AgrIntel: Spatio-temporal profiling of nationwide plant-protection problems using helpline data

Samarth Godara[1,2], Durga Toshniwal[2], Ram Swaroop Bana[3], Deepak Singh[1], Jatin Bedi[4,*], Rajender Parsad[1], Jai Prakash Singh Dabas[3], Abimanyu Jhajhria[5], Shruti Godara[6], Raju Kumar[1], and Sudeep Marwaha[1]

[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India
[2]Indian Institute of Technology Roorkee, Uttarakhand, India
[3]ICAR-Indian Agricultural Research Institute, New Delhi, India
[4]Thapar Institute of Engineering and Technology, Patiala, India
[5]ICAR-National Institute of Agricultural Economics and Policy Research, New Delhi, India
[6]Forest Research Institute, Uttarakhand, India
[*]jatin.bedi@thapar.edu

ABSTRACT

Sustainable development of the national food system must ensure the introduction of adequate food security interventions and policies. However, several high-end technological developments remain unexplored, which can be used to gain explicit information regarding agricultural problems. To obtain precise and updated insights regarding the sector's problems, incorporating modern technological advancements in food security policy-designing is the need of the hour. In this direction, the presented work proposes AgrIntel, a framework consisting of multiple AI-based pipelines to process nationwide farmers' helpline data and obtain spatiotemporal insights regarding food-production problems on an extensive scale. The call-logs dataset used in the study is obtained from the nationwide network of farmers' helpline centers, managed by the Ministry of Agriculture & Farmers' Welfare, Government of India. The article demonstrates the Spatio-temporal profile of one of India's highest food grain-affecting diseases, i.e., "blast in rice crop", to demonstrate the utility of the AgrIntel pipelines. First, the proposed framework extracts and clusters the precise geographical locations of farmers calling for help corresponding to the target agricultural problem. Next, the temporal modeling of the problem helps extract the critical dates corresponding to the crop disease/pest spread. Furthermore, by incorporating the historical agroclimatological data, the article introduces a new medium to extract the favorable weather conditions corresponding to the targeted disease/pest outbreak. In addition, the study explores the potential of Deep Learning models (based on Artificial Neural Network, Convolutional Neural Network, Gated Recurrent Unit and Long short-term memory unit) to efficiently predict the futuristic demand for assistance regarding target problems. The obtained results expose unrevealed insights regarding food production problems, significantly boosting the food security policy-designing procedure.

Keywords: agriculture policy, deep learning, disease profiling, food security, plant-protection, spatio-temporal analysis, sustainable food production.

1 Introduction

India, one of the top food-grain producing nations, produces ≈119 million tons of the total annual global rice production (≈505 million tons). Likewise, the country grows ≈107 million tons of the ≈761 million tons of total annual wheat production worldwide[1]. However, the productivity levels of the major food crops in the country are substantially lower than other leading producers of the globe[1,2]. The primary yield-limiting factors include diseases, pests, weeds, sub-optimal extension, and policy support[3]. The current yield loss of rice and wheat in India due to various insects/diseases/weeds is estimated to be 15-25%[4]. In this scenario, the current agri-related surveys conducted by organizations (governmental and non-governmental) are mainly oriented towards yield estimation, and no robust platform exists to profile crop protection-related problems on an extensive scale. To tackle this challenge, an authentic and vigorous information system is required on the extent of the problems from the pragmatic grassroots levels. Currently, the major sources of information on the nationwide crop protection problems in India include sales data of the agri-chemicals and small-scale farm trial experiments executed by the government (All India Coordinated Research Project (AICRP), Krishi Vigyan Kendra (KVK), State Agriculture Universities (SAUs), etc.)[5] and non-government networks. However, the limitations of the existing systems include:

1. Scalability: The organizations' field trials consist of fewer samples to draw a realistic estimate of the actual scenario.

Moreover, the research experiments cant be designed to examine the production-problems pattern corresponding to all the major crops belonging to the diverse climatic regions of the country. Furthermore, in the country's remote areas, due to the non-adoption of the agri-chemicals, drawing insights regarding the target problem through the sales data is impractical.

2. Frequency: In general, extensive farm sector surveys are conducted yearly, which fails to provide timely updates on the sector. Although the sales data can be collected daily, it provides generalized information on a group of production-related issues, not specific problems.

3. Survey cost: Since conducting manual surveys is a cost-intensive process, executing such programs on a large scale is a cumbersome task for the organization.

To overcome the existing methods' limitations, we propose AI-integrated pipelines to mine critical information from the nationwide farmers' helpline records. The proposed AgrIntel framework consists of the following four pipelines:

1. AgriSpatio: The pipeline is helpful for the spatial profiling of the problem. The output can be used for locating the disease-affected hot spots of the country, along with the increasing/decreasing trend for seeking help regarding the targeted problem over the past years.

2. AgriTempor: This pipeline is helpful for the temporal profiling of the targeted problem. The output effectively extracts important dates (initial, outbreak and peak period) regarding the disease/pest-spread pattern in the identified region.

3. AgriWeath: The pipeline is beneficial for the climatological profiling of the target problem. It can extract the favorable weather conditions corresponding to the focused region's crop disease/pest outbreak.

4. AgriPred: The pipeline helps to predict the demand for assistance by the producers regarding the selected problem from the focused region of the country. The module intakes weather parameters of the particular day and outputs the corresponding number of query calls that may appear regarding the selected problem.

Each of the proposed pipelines overcomes the above-mentioned limitations of the existing systems. Firstly, each pipeline helps extract information regarding the targeted agricultural problem on an extensive scale; this solves the problem of scalability. Secondly, the pipelines use the farmers' helpline data, making it convenient to extract the information daily, solving the frequency limitation. Lastly, since extracting the information only includes executing the proposed pipelines through computer programs, the computational cost is negligible compared to the cost of the existing methods for gaining such knowledge.

The proposed AgrIntel framework uses the Big-data collected from the Indian farmers' helpline center, Kisaan Call Center[6], consisting of the farmers' query-call records since March 2013, along with the historical agroclimatological of the same period. The pipelines use multiple artificial intelligence and machine learning-based techniques to extract critical insights from the dataset, which is practical for decision-making in production-related interventions on an extensive scale. The AgriSpatio pipeline uses natural language processing and K-means clustering algorithm along with multiple other modules to produce the desired spatial insights. The AgriTempor pipeline transforms the sequential data from the helplines into time series and later extracts insights after using the Trust Region Reflective curve fitting algorithm. AgriWeath pipeline differentiates the obtained time series to extract the inflection point and acquire helpful information. Furthermore, the AgriPred pipeline uses high-end Deep Learning-based forecasting models, including Long short-term memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), and Artificial Neural Network (ANN)-based models to forecast the farmers' demand for assistance. The outputs of the proposed pipelines help in focusing the extension-related interventions on the targeted regions where the aid is needed the most. Furthermore, the temporal insights help decide the time of year when assistance is required. Similarly, the outputs can provide significant insights for other interventions, including introducing agri-products (varieties, fertilizers, seeds, etc.) to the farmers and designing agri-product marketing strategies. Moreover, the AgriWeath and AgriPred pipeline can play a significant role in production-related research planning and developing early warning systems.

The presented study offers the following research contributions:

- Spatial clustering of the farmers' query-calls to identify the agricultural problem's hot spots.
- Curve fitting on the query-calls time series for extracting important dates corresponding to the target agricultural problem.
- Extraction of the weather parameters corresponding to the disease/pest outbreak on an extensive scale using the helpline data.
- Forecasting the query-calls count corresponding to the target agricultural problem (disease/pest/weed).

As a case study, the article elaborates on the output of proposed pipelines with the "blast disease in rice crop" as input to the AgrIntel framework. In addition, the supplementary information document includes the profiling output of the termite pest in the wheat crop using the proposed framework.

The remainder of the article is organized as follows: the "Literature Survey" section delivers a literature review on related topics, and the concepts used to develop the proposed pipelines are explained in the "Methodology" section. Furthermore, the "Experiments and Results" section elaborates on the experiments performed and the results obtained in the study, followed by the "Discussion" section, which elucidates the results and their applications in the related domain. Moreover, the article gives a summary of the presented study in the "Conclusion" section.

2 Literature Survey

India, the world's second most populous country, has agriculture as the primary source of livelihood for around 70% of the rural households population. With the advancements in Information and communication technologies (I.C.T.), Indian agriculture, energy, and material domains have observed significant changes. Even though the country has seen evolutionary changes in the agriculture sector over the years, agriculture's contribution to the economy has decreased[7]. There can be several causes for the decline, such as rural migration, lack of required resources or knowledge by the farmers, and many more.

The government of India has taken several initiatives to mitigate the causes through several flagship programs[8] in the agriculture sector. Examples of such programs are the National Mission for Sustainable Agriculture, P.M.F.B.Y. (Pradhan Mantri Fasal Bima Yojana), Gramin Bhandaran Yojna, and many more. Moreover, the National Sample Survey Office (N.S.S.O.) regularly conducts surveys for several tasks, such as designing and monitoring new agricultural activities, estimating agricultural production, etc. However, the majority of these surveys are paper-based. In addition, most designed surveys focus on capturing agricultural production-related information, with very few studies aiming at capturing farmers' problems. In recent years, researchers have proposed tools and IoT sensors-enabled techniques to collect real-time data from farmers. Zipper[9] proposed a social media (Twitter) based approach to monitoring the Spatio-temporal patterns. The process was implemented on the dataset of the United States, and the performance drift achieved was 10% compared to the manual surveys. Phillips[10] investigated the reliability of social media platforms for knowledge exchanges among farmers and advisors. However, these solutions are not feasible in developing countries like India for several reasons, including the high installation cost of new devices, the requirement for stable internet connection in farmlands, the literacy rate, etc.

On the other hand, in the present study, we aim to use the data generated from the "Kisan Call Centres" (K.C.C.) scheme[11,12] for similar purposes. To harness the potential of I.C.T. in the Agriculture sector, the Indian government launched the K.C.C. scheme on January 21, 2004, by the Ministry of Agriculture and Farmers Welfare, Government of India (K.K.M.S.G. of India, 2019). This project aims to respond to farmers' questions from all over India through phone calls in their regional terminology. The K.C.C. helpline assistance is available in 22 local languages from 6.00 AM to 10.00 PM on all seven days of the week. The call centers operate in 21 different sites in India, covering all the Indian States and Union Territories (U.T.), with the toll-free number 1800-180-1551. The farmers' questions are addressed by K.C.C. agents known as Farm Tele Advisors (F.T.A.s). They are graduates or above (i.e., Post-Graduates or Doctorates) in agriculture or associated sectors (including Aquaculture, Poultry, Bee-Keeping, Animal Husbandry, Fisheries, Horticulture, Bio-Technology, Sericulture, Agricultural Engineering, Agricultural Marketing, etc.). Moreover, each received query call is logged by the F.T.A.s, including the details regarding the farmer's location, time and date of the question, the complete question asked, the answer delivered, crop, category, query type, etc. These logs are maintained and made publically available by the "Kisan Knowledge Management System" (K.K.M.S.) and at an open data platform (data.gov.in) under the Ministry of Agriculture, Government of India (data.gov.in reference). The publicly accessible data can be downloaded in .json format files, where a separate file is maintained every month, corresponding to each state/U.T.

In the past few years, researchers have been exploring advanced machine learning and deep learning-based techniques in image processing[13,14], remote sensing[15], and disease diagnosis[16–22] to extract useful insights from the dataset. Furthermore, in recent years, many researchers worldwide have explored the potential of helpline services data to analyze public behavior in different application domains[23–27] such as healthcare, suicide prevention, emergency services, and agriculture. In the agricultural aspect, Bashir et al.[28] Employed a sampling-based technique to examine the farmers' willingness to access the information from the farmers' helpline services in A.B.U. Zaria. Arfan et al.[29] Performed a comparative analysis of Punjab's agricultural helpline data and other agricultural information sources. The analysis of helpline services data suggests that the information extracted is very insightful and can be reliably utilized for critical future planning activities. Many researchers have recently implemented data mining techniques on K.C.C. data to extract meaningful insights. Viswanath et al.[30] Proposed a MapReduce approach to analyze the K.C.C. data. The approach implemented Natural Language Processing (N.L.P.) techniques to provide valuable theoretical and practical insights regarding critical parameters such as peak timestamp interval of query calls, the most queried crops, and query grouping. Mohapatra and Upadhyay[31] proposed a topic modeling approach integrating LDA (Latent Dirichlet Allocation), TF-IDF (Term Frequency & Inverse Document Frequency), and Latent Semantic Indexing

(LSI) to extract and analyze topics (area-wise significant problems faced by farmers, crop-wise problems faced by farmers and time-space analysis) from the Kisan Call Center data. Gandhi and Johnson[32] analyzed the K.C.C. national data to identify and present the area for improvements in the K.C.C. scheme. Sharma et al.[33] Performed exploratory analysis of the K.C.C. data to identify hidden patterns and valuable insights into the data. The authors suggested the need for a better visualization tool for the representation of data insights. A recent study[34] in the K.C.C. data analysis domain reviewed different approaches using TF-IDF, Word encodings, and clustering for query text and query type classification. Momaya et al.[35] Proposed a chatbot system (Krushi) to answer farmer agricultural queries. The scholars used Machine learning, and Artificial Intelligence approaches to answer questions about problems such as weather conditions, plant protection, fertilizers, soil testing, and market price.

Although multiple analysis works exist on the K.C.C. dataset, until now, researchers are unable to extract data up to the problem level, i.e., Spatio-temporal insights regarding the particular disease, pest, or weed. In 2020, Godara and Toshniwal extracted sequential patterns from the dataset to discover issues that arise simultaneously[36]. Whereas, in the presented study, we can cluster the nationwide locations of the query calls to point out the problems' hot spots. Moreover, in 2022, Godara and Toshniwal developed multiple deep learning-based query count forecasting models designed to use previous time series data points[37]. In the presented study, forecasting models are developed that take the historical weather information along with the earlier time series to forecast the target problems precisely. Furthermore, several existing studies use multiple agricultural trials to find the favorable weather conditions corresponding to the outbreak of a particular disease/pest[38]. The presented research extracts similar insights using the farmers' helpline centers' data. The algorithms proposed in the study use multiple mathematical models including curve fitting and local-maxima calculation to obtain the desired insights.

3 Methodology

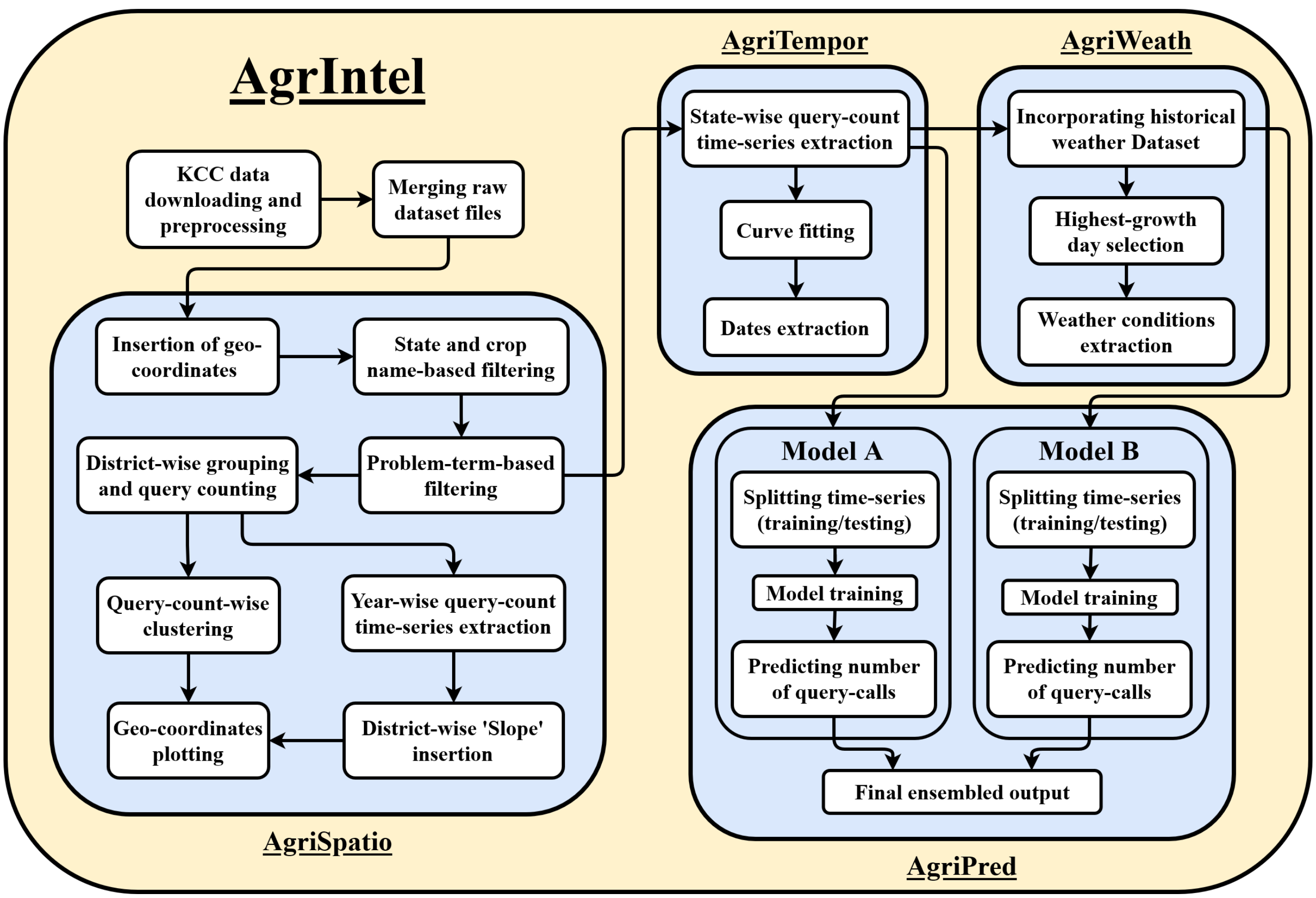


Figure 1. Block diagram of the AgrIntel framework

The proposed pipelines make use of multiple algorithms to achieve the defined objectives. To extract spatial insights, the framework first inserts the geo-locations corresponding to each query call and later performs clustering and other operations to highlight the required information. In contrast, the temporal insights are extracted by first converting the transactional database into a time series and fitting a bell-shaped curve to obtain the critical dates regarding the spread of the target agricultural problem. In addition, the favorable weather conditions corresponding to the outbreak of the disease/pest are obtained by first incorporating the historical weather information along with the query-calls time series. Later, the inflection point or the outbreak day is determined by differentiating the time series to obtain the required weather-related information. Moreover, the forecasting models for predicting the farmers' demand for assistance are developed using two sub-models, i.e., a) model that use the previous query-calls time series to generate the forecast and b) model that use the previous weather information to generate the required forecast. Later, the outputs from the two models are ensembled to have a final precise prediction. Details regarding each phase of the proposed pipelines are given in the remainder of the section.

3.1 Data Collection and preprocessing

Since the KCC helpline centers maintain a separate file for the district-wise call-log records corresponding to each month, in this work, the records are automatically crawled for files corresponding to March 2013 till November 2021 from the KCC servers through a web-server crawler python script. The downloaded files are cleaned for any erroneous characters caused by the manual entry, storage, or communication error. This step removes all characters other than the alphanumerics from the record files. All the downloaded record files are combined to form a single .csv file for easy execution of data operations that consists of over 26 million call-log records. There are nine attributes in the raw dataset (supplementary information, section 2), in which only six attributes (CreatedOn, BlockName, DistrictName, StateName, KCCQuestion, and Crop) are relevant to our study. The rest of the attributes remain unused in the present work. Figure 1 visualizes the internal working of the complete AgrIntel framework, including the different pipelines present in it.

3.2 AgriSpatio

First, the geo-coordinates corresponding to each address (block name, district name, and state name) present in the dataset are extracted and incorporated using two python APIs, i.e., opencage and geopy. Call-log records corresponding to the input state names and crop names are selected by matching the values present in the respective attributes. Approximate matching is performed between the input problem-terms and each word present in the 'KCCQuestion' attribute of the call-log records using the Levinstien Distance function[39] (equation 1). Call-log records with $l < 2$ are selected for further processing ($l < 3$ is used for longer problem-terms).

Choosing a smaller threshold value (< 1) results in a reduced dataset and a threshold value of more than 3 leads to an enlarged dataset due to the acceptance of irrelevant questions including words dissimilar to the target problem.

$$l(x,y) = \begin{cases} |x| & \text{if } |x| = 0, \\ |y| & \text{if } |y| = 0, \\ l(tail(x), tail(y)) & \text{if } if|x| = |y|, \\ 1 + min \begin{cases} l(tail(x), y) \\ l(x, tail(y)) \\ l(tail(x), tail(y)) \end{cases} & \text{otherwise} \end{cases} \tag{1}$$

Here, x, y are the input character strings between whom the Levinstien distance is to be calculated, $|x|$ represents the length of string x, and $tail(x)$ is the string x without the first character.

In the next step, the call-log records are grouped based on the districts, and the number of records is counted for each group. Based on the number of call-log records present, district-groups are clustered into five clusters using K-means clustering algorithm[40]. The clustering algorithm aims to optimize the objective function given a set of data points presented in equation 2.

$$J = \sum_{j=1}^{K} \sum_{i \in C_j} ||o_i - c_i||^2 \tag{2}$$

Here, J is the criterion function, o_i is the i^{th} observation, c_j is the j^{th} cluster center, C_j is the object set of the j^{th} cluster and K represents the number of clusters. Any norm representing the distance between the data object and the cluster's center is denoted by $||*||$.

Subsequently, the year-wise cardinality of each group is calculated to track the increasing or decreasing rate of calls over the past years. Using Linear Regression[41] (equation 3), the slope corresponding to the district-wise query index (query count divided by the total queries in the corresponding year) time-series is extracted and incorporated into the dataset.

$$Y = \alpha + \beta X \tag{3}$$

Here, Y is the explanatory variable (query index), and X is the dependent variable, i.e., 'year' in our case. The slope of the line is β, and α is the intercept. Moreover, the pipeline plots the geographical maps representing the district-wise cluster tags and time-series slopes separately (figure 2).

3.3 AgriTempor

The extracted call-log records corresponding to the input problem terms are first transformed into day-wise query-count time series. Each data point of the output time series represents the average number of the call-log records corresponding to the particular day of the year. Using the Trust Region Reflective algorithm[42], a Gaussian curve (equation 4) is fitted on the extracted day-wise state-wise query-count time-series.

$$g(t) = h \times e^{-\frac{(t-\mu)^2}{2\sigma^2}} \tag{4}$$

Here, h represents the curve's height, e is the Euler's number, μ is the position of the center of the peak, and σ represents the standard deviation. The peak date (μ) is directly noted from the optimized parameters, whereas the date corresponding to the beginning of the curve (starting date) is calculated as $(\mu - 2\sigma)$.

3.4 AgriWeath

The historical agroclimatological data is collected, corresponding to the approximate center points of each Indian state. The daily time series data is collected from NASA's Prediction of Worldwide Energy Resources (POWER) project[43]. Moreover, the data is incorporated into the daily query-count time series extracted in previous steps. The extracted query-count time series is differentiated, and a window of ten days from the maxima point m is selected. Where, m is the global maxima point of the first differential of $g(z)$, i.e., if $(\forall z \in \mathbb{R})$, $g'(m) \geq g'(z)$. The day with the highest humidity is selected from the window as the outbreak day for the selected disease/pest. Weather conditions corresponding to the outbreak day of each year are averaged and reverted as the most favorable weather conditions for the input disease/pest outbreak.

3.5 AgriPred

To achieve better accuracy, the AgriPred pipeline considers two models for making predictions. The final forecast is produced by ensembling the output of these two forecasting models, i.e., Model A and Model B, as given in equation 5.

$$P = (a \times P_A) + (b \times P_B) \tag{5}$$

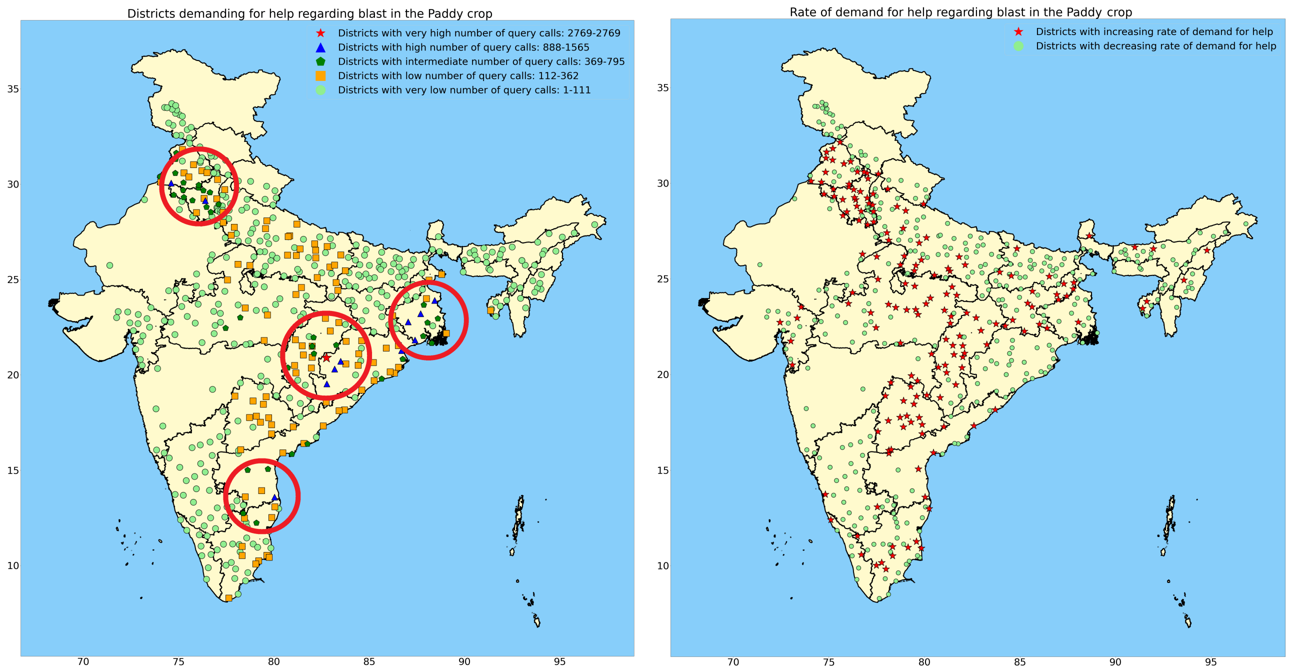
Where, P represents the final output of AgriPred, P_A and P_B are the predictions of Model A and Model B respectively, and a, b represent the weightage given to the output of each model, where $a, b \leq 1$ and $a + b = 1$. Model A uses the previous query-count time series to forecast the futuristic time series. In contrast, Model B uses the agroclimatological-data incorporated query-count time series to produce the forecast. The collected data is firstly divided into two parts for training and testing prediction models. The dataset corresponding to 2014-2019 is used as training data, whereas the time series corresponding to 2020 is used as testing data. Hyperparameters of four prediction models (ANN, CNN, GRU, and LSTM) are tuned using the grid-search approach[44]. The models' prediction accuracy (3-days and 7-days forecast) is assessed using RMSE and MAE in three states (Uttar Pradesh, Punjab, and West Bengal) corresponding to the blast disease in rice crop. The models are assessed based on their forecasting performances of the query count on the testing data.

4 Experiments and Results

4.1 AgriSpatio

Figure 2 represents the output of the AgriSpatio pipeline corresponding to the input combination of the blast disease in the rice crop. From the figure 2a, it is noted that from March 2013 to November 2021, Indian farmers requested help regarding the disease from all the rice-growing regions of India. Nonetheless, the query-count-based clusters show that there exist four major hot spots (circled in red) for the blast disease in India, i.e.,

1. The border area of Chattisgarh and Odisha,
2. South and Southwestern Haryana, and adjoining Punjab region,
3. Coastal Andhra Pradesh, and adjoining areas of Tamil Nadu, and
4. The West Bengal state.



(a) Clusters based on number of query-calls

(b) Increasing/Decreasing rate of query-calls

Figure 2. AgriSpatio Output: Spatial insights regarding the blast disease in Paddy crop; hot spots circled in red.

A reasonable explanation behind the findings is that the hot spots of blast diseases are the regions where the farmers have grown the high-yielding rice varieties for the past decades[45]. Furthermore, in the existing studies, it is noted that the breakdown of blast disease resistance in many high-yielding rice varieties is the major cause of blast spread in these regions[46].

From the figure 2b, it is observed that farmers from the states of Haryana, Northern Punjab, Western Madhya Pradesh, Telangana, and Chattisgarh have been seeking help for the blast disease increasingly in the past few years.

A potential reason behind the discovery is that the high-yielding (blast susceptible) rice varieties in these regions dominate the agricultural market, and the new blast-resistant varieties have not been established in the market of these regions yet[47].

Furthermore, farmers from the states of Uttar Pradesh, West Bengal, Bihar, and Odisha have decreased interest in the control solutions regarding the blast. There are three reasonable explanations; firstly, since the blast is a disease that appears majorly in upland-rice regions[48], the Indian government has been promoting other cropping systems in such areas for years[49]. Additionally, the adoption of resource conservation technologies and conservation agriculture has assisted in increasing the interest of the farmers in other crops[49,50]. Secondly, the government has been focusing on the awareness of seed treatment to prevent such

Table 1. AgriTempor Output: Important dates corresponding to the blast disease of the various Indian states

S.No.	Season	State	Starting Date	Outbreak date	Peak Date	σ
1	Rabi	Telangana	28-Dec	29-Jan	26-Feb	29.01
2	Rabi	Odisha	10-Jan	20-Feb	15-Mar	31.89
3	Rabi	West Bengal	22-Jan	17-Feb	17-Mar	26.34
4	Kharif	Bihar	01-Jul	28-Aug	15-Sept	37.34
5	Kharif	Uttar Pradesh	13-Jul	17-Aug	13-Sept	30.33
6	Kharif	Chhattisgarh	18-Jul	18-Aug	16-Sept	29.09
7	Kharif	West Bengal	19-Jul	24-Aug	25-Sept	33.12
8	Kharif	Odisha	24-Jul	24-Aug	22-Sept	29.76
9	Kharif	Punjab	02-Aug	26-Aug	21-Sept	24.0
10	Kharif	Haryana	03-Aug	27-Aug	22-Sept	24.0
11	Kharif	Telangana	04-Aug	31-Aug	01-Oct	28.16

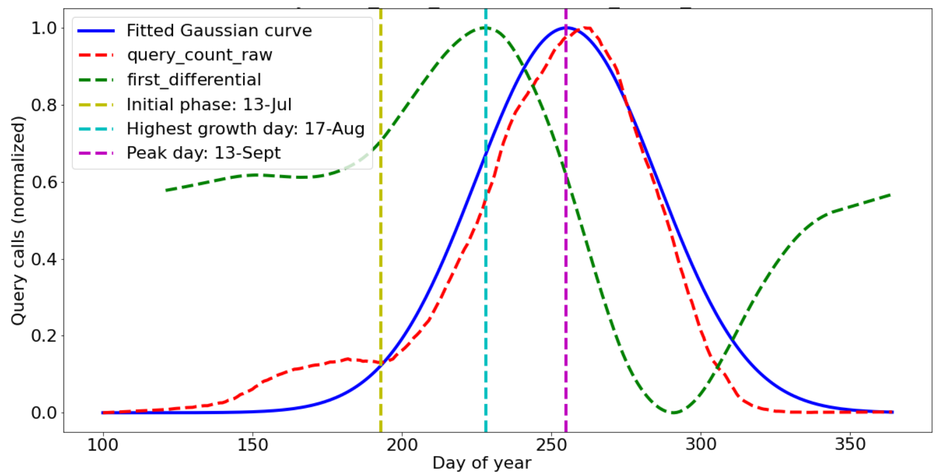


Figure 3. AgriTempor Output: Temporal insights regarding the blast disease in Paddy crop from the Uttar Pradesh state

diseases in these regions for the past few years[51]. Thirdly, the Indian government has been introducing blast-resistant varieties in these regions since 2016[52].

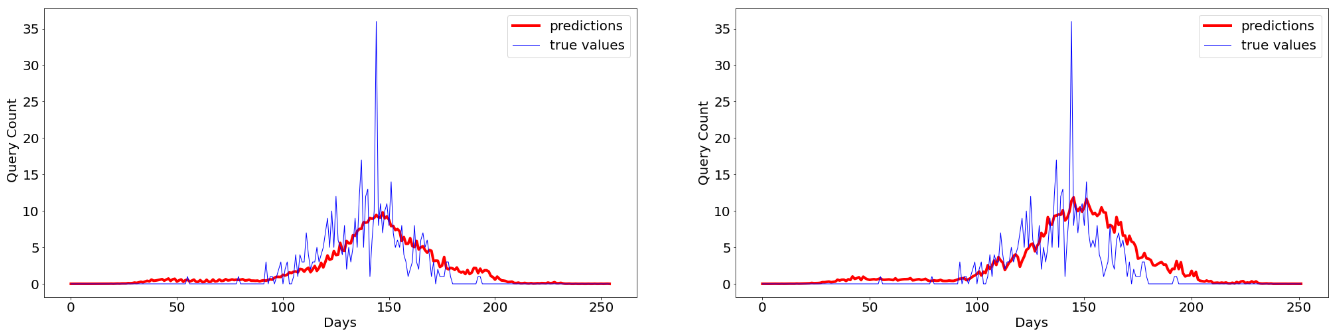
Interestingly, it is noticed that the farmers from the Western districts of Madhya Pradesh have been asking for help regarding the blast disease increasingly in the past years. Even though the high-yielding rice varieties are not recommended to farmers

Table 2. AgriWeath output: Weather conditions corresponding to the highest-growth rate of blast disease in states of Uttar Pradesh, Punjab, Madhya Pradesh, Haryana

S.No.	Weather Parameter	Mean	Standard Deviation
1.	Temperature at 2 Meters ($^{\circ}C$)	27.63	1.59
2.	Dew/Frost Point at 2 Meters ($^{\circ}C$)	24.72	1.08
3.	Wet Bulb Temperature at 2 Meters ($^{\circ}C$)	26.17	1.08
4.	Earth Skin Temperature ($^{\circ}C$)	27.79	1.88
5.	Temperature at 2 Meters Range ($^{\circ}C$)	5.86	1.74
6.	Temperature at 2 Meters Maximum ($^{\circ}C$)	30.8	2.33
7.	Temperature at 2 Meters Minimum ($^{\circ}C$)	24.94	1.22
8.	All Sky Surface Longwave Downward Irradiance (W/m^2)	438.38	8.77
9.	Specific Humidity at 2 Meters (g/kg)	20.25	1.24
10.	Relative Humidity at 2 Meters (%)	85.2	7.38
11.	Precipitation Corrected (mm/day)	27.00	23.83
12.	Surface Pressure (kPa)	97.1	1.27
13.	Wind Speed at 2 Meters Maximum (m/s)	3.51	1.11
14.	Wind Speed at 2 Meters Minimum (m/s)	1.16	0.67
15.	Wind Speed at 2 Meters Range (m/s)	2.35	0.74
16.	Wind Direction at 2 Meters (Degrees)	178.35	68.97
17.	Wind Speed at 10 Meters (m/s)	3.30	1.17
18.	Wind Speed at 10 Meters Maximum (m/s)	4.93	1.6
19.	Wind Speed at 10 Meters Minimum (m/s)	1.93	0.96
20.	Wind Speed at 10 Meters Range (m/s)	3.00	1.09
21.	Wind Direction at 10 Meters (Degrees)	178.35	68.94
22.	Wind Speed at 2 Meters (m/s)	2.24	0.83
23.	Surface Soil Wetness (1)	0.76	0.16
24.	Root Zone Soil Wetness (1)	0.72	0.17
25.	Profile Soil Moisture (1)	0.7	0.17

Table 3. Hyperparameters of the developed forecasting models (X_N, X_C, X_G and X_L represents X number of neurons, filter size of 1D convolutional layer, number of GRU cells and number of LSTM units present in the corresponding layer of the model, respectively).

Model	Input Layer	Hidden Layer(s)	Output Layer
ANN	365	365_N, 250_N, 50_N	$3_N/7_N$
CNN	365	64_C, 32_C, 16_C, 10_N	$3_N/7_N$
GRU	365	365_G	$3_N/7_N$
LSTM	365	365_L	$3_N/7_N$



(a) 3-Days Forecasting results of CNN-based model

(b) 7-Days Forecasting results of CNN-based model

Figure 4. Forecasting results of CNN-based model for the blast disease-related queries from the Uttar Pradesh state

from this region, farmers still grow them due to the dominance of these varieties in the market.

Figure 2 shows that districts from the western side of the 100 cm isohyets of India generally fall in the lowest query-call cluster. Because of a comparatively dry region, rice is an unpopular crop in these areas. Moreover, if rice is grown, it is irrigated, with minimum biotic and abiotic stress, leading to minor blast problems[53].

4.2 AgriTempor

Since rice is grown only in the Kharif season in the northern states of India (including Uttar Pradesh, Bihar, Chattisgarh, Punjab, and Haryana), temporal insights from these states are extracted for only one season. On the other hand, in the presented study, temporal insights are extracted for both the Rabi and Kharif seasons for the states of Telangana, Odisha, and West Bengal (table 1). The pipeline's output corresponding to the blast disease in the northern states shows that the starting date of blast disease spread is 1^{st} July - 4^{th} August. Moreover, the peak dates of the disease spread in Kharif rice is ten days, i.e., 15^{th} September - 25^{th} September.

From the output of the AgriTempor pipeline corresponding to the major rice-producing states of India (table 1), it seems that the initial phase of the curve (Starting dates, example: figure 3) includes the precautionary queries from the farmers, as these are asked in the early stage of the season. Moreover, it is noted that farmers start asking questions in the initial phase of the season in states including Bihar, Uttar Pradesh, Chhattisgarh, West Bengal, Odisha, and Telangana. In Punjab and Haryana, farmers start asking questions later in the crop season. Furthermore, the later parts of the curve (highest growth-rate dates and the peak dates, figure 3) seem to represent the numbers when the disease is visible in the fields. Consequently, the variation in the starting dates over the target states is more (≈30 days) than the variation in the highest growth-rate dates and peak dates (≈15 days).

This phenomenon has been observed by the fitted Gaussian curve's curve-width parameter (σ). From the pipeline output, it has been noted that the farmers from the states where the curve's width is comparatively larger ($\sigma > 25$) start asking questions in the early stage of the crop.

A potential reason behind a longer-width curve (a more extended period of the disease) is that the farmers grow diverse varieties of rice in these states. For example, farmers from Western Uttar Pradesh majorly grow Basmati rice, whereas Eastern Uttar Pradesh farmers are more inclined towards the conventional, non-Basmati rice varieties[54]. Furthermore, in the states with a comparatively shorter curve-width ($\sigma < 25$, Punjab and Haryana), farmers start enquiring at the later stage when the disease is visible in the fields. Since farmers from these states ask for help in the later stage of the crop, the severity of the disease is noted more in these areas from the AgriSpatio output, as farmers do not seem to emphasize preventive measures.

Table 4. RMSE and MAE (query-call count) comparison of the forecasting models

	3-Days Forecasting							
	RMSE				MAE			
State \Model	**ANN**	**CNN**	**GRU**	**LSTM**	**ANN**	**CNN**	**GRU**	**LSTM**
Punjab	2.958	1.903	2.65	**1.79**	1.999	**0.97**	1.624	1.146
Uttar Pradesh	3.044	**2.42**	2.685	2.516	2.143	**1.127**	1.339	1.208
West Bengal	2.696	1.933	**1.453**	2.534	2.304	1.594	**0.915**	1.714
	7-Days Forecasting							
	RMSE				**MAE**			
State \Model	**ANN**	**CNN**	**GRU**	**LSTM**	**ANN**	**CNN**	**GRU**	**LSTM**
Punjab	2.506	**1.567**	2.159	1.893	1.286	**0.85**	1.451	1.042
Uttar Pradesh	2.573	2.623	**2.493**	2.91	1.24	1.36	**1.2**	1.574
West Bengal	1.83	1.906	1.662	**1.551**	1.051	1.15	1.175	**0.977**

Table 1 gives the output of the AgriTempor pipeline corresponding to the blast disease in rice crop for various states. From the table, it has been noted that in Telangana, the peak of the blast disease spread occurs on 1^{st} October, approximately one week later than in the other states. It is because the Telangana state falls in the tropical climatic region where generally, rice is grown 15-20 days later than in the northern and western Indian states. The results show that the states with early peaks mainly fall in the continental regions (landlocked states, including Punjab, Harayana, Uttar Pradesh, Bihar, and Chhattisgarh). Besides, the states with the later peaks are the coastal states (West Bengal, Telangana, and Odisha).

From the temporal insights corresponding to the blast disease in the Rabi season, it is observed that in Telangana, farmers start seeking help regarding the disease ≈15 days earlier than in the other Rabi-rice growing states.

The reason is that the region's relatively higher humidity and temperature, the climatic conditions in Telangana are more favorable to the blast disease in the early stage of the season[55]. Whereas, in the eastern states of India (West Bengal and Odisha), cold climatic conditions in January prevent the disease growth. As the conditions become warm and humid at the end of February, the disease starts appearing in the fields.

4.3 AgriWeath

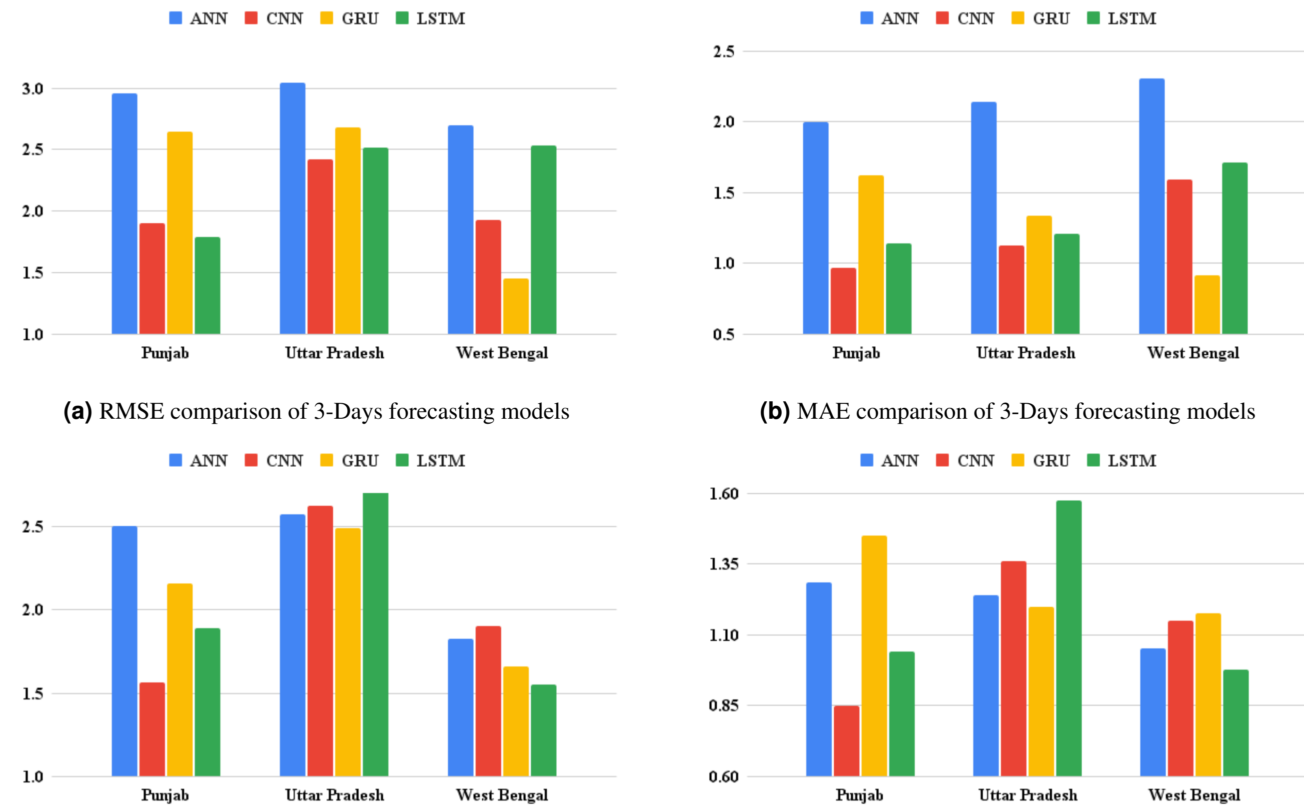
Table 2 (the output of the AgriWeath pipeline) gives the combination of the weather parameters that provides the most favorable environment for the outbreak of blast disease in the states of Uttar Pradesh, Punjab, Madhya Pradesh, and Haryana. Due to different climatic regions, the weather parameters corresponding to the states of West Bengal, Odisha, Andhra Pradesh, Bihar, and Chhattisgarh, are extracted separately (Supplementary information).

The weather conditions extracted by the pipeline for the target states include the temperature of $27.6°C$ ($\pm1.59°C$, at 2m), along with the slight fluctuation of the diurnal temperature (range of $5°C$). Furthermore, it is observed that the 'All Sky Surface Longwave Downward Irradiance' parameter is on the lower side ($438.38W/m^2 \pm 8.77W/m^2$), which indicates the cloudy weather. Moreover, the relative humidity indicated by the pipeline is very high, i.e., $85.2\% \pm 7.3\%$. In contrast, the Surface pressure is noted to be on the lower side, i.e., $97.1kPa$. Interestingly, the wind speed observed corresponding to the disease outbreak is very low ($2.24m/s \pm 0.83m/s$, at 2m), and the soil moisture is on the higher side, i.e., $70\% \pm 0.17\%$.

4.4 AgriPred

Using the historical agroclimatological time-series incorporated with the query-count time-series corresponding to the six years (2014-2019), four machine learning and deep learning models, including Artificial Neural Network[56] (ANN), Convolutional Neural Network[57] (CNN), Gated Recurrent Unit[58] (GRU), and Long short-term memory unit[59] (LSTM) are trained to make predictions of the number of query calls. The models predicted the query-count time series for 2020 by utilizing the weather details and the previous year's query counts. To compare the models' performances, each model is trained to produce two types of forecast, i.e., 3-days and 7-days forecast. Moreover, the models are trained on the query-count time series corresponding to three states (Uttar Pradesh, Punjab, and West Bengal) regarding the blast disease in rice.

The optimized ANN-based models developed in the presented study included five layers, with each layer consisting of 365 (input layer), 365 (hidden layer 1), 250 (hidden layer 2), 50 (hidden layer 3), and 3 or 7 (output layer) neurons, respectively (Table 3). Furthermore, the CNN-based models had three 1D convolutional layers consisting of filters of size 64 (hidden layer 1), 32 (hidden layer 2), and 16 (hidden layer 3), followed by two densely connected neuron layers with 10 (hidden layer 4) and 3 or 7 (output layer) neurons in each layer. Whereas both the LSTM and GRU-based models consisted of two layers, i.e., the first layer including 365 (LSTM/GRU) cells that are densely connected with an output layer of 3/7 neurons. Figure 4 plots the



(a) RMSE comparison of 3-Days forecasting models

(b) MAE comparison of 3-Days forecasting models

(c) RMSE comparison of 7-Days forecasting models

(d) MAE comparison of 7-Days forecasting models

Figure 5. Performance comparison of query-count forecasting models

prediction results (3 and 7-days) along with the actual number of query calls related to the blast disease in the Punjab state (for the year 2020) using the CNN-based model.

From the figure, it is noted that the models can generate the 3-day forecast with less fluctuation as compared to the 7-day forecast. Figure 5 and table 4 gives the forecasting error rate of the models in terms of RMSE and MAE. The experimental results show that the models based on CNN, GRU, and LSTM show comparable forecasting performance. In contrast, ANN-based forecasting is found to have the highest error rates in most of the experiments. Overall, it is observed that these models can be used to provide a reliable forecast of the demand for help related to the particular agricultural problem from the targeted region.

5 Discussion

The profiling of the food production problems done by the proposed pipelines is helpful in multiple ways. From the geographical perspective, the extracted hotspot regions of the blast disease should be given more attention while designing the related policies. Furthermore, preventive measures regarding the problems should be introduced in the hotspot areas, whereas advisory measures can be introduced in other regions. Moreover, for the optimal return of investment, the package of practices can be altered according to the disease severity in the region.

The extracted spatial insights are also valuable for future research planning. Scholars must orient their works towards the regions where demand for assistance regarding the target problem is increasing. Due to the cultural differences and differences in the agricultural-markets scenario in the hotspot regions, the government should intervene, taking into account all the parameters. Moreover, the disease-resistant varieties should be introduced to the regions with a higher degree of problem.

The critical dates extracted by the AgriTempor Pipeline give essential insights for deciding the intervention time for the disease control solutions. The Indian government should execute farmers' training and extension activities between the respective states' extracted starting and peak dates. The information is also helpful in agronomic research planning from a temporal perspective. The sowing window can be shifted accordingly so that the disease-prone phase of the crop can be moved out of the disease window.

The output of AgriWeath is helpful to cope with climate change, as the weather fluctuations make it trickier to predict the disease-growth pattern. Here, the extracted weather details corresponding to the target disease/pest give precise environmental information, which leads to the disease/pest outbreak. Such information is also valuable in varietal development, as these conditions are beneficial in screening disease-resistant varieties. Furthermore, to control the disease in protected cultivation systems, producers should be advised to deviate from the extracted weather conditions as much as possible.

Moreover, using the predictions from the AgriPred pipeline, early warning systems can be designed to alert the farmers of the target state. In addition, recommendations can be made on preventive measures in high-value crops using the forecasted demand for help. Furthermore, the output can also play a key role in developing agri-market strategies and planning extension activities in the target regions.

Compared to the existing methods for obtaining similar insights regarding the target plant protection-related problems, the proposed pipelines require minimum human resources and time. Single personnel can operate the whole framework, and the execution time required for profiling an individual target problem is generally 2-3 hours. In contrast, the existing methods (including manual surveys, field experiments, etc.) require months for experimentation, data collection, and data analysis, with the expense of multiple workforces for individual objectives. Moreover, the presented methodology considers the farmers' helpline call-logs dataset, which is updated every day by the Indian government bodies. Therefore, the proposed AgrIntel framework can deliver precise information on an extensive scale for real-time monitoring purposes, an unachievable feature of the existing methods.

6 Conclusion

The presented study incorporates modern technological developments in food security policy design. In this direction, the authors presented AgrIntel, an AI-based framework consisting of multiple pipelines to process nationwide farmers' helpline data and obtain spatiotemporal insights regarding food-production problems on the national scale. The article elaborates on one of India's highest food grain-affecting diseases, i.e., "blast in rice crop". First, four hot spots of the target diseases are identified using the farmers' helpline dataset. Next, the AgriTempor pipeline extracts the critical dates corresponding to the target crop disease/pest spread. Moreover, by incorporating the historical agroclimatological data, the favorable weather conditions corresponding to the targeted disease/pest outbreak are extracted using AgriWeath. In addition, four Deep Learning-based forecasting models are developed and compared to predict the demand for help regarding target problems from the selected regions of India.

The results reveal that the Deep Learning models, including CNN, LSTM, and GRU-based forecasting models, deliver better forecast precision than the ANN-based models, depending on the training time series dataset. In the future scope, the authors intend to develop Deep Learning integrated Natural Language Processing models to extract novel insights to help policymakers achieve the goal of sustainable production.

7 Author contributions statement

S.G.[1], D.T., R.S.B., D.S., J.B. and R.K. conceived the experiment(s), S.G.[1], and J.B. collected the data set, S.G.[1], R.S.B., D.S. and J.B. conducted the experiment(s), D.T., R.P., J.P.S.D., A.J., S.G.[5], and S.M. supervised the experiment(s), analysed and interpreted the results. S.G.[1], R.S.B., and J.B. wrote the manuscript, all authors reviewed the manuscript.

References

1. FAO. Crops and livestock products. https://www.fao.org/faostat/en (2021). Accessed: June 2022.

2. Yuan, S. *et al.* Southeast asia must narrow down the yield gap to continue to be a major rice bowl. *Nat. Food* **3**, 217–226 (2022).

3. Bana, R. S. *et al.* Weed control and rice yield stability studies across diverse tillage and crop establishment systems under on-farm environments. *Soil Tillage Res.* **204**, 104729 (2020).

4. Aayog, N. Raising agricultural productivity and making farming remunerative for farmers (2015).

5. ICAR. Indian council of agricultural research. https://icar.org.in (2022). Accessed: June 2022.

6. GOI. Kisaan call center. https://dackkms.gov.in (2022). Accessed: June 2022.

7. Nations, F. U. Fao.org.

8. Ministry, G. Ministry major schemes: Department of agriculture & farmers welfare: Mo a & fw: Goi.

9. Zipper, S. C. Agricultural research using social media data. *Agron. J.* **110**, 349–358 (2018).

10. Phillips, T., McEntee, M. & Klerkx, L. An investigation into the use of social media for knowledge exchange by farmers and advisors. *Rural. Ext. Innov. Syst. J.* **17**, 1–13 (2021).

11. Kisan Call Center, K. o. I.

12. Centre, N. I. Open government data platform india. https://data.gov.in/ (2021).

13. Wu, Y. *et al.* Semi-supervised hyperspectral image classification via spatial-regulated self-training. *Remote. Sens.* **12**, 159 (2020).

14. Wu, Y. *et al.* Commonality autoencoder: Learning common features for change detection from heterogeneous images. *IEEE Transactions on Neural Networks Learn. Syst.* (2021).

15. Wu, Y. *et al.* A two-step method for remote sensing images registration based on local and global constraints. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **14**, 5194–5206 (2021).

16. Altaf, I., Butt, M. A., Zaman, M. *et al.* Hard voting meta classifier for disease diagnosis using mean decrease in impurity for tree models. *Rev. Comput. Eng. Res.* **9**, 71–82 (2022).

17. Butuner, R. & Yuksel, H. Diagnosis and severity of depression disease in individuals with artificial neural networks method. *Int. J. Intell. Syst. Appl. Eng.* **9**, 55–63 (2021).

18. Sehirli, E. & Turan, M. K. A novel method for segmentation of qrs complex on ecg signals and classify cardiovascular diseases via a hybrid model based on machine learning. *Int. J. Intell. Syst. Appl. Eng.* **9**, 12–21 (2021).

19. Vedanarayanan, V., Narmadha, R., Amirthalakshmi, T., Velan, B. *et al.* Classification of covid-19 cases using deep neural network based on chest image data through wsn. *Int. J. Intell. Syst. Appl. Eng.* **10**, 28–34 (2022).

20. Patil, V. N. & Ingle, D. R. A novel approach for abo blood group prediction using fingerprint through optimized convolutional neural network. *Int. J. Intell. Syst. Appl. Eng.* **10**, 60–68 (2022).

21. Peker, N. & Kubat, C. A hybrid modified deep learning data imputation method for numeric datasets. *Int. J. Intell. Syst. Appl. Eng.* **9**, 6–11 (2021).

22. Paithane, P. M. & Kakarwal, S. Automatic pancreas segmentation using a novel modified semantic deep learning bottom-up approach. *Int. J. Intell. Syst. Appl. Eng.* **10**, 98–104 (2022).

23. Spike, C. & Unit, E. A. P. *The EAPU Helpline: Results of an investigation of five years of call data* (Elder Abuse Prevention Unit, UnitingCare Community, 2015).

24. O'Neill, S. *et al.* Data analytics of call log data to identify caller behaviour patterns from a mental health and well-being helpline. *Heal. informatics journal* **25**, 1722–1738 (2019).

25. Nakubulwa, M. A. *et al.* To what extent do callers follow the advice given by a non-emergency medical helpline (nhs 111): A retrospective cohort study. *Plos one* **17**, e0267052 (2022).

26. Mesfin, Y. M., Cheng, A. C., Enticott, J., Lawrie, J. & Buttery, J. P. Use of telephone helpline data for syndromic surveillance of adverse events following immunization in australia: A retrospective study, 2009 to 2017. *Vaccine* **38**, 5525–5531 (2020).

27. Roy, A. *et al.* Mental health implications of covid-19 pandemic and its response in india. *Int. J. Soc. Psychiatry* **67**, 587–600 (2021).

28. Bashir, M. B. *et al.* Rural farmers' readiness to access information from the national farmers helpline centre, national agriculture extension and research liaison services abu zaria, kaduna state, nigeria. *J. Agric. Ext.* **26**, 120–128 (2021).

29. Arfan, M., Ali, S., Khan, F. U. & Khan, G. A. Comparative analysis of punjab agriculture helpline and other agricultural information sources for the farmers in district lahore. *J. Agric. Res* **51**, 473–478 (2013).

30. Viswanath, V. K., Madhuri, C. G. V., Raviteja, C., Saravanan, S. & Venugopalan, M. Hadoop and natural language processing based analysis on kisan call center (kcc) data. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1142–1151 (IEEE, 2018).

31. Mohapatra, S. K. & Upadhyay, A. Query answering for kisan call centerwith lda/lsi. In *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 711–716 (IEEE, 2018).

32. Gandhi, V. P., Johnson, N. *et al.* Decision-oriented information systems for farmers: a study of kisan call centres (kcc), kisan knowledge management system (kkms), farmers portal, and m-kisan portal. *Agric. Situat. India* **76**, 29–32 (2019).

33. Sharma, N., Ghosh, S. & Saha, M. Analysis and visualization of farmer call center data. In *Open Data for Sustainable Community*, 151–167 (Springer, 2021).

34. Sharma, N., Ghosh, S. & Saha, M. An approach for exploring new frontiers for optimizing query volume for farmer call centre—kcc query pattern. In *Open Data for Sustainable Community*, 169–180 (Springer, 2021).

35. Momaya, M., Khanna, A., Sadavarte, J. & Sankhe, M. Krushi–the farmer chatbot. In *2021 International Conference on Communication information and Computing Technology (ICCICT)*, 1–6 (IEEE, 2021).

36. Godara, S. & Toshniwal, D. Sequential pattern mining combined multi-criteria decision-making for farmers' queries characterization. *Comput. Electron. Agric.* **173**, 105448 (2020).

37. Godara, S. & Toshniwal, D. Deep learning-based query-count forecasting using farmers' helpline data. *Comput. Electron. Agric.* **196**, 106875 (2022).

38. DeAngelis, D. L., Post, W. M. & Travis, C. C. Disease and pest outbreaks. In *Positive Feedback in Natural Systems*, 201–219 (Springer, 1986).

39. Mullin, R. Time warps, string edits, and macromolecules: The theory and practice of sequence comparison. edited by d. sankoff and jb kruskal. addison-wesley publishing company, inc., advanced book program, reading, mass., don mills, ontario, 1983. 300 pp. isbn 0-201-07809-0 (1985).

40. Likas, A., Vlassis, N. & Verbeek, J. J. The global k-means clustering algorithm. *Pattern recognition* **36**, 451–461 (2003).

41. Su, X., Yan, X. & Tsai, C.-L. Linear regression. *Wiley Interdiscip. Rev. Comput. Stat.* **4**, 275–294 (2012).

42. Conn, A. R., Gould, N. I. & Toint, P. L. *Trust region methods* (SIAM, 2000).

43. Stackhouse Jr, P. W. *et al.* The prediction of worldwide energy resources (power) project. In *United Nations Climate Change Conference, Community of Practice 26* (2021).

44. Jiménez, Á. B., Lázaro, J. L. & Dorronsoro, J. R. Finding optimal model parameters by discrete grid search. In *Innovations in Hybrid Intelligent Systems*, 120–127 (Springer, 2007).

45. Sharma, T. *et al.* Rice blast management through host-plant resistance: retrospect and prospects. *Agric. Res.* **1**, 37–52 (2012).

46. Khush, G. S. & Jena, K. Current status and future prospects for research on blast resistance in rice (oryza sativa l.). *Adv. genetics, genomics control rice blast disease* 1–10 (2009).

47. Bishwajit, G. *et al.* Self-sufficiency in rice and food security: a south asian perspective. *Agric. & Food Secur.* **2**, 1–6 (2013).

48. Fahad, S. *et al.* Major constraints for global rice production. In *Advances in rice research for abiotic stress tolerance*, 1–22 (Elsevier, 2019).

49. Das, T. *et al.* Conservation agriculture in an irrigated cotton–wheat system of the western indo-gangetic plains: Crop and water productivity and economic profitability. *Field Crop. Res.* **158**, 24–33 (2014).

50. Bana, R. S. *et al.* Planting period effects on wheat productivity and water footprints: Insights through adaptive trials and apsim simulations. *Agronomy* **12**, 226 (2022).

51. Simkhada, K. & Thapa, R. Rice blast, a major threat to the rice production and its various management techniques. *Turkish J. Agric. Sci. Technol.* **10**, 147–157 (2022).

52. Yadav, M. K. *et al.* Blast resistance in indian rice landraces: Genetic dissection by gene specific markers. *Plos one* **14**, e0211061 (2019).

53. Gautam, R. & Bana, R. S. Drought in india: Its impact and mitigation strategies-a review. *Indian J. Agron.* **59**, 179–190 (2014).

54. Singh, V., Singh, U. & Khush, G. Basmati rice of india. *Aromat. rices* **135** (2000).

55. Nayaka, S. *et al.* Current scenario and integrated approaches for management of finger millet blast (magnaporthe grisea). *Blast Dis. Cereal Crop.* 27 (2021).

56. Hassoun, M. H. *et al.* *Fundamentals of artificial neural networks* (MIT press, 1995).

57. Chen, Y., Kang, Y., Chen, Y. & Wang, Z. Probabilistic forecasting with temporal convolutional neural network. *Neurocomputing* **399**, 491–501 (2020).

58. Shen, G., Tan, Q., Zhang, H., Zeng, P. & Xu, J. Deep learning with gated recurrent unit networks for financial sequence predictions. *Procedia computer science* **131**, 895–903 (2018).

59. Yu, Y., Si, X., Hu, C. & Zhang, J. A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation* **31**, 1235–1270 (2019).