ijRASET

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Cloud Type Classification for Prediction of Rain Fall using CNN

Anuska Sarkar[1], Riya Bhowmik[2], Chetan Shaw[3], Arijit Bhattacharjee[4], Ipsita Saha[5], Srabani Kundu[6], Sayani Chandra[7], Sourish Mitra[8]

[1, 2, 3, 4]*B.Tech Students,* [5, 6, 7, 8]*Assistant Professor, Department of Computer Science & Engineering, Guru Nanak Institute of Technology*

Abstract: In order to forecast the climate, clouds are essential. Prediction of rainfall is also greatly influenced by the quantity and nature of clouds in the atmosphere. As a result, the most fascinating and important subject in meteorology is cloud identification, which also draws the most researchers from other fields. The transfer learning method used in this article to forecast rainfall using ground-based cloud images is presented. By classifying the type of cloud using cloud image input, it will be able to forecast the expected amount of rainfall. Based on the associated Precipitation that caused the appropriate Rainfall, the cloud images in the dataset are divided into three groups (classes) labeled as no rain to very low rain, low to medium rain, and medium to high rain. In this paper using CNN the classification of rainfall was done
Keywords: Rainfall, Forecast, Dataset, CNN, Classification

I. INTRODUCTION

Without human intervention, this article depicts the weather classifications model. In the article, a model for intensified climate foresight using data mining to predict the weather is introduced. While other data mining-based algorithms generate specific traits, the model can infer any characteristic possible in the dataset. It operates effectively on mobile devices like the Android ecosystem and is computationally useful (Abrar et al., 2014). It isolates the cloudy, rainy, shiny, and foggy designs and reveals the various climatic conditions. The environmental factors and surface energy supplies in this instance, along with considerations for the air-mass review, are used to classify the genetic traits (Arnfield, 2016; 3. CC. Lee, 2020). In order to facilitate interpretation in city atmosphere considerations, it provides a system that supports a regional weather type class process. Here, this aids in analysing how certain environmental circumstances, such as wind speed changes from low to high or low to increased, hot to cold, or hot to cold, change.

It aims to define future observations correctly, as it does frequently (Hidalgo, R. Jougla, 2018). It technically demonstrates the peculiarities of climate change on the basis of high-resolution baseline climate forecasts. Both times assess the assemblage range's confidence levels, providing pertinent consequences for the authenticity, such as the classifications' accuracy (Beck et al., 2018). In order to sense and simulate temperature and precipitation according to zones for different climates, this system satisfies probability density functions (PDF). The study looks at variations in weather and precipitation. In order to improve performance, the study analyses temperature and precipitation variations over the most recent weather findings. To obtain accuracy with results, this system combines numerous models from various regions (Remedio et al., 2019).

The development of real-time applications, such as computer vision, image processing, speech recognition, text processing, etc., benefits from the use of machine learning methods.Additionally, it offers stunning outcomes for categorising text data in accordance with various climatic factors like temperature and humidity. (Gad & Doreswamy,2020). An artificial neural network (ANN) designed for image pattern recognition is called a convolution neural network (CNN). Identifying a picture's features using feature maps is appropriate for image analysis. Convolutional (ConV) layers are the hidden levels that make up CNN. These ConV layers take input, modify it into a certain mode, and output the modified input. The same process is now providing data to the following layer. This adjustment, which is a convolution procedure, makes it easier to precisely and precisely identify patterns from images. The quantity of filters must be specified for each convolution layer.

Edge filters and object filters, respectively, play a crucial part in the detection of edges and objects. In order to identify the pattern of the weather image dataset for this proposed study (Ketkar, 2027), the CNN method has been used. This study examines how well the designed and built deep learning model performs on the original Kaggle.com weather image dataset using the CNN architecture with the Keras framework and TensorFlow system-defined libraries. Finally, classification and temperature prediction are the two main functions of these deep learning algorithms.

II. LITERATURE REVIEW

A thorough literature review on the prior works on cloud type classifications done with CNN or other various methods before we could move forward with our project work. Thus, we could more clearly define our objective and way of operation. There has been numerous Cloud type detection models suggested in recent years based on various technical methods, as they are one of the most crucial parameters of the Weather detection process. The difference has also been seen in the datasets that have been used, in addition to the methodical variations. In addition to the conventionally popular satellite images, whole sky imagers, ground-based sky imagers, infrared images, and meteorological images have all been used to create prediction models. A cloud classification algorithm was put forth by Jun Li et al. in 2002 using MODIS multi-spectral beam measurements. The entire procedure was very effective. The entire procedure was fairly effective, but in the area of accuracy rate, there are some problems because First, a particular scene type might not be identified or distinguished; this typically occurs when the class in MODIS visible and infrared imagery appears to be very near to another class.Snow can sometimes be hard to tell apart from low clouds because it resembles them so closely; it might even be mistaken for them. The crew also noticed a problem with the efficiency of the computations, which was slow. The efficiency of the computation is crucial because the model uses real-time input. The MODIS model, which was implemented on a 2000 computer, required several minutes to calculate on a silicon graphics. According to the finding, a stable classification typically requires several iterations.

As previously stated, new models have been created using various imagery datasets, with whole sky imager being one of them.To identify the sort of cloud, Kenneth A. Buch Jr. et al.[2] used the whole sky imager dataset.It is a pioneering attempt to use WSI in this area. Decision Their systematic technique was Binary Tree. They concentrated on the conventional classification of clouds, which is cirrus, altocumulus, altocumulus, and stratus. This classification was arbitrary, and representative images from each type were used as training data, while the remaining samples from each type were selected to assess the precision of the classification produced by the testing of the data.This model, which was one of the earliest attempts, produced a very promising result for being built on WSI data the transformation of unclassified raw pictures.

The processing of unclassified raw images contributed to a 10% improvement in efficiency overall.When looking at the results of these techniques, the primary issue became apparent, which was that half of the input cloud images were labelled as cirrus, which worsened the test data results. The developer team stated that the context of the error is that the cirrus cloud is smoother than the other kinds without having less visual variation. Re-substitution error is very minimal in the decision tree's center, but it considerably increased in the outer regions of the tree due to the lower resolution. One of the creative approaches is to use the textural characteristics of satellite images to identify the sort of cloud.Contribution from Z. Ameur al. (2021). When applying the Karhunen-Loeve transformation (KLT) and K-Means algorithm to images captured by Meteostat 4's during December 1994 that represented Africa, Z. Ameur al. (2021) made a contribution to the area.This procedure is very simple to use, and it has classification rates that are typically higher than 96%. It also has computation rates that are three times faster than those of some other techniques, such as GLCM and GLDV.In the paper, the team discussed some of its limitations, such as how the classes in this method can't yield full information about the observed cloud type.The classify accuracy peaked at 90.43% using the Lei Liu et al. (2011) method based on the structural features of cloud's infrared images as they used cloud edge segmentation and edge detection procedure after their earlier work of extracting the features of cloud images(Sun et al. 2009).

This procedure has a much lower misclassification rate than other processes. To identify the sort of cloud, Yuhang Jiangetal. Proposed amodel implemented on FY-4ASatellite image with the CNN. The method was able to obtain 84.4% classification accuracy with the CNN-CLP network implemented, and the computation process only requires 0.9s. Regarding the drawbacks, it is stated in the paper that it was challenging to implement a more accurate classification due to the orbit's running speed and time. The most popular technique in meteorology is identifying the sort of cloud based on satellite images. The suggested model by Keyang Cai et al. (2017) combined the conventional method of using satellite images with the most recent advances in deep learning to produce results that were more precise and accurate. We made the decision to state our issue based on the review of the literature and observation of the successes and shortcomings.

The classification systems for climate developed by Kappen were examined. It makes it easier to analyse weather variations according to geographic locations and atmospheric cycle times. Due to uncertainty in the classification models, measurements and forecasts are used here to determine climate zone variation.The researchers explained the pressure Laplacian and geopotential Laplacian algorithms for categorising climates, and they represented the results using a simulation model for regional climate models (RCM). Any extreme weather, such as drought, floods, strong winds, heavy snowfall, etc., harms biodiversity. (Sun et al.,2009).Understanding the different kinds of global weather is extremely difficult, according to the researchers.

It is challenging to validate manual classification of various weather types according to temperature and precipitation for various longitude, latitude, altitude, etc. The availability of data resources according to areas makes it a difficult job. As a result, as was already stated, adverse weather conditions encourage additional research on weather classifications.(Shields et al.,1998). On multi-class datasets, the authors contrasted a number of highly effective machine learning models. In order to accurately classify the weather, various systems using various algorithms, such as decision tree CART, gradient boosting, KNN, linear regression, lasso, ridge, MLP, deep learning, SVM, and random forest, have been created. There is some overfitting in these algorithms. The linear regression method is very effective for a prediction model, according to the testing phase. Additionally, the efficacy of the k-nearest neighbours algorithm is inferior. (Gad & Doreswamy,2020).The additional authors went on to describe how a global rule was added to a gridded weather typing classification (GWTC) design in terms of space. Several techniques are used to strengthen the system's ability to classify the weather and identify extreme climatic conditions. (CC Lee,2020).

The researchers provided novel weather datasets that were compared to models of REMO from the CORDEX structure using the current experimental setup. These models frequently examine the various categories of climate. In dry and arctic zones with low skill, the REMO generates relatively large annual precipitation and temperature inclinations. Any assessment of a weather data result is always subject to the bias outcome. (Armelle,2019).The biodiversity problem-solving techniques are the centre of this study. The author researched weather changes for an area on the map that was one (01) KM in size. The results are more precise, and the climate is ideal for growing plants and spices. (Arnfield,2019; Beck et al.,2018).Additionally, the intellectuals explore the local weather type (LWT) method to weather classification. Based on statistical methods and atmospheric statistics, this approach is becoming more direct. Analyzing the atmospheric parameters close to the earth is made easier by it. Examining potential issues now is a tried-and-true method of preventing them. (Hidalgo & Jougla,2018).

The distinction between modified KNN (MKNN) and k-nearest neighbours (KNN) for classifying weather was noted by the researchers. This demonstrates that the MKNN offers more trustworthy outcomes than other data mining projection techniques. (Abrar et al.,2014).The authors compared a number of extremely successful machine learning models on multi-class datasets. Numerous systems have been developed to correctly categorise the weather using a variety of algorithms, including decision tree CART, gradient boosting, KNN, linear regression, lasso, ridge, MLP, deep learning, SVM, and random forest. These systems exhibit some overfitting. According to the testing phase, the linear regression approach is very successful for a prediction model. Furthermore, the k-nearest neighbours method performs less well. (Gad & Doreswamy,2020).The subsequent writers explained how a global rule was included in a gridded weather typing classification (GWTC) design with regard to spatial organisation. To improve the system's ability to categorise the weather and recognise extreme climatic conditions, several methods are used. (CC Lee,2020).

III. PROPOSED WORK

A. Problem Statement

To determine the cloud type of a certain area from the image of the clouds this helps to predict the weather. As CNN is more in accurate in image analysing, we are going to use it on this problem to get an accurate classification of a cloud of the certain place.

1) VGG 16 Model

The Very Deep Convolutional Networks for Large-Scale Image Recognition research paper by A. Zisserman and K. Simonyan from the University of Oxford is where the VGGmodel, or VGGNet, was first suggested.This model is known as VGG-16 because it has 16 levels. This model is the best-performing model in ImageNet because it classifies millions of pictures with nearly 92.7% top-5 test accuracy.

2) VGG-16LAYERS

a) Input: The VGGNet accepts 224224-pixel images as data. To maintain a consistent input size for the ImageNet competition, the model's developers cropped out the central 224224 patch in each picture.

b) Convolutional Layers: VGG's convolutional layers use the smallest receptive field possible—33—to record up, down, and left-to-right movement. Additionally, 11 convolution filter sacting is used to transform the data linearly.The next component is a ReLU unit, a significant advancement from AlexNet that shortens training time. Rectified linear unit activation function, or ReLU, is a piecewise linear function that, if the input is positive, outputs the input; otherwise, the output is negative.

In order to maintain the spatial resolution after convolution, the convolution stride is set at 1 pixel. (stride is the number of pixel shifts over the input matrix).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

c) Hidden Layers: The VGG network's secret layers all make use of ReLU. Local Response Normalization (LRN) is typically not used with VGG as it boosts memory usage and training time. Furthermore, it doesn't increase total accuracy.

d) Fully-Connected Layers: The VGG Net has three layers that are all totally connected. The first two levels each have 4096 channels, and the third layer has 1000 channels with one channel for each class.

VGG-16ARCHITECTURE:
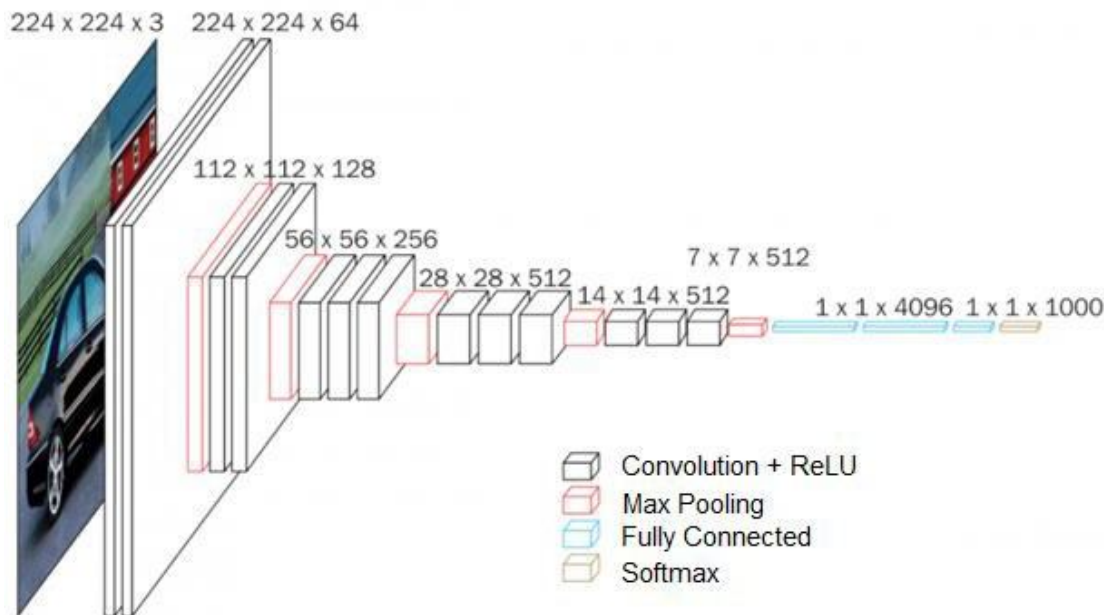
Fig1.VGG16 model architecture

B. Database

For proceeding with our blueprint of the project, we decided to make our model use different type of dataset , which will be easily available and collection process will budget friendly as the majority of Weather related predictions depends on satellite images, the collection of which on daily basis will be expensive , that's why we decided to approach with images collected in ground based way.

The SWIMCAT dataset contains 784 images of sky/cloud patches, categorized in 5 distinct categories, as specified in Table1.

Table 1: Dataset Summary

Number of Images	Type
224	Clear Sky
89	Patterned clouds
251	Thick dark clouds
135	Thick white clouds
85	Veil Clouds
Total:784	

All images have a dimension of 125 X125 pixels, and were captured using WAHRSIS, a calibrated ground-based whole sky imager. The imager is located at Nanyang Technological University in Singapore (1:34°N, 103:68°E). The images were captured over a period of 17 months from January 2013 till May 2014. They were selected based on visual characteristics, and were categorized after consultation with experts from Singapore Meteorological Services

C. PVGG16MODEL

To improve output, the standard VGG-18 was modified to have "smaller vgg net" as one of its characteristics, and it is now known as the PVGG-18 model.
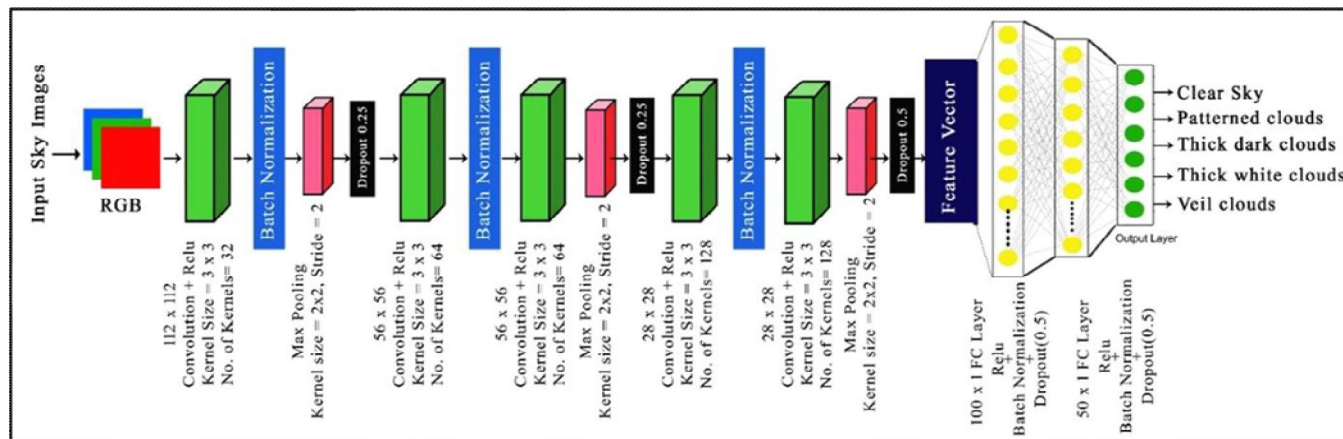
Fig2 .: Proposed model

MODEL SUMMARY OF smaller VGG NET

Fig. 3: Model Summary of Smaller VGG NET

D. Softmax Classifier

One of the popular choice of classifier is the Softmax classifier, which has a different loss function. The Softmax classifier is a generalization of the binary form of Logistic Regression in multiple classes. Unlike the SVM which treats the outputs $f(x_i, W)$ as (uncalibrated and possibly difficult to interpret) scores for each class, the Softmax classifier gives as lightly more intuitive output (normalized class probabilities) and also has a probabilistic interpretation. Softmax classifiers give you *probabilities* for each class label while hinge loss gives you the margin.

Softmax classifier provides "probabilities" for each class

E. Softmax Function

Softmax is a mathematical function that takes as input a vector of numbers and normalizes it to a probability distribution, where the probability for each value is proportional to the relative scale of each value in the vector.

It's denoted by

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

σ = softmax

$\vec{z}$ = input vector

e^{z_i} = standard exponential function for input vector

K = number of classes in the multi-class classifier

e^{z_j} = standard exponential function for output vector

e^{z_j} = standard exponential function for output vector

IV. FINDINGS

A. Cloud Type Classification Output

OUTPUT OF TRAINING CODE

1) Two files will be generated after the compiling the code i.e. Model and. Pickle file
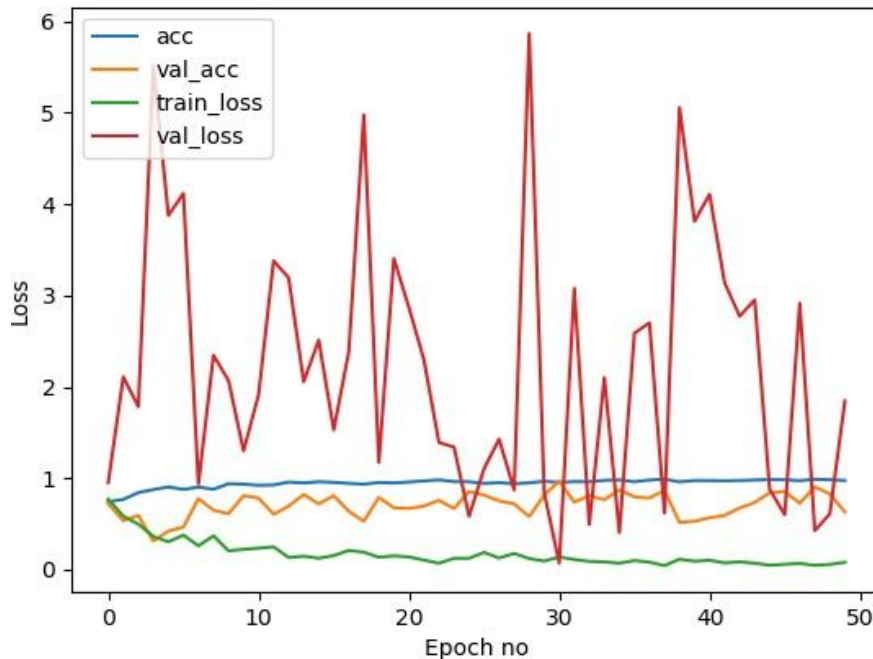
2) A graph will be generated.

Fig. 4: Output graph of Training Dataset

In this graph,as the training is loss is almost less than 1,it is favourable for the final output. The validation loss is too high which will be fixed in future procedure.

FINDINGS

For the output of data testing or the final output, we collected cloud's photos of different

Season from different side of world and used those at random input of testing process.

INPUT IMAGE 1

OUTPUT IMAGE 1

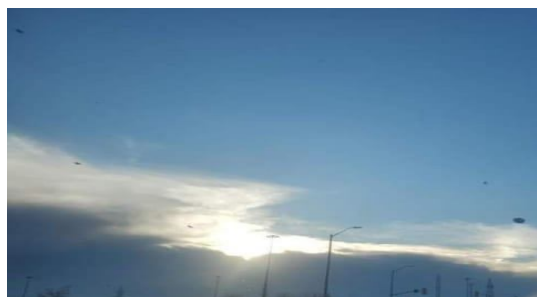[Taken in 2022 summer in West Bengal]

INPUT IMAGE 2

OUTPUT IMAGE 2

[Taken in 2022 summer in West Bengal]

INPUT IMAGE 3

OUTPUT IMAGE 3

[Taken in 2022 summer in Canada]

INPUT IMAGE 4

OUTPUT IMAGE 4

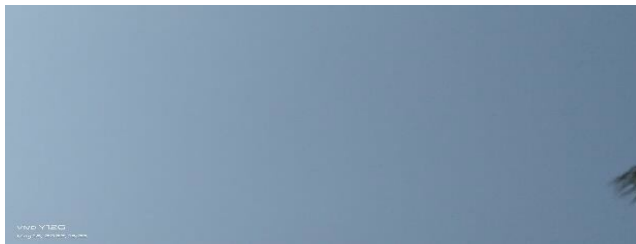[Taken in Sikkim In Summer 2022]

INPUT IMAGE 5   OUTPUT IMAGE 5
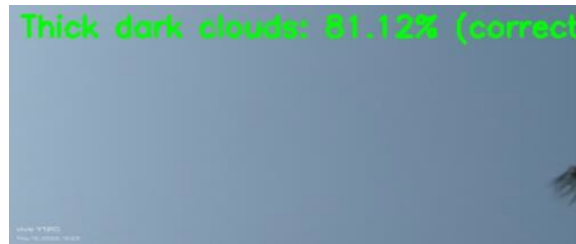
[Taken in Hyderabad in 2022summer]

INPUT IMAGE 6   OUTPUT IMAGE 6

[taken in Kolkata in 2022 summer]
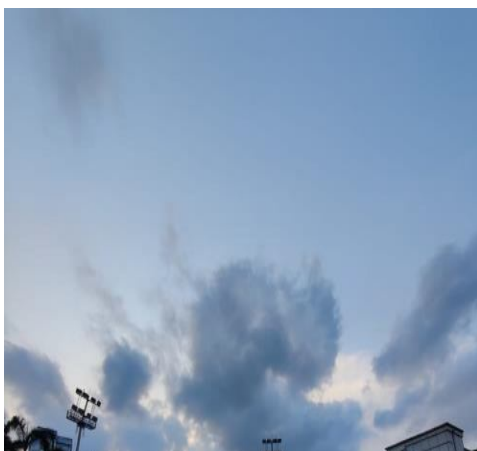

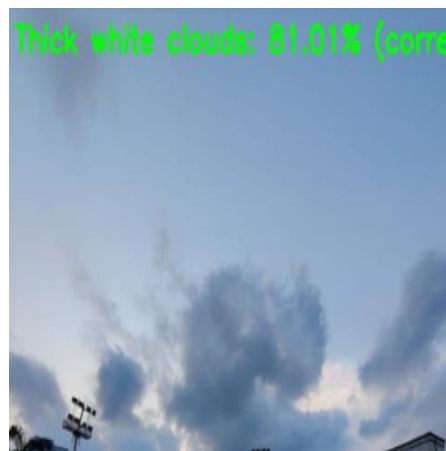INPUT IMAGE 7   OUTPUT IMAGE 7

[taken in USA in 2022 summer]

INPUT IMAGE 8   OUTPUT IMAGE 8

[taken in Kolkata in 2022 summer]

| INPUT IMAGE 9 | OUTPUT IMAGE 9 |

[Taken in Dubai in 2021 Winter]

V. RAINFALL PREDICTION OUTPUT

1) As the final step of the process depends on Softmax classifier, the output we after data testing step does belongs to the image class(among of 5 classes of input dataset) having highest similarities

	A	B	C	D	E
1	Date	Time	Cloud Type	Pridicted Result	Rainfall Prediction
2	14/05/2022	22:53:24	Veil clouds: 98.50% (correct)	98.5037744	Something went wrong
3	14/05/2022	22:53:59	Veil clouds: 98.50% (correct)	98.5037744	Rain or snow will arrive in 24 hour
4	15/05/2022	19:51:11	Veil clouds: 98.50% (correct)	98.5037744	Rain or snow will arrive in 24 hour
5	18/05/2022	12:16:33	Thick white clouds: 99.10% (corre	99.09594059	Modarate of Rain
6	18/05/2022	12:45:21	Clear Sky: 99.97% (correct)	99.97439981	No Rain Sky is full clear
7	18/05/2022	13:04:14	Thick dark clouds: 98.85% (correc	98.85053039	Maximum Possibility of Rain
8	18/05/2022	13:14:02	Patterned clouds: 90.03% (correct	90.02612233	Minimum Possibility of Rain
9	18/05/2022	13:26:43	Patterned clouds: 82.65% (correct	82.64937997	Minimum Possibility of Rain
10					
11	18/05/2022	14:52:44	Thick white clouds: 99.97% (corre	99.96714592	Modarate of Rain
12	18/05/2022	15:01:42	Patterned clouds: 82.65% (correct	82.64937997	Minimum Possibility of Rain
13	18/05/2022	15:13:32	Thick white clouds: 63.39% (corre	63.3931756	Modarate of Rain
14	18/05/2022	16:07:36	Thick white clouds: 63.39% (corre	63.3931756	it doesn't often rain with this clouds
15	18/05/2022	16:16:20	Thick white clouds: 63.39% (corre	63.3931756	it doesn't often rain with this clouds
16	19/05/2022	18:31:36	Thick dark clouds: 81.12% (correc	81.12164736	Maximum Possibility of Rain
17	19/05/2022	18:37:52	Thick white clouds: 99.60% (corre	99.60404634	it doesn't often rain with this clouds
18	19/05/2022	18:41:43	Thick white clouds: 81.01% (corre	81.01093769	it doesn't often rain with this clouds
19	19/05/2022	19:33:31	Thick white clouds: 99.46% (correc	99.45768118	Maximum Possibility of Rain
20	19/05/2022	19:48:25	Thick dark clouds: 99.46% (correc	99.45768118	Maximum Possibility of Rain
21	19/05/2022	20:29:12	Thick white clouds: 99.94% (corre	99.94090796	it doesn't often rain with this clouds
22	20/05/2022	07:00:37	Thick dark clouds: 96.02% (correc	96.01824284	Maximum Possibility of Rain
23	20/05/2022	07:02:04	Thick white clouds: 67.98% (corre	67.97841787	it doesn't often rain with this clouds
24	20/05/2022	07:02:20	Thick white clouds: 98.18% (corre	98.18027616	it doesn't often rain with this clouds
25	20/05/2022	07:02:33	Thick white clouds: 98.09% (corre	98.0935216	it doesn't often rain with this clouds
26	20/05/2022	07:07:53	Thick dark clouds: 99.22% (correc	99.21667576	Maximum Possibility of Rain
27	20/05/2022	07:21:47	Thick white clouds: 90.32% (corre	90.32122493	it doesn't often rain with this clouds

Fig. 5: DECISION TREE OUTPUT

2) With the training of our proposed VGG16 model,the training loss is less than 1which increases the accuracy of the output to highest.

3) As our one of the main focus of the project was to use different type of images other than satellite images, we used SWIMCAT's dataset in which photos are taken from ground based sky imager and used photos taken with mobile phone as input data which gives very accurate result.

4) The addition of the decision tree, the output this model gives can be use to have a primary concept of prediction of the rainfall of a certain area solely based on the cloud type.

VI. CONCLUSION

Most weather forecasting techniques used today rely on satellite imagery. Even though gathering satellite images is a very expensive procedure, identifying the cloud type is crucial for forecasting weather or rainfall. That's why we made the decision to look for a reasonable substitute. Our model can be a productive and cost-effective part of the non-satellite image-based weather forecasting process because it relies primarily on pictures from ground-based imagers or cameras.

This model will make weather detection and forecasting for a small region more effective and time-efficient. We chose to keep the classification type and nomenclature simple enough that it wouldn't be necessary for someone to be familiar with all of the geographical terminology to understand.

In order to provide a more noticeable result on rainfall, we can add more features to our model that can work on other parameters linked to rainfall and cloud. This model won't function if any satellite image is provided as a dataset or input because one of our major objectives in developing it was to find a less expensive alternative to the satellite images in determining cloud type. The cloud type is the only factor taken into account by the rainfall prediction algorithm; it is not affected by any other rainfall parameters. The calculation took longer than we anticipated.

REFERENCES

[1] JunLi,Wisconsin,W.PaulMenzel,ZhongdongYang,*RichardA.Frey,AndStevenA.Ackerman,2002,High-Spatial-Resolution Surface and Cloud-Type Classification from MODIS Multispectral Band Measurements, institute for Meteorological Satellite Studies, University of Wisconsin

[2] KennethA.Butch,Jr & Chen-Hui,Cloud Classification Using Whole Sky Imager Data,Sandia National Laboratoris ,Livermore,CA

[3] Z.Ameur, S. Ameur, A. Adane, Henri Sauvageot, K. Bara. Cloud classification using the textural features of Meteos at images.International Journal of Remote Sensing,Taylor & Francis,2004,25 (21), pp.4491-4503. ff10.1080/01431160410001735120ff. ffhal-00136464f

[4] LeiLiu,XuejinSun,FengChen,ShijunZhao,andTaichangGaoInstituteofMeteorology,Cloud Classification Based on Structure Features of Infrared Images ,PLA University of Science and Technology, Nanjing, China

[5] Jiang,Y.;Cheng,W.;Gao,F.;Zhang,S.;Wang,S.;Liu,C.;Liu,J.ACloudClassificationMethodBasedon a Convolutional Neural Network for FY-4A Satellites. Remote Sens. 2022, 14, 2314.

[6] K. Cai and H. Wang, "Cloud classification of satellite image based on convolutional neural networks," 2017 8TH IEEE International Conference On Software Engineering and Service Science (ICSESS), 2017, pp. 874-877, doi: 10.1109/ICSESS.2017.8343049.

[7] XiaoyingChen,AiguoSong,JianqingLi,YiminZhu,XuejinSun,andHongZeng,Sep2014TextureFeature Extraction Method for Ground Nephogram Based on Hilbert Spectrum of BidimensionalEmpirical Mode Decomposition

[8] Long,C.N.,SlaterD.W.,andToomanT.,2001:TotalSkyImagerModel880statusandtestingresults. Atmospheric Radiation Measurement Program Tech. Rep. DOE-SC/ARM/TR-006, 36 pp.

[9] Shields, J.E.,KarrM.E.,ToomanT.P.,SowleD.H.,andMooreS.T. ,1998:The wholeskyimager—Ayearof progress. Proc. Eighth Atmospheric Radiation Measurement (ARM) Science Team Meeting, Tucson, AZ, ARM.

[10] Shaw,J.A.,ThurairajahB.,EdqvistE.,and Mizutani K.,2002:Infrared cloud imager deployment at the north slope of Alaska during early2002.Proc.12th Atmospheric Radiation Measurement(ARM)ScienceTeamMeeting, Washington, D.C., ARM

[11] Sun, X. J., Liu L. , Gao T. C. , Zhao S. J. , Liu J. , and Mao J. T. , 2009b: Cloud classification of the whole sky infraredimage basedonthe fuzzyuncertaintytexturespectrum(inChinese).J.Chin. Appl.Meteor.Sci.,20,157– 163.

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ☉ (24*7 Support on Whatsapp)