A Comparative Study of Different Pre-Trained DeepLearning Models and Custom CNN for Pancreatic Tumor Detection

Muhammed Talha Zavalsız
Department of Computer Engineering,
Istanbul Medipol University, Turkey

Sleiman Alhajj
International School of Medicine,
Istanbul Medipol University, Turkey

Kashfia Sailunaz
Department of Computer Science,
University of Calgary, Canada

Tansel Ozyer
Department of Computer Engineering,
Ankara Medipol University, Turkey

Reda Alhajj
Department of Health Informatics,
University of Southern Denmark,
Denmark

Abstract: *Artificial Intelligence and its sub-branches like MachineLearning (ML) and Deep Learning (DL) applications have the potential to have positive effects that can directly affect human life. Medical imaging is briefly making the internal structure of the human body visible with various methods. With deep learning models, cancer detection, which is one of the most lethal diseases in the world, can be made possible with high accuracy. Pancreatic Tumor detection, which is one of the cancer types with the highest fatality rate, is one of the main targets of this project, together with the data set of computed tomography images,which is one of the medical imaging techniques and has an effective structure in Pancreatic Cancer imaging. In the field of image classification, which is a computer vision task, the transferlearning technique, which has gained popularity in recent years, has been applied quite frequently. Using pre-trained models werepreviously trained on a fairly large dataset and using them on medical images is common nowadays. The main objective of this article is to use this method, which is very popular inthe medical imaging field, in the detection of PDAC, one of the deadliest types of pancreatic cancer, and to investigate how it per-forms compared to the custom model created and trained from scratch. The pre-trained models which are used in this project areVGG-16 and ResNet, which are popular Convolutional Neutral Network models, for Pancreatic Tumor Detection task. With the use of these models, early diagnosis of pancreatic cancer, which progresses insidiously and therefore does not spread to neighboring tissues and organs when the treatment process is started, may be possible. Due to the abundance of medical imagesreviewed by medical professionals, which is one of the main causes for heavy workload of healthcare systems, this applicationcan assist radiologists and other specialists in Pancreatic Tumor detection by providing faster and more accurate method.*

Keywords: *Deep learning, medical image, pancreatic tumor detection, convolutional neural networks, pancreatic ductal adenocarcinoma.*

Received April 29, 2023; accepted May 10, 2023
https://doi.org/10.34028/iajit/20/3A/9

1. Introduction

Cancer is the name given to malignant tumors that are formed by the irregular division and rapid proliferation of cellsin an organ or tissue. In more general terms, cancer occurs as aresult of the proliferation of cells in various parts of our body, proliferation is an out-of-control proliferation, the abnormal cells begin to multiply out of control. Cancer can cause seriousillness, as well as cause death [12].

Pancreas cancer is a disease that occurs as a result of malignant tumor formation in the pancreas. According to the reports, Pancreatic Ductal Adenocarcinoma, (PDAC), is the most common type of this disease. This tumor accounts for 90 percent of all pancreatic cancer cases, and 95 percent according to some sources. PDAC most commonly occurs in the head of the gland. Although this disease is not in the top ranks in terms of the frequency of its occurrence, it is in the top ranks in terms of lethality [18, 24].

USA ranks third in terms of cancer deaths. It ranks 10th among cancer types in terms of its incidence. The pancreasis an organ that has a very important place in metabolic and digestive activities, so sudden weight loss, loss of appetiteand fatigue are observed among the symptoms of cancer. Some vital body activities may be disrupted as a result of the endocrine and exocrine glands in the pancreas being affected by the tumor. The part of the pancreas where PDAC occurs most is the pancreatic head, with a rate of about 66 percent. The most basic reason for the lethality of pancreatic cancer is its insidious progression, the fact that the patients whoare diagnosed generally show metastatic symptoms, cause the treatment to be applied against the disease to start late andthe death rate to increase. Computed Tomography (CT), one of the medical imaging techniques, is a method frequently used in PDAC diagnosis. The detection sensitivity of the method is high, but the technique used and the experience of the radiologists are also important factors. One study found that 19 percent of cases were misinformed by medical professionals that they did not have cancer, despite having the disease as a result of the images and examinations obtained. PDAC is a disease whose lethality can be reduced if detected early and the treatment process can be applied more successfully [22].

In recent years, deep learning applications have played a pioneering role in the field of medical imaging, as in other fields. Disease diagnoses made by domain experts carry risk factors in many ways. The necessity of supporting health systems with automated systems due to reasons such as increasingpopulation and heavy workload has become quite noticeablein recent years. Since pancreatic cancer is a type of cancer that progresses insidiously and causes damage very quickly, its early diagnosis is very important. The aforementioned negative factors cause the detection of the disease in the general population to be detected at a slower and later stages. Deep learning systems are of great importance in the field of medical imaging and disease detection; Therefore, the use of such methods is critical to increase the accuracy of disease detection and to make health systems work faster and more efficiently [7, 31].

Chen *et al*. [6] recently published a study on pancreatic cancer detection from CT images with DL models. They applied a CNN model to detect the tumors from manually segmented images and achieved about 97% accuracy with their Computer Aided Design (CAD) tool. The CAD tool also evaluated the tumor size and stage of cancer with 50% to 93% sensitivity for various stages and sizes. Another similar automated system for pancreatic tumor classification from CT images with Deep Learning (DL) was introduced in [28]. The images were preprocessed with Gabor filtering and an Emperor Penguin Optimizer algorithm with multilevel thresholding was used for segmentation. After extracting the features with MobileNet model, an Auto Encoder classified the images as tumor or non-tumor images with more than 99% accuracy.

A complete pancreatic tumor diagnosis system was pro- posed in [25]. A ResNet model was used to locate the pancreatic tumor, then a U-Net model was used for pancreas segmentation and finally another ResNet was applied for the final classification. The model achieved 89% to 99% accuracy, but the computation cost was increased due to different DL models for different phases of the framework. Alves *et al*. [2] proposed a similar network with U-Nets for Pancreatic tumor detection. The input CT images were cropped to focus on Region of Interest (ROI) and different U-Net models were implemented to segment the pancreas and tumor with other nearby tissues. Most of the existing DL model-based Pancreatic tumor or cancer detection frameworks achieved high accuracy with some limitations like high computation cost, model overfitting or underfitting, etc. in static systems. The lack of user-friendly web applications with high performance DL models to identify pancreatic cancer from medical images in existing literature motivated the research of this paper.

In this paper, a DL based classification model is proposed for detecting the cancerous CT images from healthy CT images with a dataset of pancreatic ductal adenocarcinoma CT images containing 5968 cancerous and non-cancerous/healthy images. A web application is developed to upload and process CT images for cancer detection with a Convolutional Neural Networks (CNN). The framework pre-processes the images with standard image pre-processing, data augmentation and then the CNN model extracts the features from the image to classify it into one of two classes (i.e., cancerous or non- cancerous). The proposed model achieved about 96% detection accuracy with 97% precision without overfitting or under- fitting the DL model. The major contribution of this paper is developing a web application for medical professionals to detect pancreatic cancer from CT images with an efficient, accurate and precise DL model. Three DL models namely VGG-16, ResNet and custom CNN are applied on 5968 publicly available CT images for pancreatic cancer detection. Basic pre-processing steps are used on the collected data to remove noises and enhance the image features and then the three DL models are applied to classify the images into cancer and non-cancer classes.

The rest of the paper organized as follows-backgrounds on medical imaging, CNN models, hyperparameters and implementation components are discussed in the next section. The system architecture is explained in the Methodology section and the implementation setup and experimental results are mentioned in the Results section. Finally, the contributions, discussions and possible future research directions are discussed in the Conclusions.

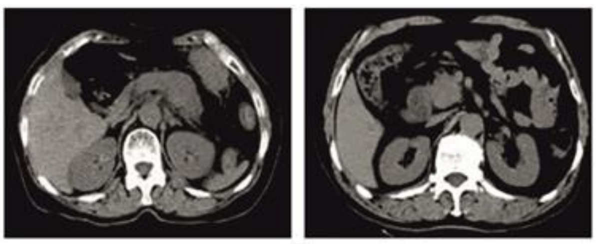
2. Background

2.1. Medical Imaging

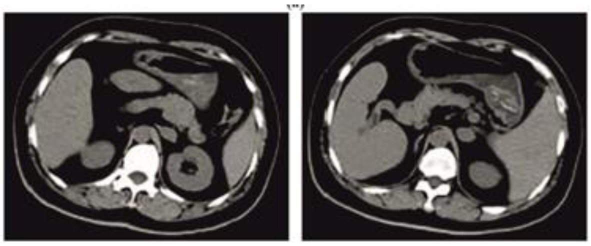
The health sector has different characteristics and requirements from other fields due to the fact that human life isthe primary priority. Medical imaging is also a field of health sciences. Thanks to medical imaging, it is possible to obtain images of the inner parts of the human body with various processes and methods, and to present medical analyzes and solutions in this way. From past to present, interpreting medical data is done by medical professionals. The human factor isimportant here as well as in other areas, some reasons such as the experience, knowledge and intense work schedule of the medical analysis specialist affect the accuracy of the analysis to be made [32].

Pancreatic cancer is a difficult disease in terms of early diagnosis, tumors found at an early stage during normal examinations may not be noticed by the relevant healthcare professional. In addition, this disease is more insidious inshowing symptoms than other types of cancer, most peopledo not show symptoms unless the cancer has grown too large or spread. Several medical imaging systems can be used for detection of pancreatic cancer, but CT is the initial medical imaging method for pancreas cancer, whereas another imaging technique used quite frequently is magnetic resonance imaging, [20].

CT is a non-invasive medical imaging method used in imaging the human body, with this method it is possible to obtain images of body sections. In addition to detecting pancreatic cancer with CT, it can also be determined whether the cancer has metastasized. CT is a technique that examines the body with X-rays in the form of thin slices (3-10 mm). In this way,it helps us to recognize the diseases of the organs in ourbody. The speed of multi-section CT devices offered by the developing technology has increased, and the section thickness they can take has decreased. During the examination, the patient should lie still on the computerized tomography table. As the table moves towards the middle section of the device, multiple cross-sectional images are taken by the device todisplay the relevant body section. There is no application that will cause pain or a feeling of pain during the examination. In abdominal CT examinations, medicated water, which is usually drunk orally, is provided to fill all the intestines with dyed water, making it easier to distinguish the masses in the intestine or its wall from other tissues. By dyeing the blood with the drug, which is given quickly by a vein pump, the condition of the veins, the relationship of the mass and the veins, the blood supply characteristics of the mass and some masses are made visible in the same examination [9, 16]. Figure 1 shows examples of CT images with and without pancreatic tumor.



a) Pancreatic tumor.



b) Non-tumor.

Figure 1. Sample CT images of [6].

2.2. Convolutional Neural Networks

Deep Learning is an approach which uses a dataset to predict outcomes. AI systems may be trained using supervised and unsupervised learning methods. Deep learning generally has 3 or more layers, we can think of this technique as a neural network, together with these networks, it is aimed to simulate the behavior of the human brain through this model. DL generally requires huge amounts of data. Optimization and improvement of the predictions to be made with the layers between the first and last layer, which wecall the hidden layer, are made. DL, which has many application areas, comes to the fore in areas such as smart life items, self-driving car and is widely used. The difference between Deep Learning and machine learning is that while ML uses labeled data to make predictions, thisis not the case for Deep Learning. DL includes many different concepts and parameters. When it comes to deep learning, it is not possible to understand Deep Learning without understanding the concept of neural networks. A DeepLearning model is a model that includes concepts such as neural networks, biases, and weights in general, and tasks suchas recognition and classifying are fulfilled with the use of thesemathematical models [15].

CNN, which is a DL architecture, is a strong candidate for image-related problems, and the contributionof CNN models to the increasing interest in deep learningin recent years is undeniable. AlexNet, which was revealed about 10 years ago, caused an increase in the interest in this model. In the following periods, there was a great increasein the number of layers used. It also has a huge effect forthe domains like recommendation systems, natural language processing. Compared to previous similar systems, CNN contains an important feature, it does not need human supervision in feature determination process. CNN, which is a model computationally efficient, performs parameter sharing, also uses special convolution and pooling algorithms. CNN modelscan now be used on any platform, making it more attractiveto a wider audience [19].

Although CNN has many basic building blocks, the importance of convolution layers is one of the most important elements to understand this structure. In terms of mathematicalmeaning, this process summarizes the combination of two existing datasets, and in deep learning applications, it is a concept for the dataset given to the model to be subjected to this mathematical processing with the use of the convolution filter. Pooling is usually done after a convolution operation to reduce dimensionality. This shortens the training time [8].

To briefly summarize the general operation, DL layers are built on top of each other and calculations proceed over this network. The input layer is the layer where the input to be processed is received, and the output layer is the layer where the estimation or different tasks reach the end [3].

2.3. Hyperparameter Tuning

While designing Machine Learning and Deep Learning models, the selection of the algorithm used in the model and the hyperparameters of these models are also important factors that will affect the success of the project [30]. These hyperparameters differ according to the types of algorithms, while the k value is a hyperparameter in the KNN algorithm, the type of kernel function is a hyperparameter in the Support Vector Machine (SVM) algorithm. There are hyperparameters such as dropout, number of layers, number of neurons in deep learning models andthey directly affect the success of the project. The selection of these hyperparameters is not clear at the very beginning of the project, these hyperparameters can be changed by looking at the preliminary results obtained in the later stages of the project, as well as the characteristics such as the type and requirements of the problem, the size and complexity of the data set. These hyperparameters are determined by considering the mentioned elements and it is up to the person who made this project. The high performance of the model does not only depend on a single combination of hyperparameters, more than one group of hyperparameters can provide high performance, these different groups can be used in designing the model [14]. Selecting the most appropriate hyperparameter by looking at the requirements of the project is one of themost important factors that will affect the success rate of the project and should be paid attention to. In addition to the technical factors involved in the selection of hyperparameters,there are designer-based critical stages such as the intuitionof the designer who will create the model, the experiencegained from the problems he has encountered before, and the reflection of the problems in different fields on the target problem. Although these are general methods, there are some techniques to find the optimal hyperparameter structure. Deep learning is a form of learning that requires high dataset size and diversity. The high learning rate is directly proportionalto the size of the dataset in general, but the training time and the size of the model will also increase with the large dataset. These factors should be taken into account in projects where elements such as storage space and the number of trainingsare intense. However, this situation can be ignored in cases where the number of trainings will not be very frequent and storage problems will not be experienced. In addition to the large data set, the diversity of the data is one of the factors thatincrease the success of the model [1]. With this information, the huge increase in the number of data does not increase the success rate of the project at the same rate, in general,the success rate increases little by little at certain rates. Ifthe similarity rate of the classes in the dataset is high and perhaps the number of noisy data is high, the performance graph will have a bumpy structure. Batch value change can be used to solve this problem without going the data exchange route. In addition, data can be augmented synthetically, thisis called data augmentation. As the increase in the number of data will increase the success rate as mentioned above, it will be useful to apply this technique. Deep learning applications can be costly in terms of time. In order to overcome thisproblem, the mini-batch parameter can be used, which means: According to the assigned mini-batch parameter, the numberof data that the model will process at the same time can be determined, and in this way, time can be saved. While the model is being trained, not all of the data is included in the training at the same time, the first piece is trained first and the weights are updated with the backpropagation after the model performance is tested [5]. Then the model is retrained with the new training set and the weights are updated again. This process is repeated at each training step to calculate the most appropriate weight values for the model. Each of these training steps is called an "epoch". Since these weight values are calculated step by step, the success rate will be low in the first epochs, and the success rate will increase in the epochs that progress with the backpropagation processes. The model usually takes a long time to train; there are models that take days or months to train. This is common in deep learning. For this reason, it is tried to shorten the training process as much as possible with other hyper parameters. As the number of epochs increases, the performance increases. The numberof epochs depends on the type of problem and requirements. Training can be terminated at these points, as performance will increase in very small units after a certain epoch [21].

Activation functions have a very important place in deep learning models. These functions add non-linearity to themodels. The linear function in the hidden layers turns into a non-linear value with matrix multiplication. The reason for this is that the deep learning method is better than other methods in nonlinear problems, and real world examples are generally non-linear problems. The conversion of the value obtained asa result of matrix multiplication to non-linear is done with activation functions. Some activation functions used in deep learning models are: sigmoid, tanh, ReLu, PreLu [11, 29].

Dropout is a regularization technique which is utilizable when overfitting occurs. It has been observed that forgetting weak information increases learning [4].

2.4. Environments, Libraries and Frameworks

It is possible to say that Artificial Intelligence applications have become more mobile with the devices being equipped with much more powerful elements. AI has become a phenomenon that we see more frequently, and the number ofapplications currently developed is expected to increase inthe coming years [27]. Google Colab was utilized in this project because of the free GPU support it offers. One ofthe main libraries used in this project is TensorFlow. As an open source library, Tensorflow facilitates the implementationof AI branches such as Deep Learning and Machine Learning,and supports this with its flexible architecture and GPU support [26].

2.5. Pancreatic Tumor Image Analysis

Researchers have been experimenting with different machine learning and deep learning models on medical images for pancreatic tumor detection to improve the performance of the classification task. Gupta *et al*. [10] provided a review on pancreatic cancer with the types, stages and severity of pancreatic cancers and summarized the machine learning and deep learning models used in pancreatic cancer detection. They included the techniques, datasets, performances, limitations of artificial neural network, Bayesian model, random forest model and genetic algorithms used in different research for pancreatic cancer detection and showed the comparisons between the performances (i.e., AUC, specificity and sensitivity) of the existing researches. A CNN based pancreatic cancer detection and segmentation system was proposed in [6] and the model was tested on real data collected locally and nationwide in Taiwan. They used an end-to-end CNN model to segment the pancreas and tumor and then applied an ensemble model containing five CNN models to decide the cancer or non-cancer class. The outputs were validated by professional radiologists and their detection achieved 91% to 98% accuracy for the two local and one nationwide dataset. The sensitivity scores of the proposed system and radiologists showed 1 to 6 score differences proving the performance accuracies of their system. Ma *et al*. [17] also proposed a CNN based pancreatic cancer detection model. They created three datasets from 3494 CT images based on their phases and implemented a binary classifier to detect cancer and non-cancer classes, and a ternary classifier to detect non-cancer, cancer at tail or body, and cancer at head or neck of pancreas classes. They used a simple CNN with three convolution layers each followed by batch normalization and max pooling and added a fully connected layer for output generation. The binary classifier and ternary classifier achieved 95% and 82% accuracy respectively. Another CNN based pancreatic cancer detection model was provided in [23] that used Gaussian Mixture Model (GMM) and Expectation-Maximization (EM) model for feature extraction. They lump feature extraction algorithm to observe features in the ROIs using size, width, depth and shape scores and created a Region of Interest Database (RID). The extracted features with GMM and EM were stored into another database called Feature Database (FD). Then a basic CNN model was applied for classification and tumor spread. Their proposed model provided a future research scope to specify and track the tumor spread in continuous evaluation of patients.

Vaiyapuri *et al*. [28] proposed an intelligent deep-learning-enabled decision-making medical system for pancreatic tumor classification called IDLDMS-PTC to detect pancreatic tumor from CT images. The CT images were pre-processed with Gabor filtering technique and the emperor penguin optimization with multilevel thresholding was used to segment the pancreatic tumor. The image features were then extracted with a MobileNet algorithm, and an auto encoder was applied for the classification task on the extracted features. A multileader optimization method was used for fine tuning the classifier. A dataset with 500 CT images (250 tumorous, and 250 healthy) was used to train and test the proposed model and gained almost 99% accuracy outperforming similar recent research efforts. Although the proposed model was applied on a smaller dataset, the high performance provided the possibility of exploring the model more with benchmark datasets to test the reliability of the detection outputs.

A pancreatic tumor detection model with Feature Pyramid Networks (FPN) and R-CNN was proposed in [33] for enhancing the feature extraction process to improve the detection of the tumor. A pre-trained ResNet-101 was used for feature extraction from the input CT images to create a feature pyramid. Then a bottom0up approach was applied for feature hierarchy enhancement and finally a Region Proposal Network (RPN) was applied on the enhanced feature pyramid for self-adaptive feature fusion to extract multi-level information from enlarged ROIs. 2890 CT images collected from the affiliated hospital of Qingdao University were used for training and testing the proposed model. The approach was able to classify tumor images with 90% accuracy outperforming R-CNN, mask R-CNN, cascaded R-CNN, YOLO, DetNet and RetinaNet by almost 15%. The comparison between applying the proposed feature enhancement and basic detection classifier showed about 15% improvement for the proposed model due to the pre-processing for feature extraction and enhancement to generate more detailed information on the ROIs. Hussein *et al*. [13] proposed a novel supervised and unsupervised 3D CNN model for both lung and pancreatic tumor detections from CT and MR images. They provided two models- one supervised module with CNN and Graph Regularized Sparse multitask learning and an unsupervised module with clustering and proportion-SVM and trained both modules for lung nodules characterization and Intraductal Papillary Mucinous Neoplasms (IPMN) cysts. The supervised module provided a malignancy score based on the input images classified with the multitask model that used the outputs generated through multiple 3D CNN feature extractions. The unsupervised module on the other hand applied feature extraction using the clustering model to generate initial labels for the images and then after applying label proportions, the images were finally classified by a SVM model. The supervised module achieved 91% accuracy whereas the unsupervised module provided 58% accuracy for IPMN classification and 78% accuracy for lung nodule detection.

3. Methodology

The methodology employed in this project involves three key steps: Data Acquisition, Data Pre-processing, and Model Training. Firstly, the Data Acquisition phase involved obtaining the relevant data required for the project. Subsequently,the Data Pre-processing step was executed, which proved crucial in ensuring that the training process was carried out efficiently. This involved converting the data from DICOM format to JPEG format, resizing the input images to 224x224 pixels, and utilizing the Keras preprocessing module to apply data augmentation techniques. The third step, Model Training,involved training the VGG-16, Resnet, and Custom CNN models, along with hyperparameter tuning. During this stage, the accuracy and loss values of both the training and validationsets were monitored to detect any potential overfitting or underfitting issues. It is important to note that the steps per epoch value is detected automatically based on the number of augmented data by Keras, which was 17, and the batch size was 256 for training dataset. Finally, the Custom CNN and pre-trained models were utilized on the test set to evaluate their effectiveness in addressing the project's objectives.

3.1. Data Acquisition

The ultimate goal in this project was to detect pancreatic tumor with high accuracy. For this, first of all, a suitabledata set must be found. The primary purpose is to use the public resource for reasons arising from Privacy Issues, there are Pancreatic Cancer CT Images shared publicly on many sites on the internet. CT images were taken from public data sources were used in this project. The dataset includes 2985 non-cancerous and 2983 cancerous pancreatic CT images. In order to avoid the imbalanced dataset problem, almost equal images are used for each class.

3.2. Data Pre-Processing

In order to efficiently carry out the training process, Data Pre-processing is a crucial step in deep learning applications. To optimize the efficiency of the model training process, the CT data was first converted from DICOM format to JPEG format, which allowed for reduced file sizes. In addition, the input images were resized to 224x224 pixels, which expeditedthe model training process. This class enabled the application of various data augmentation techniques, such as rotation, width and height shifting, shearing, and zooming, to the training data. Moreover, horizontal flipping was applied to the images in the ImageDataGenerator to artificially increase the size of the training set. Additionally, rescaling was per- formed to reduce the range of pixel intensity values in the images. The preprocessed training data was obtained usingthe ImageDataGenerator class, which was used to convert the images into batches. The target size of the images wasset to (224, 224) and the batch size was set to 256. The obtained batch data was then used for training the deep learning models. Similarly, the test and validation data were preprocessed using the ImageDataGenerator class with a targetsize of (224, 224) and a batch size of 32. It is noteworthy that the shuffle parameter was set to True for the training and validation data, which randomizes the order of the imagesin each epoch. Moreover, the seed parameter was set to 42 to ensure reproducibility of the results. The class indices ofthe image batches were printed to confirm the mapping ofthe class labels to the corresponding numerical values. The preprocessed training, test, and validation data was then used to train and evaluate the VGG-16, Resnet, and Custom CNN models, and hyperparameter tuning was carried out to improvethe performance of the models.

3.3. Model Training

The dataset used in this project was split into three sets instead of the conventional two, as this partitioning resultedin better outcomes. Specifically, the dataset was separated into 70 percent training, 15 percent validation, and 15 percenttesting sets. Basic modeling experiments were carried out on the training set, with the validation set being used for model selection and refinement. To this end, hyperparameter tuning was employed to determine the optimal coefficients/weights for the chosen model. To evaluate the model's performance, the validation accuracy and loss values were graphically compared with the accuracy and loss values. This comparison allowed for the assessment of whether overfitting occurred during the training process. This procedure was followed for all the models which are used for this project.

- *Custom CNN Model:* the model is instantiated by creating a Sequential object from the Keras API. The Sequential object is utilized as a linear stack of layers, which are addedto in order to build the CNN. Firstly, a Conv2D layer isadded with 4 filters, a kernel size of 3x3, and a ReLU activation function. The input has the shape of (224, 224, 3),representing 224x224 RGB images. A MaxPooling2D layeris then added with a pool size of 2x2 to downsample thefeature maps and reduce the spatial dimensions of the output. Next, another Conv2D layer with 8 filters and a kernel size of 3x3 is added, followed by another

MaxPooling2D layer. This process is repeated with a Conv2D layer with 16 filters and a kernel size of 3x3, followed by another MaxPooling2D layer. Afterwards, a Dropout layer with a dropout rate of 0.7 is added to prevent overfitting. The Dropout layer randomly drops some of the connections between neurons during training, forcing the network to learn more robust features. The output of the last MaxPooling2D layer is flattened, and a dense layer with 64 units and a ReLU activation function is added. Another Dropout layer is then added with a dropout rate of 0.8 to further prevent overfitting. Finally, a dense layer with a singleunit and a sigmoid activation function is added. This will output a probability between 0 and 1 indicating the likelihood that the input image belongs to a certain class. The model is then compiled using the binary crossentropy loss function, the Adam optimizer, and the accuracy metric.

- *VGG-16 Model:* the pre-trained VGG-16 model was used in the transfer learning process using the Keras API. The include top parameter is set to False to not include the upper layers of the VGG-16 model, which is a pre-trained model. The weights that were previously trained on the imagenet dataset were started, and they were also used to monitor the validation loss values in the early stopping callback training process and to prevent overfitting. To avoid retraining the pre-trained layers of VGG-16, a pre-trained model, they werefrozen with an iterative loop on each layer of this model.Custom layer addition was made on top of the pre-trained model, the first layer was the flattened layer, the dense layer with ReLU activation function was installed after the flatten layer. To find the optimal hyperparameters, Optuna is used,it was decided to assign the dropout layer with a value of 0.8. Two fully connected layer is used in this model, firstone is 32 neurons. After the normalization layer was added,a dense output layer, which is the second fully connected layer, using the sigmoid activation function was added to the architecture, this *was* done in the name of binary classification. The resulting transfer learning model was compiled with the binary-cross entropy function after these stages, using the optimum learning rate value found as 5.194856185393038e- 05 as a result of grid-search. Evaluation metrics are used to monitor the performance of the model, in this process accuracy was used as an evaluation metric. This model provided atraining process with a high success rate by utilizing pre-trained weights using the transfer learning technique.
- *ResNet:* ResNet50 is a very popular pre-trained deep learning model. ResNet, one of the models used in this project, can show very high success rates in image classification processes. In this project, the pre-trained layers were frozenin the first stage, then custom layers were added and adjusted to be successful in PDAC detection. Custom layer addition was made on top of the pre-trained model, the first layer was the flattened layer, the dense layer with ReLU activation function was installed after the flatten layer. Custom layers consisting of two fully connected layers, one is with 133 neurons whereas the other has 47 neurons, and dropout layer is added with the rate of 0.17359436988893012. The learning rate for compiling the transfer learning model with Adam optimizeris set to 0.0008158991855105919. The hyperparameters are determined by using Optuna, which is a very useful tool tofind optimal hyperparameters in Machine Learning and Deep Learning applications.

4. Result

When we look at the literature, the general accuracy value ofpancreatic cancer detection models is around 80 percent, and the data set sizes are generally smaller. The aim of this projectwas to increase this accuracy value and not to encounter unwanted situations such as overfitting and underfitting. In this project, 5968 CT images were used, this dataset was divided into training, validation and test set. The distribution of datasetinto these sets is as follows: 70 percent for Training, 15 percent for Test and 15 percent for Validation Set. In all models, tables comparing the training and validation accuracy and lossvalues were obtained and observed. Also, to observe overfitting and underfitting issues training and validation accuracy and loss values are visualized. Since there was no overfitting or underfitting issues, Confusion matrix for all of these three models is obtained and then recall, precision and F1 scorevalues are declared.

4.1. Custom CNN

As it can be seen from Figure 2, the training and validation accuracy is obtained for Custom CNN model in5 epochs. The epoch number is determined after finding the optimal one which does not cause overfitting and underfitting issues and also gives good accuracy values.

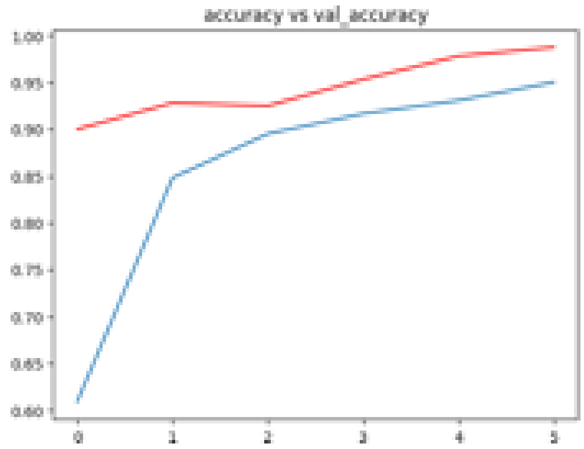


Figure 2. Training and validation accuracy plot of the custom CNN model.

In Figure 3, confusion matrix which was obtained aftertesting the model can be seen. Confusion matrix is used to evaluate the model in a classification process, the results are evaluated in 4 main parts: true positive, true negative, false positive and false negative. After testing the model, actual labels and predicted labels are compared. This provides the ability to calculate evaluation metrics. It can be seen from this figure that the model gives an accuracy value which is relatively high, but some mispredictions are made on positive and negative class. The TP predictions are 439 whereas TN value is 430, model gives false negative values which are indicating the actual label was positive but by the model these samples are predicted as negative, which is unwanted situationfor cancer prediction tasks, since it is wanted that the model toaccurately label positive-label cancerous images with as muchaccuracy as possible.

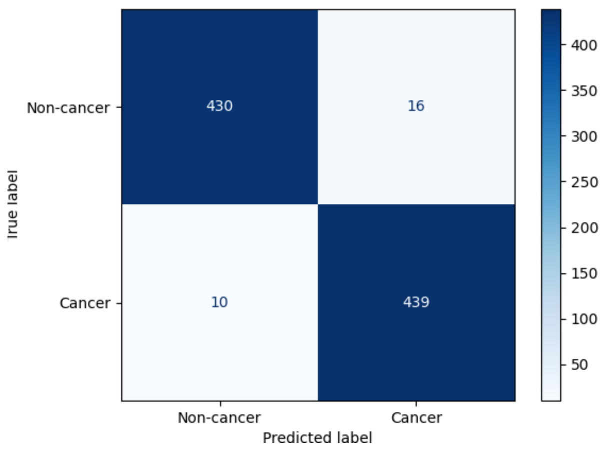


Figure 3. Confusion matrix of the custom CNN model.

4.2. VGG-16

As it can be seen from Figure 4, the training and validation accuracy is obtained for VGG-16 model which is trained for 5 epochs. The epoch number is determined after finding the optimal one which is not causing overfitting and underfitting issues and also gives good accuracy values. The steps per epochnumber is determined automatically from Keras API sinceData Augmentation was used as a method to increase the dataset artificially. When we look at the figure, the training and validation accuracy values show good performance, as the number of epochs increases, the model learns better, and the values increase. At the time the model started training, the validation accuracy was relatively high compared to the training accuracy, but then during the model learning process, by learning better in each epoch, the model learns better. Training accuracy was 0.9592, and the lowest loss was 0.13. Training and validation loss values were also observed, and it was observed that it started from 0.7 levels and decreased to 0.1 levels in both. Considering all these, it can be said in general that the model does not encounter overfitting problems, since the difference between the training and validation accuracy and loss values is not large, the model learns more in each epoch, and the loss values decrease in general.

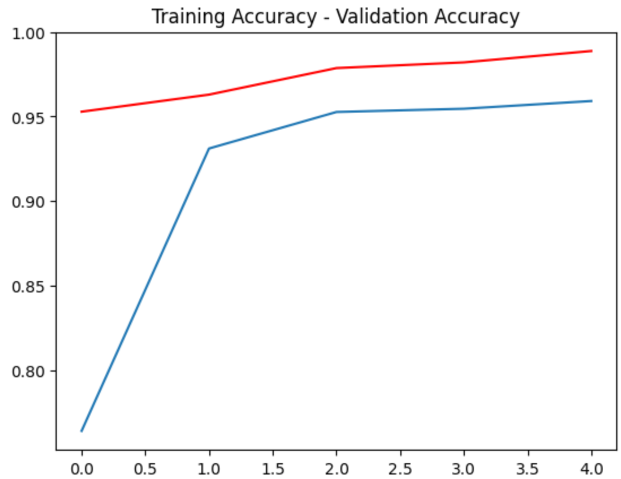


Figure 4. Training and Validation Accuracy Plot of the VGG-16 Model.

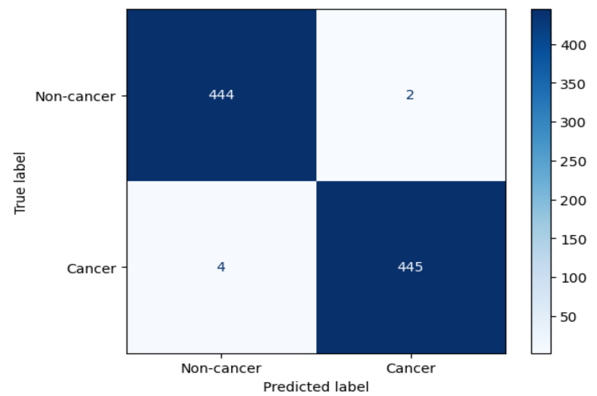


Figure 5. Confusion matrix of the VGG-16 model.

In Figure 5, confusion matrix was obtained after testing the VGG-16 model. As it can be seen from the figure, prediction made by VGG-16 model is fairly accurate, and the mispredictions are made so infrequently.

4.3. ResNet

The training and validation accuracy for the ResNet-16 model, which is trained for 5 epochs, is shown in figure 6. Theepoch number is established after determining the ideal one that does not cause overfitting or underfitting concerns while still providing high accuracy values. Since Data Augmentationwas employed to artificially augment the dataset, the number of steps per epoch is chosen automatically using the Keras API. The training and validation accuracy values in the figure demonstrate strong performance; as the number of epochsgrows, the model learns better and the values improve. The validation accuracy was relatively high before the model began training compared to the training accuracy, however the model learns better during the model learning process by learning better in each epoch. The discrepancy in validation accuracy has been reduced. The training accuracy was 0.9748, with a loss of 0.0867. Training and validation loss values were also measured, and it was discovered that the trend and the valueof these are almost same. Taking all of this into account, itis safe to say that the model does not suffer from overfitting because the gap between training and validation accuracy and loss values is small, the model learns more in each epoch, and the loss values decrease in general.

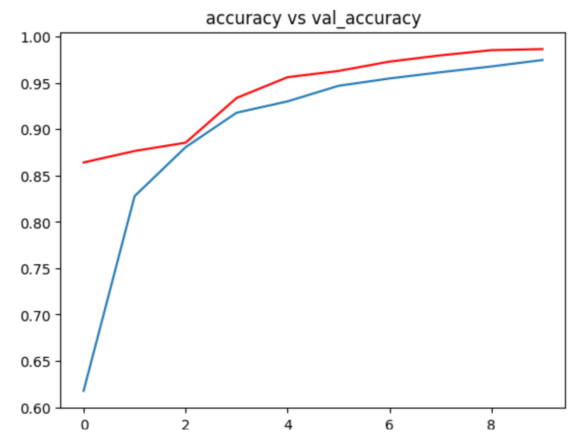


Figure 6. Training and Validation Accuracy Plot of the ResNet Model

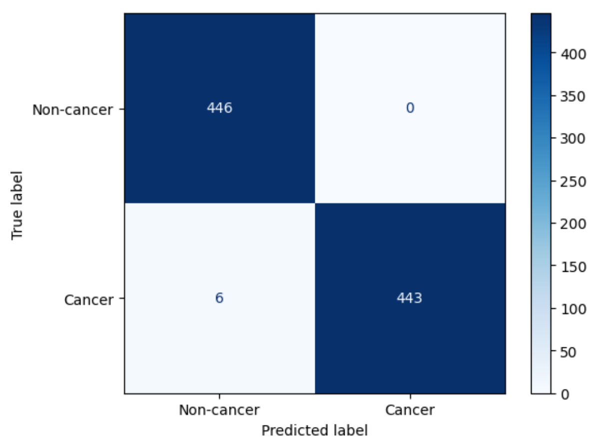


Figure 7. Confusion matrix of the ResNet model.

The confusion matrix in Figure 7 was derived after testing the ResNet model. As seen in the graph, Resnet model predictions are fairly accurate, with mispredictions occurring very seldom.

Although the most commonly used evaluation metric is accuracy, there are other metrics that should be specifically examined in classification models, such as precision, recall, and F1 score. The formulas used in the calculations of these metrics can be observed below.

$$Accuracy\ TP + TN / (TP + TN + FP + FN) \tag{1}$$

Accuracy is an evaluation metric used to measure the correct classification frequency of an ML or Deep Learning classification model. The formula of accuracy can be seen in Equation (1).

$$Precision = TP / (TP + FP) \tag{2}$$

Precision is an evaluation metric based on the ratio of actual positives to all positives when evaluating samples of a model. The formula of precision can be seen in Equation (2).

$$Recall = TP / (TP + FN) \tag{3}$$

Recall is an evaluation metric that shows how well the generated model can detect positive samples. The formula of recall can be seen in Equation (3).

$$F\,1Score = 2\ Precision\ \ Recall/(Precision + Recall) \tag{4}$$

F1 Score is an evaluation metric that gives the harmonic meanof the recall and precision values and takes the weight of thesetwo metrics the same. The calculation of F1 Score can be seenin Equation (4)

According to the formulas given above, each evaluationmetric value can be seen in the table below for all the models.

Table 1. Evaluation metric scores.

	Model		
	Custom CNN	**VGG-16**	**ResNet**
Accuracy	0.9709	0.9933	0.9933
Precision	0.964	0.9955	1.0000
Recall	0.977	0.9911	0.9866
F1 Score	0.9712	0.9933	0.9933

Table 1 shows the evaluation metric scores for the three different models used in this project. Precision, Recall, F1 Score and Accuracy are the evaluation metrics used to evaluate the success of this project. Starting with the Accuracy values, the custom CNN model has an accuracy value of 0.9709, besides, the accuracy values of the VGG-16 and ResNet models are the same, showing 99.33 percent accuracy, these two models made a better prediction than the custom CNN model. The classifications were made more accurately by these two pre-trained models and at this point they became more preferable than the custom model. When we look atthe precision values, the custom CNN model got a score of 0.964, this precision value shows us that 96.4 percent of the data that the model labeled as positive label after the testing period had positive actual labels. The precision score of the VGG-16 model was obtained as 0.9955, indicating that it is more successful at this point than the non-pre-trained model.It is worth emphasizing that the precision score of the ResNet model was obtained as 1.0, which indicates that all samples that the model has labeled positively also have true labels,at which point ResNet performed well. We can say that the recall metric is a metric that is essential to be used in the detection of diseases such as cancer and is more important thanrelative precision, the reason for this is the meaning underlying the recall metric. Recall is the ratio of the samples that the model predicts positively to the actual number of samples that are positively labeled. A more detailed explanation may be required at this point, to exemplify the situation, the fact thata patient has PDAC despite being diagnosed as not having cancer is the most undesirable situation due to the nature of thedisease. At this point, with the recall score, it can be seen how far away from these unwanted situations is. The Custom CNN model achieved a precision score of 0.977, while the VGG-16 achieved a precision score of 0.9911 and the ResNet model achieved a precision score of 0.9866. At this point, it would not be wrong to say that VGG-16 made the most successful prediction. Finally, the F1 score of all models was obtained, another metric obtained by taking the harmonic average of theF1 score Recall and Precision metrics. At this point, ResNet and VGG-16 achieved an F1 score of 0.9933, outperforming the custom CNN model on this metric basis. In summary, whenthe table is examined, we see that the VGG-16 and ResNet models achieve very high metric values. Although the CustomCNN model also has relatively high values, it can be said that it is more unsuccessful compared to these two models.

5. Conclusions and Future Work

This paper introduces three CNN models that are trained to detect pancreatic ductal adenocarcinoma with high accuracy and in an automated and much faster way than traditional healthcare systems. In this project, a two-class dataset including pancreatic ductal adenocarcinoma, that is, cancerous and healthy, was used, a special CNN architecture was designed,as well as transfer learning was performed using VGG-16and ResNet pre-trained models. A general CNN architecture that has proven itself in image-related problems and givesgood results has evaluated CT images with at least 95 percent accuracy in each model. During this project process, recent methods and projects with the same purpose were frequently examined in the literature, and the data set was obtained from public sources on the internet, thus avoiding privacy issues. The generalizable aim of the project is to create a custom model, to obtain accuracy and other evaluation metrics, and tocompare the performance rate and other metrics of pre-trained model-based models by using the transfer learning technique to be a source for future research in this area and to contribute to the literature.

This project has tested how different models, more specifically, three different architectures and two separate methods perform on Pancreatic cancer detection on a given set of tests, but for future studies this test can be extended or how models perform with the use of different test sets. Observable. Ultimately, this system can be turned into an end-to-end application in order to reduce the workload in healthcare systems and help radiologists at this point, and how it performsin real life scenarios can be monitored more professionally andaccurately.

References

[1] Albawi S., Mohammed T., and Al-Zawi S., "Understanding of a Convolutional Neural Network," *in Proceedings of International Conference on Engineering and Technology*, Antalya, pp. 1-6, 2017. DOI: 10.1109/ICEngTechnol.2017.8308186.

[2] Alves N., Schuurmans M., Litjens G., Bosma J. S., Hermans J. and Huisman H., "Fully Automatic Deep Learning Framework for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography," *Cancers*, vol. 14, no. 2, pp. 376, 2022. https://doi.org/10.3390/cancers14020376

[3] Baraniuk R., Donoho D., and Gavish M., "The Science of Deep Learning," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48. pp. 30029-30032, 2020. DOI: https://doi.org/10.1073/pnas.202059611.

[4] Cai S., Shu Y., Chen G., Ooi B., Wang W., and Zhang M., "Effective and Efficient Dropout for Deep Convolutional Neural Networks," *arXiv preprint arXiv:1904.03392*, 2019. https://doi.org/10.48550/arXiv.1904.03392

[5] Cardoza I., García-Vázquez J., Díaz-Ramírez A., and Quintero-Rosas V., "Convolutional Neural Networks Hyperparameter Tunning for Classifying Firearms on Images," *Applied Artificial Intelligence*, vol. 36, no. 1, 2022. https://doi.org/10.1080/08839514.2022.2058165.

[6] Chen P., Wu T., Wang P., Chang D., Liu K., Wu M., Roth H., Lee P., Liao W., and Wang W., "Pancreatic Cancer Detection on CT Scans with Deep Learning: A Nationwide Population-Based Study," *Radiology*, vol. 306, no. 1, pp.172-182, 2023. https://doi.org/10.1148/radiol.220152.

[7] England J. and Cheng P., "Artificial Intelligence for Medical Image Analysis: A Guide for Authors and Reviewers," *American Journal of Roentgenology*, vol. 212, no. 3, pp. 513-519, 2019. doi: 10.2214/ajr.18.20490.

[8] Erickson B.J., Korfiatis P., Akkus Z., and Kline T., "Machine Learning for Medical Imaging," *RadioGraphics*, vol. 37, no. 2, pp. 505-515, 2017. DOI: https://doi.org/10.1148/rg.2017160130.

[9] Francis I., "Role of CT and MR in Detection and Staging of Pancreatic Adenocarcinoma," *Cancer Imaging*, vol. 4, no. 1, pp. 10-14, 2003. DOI: 10.1102/1470-7330.2003.0026.

[10] Gupta A., Koul A., and Kumar Y., "Pancreatic Cancer Detection Using Machine and Deep Learning Techniques," *in Proceedings of the 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Gautam Buddha Nagar, pp. 151-155, 2022. DOI: 10.1109/ICIPTM54933.2022.9754010.

[11] Hao W., Yizhou W., Yaqin L., and Zhili S., "The Role of Activation Function in CNN," *in Proceedings of 2nd International Conference on Information Technology and Computer Application*, Guangzhou, pp. 429-432, 2020. DOI: 10.1109/ITCA52113.2020.00096.

[12] Hassanpour S. and Dehghani M., "Review of Cancer from Perspective of Molecular," *Journal of Cancer Research and Practice*, vol. 4, no. 4, pp. 127-129, 2017. https://doi.org/10.1016/j.jcrpr.2017.07.001.

[13] Hussein S., Kandel P., Bolan C., Wallace M. B. and Bagci U., "Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1777-1787, 2019. DOI: 10.1109/TMI.2019.2894349.

[14] Indolia S., Goswami A., Mishra S., and Asopa P., "Conceptual Understanding of Convolutional Neural Network-A Deep Learning Approach," *Procedia Computer Science*, vol. 132, pp. 679-688, 2018. https://doi.org/10.1016/j.procs.2018.05.069

[15] LeCun Y., Bengio Y., and Hinton G., "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015. https://doi.org/10.1038/nature14539.

[16] Lee J. and Yoon J., "Imaging Diagnosis of Pancreatic Cancer: CT and MRI," *Pancreatic Cancer. Springer Berlin Heidelberg*, pp. 95-114, 2017. DOI: 10.1007/978-3-662-47181-4_7.

[17] Ma H., Liu Z., Zhang J., Wu F., Xu C., Shen Z., Yu C., and Li Y., "Construction of A Convolutional Neural Network Classifier Developed by Computed Tomography Images for Pancreatic Cancer Diagnosis," *World Journal of Gastroenterology*, vol. 26, no. 34, pp. 5156-5168, 2020. doi: 10.3748/wjg.v26.i34.5156

[18] McGuigan A., Kelly P., Turkington R., Jones C., Coleman H., and McCain R., "Pancreatic Cancer: a Review of Clinical Diagnosis, Epidemiology, Treatment and Outcomes," *World Journal of Gastroenterology*, vol. 24, no. 43, pp. 4846-4861, 2018. doi: 10.3748/wjg.v24.i43.4846

[19] Mishra C. and Gupta D., "Deep Machine Learning and Neural Networks: An Overview," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 6, no. 2, pp. 66-73, 2017. DOI: 10.11591/ijai.v6.i2.pp66-73.

[20] Miura F., Takada T., Amano H., Yoshida M., Furui S., and Takeshita K., "Diagnosis of Pancreatic Cancer," *HPB*, vol. 8, no. 5, pp. 337-342, 2006. https://doi.org/10.1080/13651820500540949.

[21] Neary P., "Automatic Hyperparameter Tuning in Deep Convolutional Neural Networks Using Asynchronous Reinforcement Learning," *in Proceedings of IEEE International Conference on Cognitive Computing*, CA, pp. 73-77, 2018. DOI: 10.1109/ICCC.2018.00017.

[22] Orth M., Metzger P., Gerum S., Mayerle J., Schneider G., Belka C., Schnurr M., and Lauber K., "Pancreatic Ductal Adenocarcinoma: Biological Hallmarks, Current Status, and Future Perspectives of Combined Modality Treatment Approaches," *Radiation Oncology*, vol. 14, no. 1, pp. 1-20, 2019.

[23] Sekaran K., Chandana P., Krishna N., and Kadry S., "Deep Learning Convolutional Neural Network (CNN) with Gaussian Mixture Model for Predicting Pancreatic Cancer," *Multimedia Tools and Applications*, vol. 79, no. 15-16, pp. 10233-10247, 2020.

[24] Shah J., Surve S., and Turkar V., "Pancreatic Tumor Detection Using Image Processing," *Procedia Computer Science*, vol. 49, pp. 11-16, 2015. https://doi.org/10.1016/j.procs.2015.04.221.

[25] Si K., Xue Y., Yu X., Zhu X., Li Q., Gong W., Liang T., and Duan S., "Fully End-To-End Deep-Learning-Based Diagnosis of Pancreatic Tumors," *Theranostics*, vol. 11, no. 4, pp. 1982-1990, 2021. doi: 10.7150/thno.52508

[26] Singh A. and Bhadani R., *Mobile Deep Learning with Tensorflow Lite, ML Kit and Flutter: Build Scalable Real-World Projects to Implement End-To-End Neural Networks on Android and IOS*, Packt Publishing, 2020.

[27] Tecuci G., "Artificial Intelligence," *WIREs Computational Statistics*, vol. 4, no. 2, pp. 168-180, 2012. https://doi.org/10.1002/wics.200

[28] Vaiyapuri T., Dutta A., Punithavathi I., Duraipandy P., Alotaibi S., Alsolai H., Mohamed A., and Mahgoub H., "Intelligent Deep-Learning-Enabled Decision-Making Medical System for Pancreatic Tumor Classification on ct Images," *Healthcare*, vol. 10, no. 4, 2022. https://doi.org/10.3390/healthcare10040677.

[29] Wang Y., Li Y., Song Y., and Rong X., "The Influence of the Activation Function in a Convolution Neural Network Model of Facial Expression Recognition," *Applied Sciences*, vol. 10, no. 5, 2020. https://doi.org/10.3390/app10051897.

[30] Yamashita R., Nishio M., Do R., and Togashi K.,

"Convolutional Neural Networks: An Overview and Application in Radiology," *Insights into Imaging*, vol. 9, pp. 611-629, 2018. DOI: 10.1007/s13244-018-0639-9.

[31] Yoon H., Jeong Y., Kang H., Jeong J., and Kang D., "Medical Image Analysis Using Artificial Intelligence," *Progress in Medical Physics*, vol. 30, no. 2, pp. 49-58, 2019. DOI: https://doi.org/10.14316/pmp.2019.30.2.49.

[32] Zhang Y., Gorriz J., and Dong Z., "Deep Learning in Medical Image Analysis," *Journal of Imaging*, vol. 7, no. 4, 2021. https://doi.org/10.3390/jimaging7040074.

[33] Zhang Z., Li S., Wang Z., and Lu Y., "A Novel and Efficient Tumor Detection Framework for Pancreatic Cancer via CT Images," *in Proceedings of 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Montreal, pp. 1160-1164, 2020. DOI: 10.1109/EMBC44109.2020.9176172.

Muhammed Talha Zavalsiz completed his BSc in Biomedical Engineering at Istanbul Medipol University, Istanbul, Turkey. He is currently a graduate student in Computer Engineering at Istanbul Medipol University. His research interests include medical image analysis, data science, and machine learning.

Sleiman Alhajj is currently a student in the International School of Medicine at Istanbul Medipol University, Istanbul, Turkey. His research interests include cancer, infectious diseases, machine learning, social network analysis, and medical image analysis.

Kashfia Sailunaz completed her MSc in Computer Science at the University of Calgary, Alberta, Canada. She is currently a PhD candidate in Computer Science at the University of Calgary, Alberta, Canada. She received her BSc from Military Institute of Science and Technology, Dhaka, Bangladesh. She also worked as a research assistant in Advanced Intelligent Multidisciplinary Systems (AIMS) Lab, Dhaka, Bangladesh. Her research interests include machine learning, social media analysis, and medical image analysis.

Tansel Ozyer is a professor of Computer Engineering at Ankara Medipol University, Ankara Turkey. He completed his BSc in Computer Engineering and Information Sciences from Bilkent University, Ankara, Turkey, MSc in Computer Engineering from Middle East Technical University, Ankara, Turkey and PhD in Computer Science from University of Calgary. He is an associate editor SNAM journal and NetMAHIB journal. He is the coordinator of ASONAM conference. His research interests include data science, social network analysis, machine learning, homeland security, health informatics and image processing.

Reda Alhajj is a professor of Computer Science at the University of Calgary, Alberta, Canada. He is also affiliated with Istanbul Medipol University, Istanbul, Turkey, and University of Southern Denmark. He is a co-editor of the Encyclopedia on social network analysis and mining by Springer, editor in chief of SNAM journal and NetMAHIB journal. He is founding steering chair of ASONAM conference. His research interests include data science, social network analysis, machine learning, homeland security, health informatics and image processing.