IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ESTIMATION OF DROUGHT PREDICTION ALONG WITH WATER SALINITY OF LAND USING DEEP LEARNING MODELS

[1]Dr.J.Maria Arockia Dass, [2]B.Mamatha, [3]C.Om Shanti, [4]J.G.Mano Bala, [5]P.Lakshmi
[1]Associate Professor, [2-6] UG Scholar
[1-6] Department of Computer Science and Engineering,
[1-6] Siddharth Institute of Engineering & Technology, Puttur, India

Abstract: Most people's livelihoods depend directly or indirectly on the agricultural sector. In recent decades, it has been discovered that soil fertility is linked to sustainable agriculture, and that preserving soil fertility can increase present crop output levels. In agriculture, a decision support system is necessary due to a variety of aspects, including the desired yield. Water use may be optimized to a significant extent by monitoring soil moisture as the water table drops day by day. Because moisture content is helpful to agricultural development, the steps involved in crop production can be improved when we can accurately anticipate moisture levels. We need to know the moisture content of any area or place so that we can estimate how likely it is to face a famine. CNN model has been applied for classification of real time image data obtained from satellite resources. Predicting droughts with a lengthy lead time and in a changing environment due to both natural and manmade causes still presents difficulties. Hence the Deep Learning models like CNN and VGG16 transfer models are implemented and compared for their performances.

Index Terms - Deep Learning, Fuzzy logic, RF Algorithm, CNN Classifier, VGG16 Model

I.INTRODUCTION

Extreme human and environmental harm may result from drought, which is well recognised as a frequent and recurrent meteorological event. The frequency and severity of this least-understood natural occurrence varies greatly owing to regular global climate change[1, making it exceedingly difficult to identify. Although droughts are often characterised by a lack of precipitation, they can be broken down into four distinct types based on the mechanisms and types of ecosystems they affect. These sorts are the meteorological dry spell, the hydrological dry season, the horticultural dry spell, and the financial dry season. Soil dampness lack, known as horticultural dry spell, is the consequence of a flowing impact of meteorological dryness on groundwater, hydrological dry season, and ultimately rural dry season. A socioeconomic drought results from this mismatch between supply and demand for water.

It is pivotal to comprehend the turn of events and conceivable impact of dry spell to appropriately plan strategy, since the numerous dry season classes have shifted highlights and estimating procedures. About every three years, a major drought strikes South Asia, bringing with it water shortages and food poverty that have a lasting effect on the region's developing economy. The drought in India has caused a serious water scarcity impacting 33 million people, while in Pakistan it has a negative impact on agricultural development of 2.6%. Bangladesh's agriculture takes a major hit from the drought, with an average 40% reduction in output and the risk of starvation for 53% of the population. Drought has had a devastating effect on Sri Lanka, where over 0.2 million people are living in dire poverty. Consequently, there are significant functional ramifications for agrarian water the executives coming from dry season observing utilizing a precise and close constant distinguishing proof framework.

Lately, endeavors have been made to examine the chance of involving computer based intelligence calculations for mimicking water quality. These algorithms probe the intricate webs of correlation between input and output data, and then construct models that most accurately capture the underlying connections. Artificial intelligence (AI) models enjoy a few upper hands over additional customary physical and measurable models, for example, the information expected for the simulated intelligence models can be gathered with little exertion, at times through remote detecting stages; simulated intelligence models are less delicate than traditional strategies to missing information; man-made intelligence model designs are adaptable, non-straight, and strong; and simulated intelligence models can deal with monstrous measures of information and information at different scales.

The following method was used to create the SDAP model. We initially determined four surface parameters that impact the maintenance and decline of the SMI during non-rainfall seasons. As a result, we chose input variables that matched this class of surface characteristics. The RF algorithm's use of input variables and, three months later, SMI for training agricultural drought allowed us to build a drought function. Moreover, we have determined the relative weight of each input variable in terms of its impact on drought training (regression) (features). The f(x) drought function used to forecast the SMI in the research region. Using the real SMI value, we checked the drought function's training and prediction performance.

RELATION TO THE EFFECTS OF DROUGHT

Drought is a natural phenomenon, but trees and other plants have developed strategies to survive it. Certain vegetation (like grasses) may stop growing or even turn brown in an effort to save water. Trees may reduce water loss via transpiration by dropping their leaves before it's fully summer. There are several plant species that have evolved to survive for extended periods of time in dry conditions. The yucca, for example, has extensive roots that can efficiently find water even in dry conditions. Cacti feature hairy, spiny, prickly spines, spikes, or leaves on all four of its stem segments, which reduces water loss via evaporation. Dehydration has no effect on mosses. Trees of the juniper family are capable of performing self-pruning by directing water exclusively to the essential branches.

Plants that need more water to thrive are limited in their expansion. Their seeds may wait for years before sprouting, even if the earth is parched. Economic and social disruptions of substantial magnitude may also be brought on by drought. Loss of crops, falling land values, and increased unemployment caused by output decreases may all be caused by a lack of rain. Significant economic and societal issues may also be brought on by drought. Crop failure, falling land values, and increased unemployment may all be caused by a lack of rainfall. Problems with water supply might arise if rivers and lakes lose water. It's possible that they might lead to more societal issues.

Lack of water, poor nutrition, and starvation are only a few of the health issues that contribute to these larger concerns. Conflicts over scarce resources like water and food have also contributed to mass exodus from drought-affected regions. While drought is an inevitable element of the weather cycle, it is possible to mitigate the consequences of drought by human action. Others argue that drought-proneness has increased as a result of current farming techniques. Although modern irrigation methods have greatly expanded the quantity of arable land available, they have also increased farmers' reliance on this finite resource.

By alternating which areas are used for crop cultivation and animal grazing throughout the year, traditional farming methods provide land a chance to "relax" between uses. Overpopulation and a lack of farmland have become major problems in many parts of the globe, meaning that there is frequently not enough arable land to support sustainable practises. Soil compacted and unable to contain water as a result of overuse from farming or grazing. Erosion becomes more of a problem when soil moisture decreases. It's possible for formerly productive terrain to go through desertification as a result of this. This extended drought, exacerbated by agricultural methods that cause overgrazing, contributed to the desertification of the Sahel in North Africa. Deforestation and other human activities have been linked to worsening drought in Kenya. Trees are important for absorbing rainwater and preventing soil erosion. Ten million people in the nation were affected by drought in 2009, when reports surfaced that a fourth of a designated forest reserve had been removed for farming and logging.

Several machine learning (ML) techniques have been investigated in water quality modelling over the last couple of decades. Artificial neural networks (ANN) such backpropagation neural networks (BPNN), multilayer perceptrons (MLP), and feed-forward neural networks (FFNN) have been utilized to measure, model, and figure saltiness in soils, groundwater, and seas. Additionally, estimating river salinity has only been addressed by a few number of research. Salinity of the Apalachicola River in Florida was predicted using an ANN model developed by Huang and Foo. For EC forecasting in the Johor River in Malaysia, Najah et al. also employed artificial neural networks. The most popular ML approach, artificial neural networks (ANN), has weak prediction capabilities, particularly when utilising small datasets or when validation data fall outside the range of the training data. In order to address this shortcoming, an adaptive neuro-fuzzy inference system (ANFIS) was suggested, which combines ANN with fuzzy logic (FL).

Nevertheless, ANFIS was not able to correctly estimate the weight of the membership function, despite being successful and having greater prediction accuracy than both ANN and FL. Automated weight determination using bioinspired (or metaheuristic) algorithms has been developed as a solution. When metaheuristic algorithms are used with ANFIS, it has been observed that the resulting hybrid algorithms offer more predictive abilities than either one alone. As opposed to algorithms with hidden layers (such as artificial neural network (ANN) and artificial neural network fuzzy inference system (ANFIS) and support vector machine (SVM) algorithms), decision tree techniques are superior predictors. The predictive abilities of ANFIS and two other models, the extreme earning machine (ELM) and the wavelet-extreme learning machine hybrid (WA-ELM), were examined by Barzegar et al. for EC in the Aji-Chay River watershed. They concluded that WA-ELM was the most effective of the available models. Although ELM is widely used and well-known for its rapid training times, it experiences the restriction of being not able to encode more than one layer of reflection.

II.RELATED WORKS

A brief overview of the contents of this article follows. In the section labelled "Related Works," we compare and contrast previous research on DDOS attacks with other studies in the same subject.

One of the most difficult and devastating natural calamities, drought occurs often in almost every climate (Zarch et al., 2015). Droughts may be broken down into four categories based on their causes: weather-related, agricultural, hydrological, and societal/economic (Wilhite, 2005). These several droughts are connected. In most nations, agricultural drought is seen as the greatest threat to food security, economic growth, and social stability, making it crucial to conduct an assessment of the situation (Hazaymeh and Hassan, 2016; He et al., 2013; Mottaleb et al., 2015). When precipitation is below average and/or evaporation and transpiration are high, soil water levels drop, leading to a drought in the agricultural sector (Dai, 2011; Quiring and Papakryiakou, 2003). This catastrophe is localised to a certain time and place (Mishra and Singh, 2010). Subsequently, pinpointing the beginning and end of a farming dry season episode in a particular location is frequently troublesome.

As a corollary, the intensity and expected temporal and geographical fluctuations may vary considerably over time and space (Touma et al., 2015). As a result, agricultural drought is a significant risk to crop output and a leading cause of agricultural losses worldwide (Piao et al., 2010; Wang et al., 2018b). For instance, in Australia, winter wheat crop output dropped by 36% due to the drought of 2006, costing the economy roughly AUD$3.5 billion (Wong et al., 2009). Subsequently, fast response moves toward alleviate the adverse results of dry season need dependable discovery of horticultural dry spell, especially ongoing or close to constant dry spell checking. Utilizing appropriate dry spell markers is the main methodology of checking and investigating agrarian dry season.

In most cases, the length and severity of a drought, as well as the extent to which it has affected a region, may be estimated with the use of a drought index. Dry spells may be defined in a variety of ways, making it challenging to use a single index to completely and consistently assess drought features throughout an area. This is why scientists have created 160 different drought indexes (Niemeyer, 2008). Ground-based indices and remotely-sensed indices are two broad categories that describe these measurements. Drought indicators calculated at ground level are often based on local observations of weather conditions like rainfall and temperature. Drought near weather stations may be tracked with remarkable precision (Rhee et al., 2010).

The Palmer Drought Severity Index (Palmer, 1965), the Crop Drought Identification Index (Wu et al., 2018), the Standardized Precipitation Index (SPI; McKee et al., 1993), and the Standardized Precipitation Extremes Index (SPEI) are just a few examples (VicenteSerrano et al., 2010). The SPEI is one such index since it can track several forms of drought across different locations because it takes both precipitation and temperature into consideration. Researches generally agree on it, thus it must be true (Gao et al., 2017; Wang et al., 2015). While useful, these station-based indicators of drought's spatial distribution at the regional scale are unable to capture finer details (Park et al., 2016). Drought conditions in unmeasured areas may be estimated using spatial interpolation techniques based on geographical information systems (GIS), such as inverse distance and kriging; however, vulnerabilities might exist in added regions because of the addition calculations utilized and complex geological circumstances (Lover et al., 2011). Somewhat detected records have been laid out utilizing satellite information to address the geological extent of ground-based dry season markers.

Precipitation, temperature, evapotranspiration, and vegetation information are instances of the constant inclusion and ongoing nature of satellite remote detecting information. As this is the situation, remote detecting dry spell files can precisely mirror the exceptional topographical elements of dry season. Some remotely-detected dry season files have been made in ongoing many years. They include the Normalized Difference Vegetation Index (NDVI; Rouse Jr et al., 1974), the Normalized Difference Drought Index (NDDI; Gu et al., 2007), and the Normalized Multiband Drought Index (NMDI) (Wang and Qu, 2007). Although these indices do address the issue of geographic coverage that plagued traditional ground-based drought indices, their limited history prevents them from fully substituting for ground-based observations in drought monitoring (Lai et al., 2019).

However, meteorological circumstances (such as clouds) and retrieval techniques might reduce the quality of remotely-sensed drought indexes (Zhang and Jia, 2013). Hence, there is still some doubt over whether or not remotely sensed drought indexes are accurate and reliable (Alizadeh and Nikoo, 2018). Ground-based drought indices are often regarded as more accurate than remotely-sensed drought indices when it comes to monitoring agricultural drought. This is mostly due to the higher reliability of climatic variables gleaned through on-the-ground observations. Also, not at all like roundabout vegetative information, most ground-based dry spell markers track the dirt water balance straightforwardly.

In order to confirm their validity, several recently created remotely-sensed drought indicators must be tested against indices derived from the ground (Rhee et al., 2010). The most popular remotely sensed drought indicators are seldom in accord with ground-based drought indices (Bayarjargal et al., 2006). In the United States, counties in east Texas had a weak relationship between NDVI-derived VCI (Vegetation Condition Index, Kogan (1995)) and ground-based drought indices (R2 values around 0.1). (Quiring and Ganesh, 2010). Analysts have looked to utilize information driven models, like counterfeit brain organizations (Morid et al., 2007) and autoregressive incorporated moving normal models, to more readily screen dryness by reproducing ground-based dry season pointers (Belayneh et al., 2014). Unfortunately, these models often have difficulties accounting for drought estimates that include nonlinearities or non-stationarities. A rise in popularity of cutting-edge, flexible machine learning techniques may be traced back to the advancements in AI. In general, machine learning approaches outperform traditional linear regression models because of their ability to analyse hierarchical and nonlinear interactions between independent factors and the dependent variable (Belayneh et al., 2014; Guzmán et al., 2018).

In 2016, Park et al. fused 16 remotely sensed drought parameters in order to create a meteorological drought index using three different machine learning techniques (SPI). Findings revealed that machine learning methods captured > 90% of the SPI variance in the southern United States. In Iran, Alizadeh and Nikoo (2018) used five different machine learning techniques and found comparable outcomes. As a result of combining remote sensing data with ground-based drought indices, we may get a more comprehensive picture of the spatial aspects of a drought event and apply point-based indices over a larger area. Although Australia is a region prone to drought, our research has shown that few studies have utilised comparable methodologies to assess dryness there. Australia is no stranger to drought, and the current drought of 2018 has already lasted for many months. Drought conditions have afflicted most of southeast Australia, which might lead to a 45 percent drop in winter wheat output (https://www.dpi.nsw.gov.au/about-us/publications/pdi/2018/wheat). Thus, it is critical to monitor drought conditions on a regional scale and in a timely manner in order to devise effective methods for risk reduction. The motivation behind this exploration was to survey the suitability of utilizing an assortment of somewhat detected dry spell pointers to portray rural dry season in south-eastern Australia. This examination expected to: 1) assess the exactness of three AI techniques (bias-corrected random forest (BRF), multi-layer perceptron neural network (MLP), and support vector machine (SVM)) in repeating ground-put together SPEI based with respect to different remotely-detected dry spell factors in dry and wet regions; 2) decide the overall significance of remotely-detected dry season factors in deciding SPEI; and 3) contrast model-anticipated dry spell maps and corresponding perceptions.

III.PROPOSED METHODOLOGY

The Deep Learning models like CNN and VGG16 transfer models are implemented and compared for their performances. The time and effort spent training a new model may be saved by using transfer learning, a technique for feature representation based on a previously learned model. Pre-trained models are often trained on large datasets like Image Net, and the resulting weights may be utilised in your own bespoke neural network implementation for any purpose. These freshly constructed models may be utilised for either direct prediction on relatively novel problems or for training purposes in related applications. This method not only speeds up the learning process but also minimises the overall mistake rate.

Commonly used Transfer Learning Models

❖ Inception
❖ Xception
❖ VGG family
❖ Res Net

1. CNN CLASSIFIER

The CNN (Convolutional Neural Networks) is used in this research for face covering detection (i.e., helmets, scarves, and masks). The method is used to determine whether or not a person's face is covered by training and testing it with photos of individuals wearing helmets, scarves, and masks. CNNs are a popular kind of Neural Networks used for identifying and categorising images. Supervised learning is used in CNN. Filters and neurons in CNN 10 are biassed or have weights. Every filter operates by performing a convolution on the data that it has collected. A Convolutional, Pooling, Rectified Linear Unit (ReLU), and Fully Connected layer make up the CNN classifier.

I. Convolutional layer The picture is sent into this layer, and features are extracted to use as input. The output picture from this layer is used as input for the subsequent convolutional layer, and it contains a feature map generated by the neurons that were involved in processing the input image. A CNN's primary component is its convolutional layer. It has specific channels (or pieces) in it whose settings will be sorted out as you train. Channels frequently have a more modest impression than the last picture. Each channel is utilized to create an actuation map by convolving with the info picture.

II. Pooling layer Using this layer, we can reduce the size of the feature map without losing any of the key details. It is common practise to insert this layer in between two convolutional layers.

III. ReLu layer By using the non-linear operation of ReLu, all negative values in the feature map are changed to zero. This is an atomic operation. It's another name for the concept is transfer function.

IV. Fully Connected layer By definition, FLC implies that all filters in one layer are linked to all filters in the next layer. This is used to assign a label to each feature in the input picture based on the training dataset. In its four stages, they are as follows: Constructing a Model, Number Eleven (2) Instruction in the use of models Third, verifying the model 4. Assessment of the Model

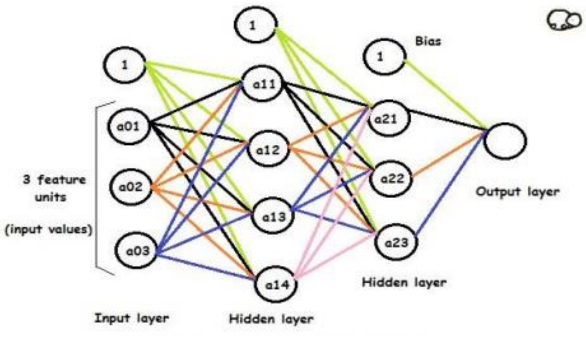

Figure1.Fully Connected Layer

2.VGG16 MODEL

To apply a model that has been trained on one issue to solve another, similar problem, is an example of transfer learning. Transfer learning is a method used in deep learning in which a neural network model is pre-trained on an issue with comparable characteristics to the target problem. After the model is trained, it may be utilised to inform the training of a new model for the same issue. The training duration of a neural network model may be reduced using transfer learning, and in certain cases, generalisation error can be reduced as well. When tackling a new issue, the weights in previously utilised layers might serve as a baseline from which to train a solution. In this context, transfer learning is understood to be an alternate method of setting the weights. At the point when the principal related issue contains much more marked information than the issue of interest, and the closeness in the design of the issues might be pertinent in the two subjects, this may be a useful technique.

3.CONFUSION MATRIX

As a result of its intuitive nature and its use as a starting point for computing other crucial measures like accuracy, recall, precision, etc., it has become the de facto standard for evaluating predictive models. The whole exhibition of a model on a given dataset is depicted by a NxN lattice, where N is the complete number of class marks in the order task. The sum of every correct prediction divided by every false one, and vice versa. The table below displays the percentages of correct predictions (True Positive [TP], True Negative [TN], and False Positive [FP]) made by each algorithm. As the term "true positive" (TP) suggests, the model accurately predicted the number of observable positive classes in the training set. When a positive class is incorrectly forecasted as a negative class, this is called a false negative (FN), and the model accurately anticipated the number of negative classes in the sample. The term "false positive" (FP) refers to the situation in which the model incorrectly predicts a larger number of positive classes of data than the actual number of negative classes.

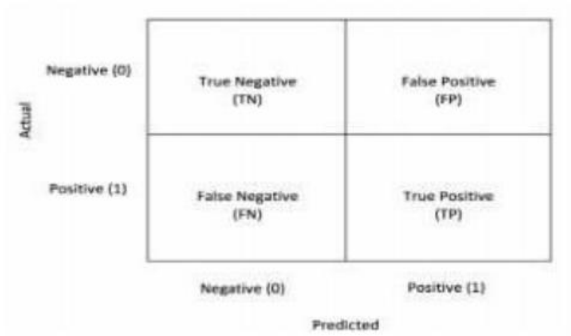

Figure2.Confusion Matrix

4.MACHINE LEARNING APPROACHES

Forestry Technique that Accounts for Bias As a kind of integrated learning algorithm, Random Forest (RF) takes use of random sampling and integration techniques to build a random forest out of several decision trees. Using the sample set produced by bootstrap, RF initially produces a number of separate trees. If you have a big enough training sample, around 37% of it will be kept for out-of-bag validation. In order to establish a final result for each tree in the forest, RF uses the bootstrap approach to generate a random sample from the whole training set. The median values of the trees are the end result of RF methods. As a consequence, as compared to standard tree-based algorithms, RF is able to reduce the variance and achieve more accurate prediction outcomes (Remote Sens. 2022, 14, 6398, 9 of 21). It may, however, introduce bias when attempting to anticipate outlying findings. The predictions of RF tend to overestimate when the number of observations is small and to underestimate when the number of observations is big. As part of our research, we used bias correction strategies to accurately quantify and adjust for RF bias in the regression. These are the specifics of this method for correcting for bias:

(1) Create the RF model by first training on the dataset Ytrain = RF (Xtrain), where Xtrain and Ytrain are the independent and dependent variables, respectively.

(2) Determine the residual and estimated value by using the formula rtrain = Ytrain Ypredict, where rtrain is the residual and Ypredict is the predicted value.

(3) Fit the random forest model, rtrain = rfres, where the residuals from (2) are the dependent variable and the training dataset from (1) is the independent variable (Xtrain, Ytrain). The residual of the test dataset is estimated here.
(4) Using the RF model derived in step (1) and the test dataset Xtest, (4) get the estimated value Ytest = RF (Xtest).
(5) Step 3's rfres model, Step 4's estimated value, and the test dataset's independent variables are used in Step 5 to get the estimated residual, rtest = rfres (Xtest, Ytest).
(6) To adjust for variance, we add the estimated residual rtest to the estimated value Ytest: Ybias-correction = Ytest + rtest.
I.XGBOOST
The XGBoost tree is a very powerful gradient lifter. With its various algorithmic and technical enhancements, it effectively implements the gradient boosting decision tree (GDBT) technique for addressing the illness. When adopting data, XGBoost takes a more random forest-like approach than the standard GBDT method. To further enhance the model's generalisation capabilities and minimise over-fitting, XGBoost includes a rule term to regulate the model's complexity. These are the specifics of XGBoost's method:

(1) One way to expand a tree is to keep adding new trees and splitting off features. To account for the change in the residual between estimates introduced by each new tree, a new function f(x) is learnt.

Minimizing this loss function yields the ideal model. obj(t) = ∑ n i=1 l(yi , yˆi) + Ω(f(t)) + Constant.

(1) After XGBoost has been taught to produce k trees, it must estimate the outcome of a sample. As a general rule, the example will arrive on a solitary leaf hub for every tree, with each leaf hub addressing an alternate score. Toward the end, XGBoost will include the discoveries for each tree to give a gauge to the entire, composed as yˆ (k) I = ∑ k γkhk(xi), where K is the absolute number of trees, k is the kth tree, k is the heaviness of this tree, and hk is the assessment for this tree.

II.SUPPORT VECTOR MACHINE
It's safe to say that support vector machine (SVM) is one of the most popular machine learning techniques out there. Strong learners with classification and regression abilities, SVM algorithms have their roots in statistical learning theory. SVM is used to find a sample distribution that can be partitioned into two or more hyperplanes. The goal of segmentation is to maximise the interval, which becomes a convex quadratic programming problem at the end of the process. As a machine learning technique, SVM is the most similar to deep learning. As compared to a two-layer neural network, nonlinear SVM performs similarly. In order to model a multi-layer neural network, nonlinear SVM may be supplemented with several kernel functions. In this research, we use Python's Scikit-learn machine learning toolkit to run the support vector regression model.

III.ACCURACY EVALUATION
In this research, we improve the effectiveness of machine learning strategies by iteratively discovering the model-stability-influencing parameters and then determining the best values for those parameters using cross-validation. A small subset of the whole dataset is chosen at random and split into a training set and an evaluation set. To ensure accuracy, this procedure is repeated a total of 100 times. Effectiveness of the model is measured by the determination coefficient (R2) and the mean square error (RMSE):

$$R^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})}\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})}}\right)^2$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}}$$

Where O and P are the method for the noticed and assessed values, n is the quantity of tests, and Oi and Pi are the noticed and assessed values, separately. As a guideline, the better the model's presentation is, the more prominent R 2 and the lower the RMSE. Also, we estimated drought conditions over time by using a number of different models and then used station retention cross validation to determine which ones were the most reliable.

IV.DESCRIPTION OF THE MODEL

Cleaning and organising data is the first step in every successful machine learning endeavour. Data collection, selection, preparation, and transformation are all integral parts of the second stage of project execution. A number of sub-steps make up each of these stages.

COLLECTION OF DATA: -

• A data analyst must now take charge in directing the rollout of machine learning. Data analysts are responsible for identifying methods and resources for gathering data, evaluating the data using statistical methods, and drawing conclusions from the findings.
• The data you use to make your predictions should reflect the nature of your target phenomenon.
Due to the fact that every machine learning issue is different, answering the question "How much data is needed?" with precision is impossible. Thus, the predictive value of characteristics determines how many features data scientists will utilise when constructing a predictive model.
A "more is better" mentality is acceptable at this time. Data scientists have suggested that maybe only a third of the information gathered is really relevant. Before model training starts, it is impossible to predict which subset of the data will provide the most reliable findings. For this reason, it is crucial to gather and retain any and all data, whether private and public, organised and unstructured.
Industry and company structure will determine the best methods for gathering internal data. If you have an online-only firm and want to begin a personalisation campaign, some web analytic tools you may consider include Mix panel, Hot jar, Crazy Egg, the ubiquitous Google Analytics, etc. In addition, a website's log file might serve as a reliable resource for gathering confidential information. Visitation times, page views, and clicked-on links are just some of the data that may be stored.
To supplement their own data, businesses may also use publicly accessible datasets. Kaggle, Github, and Amazon Web Services (AWS) are just a few examples of places to get free datasets for research.

DATA PREPROCESSING:-

The goal of data preprocessing is to prepare the data for machine learning. If a data scientist has access to clean, well-organized data, the resulting machine learning model will be more accurate. The method involves collecting, organising, and sampling data.

FORMATTING DATA: -

As information is gathered from several persons and places, its presentation takes on more significance. A data scientist's first order of business is to standardise data file formats. An expert verifies that the variables used to record each characteristic are consistent with one another. Product names, pricing, date formats, and mailing addresses are all instances of variables. Consistency in data is especially important for qualities that take the form of numeric ranges.

CLEANING THE DATA: -

Using these methods, we can clean up the data and address any problems that have cropped up. A data scientist may use imputation methods, such as replacing missing values with mean characteristics, to complete incomplete datasets. Experts can also spot outliers, or data points that are drastically different from the norm. A data scientist will discard or fix potentially inaccurate data if an outlier shows it is there. At this phase, any data items that are either incomplete or otherwise worthless are also deleted.

ANONYMIZATION OF DATA: -

Periodically, an information researcher might have to eliminate or cover characteristics that could uncover private data (for example while working with medical care and banking information).

SAMPLING DATA: -

Analysis of large datasets takes more time and computing resources. Using data sampling is the best option when working with a massive dataset. In order to construct and execute models more quickly while maintaining accuracy, data scientists often utilise this method to choose a smaller-than-average yet representative sample of available data.

PROCESSING OF IMAGES: -

Several types of image processing include analogue and digital methods.

The term "digital image processing" refers to the processing of digital pictures using computational techniques. When contrasted with its simple ancestor, computerized picture handling — a part of computerized signal handling — offers various advantages. The objective of advanced picture handling is to improve the image information (highlights) by stifling undesired contortions or potentially upgrading specific basic picture characteristics, so our man-made consciousness and PC vision models can utilize the upgraded information.

• READ IMAGES: -

Here, we put up a variable to hold the location of our picture collection and a method to import image folders into arrays.

• IMAGE RESIZE: -

To see the shift, we'll now make two picture-display functions: one to show only one image, and another to show both at once. Next, we construct a processing function that takes in a set of photos as its only argument. The need for resizing stems from the fact that the sizes of the pictures taken by cameras and input into our AI system are not uniform.

DATA AUGMENTATION:-

Computer vision tasks including picture classification, object recognition, and segmentation are among the most often used examples of deep learning's effectiveness. In such cases, data augmentation may be employed to train the DL models more efficiently. Simple picture manipulations include geometrical operations like flipping, rotating, translating, cropping, and scaling, and alterations to the colour space like colour casting, varying brightness, and noise injection.

DATA SPLITTING:-

If you're going to employ machine learning, you'll need to split your dataset into three parts: training, test, and validation.

TRAINING SET: -

With the use of a training set, a data scientist may teach a model its "best guesses" for its various settings, or the "parameters" it must learn to utilise.

TESTING SET: - To judge how well an educated model generalises, one needs a test set to put it through its paces. This second definition refers to a model's capacity to recognise similarities or differences in data that it has never seen before but has been taught to recognise in its training data. To prevent model over fitting, the inability for generalisation as discussed previously, it is essential to employ diverse subsets for training and testing.

MODELING:-

Here, the data scientist trains many models to see which one yields the most reliable predictions.

MODEL TRAINING:-

Now is the time to train the model using the available data. Fast.ai's wide selection of available architectures simplifies the process of implementing transfer learning. The pretrained models that function for majority of the applications/datasets may be used to build a convolutional neural network (CNN) model. As ResNet architecture is quick and accurate across a wide variety of datasets and challenges, we will be using it. The "18" in resnet18 refers to the network's total number of hidden layers. We also provide the metric for gauging the accuracy of the model's predictions with the help of the data loader's validation set. We are monitoring how often the model is wrong with its predictions using error rate. Comparable to the fit() function in other ML libraries, fine tune is used to fine-tune the model. To train the model, we must now define how many times (epochs) the model will be exposed to each input picture.

V.SYSTEM DESIGN

While designing a system, it is necessary to establish the interface, modules, and data in order to specify the requirements the system must meet. It might be said that system design is the practical application of system theory. The essential objective of framework configuration is to make the framework design by giving the raw numbers expected for the framework's genuine execution.

ARCHITECTURE DIAGRAM:

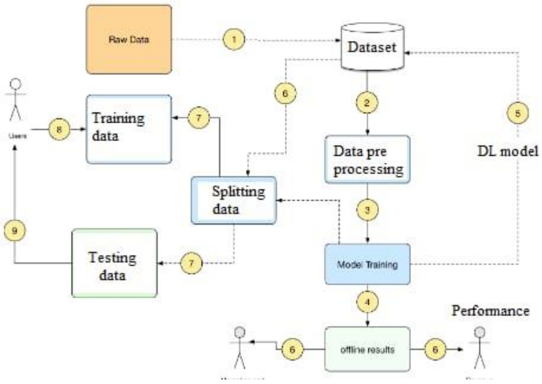

Figure3.Architecture Diagram

SEQUENCE DIAGRAM:

If you look at the sequence diagram of a system, you can see how the interactions between its many entities are arranged in a certain time frame. In this way, it plans out the classes and articles engaged with the plot, as well as the succession of messages that should be sent between the body to accomplish the situation's objectives.

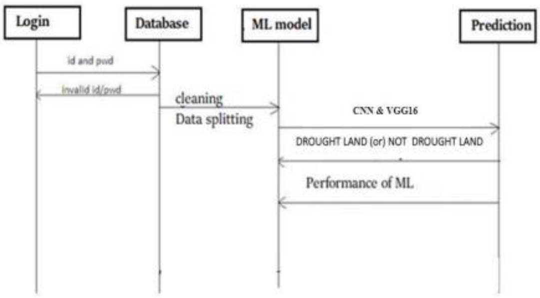

Figure4. Sequence Diagram

CLASS DIAGRAM:

The class outline is an immobile portrayal. It subs for the constant point of view of a program. Notwithstanding its more clear applications in framework representation, documentation, and portrayal, class outlines are likewise utilized in the genuine development of the program's code.

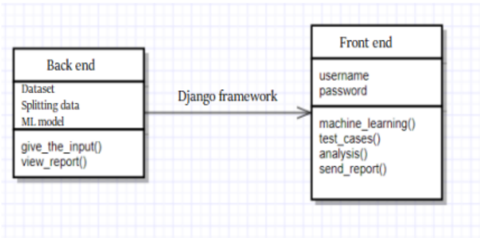

Figure5.Class Diagram

ER DIAGRAM

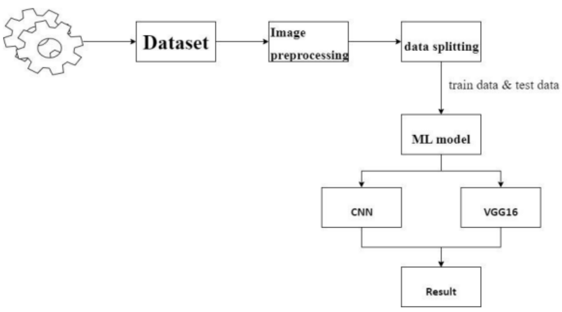

Figure6.ER Diagram

VI. CONCLUSION AND FUTURE ENHANCEMENT

In this article, we use a real-world dataset to undertake an in-depth analysis of deep learning models for drought prediction. The effects of drought on agriculture and the economy are felt not only in India, but across the globe. Using data analysis, we offer a strategy in this study for drought forecasting. Using the Deep Learning algorithm, this system explains how to anticipate drought. Use of explainable DL techniques for comprehension improvement. In order to deal with the unpredictable nature of unfavourable natural phenomena as they emerge and evolve throughout varying climates and seasons, cutting-edge technology and methods are used.

This study's methodology may be used elsewhere, particularly in regions where observational data is scarce but where remote sensing satellites can provide information on the geographical distribution of drought severity. There are several factors that contribute to drought, including altitude and the kind of plants covering the land. Future research must take into account these elements to improve the accuracy of the models and hence improve drought monitoring. The BRF model also has several drawbacks, such as an inability to accurately anticipate the occurrence of intense drought and a propensity to underestimate the severity of drought. In order to increase the models' ability to identify severe droughts, future research should take into account both improved machine learning models and additional elements that contribute to drought. In this investigation, hybrid models fared better than their separate counterparts. Overall, with the help of these improved machine learning algorithms and a number of easily obtainable input characteristics, river water quality may be properly predicted.

Although the current model is only capable of detecting the localised effects of drought, it has the potential to be expanded in the future to monitor broader land changes and warn of impending disasters. The salinity of the water in a drought-stricken area may be determined by using cutting-edge hybrid deep learning models.

REFERENCES

[1] Ali, S.; Tong, D.M.; Xu, Z.T.; Henchiri, M.; Wilson, K.; Shi, S.Q.; Zhang, J.H. Characterization of Drought Monitoring Events through Modis-and Trmm-Based Dsi and Tvdi over South Asia During 2001–2017. Environ. Sci. Pollut. Res. 2019, 26, 33568–33581. [CrossRef] [PubMed]

[2] Quiring, S.M.; Papakryiakou, T.N. An Evaluation of Agricultural Drought Indices for the Canadian Prairies. Agric. For. Meteorol. 2003, 118, 49–62. [CrossRef]

[3] Wei, W.; Pang, S.F.; Wang, X.F.; Zhou, L.; Xie, B.B.; Zhou, J.J.; Li, C.H. Temperature Vegetation Precipitation Dryness Index (Tvpdi)-Based Dryness-Wetness Monitoring in China. Remote Sens. Environ. 2020, 248, 111957. [CrossRef]

[4] Yao, N.; Li, Y.; Lei, T.J.; Peng, L.L. Drought Evolution, Severity and Trends in Mainland China over 1961–2013. Sci. Total Environ. 2018, 616, 73–89. [CrossRef]

[5] Trenberth, K.E.; Dai, A.; Van Der Schrier, G.; Jones, P.D.; Barichivich, J.; Briffa, K.R.; Sheffield, J. Global Warming and Changes in Drought. Nat. Clim. Change 2014, 4, 17. [CrossRef]

[6] Dai, A.G. Erratum: Drought under Global Warming: A Review. Wiley Interdiscip. Rev.-Clim. Chang. 2012, 3, 617. [CrossRef]

[7] Vicente-Serrano, S.M.; Quiring, S.M.; Pena-Gallardo, M.; Yuan, S.S.; Dominguez-Castro, F. A Review of Environmental Droughts: Increased Risk under Global Warming? Earth-Sci. Rev. 2020, 201, 102953. [CrossRef]

[8] Daryanto, S.; Wang, L.X.; Jacinthe, P.A. Global Synthesis of Drought Effects on Food Legume Production. PLoS ONE 2015, 10, e0127401. [CrossRef]

[9] Daryanto, S.; Wang, L.X.; Jacinthe, P.A. Global Synthesis of Drought Effects on Maize and Wheat Production. PLoS ONE 2016, 11, e0156362. [CrossRef]

[10] Loon, V.; Anne, F. Hydrological Drought Explained. Wiley Interdiscip. Rev. Water 2015, 2, 359–392. [CrossRef]

[11] Allen, C.D.; Macalady, A.K.; Chenchouni, H.; Bachelet, D.; Mcdowell, N.; Vennetier, M.; Kitzberger, T.; Rigling, A.; Breshears, D.D.; Hogg, E.H.; et al. A Global Overview of Drought and Heat-Induced Tree Mortality Reveals Emerging Climate Change Risks for Forests. For. Ecol. Manag. 2010, 259, 660–684. [CrossRef]

[12] Ciais, P.; Reichstein, M.; Viovy, N.; Granier, A.; Ogee, J.; Allard, V.; Aubinet, M.; Buchmann, N.; Bernhofer, C.; Carrara, A.; et al. Europe-Wide Reduction in Primary Productivity Caused by the Heat and Drought in 2003. Nature 2005, 437, 529–533. [CrossRef] [PubMed]

[13] Zhao, M.S.; Running, S.W. Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 through 2009. Science 2010, 329, 940–943. [CrossRef]

[14] Aadhar, S.; Mishra, V. High-Resolution near Real-Time Drought Monitoring in South Asia. Sci. Data 2017, 4, 170145. [CrossRef] [PubMed]

[15] He, B.; Wu, J.J.; Lu, A.F.; Cui, X.F.; Zhou, L.; Liu, M.; Zhao, L. Quantitative Assessment and Spatial Characteristic Analysis of Agricultural Drought Risk in China. Nat. Hazards 2013, 66, 155–166. [CrossRef]

[16] Mottaleb, K.A.; Gumma, M.K.; Mishra, A.K.; Mohanty, S. Quantifying Production Losses Due to Drought and Submergence of Rainfed Rice at the Household Level Using Remotely Sensed Modis Data. Agric. Syst. 2015, 137, 227–235. [CrossRef]

[17] Prodhan, F.A.; Zhang, J.H.; Yao, F.M.; Shi, L.M.; Sharma, T.P.P.; Zhang, D.; Cao, D.; Zheng, M.X.; Ahmed, N.; Mohana, H.P. Deep Learning for Monitoring Agricultural Drought in South Asia Using Remote Sensing Data. Remote Sens. 2021, 13, 1715. [CrossRef]

[18] Yang, X.; Li, D. Temporal and Spatial Evolution Characteristics of Strong Drought Events in North and Northeast China. Arid Land Geogr. 2019, 42, 810–821.

[19] Ren, J.; Zhang, T. Evolution Characteristics of Drought and Flood in Shandong Province in Recent 45years Based on Standardized Precipitation Index. Res. Soil Water Conserv. 2021, 28, 149.

[20] Zhang, J.; Mu, Q.Z.; Huang, J.X. Assessing the Remotely Sensed Drought Severity Index for Agricultural Drought Monitoring and Impact Analysis in North China. Ecol. Indic. 2016, 63, 296–309. [CrossRef]