Article

Input Use Efficiency Management for Paddy Production Systems in India: A Machine Learning Approach

Priya Brata Bhoi [1], Veeresh S. Wali [2], Deepak Kumar Swain [3], Kalpana Sharma [4], Akash Kumar Bhoi [4,5,*], Manlio Bacco [5] and Paolo Barsocchi [5,*]

[1] Department of Economics and Sociology, Punjab Agricultural University, Ludhiana 141004, Punjab, India; priyabrata@pau.edu
[2] Indian Institute of Millets Research, Hyderabad 500030, Telangana, India; waliveeresh92@gmail.com
[3] Department of Agricultural Statistics, Faculty of Agricultural Sciences, Siksha 'O' Anusandhan University, Bhubaneswar 751003, Odisha, India; deepakkswain@soa.ac.in
[4] Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar 737136, Sikkim, India; kalpana.s@smit.smu.edu.in
[5] Institute of Information Science and Technologies, National Research Council, 56124 Pisa, Italy; manlio.bacco@isti.cnr.it
* Correspondence: akash.b@smit.smu.edu.in (A.K.B.); paolo.barsocchi@isti.cnr.it (P.B.)

Abstract: This research illustrates the technical efficiency of the pan-India paddy cultivation status obtained through a stochastic frontier approach. The results suggest that the mean technical efficiency varies from 0.64 in Gujarat to 0.95 in Odisha. Inputs like human labor, mechanical labor, fertilizer, irrigation and insecticide were found to determine the yield in paddy cultivation across India (except for Chhattisgarh). Inefficiency in the paddy production in Punjab, Bihar, West Bengal, Andhra Pradesh, Tamil Nadu, Kerala, Assam, Gujarat and Odisha in 2016–2017 was caused by technical inefficiency due to poor input management, as suggested by the significant $\sigma^2 U$ and $\sigma^2 v$ values of the stochastic frontier model. In addition, most of the farm groups in the study operated in the high-efficiency group (80–90% technical efficiency). No specific pattern of input use can be visualized through descriptive measures to give any specific policy implication. Thus, machine learning algorithms based on the input parameters were tested on the data in order to predict the farmers' efficiency class for individual states. The highest mean accuracy of 0.80 for the models of all of the states was achieved in random forest models. Among the various states of India, the best random forest prediction model based on accuracy was fitted to the input data of Bihar (0.91), followed by Uttar Pradesh (0.89), Andhra Pradesh (0.88), Assam (0.88) and West Bengal (0.86). Thus, the study provides a technique for the classification and prediction of a farmer's efficiency group from the levels of input use in paddy cultivation for each state in the study. The study uses the DES input dataset to classify and predict the efficiency group of the farmer, as other machine learning models in agriculture have used mostly satellite, spectral imaging and soil property data to detect disease, weeds and crops.

Keywords: paddy; stochastic frontier; machine learning; *k*-nearest neighbour (KNN); support vector machine (SVM); random forest (RF)

1. Introduction

Of the many facets of agrarian distress in India, the input management factor carries the highest weight among all. Input management is the process of employing inputs, such as chemicals, in optimal quantities to increase yield and destroy pests, etc. The published agriculture statistics of India's Government show an apparent disparity among the major paddy producing states with regard to their input application rates and productivity over the years (Department of Economics and Statistics (DES), Ministry of Agriculture, GoI, reports on the Cost of Cultivation Surveys). With the advent of the 21st century,

agriculture has witnessed technological growth like all other sectors of the economy [1]. India got its share of technological augmentation in the agricultural sector in the "Green Revolution," which spanned from the 1960s to the 1990s, with long-term effects on the productivity growth of major crops like wheat and paddy. Rice, which is the final product of paddy crop, is the staple food of the majority of the population in Asia and half of the world's population. Asia accounts for 90 percent of global rice consumption (https://ricepedia.org/rice-as-food/the-global-staple-rice-consumers (accessed on 20 January 2021)) and the demand continues to rise. Paddy cultivation covers nearly 43 million hectares of arable land, which is almost 27 percent of the total 159 million hectares of arable land in India; rice is the staple food grain for nearly 50 percent of the Indian population [2], and it covers all of the states and agroclimatic zones. Thus, it is one of the most important crops for food security, with the income of about 59 percent of the Indian population [2] engaged in agriculture. The productivity of this crop has steadily increased decade by decade from 1961 to 2001; however, after 2001, there was stagnation in productivity, and a productivity plateau can be observed after the year 2005 [2]. The cost of cultivation per hectare of paddy has seen a steep growth from 2001–2002 to the latest 2016–2017, as estimated by the DES, Government of India. In 2016–2017, the total production of paddy was 109 million tonnes, of which 67 percent came from seven major producing states out of the 31 states on which DES collected data.

Furthermore, 37 percent of paddy production in India came just from West Bengal, Uttar Pradesh and Punjab, which incidentally covers 34 percent of the country's total paddy area [3]. Thus, the major production is concentrated in certain country regions due to their technological and policy advantages. Regional productivity figures demonstrate that certain regions in the country are far better than other regions for paddy farming. The northern plains recorded a productivity average of 2831 kg per hectare, followed by 2665 kg per hectare in the southern states, 2286 kg per hectare in the eastern states, and the lowest productivity was observed in the northern hilly regions, at 2133 kg per hectare. Thus, regional disparity plays a crucial role in determining future strategies for sustainable paddy farming across India. Capital intensive agriculture should have penetrated all states after the green revolution in the States of Punjab and Haryana. However, certain states still practice labor-intensive practices, and subsistence agriculture thus has low productivity figures. Thus, an empirical approach should be taken to ascertain the causes of yield stagnation, which could cause food security issues in the future.

For economists and policymakers across India, the policy challenge is to delineate a strategy to augment yield levels from the current stagnation and enhance the shrinking profit margin. Even in states like Punjab, which reaped the benefits of the green revolution, the intensive resource exploitation, the partial adoption of production technology, and the ineffective policy formulation have led to stagnation in paddy cultivation [4]. The new economic policies that proposed the removal of subsidies on crucial farm inputs, like fertilizer, have put upward pressure on the cost of cultivation and can lead to wash-off profits from paddy production. The high input requirement, rising cost of inputs and slow increase in assured prices cumulatively lead to the lower profit margins of the farmer. In such a technological setup, the only thing that remains under the control of farmers is the efficient use of input to obtain the maximum potential yield. Against the backdrop of these studies, an attempt has been made to examine the paddy cultivation status across all of the major growing states of India. The stochastic frontier analysis provides ample scope to minutely analyze the states' efficiency dynamics in paddy cultivation and the provision of improving efficiency and yield levels to the highest possible levels at the present level of technology.

Some major efficiency studies on paddy farming in India show the non-profitable status across various states of India. In Rajasthan, the share of operational and fixed costs increases in the same proportion in the total cost of cultivation [5]. The pivotal factors for the increase in operational cost were the high wage rates, the increased mechanization, and the steep increases in seed and fertilizer prices. In a pan-India study, it was found

that of the seven-time periods under study, only in two periods could farmers make some profit over the total cost of cultivation, namely C2, considering both the fixed and variable cost components per crop [6]. The reports of the Commission of Agriculture Cost and Prices (CACP) accentuated the fact that in some of the major paddy-producing states like Kerala, Tamil Nadu and Odisha, profitability hovered around ten percent in 1999–2000 and 2010–2011; the varying degree of loss was reported in other periods [7]. In Andhra Pradesh, a trend of higher input use with an increase in farm size was reported [8]. In recent studies, the higher incidence of farmer suicides has been attributed to higher production costs and low profits due to low prices [9–12].

Efficiency studies in crop production help us to understand the current production system's potential yield levels and thus improve the actual yield, e.g., how to achieve better productivities without increasing input application [13–15], or how to better use the current technology and institutional reforms to accommodate innovations and investment in rural infrastructure to increase production growth [16]. A national study designed to compare the production efficiency among various pre-classified farm categories can give policymakers an outline to properly allocate resources for the achievement of the maximum productivity potential. These studies are essential to exploit the potential of current technologies and bring productivity reforms [13]. According to Kalirajan et al. [17], developing countries face a two-fold problem: scarce resources and a lag in technological growth. In such a setup, efficiency studies provide an excellent base to achieve productivity growth through the improvement of current technologies and avoiding costly technological reforms in the short run. As Shanumugan et al. [18] proposed, it is possible to raise crops' productivity without raising the input application. On the backdrop of this research, an efficiency study for the rice crop will have pivotal implications to improve productivity by improving current technology and acquiring knowledge about the status of technologies in different Indian states.

Smart decision making in agriculture is based on four key areas, namely (a) optimal natural resource management, (b) the conservation of the ecosystem, (c) the development of adequate services, and (d) the utilization of modern technologies [19]. Various studies use different kinds of datasets, from satellite to multispectral images and generic field observation to extract information for smart agriculture applications. There are important studies for soil fertility prediction [20], soil moisture [21–23], clay prediction by portable multispectral cameras [24], prediction for the condition of indoor plants through partial least squares [25], disease detection [26–28], and weed detection [29,30]. These models use an array of machine learning algorithms, including artificial neural networks (ANN), SVM, RF, KNN, multiple linear regression (MLR), etc., for various crops to predict, including ANN, SVM, RF, KNN, and MLR, etc., for various crops to predict their yield. Specifically for paddy/rice crop yield prediction, RGB and UAV data [31–33]; satellite spectral data [34]; weather data [35,36]; weather and soil data [37,38]; and weather, irrigation, planting, and fertilizer data [39,40] have been used in various studies. Random forest models [33,38–40], SVM models [34,39,40] and KNN models [39,40] have been suggested in notable studies for yield prediction in paddy crops. However, no notable studies used these models to predict the efficiency level of the farmers based on inputs like human labor, machine labor, irrigation, fertilizer, crop area and size group. The classification of farmers through efficiency levels helps us to understand the levels of input utilization and the level of technology that the farmers are rendering. The further analysis helps us to know the farmers' size groups, which provides us the scope to improve the achievable efficiency by suggesting changes in the input management. The research problem of classifying farmers into different technical efficiency levels is addressed herein through a Stochastic Frontier Approach [20,21]. Furthermore, three machine learning models—i.e., k-nearest neighbors (KNN), support vector machine (SVM), and Random Forest (RF)—have been used to predict the efficiency group of the farmer based on the input variables and size group. The relatively accurate prediction model will be suggested for each state of the nation in order to advise on the appropriate policy measures for each state on input management.

Thus, the study proposes to study the regional disparity in paddy cultivation across India, and to establish that the input management capacity of the farmers across various states plays a pivotal role in determining the productivity and efficiency difference. In addition, the study aims to build an efficiency group classification cum prediction model for each state individually, in order to help policymakers decide on an effective input management strategy to keep the farmers at the highest level of efficiency.

2. Data and Methodology

2.1. Data Acquisition

The study used data published by the Department of Economics and Statistics (DES) under the Ministry of Agriculture and Farmers Welfare of the Government of India. The data was collected at state nodal centers under the scheme 'Cost of Cultivation of Principal Crops of India' [41]. The data used in this study came from the 2016–2017 period, the latest available one. The data is a plot-level summary of selected farmers in each state encompassing input use in paddy cultivation. The workflow for the study is illustrated in the workflow diagram below.

A three-stage stratified random sampling coupled with a probability proportional sampling method was used to collect the data. A detailed description of the complex sampling techniques can be found in the Manual of Cost of Cultivation Surveys; (2008) published by the DES, Government of India [42]. The unique feature of those data is that the farmers record and collect it carefully during the production process, so that the data accuracy remains high. The data covers varying land sizes across states as 10 farmers from each tehsil (township) of the considered states, so the bigger the state is or the higher the number of tehsils is, the larger the sample size. The farmers in the data are classified according to their farm size, and there are five size categories: Marginal (<1 ha), Small (1–2 ha), Semi-Medium (2–4 ha), Medium (4–6 ha) and Large (≥6 ha).

2.2. Data Pre-Processing

Data cleaning was performed before their use in this study. The plot-level summary data was first summarized to a farm level, which was used in our study. This study also used the cost of cultivation. We then described the methodological framework of this study, aiming to identify the loopholes in paddy cultivation technology across different Indian states and suggest appropriate mitigation measures.

The model for each state represents the technology level. They are not readily comparable because they represent the technology frontier for the respective state. For each input used by farmers in paddy production, the corresponding variable in our model is zero if that input type is not used, or the variable is removed from the model if the input is not of common use (>90% of the cases). For cases where the input is 0, we have put 0.01 because the stochastic frontier model uses a log-linear form, and a logarithm of 0 is impossible. Then, the variables are filtered again through the Ramsey Reset Test validity to obtain a well-fitted model for each Indian state.

2.3. Stochastic Frontier Algorithm

Firstly, the individual farm-level technical efficiency was estimated through the stochastic frontier approach. The model was theorized by Meeusen and Van Den Broeck, and Aigner, Lovell and Schmidt in two different seminal papers published in 1977 [43,44]. The stochastic frontiers model was then developed and applied to many sectors, including the agricultural sector, and a model based on these studies was applied to this study. The stochastic frontier models are not affected by the outliers or the extreme observations, as they require normalized logarithmic values for the estimation procedure [44].

Various researchers have carried out considerable improvements in the model since then. According to Battese and Coelli, applying stochastic frontier models to cross-sectional and panel data models to estimate individual farm level efficiency is very important. The specification of this model is such that model the error term (Ei) is divided into a stochastic

term (vi) and an inefficiency term (ui) [45,46]. This inefficiency term is of prime importance for this study. The R Frontier package 1.1–8 [47] was used in the stochastic frontier model estimation to predict individual farm efficiencies.

Generally, a Cobb–Douglas production function is represented by the following equation:

$$Y = \beta_0 \times \prod_{i=1}^{n} X_i^{\beta_i} \times e^u \tag{1}$$

where

Y = the yield or any variable representing the productivity per unit area.
X_i = the vector of inputs used in production.
β_i = the estimated coefficient of the ith input.
u = the error term.

The Cobb–Douglas production function is expanded to carry the inefficiency term in the following form of the equation, which is known as the stochastic frontier production function, and is given by

$$Y = \beta_0 \times \prod_{i=1}^{n} X_i^{\beta_i} \times e^{(v-u)} \tag{2}$$

where

Y = the yield or any variable representing the productivity per unit.
X_i = the vector of inputs (the same as Equation (1));
β_i = the estimated coefficient of the ith input.
v_i = an asymmetrical random term or stochastic noise, assumed with a normal distribution $[N(0, \sigma_v^2)]$
u_i = the individual farm level technical inefficiency assumed to be half-normally distributed.

For the current study, the variable specification for the study is as follows:

Y_i = Output/Yield (quintals per hectare)
X_1 = Total human labor (Man-hours)
X_2 = Total animal labor (Hours)
X_3 = Total machine labor (Hours)
X_4 = Total Fertilizer (kg.)
X_5 = Total insecticide (Rupees).

Each farm has its own production frontier $f(X_i, \beta)e^{v_i}$ composed of a deterministic part $f(X_i, \beta)$ common to all producers, and a farm-specific part $e^{(v_i)}$. The following equation provides the farm-level technical efficiency:

$$TE_i = \frac{f(X_i, \beta)\exp(v_i - u_i)}{f(X_i, \beta)\exp(v_i)} = \exp(-u_i) \tag{3}$$

where

f = the Cobb–Douglas type production function.
TE = the technical efficiency of an individual farm $(0 < TE_i \leq 1)$.

The efficiency levels obtained from the stochastic frontier analysis will classify the farmers into four different groups, as discussed in Figure 1 and mentioned in Table 1.
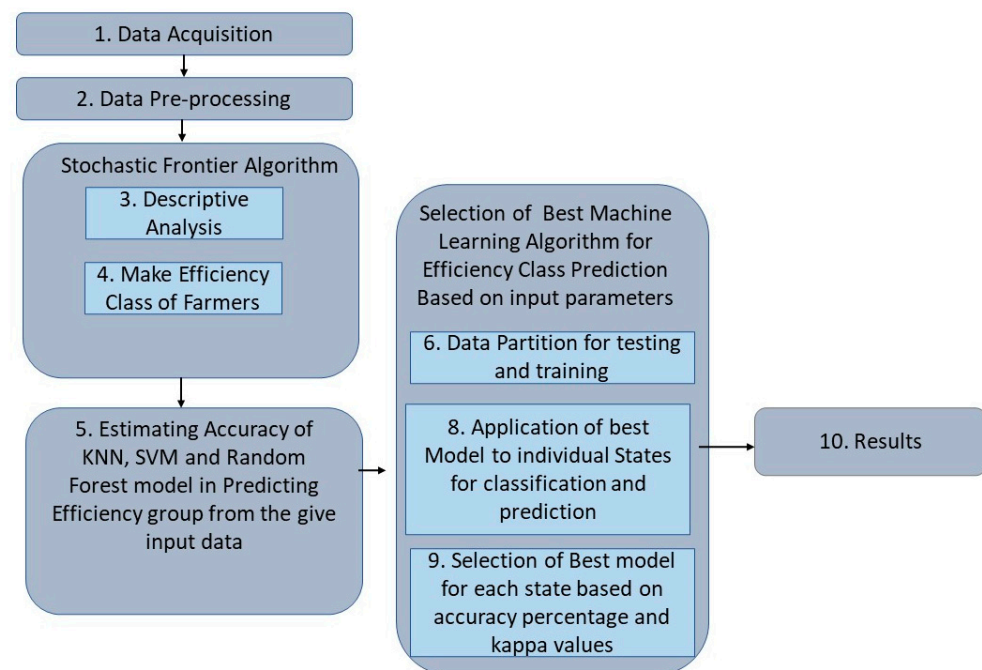
Figure 1. Systemic workflow of the research paper.

Table 1. Efficiency group cut-offs in Stochastic Frontier Analysis.

Efficiency Class	Efficiency Score Range
Very High	1.0 to 0.90
High	0.90 to 0.80
Medium	0.80 to 0.70
Low	<0.70

2.4. Machine Learning Algorithms for the Prediction of Efficiency Classes

All of the inputs used in the stochastic frontier model and the size group will be used to predict the efficiency classes. For this task, the "nnet" [48] and "caret" [49] packages provided in the R computing environment were used. The KNN, SVM and RF algorithms are run after the data partitioning for the train and test ratio. As illustrated in Figure 1, a train test ratio of 80:20 has been used for the data sets of each state, and state-wise classification algorithms were run with 10-fold cross-validation. A comparative table of classification and a prediction algorithm used in the agricultural study is given below.

Table 2 gives a comparative view of the various datasets and models used in the prediction of paddy yield, and also the use of KNN, SVM and RF algorithms in agriculture. As discussed in an earlier section, our dataset is unique for this set up, as we have used production input data to classify the efficiency groups of farmers.

Table 2. Comparison of paddy efficiency group classification and prediction with relevant works.

Author	Classifiers/Predictors	Crop	Classification/Prediction Problem	Machine Learning Algorithm	Model Selection Parameters
Gopal, M et al. [40]	Weather data, irrigation, planting area, fertilization	Paddy	paddy crop yield	ANN, SVR, KNN, RF	RF: RMSE = 0.085, MAE = 0.055, R = 0.93
Gopal, M et al. [39]	Weather data, irrigation, planting area, fertilization	Paddy	paddy fields yield	ANN, MLR, SVR, KNN RF	ANN-MLR: R = 0.99, RMSE = 0.051 MAE = 0.041

Table 2. *Cont.*

Author	Classifiers/Predictors	Crop	Classification/Prediction Problem	Machine Learning Algorithm	Model Selection Parameters
Shidnal, S et al. [31]	RGB leaf images	Paddy	nutrient deficiencies (P, N, K)	ANN	Accuracy = 77%
Khosla, E et al. [36]	Weather data	Rice, maize, millet, ragi	kharif crops yield	MANN, SVR	Overall RMSE = 79.85%
Amaratunga, V et al. [35]	Weather data	Paddy	Paddy yield	ANN	R = 0.78–1.00, MSE = 0.040–0.204
Wan, L et al. [33]	Multispectral images from UAV	Rice	rice grain yield	RF	RMSE = 62.77 kg·ha^{-1}, MAPE = 0.32
Ramesh, S et al. [26]	RGB images	Rice	Recognition and classification of rice infected leaves	KNN, ANN	ANN: Accuracy = 90%, Recall = 88%
Gomez Selvaraj et al. [50]	Satellite spectral data, Multispectral images from UAV, RGB images from UAV	Banana	Detection of banana diseases in different African	RF, SVM	RF: Accuracy = 97%, omissions error = 10%; commission error = 10%. Kappa coefficient = 0.96
Gao, J et al. [51]	RGB images from UAV	Wheat	Detection of weeds in early season maize fields	RF	Overall Accuracy = 0.945, Kappa = 0.912

The KNN algorithm is a non-parametric classification model, which is simple and effective [52]. The support vector machine has applications ranging from time series prediction [53] to biological data processing for medical diagnosis [54], and can be applied to our study for efficiency group classification. The random forest algorithm is one of the most efficient decision tree-based algorithms proposed by Leo Breiman, and it has been used to predict discrete classes [55]. The most accurate models obtained through this experiment on the basis of their accuracy percentage and kappa values for individual states can be used to classify and predict efficiency levels given the input parameters; this means that new strategies in input management can be evaluated thanks to our approach before being applied. The models used in the study are simple and are performed through preexisting modules in the R computing environment. For simplicity, we have not included the detailed mathematical explanation of the algorithms; however, the performance evaluation of the models will be based on precision, recall, accuracy, sensitivity and specificity measures. These are measured from the true positive (TP), true negative (TN), false positive (FP) and false-negative (FN) values obtained from the model.

3. Results and Discussion

This section presents the results following the approach proposed in this work. In particular, we first describe the status of paddy farming in India, and later we analyze the regional disparity in productivity. We explore the farm-level technical analysis to find the reason behind input mismanagement in the selected states. We conclude this section by analyzing three standard classification algorithms that take input data and size group labels, and predict the efficiency group for specific states.

3.1. The Status of Paddy Farming in India

Paddy farming is covered all over India, with variations in area, production and productivity, as shown in Table 3. It provides an overview of the paddy cultivation area, production and productivity statistics across all of India's major growing states in 2016–2017. The data suggest that the production percentage has surpassed the area percentage in the states with higher average productivity (see the first six rows of Table 2), indicating that more food per unit of land is produced. Thus, the disparity in productivity

must be studied at a micro level in order to ascertain the causes and prescribe remedial measures.

Table 3. State-wise area, production and productivity for paddy crops in India for 2016–2017.

States	Production Percentage to All India Production	Area as a Percentage of All India Area	Average Productivity (kg/Hectare)
Punjab	10.56	6.59	3998.00
Andhra Pradesh	6.79	4.78	3540.00
Haryana	4.06	3.15	3213.00
West Bengal	13.95	12.49	2784.00
Kerala	0.40	0.39	2550.00
Karnataka	2.37	2.35	2519.00
Bihar	7.51	7.59	2467.00
Uttarakhand	0.57	0.59	2414.00
Gujarat	1.76	1.90	2306.00
Uttar Pradesh	12.54	13.62	2295.00
Jharkhand	3.50	3.90	2241.00
Odisha	7.59	8.76	2160.00
Chhattisgarh	7.34	8.71	2101.00
Maharashtra	2.83	3.49	2025.00
Himachal Pradesh	0.13	0.17	1968.00
Assam	4.31	5.61	1916.00
Madhya Pradesh	3.85	5.20	1847.00
Tamil Nadu	2.16	3.28	1642.00

Source: Handbook of Statistics on the Indian States, RBI Publication (2018–2019).

Analyzing the cross-sectional plot-level data for farmers across various states (Table 4), it is evident that, on average, the proportion of the operational cost in paddy farming remained on a higher side than the fixed cost. However, in states like Punjab and Haryana, the proportion of fixed costs remained higher.

Table 4. Operational cost and fixed cost as a percentage of the total cost of cultivation in different states of India in 2016–2017.

States	Operational Cost (%)	Fixed Cost (%)
Andhra Pradesh	60.9	39.1
Assam	74.4	25.6
Bihar	67.7	32.3
Chhattisgarh	68.6	31.4
Gujarat	72.6	27.4
Haryana	55.6	44.4
Himachal Pradesh	69.6	30.4
Jharkhand	66.3	33.7
Karnataka	62.3	37.7
Kerala	73.8	26.2
Madhya Pradesh	71.7	28.3

Table 4. *Cont.*

States	Operational Cost (%)	Fixed Cost (%)
Maharashtra	79.7	20.3
Odisha	75.7	24.3
Punjab	47.2	52.8
Tamil Nadu	71.5	28.5
Uttar Pradesh	66.8	33.2
Uttarakhand	67.9	32.1
West Bengal	75.3	24.7

Source: Estimated from the DES Publication of 2016–2017.

From the development era of the green revolution to Punjab and Haryana's highly commercialized farm economy, it is apparent that the fixed-cost investment capacity remained high in these states. Some southern states like Andhra Pradesh and Karnataka are also catching up with the trend of investment in higher fixed costs. Linking these factors with the study shown in Table 1, it may be suggested that higher fixed cost investment may lead to higher productivity gains, and may act as a good policy implication.

From the input management perspective, Table 5 provides evidence that human labor remains the single largest input in the total operational cost, with a minimum 41 percent (Madhya Pradesh) to a maximum 74 percent (Himachal Pradesh) contribution to the total operation cost in paddy cultivation in all major paddy growing states of India. Thus, human labor wages in these states represent a crucial factor in determining the total cost of cultivation. The data illustrates that the states with higher human labor utilize fewer machines, as expected. For this study, we focused only on the input factors like human labor, machine labor, fertilizer, irrigation and insecticide, which make up nearly 90% of the total input cost in paddy cultivation across all of the states under study. The effective management of these inputs to obtain higher productivities will be crucial for paddy cultivation and these states' agrarian economy.

Table 5. Proportion of different input costs in the total operational/variable cost in paddy cultivation in India for 2016–2017.

States	Human Labor	Animal Labor	Machine Labor	Seed	Fertilizer	Insecticide	Irrigation
Andhra Pradesh	47.8	1.6	20.3	4.1	14.5	5.6	2.3
Assam	57.5	24.0	9.2	2.7	2.0	0.1	1.1
Bihar	56.0	0.4	13.6	6.6	10.1	0.1	10.0
Chhattisgarh	44.2	8.8	20.2	5.0	9.7	3.1	1.2
Gujarat	47.5	0.6	15.2	11.9	11.2	2.5	6.1
Haryana	52.3	0.0	12.9	3.0	10.2	5.0	14.0
Himachal Pradesh	74.2	7.9	6.5	6.6	1.1	1.5	0.2
Jharkhand	58.7	4.4	14.3	8.2	10.7	0.0	0.1
Karnataka	44.2	10.2	11.2	6.5	14.0	4.9	2.2
Kerala	56.2	0.0	19.6	5.6	8.3	3.3	0.2
Madhya Pradesh	41.7	9.0	19.2	6.0	9.4	3.5	2.3
Maharashtra	51.6	10.1	10.9	4.9	6.0	1.0	2.6
Odisha	66.4	7.1	11.3	2.6	6.0	0.8	0.3
Punjab	45.5	0.1	17.7	4.8	9.2	12.3	6.7
Tamil Nadu	41.5	0.2	18.2	12.7	10.8	2.8	7.4
Uttar Pradesh	49.4	1.6	11.5	10.4	11.5	0.9	12.3

Table 5. *Cont.*

States	Human Labor	Animal Labor	Machine Labor	Seed	Fertilizer	Insecticide	Irrigation
Uttarakhand	46.2	10.6	14.5	11.0	10.0	2.4	1.9
West Bengal	64.0	2.7	8.4	3.5	8.9	2.9	5.1

Source: Estimated from the DES Publication of 2016–2017.

3.2. Regional Disparity in Productivity and Input Use

Table 6 gives a lucid picture of India's various states' average input use pattern in 2016–2017. The highest yield was observed in Punjab (67.13 kg/ha), and the lowest was recorded in Himachal Pradesh (22.72 kg/ha). Furthermore, all of the eastern states except West Bengal in the study area were below the average yield of 41.31 kg per hectare in the study area, which was below the average yield of the southern region (49.68 kg/ha) and northern region (46.84 kg/ha). This may be attributed to various geographic, biotic, abiotic factors coupled with input management practices. This disparity calls for a targeted approach in these areas in terms of varietal development and input management. In chemical inputs like fertilizer, the average application was 143.30 kg per hectare over the study area. However, only West Bengal in the eastern region applied over this average (171.94 kg/ha). The rest of the eastern region states were well below it, with an average 92 kg per hectare application rate. Both the northern region (except Himachal Pradesh) and southern region had a more considerably high (more than 1.5–2.5 times) application rate than the eastern region. Insecticide use in the northern region was Rs.2069.30 per hectare, second to the southern region (Rs.2127.51 per hectare). The least insecticide use was reported in eastern region (Rs.980.60 per hectare). The average insecticide use in the study area was Rs.1630.23 per hectare.

Table 6. Descriptive statistics of the parameters used in the stochastic frontier model estimated from cross-sectional farm level data for AY2016–17.

State	Particulars	Yield (Qtls/ha)	Fertilizer (kg/ha)	Insecticides (Rs/ha)	Human Labor (Person Hours/ha)	Animal Labor (Hours/ha)	Machine Labor (Hours/ha)	Irrigation (Hours/ha)
Andhra Pradesh	Average	60.08	241.75	2810.16	541.35	22.27	23.95	293.82
	Minimum	12.5	66.5	93.2	177.83	0.5	0.91	2.47
	Maximum	110.89	590.78	22,500	1325	164.34	150	1206.25
	Coefficient of Variation	0.21	0.35	0.91	0.39	1.26	0.89	0.77
Assam	Average	33.65	47.4	829.94	668.98	180.43	65.03	92.03
	Minimum	16.68	4.83	89.55	321.16	3.9	12.64	11.56
	Maximum	69	349.89	1940.3	1415.92	424.53	156.72	229.85
	Coefficient of Variation	0.27	0.92	0.6	0.26	0.5	0.43	0.64
Bihar	Average	31.31	110.6	490.18	597.48	56.83	13.2	37.79
	Minimum	15.91	23	297.62	274.82	15.66	9	4.4
	Maximum	52.44	244.03	851.85	1192	90	30.11	80
	Coefficient of Variation	0.19	0.35	0.42	0.21	0.68	0.34	0.4
Chhattisgarh	Average	34.86	120.76	1158.67	456.14	39.61	21.84	40.36
	Minimum	13.16	31.94	53.33	99.32	1.44	15.56	8
	Maximum	48.15	220	2912.81	1040.83	192.42	30.42	141.29
	Coefficient of Variation	0.19	0.33	0.6	0.4	0.94	0.21	0.63
Gujarat	Average	37.21	153.6	1095.93	833.9	37.41	24.88	65.06
	Minimum	0.87	13.89	91.4	170.83	3.19	3.51	0.67

<div align="center">

Table 6. *Cont.*

</div>

State	Particulars	Yield (Qtls/ha)	Fertilizer (kg/ha)	Insecticides (Rs/ha)	Human Labor (Person Hours/ha)	Animal Labor (Hours/ha)	Machine Labor (Hours/ha)	Irrigation (Hours/ha)
Gujarat	Maximum	77.05	435.76	4017.22	2634.92	100	90.44	383.33
	Coefficient of Variation	0.43	0.45	0.96	0.44	0.83	0.78	1.07
Himachal Pradesh	Average	22.72	64.58	1226.39	423.48	86.47	20.45	65.73
	Minimum	6.25	14.38	388.89	200	1.39	5.56	40.62
	Maximum	52.5	191.67	4900	954.16	172.92	44.58	103.12
	Coefficient of Variation	0.49	0.9	0.69	0.32	0.41	0.49	0.39
Kerala	Average	41.01	139.79	1994.45	460.38	6.54	16.71	25.53
	Minimum	7.21	10.06	56.25	74.17	3.02	14.54	12.73
	Maximum	89.2	442.94	13550	1383.32	7.44	19.67	35.42
	Coefficient of Variation	0.4	0.59	1.02	0.53	0.3	0.12	0.27
Odisha	Average	36.77	93.22	642.26	965.91	154.52	26.74	12.62
	Minimum	15.28	23.16	25.96	507.22	1.6	0.35	0.62
	Maximum	56.18	198.49	4250	1408.67	365	58.33	32.81
	Coefficient of Variation	0.17	0.26	1.18	0.17	0.72	0.52	0.75
Punjab	Average	67.13	183.1	4122.04	363.94	2.08	24.43	251.7
	Minimum	23.64	64.94	500	247.71	0.09	1.33	26.67
	Maximum	109	344.03	11,644.78	779.69	46.51	58.81	612.5
	Coefficient of Variation	0.21	0.25	0.57	0.22	2.73	0.35	0.32
Tamil Nadu	Average	47.96	228.17	1577.93	508.96	7.17	20.97	224.8
	Minimum	13.92	103.57	129.95	168.25	0.63	3.47	38.33
	Maximum	93.75	741.67	4938.02	1281.25	40	98.77	876.47
	Coefficient of Variation	0.22	0.27	0.65	0.36	0.92	0.65	0.59
Uttar Pradesh	Average	36.19	164.7	1832.87	683.84	48.61	18.19	65.91
	Minimum	12.5	28.75	294.64	291.86	3.03	5.49	10.39
	Maximum	64.29	327.42	8767.12	1396.43	137.14	370.83	790.62
	Coefficient of Variation	0.21	0.34	1.08	0.28	0.8	1.75	0.73
West Bengal	Average	46.86	171.94	1781.97	1021.94	46.15	40.18	101.57
	Minimum	23.58	14.62	30.61	448.13	0.37	1.71	1.28
	Maximum	70.8	533	9066.07	2109.09	245.37	146.34	450
	Coefficient of Variation	0.19	0.39	0.98	0.26	0.89	0.7	0.88

Note: In the models, the yield is represented per hectare; other inputs are represented in a per-farm format.

The most crucial component of the cost of paddy cultivation, i.e., human labor, has an average application of 627.19 person-hours per hectare in the study area. The eastern states used nearly 742 person-hours per hectare, and the northern states engaged 576.29 person-hours per hectare, while the southern states used only 503 person-hours per hectare. This indicates that eastern states are more labor-intensive. Higher agricultural wages in the southern region (Rs.393 per person-day), Rs.274 per person-day in the northern region, and Rs.208 per person-day in the eastern region were recorded. As such, eastern states can easily employ higher human labor to increase production with the same capital.

The analysis illustrates that India's eastern region has a lot of potential for yield, production and productivity through higher input use. For further technical analysis, we applied the stochastic frontier approach to the assessment of the individual farmers' technical efficiency in different states in the study to obtain a comparative view of the potential yield improvement and efficiency distribution.

3.3. The Stochastic Frontier Approach of Technical Efficiency Estimation

The analysis in the previous section shows a clear disparity among Indian states in paddy cultivation methods. A large part of the paddy-producing area is incurring a loss. In order to address this problem, the study first tried to explore the farm-level technical analysis to find the reason behind input mismanagement in those states for 2016–2017. Stochastic frontier models were specified for each state under their technology's present level to determine the paddy production's technical efficiency. The models were specified based on the variables' availability and the Ramsey Reset test specified in the methodology section.

Perusing the stochastic frontier analysis results as presented in Tables 6 and 7, it was observed that, in the Uttar Pradesh area, increase has led to improvement in yield levels, while in the Bihar, Odisha Tamil Nadu, Assam, and Gujarat areas increase has significantly reduced the productivities. Thus, both positive and negative instances of the land size and productivity relationship exist in paddy production across various states of India. Human labor use has shown the highest positive and significant elasticity in Tamil Nadu (0.193), Bihar (0.145) and Odisha (0.127), followed by Gujarat and Uttar Pradesh, which indicates the excess use of human labor in these states, which would have been optimized for the improvement of paddy production. However, in Punjab (−0.271) and Kerala (−0.169), human labor was found to have negative elasticities. Mechanical labor showed significant negative elasticities in West Bengal, Odisha and Assam, mainly due to higher reliance on animal labor, while the coefficient was positively significant only for Tamil Nadu (0.014), where it had a slightly higher contribution in productivity.

Table 7. Maximum likelihood estimates for the Cobb–Douglas type stochastic frontier production function for major paddy cultivating states in India for the AY 2016–2017.

Variables/States	Punjab	Bihar	Uttar Pradesh	West Bengal	Odisha	Andhra Pradesh	Tamil Nadu	Kerala	Assam	Gujarat	Chhattisgarh
(Intercept)	4.363 *** (0.389)	2.527 *** (0.274)	2.817 *** (0.257)	4.081 *** (0.212)	0.872 *** (0.206)	3.994 *** (0.212)	2.641 *** (0.293)	4.571 *** (0.268)	3.448 *** (0.321)	2.408 *** (0.451)	3.744 *** (0.984)
Area under crop (hectare)	−0.011 ns (0.065)	−0.178 *** (0.045)	0.167 *** (0.039)	−0.013 ns (0.029)	−0.524 *** (0.031)	−0.009 ns (0.033)	−0.244 *** (0.046)	0.059 ns (0.046)	−0.089 * (0.042)	−0.345 *** (0.077)	−0.016 ns (0.586)
Human labor (man-hours)	−0.271 *** (0.062)	0.145 *** (0.043)	0.076 * (0.036)	−0.028 ns (0.031)	0.127 *** (0.027)	−0.014 ns (0.027)	0.193 *** (0.035)	−0.169 *** (0.043)	0.076 ns (0.050)	0.125 * (0.068)	0.044 ns (0.399)
Mechanical labor (Hours)	−0.004 ns (0.004)	0.001 ns (0.003)	0.003 ns (0.003)	−0.005 * (0.002)	−0.008 *** (0.001)	−0.004 ns (0.003)	0.014 *** (0.003)	0.002 ns (0.013)	−0.012 ** (0.004)	0.003 ns (0.009)	−0.012 ns (0.020)
Fertilizer (kg.)	0.061 *** (0.010)	0.049 * (0.024)	0.087 *** (0.025)	0.019 *** (0.005)	0.420 *** (0.020)	0.065 * (0.028)	0.042 ns (0.044)	0.080 *** (0.014)	−0.007 * (0.003)	0.126 *** (0.023)	−0.076 ns (0.607)
Irrigation (Hours)	0.152 *** (0.033)	−0.011 *** (0.003)	0.001 (0.004)	0.004 * (0.002)	−0.003 *** (0.003)	0.005 * (0.002)	0.002 ns (0.003)	−0.014 ns (0.014)	0.039 *** (0.004)	0.022 * (0.010)	−0.003 ns (0.047)
Insecticide (Rupees)	0.064 *** (0.014)	0.024 *** (0.004)	0.011 *** (0.002)	0.010 *** (0.002)	0.005 *** (0.001)	0.002 *** (0.003)	0.008 * (0.004)	0.020 *** (0.005)	−0.004 ns (0.004)	0.057 *** (0.008)	0.028 ns (0.019)
Sigma Square (σ^2)	0.100 *** (0.011)	0.045 *** (0.009)	0.063 *** (0.010)	0.084 *** (0.007)	0.011 *** (0.003)	0.116 *** (0.011)	0.121 *** (0.012)	0.286 *** (0.038)	0.140 *** (0.014)	0.505 *** (0.085)	0.097 ns (0.460)
Gamma (γ)	0.971 *** (0.012)	0.617 *** (0.167)	0.548 *** (0.144)	0.890 *** (0.022)	0.282 *** (0.365)	0.924 *** (0.023)	0.953 *** (0.017)	0.909 *** (0.039)	0.909 *** (0.029)	0.968 *** (0.028)	0.990 ns (0.974)
Sigma Square U ($\sigma^2 U$)	0.097 *** (0.011)	0.028 * (0.013)	0.035 * (0.014)	0.075 *** (0.007)	0.003 *** (0.005)	0.107 *** (0.012)	0.116 *** (0.013)	0.260 *** (0.044)	0.128 *** (0.016)	0.489 *** (0.092)	0.096 ns (0.443)
Sigma Square V ($\sigma^2 v$)	0.003 ** (0.001)	0.017 *** (0.004)	0.029 *** (0.005)	0.009 *** (0.002)	0.008 *** (0.002)	0.009 *** (0.002)	0.006 ** (0.002)	0.026 ** (0.009)	0.013 *** (0.003)	0.016 ns (0.012)	0.001 ns (0.096)
Lambda (λ)	5.799 *** (1.200)	1.269 ** (0.449)	1.101 *** (0.320)	2.846 *** (0.327)	0.626 *** (0.565)	3.476 *** (0.562)	4.514 *** (0.885)	3.165 *** (0.748)	3.160 *** (0.554)	5.484 * (2.433)	9.737 ns (459.040)
Log Likelihood	79.409	154.593	86.508	165.151	425.631	66.746	55.821	−78.10	17.204	−65.366	58.189
Mean Technical Efficiency	0.801	0.879	0.868	0.819	0.958	0.793	0.784	0.699	0.768	0.639	0.801
Number of Observations	260	401	487	596	449	422	317	248	448	129	149

Note: "***", "**" and "*" represent significance at the 1%, 5% and 10% levels, respectively. "ns" represents non-significant estimates. Figures in the parenthesis represent the standard error of the estimates.

In Odisha, Tamil Nadu, Assam, Gujarat and Chhattisgarh, human labor contributed positively and significantly to the paddy yield in 2016–2017. However, the states like Punjab, Andhra Pradesh and Kerala showed a negatively significant value, indicating the need to reduce human labor production. Furthermore, Tamil Nadu (0.19) showed the highest elasticity, followed by Bihar (0.132), Gujarat (0.13) and Odisha (0.10), indicating the scope for improvement in these states. In contrast, there was negative elasticity for human labor in Punjab (−0.27), Andhra Pradesh (−0.05) and Kerala (−0.18). Similarly, animal labor was found to have significantly contributed to the paddy yield in states like West Bengal, Odisha, Andhra Pradesh, Tamil Nadu, Assam, Gujarat and Chhattisgarh in the years 2016–2017.

The fertilizer application was found to have significantly contributed to states like Punjab, Bihar, Uttar Pradesh, West Bengal, Odisha, Andhra Pradesh, Kerala and Gujarat, indicating the scope for an enhanced level of fertilizer application for improved paddy yield in 2016–2017. However, Assam showed a negative value, indicating the need to reduce the fertilizer application in paddy production. Furthermore, the magnitude of elasticities shows that the highest value was observed in Odisha (0.420), followed by Uttar Pradesh (0.08), Gujrat (0.087), Andhra Pradesh (0.065), Kerala (0.080), Bihar (0.049) and Punjab (0.061), indicating the scope of improvement of fertilizer use in these states. The negative elasticity in Assam (−0.007) indicates the excess use of fertilizer application, which could have been optimized to improve the paddy yield. The studies of Shanumugan and Venkatramani [18], Bhende and Kalirajan [56], and Dung et al. [57] conform to the results of our study, e.g., that fertilizer and human labor have positive production elasticity in case of paddy production. Except for West Bengal, Odisha and Assam, mechanical labor has not contributed to a variation in paddy yield in 2016–2017 in other paddy-producing states of India. However, in these states there was also negative elasticity for mechanical labor.

As mechanical labor consists of both animal labor and machine labor, it was not directly interpretable. An increase in irrigation hours would have significantly augmented the yield in states like Punjab (0.152), West Bengal (0.004), Andhra Pradesh (0.005), Assam (0.039) and Gujarat (0.022), while increased irrigation hours in Bihar and Odisha would have reduced the yield. In small farms of central Gujarat, a study by Narala and Zala [58] found positive elasticity for irrigation in paddy production, which conforms to our study. Here, it should be noted that in Punjab, more than 98 percent of the irrigation for paddy crops is there, while other states lag behind in irrigation infrastructure development. Thus, the elasticity of the irrigation remained high for Punjab compared to other states. All of the states under study except Assam and Chhattisgarh showed significantly positive estimates for insecticide use, indicating the prevalence of insect pests throughout the country in paddy crop significantly determining yield.

The estimated variance parameters σ^2U and σ^2v in Table 6 are significantly different from zero, which suggested that the difference in the variation of the yield in the paddy production in Punjab, Bihar, West Bengal, Andhra Pradesh, Tamil Nadu, Kerala, and Assam in 2016–2017 was not caused by stochastic error alone but also involved technical inefficiency or inefficiencies in input management. Further, the significant value of γ for Punjab, Bihar, Uttar Pradesh, West Bengal, Odisha, Andhra Pradesh, Tamil Nadu, Kerala, Assam, and Gujarat shows the presence of dominant inefficiency effect over the random error term in all of the states. Among all of the states, Chhattisgarh showed the highest difference of 98 percent between the observed and frontier outputs, followed by Punjab (97%), Gujarat (97%), Tamil Nadu (95%), Assam (90%), Kerala (89%), Andhra Pradesh (93%), West Bengal (89%), Bihar (61%), Uttar Pradesh (54%) and Odisha (28%), which was mainly due to the inefficient use of resources by the farmers in these states. The value of γ also highlighted the percentage of inefficiency due to the factors under the farmers' control. It can be inferred from the estimate that states with a high level of γ have very little opportunity left to adjust production factors. Their yield can only be ameliorated through a complete change in technology in the form of a new variety or some hi-tech production measures. In contrast, states with a lower technical efficiency need to improve

their technical efficiency in order to improve management to achieve potential yield levels in paddy. Lambda (λ), which measures the degree of asymmetry in the distribution of the composite error term ($E_i = V_i - U_i$), was found to be significantly more than one for all of the states except Chhattisgarh in our study. The value of λ illustrates technical inefficiency and a higher magnitude of the one-sided error component U_i in E_i.

The stochastic frontier analysis suggests that in all of the major paddy-growing states except for Chhattisgarh, input management practices entailed the inefficiency in paddy production to varying degrees in 2016–2017. Each input showed a different degree of responsiveness to paddy production, and management must be aimed to optimize the input application. Consequently, a profitable level of paddy production can be achieved in the future. The farm-level technical efficiency estimated from this analysis reveals that India's mean technical efficiency varies from 0.64 in Gujarat to 0.96 in Odisha. The results from the pan-India study across all states by Shanumugan and Venkatramani [18] found that the technical efficiency ranged from 0.77 in Madhya Pradesh to 0.84 in Odisha in 1990–1991. The fact that Odisha farmers are more efficient in utilizing farm resources is due to high cropping intensity [18]. Table 7 divides technical efficiency into four efficiency groups, as delineated in the methodology section across all of the states and size groups of farmers.

The heatmap in Figure 2 illustrates that the marginal farmers of Uttar Pradesh and West Bengal were operating at the highest efficiency level, while Andhra Pradesh and Kerala were at the lowest efficiency level. Small farmers of Andhra Pradesh and Uttar Pradesh showed the highest efficiency in paddy production, while those of Punjab and Tamil Nadu had the lowest efficiency. In semi-medium farm groups, only Kerala showed the lowest efficiency, while Chhattisgarh, Bihar and West Bengal were operating at the highest efficiency level. Medium farms of Punjab and West Bengal were the least efficient in paddy production, while Kerala and Uttar Pradesh employed the highest efficiency level. In large farms, Kerala, Punjab and West Bengal were running at the lowest efficiency level, while those of Tamil Nadu were operating at the highest efficiency level. Overall, we can deduce that ten farm groups performed at the lowest efficiency level, ten at the very-high efficiency level, 16 at the high-efficiency level, and 13 at the medium efficiency level. The state-specific analysis showed that Kerala has the highest instance of low-efficiency farms, while Uttar Pradesh has the highest number of very-high efficiency farmers. Thus, the distribution of technical efficiency suggests that there is a need to improve the efficiency of a significant proportion of farmers, and they belong to any of the farmer classes. The study concludes that efficiency is not concentrated on any specific farm group; instead, it is a discrete phenomenon.

The distribution graph (Figure 3) suggests that in all of the states under study, the proportions of farmers operating at high and very high technical efficiency levels were high except for Gujarat and Kerala, where a significant chunk of farmers was operating at the lowest technical efficiency level. Skewed distribution can be seen in states like Uttar Pradesh, West Bengal, Bihar, Gujarat and Kerala, showing a high level of instability in input management practices.

In the following graphs (Figures 4–8), the study tries to overview the input use of farmers operating in the four efficiency groups. The charts show that the highest efficiency group also has the highest level of yields in paddy production across all of the states of India. The graph shows that, as such, policymakers cannot go for one input management policy because there is no specific pattern of input use among the different efficiency groups which can be standardized for all of the states. Thus, a specific classification cum prediction model to identify efficiency groups should be developed for appropriate input management policy before the cropping season. This may act as a basis to advise on the optimum input levels for specific states.
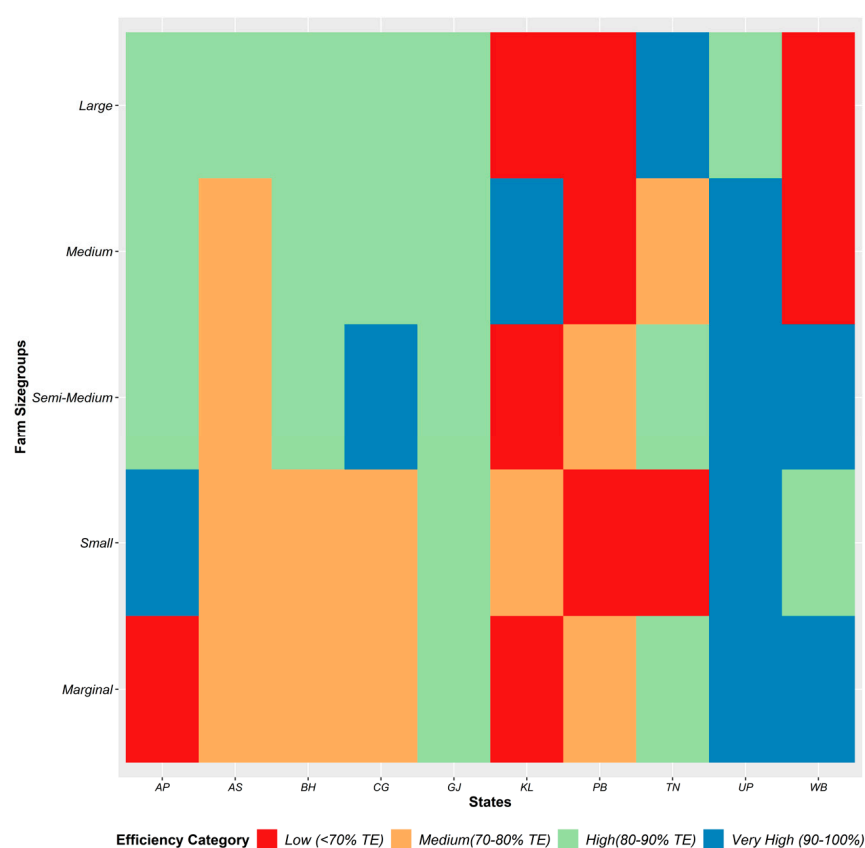
Figure 2. Heatmap showing the distribution of the technical efficiency across the states and farmer classes in paddy production in India in the AY 2016–2017.
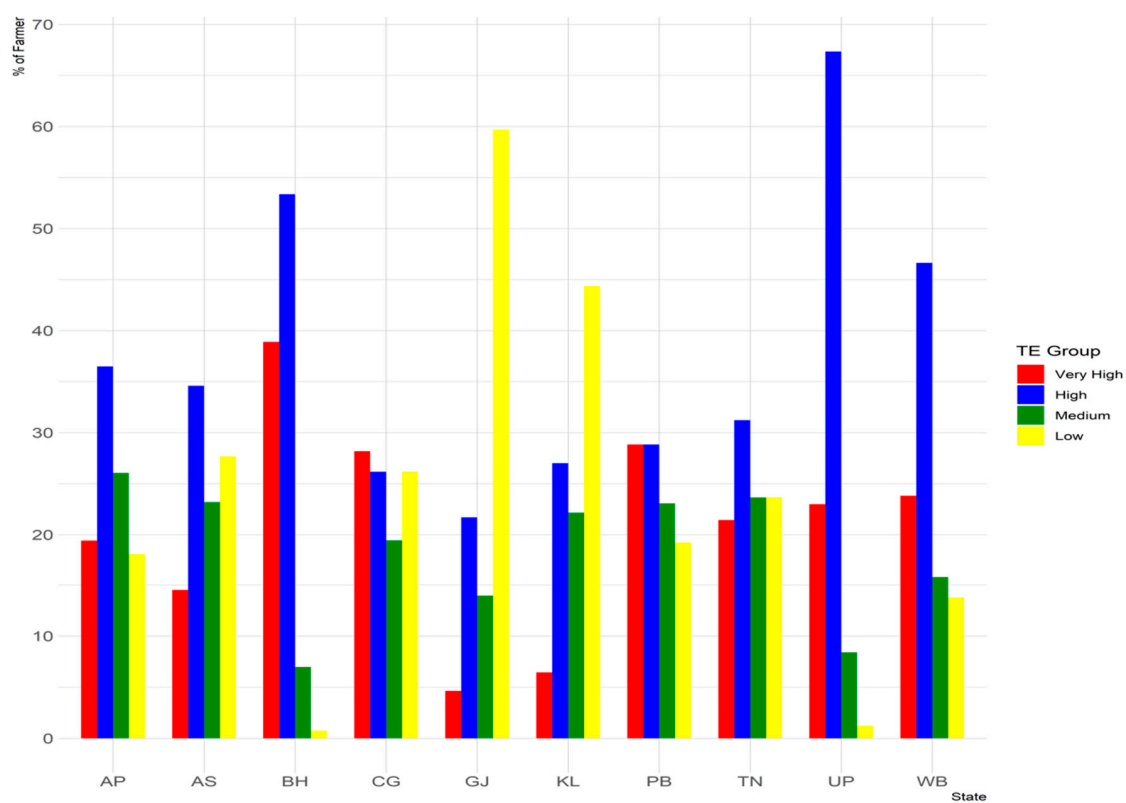

Figure 3. Distribution of technical efficiency in paddy production across various states of India in AY 2016–2017.
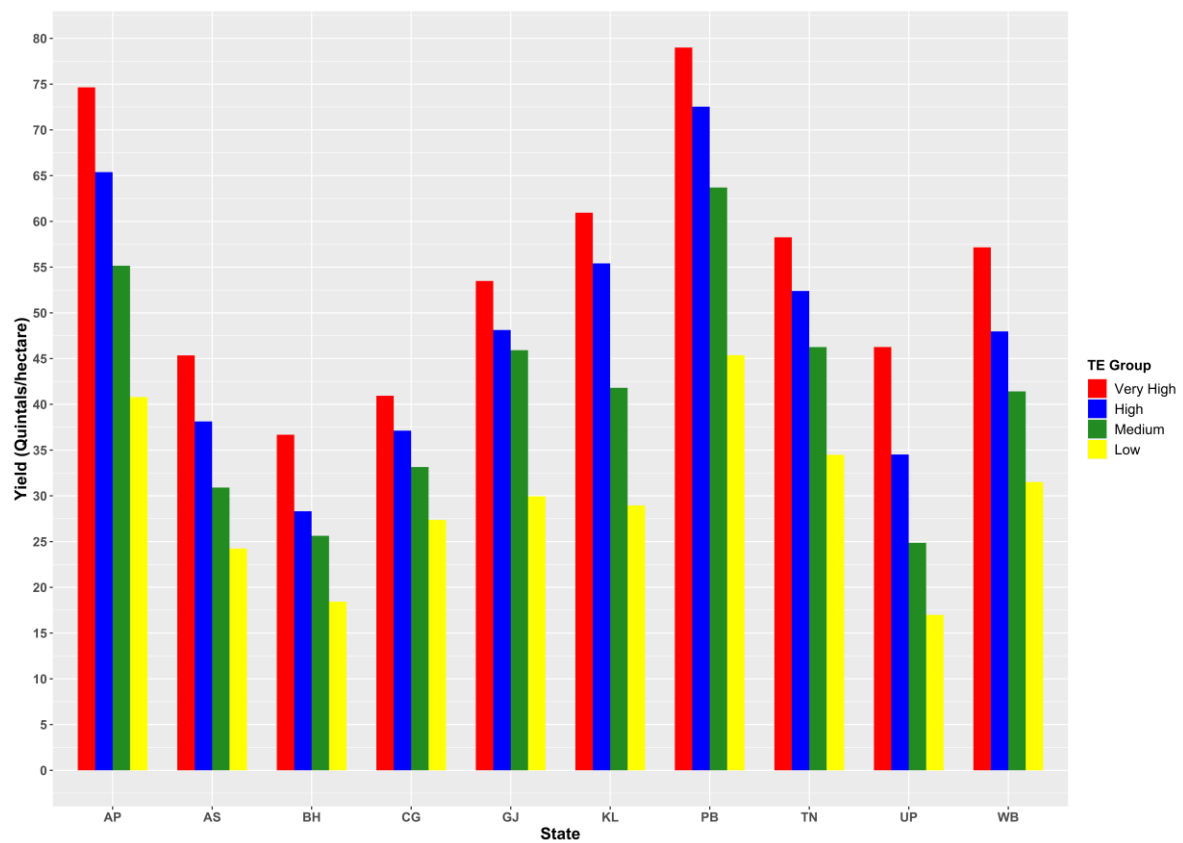
Figure 4. Yield of paddy (quintal/hectare) across various states and efficiency groups.
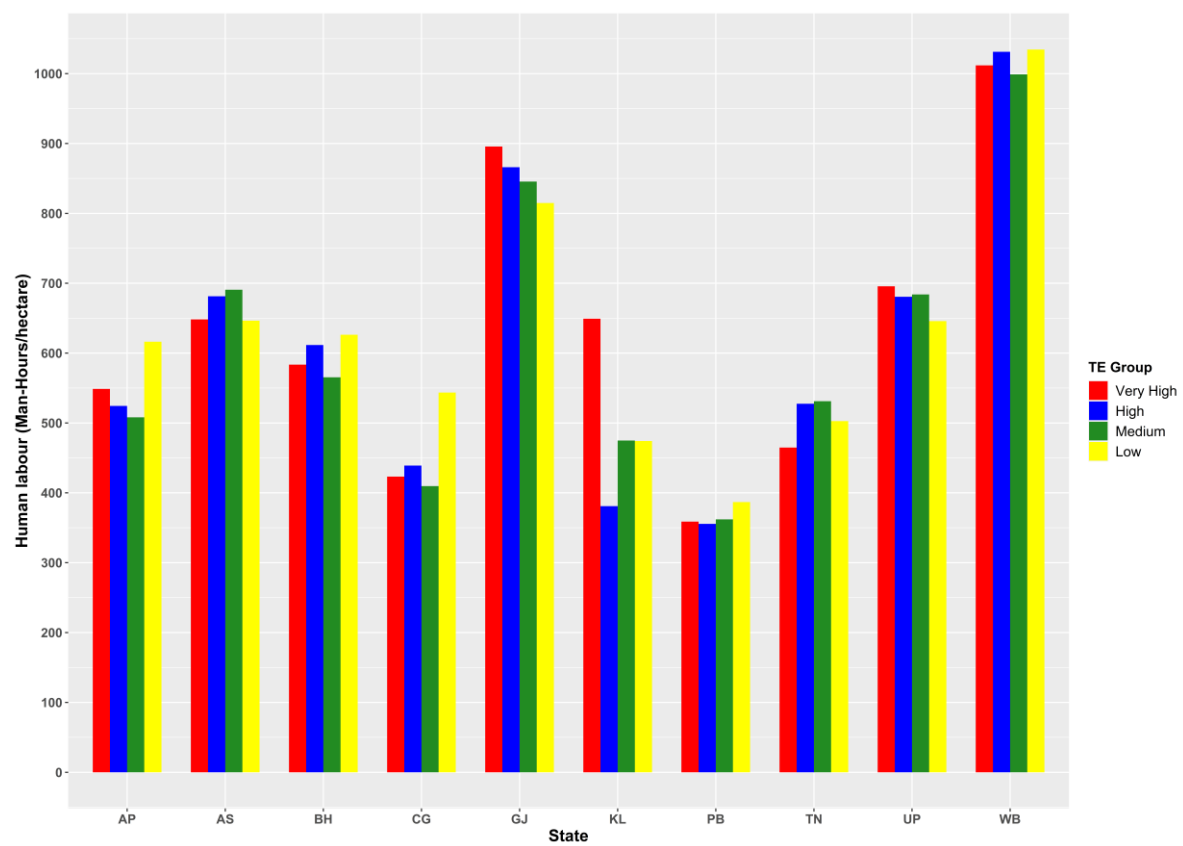

Figure 5. Human labor use in paddy (man-hours/hectare) across various states and efficiency groups.
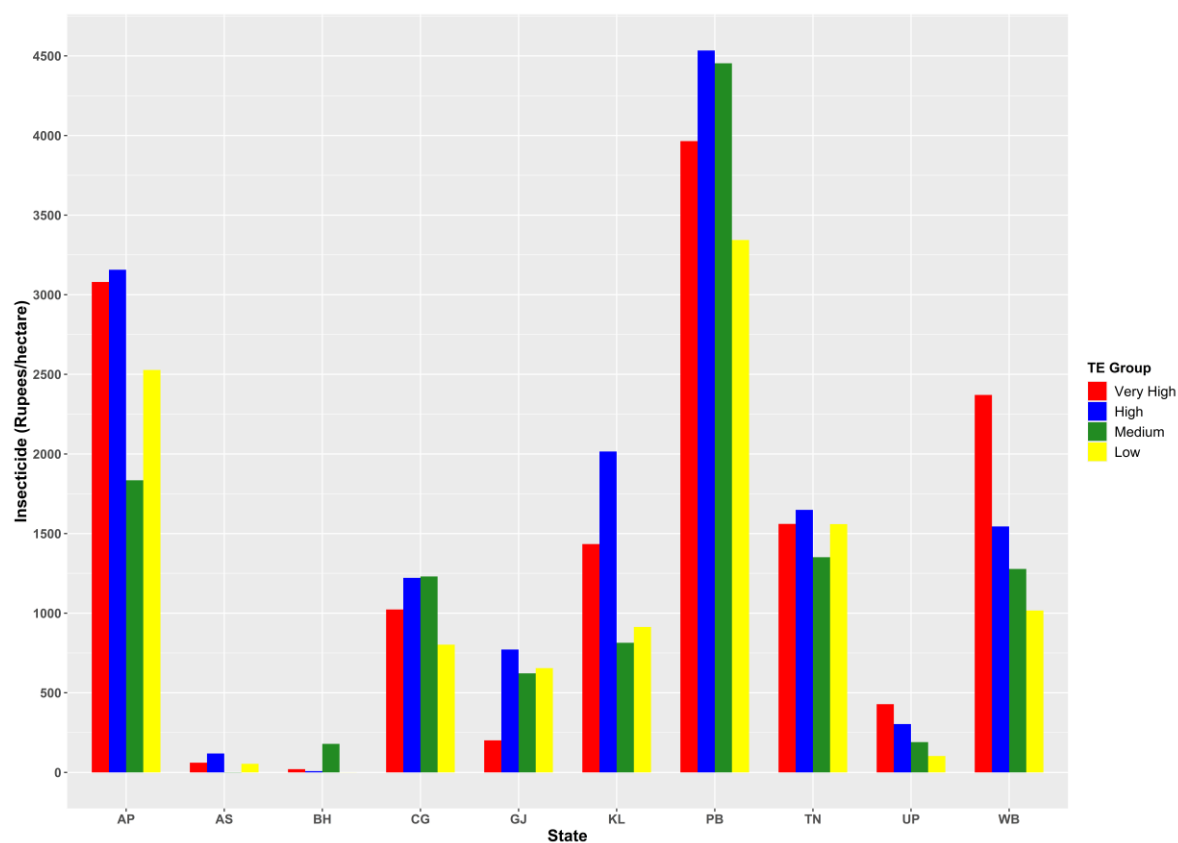
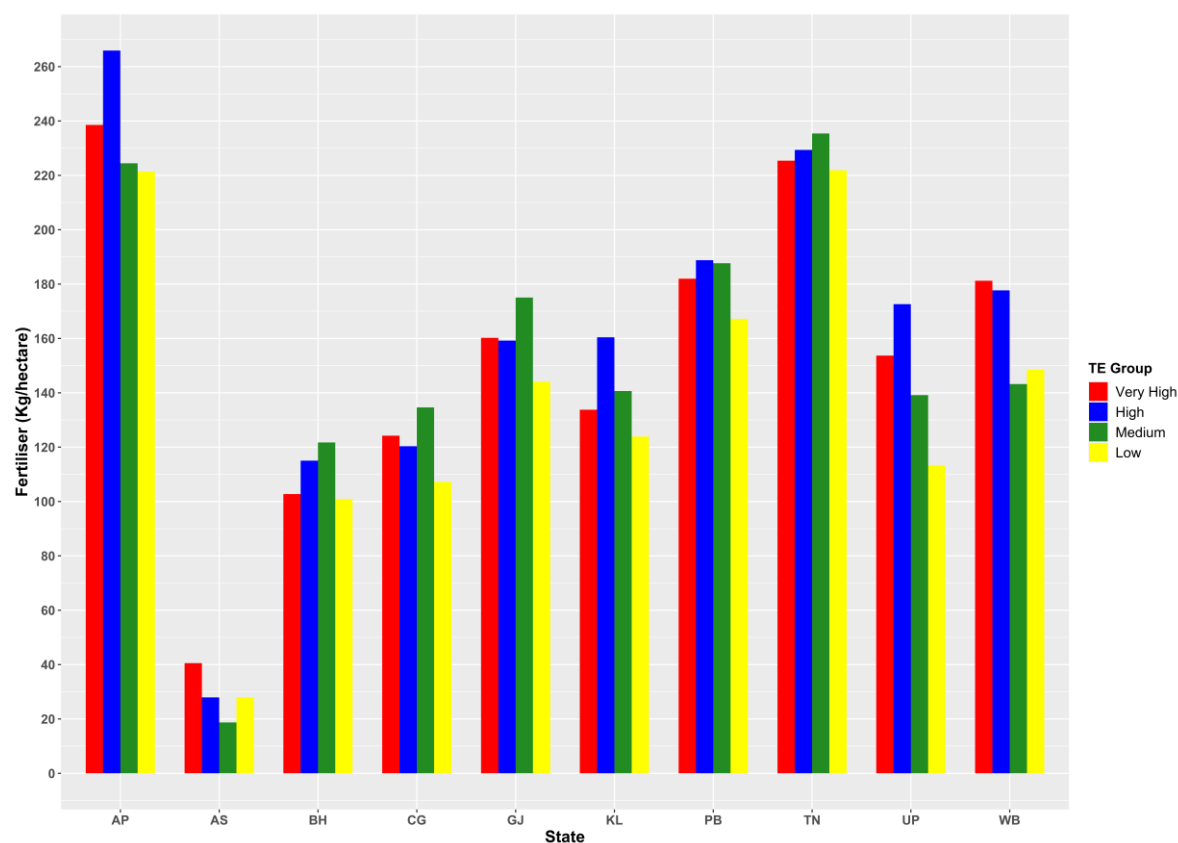**Figure 6.** Insecticide use in paddy (rupees/hectare) across various states and efficiency groups.

Figure 7. Fertilizer use in paddy (kg/hectare) across various states and efficiency groups.
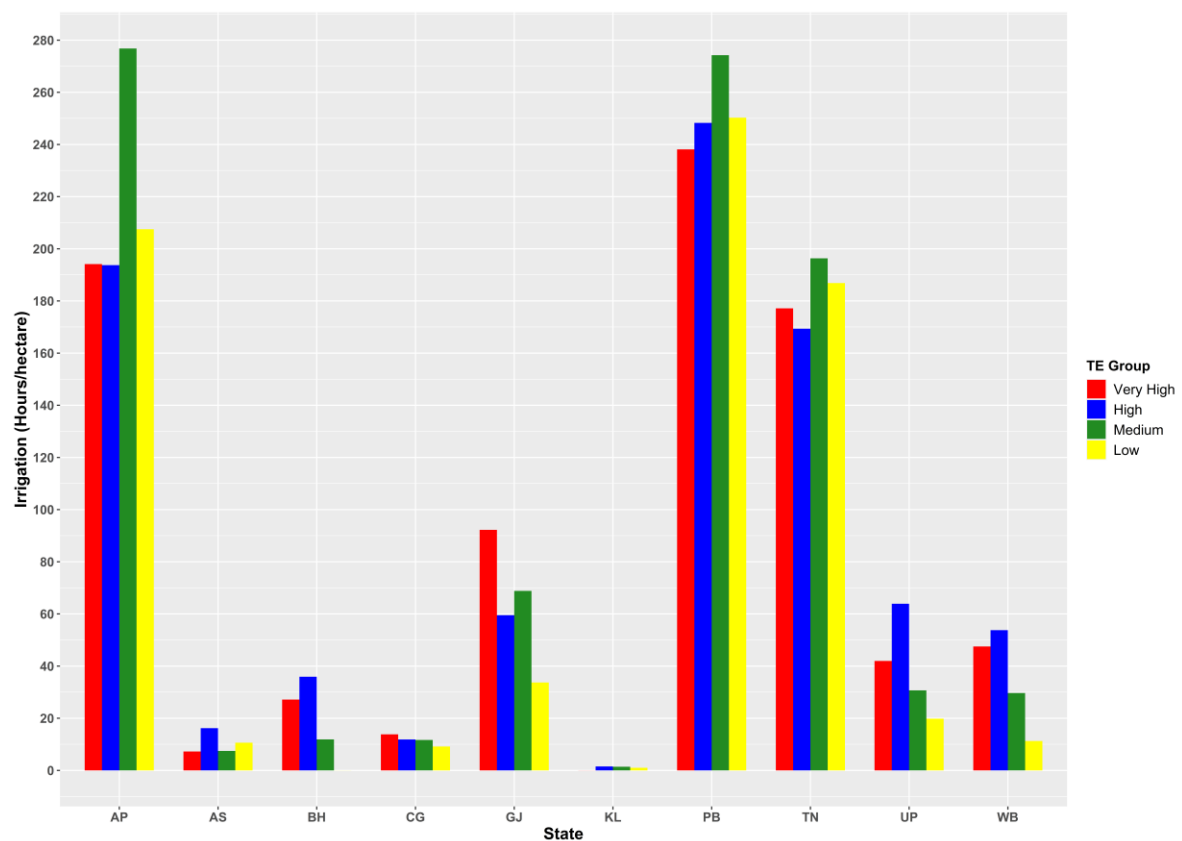
Figure 8. Irrigation in the paddy (hours/hectare) across various states and efficiency groups.

Figure 4 shows that the yield levels directly vary with the technical efficiency group. The study confirms the disparity in yield among the states of India in paddy production. Figure 5 shows that there is no pattern and no striking variation of human labor in relation to the efficiency group; Gujarat and West Bengal use the highest human labor hours among all of the states.

Figure 7 shows that there is very high variation in insecticide use among the states under study, with Andhra Pradesh and Punjab being the highest user of insecticide. There is no distinctive pattern of insecticide use among various size groups of farmers across the states. In the case of fertilizer, there seems to be no specific pattern of difference to classify the technical efficiency group (Figure 7), and the same can be observed in the case of irrigation hours (Figure 8).

Thus, the visualization of the efficiency group and related input parameters is insufficient to provide a classification based on technical efficiency, and more sophisticated methods are needed to map the pattern of input use with respect to the efficiency group. The next section employs machine learning algorithms on various parameters discussed in the methodology section to find an accurate solution to the classification problem.

3.4. Machine Learning Models for Efficiency Group Prediction

The previous analysis suggests that there is disparity among the states regarding paddy production technology, which leads to various levels of yield. The intra-state variation of yield among various farm size groups was also found from the study. Thus, input management that forms a major policy issue to target farmers needs to be tailored to state and size groups. The stochastic frontier approach concluded that there exist four efficiency levels of which the input management and yield levels differ. Scientists have employed linear programming models to determine the input levels in the past. Still, as new methodologies are being introduced, we have to check their applicability in input management in agriculture. This will open new avenues for intelligent decision-making

in agriculture. Thus, a machine learning model predicting the efficiency level of a farm, given the input levels, would be advantageous to manage farm inputs to achieve yield-augmenting objectives of the states.

Considering all of these advantages, the current study employed three standard classification algorithms that take input data and size group labels and predict the efficiency group for specific states. The tenfold cross validation method was used to compare the model accuracy of the KNN, SVM and RF algorithms, and the mean accuracy and kappa statistics are presented in Table 7. The mean accuracy for the KNN method ranges between 0.306 for Punjab to 0.685 in Uttar Pradesh, while that of SVM was in the range of 0.518 in Tamil Nadu to 0.848 in Uttar Pradesh. The accuracy statistics of the random forest model varied between 0.729 in Punjab to 0.943 in Uttar Pradesh. Overall, the random forest model was the best model for our dataset across all of the states. The dataset for Uttar Pradesh had the best response to all of the three models, while that of Punjab was the worst. The random forest model is the most accurate for the classification and prediction objectives in our case, as shown in Table 8.

Table 8. Comparison of the KNN, SVM and random forest algorithms for their accuracy in classifying efficiency groups in paddy production across major paddy-producing states of India in AY 2016–2017.

	Mean Accuracy from 10 Resamples			Mean Kappa Values from 10 Resamples		
State/Models	KNN	SVM	Random Forest	KNN	SVM	Random Forest
PB	0.306	0.595	0.729	0.072	0.451	0.646
BH	0.514	0.802	0.857	0.086	0.62	0.744
UP	0.685	0.848	0.943	0.247	0.644	0.882
WB	0.449	0.797	0.916	0.17	0.689	0.876
AP	0.399	0.767	0.843	0.149	0.671	0.784
TN	0.34	0.518	0.701	0.104	0.335	0.597
KL	0.478	0.611	0.800	0.200	0.38	0.706
AS	0.353	0.779	0.874	0.094	0.692	0.828
GJ	0.579	0.632	0.786	0.086	0.186	0.628
CG	0.334	0.531	0.795	0.108	0.361	0.725

Note: Due to model misspecification for Odisha in the standard production function, only the highest efficiency class was present and hence excluded from the classification study.

Table 8 shows that the random forest model's mean accuracy and kappa values (with 10-fold cross-validation) across all of the states remained higher than the KNN and SVM algorithms. Thus, the random forest model was chosen for classification and prediction of the efficiency groups across the selected states of India in our study. Detailed accuracy statistics are presented in Table 9. A detailed performance evaluation measure for the KNN, SVM and RF models can be found in the Table A1.

Table 9 confirms that the random forest algorithm with 10-fold cross-validation predicted the efficiency group, given the input data in nine of the ten states in the classification study, with a mean accuracy of 80 percent. The highest accuracy of the RF model was observed for the data of Bihar (0.730), and the lowest was observed in the case of Gujarat (0.667). The highest variation in accuracy was observed in case of case of Gujarat (0.45–0.84), followed by Chhattisgarh (0.50–0.86) and Punjab (0.59–0.84). The model can be further improvised by taking more features from soil fertility, soil properties, soil moisture, weather and satellite data for all of the states across time in order to improve the accuracy and reduce the NIR. This kind of model will predict the efficiency level and augment the yield in upcoming crops by making suggestions to farmers on their level of input use, and is hence a very effective tool in the hand of stakeholders to mitigate risk in agriculture.

Table 9. Accuracy statistics of random forest models in the classification of the efficiency groups in paddy production across the major producing states of India in AY 2016–2017.

States	Accuracy	95% CI	NIR	Kappa
PB	0.730	(0.589,0.844)	0.289 ***	0.638
BH	0.910	(0.824,0.963)	0.539 ***	0.834
UP	0.885	(0.804,0.942)	0.677 ***	0.740
WB	0.863	(0.787,0.919)	0.470 ***	0.801
AP	0.880	(0.789,0.941)	0.361 ***	0.835
TN	0.776	(0.634,0.882)	0.449 ***	0.667
KL	0.710	(0.581,0.818)	0.301 ***	0.614
AS	0.875	(0.787,0.936)	0.352 ***	0.827
GJ	0.667	(0.447,0.844)	0.625 NS	0.407
CG	0.704	(0.498,0.863)	0.296 ***	0.598

Note: NIR means no information rate, and is significant when accuracy > no information rate. "***" means significance at the 5% level. "NS" represents non-significant estimates. Figures in the parenthesis represent the standard error of the estimates.

4. Conclusions

Among the major paddy-growing states in the study, the production percentage is greater than the area percentage with higher productivity. The analysis suggests that higher investments in fixed-cost components like mechanization have strongly contributed to higher productivity, proving that the rural infrastructure is crucial for productivity; in other words, investment policies are bearing fruit in the areas that have benefitted from them. Input management and other efficiency-related measures can also be used to raise productivity in the states that are instead falling behind. In our analysis, all of the eastern states, except West Bengal, were below the average national yield of 41.31 kg per hectare; the southern region's average yield is 49.68 kg/ha, and the northern region's is 46.84 kg/ha. This disparity calls for a targeted approach in terms of varietal development and input management. The capital-intensive northern (except Himachal Pradesh) and southern regions have considerably higher fertilizer application rates than the eastern region (more than 1.5–2.5 times). Furthermore, the insecticide use in the northern region is Rs.2069.30 per hectare, second to the southern region (Rs.2127.51 per hectare). The least insecticide use was reported in the eastern region (Rs.980.60 per hectare). However, paddy cultivation labor hours were considerably higher in eastern states than in the northern and southern regions because of lower mechanization and the higher use of human labor. Thus, the policy for input management should be tailored to the context—i.e., capital-intensive north and southern regions, and the labor-intensive eastern part. Our analysis suggests that the yield can be increased by means of a more efficient input management. The yield can be improved for different states in the range of 4.2 percent in Odisha to 36.1 percent in Gujarat with the optimum use of inputs under the current level of technology of the specific states.

Input management inefficiency is responsible for lower yields. The technical efficiency figure of individual farmers suggests that the very-high efficiency level was achieved in our sample by ten farm groups, the high-efficiency level by 16 groups, the medium efficiency by 13 farm groups, and the lowest efficiency by ten farm groups. Low efficiency is more common in medium and large farm groups due to input management issues, rather than input availability. Overall, the study concludes that efficiency is not concentrated on any specific farm group; rather, it is a discrete phenomenon. From the maximum likelihood estimates of all of the states (except Chhattisgarh), significant inefficiency due to input management is visible, causing yield variation. In addition, many states are operating at a very high technical efficiency level, and saturation in the current status of technical efficiency has already occurred, as confirmed from high gamma estimates, which means that there is little room for improvement. In states like Odisha, Bihar and Uttar Pradesh, the

gamma values are low enough to accommodate higher inputs considering the technology level. The states with high gamma values need to improve their technology to increase their yield. Thus, the study suggests a targeted approach for states and regions regarding the input management in the short term, and a technological shift in the next years to keep farms at a profitable level. The study suggests that random forest algorithm is best suited for this dataset across all of the states under study. The random forest algorithm we used suggests that 66 percent of farmers in Gujarat to 91 percent in Bihar can be correctly associated with the achieved efficiency levels using only the farmers' input and size group features. The random forest algorithm is highly significant in predicting the efficiency levels in nine of the ten states. In the future, the development of a targeted random forest algorithm can be considered for each state to achieve higher accuracies, especially considering additional features. As the scope of the dataset of this study is limited, we recommend using more features from published datasets on soil fertility, soil moisture, satellite data and weather data to improve the accuracy and NIR. Future studies can use this dataset across time to develop other algorithms used in this field of work. Such a study can help policymakers predict farmers' efficiency levels through input application data before the cropping season, in turn providing support for policies for targeted input management for each specific state operating under different levels of technology.

Appendix A

Table A1. Performance Evaluation Measures of the KNN, SVM and RF models for various states for the classification and prediction of low, medium, high and very high efficiency classes.

State	Punjab											
Model	KNN				SVM				RF			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	0.167	0.770	0.190	0.455	0.900	0.969	0.667	0.936	0.900	0.985	0.800	0.968
Recall	0.100	0.877	0.267	0.484	1.000	0.710	0.979	0.825	0.976	0.710	0.979	0.947
Sensitivity	0.100	0.877	0.267	0.484	1.000	0.875	0.909	0.744	0.900	0.877	0.923	0.909
Specificity	0.881	0.452	0.638	0.684	0.900	0.969	0.667	0.936	0.900	0.985	0.800	0.968
State	Andhra Pradesh											
Model	KNN				SVM				RF			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	NA	0.508	0.682	0.000	NA	0.855	0.864	0.200	NA	0.873	0.864	0.600
Recall	NA	0.564	0.682	0.000	1.000	0.823	0.741	0.842	0.987	0.919	0.889	0.790
Sensitivity	NA	0.564	0.682	0.000	NA	0.810	0.731	0.250	0.000	0.906	0.864	0.429
Specificity	1.000	0.516	0.741	0.947	NA	0.855	0.864	0.200	NA	0.873	0.864	0.600
States	Assam											
Model	KNN				SVM				RF			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	0.536	0.333	1.000	0.000	0.871	0.818	0.000	0.000	0.903	0.818	1.000	0.400
Recall	0.484	0.273	0.333	0.000	0.957	0.869	1.000	1.000	0.957	0.967	1.000	1.000
Sensitivity	0.484	0.273	0.333	0.000	0.931	0.692	NA	NA	0.933	0.900	1.000	1.000
Specificity	0.723	0.803	1.000	0.864	0.871	0.818	0.000	0.000	0.903	0.818	1.000	0.400
States	Bihar											
Model	KNN				SVM				RF			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	0.111	0.529	0.300	0.000	0.500	0.773	0.467	0.538	0.667	0.727	0.800	0.846

Table A1. *Cont.*

State	Bihar											
Model	**KNN**				**SVM**				**RF**			
Recall	0.083	0.409	0.200	0.000	0.875	0.987	0.787	1.000	0.900	1.000	0.766	1.000
Sensitivity	0.083	0.409	0.200	0.000	0.546	0.944	0.412	1.000	0.667	1.000	0.522	1.000
Specificity	0.800	0.892	0.851	0.987	0.500	0.773	0.467	0.538	0.667	0.727	0.800	0.846
States	**Chhattisgarh**											
Model	**KNN**				**SVM**				**RF**			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	0.200	0.455	0.227	0.286	0.500	0.688	0.650	0.500	0.625	0.938	0.700	0.625
Recall	0.125	0.313	0.250	0.250	0.977	0.985	0.897	0.684	0.977	0.955	0.956	0.895
Sensitivity	0.125	0.313	0.250	0.250	0.667	0.917	0.650	0.400	0.714	0.833	0.824	0.714
Specificity	0.955	0.910	0.750	0.737	0.500	0.688	0.650	0.500	0.625	0.938	0.700	0.625
States	**Gujarat**											
Model	**KNN**				**SVM**				**RF**			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	NA	0.535	0.438	0.250	0.000	0.900	0.833	0.286	0.000	0.833	0.917	0.714
Recall	0.000	0.767	0.583	0.286	1.000	0.830	0.969	0.700	1.000	0.906	0.922	0.900
Sensitivity	0.000	0.767	0.583	0.286		0.750	0.909	0.250	NA	0.833	0.815	0.714
Specificity	1.000	0.623	0.719	0.700	0.000	0.900	0.833	0.286	0.000	0.833	0.917	0.714
States	**Kerala**											
Model	**KNN**				**SVM**				**RF**			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	0.638	0.273	0.444	0.556	0.952	0.600	0.692	0.714	0.952	1.000	0.846	1.000
Recall	0.714	0.200	0.615	0.714	0.778	1.000	0.889	0.800	0.889	1.000	0.917	0.800
Sensitivity	0.714	0.200	0.615	0.714	0.833	1.000	0.692	0.556	0.909	1.000	0.786	0.636
Specificity	0.528	0.882	0.722	0.800	0.952	0.600	0.692	0.714	0.952	1.000	0.846	1.000

Table A1. *Cont.*

States	Tamil Nadu											
Model	**KNN**				**SVM**				**RF**			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	0.000	0.393	0.125	NA	0.200	0.786	0.455	0.000	0.600	0.893	0.455	1.000
Recall	0.000	0.393	0.091	0.000	1.000	0.978	0.868	1.000	1.000	0.966	0.868	1.000
Sensitivity	0.000	0.393	0.091	0.000	1.000	0.917	0.500	NA	1.000	0.893	0.500	1.000
Specificity	0.959	0.809	0.816	1.000	0.200	0.786	0.455	0.000	0.600	0.893	0.455	1.000
States	**Uttar Pradesh**											
Model	**KNN**				**SVM**				**RF**			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	0.353	0.333	0.320	0.682	0.333	0.875	0.684	0.867	0.733	1.000	0.474	0.800
Recall	0.400	0.313	0.421	1.000	0.757	0.980	0.558	0.333	0.811	0.960	0.907	0.778
Sensitivity	0.400	0.313	0.421	1.000	0.357	0.875	0.406	0.684	0.611	0.800	0.692	0.857
Specificity	0.703	0.901	0.605	0.222	0.333	0.875	0.684	0.867	0.733	1.000	0.474	0.800
States	**West Bengal**											
Model	**KNN**				**SVM**				**RF**			
Class	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE	Low TE	Medium TE	High TE	Very High TE
Precision	0.300	0.385	0.333	0.000	0.467	0.667	0.077	0.000	0.667	0.667	0.846	0.000
Recall	0.400	0.278	0.154	0.000	0.703	0.929	0.980	0.952	0.946	0.960	0.959	0.905
Sensitivity	0.400	0.278	0.154	0.000	0.389	0.632	0.500	0.000	0.833	0.750	0.846	0.000
Specificity	0.622	0.919	0.918	0.952	0.467	0.667	0.077	0.000	0.667	0.667	0.846	0.000

Table A2. Abbreviations for states.

Abbreviation	State Name
AP	Andhra Pradesh
AS	Assam
BH	Bihar
CG	Chhattisgarh
GJ	Gujarat
KL	Kerala
PB	Punjab
TN	Tamil Nadu
UP	Uttar Pradesh
WB	West Bengal

Table A3. Abbreviations for institutions, machine learning models and other technical terms.

Abbreviation	Full Form
TE	Technical Efficiency
DES	Department of Economics and Statistics
CACP	Commission for Agricultural Cost and Prices
FAO	Food and Agriculture Organization
KNN	K- Nearest Neighbor
SVM	Support Vector Machine
RF	Random Forest
ANN	Artificial Neural Network
MLR	Multiple Linear Regression
SVR	Support Vector Regression
MANN	Modular Artificial Neural Networks
RGB	Red, Green, Blue
UAV	Unmanned Aerial Vehicle

References

1. Bacco, M.; Barsocchi, P.; Ferro, E.; Gotta, A.; Ruggeri, M. The Digitisation of Agriculture: A Survey of Research Activities on Smart Farming. *Array* **2019**, *3–4*, 100009. [CrossRef]
2. FAO. *AQUASTAT Core Database*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2018.
3. *Handbook of Indian Economy*; Reserve Bank of India Publications: Mumbai, India, 2018.
4. Singh, S. A study on technical efficiency of wheat cultivation in Haryana. *Agric. Econ. Res. Rev.* **2007**, *20*, 127–136.
5. Gurjar, M.L.; Varghese, K.A. Structural Changes over Time in Cost of Cultivation of Major Rabi Crops in Rajasthan. *Indian J. Agric. Econ.* **2005**, *60*, 249–263.
6. Narayanamoorthy, A. Profitability in crops cultivation in India: Some evidence from cost of cultivation survey data. *Indian J. Agric. Econ.* **2013**, *68*, 104–121.
7. Guptha, C.; Raghu, P.T.; Aditi, N.; Kalaiselvan, N.N. Comparative trend analysis in cost of paddy cultivation in profitability across three states of India. *Eur. Sci. J.* **2014**, 271. [CrossRef]
8. Reddy, E.L.; Reddy, D.R. A Study on Resource use Efficiency of Agricultural Input Factors with Reference to Farm Size in Three Revenue Mandals of Nellore District: Andhra Pradesh. *Glob. J. Manag. Bus. Res.* **2014**, *17*, 48–55.
9. Kalamkar, S.S.; Narayanamoorthy, A. Impact of Liberalisation on Domestic Agricultural Prices and Farm Income: An Analysis across States and Crops. *Indian J. Agric. Econ.* **2003**, *58*, 353–364.
10. Narayanamoorthy, A. Relief Package for farmers: Can it stop suicides? *Econ. Polit. Wkly.* **2006**, *41*, 3353–3355.

11. Narayanamoorthy, A. Deceleration in Agricultural Growth: Technology Fatigue or Policy Fatigue? *Econ. Polit. Wkly.* **2007**, *42*, 2375–2379.

12. Sainath, P. Farm Suicides: A 12-Year Saga. *The Hindu*, 25 January 2010.

13. Ali, M.; Chaudhry, M.A. Inter-Regional Farm Efficiency in Pakistan's Punjab: A Frontier Production Function Study. *J. Agric. Econ.* **1990**, *41*, 62–74. [CrossRef]

14. Umesh, K.B.; Bisaliah, S. Efficiency of groundnut production in Karnataka: Frontier profit function approach. *Indian J. Agric. Econ.* **1991**, *46*, 20–33.

15. Gaddi, G.M.; Mundinasmani, S.M.; Hiremath, G.K. Resource use efficiency in groundnut production in Karnataka—An economic analysis. *Agric. Situat. India* **2002**, *58*, 517–522.

16. KalirajanK, P.; Shand, R.T. A generalized measure of technical efficiency. *Appl. Econ.* **1989**, *21*, 25–34. [CrossRef]

17. Kalirajan, K.; Obwona, M.; Zhao, S. A Decomposition of Total Factor Productivity Growth: The Case of Chinese Agricultural Growth before and after Reforms. *Am. J. Agric. Econ.* **1996**, *78*, 331–338. [CrossRef]

18. Shanmugam, K.R.; Venkataramani, A. *Technical Efficiency in Agricultural Production and Its Determinants: An Exploratory Study at the District Level*; Madras School of Economics: Tamil Nadu, India, 2006; Volume 10.

19. Zecca, F. The Use of Internet of Things for the Sustainability of the Agricultural Sector: The Case of Climate Smart Agriculture. *Int. J. Civ. Eng. Technol.* **2019**, *10*, 494–501.

20. Helfer, G.A.; Victória Barbosa, J.L.; dos Santos, R.; da Costa, A. Ben A computational model for soil fertility prediction in ubiquitous agriculture. *Comput. Electron. Agric.* **2020**, *175*, 105602. [CrossRef]

21. Ge, X.; Wang, J.; Ding, J.; Cao, X.; Zhang, Z.; Liu, J.; Li, X. Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. *PeerJ* **2019**, *7*, e6926. [CrossRef]

22. Yamaç, S.S.; Şeker, C.; Negiş, H. Evaluation of machine learning methods to predict soil moisture constants with different combinations of soil input data for calcareous soils in a semi arid area. *Agric. Water Manag.* **2020**, *234*, 106121. [CrossRef]

23. Sanuade, O.A.; Hassan, A.M.; Akanji, A.O.; Olaojo, A.A.; Oladunjoye, M.A.; Abdulraheem, A. New empirical equation to estimatethe soil moisture content based on thermal properties using machine learning techniques. *Arab. J. Geosci.* **2020**, *13*, 377. [CrossRef]

24. Helfer, G.; Barbosa, J.; Alves, D.; da Costa, A.; Beko, M.; Leithardt, V. Multispectral Cameras and Machine Learning Integrated into Portable Devices as Clay Prediction Technology. *J. Sens. Actuator Netw.* **2021**, *10*, 40. [CrossRef]

25. Martini, B.; Helfer, G.; Barbosa, J.; Modolo, R.E.; da Silva, M.; de Figueiredo, R.; Mendes, A.; Silva, L.; Leithardt, V. IndoorPlant: A Model for Intelligent Services in Indoor Agriculture Based on Context Histories. *Sensors* **2021**, *21*, 1631. [CrossRef]

26. Ramesh, S.; Vydeki, D. Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. *Inf. Process. Agric.* **2020**, *7*, 249–260. [CrossRef]

27. Li, D.; Wang, R.; Xie, C.; Liu, L.; Zhang, J.; Li, R.; Wang, F.; Zhou, M.; Liu, W. A Recognition Method for Rice Plant Diseases and Pests Video Detection Based on Deep Convolutional Neural Network. *Sensors* **2020**, *20*, 578. [CrossRef]

28. He, Y.; Zhou, Z.; Tian, L.; Liu, Y.; Luo, X. Brown rice planthopper (*Nilaparvata lugens* Stal) detection based on deep learning. *Precis. Agric.* **2020**, *21*, 1385–1402. [CrossRef]

29. Huang, H.; Deng, J.; Lan, Y.; Yang, A.; Deng, X.; Wen, S.; Zhang, H.; Zhang, Y. Accurate Weed Mapping and Prescription Map Generation Based on Fully Convolutional Networks Using UAV Imagery. *Sensors* **2018**, *18*, 3299. [CrossRef] [PubMed]

30. Dadashzadeh, M.; Abbaspour-Gilandeh, Y.; Mesri-Gundoshmian, T.; Sabzi, S.; Hernández-Hernández, J.L.; Hernández-Hernández, M.; Arribas, J.I. Weed Classification for Site-Specific Weed Management Using an Automated Stereo Computer-Vision Machine-Learning System in Rice Fields. *Plants* **2020**, *9*, 559. [CrossRef]

31. Shidnal, S.; Latte, M.V.; Kapoor, A. Crop yield prediction: Two-tiered machine learning model approach. *Int. J. Inf. Technol.* **2019**, 1–9. [CrossRef]

32. Yang, Q.; Shi, L.; Han, J.; Zha, Y.; Zhu, P. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crop. Res.* **2019**, *235*, 142–153. [CrossRef]

33. Wan, L.; Cen, H.; Zhu, J.; Zhang, J.; Zhu, Y.; Sun, D.; Du, X.; Zhai, L.; Weng, H.; Li, Y.; et al. Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer—A case study of small farmlands in the South of China. *Agric. For. Meteorol.* **2020**, *291*, 108096. [CrossRef]

34. Son, N.T.; Chen, C.F.; Chen, C.R.; Guo, H.Y.; Cheng, Y.S.; Chen, S.L.; Lin, H.S.; Chen, S.H. Machine learning approaches for rice crop yield predictions using time-series satellite data in Taiwan. *Int. J. Remote Sens.* **2020**, *41*, 7868–7888. [CrossRef]

35. Amaratunga, V.; Wickramasinghe, L.; Perera, A.; Jayasinghe, J.; Rathnayake, U. Artificial Neural Network to Estimate the Paddy Yield Prediction Using Climatic Data. *Math. Probl. Eng.* **2020**, *2020*, 8627824. [CrossRef]

36. Khosla, E.; Dharavath, R.; Priya, R. Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. *Environ. Dev. Sustain.* **2020**, *22*, 5687–5708. [CrossRef]

37. Nesarani, A.; Ramar, R.; Pandian, S. An efficient approach for rice prediction from authenticated Block chain node using machine learning technique. *Environ. Technol. Innov.* **2020**, *20*, 101064. [CrossRef]

38. Elavarasan, D.; Vincent PM, D.R.; Srinivasan, K.; Chang, C.-Y. A Hybrid CFS Filter and RF-RFE Wrapper-Based Feature Extraction for Enhanced Agricultural Crop Yield Prediction Modeling. *Agriculture* **2020**, *10*, 400. [CrossRef]

39. Gopal, P.M.; Bhargavi, R. A novel approach for efficient crop yield prediction. *Comput. Electron. Agric.* **2019**, *165*, 104968. [CrossRef]

40. Maya Gopal, P.S.; Bhargavi, R. Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms. *Appl. Artif. Intell.* **2019**, *33*, 621–642.

41. Cost of Cultivation Plot Level Summary Reports. Available online: https://eands.dacnet.nic.in/Plot-Level-Summary-Data.htm (accessed on 5 January 2021).

42. *Manual of Cost of Cultivation Surveys*; Ministry of Statistics and Programme Implementation, Government of India: New Delhi, India, 2008.

43. Meeusen, W.; van den Broeck, J. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *Int. Econ. Rev.* **1977**, *18*, 435–444. [CrossRef]

44. Aigner, D.; Lovell, C.A.K.; Schmidt, P. Formulation and estimation of stochastic frontier production function models. *J. Econ.* **1977**, *6*, 21–37. [CrossRef]

45. Battese, G.E.; Coelli, T.J. Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *J. Prod. Anal.* **1992**, *3*, 153–169. [CrossRef]

46. Coelli, T.; Rao, D.S.; Battese, G.E. *An Introduction to Efficiency and Productivity Analysis*; Kluwer Academic Publishers: New York, NY, USA, 1998.

47. Coelli, T.; Henningsen, A.; Henningsen, M.A. Package 'Frontier'; R Package Version 1.1-8.2020. Available online: https://CRAN.R-Project.org/package=frontier (accessed on 12 January 2021).

48. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002; ISBN 0-387-95457-0.

49. Max Kuhn.caret: Classification and Regression Training. R Package Version 6.0-86. 2020. Available online: https://CRAN.R-project.org/package=caret (accessed on 12 January 2021).

50. Selvaraj, M.G.; Vergara, A.; Montenegro, F.; Ruiz, H.A.; Safari, N.; Raymaekers, D.; Ocimati, W.; Ntamwira, J.; Tits, L.; Omondi, A.B.; et al. Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 110–124. [CrossRef]

51. Gao, J.; Nuyttens, D.; Lootens, P.; He, Y.; Pieters, J.G. Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. *Biosyst. Eng.* **2018**, *170*, 39–50. [CrossRef]

52. Guo, G.; Wang, H.; Bell, D.; Bi, Y.; Greer, K. KNN Model-Based Approach in Classification. In *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003*; Meersman, R., Tari, Z., Schmidt, D.C., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2003; Volume 2888.

53. Fernandez, R. Predicting time series with a local support vector regression machine. In Proceedings of the ACAI 99, Crete, Greece, 5–16 July 1999.

54. Veropoulos, K.; Cristianini, N.; Campbell, C. The Application of Support Vector Machines to Medical Decision Support: A Case Study. In Proceedings of the ACAI 99, Crete, Greece, 5–16 July 1999.

55. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

56. Bhende, M.J.; Kalirajan, K.P. Technical efficiency of major food and cash crops in Karnataka (India). *Indian J. Agric. Econ.* **2007**, *62*, 177–192.

57. Dung, K.T.; Sumalde, Z.M.; Pede, V.O.; McKinley, J.D.; Garcia, Y.T.; Bello, A.L. Technical efficiency of resource-conserving technologies in rice-wheat systems: The case of Bihar and eastern Uttar Pradesh in India. *Agric. Econ. Res. Rev.* **2011**, *24*, 201–210.

58. Narala, A.; Zala, Y.C. Technical Efficiency of Rice Farms under Irrigated Conditions in Central Gujarat. *Agric. Econ. Res. Rev.* **2010**, *23*, 375–381.