TEST Engineering & Management

March - April 2020
ISSN: 0193-4120 Page No. 8831 - 8839

Prediction of crop yardsticks to evaluate crop proliferation using a continual Data Mining solution

[1]Ananthi.N, [2]Abishek.B, [3]Akash Eswar, [4]Dheeban.K

[1,2,3,4]Department of Information Technology, Easwari Engineering College
Chennai, India
[1]nandhura@gmail.com, [2]abishek.balaji97@gmail.com, [3]immakash@gmail.com. [4]Dheepan19@gmail.com

Article Info
Volume 83
Page Number: 8831 - 8839
Publication Issue:
March - April 2020

Article History
***Article Received**: 24 July 2019*
***Revised**: 12 September 2019*
***Accepted**: 15 February 2020*
***Publication**: 09 April 2020*

Abstract

Quantifying crop growth at various stages of its life cycle is a daunting task which has to be done by farmers in order to get an idea about how the end yield will manifest. This task has been haunting farmers for more than centuries and analysts all over the world are trying to find out new methods to quantify crop yield and demystify the evaluation of various crop yardsticks. Helping farmers to take effective decisions in order to help them increase crop yield is the main focus of many current methodologies. The quality of the crops however, has not been suitably quantified and this has resulted in low crop sales in certain places. In India, crop quality is of utmost importance and many buyers refrain from purchasing low-quality crops directly from farmers thus resulting in inadequate food supply at various places. The proposed system looks to analyze various parameters affecting crop yield and quality and evaluate the end quality and yield with the currently available data. Various factors are considered for accurate estimation which includes climatic factors such as temperature, light usage efficiency, effective rainfall and wind speed along with soil, groundwater and various other factors. It is unprecedented in the fact it is intended to be while continuously learning and improving its accuracy over time. It is also an integrated solution applied as a continual solution which when implemented as a real-time system can collect real-time data and make predictions which aims to try and evaluate how the resulting crop yield and quality might be with the help of data mining techniques that work upon datasets which can contain historical and real-time data as the situation demands.

Keywords; *Data Mining;Agriculture;cropyield;cropquality;Krishnagiri.*

I. INTRODUCTION

Many have asserted the importance of continuous growth in the agricultural sector and Indian economists have referred to the cliché “Agriculture is the backbone of Indian economy” many a time. With a continuously growing population, the demand for higher yields isn’t going to cease in the near future. Over the years it can be seen that several factors play a key role in influencing the crop yield. These include climatic factors such as precipitation effective rainfall, wind speed, temperature and many more factors such as soil factors, fertilizers used and water supply characteristics. Finding the effective rainfall is very important for determining the water needs on an evenly considered time basis be it daily or weekly basis[3]. This helps in reducing the risk of underestimating the amount of water required for irrigation [1]. When considering an accurate estimate of the yield and quality, the thermodynamic definition of crop production plays an important role [2]. The ability of the crop to absorb the incident radiation helps in determining the yield which can be achieved. The soil quality also plays a major role along with the fertilizers and water availability which are very important factors in determining crop quality and yield [4]

India is the second largest producer of potato and it is worth noting that potato is the 4th most important

March - April 2020
ISSN: 0193-4120 Page No. 8831 - 8839

crop in the world. Tamil Nadu is one of the leading agricultural producers with respect to several crops such as banana and cassava. One crop whose production when high could lead to massive returns but, still has not yet reached the peak production-wise is the potato. One of the major producers of potato in Tamil Nadu is Krishnagiri and optimizing potato crop production would be very beneficial for various reasons. Krishnagiri also has a tropical climate thus making very conducive to potato growth. Krishnagiri is one among the prominent producers of the crop and if it is to have a major say in the potato market, then potato yields must surpass previous results. Despite the implementation of various governmental policies and precision farming on various scales after careful analysis, there is still much room for development [6]. With the intention of analyzing the factors which greatly influence potato yield and quality, a suitable method has been devised and its workflow is shown in Diagram I. Data mining techniques are used to simulate current conditions in order to predict the end crop quality and yield. The results can help farmers understand the current situation better and take better decisions based on the circumstances. Since the high production potato growth seasons usually occur between the periods April/May-August/September and August/September to December/January these growth seasons alone have been simulated for analyses purposes. It is also important to make sure that required measures are taken to ensure that the crop yield produces high-quality crops. Farmers will not be able to sell their crops to the middlemen who almost invariably buy their crops if the crops are not of high quality. The system proposed will act as a decision support system that will urge farmers to take decisions based on the results derived from the analysis [10].

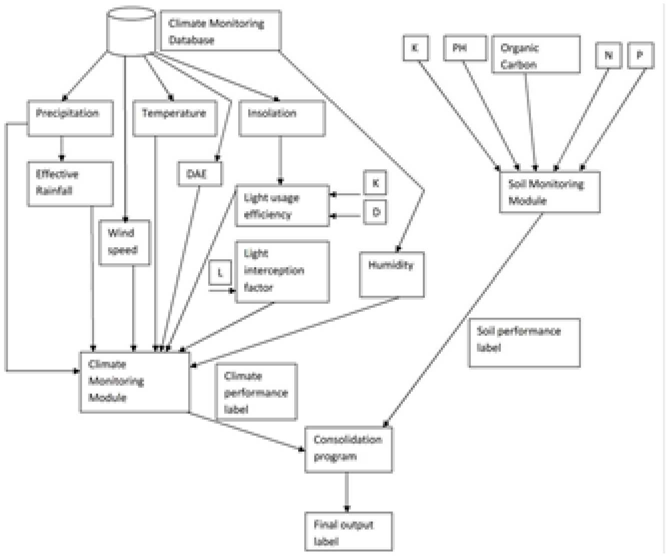

Fig. 1.Workflow diagram to depict the processes involved in the model

II. RELATED WORKS

Predicting the end crop yield and quality has been an extremely difficult task for many ages. There have been many models proposed by Britain researchers which characterized yield by the amount of dry matter produced [2] [5]. These thermodynamic based characterizations have led to the creation of models that have formed the basis for estimating crop yield based on radiation use efficiency. These models have stood the test of time and are still used for estimating crop yields. These models, however, fail to capture the effects of genetic variation and are unable to include cultivar specific information in yield estimation. The methods [7] which have incorporated cultivar specific information into their modeling process have not covered the influence of soil characteristics in detail as well. The quality of the end crops as well has not been touched upon in great detail. Researchers in China have estimated wheat yields for the winter season based on biomass estimation using multi-source data [12]. This method is also based on dry matter production and is efficient but, it fails to capture the quality of the resulting yield.

Another group of Researchers from China has predicted the change in crop yields with the

Published by: The Mattingley Publishing Co., Inc.

variation in climatic conditions [15]. They concluded that the climatic factors that majorly influenced potato yields were the temperature range, precipitation, radiation, and crop evapotranspiration. Researchers from Pakistan [4] have also characterized crop yield as a function of water availability, fertilizer consumptions, and credit distribution but they have not considered the importance of climatic variables which makes it ineffective in various circumstances.

Mining the available agricultural data to gain useful information such as interrelationships and correlation between variables has proven to be very useful as suggested by the authors in [11]. Clustering techniques such as DBSCAN, PAM, and CLARA have been used to determine a model consisting of the most influential agricultural factors and Multiple Linear Regression is used to predict the annual crop yield. This method primarily focuses on climatic factors and has not incorporated other influential factors such as soil, fertilizers, and other factors etc. Researchers from Mexico have validated the effectiveness of machine learning techniques for predicting crop yield and have come to the conclusion that the M5-Prime is an efficient method for predicting yield with high accuracy and precision [14].

Various Data mining methods such as clustering, classification, and Decision tree-based methods such as J48, LWL and LAD Tree etc. [10] have been used to gain interesting results. Similar algorithms will be applied throughout the scope of analysis to help farmers gain better insights about the resulting crop yield and quality. From the survey, we can conclude that the crux of our method is fundamentally very strong and integration with soil factors will provide a better estimation of the crop yields and quality.

III. PRELIMINARY ANALYSIS

The project kicks off the analysis process by taking into consideration all of the various factors that could influence crop yield and quality. The sample space of factors taken into consideration includes soil monitoring, manure monitoring, water availability monitoring, climate monitoring, and pesticide monitoring.

Upon close inspection, it was observed that middlemen acted as the main points of sale for farmers and these middlemen decided the rate for the crops. Since the rates were fixed by middlemen the profits made by the farmers were solely determined by the middlemen. So the analysis of the market situation in Krishnagiri has been rendered futile by the business model followed by the middlemen. Pesticides are something that is usually used by farmers only when needed on a reactive basis. Pests and crop diseases are usually identified by image processing techniques and with the advent of precision farming in Krishangiri, the development of high-quality water management practices and nutrient management practices have avoided crop losses due to pests and diseases along with correct planning of sprays [9][6]. Thus pesticide monitoring can be deemed to be of much lesser importance given current circumstances.

The addition of manure and fertilizers are important for replenishing soil nutrients and is also vital for the growth of potato crops. The main factors which characterize the effectiveness of manure and fertilizers are the nitrogen and phosphorus contents, ph of the soil and amount of organic matter [13]. The effectiveness of both manure and fertilizers has been considered under the soil monitoring module since it shapes the soil nutrient base and its effectiveness is determined based on the resulting soil profile. The various aspects of applying manure in different seasons are summarized in Table I. It is assumed that there will be no hindrances in receiving continuous water supply for irrigation and hence water availability has not been considered for further analysis.

March - April 2020
ISSN: 0193-4120 Page No. 8831 - 8839

TABLE. I. Effects of manure application during spring and fall

Application during Fall		Application during Spring	
Positive	Negative	Positive	Negative
Nutrient availability can be better estimated	Chances for N, P content loss over the winter period	Reduced risk of core nutrient loss after its application	The release of nutrients can be slower when a cool climate prevails
Risk of diseases becomes lower	Soil becomes bare after application of manure	Increased availability from producers of livestock	Higher chances for the prevalence of diseases

IV. PROPOSED METHODOLOGY

The sample space of factors after the preliminary analysis consists of climate monitoring, soil monitoring. Both modules produce an efficiency score in the form of labels which can take values 'A','B','C' respectively on an independent basis and these scores are combined in a consolidation program to give the final output. Both the modules are described in detail below.

A. Climate Monitoring

The Climate Monitoring module involves analyzing various factors such as temperature, precipitation received, effective rainfall, insolation, light usage efficiency, wind speed, and humidity. The methodology employed involves the use of theoretical considerations for estimating effective rainfall. These results are used along with a model for characterizing crop growth with light usage efficiency that is in turn coupled with other climatic factors such as humidity and wind velocity. The effective rainfall is found on a daily basis using very involved methods in the paper [1]. It has been found that depending upon the method applied in India that the effective rainfall is usually 50-80% of the received precipitation [3]. Experimental methods have observed that the effective rainfall received is approximately 70% of the average seasonal rainfall received. Thus it is assumed here that the weekly effective rainfall is nearly 70% of the received precipitation. This is an accurate measure of the effective rainfall but still not as accurate as the method provided in the paper [1]. The purpose of calculating the effective rainfall is to determine how much rainfall is effectively absorbed by the ground. This can be used to characterize the crop growth as more the amount of effective rainfall, the better the water needs of crops are satisfied and thus better crop yields can be observed as water stress is reduced.

The second subcomponent model is an improvised model suggested by Monteith in his publication [2]. This is a very effective model for characterizing crop growth of tropical crops and hence has been used to characterize the growth of potato. This model also assumes that light interception is complete throughout the year. The yield is characterized by the amount of dry matter produced which in turn depends on the light intercepted by the foliage and the light usage efficiency is characterized by the formula in equation 2.

$$\mathrm{E} = (K * D) / I - (2)$$

Here 'K' is a transmission coefficient which was characterized effectively by the considering Beers law equation for light transmission. 'D' is the dry matter that has been accumulated by the crop and 'I' is the light intercepted by the foliage of the crop.

This model analyzes crop growth in terms of light usage efficiency and light interception factor. The light interception factor is found to be an exponential function of 'L' which is the leaf area index. 'L' differs with crop growth periods and 'K' also varies proportionately with 'L'. The variation of leaf area index has been characterized in the paper [16] and this is used to determine the values of 'K'

Published by: The Mattingley Publishing Co., Inc.

March - April 2020
ISSN: 0193-4120 Page No. 8831 - 8839

which rise nearly proportionately with ‘L’. The first two stages of crop growth involve the use of the lower parabolic curve given in paper [16] and the latter two stages involve the use of the upper parabolic curve given in paper [16] for finding the values of the leaf area index using the equations given for each curve. K varies between values of 0.7(the value assumed during the application of the original model) and 0.3(the ‘K’ value for erect leaves. Potato leaves upon growth are predominantly erect and the values fixed here are based upon experimental results conducted by Monteith [2]. The insolation values are derived from experimental results in the publication [17]. The duration of the different crop stages are mapped accordingly to the crop seasons considered and this is used to help in calculating the leaf area index values characterized in the paper [16], which in turn is used to find out the light interception factor. It is to be noted that the calculation of light interception values does not take into consideration the effects of seasonal change. It was also found that when the light interception factor is low, measures such as net assimilation rate, leaf area duration, and relative growth rate can be used to characterize crop growth so lower values imply that the effect of mutual shading is less on the final crop yield and quality.

According to experimental results conducted Wheeler and his troops [18], it was observed that at higher humidity levels the yields obtained were higher but at lower humidity levels leaf areas were greater. Therefore it was observed that higher humidity levels result in higher yields but the quality is somewhat compromised and at lower levels quality increases but yield decreases. This tradeoff must be balanced and thus the optimal range for humidity may stretch between 50 and 80. The presence of precipitation implies lesser water is required for yield and thus its mere presence itself is beneficial. The optimal ranges/values for the various climatic parameters are given in Table II. The optimal value for wind speed is taken as 2m/s [20]. This value was derived by the FAO and is a good accurate estimate of the optimal speed. Strong winds are said to damage the shoots. Higher wind speeds increase the turbulence and thus increase the amount of available carbon dioxide available and thus increase the crop growth. Beyond a certain threshold value, the rate of photosynthesis becomes constant as well.

TABLE. II. Optimal Ranges/values of the various climatic parameters considered on a weekly basis.

Parameter	Optimal Range
Temperature	18 °C-20 °C
Precipiation	19-25mm
Insolation	3.3 GJ/m²
Light usage efficiency	0.28
Light interception factor	0.37
Humidity	50-85%
Effective Rainfall	13-22mm
Wind Speed	2m/s

B. Soil Monitoring

The soil monitoring module involves analyzing the role of various nutrients such as nitrogen, potassium, calcium, magnesium, zinc and phosphorus and determining how they influence crop growth. The ph and the amount of organic carbon are also taken into consideration. A quantitative analysis of the various factors is performed and the efficiency score found after the meticulous analysis in integrated with the score from the climate monitoring module to give the final output.

It was found that many of the soil testing laboratories do not have the required facilities to test for the presence of boron and sulfur. These could have hurt farmers in the long run as continual monitoring could have helped them to change the fertilizer dosage added. It has been observed that the micronutrients are in sufficient quantities in various

Published by: The Mattingley Publishing Co., Inc.

places in and around Krishnagiri and thus micronutrient sufficiency is not a problem.

The main soil factors that limit crop growth are the ph, nitrogen, phosphorus and amount of organic matter and these are determined by the manure or fertilizers that are added to soil [13]. Before delving into these factors it is to be noted that the soil quality in this model is characterized by the environmental conditions and interactions only and the genetic makeup is not considered in this model [19].

Nitrogen levels have been found to be very low at various places. This problem is also compounded by the fact that the nitrogen requirements of potato along with other nutrient requirements are very but the utilization of the supplied nutrients is comparatively low [8]. Thus a heavy dosage of nitrogenous fertilizers may seem a good solution but the resulting effects are not very promising. It was also found that the leaf area index and tuber dry weight increase as nitrogen is supplied to the plants in sufficient amounts [5]. However, this positive effect also wears as the nitrogen supply is continuously increased. The leaf area index and the tuber weight heavily influence the crop quality and yield and both are vital for achieving high standards. Thus it can be inferred that a sustained effort is required to improve the overall nitrogenous content in the soils of Krishnagiri. The phosphorus contents seem to be better off in several places and though it may be better in amounts compared to nitrogen improvement is still needed in and around Krishnagiri if we are to obtain higher yields. Current trends show a varied presence of phosphorus in various areas of Krishnagiri. It has been observed that an increase in dosage of phosphorus results in higher starch content but at the same time decreases the protein content. Thus higher dosage implies that crop yields may be higher but crop quality may be compromised.

Potassium levels are fairly high in various areas of Krishnagiri and thus it is not an issue when it comes to crop quality and yield. Higher potassium levels imply higher tuber size and dry matter production and also result in higher starch content. Thus the crop quality and crop growth increase with an increase in potassium levels.

Potatoes prefer acidic soils and it can be seen that the soil ph in Krishnagiri ranges from 5.1 to 8.4 approximately. Nearly 50% of the soil in Krishnagiri is moderately alkaline. Only 17% of the land has an ideal ph range in Krishnagiri. Thus the high variation of ph over the various land areas and slightly alkaline nature of soil could be responsible for the decrease in crop yields over time. The soils of Krishnagiri are also very much deficient in the organic matter. The increase in organic matter results in higher water holding capacity and slow release of nutrients. It also helps in maintaining the ph levels of the soil.The optimal ranges/values for the various soil parameters are given in Table III. These requirements are used to help in determining if the fertilizers added are sufficient for crop growth.

TABLE. III. Optimal Ranges/values of the various soil parameters considered on a seasonal basis.

Parameter	Optimal Range
PH	5.2-6.4
Nitrogen	187.5-204 kg/ha
Phosphorus	55.5-61 kg/ha
Potassium	296-326 kg/ha
Calcium	45.25-48.5 kg/ha
Magnesium	30.25-32.5 kg/ha
Organic Carbon	>0.75%
Zinc	44.92-112.3kg/ha

C. Consolidation Program

This part of the model deals with finding the relationship between the soil and climatic efficiency scores in determining the final yield and quality achieved. The yield and quality ranges are classified in three categories as High, Medium and Low. The output labels from this program correspond to the yield ranges displayed in Table IV. These ranges

Published by: The Mattingley Publishing Co., Inc.

were derived from historical potato yield production stats. The efficiency score from the soil module is fixed with the ‘B’ label and is expected to constant for at least the upcoming crop cycles. So the only remaining variant is the efficiency score generated by the climate module. From various analytical surveys and methodologies, it has been found that the potato crop has the ability to adapt to different climatic conditions. Moreover the resulting crop yield and quality reduce drastically only when the climatic conditions change adversely. Thus the outputs that can be derived are tabularized and are shown below in Table V.

TABLE. IV. Yield ranges corresponding to various outputs.

Final Output	Range
High Yield	40745-45368 kg/ha
Medium Yield	18749-40745 kg/ha
Low Yield	<18749 kg/ha

TABLE. V. Outputs corresponding to the different efficiency scores.

Efficiency score from climate module	Efficiency score from soil module	Output
A	B	High yield, medium quality
B	B	Medium yield, medium quality
C	B	Low yield and medium quality

V. EXPERIMENTAL ANALYSIS

It was observed that the Random Forest algorithm was most suitable for making accurate predictions about the resulting crop yields and quality overall, after a thorough analysis of various suitable algorithms using Weka. Thus the Random Forest algorithm forms the core of the predictive capabilities exhibited by our model. The efficiency scores predicted during each week after supplying the required input features are fed to the consolidation module which generates the required output. The experimental analysis of the various algorithms is shown in Table VI. This model also comes to the conclusion that the efficiency score generated by soil module will be of the label ‘B’. This is a conclusion that has been arrived upon after thorough analysis and will remain fixed at least for the next upcoming crop cycles. Continuous monitoring must be in place in order to check the suitability of the soil. Proper crop rotation practices must also be put in place in order to prevent deficiencies of nutrients in the future.

The trends observed for the crop seasons of the year 2018 were in line with the recorded yield and quality statistics. A decline in potato yields has been observed in recent years in Krishnagiri. This decline is due to several factors. It was found that the soil health of various places in Krishnagiri was very poor and many essential nutrients were depleted. The ph and nitrogen levels severely restrict the growth of potato in this region. These are limiting factors and unless worked upon in the future, the soil quality cannot be improved. It was also observed that the light usage efficiency values were comparatively higher than that of the values observed in Britain and thus gives Krishnagiri an edge in producing better crop yields.

Published by: The Mattingley Publishing Co., Inc.

TABLE. VI. Experimental analysis of the different classifiers used.

Classifier Used	Parameter			
	Root Mean Squared Error	*Recall*	*Precision*	*F-Measure*
JRip for finding the efficiency score	0.3317	0.875	0.875	0.933
J48 for finding the efficiency score	0.3536	0.875	0.875	0.933
Logistic Regression for finding the efficiency score	0.3537	0.875	0.938	0.891
Random Forest for finding the efficiency score	0.2982	0.875	0.875	0.933

VI. WORK CONCLUSION AND FUTURE

This project was developed with the aim to develop a respectable model to predict the end quality and yield of potato crops grown in Krishnagiri. The model provides a base upon which further improvements can be made and if integrated with other soil monitoring systems can produce better results. The model is expected to bring more areas into the mix such as genetic makeup of the crops and the effectiveness of the water supply. The dataset is also small to make its use effective on a large scale. This model does, however, prove to be a suitable solution when applied on a small scale and can be the start of many great models to come.

VII. ACKNOWLEDGEMENT

This work was supported by DST-FIST Programme No.SR/FST/College-110/2017, Government of India

REFERENCES

[1]. Abishek.B, R.Priyatharshini,AkashEswar and P. Deepika, “Prediction of effective rainfall and crop water needs using data mining techniques.”,IEEE Publications 2017

[2]. J.L.Monteith,F.R.S,”Climate and efficiency of crop production in Britain”,Phil.Trans.R.Soc.Lond.B.281,277-294,1977

[3]. N.G Dastane,Natural Resources Management and Environment Department,”Effective rainfall in irrigated agriculture”,Food and Agriculture organization of the United Nations publications,1978

[4]. Abdul Rehman,Abbas Ali Chandio,ImranHussain,LuanJingdon,”Fertilizer consumption, water availabilty and credit distribution: Major factors affecting agricultural productivity in Pakistan,Journal of the Saudi Societty of Agricultural Sciences,2017

[5]. E.J.Allen,R.K.Scott, “An analysis of the growth of potato crop” J.agric.Sci.,Camb.,1980

[6]. Shibusawa,”Precision farming approaches to small-farm agriculture”Agro-Chemicals Report Vol.II,2002

[7]. J.T.Richie,U.Singh,D.C.Godwin,W.T.Bowen,” Cerealgrowth,development and yield”,Systems Approaches for Sustainable Agricultural Development,Springer Publications

[8]. NeBambiLutaladio,OscarOrtiz,Antonhaverkort ,DanielCaldiz,”Sustainable Potato Production,Guidelines for developing countries”,Food and agriculture organization of the United Nations Publications

[9]. Prof. SonaPowara,DnyaneshNawale,KunalPatil,RakeshMahajan,”Early Detetection of Pomegranate Diseases Using Machine Learning and Internet Of Things”,3rd International Conference for convergence in Technology,2018

[10]. ShrutiMishra,PriyankaPaygude,SnehalChaudary,SonaliIdate,"Use of Data Mining in Crop Yield Prediction",Second International Conference on Inventive Systems and Control,2018
[11]. JharnaMujamdar,SnehaNaraseeyappa,ShilpaAnkalaki,"Analysis of agriculture data using data mining tecnhniques:application of big data",Journal of big data,Springer Publications,2017
[12]. JianqangRen,SuLI,ZhongxinChen,QingbouZhou,HuajunTang,"Regional yield prediction for winter wheat based on biomass estimation using multi-source data
[13]. S.S.Snapp,J.Nyiraneza,M.Otto,W.W.Kirk,"Managing Manure in Potato and Vegetable Systems",Department of Horticulture,MSU,2003
[14]. Alberto Gonzalez-Sanchez,JuanFrosto-Solis,Waldo Ojeda Bustamante,"Predictive ability of machine learning methods for massive crop yield prediciton"Spanish Journal of Agricultural Research,2014
[15]. JunfangZhao,YanhongZhang,YonglanQian,ZhihuaPan,YujieZhu,YiZhang,JianpingGuo,LinglingXu,"Coincidence of variation in potato yield and climate in northern China",Science of the Total Environment,2016
[16]. Pedro Manuel Villa,LinaSarmiento,Fermin Jose Rada,DanielMachado,Alice Cristina Rodrigues,"Leaf area index of potato crop under three nitrogen fertilization treatments",Agronomia Colombiana,2017
[17]. Frank Samuel,M.Venkateshkumar,G.Anandhakumar,"Analysis of Solar Irradiation and modeling of large scale PV powe plant implementation in North Tamil Nadu",Jounal of Chemical and Pharmaceutical Sciences"2016
[18]. R.M.Wheeler.T.W.Tibbitts,A.H.Fitzpatrick,"Potato growth in response to relative humidity",American Society for Horticultural Science,1989
[19]. EgataShunkaTolessa,"Importance,Nutrient content and factors affecting nutrient content of potato",American Journal of Food,Nutrition, and Health,2018
[20]. A.Abdul Harris,V.Chhavra,B.T.Bhatt,A.K.Sikka,"Yield and duration of potato crops in Bihar under projected climate scenarios",Journal of Agrometrology,2015