



Early Detection of Pancreatic Ductal Adenocarcinoma (PDAC) With Deep Learning

Written By: Sayandeep Dey

Edited by: Aaryan Patel

Layout By: Sarah Derikx

Awards: Gold Medalist at Calgary Youth Science Fair (CYSF) - Winner of Director's award for outstanding science fair project at CYSF - Winner of Top 15 projects from around 1000 submissions at CYSF - Winner of Alternate Canada wide Science Fair (CWSF) Award - Selected Science project to represent Canada at the 2023 MILSET Expo-Sciences International (ESI)

Early detection of cancer is currently one of the most researched topics in the medical field. One of the most prevalent cancers where treatment planning would benefit from early detection is pancreatic cancer, which ranks fourth in Western countries in terms of cancer-related fatalities after lung, breast, and colon cancer [1]. However, pancreatic cancer is projected to surpass breast and colorectal cancers to become the second leading cause of cancer-related death by 2030 after lung cancer [1]. Although pancreatic cancer only accounts for 3% of all cancers in the USA, it has a disproportionately high death rate of 11.1 per 100,000 men and women per year [2], [3].

Based on 2022 statistics published by the American Cancer Society on the death rates of people affected by pancreatic cancer, Fig. 1. represents the death severity of pancreatic cancer [2]. As can be seen, most of the people who are diagnosed with pancreatic cancer are projected to die as well, as the combined death-to-diagnosis ratio for males and females is 88%.

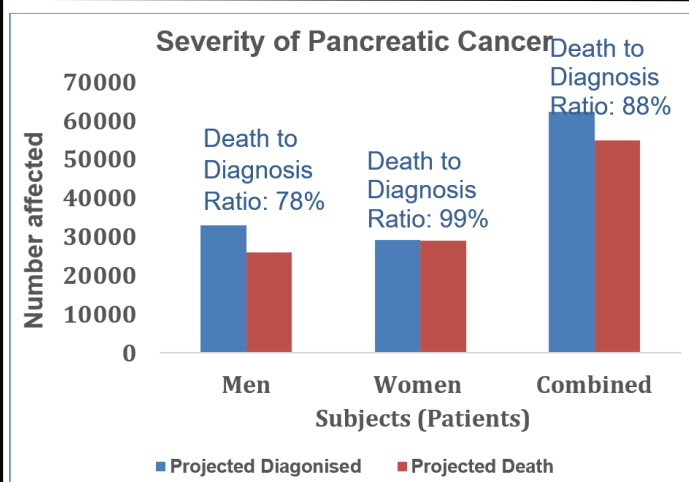


FIG. 1: THE SEVERITY OF PANCREATIC CANCER IN MEN AND WOMEN. ALL THE DATA WAS MODIFIED FROM [2]. THE BLUE BAR REPRESENTS THE NUMBER OF PEOPLE WHO WERE DIAGNOSED WITH PANCREATIC CANCER AND THE ORANGE REPRESENTS THE NUMBER THAT DIED OF PANCREATIC CANCER AFTER BEING DIAGNOSED.

PDAC is distinguished by the late onset of symptoms and subsequent rapid progression to death, which accounts for its high mortality and low survival rate [6]. Traditional imaging has not proven a useful tool for detecting early-stage PDAC, and only invasive methods such as endoscopic ultrasonography (EUS) are now available to detect early illness. There are no proven

biomarkers available to diagnose early-stage PDAC. The structure of research in this field has hampered the possibility of breakthrough advances, resulting in little collaboration and information exchange. As a result, the death rate among PDAC patients continues to climb [6].

Histopathology and/or cytology are the "gold standard" for pancreatic cancer diagnosis [5], however, they are ineffective in early-stage detection. Table 1 shows the existing diagnosis methods for 'finding' the precursor lesions, including early identification by proven procedures [5], [7].

TABLE 1: TYPE OF CURRENT DIAGNOSIS OF PANCREATIC CANCER AND THEIR LIMITATIONS [5], [7]

Type of Diagnosis	Current limitations
Biomarkers	- An overall sensitivity in the range of 25% to 50% in early-stage disease identifying credentialed biomarkers for early detection of PDAC remains an area of great unmet need. - Is not always reliable for detecting early pancreatic cancer in general and performs poorly in detecting early-stage pancreatic cancer.
Endoscopic ultrasonography	- Mainly used to diagnose metastatic disease. - The potential for early detection of asymptomatic pancreatic neoplasms in high-risk individuals (HRI) using an endoscopic approach was first reported in 1999 but didn't progress much. - Invasive procedure which comes with risks such as bleeding or infection
Computed tomography (CT) guided biopsy	- Mainly used to dragonize metastatic disease. - The sensitivity of CT scan findings suspicious for PDAC at ~18 months was only 15%. Even 6 months before diagnosis, it was only ~50%. At 3 months, it was 85%. The patients were also exposed to radiation repeatedly. - Poor at diagnosing early stages of pancreatic cancer and also an invasive procedure which can cause bleeding or infections

Early-stage pancreatic cancer is typically asymptomatic until it invades nearby tissues or spreads to distant organs [4]. Most individuals presenting with symptoms already have advanced cancer. Abnormalities





suggestive of pancreatic cancer can be detected up to a year before symptoms appear in individuals who undergo abdominal CT scans for unrelated reasons, indicating missed opportunities for early identification [4].

Improved survival through early detection relies on identifying high-risk populations that may harbour pancreatic lesions, although screening the general population remains expensive and challenging [5]. Existing diagnostic methods for pancreatic cancer, including CT scans, have limitations in detecting the disease at an early stage [5].

Deep Learning (DL) offers advantages in analyzing large volumes of medical data, particularly medical images such as X-rays, CT scans, MRI scans, and ultrasound images. Convolutional Neural Networks (CNNs) are commonly used for image analysis in cancer detection [7].

The objective is to integrate an AI algorithm into the radiology workflow as a "second reader" to enhance diagnostic confidence and potentially detect early pancreatic cancer missed by radiologists.

To achieve this, several goals must be accomplished. Suitable datasets must be obtained and cleaned before model development. Data extraction, compatibility with the development platform, and setup of ML and DL libraries are crucial steps. The DL architecture consists of an image segmentation model and a classification model that work together to segment and classify images for pancreatic cancer detection.

The hypothesis is that a DL-based model can outperform human experts in early detection, potentially improving pancreatic cancer treatment. The significance lies in providing an accurate and efficient system for early detection, considering the limitations of current methods [5]. Pancreatic cancer has a high mortality rate, but early detection can significantly improve prognosis, with some patients becoming disease-free after early treatment [15].

MATERIALS AND METHODS:

The project consists of two main stages: image segmentation model development and classification model development. These stages work together to create a system that accurately reads, analyzes, and classifies the provided data. The overall DL architecture is designed as a dual program system for training and fine-tuning. The two models, image segmentation and classification, are separate programs that collaborate to segment the images and determine if they contain pancreatic cancer.

The image segmentation model is responsible for isolating the pancreas in the images by removing irrelevant parts, generating a "mask" of the pancreas. This segmented image is then used as input for the classification model. The classification model is trained and validated using these segmented images to accurately predict the presence of pancreatic cancer. The methodology and workflow of the project are depicted in Figure 2. Further details regarding each stage are provided below.

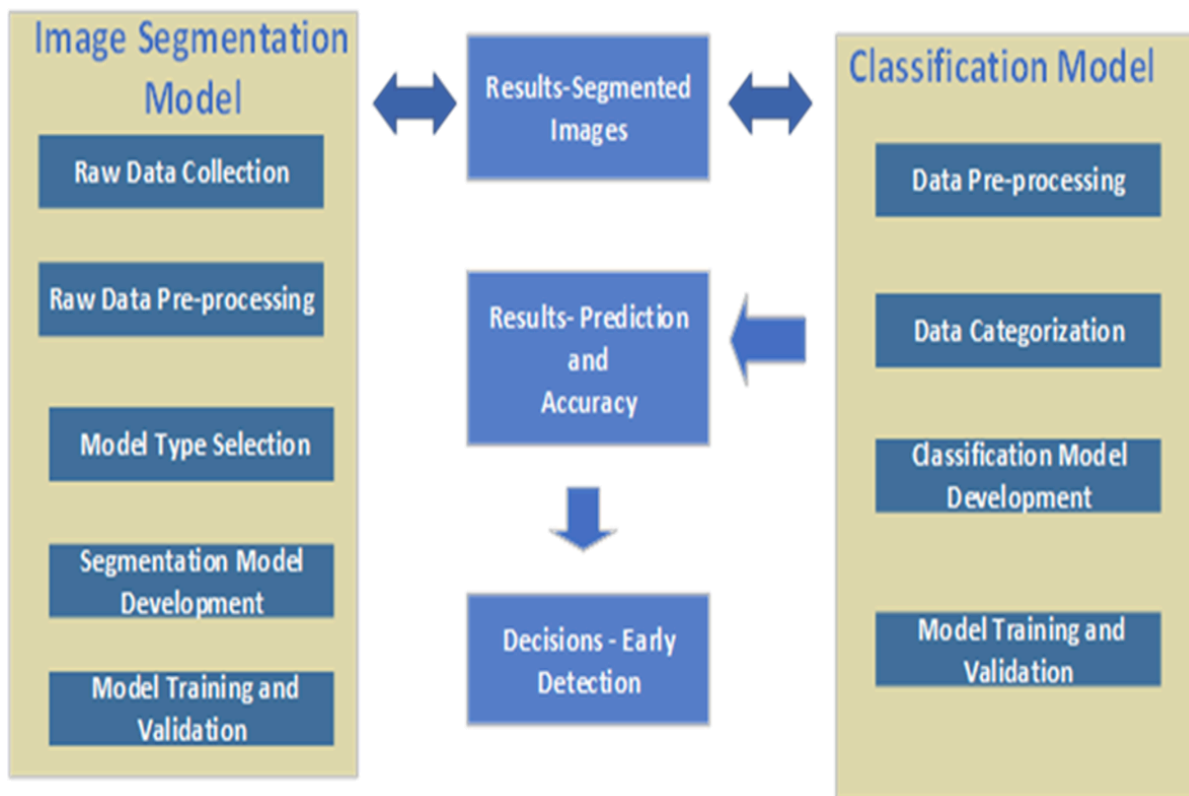


FIG. 2: DIAGRAM OF IMAGE SEGMENTATION AND CLASSIFICATION MODEL. THE FLOWCHART REPRESENTS THE FLOW OF THE DUAL MODEL SYSTEM WHERE THE INPUT IMAGE FIRST BEGINS WITH THE SEGMENTATION MODEL, GETS SEGMENTED, THEN PASSES ON TO THE CLASSIFICATION MODEL WHERE FINALLY THE MODEL MAKES A PREDICTION ON THE IMAGE AND THE ACCURACY IS OBTAINED.



1.1 Segmentation Model

Semantic segmentation is an important task in medical imaging and is useful in identifying the different structures within an image. Pancreas segmentation is a specific type of semantic segmentation that is used in medical imaging to locate and isolate the pancreas in medical scans such as computed tomography (CT) or magnetic resonance imaging (MRI) images. The goal of pancreas segmentation is to accurately identify the pancreas and separate it from the surrounding background, allowing for further analysis and assessment of the pancreas, such as the detection of abnormalities or the measurement of its size and shape. In this project, the images of the pancreas were segmented by an ML algorithm to only show the region of interest (ROI). The steps below show the process of the segmentation model.

1.1.1 Step 1: Raw Data Collection

Data collection is a crucial step in the development of a semantic segmentation model for images of the pancreas. The quality and quantity of data can greatly impact the performance of the model. Firstly, two large datasets were collected from the cancer imaging archive (National Cancer Institute). The details of the datasets are summarized below:

TABLE 2: SPECIFICATIONS OF EACH DATASET USED [16, 17]

Dataset	Specifications
Unhealthy Dataset	- Image type: CT scan 421 images - Every patient has cancer (a tumour is present), and they are all pre-treated. 512x512 pixel images
Healthy Dataset	- Image type: CT scan 80 patients (53 male and 27 female) - About 240 images per patient - All healthy pancreas without any cancer - Mean age: 46.8 +/- 16.7. 512x512 pixel images

The datasets were first imported to a CSV file in the same folder as the program so that the data could be implemented. The data was then extracted into the following subfolders: image_path, annotations_path, patient_index, image_index, and imageAndLabels. The data within all the subfolders were then processed(read) using SimpleITK.

1.1.2 Step 2: Cleaning and Preprocessing

The preprocessing of the data plays a crucial role in the success of any DL task, and semantic segmentation is no exception. The quality and type of preprocessing can greatly impact the accuracy and efficiency of the final model. The segmentation model took all the images in the healthy dataset as input and was trained upon those images only. After training, all the images from the healthy and unhealthy datasets were passed through the segmentation model, and the output (segmented images) was then collected. In the case of semantic segmentation of the pancreas, the following preprocessing steps were performed:

Classes: The brief description of each function that was used under the class is described below.

TABLE 3: DATA PREPROCESSING CODE FUNCTIONS AND THEIR PURPOSE

<u>len</u> class	Gets the length of a subfolder, i.e., the number of samples of data in the subfolder
<u>getitem</u> class	Creates a mask and an image of the input provided, taking in the array of the image as well as the input image
Mask	Isolates a specific portion of an image that has been assigned a label in a computer vision model
Transform function	Used to get a mask of an image
<u>Read_image</u> function	Reads the image from the provided path and converts it into an array of numbers
Visualize function	Plots or shows an image using the matplotlib library
<u>Dice_metric</u> function	Measures the accuracy of the model after training using the Dice coefficient, which calculates the overlap between predicted and ground truth segmentations

1.1.3 Step 3: Model Selection and Development

A pre-trained U-net model, "resnet34," was utilized for the segmentation model in this project. ResNet34 is a variant of the ResNet architecture with 34 layers. It incorporates residual blocks, consisting of two convolutional layers with batch normalization and ReLU activation, followed by an identity mapping that bypasses the activation function. This residual connection facilitates the flow of gradients and helps mitigate the vanishing gradient problem. In the U-Net model, the ResNet34 network replaces the encoder network, extracting features at different scales from the input image. These features are then fed into the decoder network, which performs upsampling and combines the features to generate a segmentation map. The ResNet34 network's strong feature extraction capabilities enable accurate segmentation of complex images. To optimize accuracy, pre-trained weights from the "imagenet" dataset were employed for the encoder network. These weights capture general patterns and features present in natural images. By starting with pre-trained ImageNet weights, the ResNet model leverages this prior knowledge, allowing for more efficient and effective learning on the specific segmentation task. Training the model was performed on a GPU to expedite the process and reduce training time. The computational power of a GPU enables faster model training and optimization, facilitating the development of accurate segmentation models. By employing the pre-trained ResNet34 model and leveraging GPU acceleration, the project aimed to achieve accurate and efficient segmentation of images, contributing to improved analysis and understanding of the target data.

1.1.4 Step 4: Model Training

The model ran for 15 epochs in total, where the model's validation and training loss were measured each epoch. The model was then evaluated on the validation set, and the performance was evaluated using three metrics: validation loss, training loss and dice metric. Validation loss is calculated by feeding the validation dataset through the trained model and comparing the predicted output to the actual output. The difference between these two values is then averaged across all examples in the validation dataset to obtain a single loss value. Training loss is a metric



used to evaluate the performance of a ML model during the training process. It measures the difference between the predicted output of the model and the actual output on the training dataset. Training loss is calculated by feeding the training dataset through the model and comparing the predicted output to the actual output. The difference between these two values is then averaged across all examples in the training dataset to obtain a single loss value. Finally, the dice metric or dice coefficient was measured using the "dice_metric" function mentioned above.

1.2 Classification Model

The classification model's job is to then use the segmented images from the segmentation model and learn how to classify images of the pancreas (if they are cancerous or not). The model's architecture is a base CNN structure, and it uses the segmented images as an input for training and its output is the final accuracy of the model.

1.2.1 Step 1: Segmented Data Pre-processing

After the segmentation model was trained and had achieved optimal accuracy, all the images in the unhealthy and healthy datasets were fed into it to be segmented. After the segmentation of all the images from the two different datasets, the new images were collected and put into a CSV file yet again. The classification model had the same functions and classes as the segmentation excluding the transform functions. The data from the segmented unhealthy and healthy datasets were then combined and split into training and validation sets (80/20 split). The images were also shuffled to increase randomness and better the diversity of both sets. These datasets were used to train, validate, and ultimately refine the classification model's predictions. The training set similar to the segmentation model was used to train the model and the validation set was used to validate the model after training.

1.2.2 Step 2: Model Development

For the classification model, a pre-trained model was also used (GoogleNet) with the weight: "IMAGENET1K_V1" to optimize accuracy. The GoogleNet CNN model consists of 22 layers, including 9 Inception modules. Inception modules are a combination of various convolutions, pooling, and concatenations. These modules allow the network to capture features at multiple scales and resolutions in parallel, which leads to higher accuracy and faster training time. One of the notable features of GoogleNet is the usage of 1x1 convolutions, which are used to reduce the dimensionality of feature maps before applying larger convolutions. This technique reduces the number of parameters and computational costs, making it easier to train deep networks. GoogleNet achieved state-of-the-art performance on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2014, which is a benchmark for image classification. It demonstrated the power of DL in computer vision and inspired the development of other advanced architectures such as ResNet, DenseNet, and EfficientNet. It was also trained upon the GPU where the loss was measured in CrossEntropyLoss, and similar to the segmentation model, the optimizer was Adam. It was trained upon 100 epochs and had an overfitting safety eject mechanism where if the model started to overfit

within 5 epochs, the training would automatically shut down and the epoch with the best model accuracy and loss would be used. The pre-trained weights for IMAGENET1K_V1 were used as a starting point for transfer learning. Transfer learning is a technique used to fine-tune pre-trained models on a new dataset with a different set of classes. By starting with pre-trained weights, the model learned to recognize new classes of objects more quickly and with less data.

1.2.3 Step 3: Training and Validation

It was trained upon 100 epochs and had an overfitting safety eject mechanism where if the model started to overfit within 5 epochs, the training would automatically shut down and the epoch with the best model accuracy and loss would be used. Therefore, sometimes after the model reached optimal accuracy, it stopped training, and the model ran for fewer epochs than what was provided. Whilst training, the validation loss, validation accuracy, training accuracy and training loss were all calculated. The validation loss and training loss were calculated in the same fashion as the segmentation model.

RESULTS

The main goal of this study was to see if the AI system that was developed could outperform current methods of early detection of pancreatic cancer. After making both the segmentation model and classification model and making sure they were compatible with each other, they were both trained. At the end of the training process for the segmentation model, approximately 19,000 images from the healthy dataset [16] were used to train the model with the training loss and validation loss both being measured. After the training process, 19,000 images from the healthy dataset [16] as well as 421 images from the unhealthy dataset [17] were all passed through the segmentation model. The segmented images were then passed on to the classification model which was used to train it and then the confusion matrix, training and validation losses and accuracies were all calculated. The results obtained provide valuable insights into the research question, and the following section will showcase these findings in depth.

1.1 Segmentation Model

The training and validation loss were monitored during the training of the segmentation model. The loss is a measure of how well the model can predict the correct segmentation mask for each input image. All the loss functions explicated below were calculated by a simple loss function within both models called "binary cross entropy."

1.1.1 Training Loss:

The training loss refers to the loss calculated on the training set during the training process. A lower loss indicates the model is performing well and generalizing well to new data, whereas a high loss value indicates the model is not performing well and there is an internal flaw in the training process or the data provided. The segmentation model ran for 18 total epochs for the training phase, meaning that it cycled through the data in the training set 18 total times before coming to a stop. The loss for the training is depicted in Fig. 4. as shown below. Looking at the training loss values, it is evident that the model is



improving as the epochs progress. The initial loss value for the first epoch is higher than the subsequent epochs, indicating that the model is starting with a higher loss and gradually improving over time. As shown below, the training loss is initially at a high of approximately 0.034 during the first few epochs of the training phase. It then instantly decreases to a value of 0.007 in the second epoch and continues to gradually decrease as the epochs continue. The training loss value comes to a plateau in the 8-18 epoch range with a value of 0.001. There is no significant overfitting observed in the training loss values, as the loss values decrease with the increase of epochs, and there are no significant increases in loss values in later epochs.

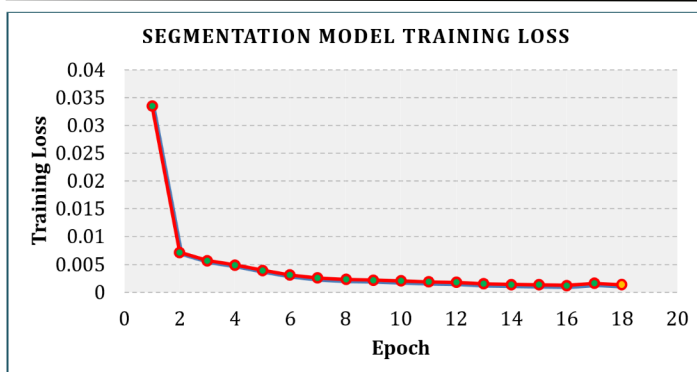


FIG. 3: SEGMENTATION MODEL TRAINING LOSS FOR EACH RUN. THE RED LINE REPRESENTS THE TRAINING LOSS VALUES FOR THE SEGMENTATION MODEL PER EPOCH AND THE DOTS REPRESENT THE EXACT VALUE ON THE SPECIFIC EPOCH OF TRAINING. THERE ARE 18 EPOCHS, MEANING THAT THE MODEL CYCLED OVER THE TRAINING DATA 18 TIMES. THE Y-AXIS REPRESENTS THE TRAINING LOSS VALUES ITSELF AND IT DOES NOT HAVE A UNIT, BUT INSTEAD IS ON A SCALE OF 0 TO 1 WITH 0 BEING THE LOWEST AND 1 BEING THE HIGHEST LOSS.

1.1.2 Validation loss:

The validation loss refers to the loss calculated on a separate validation set that was not used during the training process. This validation loss was calculated after the segmentation model was done training on the training set. The purpose of the validation loss was to see how well the model performed on new data it had never seen before as all the images in the validation set were new to the model. It is an essential metric used to evaluate the model's generalizability ability and if it performs segmentation on new images just as well as the images it was trained upon. The validation loss values are depicted in Fig. 5. As shown in the graph below, the segmentation model ran through the images in the validation set 18 times as there are 18 total epochs. The validation loss starts at 0.0103 during the first epoch and drops significantly to 0.0073 during the second epoch, indicating a good start for the model. The validation loss continued to decrease gradually until the seventh epoch, reaching a low of 0.004. From the eighth epoch, the validation loss fluctuates slightly but remains within a relatively consistent range until the end of training. The final validation loss at epoch 18 is 0.0052, which is higher than the lowest validation loss achieved at epoch

seven, indicating a slight drop in performance in the latter stages of training. The general trend of the validation loss is in a decrease and all the values remain within the low range of 0.01 to 0.004. It shows a similar trend to the training loss however there is some overfitting present in the validation phase. Overfitting occurs when the model gets used to the training data so much that it does not perform as expected on new data as instead of analyzing the patterns within the training images, the model memorizes the images themselves. This leads to slight bumps in the validation loss as shown in epochs 9 and 14.

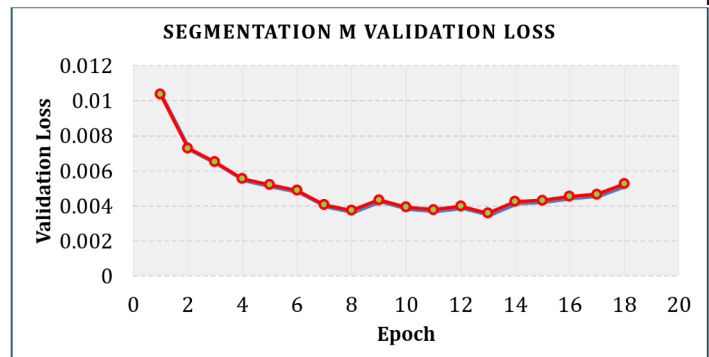


FIG. 4: SEGMENTATION MODEL VALIDATION LOSS FOR EACH RUN. THE WHITE LINE REPRESENTS THE VALIDATION LOSS VALUES FOR THE SEGMENTATION MODEL PER EPOCH AND THE DOTS REPRESENT THE EXACT VALUE ON THE SPECIFIC EPOCH OF TRAINING. THERE ARE 18 EPOCHS, MEANING THAT THE MODEL CYCLED OVER THE VALIDATION DATA 18 TIMES. THE Y-AXIS REPRESENTS THE TRAINING LOSS VALUES ITSELF AND IT DOES NOT HAVE A UNIT, BUT INSTEAD IS ON A SCALE OF 0 TO 1 WITH 0 BEING THE LOWEST AND 1 BEING THE HIGHEST LOSS.

1.2 Classification Model

The classification model measured the following: confusion matrix, training loss and accuracy and validation loss and accuracy. Each evaluation metric provides a detailed description of how the model performs in classifying specific types of data and provides the final accuracy of the entire DL system in predicting pancreatic cancer.

The precision, sensitivity/recall, F1-score and specificity were all obtained from the classification model after training and are shown below. The TN, TP, FN and FP were all calculated with the help of the program which utilized the formulas shown below and then a final confusion matrix was made using the data extracted from the classification model.

From the classification model, the precision, sensitivity, F1-score and specificity that were obtained were all 0.9375 or 93.75%. With this, TP and FP can be calculated as:

$$TP / (TP+FP) = 93.75\%$$

*The TP is calculated to be 93.75% and FP to be 6.25%
From the classification model, the sensitivity that was obtained was 0.9375. With this, FN can be calculated as:*

$$TP / (TP+FN) = 93.75/100 = 93.75\%$$



The FN is calculated to be 6.25%.
The true negative can be calculated as:

$$TN = 100 - FN, \text{ which leads to } TN = 100 - 6.25 = 93.75$$

With this, the confusion matrix in Fig. 6. was made. As shown in the confusion matrix below, the TN and TP rates for the model are approximately 94% and the FP and FN rates are 6.2%. The colour scale on the right is used to represent the depth of the values for FP, FN, TN and TP, and it shows a contrast between black and white for each of the values. The TN and TP values show that the model can accurately predict images that have cancer and do not have cancer 94% of the time and have no problems within a certain category of images.

1.2.1 Final Confusion Matrix

From all the data and calculations done above, this is the confusion matrix for the classification model:

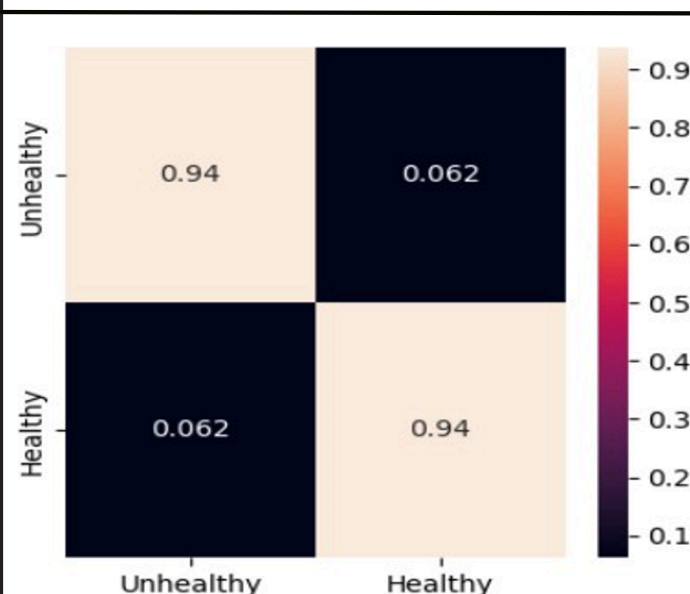


FIG. 5: FINAL CONFUSION MATRIX FOR CLASSIFICATION MODEL. THE TP AND TN ARE SHOWN ON THE TOP LEFT AND BOTTOM RIGHT RESPECTIVELY AND THE FP AND FN ARE SHOWN ON THE TOP RIGHT AND BOTTOM LEFT RESPECTIVELY. THE COLOUR GRADIENT ON THE RIGHT REPRESENTS HOW HIGH OR LOW THE VALUES ARE WITHIN THE MATRIX, WITH BLACK BEING EXTREMELY LOW AND WHITE BEING HIGH.

1.2.2 Training Loss vs Accuracy

The loss is a measure of how well the model can predict the output compared to the actual output whereas the accuracy of the classification model is a measure of how well the DL model can correctly classify both positive and negative cases. It is calculated as the proportion of correct classifications out of all cases. The training loss and accuracy are the loss and accuracy in the training phase (only within the training dataset). The model performance data are presented in Fig. 7. The training loss and training accuracy results for the classification model were presented over 15 epochs as shown below. As evident

from the graph, the training loss starts at 0.40 during the first epoch and drops significantly to 0.23 during the second epoch, indicating a good start for the model. The training loss continues to decrease gradually until the 15th epoch, reaching a low of 0.15. The loss plateaus at the 8 epoch range with a value of 0.15. The training accuracy of the model also shows improvement with each epoch. The accuracy starts at 0.84 during the first epoch and increases to 0.9 by the second epoch. The accuracy remains consistently high throughout the remaining epochs at around 0.95. The final training loss is at the low value of 0.15 and the final accuracy for the training stage is 95% meaning that the model was able to successfully identify 95% of the images in the training dataset throughout the 18 epochs of training.

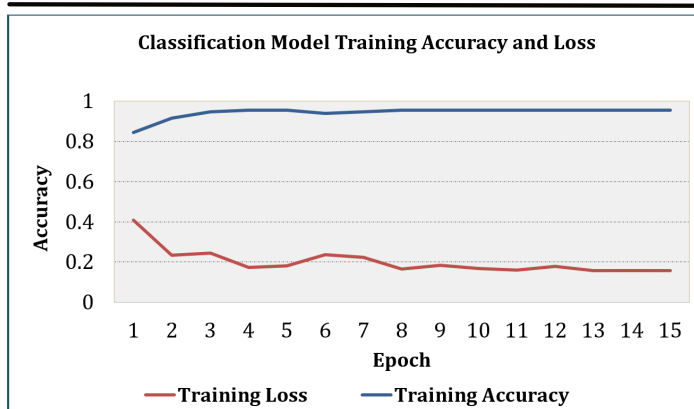


FIG. 6: CLASSIFICATION MODEL TRAINING LOSS AND ACCURACY. THE TRAINING ACCURACY IS REPRESENTED IN THE YELLOW AND THE TRAINING LOSS IS REPRESENTED IN THE RED. THE Y-AXIS REPRESENTS THE ACCURACY, WHICH IS ON A SCALE OF 0 TO 1, 0 BEING THE LOWEST AND 1 BEING THE HIGHEST. THE MODEL RAN FOR 15 EPOCHS, MEANING THAT IT COVERED THE TRAINING DATA 15 TIMES OVER.

1.2.3 Validation Loss vs Accuracy

Similar to training loss and accuracy, validation loss and accuracy reflect similar performance metrics but are now on the validation set. This validation set is filled with images the model has never evaluated before. It's used to see if the model can have similar success with new images as it had with images it had previously been trained with. The loss and accuracy are the same as described in the training loss and accuracy section, but it is based on the images in the validation set instead of the training set. The model performance data are presented in Fig. 8. The validation loss and accuracy results for the classification model were presented over 15 epochs as shown below. As evident from the graph, the validation loss starts at 0.30 during the first epoch and fluctuates between 0.22 and 0.42 throughout the remaining epochs. It is not as consistently decreasing as the training loss, which suggests that the model may be overfitting to the training data. However, the general trend of the loss is downward, indicating the model is performing well and generalizing to the new data. The validation accuracy starts at 0.90 during the first epoch and fluctuates between 0.84 and 0.94



throughout the remaining epochs. The accuracy is not as consistently high as the training accuracy, which suggests slight overfitting. Nevertheless, the validation data achieves a high of approximately 94% similar to the training accuracy and shows an upward trend. The final validation accuracy of 94% is the final accuracy of the DL system of the segmentation and classification model. This number of 94% means that the system can correctly classify images based on whether they have cancer or not 94% of the time.

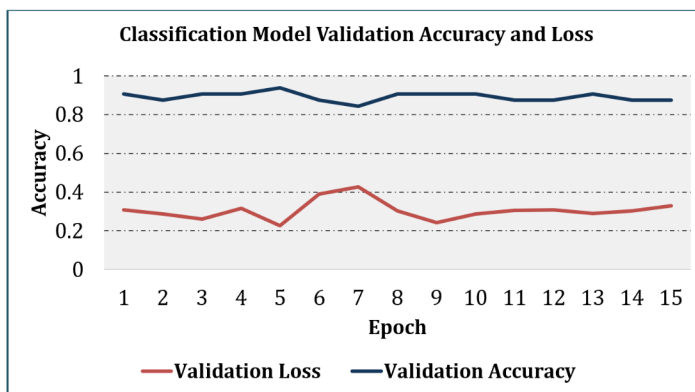


FIG. 7: CLASSIFICATION MODEL VALIDATION LOSS AND ACCURACY. THE VALIDATION ACCURACY IS REPRESENTED IN YELLOW AND THE VALIDATION LOSS IS REPRESENTED IN RED. THE Y-AXIS REPRESENTS THE ACCURACY, WHICH IS ON A SCALE OF 0 TO 1, 0 BEING THE LOWEST AND 1 BEING THE HIGHEST. THE MODEL RAN FOR 15 EPOCHS, MEANING THAT IT COVERED THE TRAINING DATA 15 TIMES OVER.

DISCUSSION

The results obtained from the segmentation and classification models provide valuable insights into the research question of accurately identifying pancreatic cancer. The segmentation model, which focuses on predicting the correct segmentation mask for each input image, showed promising performance during the training process. The training loss values demonstrated a clear improvement as the epochs progressed, indicating that the model was learning and adapting to the data. This finding aligns with previous studies that have shown the effectiveness of DL models in image segmentation tasks for various medical applications [22]. The gradual decrease in training loss suggests that the model was able to capture and learn the underlying patterns and features related to pancreatic cancer in the healthy dataset.

The validation loss, which measures the model's performance on new, unseen data, also showed a decreasing trend, although with some fluctuations. The initial drop in validation loss during the first few epochs indicates that the model was able to generalize well to new images. Generalizability refers to the model's ability to adapt properly to new, previously unseen data. However, the slight increase in validation loss towards the later epochs suggests a potential overfitting issue. Overfitting occurs when the model becomes too specific to the training data and fails to generalize to new data [20]. This phenomenon has been observed in other medical imaging studies as

well [24]. To address this issue, regularization techniques such as dropout or early stopping could be employed in future iterations of the model to improve generalization performance [21].

The training loss and accuracy of the classification model further demonstrate the model's capability to learn and classify the training data effectively. The decreasing trend in training loss and the consistently high accuracy values throughout the training epochs indicate that the model successfully captured the underlying patterns and features of the training dataset. This is consistent with previous studies that have utilized DL models for cancer classification tasks [18, 23].

The validation loss and accuracy, which evaluate the model's generalization performance, showed a slightly fluctuating trend. While the validation loss remained relatively stable with some minor fluctuations, the accuracy ranged from 0.84 to 0.94. Although there were signs of overfitting, the validation accuracy reached a high of approximately 94%, indicating that the model could generalize well to new, unseen images. However, further efforts to address the overfitting issue could be explored in future iterations of the model to improve its performance on unseen data.

CONCLUSION

In conclusion, the results obtained from both the segmentation and classification models provide valuable insights into predicting pancreatic cancer. The segmentation model demonstrated a gradual improvement in training loss, indicating its ability to learn the underlying patterns associated with pancreatic cancer. These results are supported by previous studies [20, 24] that have explored the application of DL models in medical imaging and cancer classification tasks. Although some overfitting was observed, the models showcased strong potential for pancreatic cancer prediction and warrant further exploration and refinement in future studies.

REFERENCES

- [1] Rahib, L., Smith, B. D., Aizenberg, R., Rosenzweig, A. B., Fleshman, J. M., & Matrisian, L. M. (2014). Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Research*, 74(11), 2913–2921. doi: 10.1158/0008-5472.CAN-14-0155.
- [2] American Cancer Society. (2022). Key statistics for pancreatic cancer. Retrieved from <https://www.cancer.org/cancer/pancreatic-cancer/about/key-statistics.html>.
- [3] Surveillance, Epidemiology, and End Results Program. (2022). Cancer of the pancreas - cancer stat facts. Retrieved from <https://seer.cancer.gov/statfacts/html/pancreas.html>.
- [4] Vincent, A., Herman, J., Schulick, R., Hruban, R. H., & Goggins, M. (2011). Pancreatic cancer. *Lancet*, 378(9791), 607–620. doi: 10.1016/S0140-6736(10)62307-0.
- [5] Kenner, B., et al. (2021). Artificial Intelligence and Early Detection of Pancreatic Cancer: 2020 Summative Review. *Pancreas*, 50(3), 251–279. doi: 10.1097/MPA.0000000000001762.
- [6] Kenner, B. J., et al. (2016). Early Detection of Pancreatic Cancer-a Defined Future Using Lessons From Other Cancers: A White Paper. *Pancreas*, 45(8), 1073–1079. doi: 10.1097/MPA.0000000000000701.
- [7] Tamashiro, A., et al. (2020). Artificial intelligence-based detection of pharyngeal cancer using convolutional neural networks. *Digestive Endoscopy*, 32(7), 1057–1065. doi: 10.1111/den.13653.
- [8] Mishra, M. (2020). Convolutional Neural Networks, Explained. *Towards Data Science*. Retrieved from <https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939>.
- [9] Lang, N. (2021). Using Convolutional Neural Network for image classification. *Towards Data Science*. Retrieved from <https://>



ABOUT THE AUTHOR SAYANDEEP DEY

Sayan is a high school student with background in approaching STEM subjects creatively. With strong grasp in computational techniques, algorithms and mathematical theory, Sayan is avid programming enthusiast who loves to indulge himself into computer science and discover its practical application into various fields. An enthusiastic learner eager to contribute team success who wants to grow and excel in STEM field.



- towardsdatascience.com/using-convolutional-neural-network-for-image-classification-5997bfd0ede4.
- [10] Bhatia, R. (2018). Why convolutional neural networks are the go-to models in deep learning. *Analytics India Magazine*. Retrieved from <https://analyticsindiamag.com/why-convolutional-neural-networks-are-the-go-to-models-in-deep-learning/>.
- [11] Baruch, L. (2020). Convolutional Neural Network (CNN) questions. OpenGenus IQ: Computing Expertise & Legacy. Retrieved from <https://iq.opengenus.org/cnn-questions/>.
- [12] Zheng, J., Lin, D., Gao, Z., Wang, S., He, M., & Fan, J. (2020). Deep Learning Assisted Efficient AdaBoost Algorithm for Breast Cancer Detection and Early Diagnosis. *IEEE Access*, 8, 96946–96954. doi: 10.1109/ACCESS.2020.2993536.
- [13] Welikala, R. A., et al. (2020). Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer. *IEEE Access*, 8, 132677–132693. doi: 10.1109/ACCESS.2020.3010180.
- [14] American Cancer Society. (2022). Can pancreatic cancer be found early? Retrieved from <https://www.cancer.org/cancer/pancreatic-cancer/detection-diagnosis-staging/detection.html>.
- [15] Johns Hopkins Medicine. (2022). Pancreatic Cancer Prognosis. Retrieved from <https://www.hopkinsmedicine.org/health/conditions-and-diseases/pancreatic-cancer/pancreatic-cancer-prognosis>.
- [16] Fevrier-Sullivan, B. (2023). Prediction of sunitinib efficacy using computed tomography in patients with pancreatic neuroendocrine tumours (ctpred-sunitinib-pannet). The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. Retrieved from <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=133071846>.
- [17] Berryman, S. (2023). Pancreas-CT. The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. Retrieved from <https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT>.
- [18] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. doi: 10.1038/nature21056.
- [19] Gruener, R., Bonab, A. A., Schwabe, R. F., & Kumar, V. (2020). Computer-aided diagnosis of pancreatic cancer with 3D convolutional neural networks. *Scientific Reports*, 10(1), 1–11. doi: 10.1038/s41598-020-57743-w.
- [20] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. *Springer Science & Business Media*. doi: 10.1007/978-0-387-84858-7.
- [21] Kang, K., Choi, D. H., Lee, J. H., Jang, J. Y., & Kang, S. (2021). Deep learning for the diagnosis of pancreatic ductal adenocarcinoma: a systematic review. *Korean Journal of Radiology*, 22(2), 157–165. doi: 10.3348/kjr.2020.0207.
- [22] Litjens, G., Kool, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. doi: 10.1016/j.media.2017.07.005.
- [23] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., et al. (2017). CheXnet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225. doi: 10.1109/CVPR.2017.441.
- [24] Wang, S., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2019). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097–2106). doi: 10.1109/CVPR.2019.00219.
- [25] Yang, L., Jin, R., Xie, Y., Hu, H., Han, Y., Qin, Y., et al. (2020). Combining deep learning and handcrafted features for improved breast cancer classification on dynamic contrast-enhanced MR images. *Medical Image Analysis*, 59, 101570. doi: 10.1016/j.media.2019.101570.