

# **A REGRESSION-BASED EXPLAINABLE CONVOLUTIONAL NEURAL NETWORK FOR YIELD ESTIMATION OF SOYBEAN**

ARUN VISHWANATH VENUGOPAL

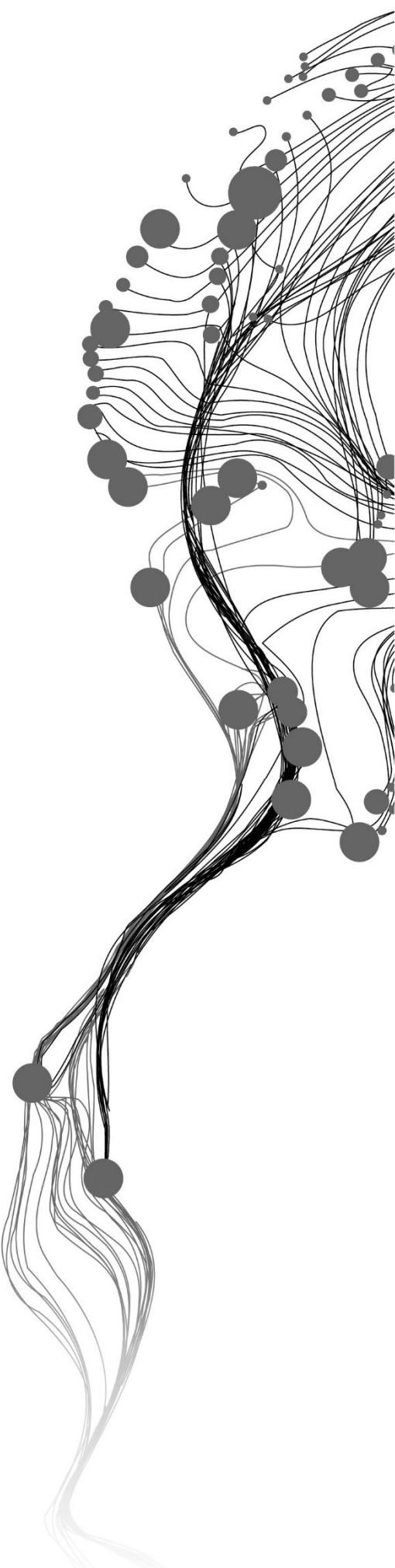
June, 2023

SUPERVISORS:

Dr. Mahdi Farnaghi

Prof. Dr. Raul Zurita Milla





# **A REGRESSION-BASED EXPLAINABLE CONVOLUTIONAL NEURAL NETWORK FOR YIELD ESTIMATION OF SOYBEAN**

**ARUN VISHWANATH VENUGOPAL**  
Enschede, The Netherlands, June, 2023

A thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.  
Specialisation: Geoinformatics

**SUPERVISORS:**

Dr. Mahdi Farnaghi  
Prof. Dr. Raul Zurita Milla

**THESIS ASSESSMENT BOARD:**

Prof. Dr. M. J. Kraak (Chair)  
Dr. C. Paris (External Examiner, Department of Natural Resources)

#### DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author and do not necessarily represent those of the Faculty.

## ABSTRACT

Crop yield estimation is essential for decision-making and ensuring food security. This MSc thesis explores the explainability of a regression-based Convolutional Neural Network (CNN) model based on Earth Observation data. This is a relatively underdeveloped area because most Earth observation research is focused on classification tasks. Understanding the model's behaviour and analysing its performance and result using saliency maps is also essential to figure out if there is a bias in the results. This research focuses on developing an explainable regression-based CNN model for crop yield estimation. The soybean yield is taken as the target to be estimated by the model. The input data is Sentinel-2 imagery, downloaded from Google Earth Engine from 2017 to 2021. A CNN model is trained and focuses on the explainability of the model and how the model behaves or interprets the data to estimate crop yield. The top soybean-producing states from 2017 to 2021 in the United States are taken as the study region. The target yield values at the county level are taken from the United States Department of Agriculture. The areas that are covered by soybean are identified by the Cropland Data Layer. This is added to the input as a mask layer. The dataset is prepared for training and linear regression and CNN models are trained. The results are compared by training one CNN model with the mask layer and the other CNN without the mask layer. The model with the mask layer has better accuracy of 98%, while the model without the mask layer has 72%. Different saliency maps, such as gradCAM, gradient, smoothGrad, Guided Back Propagation, Layerwise Relevance Propagation, Deep Taylor, Integrated Gradients, etc., are generated from the test dataset. These saliency maps are then evaluated by performing a perturbation analysis. The input image is gridded, and these grids are perturbed by providing Gaussian noise based on the order of importance for each grid. The difference in the results of the perturbed input and the true value is compared for all the explainable methods. The area under the curve of each perturbation plot is used to quantify the perturbation analysis for all the patches. The results are also critically analysed spatially with vegetation indices and land use maps to explain why the model focuses on specific regions. This demonstrates that the model focuses on regions with higher vegetation indices without the mask layer as the input. Our findings also show that the mask layer is significant when estimating the yield without bias.

## ACKNOWLEDGEMENTS

I am immensely grateful to my first supervisor, Dr. Mahdi Farnaghi, whose unwavering support, expertise, and guidance have been the cornerstone of my journey in completing this master's thesis. His profound knowledge, insightful feedback, and constructive criticism have been instrumental in shaping the direction, depth, and quality of this research. I am truly indebted to his mentorship, dedication, and patience throughout the entire process. His input was valuable, and he was patient with me when we ran into technical difficulties. When I ran into trouble training the model in the middle of the thesis, he offered substantial assistance. I will always be thankful for the time we spent working together to resolve the challenges.

I would also like to extend my heartfelt appreciation to Prof. Dr. Raul Zurita Milla. His expertise, critical insights, and attention to detail have been immensely valuable in refining the research methodology, analysing the data, and interpreting the results. His guidance, encouragement, and intellectual contributions have significantly enriched the overall research process. Always inspiring me to think outside the box, his insightful ideas about crop yield estimation and how to form a correlation with the results gave me a fresh perspective to consider when working on my thesis.

I would also like to convey my sincere thanks to the members of the thesis assessment board, prof. Dr. M.J. Kraak and Dr. rer. nat. F. Mocnik. Their advice and feedback on the visualisations and ethical assessments were vital to the success of this study. I would also like to extend my heartfelt appreciation to the esteemed faculty members of ITC, University of Twente, whose wisdom, expertise, and intellectual contributions have greatly enriched my understanding of the subject matter. Their thought-provoking discussions, valuable insights, and scholarly guidance have played a pivotal role in shaping the ideas and concepts presented in this thesis. I thank my internship supervisor Dr. Scarlet Stadler, whose creative suggestions at the beginning of my thesis paved the direction for the rest of the research period. I thank the CRIB department for its geospatial computing unit. It helped in training my model and evaluating the results. I would also like to express my appreciation to the developers of the Innvestigate module in Python, which enabled me to analyse the explainability of the results efficiently.

Lastly, I would like to express my deepest gratitude to my family and friends at ITC for their unwavering support, encouragement, and belief in my abilities. Their motivation gave me the strength and resilience to overcome challenges and persevere through the demanding phases of this academic endeavour.

## TABLE OF CONTENT

1.	Introduction.....	1
1.1.	Importance of Research.....	1
1.2.	Literature Review.....	2
1.3.	Research Gap.....	4
1.4.	Objectives and Research Questions.....	6
2.	Study Area and Datasets.....	7
3.	Methodology.....	11
3.1.	Dataset Preparation.....	12
3.2.	Model Architecture.....	17
3.3.	Model Training and Evaluation.....	19
3.4.	Explainable methods.....	20
3.5.	Perturbation Analysis.....	21
3.6.	Analysis of Explainable Methods.....	22
3.7.	System Configuration and Specifications.....	24
4.	Results.....	25
4.1.	Loss Curves.....	25
4.2.	Model Evaluation.....	25
4.3.	Difference Map.....	26
4.4.	Outliers in the scatter plot.....	27
4.5.	Feature importance of Linear Regression.....	28
4.6.	Saliency maps.....	29
4.7.	Perturbation Analysis.....	31
4.8.	Analysis of saliency maps.....	34
5.	Discussion.....	39
5.1.	Limitations.....	40
6.	Conclusions and Recommendations.....	41
6.1.	Answers to the research questions.....	41
6.2.	Recommendations.....	42
7.	References.....	43

## LIST OF FIGURES

Figure 1: Map of Study Region .....	7
Figure 2: Map of the Counties .....	8
Figure 3: Cropland calender for Soybean.....	9
Figure 4: Cropland Data Layer for Soybean (2021).....	10
Figure 5: Methodology workflow .....	11
Figure 6: Flowchart depicting the dataset preparation process .....	13
Figure 7: Soybean Yield (Bu/acre) at the county level for 2021.....	14
Figure 8: The sentinel bands and croplands of soybean (Band 13) of a patch.....	15
Figure 9: Example showing how the target yield is prepared for a patch.....	16
Figure 10: Target yield for patches .....	17
Figure 11: CNN model architecture.....	18
Figure 12: Perturbation Analysis workflow.....	22
Figure 13: Training and Validation Loss curves for CNN .....	25
Figure 14: Scatter plot of True vs predicted .....	26
Figure 15: Difference map of True vs Predicted for test dataset (Indiana).....	27
Figure 16: Mask layers of the outliers in True vs Predicted .....	28
Figure 17: Feature importance of Linear Regression Model.....	28
Figure 18: Sentinel bands mean correlation heatmap.....	29
Figure 19: Mask Layer (Left), gradCAM of CNN model with mask (middle) and gradCAM of CNN model without mask (Right) of a patch from Indiana.....	30
Figure 20: Map of saliency values for Indiana from 2017 to 2021. The top row is the saliency map of the model without the mask. The middle Row is the saliency map of the model with the mask layer. The bottom Row is the soybean yield values.....	30
Figure 21: Scatter plot of the saliency values vs Yield values.....	31
Figure 22: Perturbation Analysis based on the ranking of saliency.....	32
Figure 23: Perturbation plot – the difference in accuracy from the True value for each iteration .....	33
Figure 24: Area under the curve (AUC) plot for perturbation analysis.....	33
Figure 25: GradCAM and indices of a patch from CNN model without mask .....	35
Figure 26: Area coverage by crop type for the patch .....	35
Figure 27: Crop Land cover for the patch .....	36
Figure 28: Line plot comparing the two models’ saliency and indices. The top row shows the saliency map of the patch for the CNN model without a mask, and the bottom row is the same patch for the CNN model with a cover.....	37
Figure 29: Boxplot of saliency values w.r.t crop type for CNN model without mask.....	38
Figure 30: Boxplot of saliency values w.r.t crop type for CNN model with mask.....	38



## LIST OF TABLES

---

Table 1: Literature Review .....	5
Table 2: Datasets Used .....	8
Table 3: Sentinel-2 Bands (Source: Google Earth Engine).....	9
Table 4: Number of counties having yield values per year .....	10
Table 5: CNN Model architecture .....	18
Table 6: Hyperparameters of the CNN model.....	19
Table 7: System Specifications.....	24
Table 8: RMSE and R <sup>2</sup> of the models .....	26
Table 9: Area under the curve values for 20 patches.....	34

## LIST OF EQUATIONS

---

Equation 1: Target Yield calculation.....	15
Equation 2: NDVI Formula.....	22
Equation 3: EVI Formula.....	23
Equation 4: NDMI Formula.....	23
Equation 5: WDRVI Formula.....	23
Equation 6: SAVI Formula.....	23

# 1. INTRODUCTION

## 1.1. Importance of Research

Crop yield estimation is crucial for food security owing to the rising demand caused by the increasing population (Wolanin et al., 2020a). Due to uncertainty in the weather and the use of conventional agricultural techniques, farmers in emerging economies face challenges in increasing their harvests (Johnson et al., 2021). The annual crop yield estimate grossly depends on several weather characteristics and climate change-induced phenomena (Johnson et al., 2021). To maintain food security and enable effective decision-making, it is necessary to have an accurate estimation for the subsequent years (Srivastava et al., 2022). Several methods have been researched and implemented over the past years for estimating crop yield, varying from simulation models to statistical methods based on data availability (Srivastava et al., 2022).

Predicting crop yield via statistical methods considers the spatio-temporal heterogeneity within a local area (Yang et al., 2022). For instance, meteorological parameters, which vary dynamically over time and space due to different topographies, influence the yield (Yang et al., 2022). To address this spatio-temporal heterogeneity in the local region, Geographically Weighted Regression (GWR) models were utilised to identify the spatial correlation between features (Imran et al., 2015). However, these models cannot fit attributes that have complex non-linear relationships (Brunsdon et al., 2010). An essential assumption is that the irregularities in the GWR model are uncorrelated and share variance (Leung et al., 2000). Non-linear models like Deep Neural Networks resolve this issue.

Over the past decade, there has been an increase in research conducted by utilising machine learning models for crop yield prediction (Klompenburg et al., 2020). The most frequently used deep learning algorithm is Convolutional Neural Networks (CNN), followed by Long-Short Term Memory (LSTM) (Klompenburg et al., 2020). However, one major problem in using Neural Networks is their limitation in interpreting the model. They are composed of several hidden layers (Rudin, 2019a). These models are called black-box models as they do not provide an interpretation as to how they arrived at their prediction. Recently, there has been some research to interpret the black box models (Casas-Roma & Conesa, 2021). The study led to the development of Explainable Artificial Intelligence, also called XAI (Explainable AI).

Explainability is necessary to offer the user transparency and trust, particularly in intricate deep-learning algorithms (Molnar, 2022). XAI models provide qualitative and quantitative accuracy regarding the model's performance and architecture. There are two different concepts in XAI, "Interpretation" and "Explanation". The term "Interpretation" refers to the ability to connect an abstract concept to a field that humans can comprehend (Montavon et al., 2017). An "Explanation" refers to the set of distinct features that have played a role in generating a decision, such as classification or regression (Montavon et al., 2017).

Interpretable or Explainable models allow a human to understand a model's decision for arriving at the result (Molnar, 2022). Interpretable Deep learning models are required to address the incompleteness in the modelling process (Doshi-Velez & Kim, 2017). In contrast, explainability provides analytical insight into the decision that was taken by the model, helping to identify incorrect or biased outcomes, etc. (Miller, 2019).

In the case of crop yield, it is crucial to determine if the model is effectively estimating the yield by utilising the input features accurately (Wolanin et al., 2020a). CNN models generate feature maps as intermediate layers, extracting relevant attributes from the input image to get the output layer. A critical aspect of explainable CNN models is providing insight into how these attributes are extracted and correcting any bias in the feature extraction process. This also ensures that the model performs correctly for the provided input dataset and can be trusted to perform effectively for future datasets. This research mainly focuses on critically analysing the explainability of a CNN model to enhance the performance and quality of the spatial estimation of soybean yield.

## 1.2. Literature Review

A comprehensive literature review by Klompenburg et al. (2020) found that most of the research conducted in crop yield estimation has utilised CNN and LSTM models. Linear Regression models are used as a benchmark to compare the performance of the deep learning models. Root Mean Square Error (RMSE) and  $R^2$  score are the commonly used validation parameters in several studies (Klompenburg et al., 2020). To identify the research gap and explore a new direction in the domain of crop yield estimation, we initially identified previous research that utilised deep learning algorithms for crop yield estimation. Subsequently, we look at scholarly articles where CNN models were implemented for crop yield estimation. This will further be narrowed down to focus on research that specifically uses remote sensing images as input data where explainable and interpretable methods are implemented for deep learning models predicting crop yield. We also focus on research that implements explainable methods for Earth observation data. Nevertheless, there is a scarcity of research utilising the explainability of the CNN models in crop yield estimation.

A few studies that use deep learning algorithms for crop yield estimation are described.

- In a work by Schwalbert et al. (2020), an LSTM model was implemented to forecast soybean yield in southern Brazil (Schwalbert et al., 2020). Data used include vegetation indices like NDVI (Normalised Difference Vegetation Index) and EVI (Enhanced Vegetation Index) derived from MODIS (Moderate-resolution Imaging Spectroradiometer), and weather parameters (temperature, precipitation). The LSTM model was compared with Random Forest and Linear Regression. The study concludes that the results from LSTM had a higher accuracy than the other models.
- Another study using a deep learning transfer model was implemented by Wang et al. (2018) to predict soybean yield in Brazil and Argentina. MODIS surface reflectance product was taken as the input data from 2012 to 2016. The Ridge Regression model was taken as the benchmark model. An LSTM model was used for Argentina, and transfer learning was utilised for this model to train for Brazil (Wang et al., 2018). The research concludes that the results from transfer learning had a higher accuracy.

A drawback of the above mentioned studies is that there is little exploration into the spatial importance of the model. More emphasis is provided on temporal analysis and forecasting.

A study by Khaki & Wang (2019) used a deep neural network model to estimate crop production of corn hybrids from 2008 to 2016 throughout the United States and Canada. The input data that they utilised was tabular and had no spatial information. Other Machine Learning models, including Least Absolute Shrinkage and Selection Operator (LASSO), Shallow Neural Network (SNN), and Regression Tree, were compared to the findings. It was discovered that the prediction and performance of the Deep Neural Network model had higher accuracy. One of the limitations noted by the authors is the model's black box mechanism, which does not offer the user any transparency (Khaki & Wang, 2019).

Some relevant papers that specifically implement CNN models for crop yield estimation using remote sensing images are mentioned below.

- A study by Nevavuori et al. (2019) implemented a CNN model to predict wheat and barley yield using RGB and NDVI images collected from UAVs (Unmanned Aerial Vehicles). K- Fold cross-validation is performed to optimise the training dataset, and the Adadelta algorithm was chosen as the best-performing optimisation algorithm. The results indicated that the model provided higher accuracy from the RGB layers than the NDVI layers (Nevavuori et al., 2019).
- In a research conducted by Yang et al. (2019) for estimating rice grain yield, RGB and multi-spectral images were acquired from a UAV for the study region Binyang County, Guangxi Province, China (Yang et al., 2019). The spatial resolution for the images was 0.04 m for RGB and 0.2 m for the multi-spectral layers. Images were captured on six different days from August 2017 to November 2017, depending on the phenological stages of the crop. The CNN architecture implemented used RGB images and multi-spectral images separately. The Stochastic Gradient Descent (SGD) with momentum was used as the optimisation algorithm. The model results showed a higher weightage for the RGB images than the multi-spectral images.
- Terliksiz & Altlyar's (2019) work used a 3D-CNN model to predict soybean yield in Lauderdale County, Alabama, USA (Terliksiz & Altlyar, 2019). Satellite imagery from MODIS products, like surface reflectance, LST (Land Surface temperature) and Land-Cover maps, were provided as input from 2003 to 2016. The training samples were collected from the USDA (United States Department of Agriculture). The results from the model provided an accuracy of 0.81 RMSE for a 20% cropland coverage (Terliksiz & Altlyar, 2019). One of the limitations revealed in this research is that the cropland coverage changes every year, which means that the data frame for each year cannot be set to a fixed dimension.

Even though these studies utilised remote sensing imagery and provided emphasis to spatial dataset, there was not much focus on interpreting the deep learning models.

A few studies that focused on the explainability of CNN models for yield prediction using Earth Observation data are stated.

- In a study by Wolanin et al. (2020), an explainable 1D CNN model was implemented to estimate the crop yield along the Indian wheat belt (Wolanin et al., 2020b). The input features were arranged temporally for the training dataset. Regression Activation Mapping (RAM), a post-hoc method, interpreted the model's feature characteristics and importance. Other interpretation

techniques were not explored. Since the approach utilised a 1D CNN model along the temporal dimension, the spatial heterogeneity of the region was also not considered.

- Another study by Srivastava et al. (2022) predicted the yield of wheat for winter utilising a CNN model with a fully connected neural network, considering the temporal aspects of the data layers (Srivastava et al., 2022). The input features considered included weather and soil parameters, along with the phenological data like sowing, flowering and harvest seasons of wheat. The resulting performance of the CNN model was better than the other models like Random Forest, K-Nearest Neighbour, Lasso and Ridge Regression, Support Vector Regression, XGBoost, and Deep Neural Networks (DNN). The SHAP (Shapley Additive exPlanations) module was used as a post-hoc interpretation of the feature characteristics. However, this study does not explore other means of explainability utilising saliency maps or intrinsic methods.
- In another study by Stomberg et al. (2021), an intrinsic approach was implemented to classify wilderness from remote sensing data using CNNs. Since nature has no proper classification samples, the research focused on clustering the activation maps from the annotated dataset as potential training samples (Stomberg et al., 2021). The results indicated that the model classified non-wilderness regions better than wilderness regions.

The recent research conducted by Wolanin et al. (2020b), Srivastava et al. (2022) and Stomberg et al. (2022) focused on the interpretation and explainability of the deep learning model. However, Wolanin et al. (2020b) focused on the temporal aspect, where the data was prepared for a 1D CNN model. On the other hand, Srivastava et al. (2022) placed greater emphasis on the feature importance and conducted an analysis to assess the impact of input features on the model. There has been a lack of spatial explainability when analysing the model's results. However, Stomberg et al. (2022) utilised Class Activation Maps (CAM) and performed sensitivity analysis to explore how the model behaves when some input features were masked. Nevertheless, the analysis was performed for the discrete classification of wilderness, while our research will attempt to estimate the crop yield, which is a continuous value. Table 1 provides an overview of the relevant background research conducted related to crop yield estimation.

### **1.3. Research Gap**

The field of crop yield estimation and earth observation is currently characterised by an absence of Explainable Convolutional Neural Network (CNN) models. Several studies have been conducted to explore the implementation of post-hoc interpretation techniques in the context of crop yield estimation (Srivastava et al., 2022; Wolanin et al., 2020b). The utilisation of post-hoc techniques offers valuable analytical insights into both the significance of features and the functioning of the model. However, the majority of the post-hoc methodologies employ temporal analysis in prior studies. The current state of research indicates a lack of spatial analysis techniques that incorporate the use of saliency maps. To our knowledge, this is the first time research has been performed utilising explainable methods to analyse yield estimation spatially. As stated in section 1.2, previous research employed models specifically designed for temporal forecasting in order to make predictions about crop yield. Explanation is also conducted within a temporal framework. It is vital to comprehend the spatial characteristics that impact the estimation of yield. The importance of identifying the specific spatial regions that directly influence the decision-making

process of a model should be emphasised, even if the model exhibits strong performance and high accuracy. This is crucial in order to address any potential biases that may arise in the outcomes.

Author	Methodology	Research Gap
<b>(Schwalbert et al., 2020)</b>	LSTM to forecast soybean yield	CNN models are not used. Also, Interpretation techniques are not explored.
<b>(Khaki &amp; Wang, 2019)</b>	Deep Neural Network to estimate corn hybrids	
(Wang et al., 2018)	LSTM transfer learning model to predict soybean yield	
(Nevavuori et al., 2019)	CNN model from UAV images to estimate wheat and barley yield	Interpretation or explainable techniques are not implemented.
(Yang et al., 2019)	CNN model from UAV images (RGB and multi-spectral) to estimate rice yield	
<b>(Terliksiz &amp; Altylar, 2019)</b>	3D CNN model using MODIS imagery to estimate soybean yield	
(Wolanin et al., 2020a)	Explainable 1D-CNN to estimate wheat yield	Only Regression Activation Maps were used. The spatial aspect is not considered, as only the temporal dimension is used.
(Srivastava et al., 2022)	CNN+FC to predict winter yield	No explainable technique was implemented. Only the SHAP module was used for interpreting the feature characteristics.
(Stomberg et al., 2021)	Intrinsic Explainable CNN To detect wilderness	Used only for discrete classification, not for a regression model

Table 1: Literature Review

## 1.4. Objectives and Research Questions

The main objective of this research is to develop an explainable Convolutional Neural Network Model to estimate crop yield.

To achieve the main objective, we need to work towards the following sub-objectives:

1. To evaluate the accuracy of a linear CNN model for crop yield estimation with a baseline linear regression model.
2. To quantify the explainable methods and identify which are suitable for crop yield estimation.
3. To critically analyse the use of explainable methods and detect bias or irregularities in the yield estimation.

To achieve the main objectives and sub-objectives, we must answer the following research questions.

Research Questions for Sub Objective 1:

- What is the level of accuracy for the CNN model compared to the linear regression model?
- Which characteristics are essential for estimating crop yield?

Research Questions for Sub Objective 2:

- What differences can be observed between the explainable methods regarding their performance and accuracy?
- Which explainable methods are ideal for crop yield estimation?

Research Questions for Sub Objective 3:

- How is explainability valuable in the case of crop yield estimation?



## 2. STUDY AREA AND DATASETS

Soybean production of the five leading states in the United States is considered for the study region. The United States is one of the primary producers of the world's soybean (USDA & National Agricultural Statistics Service, 2022). In 2021, the total production of soybeans in the United States reached 4.44 billion bushels, where 1 bushel equals 60 lbs (27.216 Kg). Compared to 2021, production and harvest area increased by 5% in 2022 (USDA & National Agricultural Statistics Service, 2022). There was also an increase, up to 0.4%, in the average yield (USDA & National Agricultural Statistics Service, 2022).

Since Illinois, Iowa, Minnesota, Nebraska, and Indiana are the five states accounting for more than 50% of the soybean output for the United States (USDA & National Agricultural Statistics Service, 2022), these states were chosen for the study region, highlighted in red in Figure 1. Hence, an accurate prediction of the yield in the following years is essential to decide how to implement further action. An analysis of the temporal change over the years also determines that the soybean yield has steadily increased over the past six years (USDA & National Agricultural Statistics Service, 2022). However, the yield rate needs to be proportionate to the demands of the increasing population (Antony, 2021).

It is also important to note that the spatial and temporal distribution of the datasets is adequate for this study area. This is necessary to provide a reasonable explanation of the results.



Figure 1: Map of Study Region

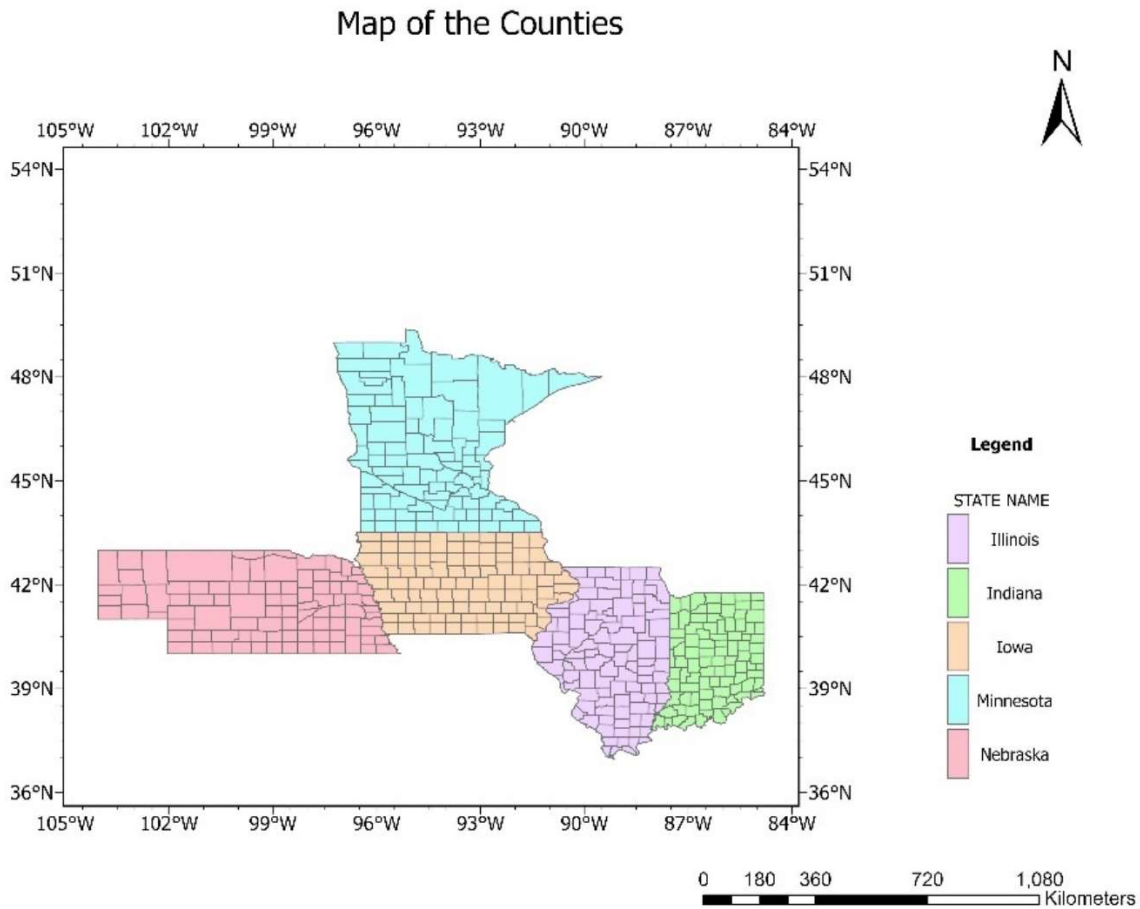


Figure 2: Map of the Counties

Dataset	Source
<b>Cropland Data Layer (CDL)</b>	GEE (Google Earth Engine), USDA
<b>Sentinel-2 Multi-Spectral Images (2017 – 2021)</b>	GEE (Google Earth Engine)
	2017-2021 (Level-2A orthorectified atmospherically corrected surface reflectance)
<b>Crop Yield (2017 to 2021) for counties</b>	USDA, NASS ( <a href="https://www.nass.usda.gov/">https://www.nass.usda.gov/</a> )

Table 2: Datasets Used

Table 2 displays the datasets that will be used in this study. The input images provided are from Sentinel 2 MSI (Multi-Spectral Instrument). The images for 2017 to 2021 are acquired from Level 2A orthorectified atmospherically corrected surface reflectance. The sentinel-2 data consists of 12 bands, as shown in

**Table 3.** The cloud cover percentage is set to 10% when downloading the sentinel images. The aggregate of the sentinel images for each year is taken in July. The month of July is selected as the time interval due to the fact that soybean typically reaches its mid-growth stage during this period, as indicated by the crop calendar depicted in Figure 3. Initially, the cropland layers are downloaded and clipped to the geographical boundaries of the study region. Afterwards, the sentinel images are downloaded, also clipped to the study region. Due to the substantial volume of data that is being downloaded, Google Earth Engine splits the images into tiles. Later, these are merged accordingly to encompass the study area.

Band	Description	Pixel Size
<b>B1</b>	Aerosols	60 meters
<b>B2</b>	Blue	10 meters
<b>B3</b>	Green	10 meters
<b>B4</b>	Red	10 meters
<b>B5</b>	Red Edge 1	20 meters
<b>B6</b>	Red Edge 2	20 meters
<b>B7</b>	Red Edge 3	20 meters
<b>B8</b>	NIR	10 meters
<b>B8A</b>	Red Edge 4	20 meters
<b>B9</b>	Water vapour	60 meters
<b>B10</b>	Cirrus	60 meters
<b>B11</b>	SWIR 1	20 meters
<b>B12</b>	SWIR 1	20 meters

Table 3: Sentinel-2 Bands (Source: Google Earth Engine)

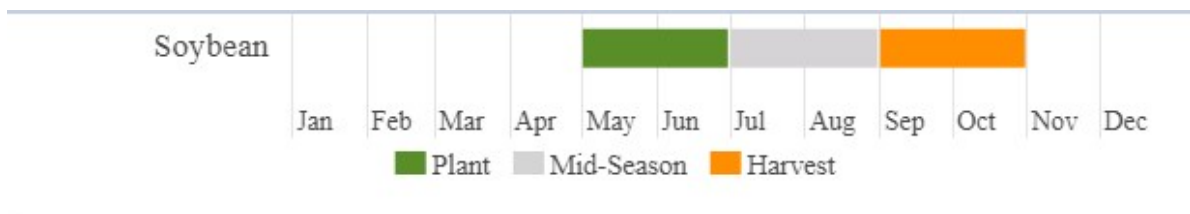


Figure 3: Cropland calendar for Soybean

(Source: USDA & National Agricultural Statistics Service, 2022)

A temporal extent from 2017 to 2021 is taken for the soybean yield value, spatially distributed over the counties of the five states, having 1956 observations. Figure 2 displays the counties for the selected extent. In total, there are 473 counties. The period is chosen based on the availability of the sentinel 2 dataset. Table 4 displays the resulting observations acquired for each year.

Year	Observations (Number of counties)
2021	345
2020	421
2019	370
2018	401
2017	419

Table 4: Number of counties having yield values per year

The CDL (Cropland Data Layers) are compiled maps of the classification of croplands in the United States by USDA (United States Department of Agriculture) from 2017 to 2021. They have a spatial resolution of 30 meters. These maps can be downloaded from Google Earth Engine to identify the regions where soybean is cultivated. Figure 4 displays the cropland covered by soybean in 2021. Since all the bands in sentinel-2 do not have the same resolution (Refer to Table 3), they are all resampled to 60 meters. The CDL maps are also resampled from 30 to 60 meters to be overlaid with the sentinel bands. Bilinear interpolation technique is used for resampling since it provides a close approximation to the original values.

Cropland Data Layer for Soybean (2021)

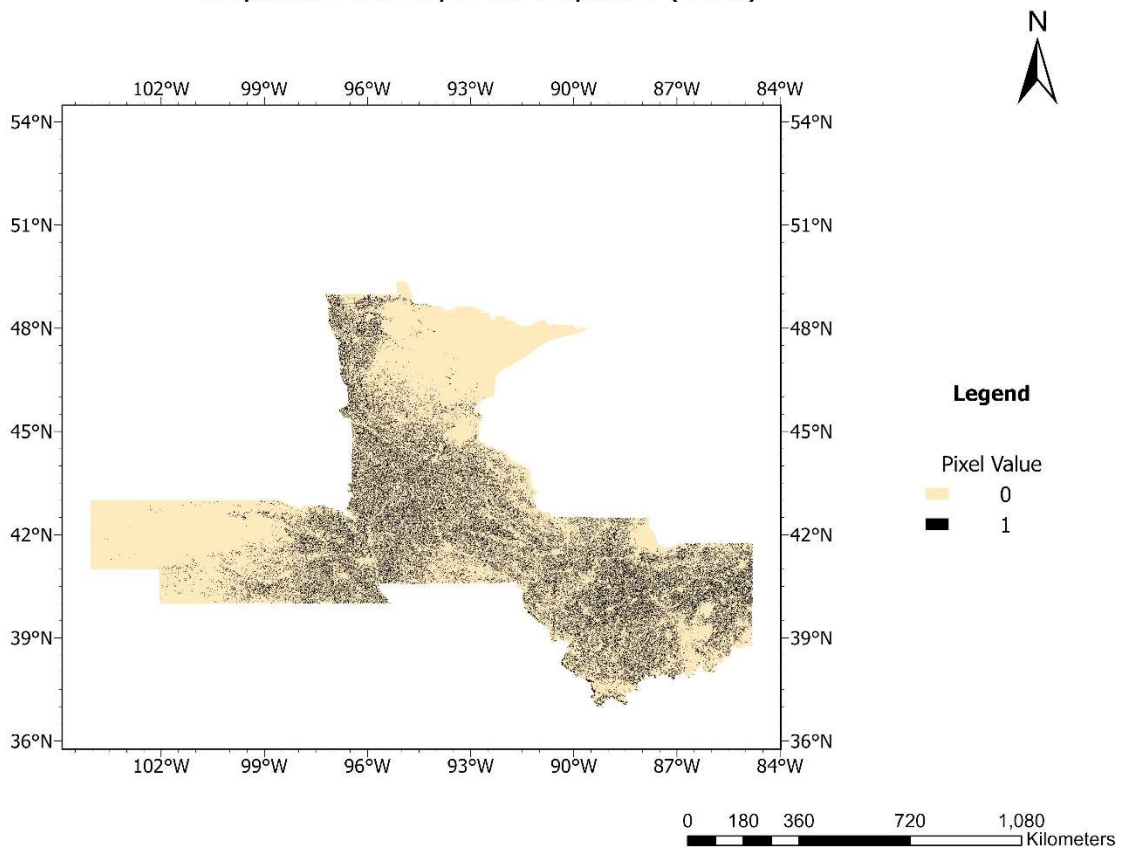


Figure 4: Cropland Data Layer for Soybean (2021)

### 3. METHODOLOGY

The workflow of this thesis is shown in Figure 5. It starts with downloading the sentinel images and cropland data layer for the five states from 2017 to 2021. Afterwards, the dataset is prepared by generating patches with a smaller area and a target yield for each patch by downscaling the yield values at the county level. This is done so that there could be sufficient amount of patches to train the model. Next, two CNN models are trained, one with the cropland data layer as a mask and the other without the cropland data layer. The training of two models is done to compare how effectively the model could extract relevant information with respect to the mask layer. A linear regression model is also fitted and used as a baseline comparison to the CNN models. The dataset is split into training, validation and testing. Subsequently, the test dataset is used for evaluating the models. Regarding the explainability of the models, different saliency maps are generated for the CNN models. The saliency maps provide an understanding of which regions from the input patches are considered important by the CNN models when estimating the crop yield. A perturbation analysis is performed to test their sensitivity and identify which explainable method has the best performance. Later, to check if the models have any bias while estimating the yield, the saliency maps are analysed by comparing with various vegetation indices and the landuse map to find any correlation.

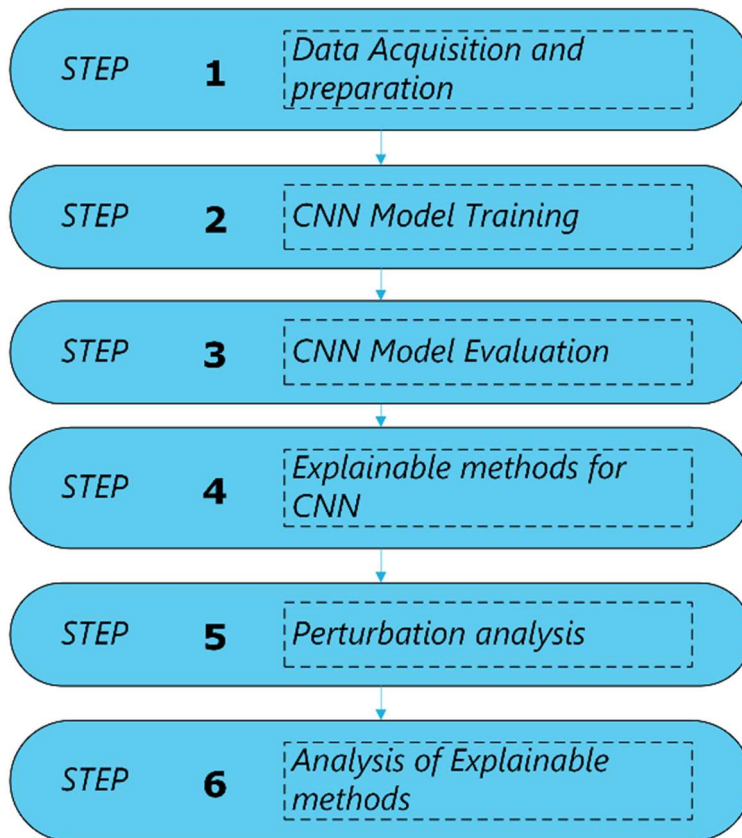


Figure 5: Methodology workflow

### 3.1. Dataset Preparation

Figure 6 provides an overview of the dataset preparation for the CNN model. The first dataset selected is the Crop Land Data Layer of soybean for 2017 to 2021. These layers are then clipped to the study region comprising the five states. Additionally, the sentinel 2 dataset from 2017 to 2021 is selected and also clipped to the study region. The clipped layers of the sentinel bands and the cropland data layer are merged, resulting in a merged dataset having 13 channels. Next, the pixel values of the sentinel-2 bands are normalised from 0 to 1 and patches having dimension of  $256 \times 256 \times 13$  are created from the normalised and merged dataset. These will serve as the input features for the CNN model. A patch size of  $256 \times 256$  pixels was chosen as it provided a sufficient number for training the CNN model. In total, 18816 patches are created. Now, for each patch, we require a target yield value. In order to prepare the target, the yield values are first selected for all the counties within the five states. These yield values are converted to metric units and scaled down by dividing with  $10e7$ . This scaling down is performed instead of standard scaler to retain the target values in Kg. Since each county has a larger spatial extent than a patch of  $256 \times 256$  pixels, the yield for each patch is calculated by downscaling the county level yield based on the area of soybean field covered in that patch. This will provide a dataset having a target soybean yield for each patch, where each patch consists of the 12 bands from Sentinel-2 and a mask layer of the soybean fields (Refer to Figure 8).

Originally, the yield is measured in bushel/acre, where 1 Bushel is 27.2 Kg and 1 acre is 4046.86 sq. m, which is  $63 \times 63$  meters approximately. The values are converted to metric units of Kg/ sq. meters.

In total, there are 1956 crop yield observations per county from 2017 to 2021 (Refer to Table 4). A map for each year is first generated by grouping the yield values of all the counties per year.

To ensure that no bias occurs in the model training due to the varying range of the bands, all the bands and predictor values are normalised using the StandardScaler method from sci-kit Learn (Pedregosa et al., 2011).

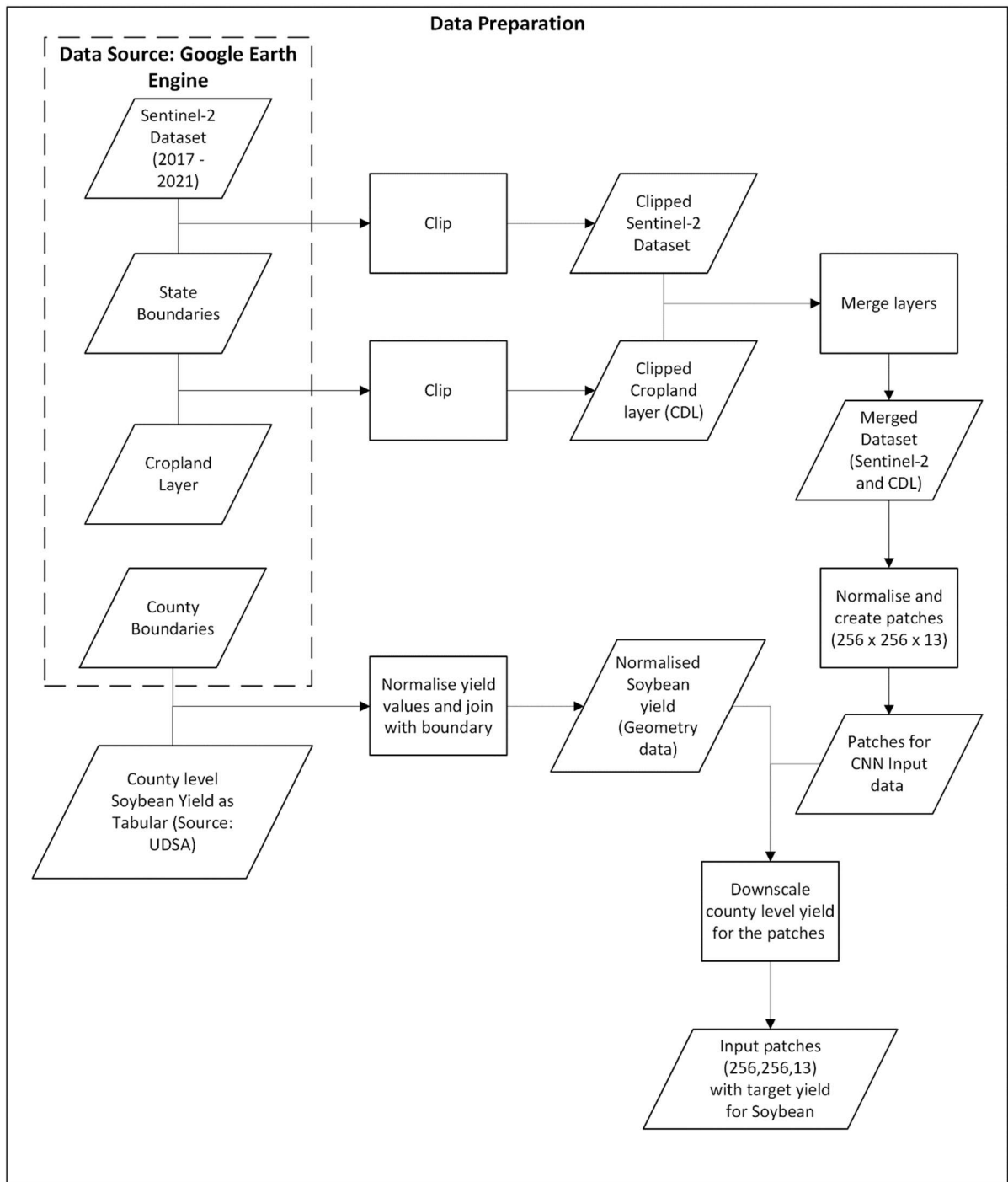


Figure 6: Flowchart depicting the dataset preparation process

Figure 7 shows the yield per county for 2021 in Bushels/acre. From Figure 7, we can see that the yield is missing for specific counties. This is due to a lack of insufficient information such that USDA does not estimate the yield for that year.

### Soybean Yield at County level (2021)

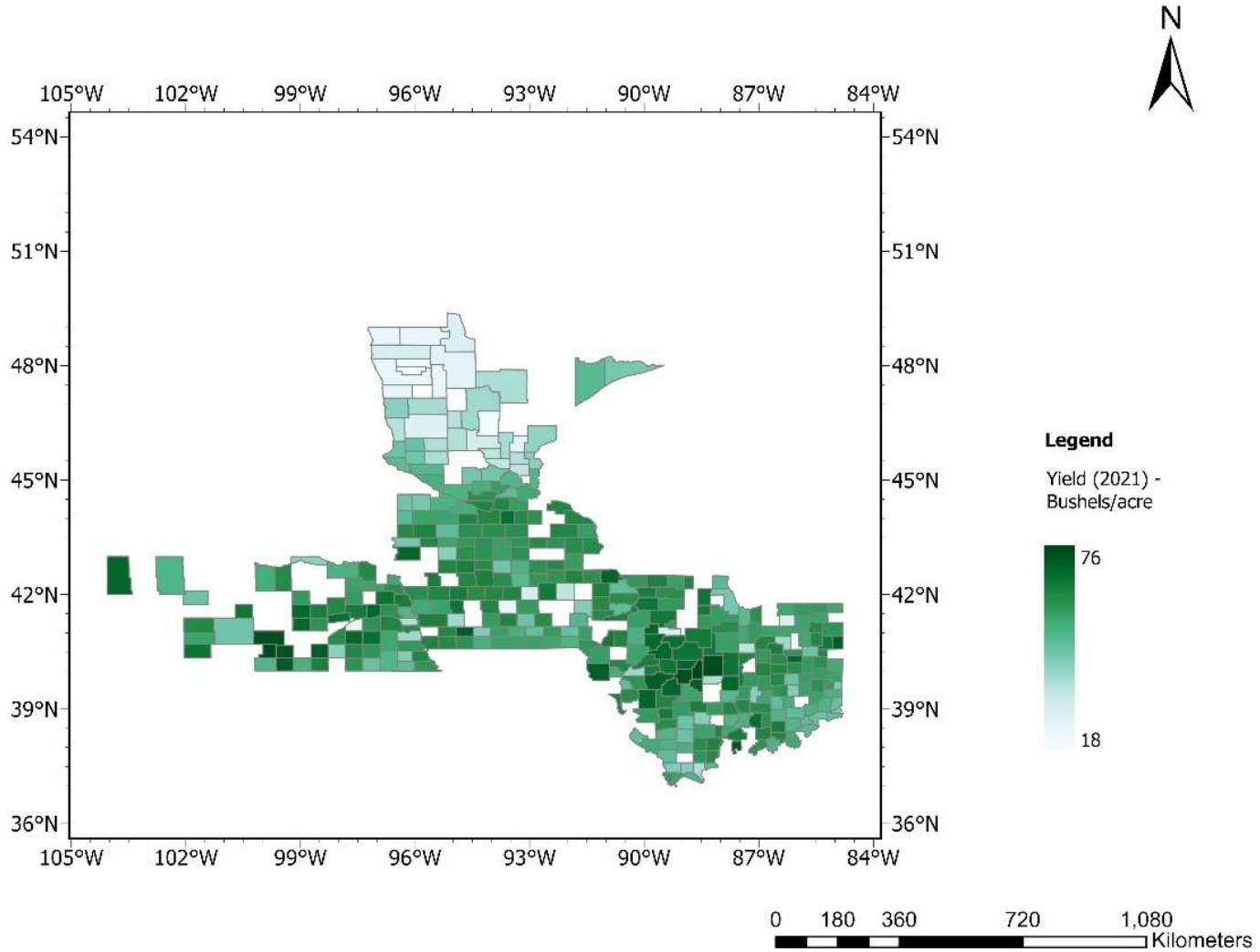


Figure 7: Soybean Yield (Bu/acre) at the county level for 2021

Patches are created for the sentinel 2 dataset by creating 256 x 256 pixel grids. The target yield for each patch is computed by distributing the yield per county with respect to the area of cropland utilised for soybean within that patch. Figure 8 shows how one patch of 256 x 256 pixels look. Each pixel is 60 x60 m. Thus the extent of one patch is 15.36 Km



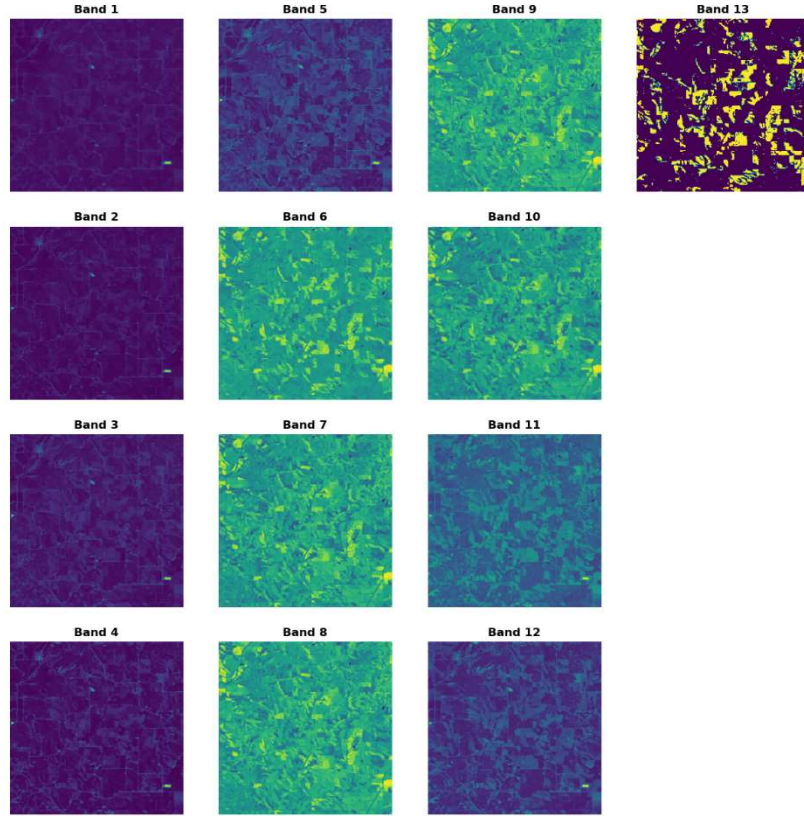


Figure 8: The sentinel bands and croplands of soybean (Band 13) of a patch

The target yield for each patch is calculated as follows.

$$Y = \sum_{j=1}^{j=n} y_j * A_j$$

Equation 1: Target Yield calculation

$Y$  -> total yield for that patch (Kg)

$n$  -> Total number of counties that covers the patch

$y_j$  -> Average yield of county  $j$  (Kg/sq. m)

$A_j$  -> Area covered by the cropland for county  $j$  in sq. m

For example, let there be a patch that covers 5 counties. This implies that there will be five different yield values ( $Y_1$  to  $Y_5$ ), one per county, for that patch, as shown in Figure 9. The yield for this patch will then be calculated as follows:

$$Y = Y_1 * A_1 + Y_2 * A_2 + Y_3 * A_3 + Y_4 * A_4 + Y_5 * A_5$$

Where  $A_1$  to  $A_5$  are the area of cropland covered by that portion of the patch for the corresponding county.

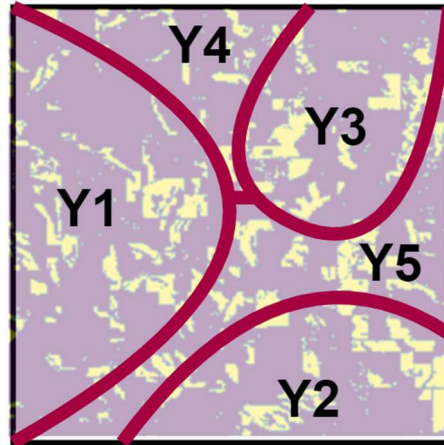


Figure 9: Example showing how the target yield is prepared for a patch

There are 18816 patches in total. Figure 10 displays the target yield for the patches from 2017 to 2021 for the study region. Something that can be inferred from these plots is that the yield value is higher in Illinois and Indiana. Also, the north-eastern part of Minnesota and large portions of Nebraska have lower yields in general for all the years from 2017 to 2021. This implies that compared to Illinois and Indiana, soybean is not predominantly grown in Minnesota and Nebraska. Performing a temporal analysis on this dataset will also not be possible for all the patches, since the yield values are missing for a few patches at the same location per year.

Target Yield for patches from 2017 to 2021

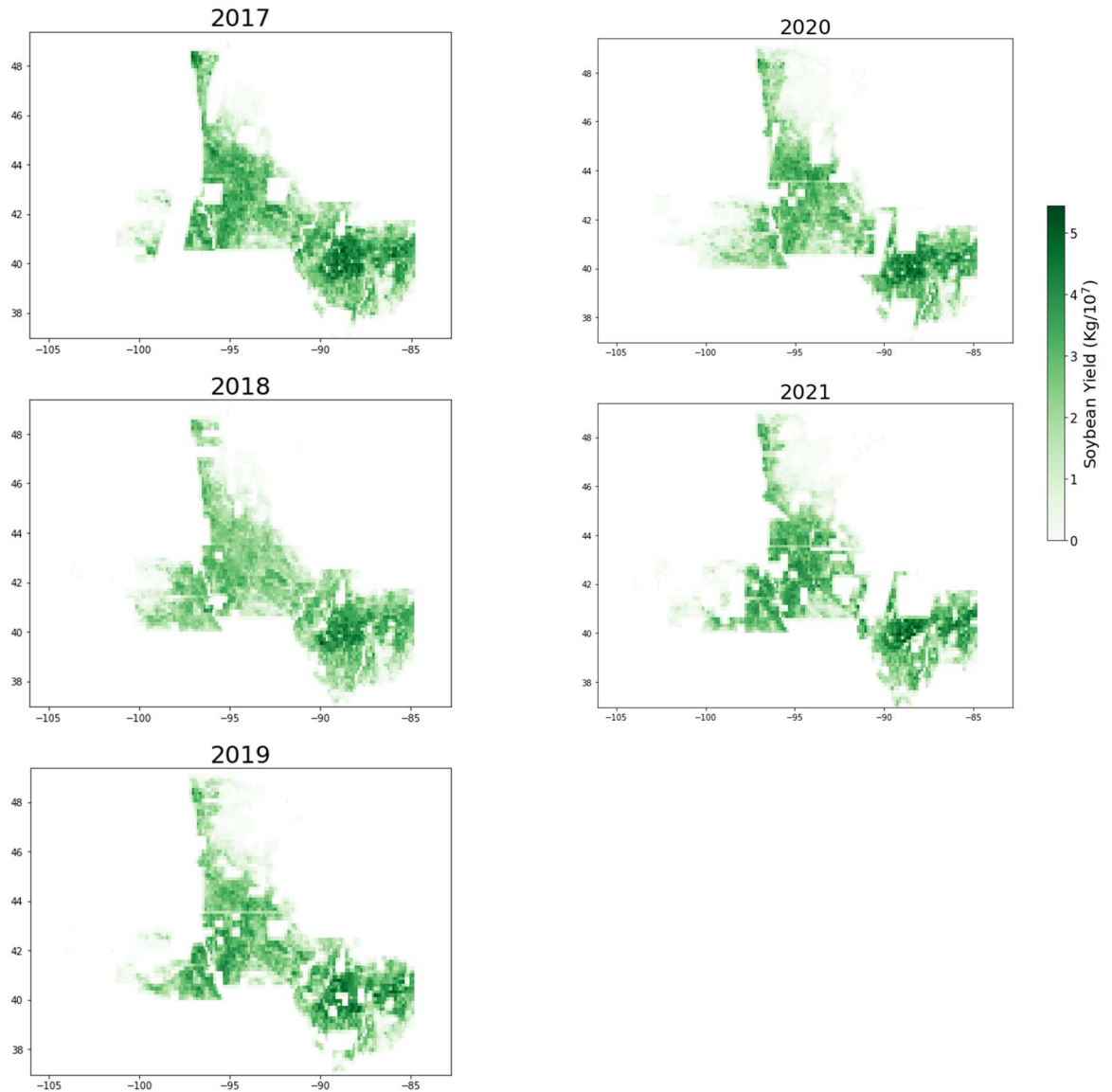


Figure 10: Target yield for patches

### 3.2. Model Architecture

The model architecture is illustrated in Figure 11 and shown in Table 5. A simple CNN model architecture is initially designed and tested. The model has three convolutional layers with a kernel size of 3 x 3 and three max-pooling layers with a kernel size of 2 x 2. Max pooling is used in CNN models to downsample feature maps, reducing their spatial dimensions while retaining the most salient features by selecting the maximum value within each pooling region, aiding in translation invariance and efficient feature extraction (Lecun et al., 1998). After the third max-pooling layer, the model flattens the next layer and is connected to a densely connected network having two layers of node size 64 and 32. All the layers have ReLU (Rectified Linear Unit) activation functions except the output layer. ReLU introduces non-linearity to the

network, which is essential for learning complex patterns and making the CNN capable of approximating any arbitrary function. ReLU is a simple function that only activates if the input is positive, effectively introducing non-linearity by breaking the linearity of the input range (Glorot et al., 2011). The final output layer has one node with a linear activation layer since the output is a continuous variable.

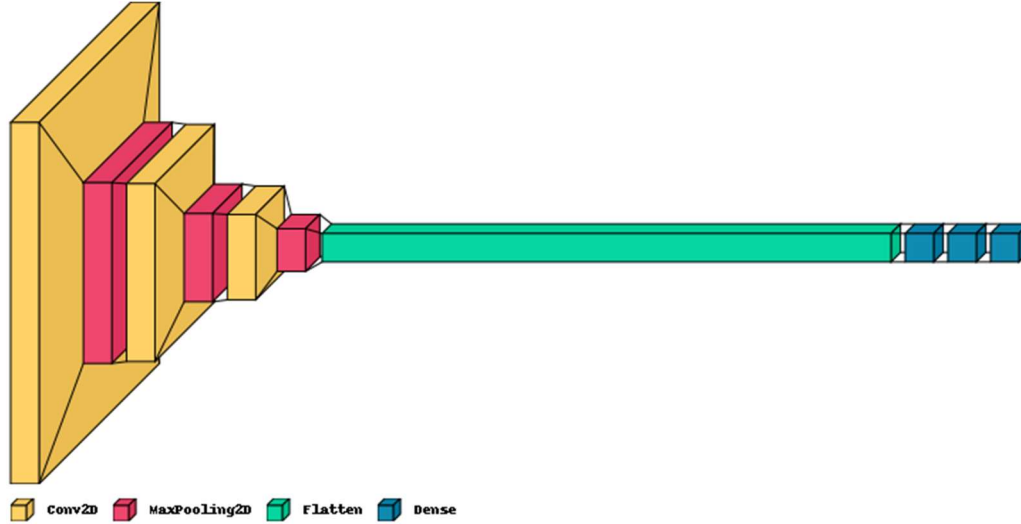


Figure 11: CNN model architecture

Type	Kernel Size	Output Shape
InputLayer	None	256,256,13
Conv2D	3x3	254, 254, 32
MaxPooling2D	2x2	127, 127, 32
Conv2D	3x3	125, 125, 64
MaxPooling2D	2x2	62, 62, 64
Conv2D	3x3	60, 60, 64
MaxPooling2D	2x2	30, 30, 64
Flatten	None	57600
Dense	None	64
Dense	None	32
Dense	None	1

Table 5: CNN Model architecture

The defined model architecture is very simple. Initially, the idea is to find out how well the model performs in estimating the yield. A comparison is made to find out how the mask layer affects the yield estimation. Since we compare the CNN model’s accuracy based on the influence of the cropland mask, further layers are not added. Subsequently, a linear regression model is also implemented using Sci-kit Learn (Pedregosa et al., 2011) to assess the accuracy of the CNN models.

### 3.3. Model Training and Evaluation

Table 6 shows the hyperparameters that were used to train the CNN models. Two CNN models were trained, one including the mask layer and the other without the mask layer. This can be noted in the row showing patch dimension, where the shape (256,256,13) includes the mask layer and the 12 sentinel bands.

Hyperparameters	Value	
Batch Size	64	
Epochs	50	
Learning Rate	0.0001	
Optimiser	Adam	
Loss Function	MSE (Mean Squared Error)	
Metrics	MAE (Mean Absolute Error)	
Number of patches	16225	
Patch dimension	256,256,13	256,256,12
Years	2017-2021	

Table 6: Hyperparameters of the CNN model

The batch size is set to 64. This means that for each epoch, 64 samples are trained before the model's parameters are updated. Increasing the batch size will result in a better loss curve, which may lead to a model having better accuracy. However, it takes more computational power, which would result in higher memory consumption.

An epoch refers to a single pass through the entire training dataset, during which each training sample is used to update the model's parameters. The number of iterations within an epoch is equal to the number of patches for training divided by the batch size. With a batch size of 64, one epoch will have 190 iterations.

The Adam optimiser is a popular optimisation algorithm used in CNN models (Kingma & Ba, 2014). It is chosen since it combines the benefits of both Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp) by adapting the learning rate for each parameter based on its past gradients, resulting in efficient and effective parameter updates during training (Kingma & Ba, 2014).

16225 patches from four states (Iowa, Illinois, Minnesota and Nebraska) are taken for training the CNN model. While training the CNN model, 25% of the patches from the four states will be randomly chosen to improve the validation loss curve. The remaining 2591 patches from Indiana are kept aside for testing and evaluating the model. By selecting Indiana as the test dataset, we ensure that there is no potential for overfitting during model evaluation. This is due to variations in the spectral values and spatial texture between the patches obtained from Indiana and those acquired from other states.

After training the model, the run with the best epoch is chosen to evaluate the test data. The best epoch is when the validation loss curve is close to the training loss.  $R^2$  score and RMSE (Root Mean Square Error) are used as evaluation metrics to compare the accuracy of the estimated values with the true values.

### 3.4. Explainable methods

Several explainable methods are utilised for CNN models. The results are heatmaps/saliency maps to explain how the CNN model estimates the yield. Heatmaps/saliency maps for a CNN is a visual representation that highlights the most important regions or features in an input image that contribute to the network's prediction (Molnar, 2022). It identifies the areas of the image that has the most significant influence on the output of the network, providing insights into which parts of the image the CNN model focuses on to make its decision. Different types of explainable methods can be generated based on various algorithms. Below is a brief explanation of the algorithms used to explain the CNN model.

**Gradient:** A feature map is generated by computing the gradient of the output layer to the input layer (Alber et al., 2019).

**Smooth Grad:** SmoothGrad is used to perturb the input data by adding random noise and then averaging the gradients obtained from multiple noisy samples (Smilkov et al., 2017). This process helps to reduce the influence of random noise on the gradient estimates and provides a smoother and more stable interpretation.

**Deep Taylor:** The Deep Taylor method is based on the Taylor decomposition formula, which is used to simplify a complex formula by expanding the function as a series (Montavon, Lapuschkin, et al., 2017). When implemented for CNN, this method considers the relationship between the input features and the estimated yield. This is done by decomposing the predictions made by the CNN model based on the contributions from each input feature. The process in which Deep Taylor works is explained below:

- First, the output yield of a patch is obtained from the input bands.
- Then, the gradient of the output to the input is calculated by back-propagating the error of the output to the input
- An importance score is set to the input features based on the gradient
- The importance score is distributed to all the layers of the model network. Each layer has an importance level propagated via the activation layers to obtain an accumulated score that signifies the importance level of the input features.

**LRP (Layer-wise Relevance Propagation):** The weights of each neuron in a layer are propagated from the last layer to the input to identify which pixel contributed the most to the output (Bach et al., 2015). The process of LRP is similar to Deep Taylor in terms of propagating the relevance amongst the layers. The difference lies in the method in which the propagation occurs. While Deep Taylor directly assigns relevance to neurons, LRP emphasises the flow of relevance through the network layers, making it more suitable for understanding the overall importance and information flow (Montavon et al., 2019).

There are different variants of LRP, including LRP-A and LRP-B, which differ in their propagation rules and methods of relevance assignment.

LRP-A, also known as “Simple LRP,” is a conservative variant of LRP that aims to preserve relevance conservation. It distributes relevance based on the positive contributions of the neurons in the forward pass. LRP-A divides the relevance among the input features in proportion to their positive contributions in the forward pass (Bach et al., 2015).

LRP-B, also known as “Epsilon LRP,” is a more generalised variant of LRP that allows relevance to flow both through positive and negative contributions (Bach et al., 2015). LRP-B assigns relevance based on

the combination of positive and negative contributions, taking into account both excitatory and inhibitory influences of the neurons (Bach et al., 2015).

**Guided Back Propagation:** Guided Back Propagation is a technique used in deep learning to understand which input features contribute positively or negatively to the final prediction (Springenberg et al., 2014). It modifies the traditional backpropagation algorithm by only allowing gradients to flow through activation units with positive values, highlighting essential features while suppressing irrelevant or negative influences (Springenberg et al., 2014).

**Input\*gradient:** A variation of the gradient method where the input features are multiplied by the gradient (Alber et al., 2019)

**Integrated Gradient:** Integrated Gradient is used to attribute importance to input features by quantifying their contributions towards a model's prediction. It computes the accumulated gradients along a straight line path from a baseline input (typically an input with few or no features) to the actual input, considering varying levels of feature presence (Sundararajan et al., 2017). Integrated Gradients assign relevance scores to each feature, signifying their impact on the model's output by integrating these gradients. This method provides a deeper comprehension of feature importance than simple gradient-based methods (Sundararajan et al., 2017).

**gradCAM:** This is a variation of Class Activation Maps (CAM). In CAM, a weighted activation map is calculated from every layer. CAM focuses on the final convolutional layer of a CNN and calculates the class activation map by taking a weighted average of the feature maps. It assigns importance to each spatial location in the feature maps based on the learned weights, highlighting the regions that contribute most to the predicted class (Selvaraju et al., 2016). However, CAM requires the network to have global average pooling layers to obtain the final class scores.

On the other hand, Grad-CAM extends the idea of CAM by incorporating gradient information from the target class (Selvaraju et al., 2016). Instead of relying solely on the final convolutional layer, Grad-CAM calculates gradients of the target class score with respect to the feature maps of the last convolutional layer (Selvaraju et al., 2016). These gradients represent the importance of each feature map for the target class. Grad-CAM then combines these gradients with the feature maps to generate a heatmap highlighting the image's important regions.

Apart from gradCAM, all the other explainable methods are implemented using the Innvestigate module (Alber et al., 2019). The saliency maps are generated for all the patches and then merged together to visualise the region of importance that the CNN model focuses on to estimate the target yield.

### 3.5. Perturbation Analysis

Since several explainable methods are implemented to assess the behaviour of the CNN model, a quantitative analysis is required to identify which method is the most suitable. One way of quantifying the explainable methods is through perturbation analysis of the input based on the ranking of the importance, as shown in Figure 12. The degree to which the explainability differs quantitatively provides an idea of the method that can be selected for further analysis and modelling (Yeh et al., 2019). Several saliency maps were generated for different explainable methods like LRP (Layerwise Relevance Propagation), Smooth

Gradient, Deep Taylor and gradCAM, shown in Figure 22. The patch is split into 64 grids (8 x 8), each grid having a shape of 32 x 32 pixels. These grids are ranked based on the aggregation of the corresponding explainable method. Based on the ranking order, the grids are perturbed with Gaussian noise, and the patch yield is estimated by passing it to the CNN model. The difference between predicted and actual yield values is saved to a list. This process is continued iteratively by perturbing the grids based on the ranking order. To generalise the perturbation analysis for the entire study region, the area under the curve for each explainable method is utilised.

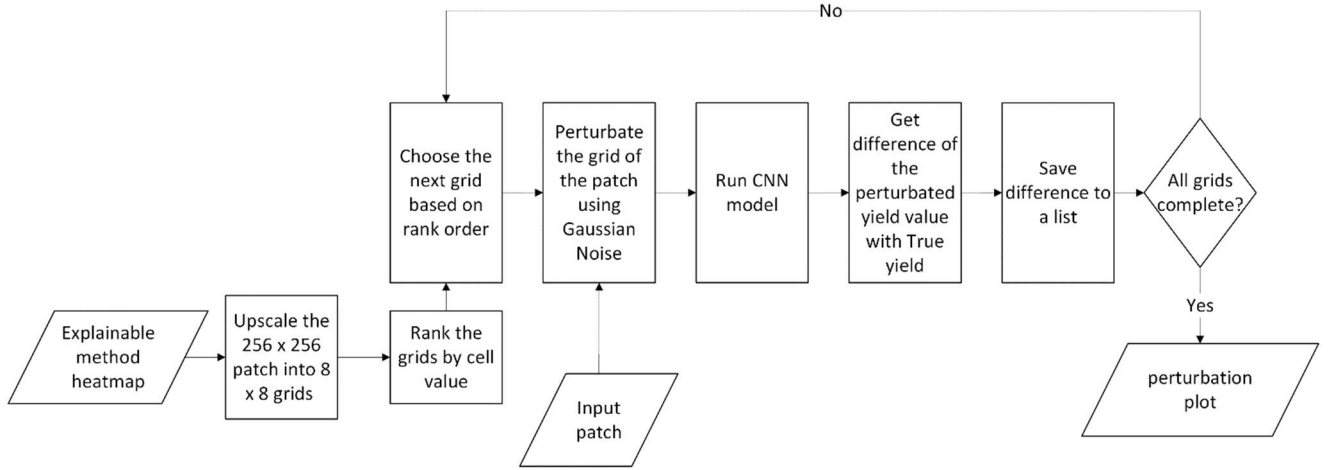


Figure 12: Perturbation Analysis workflow

### 3.6. Analysis of Explainable Methods

To understand whether there is any relation between vegetation indices, landuse and the crop yield, we try to find a connection with the saliency maps. Vegetation indices that are specifically focused on crop/agriculture applications are used to compare with the saliency maps. Crop yield estimation strongly correlates with the vegetation indices and depends on the crop growth cycle (Sakamoto, 2020). A short description of these vegetation indices is provided below.

**Normalised Difference Vegetation Index (NDVI):** This index calculates the vegetation index and is a ratio between NIR (Near Infra Red) and Red wavelength. For sentinel-2, band 8 is NIR and band 4 is Red (Refer to Table 3)

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

Equation 2: NDVI Formula

**EVI (Enhanced Vegetation Index):** Enhanced Vegetation Index (EVI) estimates vegetation density. It is designed to improve the limitations of the Normalized Difference Vegetation Index (NDVI), especially in areas with high levels of aerosols, canopy cover, and soil brightness.



The EVI combines information from the Red (Band 4), Blue (Band 2), and NIR (Band 8) wavelengths of the Electromagnetic spectrum to calculate the index.

$$EVI = 2.5 * \frac{NIR - Red}{(NIR + 6.0 * Red - 7.5 * Blue)}$$

Equation 3: EVI Formula

**NDMI (Normalised Difference Moisture Index):** NDMI assesses the vegetation's moisture content or water stress. It is calculated using the NIR (Band 8) and SWIR (Band 11) wavelengths.

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR}$$

Equation 4: NDMI Formula

**WDRVI (Wide Dynamic Range Vegetation Index):** WDRVI is a vegetation index designed to enhance the sensitivity to vegetation changes in areas with a wide range of vegetation cover and varying atmospheric conditions. It is an improvement over the Normalized Difference Vegetation Index (NDVI) in situations with significant differences in vegetation density or when the canopy cover is not uniform.

$$WDRVI = \frac{0.1 * NIR - Red}{0.1 * NIR + Red}$$

Equation 5: WDRVI Formula

**SAVI (Soil Adjusted Vegetation Index):** The Soil-Adjusted Vegetation Index (SAVI) is a vegetation index that aims to minimise the influence of soil background on vegetation analysis, especially in areas with sparse vegetation or high soil brightness. It is an enhancement of the Normalized Difference Vegetation Index (NDVI) that attempts to correct soil reflectance effects.

$$SAVI = \left( \frac{NIR - Red}{NIR + Red + L} \right) * (1.0 + L)$$

Equation 6: SAVI Formula

L -> soil brightness correction factor that could range from (0 -1)

Here, L is set to 0.5 to minimise the soil brightness factor

To understand how including and excluding the mask layer affects the yield estimation, the vegetation indices and the saliency values are aggregated according to the cropland type, and plotted to check for any correlations. This aggregation is done for each patch and also analysed for all the patches in the test region to understand whether the findings could be generalised.

### 3.7. System Configuration and Specifications

The model training and computation analysis is performed on the geospatial computing platform provided by the Faculty of Geoinformation Science and Earth Observation, University of Twente. The system specifications are as follows:

Unit	Architecture	CPU	Max. Speed (GHz)	Cores	Threads	Memory (GB)	GPU	# of Units
PowerEdge R730	Intel x86-64	E5-2695 v4	3.3	2 x 18	72	768	NVIDIA Titan XP (CC 6.1)	1

Table 7: System Specifications

More information can be found at this link: <https://support.crib.utwente.nl/kb/faq.php?id=19>

By default, TensorFlow attempts to train the CNN model using the system's GPU. However, since the dataset is huge, the server's GPU could not allocate sufficient memory to train the model. Hence, TensorFlow is set to use the server's CPU memory, which has a capacity of 768 GB.

## 4. RESULTS

Three models are trained on the study region for four states (Iowa, Illinois, Minnesota, and Nebraska), and their accuracy and performances are evaluated on the test data (Indiana). One is a Linear Regression model, and the other two models are CNNs with and without the mask layer. The feature importance of the bands is taken from the linear regression model, and the saliency maps are derived from the CNN models. The saliency maps are then utilised to implement a perturbation analysis to determine the robustness and sensitivity of a model spatially.

### 4.1. Loss Curves

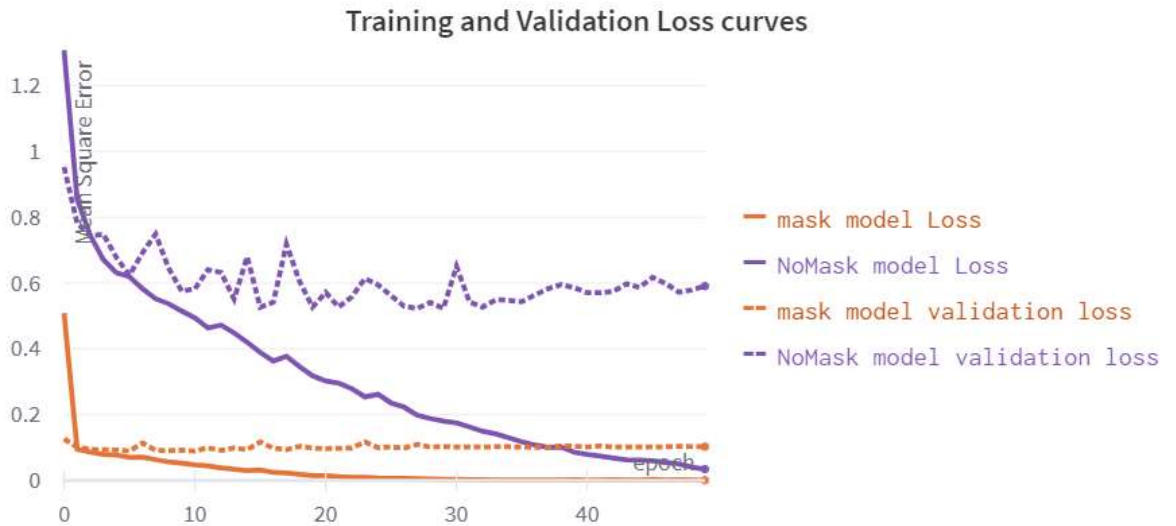


Figure 13: Training and Validation Loss curves for CNN

Figure 13 displays the training and validation loss curves of the two CNN models, one trained with just the sentinel bands (purple) and the other trained with including the cropland layer of Soybean as a mask layer (orange). The curve for the model with a mask layer is much steeper, indicating a better accuracy and fit with the dataset.

### 4.2. Model Evaluation

The models are then implemented on the test data. Figure 14 displays the true vs predicted scatter plot for the three models (Linear Regression, CNN model without mask and CNN model with mask). The CNN model with the mask layers has the closest fit with the data, followed by the model without a mask and the

linear regression model. Table 8 shows the RMSE and  $R^2$  scores of the models. Linear Regression has the lowest  $R^2$  score and highest RMSE, while the CNN models have higher  $R^2$  scores and lower RMSE.

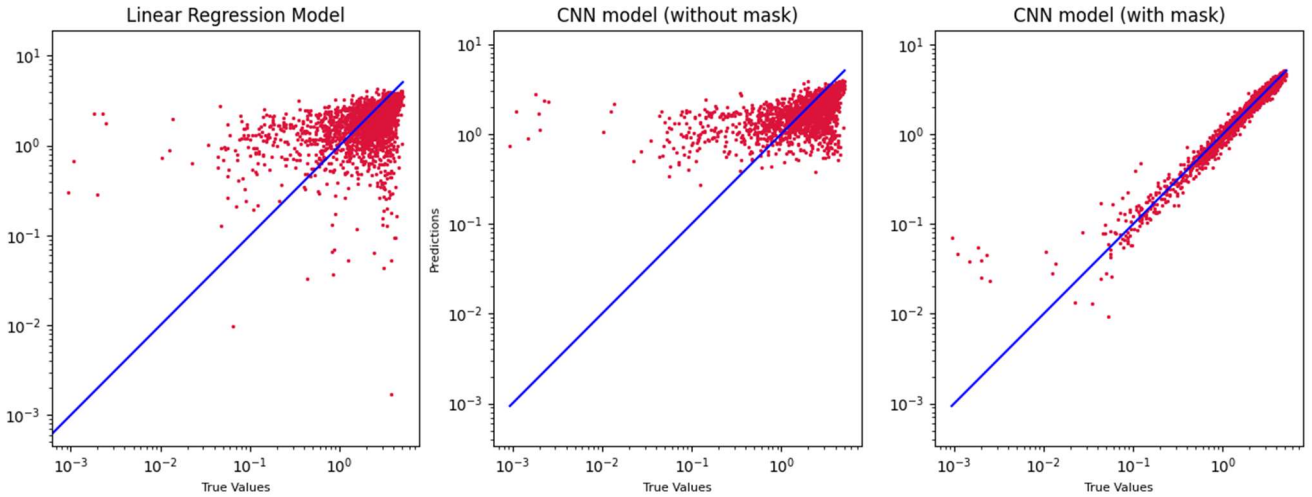


Figure 14: Scatter plot of True vs predicted

Metric	Linear Regression	CNN Model (without Mask)	CNN Model (with mask)
RMSE (Root Mean Squared Error)	1.093	0.641	0.055
$R^2$ score	0.617	0.729	0.982

Table 8: RMSE and  $R^2$  of the models

The evaluation metrics from Table 8 indicate that the CNN models without masks and with masks have higher accuracy and perform 11% and 37% better than the linear regression model on the test dataset.

### 4.3. Difference Map

The difference map is plotted between the true and predicted yield values for the state of Indiana, as shown in Figure 15. The maps are from the years 2017 to 2021. The first row shows the difference map for the linear regression model, the second row for the CNN model without mask and the third row for the CNN model with mask layer.

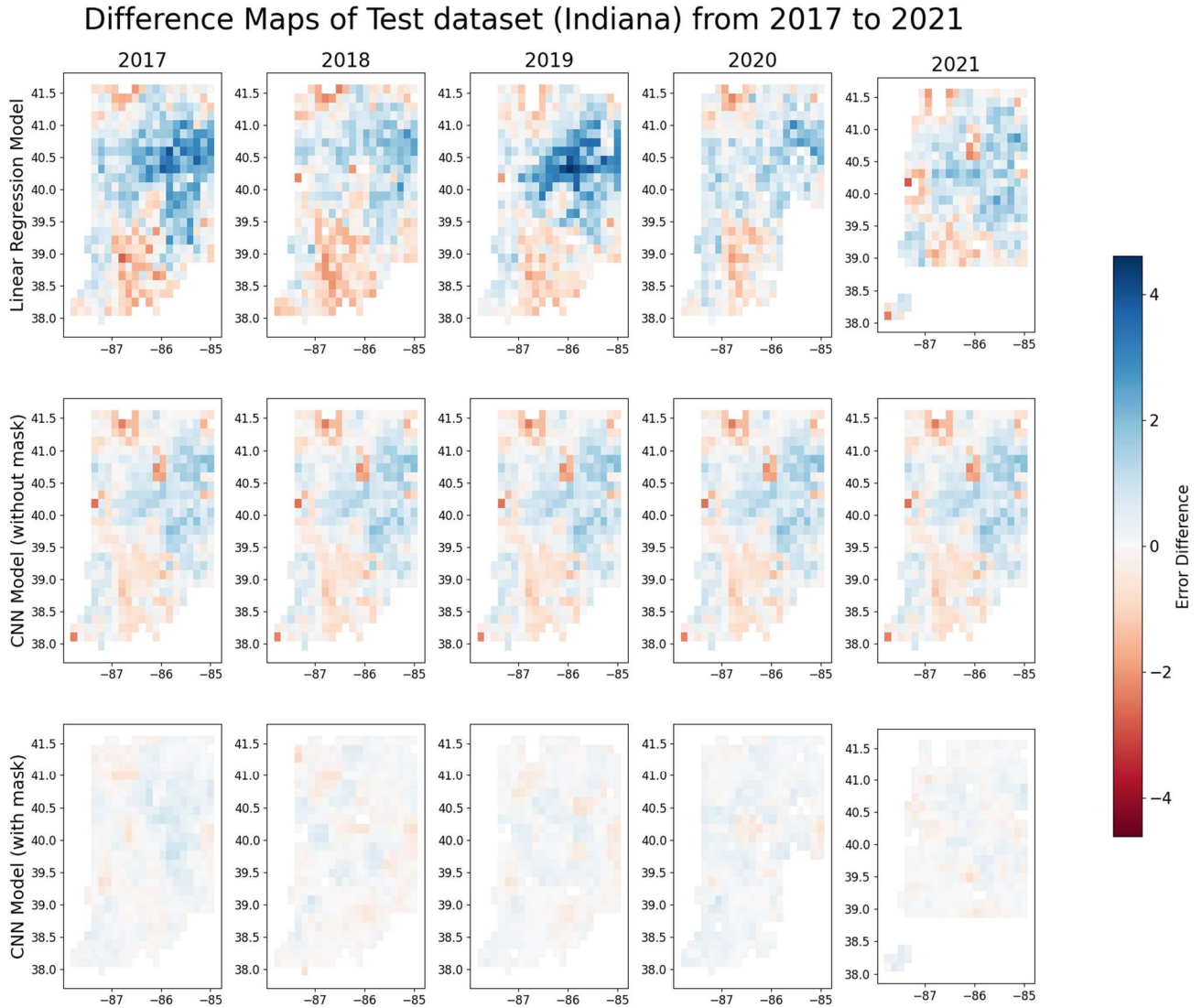


Figure 15: Difference map of True vs Predicted for test dataset (Indiana)

The linear regression model has the highest degree of error, followed by the CNN model without the mask layer. The margin of error is the lowest for the CNN model with a mask layer. There also seems to be a higher range of errors for the years 2017 and 2019. But the error seems to be random. There is no discernible pattern that could be identified.

#### 4.4. Outliers in the scatter plot

Some patches differ considerably from the predicted yield to the true yield. This can be seen as outliers in the scatter plot. Upon looking at these patches, they scarcely have any soybean fields, as depicted in Figure 16. These regions have relatively very few pixels that are identified as Soybeans. Since some pixels are

considered Soybeans, they are also not disregarded from the training and test dataset since they contribute at least a small yield value for the patch.

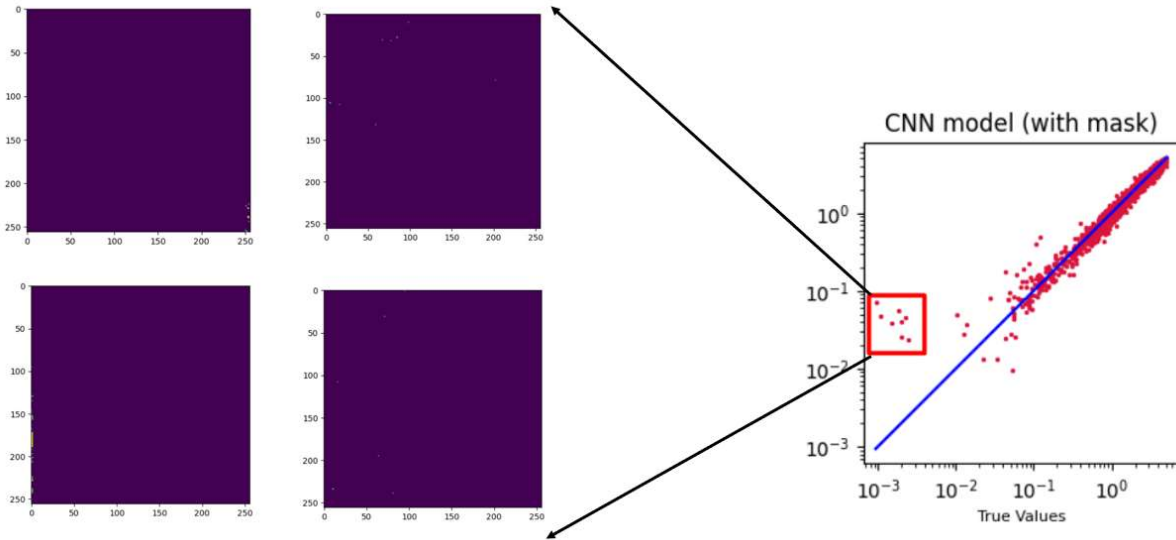


Figure 16: Mask layers of the outliers in True vs Predicted

#### 4.5. Feature importance of Linear Regression

Figure 17 displays the linear regression model's coefficient weights assigned to the sentinel bands. This signifies which features are most significant for predicting the yield. Band 7 (Red Edge 3) has the highest weightage. Interestingly, the model gives Band 8 (Near Infra Red - NIR) low importance, even though the vegetation's spectral reflectance is higher at NIR. After Band 7, bands 5 (Red Edge 1) and 9 (Red Edge 4) have negative significance, and Band 4 (Red) has positive significance. Bands 3 (Green), 6 (Red Edge 2) and 10 (Water Vapour) are given the lowest importance. The lowest weightage is given to bands 3, 6 and 10. Since the linear regression model does not give the best accuracy, these weights are not considered to be optimal for soybean yield estimation. However, the weights are distributed randomly across the bands and are not biased towards specific bands.

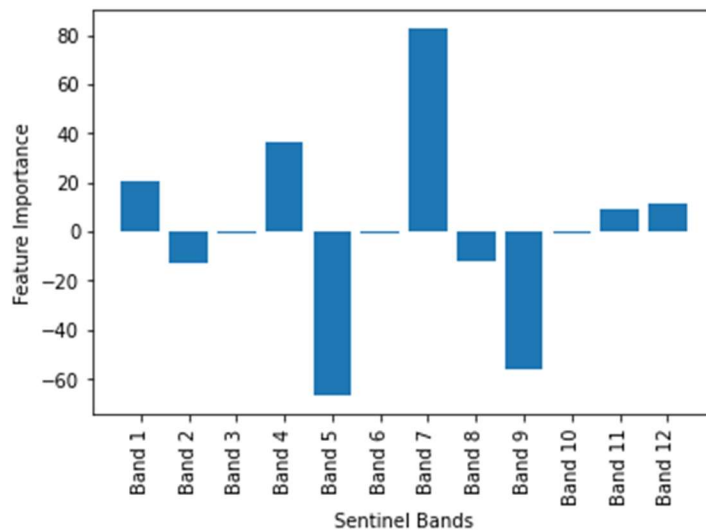


Figure 17: Feature importance of Linear Regression Model

To further analyse the relationship between the sentinel bands, a correlation heatmap is plotted for the mean value of the sentinel bands for all the patches. This heatmap is shown in Figure 18. The correlation between the bands forms an interesting pattern here. Initially, bands 1 to 5 have a high correlation, then bands 6 to 9 have a high correlation, and bands 10 and 11 correlate.

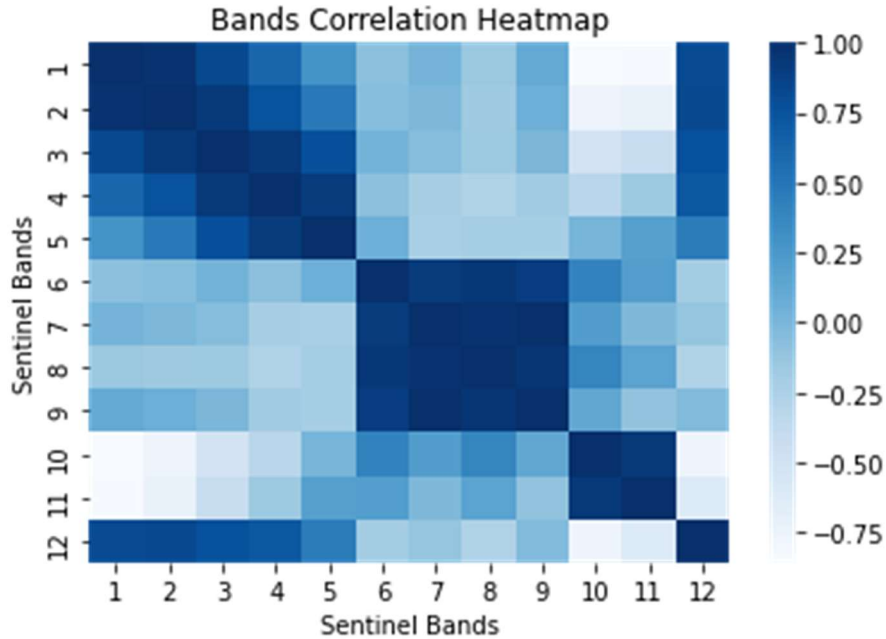


Figure 18: Sentinel bands mean correlation heatmap

From Figure 17 and Figure 18, we can conclude that the dataset chosen has a good relation, and the band's feature importance is well distributed so that it could be provided as input to the CNN model.

#### 4.6. Saliency maps

Several saliency maps using various methods were generated. Figure 19 shows the heatmap generated by gradCAM for one patch taken from Indiana. The red regions (close to 1) are more critical for predicting the yield. Values close to 0 (blue regions) are less important. The gradCAM generated by the CNN model with mask layer mainly focuses on the areas cultivated by soybean.

Figure 20 shows Indiana's soybean yield and gradCAM saliency map of both models (with and without mask layer) from 2017 to 2021. The gradCAM was aggregated by mean for each patch. The scatter plot of the saliency values and the soybean yield is shown in Figure 21. The  $R^2$  score is 0.8307 for the model with a mask layer and 0.1925 for the model without a mask layer. This indicates that the model properly utilises the mask layer, and the pixels of the saliency map correlates highly with the target yield. On the other hand, the gradCAM for the CNN without the mask layer is not focused on the soybean fields due to the lack of the mask layer as an input. Without the mask layer, the model focuses on regions having a higher vegetation index. Further explanation can be found in Section 4.8.



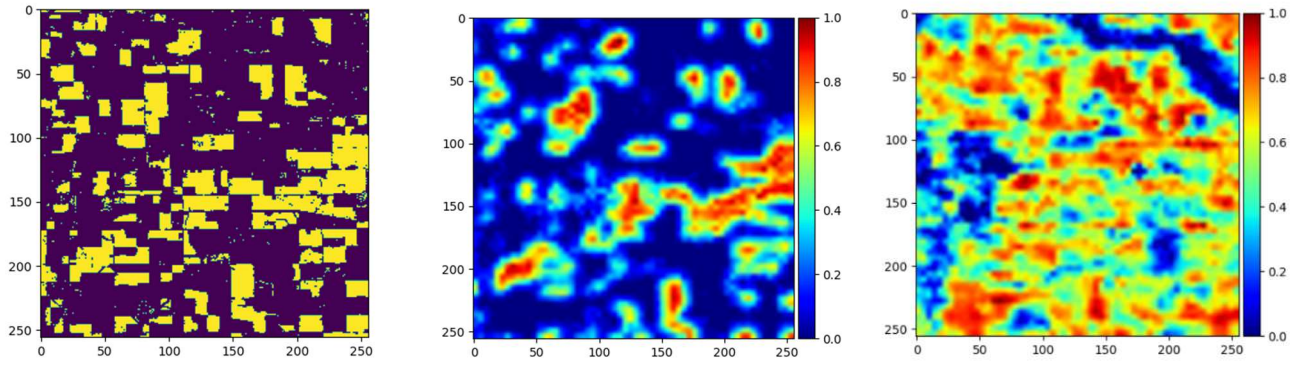


Figure 19: Mask Layer (Left), gradCAM of CNN model with mask (middle) and gradCAM of CNN model without mask (Right) of a patch from Indiana

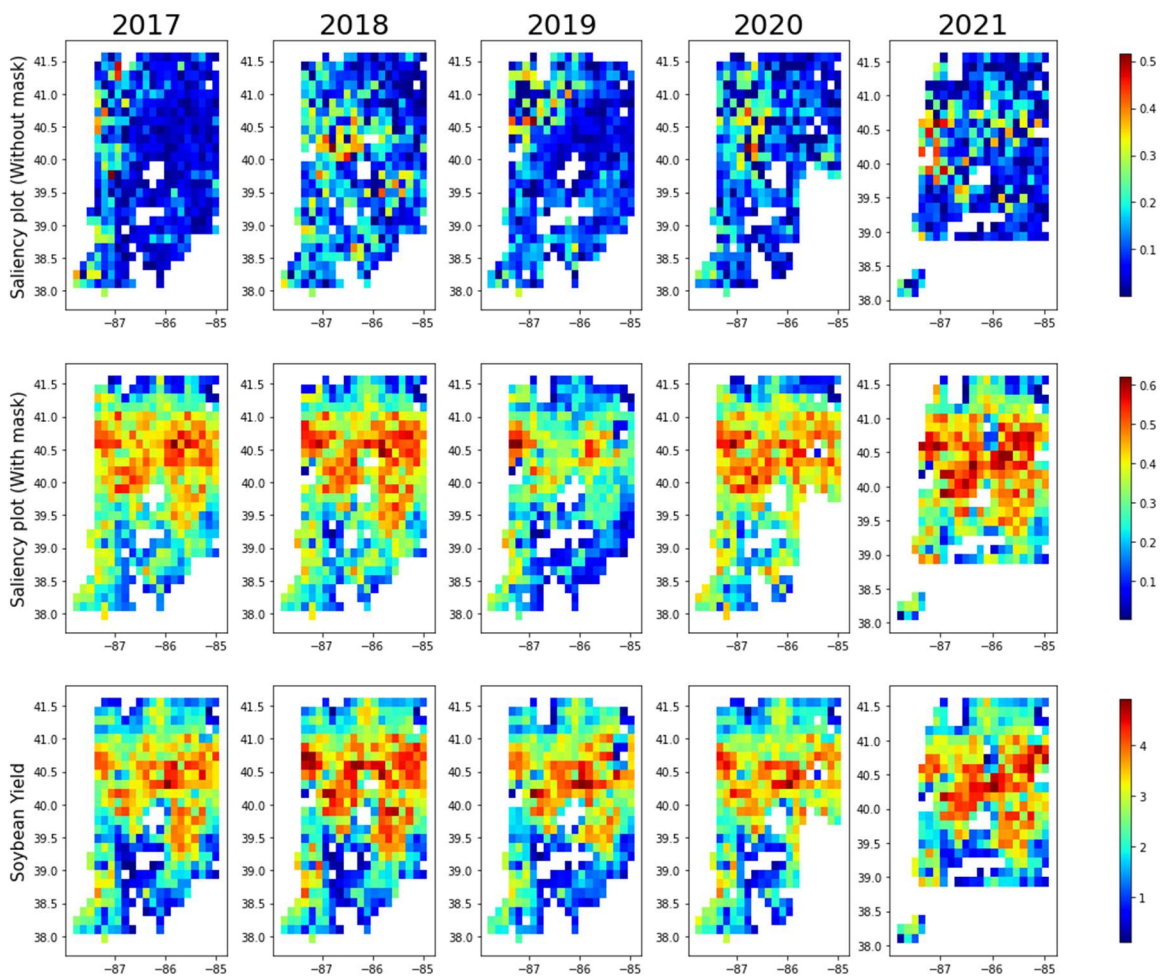


Figure 20: Map of saliency values for Indiana from 2017 to 2021. The top row is the saliency map of the model without the mask. The middle Row is the saliency map of the model with the mask layer. The bottom Row is the soybean yield values.



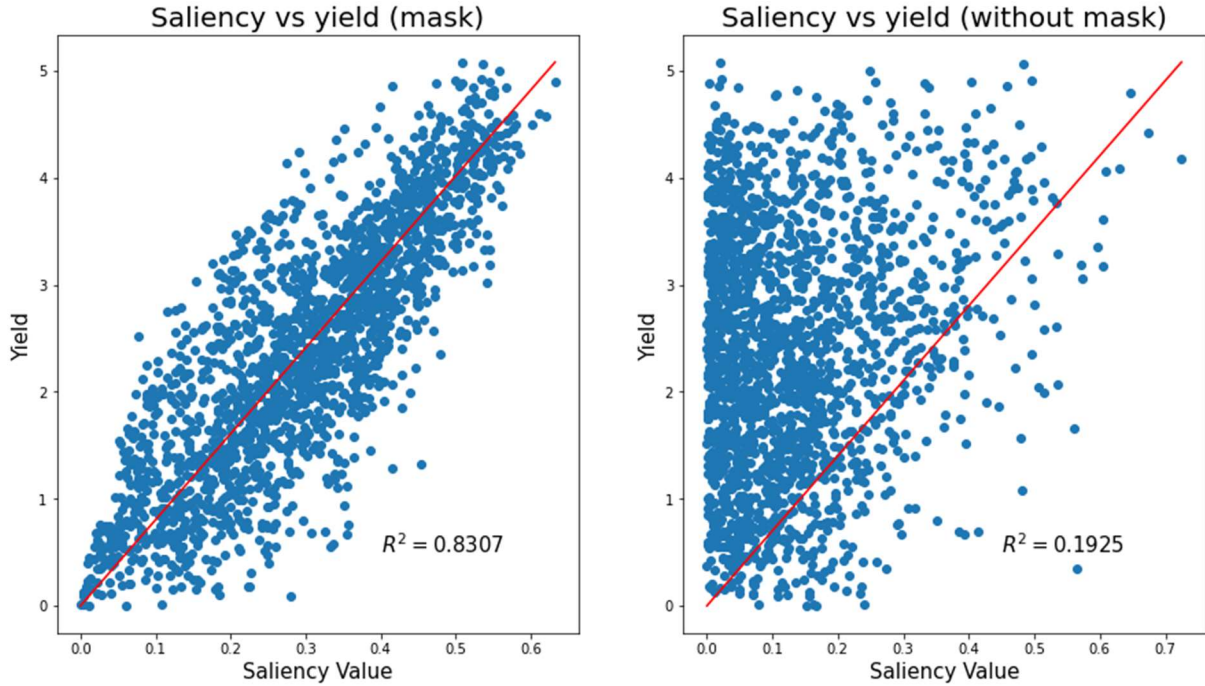


Figure 21: Scatter plot of the saliency values vs Yield values

#### 4.7. Perturbation Analysis

Perturbation analysis is performed to figure out which explainable method is better quantitatively. Figure 22 shows the saliency maps of one patch from Indiana for both the CNN models (with and without mask) and the ranking on which the perturbation takes place. We can see that the order in which the grids are masked differs for each explainable method.

Figure 23 shows the perturbation plot where the x-axis represents the number of grids (64, in this case), and the y-axis depicts the difference in accuracy between the perturbed patch and the actual value for each iteration. With each iteration, a grid is perturbed based on the ranking order, and the difference in accuracy is plotted. Some explainable methods are ignored for the perturbation process based on the heatmaps generated to fasten the process. Currently, there is no quantitative method that is used to ignore these methods. Visually, we can see that GBP (Guided Back Propagation), gradient, input\*gradient, and integrated gradients do not indicate any importance in the patches, as shown in Figure 22. Hence, these methods are not selected for the perturbation analysis. LRP (A and B), Deep Taylor, SmoothGrad and gradCAM are chosen. The lower and steeper the curve for each iteration, the better the explainable method since it indicates that the important regions are being perturbed correctly. Figure 23 indicates that smoothGrad is not a suitable explainable method for this model and dataset because of its higher curve shape. This means that the grids that are being perturbed are not significant enough for the resulting yield to drop from its true value. This implies that smoothGrad is not a good method to explain the model's decision-making when it comes to estimating the soybean yield. LRPa and LRPb have the steepest decline in accuracy, signifying that their saliency maps are better at explaining the spatial features. gradCAM and deep Taylor methods also have a steep decline in their curves. Further explanation of why LRP gives better result is provided in section 5.

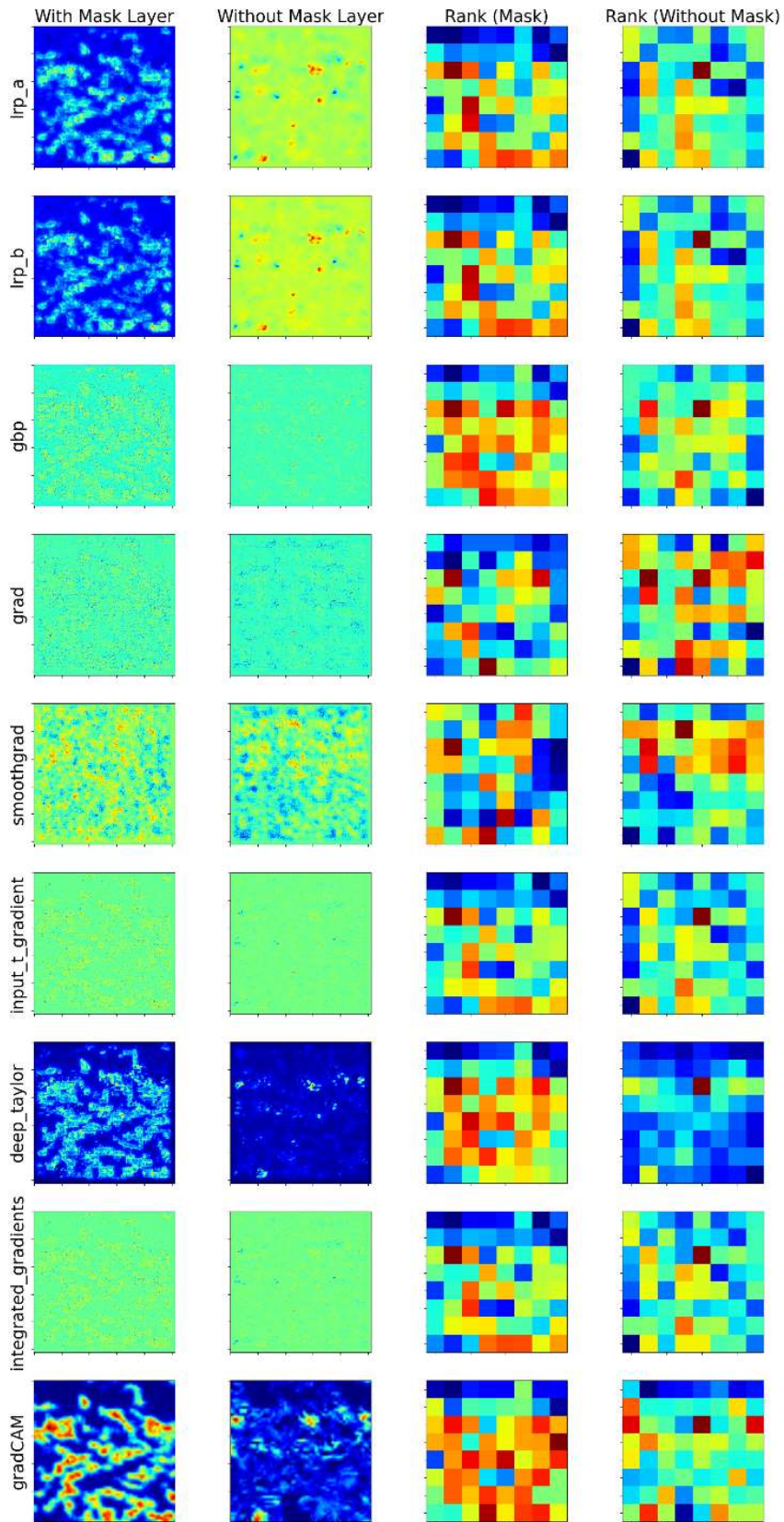


Figure 22: Perturbation Analysis based on the ranking of saliency

## Perturbation Plot

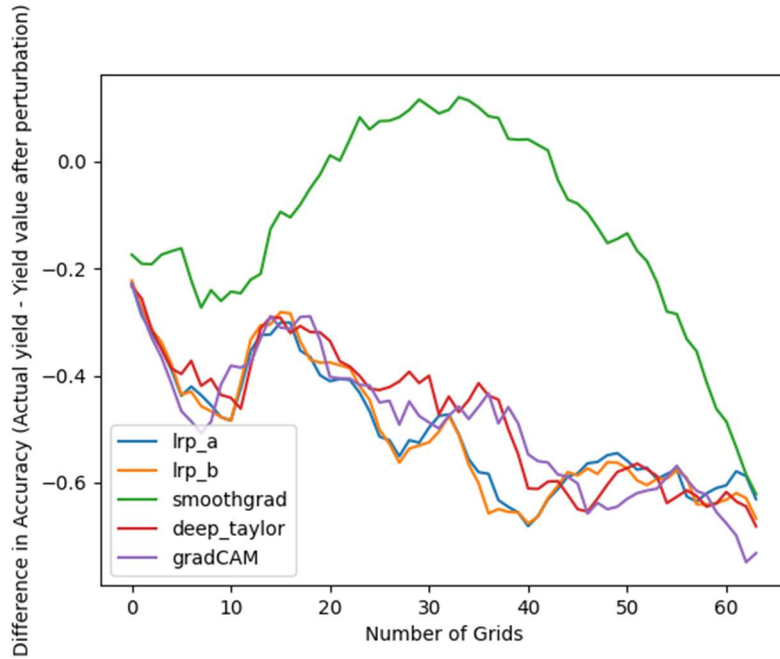


Figure 23: Perturbation plot – the difference in accuracy from the True value for each iteration

Figure 24 displays the plot of AUC (Area Under the Curve) for twenty patches. The perturbation analysis was only implemented for 20 patches due to the high computational time it took to complete. The higher the area under the curve, the lower the explainable method's quality. From the AUC plot, smoothGrad generally has a higher area than the other explainable methods. The other explainable methods have similar AUC values, implying that they provide similar levels of accurate information, with LRP being the best explainable method for these 20 patches.

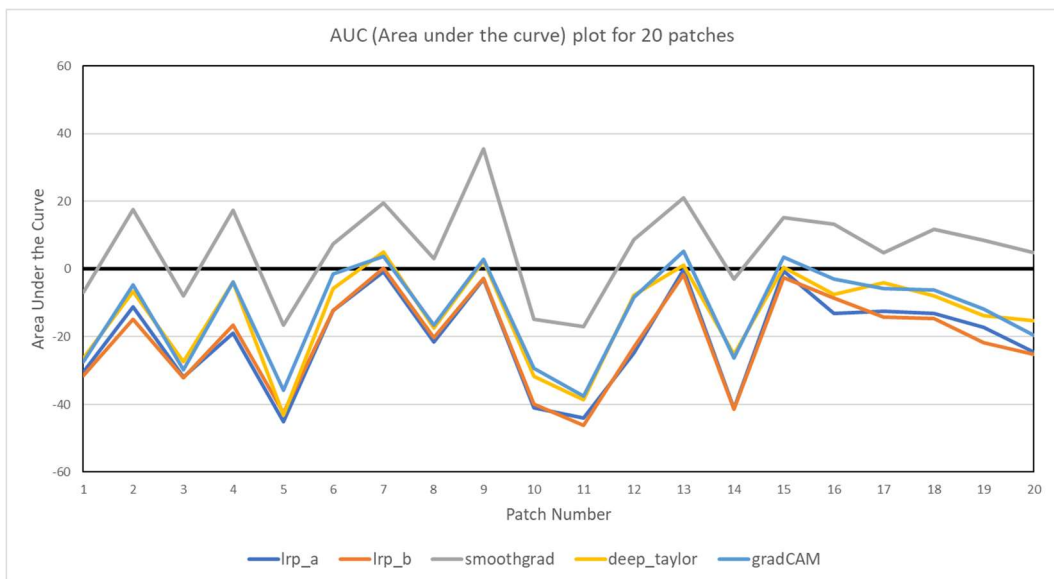


Figure 24: Area under the curve (AUC) plot for perturbation analysis

Table 9 displays the area under the curve values for each patch, highlighted as a heatmap. Red values have a lower area, and blue values have a high area. To be more specific, smoothgrad has the highest AUC for each patch, and its average is also the highest when compared to the other methods, having 6.16 AUC. LRPa and LRPb have the lowest value of -20.33 and -20.56, respectively. Both of them are relatively close, thus making them both a suitable method for these patches. We must note that the AUC value could change when the perturbation analysis is implemented for all the patches in the test dataset.

Patch Number	lrp_a	lrp_b	smoothgrad	deep_taylor	gradCAM
1	-30.395949	-31.395496	-6.863898	-26.240199	-27.341866
2	-11.189354	-14.761003	17.549556	-6.551421	-4.662912
3	-31.873941	-32.173338	-7.806655	-27.310313	-29.816895
4	-18.892769	-16.484184	17.467263	-3.695225	-3.845658
5	-45.058501	-42.523065	-16.647276	-43.069404	-35.732861
6	-12.137823	-12.164325	7.48268	-5.766113	-1.466209
7	-0.663962	0.340636	19.631708	5.090627	3.716439
8	-21.423821	-20.245256	3.056862	-17.358993	-16.596916
9	-2.829099	-2.785724	35.458661	2.520518	2.891678
10	-40.878388	-39.796446	-14.853174	-31.68489	-29.292054
11	-44.048587	-46.186179	-17.041757	-38.516637	-37.542692
12	-24.740844	-23.132259	8.830847	-7.712482	-8.444395
13	-0.176562	-1.649899	21.147656	1.113495	5.309906
14	-41.120783	-41.411846	-2.882156	-25.206237	-26.361757
15	-0.628532	-2.4589	15.266421	0.535599	3.585054
16	-13.08346	-8.621461	13.327037	-7.490797	-3.025272
17	-12.515818	-14.075568	4.939622	-3.947319	-5.825843
18	-13.177762	-14.565322	11.721566	-7.810019	-6.246242
19	-17.290297	-21.734098	8.537443	-13.75546	-11.875997
20	-24.534363	-25.286535	4.866935	-15.21665	-19.612065
Mean	-20.33303075	-20.5555134	6.15946705	-13.603596	-12.60932785

Table 9: Area under the curve values for 20 patches

#### 4.8. Analysis of saliency maps

Various vegetation indices are taken for analysing the results of the saliency maps. The vegetation indices taken includes NDVI (Normalised Difference Vegetation Index), NDMI (Normalised Difference Moisture Index), WDRVI (Wide Dynamic Range Vegetation Index), EVI (Enhanced Vegetation Index), SAVI (Soil Adjusted Vegetation Index)



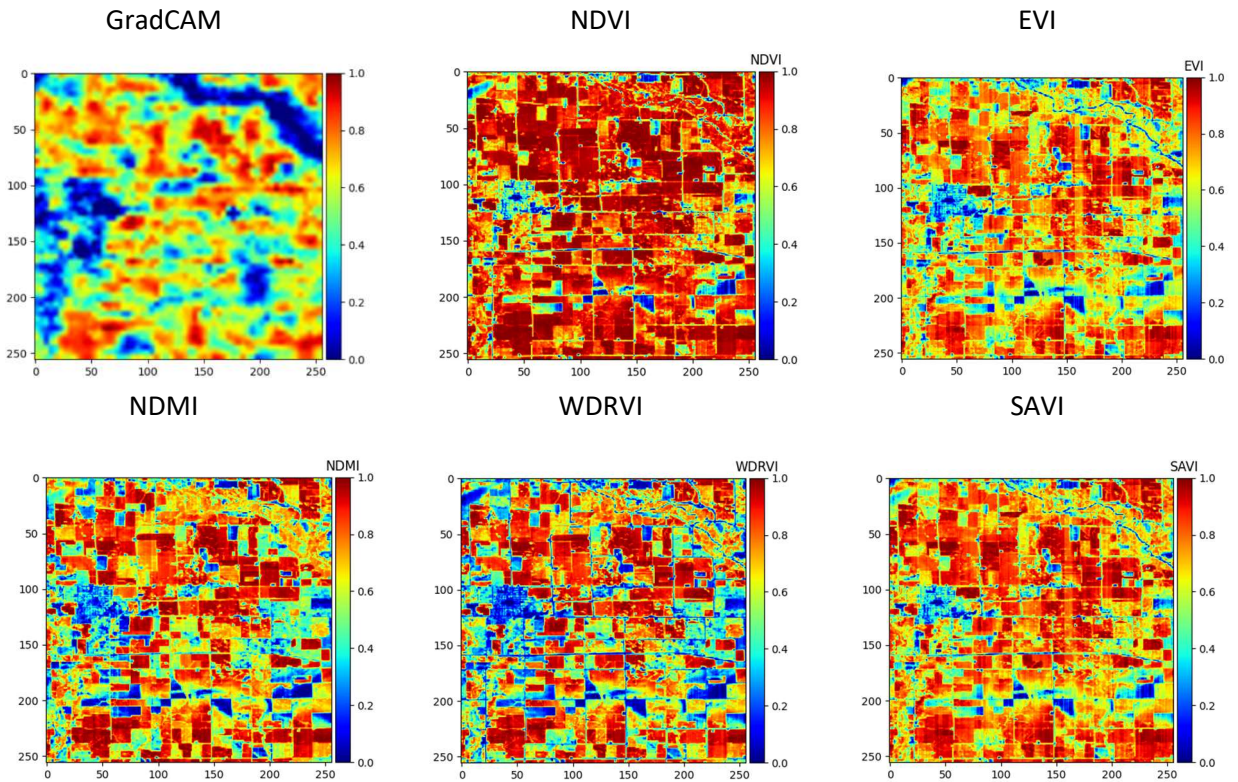


Figure 25: GradCAM and indices of a patch from CNN model without mask

Figure 25 displays the indices and the gradCAM saliency of a patch. This gradCAM is the saliency map of the CNN model with just the sentinel bands. Visually, a pattern is seen spatially where the saliency is high for regions having higher indices. NDVI and SAVI have a low correlation, while WDRVI and EVI have a higher correlation. To further analyse the results, the land use of the patch is taken, and the saliency map and the various indices are grouped according to the land use type.

Figure 26 and Figure 27 displays the area covered by the cropland and the cropland cover. Corn has higher coverage, followed by Soybean, Grassland and woody wetlands for this particular patch.

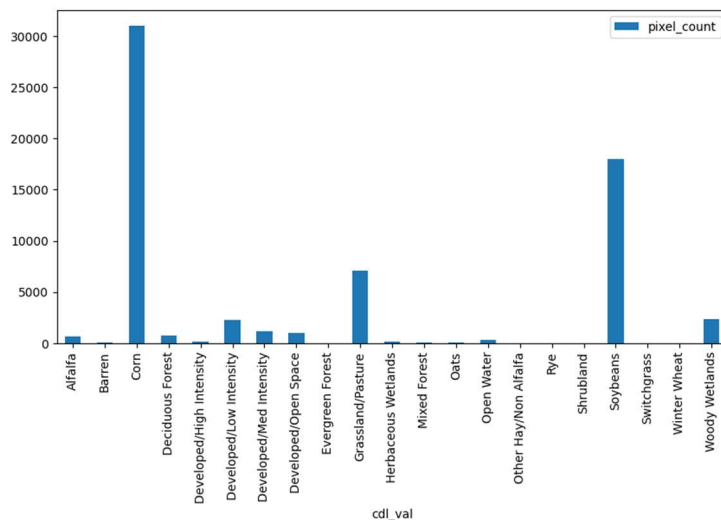


Figure 26: Area coverage by crop type for the patch

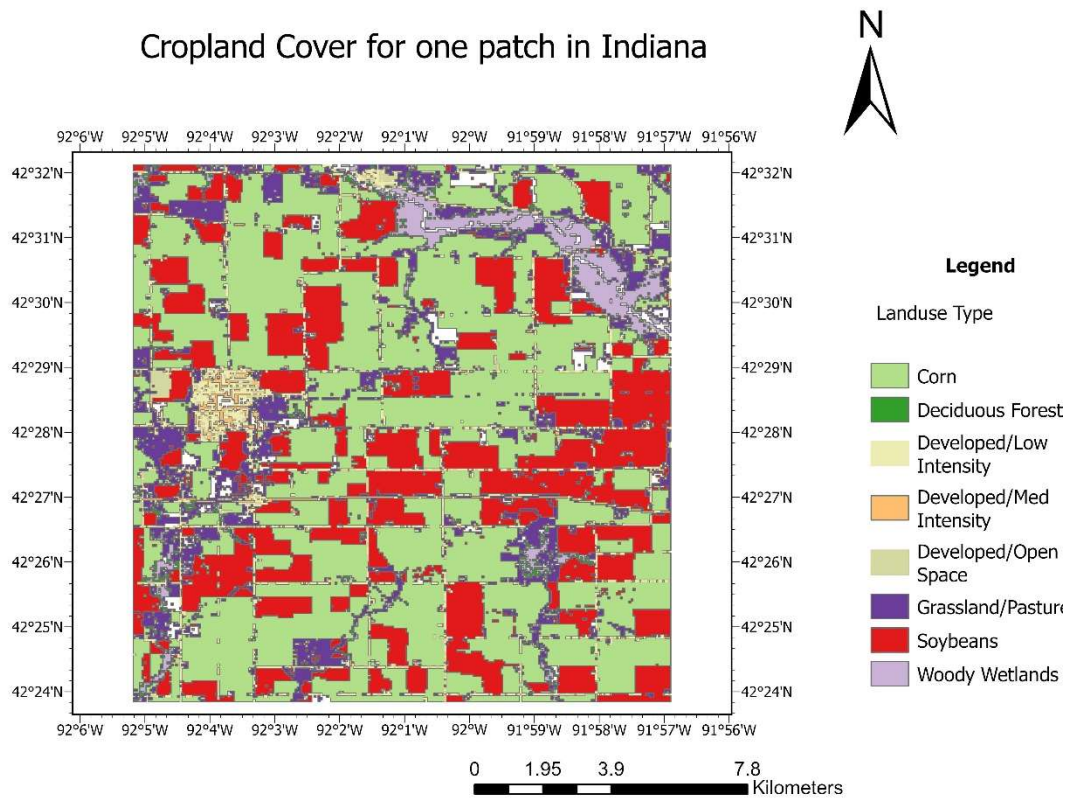


Figure 27: Crop Land cover for the patch

Figure 28 shows a correlation between area coverage, crop type and, indices & saliency level. For example, corn has the highest area covered and higher indices level for this patch. When the CNN model is trained without the mask layer, the model focuses on regions where Corn is cultivated instead of soybeans. This highlights the importance of the mask layer and provides insight into how the CNN model performs with just the sentinel bands. Even though the model might estimate the correct yield value, it cannot be implemented since it focuses on the incorrect regions. On the other hand, for the CNN model with the mask layer, the accuracy of the plot indicates that the saliency is highest for soybean, while less so for the other crop types. This is also reflected by the saliency map, where the focus is highest for soybean regions.

This analysis is performed only for one patch. It should be noted that the cropland type could differ for other patches. Hence the research is also conducted on all the patches of the study region to ensure which crop type the model focuses on.

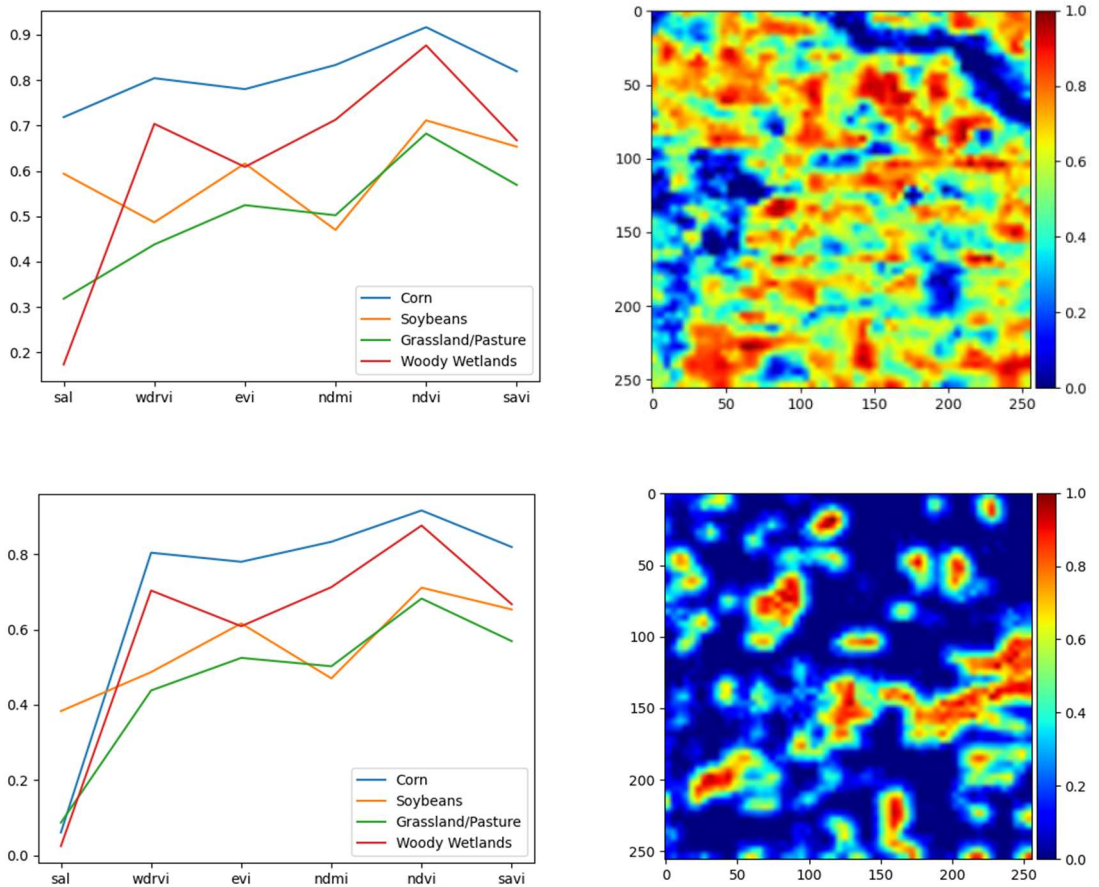


Figure 28: Line plot comparing the two models' saliency and indices. The top row shows the saliency map of the patch for the CNN model without a mask, and the bottom row is the same patch for the CNN model with a cover.

Figure 29 and Figure 30 show the boxplot of saliency values for each crop type across the test dataset for the models without and with the mask layer, respectively. From Figure 29, it is interesting that the boxplots indicate that Soybean has a higher range of values. However, the mean is higher for sweet corn, sunflower and developed/ Open Space, showing that the model gives these regions more significance. Figure 30 tells a different story, where Soybean has the highest mean, showing that the mask layer influences the model's decision. However, next to Soybeans, Sugarbeets and Developed/Open spaces have a significant level of importance. This occurs because the saliency maps are not a binary representation of the cropland, where only soybean is grown. They also consider other regions important, depending on the indices level and if these lands are close to soybean fields.

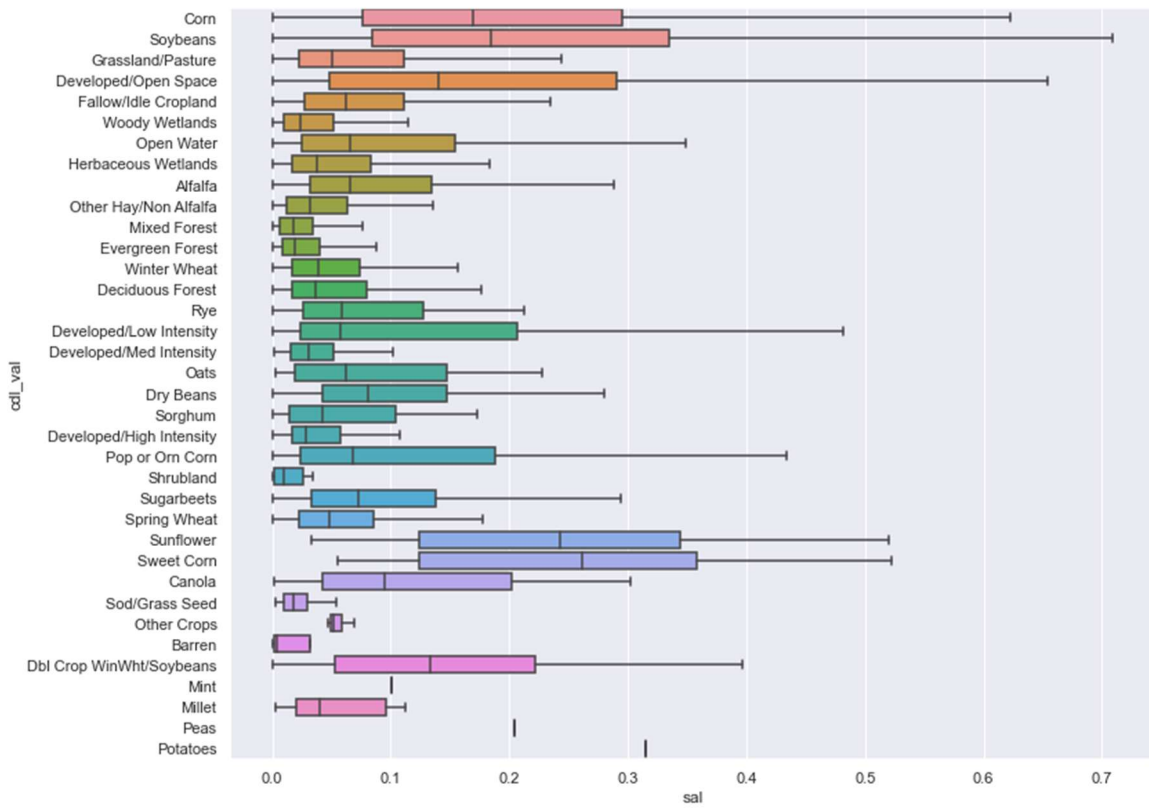


Figure 29: Boxplot of saliency values w.r.t crop type for CNN model without mask

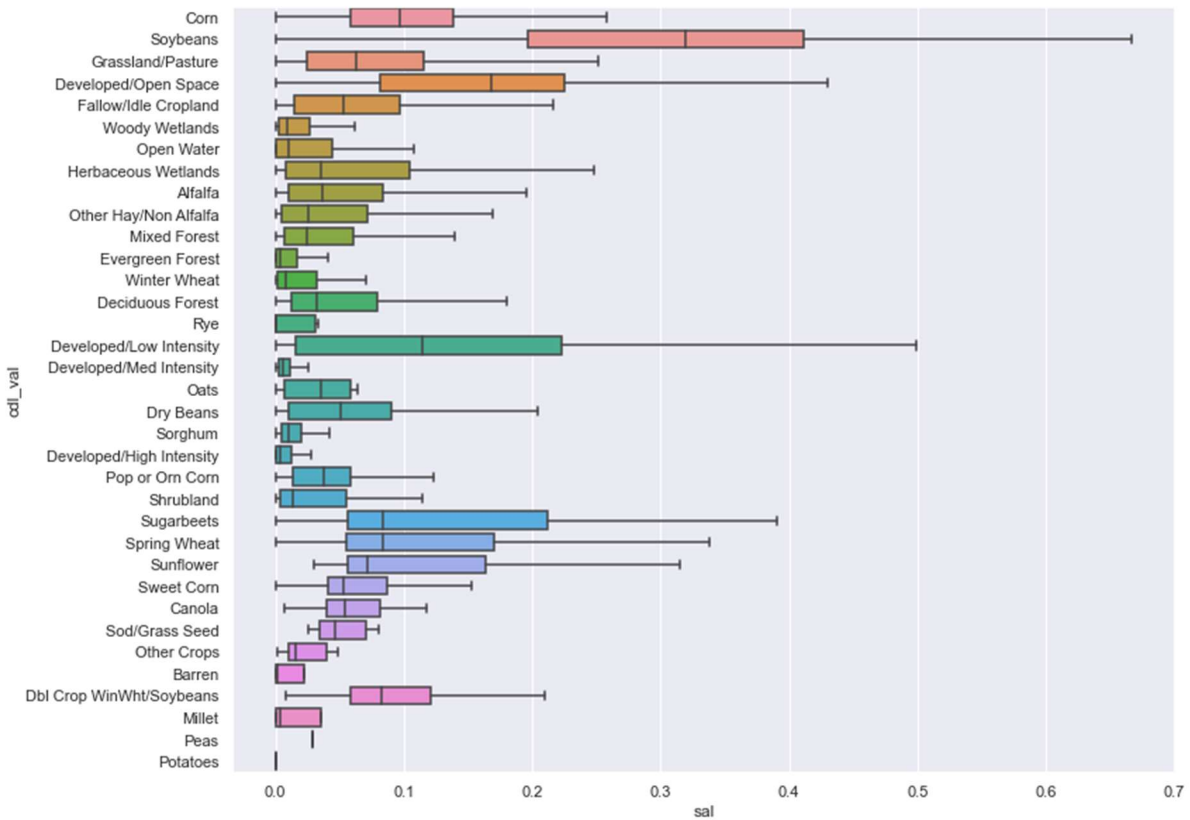


Figure 30: Boxplot of saliency values w.r.t crop type for CNN model with mask



## 5. DISCUSSION

From analysing the results, we could safely infer a bias in the model when the mask layer is not added to the input dataset. Hence, in a scenario where the input data is without the mask layer, the CNN model might perform better and have good accuracy but incorrectly point to a wrong region. This is similar to other research that has explored the bias in CNN models using saliency maps. Ribeiro et al. (2016) provide an example in which a logistic regression classifier detects the image of a dog (husky) incorrectly as a wolf. This was purposefully done by training the classifier with images of wolves with snow in the background and dogs without snow. While testing the model, an image of a dog with snow in the background was provided, and the model classified it as a wolf (Ribeiro et al., 2016). Looking at the explanation for that image, it seems the model focused on the snow for classification rather than the dog’s features. Another example in the medical industry was discovered by (Zech et al., 2018), where CNN models were used to detect pneumonia through X-ray images. However, more cases of pneumonia were detected from images scanned using portable X-ray scanners (Rudin, 2019b; Zech et al., 2018). Upon interpreting these images, it was found that the model was focusing on the text ‘portable’ rather than the X-ray scan of the chest. Such cases of bias occurring in deep learning models would be difficult to point out without explainable methods. We can see a relation to our research where the mask layer is vital in correctly estimating the yield. Previous studies conducted in crop yield estimation have rarely focussed on explainability. Sun et al. (2020) uses MODIS satellite imagery along with soil characteristics to estimate the corn yield by implementing a multi-level deep learning model but do not assess the interpretability of the model. They utilise the cropland layer to mask the pixels that do not belong to corn fields. However, the spatial information is not considered as the data is transformed into histogram-based tensors (Sun et al., 2020). Though they achieve an  $R^2$  score of 0.75, we cannot draw more parallels to our results apart from the fact that utilising the mask layer is essential. Srivastava et al. (2022) and Wolanin et al. (2020a) both utilise a 1D CNN model to forecast wheat yield. Srivastava et al. (2022) acquire point samples, while Wolanin et al. (2020a) reshape MODIS imagery along the temporal axis. Nevertheless, we can gather that both studies include data about the regions where the crops are cultivated. Also, using the sentinel-2 dataset to estimate soybean yield has not been previously explored. This could be due to the lack of yield data smaller than county-level. Further explanation is provided in Section 5.1. It should also be pointed out that running a perturbation analysis helps identify which explainable method is most reliable. We can identify whether the method is meaningful qualitatively through just visualisation. However, in some instances, if we want to perform further experimentations using saliency maps, it is helpful to have a quantitative value. Zhang et al. (2020) perform a perturbation analysis on explainable methods for CNN to identify the best method for estimating fetal head circumference. Their findings showed that LRP and Input\*gradient were highly sensitive to perturbation and had a higher error rate (Zhang et al., 2020). Similarly, Kakogeorgiou & Karantzalos (2021) assessed several explainable methods for multi-label classification in Earth Observation. They also compared the saliency map methods by doing a sensitivity analysis and AUC-MoRF (Area Under the Curve Most Relevant First), which is also a form of perturbation analysis (Kakogeorgiou & Karantzalos, 2021). They discovered that Occlusion, LIME (Local Interpretable Model-agnostic Explanations) and gradCAM were the most suitable explainable methods (Kakogeorgiou & Karantzalos, 2021).

In our case, gradCAM performs reasonably well, while LRP gives the best results. However, Kakogeorgiou & Karantzalos (2021) were running the model for classification while our task is to train it as a regression model. The reason why LRP has better results might be due to the process in which each

layer in the model has relevance. This is different from the other explainable methods like Deep Taylor and gradCAM. Though Deep Taylor provides an accumulated score of a pixel by propagating through the layers, LRP retains the importance level provided by each layer, and this plays a role in getting the final score for the pixel. On the other hand, gradCAM takes the gradient of the weighted average of the feature maps, which makes it more suitable for visual interpretation. This is also different from LRP, where each individual pixel is retained and does not lose its value by calculating the weighted average. SmoothGrad perturbs the input image multiple times and takes the average of the gradients from the perturbed images. This is also completely different from LRP, where random noise is not added to test the model's decision making process.

There are not many cases in which 2D-CNN models have been used for regression-based estimation or prediction (Letzgus et al., 2022). In a real-world application, Letzgus et al. (2022) demonstrate how a CNN model is used as a regression model to detect a person's age from a facial photograph. Saliency maps are generated using LRP (Layerwise Relevance Propagation), and breakpoints are set for the age to test the relation in which saliency changes when the age deviates from the true value. The results are intriguingly similar to what we got regarding yield estimation. For instance, if the image was of a child around age 10 but the model was forced to provide a saliency map for age 80, the heatmap highlighted indiscernible features. This is equivalent to how we get the saliency maps when running the model without a mask layer. Interestingly, when the age gap gets closer, where the image had a person of age 40, and the model was set to age 50, the saliency maps gave distinguishable features, highlighting the eyes, nose, mouth, etc. (Letzgus et al., 2022). Since our research offers a regression model, we also need to consider if there is a correlation between the saliency level and the target value. Though we did not test it by setting breakpoints, an overview of the correlation was checked. Figure 20 and Figure 21 show us that when the mask layer is included, the saliency values are closely correlated with the target yield. Otherwise, it is not possible to form a valid correlation.

## 5.1. Limitations

The target variable for each patch is feature engineered from existing yield at the county level. The area of the cropland covered by soybean in each patch is used to distribute the yield value spatially. The main limitation in preparing the dataset this way is that there are no means to confirm the productivity of the croplands in that patch. Hence it could vary for each patch, even though the weighted mean of the yield would be the same at the county level. To overcome this limitation, we might have to resort to using a low resolution satellite imagery, where each patch must cover multiple counties. MODIS (Moderate Resolution Imaging Spectroradiometer), having a resolution of 500 meters could be an option. A new dataset can be prepared, where the target yield need not be downscaled. However, the dataset for the selected study region will be lower if only five states are selected. An approach to compensate for this would be to choose a larger extent for the study area to prepare a dataset sufficient for training the model.

## 6. CONCLUSIONS AND RECOMMENDATIONS

Initially, a linear regression model was fitted to the data. The bands of each patch were aggregated and fitted for a linear regression model. It had an accuracy of 60% when tested with Indiana. However, more information was required other than the linear model. From the weights of the bands, bands 7,9, 5 and 4 are given higher importance, whereas bands 3,6 and 10 are given minor importance. Further explanation of how the model interprets the results could not be provided spatially. Further analysis was performed using saliency maps generated by running the CNN models. Two CNN models were trained with the same hyperparameters but different datasets. The mask layer of soybean classified regions was provided along with the sentinel bands for one model, whereas only the sentinel bands were given to the second model. Upon evaluation, the model with the mask layer demonstrated an excellent accuracy of 98%, while the model with just the sentinel bands had 70% accuracy. This implied a perfect correlation with the features in the mask layer. Those pixels were the regions where soybean was being cultivated. This was further supported by analysing the saliency maps of the patches. Higher importance was placed on the mask region, which indicates that the CNN model utilised the classified pixels of soybean to estimate the yield effectively. Different explainable methods were then compared by performing a perturbation analysis. LRP, Deep Taylor and gradCAM were identified as the better explainable methods, with LRP giving the lowest AUC score. Further analysis of the explainable methods was performed by comparing them with cropland cover maps. This indicated a bias in the CNN model that trained with only the sentinel-2 bands, focusing on corn and other crop types rather than soybeans. Further analysis may be required, but this showed the value of explainability in Earth Observation.

### 6.1. Answers to the research questions

- **What is the level of accuracy for the CNN model compared to the linear regression model?**

The linear regression model has a lower accuracy than both CNN models. For this research, the CNN model without the mask layer had a 12% higher accuracy than the linear regression model. In contrast, the CNN model, including the mask layer, had a 37% higher accuracy (Refer to Table 8). This indicates that the CNN model performs better since it detects the spatial features from the input data.

- **Which characteristics are essential for estimating crop yield?**

The mask layer representing the Soybean fields highly influences the model's capability to estimate the yield, giving an almost perfect accuracy of 98%. We also find that without the mask layer, the model focuses on the crops having higher vegetation indices. This implies that a combination of the mask layer and the bands are essential to compute vegetation indices like WDRVI, EVI, NDMI, NDVI and SAVI. These include Bands 2, 4, 8 and 11 (Refer to Section 3.6).

- **What differences can be observed between the explainable methods regarding their performance and accuracy?**

Every explainable method gives a different heatmap/saliency map. As discussed in Section 4.7, we can identify which method performs better through perturbation analysis. In Figure 23, a steeper curve indicates better performance since important information is lost, resulting in a dip in accuracy. Also, the explainability for the CNN model with the mask layer is generally better as it mainly highlights the soybean fields. In contrast, the model with only the sentinel-2 bands has less clarity in its saliency maps (Refer to Figure 19).

- **Which explainable methods are ideal for crop yield estimation?**

It is noticed that explainable methods like LRP, Deep Taylor and gradCAM provide better results while others like smoothGrad and the Gradient methods are less suitable for crop yield estimation (Refer to Section 4.7).

- **How is explainability valuable in the case of crop yield estimation?**

Explainability provides insight into how the model has arrived at the estimated yield. In this research, the explanation is done via saliency maps. Landuse maps are compared with saliency maps to assess the relationship with crop yield. When the model is trained without the mask layer, a bias is seen through the saliency maps where corn fields are given higher importance (Refer to Figure 28). This emphasises the need for explainability to analyse the model's performance and bias in Earth Observation, not just crop yield estimation.

## 6.2. Recommendations

Utilising explainable methods provided insight that the mask layer is also essential. This can be further extended into improving the model to train for the classification of soybean fields initially and then estimating their yield. This will lead to development of a multi-model CNN that performs both classification and regression.

Implementing a 3D CNN model to predict and forecast the yield will also be helpful. The current model is only explored spatially. 3D CNN would also pave the road for temporal analysis using XAI. The existing CNN model is built from scratch. It could be extended and compared with existing CNN model architectures like Inceptionv3 and ResNet for a better assessment.

Also, enhancing the model to be transferable to another study region and datasets could be explored. This will provide a flexible model that could be implemented for any region with different datasets.

Vision Transformers has also gained popularity recently, and extensive research is being conducted in Computer Vision. Explainability using self-attention maps could provide additional insights into the model's behaviour and relation between the datasets that could be worth exploring.

Finding correlations with other datasets like weather variables and soil moisture would also provide new insights into the interpretation and better yield estimation. These could be provided as additional datasets when developing the model. It could be compared with statistical methods like multi-scale GWR (Geographical Weighted Regression) as a baseline since several parameters will influence the yield estimation.

## 7. REFERENCES

- Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., Samek, W., Müller, K.-R., Dähne, S., Kindermans, P.-J., & Müller, K.-R. (2019). iNNvestigate Neural Networks! In *Journal of Machine Learning Research* (Vol. 20). <http://jmlr.org/papers/v20/18-540.html>.
- Antony, B. (2021). Prediction of the production of crops with respect to rainfall. *Environmental Research*, 202, 111624. <https://doi.org/10.1016/J.ENVRES.2021.111624>
- Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS ONE*, 10(7), e0130140. <https://doi.org/10.1371/journal.pone.0130140>
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (2010). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), 281–298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
- Casas-Roma, J., & Conesa, J. (2021). A literature review on artificial intelligence and ethics in online learning. *Intelligent Systems and Learning Data Analytics in Online Education*, 111–131. <https://doi.org/10.1016/B978-0-12-823410-5.00006-1>
- Doshi-Velez, F., & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. <https://doi.org/https://doi.org/10.48550/arXiv.1606.05386>
- Glorot, X., Bengio, Y., Bordes, A., & Bengio, Y. (2011). *Deep Sparse Rectifier Neural Networks Oracle Performance for Visual Captioning View project Parsing View project Deep Sparse Rectifier Neural Networks*. <https://www.researchgate.net/publication/215616967>
- Imran, M., Stein, A., & Zurita-Milla, R. (2015). Using geographically weighted regression kriging for crop yield mapping in West Africa. *International Journal of Geographical Information Science*, 29(2), 234–257. <https://doi.org/10.1080/13658816.2014.959522>
- Johnson, N., Santosh Kumar, M. B., & Dhannia, T. (2021). A survey on Deep Learning Architectures for effective Crop Data Analytics. *2021 International Conference on Advances in Computing and Communications (ICACC)*, 1–10. <https://doi.org/10.1109/ICACC-202152719.2021.9708193>
- Kakogeorgiou, I., & Karantzalos, K. (2021). Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102520. <https://doi.org/10.1016/J.JAG.2021.102520>
- Khaki, S., & Wang, L. (2019). Crop Yield Prediction Using Deep Neural Networks. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.00621>
- Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. <http://arxiv.org/abs/1412.6980>
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- Letzgus, S., Wagner, P., Lederer, J., Samek, W., Müller, K.-R., & Montavon, G. (2022). Toward Explainable Artificial Intelligence for Regression Models: A methodological perspective. *IEEE Signal Processing Magazine*, 39(4), 40–58. <https://doi.org/10.1109/MSP.2022.3153277>
- Leung, Y., Mei, C.-L., & Zhang, W.-X. (2000). Testing for Spatial Autocorrelation among the Residuals of the Geographically Weighted Regression. *Environment and Planning A: Economy and Space*, 32(5), 871–890. <https://doi.org/10.1068/a32117>

- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. <https://christophm.github.io/interpretable-ml-book/>
- Montavon, G., Binder, A., Lapuschkin, S., Samek, W., & Müller, K.-R. (2019). *Layer-Wise Relevance Propagation: An Overview* (pp. 193–209). [https://doi.org/10.1007/978-3-030-28954-6\\_10](https://doi.org/10.1007/978-3-030-28954-6_10)
- Montavon, G., Lapuschkin, S., Binder, A., Samek, W., & Müller, K.-R. (2017). Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognition*, 65, 211–222. <https://doi.org/10.1016/j.patcog.2016.11.008>
- Montavon, G., Samek, W., & Müller, K.-R. (2017). *Methods for Interpreting and Understanding Deep Neural Networks*. <https://doi.org/10.1016/j.dsp.2017.10.011>
- Nevavuori, P., Narra, N., & Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. *Computers and Electronics in Agriculture*, 163, 104859. <https://doi.org/10.1016/J.COMPAG.2019.104859>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Cournapeau, D., Brucher, M., & Perrot, M. (2011). Scikit-learn: Machine Learning in Python Pedregosa, Varoquaux, Gramfort et al. In *Journal of Machine Learning Research* (Vol. 12). <http://scikit-learn.org>.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Rudin, C. (2019a). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Rudin, C. (2019b). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Sakamoto, T. (2020). Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 208–228. <https://doi.org/10.1016/J.ISPRSJPRS.2019.12.012>
- Schwalbert, R. A., Amado, T., Corassa, G., Pott, L. P., Prasad, P. V. V., & Ciampitti, I. A. (2020). Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*, 284, 107886. <https://doi.org/10.1016/J.AGRFORMET.2019.107886>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2016). *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. <https://doi.org/10.1007/s11263-019-01228-7>
- Smilkov, D., Thorat, N., Kim, B., Viégas, F., & Wattenberg, M. (2017). *SmoothGrad: removing noise by adding noise*. <http://arxiv.org/abs/1706.03825>
- Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). *Striving for Simplicity: The All Convolutional Net*. <http://arxiv.org/abs/1412.6806>
- Srivastava, A. K., Safaei, N., Khaki, S., Lopez, G., Zeng, W., Ewert, F., Gaiser, T., & Rahimi, J. (2022). Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. *Scientific Reports*, 12(1), 3215. <https://doi.org/10.1038/s41598-022-06249-w>

- Stomberg, T. T., Stone, T., Leonhardt, J., Weber, I., & Roscher, R. (2022). *Exploring Wilderness Characteristics Using Explainable Machine Learning in Satellite Imagery*. <http://arxiv.org/abs/2203.00379>
- Stomberg, T., Weber, I., Schmitt, M., & Roscher, R. (2021). JUNGLE-NET: USING EXPLAINABLE MACHINE LEARNING TO GAIN NEW INSIGHTS INTO THE APPEARANCE OF WILDERNESS IN SATELLITE IMAGERY. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *V-3-2021*, 317–324. <https://doi.org/10.5194/isprs-annals-V-3-2021-317-2021>
- Sun, J., Lai, Z., Di, L., Sun, Z., Tao, J., & Shen, Y. (2020). Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Corn Belt. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 5048–5060. <https://doi.org/10.1109/JSTARS.2020.3019046>
- Sundararajan, M., Taly, A., & Yan, Q. (2017). *Axiomatic Attribution for Deep Networks*. <http://arxiv.org/abs/1703.01365>
- Terliksiz, A. S., & Altıylar, D. T. (2019). Use Of Deep Neural Networks For Crop Yield Prediction: A Case Study Of Soybean Yield in Lauderdale County, Alabama, USA. *2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 1–4. <https://doi.org/10.1109/Agro-Geoinformatics.2019.8820257>
- USDA, & National Agricultural Statistics Service. (2022). *Crop Production 2021 Summary*. <https://usda.library.cornell.edu/concern/publications/k3569432s>
- van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709. <https://doi.org/10.1016/J.COMPAG.2020.105709>
- Wang, A. X., Tran, C., Desai, N., Lobell, D., & Ermon, S. (2018). Deep Transfer Learning for Crop Yield Prediction with Remote Sensing Data. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, 1–5. <https://doi.org/10.1145/3209811.3212707>
- Wolanin, A., Mateo-García, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liangzhi, Y., & Guanter, L. (2020a). Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environmental Research Letters*, *15*(2), 024019. <https://doi.org/10.1088/1748-9326/ab68ac>
- Wolanin, A., Mateo-García, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liangzhi, Y., & Guanter, L. (2020b). Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. *Environmental Research Letters*, *15*(2), 024019. <https://doi.org/10.1088/1748-9326/ab68ac>
- Yang, B., Wu, S., & Yan, Z. (2022). *Geo-Information Effects of Climate Change on Corn Yields: Spatiotemporal Evidence from Geographically and Temporally Weighted Regression Model*. <https://doi.org/10.3390/ijgi11080433>
- Yang, Q., Shi, L., Han, J., Zha, Y., & Zhu, P. (2019). Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Research*, *235*, 142–153. <https://doi.org/10.1016/J.FCR.2019.02.022>
- Yeh, C.-K., Hsieh, C.-Y., Suggala, A. S., Inouye, D. I., & Ravikumar, P. (2019). *On the (In)fidelity and Sensitivity for Explanations*. <http://arxiv.org/abs/1901.09392>
- Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLOS Medicine*, *15*(11), e1002683. <https://doi.org/10.1371/journal.pmed.1002683>

Zhang, J., Petitjean, C., Yger, F., & Ainouz, S. (2020). *Explainability for Regression CNN in Fetal Head Circumference Estimation from Ultrasound Images* (pp. 73–82). [https://doi.org/10.1007/978-3-030-61166-8\\_8](https://doi.org/10.1007/978-3-030-61166-8_8)