24] N. Tadjbaksh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and L. Jianming, “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?,” IEEE Trans Med Imaging, vol. 35, no. 3, pp. 1289–1302, 2016.
[25] S. S. Yadav and S. M. Jadav, “Deep convolutional neural network based medical image classification for disease diagnosis,” Journal of Big Data, vol. 6, no. 1, pp. 1–18, 2019.
[26] D. Wang, G. Hu, and C. Lyu, “Multi-path connected network for medical image segmentation,” Journal of Visual Communication and Image Representation, vol. 71, 2020.
[27] P. Elfert, S. Siggelkow, M. Eichelberg, and A. Hein, “Towards an Ambient Estimation of Stool Types to Support Nutrition Counseling for People affected by the Geriatric Frailty Syndrome,” in 2020 IEEE Symposium on Computers and Communications (ISCC), pp. 1–6, 2020.
[28] L. Leng, Z. Yang, C. Kim, and Y. Zhang, “A Light-Weight Practical Framework for Feces Detection and Trait Recognition,” Sensors (Basel), vol. 20, no. 9, 2020.
[29] D. Huchuel, A. Jha, D. Estrin, A. Martinez, K. Staller, and C. Velez, “Aug- menting Gastrointestinal Health: A Deep Learning Approach to Human Stool Recognition and Characterization in Macroscopic Images,” arXiv, 2019.
[30] Z. Yang, L. Leng, and B.-G. Kim, “StoolNet for Color Classification of Stool Medical Images,” Electronics, vol. 8, no. 12, 2019.
[31] D. Wang, G. Hu, and C. Lyu, “FRNet: an end-to-end feature refinement network for medical image segmentation,” The Visual Computer, pp. 1101–1112, 2020.
[32] M. M. Thaha, K. P. M. Kumar, B. S. Murugan, S. Dhanasekeran, P. Vijayakarthick, and A. S. Selvi, “Brain Tumor Segmentation Using Convo-
et al.: Detection and Classification of Human Stool Using Deep Convolutional Neural Networks

YIN PUI CHOY is currently a postgraduate student at the South China University of Technology, China. He received the bachelor’s degree in mechanical engineering with honors in 2018 from the University of Adelaide.

His research interests are in computer vision and developing intelligent control systems based on image analysis.

GUOQING HU is a Professor and Ph.D. student supervisor with the School of Mechanical and Automotive Engineering, South China University of Technology, China. He received an M.S. degree from Northwestern Polytechnical University, China, and a Ph.D. degree from Sichuan University, China.

He was a Professor at Xiamen University, China. He was an advanced visiting scholar at the Chinese University of Hong Kong and the University of Nottingham. He was completed and participated in more than 90 projects including the National 863 project, the National Natural Science Foundation project, the national major projects, international cooperation projects, provincial key projects, the province fund cooperation projects. He has published 248 papers, 22 patents, and two textbooks. His research interests include amphibious flying machines, intelligent robots, industrial image processing, automation and industrial robot, electromechanical integration, and advanced sensor technology.

JIA CHEN is an Associate Professor with the School of Chinese Medicine, Guangdong Pharmaceutical University, China.

He is a professional in the field of Clinical Aspects of Integration of Traditional Chinese and Western Medicine. His research interests are in the application of artificial intelligence and deep learning to the diagnosis and treatment of functional bowel disorders diseases.

---

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/