Author's Thanks:

We sincerely appreciate the reviewer's dedicated time and expertise in critically evaluating our work. The constructive feedback has prompted essential refinements to both the scholarly substance and structural clarity of this manuscript, significantly elevating its academic contribution. Below we provide a systematic point-by-point response to each comment. The italicized content represents the modifications made in the manuscript.

RC2 Comment 1:

Abstract: Lines 13–14 provide a good overview of the methodology, including the data and model used. However, the workflow of ARE is unclear to readers unfamiliar with it. Adding one or two sentences explaining its mechanism would be beneficial. The explanation in lines 17–18 is too general. Readers may struggle to grasp the concept just from the abstract. Line 22 presents rice expansion values but lacks a specified duration. I recommend revising this sentence and the following one for better readability and clarity.

Author's Response 1:

Thank you for your valuable comment. To improve readability, we have supplemented an explanation of the ARE method. The corresponding content is as follows: "*ARE integrates differences in category probability and confidence levels of the FR-Net across phenological stages, effectively reducing classification uncertainty. This approach could mitigate the impact of limited training sample on large-scale and across-sensors paddy rice mapping.*"

In addition, we have specified the time range of expansion in paddy rice cultivation area in the manuscript, as follows: "*The study revealed that the area used for paddy rice cultivation in Northeast China increased from* 1.11×10^4 km^2 *to* 6.45×10^4 km^2 *between 1985 and 2023.*"

RC2 Comment 2:

Introduction: This section is well-structured and easy to follow. In line 63, the phrase "multiple annual results" is unclear. Using plain language would improve readability.

Author's Response 2:

According to your suggestion, we have revised "However, determining the final mapping result from multiple annual results remains a challenge for large-scale paddy rice mapping." to *"However, determining the final mapping result for a specific year from multiple intermediate maps remains a challenge for large-scale paddy rice mapping.*" in the manuscript. Thanks for your valuable comment.

RC2 Comment 3:

Methods: This section requires substantial improvement for better clarity. Adding a workflow diagram would significantly enhance readability and coherence. The sections are loosely connected. The authors should clearly specify: Which datasets were used to feed the model? Which results were used for ARE? Which datasets were used for validation and how they relate to the modeling outputs or ARE results?

Author's Response 3:

Thank you for your significant comment. We have integrated a workflow diagram into the manuscript and provided a concise description of the process. Additionally, we have revised the text to clarify the role of each dataset in the respective sections, with enhanced explanations of their relationships with the modeling outputs and ARE results. The supplementary details are summarized below:

2.2.1 Workflow of the study

The workflow for paddy rice mapping in Northeast China is illustrated in Figure 2. Firstly, Landsat data, paddy and non-paddy samples derived from Google Earth data and field survey, and the paddy cultivation area from agricultural statistics data between 1985 and 2023 were compiled to validate the accuracy of the annual paddy rice mapping results in this study. Secondly, a cross-sensor dataset containing 115 scenes paddy rice maps was generated using the XGBoost classifier combined with manual visual correction. This cross-sensor dataset served as training and testing datasets for developing the FR-Net model. Thirdly, based on the cross-sensor dataset, we employed the FR-Net and ARE methods to account for category probability differences across different phenological periods within a year to reconstruct annual paddy rice maps for Northeast China from 1985 to 2023, and the paddy rice maps were systematically validated with validation data. Finally, the paddy rice mapping results in this study were compared with representative products, and we analyzed the spatial and temporal dynamic characteristics of paddy rice in Northeast China.

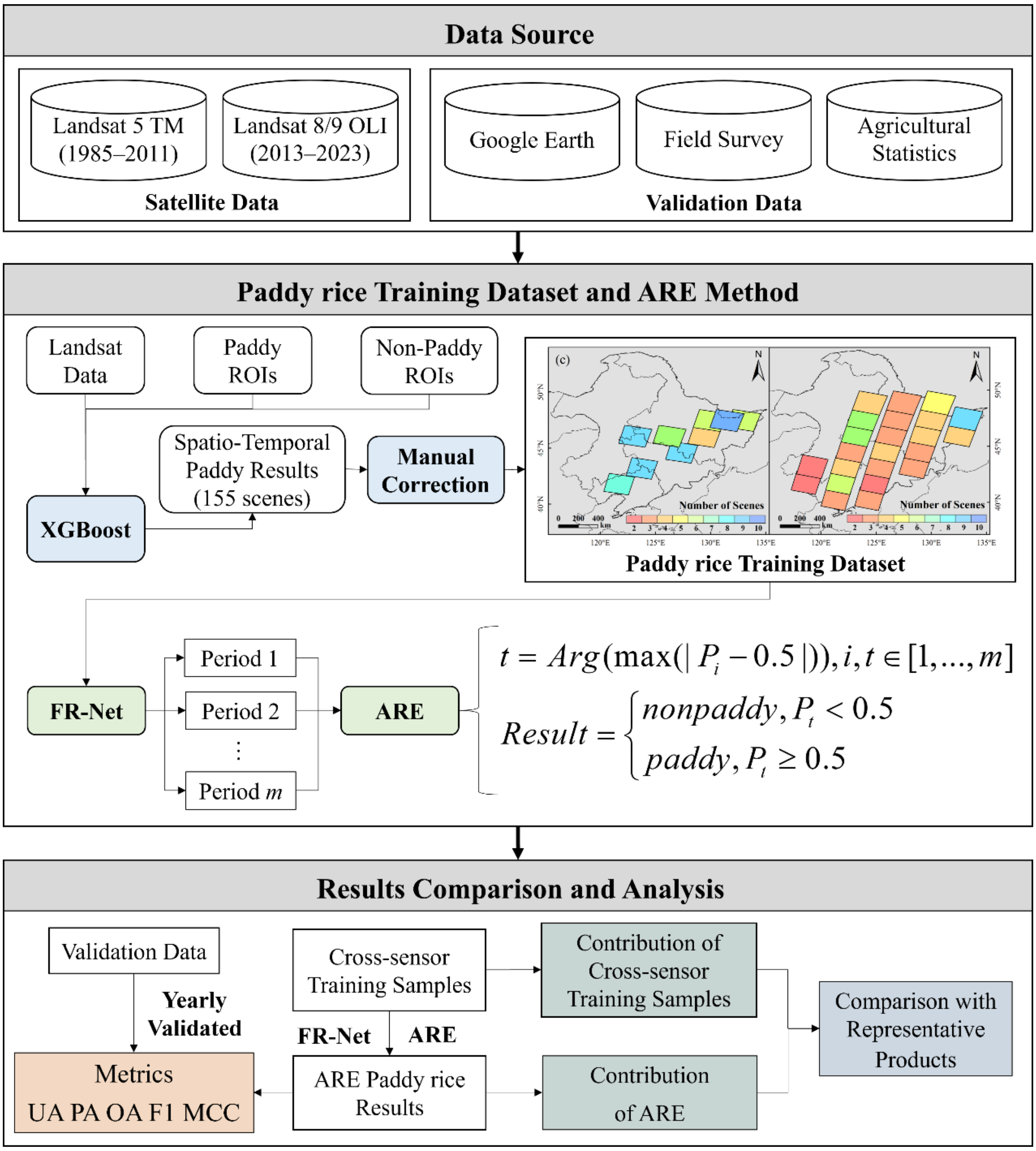

Figure 2: Workflow of the study.

Which datasets were used to feed the model and how they relate to the modeling outputs or ARE results?

Response: The 115 scenes cross-sensor paddy rice maps, generated by the XGBoost classifier and refined through manual visual correction, were fed into the FR-Net model as input data.

Which results were used for ARE and how they relate to the modeling outputs or ARE results?

Response: The ARE framework enhances all intermediate maps generated by the FR-Net model within a year.

Which datasets were used for validation and how they relate to the modeling outputs or ARE results?

Response: The paddy and non-paddy samples derived from Google Earth and field surveys were utilized to validate the accuracy of ARE results, while agricultural statistical data were employed to assess the accuracy of the ARE results at the district, municipal, provincial, and entire study area levels. Besides, we have supplemented verification of paddy rice cultivation areas at the district, municipal, and provincial levels in the manuscript.

RC2 Comment 4:

2.2.1: As this is the most critical section for mapping, more details on FR-Net are necessary. Although Xia et al. (2022) describes the model, a concise explanation of its working principles, strengths, and weaknesses is still needed. This will provide readers with a foundational understanding, allowing them to refer to the cited work for further details.

Author's Response 4:

Thank you for your important comment. We have added a concise explanation of FR-Net, including working principles, strengths, and weaknesses. The revised content is as follows: *The multi-resolution feature fusion unit (MRFU) serves as the core component of FR-Net, specifically designed to achieve high-resolution semantic segmentation while maintaining precise output quality. The MRFU regulates feature propagation through controlled information flow, integrates multi-scale feature representations via resolution-specific streams, and preserves spatial fidelity through hierarchical resolution retention. Its architecture comprises two distinct pathways: the horizontal stream, which preserves native resolution through identity mapping operations, and the vertical stream, which doubles channel capacity while halving spatial resolution. Additionally, the framework incorporates 3×3 convolutional layers with a stride of 2, a batch normalization (BN) layer, and a rectified linear unit (ReLU) activation layer. These components work together to control and fuse feature streams with different resolutions. FR-Net has a simple structure and requires minimal computational resources, making it suitable for extracting characteristic information from Landsat data and mitigating the issue of gradient disappearance. However, despite its straightforward design, the cascading operation of multi-resolution feature fusion may result in computational delays, which could hinder its ability to meet the near real-time requirements for agricultural monitoring.*

RC2 Comment 5:

2.2.2: In line 113, it would be helpful to first explain why multiple mapping results exist and how they are produced. The ARE method is not clearly explained, which raises concerns. Equation 2 suggests that a good map ($Pt > 0.5$) wins only when it has a greater distance from 0.5. However, this approach may not be fair in all situations. For instance, probability values at the start and end of the growing season may be less reliable than those in mid-season, potentially leading to misclassification of rice pixels as non-rice. More details on this method and additional case studies under different conditions would be beneficial.

Author's Response 5:

Thank you for your valuable comment. We have explained the reasons for the existence of multiple mapping results and how they were generated, the content is as follows: "*During the growth period of paddy rice, multiple Landsat images of the same area can be obtained, allowing for the identification of each image and resulting in multiple mapping results.*".

In addition, we have supplemented the explanation of ARE method, the content is as follows: "*Therefore, based on the distinct differences in spectral and texture characteristics of paddy rice across growth stages, we developed an annual result enhancement (ARE) method to address this limitation. ARE integrates differences in category probability and confidence levels of the FR-Net across phenological stages, effectively reducing classification uncertainty. This approach mitigates the impact of limited training sample on large-scale and across-sensors paddy rice mapping.*"

We agree with you on the probability values at the start and end of the growing season may be less reliable than those in mid-season. To minimize the impact of the start and end of the rice growing season on mapping accuracy, we selected Landsat images from May to September for paddy rice mapping.

RC2 Comment 6:

2.3.1: How is the growing season defined? How is the model trained by using these bands? The band meaning and numbers vary across years and satellite products, how are they handled? I suggest merging 2.3.3 to this section.

Author's Response 6:

Thank you for your valuable suggestion. We agreed to consolidated Section 2.3.3 into Section 2.3.1 to enhance structural coherence, as recommended in the review feedback, and we modified the title of 2.3.1 to "Acquisition and processing of Landsat images".

How is the growing season defined?

Response: The growing season for paddy rice was defined based on region-specific phenological patterns and time-series spectral signatures derived from Landsat images. Specifically, this season spans from pre-transplanting flooding to post-harvest senescence, and in the study its May to September.

How is the model trained by using these bands?

Response: We selected the Blue, Green, Red, Near Infrared (NIR), Shortwave Infrared (SWIR) 1, and Shortwave Infrared (SWIR) 2 bands of Landsat 8/9 OLI and Landsat 5 TM images to train the FR-Net model.

The band meaning and numbers vary across years and satellite products, how are they handled?

Response: We rotated and patched these bands data as inputs for XGBoost and FR-Net models, and used them for subsequent analysis.

RC2 Comment 7:

2.3.2: This dataset is crucial for validating the study's product and holds significant value for broader research communities in the study area. It is necessary to publish the relevant dataset for validation checking and a broader use.

Author's Response 7:

Thank you for your important comment. This study leverages a multi-temporal ground truth dataset to validate the accuracy of our paddy rice mapping product. We are aware that the validation dataset is invaluable for assessing the study's product and holds significant importance for broader research communities in Northeast China. The validation dataset used in this study was contributed by multiple collaborating institutions (including the authors' affiliations), and we have not obtained explicit authorization from other participating entities to publicly release the complete dataset. Researchers who require access for academic purposes may contact the corresponding author to request a subset of the validation data.

RC2 Comment 8:

2.3.3: This section is very confusing and needs to be recontructed. First, what is the connect of XGBoost to the DL model? Given it can generate the paddy and non-paddy maps, what are the differences between its results and the DL model? Second, The ROIs in Fig 1(c)&(d) are very large. From my understand, they indicate paddy and non-paddy. Does it mean that within the ROI, all pixels are either paddy or non-paddy? Third, how was the manual correction conducted? Forth, What does the mask mean in line 167?

Author's Response 8:

Thank you for your valuable comment.

First, what is the connect of XGBoost to the DL model? Given it can generate the paddy and non-paddy maps, what are the differences between its results and the DL model?

Response: The XGBoost classifier generates preliminary rice classification results based on visually interpreted Regions of Interest (ROIs). This initial output undergoes manual rectification to address commission errors (e.g., non-paddy rice areas misclassified as paddy rice) and omission errors (e.g., undetected paddy rice cultivation areas). The DL (FR-Net) model utilizes expert-refined paddy rice results derived from manual correction of XGBoost-generated preliminary maps. Specifically, the XGBoost results after manual corrected serve as training inputs for the DL model, while the DL outputs are further processed through ARE method to generate the final paddy rice product presented in this study.

Second, The ROIs in Fig 1(c)&(d) are very large. From my understand, they indicate paddy and non-paddy. Does it mean that within the ROI, all pixels are either paddy or non-paddy?

Response: Fig.1(c) and Fig.1(d) present the paddy rice mapping results for 115 scenes across different geographical regions and times based on XGBoost classification with subsequent manual visual correction. The regions displayed in these figures correspond to the spatial positions of input data within the DL model, which contain both paddy rice pixels and non-paddy rice pixels.

Third, how was the manual correction conducted?

Response: manual correction is the process of manually modifying the misclassification and omission errors of paddy rice to make them the correct category.

Forth, What does the mask mean in line 167?

Response: the term 'mask' in the manuscript refers to the processed paddy rice classification results derived from 115 original Landsat images through XGBoost classifier and manual visual correction. The raw satellite imagery and corresponding paddy rice masks underwent standardized preprocessing operations including geometric rotation and patching, ultimately generating paired 256×256 pixel patches. These aligned image-mask pairs maintain spatial correspondence between the preprocessed Landsat images (images) and their associated paddy rice mapping results (masks).

Furthermore, we have merged 2.3.3 to 2.3.1, and modified the title of 2.3.1 to "Acquisition and processing of Landsat images".

RC2 Comment 9:

2.5: There are no clear criteria for model constraints, such as loss functions. This should be explicitly mentioned.

Author's Response 9:

Thank you for your valuable comment. We have added the loss function in section 2.5 of this manuscript. The content is as follows: "*To mitigate class imbalance during model training, the Dice loss function was employed in this study. This metric, derived from the Dice similarity coefficient (DSC; Eq.3), demonstrates inherent robustness against skewed class distributions by equivalently weighting false positive and false negative errors during optimization, thereby addressing prevalent challenges in imbalanced semantic segmentation tasks.*

$$DSC = \frac{2\times|A\cap B|}{|A|+|B|}, \quad (3)$$

where $|A \cap B|$ *quantifies the intersection cardinality between the predicted paddy rice pixels (A) and ground truth paddy rice pixels (B);* $|A|$ *and* $|B|$ *represent the quantity of paddy rice pixels in A and B, respectively.*"

RC2 Comment 10:

Technique comments:

The authors need to update the caption for figures. What is the scale of the dots in Fig 4, district, county, or province? In (a), the dots are for one year or multiple years? In Fig 7, what is the difference between paddy and interpolated paddy? In Fig 8, what does the trend mean? Are the values on map the change rate?

Author's Response 10:

Thank you for your valuable comment.

What is the scale of the dots in Fig 4, district, county, or province?

Response: In Fig.4, we validated the total area of paddy rice from 1985 to 2023 using agricultural statistical data across the entire study area. In addition, to further confirm the accuracy of the paddy rice maps presented in this study, we utilized all publicly available agricultural statistical data from the study area to validate the paddy rice mapping results at the provincial, municipal, and district levels in the manuscript.

In (a), the dots are for one year or multiple years?

Response: The dots in Fig.4(a) represent data for individual years.

In Fig 7, what is the difference between paddy and interpolated paddy?

Response: In Fig.7, the term 'paddy' denotes the paddy rice result directly classified using clear-sky observations, while the term 'interpolated paddy' refers to the paddy rice result derived using a multi-year comprehensive method, based on the historical phenological patterns from the nearest available clear-sky year's image. To avoid ambiguity, we have replaced 'interpolated paddy' with 'gap-filled paddy' throughout the manuscript.

In Fig 8, what does the trend mean? Are the values on map the change rate?

Response: In Fig.8, the trend means the whether there have been changes in the paddy rice cultivation areas in 1985 and 2023, and the values on maps refer to the areas where paddy rice cultivation increased, decreased, and remained unchanged in 2023 compared 1985.