Discover Imaging

Research

Deep learning imaging analysis to identify bacterial metabolic states associated with carcinogen production

Maysam Orouskhani[1] · Sarwesh Rauniyar[1] · Norma Morella[1] · Daniel Lachance[1] · Samuel S. Minot[2,3] · Neelendu Dey[1,2,4,5]

Received: 11 September 2024 / Accepted: 17 February 2025
Published online: 10 March 2025

Abstract

Background Colorectal cancer (CRC) is a globally prevalent cancer. Emerging research implicates the gut microbiome in CRC pathogenesis. Bacteria such as *Clostridium scindens* can produce the carcinogenic bile acid deoxycholic acid (DCA). It is unknown whether imaging methods can differentiate DCA-producing and DCA-non-producing *C. scindens* cells.

Methods Light microscopy images of anaerobically cultured *C. scindens* in four conditions were acquired at 100× magnification using the Tissue FAX system: *C. scindens* in media alone (DCA-non-producing state), *C. scindens* in media with cholic acid (DCA-producing state), or *C. scindens* in co-culture with one of two *Bacteroides* species (intermediate DCA production states). We evaluated three approaches: whole-image classification, per-cell classification, and image segmentation-based classification. For whole-image classification, we used a custom Convolutional Neural Network (CNN), pre-trained DenseNet, pre-trained ResNet, and ResNet enhanced by integrating the Digital Images of Bacterial Species (DIBaS) dataset. For cell detection and classification, we applied thresholding (OTSU or adaptive thresholding) followed by a ResNet model. Finally, image segmentation-based classification was performed using nnU-Net.

Results For whole-image analysis, DIBaS-enhanced ResNet models achieved the best performance in distinguishing *C. scindens* states in monoculture (accuracy 0.89 ± 0.006) and in co-cultures (accuracy 0.86 ± 0.004). Per-cell analysis was optimal at a C constant value of 3, with the ResNet model achieving 62–74% accuracy for *C. scindens* states in monoculture. Segmentation-based analysis using nnU-Net resulted in Dice coefficients of 87% for *C. scindens* and 74–76% for the *Bacteroides* species.

Conclusions This study demonstrates feasibility of image-based deep learning models in identifying health-relevant gut bacterial metabolic states.

Keywords Colorectal cancer · Microbiome · Deoxycholic acid · Deep learning

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s44352-025-00006-1.

✉ Neelendu Dey, ndey@fredhutch.org | [1]Translational Science and Therapeutics Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. [2]Microbiome Research Initiative, Fred Hutchinson Cancer Center, Seattle, WA, USA. [3]Data Core, Fred Hutchinson Cancer Center, Seattle, WA, USA. [4]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA. [5]Department of Medicine, Division of Gastroenterology, University of Washington, Seattle, WA, USA.

1 Introduction

Colorectal cancer (CRC) is among the most prevalent cancers globally. Early-stage CRC often presents no symptoms, emphasizing the importance of screening. Emerging data implicates the gut microbiome as a driver of CRC, thereby raising the prospect of a microbiome-based screening tool [1–3]. Gut bacteria can influence one's risk of developing CRC via the production of metabolites that can be carcinogenic or chemoprotective. One such metabolite is deoxycholic acid (DCA), a pro-inflammatory genotoxic bile acid which is produced from cholic acid by a subset of gut bacteria including *Clostridium scindens* [4–7]. While time and cost constraints limit the broad use of sequencing or mass spectrometry in screening, imaging methods could potentially be more easily deployed to identify microbial features linked to CRC risk.

High-resolution light microscopy is a commonly used tool for visualizing bacteria and their morphological cellular changes at a fine scale. Higher magnification imaging technology such as electron microscopy (EM) offers subcellular resolution and potentially more accurate deep learning classifiers, but EM is impractical as a routine clinical tool given cost and time requirements. As such, we restricted our study to light microscopy images because its generation is cheaper, faster, and reliant on equipment that is typically already available in clinical labs.

We asked whether deep learning models can classify different bacterial species or metabolic states based on light microscopy imaging data alone. Unlike traditional methods that rely on manual feature extraction, deep learning algorithms can automatically learn hierarchical representations of data directly from complex but structured grid data such as images. Convolutional neural networks (CNNs) utilize multiple layers to progressively extract higher-level features from raw input data, capturing essential features like edges, textures, and shapes. Pre-trained models like ResNet [8] and DenseNet [9] leverage large-scale image datasets to further enhance the performance of deep learning models.

We evaluated whole-image classification, per-cell classification, and classification following image segmentation for delineation of regions of interest using models such as U-Net [10] and nnU-Net [11], which have been applied effectively in cell counting [12] and cell structure analysis [13]. We reasoned that these methods might be used to extract subtle patterns and features in bacterial images that are not discernible to the human eye or with traditional analytical methods and utilize them for accurate classification of bacteria in different metabolic states.

2 Methods

2.1 Bacterial culturing and imaging

Bacterial cultures were cultured in rich bacterial growth media [14] for 24 h at 37 °C in an anaerobic chamber. We cultured *C. scindens* in growth media alone, *C. scindens* in media with cholic acid (100 µM), *C. scindens* in co-culture with *B. thetaiotaomicron* and taurocholic acid (100 µM), or *C. scindens* in co-culture with *B. vulgatus* and taurocholic acid (100 µM). (Cholic acid is the substrate required for DCA production. Images from cultures with cholic acid were considered cancer-associated states of *C. scindens*, while those from media-alone cultures represented health-associated states. *B. thetaiotaomicron* and *B. vulgatus* can produce cholic acid from taurocholic acid.) Overnight cultures with an OD_{600} of approximately 0.2 were used for slide preparation. Each culture (150 µl) was deposited onto a glass slide using the CytoSpin™ system, spun at 500 RPM for 5 min, and allowed to dry. The slides were briefly passed through a flame to fix the cells and then gram stained using a Fisher Scientific Gram Stain kit. After drying, coverslips were fixed with a drop of DEPEX and left overnight to harden. To acquire microscopy images for cell classification and segmentation, we utilized the TissueFAX microscopic imaging system in the Fred Hutchinson Cellular Imaging Core Facility to capture detailed images of *C. scindens* at 100× magnification. Images were captured at a z-stack depth of 0.2 µm, with a conversion rate of 0.064 µm per pixel.

2.2 Bile acid quantification in cultures

Supernatant from anaerobic co-cultures of *C. scindens* with either *B. thetaiotaomicron* or *B. vulgatus* were analyzed by liquid chromatography mass spectrometry using previously reported methods [15].

2.3 Image loading and preprocessing

For all analyses, bacterial images were loaded into memory and preprocessed using the 'preprocess_input' function from TensorFlow's 'ImageDataGenerator' class from the Keras API, which facilitates efficient handling of large datasets stored on disk. The 'ImageDataGenerator' was configured to normalize the pixel values of the images and split the dataset into training and test subsets (80:20 ratio). For training models, 25% of the training dataset was set aside and utilized for initial validation. Images were resized to 224×224 pixels to match the input dimensions required by the CNN model. Labels were assigned automatically based on the directory structure (e.g. presence/absence of cholic acid or different bacterial species names).

2.4 Whole-image classification with CNN models (Approach 1)

2.4.1 Training a CNN from scratch

2.4.1.1 Algorithm for training from scratch The training algorithm was implemented in a GPU-enabled environment using TensorFlow and Keras, which are Python libraries for building and training deep learning models. A custom CNN was developed from scratch, utilizing layers defined in the sequential API. The model was trained for 50 epochs, using the Adam optimizer for adaptive learning rate adjustment, and binary cross-entropy as the loss function to optimize the binary classification task. The training process involved iterating through batches of images and updating the model weights using backpropagation, guided by the gradients computed during each epoch.

2.4.1.2 CNN architecture and learning parameters The CNN model was composed of 6 layers: 3 convolutional layers interweaved with max-pooling layers, 1 flattening layer, and 2 fully connected (dense) layers. The first convolutional layer used 32 filters, the second used 64 filters, and the third used 128 filters, with all convolutional layers employing a kernel size of 3×3 and the ReLU activation function. The pooling layers used a 2×2 window for downsampling. The dense layers included one hidden layer with 128 neurons, activated by ReLU, and an output layer with a single neuron, activated by a sigmoid function for binary classification. The Adam optimizer was used with default parameters (learning rate = 0.001, beta_1 = 0.9, beta_2 = 0.999).

2.4.2 Pre-trained model (ResNet50)

2.4.2.1 Pre-trained ResNet50 integration The ResNet50 model, pre-trained on the ImageNet dataset, was imported from the Keras Applications library. The pre-trained model was loaded without its fully connected (top) layers. This allowed the use of ResNet50's convolutional base as a feature extractor. The weights of the ResNet50 convolutional layers were frozen to retain the pre-trained knowledge during the initial training phase. This approach leveraged ResNet50's powerful feature extraction capabilities without requiring additional computational resources to retrain its deep convolutional layers.

2.4.2.2 Custom classification model On top of the 'frozen' ResNet50 base model, a custom classification head was constructed comprised of (i) a flatten layer to transform the 4D tensor output of the ResNet50 base into a 1D tensor suitable for dense layers; (ii) a fully connected dense layer with 128 neurons and ReLU activation to introduce learnable parameters tailored to the specific classification task; and, (iii) a single-node dense layer with a sigmoid activation function for binary classification, outputting probability scores between 0 and 1 for the target classes.

2.4.2.3 Training and optimization The model was compiled using the Adam optimizer with default parameters (learning rate = 0.001), which adaptively adjusts learning rates during training. Binary cross-entropy was selected as the loss function, appropriate for binary classification tasks. The model was trained for 30 epochs. The training process utilized a GPU-enabled environment to accelerate computations, particularly beneficial for the large number of operations in the ResNet50 convolutional base.

2.4.3 Pre-trained model (ResNet50 and Digital Images of Bacterial Species)

The ResNet50 model was fine-tuned by incorporating the Digital Images of Bacterial Species (DIBaS) dataset, which comprises images of 33 diverse bacterial species [16], into the training pipeline, thereby allowing the model to adapt its feature extraction capabilities to bacterial morphologies. The resulting model was then used as the starting point for training on our specific dataset.

2.5 C. scindens cell detection and automated cell classification (Approach 2)

2.5.1 Adaptive thresholding

After image loading as above, initial preprocessing entailed detecting individual cells in each image using Adaptive Thresholding with a constant value C = 3. Unlike global thresholding, which applies a single threshold value to the entire image, adaptive thresholding calculates the threshold for smaller regions of the image, making it effective for images varying in brightness. In adaptive thresholding, the constant C is a parameter subtracted from the calculated mean or weighted mean intensity of the neighboring pixels, playing a crucial role in fine-tuning the thresholding process. This technique segmented the images by calculating the threshold for small regions, effectively isolating each cell based on local pixel intensity variations. The detected cells from all images in the dataset were then processed to classify them into two categories: *C. scindens* with cholic acid and *C. scindens* without cholic acid. Further image preprocessing was then performed as above. Pixel intensity values were normalized from TensorFlow's ResNet50 pre-trained model for consistent feature scaling.

2.5.2 Pre-trained ResNet50 integration

The ResNet50 model, pre-trained on the ImageNet dataset, was employed as a feature extractor. The pre-trained convolutional layers were used without modification, while the fully connected (top) layers were removed to adapt the model to the new classification task. This approach leveraged the pre-trained weights of ResNet50, which encode general-purpose visual features, providing a strong starting point for classifying cell images. The convolutional base of ResNet50 was frozen during initial training to retain the pre-trained knowledge and prevent overfitting, given the limited size of the dataset.

2.5.3 Custom classification model

A custom classification head was constructed and added on top of the frozen ResNet50 convolutional base. This consisted of a flattened layer to transform the high-dimensional feature maps output by ResNet50 into a 1D tensor. A fully connected dense layer with 128 neurons and ReLU activation to capture class-specific patterns, and an output layer with a single neuron and sigmoid activation, predicting the probability of the image belonging to either class.

2.5.4 Training and optimization

The model was compiled using the Adam optimizer, which adaptively adjusts the learning rate during training, ensuring stable and efficient convergence. Binary cross-entropy was used as the loss function, as the task involved binary classification. The training was conducted for 30 epochs, with a batch size of 16. To improve model generalization, data augmentation techniques such as random rotations were applied during training. The trained model was evaluated on the 20% test set, ensuring a robust assessment of its classification accuracy.

2.6 Automated image segmentation with nnU-Net (Approach 3)

2.6.1 Image segmentation

Images were annotated manually to create ground truth segmentation masks. These annotated images and masks were then loaded and preprocessed as above, with two additional preprocessing steps: z-score normalization to ensure

that image intensities followed a normal distribution and random cropping using 'DataLoader' to augment the dataset. The resulting image dataset served as the training dataset for nnU-Net [11], a self-configuring method for image segmentation.

2.6.2 Algorithm for training with diverse data

During training, the model learns to identify the boundaries and structures of bacterial cells by adjusting its parameters to minimize the difference between the predicted segmentations and the ground truth masks through the combined loss function of Dice and Cross-entropy loss. The training began with nnU-Net's default configuration, employing a 2D U-Net architecture for segmentation. In the 2D U-Net architecture, both the encoder and decoder consisted of five 3×3 convolutional blocks, each followed by instance normalization and leaky rectified linear unit (ReLU) activation. Skip connections were used to link encoder and decoder layers, preserving spatial information and enhancing localization. The training algorithm combined two loss functions: cross-entropy loss, which optimized pixel-wise classification, and Dice loss, which maximized overlap between predicted and ground truth masks. The optimization was performed using Stochastic Gradient Descent with an initial learning rate of 0.01. The training spanned 100 epochs (batch size of 8) and utilized five-fold cross-validation to ensure generalizability and robust performance to a new, unseen set of image data.

2.7 Gradient-weighted Class Activation Mapping

Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized to visualize class-discriminative regions in the input image. Grad-CAM computes the gradient of the class score y^c with respect to the feature maps A^k of a convolutional layer. The importance weights α_k^c for each feature map are calculated as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

where Z is the number of pixels in the feature map, and the indices i and j represent the spatial coordinates representing the height and width positions of each pixel.

The class activation map $L^c_{\text{Grad-CAM}}$ is then obtained by:

$$L^c_{\text{Grad - CAM}} = \text{ReLU}\left(\sum_k \alpha_k^c A^k \right)$$

This produces a heatmap that highlights the most influential regions from the input image for class prediction. Finally, Grad-CAM heatmaps were overlaid on the original microscopy images.

3 Results

Using light microscopy images of bacterial cultures with *C. scindens*, we evaluated deep learning imaging analysis of entire fields of view (whole-image classification), individual cell-based classification (cell detection and classification), and image segmentation (Fig. 1). We trained three distinct CNN architectures tailored for specific tasks and leveraging various methodologies to enhance classification accuracy and robustness.

3.1 Whole-image classification

For whole-image classification, three distinct classification methodologies were employed to assess the presence and metabolic states of *C. scindens* under various conditions. The first methodology involved developing a Convolutional Neural Network (CNN) from scratch. This methodology involves defining the network layers, activation functions, and learning parameters from the ground up. By starting with a clean slate, this model can learn features directly relevant to the dataset of interest, optimizing performance for the specific task of differentiating between different states of *C. scindens*. The second methodology utilized the pre-trained models ResNet and DenseNet, which are conceptually related neural network architectures differing in the density of connections between layers. Like

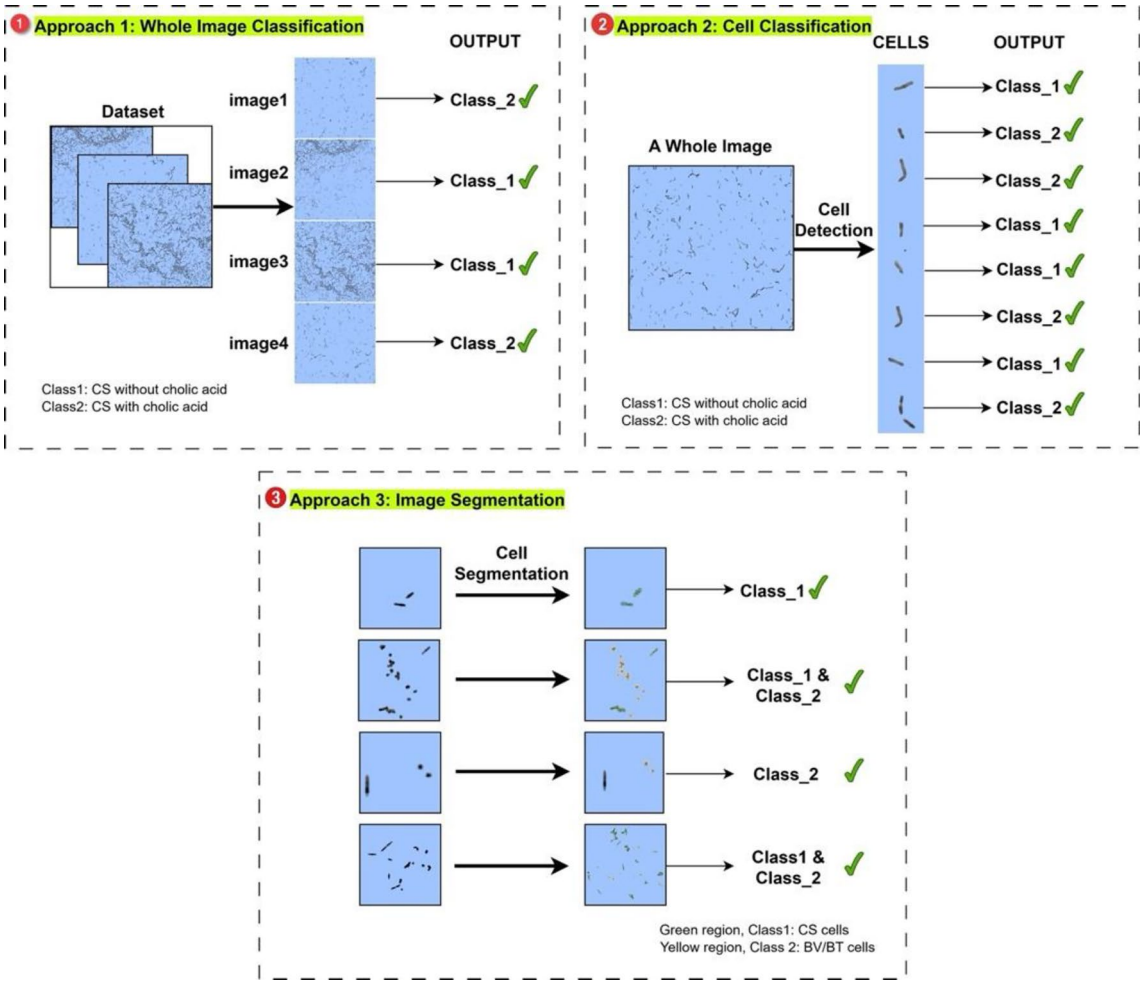

Fig. 1 This research comprises three distinct approaches. The first approach employs deep learning models to classify entire images based on overall characteristics. The second approach introduces cell detection prior to classification, identifying and classifying individual cells within each image. In the final approach, deep learning segmentation model is applied to do the pixel-to-pixel classification and isolate each cell, enabling precise identification of bacterial cells

the second methodology, our third methodology also utilized transfer learning, additionally incorporating diverse imaging data of 33 bacterial species from the Digital Images of Bacterial Species (DIBaS) repository into a ResNet framework. The performance of these models was evaluated using Accuracy, F1 score, Precision, and Recall. Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total instances (TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision indicates the proportion of true positive predictions among all positive predictions, reflecting the model's ability to avoid false positives:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall, or sensitivity, measures the proportion of true positive instances correctly identified by the model, indicating its ability to capture all relevant instances:

Table 1 Classification performance metrics of different methods

Model	Classification problem			
	(*C. scindens* with cholic acid) vs (*C. scindens* without cholic acid)		(*C. scindens* + *B. thetaiotaomicron*) vs (*C. scindens* + *B. vulgatus*)	
	Accuracy	F1 score	Accuracy	F1 score
CNN from scratch	0.34 ± 0.031	0.39 ± 0.035	0.31 ± 0.027	0.27 ± 0.029
Pre-trained with DenseNet	0.75 ± 0.0064	0.85 ± 0.006	0.43 ± 0.029	0.60 ± 0.022
Pre-trained with ResNet	0.85 ± 0.0015	0.91 ± 0.0023	0.81 ± 0.0031	0.72 ± 0.0038
Pre-trained with ResNet + DIBas dataset	0.89 ± 0.0056	0.90 ± 0.0037	0.86 ± 0.0042	0.83 ± 0.0049

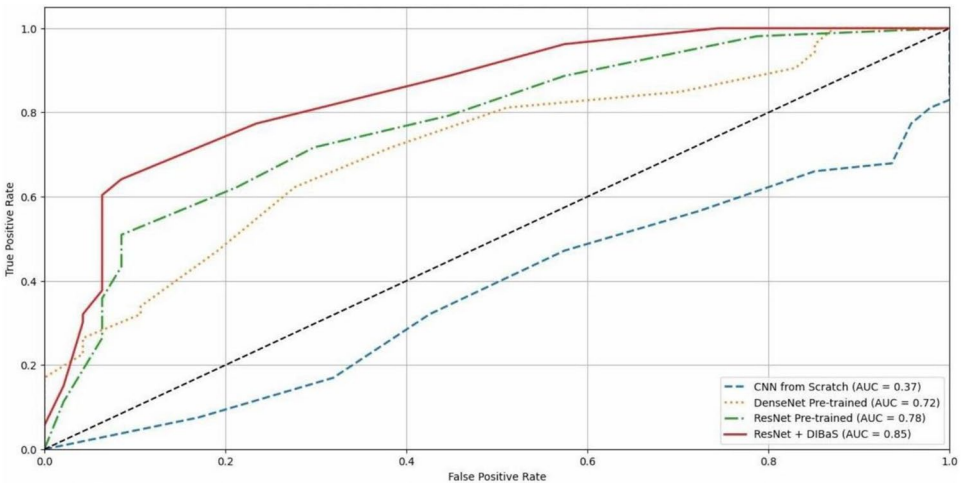

Fig. 2 ROC for distinguishing *C. scindens* with cholic acid from *C. scindens* without cholic acid. The x-axis, False Positive Rate (FPR), indicates the proportion of negative instances that are incorrectly classified as positive, and y-axis, True Positive Rate (TPR), shows the proportion of actual positive instances that are correctly classified by the model. A higher TPR indicates better sensitivity

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1 score is the harmonic mean of Precision and Recall, providing a single metric that balances both concerns, particularly useful when dealing with imbalanced datasets:

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

The classification performance metrics of the various methods we tested are summarized in Table 1, highlighting the comparative accuracies and predictive capabilities. When distinguishing *C. scindens* states with and without cholic acid, representing the DCA-producing cancer-associated state, the de novo CNN model achieved an accuracy of 0.34 and an F1 score of 0.39. In contrast, DenseNet showed a significant improvement with an accuracy of 0.75 and an F1 score of 0.85. ResNet further enhanced these metrics, achieving an accuracy of 0.85 and an F1 score of 0.91. Integrating the DIBaS dataset into the ResNet framework resulted in even better performance, with an accuracy of 0.89 and an F1 score of 0.90. In the task of discriminating between different co-culture scenarios, namely "*C. scindens* + *B. thetaiotaomicron*" and "*C. scindens* + *B. vulgatus*," the ResNet model augmented with the DIBaS dataset demonstrated superior performance. It achieved an accuracy of 0.86 and an F1 score of 0.83. These results indicate the superior performance of pre-trained models over the de novo CNN model, with DenseNet and ResNet models significantly improving accuracy and F1 scores. In addition, Fig. 2 illustrates the ROC curve for distinguishing *C. scindens* with cholic acid from *C. scindens* without cholic acid. The figure highlights that ResNet + DiBAS achieving the highest performance (AUC = 0.85).

The statistical analysis of classification performance using the Mann–Whitney U Test reveals significant differences in accuracy and F1 score among the different models (Tables S1, S2). The classification of *C. scindens* with cholic acid vs *C. scindens* without cholic acid reveals significant differences in accuracy and F1 score among the various models. The ResNet model pre-trained with the DIBaS dataset demonstrated significantly higher accuracy compared to the standard ResNet model (Mean ± Std: 0.89 ± 0.0056 vs. 0.85 ± 0.0015, p = 0.037), DenseNet (Mean ± Std: 0.89 ± 0.0056 vs. 0.75 ± 0.0064,

p = 0.0063), and the custom CNN (Mean ± Std: 0.89 ± 0.0056 vs. 0.34 ± 0.031, p = 0.00021). For the F1 score, the DIBaS-enhanced ResNet significantly outperformed DenseNet (Mean ± Std: 0.90 ± 0.0037 vs. 0.85 ± 0.006, p = 0.0090) and the custom CNN (Mean ± Std: 0.90 ± 0.0037 vs. 0.39 ± 0.035, p = 0.00020), while showing a non-significant trend towards improvement over the standard ResNet (Mean ± Std: 0.90 ± 0.0037 vs. 0.91 ± 0.0023, p = 0.167).

Similarly, accuracy and F1 score varied significantly between models classifying *C. scindens* + *B. thetaiotaomicron* vs. *C. scindens* + *B. vulgatus*. The DIBaS-enhanced ResNet model demonstrated significantly higher accuracy compared to the standard ResNet model (Mean ± Std: 0.86 ± 0.0042 vs. 0.81 ± 0.0031, p = 0.015), DenseNet (Mean ± Std: 0.86 ± 0.0042 vs. 0.43 ± 0.029, p = 0.00039), and the custom CNN (Mean ± Std: 0.86 ± 0.0042 vs. 0.31 ± 0.027, p = 0.00033). The standard ResNet also showed significantly higher accuracy than DenseNet (Mean ± Std: 0.81 ± 0.0031 vs. 0.43 ± 0.029, p = 0.00051) and the custom CNN (Mean ± Std: 0.81 ± 0.0031 vs. 0.31 ± 0.027, p = 0.00035). DenseNet's accuracy was significantly higher compared to the custom CNN (Mean ± Std: 0.43 ± 0.029 vs. 0.31 ± 0.027, p = 0.0048). For the F1 score, the ResNet model pre-trained with the DIBaS dataset achieved a significantly higher F1 score compared to the standard ResNet (Mean ± Std: 0.83 ± 0.0049 vs. 0.72 ± 0.0038, p = 0.0088), DenseNet (Mean ± Std: 0.83 ± 0.0049 vs. 0.60 ± 0.022, p = 0.00040), and the custom CNN (Mean ± Std: 0.83 ± 0.0049 vs. 0.27 ± 0.029, p = 0.00018).

The integration of the DIBaS dataset into the ResNet framework boosted the model's accuracy and F1 score, demonstrating the benefits of using a diverse and extensive dataset for training, even though the additional training data did not include images of *C. scindens* per se. The enhanced ResNet model's ability to effectively discriminate between different co-culture scenarios underscores its robustness and effectiveness in complex classification tasks. Such significant improvements can be attributed to several factors. The diversity of the DIBaS dataset, which includes images from 33 different bacterial species, enhances the model's ability to generalize across different bacterial morphologies. This diversity enables the ResNet model to learn more robust and transferable features, improving its performance on unseen bacterial species like *C. scindens*. Transfer learning from the DIBaS dataset allows the ResNet model to develop a deep understanding of various bacterial characteristics before fine-tuning on the target dataset, resulting in an enriched feature space that enhances classification capability. The significant boosts in accuracy and F1 score highlight the effectiveness of using diverse and comprehensive datasets for pretraining deep learning models.

3.2 Cell detection and classification

In this task, we employed a two-step approach for cell detection and classification from microscopy images of *C. scindens*, both with and without cholic acid. Initially, we utilized the OTSU and adaptive thresholding method for automated cell detection, leveraging its capability to threshold grayscale images effectively without requiring manual intervention. This step enabled us to identify individual cells within the microscopy images accurately. Subsequently, the detected cells were categorized into one of two groups: *C. scindens* cells cultured with cholic acid and those cultured without. To perform the classification task, we deployed a ResNet model pre-trained on a diverse dataset, optimizing it to differentiate between these two cell groups. ResNet's deep architecture and feature extraction capabilities facilitated precise classification by capturing nuanced differences in cell morphology and staining patterns associated with cholic acid.

For cell detection, we applied adaptive thresholding with different C constant values to achieve optimal performance using images of *C. scindens* without cholic acid (Fig. 3). Using a small C value results in overly sensitive detection, identifying noise along with many cells. In contrast, using C = 3 provides a more refined detection, balancing sensitivity and specificity by reducing noise while still identifying many cells. This visual assessment was corroborated by our statistical metrics (Table 2). For C = 1, the total detected cells are 8375, with a precision of 0.88, recall of 0.76, and an F1 score of 0.81. For C = 3, the total detected cells are 9252, with a precision of 0.91, recall of 0.86, and an F1 score of 0.89, indicating the best performance. For C = 4, the total detected cells drop to 7180, with a precision of 0.82, recall of 0.61, and an F1 score of 0.70. For C = 5, the total detected cells are 5500, with a precision of 0.73, recall of 0.41, and an F1 score of 0.52. For C = 8, the total detected cells further decrease to 2050, with a precision of 0.71, recall of 0.15, and an F1 score of 0.25, indicating the lowest performance. Overall, C = 3 provides the most accurate and well-separated cell labels, with the highest F1 score of 0.89. Therefore, we used C = 3 to create reliable pseudo-labeling, and we fed these adaptive thresholding-derived labels into our deep learning model for training.

Training, internal validation over 50 epochs (Figure S1) using fivefold cross-validation (Fig. 4), and external testing was performed (Table 3).

The training set for *C. scindens* without cholic acid comprised 20 images containing 35,990 total cells, of which 27,353 *C. scindens* cells were detected. The model successfully classified 23,685 cells as *C. scindens* without cholic acid and 3668 cells as *C. scindens* with cholic acid. In the internal validation phase (a supervised problem), of 13,806 total cells from 5

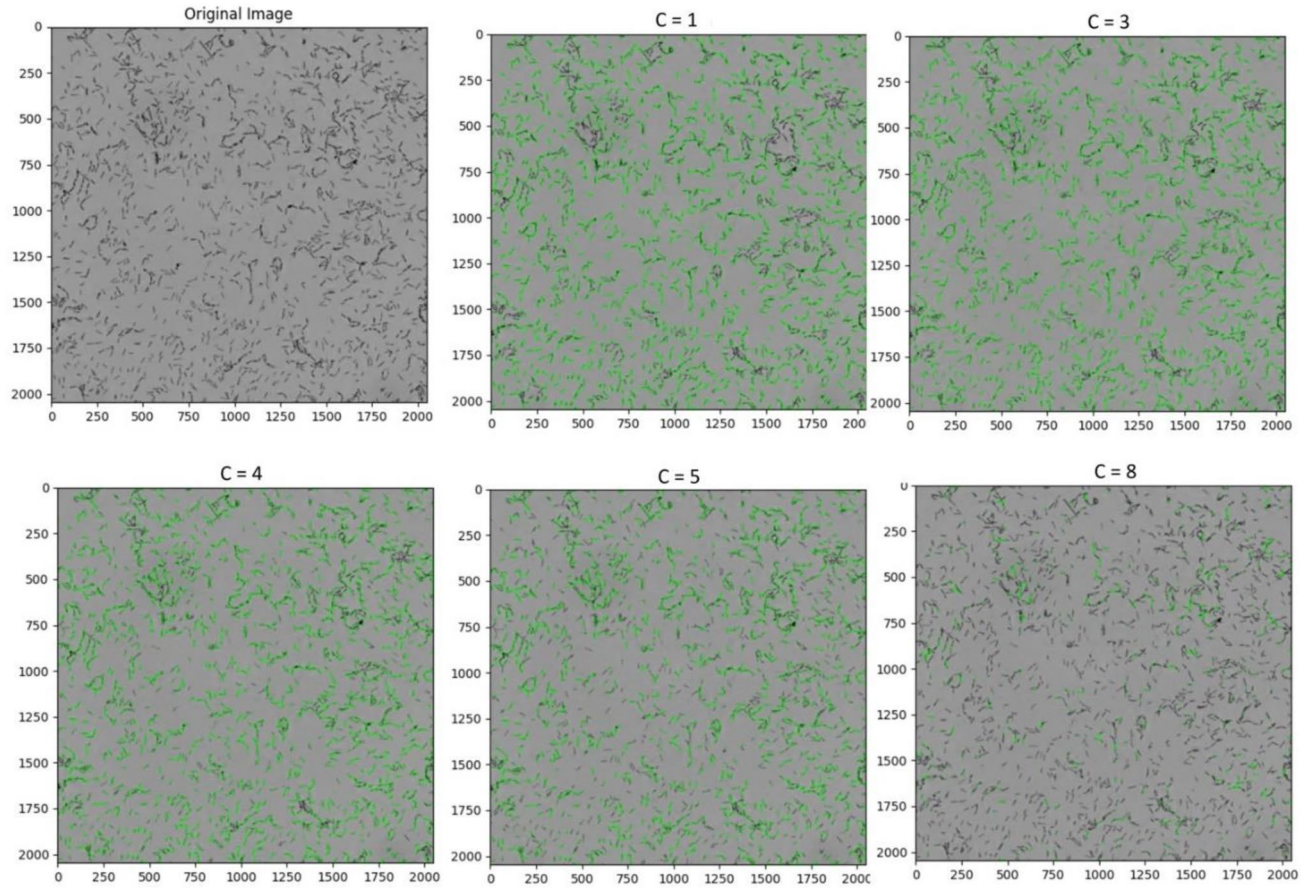

Fig. 3 Cell detection in *C. scindens* without cholic acid using adaptive thresholding with different values of the C constant. The original image is displayed in the top-left corner. The subsequent images demonstrate how varying the C constant (1, 3, 4, 5, and 8) affects the detection of cells, highlighted in green. C = 1. The detection appears to be very sensitive, identifying a large number of cells but possibly including noise and non-cell artifacts. C = 3: The detection is more refined, with a balance between sensitivity and specificity, reducing some noise while still identifying many cells. C = 4: The results are similar to C = 3, maintaining a good balance between cell detection and noise reduction. C = 5: The detection becomes slightly more conservative, possibly missing some cells but reducing noise further. C = 8: The detection is much more conservative, with fewer cells identified, suggesting that some true cells might be missing. Overall, the comparison indicates that lower C values (e.g., 1) result in more sensitive detections, while higher values (e.g., 8) lead to more conservative results, potentially missing some cells. Values like 3, 4, and 5 seem to provide a balanced detection, minimizing noise while capturing a significant number of cells

Table 2 Performance metrics of adaptive thresholding for cell detection at varying C constants

C constant	Total detected cells	TP	FP	FN (missed)	Precision	Recall	F1 score
C = 1	8375	7388	987	2337	0.88	0.76	0.81
C = 3	**9252**	**8455**	**797**	**1270**	**0.91**	**0.86**	**0.89**
C = 4	7180	5980	1290	3745	0.82	0.61	0.70
C = 5	5500	4025	1475	5700	0.73	0.41	0.52
C = 8	2050	1475	575	8250	0.71	0.15	0.25

Bold indiactes optimal precision, recall, and F1 score were seen with a C constant value of 3

We randomly selected 10 images from the dataset (total number of cells was 9725) for this analysis

images, 8622 *C. scindens* cells were detected. Of these, 6439 (74%) were correctly classified as *C. scindens* without cholic acid, while 2183 (26%) were misclassified as *C. scindens* with cholic acid. The external test set (also a supervised problem), consisting of 5 images with 13,559 total cells, yielded 8405 detected *C. scindens* cells. The model correctly classified 5622 (66%) as *C. scindens* without cholic acid and misclassified 2783 (34%) as *C. scindens* with cholic acid.

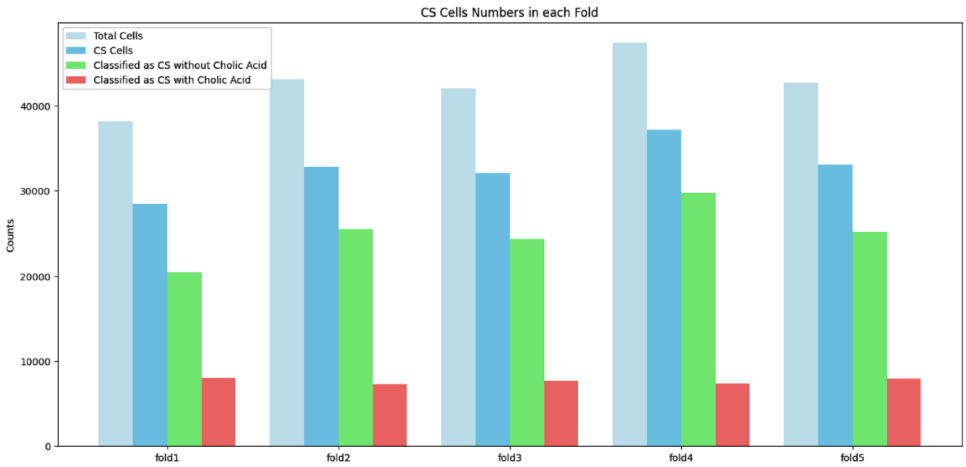

Fig. 4 The bar chart displays the total and detected C. scindens (CS) cells across five different folds, illustrating the performance of cell detection and classification in each fold. The total number of cells remained relatively constant across all folds. The number of detected CS cells also shows a consistent pattern, suggesting stable detection performance. The classification into cholic acid-producing and non-producing states is consistent across folds, with a larger proportion classified as non-producing. This consistency across folds indicates that the detection and classification models are performing uniformly, without significant variation in performance. The detection of CS cells is robust, capturing a significant portion of the total cells, and the classification results show a clear distinction between cholic acid-producing and non-producing states

Table 3 Training, internal validation, and external testing

State	Phase	# images	# total Cells	# CS cells detected	# CS cells classified as CS without cholic acid	# CS cells classified as CS with cholic acid
CS without cholic acid	Training	20	35,990	27,353	23,685	3668
	Validation	5	13,806	8622	6439 (74%)	2183 (26%)
	Test	5	13,559	8405	5622 (66%)	2783 (34%)
CS with cholic acid	Training	20	6731	5398	1383	4015
	Validation	5	2458	1776	540 (31%)	1236 (69%)
	Test	5	2353	1557	608 (38%)	949 (62%)

We performed fivefold cross-validation. Total cells, detected cells, and classified *C. scindens* (CS) cells represent the average across the five-folds

For *C. scindens* with cholic acid, the training set included 20 images containing 6731 total cells, with 5398 *C. scindens* cells detected. The model correctly classified 4015 cells as *C. scindens* with cholic acid in training. Internal validation accuracy was 69%, and external testing accuracy was 62%.

To identify the salient features used by the deep learning models for its predictions (i.e., to understand what the model is "seeing"), gradient-weighted Class Activation Maps (Grad-CAM) were deployed. Interestingly, the most class-discriminative areas were not intracellular (i.e., where DCA is produced), but rather, the Grad-CAM heatmaps (Fig. 5) highlighted the cell membrane and surrounding halo as critical regions for the ResNet model's classification of bacterial metabolic states.

We next tested the deep learning model in an unsupervised setting using unseen images of C. *scindens* in mixed bacterial populations. In co-cultures with one of two *Bacteroides* species, *B. thetaiotaomicron* or *B. vulgatus* (neither of which is capable of producing DCA), we observed significantly higher DCA production by *C. scindens* in co-culture with *B. vulgatus* (6.89e6 ± 2.55e5 ng/mL) compared to co-culture with *B. thetaiotaomicron* (1.61e5 ± 5.49e4 ng/mL) (Fig. 6). Therefore, testing images from these co-cultures offered an opportunity to test whether our model would accurately predict higher levels of DCA-producing *C. scindens* cells in the co-culture with *B. vulgatus*.

First, to establish ground truth, cells were annotated manually through visual inspection, which was feasible because of obvious morphological differences between *C. scindens* and the *Bacteroides* species. Then, training was

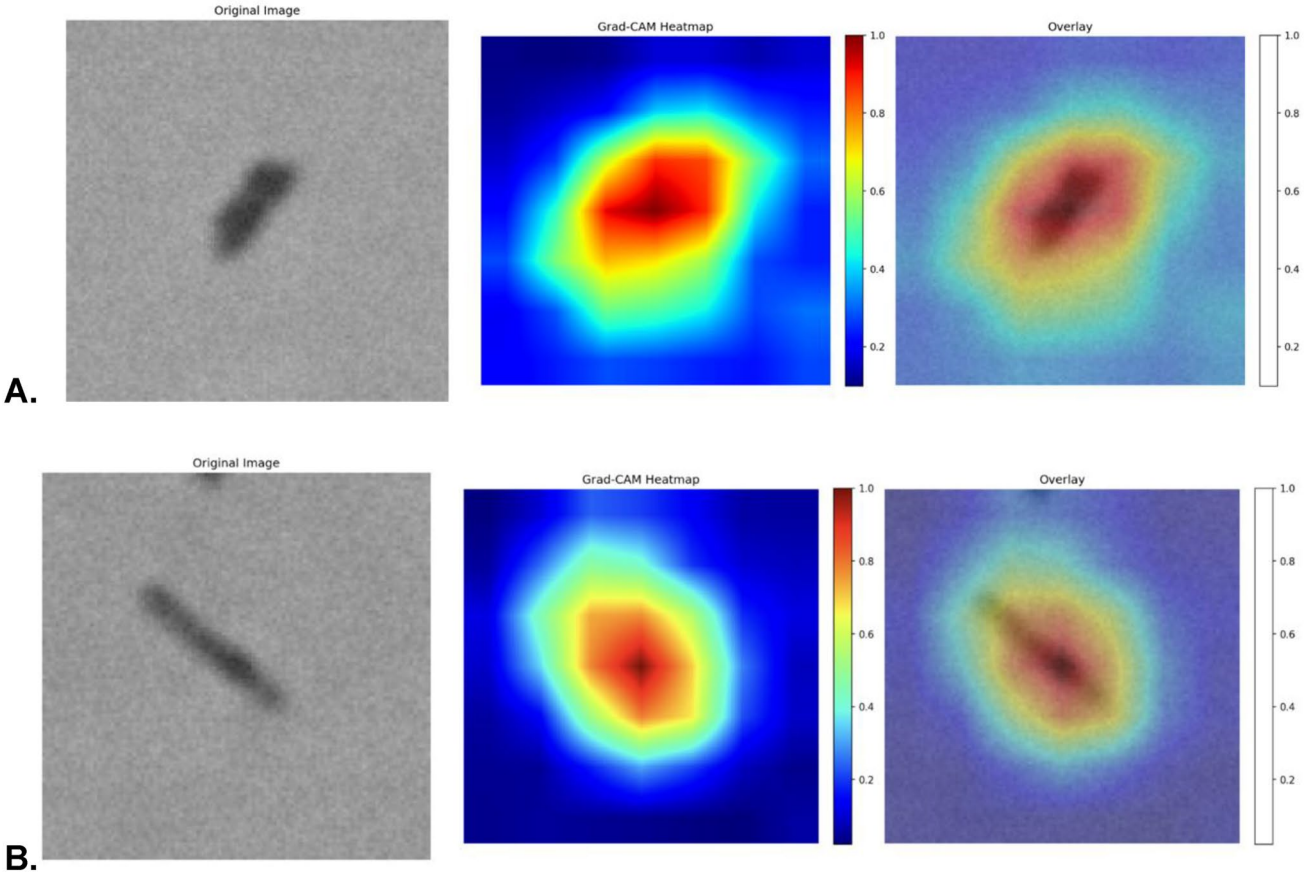

Fig. 5 *C. scindens* cultured with cholic acid (**A**) vs without cholic acid (**B**), visualized under high-resolution microscopy. The Gradient-weighted Class Activation Map overlay focuses on the cell membrane and the surrounding halo region indicating the importance of these regions for the ResNet model to accurately classify the correct bacterial metabolic state. Warmer colors (red and yellow) represent regions that greatly influenced the model, while cooler colors denote less important areas for classification

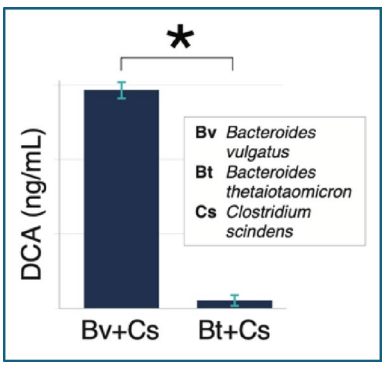

Fig. 6 Differential DCA production associated with the two co-cultures imaged. Despite comparable levels of C. scindens in both co-cultures, we observed differences in DCA production between these species.

performed across 50 epochs (Fig. 7). The observed trends demonstrate that the classifier effectively learned over time, improving its accuracy in distinguishing between the two *C. scindens* metabolic states.

In the testing phase, of the 1538 cells from 20 microscopy images of *C. scindens* + *B. thetaiotaomicron* that we analyzed (Table 4, Fig. 8), 1021 cells were *C. scindens* and 517 were *B. thetaiotaomicron*. The model detected 529 *C. scindens* cells, demonstrating a detection rate of 51.8% (529/1021). Among the detected *C. scindens* cells, 382 (72.2%) were classified as *C. scindens* without cholic acid, while 147 (27.8%) were classified as *C. scindens* with cholic acid.

For the *C. scindens* + *B. vulgatus* condition, 1958 total cells from 20 images were analyzed. This set contained 1372 *C. scindens* cells and 586 BV cells. The model detected 705 *C. scindens* cells, yielding a detection rate of 51.4% (705/1372).

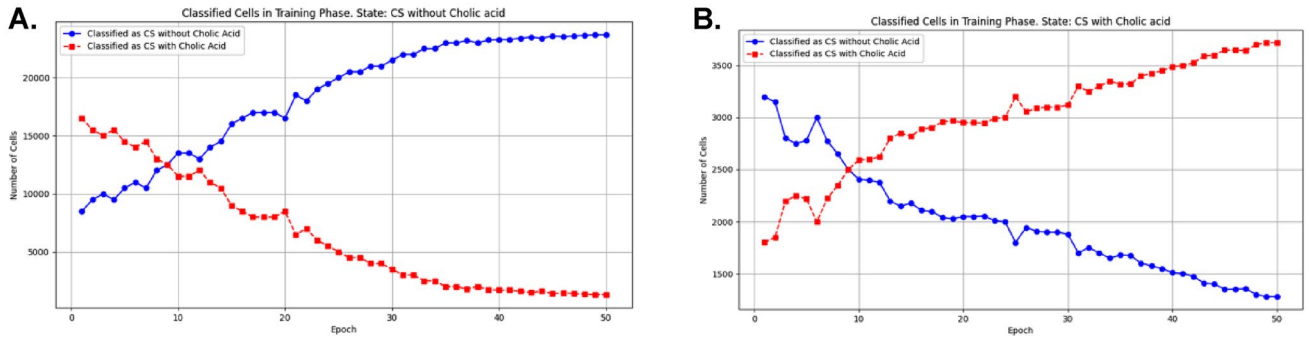

Fig. 7 Classification of C. scindens (CS) cells during the training phase, separated into states (**A**) without cholic acid and (**B**) with cholic acid. In each chart, the number of cells classified correctly increases steadily, stabilizing at a high count by the end of the training period, while the number of cells misclassified decreases consistently

Table 4 Detection and classification of bacterial cells in images from mixed bacterial populations during testing phase (CS: *C. scindens*, BV: *B. vulgatus*, BT: *B. thetaiotaomicron*)

State	# images	# CS	# BV/BT	# total cells	# Detected CS cells	# CS cells classified as CS without cholic acid	# CS cells classified as CS with cholic acid
CS + BT	20	1021	517	1538	529	382	147
CS + BV	20	1372	586	1958	705	309	396

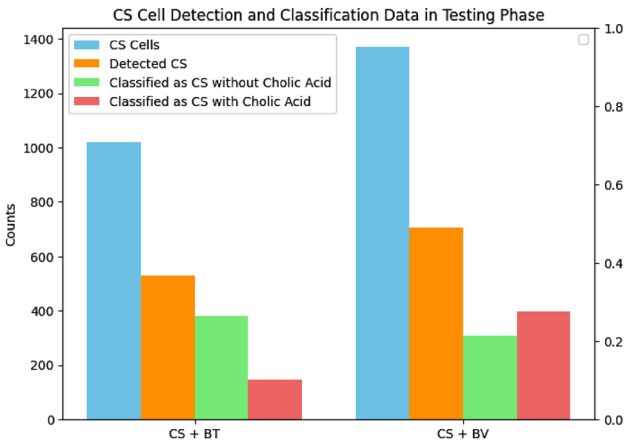

Fig. 8 Detection and classification results for *C. scindens* (CS) cells comparing two conditions: *C. scindens* with *B. thetaiotaomicron* (CS + BT) and *C. scindens* with *B. vulgatus* (CS + BV) (testing phase)

Of these detected *C. scindens* cells, 309 (43.8%) were classified as *C. scindens* without cholic acid, and 396 (56.2%) were classified as *C. scindens* with cholic acid.

Overall, while the *C. scindens* cell detection rate was modest, the majority of cell-level predictions regarding *C. scindens* metabolic states was correct in both co-cultures: in the presence of *B. thetaiotaomicron* (associated with lower DCA production), the model showed a stronger tendency to classify *C. scindens* as 'without cholic acid' (72.2%), whereas in the presence of *B. vulgatus* (associated with higher DCA production), the majority of *C. scindens* cells was classified as 'with cholic acid' (56.2%).

3.3 Automatic segmentation

We utilized nnU-Net, a specialized deep-learning framework designed for medical image segmentation, to delineate *C. scindens*, *B. vulgatus*, and *B. thetaiotaomicron* cells from microscopy images. The nnU-Net architecture (Fig. 9) follows a 2D U-Net structure, featuring encoding and decoding pathways with five convolutional blocks each. These blocks are augmented with instance normalization and employ leaky rectified linear unit activation functions, optimizing

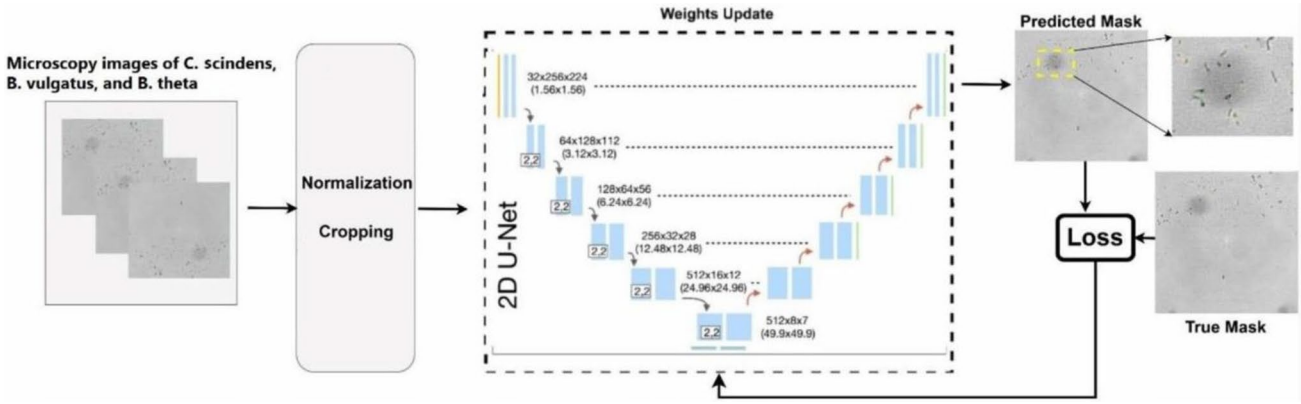

Fig. 9 The 2D nnU-Net's architecture, a neural network model designed for segmentation tasks. First, bacterial microscopy images undergo preprocessing steps of normalization and cropping. The 2D nnU-Net includes an encoder-decoder architecture with several layers, each comprising multiple convolutional operations followed by ReLU activations and pooling layers. The encoder compresses the spatial dimensions while capturing the context, and the decoder progressively reconstructs the spatial dimensions to produce a detailed segmentation mask. The network also utilizes skip connections that transfer feature maps from the encoder to the corresponding decoder layers, ensuring the preservation of fine-grained details

the network's ability to extract meaningful features from microscopy images. Image preprocessing steps included cropping and Z-Score normalization, ensuring standardized input data and improving segmentation accuracy. The nnU-Net framework was initialized with a hybrid loss function incorporating both Dice coefficient, which measures spatial overlap, and Cross-entropy, which evaluates pixel-wise classification accuracy, to enhance the model's capability in capturing intricate features of and boundaries between bacterial cells.

In nnU-Net training process, we utilized a dataset consisting of 50 images for training and internal validation, with fivefold cross-validation (dividing the training data into five subsets, where each subset is used as a validation set while the remaining subsets serve as the training set, rotating through each subset), and 25 additional images for external testing. The Dice coefficient, a commonly used metric in image segmentation tasks, quantifies spatial overlap between predicted and ground truth segmentations. It is defined as

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{5}$$

where TP (True Positive) represents correctly segmented pixels, FP (False Positive) denotes incorrectly segmented pixels, and FN (False Negative) indicates pixels in the ground truth that were not captured by the segmentation.

Model performance metrics during validation over 100 epochs of training are shown in Figure S2. During validation and testing, nnU-Net achieved a Dice coefficient of 98% and 87% for *C. scindens*, 85% and 76% for *B. vulgatus*, and 81% and 74% for *B. thetaiotaomicron*, respectively (Table 5). The high Dice coefficients obtained indicate that nnU-Net excels in accurately delineating bacterial cell boundaries, crucial for subsequent quantitative analyses and interpretations in microbiome research. Table 6 provides a comprehensive overview of the detection performance of the nn-Unet across two co-culture conditions (CS + BT and CS + BV). The results demonstrate high accuracy for detecting *C. scindens* (0.87–0.90) and comparable accuracy for *B. vulgatus* and *B. thetaiotaomicron* (both 0.80), showcasing the model's efficacy in differentiating bacterial cell states. These results validate nnU-Net's efficacy in automated bacterial cell segmentation from microscopy images (Fig. 10), offering potential advantages over traditional manual segmentation methods.

Table 5 Average segmentation performance from fivefold cross-validation

Segmentation metric	Test	C. scindens	BV	BT
Dice coefficient (pixel-by-pixel classification)	Validation (internal)	0.98	0.85	0.81
	Testing (external)	0.87	0.76	0.74

Table 6 Detection performance of nnU-Net

State	# images	# CS cells	# BV or BT cells	# total cells	# detected CS cells	# detected BV cells	# detected BT cells	Accuracy (CS)	Accuracy (BT)	Accuracy (BV)
CS + BT	10	560	235	795	488	–	190	0.87	0.80	–
CS + BV	15	930	265	1195	837	212	–	0.90	–	0.80

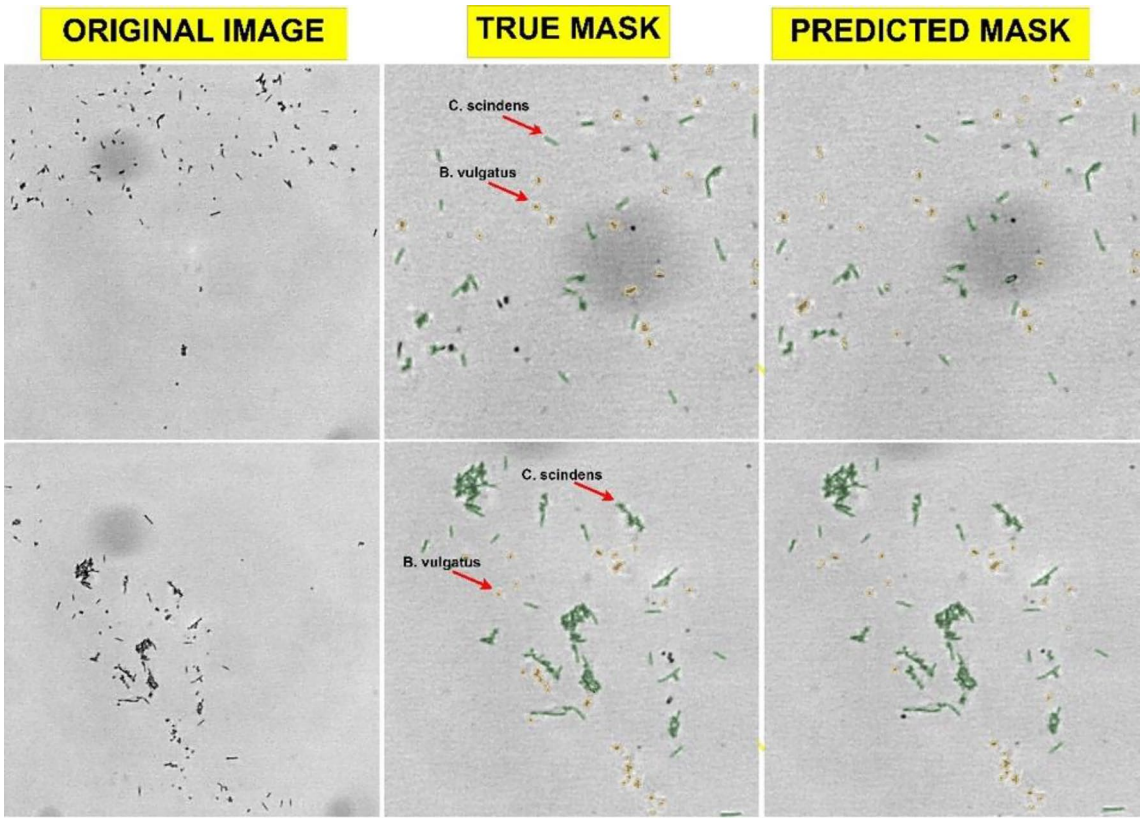

Fig. 10 Examples of segmentation results of *C. scindens* and *B. vulgatus* using nnU-Net. The original images show raw microscopy visuals of the bacteria. The true masks display manually annotated segmentations with *C. scindens* and *B. vulgatus* distinctly labeled. The predicted masks, generated by nnU-Net, closely match the true masks, indicating the model's high accuracy in segmenting and differentiating the two bacterial species

4 Discussion

Our study offers proof-of-principle for the potential of deep learning models in advancing clinically relevant microbiome imaging, potentially in the context of developing targeted diagnostic tools for CRC screening. We explored the capabilities of a deep learning model for whole-image classification, cell detection and classification, and automatic cell segmentation. The main contributions of this work encompass applications of deep learning in microbiome imaging analysis to advance diagnostic capabilities in identifying CRC-associated bacterial metabolic states.

We employed neural networks and pre-trained models to distinguish between CRC-associated and health-associated metabolic states of *C. scindens*. Importantly, we found that the external image data contributed to the model's ability to generalize better to bacterial species that were not present in the training data, suggesting generalizability in features critical to bacterial image analysis by deep learning models. Further, leveraging pretrained models and utilizing previously learned features and patterns reduces the amount of new data required for effective training, thereby saving time and resources.

Finally, using an established deep learning-based image segmentation framework, we were able to automate segmentation and highly accurate classification of microbial images, enabling analysis of bacterial morphology and spatial distribution. The performance of nnU-Net can be attributed to its sophisticated architecture and the use of a combined Dice + Cross-Entropy (CE) loss function, which balances the accuracy of segmentation and penalizes misclassifications.

Our current approach to bacterial cell detection and classification using thresholding methods represents a foundational step but comes with inherent limitations. The reliance on traditional image processing methods like OTSU restricts our ability to achieve optimal detection performance, particularly in complex microbial environments and mixed bacterial populations where cells may exhibit varying morphological and intensity characteristics. Additionally, the lack of advanced preprocessing steps and limited imaging resolution restricts the model's ability to capture fine details, which could improve detection accuracy.

To further improve these results, future studies should consider using higher-resolution imaging and advanced microscopes to capture more detailed images. Second, implementing preprocessing techniques to enhance image quality could aid in improving model performance. Third, expanding the dataset by capturing more images would also provide the model with a larger and more diverse set of samples, enhancing robustness and generalizability. Fourth, implementing data augmentation techniques could help by artificially increasing the variety of training images, thereby reducing overfitting. Fifth, exploring alternative loss functions or fine-tuning the current Dice + CE loss function to handle imbalanced data could further enhance segmentation accuracy.

Future research should explore advanced architectures, such as an enhanced nnU-Net that integrates Residual, Dense, and Inception Blocks, along with attention-based models like Transformers, to improve cell segmentation accuracy by capturing complex spatial and contextual relationships within bacterial images. These architectures could help the model adapt more effectively to mixed populations by focusing on nuanced differences in cell structure and appearance. Additionally, implementing an end-to-end framework for automatic cell detection and classification—using the YOLO deep learning model [17] for automatic cell detection followed by a ResNet classification model—could streamline the workflow and improve both speed and accuracy. Such an approach would allow the model to perform fine-grained identification of bacterial states within diverse microbial communities, addressing the challenges associated with mixed populations. By leveraging these advanced models, along with enhanced imaging techniques, we anticipate a significant enhancement in our ability to detect and classify bacterial cells with high precision in complex and heterogeneous samples.

Acknowledgements We thank Naisi Li, James Soetedjo, and other members of the Dey lab for their helpful suggestions in the preparation of this paper.

Author contributions N.M.M. and N.D. designed the study. N.M.M. and D.M.L. generated the data. M.O., S.R., S.S.M., and N.D. performed the analysis and wrote the manuscript. All authors reviewed the manuscript.

Funding This research was supported by National Institutes of Health K08 DK111941, National Institutes of Health U54 CA274374, and the Cellular Imaging Shared Resource RRID: SCR_022609 of the Fred Hutch/University of Washington/Seattle Children's Cancer Consortium (P30 CA015704).

Data availability Image datasets are available at EMBL-EBI BioImage Archive (accession number S-BIAD1671).

Code availability Code used are available at GitHub (https://github.com/DeyLab/Orouskhani_Rauniyar_et_al_2025).

Declarations

Competing interests The authors declare no competing interests.

References

1. World Health Organization. Colorectal cancer. WHO; 2023.
2. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. Gut. 2023;72(2):338–44.
3. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol. 2017;19(1):29–41.
4. Reddy BS, Narasawa T, Weisburger JH, Wynder EL. Promoting effect of sodium deoxycholate on colon adenocarcinomas in germfree rats. J Natl Cancer Inst. 1976;56:441–2.
5. Ochsenkühn T, et al. Colonic mucosal proliferation is related to serum deoxycholic acid levels. Cancer. 1999;85:1664–9.
6. Flynn C, Nakanishi M, Montrose D, Johnson M, Rosenberg DW. Promotional effect of deoxycholic acid in murine model resistant to colorectal cancer. Cancer Res. 2005;65:279–279.
7. Buffie CG, et al. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. Nature. 2015;517:205–8.
8. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.
9. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–8.
10. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18. Springer International Publishing; 2015. p. 234–41).
11. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203–11.
12. Pham B, Gaonkar B, Whitehead W, Moran S, Dai Q, Macyszyn L, Edgerton VR. Cell counting and segmentation of immunohistochemical images in the spinal cord: comparing deep learning and traditional approaches. In: 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2018. p. 842–5.
13. Ayanzadeh A, Yağar HO, Özuysal ÖY, Okvur DP, Töreyin BU, Ünay D, Önal S. Cell segmentation of 2d phase-contrast microscopy images with deep learning method. In: 2019 medical technologies congress (TIPTEKNO). IEEE; 2019. p. 1–4.
14. Goodman AL, Kallstrom G, Faith JJ, Reyes A, Moore A, Dantas G, Gordon JI. Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. Proc Natl Acad Sci USA. 2011;108:6252–7. https://doi.org/10.1073/pnas.1102938108.
15. Li N, Koester ST, Lachance DM, Dutta M, Cui JY, Dey N. Microbiome-encoded bile acid metabolism modulates colonic transit times. iScience. 2021;24: 102508. https://doi.org/10.1016/j.isci.2021.102508.
16. Zieliński B, Plichta A, Misztal K, Spurek P, Brzychczy-Włoch M, Ochońska D. Deep learning approach to bacterial colony classification. PLoS ONE. 2017;12(9): e0184554.
17. Redmon J. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.