Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2024) : –
© TÜBİTAK
doi:

DPafy-GCaps: Denoising patch-and-amplify Gabor capsule network for the recognition of gastrointestinal diseases

Henrietta Adjei POKUAA[1,2*], **Adebayo Felix ADEKOYA**[3],
Benjamin Asubam WEYORI[4], **Owusu NYARKO-BOATENG**[1,5]

[1]Department of Computer Science and Informatics, School of Sciences,
University of Energy and Natural Resources, Sunyani, Ghana
[2]Department of Computer Science, Faculty of Applied Science and Technology,
Sunyani Technical University, Sunyani, Ghana
[3]Department of Computing and Information Sciences, Faculty of Computing Engineering and Mathematical Science,
Catholic University of Ghana, Sunyani, Ghana
[4]Department of Computer and Electrical Engineering, School of Engineering,
University of Energy and Natural Resources, Sunyani, Ghana
[5]School of Information Technology, University of Cincinnati, Ohio, USA

Received: 20.10.2023 • **Accepted/Published Online:** 27.04.2024 • **Final Version:** ..2024

Abstract: Deep learning (DL) models have performed tremendously well in image classification. This good performance can be attributed to the availability of massive data in most domains. However, some domains are known to have few datasets, especially the health sector. This makes it difficult to develop domain-specific high-performing DL algorithms for these fields. The field of health is critical and requires accurate detection of diseases. In the United States Gastrointestinal diseases are prevalent and affect 60 to 70 million people. Ulcerative colitis, polyps, and esophagitis are some gastrointestinal diseases. Colorectal polyps is the third most diagnosed malignancy in the world. This work, therefore proposes a variant Capsule Network (CapsNet) termed Denoising patch-and-amplify Gabor Capsule Network for detecting gastrointestinal tract diseases. The proposed model leverages the advantages of patching, feature amplification, and Gabor filters to enable the learning of meaningful information (from small datasets) needed to achieve intelligence in domain-specific models. During the experimental procedure, the proposed model achieved 95.10%, 85.50%, and 96.80% recognition accuracies on Fashion-MNIST, Cifar10, and Kvasir-v2 datasets. The proposed model performs comparably well with CapsNet models in the domain of gastrointestinal recognition.

Key words: Capsule Network, gastrointestinal tract, Gabor filters, deep learning

1. Introduction

In the USA, 60 to 70 million people are diagnosed with gastrointestinal diseases every year [1]. The endoscopic examination is performed by a trained gastroenterologist. This approach requires a careful analysis. The susceptible nature of gastrointestinal diseases, and the need to achieve complete certainty on the diagnoses of such diseases, has led to the requirement and utilization of intelligent decision-making algorithms. Artificial Intelligence algorithms have excelled in this decision-making terrain. For instance, deep learning algorithms

*Correspondence: henrietta.opokuaa@stu.edu.gh

[1] NIH (2020). Digestive Diseases Statistics for the United States [online]. Website https://www.niddk.nih.gov/health-information/health-statistics/digestive-diseases [accessed 30 October 2023]

such as Convolutional Neural Networks (CNNs) have been applied in health image recognition, detection, and classification tasks. However, to achieve close to human recognition accuracy CNNs require large training data [1, 2]. A large dataset is an issue in the domain of health [3, 4]. This limitation, therefore, requires the application of diverse data augmentation techniques to the existing small health datasets to increase the size of the dataset. However, data augmentation approaches are time-consuming and laborious, and annotation requires experts.

To address these drawbacks of CNNs, Capsule Network [5] was proposed. Capsule Networks are embedded with data variation abilities. In that, CapsNet learns to infer pose parameters from images. This, therefore, makes the data augmentation procedure an optional task. Furthermore, CapsNet does not require large datasets and is convenient for the existing small health datasets. Though CapsNet addresses the drawbacks of CNNs, they also have some limitations attached to them. The CapsNet algorithm performs poorly on complex images, thus, images with varied backgrounds. This can be attributed to their weak encoder network [6] that extracts every feature present in an image and has a weak selection of relevant features.

This work, therefore, proposes a modified Capsule Network algorithm with a robust encoder network to enable them to generalize well on complex health images and assist in diagnosing gastrointestinal diseases. A technique for feature enhancement termed Denoising Patch-and-Amplify is proposed. This technique involves the amplification of features, removal of noise, and the splitting of images into two halves. The algorithm aims to make relevant features dominant, remove noise, and make the proposed model to be more focused on the amplified features. Gabor filters are embedded into the network to assist the model in selecting specific frequency content in the images. To encourage deployment and contribute to the domain of explainable Capsule Networks (XCaps), the outputs of the model are explored via diverse visualizations.

The paper is sectioned as follows; Section 2 presents related works in the domain of Gastrointestinal detection via Convolutional Neural Networks and Capsule Networks. Section 3 presents the proposed feature enhancement technique, the proposed model, and details about the experimental procedure. Section 4 delves into the results and analysis. The article is concluded in section 5.

2. Related work

The popular deep learning methods for detecting anomalies in the digestive tract are CNNs. For instance, Sharif et al.[7] proposed a CNNs model based on a hybrid contrast stretching and feature fusion technique to detect gastrointestinal diseases. The Hybrid contrast stretching methods enhances the global and local contrast of the gastrointestinal images. The diseased parts are then segmented and the geometric features are computed. Redundancy in features is handled by the fusion of the geometric features via Euclidean Fisher Vector (EFV). The proposed CNNs model when applied to a private database achieved a classification accuracy of 99.42%. Jia and Meng [8] proposed a smaller-scale CNN model trained with handcrafted features extracted from Gastrointestinal images. The CNNs model achieved a recall value of 91%, a precision value of 94.79%, and an F1 score of 92.85%. Saban and Umut [9] proposed a Long Short Term Memory (LSTM) based CNNs model for classifying gastrointestinal tract infections. Several model architectures (CNNs + Artificial Neural Network and CNNs + support vector machine) were explored with 7500 data samples. However, the architecture combination of LSTM and CNNs attained the highest recognition accuracy of 97.90%. Currently, there is limited work on Capsule Network models in the domain of Gastrointestinal disease detection. Afriyie et al.[10] proposed a denoising Capsule Network to recognize gastrointestinal diseases. The fast denoising for multicolored method from OpenCV was implemented and infused in the CapsNet. The proposed CapsNet model was trained on five classes of the Kvasir-V2 dataset achieving an overall recognition accuracy of 94.16%. Sarsengeldin et al.[11] proposed a hybrid network comprising a VGG and a Capsule network. The feature extractor of the VGG network was connected to the primary capsule layer. This proposed VGG+CapsNet model was trained on five classes of the Kvasir dataset achieving 87% recognition accuracy. Ayidzoe et al.[12] proposed a Gabor Capsule Network model with a feature enhancement technique termed custom preprocessing blocks. This feature enhancement technique was implemented as a layer and infused in the encoder network of CapsNets. The proposed model achieved a recognition accuracy of 91.50% on the Kvasir dataset. The above works are state-of-the-art works in the domain of Gastrointestinal disease detection using Capsule Networks.

3. Materials and methods

This section presents the proposed feature enhancement technique, proposed model, dataset description and preprocessing, and the experimental settings adopted in the training and testing of the proposed model.

3.1. Proposed feature enhancement techniques

The proposed feature enhancement technique is termed denoising patch-and-amplify (See Algorithm 1). The algorithm is made up of three key tasks that are applied to both the training and test datasets. The first part involves amplifying the features. The second part involves noise removal. The third part involves slicing the images into two halves. The algorithm aims to efficiently remove noise from and amplify the extracted features which is achieved by applying a fast-denoising technique on the image and creating a shadow of the version of the image or feature map. This reduces the influence of irrelevant features and intensifies the strength of the relevant features in the network. Figure 1 presents the results of the denoising patch-and-amplify task and shows a reduction of irrelevant features and, the expansion and intensification of relevant features evident in the pixel plots. The patching aims to enhance the focus of the model towards features with detailed information and improve the prediction output of the model. The denoising patch-and-amplify algorithm is implemented as a layer and infused in the Capsule Network architecture. In Algorithm 1, $L^l_{m,n}$ refers to the training dataset, $Z^l_{m,n}$ refers to the test dataset, p is a variable used to hold the dimension of images in the training dataset, b refers to the slicing weight, $X^i_{m,n}$ refers to amplified feature maps, $J^i_{m,n}$ refers to a sliced part of the feature map and $Q^i_{m,n}$ refers to another sliced part of the feature map. These parameters do not add to the total number of parameters.

Algorithm 1 Algorithm 1. Denoising patch-and-amplify algorithm.

```
$L^l_{m,n} = \{l^0_{m,n}, l^1_{m,n}, ..., l^{1-i}_{m,n}\} \triangleleft$ Train images
$Z^l_{m,n} = \{z^0_{m,n}, z^1_{m,n}, ..., z^{1-i}_{m,n}\} \triangleleft$ Test images
$p = L^l_{m,n}.shape[1] \triangleleft$ Getting the shape of images in $L^l_{m,n}$
$b = \frac{p}{2} \triangleleft$ Slicing weight
To preprocess features, $\forall L^l_{m,n}, Z^l_{m,n}$
for $i$ in $L^l_{m,n}, Z^l_{m,n}$
    $X^i_{m,n} \implies L^i_{m,n} - \alpha L^i_{m,n} \triangleleft$ where $\alpha = [2, 4]$, amplify
    $X^i_{m,n} \implies$ fastNIMeansDenoisingColored($X^i_{m,n}$)$\triangleleft$ Denoising
    $J^i_{m,n} \implies X^i_{m,n}[:, : b, : b, :] \triangleleft$ Patching
    $Q^i_{m,n} \implies X^i_{m,n}[:, b :, b :, :] \triangleleft$ Patching
    return $J^i_{m,n}, Q^i_{m,n}$
```

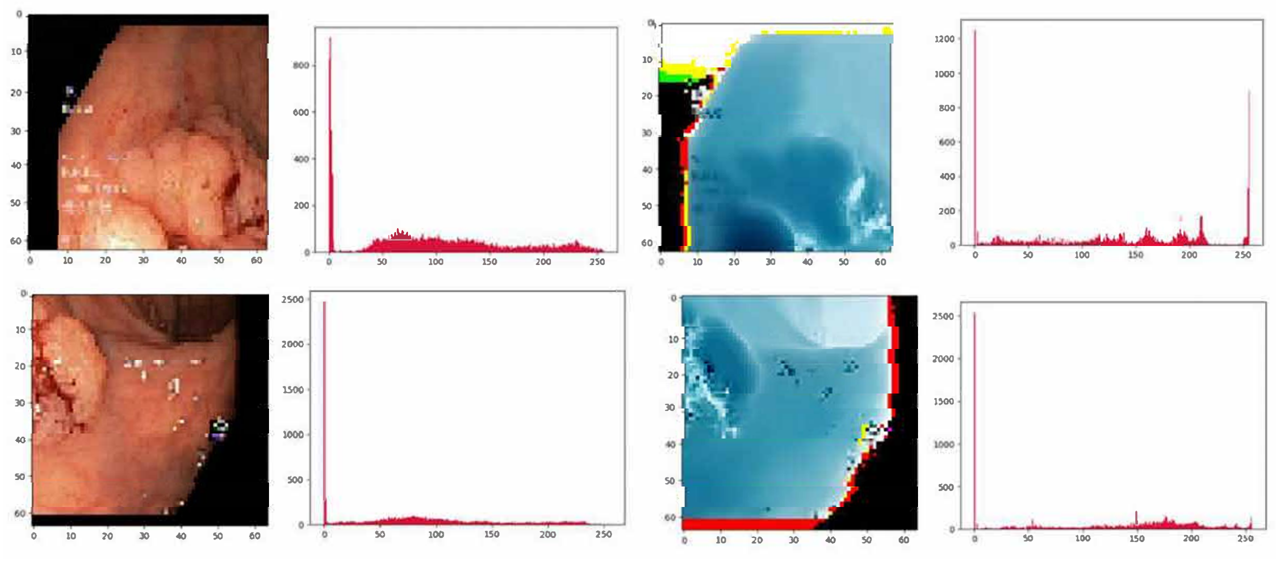


Figure 1. Visualization of the pixel intensity of the original image and a modified image (using the patch-and-amplify algorithm).

3.2. Proposed model

The proposed model is presented in Figure 2. It comprises a dual-lane capsule network. The first half of the input image passes through the first lane of the encoder network. The second half of the image passes through the second half of the network. Each lane consists of one denoising patch-and-amplify layer, two denoising amplify layers, one convolutional layer, and two Gabor convolutional layers. The first, and second convolutional layer comprises 64 kernels of size 3x3. The first, second, third, and fourth Gabor layers comprise 64, 128, and 256 kernels of size 3x3, respectively. The primary capsule has 16 component capsules each with dimension eight. The class capsule is made up of five capsules consistent with the number of classes in the dataset. The decoder network involves three fully connected layers of 512, 256, and 3072 neurons.

3.3. Feature enhancement, extraction, and processing

Every neural network model comprises an input layer, a hidden layer, and an output layer. The input layer provides a path for the images to be submitted to the hidden layer. The hidden layer is responsible for feature extraction, preprocessing, and classification. The output layer presents the prediction of the neural network. The encoder network comprises of all three layers. Firstly, an image is passed through the input layer to the denoising patch-and-amplify layer where the batch of images are enhanced via amplification. This amplification technique is done by applying a multiplicative effect on the images. For this work, the weight of 2 was used to amplify the features. The next step involves the removal of noise from the amplified features via the FastNIMeansDenoisingColored (eqn 1). FastNIMeansDenoisingColored is a denoising technique in OpenCV that is used to remove noise in colored images. It was adopted in this work because the images used for the experiments are colored images.

$$NLu(p) = \frac{1}{C(m)} \int_{\Omega} e - \frac{(G_b * |p(m+.) - p(n+.)|^2(0)}{i^2} p(y)dy \tag{1}$$

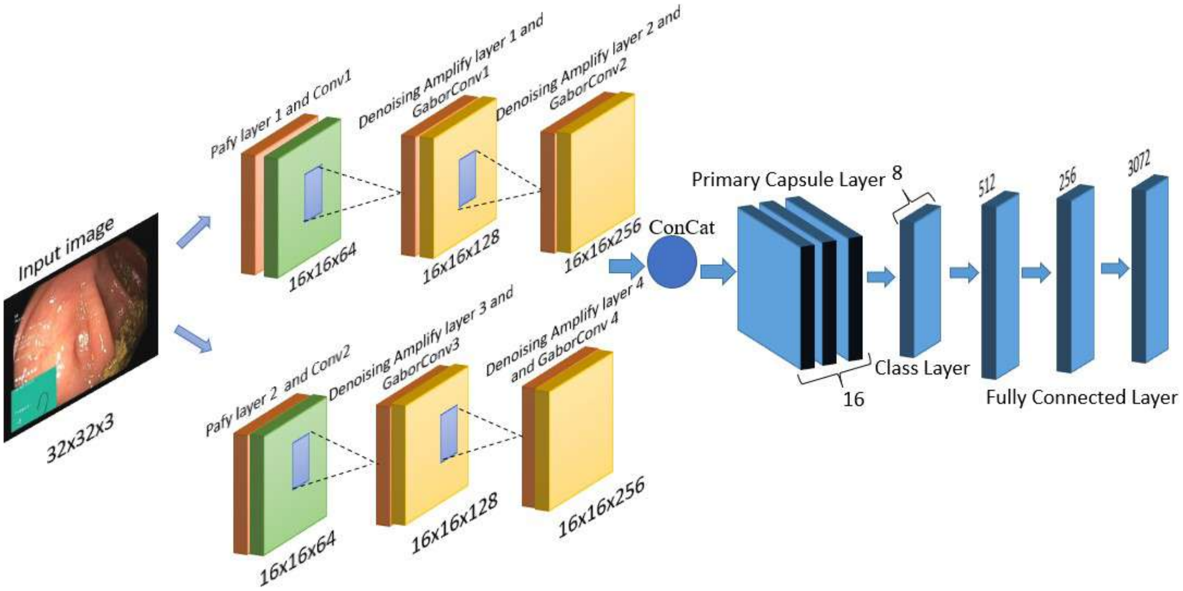


Figure 2. The proposed CapsNet model. The Pafy layers refer to the denoising patch-and-amplify layers and Conv refers to the convolutional layer.

Where p is one image $\Omega \subset \mathbb{R}^2, x \in \Omega, G^a$ is a kernel of standard deviation that is Gaussian, b and i are refining parameters, and $C(m) = \int_\Omega e - \frac{(G_b * |p(m+.) - p(n+.)|^2(0)}{i^2} d(z)$ is the standardizing element where m, n are pixels.

The last step involves the splitting of the images into two halves. Each of these image halves is passed through a branch of the network each comprising of a convolutional neural network, two Gabor layers, and two amplify layers. The denoising amplify layers are just a modified version of the denoising patch-and-amplify algorithm that omits the patch function. Both the convolutional neural network and the Gabor layers extract and preprocess features. Each branch of the network submits its batch of feature maps to an add layer that performs concatenation of the batch of feature maps. The aim of the add layer is to submit a stack of uniform feature maps to the primary capsule layer for coupling based on feature similarity. The primary capsule layer comprises of several capsules each attaches itself with features coming from the add layer. A capsule refers to a class of neurons whose vector values represent an entity of parts of an entity. Capsules u_i in the primary capsule layer, is transformed using weights w_{ij} to generate prediction vectors $\hat{u}_{i|j}$ (eqn 2).

$$\hat{u}_{i|j} = u_i.w_{ij} \tag{2}$$

Each prediction vector predicts the capsule in the class capsule layer. The strength of the prediction determines the level of similarity that exists between these two capsules. Coupling coefficients (c_{ij}) are calculated to assist in confirming the prediction and are updated by an iterative dynamic routing process. The coupling coefficients are generated by passing some log prior logits b_{ij} through a SoftMax activation function (eqn 3).

$$c_{ij} = \frac{exp(b_{ij})}{\sum_k exp(b_{ik})} \tag{3}$$

The log priors can be referred to as learnable weights. Each capsule in the class capsule layer can be predicted by more than one capsule. Therefore, a weighted sum over all the prediction vector needs to be calculated (eqn 4) for each capsule (s_j) in the class capsule layer. The result is then squashed using a non-linear squashing function (eqn 5). The aim of the squashing is to shrink short vectors to almost zero and long vectors to slightly below 1.

$$s_j = \sum_i c_{ij} \hat{u}_{i|j} \tag{4}$$

$$v_j = \frac{||s_j||^2}{1 + ||s_j||^2} \frac{s_j}{||s_j||} \tag{5}$$

Where v_j is the vector output and s_j is the total input of a capsule in the class capsule layer.

3.4. Dataset description and preprocessing

Kvasir-V2 [13] is made up of 8000 colored images of size 1024x1024 apportioned into eight classes. Five classes were used for the experiment reported in this article. The selection was based on the fact that some of the images shared similar characteristics. The classes selected were esophagitis, polyps, ulcerative colitis, normal-pylorus, and normal-cecum. The images in these classes were resized to 32 × 32 × 3 and reapportioned to 80:20 leave-out approach.

Fashion-MNIST [14] consists of 70,000 grayscale images of size 28x28. It comprises ten classes: 0: T-shirt, 1: trousers, 2: pullover, 3: dress, 4: coat, 5: sandals, 6: shirt, 7: sneakers, 8: bags, and 9: ankle boot.

CIFAR-10 [15] is made up of 60,000 colored images each of size 32x32. It comprises ten classes: 0:airplane, 1: automobile, 2: bird, 3: cat, 4: deer, 5: dog, 6: frog, 7: horse, 8: ship, and 9: truck.

3.5. Experimental settings

The proposed model was designed, developed, and implemented on a 64-bit Windows machine with 8 gigabyte read access memory. The model was trained on a Nvidia GeForce 1060 graphic processing unit with 8 gigabytes read access memory. All codes for training and evaluation were written in Keras with TensorFlow backend. The code at [2] was modified for the experiments. The batch size is set to 100, the learning rate is set to 0.001, and epochs are set to 200. The margin loss is adopted in this work.

4. Results and discussion

This section presents the performance of the proposed model on three datasets, namely, Kvasir-V2, Fashion-MNIST, and CIFAR 10. We present the training curves and confusion matrix. Several ablation tasks are performed on the proposed model to test for robustness and generalizability. Furthermore, diverse visualizations are explored to make the results of the proposed model understandable.

4.1. Experimental curves

Figure 3 presents plots of the experimental curves for the Kvasir-V2 dataset. The proposed model achieved an accuracy of 96.80% while the baseline model achieved 83.50% recognition accuracy. The experimental curves

[2] Xigenfu Guo (2018). Capsule Network code [online]. Website https://github.com/XifengGuo/CapsNet-Keras [accessed 25 August 2023]

of the proposed model show smoother curves signifying the robust nature of the network. These smooth curves signify that the proposed model can handle the complexity that exists in the Kvasir-v2 dataset. The baseline model, on the other hand, shows continuous spikes signifying the negative impact of the existing data complexity and the weakness of the model to handle these complexities thereby affecting the overall recognition accuracy. Figure 4 presents the confusion matrix of the proposed model evaluated on the Kvasir-V2 dataset. Table 1 presents an in-depth analysis of the confusion matrix (Figure 4). The values under the true positive, false positive, false negative, and true negative represent the total number of images that fulfill the column title criteria. The precision, sensitivity, and specificity values depict the confidence of the proposed model in correctly or wrongly classifying an image. In Figure 4 and Table 1, considering the per-class accuracy and the other parameters (thus, precision, sensitivity, and specificity) it can be observed that the proposed model is good at identifying esophagitis disease compared to the normal pylorus.

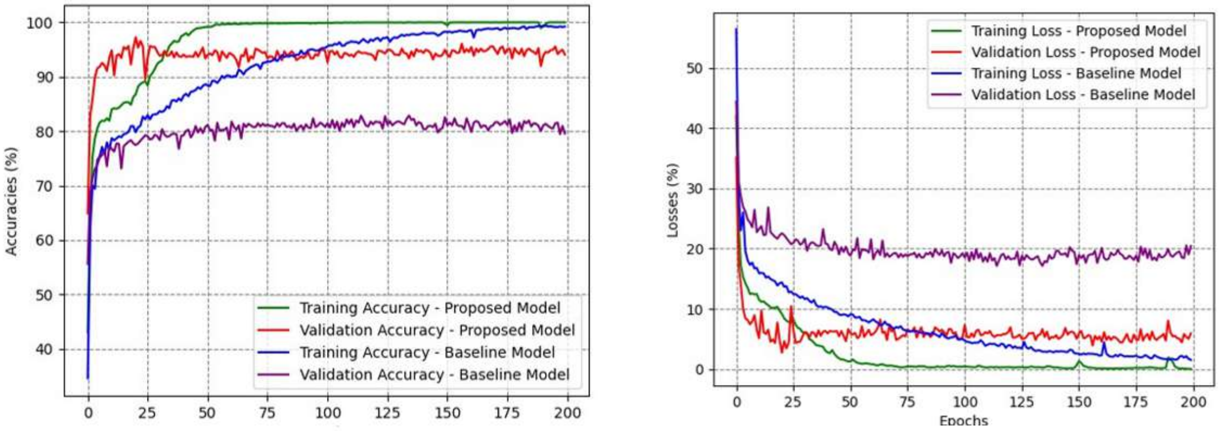


Figure 3. Accuracy and loss curves of the proposed and baseline models. This shows the train and test accuracies and losses of each model at every epoch.

Table 1. Confusion matrix analysis of the proposed model trained with the Kvasir-V2 dataset. TP = true positive, FP = false positive, FN = false negative and TN = true negative.

Class	TP	FN	FP	TN	Precision	Sensitivity	Specificity	Class per Accuracy
Esophagitis	195	5	2	798	0.990	0.975	0.998	0.993
Normal-cecum	192	8	6	794	0.970	0.960	0.993	0.986
Normal-pylorus	196	4	12	788	0.952	0.960	0.985	0.984
Polyps	192	3	8	797	0.960	0.985	0.990	0.989
Ulcerative-colitis	193	7	3	797	0.985	0.965	0.997	0.990

4.2. Ablation studies

Ablation study involves the pruning of some portion of a model and evaluating the performance compared to the performance of the unpruned model to assess the robustness of the model. In this work, the denoising amplify layers, the GaborConv layers and some of the convolutional layers are pruned for analysis. The analysis is extended to the other benchmark datasets (Fashion-MNIST and CIFAR 10). From the ablation analysis

(Table 2), it was observed that the removal of any of the denoising amplify layers, pafy layers and GaborConv layers affect the performance of the proposed model negatively. This analysis depicts that the existence of these layers in the network of the proposed model contributes massively to its performance.

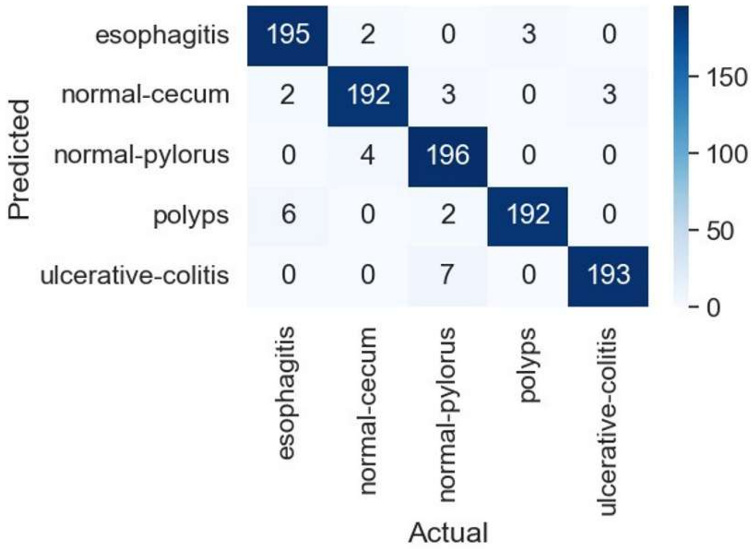


Figure 4. Confusion matrix of the proposed model trained with the Kvasir-V2 dataset.

Table 2. Ablation study of the proposed model. “-“ refers to the removal of a layer.

Layers	Accuracy (%)		
Layers	Kvasir-V2	Fashion-MNIST	Cifar-10
-All Pafy layers	91.80	90.78	75.65
-Denoising amplify layer 1 - GaborConv 1	92.35	92.09	81.90
-Conv1	96.10	94.67	84.05
-Conv2, -Denoising amplify layer 3	93.67	92.05	82.90
-Conv1, -Denoising amplify layer 1	93.90	92.10	81.29
-GaborConv4, -Denoising amplify layer 4	92.87	91.99	82.33
-Denoising amplify layer 3, - GaborConv 3	92.94	92.12	82.78
-Conv2	95.99	94.93	84.20

4.3. Number of parameters analysis

Currently, the deep learning terrain is moving towards deployment. This, therefore, requires the development of lightweight models for installation on devices with low memory such as microcontroller boards, smartphones, and many more. Development of models with low memory will encourage their deployment and usage as most state-of-the-art models [16] in deep learning is known to have a huge number of parameters[17],[11]. In this work, the first capsule network model with dynamic routing is implemented as a baseline model, and its performance is compared to the proposed model in terms of parameters and size in memory in the domain of Gastrointestinal diseases, Fashion-MNIST, and CIFAR 10. From the comparison, we observe that the proposed model produces a smaller number of parameters and occupies less space compared to the baseline model (see Table 3).

Table 3. Number of parameters analysis.

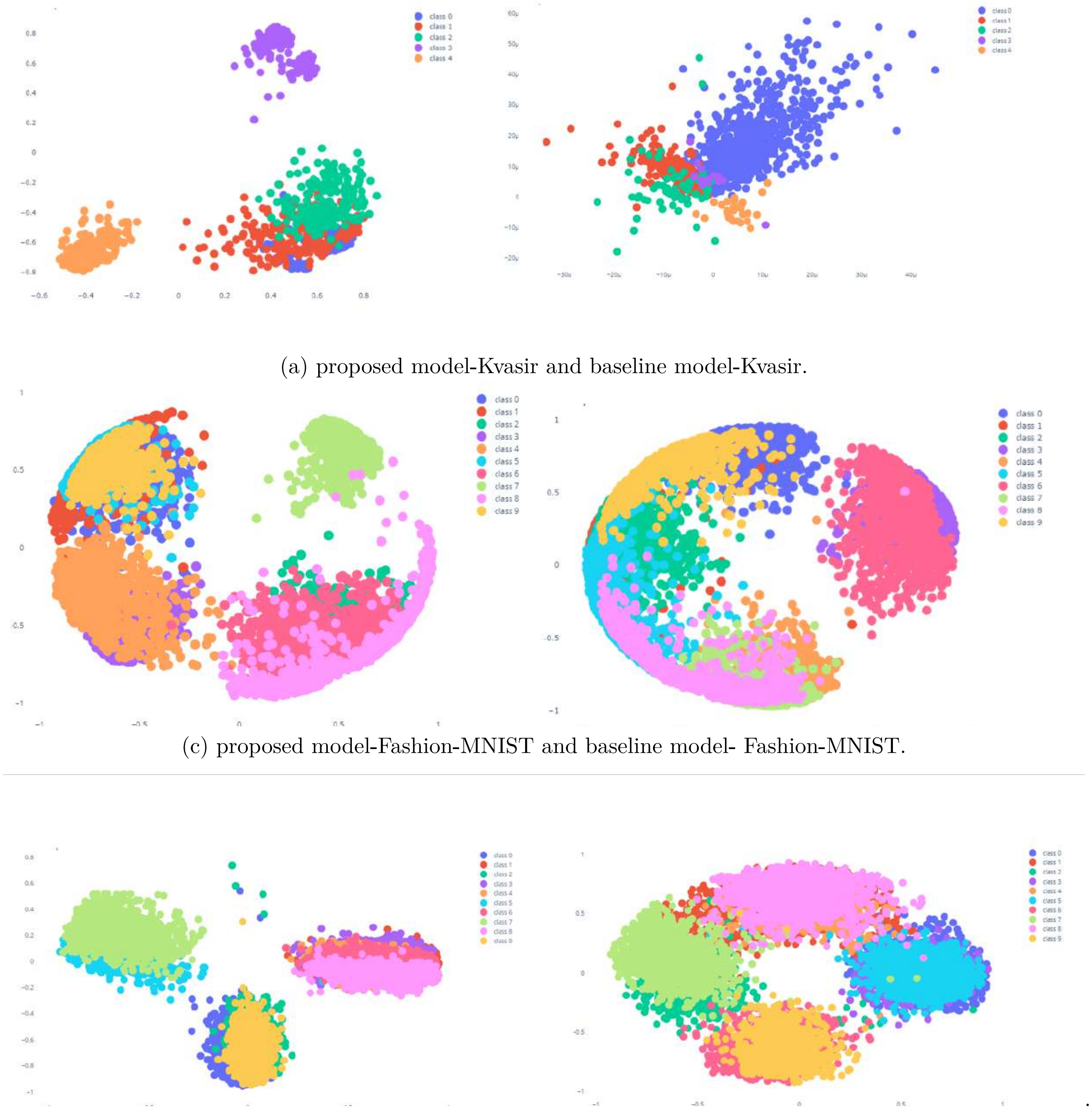
Dataset	Number of parameters (million)		Size in Memory (MB)	
Dataset	Proposed Model	Baseline Model	Proposed Model	Baseline Model
Kvasir-V2	3.65	8.7	13.25	33.40
Fashion-MNIST	2.07	7.4	8.67	28.79
CIFAR 10	4.20	10.4	15.70	37.65

4.4. Visualizations analysis

For a deep learning model to qualify for real-life testing, it needs to be a white box. Most deep learning models are said to be black box models. This makes it difficult to tell what goes on inside the model and how predictions are made. In literature, several methods are used to achieve interpretability in deep learning models. Mathematical models, Activation maps, and many more are some examples of interpretability techniques employed to give clarity to the modeler and users of the deep learning model. In this work, the visualization part of interpretability is explored to assist in making the proposed model open, its prediction understood and further evaluate its performance. For a modeler to understand if a model has enough knowledge of the domain area, the clusters produced at the class capsule layers need to be analyzed. In Figure 5 the clusters produced by the proposed model on each dataset are very compact and have fewer outliners compared to clusters produced by the baseline model. This depicts a model that has learned the needed knowledge in the application domain. It also shows that the proposed model can distinctively classify between features. This characteristic leads to the achievement of high accuracy and makes the model an essential assistant to a gastroenterologist. The next visualization area is the reconstruction domain of CapsNet. The CapsNet algorithm has a decoder layer that has a responsibility of regenerating the input image and makes predictions based on these regenerated images. Figure 6 presents the regenerated images alongside their predictions. The proposed CapsNet network out of ten predictions got 8 correct and 2 wrong while the baseline model out of ten predictions got 5 correct and 5 wrong (see Figure 6). The ability of the decoder layer to clearly reproduce the input images and attain a good number of predictions shows a model with enough domain knowledge.

4.5. Comparison with literature

In Table 4, the performance of the proposed model is compared to the state-of-the-art models in the literature in the domain of Gastrointestinal diseases, Fashion-MNIST, and CIFAR 10. The following works [19, 21, 25] were focused on modifying some parameters (such as the activation vectors) in their routing algorithms whereas the proposed model focuses on enhancing the encoder layer to facilitate the extraction of features, removal of noise (feature enhancement), and amplification of relevant features. The modifications implemented in the encoder layer improve the performance of the proposed model on the Fashion-MNIST dataset and the Kvasir-V2 dataset (thus, the Gastrointestinal disease dataset). Modification of the encoder layer aims to prevent the increase in the number of parameters. Though we experimented on Capsule Network with dynamic routing, we, however, extended our comparison to other CapsNet models with different routing algorithms.



(a) proposed model-Kvasir and baseline model-Kvasir.

(c) proposed model-Fashion-MNIST and baseline model- Fashion-MNIST.

(e)proposed model-CIFAR 10 and baseline model- CIFAR 10

Figure 5. Clusters generated at the class capsule layer of the proposed and baseline model. Each color represents a class and each data point in the clusters represents the features extracted.

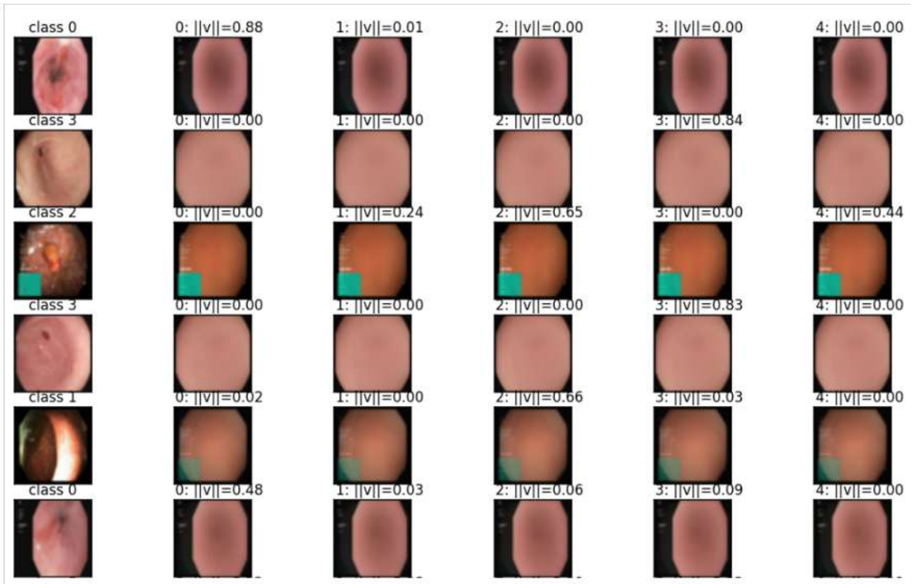


Figure 6. Reconstructed images of the proposed models. The reconstructed images are produced by the decoder layer of the proposed model.

Table 4. Comparison analysis with the state-of-the-art in the literature. "*" is used to replace unavailable values.

Method	Cifar-10	Fashion-MNIST	Kvasir-V2
Baseline model[5]	61.78	90.18	83.50
Denoising CapsNet[10]	84.57	94.93	94.16
Gabor Preprocessing blocks[12]	85.24	94.78	91.50
RS-CapsNet[18]	89.81	93.51	*
Max-Min[19]	75.92	92.07	*
ResCapsNet[20]	78.50	*	*
Quaternion CapsNet [21]	82.21	90.26	*
MS-CapsNet[22]	75.70	92.70	*
STAR-CapsNet[23]	91.23	*	*
Fsc-CapsNet[24]	78.73	93.58	*
Fast Inference [25]	70.33	91.52	*
Proposed model	85.50	95.10	96.80

5. Conclusion

In this paper, a variant Capsule network termed Denoising patch-and-amplify Gabor capsule network is proposed. The proposed model leverages the benefits of patching, Gabor filter, and feature amplification to enable the learning of meaningful hierarchical information. This makes the proposed model perform comparable well with literature. The proposed model achieved recognition accuracy of 85.50% on the Cifar10 dataset, 95.10% on the fashion-MNIST dataset, and 96.80% on the Kvasir-v2 datasets. The proposed model is scalable and adaptable for domains with smaller datasets like health. In future, the concept of uncertainty, and the human-in-the-loop will be incorporated into the proposed model and deployed.

References

[1] Gu S, Pednekar M, Slater R. Improve Image Classification Using Data Augmentation and Neural Networks. SMU Data Science Review 2019; 2 (2).

[2] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y et al. Review of deep learning: concepts, CNNs architectures, challenges, applications, future directions. Journal of Big Data, 2021; 8 (1), https://doi.org/10.1186/s40537-021-00444-8

[3] Shaikhina T, Khovanova NA. Handling limited datasets with neural networks in medical applications: A small-data approach. Artificial Intelligence in Medicine 2017; 75, 51–63. https://doi.org/10.1016/j.artmed.2016.12.003

[4] Oz E, Yigit OE, Sakarya U. DNA Chromatogram Classification Using Entropy-Based Features and Supervised Dimension Reduction Based on Global and Local Pattern Information, International Journal of Pattern Recognition and Artificial Intelligence 2023. https://doi.org/10.1142/S0218001423560190

[5] Sabour S, Nicholas F, Hinton GE. Dynamic Routing Between Capsules. Advances in Neural Information Processing Systems, 2017; 3856–3866.

[6] Cao S, Yao Y, An G. E2-Capsule Neural Networks for Facial Expression Recognition Using AU-Aware Attention. IET Image Processing 2019; 14 (11) 2417-2424.

[7] Sharif M, Khan MA, Rashid M, Yasmin M, Afza F et al. Deep CNNs and Geometric Features-Based Gastrointestinal Tract Diseases Detection and Classification from Wireless Capsule Endoscopy Images, Journal of Experimental & Theoretical Artificial Intelligence, 2019.

[8] Jia X, Meng M Q-H. Gastrointestinal Bleeding Detection in Wireless Capsule Endoscopy Images Using Handcrafted and CNNs Features, 2017; 10.0/Linux-x86_64.

[9] Özturk S, Özkaya U. Gastrointestinal tract classification using improved LSTM based CNNs, Multimed Tools Appl 2020. https://doi.org/10.1007/s11042-020-09468-3

[10] Afriyie Y, Weyori AB, Opoku A. Gastrointestinal tract disease recognition based on denoising capsule network. Cogent Engineering 2022; 9:1. https://doi.org/10.1080/23311916.2022.2142072

[11] Sarsengeldin M, Imatayeva S, Abeuov N, Naukhanov M, Erdogan AS et al. Gastrointestinal Disease Diagnosis with Hybrid Model of Capsules and CNNs. In: IEEE International Conference on Electro Information Technology, 2023; 143–146. https://doi.org/10.1109/eIT57321.2023.10187250

[12] Ayidzoe MA, Yu Y, Mensah PK, Cai J, Adu K et al. Gabor Capsule Network with Preprocessing Blocks for the Recognition of Complex Images. Machine Vision and Applications, 2021; 32. https://doi.org/https://doi.org/10.1007/s00138-021-01221-6

[13] Pogorelov K, Randel KR, Griwodz C, Eskeland SL, de Lange T et al. Kvasir: A multi-class image dataset for computer-aided gastrointestinal disease detection. In Proceedings of the 8th ACM on Multimedia Systems Conference (pp. 164-169), 2017.

[14] Xiao H, Rasul K, Vollgraf R. Fashion-mist: a novel image dataset for benchmarking machine learning algorithms. ArXiv preprint, 2017; arXiv:1708.07747.

[15] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. 2009.

[16] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. COMMUNICATIONS OF THE ACM 2012; 60 (6): 84–90. https://doi.org/10.1145/3065386

[17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings 2015; 1–14.

[18] Yang S, Lee F, Miao R, Cai J, Chen L et al. RS-CapsNet: An Advanced Capsule Network. IEEE Access 2020; 8, 85007–85018. https://doi.org/10.1109/ACCESS.2020.2992655

[19] Zhao Z, Kleinhans A, Sandhu G, Patel I, Unnikrishnan KP. Capsule Networks with Max-Min Normalization. Arxiv 1–15, 2019. http://arxiv.org/abs/1903.09662

[20] Deborshi G, Sun R. Application of Capsule Networks for Image Classification on Complex Datasets. PhD, University of Illinois at Urbana-Champaign; 2019.

[21] Özcan B, Kınlı F, Kıraç F. Quaternion Capsule Networks. In: 2020 25th International Conference on Pattern Recognition (ICPR), 2021;6858–6865. IEEE. http://arxiv.org/abs/200.04389

[22] Xiang C, Zhang L, Zou W, Tang Y, Xu C. MS-CapsNet : A Novel Multi-Scale Capsule Network. IEEE Signal Processing Letters 2018; 1. https://doi.org/10.1109/LSP.2018.2873892

[23] Ahmed K, Torresani L. STAR-CAPS: Capsule Networks with Straight-Through Attentive Routing. NeurIPS 2019; 1–10.

[24] Han T, Sun R, Shao F, Sui Y. Feature and spatial relationship coding capsule network. Journal of Electronic Imaging, 2020; 29 (2): 1. https://doi.org/10.1117/1.jei.29.2.023004

[25] Zhao Z, Kleinhans A, Sandhu G, Patel I, Unnikrishnan KP. Fast Inference in Capsule Networks Using Accumulated Routing Coefficients. 2019;1–13 http://arxiv.org/abs/1904.07304