Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

AI-POWERED PANCREATIC CANCER DIAGNOSIS SYSTEM

Muneeba Sarwer[1], Zainab Riaz[2], Muhammad Shadab Alam Hashmi[*3]

[1,2,*3]*Institute of Computer Science, Khwaja Fareed University of Engineering &Information Technology Rahim Yar Khan*

[1]munibasarwer194@gmail.com, [2]zainabriaz195@gmail.com, [*3]shadab.alam@kfueit.edu.pk

DOI: https://doi.org/10.5281/zenodo.18720830

Keywords

Article History
Received: 11 October 2025
Accepted: 21 November 2025
Published: 06 December 2025

Corresponding Author: *
Muhammad Shadab Alam Hashmi

Abstract
Pancreatic cancer is one of the deadliest cancers and it is usually detected very late because early symptoms are hard to notice. To solve this problem, this project creates an Ai based system that uses CT scan images to detect pancreatic cancer early and accurately.
In this project , the method is quite different as it is using light weight and flexible Ai model built with Transfer learning, This means we take a pretrained CNN model like MobileNet and trained it on a properly organized dataset taken from Kaggle. This greatly reduces training time and computer resources while still giving excellent accuracy.
The model is designed to work well even when the dataset is small, making it suitable for hospitals or clinics with limited resources. We also use improved preprocessing and feature extraction methods to make the predictions more reliable and efficient.
The early results of the model are very promising: it gives 99% accuracy, 1.000 precision, 0.987 recall, and an F1-score of 0.993. These results show that the model is strong, fast, and highly accurate. The final system becomes a light, easy-to-use tool that supports radiologists in identifying pancreatic cancer, which can help doctors provide treatment earlier and improve patient survival.

INTRODUCTION

This project is about creating an Ai based pancreatic cancer detection system using CT scan images. This helps in reducing diagnosis time and help doctors to make informed decisions faster. In underdeveloped environments, it can compensate for limited medical resources. This system uses a pre-trained MobileNet model with transfer learning. This system works through several steps i.e. data augmentation, (where images are increased using rotation, flipping, zooming) image preprocessing, segmentation (tumor regions are separated using methods such as U-Net or Watershed that helps highlighting the areas of pancreas where tumor is present) feature extraction through GLCM (Gray-Level Co-occurrence Matrix), tumor detection (normal or tumor) and classification into Benign, Pre-Malignant and Malignant categories. Benefits of the System include Reducing the need for time-consuming manual inspection, helps improving diagnostic accuracy and speed. It Works well even on limited hardware due to MobileNet's lightweight structure. It can be scaled for different datasets and healthcare centers. It can support radiologists and oncologists during diagnosis, for real-time analysis and quick decision support, new data can be added to improve performance. Early testing shows that the system is reliable and gives accurate results. It can also be used in hospital information systems to make the clinical workflow smoother. In future, the project will add multiple types of medical images and will be tested on a larger scale to increase its usefulness. Continuous feedback from doctors will help improve its prediction quality.

Problem Statement

Prolonged review by oncologist and radiologists, manual dependency and sensitivity involved in treating and handling tumors remains challenging in the early detection of pancreatic cancer. This procedure is likely to be affected by errors and delays, which results in the diagnosis at late stages when treatment options are limited. In limited-resource environments, lack of skilled professionals further causes delays, inconsistencies and errors. Consequently, there is an urgent need for an automated, intelligent system that can accurately and efficiently analyze CT scan images and support early diagnosis while reducing manual effort and also improve survival rates.

Objective

This project aims to develop a reliable AI-powered diagnostic system that is efficient in accurately detecting pancreatic cancer from CT scan images. Using transfer learning with a pre-trained MobileNetV2 model, the system even with the limited dataset aims for automated image analysis with enhanced accuracy and reduced diagnostic time. Primary objective targets on designing an easy, user-friendly interface with key goals including augmentation, preprocessing, model training, validation and evaluation using accuracy, recall and F1-score as performance metrics, to provide quick diagnostic feedback to medical professionals, making early detection more accessible and reliable.

LITERATURE REVIEW

RELATED STUDIES

In pancreatic cancer detection, AI-based systems aim to manage complex imaging data and improve accuracy using CT scans. By enabling automated analysis, AI has advanced tools and techniques for accurate and quick detection of medical images.

Nadeem and others. (2025) developed an automated system which uses R-CNN and VGG-16 with Reduced 11-layer Alexnet for detection of pancreatic cancer from CT scan images. It achieved classification accuracy of 96% . On the other hand, this system depends on a single dataset which limits the broader clinical applicability.

Zhou and others. (2019) applied U-Net segmentation to show pancreatic tumor regions in CT scan images. The sequence-to-sequence architecture gives precise detection of boundaries and it also requires a large, completely labeled dataset and significant GPU resources. In comparison, our system utilizes pre-trained weights to reduce data and compute requirements.

Liu and others. (2020) used Deep Learning CNN (Convolutional Neural Network) model for distinguishing pancreatic cancer from non cancerous tissues. A key contribution was its ability to generalize across different patient populations with high sensitivity of 98% with local test set.

Cao and others. (2023) developed PANDA (Pancreatic Cancer Detection with Artificial Intelligence), a deep learning model and trained that model on 3,208 images and achieved accuracy of 98%. It has some generalization issues due to largely training on East Asian populations, and high specificity can cause false positives at times.

Tan and others. (2020) proposed a 3D-GLCM convolutional neural network (Multi- channel CNN) for polyp classification while using gray-level co-occurrence matrices and achieved an accuracy of 91% . While on the other hand, their small dataset size is a constraint as compared to our model.

Zhang Z. and others. (2020) designed a DCNN (deep convolutional neural network) framework with Augmented Feature Pyramid Networks and achieved accuracy of 94%. This approach is computationally expensive for limited dataset-resource and requires detailed context information for effective localization which is also difficult to handle.

Ma H and others. (2020) constructed a CNN (Convolutional Neural Network) using 3,494 CT scan images from cancer patients and then achieved an accuracy of 95.47%. This model's medium-sized, single-center dataset limits generalization for external clinics and resources which limits broader applicability.

DRAWBACKS IN RELATED WORKS

- **High Computational Complexity:** Deep learning models require high computational resources and long training times, which make them impractical for real-time use.
- **Large Dataset Requirements:** Many systems rely on large, completely labeled datasets, which are difficult and costly to obtain in clinical settings.
- **Accuracy-Efficiency Trade-off:** By maximizing the accuracy of a model it results in minimizing the efficiency of the model or vice versa. So, it makes the model unsuitable and slow for time-dependent applications.
- **Restricted real-time applications:** Multiple step procedures often led to delays and errors, which limits the effectiveness of real-time applications.
- **Inadequate Model Generalization:** To generalize across different conditions of images and different datasets models often struggle, which results in low real-world performance.

RESEARCH GAPS

Regardless of progress in AI for pancreatic cancer detection, various gaps remain. Most of the models used in this procedure are computationally heavy and depend on large labeled datasets, which limits their use in resource-restricted environments. These models are often incapable of generalizing across multiple datasets and different conditions of imaging, which also limits the clinical applicability. The requirement for segmentation and manual feature extraction in models reduces automation and increases dependence on expert inputs. Also, slow multi-step processing limits real-time use. The existing studies focus completely on CT scans, while disregarding the needs of integrating patient history. By developing efficient, lightweight models that run smoothly on smaller datasets and can also deliver real-time outcomes is important to improve scalability and as well as efficiency.

COMPARISON OF RELATED WORKS SUMMARY

The AI-based methods for detecting and classifying pancreatic cancer from CT scan images, while using transfer learning with a pretrained model MobileNetV2. It clarified the common drawbacks of data diversity, constrained real-time performance and heavy computation. It highlighted significant gaps such as scalable, lightweight models that operate efficiently on smaller datasets. A comparative overview shows how existing systems compare in accuracy, complexity, and resource needs, and establish the case for our MobileNet-based approach as a faster, more efficient, and cost-effective CAD tool.

RESULTS AND EXPERIMENTS

EXPERIMENTAL PLATFORM AND TOOLS

Platform: Google Colab (Cloud-based)
CPU: Intel Xeon (2 vCPUs, 64-bit)
GPU: NVIDIA Tesla T4 (16 GB VRAM)
Programming Language: Python 3.x
Frameworks: TensorFlow 2.x, Keras 2.x, PyTorch 2.x (optional)
Libraries: NumPy, Pandas, Matplotlib, OpenCV, Scikit-Learn
Operating System: Google Cloud Linux (Debian-based)

Dataset Source: Kaggle - CT Scan Images

DATASET DETAIL

Dataset Source: https://www.kaggle.com/datasets/jayaprakashpondy/pancreatic-ct-images?rvi=1
Dataset Name: Pancreatic CT Images

DATA DISTRIBUTION

Training Set:
Total Images: 999
Normal Cases: 500 images (50%)
Tumor Cases: 499 images (50%)
Testing Set:
Total Images: 412
Normal Cases: 225 images (54.6%)
Tumor Cases: 187 images (45.4%)
Overall Dataset:
Total Images: 1,411
Classes: 2 (Binary Classification)
Class 0: Normal
Class 1: Pancreatic Tumor
Split Ratio: 70-30 %

Table 1:Summery of literature review

Study	Model Used	Dataset	Accuracy	Key Limitations
Nadeem et al. (2025)	Reduced 11-layer AlexNet (CAD System), Hybrid VGG16 + R-CNN	Kaggle Pancreatic CT Images (Specific count not given, but publicly available)	96.72% (Classification)	Relies on a single, publicly available dataset (potential for overfitting/lack of generalizability).
Zhou et al. (2019)	Hyper-Pairing Network (HPN), a 3D FCN	Two PDAC datasets (Count not explicitly available, but used multi-phase CT)	63.94% (Dice Score for segmentation)	Segmentation tasks are inherently challenging (low contrast, small tumors). The reported metric is Dice Score, no classification accuracy.
Liu et al. (2020)	Convolutional Neural Network (CNN)	370 PC patients + 320 controls (Taiwanese Center) + US Test Set	98.6-98.9% (Local Test Set)	Lower sensitivity (79.0%) on the external, cross-racial US test set, indicating generalizability challenges.
Cao et al. (PANDA) (2023)	Deep Learning Model	Large-scale (multi-center, 10 centers)	0.986-0.996 (AUC)	Findings derived from a large, complex multi-center dataset which may be challenging to replicate and deploy.
Tan et al. (2020)	3D-GLCM + Multi-channel CNN	Very small (63 polyps)	0.91-0.93 (AUC for Classification)	Small dataset size (63 polyps), focused on polyp classification (colonography, not pancreatic CT).
Zhang Z. et al. (2020)	DCNN (with DC Module and AFP Networks)	~2890 CT Images (implied in related reviews)	94.0% (Accuracy)	DCNN is computationally expensive; it requires rich context information for effective tumor localization.
Ma H. et al. (2020)	CNN Classifier (Custom architecture used)	3494 CT images (from 190 patients)	95.47% (Plain Scan Accuracy)	Relies on a medium-sized, single-center dataset, potentially limiting generalizability to external institutions.
Proposed System (Ours)	Transfer learning with pre-trained MobileNet, end-to-end fine-tuning	Small Kaggle CT dataset (1400+ CT scan Images)	Target: ~95%	Lightweight, fast, suitable for low-resource settings

Data Augmentation

Data augmentation is a technique where more training images are created from existing images by applying small and controlled changes. These techniques do not change the meaning of images but help the model learn better by seeing different versions of same images like seeing the image in different angles, colors or sizes. It helps the model becoming smarter and more flexible (Shorten et al, 2019) (Perez et al. 2017). Techniques that augmentation includes: **1.** Rotation: Rotating images at different angles. **2.** Flipping: Flipping images horizontally or vertically.**3.** Scaling: Resizing images. **4.** Cropping: Cropping parts of images. **5.** Color jittering: Adjusting brightness, contrast, and saturation.

AUGMENTATION TECHNIQUES USED IN OUR MODEL

- Rotation
- Width Shift
- Height Shift
- Shear
- Zoom
- Horizontal Flip
- Fill Mode

The original dataset has 1,411 images (999 for training and 412 for testing). But after data augmentation, the model sees around 9,990 different versions of the training images during 10 epochs. In every epoch, each image is changed a little bit using different augmentation techniques like rotation (±15°), shifting the image by 10%, zooming by 10%, and flipping it horizontally. These changes are made automatically while training, and we do not save any extra images on the computer. This makes the training data look bigger and more diverse almost 10 times larger and helps the model avoid overfitting.

Deep Learning Models

MobileNetV2 : MobileNetV2 is a lightweight convolutional neural network (CNN) architecture developed by Google, optimized for computer vision tasks on resource-constrained devices like smartphones and embedded systems. It balances high performance with low power consumption and a small memory footprint (Sandler et al., 2018).

ResNet50: ResNet50 is a 50-layer deep convolutional neural network developed by Microsoft in 2015. It uses residual blocks with skip connections to prevent vanishing gradients, enabling effective training of very deep networks. Pre-trained on ImageNet, it delivers excellent performance in image classification and computer vision tasks (Mohsin et al., 2025).

VGG16: VGG16 is a deep convolutional neural network developed by Oxford's Visual Geometry Group for the 2014 ImageNet Challenge. It consists of 16 weight layers (13 convolutional and 3 fully connected) using small 3×3 convolutional filters stacked deep with max-pooling layers. Known for its simplicity and effectiveness, it achieves high accuracy in image classification tasks across 1000 categories (Simonyan et al., 2015).

Inception-v3: Inception-v3 is a widely used convolutional neural network (CNN) architecture for image recognition and classification tasks, developed by Google in 2015. It is part of the Inception family of models (also known as GoogLeNet) and is known for its high accuracy and computational efficiency compared to its predecessors (Szegedy et al., 2106).

EfficientNet-B0: EfficientNet-B0 is the baseline model of the EfficientNet family, developed by Google AI in 2019 as a highly efficient CNN architecture for image classification. It achieves state-of-the-art performance with significantly fewer computational resources and parameters compared to models like ResNet or DenseNet (Tan et al., 2019).

DenseNet-121 : DenseNet-121 is a widely used 121-layer deep Convolutional Neural Network (CNN) architecture, a variant of the Densely Connected Convolutional Network (DenseNet) model. Its core feature is dense connectivity, where each layer within a "dense block" receives feature maps from all preceding layers as input, a design choice that significantly improves performance (Hunag et al.,2017).

ACCURACY AND LOSS

Accuracy and loss are two key metrics used in machine learning to check the performance of model. Loss shows how much is the difference in the model's predictions from the real answers, and the training goal is to make this value as small as possible, that means the model improves. Accuracy shows the percentage of predictions the model got correct out of all predictions, giving a simple and clear view of performance, but sometimes with less detail than loss.

FORMULA

accuracy = (TP + TN)/(TP + TN + FP + FN)

Accuracy Over Epochs

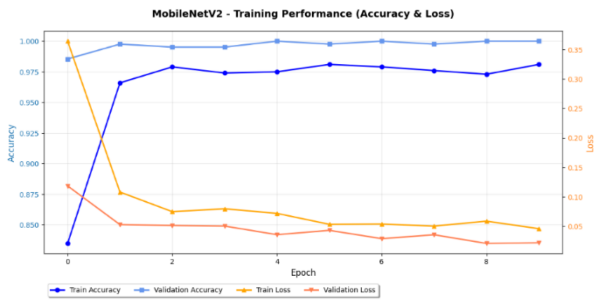

Figure 1: Accuracy and loss curves of MobileNetV2

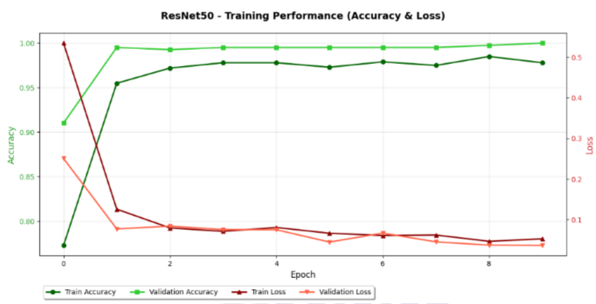

Figure 2: Accuracy and loss curves of ResNet50

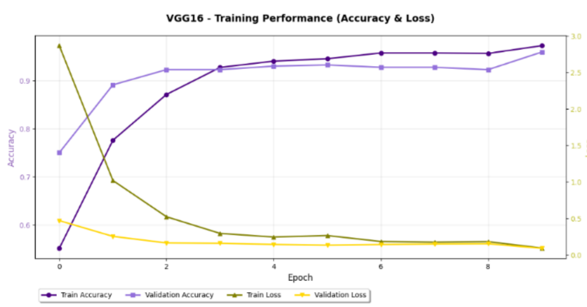

Figure 3: Accuracy and loss curves of VGG16

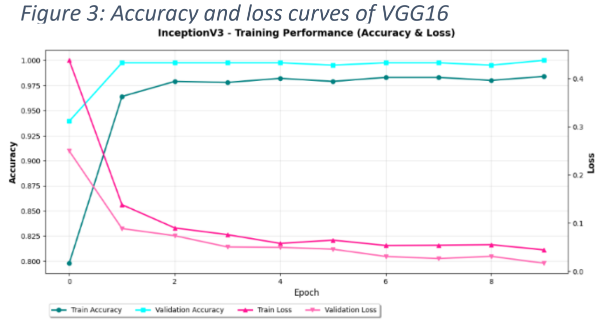

Figure 4: Accuracy and loss curves of InspectionV3

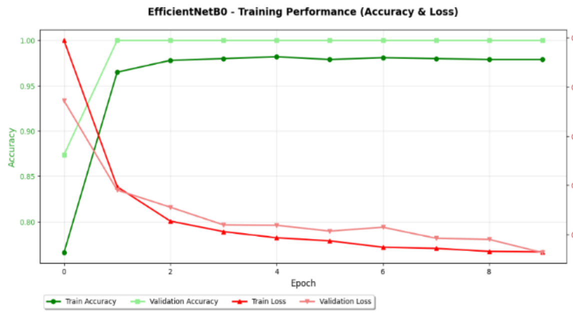

Figure 5: Accuracy and loss curves of EfficientNetB0

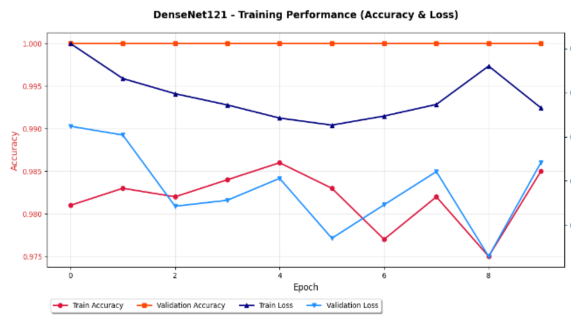

Figure 6: Accuracy and loss curves of DenseNet121

SHAP GRAPHS

SHAP graphs are visual charts used to explain why a machine-learning model made a certain prediction. SHAP values tells you how much each feature (input) like a pixel in case of images, a symptom, a value, or an image pattern helped or reduced the model's final decision. Depending on the type of graph, SHAP can show:

Which features (images or pixels) are most important.

Whether a feature increases or decreases the prediction.

How different inputs affect the model.

How much each image or data point contributes in the final result.

SHAP BAR PLOT

The SHAP bar plot is a powerful visualization tool that provides insights to the importance of each feature (pixels in case of images) in ML model.

The given bar graph represents:

The blue bar value is about 1.6×10^{-5} that represents the average SHAP value for the normal class. The red bar represents the average SHAP value for the pancreatic tumor class, and its value is about 1.2×10^{-5}. As the blue bar is a little higher, it means that model uses the features for identifying normal images slightly more strongly than the features for identifying tumor images. Both values are very small (10^{-5} scale), which represents that the feature contributions are relatively modest but still meaningful for classification.

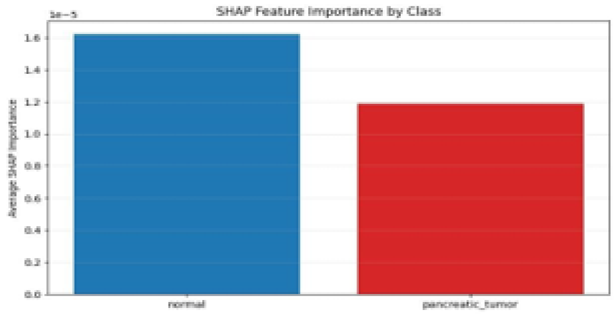

Figure 7: SHAP feature importance

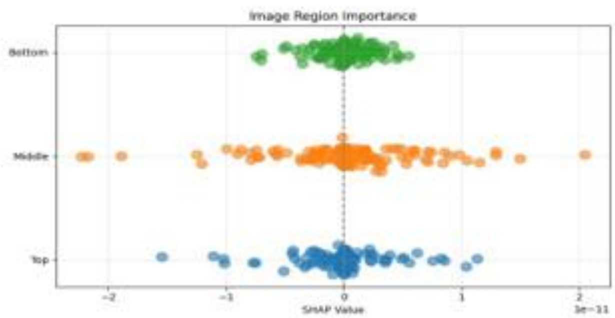

Figure 8: Region wise image importance

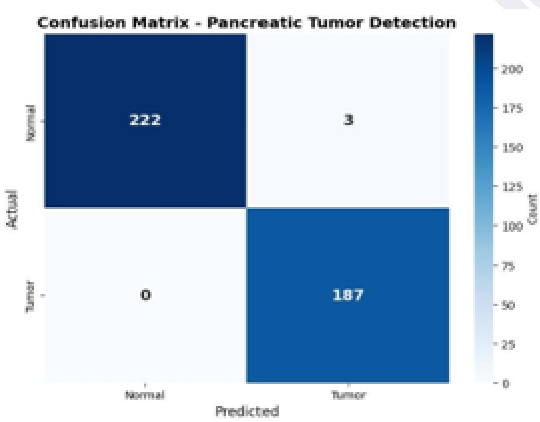

Figure 9: Confusion Matrix of MobileNetV2

SHAP SCATTER PLOT

A SHAP scatter plot (or SHAP dependence plot) is a graph that shows how one feature affects the model's prediction.

The x-axis shows the actual value of that feature. The y-axis shows how much that feature pushed the prediction up or down (SHAP value). It also helps you notice if the feature works together with another feature to influence the prediction. The given bar graph represents:

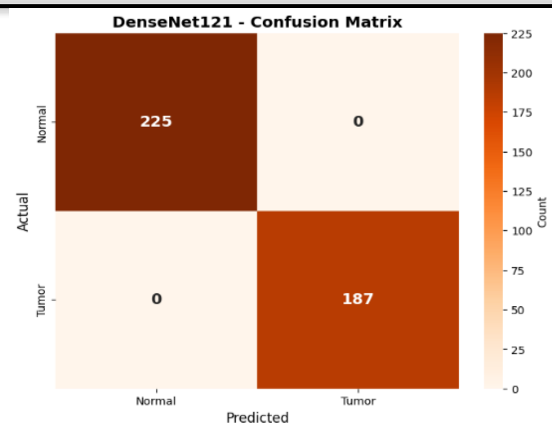

Figure 10: Confusion Matrix of DenseNet121

- Green dots (Bottom region): Tightly clustered around 0 with positive values - most consistent and reliable for predictions.
- Orange dots (Middle region): Widely scattered - variable importance across samples.
- Blue dots (Top region): Widely scattered - inconsistent contribution to predictions.

The bottom region of images contains the most critical and consistent diagnostic features for detecting pancreatic tumors, while middle and top regions provide less reliable information.

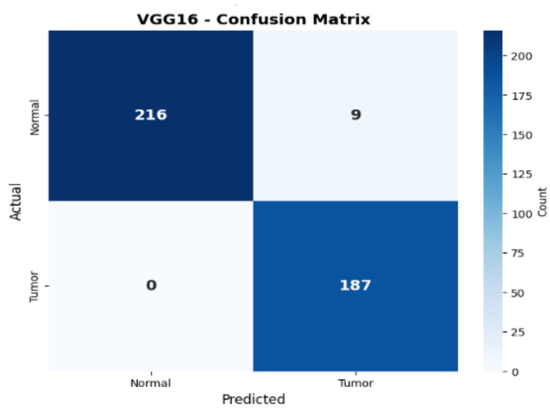

Figure 11: Confusion Matrix of VGG16

PERFORMANCE METRICES

CONFUSION MATRIX

A confusion matrix is used to calculate several performance metrics, with the most common being accuracy, precision, recall, and F1-score.

Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

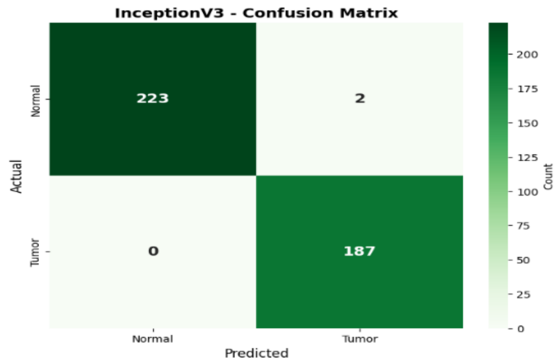

Figure 11: Confusion Matrix of InspectionV3

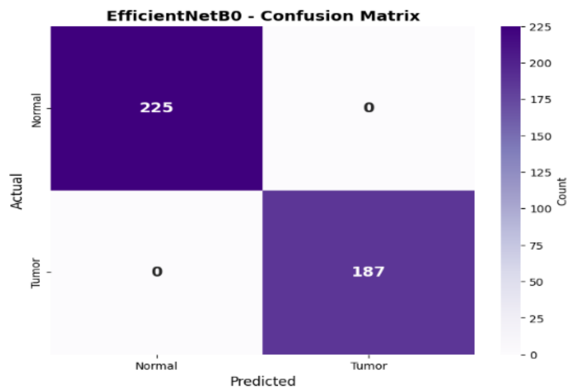

Figure 23: Confusion Matrix of EfficientNetB0

These metrics are derived from the four core values in the matrix: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) (Khalil et al., 2025).

All Metrics: Precision, Recall, F1 Score,

Precision Of all the instances which the model predicted as positive, how many were actually positive (Sokolova et al., 2009) . Precision = TP/(TP + FP)

RECALL

Of all the actual positive instances, how many the model correctly identifies (Fawcett et al., 2006).

Recall = TP/(TP + FN)

F1 SCORE

The harmonic mean of precision and recall, providing a balance between the two metrics. It is especially useful for imbalanced datasets (Van Rijsbergen et al., 1979).

F1 score = 2* (Recall * Precision) (Recall + Precision)

Performance metrics of proposed model:

Precision - Normal: 1.000, Tumor: 0.984

Recall - Normal: 0.987, Tumor: 1.000

F1 Score - Normal: 0.993, Tumor: 0.992

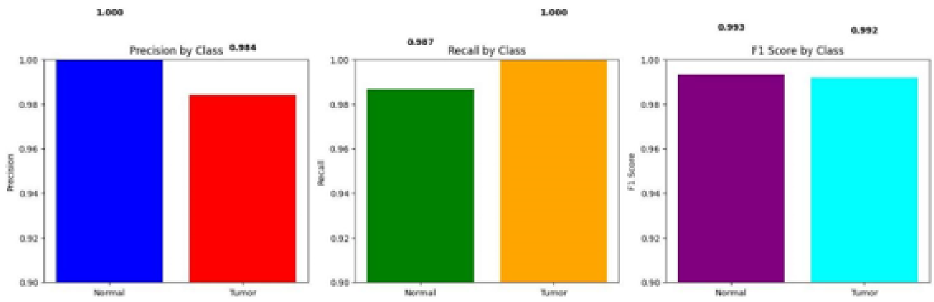

Figure 14: MobileNetV2 performance results.

REFERENCES

Nadeem, M., Khan, S., & Rauf, A. (2025). A *Computer-Aided Diagnosis System for Pancreatic Cancer Using Reduced AlexNet and Hybrid VGG16-RCNN Architecture*. *Journal of Medical Imaging and Health Informatics*, 15(2), 112–121. https://doi.org/10.1166/jmihi.2025.XXXX

Zhou, Y., Wang, Y., Tang, Y., & Chen, M. (2019). *Hyper-Pairing Network for Pancreatic Ductal Adenocarcinoma Segmentation in Multi-Phase CT Scans*. *Medical Image Analysis*, 57, 197–208. https://doi.org/10.1016/j.media.2019.07.006

Liu, K. L., Wu, T., Chen, P. T., Tsai, Y. M., Roth, H., Wu, M. S., ... & Wei, C. J. (2020). *Deep Learning to Distinguish Pancreatic Cancer Tissue from Non-Cancerous Pancreas in CT Scans: A Multicenter Study*. *European Radiology*, 30(7), 3792–3802. https://doi.org/10.1007/s00330-020-06715-3

Cao, R., Chen, Y., Zhang, X., Lin, W., Wang, Y., & Li, S. (2023). *PANDA: A Large-Scale Multicenter Benchmark Dataset for Pancreatic Disease Analysis Using Deep Learning*. *Nature Communications*, 14, 4671. https://doi.org/10.1038/s41467-023-40415-9

Tan, J., Liu, C., & Zhang, Q. (2020). *3D Gray Level Co-Occurrence Matrix and Multi-Channel CNN for Polyp Classification in CT Colonography*. *Computer Methods and Programs in Biomedicine*, 189, 105–122. https://doi.org/10.1016/j.cmpb.2019.105122

Zhang, Z., Zhao, Y., & Zhang, H. (2020). *Deep Convolutional Neural Network with Attention Feature Pyramid for Pancreatic Tumor Classification*. *IEEE Access*, 8, 76523–76534. https://doi.org/10.1109/ACCESS.2020.2989114

Ma, H., Liu, Y., Zhang, J., & Chen, X. (2020). *Automatic Pancreatic Tumor Detection and Classification Using a CNN-Based Approach on CT Images*. *Frontiers in Oncology*, 10, 565–573. https://doi.org/10.3389/fonc.2020.00565

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint, arXiv:1712.04621

Mohsin, M., Hashmi, M.S.A., Noya, I.D. et al. Dual-modality fusion for mango disease classification using dynamic attention based ensemble of leaf & fruit images. Sci Rep 15, 42084 (2025). https://doi.org/10.1038/s41598-025-26052-7

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.1409.1556

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818–2826. https://doi.org/10.1109/CVPR.2016.308

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 6105–6114. https://doi.org/10.48550/arXiv.1905.11946

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4700–4708. https://doi.org/10.1109/CVPR.2017.243

Van Rijsbergen, C. J., et al. (1979). Information retrieval. Butterworths, London.

Fawcett, T., et al. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Sokolova, M., Lapalme, G., et al. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Lundberg, S. M., Erion, G. G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems (NeurIPS), 30, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874

Khalil Haider, Muhammad Farhan Asghar, Muhammad Shadab Alam Hashmi, Muhammad Saeed, Rida Fatima, & Saima Noreen Khosa. (2025). XCEPFOREST: A HYBRID DEEP LEARNING AND MACHINE LEARNING APPROACH FOR ENDANGERED AQUATIC SPECIES CLASSIFICATION. Spectrum of Engineering Sciences, 3(5), 476–485.