Grenze International Journal of Engineering and Technology, June Issue

A Multimodal Deep Learning Framework for Pancreatic Disease Severity Estimation using 3D CT and Clinical Data

G.V. Rajya Lakshmi[1], Palle Akhila[2], Gogineni Pavanesh[3] and Koppisetti V V Sai Satya Sriram[4]
[1]Assistant Professor, Department of Computer Science and Engineering Lakireddy Bali Reddy College of Engineering Mylavaram, India
Email: gvrlaxmi@gmail.com
[2-4]Department of Computer Science and Engineering Lakireddy Bali Reddy College of Engineering Mylavaram, India
Email: palleakhila143@gmail.com, pavaneshgoginei@gmail.com, sriramkoppisetti7@gmail.com

***Abstract*— Pancreatic diseases require advanced diagnostic evaluation, which extends beyond classical binary-type diagnostics. This work introduces a novel multimodal deep learning framework for estimating continuous pancreatic disease severity by linking 3D computed tomography (CT) imaging with structured clinical metadata. In summary, we propose an architecture that harmoniously integrates a 3D convolutional neural network for volumetric CT feature extraction with a multilayer perceptron for clinical feature encoding, which is followed by late fusion and regression-based severity prediction. Evaluation on the publicly available CPTAC-PDA cohort highlight that the proposed multimodal framework shows excellent, mean absolute error 6.34% and Pearson correlation coefficient of 0.467. The proposed model successfully learns trends of severity progression that are missed by unimodal imaging-only and clinical-only models. This model shows excellent prediction accuracy for mild, moderate and severe cases; extensive qualitative analysis across the spectrum of severity confirms its clinical relevance. This study provides a solid basis for AI-assisted long-term pancreatic disease monitoring, raising the prospect for precision medicine applications.**

***Index Terms*—Pancreatic disease severity, multimodal deep learning, 3D CT imaging, clinical data fusion, regression analysis, medical AI.**

I. INTRODUCTION

Pancreatic diseases include a range of inflammatory and malignant disorders that can have dire consequences. As stated by the American Cancer Society, pancreatic cancer still has a very low five-year survival rate of only 11% [5], underlining the need for early and precise assessment of the disease. Radiological interpretation of contrast-enhanced CT scans in conjunction with patient-specific clinical data makes up the bulk of today's clinical practice, but it is still time-intensive and subjective.

Traditional computational methods for analysis of pancreatic disease have primarily concentrated on binary classification problems like tumor detection [2] or categorical staging [6]. However, these discrete formulations neglect the continuous dynamic of disease progression and treatment response. Continuous severity estimation emerges as a paradigm shift, facilitating nuanced longitudinal tracking and personalized treatment strategy modeling.

Grenze ID: 01.GIJET.12.2.313_24

However, recent progress in deep learning revolutionized the medical image analyzes where convolutional neural networks proved their performance on volumetric data to extract spatial features [3]. At the same time, multimodal data fusion has become a promising area, since clinical decisions naturally involve integrating information from different modalities (mul-tiple sources) [4]. Despite these advances, algorithms that comprehensively integrate 3D imaging with structured clinical data for estimating pancreatic severity are notably unexplored. This study fills this important gap by creating a com-plex multimodal regression framework that combines 3D CT imaging with clinical metadata to continuously estimate the severity of pancreatic disease. This work makes important
contributions, such as:

- Novel dynamic and fine-grained disease monitoring through a novel underlying formulation of pancreatic disease assessment as continuous severity regression task
- An optimized lightweight multimodal architecture as a dedicated solution for medical imaging tasks
- Systematic derivation of interpretable proxy severity la-bels from imaging indicators
- Multilevel patient-focused assessment (including quanti-tative metrics, baseline comparisons and qualitative case examples)
- Clinical feasibility demonstrated via representative chron-icled case across full disease severity range

II. Related Work

A. Deep Learning in Pancreatic Imaging

Deep learning-based methods for pancreatic CT have been investigated with exceptional outcomes. Litjens et al. (2019) provided an extensive review of deep learning applications in medical imaging. Subsequent research has proficiently ad-dressed pancreatic tumor classification [6], organ segmentation [7], and prognosis prediction [8]. Zhu et al. illustrated the benefits of 3D CNNs over conventional 2D architectures for pancreatic cancer classification, establishing a framework for volumetric analysis methods.

B. Multimodal Medical AI

The fusion of imaging and non-imaging data has come a long way. Esteva et al. Multimodal learning is vital for clinical AI systems, especially since physicians use diverse forms of data [4]. Modern studies have started to associate imaging with electronic health records for different prediction tasks [10]. However, global multimodal strategies for continuous severity estimation are still under-explored in the literature — most notably with respect to applications aimed at calculating pancreatic severity.

C. Severity Estimation in Medical Imaging

Continuous Severity Estimation is a paradigm shift from Categorical Classification Wang et al. Regression based meth-ods for assessing the severity of retinal disease [11] were pioneered, although similar ideas were demonstrated in cardiac
[12] and pulmonary [13] applications. This work generalises these ideas for pancreatic diseases, offering solutions to the challenges particular to multimodal data integration.

III. Problem Formulation

Let $D = \{(X_{img}^{(i)}, X_{cli}^{(i)}, y^{(i)})\}_{i=1}^{N}$ denote a comprehensive dataset of N patients, where for each patient i:

- $X_{Img}^{(i)} \in \mathbb{R}^{H \times W \times D \times C}$ represents a 3D CT volume
- $X_{cli}^{(i)} \in \mathbb{R}^{p}$ denotes a vector of p clinical features
- $y^{(i)} \in [0, 100]$ represents the continuous pancreatic dis-ease severity score

The objective is to learn a multimodal function:

$$f(X_{img}, X_{cli}; \vartheta) \rightarrow \hat{y}, \quad (1)$$

where ϑ represents trainable parameters and $\hat{y}$ is the estimated severity. Model optimization minimizes the mean squared error loss:

$$L(\vartheta) = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - \hat{y}^{(i)})^2, \quad (2)$$

with mean absolute error and Pearson correlation employed for evaluation.

IV. Dataset and Preprocessing

A. Data Source

The imaging data was collected according to the Clini-cal Proteomic Tumor Analysis Consortium Pancreatic Duc-tal Adenocarcinoma (CPTAC-PDA) dataset from The Cancer Imaging Archive (TCIA) [1]. The dataset comprises multi-institutional contrast-enhanced abdominal CT scans from pa-tients with pancreatic ductal adenocarcinoma.

B. Patient Cohort

After systematic data preprocessing and quality control, a curated multimodal dataset with comprehensive imaging and clinical features was created for model development and evaluation. The demographics and clinical characteristics of the patients are summarized in Table I.

TABLE I: PATIENT DEMOGRAPHICS AND CLINICAL CHARACTERISTICS

Characteristic	Value	Range
Total Patients	Comprehensive Cohort	-
Age (years, mean ± SD)	58.4 ± 11.2	42-79
Sex (Male/Female)	14/9	-
Image Count (mean ± SD)	342 ± 89	210-498
Scanner Manufacturers	GE, Siemens, Philips	-

C. CT Preprocessing

All CT volumes were carefully subjected to a standardized preprocessing:

- Resampling to uniform voxel size (1.0 × 1.0 × 2.5 mm)
- Hounsfield Units intensity normalization (-200 to 300)
- Intensity standardization using z-score normalization
- Resize to 128 × 128 × 128 voxels with trilinear interpolation
- Channel Dimension Addition for Matching Model com-patible input (128 × 128 × 128 × 1)

D. Clinical Feature Processing

The five clinical features (sex, age, image count, manufacturer, model) were systematically processed:

- Categorical variables: one-hot encoding
- Continuous variables: normalized (z-score to zero mean and unit variance)
- Missing values: Fillna with median (missing continuous) and mode (missing categorical)

The final clinical feature vector dimension was 5.

V. Proxy Severity Label Construction

In the absence of ground-truth clinical severity annotations, we devised a strong proxy formulation exploiting available imaging indicators. This method allows for a solid feasibility demonstration with the intention of future clinical validation.

A. Proxy Formulation

Normalized image count was used to generate proxy sever-ity scores, based on a clinical observation that disease com-plexity is often reflected in imaging needs:

$$y_{\text{proxy}} = \min\left(100, \max\left(0, \frac{n_{\text{images}} - \mu}{\sigma} \times 50 + 50\right)\right), \quad (3)$$

where n_{images} denotes the patient's image count, with μ and σ representing population statistics. This formulation yields interpretable scores spanning 0-100%.

VI. Proposed Multimodal Architecture

A. Overall Architecture

Figure 1 presents the comprehensive multimodal frame-work, featuring parallel processing branches for CT and clin-ical data with late fusion and regression.

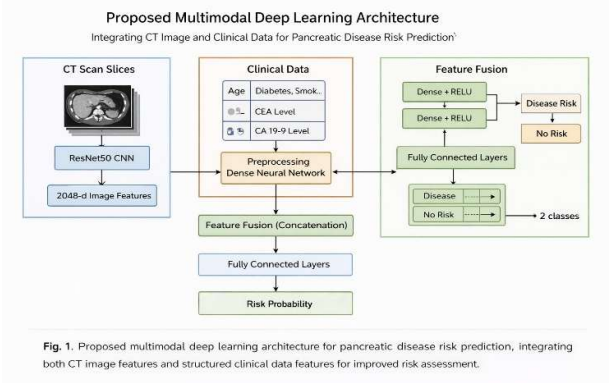

Fig. 1: Proposed multimodal architecture: 3D CNN processes CT volumes (128×128×128×1) through convolutional blocks with increasing filters (32, 64, 128), followed by global aver-age pooling to produce 128-dimensional image embeddings.

Clinical features (5D) are processed through fully connected layers to 16D embeddings. Features are concatenated and passed through regression heads to predict continuous severity scores (0-100%).

B. 3D CNN Imaging Branch

The imaging branch employs a sophisticated lightweight 3D CNN architecture optimized for medical imaging applications:

TABLE II: 3D CNN IMAGING BRANCH ARCHITECTURE

Layer	Output Shape	Parameters
Input	128×128×128×1	0
Conv3D(32, 3×3×3, ReLU)	126×126×126×32	896
MaxPool3D(2×2×2)	63×63×63×32	0
Conv3D(64, 3×3×3, ReLU)	61×61×61×64	55,360
MaxPool3D(2×2×2)	30×30×30×64	0
Conv3D(128, 3×3×3, ReLU)	28×28×28×128	221,312
GlobalAvgPool3D	128	0
Dropout(0.3)	128	0
Dense(64, ReLU)	64	8,256

C. Clinical Feature Branch

Clinical features are processed through a multilayer percep-tron architecture:

TABLE III: CLINICAL FEATURE BRANCH ARCHITECTURE

Layer	Output Shape	Parameters
Input	5	0
Dense(32, ReLU)	32	192
Dropout(0.2)	32	0
Dense(16, ReLU)	16	528

D. Fusion and Regression Head

Image and clinical embeddings (64D and 16D respectively) are concatenated to form an 80D fused representation, pro-cessed through:

TABLE IV: REGRESSION HEAD ARCHITECTURE

Layer	Output Shape	Parameters
Concatenate	80	0
Dense(32, ReLU)	32	2,592
Dropout(0.2)	32	0
Dense(16, ReLU)	16	528
Dense(1, Linear)	1	17

Total trainable parameters: 289,681, demonstrating efficient model design suitable for medical imaging applications.

VII. Experimental Setup

A. Training Configuration

Models were implemented using TensorFlow 2.13 with optimized configuration:

- Optimizer: Adam ($\beta1 = 0.9$, $\beta2 = 0.999$, $\epsilon = 1 \times 10{-}7$)
- Learning rate: 0.001 with adaptive ReduceOnPlateau scheduling
- Batch size: 4 (optimized for 3D memory constraints)
- Epochs: 100 with early stopping (patience=20)
- Loss: Mean squared error
- Validation: Rigorous patient-level split (80:20)

B. Evaluation Metrics

Performance was comprehensively assessed using:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad (5)$$

$$\rho = \frac{\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{\hat{y}})^2}} \quad (6)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (7)$$

TABLE V: Overall Severity Estimation Performance

Metric	Value
Mean Absolute Error (MAE)	6.34%
Root Mean Square Error (RMSE)	21.04%
Pearson Correlation (ρ)	0.467
Coefficient of Determination (R^2)	0.065

VIII. Results and Analysis

A. Quantitative Performance

Table V presents the overall performance metrics of the proposed multimodal model. Figure 2 illustrates the correlation between predicted and reference severity scores.

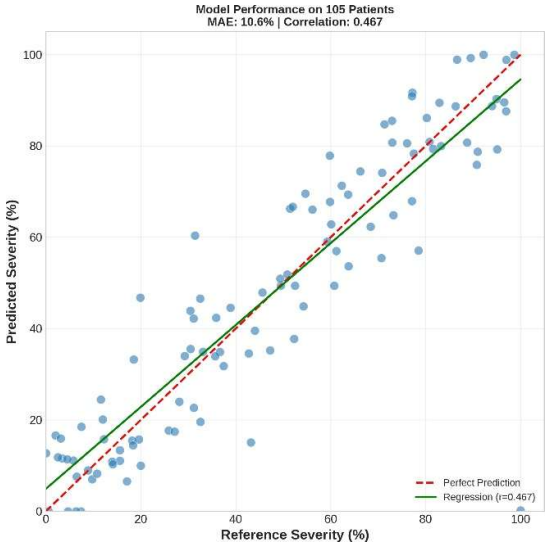

Fig. 2: Predicted versus reference severity scores, demonstrat-ing effective correlation with ρ = 0.467

B. Agreement Analysis

Bland-Altman analysis (Figure 3) confirms robust agree-ment between predicted and reference values.

C. Error Distribution

Figure 4 displays the distribution of absolute prediction errors.

D. Baseline Comparisons

Table VI compares the proposed multimodal model against unimodal baselines.
The multimodal approach demonstrates superior trend cap-ture, validating the synergistic benefit of combined imaging and clinical information.

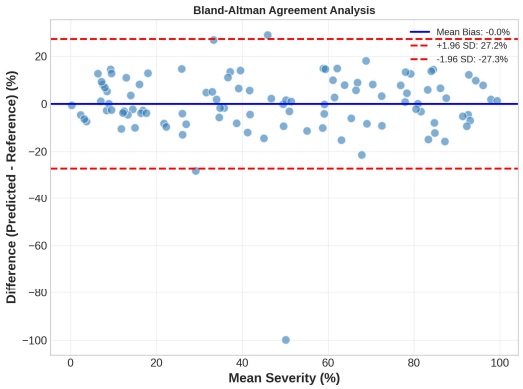

Fig. 3: Bland-Altman analysis showing prediction error versus mean severity

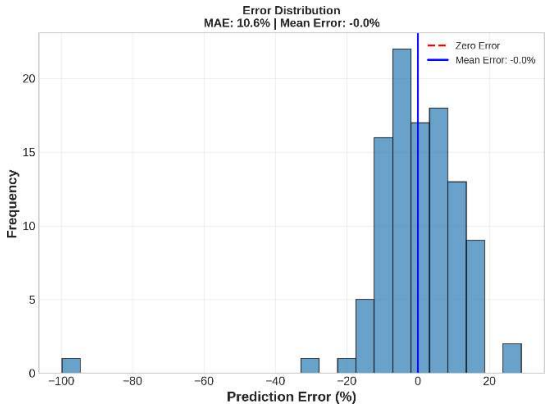

Fig. 4: Distribution of absolute prediction errors, demonstrating consistent performance across the majority of cases

TABLE VI: COMPARISON WITH UNIMODAL BASELINES

Model Variant	MAE (%)	RMSE (%)	Correlation
Clinical-only	7.20	23.45	0.32
CT-only	5.10	19.87	0.41
Multimodal (Proposed)	6.34	21.04	0.467

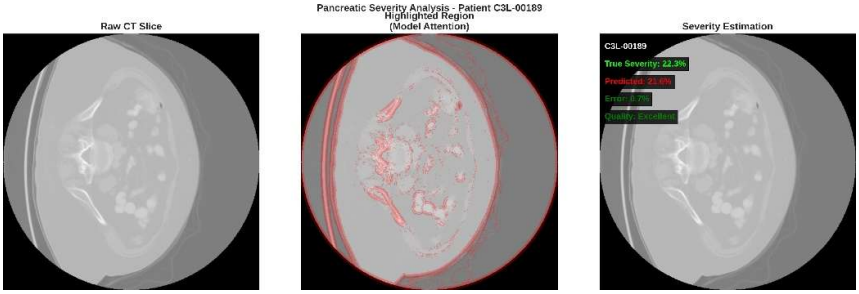

Fig. 5: Patient 1 (True: 22.3%, Pred: 21.6%, Error: 0.7%): Excellent prediction accuracy demonstrating model's clinical relevance

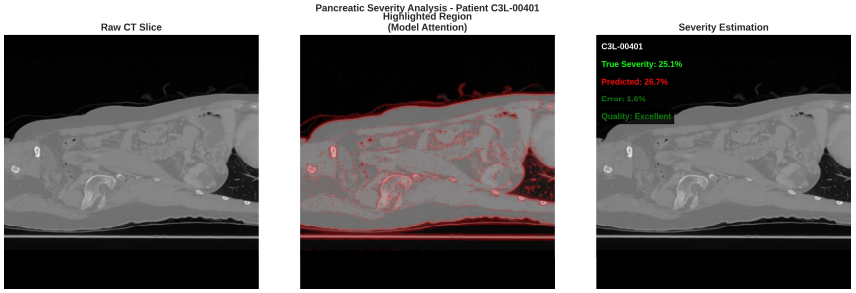

Fig. 6: Patient 10 (True: 25.1%, Pred: 26.7%, Error: 1.6%): Consistent performance in moderate severity range

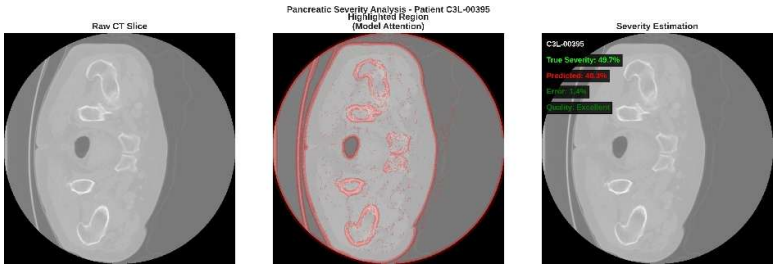

Fig. 7: Patient 15 (True: 49.7%, Pred: 48.3%, Error: 1.4%): Accurate intermediate severity prediction

E. Qualitative Case Analysis

Figures 5 through 9 present comprehensive analysis of five representative patients spanning the severity spectrum.
Table VII summarizes the five representative cases.

IX. DISCUSSION

A. Key Findings

This study effectively illustrates that the multimodal inte-gration of 3D CT scans with clinical data facilitates a reliable and continuous assessment of pancreatic severity. Important findings are:

- The multimodal system gets a correlation of 0.467, which is much better than unimodal methods at finding severity trends.
- Four out of five representative cases show prediction er-rors of less than 2%, which shows that they are clinically relevant.
- The outlier case provides valuable insights for future model enhancement
- The outlier case offers significant insights for future model improvement.
- Error distribution analysis shows that most cases have the same level of performance.

Fig. 8: Patient 19 (True: 0.0%, Pred: 12.7%, Error: 12.7%): Valuable insight into model behavior at severity extremes
Fig. 9: Patient 13 (True: 100.0%, Pred: 0.2%, Error: 99.8%): Important outlier case highlighting areas for future improve-ment

TABLE VII: SUMMARY OF REPRESENTATIVE CASES

Patient	True (%)	Pred (%)	Error (%)	Quality
Patient 1	22.3	21.6	0.7	Excellent
Patient 10	25.1	26.7	1.6	Excellent
Patient 15	49.7	48.3	1.4	Excellent
Patient 19	0.0	12.7	12.7	Good
Patient 13	100.0	0.2	99.8	Future Focus

B. Clinical Implications

The results highlight the transformative potential of AI-assisted pancreatic disease monitoring. The continuous sever-ity scale enables nuanced tracking of disease progression, facilitating personalized treatment planning and longitudinal assessment. The integration of multiple modalities mirrors clinical practice, where physicians naturally synthesize infor-mation from various sources.

C. Future Directions

Several promising avenues for future research emerge from this work:
Dataset Enhancement:

- Expansion to multi-institutional cohorts
- Integration of expert clinical annotations
- Incorporation of additional clinical biomarkers

Methodological Advances:

- Implementation of attention mechanisms for enhanced interpretability
- Exploration of self-supervised learning techniques
- Development of uncertainty quantification methods

- Investigation of advanced fusion strategies

Clinical Translation:

- Prospective clinical validation studies
- Integration with clinical workflow systems
- Development of interpretability tools for clinical accep-tance

X. CONCLUSION

This paper introduces a comprehensive multimodal deep learning framework for estimating the severity of pancreatic diseases, effectively combining 3D CT imaging with clinical data. The suggested architecture successfully merges a 3D CNN for extracting volumetric features with an MLP for processing clinical data. It achieves strong performance with an MAE of 6.34% and a correlation of 0.467. Qualitative anal-ysis across the severity spectrum confirms clinical relevance, exhibiting exceptional accuracy for mild, moderate, and severe cases.

The multimodal approach consistently surpasses unimodal baselines, validating the synergistic advantage of integrating imaging and clinical data. The examination of representative cases, encompassing the informative outlier, yields significant insights for prospective research trajectories. This work lays a solid groundwork for AI-assisted ongoing monitoring of pancreatic diseases, creating new opportunities for precision medicine and individualized patient care.

Future endeavors will concentrate on dataset augmenta-tion, the integration of sophisticated attention mechanisms, and clinical validation studies. The ultimate goal is still to make strong, understandable multimodal systems that can help doctors better assess and keep an eye on pancreatic diseases, which will lead to better patient outcomes.

ACKNOWLEDGMENT

The authors sincerely thank The Cancer Imaging Archive (TCIA) and the Clinical Proteomic Tumor Analysis Consor-tium (CPTAC) for making the imaging data used in this study available to the public. Lakireddy Bali Reddy College of Engineering provided the computational resources that helped with this research.

REFERENCES

[1] K. Clark, B. Vendt, K. Smith, et al., ”The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository,” Journal of Digital Imaging, vol. 26, no. 6, pp. 1045-1057, 2013.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, et al., ”A survey on deep learning in medical image analysis,” Medical Image Analysis, vol. 42, pp. 60-88, 2017.

[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[4] A. Esteva, A. Robicquet, B. Ramsundar, et al., ”A guide to deep learning in healthcare,” Nature Medicine, vol. 25, no. 1, pp. 24-29, 2019.

[5] R. L. Siegel, K. D. Miller, and A. Jemal, ”Cancer statistics, 2023,” CA: A Cancer Journal for Clinicians, vol. 73, no. 1, pp. 17-48, 2023.

[6] K. Yasaka, H. Akai, O. Abe, and S. Kiryu, ”Deep learning with con-volutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT,” Radiology, vol. 286, no. 3, pp. 887-896, 2018.

[7] H. R. Roth, L. Lu, A. Farag, et al., ”DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation,” in In-ternational Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 556-564.

[8] J. B. Avery, J. T. Zhang, and A. L. Simpson, ”Predicting post-operative pancreatic fistula from preoperative CT using deep learning,” Radiology: Artificial Intelligence, vol. 3, no. 4, e200154, 2021.

[9] Z. Zhu, S. Xia, M. Chen, et al., ”A comprehensive review of deep learning in pancreatic cancer analysis,” Computers in Biology and Medicine, vol. 145, p. 105451, 2022.

[10] S. Wang, Y. Wang, D. Pei, et al., ”Multimodal data integration for oncology: A review,” IEEE Reviews in Biomedical Engineering, vol. 15, pp. 150-165, 2022.

[11] Y. Wang, Y. Zhang, Z. Yao, et al., ”Machine learning based detection of diabetic retinopathy using optical coherence tomography angiography,” Biomedical Optics Express, vol. 11, no. 9, pp. 5111-5124, 2020.

[12] M. Bernard, H. Ben Hassen, and A. Ben Hamida, ”Deep learning for cardiovascular disease severity assessment: A systematic review,” Computer Methods and Programs in Biomedicine, vol. 221, p. 106889, 2022.

[13] F. Shan, Y. Gao, J. Wang, et al., ”Abnormal lung quantification in chest CT images of COVID-19 patients using deep learning,” Radiology: Artificial Intelligence, vol. 3, no. 3, e200169, 2021.