Multimodal Deep Learning for Pulmonary Embolism Prognosis Prediction

Jayadev R[1*], Muhammed Rasin O M[2*], Subin A[3*], Prof. Rotney Roy Meckamalil[4*]
*Department of Computer Science and Engineering
Mar Athanasius College of Engineering, Kothamangalam, Kerala
[1]jayadev2002r@gmail.com, [2]rasinbinabdulla@gmail.com, [3]subinkoloth7@gmail.com, [4]rotney.rotney@gmail.com

***Abstract*—Pulmonary embolism (PE) denotes a crucial medical condition marked by the sudden blockage of the pulmonary artery due to blood clots, demanding immediate attention. Despite progress in clinical methodologies utilizing deep learning models focused on pixels, valuable insights contained in Electronic Health Records (EHRs) often remain unexplored. When employing deep learning to categorize PE in computed tomography pulmonary angiography (CTPA), significant hurdles arise due to the intricacy of CTPA assessments. To tackle this challenge, integrating clinical and imaging data using diverse methods becomes essential. Drawing from the 'RadFusion Dataset,' our research investigates the effectiveness of combining TabNet for clinical data and PENet for imaging data, with the aim of surpassing current benchmarks. Advanced fusion methods are utilized to improve prediction accuracy and comprehensibility. An analysis of feature importance is conducted to create a prototype for real-time clinical decision support. This endeavor aims to enhance the accuracy of PE prognosis, bridging the divide between deep learning models and clinical applications, thereby propelling AI utilization in medical research and consultations.**

***Index Terms*—Pulmonary Embolism, Deep Learning, Multimodal fusion, Clinical Decision Support**

I. Introduction

Pulmonary embolism (PE) presents a significant danger, characterized by the abrupt blockage of pulmonary arteries, typically triggered by blood clots originating from various parts of the body, notably the legs. This blockage leads to serious respiratory and cardiovascular complications, necessitating timely and strategic intervention for effective treatment. A crucial element of this approach involves risk assessment, which guides diagnostic and therapeutic decisions. In everyday clinical settings, healthcare professionals heavily rely on patients' electronic health records (EHRs) to provide context for interpreting medical imaging results. Traditionally, practitioners consider either clinical or imaging data alone. However, prevalent deep learning models in radiology often concentrate solely on pixel values, overlooking valuable insights within comprehensive health records.

Although deep learning has shown promise in numerous medical imaging applications, its application to automated pulmonary embolism (PE) classification in computed tomography pulmonary angiography (CTPA) studies poses distinctive challenges. CTPA examinations, with their larger scale compared to traditional medical imaging tests, necessitate a nuanced approach. The pixel data associated with PE findings represents only a small portion of the extensive 3D CTPA volume. Therefore, integrating patient electronic health record (EHR) data with imaging data becomes essential for efficacy. This underscores the importance of employing multimodal approaches in medical research.

The incorporation of multimodal methodologies, such as amalgamating clinical and imaging data in pulmonary embolism (PE) prediction models, carries profound importance. These models provide a holistic insight into patient conditions by amalgamating a range of information sources, thereby refining diagnostic precision and guiding personalized treatment approaches. Nonetheless, the scarcity of openly accessible datasets presents a hurdle in crafting efficient PE prediction models. Additionally, many prevailing deep learning frameworks are trained on limited datasets.

The paper titled 'RadFusion' addresses this void by introducing a comprehensive benchmark dataset named 'RadFusion' for detecting pulmonary embolism, integrating 3D medical imaging with electronic health record (EHR) data from 1794 patients. To the best of our knowledge, this dataset is currently the largest publicly accessible multimodal dataset. The authors advocate for exploring multimodal fusion techniques for pulmonary embolism prognosis, a domain where research lags behind that of other prevalent diseases. Their research indicates that instead of solely enhancing medical image representation models, there should be a concerted effort to improve multimodal fusion models. However, a cursory examination reveals that many other multimodal fusion models showcasing superior performance are often trained and evaluated on smaller datasets, as the limited availability of confidential medical data remains a significant obstacle for researchers in this field.

This paves the way for our project, which seeks to tackle specific challenges in pulmonary embolism prognosis through an exploration of top-performing deep learning models in their respective domains—namely TabNet and PENet. Our objective is to extract intricate features from both clinical and medical imaging data, revealing hidden patterns crucial for prognosis. By harnessing the strengths of TabNet and PENet, our proposed deep learning model employs fusion mechanisms to enhance overall predictive capacity. Our endeavor is to construct an effective deep learning framework, aiming to assess and surpass the benchmark established by the RadFusion dataset.

In our endeavor, we endeavor not only to construct a resilient multimodal fusion model for effective pulmonary embolism (PE) prognosis prediction but also to craft a prototype tool designed to assist clinical decision-making. This tool harnesses the capabilities of natural language processing (NLP) through the aid of large language model (LLM) tools, empowering healthcare practitioners to seamlessly integrate electronic health record (EHR) data into the predictive framework. The envisioned interface enables physicians to input the EHR report, upload pertinent computed tomography pulmonary angiography (CTPA) data, and initialize the model.

II. Related Research

Previous studies in the detection of pulmonary embolism (PE) have primarily focused on utilizing either imaging data or clinical records in isolation, which may lead to constraints in diagnostic accuracy and interpretability. Consequently, recent investigations have delved into the amalgamation of Electronic Health Records (EHRs) with imaging data through multimodal fusion methodologies.

Several earlier works have explored the utilization of deep learning techniques for the detection of pulmonary embolism (PE) using diverse imaging modalities such as computed tomography pulmonary angiography (CTPA). For instance, Tajbakhsh et al. [1] showcased the effectiveness of 2D convolutional neural networks in PE detection. Subsequent studies have employed various approaches, ranging from 3D convolutional neural networks to architectures based on ResNet and DenseNet. Recent advancements in deep learning architectures have led to the development of specialized models for analyzing medical imaging data, promising improved accuracy and efficiency. Huang et al. [5] introduced PENet, a cutting-edge deep learning model optimized for pulmonary embolism detection in CTPA images, demonstrating superior performance compared to conventional methods.

Multimodal fusion techniques have emerged as a promising avenue to harness complementary information from diverse data sources, thereby enhancing the robustness and interpretability of predictive models. Chen et al. [9] utilized deep learning and machine learning for type 2 diabetes risk prediction using various modalities. Similarly, Zhou et al. [10] proposed a multi-layer framework for Alzheimer's detection and progression employing multiple modalities (imaging, clinical, etc.) with a focus on explainability. The integration of Electronic Health Data with Imaging CTPA data was highlighted by Zhou et al. [7].

Tabular data processing techniques have found widespread application in medical diagnosis tasks, particularly for managing structured clinical data extracted from EHRs. Arik et al. [11] introduced TabNet, a novel deep learning architecture customized for tabular data, demonstrating its efficacy in disease prediction tasks with high-dimensional clinical data. Given that the dataset provided by the RadFusion study is in CSV file format, our objective was to enhance predictions using a model that achieves state-of-the-art performance in processing tabular data.

Ensemble learning methodologies have been deployed for decision-making and classification tasks. Mienye and Sun [14] conducted a comprehensive survey on ensemble learning, offering insights into concepts, algorithms, applications, and future prospects. In our investigation, we employed a decision-level late fusion strategy using stacking, wherein multiple meta-learner models were utilized and the one yielding the highest accuracy was selected.

III. Methodology

In this methodology, we present a comprehensive approach for developing a multimodal fusion model aimed at pulmonary embolism (PE) classification, integrating information from both imaging data and electronic health records (EHR).

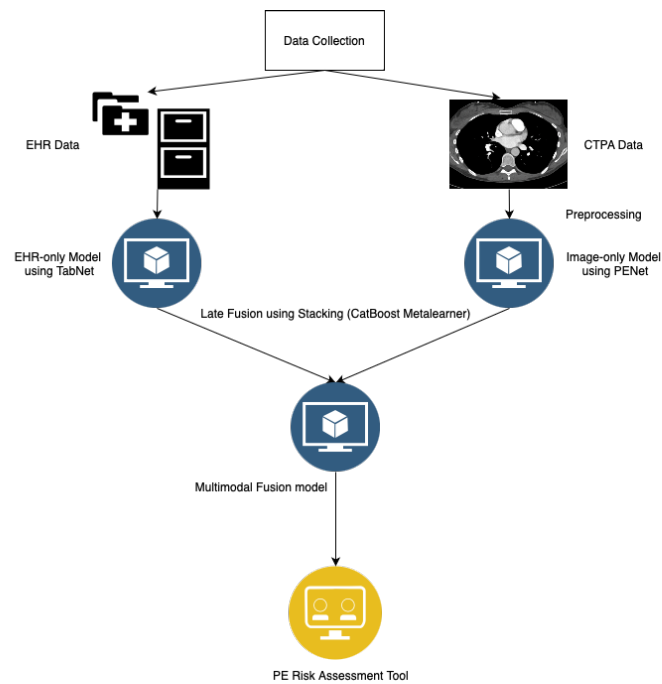

Fig. 1. Proposed Method

A. Dataset

The dataset utilized in this study was obtained from the RadFusion database [7], a comprehensive collection of medical data curated for research purposes, focusing on pulmonary embolism (PE) detection. Approval for dataset access was obtained from Stanford University Medical Center (SUMC) with the necessary ethical clearances from the Stanford Institutional Review Board (IRB). The RadFusion dataset includes both imaging data from CT scans and

Electronic Health Records (EHRs) of patients. The dataset provides detailed information on patient demographics, vitals, medications, International Classification of Diseases (ICD) codes, lab test results, and annotations related to PE diagnosis. Annotations include the classification of PE subtypes and the identification of slices containing PE lesions. The dataset was then partitioned into training, validation, and testing sets in an 80% /10%/10% split, ensuring no patient overlap between the sets. The RadFusion dataset serves as a valuable resource for developing and evaluating machine learning algorithms for PE detection. It enables the exploration of multimodal fusion models using both CT scans and patient EHRs, contributing to advancements in medical image interpretation and clinical decision support systems. Researchers can utilize the dataset to study model fairness across different demographic groups and investigate algorithmic bias in medical AI applications.

1) Data Characteristics: Table I summarizes the data characteristics of the RadFusion dataset, providing statistics for the training, validation, and testing sets. These characteristics include the number of studies, patients, slices, and the distribution of positive and negative PE cases. Additionally, vital signs, D-dimer tests, and BMI information are provided, offering insights into patient health records. The dataset is enriched with patient demographics, enabling studies on model fairness and demographic-based analysis.

B. Image-only Model

For the image-only model, we utilize the PENet architecture. PENet excels in processing volumetric computed tomography pulmonary angiography (CTPA) scans, leveraging spatial information across multiple slices to detect subtle abnormalities indicative of PE. The architecture comprises four key components: the PENet unit, Squeeze-and-Excitation (SE) block, PENet bottleneck, and PENet encoder.

The PENet unit serves as the fundamental building block, employing 3D convolutions, group normalization, and LeakyReLU activation to extract features from the input data. The SE block enhances feature recalibration by adaptively weighting channel-wise information, promoting discriminative feature learning. Furthermore, the PENet bottleneck aggregates multiple PENet units to form a hierarchical feature extractor, facilitating the representation of complex spatial patterns.

The PENet encoder, consisting of stacked PENet bottlenecks followed by GapLinear activation, culminates in the final prediction. Notably, the depth of the network is optimized through cross-validation, ensuring a balance between model complexity and generalization performance.

In preparation for model training, we preprocess the CTPA image data to enhance model efficacy. Each CT scan undergoes preprocessing steps, including resizing each slice to 224 × 224 pixels, and applying an optimized viewing window centered around pulmonary arteries (window center = 400, window width = 1000). Additionally, we clip Hounsfield Units to the range of -1000 to 900 and normalize each CT scan to be zero-centered. The training strategy incorporates techniques to address class imbalance, such as binary cross-entropy focal loss and up-sampling of positive windows. Data preprocessing methods, including random cropping, rotation, and jittering, are employed to improve model robustness.

C. EHR Model

In developing the EHR model, we employ the TabNet algorithm, an advanced deep learning framework tailored specifically for processing tabular data. Distinguished from conventional regression models, TabNet features a unique architecture that seamlessly integrates feature selection and decision-making mechanisms, making it particularly adept at handling high-dimensional and sparse input features characteristic of EHR datasets. By leveraging the regularization penalties of L1 and L2, TabNet strikes an optimal balance between model complexity and generalization capacity, thereby enhancing its proficiency in discerning meaningful patterns inherent in the data.

Furthermore, the EHR data from disparate CSV files are consolidated, followed by an initial training phase wherein a TabNet model is trained. Subsequently, leveraging the feature importance matrix, non-contributing columns are pruned, and the TabNet model is retrained. This iterative process optimizes the model's efficacy by focusing solely on salient features conducive to accurate predictions. Finally, the feature importance analysis is conducted to ascertain the significance of individual features in influencing model predictions, thereby offering valuable insights for clinical decision-making.

D. Multimodal Fusion

In this section, we outline the construction of our multimodal fusion model, which integrates predictions from individual models trained on both imaging data and Electronic Health Records (EHRs). Our primary fusion strategy revolves around late fusion, also known as decision-level fusion, wherein the predictions from individual models are combined using various meta-learning algorithms, including TabNet, XGBoost, and CatBoost. This approach enables us to leverage the complementary strengths of each modality while mitigating their respective limitations, thereby enhancing overall prediction performance.

Late fusion involves aggregating the predictions generated by individual models at the decision level, typically by averaging or applying a weighted combination of the model outputs. We experiment with several late fusion techniques to determine the most effective fusion strategy for our multimodal model. Additionally, we explore the use of ensemble learning methods, such as stacking, to further improve prediction accuracy and robustness.

Category	Sub-category	Overall	Train	Validation	Test
CTPA exams	# of studies	1837	1454	193	190
	# of patients	1794	1414	190	190
	Median # of slices (IQR)	386 (134)	385 (136)	388 (132)	388 (139)
PE	# of negative PE	1111(60.48%)	946 (65.06%)	85 (44.04%)	80 (42.10%)
	# of positive PE	726 (39.52%)	508 (34.94%)	108 (55.96%)	110 (57.89%)
	Central	257(35.40%)	202 (39.76%)	27 (25.00%)	28 (25.45%)
	Segmental	387(53.31%)	281 (55.31%)	52 (48.15%)	54 (49.09%)
	Subsegmental	82 (11.29%)	25 (4.91%)	29 (26.85%)	28 (25.45%)
Vitals	BMI (mean: std)	28.37 : 9.65	28.36 : 10.03	27.11 : 6.78	29.60 : 9.22
	Pulse (mean: std)	81.62 : 14.99	81.53 : 15.64	83.05 : 11.86	80.50 : 13.06
D-dimer	D-dimer test taken	580 (30.62%)	461 (30.90%)	58 (28.71%)	61 (30.50%)
	D-dimer positive	496 (26.18%)	389 (26.07%)	51 (25.25%)	56 (28.00%)

TABLE I
DATA CHARACTERISTICS OF THE RADFUSION DATASET.

Algorithm 1 outlines the late fusion process used in our multimodal model. Given the predictions from individual models for a given sample, the algorithm combines these predictions using a specified fusion method, such as simple averaging or weighted averaging based on model performance. The fused prediction is then used as the final output of the multimodal model.

Algorithm 1: Stacking Method for Multimodal Fusion

Input: Training datasets: D_{EHR}, D_{CTPA}
Output: Fused prediction: P_{fused}

Train EHR Model: Train the EHR model using the TabNet algorithm on the EHR training dataset D_{EHR};

Train CTPA Model: Train the CTPA model using the PENet architecture on the CTPA training dataset D_{CTPA};

Generate Predictions: Generate predictions P_{EHR} and P_{CTPA} from the trained EHR and CTPA models, respectively, for the validation/test datasets;

Combine Predictions: Combine the predictions P_{EHR} and P_{CTPA} to create a combined feature matrix;

Train Meta-Learner Model: Train a meta-learner model (e.g., XGBoost, CatBoost) on the combined feature matrix to learn the relationship between individual model predictions and the target variable;

Generate Fused Prediction: Generate the fused prediction P_{fused} by feeding the predictions from the EHR and CTPA models into the trained meta-learner model;

Return: P_{fused}

Furthermore, we evaluate the performance of the multimodal fusion model using appropriate metrics, such as accuracy to assess its effectiveness in predicting pulmonary embolism. Through systematic experimentation and analysis, we aim to identify the optimal fusion strategy and meta-learning algorithm combination that maximizes prediction accuracy while ensuring robustness. The multimodal fusion model represents a significant advancement in leveraging heterogeneous data sources for medical diagnosis, offering potential benefits in enhancing diagnostic accuracy and clinical decision-making in the context of pulmonary embolism detection.

E. Clinical Decision Support Interface

In order to demonstrate the practical application of our multimodal fusion model for clinical decision support, we developed an interactive interface prototype. This interface serves as a tool for healthcare professionals to input patient data and obtain predictions for pulmonary embolism (PE) risk.

For the Electronic Health Record (EHR) input, we first identify the highest feature importance columns from the TabNet model. These columns represent the most influential factors in predicting PE risk based on the EHR data. We then design a set of question-answer-based inputs corresponding to these key features. Healthcare professionals can input patient information through this interface, answering questions related to vital signs, medical history, and other relevant factors.

The interface dynamically processes the user inputs and generates predictions using our multimodal fusion model, thereby eliminating the need to input all the features upon which the model is trained in. By leveraging both EHR and imaging data, our model provides a comprehensive assessment of PE risk for the given patient. The output is displayed to the user, indicating the predicted likelihood of PE occurrence along with any additional diagnostic insights.

This prototype interface demonstrates the potential of integrating advanced machine learning models into clinical practice for improved decision-making support. By leveraging patient data and state-of-the-art algorithms, healthcare professionals can make more informed decisions and enhance patient care outcomes.

IV. Result

A. Model Performance

The Multimodal Models integrate both image and EHR data for improved predictive performance. The CatBoost Fusion model outperformed other models with an accuracy of 0.93. It achieved a precision of 0.87 and a recall of 0.95. The F1 score for this model is 0.90, and the AUC is 0.96, indicating excellent discriminatory capability. Overall, the results indicate that the CatBoost Fusion model, leveraging multimodal data integration, achieves the highest performance across all evaluated metrics. This is in comparison with other decision level late fusion models using meta-learners of XGBoost, TabNet, and CatBoost models. Compared to the RAD Fusion benchmark, our multimodal fusion models demonstrated significant improvements in accuracy across all evaluated metrics. Specifically, surpassing the benchmark by 5 percentage points. This notable enhancement underscores the effectiveness of our late decision fusion approach in leveraging complementary information from different modalities to enhance predictive performance.

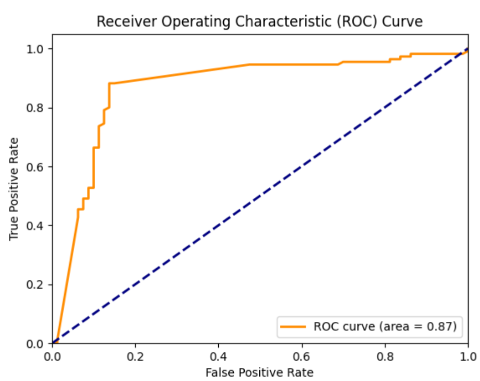

Fig. 2. Receiver Operating Characteristic (ROC) Curve

Figure 3 showcases the prediction trend for about 40 samples on how the prediction scores can be compared to their actual labels for the PENet model, the TabNet model, and the Multimodal model respectively. It can be understood that the predictions done by the multimodal model align more with the pattern of the actual labels.

B. Feature Importance

Our feature importance analysis revealed key insights into the predictive power of features extracted from the EHR data. Notably, diseases of pulmonary circulation emerged as the most influential feature, with a feature importance score of 0.041. This highlights the critical role of pulmonary

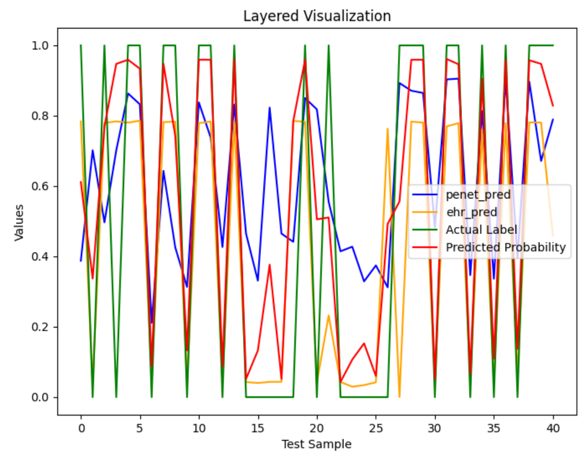

Fig. 3. Prediction trend of different validation samples using different models in comparison to their actual labels

health indicators in disease prediction and underscores the importance of monitoring and managing pulmonary-related conditions in clinical settings.

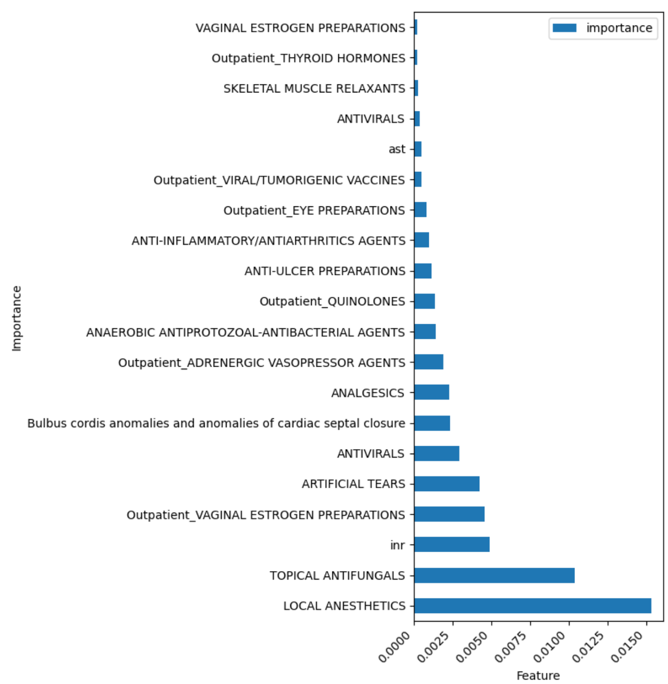

Fig. 4. Feature Importance other than Diseases of Pulmonary Circulation

Furthermore, local anesthetics and topical antifungals were identified as important features, with feature importance scores of 0.015 and 0.010, respectively. These findings suggest that medications and treatments related to specific medical conditions play a significant role in predicting disease outcomes, emphasizing the importance of incorporating medication history and treatment plans into predictive models for enhanced accuracy.

TABLE II
PERFORMANCE COMPARISON OF MODELS

Model	Accuracy	Precision	Recall	F1 Score	AUC
Image-Only Models					
PENet	**0.77**	**0.81**	**0.73**	**0.77**	**0.84**
EHR-Only Models					
TabNet	**0.85**	**0.84**	**0.90**	**0.86**	**0.87**
ElasticNet	**0.83**	**0.90**	**0.80**	**0.84**	**0.92**
Multimodal Models					
XGBoost Fusion	**0.88**	**0.83**	**0.94**	**0.86**	**0.85**
TabNet Fusion	**0.88**	**0.83**	**0.94**	**0.86**	**0.85**
CatBoost Fusion	**0.93**	**0.87**	**0.95**	**0.90**	**0.96**

Additionally, the inclusion of INR (International Normalized Ratio) as a feature with a feature importance score of 0.005 underscores the relevance of coagulation status in disease prediction, particularly in conditions where blood clotting disorders may influence clinical outcomes. Overall, our feature importance analysis provides valuable insights into the factors driving predictive performance in our EHR-based model, offering clinicians and researchers a deeper understanding of the underlying mechanisms influencing disease prognosis and treatment outcomes.

The visualization presented in the result section illustrates the alignment between the model's predictions and the actual labels using the validation set. The plot showcases the values of the PENet predictions and TabNet prediction features alongside the corresponding actual labels and predicted probabilities. This graphical representation offers a comprehensive overview of the model's performance, allowing for an intuitive understanding of its predictive capabilities. By visually comparing the predicted probabilities with the actual labels, one can assess the model's accuracy and its ability to correctly classify instances.

V. DISCUSSION

A. Limitations

1) **Availability of Multimodal Dataset**: One of the primary limitations of this study is the scarcity of publicly available multimodal datasets. While multimodal datasets offer a rich source of information and potential insights, they are often challenging to obtain due to privacy concerns, data access restrictions, and the need for integration across multiple data sources.
2) **Consistency in Features**: Another limitation is the lack of consistency in features across modalities. Different modalities may have varying levels of completeness and consistency, leading to challenges in feature engineering and model training. This inconsistency can introduce bias and affect the performance of multimodal models.
3) **Ethical and Regulatory Considerations**: Working with multimodal healthcare data involves ethical and regulatory considerations, including privacy protection, informed consent, and compliance with data protection regulations such as GDPR and HIPAA. These considerations may limit the accessibility and usability of healthcare datasets for research purposes.

B. Further Discussion and Possible Improvements

1) **Data Collection and Integration**: Efforts should be made to collect and integrate multimodal data from diverse sources, including healthcare institutions, research organizations, and data-sharing initiatives. Collaborations and partnerships can facilitate access to a broader range of data types, enabling more comprehensive analyses and model development.
2) **Feature Engineering and Harmonization**: Improving the consistency and quality of features across modalities is essential for the development of robust multimodal models. Standardization protocols, such as ontologies and data dictionaries, can be employed to ensure uniformity in data representation and semantics. Advanced feature engineering techniques and data preprocessing methods can help address discrepancies and enhance the compatibility of multimodal features.
3) **Model Robustness and Generalization**: Enhancing the robustness and generalization of multimodal models is crucial for their real-world applicability. Techniques such as transfer learning, domain adaptation, and model ensembling can leverage knowledge from related tasks or domains to improve performance on target tasks. Additionally, model interpretability methods can provide insights into the decision-making process of multimodal models, enhancing trust and transparency.
4) **Ethical and Regulatory Compliance**: Ethical and regulatory considerations must be carefully addressed when working with multimodal healthcare data. Privacy protection, informed consent, data anonymization, and compliance with data protection regulations are critical aspects that require attention to ensure the responsible and ethical use of sensitive patient information.

In conclusion, while multimodal datasets hold great promise for advancing healthcare research and improving patient outcomes, addressing the challenges associated with data availability, consistency, and ethical considerations is essential for realizing their full potential. By overcoming these limitations and implementing appropriate strategies for improvement, multimodal models can play a significant role in personalized medicine, disease diagnosis, and treatment planning.

VI. Conclusion

In this study, we explored various machine learning models for the task of PE detection. We experimented with PENet, TabNet, XGBoost, CatBoost, etc., leveraging both individual modalities and fusion techniques to improve predictive performance. Through comprehensive experimentation and evaluation, several key findings have emerged.

We conducted a thorough performance comparison of different models, evaluating metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. Our results indicate that the decision-level fusion (Late fusion) model using the CatBoost meta-learner and the PENet+TabNet individual models performs the best, achieving a high accuracy of 0.93 on the test set II.

This finding suggests that leveraging both PENet and TabNet models in combination with decision-level fusion can significantly enhance the accuracy of PE detection, offering potential benefits for clinical diagnosis and patient care.

However, it's essential to acknowledge the limitations of our study. We faced challenges related to the availability of multimodal datasets and ensuring consistency in the features across different modalities. Additionally, while our best-performing model demonstrates promising results, further research is needed to address model interpretability and generalizability issues.

In conclusion, our study highlights the efficacy of machine learning models and fusion techniques for PE detection. Future research directions may involve refining existing models, incorporating additional features or modalities, and conducting prospective clinical validation studies to assess the real-world performance of these models.

References

[1] Tajbakhsh, Nima, Michael B. Gotway, and Jianming Liang. ”Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks”. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part II 18. Springer International Publishing, 2015.

[2] Nima Tajbakhsh, Jae Y. Shin, Michael B. Gotway, Jianming Liang, ”Computer-aided detection and visualization of pulmonary embolism using a novel, compact, and discriminative image representation”, Medical Image Analysis, Volume 58, 2019.

[3] Yang, Xin et al. “A Two-Stage Convolutional Neural Network for Pulmonary Embolism Detection From CTPA Images.” IEEE Access 7 (2019): 84849-84857.

[4] Noa Cahan, Edith M. Marom, Shelly Soffer, Yiftach Barash, Eli Konen, Eyal Klang, Hayit Greenspan, ”Weakly supervised attention model for RV strain classification from volumetric CTPA scans”, Computer Methods and Programs in Biomedicine, Volume 220, 2022, 106815.

[5] Huang SC, Kothari T, Banerjee I, Chute C, Ball RL, Borus N, Huang A, Patel BN, Rajpurkar P, Irvin J, Dunnmon J, Bledsoe J, Shpanskaya K, Dhaliwal A, Zamanian R, Ng AY, Lungren MP. ”PENet-a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. NPJ Digit Med. 2020 Apr 24.

[6] Shih-Cheng Huang, Anuj Pareek, Saeed Seyyedi, Imon Banerjee, and Matthew P Lungren. ”Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines”. NPJ digital medicine, 3(1):1–9, 2020.

[7] Zhou, Y., Huang, S.C., Fries, J.A., Youssef, A., Amrhein, T.J., Chang, M., Baner- jee, I., Rubin, D., Xing, L., Shah, N., Lungren, M.P.: Radfusion: Benchmarking performance and fairness for multimodal pulmonary embolism detection from ct and ehr (2021)

[8] Miotto, Riccardo, et al. ”Deep patient: an unsupervised representation to predict the future of patients from the electronic health records.” Scientific reports 6.1 (2016): 26094.

[9] Chen, Mengyang, et al. ”Multimodal early risk prediction for major chronic diseases: A case study of type 2 diabetes.” IEEE Access 7 (2019): 183719-183732.

[10] Zhou, Tong, et al. ”A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease.” Scientific Reports 11.1 (2021): 1-17.

[11] Arik, Sercan Ö., and Tomas Pfister. ”Tabnet: Attentive interpretable tabular learning.” Proceedings of the AAAI conference on artificial intelligence. Vol. 35. No. 8. 2021.

[12] Chen, Tianqi, and Carlos Guestrin. ”Xgboost: A scalable tree boosting system.” Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016

[13] Prokhorenkova, Liudmila, et al. ”CatBoost: unbiased boosting with categorical features.” Advances in neural information processing systems 31 (2018).

[14] I. D. Mienye and Y. Sun, ”A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects,” in IEEE Access, vol. 10, pp. 99129-99149, 2022