LITERATURE SURVEY

ESTIMATE THE CROP YIELD USING DATA ANALYTICS

Scholars of science have made many contributions to the topic of crop yield and climate change. In order to determine the potential for moving forward with the work as part of our research, we attempt a brief literature assessment of the work completed by other research academics in this section.

A. L. Ismail et al.(2018) developed a system that uses machine learning to forecast how well-prepared a nation is to deal with climate change. South East Asia is the focus of the study. The predictive index is calculated using the following steps: data collection, training, testing, index prediction, index validation, and index visualisation. The study is a preventative strategy to warn the regions and use deep learning to confirm its vulnerable index.

Zhen Nan Liu, et al., (2018), The Standardised Precipitation Index (SPI) and SPEI have been calculated using various machine learning techniques that have been compared in this research. Extreme learning techniques, an online sequential extreme learning machine, and an evolutionary extreme learning machine that is self-adaptive are used after data collection. The three systems can all be successfully used for drought predictions, according to the authors. However, OS-ELM and SADE-ELM outperform ELM in terms of performance.

Chaoyun Zhang, et al., (2018) gave a thorough analysis of the connections between the two domains. Using deep learning techniques, networking applications are briefly studied. The implementation of deep learning into mobile systems is thus effectively facilitated by a number of methods and platforms, which are subsequently discussed. The emphasis of the authors is on the use of deep learning to wireless and mobile networking. The issues and difficulties with deep learning in wireless and mobile networks are exposed in this survey report.

K. G. Liakos et al., (2018) gave a thorough analysis of research on the use of machine learning in the agricultural sector. Crop management, livestock management, water management, and soil management were among the various criteria on which the work was evaluated. ML models have been used for disease diagnosis and crop yield prediction. Without fusing data from other sources, ML-based detection can be derived. According to the author, farm management systems are evolving into true artificially intelligent systems with the end goal of enhancing production. As the most fundamental requirement among all other necessities for survival, the author encourages the use of ML for the benefit of agriculture.

Crane Droesch (2018), has used data on corn yield from the USMidwest, and shown that the approach of using semi-parametric variant of deep neural network, accounting for complex non-linear relationship in high dimensional dataset, the model will outperform both classical statistical methods and fully non-parametric neural networks in predicting yield of years withheld during model training. Authors have developed a novel approach for augmenting parametric statistical model with deep neural networks, they have termed it as semiparametric neural networks. It is used as a crop yield modelling framework, the SNN achieves better out of sample predictive performance than anything else yet published. it uses prior knowledge of functional phenomenon and functional form relating them to the outcome. So the SNN improves statistical efficiency over typical neural networks. They found that combining ML with domain area knowledge from empirical studies improves predictive skills, while altering conclusions about climate change impact to agriculture.

P. Priya et al., (2018) has proposed a random Forest Algorithm for predicting the crop yield of particular area considering various parameters such as rainfall, seasonal crop (Rabi and Kharif) district-wise, temperature (max.), crop production in terms of Kgs/tonnes. Area for doing research was Tamil Nadu. Dataset record were collected from Indian Government over 15years for rice production. They proved in experimental results that prediction analysis done using Random Forest Algorithm – a supervised machine learning algorithm will help farmer to predict the yield of the crop before cultivating onto the agricultural field. This algorithm run efficiently on large databases with high classification accuracy.

N. Zhu, et al. (2018) This article summarizes DL algorithms, considering concepts, constraints, implementation, procedure of training, and sample codes, to aid researchers in agriculture to facilitate with DL techniques quickly. Research on DL applications in agriculture is summarized and analysed, and future opportunities are discussed in this paper, which is expected to help researchers in agriculture to better understand DL algorithms and learn major DL techniques quickly, and further to facilitate data analysis, enhance related research in agriculture, and thus promote DL applications effectively.

S. Rasp, et al. (2018), presents a different perspective to sub-grid parameterizations to a DDA that influences the benefits of high-resolution modelling. Challenges to overcome, but advances in computing capabilities and deep learning in recent years present novel opportunities that are just beginning to be investigated. Authors believe that machine-learning

approaches have huge potential to be explored connection with development of traditional model.

Jinyoung Rhee et al., (2018) has targeted all officials whose main duties include water resources and agricultural management. The final beneficiaries of the output are residents of the area; water users and farmers for whom decision-making can be helped by drought prediction information with finer spatial resolution The models provide spatially distributed detailed drought prediction data of the 6-month Standardized Precipitation Index for the case study area, Fiji. They used Weather Research Forecasting (WRF) model as reference data for overcoming the limitations of non-dense monitoring network. Also they used Performance measures of the mean absolute error as well as classification accuracy. The WRF outputs reflect the topography of the area. Hybrid models showed better performance than simply bias corrected forecasts in most cases. The model based on Extra-Trees trained using the WRF model outputs performed the best in most cases.

S. D. Patil et al., (2017) suggests that according to their results, direct prediction of spectral band information is highly beneficial due to the ability it provides for deriving ecologically relevant products which can be used to analyse land cover change scenarios from multiple perspective. Aim of the authors, is to enhance the use of machine learning based land cover change models to predict the spectral band information of satellite based land cover images. Experimental areas covered by authors is some portion of United States. They used data from two large sites in US to train model RF machine learning model to spectral values from bands. They used the trained model to explore the look of land cover for a climate change scenario. The demonstrative results show that the direct prediction of spectral band information is helpful for deriving ecological products. They have considered this as a major strength of their proposed approach as it has enabled the analysis of land cover change from multiple viewpoints The authors have made a comment through their literature survey, that in 2081-2100projected rise in temperature will be $1.5^{o} – 4.8^{o}$ C than 1986-2005 era. This will impact global landcover. timely and accurate prediction may provide useful solutions. Author has chosen RF model of ML in the study as it ensembles constitutes are comprised of DT models that offers variety of attractive features over other statistical learning techniques. Parameter selection was done based on their judgement of importance of factors. They urge research scholars to continue the work by analysing other predicted variables or modifying the data sources.

Annelie Holzkamper (2017) gives a systematic literature review in modelling for adaptation in planning in agricultural production systems. The author has studied five types of models namely empirical crop model, regional suitability model, Biophysical model, meta model, decision model. According to the author, the key challenge of adaptation plantation is the risk of maladaptation – adaptation that implies negative consequences in long term or in wide context. The five approaches differ in terms of their applicability for decision support in short term and long term adaptation planning. The main value lies in the ability to predict the climate change impact on yield potential at all level. For short term, reactive adaptation responses, statistical and biophysical models are less useful. Authors say adaptive management cycles should be institutionalised, within which adaptation behaviour, consequences of adaption responses and changes in impacts are continuously monitored.

Dr. Pushpa Mohan et al, (2017) has given analysis of the techniques employed and parameters achieved with limitation that every technique and experiment faced. This paper helps to have a crisp view of Regression Analysis, Linear regression by Sellam (2016). Limitations say that it is more complex to predict the optimized number of input parameters.

Evan Racah et al., (2017), The authors have used deep learning for weather prediction and climate change. for this they have used. For calculating the extreme weather projection values, the labels for extreme weather events namely Tropical Depressions (TD) Tropical Cyclones (TC), Extra-Tropical Cyclones (ETC) and Atmospheric Rivers (AR) using TECA (Prabhat et al., 2012) are identified. 3D semi-supervised learning architecture is used. For experimentation, frame-wise reconstruction is done, Detection and localization and feature exploration is done.

Zaki Ahmad Khan, et al., (2017) suggests different machine learning strategies for Wireless Sensor Networks (WSN). It presents a brief idea about supervised and unsupervised learning and its respective types. The author has suggested machine learning solutions for some operational, functional issues such as - query processing and event recognition, Medium Access Control, routing in WSN, object targeting and localization, Clustering and Data collection. Some other challenges highlighted are non-operational and application-specific challenges to address the WSN challenges.

Amir Ghaderi et al., (2017) mainly contributed to obtain forecasts of all nodes of the graph at the same time based on one framework. They studied the results of a case study on recorded time series data from a collection of wind mills in the north-east of the U.S. and shown that the

proposed DL-based forecasting algorithm significantly improves the short-term forecasts compared to a set of widely-used benchmarks models. They used LSTM and RNN for their work

Aized Amin Soofi et al., (2017) Classification is a data mining (machine learning) technique used to predict group membership for data instances. There are several classification techniques that can be used for classification purpose. Researchers use the basic classification techniques. Later usage of some major types of classification method including Bayesian networks, decision tree induction, k-nearest neighbour classifier and Support Vector Machines (SVM) with their strengths, weaknesses, potential applications and issues with their available solution had been applied. Their ultimate goal was to provide a comprehensive review of different classification techniques in machine learning. This provided platform for both academia and new comers in the field of machine learning to further strengthen the basis of classification methods.

Donghyun Lee et al., (2017) describes Artificial Intelligence and deep learning as a promising futuristic concept of technological advancements. Authors used deep learning's recurrent neural network (RNN) model algorithms to predict pro-environmental consumption index based on Google search query data. Advanced research on ANN and RNN development processes is done. 84 different datasets were used by the author for verification of reliability of data by doing repeated experiments. Authors have used the data for experimentation on different human parameters, and a comparative analysis of ANN and RNN is done.

M. Shah et al., (2016) The proposed model provides forecast of the monsoon at a long lead time which supports the government to implement appropriate policies for the economic growth of the country. The monsoon of the central, north-east, north-west, and south-peninsular India regions are predicted with errors of 4.1%, 5.1%, 5.5%, and 6.4%, respectively. The identified predictors show high skill in predicting the regional monsoon having high variability. The proposed model is observed to perform better than most of the prediction models.

Karandeep Kaur (2016), The author has tried to provide a brief overview of various machine learning applications in Indian agriculture, to help farmers advance their work manifolds. Author has described what is Machine learning and its technique such as reinforcement learning, supervised and unsupervised learning. While studying the applications in agriculture author has taken into consideration various parameters such as crop selection, crop yield prediction, weather forecasting, smart irrigation system and crop disease prediction and hence

deciding the minimum support price. Considering all these parameters, the best suited algorithms are suggested respectively with the help of literature survey. Author has concluded his research saying the high accuracy of AI machines is the result of machine learning algorithms. one of the example is sensor based farming system for increased precision. Prescriptive solution for more complex problem in case of large data and field size is yet to be done.

P. Mondal, et al., (2015) recommends use of Enhanced Vegetation Index (EVI) over other remotely sensed vegetation indices as it better adjusts for background soil and canopy reflectance. Authors used around 25 climate variables in their study and finalised the data sets in to 4 sets, Monsoon and winter season for both central and Western India. They concluded both central and western sites showed strong sensitivity day time and night time temperature for both seasons, especially to winter daytime warming. Western site was less sensitive to monsoon precipitation variability, likely due to increased access to groundwater level irrigation. This groundwater irrigation is sensitive to climate variability. Authors suggest that heat tolerant high yield varieties to be added for better crop cover.

Aditya Grover et al. (2015) explores a new way against the traditional predictive model. They call it as the hybrid approach - a combination of partially trained predictive model with Deep Neural Network (DNN) for joining weather-related variables. They have evaluated the methods by experimenting on the real- world meteorological data assuring the results found to be promising. They have collected weather forecast data from the National Oceanic and Atmospheric Administration (NOAA). For preparing logs, Integrated Global Radiosonde Archive (IGRA) is used. The authors have thrown light on the challenges faced and overcome by them e.g first challenge was relationship of the weather parameters as tightly coupled and author suggests to clear the concepts from physics-based tight statistical couplings. some excerpts from the same are pressure and temperature follows natural gas law and relative humidity and temperature follow tight relationship, second challenge was when space and time is considered, the variable dependencies may have long range impacts. Authors have generated hybrid model for weather related spatiotemporal inferences, data-driven kernel function for prediction according to physical law, efficient inference procedure and experimental results. The scope for work still lies in increasing the spatio-temporal dimension, and testing the results of the same in Maharashtra, India.

R. Kumar, et al., (2014), focuses on the impact of climate change on crop productivity. The particulars considered are geographical area, average rainfall, variation in rainfall, annual precipitation, available water resources, utilizable surface and ground water, present water utilization and per capita availability.

R.C. Deo, M. Şahin (2014) presents a drought prediction algorithm for region of eastern Australia. They have devised a novel algorithm called extreme learning machine (ELM). Prediction done by them is on Effective Drought Index (EDI). Variables considered were mean, min and max. air temperatures and rainfall. Performance of ELM was evaluated over ANN models and ELM excelled in terms of mean absolute error, root mean square errors, Coefficients of determination and Willmott's Indices of Agreement.

M. Senapati et al., (2013)

This paper considers the needed adaptation measures including changes needed for mitigation to improve agriculture sector in India. It considers the likely changes that climate change will bring in temperature, precipitation and extreme rainfall, drought, flooding, storms, sea-level rise and environmental health risks and the overall impact on agriculture. The agricultural sector is the major source of employment in India. Climate change has adverse impacts on agriculture, hydropower, forest management and biodiversity. Anticipated impacts on agriculture from climate change and its various aspects have been studied.

A. Belayneh et al., (2013) worked for the Awash river basin of Ethiopia. They chose Standard Precipitation Index (SPI) as the drought index. Artificial neural networks (ANNs), support vector regression (SVR), and coupled wavelet-ANNs, which pre-process input data using wavelet analysis (WA) was used. Using RMSE and R^2, forecasted results were compared. coupled wavelet neural network (WA-ANN) models were the most accurate models for forecasting SPI 3 (3-month SPI) and SPI 6 (6-month SPI) values over lead times of 1 and 3 months in the Awash River Basin in Ethiopia.

Dr. Parag Kulkarni, (2012) highlights on the concepts of supervised and unsupervised learning. Defines what is Machine learning and also explains the learning paradigms. It says that empirical learning method has three different approaches to modelling problems based on observation, data and partial knowledge about problem domains. They are more specific to problem domain. They are generative modelling, discriminative modelling and imitative modelling. Author has also quoted some similarities and differences between reinforcement learning and systematic learning.

K.S. Kavi Kumar (2009) The paper emphasises on agricultural impacts of spatial features that influences the agricultural climate sensitivity. The authors quote that dependent variable, net revenue at farm level and error term has significant positive spatial autocorrelation working on it can improve the accuracy of climate impact studies. The book also refers to the climate change projections for India.

Seneviratne et al., (2008) provides relative material on what are extreme events, compound events, the relevance of feedbacks for extremes. The author discusses on requirements and methods for analysing changes in climate extremes, assessments regarding changes in the climate variables, phenomena, and impacts. Analysis of regional to globally scaled data is done. The extremes are calculated based on the changes observed since 1950 till the 20^{th} century, which is really an effort to be appreciated. This chapter has really helped to understand the reasons of extreme climatic changes.

According to the literature survey done, we propose a novel approach through machine learning with parallel computation to solve the problem of more complex computation to predict the optimized solution considering selective parameters. Further we try to extend and complete the loop by using prescriptive analysis to give complete prescription to the farmer for necessary action and automatize the complete process.