RESEARCH ARTICLE OPEN ACCESS

ARMA BASED CROP YIELD PREDICTION USING TEMPERATURE AND RAINFALL PARAMETERS WITH GROUND WATER LEVEL CLASSIFICATION

R.Navin Kumar M.C.A.,M.Phil.,[1], V.Emayavarman[2]

[1]Assistant Professor, Department of Computer Applications, Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.

[2]Final MCA, Department of Computer Applications, Nandha Engineering College (Autonomous), Erode, Tamilnadu, India.

Email: [1]navinsoccer07@gmail.com, [2]emayavarman535@gmail.com

Abstract. Nowadays, wireless telecommunication networks are promising alternative for rainfall measuring instruments that complement previous monitoring devices. Due to big dataset of the rainfall and the telecommunication networks data, empirical computational methods represnt less adequate of actual data. So, deep learning models are proposed for the analysis of big data and give more accurate presentation of real measurements. This project investigated rainfall monitoring results from experimental measurements. The main aim of this study is to provide a methodology for rainfall data classification based on neural network methods based on the historical rainfall data production data. Classification based on the previous years of rainfall can help farmers take necessary steps to measure crop production in the coming season. Understanding and assessing future crop production can help ensure food security and reduce impacts of climate change. In this work, ARMA (Auto Regressive Moving Average) method is used for proposed work. Past ten years of data set is taken for rainfall and ground water level for our country. The proposed work classifies the ground water level data set records using ARIMA model to predict the model for future test record data sets. The new model will helpf for analyzing ground water levels in past and so as to find the future levels.

Keywords: Crop Yield Prediction, ARMA Model, KNN Classification, Neural Network.

I. INTRODUCTION

Real-time precipitation observations show a significant role in various aspects of human life, such as agriculture, hydrometeorology, natural disaster warning and climate monitoring. At present, rainfall monitoring methods comprise weather radar, rain gauges, and weather satellites [1], [2].

Rain gauges (RG) are used as precise ground-based rainfall estimation instrument. But, they do not provide rainfall information with more spatial resolution owing to errors made by calibration or ground winds [3-5]. Weather radars address the shortcomings of RG and provide a wide range of precipitation. The associate editor coordinating the review of this manuscript and approving it for publication is Guido Valerio distribution information, but ground clutter frequently affects it, which produces less accurate ground-level observations.

In South Korean context, radar monitoring network operated by Korea Meteorological Administration has a comparatively high density and is mounted at positions appropriate for observations of peninsular part of that country. Still, it has incomplete representation in urban areas. Therefore, it is overbearing to develop real-time, accurate, and representative rainfall measurement techniques.

Recently, wireless telecommunication links have grown attention as a promising rainfall measurement method because the power of received signals, which is extremely delicate to rainfall in microwave and millimetric frequency bands, can be measured everywhere for communications.

Several studies have shown the potential application of terrestrial radio links for rainfall forecast in different parts of the world over several different carrier frequencies (5–50 GHz). Outstanding examples are the nationwide campaign in Germany and the Netherlands, which confirmed that commercial microwave links (CML) provides precipitation data which correspond well with gauge and weather radar rainfall data. Rainfall data taken from telecommunication signal are successfully applied for the simulations of river runoff in Germany for pre-alpine catchment area, and then for urban catchment pipe flow simulation in Czech Republic.

In addition, CML-derived rainfall generation in developing country was also carried out using Burkina Faso's CML data.

This study makes the following contributions:

• To the best of our knowledge, this is the first wireless telecommunication-based monitoring system for rainfall in a) South Korea and b) Ethiopia in the frequency bands. The retrieval of rainfall data from wireless communication links has been investigated in South Korea over four terrestrial line-of-sight links and two satellite networks, and over one terrestrial line-of-sight at microwave frequency spectrum link in Ethiopia.

• The rainfall rate distribution statistical results experiment from measured rain data were compared with rain rate produced by two proposed deep learning models. In addition, past studies have been mainly the wet-dry periods classification, however, the study deals with generation of rainfall at every level and comparing the distribution over the year.

• This study is substantially grounded on the factual entered signal power and rain rate data measures, which make the findings virtually sound.

• While deep learning in many main fields has achieved several milestones, but the black box aspect of deep learning impedes its comprehension significantly because of its nonlinear nature. As a result, this study attempts to unravel every black box in the deep learning networks that have been introduced.

II. LITERATURE REVIEW

In this paper [1] the authors stated that most of greenhouse growers need a determined amount of yields in order to accurately face market requirements. The purpose of this study was to explore the neural networks dynamics for forecasting crop-yield with environmental variables; they aimed in producing accurate yield amount.

They used ANFIS which is termed as Adaptive Neuro Fuzzy Inference System. The input given to ANFIS is several parameters derived from crop growth model (Co2, temperature, vapor, radiation, yield and pressure deficit (VPD)).

ANFIS is having only one output node, which is the yield. One of the toughest issues during predicting yield is remote sensing data don't go long back in time. So, any predicting effort is forced to utilize a very restricted number of past years so as to construct a model for forecasting future values.

Their system is trained by leaving 1 year out and using the other entire data. Then they evaluated the deviation of their estimates compared to yield of the year that is residual. The procedure was applied to all the remaining years and average forecasting accuracy was given.

ANFIS is considered as a class of adaptive networks which perform, as a framework in adaptive fuzzy inference [6] systems. Generally, this is a multilayer feed forward adaptive network in which all nodes realized a particular node function of corresponding inputs and nodes in ANFIS include adaptive / fixed ones, and ANFIS is characterized with parameter set that is union of the parameter sets associated with all the adaptive nodes. The use of neuro-fuzzy system to yield crop [7] estimate had some motivating characteristics. All the variables that are input into system were associated with varying accuracy degrees. Some uncertainty comes first from measurement error and then from generality.

Using fuzzy sets instead of actual values as inputs, the authors aimed at shifting to semantics of the data rather than the measure [6]. It is known that with neuro–fuzzy modeling there is an alternative to use fuzzy set as the output. Here in this case, yield is expressed for example as low/normal of high with each of these three borders corresponding to the real fuzzy set.

The authors did not imply that seeking a crisp non-fuzzy value is the more exact approach than seeking the trend expressed in fuzzy sets (low/medium, etc.). But the accuracy of prediction is probably equal in both expressions in desired output we are more used to and feel confident in looking at numeric values rather than a membership function.

ANFIS accepts numerical inputs and yields a single out value. ANFIS is susceptible to "dimensionality curse". Training time increases exponentially with respect to fuzzy sets number per input variable used. To illustrate this, the authors considered the system with eight input variables that are encoded into two fuzzy sets (e.g.. low/high) and has 256 rules.

If the authors chose to use 3 linguistic variables instead of two (say low/medium/high), then the number of fuzzy rules is 6561. This phenomenon limits the input variables choice and expression of those variables into the meaningful fuzzy sets.

Fuzzy logic was first introduced by [6] [8] for representing and manipulating data and information in which there are different forms of uncertainty. Fuzzy rule-based systems are being used linguistic variables to reason use the series of logical rules which contain IF-THEN rules which connect i) antecedent(s) and ii) consequent(s), respectively.

An antecedent is a fuzzy clause with certain membership degree (between zero and one). Fuzzy rules may have multiple antecedents joined with AND or OR manipulators, in which all parts are calculated in parallel and resolved into the single number. Consequents are also comprised of multiple parts, which are aggregated then into a fuzzy set's single output [9]. Fuzzy inference [10] is the process of mapping a given input to an output using fuzzy set methods.

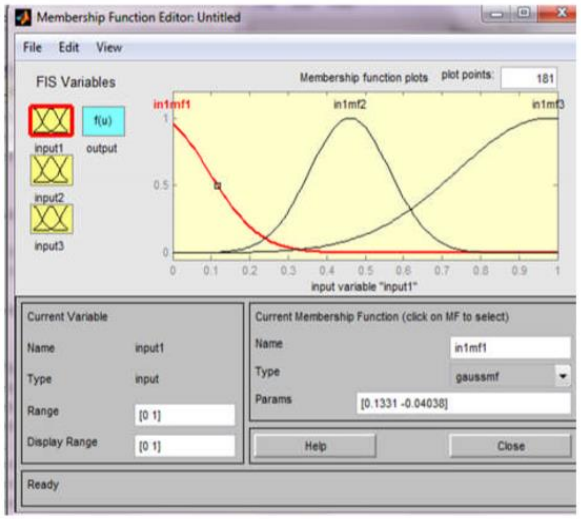

Fig 2.1 MEMBERSHIP FUNCTIONS DEGREE

Fig2.1 shows an illustration of a fuzzy conclusion system with five functional blocks. The fuzzification element transforms each crisp input variable into a class grade grounded on the class functions defined. The conclusion machine also conducts the fuzzy logic process by applying the applicable fuzzy drivers in order to gain the fuzzy set to be accumulated in the affair variable. The defuzzifier transforms the fuzzy affair into a crisp affair by applying a specific defuzzification system.

They concluded that this trouble showed that a neuro-fuzzy configuration can be used for tomato yield vaticination for the region with respectable results. In future they intend to make other trials with the same introductory configuration. They would use different parameters as inputs and alternate test more different ANFIS configurations.

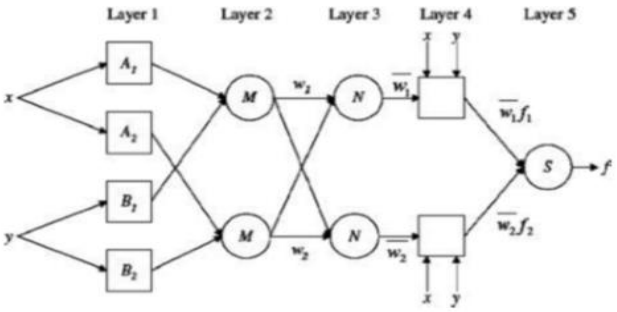

FIG 2.2 ANFIS ARCHITECTURE

They believed that parameters similar as the number of fuzzy sets, the type of class functions as well as considering different parameters per input, should admit further experimental trouble. Also a possible way is to use a inheritable algorithm to elect the optimal values.

The computation of transfer capability by using ANFIS is suitable for the operation which fleetly updates the value. Given the representation of fuzzy conclusion systems, in which knowledge is decoded as a set of unequivocal verbal rules that can be fluently understood by people without specialized moxie, it's hoped that this will allow the objectification.

In this study (2) a seasonal ARIMA model was erected using Box and Jenkins system to read long term downfall in Sylhet. For this purpose downfall data from 1980 to 2010 of Sylhet station were used to make and check the model.

Downfall data from 1980 to 2006 were used to develop the model while data from 2007 to 2010 were used to corroborate the vaticination perfection. Four introductory chronological way videlicet identification, estimation, individual checking, and soothsaying were fitted out in developing the model. Box ad Jenkins gave the validity of model which was tested by standard graphical explanation of residuals. As a alternate step of confirmation, read values of yearly downfall were checked using factual data series.

After completion of necessary checking and cast observation, the ARIMA (0, 0, 1) (1, 1, 1) 12 was plant to be the most effective to prognosticate unborn rush with a 95 confidence interval. It's anticipated that this long term vaticination will help decision makers in effective scheduling of deluge vaticination, civic planning, rainwater harvesting and crop operation.

Downfall is veritablynon-linear in nature and veritably complicated to prognosticate. Due to adverse goods of climate change downfall pattern has also been changing fleetly Short term and long term cast of downfall have significant applicability to agrarian, tourism, deluge forestallment and operation strategy and water body operation which impact the frugality of a country. To prognosticate similar event, multitudinous ways including numerical and machine literacy processes have been espoused grounded on literal time series and radar data (Chander etal. 2002, Ingsrisawang etal. 2008).

Still, presently most common methodology for downfall vaticination uses radar image data available from colorful associations and assaying them to prognosticate downfall. Still colorful statistical styles are frequently useful to prognosticate downfall (Bisgaard and Kulahci, 2011). Among which the most effective approaches for assaying time series data is the model introduced by Box and Jenkins (1976) and modified by Box, etal. (1994), also known as ARIMA (Autoregressive Integrated Moving Average).

ARIMA has extensively been exercised over the times to prognosticate the downfall trend (Mahsin etal. 2012) and force and swash modeling (Valipour etal. 2012. The system has some intriguing features that made it more desirable for experimenters. It eases the soothsaying process allowing experimenters to use only single variable time data series while also allow multiple for more complex cases. Downfall cast study for Dhaka division of Bangladesh has been done by Mahsin etal. (2012).

Still, soothsaying of downfall for Sylhet has n't been done yet. The area is one of the top sightseer attracting locales in Bangladesh. Although the area is hard the world's wettest place (Cherapunji), civic Sylhet and some

other region including Surma swash receptacle are facing a rapid-fire ground water reduction rate.

Besides, inordinate iron and arsenic impurity has an adverse impact in drinking water force. To pursue a sustainable indispensable source of force, rainwater has come a suitable option in numerous corridor of the region. Thus the study ideal of this study is to concentrate on development of a dependable soothsaying of downfall over Sylhet to manage water coffers as well as handle flash deluge effective and timely.

They concluded that then an ARIMA model has been developed. Comparing the observed and read values with 95 confidence limit, the presented model gives a reasonable result. Thus, this model could help to determine possible unborn strategy in the separate field for the Sylhet City and its near areas. Therefore, this model could help to determine possible future strategy in the respective field for the Sylhet City and its nearby areas. Model diagnostic checking presented that ARIMA (0,0,0) (1,1,1) and ARIMA (0,0,1) (1,1,1) should have significant results.

In this paper (3) the authors stated that with the impact of climate change in India, maturity of the agrarian crops are being poorly affected in terms of their performance over a period of last two decades. Predicting the crop yield well ahead of its crop would help the policy makers and growers for taking applicable measures for marketing and storehouse. Similar prognostications will also help the associated diligence for planning the logistics of their business.

Several styles of prognosticating and modeling crop yields have been developed in the history with varying rate of success, as these do n't take into account characteristics of the rainfall, and are substantially empirical. In the present study a software tool named‘Crop Advisor’has been developed as an stoner friendly web runner for prognosticating the influence of climatic parameters on the cropyields.C4.5 algorithm is used to find out the most impacting climatic parameter on the crop yields of named crops in named sections of Madhya Pradesh. This software provides an suggestion of relative influence of different climatic parameters on the crop yield, otheragro-input parameters responsible for crop yield aren't considered in this tool, since, and operation of these input parameters varies with individual fields in space and time.

Crop Product is a complex miracle that's told byagro-climatic input parameters. Agriculture input parameters varies from field to field and planter to planter. Collecting similar information on a larger area is a daunting task. Still, the climatic information collected in India at every 1sq. m area in different corridor of the quarter are tabulated by Indian Meteorological Department. The huge similar data sets can be used for prognosticating their influence on major crops of that particular quarter or place.

There are different soothsaying methodologies developed and estimated by the experimenters each over the world in the field of husbandry or associated lores. Some of the similar studies are Agrarian experimenters in Pakistan have shown that attempts of crop yield maximization throughpro-pesticide state programs have led to a dangerously high fungicide operation. These studies have reported negative correlation between fungicide operation and crop yield (11).

In their study they've shown that how data mining integrated agrarian data including pest gibing, fungicide operation and meteorological data are useful for optimization of fungicide operation. Thematic information related to husbandry which has spatial attributes was reported in one of the study (12). Their study aimed at sapient trends in husbandry product with references to the vacuity of inputs. K- means system was used to perform vaticinations of the pollution in the atmosphere (13), the k nearest neighbor was applied for bluffing diurnal hustle and other rainfall variables (14), and different possible changes of the rainfall scripts are anatomized using SVMs (15).

Data mining ways are frequently used to study soil characteristics. As an illustration, the k- means approach is used for classifying soils in combination with GPS-grounded technologies). Apples were checked using different approaches before transferring them to the request.

Spatial data mining introduced especially decision tree algorithm applying to husbandry land grading (16). He combined spatial data mining ways with expert system ways and applied them to establish an intelligent husbandry land grading information system. The author espoused decision tree C4.5 algorithm and apply with Mo2.0 and VC6.0 to make husbandry land grading expert system. The study showed the advantages of this system in addressing problems in land grading.

A decision tree classifier for agriculture data was proposed (17). This new classifier takes new data expression and deals with both complete data as well as incomplete data. In the trial run, ten-fold cross confirmation system is used to test the dataset, horse-colic dataset and soybean dataset. Their results showed the proposed decision tree is able of classifying all kinds of husbandry data. Data booby-trapping fashion for elaboration of association rules for famines and cataracts in India was applied using climate inputs (18).

In their study, a data-mining algorithm using the generalities of minimum circumstances with constraints and time lags was used to discover association rules between extreme downfall events and climatic indicators. Downfall events were read the using data miningtechniques.The circumstance of dragged dry period or heavy rain at the critical stages of the crop growth and development may lead to significant reduction in crop yield. Sugarcane yield was estimated in Brazil, using 10-day ages of SPOT foliage NDVI images and meteorological data.

Data Mining approach grounded on Spatio-Temporal data to read irrigation water demand. A set were prepared containing attributes attained from meteorological data, remote seeing images and water delivery statements. In

order to make the set data sets useful for demand soothsaying and pattern birth data sets were reused using a new approach grounded on a combination of irrigation and data booby-trapping knowledge. Decision tree ways were applied to read unborn water demand.

They concluded that the present study demonstrated the implicit use of data mining ways in prognosticating the crop yield grounded on the climatic input parameters. The developed webpage is stoner friendly and the delicacy of prognostications is over 75 per cent in all the crops and sections named in the study indicating higher accuracy of prediction. The stoner friendly web runner developed for prognosticating crop yield can be used by any stoner their choice of crop by furnishing climatic data of that place.

In this paper (4) the authors stated that analytics frequently involves studying past literal data to probe implicit trends. Weather condition is the state of atmosphere at a given time in terms of rainfall variables like downfall, pall conditions, temperature,etc., the being models use data mining ways to prognosticate the downfall. The main disadvantage of these systems is that it doesn't give an estimate of the prognosticated downfall. The system calculates normal of values and understand the state of atmosphere, which doesn't yield estimate results.

This paper represented a fine system called Linear Retrogression to prognosticate the downfall in colorful sections in southern countries of India. The Linear Retrogression system is modified in order to gain the most optimum error chance by repeating and adding some chance of error to the input values. This system provides an estimate of downfall using different atmospheric parameters like average temperature and pall cover to prognosticate the downfall.

The direct retrogression is applied on the set of data and the portions are used to prognosticate the downfall grounded on the corresponding values of the parameters. The main advantage of this model is that this model estimates the downfall grounded on the former correlation between the different atmospheric parameters. Therefore, an estimate value of what the downfall could be at a given time period and place can be plant fluently.

The operation of wisdom and technology that predicts the state of atmosphere at any given particular time period is known as Weather soothsaying. There's a numerous different styles to ride cast. Weather read notices are important because they can be used to help destruction of life and terrain.

The rainfall soothsaying styles used in the ancient time generally inferred pattern recognition i.e., they generally calculate on observing patterns of events. For illustration, it's plant that the following day has brought fair rainfall; if the antedating day evening is particularly red.

Still, all of the prognostications prove not to be dependable. Then in this system we used parameters like average temperature, pall cover to prognosticate the downfall. The data set of 100 times is taken for this design and the enforced using Numerical styles. Weather soothsaying system uses atmospheric parameters like moisture, wind and temperature and cast rainfall grounded on antedating record, therefore, this soothsaying is more dependable.

There are numerous operation that this system be used similar as Air Traffic, Agriculture, Marine, Forestry, Navy, and Military etc. The training of the data is done using a modified interpretation of direct retrogression system. The error chance between the factual and prognosticated is used to ameliorate the training set and train the data with the new inputs Therefore, with the advantages this system, they could read the issues till there's no farther enhancement in the error chance. This system can be used to read the downfall and help the destruction caused by it to the life or property.

They used a modified interpretation of Linear Retrogression to perform the vaticination of downfall in our system. The process of this system is explained in these following way

1. The input data sets are examined. The input data of training set is attained from 1901 to 2002 for each month to perform the proposed system and check the system.

2. The training and test data are formed from the input data sets. The training set contains the average temperature, downfall and pall cover from 1901 to 1970 from the input data sets. The proposed system is applied on these training sets. The test data contains the data from 1971 to 2002 on which the testing of model is done.

3. The direct retrogression is applied on the training data sets and the downfall is read using the downfall in training data as dependent variable and average temperature and pall cover as independent variables.

4. The error chance is now calculated by abating the prognosticated value from the factual value and multiplying it with 100 to get chance.

They concluded that downfall is the major cause for numerous of the natural disasters like flash cataracts, famines, surfs. So in order to help these natural disasters, we should be suitable to prognosticate the cause of the source. The proposed system can be used to estimate the downfall over the required period so that the separate authorities can take preventives to help the loss of life and property.

The proposed system uses modified direct retrogression approach to prognosticate the downfall that has lower error chance than compared to utmost data mining ways like clustering, back propagation which provides the generalized values rather than estimate values. This data is used to perform the necessary computations to prognosticate the downfall from average temperature and pall cover of that particular quarter.

They might improve this system further using multiple retrogressions which can take multiple months at a time as input and just forming a single equation which leads nearer to an accurate downfall prognosticated. The proposed approach may also be used in other operations like, in seminaries to prognosticate the average marks of their scholars, in sports to prognosticate the scores or winning brigades grounded on their former performance, in enterprises to estimate their gains,etc.,.

In this paper (5) the authors stated that prognosticating downfall is a veritably tough and indigent problem. Experimenters worldwide are trying to do it with different styles. Still no similar vaticination gives accurate result, because of unusual and doubtful changes passed. Then they tried to prognosticate the yearly downfall using Box-Jenkins Seasonal Auto Regressive Integrated Moving Average model, with 136 times of downfall data. Monitoring the environmental changes especially climatic changes is the veritably essential part of erecting future. Downfall vaticination is the most concerned variable due to its commencement and non uniformity. Numerous experimenters contributed towards the vaticination of downfall in Tamilnadu.

O.N.Dharet.al. anatomized the average downfall for the north east thunderstorm using standard styles. They anatomized trend, periodicities and variability for vaticination periodic downfall in Tamilnadu.O.N.Dhar and Rekecha attained the relation between southwest and northeast through a correlation analysis. The NEMR – STA association is displayed by Balachandranet.al. In this exploration work, they've used the classical statistical seasonal ARIMA model for yearly downfall vaticination.

Their study is within Tamilnadu region bounded on the north by Andhra Pradesh and Karnataka, on the east by the Bay of Bengal, on the south by the Indian Ocean and on the west by the state of Kerala. The state has a latitude between 8o 5'N and 13 o 35'N and longitude between 78 o 18'E and 80 o 20'E.

It receives downfall by both southwest and northeast showers. The fate of the agrarian frugality of the state is decided by the showers. Tamilnadu, located in southeast Peninsular India receives the major part of its periodic downfall during the northeast thunderstorm (the three-month period from October to December).

While littoral Tamilnadu receives about 60 of its periodic downfall, interior Tamil Nadu receives about 40 – 50 of periodic downfall during Northeast thunderstorm season. Roughly 33 of periodic downfall in Tamilnadu is from the southwest thunderstorm and 48 is from the northeast thunderstorm. They concluded that their study revealed that the time series model SARIMA is a veritably useful tool for vaticination of downfall. This is substantiated from the following figure 6 and also seeing the MAPE value which is veritably less.

III. PROPOSED METHODOLOGY

This study incudes the existing method of crop yield prediction and suggests new innovative method. In existing study, the data set is taken from Indian meteorological sites saved in excel files as 'csv' files. It contains rainfall data. These rainfall data are taken for pre-processing of text first and converted into time series data set format. Then artificial neural network is used to classify the new test records with the given about data set. It will be helpful to predict the future years' levels. Drawbacks are:

- Only classification is made among past records Future rainfall level is not predicted yet.
- Aims in classification only and helps to assess future crop yield production.
- It could not be preferred when the outlier data is more.

In proposed system, the existing methodology is also carried out. Like existing system, here also, Indian meteorological sites dataset is taken and saved in excel files as 'csv' files which contains rainfall / ground water level data. These data are then converted into time series data set format and ARMA model is used for predicting the future years' levels and then a fuzzy logic is applied for classifying crop yield in future years.

In addition, it can be applied for classification and so it is found to be suitable especially if the data set is having a greater number of records is contains outlier data. A wide variety of ground water level records can be taken for classification purpose and predicting a new model at the same time increasing the efficiency. Advantages are:

- The ARMA model yields better accuracy only if the test record is exactly matched with any of the training data records.
- ARMA is applied so could be preferred when the data set grows larger.
- ARMA can be applied even when the outlier data is more.

This study is carried out by spliting work into these modules.

1. ANN BASED CLASSIFICATION OF RAINFALL DATA

Artificial Neural Network: It includes single input layer, 1 hidden layer (can be made to two if required) and 1 output layer, whose input is instance's feature vector and output is the instances' category. It also includes one propagation process, then forward propagation and then back propagation. The process of propagation is in Algorithm 1.

ALGORITHM 1

INPUT: DESCRIPTION FOR VECTOR ENCODED (Name of State, Year and values of Annual rainfall are encoded to binary strings. Take for example, if four states are taken then first state value is 0001, second state value is

0010 and so on. Likewise if 15 years are taken, then first year value is 000000000000001, second year value is 000000000000002 and so on continuing. Likewise, the entire columns are taken for input layer neurons are concatenated and then bit strings are prepared for each neuron.

OUTPUT: 1 Output neuron with value zero to one. The category count is set to three so for any given input neurons with thousand neuron values, if value between zero and 0.333 is prepared as the output in the output neuron, then category belongs to the first (out of thousand categories), 0.334 to 0.666 belongs to next category and so on.

Fetch all records, convert into bit strings, 32 hidden layer, 3 output layer neurons are set and network is made to run.

The output is generated and displayed.

For all records, above process is made and output is displayed.

The weights / biases for hidden and output layer are recalculated for any given number of epochs (ten iterations).

The forward propagation of input layer and hidden layer uses the ‘tanh’ as activation function, while output layer uses ‘softmax’ as activation function.

Final weights / bias values are taken for futuretest description samples.

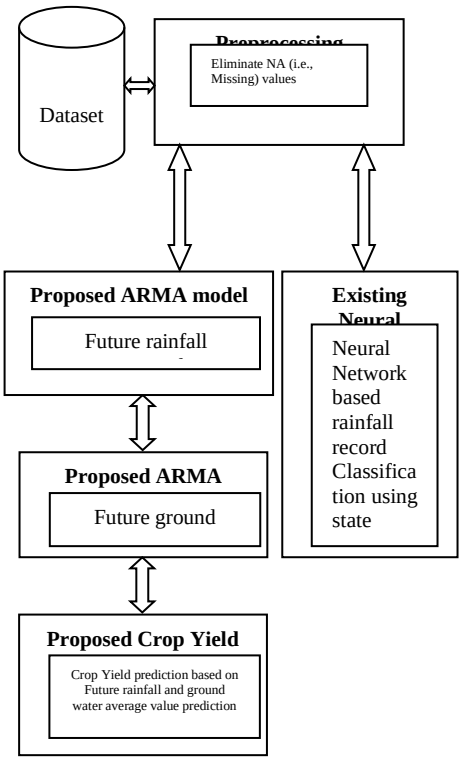

FIG 3.1 ARCHITECTURE DIAGRAM

2. ARMA MODEL BASED PREDICTION FOR RAINFALL

Here, rainfall water data set are taken for Indian rainfall data for past years. Then the data is converted into data frames and preprocessed so that zero values in the entire columns records are eliminated. The data is then converted into time-series format such that 12 records (for each month) for all years present in data set. Then using "arima’ function in R, given data set is used to predict the model and predicted after for upcoming years. Using ts.plot() upcoming years values are also plotted.

3. ARMA MODEL BASED PREDICTION FOR GROUND WATER

In this third module, groundwater level data set is read out for Indian rainfall data for past ten years. The data is converted to data frame and zero values in all columns records are removed/eliminated. The data is converted to time-series format such that 12 records (monthwise) for all years present in data set. Using ‘arima’ function, given data set is used to predict the model and then prediction is carried out for next upcoming years. Using ts.plot() next upcoming years values are plotted.

4. CROP YIELD PREDICTION BASED ON RAINFALL AND TEMPERATURE

In this fourth module, using previous module rainfall / temperature outcomes, crop yield prediction process based onfuzzy logic is carried out using the following algorithm.

IF (rain=='very good' && temp=='very good') || (rain=='very good' and temp=='good'): THEN
yield = 'very good'

IF (rain=='very good' && temp=='average') || (rain=='good' && temp=='very good') || (rain=='good' and temp=='good') || (rain=='good' && temp=='average'): THEN
yield = 'good'

IF (rain=='very good' && temp=='bad') || (rain=='very good' && temp=='very bad') || (rain=='good' && temp=='bad') || (rain=='good' && temp=='very bad') || (rain=='average' && temp=='very good') || (rain=='average' && temp=='good') || (rain=='average' && temp=='average'): THEN
yield = 'average'

IF (rain=='average' && temp=='bad') || (rain=='average' && temp=='very bad') || (rain=='bad' && temp=='very good') || (rain=='bad' && temp=='good') || (rain=='bad' && temp=='average') || (rain=='bad' && temp=='bad'): THEN
yield = 'bad'

IF (rain=='bad' && temp=='very bad') || (rain=='very bad' && temp=='very good') || (rain=='very bad' && temp=='good') || (rain=='very bad' && temp=='bad') || (rain=='very bad' && temp=='average') || (rain=='very bad' && temp=='bad') || (rain=='very bad' && temp=='very bad'): THEN
yield = 'very bad'

CLASSIFICATION	AVERAGE YEARLY RAIN	AVERAGE SEASON TEMPERATURE
Very Good	62-87 cm	20-25 C
Good	50-62 cm or 87-100cm	15-20 C
Average	25-50 cm or 100-150 cm	12-15 C or 25-33 C
Bad	10-25 cm or 150-200cm	8 -12 C or 33-40 C
Very Bad	All other values	All other values

5. GROUND WATER LEVEL CLASSIFICATION BASED ON KNN MODEL

In this fifth module, India ground water-level data set is taken for ten years old data. Then it is converted into R data frames and zero values in all records are removed. The data from i) 'MONSOON', ii) 'POMKH', iii) 'POMRB' and iv) 'PREMON' column are taken, then average value calculated, and middle value is prepaped and values lesser than middle value are updated as zero and greater than as one. These 2 values are applied for classfactor column. KNN classification is allied than for prediction using class factor column. For test runs, K value is given as 5 and then 6 and model is predicted. Then accuracy value is calculated and view for the KNN model.

IV. CONCLUSION

According to the results, temperature is best predicted by the ARIMA model and the accuracy of predictions made for rainfall by ARMA model is also good compred with neural network. Rainfall, which is an important factor for the prediction of crop yield is difficult to estimate precisely. Climate factors may change due to other remaining variables which may influence the prediction of rainfall. Also, the proposed work makes use of fuzzy logic to estimate crop yield which works on a set range rather than discrete values, therefore, the error in predicted rainfall data does not cause problems as long as the difference between actual and estimated values is not drastic. The model can successfully predict crop yield for a given year when the rainfall and temperature values for the previous years is known. The model can successfully predict ground water level for a given year when the previous years' value is known. In addition, this project classifies the ground water level data set records using KNN to predict the model for future test record data sets. It will be helpful in analyzing the ground water levels in the past and so as to predict the future levels. In future, logistic regression can be applied to further classify the data.

REFERENCES

[1] P. Kumar "Crop yield forecasting by adaptive neuro fuzzy inference system." Mathematical Theory and Modeling 1.3, pp. 1-7, 2011.
[2] Sheikh Hefzul Bari. Md. Manjurul Hussain Shourav. "Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model". December 2015. Civil and Environmental Research. www.iiste.org ISSN 2224-5790 (Paper) ISSN 2225-0514 (Online) Vol.7, No.1, 2015.
[3] S. Veenadhari, B. Misra, and C. D. Singh. "Machine learning approach for forecasting crop yield based on climatic parameters." 2014 International Conference on Computer Communication and Informatics, pp. 1-5. IEEE, 2014.
[4] S. Prabakaran, P.N. Kumar, and P.S.M. Tarun. "Rainfall prediction using modified linear regression.", 2006.
[5] Sundaram, S. Meenakshi, and M. Lakshmi. "Rainfall prediction using seasonal auto regressive integrated moving average model." Computer science 3, no. 4 (2014), pp. 58-60, 2014.
[6] L. A. Zadeh, "Fuzzy sets," Information and Control, pp. 338-353, 1965.
[7]Ehret, D. L., B. D. Hill, D. A. Raworth, and B. Estergaard. Artificial neural network modeling to predict cuticle cracking in green house peppers and tomatoes. Computers and Electronics in Agric. 61(2): 108-116. 2008.
[8]Pucheta, J., D. Patiño, R. Fullana, C.. Neurodynamic programming-based optimal control for crop growth in precision agriculture. In Proc. 16th IFAC World Congress. Available at: www.nt.ntnu.no/users/skoge/prost/ proceedings/ ifac2005/ Papers/ Paper5047.html. Schugurensky, and B.Kuchen.2005.
[9] HortiMax. 2008. Village Farms, USA, reports record tomato production. Rancho Santa Margarita, Cal.: HortiMax Growing Solutions. Schugurensky, and B.Kuchen.2005.
[10] Ehrlich, H., M. Kühne, and J. Jäkel. Development of a fuzzy control system for greenhouses. Acta Hort. 406: 463-470. 1996.
[11]. Abdullah, A., Brobst, S, Pervaiz.I., Umer M.,and A.Nisar.2004. Learning dynamics of pesticide abuse through data mining. Proceedings of Australian Workshop on Data mining and Web Intelligence, New Zealand, January.
[12] Kiran Mai,C., Murali Krishna, I.V., and A.VenugopalReddy,2006.Data Mining of Geospatial Database for Agriculture Related Application. Proceedings of Map India. New Delhi.
((http://www.gisdevelopment.net/proceedings/mapindia/2006/agriculture/mi06agri_12 4.htm).
[13] Jorquera H, Perez R, Cipriano A, Acuna G(2001). Short term forecasting of air pollution episodes. In. Zannetti P (eds) Environmental Modeling 4. WITPress, UK.
[14] Rajagopalan B. Lall U (1999) A k- nearest- neighbor daily precipitation and other weather variables. WatResResearch35(10) :3089 – 3101.
[15] Tripathi S, Srinivas VV, Nanjudiah RS (2006).Down scaling of precipitation for climate change scenarios: a support vector machine approach, J. Hydrology 330-337.
[16] Zelu Zia (2009). An Expert System Based on Spatial Data Mining used Decision Tree for Agriculture Land Grading. Second International Conference on Intelligent Computation Technology and Automation.Oct10-11,China
[17] Jun Wu, Anastasiya Olesnikova, Chi- Hwa Song, Won Don Lee (2009).The Development and Application of Decision Tree for Agriculture Data. IITSI :16-20.
[18] Dhanya,C.T. and D. Nagesh Kumar, 2009. Data mining for evolution of association rules for droughts and floods in India using climate inputs. J. of Geo. Phy.Res.114:1-15.