European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 08, 2020

A Noval Deep Learning Approach For Semantic Information Extraction From Medicinal Crops

Sunil Kumar[1], Hanumat Sastry G[2], Venkatadri Marriboyina[3],Dinesh Goyal[4], Madhushi Verma[5]
[1,2] School of Computer Science, University of Petroleum & EnergyStudies, Dehradun, India
[3]Amity School of Engineering &Technology, Amity University, Gwalior , India
[4]Poornima Institute of Engineering & Technology,Jaipur, India
[5]Department of Computer Science Engineering,Bennett University, Greater Noida, India
(e-mail: skumar@ddn.upes.ac.in*,hsastry@ddn.upes.ac.in, venkatadri.mr@gmail.com, dinesh.goyal@poornima.org, madhushi.verma@bennett.edu.in)

Abstract— **To extract significant information from large amount of data is an essential task of natural language processing. Every Information Extraction technique design different rules for tuning the domain based raw data to extract semantic information. Huge amount of unstructured data in agricultural domain increases the complexity of information extraction techniques. The paper presented an algorithm of semantic information extraction from a text article on health benefits from medicinal plants. The proposed algorithm apply deep leaning techniques to extract semantic (relational) information from medicinal crops corpus. The proposed algorithm tuned and tested on agricultural and weather data collected from DACFW, Government of India. The experimental results stated that the proposed deep learning based method achieved nearly 95% prediction accuracy. Proposed method also compared with existing techniques like Self-Organizing Map (SOM) and Ensemble Neural Network (ENN).**

Keywords—Natural language processing, agriculture, ENN, Deep learning, Information Extraction.

I. Introduction

The information extraction from agricultural sector plays crucial role in Indian economy. Agricultural domain mainly includes soil sampling, field mapping, farm planning, crop scouting, tractor guidance, yield mapping and variable rate applications etc.[1]. This sector also includes various other categories likes Herbs, Shrubs, Trees, Climbers and Creepers plants etc[2]. Plant have an integral part of human life because it provides us medicine, oxygen, wood, fuel, food etc. [3]. Important information of flora permits to improvise in agricultural productivity, ecosystem balance, biodiversity protection, planning and minimize the effects of change climate. Currently, the information extraction techniques considered as a promising solution to extract various entities like named entity, objects, semantics and other meaningful information from medicinal plants

Bonnet et al. [4], experimental result for identification of plants are not convincing as compared to the best existing techniques but as far as plant-taxonomy concerned, the authors show results that are far exceeds those of beginners or amateurs agricultural scientists. Based on this automatic plant identification techniques, various application have been designed and deployed that are widely used as Pl@ntnet, Leafsnap, MOSIR and VnMed[5, 6, 7, 8]. The said systems designed for a particular region species like Pl@ntnet works for the France plants where Leafsnapfocus on US and Canadian plants.

S.Verma et.al. [9], applied Rule Based Open Information Extraction in the field of medical data from tweets during mass emergency and obtained the accuracy of 75%. S. Vieweg et.al. [10], applied framework for Information Extraction on data collected from twitter on natural hazards emergency. They obtained the accuracy of 78%.

Agricultural Named Entity Extraction (AGNER) has gained much importance in the field of Agriculture as well. In early 80's, Food and Agriculture Organization of UN and European Countries brought AGROVOC concept server and Agropedia[11]. It was initially developed in English Language but later on due to its popularity, it was further translated into other four languages: Chinese, Spanish, Arabic and French. It was based on how controlled vocabularies are of limited semantics and how they can be improved for IE by doing reengineering. Data from soil and weather are used for IE for type of crops. Researchers have word on the data of land use, with respect to different regions to extract the information about suitable crops for that region.

O. Medeleyan et.al. [12] proposed a new algorithm for extracting index terms from agriculture related documents and obtained the accuracy of 86%. Information Extraction in the field of Agriculture plays a very important role because the agriculture production depends a lot on various factors like weather, temperature, soil etc. So, gathering information from the aforementioned data and then performing event based for analysing the crops is a challenging task.

William H. etal. [13] demonstrate the potential of Compound Specific Stable Isotope (CSSI) for soil resource management and protection of water resources for crop-specific sediment. They worked on different crop regimes suited for different sediments. Ontology based identification of diseases in crops is presented in [14]. A lot of work has been done in extreme weather events (EWE) for agriculture sector.

From last two decades, several studies and experiments have conducted to explore the research on information extraction. The experimental outcomes based on different statistical and machine learning techniques provide expected results of IE in various application areas. However, when the transformation of datasets to corpus happens then researchers decided to change their focus from simple machine learning algorithms to deep learning based algorithms to get the semantic information extraction.

Paper organization is as follows: section 1 represented the introduction and literature review part. Section 2 described the geographical information about the research area. Section 3 discussed various tools and techniques used for designing the proposed framework that are preprocessing methods, Long Short-Term Memory (LSTM), Adam optimizer along with various post-processing tools. In section 4, proposed deep learning based IE algorithm described. Section 5 discussed the results and discussion related to the proposed algorithms and last section is the conclusion of the present research work.

II. STUDY AREA

After completed a survey on different zones (statewide/zone wide) in India, authors selected the Uttarakhand state as an area of research for present investigation. The Uttarakhand state lies between Lat. 30.0668° N and Long 79.0193°E with 53,483 square km geographical area[15]. Figure 1 shows the geographical region takenas the study area for current research work. The entire area divided into 13 districtsregion takes an input study. Datafrom various data sources like the https://data.gov.in/, Indian MetrologicalDepartment (IMD) were collected and created a database.

Figure 1 Study Area of Proposed Method[16]

III. TOOLS & TECHNIQUES USED

This section divided into three sub-parts preprocessing, long short-term memory and adam optimizer. The detailed explanation of these three sub-parts are mentioned below

A. Preprocessing tools

Input agricultural corpus collected from various sources like IMD Dehradun, KrishiVigyan Kendra, Dehradun and Open Government Data (OGD) platform of India (https://data.gov.in/).

Because input data is unstructured in nature so preprocessing techniques have to apply on it. Some data preprocessing issues as data ambiguity, data imbalanced, blank values etc. are need to be taken care. Data ambiguity generally arises while handling the large amount of data (corpus). WSD is a solution of ambiguous words arises due to distinct meaning words in different context [8]. There are two types of WSD i.e. knowledge based WSD and Corpus based WSD

1) Knowledge based WSD

While working with huge amount of lexical token like thesauri, dictionaries and corpora, knowledge based WSD became widely focused. By using the knowledgebase from corpora, it mainly seek to ignore training based on large amounts data [9]. Generally, these WSD techniques use existing structured lexical knowledge base resources different from the following

- The used lexical resource like monolingual or/and bilingual machine-readable dictionaries (MRD), thesauri, etc.
- The information mentioned in lexical resource.
- The property used to find out the relation between words and senses.

Knowledge based WSD techniques recognized as ready-to-use tools for all words because this technique do not need sense-annotated data [10].

2) Corpus Based WSD

Corpus based learning also called supervised learning in the area of NLP. The training of ML algorithms or statistical classification techniques prompted by using the semantically annotated corpus. Trained modules are enough cable to choose word sense from desired contexts. Commonly WordNet tools applied for manually tagging by using semantic class (from specific lexical semantic recourse) to corpora. Therefore, it requires maximum human intervention for training purpose [8].

3) Tools for WSD

There are several tools and web links also that can be directly use to automate the process of WSD. Following mentioned tools commonly used in research areas of NLP or computer vision. For proposed research work, NLTK preprocessing tool applied on input corpus.

a) Resource Description Framework (RDF)

RDF data model behaves similarly as ER or class diagram (classical conceptual modeling approaches) [11]. It works specially for web resources for making statements in triples expressions (subject–predicate–object). The subject, predicate and object denote the resource, traits or aspects of the resource and expresses the properties of subject respectively. RDF also uses another approach i.e entity–attribute–value. Where object (entity) used instead of subject, attribute as predicates and value as object e.g., this ink has color red. In some object-oriented approaches: entity is ink, attribute is color and value is red. As RDF and OWL demonstrate, one can build additional ontology languages upon RDF.

b) Ontology Language (OWL):

OWL (semantic web language) used to demonstrate the rich and complex information about an entity, group of entities and relation between entities. The knowledge represented in OWL can exploit by computer program, so it also called computational logic-based language. OWL documents (Ontologies) can publish in the Web and referred from (refer to) other OWL like Resource Description Framework (RDF), Resource Description Framework Schema (RDFS) and SPARQL Protocol and RDF Query Language (SPARQL), OWL is also the part of W3C's Semantic Web technology [12, 13].

c) DARPA Agent Markup Language (DAML)

Like RDF and OWL, DARPA also used in semantic web. It is a markup language based on RDF. It used to define the sets of facts for making an ontology. DAPRA had its roots in three main languages - DARPA Agent Markup Language (DAML), OIL (Ontology Inference Layer) and Simple HTML Ontology Extensions (SHOE) [14].

d) Natural Language Toolkit (NLTK)

NTLK library is an open source tool developed by Princeton University. The core functions of NLTK tool are provides training data sets, taggers, stemmers, Wordnet corpus, various tokenizers and lemmatizers [15].

B. Long Short-Term Memory (LSTM)

The advanced version of Recurrent Neural Network (RNN) known as Long Short-Term Memory (LSTM) in the area of Deep learning techniques. LSTM introduced to resolve the gradient decedent problem of RNN by adding extra memory cell per module. LSTM is a special type of RNN by which remembering information and long-term dependencies can learn by DL model for long periods. LSTM has four layers with a unique communication model as mentioned in following figure

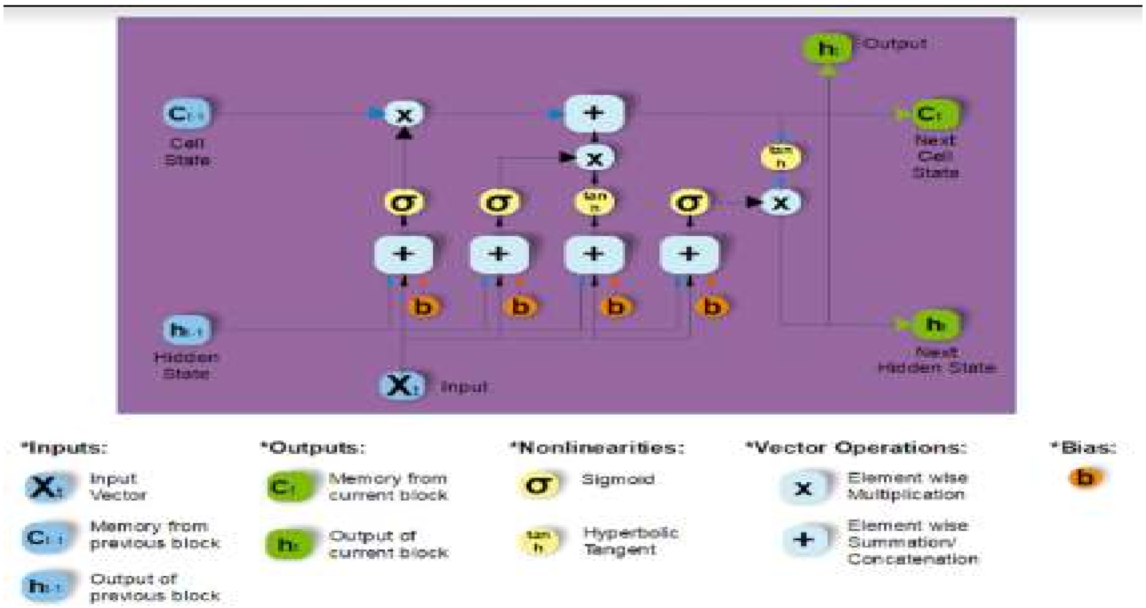

Figure 2 Structure of Long Short-Term Memory (LSTM) Deep Network [17]

In Figure 2, LSTM uses the memory cell (gates) to handle the memorizing process, so it can use to design for preventing long-term dependency problem. While constructing the LSTM, initially step is to identify unessential information and deleted from the network using memory gate. The sigmoid function uses to identify and exclude data from network. This function takes current input $X_t$at time t and output of previous LSTM h_{t-1} at time t-1. The sigmoid function also determines that which portion from the last output should omit from network. This function is processed by the forget gate (f_t).

As per each number in the C_{t-1} (cell state), value of vector f_t is ranging from 0 to 1,

$$f_t = \sigma\ (W_f[h_{t-1}, X_t] + \ b_f \qquad \text{Eq (1)}$$

In the forget gate, weight and bias are represented by W_f and b_f respectively. To decide, store and update the cell state from input X_t, following steps used into two parts that are sigmoid layer and tanh layer. Based on 0 or 1, sigmoid layer will decide that the new information should update or ignore and tanh function assign the value (-1 to 1) and decide the level of importance

After multiplication of these two values, LSTM will update the state of new cell. This updated memory added to the previous memory C_{t-1} resulting new C_t.

$$i_t = \sigma\ (W_i[h_{t-1}, X_t] + \ b_i \qquad \text{Eq (2)}$$

$$N_t = \tanh(W_n[h_{t-1}, X_t] + \ b_n \qquad \text{Eq (3)}$$

$$C_t = [C_{t-1}f_t] + \ N_t i_t \qquad \text{Eq (4)}$$

Here, at time t-1 and t cell states are Ct-1 and Ct, whereas weight matrices and bias of the cell state are denoted by W and b respectively. In last step, the h_t (output values) is based on the O_t (output of cell state). Firstly sigmoid layer function selects that which cell state part make it to the output, then O_t (sigmoid gate output) multiplied by updated Ct (Cell state) values produced by the tanh layer i.e. ranging from -1 and 1.

$$O_t = \sigma\ (W_o[h_{t-1}, X_t] + \ b_o \qquad \text{Eq (5)}$$

$$h_t = O_t \tanh(\ C_t\ ) \qquad \text{Eq (6)}$$

Here, weight matrices and bias of output gate are denoted by Wo and bo respectively.

C. Adam optimizer

The modification of SGD (stochastic gradient descent) called Adam (adaptive moment estimation)optimizer, which broadly adopted by deep learning techniques especially in the area of NLP and computer vision [18, 19]. The SGD maintains common alpha (linear learning rate) value for updating all weights and alpha does not change (or update) during the complete training process. Whereas Adam maintain the specific adaptive learning rate for the individual parameter of first and second moment of gradient. Adam is combination of AdaGrad and RMSProp.

AdaGrad (Adaptive Gradient Algorithm) Maintain the individual parameter-learning rate to upgrade the performance of sparse gradient problems [20, 21] .

RMSProp (Root Mean Square Propagation) Similar to AdaGrad,maintains the individual – parameter-learning rates i.e. average weight of recent gradient magnitudes. It is well suited to noisy online and non-stationary problem.

Like Adadelta and RMSprop, Adam also used to store an exponentially decaying average of previous gradient $m_t$and squared gradient (v_t). Getting gradients with respect to the stochastic object at t time step

$$g_t \leftarrow \nabla_\theta f_t(\theta_{t-1}) \quad \text{Eq (7)}$$

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \quad \text{Eq (8)}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \quad \text{Eq (10)}$$

$m_t$andv_t used to evaluate the first and second gradient moment i.e. mean and uncentered variance. The authors of Adam optimizer observes that m_t and $v_t$are moving towards biasing towards zero because they are initialized 0's vectors. This biasing towards zero noticed especially during the starting time steps and small decay rates (i.e. β_1 and β_2 close to 1). The biases offset defined by computing bias-corrected 1st and 2nd moment estimates:

$$m_t^{new} = \frac{m_t}{1-\beta_1^t} \quad \text{Eq (11)}$$

$$v_t^{new} = \frac{v_t}{1-\beta_2^t} \quad \text{Eq (12)}$$

The updated parameters in Adam with combination of Adadelta and RMSprop optimizer:

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{v^{new}+\epsilon}} m_t^{new} \quad \text{Eq(13)}$$

IV. PROPOSED ALGORITHMS

Pseudo Code: Long short-Term Memory with Adam Optimizer

1. **Class LSTM-RNN [_cropdata, _weights, _biases]**
2. Categorize the input data by state wise
3. Crop Data processed by three medicinal plant of Uttarakhand state
4. $D = \{x_{i,n}, y_n | i \in f \ and \ n \in N\}$ //Database with N instances and f features
5. $x'_{i,n} = \frac{x_{i,n} - min(x_i)}{max(x_i) - min(x_i)}(nMax - nMin) + nMin$ // Min-Max Normalization within the range of -1 to 1, or 0 to 1
6. X_{train}, X_{test}, Y_{train}, Y_{test} ⟵train_test_split(crop_data_processed,test_size=0.3) // spitting the data into training data and testing data for the LSTM-RNN model
7. **def**model**:**

While θ_t is not converged **do**

8. $m_0, v_0 \leftarrow 0{,}0$ (1st and 2nd moment moving initialize)
9. $\rho_\infty \leftarrow 2/(1-\beta_2) - 1$
10. while $t = \{1, \ldots, T\}$ do
    a. $g_t \leftarrow \Delta_\theta f_t(\theta_{t-1})$

b. $v_t \leftarrow \beta_2 v_{t-1} + (1-\beta_2) g_t^2$ (Update exponential moving 2nd moment)
c. $m_t \leftarrow \beta_1 m_{t-1} + (1-\beta_1) g_t$ (Update exponential moving 1st moment)
d. $\widehat{m_t} \leftarrow m_t/(1-\beta_1^t)$ (Compute bias-corrected moving average)
e. $\rho_t \leftarrow \rho_\infty - 2t\beta_2^t/(1-\beta_2^t)$ (Compute the length of the approximated SMA)
f. $\theta_t \leftarrow \theta_{t-1} - \alpha_t \widehat{m_t}$ (Update parameters with un-adapted momentum)

End while

Proposed algorithms have advantages over all the above said algorithms. Following algorithm used the advantages of LSTM techniques over RNN methods. For weight adjustment adam optimizer used.

V. RESULTS AND DISCUSSION

For this experimental scenario simulation, Python Jupiter notebook installed in a computer system with 3.2 GHz i5 processor. The proposed Deep Residual Network (DRN) method extracts the features from unstructured data of agricultural corpus with the help of NER, EEand RE forpredicting the significant medicinal crop productivity in the Uttarakhand region. The proposed DRN method performance evaluated by using parameters like precision, recall, accuracy, and F-Measure.

Out of total datasets, 20% of the validation datasets applied for testing purposes, and the remaining 80% of the datasets utilized for training purposes. For finding better crop production, the main factors like soil, season, geographical area, climate, water, input support facilities, and risk are used. However, the proposed DRN method concentrates only on season factor (to bind with weather data) with the parameters mentioned in Table 1. For proposed research work, authors selected only medicinal crops (3 only) i.e. ginger, garlic and turmeric out of 39 agricultural crops. Table 2 shows the sample data of season based (Kharif and whole year) production of said medicinal crops from three (out of 13) districts that is Dehradun, Haridwar and TehriGarhwal.

Table 1Season Based Sample Data for Medicinal Crops[22]

S.No	District Name	Season	Crop	Area	Production
1	DEHRADUN	Kharif	Ginger	449	8156
2	DEHRADUN	Whole Year	Garlic	41	82
3	DEHRADUN	Whole Year	Ginger	155	3079
4	DEHRADUN	Whole Year	Turmeric	359	747
5	HARIDWAR	Whole Year	Ginger	2	40
6	HARIDWAR	Whole Year	Turmeric	23	42
7	TEHRI GARHWAL	Kharif	Ginger	196	2415
8	TEHRI GARHWAL	Whole Year	Garlic	49	80
9	TEHRI GARHWAL	Whole Year	Ginger	286	3524

10	TEHRI GARHWAL	Whole Year	Turmeric	31	48

Table 1describes the sample data collected for medicinal plant data with their productivity of Uttarakhand region. Based on this sample data, the garlic gives more productivity like 162 kg per hectares (ha), whereas ginger gives 14124 kg and turmeric production 837 kg. Here, gingerconsidered as the most important medicinal plant for Uttarakhand region because of its productivity.

Based on the weather data, the medicinal plant being predicted by DRN to the agriculturists for better productivity in fore-coming seasons (i.e., rainfall, winter, and summer). Table 2 shows the monthly weather data from 2015 to 2019 based on five parameters i.e. Total Monthly Rainfall (TMR) in MM, Mean Monthly Maximum (MMMAX) temperaturein °C, Mean Monthly Minimum(MMMAX)temperature in °C,VaporPressure (VP) and Relative Humidity (R.H).

Table 2Sample data for Rainfall, Min temp, Max temp, Vapor pressure and Relative Humidity from 2015-19

YEAR	Parameter	January	February	March	April	May	June	July	August	September	October	November	December
2015	TMR	29.00	23.00	181.40	60.90	10.70	144.90	566.00	654.20	78.00	27.30	3.50	8.80
	MM MAX	19.33	23.38	25.86	30.32	36.36	34.54	30.68	30.50	31.66	30.15	27.08	21.87
	MM MIN	7.05	10.33	13.10	16.95	21.34	23.18	23.67	23.22	21.55	17.23	12.72	7.98
	VP	12.50	13.90	14.80	17.70	17.00	23.10	29.60	30.20	26.90	20.80	16.10	13.50
	R.H.	77.80	61.70	55.10	52.20	35.70	51.80	81.30	82.20	69.60	64.40	67.30	70.10
2016	TMR	0.00	22.20	32.00	7.80	45.30	187.40	549.10	412.90	222.60	45.90	0.00	0.00
	MM MAX	21.40	24.53	29.50	35.13	36.45	34.02	30.69	32.23	32.03	31.03	27.70	25.04
	MM MIN	7.00	9.93	14.10	18.04	21.52	23.71	23.73	23.80	22.56	17.40	10.74	8.46
	VP	11.60	12.70	14.30	18.80	21.20	29.60	30.80	30.30	29.40	23.70	18.30	15.00
	R.H.	64.00	55.60	44.60	38.90	40.90	66.20	79.90	80.10	79.70	67.70	69.50	67.90

2017	TMR	42.70	9.40	19.20	52.20	116.20	456.50	515.00	543.50	404.40	0.00	0.20	19.30
	MM MAX	21.33	25.43	28.59	34.31	35.80	33.87	31.96	31.62	31.50	31.69	26.70	23.71
	MM MIN	7.85	9.95	12.38	18.54	21.16	22.64	24.17	24.03	22.23	17.16	11.30	9.01
	VP	13.80	13.70	12.70	17.10	20.00	25.90	30.70	30.80	29.50	22.50	19.20	17.20
	R.H.	67.90	55.30	40.20	36.60	43.30	56.20	76.50	82.60	77.30	60.60	89.80	95.80
2018	TMR	25.20	7.90	14.20	24.60	17.30	206.50	663.00	958.10	285.40	6.10	7.40	2.00
	MM MAX	22.08	25.30	31.30	33.37	36.60	34.80	32.26	30.52	31.35	30.70	26.55	22.51
	MM MIN	6.14	9.90	13.70	17.62	20.30	24.00	23.34	23.80	22.20	14.60	11.55	6.39
	VP	11.90	13.00	14.50	15.60	17.10	27.40	31.60	31.80	28.40	21.10	18.40	12.70
	R.H.	62.20	52.90	40.00	36.00	35.80	57.30	80.10	86.40	77.40	63.00	72.40	66.40
2019	TMR	56.70	66.20	39.70	29.00	12.50	402.40	402.40	395.60	490.20	22.30	8.70	28.60
	MM MAX	21.08	22.10	27.10	33.40	36.50	32.00	30.94	32.37	29.83	30.14	26.80	21.20
	MM MIN	6.56	9.56	11.60	18.00	20.30	23.98	23.98	24.23	23.18	17.04	12.70	7.40
	VP	11.80	13.90	14.80	18.60	17.80	25.10	31.40	32.90	30.40	24.80	17.70	13.30
	R.H.	64.30	66.90	49.80	42.60	31.50	49.10	79.50	81.40	80.30	64.40	67.00	77.60

In the Uttarakhand region, medicinal plantslike ginger, garlic and turmeric are mostly sown at low, medium, and high rainfall. Table 3 shows the sample data of last five years for two seasons (Kharif and whole year) of above said medicinal crops.

Table 3Season based Uttarakhand Region Medicinal Crops

YEAR	Parameter	Whole Year	Kharif	Ginger	Garlic	Turmeric
2015	TMR	148.98	331.38	129.20	1834.50	77.20
	MMMAX	28.48	30.75			
	MMMIN	16.53	21.42			

	VP	19.68	26.88			
	R.H.	64.10	74.38			
2016	TMR	127.10	307.63	323.81	1742.50	1103.70
	MMMAX	29.98	31.49			
	MMMIN	16.75	21.87			
	VP	21.31	28.55			
	R.H.	62.92	76.85			
2017	TMR	181.55	365.73	331.50	2219.00	524.60
	MMMAX	29.71	31.69			
	MMMIN	16.70	21.90			
	VP	21.09	28.38			
	R.H.	65.18	74.25			
2018	TMR	184.81	478.15	157.90	1554.50	131.70
	MMMAX	29.78	31.21			
	MMMIN	16.13	20.98			
	VP	20.29	28.23			
	R.H.	60.83	76.73			
2019	TMR	162.86	327.63	161.40	1537.20	174.00
	MMMAX	28.62	30.82			
	MMMIN	16.55	22.11			
	VP	21.04	29.88			
	R.H.	62.87	76.40			

The Table IV provides the value for the accuracy, F-Measure, sensitivity, specificity for major crops of Uttarakhand Region, and the graphical representation for these parameters are describe in Figure2.

Table 4Performance of Proposed DRN method

Plant	Accuracy (%)	F-Measure (%)	Specificity (%)	Sensitivity (%)
Ginger	95.55	93.37	96.65	94.25
Garlic	92.45	93.35	97.12	94.54
Turmeric	93.32	92.36	94.0	93.43

The following figure represent the bar representation of Table 4, blue, orange and gray colored bar represent the accuracy, F-Measure, sensitivity and sensitivity of ginger, garlic and turmeric plants

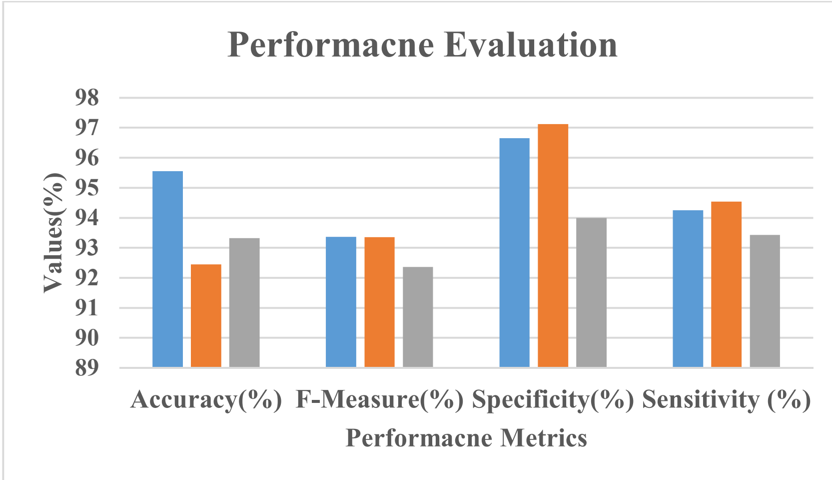

Figure 3*Performance Results of Proposed DRN Method*

From the above results, the results proved that the proposed DRN method achieved better results in performance measures such as accuracy, sensitivity, specificity, and F-Measure for major crops of the Uttarakhand region.

A. Comparative Analysis

The forecasted parameter's values obtained can use in water resource planning, hydrological model study, climate change study, and agricultural sector for various purposes. Table 5 demonstrates the performance criteria of different forecasting methods, particularly on said parameters. The proposed DRN is compared with existing techniques such as Deep Learning based Weighted SOM, and Ensemble-Neural Network (ENN) were evaluated in the combinations of testing and training percentage like 80% training and 20% testing of collected data. P. Mohan and K. K. Patil, [22] proposed a Weighted-SOM for dimensionality reduction to predict suitable crop for suitable seasons. By employing DNN, the own dimensionality reduced data is used for classification purposes. The drawback of this weighted SOM was presented a poor performance while the method used a large set of data for daily prediction of rainfall data.

Table 5Comparison of Existing with Proposed Method

Authors	Methodology	Accuracy	Sensitivity	Specificity
P. Mohan, and K. K. Patil, [22]	Weighted SOM	78.98%	83.05%	81.45%
H. Y. Kung, et al., [23]	ENN	65.12%	90.0%	90.6%
Proposed Method	**DRN**	**95.55%**	**94.25%**	**96.65%**

H. Y. Kung, et al., [23] presented the ENN method to predict the agricultural production activities. Especially in agriculture forecast analysis, this study employed stepwise regression analysis and ENN for the design guidelines. The ENN method consumed a lot of time to process because it randomly created a plurality of networks for analysis. The existing methods mainly focused on predicting the rainfall directly without extracting the useful information from the unstructured data. Hence, the performance of these methods provides poor performance in terms of less accuracy, nearly 79% only. The proposed DRN method focused on extracting the important features for predicting the rainfall amount by using NER, RE, and EE methods. The features such as cloud cover, average temperature, rainfall, vapor pressure, humidity, crop selection, etc. extracted from the unstructured data to predict the significant crop productivity in Uttarakhand state. From the above results, the proposed DRN method achieved nearly 95% accuracy when compared with ENN and weighted SOM. The ENN method achieved less accuracy when compared with SOM because of consuming more time, but delivered better performance in both sensitivity and specificity.

VI. CONCLUSION

Before ML many other research themes were in use to detect symptoms by sensing including work of Kumar et al. [24-27]. Now, approaches using IoT [28-30], and ML[31, 32].Now , a subset of ML is popular in research. The proposed experiment's scope was to enhance the significant plant productivity of garlic, ginger and turmeric plants in high rainfall areas by investigating the accurate rainfall and right quantity of crop prediction. In this scenario, a DRN method was implemented in order to predict the suitable major crop for the season in Uttarakhand region. The features are extracted with the help of fundamental tasks such as AGNER and AGEEto increase the medicinal crop productivity based on TMR, MMMAX, MMMAX, VP and R.H parameter. The evaluated output of the proposed method describes better performance than existing methods. The proposed approach attained 95.55% accuracy by properly utilizing a DRN algorithm.

The advanced scheme delivered an effective performance employing sensitivity, accuracy, specificity, and F-measures than the previous methods related to the other approaches for plant or crop productivity prediction. In the future, for improving agricultural productivity, the quantity of crop prediction may be improved by applying new strategies with several other significant factors such as soil, risk, pest and fertilizer, etc.

VII. REFERENCES:

[1] J. Hatfield, E. Takle, R. Grotjahn, P. Holden, R. C. Izaurralde, T. Mader, E. Marshall and D. Liverman, Agriculture, 2014.

[2] S. Haug, A. Michaels, P. Biber and J. Ostermann, "Plant classification system for crop/weed discrimination without segmentation," in *IEEE winter conference on applications of computer vision*, 2014.

[3] N. K. Behera and G. S. Mahalakshmi, "Medicinal Plant Information Extraction System—A Text Mining-Based Approach," in *Advanced Computing and Intelligent Engineering*, Singapore, Springer, 2019, pp. 215-226.

[4] P. Bonnet, A. Joly, H. Goëau, J. Champ, C. Vignau, J.-F. Molino, D. Barthélémy and N. Boujemaa, "Plant identification: man vs. machine," *Multimedia Tools and Applications,* vol. 75, no. 3, pp. 1647-1665, 2016.

[5] A. Joly, H. Goëau, P. Bonnet, V. Bakić, J. Barbe, S. Selmi and I. Yahiaoui, "Interactive plant identification based on social image data," *Ecological Informatics,* vol. 23, pp. 22-34, 2014.

[6] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez and J. V. Soares, "Leafsnap: A computer vision system for automatic plant species identification," in *European conference on computer vision* , Berlin, Heidelberg, 2012.

[7] K. H. Phyu, A. Kutics and A. Nakagawa, "Self-adaptive feature extraction scheme for mobile image retrieval of flowers," in *18th International Conference on Signal Image Technology and Internet Based Systems*, 2012.

[8] T.-L. Le, D.-T. Tran and V.-N. Hoang, "Fully automatic leaf-based plant identification, application for Vietnamese medicinal plant search," in *15th symposium on information and communication technology*, 2014.

[9] S. V. W. C. L. P. J. M. M. P. A. S. a. K. M. A. S. Verma, "Natural language processing to the rescue? Extracting “situational awareness” tweets during mass emergency," in *5th International AAAI Conference on Weblogs and Social Media (ICWSM 2011),*, Barcelona, 2011.

[10] A. H. K. S. a. L. P. S. Vieweg, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in *28th International Conference on Human Factors in Computing Systems, ACM,*, Newyork, 2010.

[11] M. &. R. S. &. K. J. &. T. P. &. K. A. Sini, "Smart Organization of Agricultural Knowledge The example of the AGROVOC Concept Server and Agropedia. Advances in Knowledge Organization," *Advances in Knowledge Organizations,* vol. 12, pp. 322-326, 2010.

[12] O. a. I. H. W. Medelyan, "Thesaurus-based index term extraction for agricultural documents," *EFITA/WICCA,* pp. 1122-1129, 2005.

[13] W. H. F. K. J. Blake, P. Taylor, M. A. Russell and D. E. ". c.-s. s. Walling, "Tracing crop-specific sediment sources in agricultural catchments," *Geomorphology,* Vols. 139-140, pp. 322-329, 2012.

[14] M.-x. Z. J.-l. W. Bo Jiang, "Ontology-Based Information Extraction of Crop Diseases on Chinese Web Pages," *JOURNAL OF COMPUTERS,* vol. 8, no. 1, 2013.

[15] C. K. Jain, A. Bandyopadhyay and A. Bhadra, "Assessment of ground water quality for drinking purpose, District Nainital, Uttarakhand, India," *Environmental monitoring and*

assessment , vol. 166, no. 1-4, pp. 663-676, 2010.

[16] C. I. Limited, "District Map of Uttarakhand," mapsofindia.com, [Online]. Available: https://www.mapsofindia.com/maps/uttarakhand/uttaranchal.htm. [Accessed 2020 08 22].

[17] A. Graves, "Long short-term memory," in *Supervised sequence labelling with recurrent neural networks*, Berlin, Heidelberg, Springer, 2012, pp. 37-45.

[18] J. Brownlee, "Gentle introduction to the adam optimization algorithm for deep learning," Machine Learning Mastery, 20 August 2020. [Online]. Available: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/. [Accessed 25 Nov 2020].

[19] D. P. Kingma and J. Ba., Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

[20] S. Ruder, An overview of gradient descent optimization algorithms, arXiv preprint arXiv:1609.04747, 2016.

[21] M. Devin, Q. V. Le, M. Z. Mao, M. A. Ranzato, A. Senior, P. Tucker, K. Yang and A. Y. Ng., "Large scale distributed deep networks," in *Advances in neural information processing systems*, 2012, pp. 1223-1231.

[22] O. p. o. India, "District-wise, season-wise crop production statistics," https://data.gov.in/, 13 02 2020. [Online]. Available: https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics?filters%5Bfield_catalog_reference%5D=87631&format=json&offset=0&limit=6&sort%5Bcreated%5D=desc. [Accessed 13 06 2020].

[23] G. S. a. U. C. M. Nair, "Prediction of Monthly Summer Monsoon Rainfall Using Global Climate Models Through Artificial Neural Network Technique,," *Pure and Applied Geophysics,* vol. vol. 175, no. 1, pp. 403-419, 2018..

[24] N. Kumar, A. Agrawal, R. A. Khan, “IoT based Alert Network for Flood Generated Emergencies”, in 3rd IEEE conference on Research in Intelligent and Computing in Engineering (RICE-2018), at University Don Bosko, El Salvador, Central America, Online in Nov. 2018.

[25] N. Kumar,A. K. Pandey, R.C.Tripathi, “A framework to prevent mobile sinks accessing by unauthorized nodes in WSN”, Special issue on MANET, IJCA (USA), pp.-13-17, 2010. ISBN: **0975 – 8887.**

[26] N. Kumar, A. Kumar, D. Chaudhary, “A Novel Approach to Use Nano Sensor in WSN Applications”, in the International Journal IJCA, (USA), Jan 2011. ISBN: **0975 – 8887**

[27] G.Arora, A.Kumar, Versha, N.Kumar, “Swarm Intelligence based QoS optimized routing in WSN”, Test Engineering & Management, Vol.-82, 2020.pp-12880-12885.

[28] Neeraj Kumar; PareshGoyal; GayatriKapil; AlkaAgrawal; Raees A Khan, “Flood Risk Finder for IoT based Mechanism using Fuzzy Logic”, Materials Today: Proceedings, Elsevier, 2020.

[29] N. Kumar, A. Agrawal,R. A. Khan, “METHWORK: An Approach for Ranking in Research Trends with a Case Study for IoET, Recent advances in Computer Science and Communication (formerly Recent Patents on Computer Science), 2019

[30] Kumar, Neeraj, AlkaAgrawal, and R. A. Khan. "Cost estimation of cellularly deployed IoT-enabled network for flood detection." *Iran Journal of Computer Science* , issue 2, no. 1 (2019), Springer Nature: 53-64.

[31] ManojDiwakar, AmrendraTripathi, Kapil Joshi, MinakshiMemoria , Prabhishek Singh, Neerajkumar, “Latest Trends on Heart Disease Prediction using Machine Learning and Image Fusion”, Materials Today: Proceedings,Elsevier, 2020.

[32] ParthWadhwa; Aishwarya; AmrendraTripathi; Prabhishek Singh; ManojDiwakar; Neeraj Kumar, “Predicting the Time Period of Extension of Lockdown due to Increase in Rate of COVID-19 Cases in India using Machine Learning”, Materials Today: Proceedings
[33] Elsevier, 2020.

V. Velvizhi; Satish R Billewar; GauravLondhe; PravinKshirsagar; Neeraj Kumar, “Big
[34] Data for Time Series and Trend Analysis of Poly Waste Management in India”, Materials Today: Proceedings, Elsevier, 2020.

G. Arora, A. K. Maurya,N. Kumar, A. K. Mishra, “Application of big data generated by IoT environment for HealthCare using Voice Recognition”, International journal of research in engineering, IT and Social Sciences, vol.-08, issue-11, November 2018, page. 132-136.