Project Report

Designing a Neural Network-Based Decision Support System for Optimal Crop Cultivation

Team Members:

Suprit Shivshanthkumar Patil (G37575767)

Omkar Balekundri (G41085894)

Yash Manish Bobde (G20377188)

1. Problem Statement

While looking into agricultural issues, we came across this issue that there is a lack of guidance for farmers when it comes to selecting the most suitable crops for their land based on the soil, weather parameters and market trends. Due to this problem, it results in lower yields and potential financial losses for the farmer. For instance, if a farmer owns land with acidic soil and resides in an area with minimal rainfall, it is crucial to choose crops that thrive in such conditions. In this scenario, our system would recommend crops like potatoes, which is suitable for acidic soil and require less water. Additionally, our system considers market prices and trends to assist farmers in making informed decisions about which crops will yield the best financial return. To address this issue, we decided to develop a crop recommendation system. This system will take into account a range of factors such as soil parameters, weather conditions and market price. By analysing these factors collectively, the main aim is to provide farmers with personalized optimized recommendations based on their specific agricultural environments.

2. Literature Survey

There has been various research on crop prediction using Neural networks. We have studied various research in crop prediction where they have incorporated neural networks. In[1] this research paper they have proposed five machine learning models: Random Forest, Support vector machine, XGBoost, Decision Tree, and KNN for recommending crops based on soil pH, NPK, and three climatic variables: rainfall, temperature, and humidity. They have sourced their data from Kaggle repository. They used individual sets of 10 Horticultural and 11 agricultural crops. They found that using individual datasets for each crop category generated better outcomes than combining these two crop categories. They studied how different crops require a different range of NPK, rainfall, temperature, humidity, and soil pH. For example, mung beans have the ability to grow in wide range of soil pH levels whereas typical agricultural crops require a pH range of 6-8. Among all the five models, the XGBoost had the highest accuracy level and they demonstrated how it is better than the other models. We found that soil pH, NPK, and three climatic variables: rainfall, temperature, and humidity parameters can be used in our neural network model.

The [2] paper introduces a Concurrent Excited Gated Recurrent Unit (CEGRU) which is based on deep learning. They have trained and tested this model with Kaggle datasets, from Andhra Pradesh (Chittoor district) and Maharashtra dataset. Hunter-prey optimization is introduced for tuning the parameters. The model achieves an overall accuracy of 99.2%. The CEGRU uses various performance metrics like kappa, PPV, RMSE and accuracy. They conclude that the proposed model is compared with RNN, DNN, CNN, LSTM, and GRU models and has a better accuracy performance at minimal time complexity. This paper described the use of CNN and RNN in crop recommendation system.

The [3] research paper focuses on feasibility of crop type prediction using a one-dimensional convolutional neural network (1D CNN) and decision tree algorithm. The training data used consists of historical Cropland Data Layer (CDL) of Cass County as the study area from 2008 to 2020 that has crop types, their spatial distribution over time, acreage, yield data, cropping season information, rainfall data, and other environmental factors relevant to crop growth and cultivation. The 1D CNN consists of a hidden convolutional layer followed by fully connected layers. The convolutional layer uses a 2-size kernel with ReLU activation. To prevent overfitting, a dropout of 0.2 is applied after convolution. The model includes two fully connected layers that use ReLU and SoftMax activations to predict probability. The Adam algorithm is used to optimize accuracy after verifying it with sparse categorical cross-entropy. Using historical data, the 1D CNN model accurately predicts Cass County's crop map in 2021. While corn and soybean predictions were extremely accurate, alfalfa predictions were less accurate.

The [4] research paper proposes a novel approach to predict the yield from multi-spatial images by using the dimensionality reduction method and the 3D convolutional neural network. The test data consists of remote sensing data from MODIS satellite, cropland data from the USDA website, and yearly average soybean yield at the county level. The model converts multi-spectral images into histogram matrices. These histograms are then fed into a 3D CNN with ReLU as activation function. To ensure counties with similar crop yield share similar learned features, multitask learning is used. Experimental results demonstrate superior performance compared to existing methods, with average RMSE improvements ranging from 2.7% to 31.9%.

In this [5] research study, they have implemented Artificial neural Network (ANN) for recommending a crop based on soil texture, temperature, rainfall, humidity, sunshine hours, wind speed, precipitation, and potential Evo transpiration. The data was sourced from various sources: Agro meteorology Section, University of Agricultural Sciences, Bengaluru (For the period – 2007 to 2017), National Bureau of Soil Survey and Soil Usage Planning (NBSS & LUP), Bengaluru. They have focused mainly on four crops: rice, finger millet, sugarcane, and maize. They have proposed a four-layered ANN architecture with 1 input layer with 6 neurons, 2 hidden layers with 5 neurons, and 1 output layer with 4 neurons. There are 4 classes: 1. Highly suitable, 2. Moderately suitable, 3. Marginally suitable, 4. Not suitable. Their model predicts what is the class of the crop depending upon the location and other features. They concluded that ANN performs better than the decision tree model in all the crops.

The [6] research paper proposes a hybrid CNN-RNN model consisting of convolutional neural networks (CNNs) and recurrent neural networks (RNNs). The training data used consists of four sets that are yield performance, management, weather, and soil. The W-CNN model uses one-dimensional convolution to capture the time-dependent patterns of weather data, whereas the S-CNN model uses one-dimensional convolution to capture the geographic dependencies of soil data collected at various depths underground. A fully connected layer was used to reduce the dimension of the output of the CNN models. The designed RNN model captures the time dependencies of crop yield over a number of years. Further, the results from the proposed design were compared with other three models that were random forest (RF), deep fully connected neural network (DFNN), the least absolute shrinkage and selection operator (LASSO). The research paper showcases that the test data taken for three validation years and the yields of corn and soybean for these years were predicted. The hybrid CNN-RNN model is better than others in predicting corn and soybean yields.

The [7] research paper focuses on predicting crop yield performance by using deep neural networks. The data used in the research paper was provided by the 2018 Syngenta Crop Challenge, which included three sets: crop genotype, yield performance, and environment (weather and soil) from multiple locations in the United States and Canada from 2008 to 2016. For weather prediction, the paper proposes a neural network with 4 input variables and root-mean-square-error (RMSE) as loss function. The input variable denotes the weather variable ‘w’ at location ‘l’ in year ‘y’, which had 72 different variables as per the data set used. For yield prediction two deep neural networks (21 hidden layers and 50 neurons in each layer), one for yield and the other for check yield, and then used the difference of their outputs as the prediction for yield difference. Further, the research paper compares the DNN with three other prediction models that were the least absolute shrinkage and selection operator (Lasso), shallow neural network (SNN) (having a single hidden layer with 300 neurons), and regression tree. DNN outperformed due to the ability to capture nonlinear effects, while Lasso's linear structure led to weaker performance. SNN surpassed Lasso in most measures. Regression tree performed comparably with SNN but worse on yield difference. DNN excelled in predicting yield and check yield but struggled slightly with yield differences.

3. Data Source

Data is a crucial part of our project to determine the crop. From our research, we selected 10 features that might influence the decision-making process for crop recommendation. The features are Soil parameters (Nitrogen, Phosphorus, Potassium), Humidity, Temperature, pH of the soil, Rainfall, Price of the crop ($/ CWT) where CWT is hundredweight and Year and Month of the Price of the crop.

We searched the internet to gather the data related to these features. We found two resources which can provide us with the data we need.

Data sourced from:

1. United States Department of Agriculture → https://quickstats.nass.usda.gov/
2. Crop recommendation data - Harvard Dataverse → https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/4GBWFV

The website of United States Department of Agriculture (1) contains the dataset of the crops grown in USA and it provided us with the Price of the crop ($/CWT) for specific month and year. The Harvard Dataverse (2) has optimum values of Nitrogen, Phosphorus, Potassium, Humidity, Temperature, pH of the soil, Rainfall and a label that indicates the crop. We combined these two datasets to form a single dataset which has Nitrogen, Phosphorus, Potassium, Humidity, Temperature, pH of the soil, Rainfall, Price of the crop ($/ CWT) and a label of the crop.

We got the data of only four crops: Rice, Beans, Soyabeans and apple, which have records of price received for each month and year.

As the Crop recommendation data - Harvard Dataverse [2] have optimum values of crop, we have clubbed these values with each month and year data of CWT price and created a custom dataset.

4. Proposed design and architecture of Neural network model

The following is the proposed design:

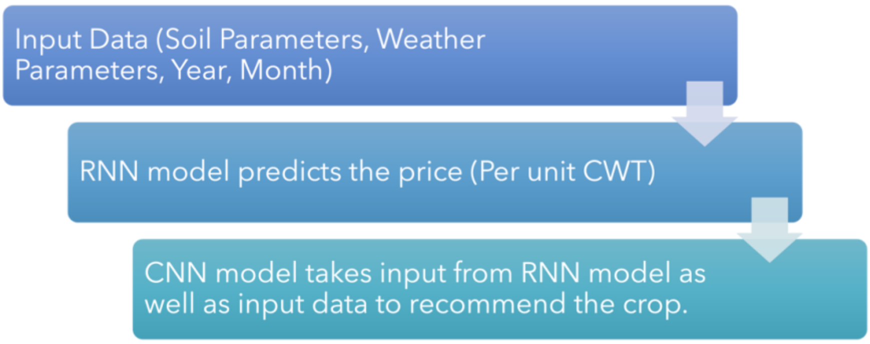
Data processing: In this project, the dataset employed is named dataset.csv. Categorical variables like 'Period' and 'label' have undergone label encoding, converting them into a format readable by machines. Furthermore, a stratification strategy was employed to divide the dataset into training, validation and testing sets, ensuring that each split accurately reflects the overall distribution of the labels i.e. crop labels.

The preparation of input data for the RNN model includes padding and sequencing. The sequence length was set to 12, indicating that each input sequence fed into the RNN consists of 12 consecutive time steps. To ensure that all sequences in our training, validation, and testing datasets are of the same length. We used padding to the data arrays. Specifically, if a sequence in the dataset had fewer than 12 time steps available, it was padded with zeros at the start to meet the required length. Adding zeros in the sequences does not affect the training or prediction of the RNN model. The model will learn from exposure to the data that the value 0 means missing data and will start ignoring the value[8].

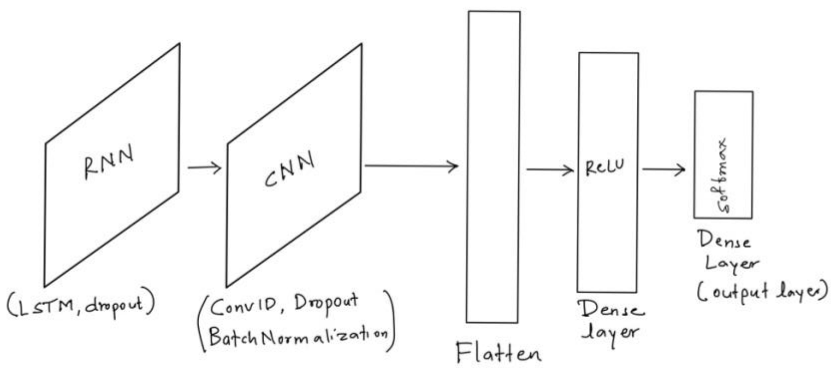
For the CNN input, the predicted price values ($/CWT) from the RNN of sequenced training, validation, and test datasets are expanded to match with the shape of the original split data set that does not have price values ($/CWT) (X_train, X_val and X_test). Further the both the expanded values and the original data splits are merged which will be used as input for the CNN model.

Feature engineering was performed using a set of variables such as 'Year', 'Period', 'N' (nitrogen), 'P' (phosphorus), 'K' (potassium), 'temperature', 'humidity', 'pH', 'rainfall', and 'cwt' (price per cwt). These features were designed to meet the input specifications of the CRNN (Convolutional Recurrent Neural Network). To effectively classify labels, they were reshaped to accommodate the time series aspect for the RNN (Recurrent Neural Network) and CNN.

Below is the flow of data in the proposed model:



Proposed model Architecture:



For our project, we have incorporated both Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). The use of this combined architecture that includes both is an approach for leveraging the distinct strengths of each neural network type. The decision to use a CRNN architecture results from the need to address various aspects of the prediction task in a step-by-step manner, with the output of one model serving as input for the next. This structured learning model enables a broader view of the data and facilitates context-dependent predictions, which are critical for making accurate crop recommendations.

The model architecture starts with using an RNN. It uses Long Short-Term Memory (LSTM) layers which is ideal for capturing temporal dependencies in sequential data. By predicting price ($/weight per hundred pounds (cwt)), the RNN creates an important feature which represents historical market trends over time. This predicted price ($/cwt) is a useful input to the subsequent CNN model. It improves the feature representation and allows for more accurate crop recommendations.

Further, the predictions from the RNN is taken and combined with the other features for the CNN as input data. The CNN model includes layers such as Conv1D, BatchNormalization, and Flatten, indicating that it is intended to process sequential data. Dropout layers are used to prevent overfitting. The final layer employs a softmax activation function, which is suitable for multi-class classification tasks.

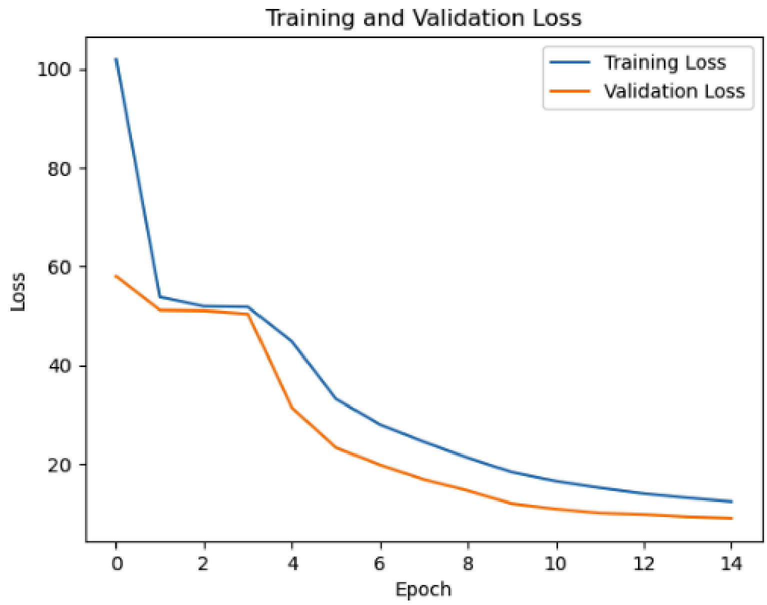
5. Performance and Results

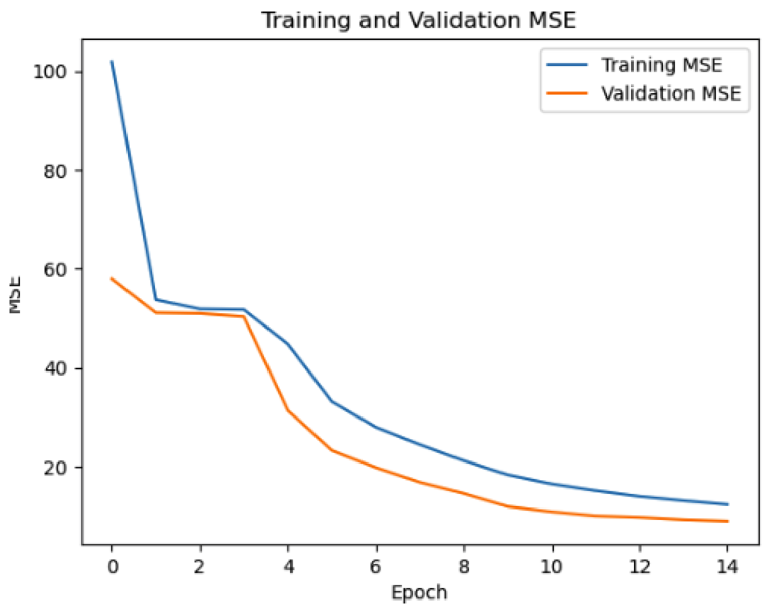
Our study used a Recurrent Neural Network (RNN) to predict the price ($ / hundredweight (CWT)) of agricultural products. The primary goal was to minimise the difference between training and validation losses, indicating a well-generalized model.

The RNN model was trained for 15 epochs, resulting in a significant reduction in both training and validation loss. Initially, the model had a significant training loss of 101.8931 and a validation loss of 57.9815. These values decreased significantly after iterative training, demonstrating the model's increasing accuracy in predicting the CWT value. By the end of the training process, the training loss and validation loss had stabilised at 12.3329 and 8.9666, respectively.

The model was further validated with a separate test dataset, yielding a test loss and mean squared error (MSE) of 9.3950. The similarity of the test loss to the final validation loss highlights the model's robustness and ability to generalise well on previously unseen data. Graphical representations of the training process show a steep decline in initial epochs that gradually plateaus, indicating that the model is nearing its peak performance. The graphs below show the loss and MSE across epochs:

Training and validation dataset results:





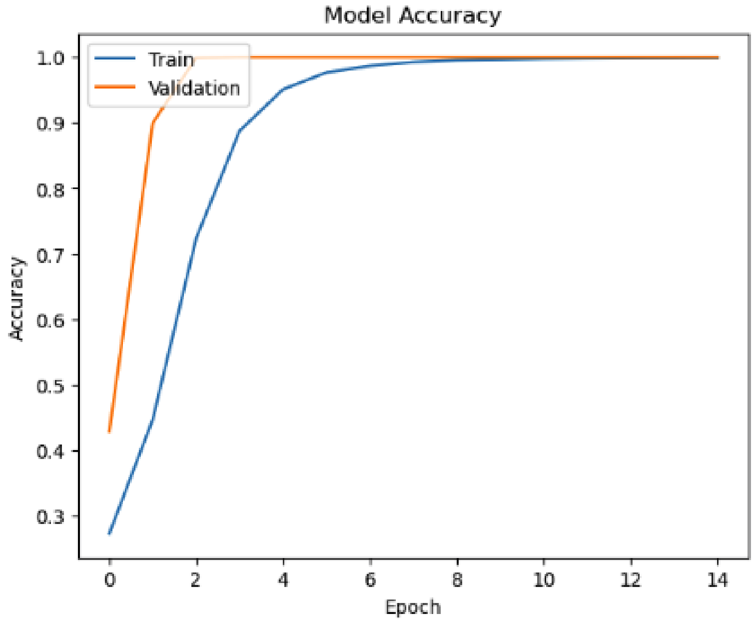
The loss curves gradually decrease and eventually plateau, indicating effective learning, with no significant overfitting observed, as the training and validation loss curves closely follow each other.

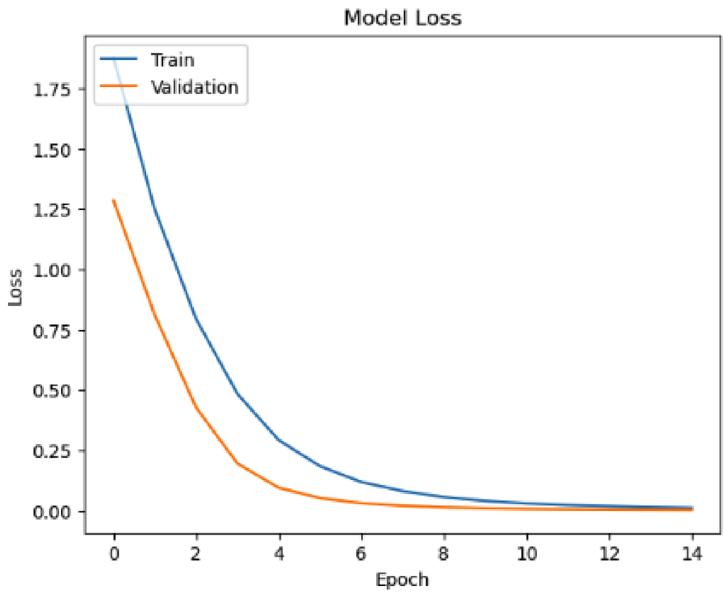
The following section of our work focuses on the use of a Convolutional Neural Network (CNN) model to recommend optimal crops by taking the predicted price ($ / hundredweight (CWT)) from a previous RNN model as input. The CNN model was set up to optimise classification accuracy. The training began with a significant difference in training and validation accuracy which quickly narrowed as the model learned from the data. The initial training accuracy was only 32.92%, while the validation accuracy was 73.49%. This rapid improvement in validation accuracy demonstrates the model's ability to learn biased features effectively.

Throughout the training sessions the CNN model shows consistent improvement eventually reaching a perfect accuracy score of 100% in validation by the fourth epoch. Training accuracy was closely followed by near perfect scores soon after. The training and validation loss followed this pattern with the validation loss dropping to 0.0016 by the final epoch. The CNN model was tested on an unseen test dataset resulting in a test loss of 0.00161 and a test accuracy of 100 percent. This performance showcases the model's robustness and ability to generalise well which is important for practical applications.

The following graphical representations show the accuracy and loss for both the training and validation phases, illustrating the model's progression over the epochs. The graphs show a quick increase in accuracy and a rapid decrease in loss, highlighting the current is powerful enough than the complexity in the relations between features of the dataset.

Training and validation dataset results:





In conclusion, we were able to use Neural Network models to improve recommendations and predictions of crop. The RNN model accurately predicts market values and the CNN uses these predictions to recommend best crops. This dual-model approach not only maximises financial returns, but it also adds a new dimension to crop recommendation.

Further, we used the trained model to predict the price ($ per hundredweight (CWT)) and classify crop for a single sample data. Below are the input data and their results:

a. Crop : Rice and price/CWT: 19.6

```
sample_data = {
    'Year': 2024,
    'Period': 'JAN',
    'N': 73,
    'P': 57,
    'K': 41,
    'temperature': 21.4,
    'humidity': 84.9,
    'ph': 5.8,
    'rainfall': 272.2,
}

predicted_price, predicted_class = pred_crop_price(sample_data, rnn_model, cnn_model, le_period, le_label)

print("\nPredicted Price per CWT:", predicted_price)
print("Predicted Class:", predicted_class)
```

```
1/1 [==============================] - 0s 73ms/step
1/1 [==============================] - 0s 15ms/step

Predicted Price per CWT: 13.806007
Predicted Class: ['rice']
```

b. Crop : Maize and price/CWT: 8.5

```
sample_data = {
    'Year': 2024,
    'Period': 'JAN',
    'N': 71,
    'P': 54,
    'K': 16,
    'temperature': 22.6,
    'humidity': 63.7,
    'ph': 5.7,
    'rainfall': 87.8,
}

predicted_price, predicted_class = pred_crop_price(sample_data, rnn_model, cnn_model, le_period, le_label)

print("\nPredicted Price per CWT:", predicted_price)
print("Predicted Class:", predicted_class)
```

```
1/1 [==============================] - 0s 9ms/step
1/1 [==============================] - 0s 8ms/step

Predicted Price per CWT: 6.3022537
Predicted Class: ['maize']
```

c. Crop : Soyabeans and price/CWT: 21.3

```
sample_data = {
    'Year': 2024,
    'Period': 'JAN',
    'N': 40,
    'P': 72,
    'K': 77,
    'temperature': 17,
    'humidity': 17,
    'ph': 7.5,
    'rainfall': 88.6,
}

predicted_price, predicted_class = pred_crop_price(sample_data, rnn_model, cnn_model, le_period, le_label)

print("\nPredicted Price per CWT:", predicted_price)
print("Predicted Class:", predicted_class)
```

```
1/1 [==============================] - 0s 12ms/step
1/1 [==============================] - 0s 8ms/step

Predicted Price per CWT: 15.498697
Predicted Class: ['Soyabeans']
```

In conclusion, the CNN classifies the crops correctly and is more than powerful for the current dataset. On the other hand, the RNN gives us good price predictions but there is a little bit of difference from the actual market prices. Even after tuning the model to its best and the restrictions of the data set and its preparations. The RNN model needs more detailed market data to enhance the accuracy and predictions. Even though, there is little discrepancies, the RNN and CNN together really helps match up the predicted prices with the best crop choices.

6. Challenges Faced

The main challenge that we faced while developing the model was to gather data. The obstacles we faced were many crops had less documented data respective to price values per month. The soil and weather parameters values are insufficient for the CRNN model training. An ideal dataset for this model is which has soil and weather parameters for specific regions and price ($/CWT) per month with respective to the crops.

7. Future work

The main issue we faced was scarcity of data. Due to this we had to restrict the weather parameters, soil parameters and price data which are co-related on each other. Considering this similar model architecture the model's feature set should be expanded to include more weather parameters like wind speed, solar radiation and temperature extremes. Similarly, a wider range of soil parameters, such as soil texture, organic matter content, soil moisture levels at various depths and compactness of the soil should be considered.

Apart from this, the price data provided by the government agencies is very dynamic, the reports sometime consist of monthly, quarterly or annually. Currently, the data set for the model is restricted by national yield prices and optimised soil and weather parameters of the crops for crop recommendation. For the future research, the model is capable of handling complex relations by increasing the granularity of the data set by including region specific parameters. This will result in more tailored accurate predictions for the farmers and the agricultural advisors.

Additionally, another feature that is essential to add in this architecture is crop rotation. Crop rotation is essential because each type of plant has its own specific needs for nutrients and can be affected by different types of pests and diseases, it's important to customize the way we take care of them. Hence, in future work crop rotation with respective to the region, soil and weather parameters will provide optimised recommendations and predictions to maximise the profit and also maintaining the soil health of the land.

Further, this requires scheduled acquisition of data from farmers and government agencies respective to specific crop growth and region. With the integration of additional weather parameters, soil parameters, price and region data into the model will offer a promising way for future research. This combination will allow for a more improved analysis of the interactions between these parameters and their effects on crop growth, yield and quality. Ultimately, resulting in better accuracy and less loss of model performance in predictions and recommendations.

Another point is the development of decision support application will be crucial for integrating for integrating the models we have developed with user friendly interfaces. These applications will provide intuition and recommendations to farmers and agricultural advisors. To ensure accessibility, scalability and reliability using the decision support system can be achieved through the development of mobile applications and web-based platforms in the coming future.

8. References

[1] Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables [February 2024]
Link: - https://www.sciencedirect.com/science/article/pii/S2405844024011435

[2] Deep learning based concurrent excited gated recurrent unit for crop recommendation based on soil and climatic conditions [January 2024]
Link: - https://link.springer.com/article/10.1007/s11042-023-18004-y

[3] Prediction of Crop Planting Map Using One-dimensional Convolutional Neural Network and Decision Tree Algorithm [July 23]
Link: - https://ieeexplore.ieee.org/document/10233466

[4] Predicting Crop Yield Using 3D Convolutional Neural Network with Dimension Reduction and Metric Learning [August 2023] Link: - https://ieeexplore.ieee.org/document/10331974

[5] Artificial Neural Networks Based Integrated Crop Recommendation System Using Soil and Climatic Parameters [May 2021]
Link:https://www.researchgate.net/publication/352308018_Artificial_Neural_Networks_Base d _Integrated_Crop_Recommendation_System_Using_Soil_and_Climatic_Parameters

[6] A CNN-RNN Framework for Crop Yield Prediction [January 2020] Link: - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6993602/

[7] Crop Yield Prediction Using Deep Neural Networks [May 2019] Link:- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6540942/

[8] Padding reference from Deep Learning with Python by François Chollet
Link: - https://stackoverflow.com/questions/52570199/multivariate-lstm-with-missing-values