

Article

Diagnosis of Pancreatic Ductal Adenocarcinoma Using Deep Learning

Fulya Kavak ¹, Sebnem Bora ^{1,*}, Aylin Kantarci ¹, Aybars Uğur ¹, Sumru Cagaptay ², Deniz Gokcay ², Anıl Aysal ², Burcin Pehlivanoglu ² and Ozgul Sagol ²

¹ Department of Computer Engineering, Ege University, 35040 Izmir, Turkey; fulya-kavak@outlook.com (F.K.); aylin.kantarci@ege.edu.tr (A.K.); aybars.ugur@ege.edu.tr (A.U.)

² Department of Pathology, Faculty of Medicine, Dokuz Eylul University, 35220 Izmir, Turkey; sumru.cagaptay@deu.edu.tr (S.C.); deniz.kabakci@deu.edu.tr (D.G.); anil.aysal@deu.edu.tr (A.A.); burcin.pehlivanoglu@deu.edu.tr (B.P.); ozgul.sagol@deu.edu.tr (O.S.)

* Correspondence: sebnem.bora@ege.edu.tr

Abstract: Recent advances in artificial intelligence (AI) research, particularly in image processing technologies, have shown promising applications across various domains, including health care. There is a significant effort to use AI for the early diagnosis and detection of diseases, offering cost-effective and timely solutions to enhance patient outcomes. This study introduces a deep learning network aimed at analyzing pathology images for the accurate diagnosis of pancreatic cancer, specifically pancreatic ductal adenocarcinoma (PDAC). Utilizing a novel dataset comprised of cases diagnosed with PDAC and/or chronic pancreatitis, this study applies deep learning algorithms to assess the effectiveness and reliability of the diagnostic process. The dataset was enhanced through image duplication and the creation of a second dataset with varied dimensions, facilitating the training of advanced transfer learning models including InceptionV3, DenseNet, ResNet, VGG, EfficientNet, and a specially designed deep neural network. The study presents a convolutional neural network model, optimized for the rapid and accurate detection of pancreatic cancer, and conducts a comparative analysis with other models to select the most accurate algorithm for a decision support system. The results from Dataset 1 show that EfficientNetB0 achieved a high success rate of 92%. In Dataset 2, VGG16 was found to have high performance, with a success rate of 92%. On the other hand, ResNet50 achieved a remarkable success rate of 96% despite a moderate training time and showed high precision, recall, F1 score, and accuracy. These results provide valuable data to demonstrate and share the relevance of different deep learning models in pancreatic cancer diagnosis.

Keywords: deep learning; convolutional neural networks; classification; health services; pathology images; pancreatic ductal adenocarcinoma



Citation: Kavak, F.; Bora, S.; Kantarci, A.; Uğur, A.; Cagaptay, S.; Gokcay, D.; Aysal, A.; Pehlivanoglu, B.; Sagol, O. Diagnosis of Pancreatic Ductal Adenocarcinoma Using Deep Learning. *Sensors* **2024**, *24*, 7005. <https://doi.org/10.3390/s24217005>

Academic Editors: Alain Tremeau, Marco Buzzelli and Flavio Piccoli

Received: 29 August 2024

Revised: 26 October 2024

Accepted: 28 October 2024

Published: 31 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial intelligence (AI) has become a main focus of research in the last decade. Today, with the rapid growth in artificial intelligence (AI) studies, interest in application areas such as object detection and classification based on image processing has increased. Medicine is one of the application domains of AI. The first purpose of health-related AI applications is to analyze the relationships between disease prevention or treatment techniques and patient treatment outcomes.

AI software has been developed for applications such as diagnosis processes, treatment protocols, personalized medicine, patient monitoring, and patient care. Moreover, AI algorithms can be used to analyze large amounts of complex data through electronic health records for disease prevention and diagnosis. AI applications provide an opportunity to increase efficiency, especially in health care in terms of early diagnosis and cost-effective solutions. The widespread use of AI in medicine will increase even more in the coming years.

Machine learning and artificial intelligence-based studies, especially in the field of diagnosis, focus on chronic diseases such as cancer, heart diseases, Alzheimer's, and kidney diseases and show that artificial intelligence-based applications will be encountered frequently in the diagnosis processes of these diseases in the near future. Although the results of initial studies are promising, there still seems to be a long road ahead for the daily use of AI in health care.

Over the past decade, modern machine learning methods have been successfully applied to the analysis of histopathology images thanks to significant advances in artificial intelligence and the increased ability to digitize histopathology slides. Despite the need for large amounts of labeled data, these models have demonstrated their potential in several applications, including cancer diagnosis, prognosis, and the prediction of genomic and molecular subtypes. The development of AI in this area not only facilitates more accurate and timely assessments of disease but also makes a significant contribution to clinical practice. This trend has led to the growing adoption of AI-based solutions in the histopathology field, thereby increasing the effectiveness of medical decision support systems [1].

Pancreatic ductal adenocarcinoma (PDAC) is the fourth leading cause of cancer-related deaths in the USA [2]. Pancreatic ductal adenocarcinomas (PDACs) comprise more than 90% of pancreatic cancer cases, and they exhibit a highly aggressive behavior. Chronic pancreatitis (CP) is a major diagnostic challenge for the pathologist and may mimic and accompany PDAC; distinguishing CP from PDAC is of the utmost importance for accurate diagnosis and optimal treatment. There are limited studies on the usage of deep learning in the diagnosis of pancreatic cancer [3,4].

Recently, anatomical structures within the pancreas have been classified using a deep learning algorithm on histopathological images [3]. In another study, it has been reported that pancreatic cancer can be identified by convolutional neural networks using deep transfer learning [5]. A deep convolutional neural network (DCNN) system was developed to segment cell clusters from patients from different hospitals and identify cancer cell clusters from cytopathology images [6]. Deepthi et al. [4] performed a comparative analysis of pancreatic tumor detection using VGG16, ResNet, and DenseNet. Dinesh et al. developed a system that uses convolutional neural networks (CNNs) and YOLO-based CNN (YCNN) models to identify critical features and cancerous growths in the pancreas. The study highlights the potential of the YCNN approach for accurate cancer detection and evaluates its effectiveness compared to other methods [7]. Saillard et al. developed a multi-stage deep learning model for the rapid molecular subtyping of PDAC using histopathological data. A tile-level analysis identified PDAC microheterogeneity and identified hybrid and intermediate tumor types, which could indicate transitional stages of disease progression [8].

In this study, we aimed to investigate the potential role of AI in distinguishing CP and PDAC and to develop a convolutional neural network to diagnose PDAC on histopathological images. Our study employs a specialized dataset of PDAC and chronic pancreatitis cases, enhanced through image duplication and varied dimensions, to train advanced models such as InceptionV3, DenseNet, VGG, EfficientNet, and ResNet.

The innovations in our study are the following:

- The development of a specialized dataset containing cases of PDAC and CP is an important and innovative aspect of this study. This dataset forms the basis of the research and provides a valuable reference for future studies.
- The use of different transfer learning models in the study helps to determine the most effective model by allowing a comparative analysis of different approaches.
- The developed convolutional neural network (CNN) model is optimized for the fast and accurate detection of pancreatic cancer and has the potential to save time and costs in clinical applications.

The aim is to assess and improve the diagnostic accuracy and reliability of a decision support system. In this context, diagnostic accuracy refers to the rate at which the

system provides accurate results, whereas reliability refers to the system's ability to provide consistent and reliable results. These two concepts are critical in decision-making processes because they enable users to make decisions with confidence. Therefore, this study aims to analyze the performance of the current system and identify potential areas for improvement.

Section 2 of this paper describes the construction of the dataset. It presents key information on the methods used and describes the training of the models. Section 3 presents the results, followed by Section 4, which discusses the results. Finally, Section 5 provides the conclusions.

2. Materials and Methods

The aim of this study was to perform classification with a visual dataset for pancreatic ductal adenocarcinoma (PDAC) and chronic pancreatitis (CP) analyses with the following steps:

- Data collection: Digital slides of PDAC and CP fields were photographed at $200\times$ magnification to obtain high-resolution images.
- Diagnostic criteria: The diagnosis of PDAC was based on major criteria (e.g., 4:1 nuclear size change, incomplete ductal structures) and minor criteria (e.g., single cell infiltration, irregular stroma, mitotic figures).
- Image preparation: Two datasets were created.
- Image processing: Captured images were de-identified for privacy and systematically numbered for easy classification.
- Image characteristics: All images at 2035×3841 were resized for a detailed analysis. Each image was subjected to a cropping procedure to ensure a balanced dataset to obtain two different image sets, 256×256 pixels and 512×512 pixels.

This methodology provides robust datasets suitable for training deep learning algorithms to increase PDAC diagnostic accuracy and improve clinical outcomes.

2.1. Dataset Preparation

First, hematoxylin–eosin-stained slides from consecutive patients with PDAC and/or CP (Figure 1) diagnosed at the Department of Pathology, Dokuz Eylul University Faculty of Medicine, were re-reviewed. Representative slides were scanned with a slide scanner (3D Histech, Panoramic 250 Flash III, Budapest, Hungary) with a $20\times$ objective lens. On the digital slides, areas of PDAC and CP were photographed at $200\times$ magnification to compile a comprehensive visual dataset. A nuclear size change among ductal cells of 4:1 or more, incomplete ductal structures, and the irregular distribution of ducts were the major criteria and single cell infiltration, irregular stroma, mitotic figures in epithelial cells, necrotic glandular debris, cribriform structures, large irregular eosinophilic nucleoli, and perineural invasion were the minor criteria used for the diagnosis of PDAC [9].

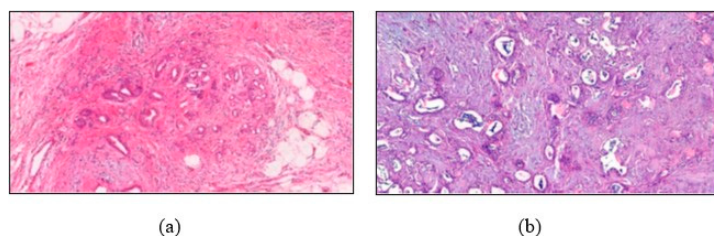


Figure 1. (a) A CP image showing the features of the dataset, and (b) a PDAC image showing different disease data. Both of these images are part of the datasets that were used in the study.

Photographs showing only malignant tumoral areas (cells or glands with these features) and photographs showing only findings of chronic pancreatitis without malignancy were prepared as two separate datasets. The images captured were de-identified at a

resolution of 2035×3841 pixels and uniquely classified through a systematic numbering approach.

A collection of 54 CP and 65 PDAC images was compiled, each with dimensions of 2035×3841 . In order to ensure a balanced dataset, each image in the dataset was generated by a rigorous cropping procedure, resulting in two different image sets of 256×256 and 512×512 (Table 1). This process facilitated the optimal utilization of visual data for the training of CNNs, illustrated in the subsequent Figure 2.

Table 1. The dataset created by the image cropping process and used in the study.

Dataset Information	Image Size	Number of Malignant Cell Images	Number of Benign Cell Images
Dataset with cropping	256×256	6825	5670
	512×512	1365	1134
Dataset in the study	256×256	6000	6000
	512×512	2000	2000

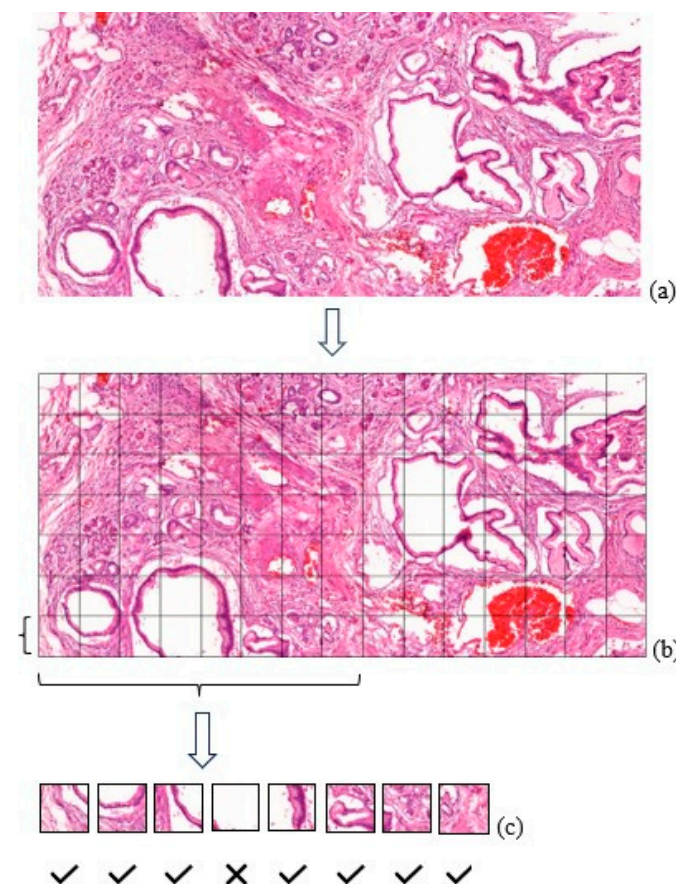


Figure 2. (a) PDAC/CP image; (b) the illustration demonstrates the selection of normalized tissue patches based on the CNN classification; (c) selected images used after the data cleanup.

The data was resized to two different sizes, 256×256 and 512×512 . After resizing, each image was preprocessed to create a 256×256 Dataset 1 consisting of 12,000 images and a 512×512 Dataset 2 consisting of 4000 images to be used in the study.

Subsequent to cropping, the images detailed in Figure 2 were re-evaluated by a pathologist (B.P.) in order to exclude any unsuitable visuals.

The images were then subjected to further augmentation—including rotation, contrast adjustments, and brightness modifications—to enrich data diversity (Figure 3).

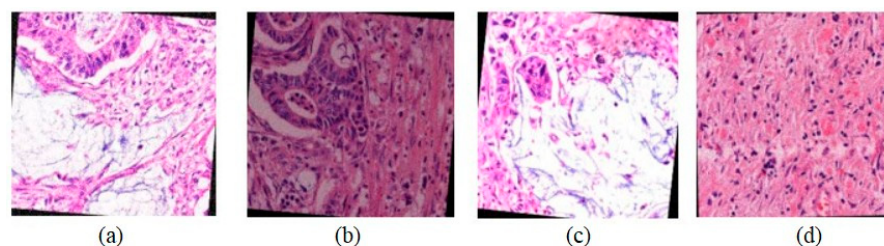


Figure 3. (a–d) are the images obtained as a result of image enhancement.

The dataset, comprising 256×256 pixel images, included 6000 images each of malignant and benign areas. Similarly, in the 512×512 pixel dataset, both malignant and benign areas were represented by 2000 images each (Table 1). The dataset allocation consists of 80% for training, 12% for testing, and 8% for validation (Table 1). The training set underwent additional augmentation processes and background refinement to enhance data quality.

A collection of 54 chronic pancreatitis and 65 pancreatic ductal adenocarcinoma images, each 2035×3841 in size, were curated. To ensure a balanced dataset, two distinct sets of images were generated with dimensions of 256×256 and 512×512 following a meticulous cropping procedure. This process facilitated the optimal utilization of visual data for the training of CNNs, as illustrated in the subsequent figure.

The dataset, comprising 256×256 pixel images, includes 6000 images each of malignant and benign cells. Similarly, in the 512×512 pixel dataset, both malignant and benign cells are represented by 2000 images each.

The dataset allocation consists of 80% for training, 12% for testing, and 8% for validation purposes. The training set underwent additional augmentation processes and background refinement to enhance data quality (Table 2).

Table 2. Distribution of Dataset 1 and Dataset 2 in the study.

Dataset	Train	Test	Validation
Dataset 1	9600	1440	960
Dataset 2	3200	480	320

2.2. Selection of Deep Learning Techniques

To discern PDAC from CP, a range of deep learning frameworks were employed. The study focused on developing a CNN model. Following the division of the dataset into training and testing subsets, the process continued with model training and evaluation on test data.

2.3. Evaluation of Predictions

Following the training and testing phases, the models underwent scrutiny based on specific evaluation metrics. The analysis encompassed observing fluctuations in metrics across both balanced and imbalanced data distributions, examining the correlation between forecasted outcomes and error rates, and assessing the success metrics of the final results.

Confusion matrix, accuracy, the F1 score, loss, and precision are used as success metrics. The collection of a model allows for a more comprehensive analysis when these metrics are evaluated together.

2.4. Used Deep Learning Methods

In this study, different deep learning methods were used to classify PDAC and CP disease. Different types of convolutional neural networks [5] were used for feature extraction from histopathology images, species classification, and the evaluation of algorithm performance. InceptionV3 [6], DenseNet121 [4], ResNet50 [10,11], VGG19 [12], VGG16 [12,13], EfficientNetB0 [14], EfficientNetB5 [14], and the designed deep learning model were used to improve the classification performance.

2.5. Architecture of the Deep Learning Model

The datasets were configured in two distinct resolutions and partitioned into 80% for training purposes, with the remaining 20% allocated to testing and validation. In the constructed CNN model, the batch size was uniformly set to 64 across both datasets, and the training was conducted over 10 epochs. Following the initial preprocessing phase, a CNN was structured to facilitate the training of the image datasets, featuring an input layer dimension of $224 \times 224 \times 3$ (Figure 4).

Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 224, 224, 64)	1792
conv2d_1 (Conv2D)	(None, 224, 224, 64)	36,928
max_pooling2d (MaxPooling2D)	(None, 112, 112, 64)	0
conv2d_2 (Conv2D)	(None, 112, 112, 128)	73,856
conv2d_3 (Conv2D)	(None, 112, 112, 128)	147,584
max_pooling2d_1 (MaxPooling2D)	(None, 56, 56, 128)	0
conv2d_4 (Conv2D)	(None, 56, 56, 256)	295,168
conv2d_5 (Conv2D)	(None, 56, 56, 256)	590,080
conv2d_6 (Conv2D)	(None, 56, 56, 256)	590,080
max_pooling2d_2 (MaxPooling2D)	(None, 28, 28, 256)	0
conv2d_7 (Conv2D)	(None, 28, 28, 512)	1,180,160
conv2d_8 (Conv2D)	(None, 28, 28, 512)	2,359,808
conv2d_9 (Conv2D)	(None, 28, 28, 512)	2,359,808
max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 512)	0
conv2d_10 (Conv2D)	(None, 14, 14, 512)	2,359,808
conv2d_11 (Conv2D)	(None, 14, 14, 512)	2,359,808
conv2d_12 (Conv2D)	(None, 14, 14, 512)	2,359,808
max_pooling2d_4 (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25,088)	0
dense (Dense)	(None, 256)	6,422,784
dense_1 (Dense)	(None, 64)	16,448
dense_2 (Dense)	(None, 2)	130
Total params: 21,154,050 (80.70 MB)		
Trainable params: 21,154,050 (80.70 MB)		
Non-trainable params: 0 (0.00 Byte)		

Figure 4. Design of the CNN architecture.

The constructed model adheres to the CNN framework, incorporating convolutional layers for feature extraction, max-pooling layers to reduce spatial dimensions, and dense layers to execute classification tasks. The model culminates in an output layer with two nodes, signifying a binary classification objective.

The intricate design is indicative of a model with substantial depth and complexity. In this model, convolutional layers were instantiated with Conv2D() functions employing 3×3 kernels across 64, 128, and 512 filters. The convolutional layers, configured with 3×3 kernel sizes and a linear activation function, were systematically organized, with the Conv2D() function applied twice per dimension. Interposed with these are max-pooling layers, implemented via the MaxPooling2D() function, each with a 2×2 scope, to efficiently downscale feature dimensions.

The model harnesses the ReLU activation function, renowned for its efficacy in neural networks, particularly within CNN architectures. Following the convolutional sequences, the flatten() function transforms the network's multidimensional output into a singular

vector, seamlessly integrated with a densely connected layer. This terminal layer, equipped with a softmax activation function, is pivotal for resolving the classification dilemma inherent in this research. The inclusion of 256 neurons via the Dense() function, alongside a fully connected layer with a linear activation response, rounds off the model's architecture. The Dense() layer is used to determine the image into one of two classes.

2.6. Computational Resources

The research was executed leveraging the Python 3.10 programming language, utilizing Jupyter Notebook and Google Colab Pro for the development environment on a personal computing device equipped with a Lenovo Yoga 11th Generation Intel(R) i7 processor. The deep learning experiments were facilitated using the Keras library [15], version 2.9.1, operating atop the TensorFlow [16] framework.

3. Results

To determine the efficacy of the trained models, we employed a suite of quantitative performance metrics, including the confusion matrix [17,18] (Figure 5), accuracy, precision, sensitivity, the F1 score, and loss, to evaluate their capabilities across assigned tasks [19,20]. Moreover, the time span of the training sessions was scrutinized. Definitions for the metrics used in our assessment are provided for a comprehensive understanding.

		Actual	
		Positive	Negative
Predict	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Figure 5. Confusion matrix.

These metrics are predicated on the following definitions:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100 \quad (1)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \times 100 \quad (2)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \times 100 \quad (3)$$

$$\text{F1} = \frac{TP}{(TP + TN + FP + FN)} \times 100 \quad (4)$$

The definition of parameters used in (1)–(4) are given as follows:

- TP (true positive): The accurate identification of a true condition.
- FP (false positive): The erroneous identification of a condition as true.
- TN (true negative): The accurate identification of a false condition.
- FN (false negative): The erroneous identification of a condition as false.

The confusion matrix visualizes the correct and incorrect classifications of the model and helps us understand which class predictions are better or worse. Accuracy indicates the overall correct prediction rate of the model, whereas precision and sensitivity measure the quality of positive class predictions and the model's ability to find true positives. The F1 score provides a combined assessment of these two measures, balancing precision and sensitivity. Loss indicates how effective the learning process is by reflecting the error rate in the model's predictions. These multifaceted metrics allow us to analyze the performance of models in a more comprehensive way.

Given the inherent strengths and weaknesses of each model, selecting the optimal one for specific tasks or datasets becomes imperative. The commendable precision, recall, F1 score, and accuracy exhibited by most models signify their superior classification prowess, with EfficientNetB0 and ResNet50 demonstrating particularly outstanding overall performance.

Analysis of the data presented in Table 3 reveals EfficientNetB0 as the frontrunner in precision (92%), closely followed by ResNet50 and VGG16. This trend persists across recall rates and F1 scores, positioning EfficientNetB0 at the apex of model accuracy, a testament to its superior predictive capability.

Table 3. Testing results for Dataset 1.

Model	Precision	Accuracy	Recall	F1 Score	Loss	Training Time
DenseNet121	88%	88%	88%	88%	0.41	28 min 58 s
EfficientNetB0	92%	92%	92%	92%	0.40	14 min 18 s
EfficientNetB5	83%	83%	82%	82%	0.52	14 min 14 s
Inceptionv3	77%	77%	77%	77%	0.57	16 min 32 s
ResNet50	92%	92%	92%	92%	0.37	26 min
VGG16	89%	89%	88%	88%	0.41	12 min 42 s
VGG19	85%	85%	85%	85%	0.48	28 min 19 s
Designed Model	85%	84%	84%	84%	0.31	19 min 49 s

The training duration varied significantly between models. As shown in Figure 6, VGG16 emerged as the most time-efficient model, requiring only 12 min 42 s (762 s) for training, whereas DenseNet121 recorded the longest training time of 28 min 58 s (1738 s).

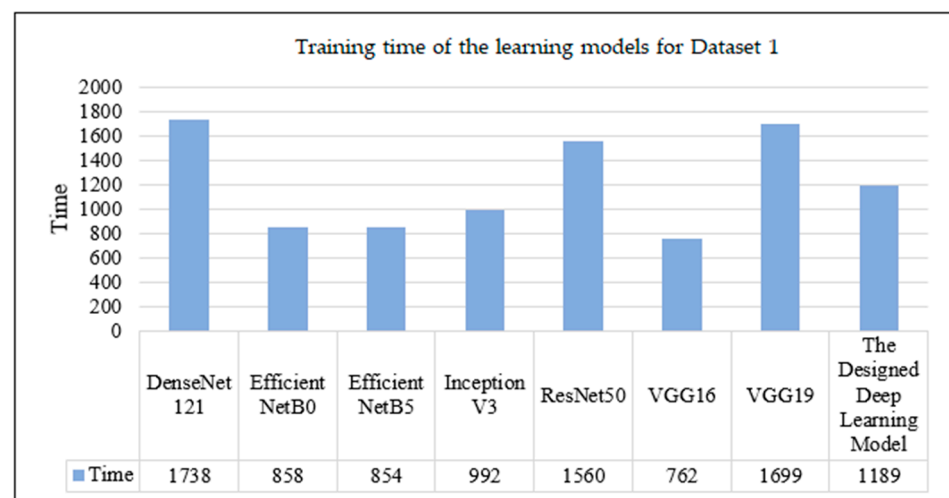


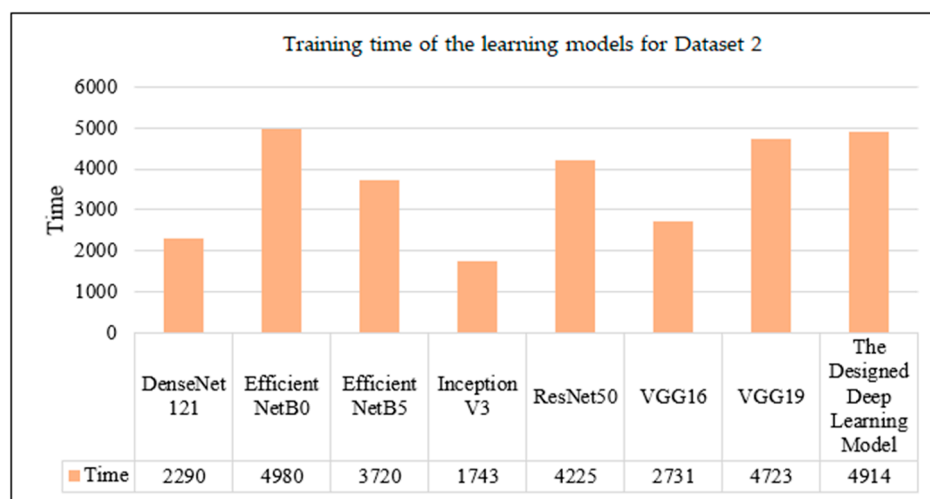
Figure 6. Training time of the learning models for Dataset 1.

An evaluation of the test results from Dataset 2 shows that VGG16 is a high-performance competitor, with a 92% success rate despite the long training time, as can be seen from the data in Table 4. On the other hand, ResNet50 stands out with an impressive 96% success rate, showcasing high precision, recall, F1 score, and accuracy, with its training time being relatively moderate in comparison. The developed deep learning model demonstrates moderate efficacy, with a training duration marginally exceeding that of its counterparts.

Table 4. Testing results for Dataset 2.

Model	Precision	Accuracy	Recall	F1 Score	Loss	Training Time
DenseNet121	92%	92%	92%	92%	0.43	38 min 10 s
EfficientNetB0	94%	94%	94%	94%	0.46	1 h 23 s
EfficientNetB5	88%	87%	87%	88%	0.67	1 h 2 s
Inceptionv3	75%	74%	74%	74%	0.90	29 min 3 s
ResNet50	96%	96%	96%	96%	0.40	1 h 10 min 25 s
VGG16	92%	92%	92%	92%	0.51	45 min 31 s
VGG19	87%	87%	87%	87%	0.56	1 h 18 min 43 s
Designed Model	86%	86%	86%	86%	0.32	1 h 21 min 54 s

Figure 7 illustrates that models like EfficientNetB0 and ResNet50 maintain superior performance even with prolonged training times. The custom model matches the performance of its peers while sustaining an average training duration.

**Figure 7.** Training time of the learning models for Dataset 2.

The CNN, upon training and testing, attained 84% accuracy, 86% precision, 86% recall, and an 86% F1 score in Dataset 1. For Dataset 2, the figures were 86% for accuracy, 85% for precision, 84% for recall, and 84% for the F1 score. The training validation loss and accuracy for Datasets 1 and 2 are depicted through graphs, with the epoch value set at 10. Accuracy stood at 84% for Dataset 1 and 86% for Dataset 2, underscoring the network's efficiency despite its relatively brief training span.

4. Discussion

Histopathology plays an important role in the diagnosis of diseases such as cancer. Deep learning algorithms are used as a powerful tool to identify pathological features in histopathological images and diagnose diseases. These algorithms are trained on large amounts of data in order to detect and analyze pre-defined features. The use of deep learning algorithms can speed up the diagnostic process for pathologists and help them achieve accurate results. In addition, this technology can reduce human error in image analysis and facilitate the early diagnosis of disease.

For the pathological diagnosis of pancreatic ductal adenocarcinoma, a 4:1 or more nuclear size change among ductal cells, incomplete ductal structures, and the irregular distribution of ducts as major criteria and single cell infiltration, irregular stroma, mitotic figures in epithelial cells, necrotic glandular debris, cribriform structures, large irregular eosinophilic nucleoli, and perineural invasion as minor criteria are helpful in reaching a diagnosis of malignancy [9]. In addition, the disruption of the normal lobular structure of the parenchyma and the presence of ductal structures in close proximity to large muscular

vessels within the interlobular connective tissue, where they should not normally be found, are other helpful findings [21]. However, in the presence of significant atrophy, inflammation, and the irregular ductal structures observed in chronic pancreatitis, microscopic evaluation can be quite difficult. Especially in pancreatic resection materials showing findings of chronic pancreatitis, in order to exclude PDAC, a large number of samples must be obtained and a large number of slides must be evaluated in detail in terms of these morphological features, which is quite time-consuming and challenging. At this point, deep learning can assist the pathologist by detecting areas with suspicious findings, which can provide a diagnostic advantage in terms of reducing time loss for patients.

Kronberg et al. reported the detection of pancreatic cancer in hematoxylin and eosin-stained slides using deep transfer learning convolutional neural networks. The ResNet18 network achieved a five-class accuracy of approximately 94% on test data images. The network was validated on independent tissue sections consisting of healthy pancreatic tissue, pancreatic ductal adenocarcinoma (PDAC), and lymph node metastases from pancreatic cancer. Patient data were divided into 80% for training, 10% as normal, and 10% as test datasets. The raw dataset included 1690 images for the test data, 14,450 for training data, and 1702 for continuous data for a healthy pancreas; 874, 8275, and 805 slices for healthy lymph node (HLN) tissue; and 2454, 20,694, and 2502 slices for PDAC. The model achieved a weighted accuracy of 90%, a weighted Jaccard score of 81%, and a weighted F1 score of 90% across all classes. Individual F1 scores were 86% for a healthy pancreas and 92% for PDAC. The Jaccard scores were 82% for HLN and 85% for PDAC [2].

Zhang et al. conducted a retrospective multicenter diagnostic study using 5345 cytopathology slide images from 194 patients. They developed a deep convolutional neural network (DCNN) to segment cell clusters and identify cancer cells in cytopathology images. The DCNN achieved F1 scores of 0.929 in internal tests and 0.899–0.938 in external tests for separating stained cells from the background. For cancer identification, it showed an AUC of 0.958 internally and 0.948–0.976 externally. The DCNN's performance surpassed that of validated endoscopists and cytopathologists in sensitivity analysis and test datasets, demonstrating its robustness and generalizability [22].

The study by Deepthi and colleagues demonstrated high accuracy for their models: VGG16 achieved 96%, ResNet 98%, and DenseNet 99%. Precision values were 95% for VGG16, 96% for ResNet, and 95% for DenseNet. Recall values were 98% for VGG16, 95% for ResNet, and 97% for DenseNet, whereas the F1 scores were 98%, 97%, and 98%, respectively. The findings indicate that VGG16, ResNet, and DenseNet are effective tools for pancreatic tumor detection, each with distinct advantages. By improving the early detection of pancreatic cancer, these models can enhance patient outcomes and inform future developments in medical image analysis [23].

Deep learning algorithms can therefore be used to diagnose histopathological images using a CNN model [24]. In this study, datasets were created using images of patients diagnosed with PDAC and CP from the Department of Pathology, Dokuz Eylul University Faculty of Medicine. The datasets were tested with DenseNet121, InceptionV3, EfficientNetB0, EfficientNetB5, ResNet50, VGG16, VGG19, and the designed network and analyzed to classify the models when extracting features from images.

The models used in the study were evaluated in terms of precision, recall, the F1 score, and accuracy. In the graphs resulting from the training, it is clear that the EfficientNetB0 model performs better than the other models for Dataset 1. It achieved 92% for precision, recall, the F1 score, and accuracy. However, it is also clear that some models (e.g., InceptionV3) perform worse than others. From the analysis of training times, it can be seen that each model has a different training time. For example, the VGG16 model has a lower performance than the other models, although it has a shorter training time. The graphs resulting from the training of Dataset 1 compare the performance of different deep learning models and provide an idea of which model is preferable. In this context, the EfficientNetB0 model can be preferred as a better deep learning model than the others. In the graphs resulting from the training on Dataset 2, it is clearly seen that the ResNet50

and EfficientNetB0 models perform better than the other models. Both models achieved 94% and above for precision, recall, the F1 score, and accuracy. There are some differences between the other models. For instance, the Inceptionv3 model showed lower performance than the others, especially for precision, recall, and the F1-score, dropping to 75%. In the analysis of the training times, it can be seen that each of the models takes a different amount of time to train. For example, the EfficientNetB0 model was trained in only 1 h and 23 min but achieved high performance. The other models required longer training times. As a result, ResNet50 and EfficientNetB0 can be preferred as deep learning models for Dataset 2 (Figure 8).

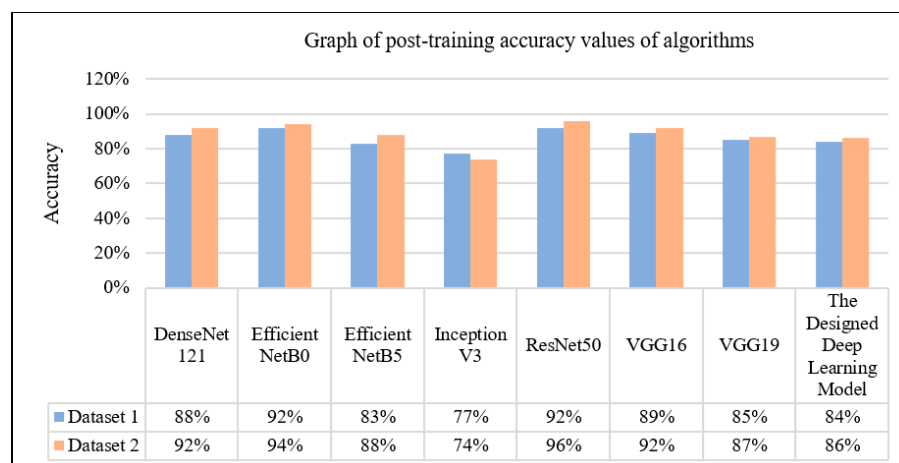


Figure 8. A comparison of the accuracy value of deep learning algorithms.

The deep learning algorithms used in the study are effective in identifying salient features to identify diseases by analyzing the histopathological dataset. This can be used to further improve the early detection of diseases such as cancer and determine appropriate treatment. However, there are some challenges associated with the use of deep learning classification algorithms.

In particular, the size and quality of datasets can affect the accuracy and generalizability of algorithms. For this reason, a dataset was created, and high-quality data was selected. To conclude, classifying histopathological images using deep learning has significant potential in the medical field. This technology can provide a faster, more accurate, and reliable approach to the diagnosis and treatment of diseases. To achieve accurate results, advanced data analysis and the continuous improvement of algorithms are required.

Furthermore, the training validation loss and accuracy plots show a steady improvement, providing information about the learning dynamics of the network. The high precision, recall, and F1 values obtained on both datasets underline the ability of the CNN to classify the data samples in an effective way. Overall, these results confirm the robustness and reliability of the CNN in processing different datasets and show its potential for real applications in different domains.

5. Conclusions

In conclusion, our study shows that the classification performance for disease identification in PDAC tissue sections can be improved by preprocessing the dataset with different convolutional networks and by varying the hyperparameters of the designed network. Further studies on the application of this approach to metastases from different primary tumors will be necessary for validation.

Author Contributions: B.P., F.K., S.B., A.K. and A.U. designed the study. F.K. and S.B. designed the deep learning algorithm and software and performed the tests. B.P., A.A., S.C., D.G. and O.S. provided and analyzed the pictures. All authors contributed to the manuscript, and S.B., A.K., A.U.,

A.A., B.P. and O.S. critically reviewed it. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study protocol was approved by the institutional ethics committee (approval no: 2022/36-17).

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available upon reasonable request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

CNN	Convolutional Neural Network
PDAC	Pancreatic Ductal Adenocarcinoma
CP	Chronic Pancreatitis
DNN	Deep Neural Network
RNN	Recurrent Neural Network
VGG	Visual Geometry Group
DENSENET	Densely Connected Convolutional Network
RESNET	Residual Network
ReLU	Rectified Linear Unit
TN	True Negative
TP	True Positive
FN	False Negative
FP	False Positive
HP	Healthy Pancreas
HLN	Healthy Lymph Node
DCNN	Deep Convolutional Neural Network
YOLO	YOLO Model-Based CNN
YOLO	You Only Look Once

References

1. Ahmadvand, P.; Farahani, H.; Farnell, D.; Darbandsari, A.; Topham, J.; Karasinska, J.; Nelson, J.; Naso, J.; Jones, S.J.M.; Renouf, D.; et al. A Deep Learning Approach for the Identification of the Molecular Subtypes of Pancreatic Ductal Adenocarcinoma Based on Whole Slide Pathology Images. *Am. J. Pathol.* **2024**. [[CrossRef](#)] [[PubMed](#)]
2. Park, W.; Chawla, A.; O'Reilly, E.M. Pancreatic Cancer: A Review. *JAMA* **2021**, *326*, 851–862. [[CrossRef](#)] [[PubMed](#)]
3. Kriegsmann, M.; Kriegsmann, K.; Steinbuss, G.; Zgorzelski, C.; Kraft, A.; Gaida, M.M. Deep learning in pancreatic tissue: Identification of anatomical structures, pancreatic intraepithelial neoplasia, and ductal adenocarcinoma. *Int. J. Mol. Sci.* **2021**, *22*, 5385. [[CrossRef](#)] [[PubMed](#)]
4. Deepthi, G.; Bamini, A.A.; Praveen, Y.J. A Comparative Analysis of Pancreatic Tumor Detection using VGG16, ResNet, and DenseNet. In Proceedings of the 2024 IEEE 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 5–7 June 2024; pp. 1303–1308.
5. Kronberg, R.M.; Haeberle, L.; Pfaus, M.; Xu, H.C.; Krings, K.S.; Schlenz, M.; Rau, T.; Pandya, A.A.; Lang, K.S.; Esposito, I.; et al. Communicator-driven data preprocessing improves deep transfer learning of histopathological prediction of pancreatic ductal adenocarcinoma. *Cancers* **2022**, *14*, 1964. [[CrossRef](#)]
6. Zhang, S.; Zhou, Y.; Tang, D.; Ni, M.; Zheng, J.; Xu, G.; Peng, C.; Shen, S.; Zhan, Q.; Wang, X.; et al. A deep learning-based segmentation system for rapid onsite cytologic pathology evaluation of pancreatic masses: A retrospective, multicenter, diagnostic study. *EBioMedicine* **2022**, *80*, 104022. [[CrossRef](#)]
7. Dinesh, M.G.; Bacanin, N.; Askar, S.S.; Abouhawwash, M. Diagnostic ability of deep learning in detection of pancreatic tumour. *Sci. Rep.* **2023**, *13*, 9725. [[CrossRef](#)]
8. Saillard, C.; Delecourt, F.; Schmauch, B.; Moindrot, O.; Svrcek, M.; Bardier-Dupas, A.; Emile, J.F.; Ayadi, M.; Rebours, V.; Mestier, L.; et al. Pacpaint: A histology-based deep learning model uncovers the extensive intratumor molecular heterogeneity of pancreatic adenocarcinoma. *Nat. Commun.* **2023**, *14*, 3459. [[CrossRef](#)]
9. Hyland, C.; Kheir, S.M.; Kashlan, M.B. Frozen section diagnosis of pancreatic carcinoma: A prospective study of 64 biopsies. *Am. J. Surg. Pathol.* **1981**, *5*, 179–191. [[CrossRef](#)] [[PubMed](#)]
10. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [[CrossRef](#)]

11. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
12. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
13. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
14. Al-Haija, Q.A.; Adebajo, A. Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network. In Proceedings of the 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Vancouver, BC, Canada, 9–12 September 2020; pp. 1–7.
15. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
16. Abbas, M.A.; Bukhari, S.U.K.; Syed, A.; Shah, S.S.H. The Histopathological Diagnosis of Adenocarcinoma & Squamous Cells Carcinoma of Lungs by Artificial intelligence: A comparative study of convolutional neural networks. *MedRxiv* **2020**.
17. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
18. Ketkar, N. *Deep Learning with Python: A Hands-On Introduction*; Apress: Karnataka, India, 2017; pp. 22–109.
19. Raschka, S.; Mirjalili, V. *Python machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow 2*, 3rd ed.; Packt Publishing Ltd.: Birmingham, UK, 2017.
20. Saito, T.; Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **2015**, *10*, e0118432. [[CrossRef](#)] [[PubMed](#)]
21. Hart, J.; Pai, R.K.; Wilcox, R.; Noffsinger, A. *Liver, Extrahepatic Biliary Tree, Gallbladder and Pancreas, in Biopsy Interpretation; The Frozen Section, Second*; Cipriana, N.A., Taxy, J.B., Husain, A.N., Eds.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2014; pp. 297–343.
22. Vujović, Ž. Classification model evaluation metrics. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 599–606. [[CrossRef](#)]
23. Albahr, A.; Albahr, M. An empirical comparison of fake news detection using different machine learning algorithms. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 146–152. [[CrossRef](#)]
24. Barisoni, L.; Lafata, K.J.; Hewitt, S.M.; Madabhushi, A.; Balis, U.G. Digital pathology and computational image analysis in nephropathology. *Nat. Rev. Nephrol.* **2020**, *16*, 669–685. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.