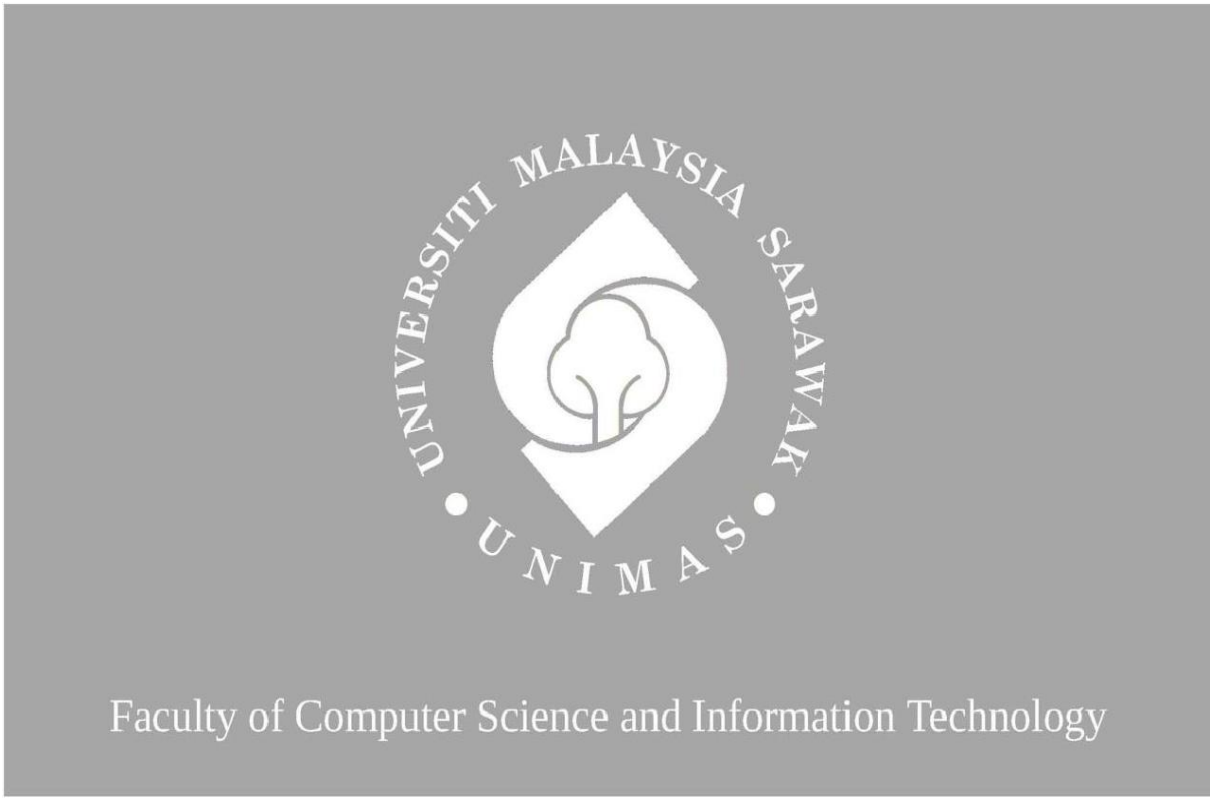

An Intelligent Hybrid Model Using CNN and RNN for Crop Yield Prediction

JUNE KHOO YAN

Bachelor of Computer Science with Honours

(Software Engineering)

2023

AN INTELLIGENT HYBRID MODEL USING CNN AND RNN FOR CROP YIELD PREDICTION

JUNE KHOO YAN

This project is submitted in partial fulfillment of the
requirements for the degree of
Bachelor of Computer Science and Information Technology
(Software Engineering)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2023

MODEL HYBRID PINTAR MENGGUNAKAN CNN DAN RNN UNTUK RAMALAN HASIL TANAMAN

JUNE KHOO YAN

Projek ini merupakan salah satu keperluan untuk Ijazah
Sarjana Musa Sains Komputer dan Taknologi Maklumat
(Kejuruteraan Perisian)

Faculty Sains Komputer dan Teknologi Maklumat
UNIVERSITI MALAYSIA SARAWAK
2023

UNIVERSITI MALAYSIA SARAWAK

THESIS STATUS ENDORSEMENT FORM

TITLE AN INTELLIGENT HYBRID MODEL USING CNN AND RNN FOR CROP YIELD PREDICTION

ACADEMIC SESSION: 2019/2020

I, JUNE KHOO YAN (70037),

(CAPITAL LETTERS)

hereby agree that this Thesis* shall be kept at the Centre for Academic Information Services, Universiti Malaysia Sarawak, subject to the following terms and conditions:

1. The Thesis is solely owned by Universiti Malaysia Sarawak
2. The Centre for Academic Information Services is given full rights to produce copies for educational purposes only
3. The Centre for Academic Information Services is given full rights to do digitization in order to develop local content database
4. The Centre for Academic Information Services is given full rights to produce copies of this Thesis as part of its exchange item program between Higher Learning Institutions [or for the purpose of interlibrary loan between HLI]
5. ** Please tick (√)

	CONFIDENTIAL	(Contains classified information bounded by the OFFICIAL SECRETS ACT 1972)
	RESTRICTED	(Contains restricted information as dictated by the body or organization where the research was conducted)
√	UNRESTRICTED	

June

(AUTHOR'S SIGNATURE)

Validated by

(SUPERVISOR'S SIGNATURE)

Permanent Address

NO 28A, LORONG 3B1, TAMAN SURIA JAYA, BATU 6 JALAN MATANG, 93050, KUCHING, SARAWAK.

Date: 15 JULY 2023

Date: 17 JULY 2023

Note * Thesis refers to PhD, Master, and Bachelor Degree

** For Confidential or Restricted materials, please attach relevant documents from relevant organizations / authorities

DECLARATION

I hereby declare that this project together with all its content is none other than of my work, except for some information taken and extracted from the other resources that have been cite respectively.

(JUNE KHOO YAN)

Faculty of Computer Science and Information

Technology, University Malaysia Sarawak

ACKNOWLEDGEMENT

First, I would like to express my sincere gratitude to my supervisor and course coordinator who has guided me through my final year project. He has given me a great supervision and useful advice and encouragement throughout my Final Year Project completion. I would also thanks to my examiner for his comments and evaluation on my final year project. Besides, I would also thanks to my fellow friends, senior and parents who had given me support and encouragements throughout this project.

ABSTRACT

AI has been successfully applied in agriculture field in terms of prediction, decision making, crops and soil monitoring and analysing. In this study, an intelligent hybrid model using CNN and RNN for crop yield prediction is proposed. The learning model that proposed is the combination of Convolutional Neural Network (CNN) with Recurrent Neural Network (RNN) models. CNN is a popular learning model used in predicting crop yield due to its high performance in feature extraction. CNN algorithm is used in this study due to its characteristic where it considers a smaller number of parameters in the network, and it has a lower chance of overfitting While RNN acts as a prediction model in this study. RNN has the nature of learning, a feedback network and can encode temporal sequence information. Due to the short time memory behaviour of RNN, RNN network is enhanced with LSTM cells which allows them to perform long-term memory tasks. LSTM presents memory blocks in solving the exploding and vanishing gradient problem rather than the uses of conventional RNN units. Besides, this study will discover the best parameter for crop yield prediction by identifying the correlation between them using python. Lastly, the performance of the hybrid model is evaluate using a few evaluation metrics.

ABSTRAK

AI telah berjaya diterapkan dalam bidang pertanian dari segi ramalan, pembuatan keputusan, pemantauan dan analisis tanaman dan tanah. Dalam kajian ini, model hybrid pintar menggunakan CNN dan RNN untuk ramalan hasil tanaman dicadangkan. Model pembelajaran yang dicadangkan ialah gabungan Convolutional Neural Network (CNN) dengan model Recurrent Neural Network (RNN). CNN adalah model pembelajaran yang popular yang digunakan dalam meramalkan hasil tanaman disebabkan oleh prestasinya yang tinggi dalam pengekstrakan ciri. Algoritma CNN digunakan dalam kajian ini adalah kerana cirinya di mana ia menganggap bilangan parameter yang lebih kecil dalam rangkaian, dan ia mempunyai peluang yang lebih rendah untuk terlalu muat dengan data yang diproses. RNN bertindak sebagai model ramalan dalam kajian ini. RNN mempunyai sifat pembelajaran, rangkaian maklum balas dan boleh mencatat maklumat urutan temporal. Disebabkan oleh tingkah laku ingatan masa yang singkat RNN, rangkaian RNN dipertingkatkan dengan sel LSTM yang membolehkan mereka melaksanakan tugas ingatan jangka panjang. LSTM mempersembahkan blok memori dalam menyelesaikan masalah kecerunan yang meletup dan lenyap daripada penggunaan unit RNN konvensional. Selain itu, kajian ini akan menemui parameter terbaik untuk ramalan hasil tanaman dengan mengenal pasti korelasi antara mereka menggunakan ular sawa. Akhir sekali, prestasi model hybrid akan dinilai menggunakan beberapa metrik penilaian.

Table of Contents

List of Figure

List of Table

Chapter 1 Introduction

1.0 Introduction

Nowadays, agricultural advancement has become a critical component in meeting global food demand while also driving global economic growth. With the advent of science and technology, Artificial Intelligence has begun to play an important role in our daily life (Talaviya et al., 2020). Therefore, Artificial Intelligent (AI), as an important branch in science and technology, began to be implemented in the agriculture sector to improve the enterprise performance and its productivity.

According to Nwanganga et al., (2020), Artificial Intelligent (AI) is a simulation of human intelligence to get the computer system to mimic human behaviour. Implementation of AI in agriculture improves efficiency like crop yield, agriculture robots, irrigation and so on. Today, an increasing number of people tend to implement AI in agriculture to monitor and manage their farm. In this case, crop prediction is a crucial process where it directly influences the production of agriculture. Yield is depending on various environmental factors such as rainfall, humidity, and temperature (Raja et al., 2022). Besides, there are many other factors that would have influenced the yield of crops, for example fertilisation, plant protection, air pollution, water pollution and so on (Raja et al., 2022).

With the rising world population, quality food demand is increasing. The society and community are facing problems in their traditional farming which could not fulfil the demand of food. The problems that the growers or farmers are facing include the growers are facing a hard time to get the best environmental parameters to train the model which causes a huge difference between the predicted and actual crops yield (Mamunur Rashid et al., 2021). Next, the non-hybrid crop yield prediction model underperformed the hybrid crop yield prediction

model (Khaki et al., 2020). Furthermore, the high demand of food and the insufficient food supply has raised the issue of food insecurity in a country (Anbananthen et al, 2021).

Therefore, in this study, Machine Learning (ML) which is a branch of AI will be discussed to build a crop yield prediction model. Machine learning can be categorised into supervised and unsupervised learning where supervised learning is to train on labelled data and unsupervised learning train unlabelled data (Nwanganga et al, 2020). Besides, a hybrid of two machine learning is used in this study to model out the crop yield prediction. Hybrid model is a combination of two different machine learning models (Rasha M. Abd El-Aziz, 2022). Hence, in this study, a combination of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) is proposed to predict the crop yield in agriculture.

1.1 Problem Statement

With the rising world population and changing dietary habits, the food demand is expected to increase by 60% by the year 2050 (Fróna et al., 2019). Hence, advanced technologies should be implemented in the agriculture sector, to meet the increased food demand worldwide to improve the productivity of agriculture. Over the years, Artificial Intelligence has been introduced in the agriculture sector to improve the crop yield and crop quality. However, there are some problems that are still existing which include i) The growers are facing a hard time to get the best environmental parameters to train the model which causes a huge difference between the predicted and actual crops yield (Mamunur Rashid et al., 2021). Besides, ii) The non-hybrid crop yield prediction model underperformed the hybrid crop yield prediction model (Khaki et al., 2020). Lastly, iii) The high demand of food and the insufficient food supply has raised the issue of food insecurity in a country (Anbananthen et al, 2021).

1.2 Objectives

With relates to the problem statements that figured out, the objectives to solve the problem of food insecurity and improve the performance in agriculture's yield are shown as followings:

i. To identify the best environment parameters to be used to predict crops yield.
ii. To design a hybrid model which combines two machine learning algorithms for crop yield prediction.
iii. To develop a prototype of crop yield prediction model for farmers and growers to make better decisions on the crops management.

1.3 Brief Methodology

In this section, the method or technique used to carry out the project will be explained briefly, to give an outline on how the project will be conducted. These processes can be divided into several sections which include data acquisition, proposed CNN-RNN model and metric evaluation.

1.3.1 Data acquisition

This dataset is sized as 12628 data and is collected from kaggel.com which is about agriculture crop production in India from 1997 to 2014, the data source is from https://data.gov.in/ fully Licensed. Based on the data available in the dataset, wheat will be chosen for this study where this is the basic food during a food crisis. The collected data will then be pre-processed to get clean and quality data. Data pre-processing is a process to remove duplicate data, fill missing values or replace them and data integration by merging data in multiple sources into a single larger data. Next, transform the data into meaningful datasets using normalisation or generalisation. Finally, the dataset will be separated into two set which

are the training set and the test set. The training set will be 80% of total instances and the testing set will be 20% of the total instances where the training set is to train the model while the test set is to test the trained model (Bali & Singla, 2021). Next, the best parameter will be identify using correlation between parameters.

1.3.2 Proposed CNN-RNN model

The proposed hybrid consists of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) which will be modelled using Python in this project. CNN in this project is to capture the spatial data and RNN is to process sequential data for prediction. Convolutional Neural Network (CNN) is a specialised neural network for processing multidimensional data (Varghese and Kandasamy, 2021). According to a study by Benos et. al (2021), Neural Network (NN) was the most frequently used machine learning which appeared in almost half of the reviewed surveys which is about 51.8% out of the studies. While CNN is categorised as one of the Neural Network, is an algorithm that provides the best output, with almost 50% out of the percentage of NNs in crop yield prediction (Benos et al, 2021). Hence, CNN is designed to learn spatial-temporal dependencies where it is used for feature selection in the hybrid model. In this study, the CNN part of the model will consist of convolutional layer and fully connected layer (FC). On the other hand, Recurrent Neural Network (RNN) is a type of neural network that is very effective in sequence modelling (Varghese and Kandasamy, 2021). Its role in the hybrid model is as a predictive model to predict the crop yields. Long Short-Term Memory (LSTM), is a familiar type of RNN where it is used to remember time series information. LSTM monitors the attributes or factors from time to time and the time series data is analysed. Besides, the hyperparameter of the CNN likes number of nodes, stride number and filter size will be tune during the training of the model to get the best prediction

performance. Therefore, the proposed model can find nonlinear and interaction from historical data to increase the performance of prediction.

1.3.3 Metric Evaluation

The prediction model will be tested with its accuracy using Root Mean Squared Error (RMSE) to find the difference of error between predicted value and the actual value. Accuracy (%) and R-Squared (R^2) will also be used to test the accuracy of the proposed model.

1.4 Scope

The scope of this project is to implement a prediction model to forecast crop yield for growers or farmers. The crop yield prediction dataset will be used where the crops available in dataset which is wheat. In this project, a combination of hybrid models is proposed to increase the performance and accuracy of the crops yield prediction. This study also helps to determine the best parameters of prediction and helps the growers to have decision support in their operation. The proposed hybrid model will be examined by the evaluation metrics which include Root Mean Square Error (RMSE), R-Squared (R^2) and Accuracy (%).

1.5 Significance of Project

The significance of this project is to investigate the best parameter and suitable machine learning hybrid model to be used in the prediction model for crop yield. This project emphasises the use of deep neural networks and time-series models to model a crop yield prediction model. This study will benefit the growers and farmers to improve the accuracy on crop yield prediction and knowing the parameters used for crop yield prediction. Finally, this study is significant towards the management to have their operational decision making.

1.6 Project Schedule

This project will be done throughout the two semesters of the fourth year of study which is session 2022/2023. The detailed schedule is represented as *Figure A.1* and *Figure A.2* under Appendix.

1.7 Expected Outcome

This study is expected to prove that Artificial Intelligence is important in the agriculture field to improve agriculture production. Besides, the best parameters and techniques will be determined to increase the crop yield prediction besides modelling a hybrid crop yield prediction model. It will also allow agriculturists and farmers to make better decisions in crop management to maximise crop yields for food sufficient supply and food security.

Chapter 2 Literature Review

2.0 Introduction

Ansarifar et al. (2021) point out that the crop yield prediction is important for global food security, however there are various factors that will become the obstacle in food security. The factors of Ansarifar et al. (2021) mentioned include genotype, environment, management, and their complex interactions. On the other hand, early crop yield prediction can significantly contribute to reducing famine by forecasting the availability of food for rising global population (Al-Adhaileh et al., 2022). According to World Health Organization, the world is still existing 820 million people who in the situation of insufficient food supply; while the United Nations' Sustainable Development Goals (SGDs) aim to eradicate hunger, achieve food security goal and promote sustainable agriculture by 2030 which are emphasis on agriculture by WHO, 2021 (Al-Adhaileh et al., 2022). Hence, this study of crop yield prediction is aimed to help in food security too. While this chapter will focus on the relevant literature review on related fields of study. Literature review summarizes and outlines previous or existing research and theories that are done by the other researcher to provide the author foundation knowledge of the study. Important concepts, experimental techniques, and research methods can be learned when exploring through a particular topic.

In this related work section, we will discuss the existing machine learning models which are used in the crop prediction field, the common parameters of the researchers used, and the evaluation method used. Besides, the discussion and result of other research papers will be discussed in this chapter to prove the CNN-RNN hybrid model will increase the performance in prediction. Furthermore, the tool, software and the programming language will be reviewed in this chapter to let us clear about which platform or tool we should take into consideration during the development phase.

2.1 Overview of the flow of crop yield prediction

To visualize the trend of the crop yield prediction, there are a few aspects that are important to observe. The aspects include the data that will be used to forecast, then the method used to do the prediction, the hyperparameters to use and tune during the forecasting process. Figure 2.1 illustrates the process of crop yield prediction. Firstly, the required data is collected from related fields then processed to clean the data. Secondly, feature selection and feature extraction were applied to extract the important features (P. S. & R., 2019). The dataset is then split into two which are the train set and test set (P. S. & R., 2019). Train set data is to train the model while test set data is to test the trained model (Bali & Singla, 2021). Then, the machine learning algorithm proposed will be applied to train the dataset then predict the crop yield. The validation and evaluation of performance is then complete by using evaluation metrics. Finally, the results in term of accuracy, errors and graph will be obtained from the forecast.

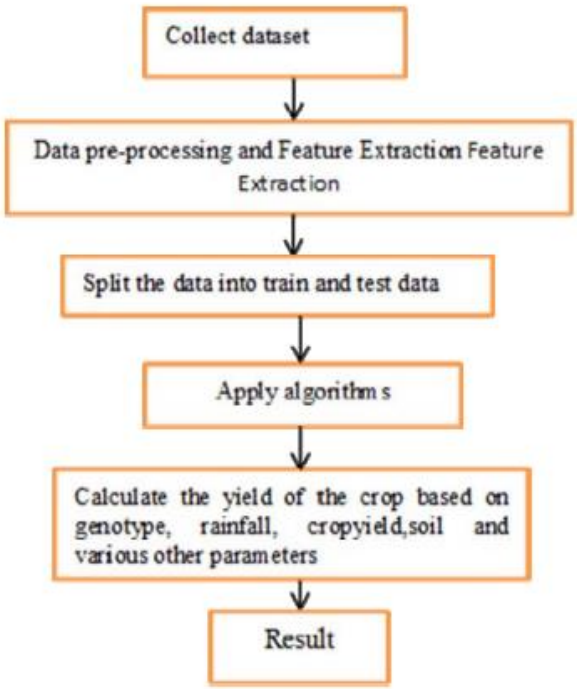

Figure 2.1: Flow Chart of Crop Yield Prediction (Pravallika et al., 2021)

2.2 Machine Learning model and algorithms used for crop yield prediction

According to Nwanganga et al., (2020), Artificial Intelligent (AI) is a simulation of human intelligence to get the computer system to mimic human behavior. There are two types of AI branches which are “Narrowed AI” and “Generalized AI”. “Narrowed AI” means the system simulates the human mind to mimic human behavior while “Generalized AI” means the system will perform intelligence to learn how to learn like a person (Davenport, 2019). While machine learning (ML) is a field that focuses on the learning of AI by developing the algorithms that represent a dataset. ML optimizes the performance of a task (Benos et al., 2021). Machine learning can generate the relationship regarding the input data (Benos et al., 2021). Machine learning learns the pattern of provided data and then predicts future knowledge from previous experience (Benos et al., 2021).

There are numerous types of crop yield prediction models and algorithms that have been introduced in the agriculture industry. However, there are some machine learning models which are more popular in the field of crop yield prediction. According to Mamunur, et al. 2021, Random Forest (RF) is the most widely used algorithm for crop yield forecasting followed by Convolutional Neural Network (CNN). On the other hand, Least Absolute Shrinkage and Selection Operator (LASSO), Long Short-Term Memory (LSTM) and hybrid model CNN-RNN or CNN-LSTM which is also algorithm that used in crop yield prediction (Mamunur, et al., 2021). Therefore, LASSO, Random Forest, CNN, LSTM, and hybrid models will be discussed in the following section. Table 2.1 shows the common parameters used in crop yield prediction and the crop that are used to forecast for wheat, maize, soybean, and rice (Amaratunga at al., 2020; Karuna et al., 2021).