Application of Explainable Machine Learning Methods in Predicting Crop Yield Using Multi-modality Datasets

A Dissertation

Submitted To the School of Information Technology

Of Murdoch University

For The Degree Of

Master of Information Technology

(Artificial Intelligence and Data Science)

November 2024

Jigme Dorji
34647859

Table of Contents

List of Figures

1. Introduction

The introduction of advanced machine learning algorithms to improve farming practices and maximize crop output projections has spurred a transformational wave at the convergence of agriculture and technology in recent years [1]. Explainable Artificial Intelligence (XAI) is a rapidly growing discipline that aims to understand the complexities of prediction models and provide insights into the underlying variables impacting agricultural outcomes [2].

By providing insights into the decision-making process of the Artificial Intelligence (AI) systems, XAI overcomes the challenge of understanding how complex it works, thereby building stakeholders' trust and confidence in the AI systems. This understanding is critical in industries where decisions made by AI can have a significant impact, like healthcare, finance, and agriculture. In agriculture, XAI can help farmers understand and comprehend the rationale against their expertise and contextual knowledge by assisting them in understanding why a specific crop management decision is recommended [3].

In crop genomics, XAI improves the effectiveness by increasing the transparency of predictive algorithms, which helps geneticists comprehend the specific genetic markers that impact qualities such as yield and disease resistance. Furthermore, XAI can detect and address model biases, which enhances the precision of genetic predictions [4]. For researchers, XAI is valuable given its ability to validate and elucidate complex predictive models, which facilitates the dissemination of findings to stakeholders and the scientific community by emphasizing the advantages and drawbacks of models [5]. Likewise, farmers benefit from XAI by acquiring understandable insights into crop management practices. Transparent AI recommendations build confidence and promote acceptance, resulting in enhanced utilization of resources and increased productivity. It assists farmers in promptly identifying issues by providing explanations for data abnormalities, therefore allowing for appropriate interventions [6].

Growing recognition of AI's potential to transform conventional farming methods has led to a noticeable change in agricultural research toward using machine learning algorithms to predict and analyze crop yields [7]. This trend emphasizes the importance of accurate predictive models that can estimate crop productivity and offer interpretable explanations for such projections. Adopting XAI approaches is a crucial step toward accomplishing this, which prioritizes interpretability and accuracy and provides insightful information about the complex dynamics of the Deep Learning (DL) model [8].

This thesis surveys recent advancements in XAI applied to agriculture, drawing upon diverse studies and methodologies. To understand the complex interactions between multimodal data and factors influencing crop yield, it explores how Deep Learning (DL) methods can decipher the relationship between this multimodal dataset. Through a comprehensive review of the literature, the study seeks to clarify XAI's important role in enhancing the understanding of agricultural predictions and enabling farmers to make well-informed decisions.

This thesis also explores the methodological frameworks used to concatenate phenotype, genetic and environmental data to predictive crop yield using DL models. By leveraging Convolutional

Neural Networks (CNNs), the thesis sought to capture the spatial dependencies inherent in genomic and phenotype and Long Short Term Memory (LSTM) to capture temporal dependencies in environmental data, thereby enabling more reliable and robust predictions. The thesis also examines how Attention-based CNN and LSTM try to clarify how individual features contribute to model predictions, improving the interpretability and trustworthiness of predictive models in an agricultural context.

This thesis hopes to provide an overview of XAI's applications in agriculture by synthesizing recent research and methodologies, highlighting the technology's potential to transform agriculture predictions, which may help farmers make better transparent decisions. The application of the Attention-based CNN and LSTM hybrid model seeks to use attention as a premise to interpret the complex internal workings of the hybrid model and explain the decision-making of the model.

This thesis is structured into three key sections to provide the overall flow and structure of the report. The first section provides a detailed literature review of XAI's core concepts and techniques and their application in agriculture and genomic data using Deep Learning. The second section highlights the methodology, predictions model, results and explanation results, followed by project significance and milestones. Finally, the last section summarizes the report with a conclusion.

2. Literature Review

2.1. Brief Overview of Crop Yield Prediction

Historically, as civilization relied heavily on agriculture for survival and economic stability, forecasting crop production for proper management of food supplies became a critical activity, given the inherent uncertainty of agricultural output due to weather conditions, diseases, and pests [9]. To overcome these challenges, interest in crop prediction was documented, which traced back to Mesopotamia, Egypt, and China, where these societies developed rudimentary methods of crop prediction based on weather patterns and environmental indicators [10]. Records show that Egyptians monitored the annual flooding of the Nile River to predict soil fertility and crop yield for the coming season [11]. The motive behind the early efforts of crop prediction was mainly for food security and economic prosperity, as it played an important role in socioeconomic stability and political governance as rulers relied on crop prediction to make decisions on taxation, allocation of resources, and famine relief efforts. As the civilization progressed, so did the prediction methods.

Before the emergence of AI, the prediction of crop output mainly depended on conventional statistical approaches and empirical models. Commonly employed methodologies, such as linear regression, time series analysis, and multivariate analysis, were utilized to discover correlations between crop yields and influential elements such as weather conditions, soil qualities, and agricultural practices [12]. Although these models were helpful, they typically did not have the necessary intricacy to consider the multiple variables and their interconnections that influence agricultural production.

The invention of newer technologies and better data collection techniques and tools has led to the development of more advanced methods, improving the precision and dependability of crop yield

forecasts. Initially, AI techniques relied on fundamental machine learning algorithms, including decision trees, support vector machines, and ensemble methods like random forests. As AI technology advances, DL algorithms [13] have led to increased prominence in predicting crop yield. These models can extract intricate patterns from datasets that encompass several modes of information, efficiently capturing the links and interactions among genotype, phenotype, and environmental factors.

As crop productivity depends on multiple factors, predicting outcomes necessitates data from multiple sources. This multisource data consists of information from various internal and external sources that impact crops. By combining and analyzing these multisource data, farmers, policymakers, governments, and businesses can develop prediction models enabling stakeholders to increase crop productivity. The multimodality data used for prediction models range from weather data, soil data, satellite images, farm management data, genotype data, phenotype data, and historical records. According to the survey done by Van Klompenburg et al. the most commonly used features for crop yield prediction were temperature, rainfall, and soil information [6].

2.2. Applications of Deep Learning in Genomics Study

Montesinos-López et al. illustrate that DL outperforms many traditional methods in prediction tasks, particularly in genome-based prediction [14]. DL's ability to capture complex patterns, support raw and multisource data, handle large and intricate datasets, and offer flexibility in network architecture to build a new model using the core architectural elements of DL models are cited as key reasons for its adoption in this context. Furthermore, Van Klompenburg et al. surveyed 50 research papers on crop yield prediction, revealing that CNN was the most frequently applied DL algorithm, followed by Long Short Term Memory (LSTM) and Deep Neural Networks (DNN) [6].

Wolanin et al. underscored the importance of DL models and interpretability in crop yield estimation [15]. Deploying CNN, they demonstrated superior performance in predicting wheat yields in the Indian Belt. Their model surpassed opaque models like Ridge Regression and Random Forest, offering explainable features related to seasonal length, temperature, and lighting conditions. Additionally, Regression Activation Mapping (RAM) was utilized as an XAI technique to provide insights into the nonlinear interactions within the input data and their relation to crop growth progression [16]. These findings highlight the potential of DL in genomic studies and agricultural applications and emphasize the importance of interpretability in leveraging its benefits effectively.

2.3. Overview of Explainable Artificial Intelligence (XAI)

Artificial Intelligence is rapidly growing; numerous technologies and methods are being tested using Machine Learning (ML) and Deep Learning models. However, these models are complex and opaque, and the internal workings of the model on how the findings and predictions were made

are not easily accessible. This lack of explanation is called the *'Black Box'* [17]. This complicates the user's understanding of the model's characteristics, ability to identify biases or errors, and confidence in the model's decision-making process. With the constant use and introduction of AI models, this explanation is becoming even more complicated, given the number of AI applications in diverse fields.

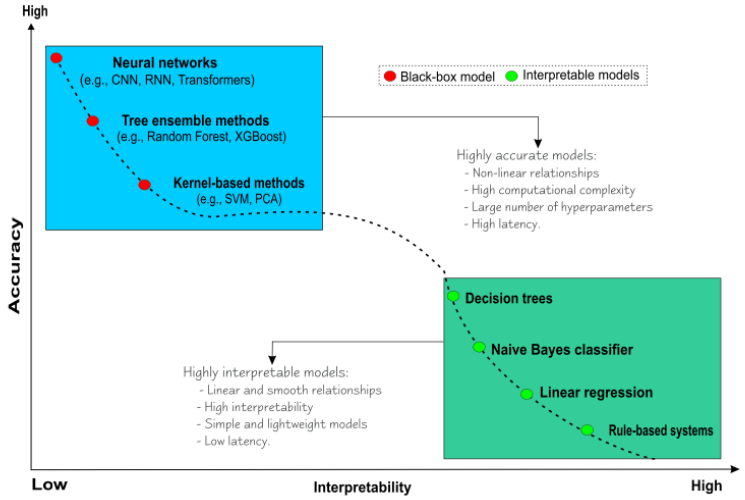

Figure 1: Models tradeoff between predictive and explanatory power [18]

In machine learning, interpretability is critical as for any model to be audited and debugged, a user should be able to quantify and explain the internal workings of the model. When an incorrect prediction is made, identifying the root cause and ratifying the error is crucial for understanding the model. Complex AI and ML algorithms, such as Random Forest, Gradient Boosting, and Artificial Neural Networks, provide much higher prediction accuracy at the cost of interpretability [19]. This is substantiated by the image in Figure 1, showing the model tradeoff between various modeling techniques and their relative strengths and weaknesses in interpretability and accuracy. The models at the bottom are linear models followed by decision trees and graphical models, which are the simplest and easiest to interpret but low on accuracy. However, the models on top, like a neural network, Random Forest, and XG Boost, which are complex, have high accuracy but low interpretability. This is where XAI comes in XAI to primarily explain the workings of the model to build confidence and trust in the model's decision-making capability.

2.4. Core concepts of XAI

Explanation of models requires an understanding of the problem, users' expectations, level of explanation, and the scope of the project, which defines the level of interpretability. Interpretability can be based on scope, agnosticism, stages, and data type, as shown in Figure 2. Interpretation based on scope has two approaches: Global Interpretability, which provides a holistic view of all the explanations provided by the model, and Local Interpretability, which explains individual instances provided by the model.

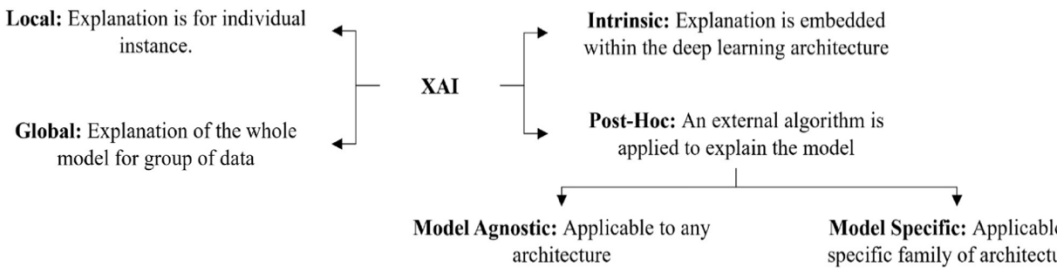

Figure 2: Core concepts of XAI [20]

2.4.1. Global Interpretability

Global Interpretability approaches aim to simplify the understanding of a model's overall reasoning and the entire rationale used to generate specific predictions. The first approach to achieve interpretability is Model Extraction, which involves training an interpretable model using the predictions made by the black box model. Two strategies are used in the creation of a model extraction algorithm. The first strategy is the Rule Extraction method, where a diverse sample of input instances is iteratively fed into the trained neural network, which processes the input data through its layers of interconnected neurons to produce a prediction for each instance. These responses are analyzed to formulate rule-based explanations by identifying common features that result in similar predictions [21]. The main drawback of this method is its inability to create models but rather explain how other complex models work. That is where Model Distillation was proposed by Hinton et al. to overcome this issue [22]. This model transfers all the insights and patterns learned by the complex neural network model during training to a simpler neural network model that becomes equally competent in prediction and interpretability.

The Model Extraction approach provides a successful global explanation of the black box at the expense of compromised accuracy for oversimplification of complex models. To overcome this challenge, Feature Based Methods were introduced, which looked at how important each input variable is in influencing the behavior of an algorithm. This was achieved by measuring the importance and interaction of the input variable's features.

Feature interaction is when an AI algorithm makes a prediction using two features, and the effect of one feature depends on the value of the other. To understand the strength of these interactions affecting the prediction, a statistic called the H-Statistic was created by [23]. This approach involved consolidating individual rules to construct a robust predictive model that effectively integrates these rules to enhance both predictive accuracy and interpretability. Feature importance is calculated based on each feature’s contribution towards the prediction.

2.4.2. Local Interpretability

Local Interpretability focuses on explaining individually why the model made a specific prediction for a single instance rather than understanding the internal workings of the black box at once. It

assigned each input variable a weight multiplied by its corresponding weight before adding it together with a bias term when making a prediction. This weight enables us to ascertain the most important feature for making model decisions. These flexible techniques can be adapted to a complex model but lack generalizability. This technique is more popular than global explanation methods [20].

Local interpretability methods are extensively applied in DNNs, offering valuable insights into the model's decision-making process. However, DNN has many features, making understanding the interaction and relationship between features is complex. Such a complex model makes it difficult to understand which features influence a given prediction most [17], and XAI helps expand the understanding by providing interpretability and insights into how models work.

2.4.3. Intrinsic Technique and Post-Hoc Technique

The timing of interpretability determines another classification of model explanation methods. This classification encompasses the intrinsic technique, consisting of traditional approaches for examining black-box models before their training, and the Post Hoc technique, which involves methods for scrutinizing black-box models after training.

The intrinsic technique applies interpretability to explain the outcome of machine learning before it is trained or used for making predictions. It is designed and integrated into the model to ensure that its decision-making process is understandable and explainable by incorporating interpretable components during the model development phase. This builds trust and understanding of the user of the model.

The explanation of the model prediction after training the model but without modifying the internal workings of the model itself follows a Post-Hoc approach where an external explainer is used to analyze the trained model and provide an explanation [24]. This post-hoc model interpretation method is generally understood through a feature removal approach by accessing the importance of features or a combination of feature importance to the model's performance [25]. This interpretation is achieved by common post hoc approaches.

2.4.4. Model-agnostic and Model Specific Techniques

Model agnostic and model specific techniques are two commonly used post hoc techniques. Model agnostic XAI techniques are flexible and do not require knowledge of the model's internal workings. It can be applied to the DL model, regardless of its architecture [26], demonstrating significant flexibility. This flexibility creates a trade-off between the precision of the explanations and their ability to be applied to a wide range of situations. Shapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are techniques that analyze a model's input-output relationships.

On the other hand, model-specific XAI techniques are restrictive as they are specifically developed to be compatible with a specific range of models that utilize the inherent structure of the model to provide explanations and insight into its internal workings. Such techniques are useful when it is important to understand how the decisions were made by the model. Techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) for CNN and Integrated Gradients for neural networks fall into this category [27]. Another model-specific XAI technique is the Attention Mechanism [28], which is usually applied to transformer-based models where the attention weights are built into the architecture of the model to provide built-in interpretability. This attention plays an important role in focusing on important aspects of the input and understanding the model's decision-making process.

2.5. XAI Techniques

Model interpretation is crucial in explaining how a black box model makes decisions such that it can be understood by humans easily. The higher the interpretability, easier it is for humans to understand. This interpretability can be achieved at an individual instance level or at a global level, which are achieved through several approaches and methods.

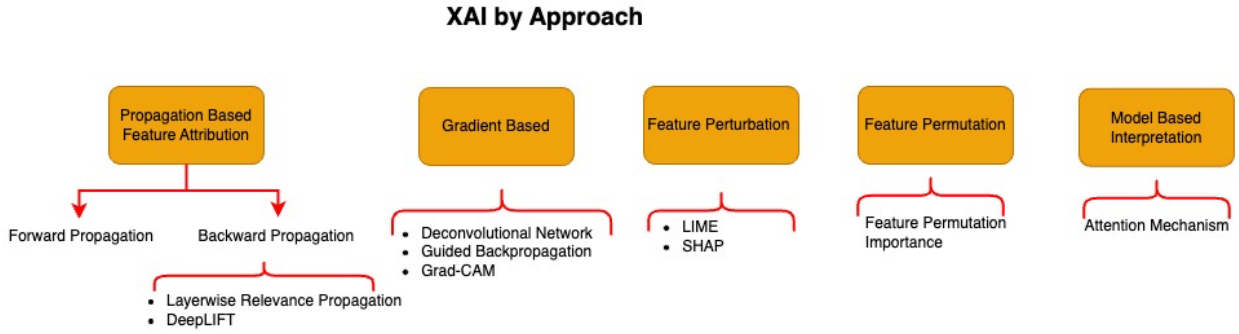

Figure 3: XAI by Approaches

2.5.1. Model Based Interpretation

One such approach is the model-based interpretation approach [29]. It examines all network components to understand the hidden patterns they represent and their impact on the model's performance. This method delves deep into the inner workings of the network, analyzing its various layers, neurons, and connections to uncover meaningful motifs and structures. Gaining insights into these components unfolds how features contribute to the model's decision-making process and predictive capabilities. The model-based interpretation approach offers a comprehensive understanding of neural networks, providing interpretability on the intricate relationships between network components and their collective influence on model behavior. This approach of reviewing every single component is ideal for small datasets; however, it is computationally expensive and time-consuming for large and complex datasets and, therefore, not recommended for such cases.

One such approach is the application of the attention mechanism in neural networks. This is done through attention, which is a complex cognitive function that is an important trait of human beings. This cognitive concept of 'attention' has inspired and evolved into a transformative feature in the deep learning model called the Attention Mechanism [48]. The attention mechanism interprets the

model by assigning weights to different levels of importance to various input features, indicating which part of the input data contributes the most to the model's decision-making process. This mechanism enables the interpretation of which features, or sequence elements are prioritized, which helps to interpret the internal workings of the model's prediction.

2.5.2. Propagation Feature Based Attribution Method

The other method is the propagation-based feature attribution method. This method directly uses altered input data with the model and observes its effect on the prediction. This method is divided into two groups: forward and backward propagation. In forward propagation, input data is sequentially passed through the network layers, from the input layer to the hidden and output layers. It features a unidirectional flow of information with no reverse pathway. At each layer, the incoming data experiences transformation facilitated by weight connections and activation functions. The flow of information is unidirectional, enabling the network to progressively refine the outcome of the input data to facilitate in making predictions.

In this method, input elements are flipped to ascertain feature importance for the trained model, whereby modifying one or more elements can greatly affect the prediction, indicating that the model has ascertained the feature as highly important for making a prediction. In genomics, this strategy is known as In Silico Mutagenesis (ISM) [30] where elements corresponding to nucleotides of biological sequences are flipped to determine important features for the trained models. The feature's importance is shown by a matrix called the attribution map [31], derived from the difference between the model's prediction of the original sequence and the alternative sequence by repeating across all the results. However, the drawback of this method is that it computationally inefficient as each perturbation requires separate forward propagation through the network, which is expensive and time-consuming.

Conversely, backward propagation is efficient since it retroactively computes the importance of features starting from the output neuron and traverses back to the input features through the network's layers in a unified process. This process begins by computing the loss between the predicted and target outputs. Subsequently, this error is propagated backward through the network, allowing for the adjustment of parameters in each layer. This adjustment is accomplished by iteratively computing the loss function's gradients for each layer's parameters. These gradients guide updating the parameters in the direction opposite to their gradient, a technique known as gradient descent. The model incrementally improves its predictions through this iterative process while minimizing the loss function.

One of the backward propagation techniques is the Layerwise Relevance Propagation (LRP) [24], which can be applied to both fully connected and convolutional layers. A standard forward pass through the network is the first step, during which the activations at each layer are recorded. Until the input is reached, each neuron receives a portion of the network's output and distributes it equally to its predecessors. The score attained at the network's output is backpropagated in the network during the second phase, which follows propagation rules based on the relevance conservation property. The relevance measures the strong connection between the input and the output without making any changes to the input. The output of the LRP technique produces a

relevant heatmap highlighting which areas of an input contribute to the output. This allows the practitioner to understand which regions in the input are most relevant for the prediction.

The other back propagation XAI technique is the DeepLIFT (Deep Learning Important Features). It is a technique developed by Springenberg et al. that dissects the influence of specific input features on the network's output prediction through backpropagation [32]. Unlike gradient-based methods that solely rely on gradients, DeepLIFT establishes a reference input to compare against the actual output. This approach enables the propagation of vital signals even when gradients are absent or discontinuous, capturing dependencies that gradient-based models might overlook.

2.5.3. Gradient Based Method

A Deconvolutional Network (Deconvert) [33] is a gradient-based method that begins with activation at a specified layer and maps the activation back to the input space. Activation inversion involves replacing the convolution layer with a deconvolution layer by transposing the filters, the max pooling layer in the forward pass with an un-pooling operation that records the locations of maxima, and the rectification units (ReLU) to only permit the positive signal to pass. It initially sets all other activations in that layer to zero to remove the interference from other neurons and then iteratively sends the modified feature map to the replaced deconvnet layers until it reaches the input space.

Guided Backpropagation [33] is another gradient-based method with simple network architecture presented by Springenberg et al., given that "deconvnet" performs poorly without max pooling layers. This method maintains accuracy by replacing the max pooling layer with a convolutional layer and zeroes out the entries with negative values in both the forward and backward passes as it propagates through the Rectified Linear Units (ReLUs).

The technique known as Grad-CAM (Gradient-weighted Class Activation Mapping) [34] is another gradient-based method used in identifying and interpreting the important regions of an input image for the predictions made by a neural network. It creates a localization map by utilizing the gradients of the target class flowing into CNN's last convolutional layer to highlight important areas in the input image. It calculates the target class's gradients of the score, applies global average pooling to these gradients to determine the weights, and then combines the feature maps using the weights that were determined. The weighted combination that results is then subjected to ReLU activation to make sure that only positive influences are taken into account.

2.5.4. Feature Perturbation Method

Perturbation-based approaches entail systematically altering the input data and observing how these modifications impact the model's output. Determining the importance and impact of various input features on the model's predictions is the main objective of these techniques. The model's decision-making process and the relative weight of various features can be understood by analyzing the changes in the output that arises from these input perturbations.

One of the techniques is the Local Interpretable Model Agnostic Explanation (LIME), which was introduced by Covert et al, to explain how the black box model makes specific predictions by creating simpler, interpretable models that imitate the behavior of the black box [27]. This is achieved by selecting an instance of interest, creating a surrogate dataset by slightly changing the original data points, and assigning weight to the perturbed samples. Sample with higher similarities are given higher weight. This perturbed dataset, along with the corresponding predictions and weights, is used to train the interpretable model, making it simpler and easier to understand. Using the results from the interpretable model, LIME provides an accurate explanation of the prediction made by the black box model.

The SHAP (SHapley Additive exPlanations) [35] is another technique used to explain the output of machine learning models. It is based on Shapely values based on cooperative game theory, which supports understanding the results by visualizing the impact of each feature on the output of the model in summary plots. Features with high positive SHAP values increase the model's performance, and features with negative values decrease the model's performance.

2.5.5. Feature Permutation Method

Another technique for determining the relative importance of specific features in a machine-learning model, is called Feature Permutation Importance (FPI). This technique assists in deciding which features are most crucial for making precise predictions by permuting the values of each feature and evaluating the effect on the model's performance. The relationship between a feature and the target variable is broken by randomly rearranging the value of a specific feature among the data points. A decrease in the model's performance indicates its importance when a feature is permuted. One of the popular methods is the Permutation Feature Importance (PFI) for Random Forest [36]. This method first makes predictions without changing anything, then it randomly shuffles the value of one feature, keeps everything else the same, and makes predictions; this enables to ascertain how much prediction error increases with the change. The higher the increase, the more important the feature is in making accurate predictions.

2.6. XAI in Agriculture

2.6.1. Applications of XAI in Agriculture

In recent agricultural research, a notable shift has been towards applying advanced machine learning techniques to predict and analyze crop yields. This trend reflects a broader recognition of the potential of AI to revolutionize farming practices and enhance productivity. A key aspect of this evolution is the adoption of XAI, which prioritizes accuracy and seeks to explain the underlying factors influencing yield variability. This is exemplified by recent studies such as those conducted by [37] and [38]. Li et al. employed Decision Tree analysis, Feature Importance Evaluation, and SHapley Additive exPlanations (SHAP) to interpret model outcomes, while Zhou et al. introduced the Bayesian Ensemble Model (BM), a statistical crop model that combines multiple weak models to form a robust predictive framework. These approaches offer both

interpretability and accuracy, providing valuable insights into crop yield patterns and contributing to more informed agricultural decision-making.

Furthermore, advancements in XAI techniques have facilitated the early classification of crops by identifying important timesteps in the data. Shams et al utilized Layer-wise Relevance Propagation (LRP) to understand the contribution of individual timesteps to the model's decision-making process, thereby uncovering meaningful insights and highlighting critical factors for classification [39]. Moreover, researchers have explored a diverse range of machine learning models, from simple linear models to complex ensemble models, better to understand agricultural practices' impact on crop yield predictions. Schwalbert et al. demonstrated the efficacy of XAI techniques such as feature importance analysis, Partial Dependence Plots, and LIME in elucidating the variables, interactions, and contextual dependencies driving these predictions [40].

Additionally, developing Glass Box methods like the Explainable Boosting Machine (EBM) has enabled transparent insights into crop yield predictions using multisource data. Çelik et al. introduced EBM, which combines accuracy with interpretability, outperforming traditional black-box methods while providing insights into the interaction between features [41]. Furthermore, the Long Short Term Neural Networks application by Chandra et al. [42] have demonstrated superior performance in predicting crop yields, leveraging extensive time series data.

DL models have emerged as powerful tools for analyzing large datasets and uncovering intricate patterns, particularly in plant phenotyping studies. As Krogh Mortensen et al. and Varshney et al. demonstrate, researchers increasingly favor DL techniques over conventional methods in this domain [43] [44]. Additionally, in the field of genomic selection, there is a growing interest in applying statistical machine-learning methods to analyze, interpret, and predict genetic data. While the application of DL in genetics and plant breeding is relatively new and underexplored, Wolanin et al. suggest that more research and applications are underway to harness its potential benefits [45].

Interpretability is important in deep learning models, as Samek et al. emphasized, where CNNs were deployed to influence large datasets and capture nonlinear relationships among features [46]. The model outperformed other opaque models (Ridge Regression and Random Forest) and demonstrated it could be applied in agriculture as it provided explainable features related to the length of the season, temperature, and lighting conditions. Regression Activation Mapping (RAM) was applied as an XAI technique that provided insights into the complex nonlinear interactions of the input data and how the weather affected the model in relation to the crop growth progression. These developments collectively represent a significant step in optimizing crop production and ensuring food security in a rapidly evolving agricultural landscape.

Humans do not process information in its entirety but rather selectively concentrate on a part of the information that is needed while disregarding other observable information [47]. This is done through attention, a cognitive concept of 'attention' that is inspired and evolved into a transformative feature in the deep learning model called the attention mechanism [48]. Adapting the cognitive capability of allocating attention to important aspects of information, this attention mechanism in deep learning emulates this core concept. This mechanism helps neural networks focus on specific parts of the input and prioritize the most important information. The core idea of

the attention mechanism adapted from human cognitive function is not to treat all input data equally but to dynamically allocate different weights to different parts of the input data based on their importance or relevance. This technology to interpret the workings of a model has been applied to a parallel structure with attention-based CNN and the LSTM model [28]. The attention mechanisms enable neural networks to dynamically prioritize and concentrate on specific input data, which enhances AI model performance and interpretability. The strength of the technology lies in processing sequential or spatial data, where the importance of individual input features varies greatly.

2.6.2. Application of XAI Techniques in Genomics Study

One of the XAI techniques in DNN is the gradient-based method, which is used in genomic studies. It interprets and elucidates DNN predictions by computing gradients of the output and assigning importance to input features through backpropagation. Popular gradient-based techniques for XAI in CNN applied in genomic data are Grad-CAM, Smooth Grad, Guided Backpropagation, and Integrated gradient. However, a primary constraint of this gradient technique is the treatment of zeros by the ReLU, an activation function in neural networks. If the input to ReLU is negative during backpropagation, ReLU converts it to zero, thus disregarding its impact on the process. Guided Backpropagation introduced by Selvaraju et al. experiences similar issues, nullifying features if the input to ReLU is negative during both forward and backward passes [33]. By transforming negative scores to zero, both techniques overlook the negative contribution of gradients to the output.

The other method applied in the genomic study is the propagation-based feature attribution method, which uses forward or backward propagation to alter input data with the model and observe its effect on the prediction. LRP is one such technique applied in various domains, including bioinformatics. This method explains the predictions made by neural networks layer by layer, redistributing relevance scores based on each layer's activation and parameters. These relevance scores, representing each neuron's contribution to the final prediction, are allocated to individual neurons within the network. As this score is retroactively propagated backward through the network, relevance is apportioned to input features according to their impact on neuron activations across layers. By highlighting the flow of relevance throughout the network, LRP facilitates a deeper understanding of the significance of individual features or neurons in shaping the model's predictions [49]. Moreover, it enhances the interpretability of neural networks by associating model decisions with specific input features or regions within the input space.

The other back propagation XAI technique is the model-specific technique specifically designed for deep neural networks called the DeepLIFT. It is a technique developed by Montesinos-López et al. that dissects the influence of specific input features on the network’s output prediction through backpropagation [50]. It achieves this by evaluating each neuron's activation value and assigning a contribution score based on discrepancies. This technique facilitates the creation of visualization tools such as heat maps to show the most important input features. This approach enables the propagation of vital signals even when gradients are absent or discontinuous, capturing dependencies that gradient-based models might overlook. Moreover, DeepLIFT efficiently

generates contribution scores in a single backward pass following a prediction via backpropagation.

DeepSHAP, a mixture of DeepLIFT and Shapley values, is an enhanced version of the DeepLIFT algorithm that approximates the conditional expectations of SHAP values using a selection of background samples. It estimates approximate SHAP values by integrating many background samples and finding the difference between the expected model output and current model output on the passed background samples [51]. This explainable approach has several advantages, including the local and global explanation of the contributions of each input variable. It also exploits the DL features to improve computational performances and extract deep information.

Attention Mechanisms have been used on a recurrent neural network model to classify images; likewise, they have also been used for translation and alignment on machine translation tasks. Novakovsky et al. performed an experiment to estimate the wheat yield using remote sensing and meteorological data by applying two forms of attention-based structure [28]. One is a parallel structure with attention-based CNN with LSTM, and the other is a series structure with attention-based CNN with LSTM. The results from the study showed that the parallel structure provided a better estimation of accuracy than the series structure, with remote sensing data contributing significantly to crop yield estimation [28].

Apart from being used in numerous data analysis tasks, attention mechanisms, a model-specific XAI technique have also been utilized for generating explanations of the prediction models' behavior, typically in the form of visualization schemes (indicating words or image areas where the primary model focuses on) or feature importance metrics [52]. However, even though the attention mechanism has been used for the explanation of prediction models in the form of feature importance metrics, some conflicting arguments have been raised on the suitability of the attention mechanism to be used for explainability. [53] have argued that the distribution between learned attention weight and gradient-based features relevance method does not provide similar results for the same prediction.

2.7. Multimodal Learning

Multiple-modal learning is integrating heterogeneous data from multiple sources to make better predictions by unveiling patterns and motifs that would not be possible using a single modality in isolation [54]. In multimodal learning, one of the key challenges is learning from the representation of input data. When dealing with tasks that require handling multiple data types (such as images and text), it is often challenging to gather corresponding data from these distinct types. Pre-trained models with prior knowledge of these specific data types greatly help resolve this issue.

The other challenge is fusing this representation of different modalities into one. Fusion integrates information from different data sources into a single multimodal representation. This process of combining data from multiple sources is called multimodal fusion [55]. The two most common multimodal representation techniques are early fusion and late fusion. In early fusion, vector representations of the information are concatenated at the initial stage before training. Meanwhile,

several independent models concerning each modality are trained in late fusion, and their outputs are concatenated [56], as shown in Figure 4.

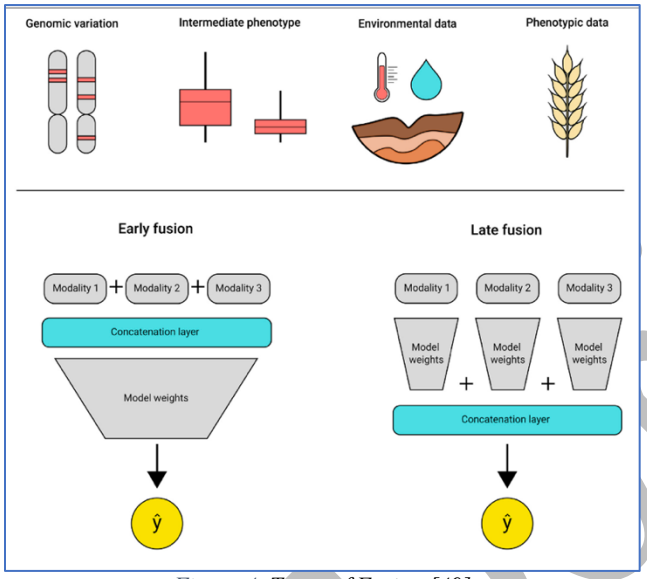

Figure 4: Types of Fusion [49]

3. Methodology

3.1. Data Overview

The dataset utilized in the study was obtained from the Western Barley Genotyping Alliance (WCGA) at Murdoch University. It encompasses field experiments conducted between 2014 and 2016, integrating multimodal sources such as phenotype and genomic data.

Genotype data

This dataset comprised genotype data of barley plants, offering insights into the genetic composition of individual species and the combination of alleles found at particular genomic locations represented by four nucleotide bases Adenine (A), Cytosine (C), Thymine (T), and Guanine (G) that make up DNA. Each nucleotide base in DNA pairs specifically with its complementary base according to the rules of base pairing: Adenine (A) pairs with Thymine (T), and Cytosine (C) pairs with Guanine (G), which are important for maintaining the structural integrity of DNA.

A total of 894 barley accessions underwent genotyping using Next-Generation-Sequencing techniques. Subsequent filtering for heterozygosity, mapping quality (20), and major allele frequency (MAF) of 0.01 yielded 30,543 high-quality Single Nucleotide Polymorphism (SNP) markers, which constituted the genotype dataset.

Phenotype data

Phenotypic data for the barley comprised 12,421 observations and 17 features collected from 2014 to 2016, encompassing physical attributes, physiological features, and other measurable traits of barley plants collected across five distinct geographical field trial sites in Western Australia: Geraldton, Merredin, South Perth, Katanning, and Esperance. These sites encompassed twelve large-scale field trial experiments conducted between 2015 and 2016. Each trial site represented unique climatic conditions and belonged to separate agricultural zones. Additional experiments assessed specific environmental factors' impact on barley phenotypes. For instance, an extended light exposure trial was conducted in 2016 at the South Perth site, comparing phenotypic performances under 18 hours of artificial lighting to those under natural light conditions. In contrast, a separate irrigation trial at the Merredin site compared barley traits under irrigated versus non-irrigated conditions.

Variety	Reason	Year	Study	Location	ZS49PlHt	ZS49 (no days)	HrvPlHt	GrYld(kg/ha)
SMBA12-2297	RES	2016	2Ir	MER	60	115	NaN	4,375.09
BmnL-28	RES	2016	2Ir	MER	70	103	NaN	2,953.28
SB99252	RES	2016	2Ir	MER	70	111	NaN	3,909.63
Astoria	RES	2016	2Ir	MER	50	115	NaN	4,029.26
IGB1244	RES	2016	2Ir	MER	75	106	NaN	4,653.19
Skiff	AUS	2016	2Ir	MER	55	115	NaN	4,243.22
WI4876	RES	2016	2Ir	MER	65	103	NaN	4,015.71
WI4547	RES	2016	2Ir	MER	85	111	NaN	3,947.74
WI4665	RES	2016	2Ir	MER	65	106	NaN	3,383.93

Features included details on plant variety, irrigation and lighting experiment management, plant location, flowering time (ZS49), flowering time plant height (ZS49PlHt), plant height (HrvPlHt), and grain yield in kilograms per hectare (GrYld(Kg/ha)). Additionally, the features incorporated the growth scale of barley related to two developmental phases: booting phase ZS49 (flowering time) and ripening phase ZS91 (harvesting time). Information on lighting and plant irrigation treatment was also provided through features such as Harvesting Plant height (HrvPlHt) and study features. Moreover, the study included the grain yield of barley measured in kilograms per hectare (GrYld(Kg/ha)) for specific species.

Environmental Data

The environmental data utilized in this study comprised three key weather parameters: rainfall, solar exposure, and temperature. These data were collected over two years, spanning from January 2015 to December 2016, from weather stations located within the study area.

Specifically, the dataset included daily rainfall measurements (millimeters), daily solar exposure readings (MJ/m²) measuring the total amount of solar energy that plants received over the course of a day, and daily maximum temperature recordings (Celsius). As the data were collected over a period of 2 years, a data reduction strategy was applied to make the analysis more manageable and aligned with the crop growth cycle.

GrYld(kg/ha)	04_temp_max	05_temp_max	06_temp_max	07_temp_max	08_temp_max	09_temp_max
2410.5	28.04	24.593548	23.313333	20.867742	21.029032	25.343333
4024.9	27.386667	23.36129	19.676667	18.916129	19.36129	20.436667
2360.5	21.9	17.651613	17.48	15.151613	15.870968	19.143333
3683.5	20.853333	17.383871	14.313333	13.574194	13.741935	15.083333
4060	22.46	18.506452	18.443333	15.880645	17.390323	20.031034

3.2. Data Pre-processing

From the phenotype dataset, the investigation looked at attributes specifically pertinent to barley yield. Among the 18 attributes initially considered, 8 were discarded due to excessive missing values and their limited relevance to the study objectives. Subsequently, only 9 crucial attributes remained for further analysis. However, it's noteworthy that these retained features exhibited missing values and a straightforward approach was adopted to address this by imputing the missing values with the mean. Notably, it's important to highlight that the grain yield of barley (GrYld(Kg/ha)) was excluded from this imputation process, as it served as the target variable for the study.

Of the 643 features initially present in the Genotype dataset, 6 features ('assembly#,' 'center,' 'protLSID,' 'assayLSID,' 'paneled,' and 'QCcode') were deemed unsuitable for the study due to their absence of values, resulting in their exclusion. Consequently, the analysis focused on the remaining 637 features, encompassing DNA sequences utilized for the study's investigation.

One-hot encoding is a prevalent method in leveraging genome SNP data for deep learning comprehension, as [57] indicated. This approach was employed explicitly in analyzing whole-genome SNP data, where each genomic position was represented by four columns corresponding to the four DNA bases. Here, a binary representation indicated the presence (1) or absence (0) of each base at a given position. Moreover, to account for non-DNA bases denoted by '-' and 'N', numerical values of -100 and 100 were respectively assigned.

Subsequently, after data preprocessing, phenotype and SNP data were fused into a single, compact multimodal representation before feeding into the model using the 'Variety' feature as the identifying column. The merged data encompassed 6,675 observations and 30,552 features, capturing a comprehensive array of plant species alongside their associated phenotype characteristics. Categorical variables such as 'Reason,' 'Study,' 'Location,' and 'Year' were transformed into the binary form using a label encoder.

Likewise, the environmental data collected over a period of 2 years were reduced to manageable levels by converting daily measurements to average monthly figures and aligned with the crop growth cycle, resulting in 24 data points for each weather parameter. Recognizing that barley had a specific growing season, the study concentrated on the six-month period critical for barley development spanning from April to September, excluding the remaining six months (October through March) from the analysis. By removing this, only 18 data points were taken for the experiment, which consisted of 6 features for each environmental data. This decision ensured that only the most relevant environmental conditions during the barley growing season were considered in the predictive model. This refined approach allowed for a more targeted analysis, focusing on the environmental factors that significantly impact barley growth and yield.

Following this process, the dataset was partitioned into labeled and unlabeled subsets, the division being contingent upon the presence of grain yield data for barley (GrYld(Kg/ha)). Out of the total 6,675 observations, 4,058 lacked grain yield values, while the remaining 2,617 observations contained this value. These 2,617 instances constituted the labeled data and were split into training and testing datasets. Specifically, the training set comprised 70% of the labeled data, with the remaining 30% reserved for the testing set. In this partitioning, grain yield (GrYld(Kg/ha)) functioned as the dependent variable, while the other features were treated as independent variables.

3.3. Proposed Models and Experimental Setup

3.3.1. Baseline Model – Random Forest Regressor

The Random Forest Regressor (RF) was used as the baseline model to provide a reference point from which to compare the performance with the proposed models. Random Forest is an ensemble method that makes predictions by averaging the results created through individual decision trees trained on a random subset of the data. This model was used as a baseline for the thesis as it is well suited for high dimensional data, which have a large number of features, capturing intricate patterns without assuming linear relationships and also providing feature importance score, which is important for explaining the working of the prediction.

The experimental setup of the model was designed in the following way. As the baseline model's performance had the opportunity for improvement, a performance improvement technique was applied. For the Random Forest Regressor, hyperparameter tuning was applied using the Optuna optimization framework [58], where the models were trained on different subsets of Gene Data and Phenotype data (excluding environmental data) to reduce overfitting and improve generalization.

Hyperparameters were set using Optuna optimization framework to improve the prediction performance of the RFR. The hyperparameters included the number of observations drawn randomly for each tree, the number of variables for each split, the splitting rule, the minimum

number of samples, and the number of trees. This was done to balance model complexity and performance, reduce overfitting, better handle noisy data, and focus on the most important features.

The optimization results yielded n_estimators of 415, max_depth of ≈ 75, min_samples_split of 5, and min_samples_leaf of 3 as the best hyperparameters. These hyperparameters indicate moderately large forests with relatively deep trees, indicating a balance between model complexity and generalization. The training involved 100 epochs with a 32-batch size.

3.3.2. Uni-Modal Models - CNN and LSTM

CNN Model for Gene Data

The Gene Data was experimented with using the CNN Model as it is efficient in processing and analyzing complex data through its layered architecture. CNN was applied to the genomic data by treating it as sequential data to understand the underlying relationships between genomic regions. The model architecture consisted of layers, filters, dense layers, dropouts, and activation functions to focus on different features of the data to capture complex patterns, reduce the dimension of the data, and prevent overfitting.

The experimental setup of the model was designed as follows. Hyperparameters of the CNN Model were optimized using the Optuna framework. The result of the optimization achieved parameters of, i.e. the number of filters (n_filters: 61), kernel size (kernel_size: 2), activation function (activation: relu), units in the dense layer (dense_units: 47), and dropout rate (dropout_rate: 0.408087511) to allow the model to capture complex patterns while maintaining the robustness and preventing overfitting.

Based on these parameters, the CNN architecture incorporated a 1D convolutional layer with 61 filters designed to capture local patterns in the input data. With 61 filters, the model was designed to identify various features at each step in the sequence, improving its ability to detect subtle and nuanced patterns. Kernel size of 2 was applied to the architecture to process two adjacent values simultaneously, which is particularly effective for time-series data. Furthermore, the ReLU activation function was applied to learn complex patterns while preventing the risk of vanishing gradients that can hinder model training. The L2 regularization was also applied on both the kernel and bias parameters (each with a regularization factor of 0.01) to introduce a degree of weight penalty, effectively discouraging the model from overfitting.

The second layer had a MaxPooling1D layer with a pool size of 2 to reduce the data's spatial dimensions and allow for computational efficiency while retaining important features. A dropout layer with a dropout rate of 0.408 was further applied to help mitigate overfitting by randomly deactivating a portion of neurons during training, forcing the model to generalize better.

With the same configuration, a second 1D convolutional layer, MaxPooling and Dropout layers, was applied to reinforce feature extraction and prevent overfitting. The output was then flattened to create a vector suitable for dense layers. A Dense layer with 47 units and ReLU activation, L2 regularization and dropout was applied to transform features to promote model robustness. Finally, the output is a Dense layer with a single unit to predict the target variable.

The training process was performed with 100 epochs with a batch size set to 32, and the full training set was 20% of the data used for the validation, followed by fine-tuning 80% of the training data. The trained model was then evaluated on the 20% held-out test set.

LSTM Model for Environmental Data

The Attention-based LSTM model was applied to the environmental data to capture its temporal nature and make accurate predictions. The model's hyperparameters were optimized using the Optuna framework, resulting in an LSTM architecture tailored to the temporal characteristics of the input data. The network's core is an LSTM layer with 50 units, designed to capture and learn long-term dependencies in the temporal data, configured to return sequences, allowing for the preservation of temporal information throughout the network. A key feature of the architecture was incorporating a Multi-Input Attention mechanism. This attention layer processed the output from the LSTM, which enabled to focus on the most relevant parts of the input sequence. The attention mechanism produced two outputs: the attention-weighted features (attention_output) and the attention weights themselves (attention_weights).

Following the attention layer, the model includes a Dense layer with 128 units and ReLU activation. This layer transforms the attention-weighted features, allowing for further abstraction and pattern recognition. A Dropout layer with a rate of 0.10808 was applied to mitigate overfitting after this dense layer. The Dense layer with a single unit and ReLU activation was applied. The training process likely involved 100 epochs with a 16-batch size, using a portion of the data for validation during training.

3.3.3. Multi-Modal Model - Attention-based Hybrid Model

Understanding complex patterns and producing accurate predictions require processing and interpreting data from various sources and types. Given that LSTM efficiently captures temporal dependencies and patterns over time and CNN effectively handles high-dimensionality data (genomic and phenotypic data), a hybrid model with an integrated attention mechanism is proposed. By combining CNN and LSTM with attention mechanisms, the model was able to learn from temporal changes and spatial relationships across different types of data, increasing its overall explainability and predictive capability. Given its advantages, the model was applied to make predictions and explain the internal workings of the model. The experimental setup of the model was designed as follows.

LSTM for Rainfall, Solar, and Temperature Data

The rainfall, solar, and temperature were processed through LSTM networks in their own LSTM pipeline as they were ideal for sequential data to capture temporal dependencies. The model applied two LSTM layers with 223 and 265 units with a dropout of 0.2 between them for each modality, allowing each branch to focus on learning meaningful temporal patterns for its specific input, helping prevent overfitting with the dropout layers.

CNN for Gene Data

Gene data and phenotype data, represented in a different structure, were processed by a CNN to extract their spatial patterns. The CNN architecture incorporated a 1D convolutional layer with 61 filters and a Kernel size of 2 into the architecture. Furthermore, the ReLU activation function was applied to learn complex patterns while preventing the risk of vanishing gradients that can hinder model training. The second layer incorporated a MaxPooling1D layer with a pool size of 2 and a dropout layer with a dropout rate of 0.408. Flattening was followed to reduce the dimensionality and create a compact feature vector.

Feature-Level Fusion

After processing each modality, the model combined the outputs from the LSTM and CNN branches through concatenation. The outputs from each LSTM branch (representing rainfall, solar, and temperature data) were concatenated into a single representation or feature vector to capture an integrated view of temporal dependencies across all weather factors, enabling the model to analyze relationships between them. Separately, the CNN output from the gene and phenotype data was concatenated with this unified LSTM feature vector in the final concatenation layer by applying early fusion, producing a comprehensive multimodal feature representation.

Attention Mechanism

This feature-level fusion enabled the model to use patterns across modalities, combining all feature representations before passing them through the attention and dense layers. After the feature level fusion, an attention layer was applied to allow the model to focus on important aspects of the fused representation. This additional attention mechanism allowed the model to weigh influential input features more heavily than others, providing an additional feature extraction layer that has the most influence on the target prediction.

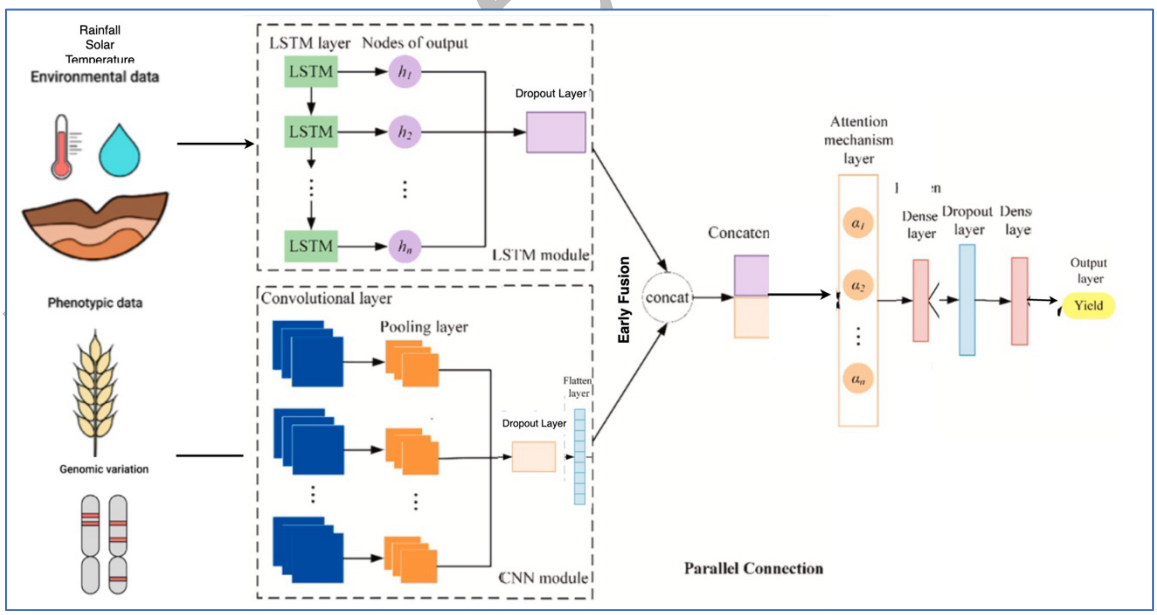

Figure 5: Attention-based CNN and LSTM model

The attention-enhanced fused representation was passed through a dense layer with 64 units, with ReLu activation function and dropout of 0.4, which was then passed through a single dense layer to generate the final results. The training process was performed with 100 epochs with a batch size set to 32, and the full training set was 20% of the data used for the validation, followed by fine-

tuning 80% of the training data. The trained model was then evaluated on the 20% held-out test set.

3.4. Evaluation Metrics

In this thesis, three performance metrics were used to evaluate the performance of the models. The primary metric was the Root Mean Square Error (RMSE), which provides an interpretable error metric in the same unit as the target variable. It was used to evaluate the accuracy and performance of the models by measuring how well the model-predicted values match the actual observed value.

The second metric, the Mean Absolute Error (MAE), was used to measure the average of the absolute differences between predicted and actual values without squaring them. It was used to help find the average prediction performance across diverse values without disproportionately focusing on the largest errors. The last metric was Mean Squared Error (MSE), which measures the average of the squared differences between actual and predicted values and penalizes larger errors more significantly. It is sensitive to outliers, so squaring larger errors contributes higher to the total error, which helps in indicating significant variation and prediction failures.

4. Results

4.1. Prediction Results

The RFR (Baseline Model) achieved a Root Mean Squared Error (RMSE) of 742.568, indicating, on average, the predictions were approximately 742.568 units of the actual values. It also indicated that the model was fairly consistent with predictive accuracy. Additionally, the model achieved a Mean Absolute Error (MAE) of 571.235, showing that the average absolute deviation between predicted and true values was around 571.235 units. Both RMSE and MAE slightly decreased with hyperparameter tuning, indicating overall reduced error, validating the positive impact of tuning on the model's performance.

The CNN model with Gene data achieved an MAE of 870.18, MSE of 1,120,369.41, and RMSE of 1058.48, suggesting that, on average, the model predication deviated from the actual values by approximately 870 units of MAE, which could be due to the complexity and size of the gene data across diverse gene sequences. RMSE of 1058.48, which is higher compared to MAE, indicated that the model was effective, but it encountered some variability in its test set prediction.

The LSTM Model with environmental data achieved MAE of 724.25, demonstrating an average prediction deviation from actual values of around 724 units, indicating a relatively accurate performance in capturing key patterns within the environmental data. Similarly, RMSE of 924.74 indicated that the attention mechanism successfully focused on relevant features, enhancing the LSTM's performance.

The Attention-Based Hybrid Model performed relatively well, achieving an RMSE of 877.314. suggesting that the model effectively captured patterns within the data, leading to accurate predictions on unseen samples. While comparing to the training set, there was a relatively small

increase in RMSE, indicating generalization and less overfitting. The summary of the results were as shown below.

Classifier	MAE	MSE	RMSE
Random Forest Regressor (Baseline Model)	571.235	551,408.604	742.568
Unimodality - CNN model (Gene Data)	870.175	1,120,369.412	1058.475
Unimodal - LSTM model (Environmental Data)	724.247	855,134.938	924.735
Multimodal - Attention Based Hybrid Model	688.614	769,681.033	877.314

From the above experiments, the Attention-based Hybrid Model, with its integrated architecture built on the strength of CNN, LSTM and an attention mechanism, captured both complex sequential and temporal data and provided a cohesive understanding of the underlying patterns that impact the model's prediction. The model achieved superior performance with a low RMSE and MAE while capturing complex patterns. It also showed better generalization with a smaller increase in RMSE from the training and testing, indicating less overfitting.

In contrast, the CNN and LSTM model's performance results were lower compared to the other models, in addition to being trained on a single data source compared to the hybrid model, and it also lacked the same level of feature integration, making it less accurate. While RFR performance metrics were superior, it had a few drawbacks. Firstly, the results provided did not present the full representation of the prediction as it did not contain environmental data in the model, which might have ignored capturing the complex relationships across multimodal data, which is important for the prediction. Secondly, even with optimization, RFR had a higher variation in RMSE from the training and testing, indicating overfitting in the model. These results showed that the attention-based hybrid model performed the best as it was the most balanced and effective model, integrating multimodal data and capturing diversity and complex patterns.

4.2. Explanation Results

After achieving the results, XAI was applied to respective experiments to understand the internal workings of the model.

4.2.1. Baseline Model - Random Forest Model

For the baseline model, Mean Decrease in Impurity (MDI) was used to measure the influence of each feature on the model's prediction and help identify which features are the most important for accurate predictions. It leveraged variance to calculate feature importance and track how often each feature is used to split nodes across all trees in the forest. The cumulative impurity reduction achieved with those splits gave a feature importance score, with higher scores indicating more influence on the model's output.

Three of the three phenotype features were identified as highly influential based on MDI scores. The model highlighted "Year" as the most critical feature, with an importance score of 0.262692, suggesting that time-related changes were essential for accurate predictions. "Location" ranked

second with an importance score of 0.118631, emphasizing the significance of geographic information in shaping model predictions. In contrast, "Study" received a comparatively low importance score of 0.023387, indicating a lesser but still relevant impact. "Year" and "Location" as top features suggest that time and location-specific environmental factors play a substantial role in the model's predictions.

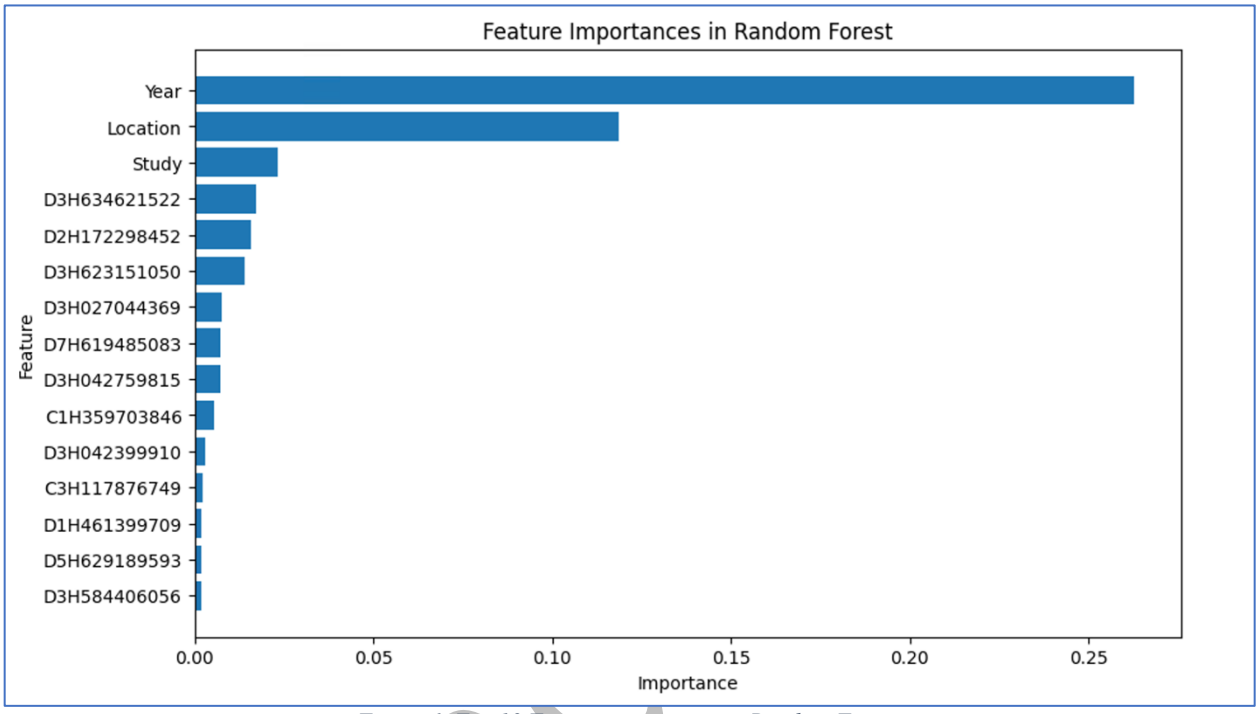

Figure 6: Top 10 Feature importance in Random Forest

In addition to these phenotype features, three genomic data features also showed importance scores exceeding 0.010, indicating their moderate but notable influence. Specifically, "D3H634621522" had an importance score of 0.017060, "D2H172298452" scored 0.015612, and "D3H623151050" scored 0.014140. While lower than "Year" and "Location," these genomic features still contribute meaningfully to the model's predictions, potentially capturing specific genetic variations that interact with the primary phenotype features. This combination of time, location, and genomic influences enhances the model's predictive capacity and offers insight into the key factors shaping the outcomes.

4.2.2. Uni-Modal Models - CNN and LSTM

CNN Model for Gene Data

For the CNN model using the Genomic data, DeepSHAP, a mixture of DeepLIFT and Shapley values, which was applied to understand the main results related to the model's explanation. This XAI was applied to highlight the importance of each input variable and explain its influence on the CNN model's decision-making. This technique assigns an importance value to each feature, which is indicated by a Shapley value; the positive Shapley value reflects an increased effect in the prediction and vice versa.

The summary bar plot describes the absolute average and the Shapley value of influence each feature has on the prediction, and the longer the blue bar, the more influence that particular variety has on the model's predictions. From the top 100 varieties, the top 10 varieties that had the most impact on the model prediction, with varieties D1H001041484 and L1H001008205, had the strongest impact on the model output with a SHAP value of more than 2.5. Followed by L1H001008155, LH001008139, and L1H001008126, with SHAP values between 2 and 2.5. L1H000995418, L1H000995388, L1H000995346, L1H000995322 and L1H000995310 had SHAP value between 1.7 and 2 as shown in the figure below.

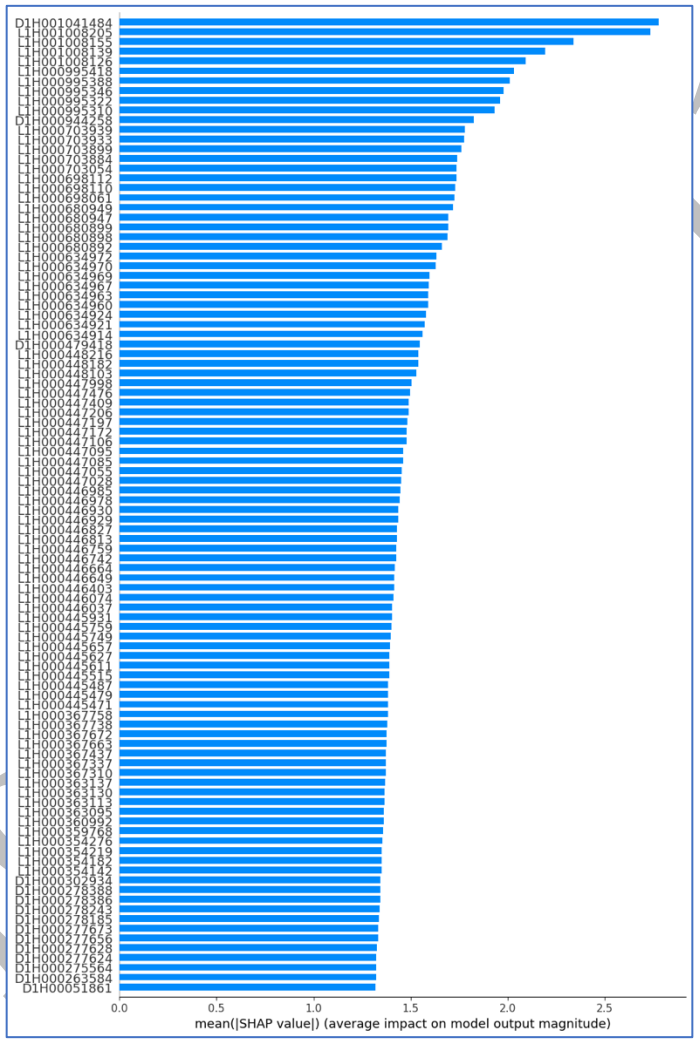

Figure 7: Global feature importance sorted from most important to least important

Attention-based LSTM for Environmental Data

An attention mechanism was used to explain the prediction for the attention-based LSTM model on the environmental data. The two bar charts below illustrate the distribution of attention weights across the environmental features used in the LSTM model with an attention mechanism. The attention mechanism enabled the model to assign different weights to input features based on their relevance to predicting the target variable, allowing it to focus on the most influential feature. The

results highlight the effectiveness of the attention-based LSTM model in identifying key environmental features along with specific timeframes that impact the prediction.

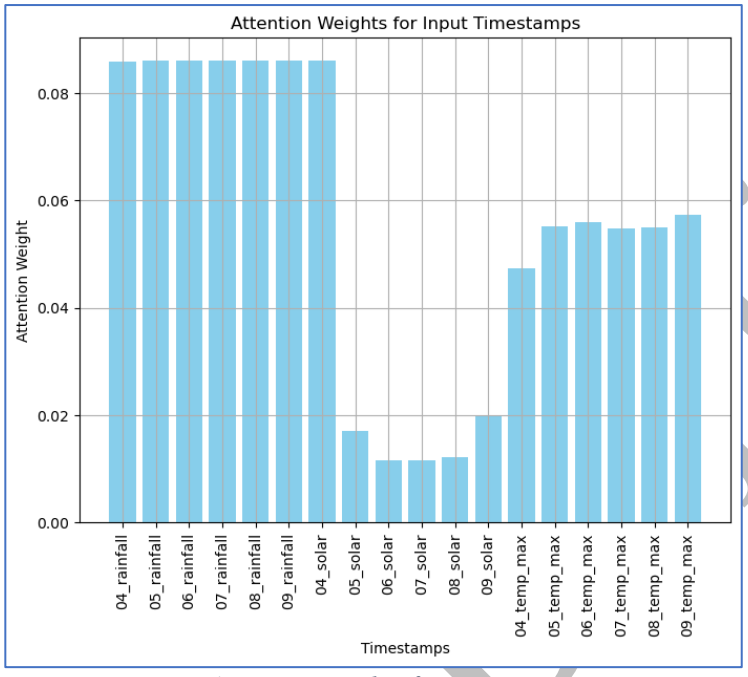

Figure 8: Attention weights for input timestamps

The above graph plots attention weights across different timestamps (April to September) for each feature type (Rainfall, Solar, and Temperature), visualizing the importance the model assigns to different months. The rainfall data shows high attention weights in the initial months (April to July), indicating that early rainfall data is considered highly relevant for predictions. This likely reflects early rainfall's importance in setting up crop yield conditions. The attention weights for solar exposure are comparatively lower than rainfall, with the weights peaking around August and September, suggesting that it plays a more nuanced role than rainfall and is likely relevant later in the season. The attention weights assigned to temperature are consistent from April to September, with a moderate increase in attention during August and September. This stability suggests that temperature plays a relatively steady role across months, perhaps influencing crop growth throughout the season.

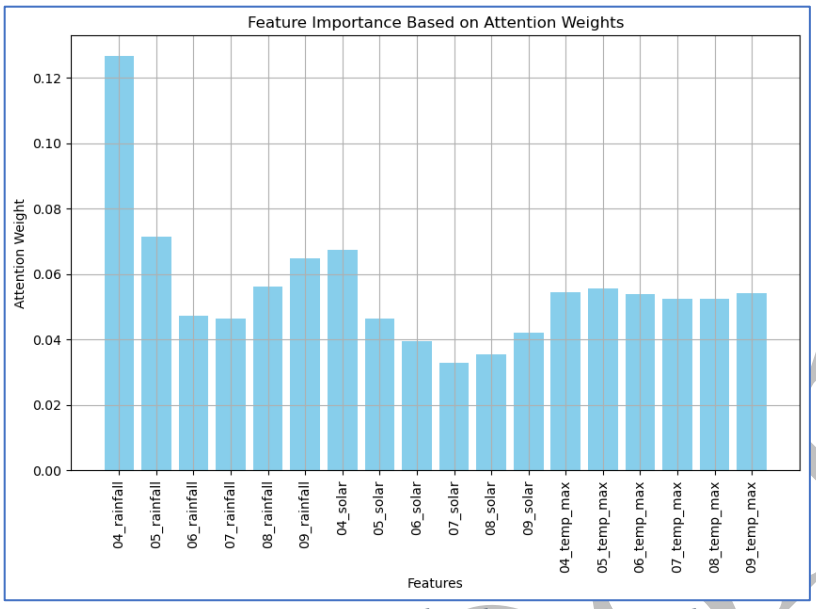

Figure 9: Feature importance based on attention weights

The second graph aggregates attention weights by feature, highlighting an overall view of each feature's importance. The graph shows that April's rainfall holds the highest weight among all features, suggesting it is the most critical factor in the model's predictions. These months correspond to autumn and early winter in the region, which is the wet season and typically experiences higher rainfall. This is essential for the crop as it provides adequate rainfall during these months, supporting soil moisture levels and facilitating strong root establishment and initial vegetative growth. The water availability during these months signifies that the crop has sufficient water supply through its growth phase, which is important for maximizing yield. Followed by this, May and June rainfall also show relatively high attention weights, confirming that early-season rainfall plays a crucial role. The other two feature types are weighted more evenly. Solar exposure peaks mid-season, while temperature's importance is consistent across months, indicating these features have a more balanced influence throughout the season.

4.2.3. Multi-Modal Model - Attention-Based CNN and LSTM Model

In the final experiment, the results from the model highlight the top 50 important features based on the attention weights of the hybrid CNN-LSTM model. From the graph, genetic markers play the most significant role in the model's predictions, with the top 50 features having the highest attention weights. The presence of many genetic markers at the top, with an attention weight of around 0.62, indicated that the model relied heavily on these key features to make predictions. While genetic markers dominate the top features, rainfall and temperature data showed relatively higher attention weights than solar, underscoring the model's emphasis on precipitation and temperature patterns in decision-making. With all the top 50 important features being genetic markers and no environmental data included, it suggested that these markers predominantly influence the model's accuracy, as shown in the graph below.

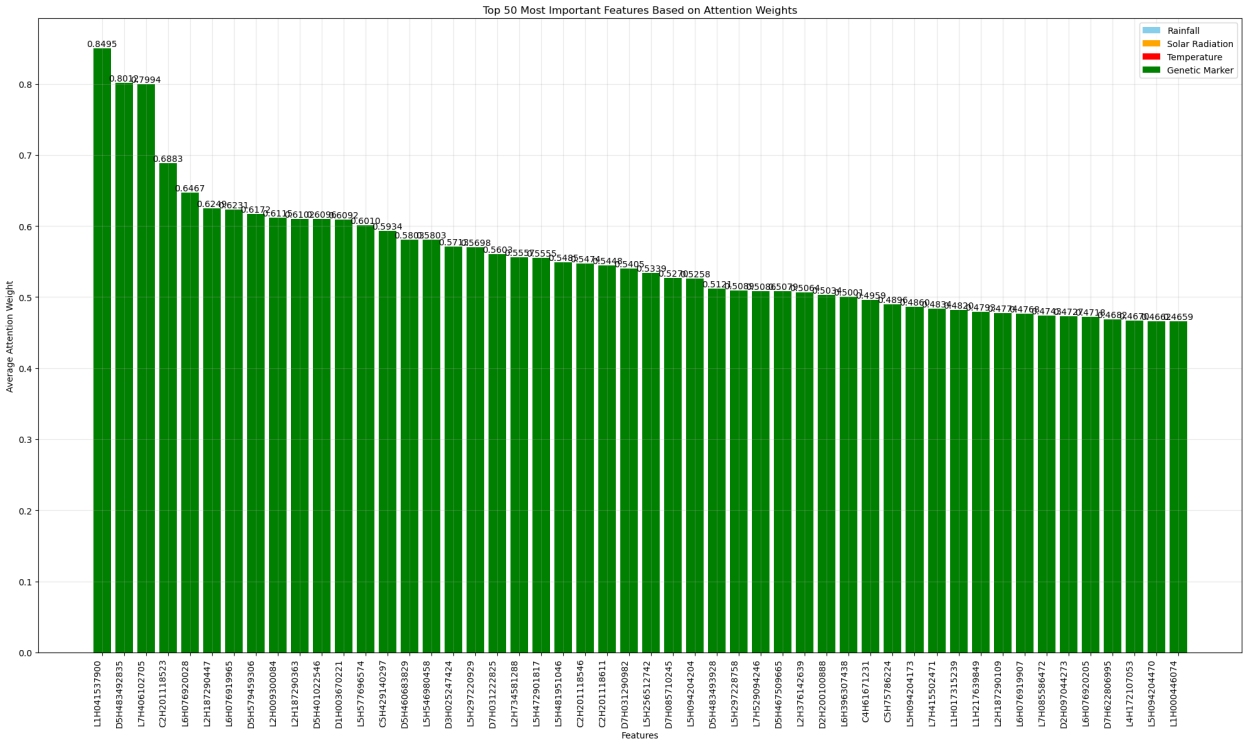

Figure 10: Top 50 most important features based on attention weights

The model identified the genetic marker 'L1H040041456' as having the highest weight, with a score of 0.8495, making it the most impactful factor in the model's predictions, followed by 'D5H480634196' and 'L7H406100741' with a weight of 0.8012 and 0.7994 respectively. The top 10 markers with the highest weights are as follows:

Genetic Marker	*Weight*
1. L1H040041456	| Weight: 0.8495 | Type: Genetic
2. D5H480634196	| Weight: 0.8012 | Type: Genetic
3. L7H406100741	| Weight: 0.7994 | Type: Genetic
4. L2H198497963	| Weight: 0.6883 | Type: Genetic
5. D6H074421386	| Weight: 0.6467 | Type: Genetic
6. D2H185515224	| Weight: 0.6249 | Type: Genetic
7. L6H073914128	| Weight: 0.6231 | Type: Genetic
8. D5H577791636	| Weight: 0.6172 | Type: Genetic
9. L2H009299134	| Weight: 0.6115 | Type: Genetic
10. L2H181338680	| Weight: 0.6102 | Type: Genetic

Rainfall for April, May, and August had higher attention weights, with April (0.1910) and August (0.0691) indicating that this feature played an important role in crop yield prediction. April and May, which marked the growing season's beginning, were assigned higher attention weights, highlighting rainfall's important role in germination and early crop growth stages. Late seasonal rainfall in August indicated a cooler season, impacting crop maturity and health.

Temperature received the highest attention weight, with a value of 0.4535 in May, surpassing other months, indicating that this month had the highest impact on the prediction for the environmental data as it is important in influencing crop growth in the initial crop development. The reason for receiving the highest attention weight across all features could be attributed to its influence on crop growth, flowering, and development. Later months after June received negligible attention weights, indicating minimal impact on the yield as it entered the maturing and harvest stages.

The highest attention weight for solar radiation was Sept (0.2758), followed by April, indicating the requirement of adequate sunlight for crop photosynthesis, which is essential for crop growth. The findings suggested that the model places the least importance on solar compared to the other two features, as shown in the graph below.

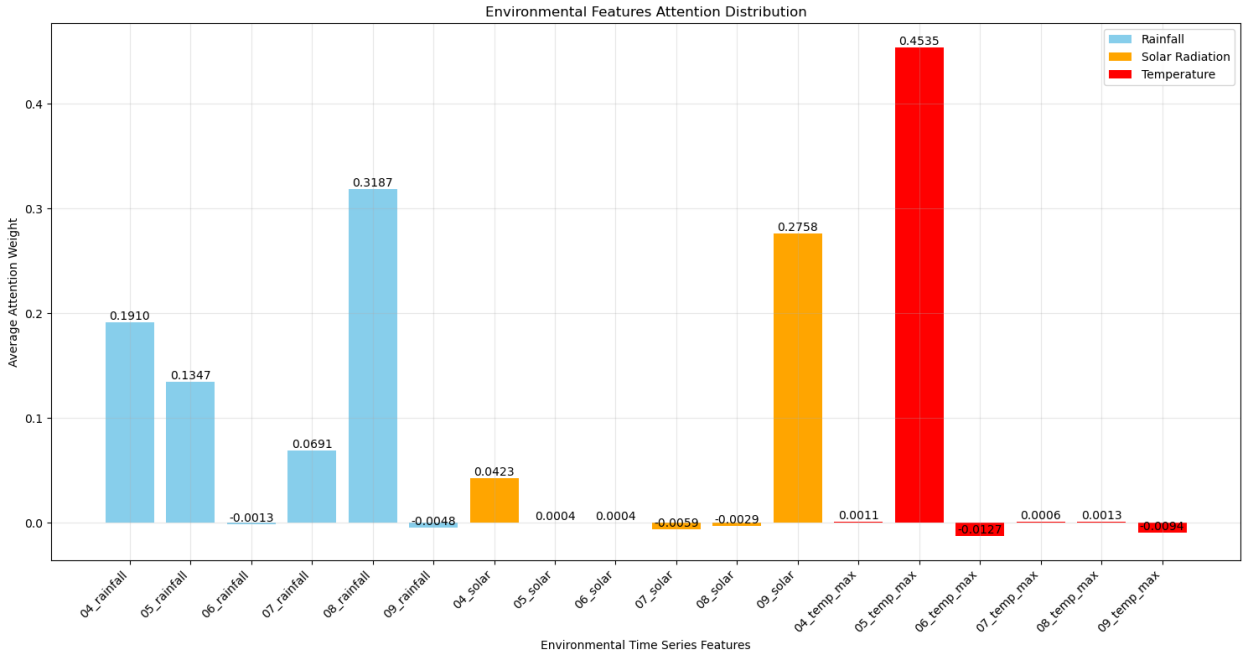

Figure 11: Environmental features based on attention weights

The dominance of attention weight to rainfall could be due to multiple reasons. Firstly, soil moisture is very important for crop growth, especially during the early stages, where consistent and sufficient rainfall is required to ensure soil remains moist; this enables seeds to germinate and establish strong roots, which are vital for the plant's overall health and resilience. Secondly, it is important for photosynthesis as it helps convert sunlight into energy and get adequate water to maintain a high rate of photosynthesis, which is important for crop growth. Thirdly, water plays an important role in transporting nutrients from the soil to different parts of the plant, which are important for the development of plant tissues and, ultimately, for yield formation. This could be the reason why many months received higher weights for rainfall.

5. Project Significance and Value

This thesis contributes to modern agriculture's existing knowledge by integrating deep learning models with Explainable AI. The thesis shows improvements in predicting crop yields by applying deep learning to multimodal data. Further, by utilizing XAI techniques, the study aims to transform the black-box nature of the model into an interpretable one, which is crucial for building trust of stakeholders and interpreting the intricate relationships between genotype and phenotype data.

Integrating multimodal data sources, such as genotype, phenotype, and environment, is a key research component. This comprehensive approach allows for a more nuanced understanding of the factors influencing crop yields. By using CNN, LSTM, and attention mechanisms, this study aims to identify complex patterns and dependencies that simpler models overlook by analyzing these multimodal datasets.

The insights this thesis provides can potentially enhance decision-making processes in agriculture, thereby allowing farmers and agricultural managers to make better decisions about crop management techniques. Furthermore, by comprehending the genetic variables affecting crop yields, breeding efforts can be directed toward developing more resilient and productive crop varieties.

6. Project Milestones

The project timeline was broken down into five major timelines outlining important milestones for completing this research paper. In March, the research began with selecting and approving a research topic. This was followed by a preliminary literature review in April and May to identify key concepts and sources, culminating in a detailed literature review by the end of May, followed by the presentation and submission of the proposal.

Subsequently, a thorough research plan was developed in August, and data analysis was carried out. The final phase focused on report development from September to November, including drafting the report, seeking feedback, and finalizing the report for submission and presentation, as shown in the project timeline below.

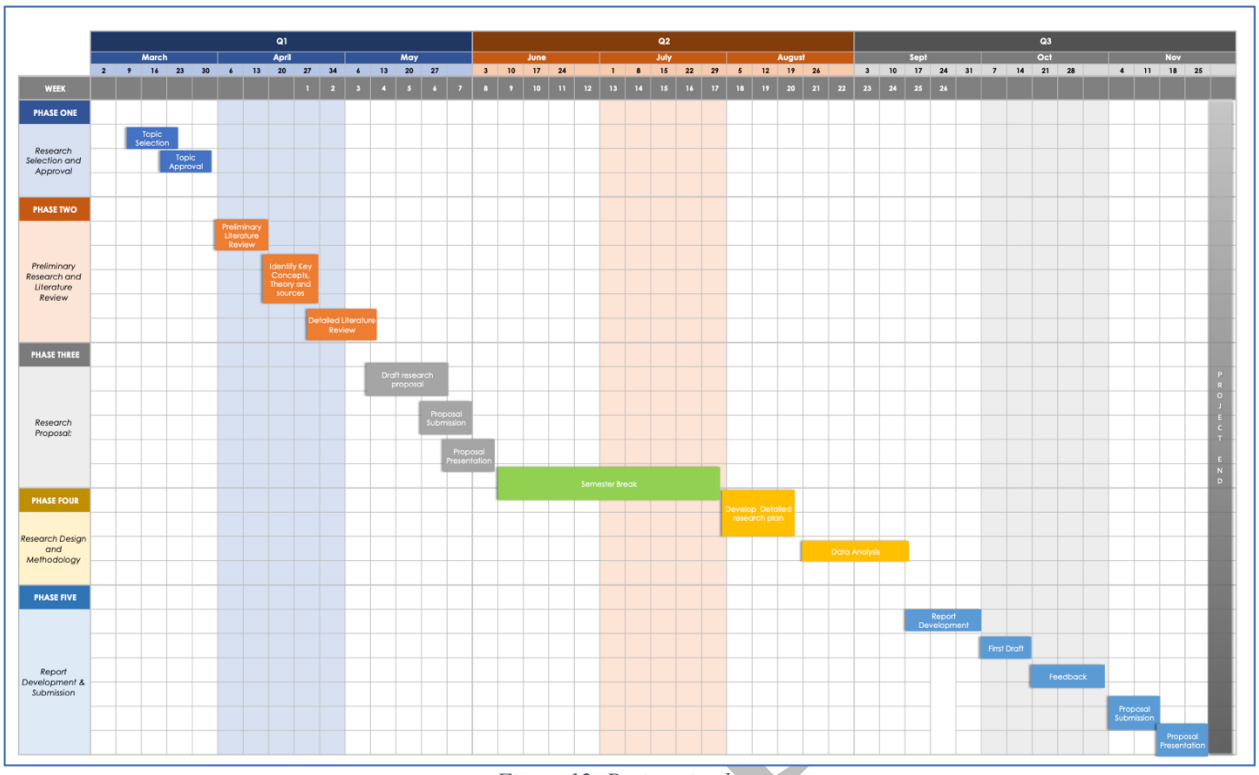

Figure 12: Project timeline

7. Discussion

Implementing attention-based CNN and LSTM models in this thesis gives insights into the interpretation of crop yield predictions, which aligns with the broader spectrum of XAI in agriculture. This hybrid model showed promising efficacy in capturing the interplay between genomic, phenotype, and environmental factors influencing crop yield. CNN was able to extract spatial dependencies, while the LSTM component captured temporal aspects of environmental data, portraying the long-term dependencies that were critical in agricultural prediction. Integrating the attention mechanism in the model allowed us to focus on the important features and time periods in the prediction process.

This interpretability provided by the attention mechanism has identified the features that significantly impact yield predictions. Secondly, the temporal attention patterns have revealed insights into important growth periods that influence crop yield. This information has provided valuable insights that would enable farmers and relevant stakeholders to optimize resource allocation and interventions at key stages of crop development. The explainability provided by the model tried to overcome the black box problem by interpreting the model's internal workings and prediction process. The combination of CNNs, LSTMs, and attention mechanisms offered a framework for predicting crop yields and understanding the underlying factors driving these predictions.

Even though the baseline model's performance is superior to the hybrid model, the hybrid model incorporated multimodal data (genomic, phenotype, environmental) into the model, unlike the baseline model, capturing both spatial and temporal dependencies, which are important for understanding hidden patterns and providing accurate predictions. Given these, the hybrid model had a larger impact on the prediction than other models, providing complete, accurate, and reliable results. The results identified the genetic markers 'L1H040041456', 'D5H480634196', and 'L7H406100741' as having the highest impact on the model's predictions. It also highlighted genomics data having the biggest impact on the prediction when compared to environmental and phenotype data, as the top 50 features were from the genomic data. Rainfall and temperature from the environmental data had lower levels of impact, showing a certain level of importance on the predictions.

There are further avenues for future research in implementing attention-based CNN and LSTM models and using attention as XAI to understand insights into the model's decision-making process. Further, investigation on the application of this approach to different crops and testing its generalizability could potentially uncover crop-specific patterns in yield prediction.

8. Conclusion

In agriculture, explainable artificial intelligence is a fast-growing discipline that is transforming our knowledge of and methods for optimizing agricultural production systems. Through a synthesis of recent research and methodologies, this thesis sheds light on the transformative potential of XAI in enhancing predictive models, understanding the intricate interactions between genotype, phenotype, and environmental data, and giving farmers valuable insights for informed decision-making. Furthermore, this thesis creates predictive models that accurately reflect agricultural systems' temporal and geographical dependencies by combining these multimodal data into a deep learning framework.

It explored the application of XAI in agriculture by using attention-based CNN and LSTM models to predict crop yield. It tries to show the intricate interaction between multimodal data influencing crop prediction, offering valuable insights into these complex interactions. Integration of CNN and LSTM enhanced by attention mechanisms has shown to be an alternative approach to handling the multimodal nature of agricultural data. One of the important aspects of this thesis was the model interpretability provided by the attention mechanisms. This allowed the model to focus on the most important features and time periods while ignoring other non-important features, providing valuable insights into the decision-making process. This hybrid model captures the spatial dependencies inherent in genomic and phenotypic data while also capturing the temporal aspects of environmental factors, resulting in more reliable predictions. This ability to identify important genetic markers and phenotypic traits will be useful in guiding farmers and stakeholders in making informed decisions in future research, allowing for more efficient crop management strategies.

In conclusion, by integrating deep learning's predictive capabilities with attention mechanisms' interpretability, this thesis created a model that predicts crop yields and offers insightful information about the underlying factors influencing these predictions. There is also huge potential for agricultural practices to be revolutionized through the convergence of XAI, sophisticated deep

learning algorithms, and multimodal data integration. More research and innovation in XAI are necessary to utilize innovation in agriculture fully.

9. Acknowledgment

I would like to thank my academic supervisor, Dr. Guanjin Wang, for her continuous effort, patience, and guidance. While undertaking this thesis, I faced many personal challenges that were very difficult to overcome; however, with her constant support and motivation, I was able to keep moving forward and finalize this thesis. Without her unwavering support, this thesis would not have been completed.

Reference

[1] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine Learning in Agriculture: A Comprehensive Updated Review," *Sensors*, vol. 21, no. 11, Art. no. 11, Jan. 2021, doi: 10.3390/s21113758.

[2] D. Minh, H. X. Wang, Y. F. Li, and T. N. Nguyen, "Explainable artificial intelligence: a comprehensive review," *Artif. Intell. Rev.*, vol. 55, no. 5, pp. 3503–3568, Jun. 2022, doi: 10.1007/s10462-021-10088-y.

[3] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[4] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," Mar. 02, 2017, *arXiv*: arXiv:1702.08608. doi: 10.48550/arXiv.1702.08608.

[5] C. Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," Sep. 22, 2019, *arXiv*: arXiv:1811.10154. doi: 10.48550/arXiv.1811.10154.

[6] T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Comput. Electron. Agric.*, vol. 177, p. 105709, Oct. 2020, doi: 10.1016/j.compag.2020.105709.

[7] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics*, vol. 8, no. 8, Art. no. 8, Aug. 2019, doi: 10.3390/electronics8080832.

[8] J. Tang, D. Zhang, and Y. Wang, "Research on Multi-Sensor Task Planning Algorithms for Air Defense Operations," in *2019 IEEE International Conference on Unmanned Systems (ICUS)*, Oct. 2019, pp. 637–643. doi: 10.1109/ICUS48101.2019.8995946.

[9] M. T. Batte, "National Research Council. Precision Agriculture in the 21st Century: Geospatial and Information Technologies in Crop Management. Washington DC: National Academy Press, 1997, 168 pp., $39.95," *Am. J. Agric. Econ.*, vol. 81, no. 3, pp. 755–756, 1999.

[10] "World Prehistory: A Brief Introduction." Accessed: Oct. 12, 2024. [Online]. Available: https://www.researchgate.net/publication/370876226_World_Prehistory_A_Brief_Introduction

[11] "Pillar of Sand. Can the Irrigation Miracle Last?: S. Postel; WW Norton & Company, New York, and World Watch Institute, 1999, 320 pages, paperback, ISBN 0-393-31937-7, $13.95 | Request PDF," *ResearchGate*, Accessed: Oct. 12, 2024. [Online]. Available: https://www.researchgate.net/publication/4748565_Pillar_of_Sand_Can_the_Irrigation_Miracle_Last_S_Postel_WW_Norton_Company_New_York_and_World_Watch_Institute_1999_320_pages_paperback_ISBN_0-393-31937-7_1395
[12] M. Hossen, N. Fahad, R. Sarkar, and M. Ruhani, "Artificial Intelligence in Agriculture: A Systematic Literature Review," *Turk. J. Comput. Math. Educ. TURCOMAT*, vol. 14, pp. 137–146, Jan. 2023.
[13] "Machine Learning Applications for Precision Agriculture: A Comprehensive Review | IEEE Journals & Magazine | IEEE Xplore." Accessed: Oct. 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9311735
[14] O. Montesinos-López *et al.*, "A review of deep learning applications for genomic selection," *BMC Genomics*, vol. 22, Jan. 2021, doi: 10.1186/s12864-020-07319-x.
[15] A. Wolanin *et al.*, "Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt," *Environ. Res. Lett.*, vol. 15, Feb. 2020, doi: 10.1088/1748-9326/ab68ac.
[16] "Learning Deep Features for Discriminative Localization." Accessed: Oct. 12, 2024. [Online]. Available: https://www.computer.org/csdl/proceedings-article/cvpr/2016/8851c921/12OmNqIhFR6
[17] V. Hassija *et al.*, "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence," *Cogn. Comput.*, vol. 16, Aug. 2023, doi: 10.1007/s12559-023-10179-8.
[18] Talal A, Mohd Soperi Mohd Zahid, and Waleed Ali, "A Review of Interpretable ML in Healthcare: Taxonomy, Applications, Challenges, and Future Directions." Accessed: Oct. 12, 2024. [Online]. Available: https://www.mdpi.com/2073-8994/13/12/2439
[19] "Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author) | Semantic Scholar." Accessed: Oct. 12, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Statistical-Modeling%3A-The-Two-Cultures-(with-and-a-Breiman/e5df6bc6da5653ad98e754b08f63326c2e52b372
[20] "Frontiers | Explainable deep learning in plant phenotyping." Accessed: Oct. 12, 2024. [Online]. Available: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1203546/full
[21] M. W. Craven and J. W. Shavlik, "Using Sampling and Queries to Extract Rules from Trained Neural Networks," in *Machine Learning Proceedings 1994*, W. W. Cohen and H. Hirsh, Eds., San Francisco (CA): Morgan Kaufmann, 1994, pp. 37–45. doi: 10.1016/B978-1-55860-335-6.50013-1.
[22] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," Mar. 09, 2015, *arXiv*: arXiv:1503.02531. doi: 10.48550/arXiv.1503.02531.
[23] "Predictive Learning via Rule Ensembles." Accessed: Oct. 12, 2024. [Online]. Available: https://www.researchgate.net/publication/23418298_Predictive_Learning_via_Rule_Ensembles
[24] "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation | PLOS ONE." Accessed: Oct. 12, 2024. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0130140

[25] I. Covert, S. Lundberg, and S.-I. Lee, “Explaining by Removing: A Unified Framework for Model Explanation,” May 13, 2022, *arXiv*: arXiv:2011.14878. doi: 10.48550/arXiv.2011.14878.
[26] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier,” Aug. 09, 2016, *arXiv*: arXiv:1602.04938. doi: 10.48550/arXiv.1602.04938.
[27] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” Dec. 03, 2019, *arXiv*: arXiv:1610.02391. doi: 10.48550/arXiv.1610.02391.
[28] H. Tian, P. Wang, K. Tansey, J. Wang, W. Quan, and J. Liu, “Attention mechanism-based deep learning approach for wheat yield estimation and uncertainty analysis from remotely sensed variables,” *Agric. For. Meteorol.*, vol. 356, p. 110183, Sep. 2024, doi: 10.1016/j.agrformet.2024.110183.
[29] “Obtaining genetics insights from deep learning via explainable artificial intelligence | Nature Reviews Genetics.” Accessed: Oct. 12, 2024. [Online]. Available: https://www.nature.com/articles/s41576-022-00532-2
[30] J. Zhou and O. Troyanskaya, “Predicting effects of noncoding variants with deep learning-based sequence model,” *Nat. Methods*, vol. 12, Aug. 2015, doi: 10.1038/nmeth.3547.
[31] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps,” Apr. 19, 2014, *arXiv*: arXiv:1312.6034. doi: 10.48550/arXiv.1312.6034.
[32] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” Nov. 28, 2013, *arXiv*: arXiv:1311.2901. doi: 10.48550/arXiv.1311.2901.
[33] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, “Striving for Simplicity: The All Convolutional Net,” Apr. 13, 2015, *arXiv*: arXiv:1412.6806. doi: 10.48550/arXiv.1412.6806.
[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” Dec. 03, 2019, *arXiv*: arXiv:1610.02391. doi: 10.48550/arXiv.1610.02391.
[35] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS’17. Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 4768–4777.
[36] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[37] “Research on Factors Affecting Global Grain Legume Yield Based on Explainable Artificial Intelligence.” Accessed: Oct. 12, 2024. [Online]. Available: https://www.mdpi.com/2077-0472/14/3/438
[38] “Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield - ScienceDirect.” Accessed: Oct. 12, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0168192323001508
[39] “XAI for Early Crop Classification | IEEE Conference Publication | IEEE Xplore.” Accessed: Oct. 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10281498

[40] "Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making | Neural Computing and Applications." Accessed: Oct. 12, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s00521-023-09391-2
[41] R. A. Schwalbert, T. Amado, G. Corassa, L. P. Pott, Pvv. Prasad, and I. A. Ciampitti, "Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil," *Agric. For. Meteorol.*, vol. 284, pp. 107886-, 2020, doi: 10.1016/j.agrformet.2019.107886.
[42] "Explainable Artificial Intelligence for Cotton Yield Prediction With Multisource Data." Accessed: Oct. 12, 2024. [Online]. Available: https://www.researchgate.net/publication/373029971_Explainable_Artificial_Intelligence_for_Cotton_Yield_Prediction_with_Multisource_Data
[43] A. L. Chandra, S. V. Desai, W. Guo, and V. N. Balasubramanian, "Computer Vision with Deep Learning for Plant Phenotyping in Agriculture: A Survey," Jun. 18, 2020, *arXiv*: arXiv:2006.11391. doi: 10.48550/arXiv.2006.11391.
[44] A. Krogh Mortensen, S. Skovsen, H. Karstoft, and R. Gislum, "The Oil Radish Growth Dataset for Semantic Segmentation and Yield Estimation," Jun. 2019, pp. 2703–2710. doi: 10.1109/CVPRW.2019.00328.
[45] R. K. Varshney *et al.*, "Fast-forward breeding for a food-secure world," *Trends Genet.*, vol. 37, no. 12, pp. 1124–1136, Dec. 2021, doi: 10.1016/j.tig.2021.08.002.
[46] A. Wolanin *et al.*, "Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt," *Environ. Res. Lett.*, vol. 15, Feb. 2020, doi: 10.1088/1748-9326/ab68ac.
[47] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48–62, Sep. 2021, doi: 10.1016/j.neucom.2021.03.091.
[48] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 02, 2023, *arXiv*: arXiv:1706.03762. Accessed: Nov. 05, 2024. [Online]. Available: http://arxiv.org/abs/1706.03762
[49] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the Visualization of What a Deep Neural Network Has Learned," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2660–2673, Nov. 2017, doi: 10.1109/TNNLS.2016.2599820.
[50] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning Important Features Through Propagating Activation Differences," Oct. 12, 2019, *arXiv*: arXiv:1704.02685. doi: 10.48550/arXiv.1704.02685.
[51] A. T. Keleko, B. Kamsu-Foguem, R. H. Ngouna, and A. Tongne, "Health condition monitoring of a complex hydraulic system using Deep Neural Network and DeepSHAP explainable XAI," *Adv. Eng. Softw.*, vol. 175, p. 103339, Jan. 2023, doi: 10.1016/j.advengsoft.2022.103339.
[52] N. Rodis, C. Sardianos, P. Radoglou-Grammatikis, P. Sarigiannidis, I. Varlamis, and G. T. Papadopoulos, "Multimodal Explainable Artificial Intelligence: A Comprehensive Review of Methodological Advances and Future Research Directions," Jun. 30, 2024, *arXiv*: arXiv:2306.05731. Accessed: Nov. 09, 2024. [Online]. Available: http://arxiv.org/abs/2306.05731
[53] S. Jain and B. C. Wallace, "Attention is not Explanation," May 08, 2019, *arXiv*: arXiv:1902.10186. Accessed: Nov. 09, 2024. [Online]. Available: http://arxiv.org/abs/1902.10186

[54] “Multimodal Intelligence: Representation Learning, Information Fusion, and Applications | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Oct. 12, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9068414
[55] M. Togninalli *et al.*, “Multi-modal deep learning improves grain yield prediction in wheat breeding by fusing genomics and phenomics,” *Bioinformatics*, vol. 39, May 2023, doi: 10.1093/bioinformatics/btad336.
[56] M. Pawłowski, A. Wróblewska, and S. Sysko-Romańczuk, “Effective Techniques for Multimodal Data Fusion: A Comparative Analysis,” *Sensors*, vol. 23, no. 5, p. 2381, Feb. 2023, doi: 10.3390/s23052381.
[57] Abelardo Montesinos-López,* Osval A. Montesinos-López,†,1 Daniel Gianola,‡ José Crossa,§,1 and Carlos M. Hernández-Suárez, “Multi-environment Genomic Prediction of Plant Traits Using Deep Learners With Dense Architecture,” *ResearchGate*, Sep. 2024, doi: 10.1534/g3.118.200740.
[58] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” Jul. 25, 2019, *arXiv*: arXiv:1907.10902. Accessed: Nov. 07, 2024. [Online]. Available: http://arxiv.org/abs/1907.10902