Attention-based Crop Yield Prediction

Pfennigbauer Johannes, 11902046

January 20, 2025

1 Introduction

In agricultural decision-making, predicting crop yield is essential for optimizing food production, land management, and economic planning. Accurate crop yield predictions help farmers and policymakers make informed decisions, ensuring food security and economic stability. However, predicting crop yield is challenging due to the complex interplay of various factors such as weather, soil conditions, and management practices.

Machine learning models, particularly deep learning approaches, have enabled more accurate and scalable crop yield predictions by integrating spatial and temporal data. A well-established method, the CNN-RNN framework developed by Khaki et al. [1], has demonstrated strong predictive capabilities using a combination of weather, soil and management data. However, existing models often overlook the varying importance of different features over time, particularly in relation to plant growth stages.

This project aims to address this problem by improving the CNN-RNN model through the incorporation of an attention mechanism. The attention mechanism allows the model to dynamically prioritize the weather attributes that are most critical to crop yield at different growth stages. By focusing on key periods such as seeding and early development, when weather conditions have a significant impact on crop establishment and future yield potential, the model will be better equipped to predict yields more accurately.

Furthermore, the project will extend the method by adding explainability to justify its decisions and provide a deeper understanding of the model itself. This will offer insights into how the attention mechanism contributes to the prediction process, thereby validating whether the improvements are due to the model's increased focus on key weather periods. By enhancing the model's interpretability, stakeholders can gain confidence in the predictions and make better informed decisions based on the outputs.

2 Methodology and Implementation

The problem of predicting crop yield is complex due to the interplay of various factors such as weather, soil conditions, and management practices. Deep learning models, particularly those integrating spatial and temporal data, have shown promise in addressing this complexity. [1]

Building on the CNN-RNN structure, a multi-head attention mechanism will be integrated after the W-CNN component, which processes weekly weather data throughout the year. This attention layer will enable the model to dynamically assign greater weight to the most relevant meteorological features during critical periods of crop development, such as the early vegetative stages. Research shows that favorable weather during these growth periods is essential for effective crop establishment, weed suppression, and nutrient uptake, all of which directly contribute to biomass production and yield potential [2]. This design mirrors the approach used by Yi-Ming and Hao [3], where a multi-head attention mechanism was successfully applied in a comparable time-series prediction task to highlight important time steps and features. The attention mechanism in this project is expected to improve the model's ability to predict crop yield by focusing on the most critical meteorological information at each growth stage.

Deep learning is a suitable solution because it can automatically learn complex patterns and relationships from large datasets, making it well-suited for integrating diverse data sources such as weather, soil, and management practices. The CNN-RNN framework, enhanced with attention mechanisms, leverages the strengths of convolutional layers for spatial feature extraction and recurrent layers for temporal sequence modeling. The attention mechanism further refines this by allowing the model to focus on the most relevant features at different times, improving prediction accuracy.

However, the CNN-RNN network faced several challenges due to a lack of adherence to standard coding principles. The original codebase was poorly structured, relied on TensorFlow 1.x, and it was not possible to run the code as is. To address these issues, the model was completely reimplemented in TensorFlow 2, ensuring compatibility with the latest frameworks and improving code clarity and maintainability.

During reimplementation, significant effort was put into structuring the code to improve readability and usability. Custom classes were defined for the CNN-RNN models, containing well-organized methods for training, prediction, evaluation, and explanation. This modular design not only streamlined experimentation but also facilitated future extensions to the model.

To improve model performance, an attention mechanism was incorporated into the framework. This mechanism aimed to highlight the most critical temporal and feature-specific components of the input data. This was done in two steps. First, a standard attention layer was incorporated directly into the W-CNN model after the first convolution to guide the models to focus on the key weeks of the growing season for each weather feature. Subsequently, a multi-head attention layer was integrated after concatenating all W-CNN outputs to assist the model in extracting the most relevant weather attributes.

The Layer-wise Relevance Propagation (LRP) [4] technique was initially considered to address the need for explainability. However, due to the unavailability of TensorFlow 2 compatible libraries for LRP, an alternative method, Local Interpretable Model-agnostic Explanations (LIME) [5], was adopted. The implementation of LIME was straightforward and allowed for a detailed analysis of the importance of the features. This method provided valuable information on the behavior of the model, including the identification of critical temporal features and the relative importance of different input attributes.

3 Results

The performance of the implemented models was evaluated using Root Mean Square Error (RMSE) and R^2 (coefficient of determination) as the primary metrics. RMSE was chosen to quantify the average size of prediction errors, with particular emphasis on penalising larger errors, which is critical for regression tasks. This metric was also used in the original paper and therefore allows for comparison. R^2 was used to measure the proportion of variance in the target variable explained by the model, providing a complementary perspective on performance.

The initial goal was to exceed the RMSE of the original paper, which was 4.32 for the validation year 2017. However, due to significant adjustments required in the model definition and data pre-processing, as well as computational constraints, this benchmark was not expected to be achieved. Instead, the target was adjusted to improve the implemented version of the original model by at least 10%, which translated into achieving an RMSE below 4.759 on the validation data. Although the multi-head attention model improved the model only slightly, the lighter model using only the standard attention layers proved very beneficial to model performance in most experiments. It achieved an RMSE of 4.610, an improvement of almost 13%. Table 1 shows a summary of the evaluation metrics.

Model	**Training RMSE**	**Validation RMSE**	**Training R^2**	**Validation R^2**
Original Model	4.125	5.288	0.845	0.684
Standard Attention Model	4.243	4.610	0.836	0.760
Multi-Head Attention Model	4.297	5.190	0.831	0.695

Table 1: Model Performance Comparison on Training and Validation data.

Across all experiments, the Standard Attention Model consistently outperformed the baseline CNN-RNN model on validation data, achieving better RMSE and R^2 values. Conversely, the Multi-Head Attention Model exhibited signs of overfitting, performing better on training metrics but worse on validation metrics compared to the Standard Attention Model. Interestingly, the original model occasionally performed as well as, or better than, the attention-based approaches, underscoring the robustness of the baseline architecture. Notably, the Standard Attention Model exhibited remarkable resistance to overfitting, with minimal differences observed between training and validation performance in most experiments.

The comparison of feature importance between the *original* and *standard attention* models using the LIME explainability method highlights the dominance of the weather and soil feature groups. Both models attribute the greatest impact to the weather group. Interestingly, the attention-based model gives it slightly *less* importance ($\sim 52\%$).

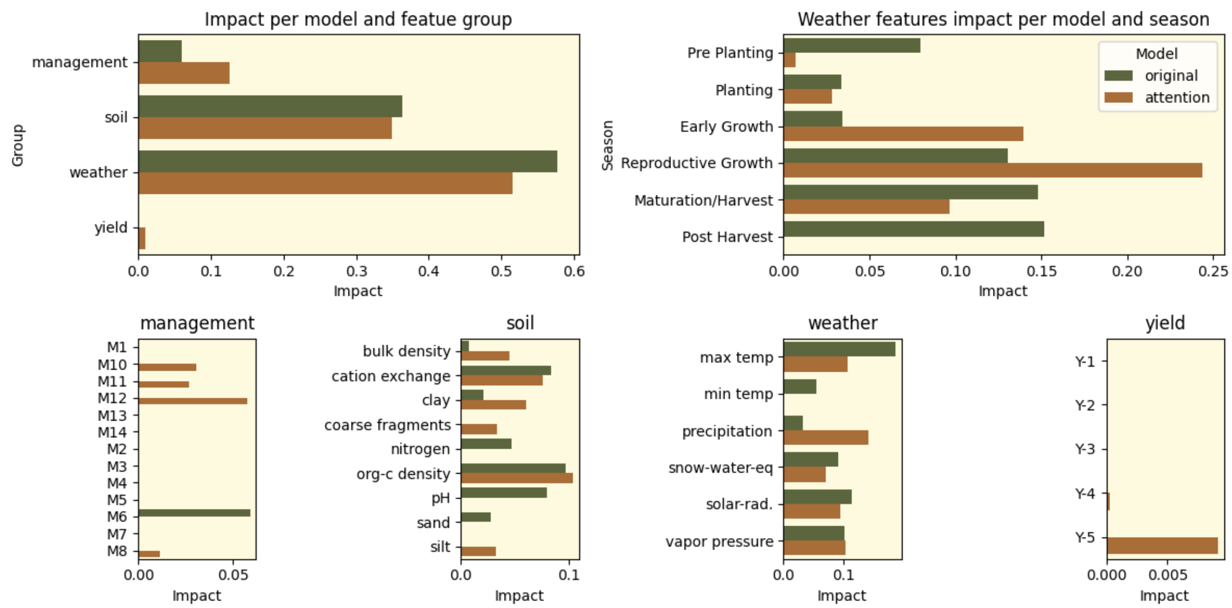

Figure 1: Comparison of the importance of feature groups, subgroups and single features in the original and standard-attention model.

However, when looking at the seasonal difference, it can be seen that the attention model gives much more importance to the growing seasons, while the original model also focuses on the post-harvest and pre-planting weeks, which can be assumed to be more noisy than useful from a practical perspective. This behaviour is intentional, as the attention mechanism was specifically implemented to focus on key weeks of the growing season. The soil group shows a similar importance in both models at around 35%, while the management and yield characteristics have a marginal impact overall, with the attention model slightly increasing the importance of management characteristics to around 15%. Details are shown in Figure 1.

4 Conclusion

The implementation of this project has been both challenging and rewarding, requiring the reengineering of an outdated CNN-RNN framework to meet modern standards. The integration of attention mechanisms yielded mixed results. While the Standard Attention Model achieved the goal by showing a 13% improvement in validation RMSE over the baseline, the Multi-Head Attention Model did not meet expectations and exhibited signs of overfitting.

One of my main takeaways from this project is the importance of carefully selecting and reviewing the code from the original paper. Ensuring that the code is well structured and compatible with modern frameworks is crucial for successful implementation. In addition, it is important to be defensive about time estimates, as unexpected challenges can arise during the reengineering process. Personally, I completely underestimated these challenges, which led to the project taking much longer than expected. However, other things, especially report writing and presentation, took much less time than I had originally estimated. All in all, I spent approximately 140 hours on the project, which is slightly more than I thought I would.

Despite these challenges, the project provided very valuable insights, and I learned a lot, probably more than from any other project so far. The hands-on experience in enhancing and analyzing a complex model has been incredibly educational. I would definitely undertake this project again and want to continue working on this topic. I personally like that I have the feeling that this could have a significant practical impact once the approach and model are mature enough.

References

[1] Saeed Khaki, Lijuan Wang, and Sotirios V. Archontoulis. A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science*, 10, 2020.

[2] Christine J. Butts-Wilmsmeyer, John R. Seebauer, Leslie Singleton, and Fred E. Below. Weather during key growth stages explains grain quality and yield of maize. *Agronomy*, 9(1):16, 2019.

[3] Zhang Yi-Ming and Wang Hao. Multi-head attention-based probabilistic cnn-bilstm for day-ahead wind speed forecasting. *Energy*, 278, Part A, 2023.

[4] Imran Ullah, Adrian Rios, Vatsal Gala, and Susan Mckeever. Explaining deep learning models for tabular data using layer-wise relevance propagation. *Applied Sciences*, 12(1):136, 2022.

[5] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ”why should i trust you?”: Explaining the predictions of any classifier, 2016.