**Rapid Rice Yield Estimation Using Integrated Remote Sensing and Meteorological Data and Machine Learning**

**INTRODUCTION**

Rice is one of the most consumed staple foods for nearly a half of the world’s population. By 2050, the population is projected to grow to 9.8 billion, and to keep up with this growth, there will be a 60% rise in food consumption. In addition, the livelihood of millions of farmers depends directly or indirectly on crop production. Over the past few years, the world’s total arable land has been diminishing as a result of rising urbanization, which has an impact on overall production and results in a persistent inability to meet the world’s demand for agricultural products. To satisfy this anticipated future demand, ending hunger, establishing food security, and promoting sustainable agriculture are expressly listed as the top priorities in the 2030 Agenda for Sustainable Development. Goals (SDG) of the United Nations (UN). In recent years, a great deal of effort has been initiated to increase rice production with modern technology. However, it is generally considered that rice production is associated with the immediate and dynamic nature of global anthropogenic changes, such as population growth and climate changes, and also technological advancement. Extreme natural and manmade events such as drought, flood, and fire/war frequently harm food production. It is anticipated that both the rates and patterns of total precipitation amount will keep changing, along with the rise in global temperatures, which is expected to have mild to severe consequences on agriculture around the world. The shortage of water and energy supply in the agriculture sector will further impose constraints on food production. In this regard, more precise crop monitoring, dependable mapping tools, and early forecasting of rice production, before the harvest time, can substantially assist the decision makers in minimizing losses and achieving desired yields. Recently, remote sensing became a popular tool to monitor crop health, growth, measurement, and to determine the optimal time for harvesting and rapid near real-time crop yield estimation with minimal cost. Remote sensing-based techniques have already been successfully applied for mapping rice-cultivated areas and demonstrated promising results in delineating accurate cultivated yield. Furthermore, the vegetation index derived from satellite images was also successfully applied for predicting rice yield before harvesting. One of the popular remote sensing products is MODIS NDVI data, which has the benefits of decadal archives and high spatiotemporal resolution and has been widely employed for regional agricultural yield assessment and forecast. The most widely used vegetation indices are the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), soil-adjusted vegetation index (SAVI), leaf area index (LAI), and the fraction of absorbed photo-synthetically active radiation index (FPAR). A substantial link between NDVI and LAI and green biomass yield before rice harvest. In this study, MODIS NDVI product along with MODIS LAI and FPAR product and several meteorological factors were used in the experimental rice yield estimation model development. Since meteorological factors such as monthly rainfall, land surface temperature, potential evapotranspiration, and soil moisture content have a substantial influence on rice crop productivity, it is crucial to take into account this information in the yield estimation workflow. Based on the literature, we initially selected NDVI, LAI, FPAR as the primary factors, then added soil moisture, rainfall, total precipitation amount, land surface temperature (LST), and evapotranspiration (ET) as auxiliary factors.

**EXISTING SYSTEM**

The existing system described in the provided information involves the use of remote sensing technology, particularly MODIS (Moderate Resolution Imaging Spectroradiometer) NDVI (Normalized Difference Vegetation Index) data, MODIS LAI (Leaf Area Index), and FPAR (Fraction of Absorbed Photosynthetically Active Radiation) product, along with various meteorological factors for developing an experimental rice yield estimation model. This system aims to address the challenges in rice production caused by factors such as population growth, climate change, and technological advancements.

**PROBLEM STATEMENT**

The lack of precise crop monitoring, mapping tools, and early forecasting mechanisms impedes informed decision-making in rice production. There is a critical need for advanced and reliable tools that enable early prediction of rice production, minimizing losses, and ensuring desired yields. The agriculture sector faces challenges from technological advancements and the consequences of climate change. Adapting to these changes and leveraging modern technology effectively are crucial aspects for ensuring the resilience of rice production systems.

Despite the growing emphasis on crop monitoring and yield estimation, there is a need for robust data integration and decision support systems. Providing decision-makers with accurate, timely, and actionable information remains a challenge that needs to be addressed for effective agricultural data management.

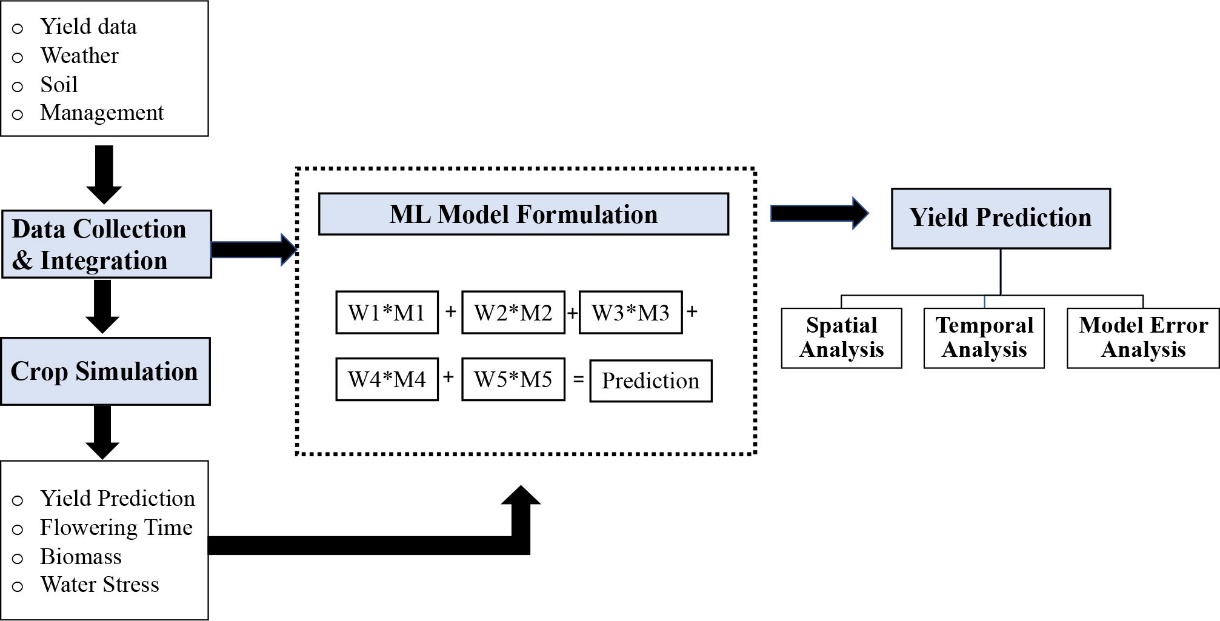
Developing an accurate rice yield estimation model, incorporating data and meteorological factors, requires thorough calibration and validation. The challenge lies in understanding and refining the relationships between these factors to enhance the reliability of the model, ensuring it accurately reflects the dynamic nature of rice production.

**OBJECTIVE**

**The main objectives is:**

* To develop an efficient and rapid workflow for estimating rice yield by integrating remote sensing and meteorological data with machine learning (ML) techniques.
* To develop a customized machine learning model tailored for rice yield prediction for Mysore region only.
* To implement data processing techniques and evaluate different machine learning models and determine the most suitable approach for rice yield estimation which can enhance the accuracy of crop yield estimation at the district level in Mysore.

**PROPOSED SYSTEM**



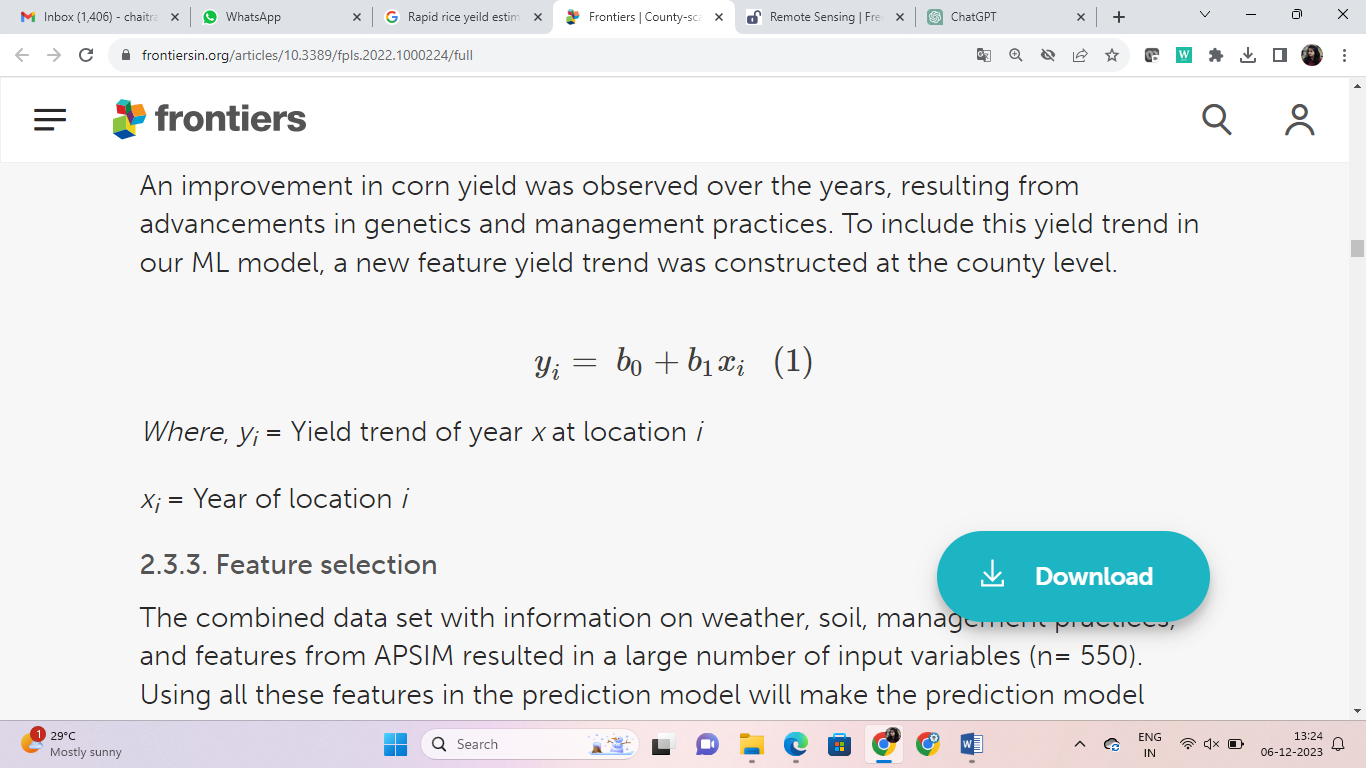
The entire workflow of crop yield estimation was implemented in several steps, including study area selection, data preprocessing, combining multiple datasets, ML model development, and model performance evaluation. Figure 1 displays the simplified workflow of the entire crop yield estimation model development process

1. **Data Collection and Preprocessing**

Prior to developing ML prediction models, the following data processing will be conducted. First, data from all sources will be aggregated while treating any missing values. Thereafter, new features are constructed to improve the model’s performance. After feature construction, a set of important features will be identified and used to develop the final prediction models.

1. **Feature construction of rice yield trend**

An improvement in corn yield was observed over the years, resulting from advancements in genetics and management practices. To include this yield trend in our ML model, a new feature yield trend was constructed at the county level.



Where, yi = Yield trend of year x at location i

xi = Year of location i

1. **Feature selection**

Feature selection is essential to ensure that the designed ML model is generic. The input features are reduced using a two-stage feature selection approach. The first stage is based on expert knowledge, and the second stage is based on the permutation feature selection approach. Then the ML models are trained accordingly.

**SOFTWARE REQUIREMENTS**

* Python Programming Language
* Python IDLE

**HARDWARE REQUIREMENTS**

Modern Operating System:

* Windows 7 or 10
* Mac OS X 10.11 or higher, 64-bit
* Linux: RHEL 6/7, 64-bit (almost all libraries also work in Ubuntu)
* x86 64-bit CPU (Intel / AMD architecture). ARM CPUs are not supported.
* 4 GB RAM
* 5 GB free disk space

**PROBLEM STATEMENT**