TIM: Temporal Decoupling with Iterative Mutual-Refinement Model for Longitudinal Radiology Report Generation

Supplementary Material

A. More Hyperparameter Analysis

To further investigate the impact of hyperparameters besides the number of perception frames and the number of refinement iterations, we conducted a comparison study on four variables: the static pathological alignment weight λ_1 (cf. Eq. (8)), the dynamic progression alignment weight λ_2 (cf. Eq. (8)), the refinement scaling factor β (cf. Eq. (11)), and the refinement loss weight λ_Δ (cf. Eq. (12)). The experimental setup follows the dataset split in Sec. 4.1 and the implementation details in Sec. 4.2. Model performance is evaluated using ROUGE-L and F1-score to measure textual fidelity and clinical accuracy.

As illustrated in Fig. 7, we vary $\lambda_1 \in \{0, 0.1, 0.5, 1.0, 2.0, 5.0\}$ and $\lambda_2 \in \{0, 0.01, 0.1, 0.5, 1.0\}$ to study their respective roles in static pathology modeling and dynamic progression alignment. $\lambda_1 = 0.5$ achieves the best balance between language quality and pathology description accuracy, while $\lambda_2 = 0.1$ provides optimal temporal evolution modeling. Values of $\lambda_1 \in [0.1, 1.0]$ and $\lambda_2 \in [0.01, 0.5]$ yield relatively strong results, indicating the robustness of our method to moderate variations. Performance deteriorates for extreme settings, where overly small weights weaken the corresponding module contributions, while overly large weights introduce excessive alignment constraints that hinder generation quality.

For the Stage II refinement hyperparameters , we investigate $\beta \in \{1, 2, 5, 10, 15\}$ and $\lambda_\Delta \in \{0.01, 0.1, 0.5, 1.0\}$ in Fig. 8. The scaling factor β regulates the sensitivity of the sigmoid function to similarity differences, effectively controlling the sharpness of the refinement loss. The model achieves its best performance at $\beta = 5$, while values within the range $\beta \in [2, 10]$ also perform well. Extremely large values destabilize optimization, while smaller values diminish the corrective signal. The refinement loss weight λ_Δ adjusts the trade-off between the refinement objective and the autoregressive generation loss. We observe that $\lambda_\Delta = 0.1$ provides the optimal balance, with values in $\lambda_\Delta \in [0.01, 0.5]$ maintaining competitive performance.

Overall, the stable performance across broad hyperparameter ranges suggests that our model is not overly sensitive to parameter tuning, making the configuration process straightforward in practice.

B. Evaluation of Spatiotemporal Consistency

Beyond report-level metrics, we quantitatively evaluate spatiotemporal consistency by extracting medical condition labels from predicted and ground-truth prior/current reports with CheXbert and deriving progression states (stable, worsening, improving) based on label transitions. As shown in Table. 4, we observe that incorporating the dynamic progression modeling branch substantially improves progression F1 scores from 0.78 to 0.81, with further gains

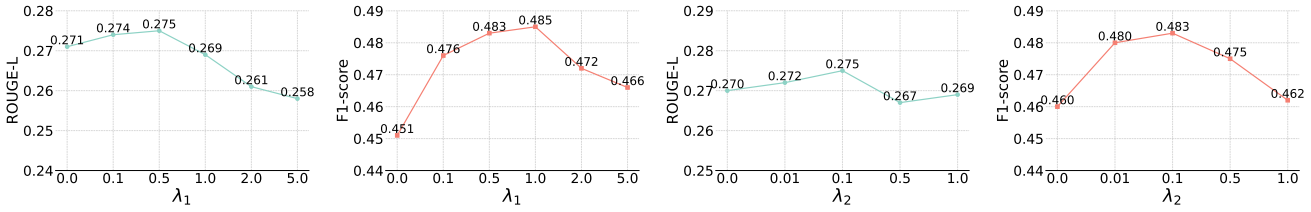


Figure 7. **Stage I hyperparameter analysis**, including λ_1 and λ_2 in Eq. (8). (a) ROUGE-L of different λ_1 values. (b) F1-score of different λ_1 values. (c) ROUGE-L of different λ_2 values. (d) F1 of different λ_2 values.

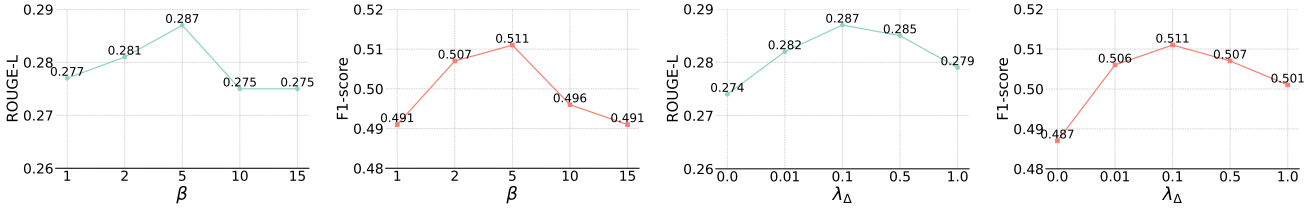


Figure 8. **Stage II hyperparameter analysis**, including β in Eq. (11) and λ_Δ in Eq. (12). (a) ROUGE-L of different β values. (b) F1-score of different β values. (c) ROUGE-L of different λ_Δ values. (d) F1 of different λ_Δ values.

when dynamic-aware textual supervision and stage II mutual refinement are applied, whose F1-score reaches 0.84, demonstrating more accurate temporal reasoning.

Table 4. **Evaluation of Spatiotemporal Consistency,** conducting an ablation study on the components related to dynamic modeling.

Model	F1-score			
	Stable	Worsening	Improving	Avg.
TIM w/o Progression Branch	0.85	0.30	0.52	0.78
TIM w/o Progression Loss	0.89	0.31	0.58	0.81
TIM-stage I	0.89	0.33	0.61	0.82
TIM-stage II	**0.91**	**0.35**	**0.65**	**0.84**

C. Example of Prompts for LLM Input

In both Stage I and Stage II, the LLM receives a textual prompt that guides the report generation process. Each prompt specifies the model's role as a radiologist, outlines the task and the expected input-output behavior, and indicates the temporal context of the report (i.e., prior or current examination). Below, we provide examples of the prompts used in two stages. The symbol "[·]" denotes the corresponding embedding inputs:

Stage I Input Prompts

<s> User: Progression: [progression]. Prior report: [Context]. Image: [Image]. You are a radiologist. Please generate a radiology report for the provided current chest X-ray image, using the visual findings from the image, the text of the previous report, and the progression between current and prior images. Assistant:

Stage II Input Prompts

<s> User: You are a radiologist. You have already written reports for the current and previous chest X-ray images of a patient, but there might be some issues in those reports. Current image: [Image]. Your report: [Report]. The following pathology was not consistent with the actual report in the generation of your previous report: [Diff]. Please pay special attention to these errors and regenerate an accurate report. Assistant:

D. Examples of Clinical Descriptions of Medical Conditions

As demonstrated in Section 3.2.1, to couple pathological features with clinical semantics, we expand each medical condition into a clinical description using external medical knowledge sources such as UMLS and PubMed. Each description summarizes the characteristic radiographic manifestation of the disease on chest radiographs and follows a standardized template in the form of "{*condition*} on chest x-ray refers to ..." to ensure consistent linguistic structure across disease categories. This formulation preserves key clinical terminology (e.g., opacity, cardiomegaly, air bronchogram) while maintaining concise semantics suitable for vision–language alignment. Table 5 presents representative examples of these disease-aware textual descriptions.

E. Case Studies

To further illustrate the effectiveness of our proposed TIM, we present four case studies. Each case compares model outputs from both Stage I and II with the ground truth reports, shows the predicted observations, and reports the ROUGE-L score for each output. The 14 observation categories are {*enlarged cardiomediastinum, cardiomegaly, lung opacity, lung lesion, edema, consolidation, pneumonia, atelectasis, pneumothorax, pleural effusion, pleural other, fracture, support devices, no finding*}. The label codes 0, 1, 2, and 3 denote {*not mentioned, positive, negative, uncertain*}.

As shown in Figs. 9–11, these qualitative evaluations demonstrate how our method captures temporal disease progression and corrects cross-time inconsistencies. In the predicted reports, red and green fonts indicate statements that are consistent with and conflicting with the ground truth, respectively. Blue numbers in the predicted observations mark classification errors.

Case 1 is shown in Fig. 9. For the prior exam, both the ground truth and Stage I agree that the endotracheal tubes are removed, and no pneumothorax is present. Stage II preserves these correct findings, but also introduces a left chest tube that is not described in the prior ground truth. For the current exam, Stage I incorrectly omits the right IJ catheter described in the ground truth and reports removal of the chest tube, which contradicts the ground truth. Stage II corrects most of these errors for the current study: it restores the presence and stability of the left retrocardiac opacity, explicitly reports bilateral small pleural effusions and mild pulmonary vascular congestion, and states that heart size is mildly enlarged but stable, all consistent with the ground truth.

Case 2 is shown in Fig. 10. For the prior examination, Stage I correctly reports increased bilateral pulmonary opacities and cardiomegaly but omitted the newly developed small right pleural effusion described in the ground truth. Stage II recovers the pleural effusion and preserves the finding of unchanged cardiomegaly. For the current examination, Stage I captures worsening bilateral opacities and stable cardiomegaly, but again misses the small right pleural effusion and provides a slightly different endotra-

cheal tube position. After refinement, Stage II restores the small right pleural effusion and retrocardiac air-space opacity reported in the ground truth.

From these cases, we observe that Stage II refinement improves both report completeness and disease-label accuracy compared with Stage I outputs, indicating that the refinement process reduces repeated errors across timepoints and yields reports that align more closely with the ground truth progression.

We also present a representative failure case in which the model does not recover the subtle temporal evolution described in the ground truth reports. As shown in Fig. 11, the prior ground truth report documents a small right-basilar consolidation, and the current GT report describes a rapid interval decrease of this opacity. However, the visual appearance of this lesion is extremely subtle and often confounded by postoperative changes, vascular markings, and acquisition variability across the two time points. As a result, both Stage I and Stage II tend to produce more conservative normal findings, which is a common behavior observed when the radiographic evidence for a minor abnormality is ambiguous. This example illustrates a challenging scenario driven by imaging subtlety. Future work may incorporate uncertainty modeling to better handle such borderline cases.

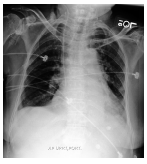
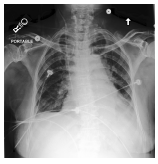
	Prior Examination	***Current Examination***
Ground Truth	 **Report:** comparison is made to the prior study from at 4 :16 am . there has been removal of the endotracheal tube. there is a right-sided ij catheter with distal lead tip at the cavoatrial junction . there is again seen some volume loss on the left side. there are no pneumothoraces. there is likely a left-sided pleural effusion as well as atelectasis. this is stable from the prior study. **Observations:** [0, 0, 0, 0, 0, 0, 0, 1, 2, 1, 0, 0, 1, 0]	 **Report:** comparison is made to the prior study from . there is a right ij catheter with distal lead tip in the proximal right atrium. heart size is mildly enlarged but unchanged. there is a left retrocardiac opacity and bilateral pleural effusions which are small. there is mild pulmonary fluid overload. overall these findings are stable. there is no visible pneumothorax. **Observations:** [0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0]
Stage I	**Report:** the endotracheal tube and nasogastric tube have been removed. there is no evidence of pneumothorax. a right internal jugular central venous catheter terminates in the lower svc . there is mild bibasilar atelectasis with small left pleural effusion. there is no focal consolidation or pulmonary edema. the mediastinal and hilar contours are normal. the heart size is enlarged but unchanged. median sternotomy wires are intact. **Predicted Observations:** [2, 1, 0, 0, 2, 2, 0, 1, 2, 1, 0, 0, 1, 0] **ROUGE-L:** 0.271	**Report:** there has been interval removal of the endotracheal tube right ij catheter and left-sided chest tube. there is no pneumothorax. there is a small left pleural effusion with adjacent atelectasis. there is mild pulmonary vascular congestion but no overt pulmonary edema. cardiomediastinal silhouette is unchanged. **Predicted Observations:** [3, 0, 0, 0, 2, 0, 0, 1, 2, 1, 0, 0, 2, 0] **ROUGE-L:** 0.322
Stage II	**Report:** there is a right ij central venous catheter with distal lead tip in the mid svc. there are no pneumothoraces. there has been interval removal of the endotracheal tube and nasogastric tube. there is a left-sided chest tube which is unchanged in position. there is some volume loss in the left lower lobe likely due to a combination of atelectasis and pleural effusion. this is stable compared to the prior study. there are no changes in the appearance of the right lung. the heart size is enlarged but stable. **Predicted Observations:** [0, 1, 0, 0, 0, 0, 0, 1, 2, 1, 0, 0, 1, 0] **ROUGE-L:** 0.361	**Report:** the heart size is mildly enlarged and stable. there is a left retrocardiac opacity which is unchanged since the prior study. there are bilateral pleural effusions which are also unchanged since the prior study. there is some mild pulmonary vascular congestion which is similar to the prior study. all other findings are stable. these findings are consistent with the clinical diagnosis of cardiomyopathy. **Predicted Observations:** [0, 1, 1, 0, 3, 0, 0, 0, 0, 1, 0, 0, 1, 0] **ROUGE-L:** 0.542

Figure 9. **Specific Case Study 1.** Red and green fonts indicate statements that are consistent with and conflicting with the ground truth, respectively. Blue numbers in the predicted observations mark classification errors.

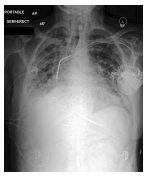
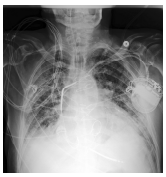
	Prior Examination	*Current Examination*
Ground Truth	 **Report:** as compared to the previous radiograph there is a severe increase in extent of the bilateral parenchymal opacities. these are strongly suggestive for severely increasing pulmonary edema. in addition a small right pleural effusion has newly occurred. there is unchanged evidence of cardiomegaly. no pneumonia retrocardiac atelectasis is present. at the time of dictation 8:27 am referring physician. was paged for notification. **Observations:** [0, 1, 1, 0, 1, 0, 2, 1, 0, 1, 0, 0, 0, 0]	 **Report:** endotracheal tube ends 3 cm from the carina. nasogastric tube ends in the stomach . left aicd device leads terminate in the appropriate positions. after ett placement there are increased lung volumes and still severe pulmonary edema. cardiomegaly is stable. small right pleural effusion is stable. retrocardiac atelectasis is unchanged. no evidence of pneumonia. sternotomy wires are seen. **Observations:** [0, 1, 0, 0, 1, 0, 2, 1, 0, 1, 0, 0, 1, 0]
Stage I	**Report:** there is a left-sided pacemaker with leads terminating in the right atrium right ventricle and coronary sinus. median sternotomy wires are intact. diffuse bilateral pulmonary opacities have increased since the prior examination. there is no pneumothorax or large pleural effusion. the cardiac silhouette is enlarged but unchanged. **Predicted Observations:** [0, 1, 1, 0, 0, 0, 0, 0, 2, 2, 0, 0, 1, 0] **ROUGE-L:** 0.232	**Report:** the endotracheal tube terminates approximately 4 cm above the carina. there is no pneumothorax. diffuse bilateral pulmonary opacities have increased in severity since the prior examination likely reflecting worsening pulmonary edema. cardiomegaly is unchanged. there is no pleural effusion or pneumothorax. sternotomy wires are intact. **Predicted Observations:** [0, 1, 0, 0, 1, 0, 0, 0, 2, 2, 0, 0, 1, 0] **ROUGE-L:** 0.251
Stage II	**Report:** as compared to the previous radiograph there is a slight increase in severity of pre-existing signs of pulmonary edema. there is also a small right pleural effusion. cardiomegaly persists unchanged. left chest wall pacemaker is in constant position. on the current radiograph retrocardiac parenchymal opacities have slightly increased in severity. no other parenchymal opacities are seen . **Predicted Observations:** [0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0] **ROUGE-L:** 0.286	**Report:** endotracheal tube is in appropriate position. left chest wall pacemaker is unchanged. sternotomy wires are intact. there has been interval increase in severity of pulmonary edema. this is likely due to increased fluid overload. cardiomegaly is stable. small right pleural effusion is unchanged. retrocardiac airspace opacity is stable. heart size is stable enlarged. **Predicted Observations:** [0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0] **ROUGE-L:** 0.288

Figure 10. **Specific Case Study 2.** Red and green fonts indicate statements that are consistent with and conflicting with the ground truth, respectively. Blue numbers in the predicted observations mark classification errors.

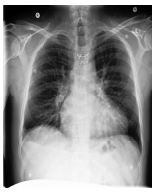
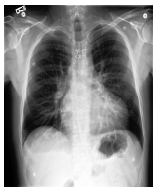
	Prior Examination	*Current Examination*
Ground Truth	 **Report:** two images of the chest shows a small consolidation at the right base most consistent with pneumonia. there are no other consolidations. there is no evidence of interstitial edema. there are no pleural effusions. the heart size is at the upper limits of normal . the mediastinal contours are normal. there are sternotomy wires in place. **Observations:** [2, 2, 0, 0, 2, 1, 1, 0, 0, 2, 0, 0, 0, 0]	 **Report:** the previously seen right lower lobe opacification has decreased substantially. there has also been a mild decrease in the amount of vascular engorgement suggesting improvement in mild biventricular heart failure. in retrospect given the rapid change the opacification likely represented fluid overload. the heart size is at the upper limits of normal. the sternal wires are intact and midline. there is longstanding midline lucency in the manubrium and upper body is due to incomplete sternal fusion there is no evidence of other incision complications. **Observations :** [0, 3, 1, 0, 1, 0, 0, 0, 0, 2, 0, 0, 1, 0]
Stage I	**Report:** there is mild pulmonary edema. there is no focal consolidation pleural effusion or pneumothorax. the cardiomediastinal silhouette and hilar contours are unchanged. sternal wires are intact. a left-sided picc terminates in the mid svc. median sternotomy wires are intact. there is no pneumothorax. **Predicted Observations:** [3, 0, 0, 0, 1, 2, 0, 0, 2, 2, 0, 0, 1, 0] **ROUGE-L:** 0.280	**Report:** there is no focal consolidation pleural effusion or pneumothorax. moderate cardiomegaly is unchanged. median sternotomy wires are intact. the mediastinal and hilar contours are normal. the lungs are clear. there is no evidence of pulmonary edema. there is no pneumothorax. bony structures are intact. **Predicted Observations:** [2, 1, 0, 0, 2, 2, 0, 0, 2, 2, 0, 0, 0, 0] **ROUGE-L:** 0.212
Stage II	**Report:** the patient is status post sternotomy. a left-sided picc line terminates in the mid svc. there is no pneumothorax. there is mild pulmonary vascular congestion and mild interstitial edema. there is no focal consolidation or pleural effusion. the cardiomediastinal silhouette is unchanged from the prior examination. **Predicted Observations:** [3, 0, 0, 0, 1, 2, 0, 0, 2, 2, 0, 0, 1, 0] **ROUGE-L:** 0.315	**Report:** the patient is status post median sternotomy and cabg. the lungs are clear. there is no pleural effusion or pneumothorax. the heart size is at the upper limits of normal and is unchanged from the previous examination. mild pulmonary vascular congestion is seen. **Predicted Observations:** [0, 3, 0, 0, 3, 0, 0, 0, 2, 2, 0, 0, 0, 0] **ROUGE-L:** 0.207

Figure 11. **Specific Case Study 3.** Red and green fonts indicate statements that are consistent with and conflicting with the ground truth, respectively. Blue numbers in the predicted observations mark classification errors.

Table 5. Examples of clinical descriptions used for disease-aware text construction.

Medical Condition	Source Literature	Annotation Prompt
Enlarged Cardiomediastinum	Klein et al. [39]	Enlarged cardiomediastinum on chest x-ray refers to the widening or abnormal enlargement of the combined cardiac and mediastinal contours on chest radiography.
Cardiomegaly	Siwik et al. [40] Cardinale et al. [6]	Cardiomegaly on chest x-ray refers to an abnormally enlarged cardiac silhouette, typically identified when the cardiac-thoracic ratio (CTR) exceeds 0.5, meaning the heart width occupies more than half of the thoracic cage width.
Lung Opacity	Turk et al. [43]	Lung opacity on chest x-ray refers to an area in the normally dark lung that appears whiter, hazy, or denser due to increased tissue density from the replacement of air by substances such as fluid, fibrosis, or other material.
Lung Lesion	Feyisa et al. [14]	Lung lesion on chest x-ray refers to pulmonary lesions presenting as localized nodules, masses, or cavitary abnormalities within the lung parenchyma, often associated with structural distortion of surrounding lung tissue.
Edema	Siwik et al. [40]	Pulmonary edema on chest x-ray refers to interstitial or alveolar fluid accumulation, often producing distortion in the hilar image, thickening of interlobar or interlobular fissures, and bilateral higher density opacities in central and basal lung areas.
Consolidation	Bankier et al. [3] Cardinale et al. [6]	Consolidation on chest x-ray refers to a homogeneous increase in attenuation caused by alveolar filling with fluid, pus, blood, or cells, resulting in homogeneous opacification of the lung parenchyma on radiographs, often accompanied by air bronchograms.
Pneumonia	Bankier et al. [3]	Pneumonia on chest x-ray refers to infection-related lung abnormalities that manifest as opacity, potentially accompanied by air bronchograms (air bronchogram), tree-in-bud opacities, or pleural effusions (pleural effusion).
Atelectasis	Klein et al. [39]	Atelectasis on chest x-ray refers to partial or complete collapse of lung tissue leading to volume loss, increased density on radiographs, and displacement of adjacent structures such as fissures or mediastinum.
Pneumothorax	Bankier et al. [3]	Pneumothorax on chest x-ray refers to the presence of pleural air typically identified on upright chest radiographs by a curvilinear visceral pleural line indicating the demarcation between pleural air and aerated lung.
Pleural Effusion	Bankier et al. [3]	Pleural effusion on chest x-ray refers to accumulation of fluid in the pleural space where small pleural effusions manifest as blunting of the costophrenic angles, often meniscus-shaped, and larger effusions as homogeneous opacities in dependent regions.
Pleural Other	Bankier et al. [3] Eisenhuber et al. [12]	Pleural other on chest x-ray refers to other pleural abnormalities presenting as pleural thickening, plaques, or irregular pleural contours, sometimes associated with calcifications or localized pleural masses.
Fracture	Dilday et al. [10]	Fracture on chest x-ray refers to rib fractures appearing as cortical discontinuities or step deformities along the rib contour and may be associated with complications such as pneumothorax or hemothorax.
Support Devices	Hunter et al. [16] Baratella et al. [4]	Support devices on chest x-ray refers to medical support devices such as endotracheal tubes, central venous catheters, pacemaker leads, and thoracostomy tubes appearing as radiopaque linear or tubular structures used to confirm correct positioning and detect potential complications.