Background

Wireless capsule endoscopy (WCE) uses a pill-shaped camera in order to view and explore the small intestine. A WCE capsule, or “pill-shaped” camera, is approximately 11 x 27 millimeters with no external wiring system and enables the patient’s body to process it naturally [1].

Over the course of eight hours, the camera travels through the alimentary canal propelled by normal muscular movements, capturing and transmitting images at a rate of two frames per second for a total of 50,000 images [1]. The capsule is then excreted and flushed down the toilet, and the external recording device is given to the physician for analysis. Use of this capsule is considered a minimally invasive method compared to upper endoscopy, push enteroscopy, or colonoscopy. Patients can access the small intestine through the natural canal of the digestive system. No anesthesia or anesthetic routine during WCE procedure [2]. The procedure may go on normal routines during WCE procedure [3].

1. Cost of WCE procedure: $1500 [4].
2. Less dietary & medication restrictions for patients.
3. The convoluted convolutional neural network architecture achieves an average accuracy, specificity, and sensitivity all over 94.68%.
4. This threshold is decided by looking at previous methods, none of the mentioned ones in the above section have achieved an accuracy, specificity, and sensitivity all over 94.00%.

Criteria for Success

1. The convolutional neural network code is executed on a 2022 MacBook Pro with 4 GB of RAM. Because most medical facilities will have more efficient computers with larger memory, this requirement demonstrates efficiency and practicability.

Constrains and Assumptions

1. For this project, one hundred images are available for use, which is a smaller dataset than used in others’ previous research. While the accuracy, specificity, and sensitivity of programs from previous research are considered in the discussion of results, it is reasonable to assume that a reliable conclusion can only be drawn by testing other methods with the same dataset, as opposed to simply comparing accuracies produced in previous research projects.

2. The dataset made available for this project is balanced evenly, with fifty bleeding and fifty non-bleeding images; it does not simulate an actual clinical setting. Because of the limited number of images, a bleeding-to-non-bleeding ratio closer to an actual WCE video would have resulted in an insufficient amount of data for training with bleeding images. However, it is reasonable to assume that the general methodology of this project can be applied to an unbalanced dataset with a modified architecture.

3. Convolutional neural networks require a unique architecture whenever the size of the data set changes or the training set changes. Determination of that architecture can only be achieved through trial and error. Three different training sets are tested with the same architecture, with the architecture optimized for each size. This is sufficient for drawing conclusions because the same training set was not tested on other machine learning programs to determine whether the convolutional neural network approach is superior or not.

General Methodology

The convolutional neural network is created, tested, and evaluated in three main parts.

1. Preprocessing and Creation of Dataset

The raw data of 100 images are converted to grayscale, sorted into the two directories based on the presence or absence of blood, and compressed into a usable dataset for the convolutional neural network.

2. Design, Training, and Testing of CNN

The optimized architectures for the three different testing sizes are developed in the machine learning library and tested, and the accuracy, sensitivity, and specificity of each architecture is retrieved by the program.

3. Comparisons of CNN Results to Other Forms of Machine Learning

The results of the convolutional neural network are compared to the results of other forms of machine learning, run with the same dataset as the CNN. Results include accuracy, sensitivity, and specificity of the method.

Preprocessing and Description of Dataset

The dataset consists of 100 wireless capsule endoscopy images (50 bleeding, 50 non-bleeding) sampled from twelve recordings of procedures on multiple patients. Each image in the dataset has been pre-labeled by a physician as bleeding or non-bleeding so the program can account for the annual report. After preprocessing, the images are grayscale, sorted into either the “bleeding” or “non-bleeding” directory, with each directory containing a sub-directory of bleeding.

Architectures and neural network architectures, such as CNN, are used to find patterns in the data and make predictions. These architectures are used to classify images, including medical images, into diagnostic categories. CNNs are particularly useful for image classification tasks, such as medical image analysis, because they can automatically learn the hierarchical features of the images.

CNN Concept

The CNN Concept involves several key ideas:

1. An input layer—a full image (256 x 256 grayscale image).
2. A convolution layer with a set of filters of a certain dimension. If a filter of certain dimension is passed over entire image, multiplying grayscale value of each pixel by a corresponding filter weight. The output of this process is called a feature map, and each feature map will contain the information that was relevant to the particular filter used.
3. A pooling layer to allow outputting, max pooling is applied to each feature map, max pooling takes only the maximum grayscale value within a certain region, creating a condensed version of the complete information in that region.
4. Fully connected layer(s) connects every neuron in the previous layer with every neuron in the next layer. This allows the CNN to learn more complex features.

2. An output layer comprised of two neurons that classify an image as bleeding or non-bleeding based on percent likelihood.

Using backpropagation, weights & biases are adjusted to minimize the error before the next image is processed.

Architectural Design, Training, and Testing

The network was tested when 90%, 80%, and 70% of the data is used for training, and the remaining for testing. This helps determine whether the program can produce a high accuracy (>90%) when there is less data available for training.

Architecture 1 (90% training set)

1. Significant changes to the original architecture is not needed due to the larger training set size.

Architecture 2 (80% training set)

1. 100% accuracy can be reached through trial and error by hand.

Architecture 3 (70% training set)

1. 100% accuracy is not able to be reached by empirical change.

Method of AlexNet and MuMMER is used to determine best architectures

Conclusions and Future Research

Two of the three CNN architectures tested have an accuracy, sensitivity, and specificity of 100%. It can be concluded that the convolutional neural network approach is the most effective out of all classification methods tested for bleeding detection. The other machine learning, methods of accuracy 93.33%, and the other machine learning methods have accuracies ranging from 40.00% to 46.70%. The lower accuracy of the CNNs can be attributed to the traversal of images pixel-by-pixel, as opposed to the CNN’s feature detection with filters. Because many CNN methods use various classification algorithms that create an optimized graph based on performing operations on training data, the estimates are inaccurate and prone to overfitting, while also having a large discrepancy between sensitivity and specificity.

Evaluation Metrics

Table 3: Comparison of architectures for various training sizes and the original model provided by TensorFlow

Comparison and Evaluation:

The results of the convolutional neural network approach are compared to other various commonly used forms of machine learning, some of which were discussed in the Introduction and Background section, in addition to a multilayer perceptron (MLP) network, which is a neural network that is not deep. The machine learning classification methods were all imported from Python’s scikit-learn library.

The learning curve of each architecture is also plotted. These curves demonstrate the process of learning for the program by plotting the number of errors or bits, for each successive epoch trained. An initial significant decline, followed by a general downward trend, is shown in all three curves, until the loss is near zero, as seen in Figure 6.

90% Architecture Learning Curve

90% Architecture Learning Curve

70% Architecture Learning Curve

80% Architecture Learning Curve

Evaluation Metrics

Table 3: The accuracy, sensitivity, and specificity of the CNN for the three training sizes

90% Architecture Learning Curve

80% Architecture Learning Curve

70% Architecture Learning Curve

Evaluation Metrics

Table 3: The accuracy, sensitivity, and specificity of the CNN for the three training sizes

90% Architecture Learning Curve

80% Architecture Learning Curve

70% Architecture Learning Curve

Evaluation Metrics

Table 3: The accuracy, sensitivity, and specificity of the CNN for the three training sizes

90% Architecture Learning Curve

80% Architecture Learning Curve

70% Architecture Learning Curve

Evaluation Metrics

Table 3: The accuracy, sensitivity, and specificity of the CNN for the three training sizes

90% Architecture Learning Curve

80% Architecture Learning Curve

70% Architecture Learning Curve

Conclusion

The performance of the CNN in detecting bleeding with a sensitivity and specificity of 100% suggests that the convolutional neural network approach is superior to other traditional machine learning methods such as QDA, SVM, and MLP. The results indicate that the CNN is a promising method for detecting bleeding in wireless capsule endoscopy images, and further research is needed to validate these findings in a clinical setting. The future work includes further investigation into the use of convolutional neural networks for medical image analysis and their potential applications in other fields such as pathology and genomics.