Classifying Weather Images using Deep Neural Networks for Large Scale Datasets

Shweta Mittal, Om Prakash Sangwan

Computer Science and Engineering, Guru Jambeshwar University of Computer Science and Engineering, Hisar, India

Abstract—**Classifying weather from outdoor images helps prevent road accidents, schedule outdoor activities, and improve the reliability of vehicle assistant driving and outdoor video surveillance systems. Weather classification has applications in various fields such as agriculture, aquaculture, transportation, tourism, etc. Earlier, expensive sensors and huge manpower were used for weather classification making it very tedious and time-consuming. Automating the task of classifying weather conditions from images will save a huge time and resources. In this paper, a framework based on the transfer learning technique has been proposed for classifying the weather images with the features learned from pre-trained deep CNN models in much lesser time. Further, the size of the training data affects the efficiency of the model. The larger amount of high-quality data often leads to more accurate results. Hence, we have implemented the proposed framework using the spark platform making it scalable for big datasets. Extensive experiments have been performed on weather image dataset and the results proved that the proposed framework is reliable. From the results, it can be concluded that weather classification with the InceptionV3 model and Logistic Regression classifier yields the best results with a maximum accuracy of 97.77%.**

Keywords—Weather classification; big data; transfer learning; deep learning; Sparkdl; convolutional networks

I. INTRODUCTION

Weather classification is a task of classifying weather conditions by looking at the weather data. Accurate weather classification enables users to organize day to day activities for instance deciding what type of clothes to wear, planning outdoor travel or sports activity, solar technologies, etc. [1]. It plays a critical role in various other fields such as in agriculture, it helps farmers to decide which fertilizers or pesticides to use, whether to turn off the sprinklers etc. based on the outdoor weather conditions. Accurate weather classification also helps in improving reliability of vehicle assistant driving and outdoor video surveillance systems [2]. Timely prediction of severe weather conditions helps saving people life and property. It helps to avoid road accidents, train derailment and ship collisions caused due to severe weather conditions like rain, fog, storm and snow [3]. It also assists in predicting natural calamities, thereby helping in saving people's life to a greater extent. Thus, timely and accurate classification of weather images is of utmost important to society and needs to be explored further by the researchers and meteorologists.

Convolutional Neural Network (CNN) is the commonly used deep learning architecture now-a-days to classify image datasets. It comprises multiple convolutional and pooling layers for better hierarchical representation and a Fully Connected Layer at the end for classification. Numerous variants of CNN such as ResNet, Inception V3, Exception, VGG-16, and VGG-19, etc. have been developed by researchers in past for improving the accuracy of networks. Several features such as sky, shadow, reflection, haze, contrast, etc. are present in images which are used by the CNN for weather classification.

In this paper, a framework based on transfer learning techniques has been implemented for classifying the weather images into its appropriate category with the help of features extracted from pre-trained deep CNN models. Automating the classification will not only save time and resources but also aids in increasing the reliability of the process. Section II highlights the work done by researchers in the field of weather image classification. Limitation and challenges of existing technologies has been presented in Section III. Section IV defines the proposed framework. In Section V, the hardware and software required to implement the proposed framework have been discussed. The experimental results have been described in Section VI and the conclusion and future work has been presented in Section VII of the paper.

II. RELATED WORK

In the past few years, a significant amount of work has been done by researchers to resolve the issue of weather classification. As the weather dataset is quite huge, researchers have implemented big data technologies such as Hadoop and Spark to process weather datasets. Alam and Ajmad proposed architecture for parallel and distributed processing of big Weather datasets using Hadoop and Map Reduce platforms [4]. Yang et al. used big data technology to implement a regression model for predicting weather conditions with a large amount of weather data [5]. Ismail et al. build a big data framework for predicting weather temperature using Map Reduce technique and the proposed framework had the advantage of improved efficiency and scalability [6].

Dhoot et al. implemented the ARIMA technique and Kalman filter to forecast the weather using a spark framework and it was found that the ARIMA model with spark resulted in a significant reduction of execution time [7]. Spark was also used by Jayanthi and Sumathi to analyze weather datasets using ipython notebook and the proposed framework overcome the drawback of Hadoop in terms of processing speed [8]. Sudmant et al. studied the values and limitations of big data for fair weather prediction and emphasis on the need for technological advances for sustainable development [9].

In past, numerous authors have experimented with different architectures of neural networks to forecast weather conditions and the results were quite promising. Reddy and Babu surveyed various models for weather forecasting and it was found that Feed Forward Neural Networks (FFNN) and Time Delay Recurrent Neural Networks (TDNN) are well suited for short-term and long-term forecasting respectively [10]. Ren et al. too surveyed deep learning-based weather forecasting models and compared their performance with conventional NWP techniques [11].

Tran et al. reviewed neural networks for air temperature forecasting and found that Multi-Layer Perceptron (MLP), FFNN, Ward Style ANN, Radial Basis Function (RBF), Generalized Regression Neural Network (GRNN), etc. were widely used networks whereas few studies also used deep learning networks which provide encouraging results [12]. A comparison has been done between traditional and emerging techniques used in weather forecasting and it was concluded that for classification, CART, Adaboost and XGBoost gave the best results while for prediction, Linear Regression gave better performance [13].

Sharma and Ismail implemented 19 layers deep CNN using Keras and Tensorflow library for classifying the weather images into 4 classes (rain, shine, sunshine, and cloud), and the model provided 94% accuracy [14]. Wang and Li applied the fusion of DenseNet and ResNet CNN for classifying the weather images into nine classes and it was concluded that the output of the proposed integrated model was better than implementing the basic models alone [15]. Xiao et al. build a weather dataset comprising 6877 images of 11 different categories and proposed MeteCNN, a deep CNN model for classification [16]. The proposed model gave 92% accuracy and its performance was found to be superior to VGG19, VGG16, ResNet18, ResNet34, and MobileNet models.

Elhoseiny et al. studied the CNN layer's performance for two pre-trained models i.e. ImageNet-CNN and Weather-Trained CNN to classify weather images and the proposed work outperformed state-of-art methods [2]. Ibrahim et al. proposed a WeatherNet framework that combines both weather and visual conditions and comprises 4 parallel deep CNN models a) NightNet to classify dawn, night, day, and dusk b) GlareNet to classify Glare and No Glare c) PrecipitationNet to detect clear, rain and snow d) FogNet to detect fog and no fog [17]. Xia et al. introduced the ResNet-15 model to classify the "Weather-Dataset 4" dataset consisting of weather images into four classes and it was concluded that the proposed model performed superior to the traditional ResNet50 model [18].

Dhananjaya et al. discussed the application of weather classification for autonomous driving and build a dataset for the classification of weather, light level, and street type into nine classes [19]. An active learning framework has been implemented to reduce redundancy in the dataset and the ResNet50 network has been used for classification. Guerra et al. too build a weather dataset comprising 5500 weather images of five different categories and further implemented CNN for feature extraction using Matlab and Caffe library. The experimental results proved that all the variants of ResNet performed better than VGG and CaffeNet [3]. Matteussi et al.

implemented a Deep Learning algorithm on the BigDL framework using spark and Tensorflow library and the results concluded that ResNet performed better than VGG in terms of speed while LeNet achieved higher accuracy [20].

Zhao et al. too implemented the CNN-RNN approach for multi-label weather recognition to extract co-related visual features [21]. CNN model is extended with the attention model for feature extraction and the RNN model is used to model the dependency among labels. The experimental result proved that the proposed model is better than VGGNet. Li et al. combined CNN features with other features such as haze, contrast, brightness, etc. to construct a high dimensional vector for multi-class weather classification and the proposed model is proved to be effective and inspiring [22].

Zhang and Ma proposed Multiple Kernel Learning (MKL) for multi-class weather image classification and the results proved that the proposed technique gave better results as compared to SVM and AdaBoost [23]. Kang et al. presented a framework for weather image recognition to classify weather images into three categories: hazy, snowy, and rainy. The performance of GoogleNet, Alexnet, and the Multiple Kernel-based Learning approach was compared and it was found that GoogleNet slightly outperformed Alexnet [24].

Chen et al. proposed a weather recognition system using MKL and active learning techniques for automatically labeling the weather images into sunny, cloudy, and overcast categories, and the proposed system achieved better performance [25]. Lu et al. proposed an approach using a collaborative learning strategy to classify a dataset of 10k images into two categories: sunny and cloudy [26]. The proposed approach has been compared with SVM, Adaboost, LLC, and ScSPM techniques and the results proved the efficiency of the proposed approach.

Chu et al. build a large-scale dataset to estimate weather from images and demonstrated the relationship between different visual and weather features [27]. Random Forest has been implemented to classify the images and the results were found to be satisfactory for sunny, cloudy, and snowy classes. For rainy and foggy, the model did not give good results. Li et al. constructed a decision tree-based SVM classifier for extracting features to automatically acquire weather conditions from images using C++ and OpenCV [28].

Wang et al. introduced a new method for multi-class weather classification which builds feature vectors from the combination of real-time weather factors and dark channel feature value derived from images [29]. The proposed model is implemented using the Adaboost classifier on 5k images and the results demonstrated that the model performed better. Transfer learning has been applied by Challa and Vaishnav using the pre-trained ResNet50 and DenseNet161 models to classify the weather images [30].

As per the review of the work done by researchers, it can be concluded that Multiple Kernel Learning and pre-trained Neural Networks such as AlexNet, ResNet, VGG, and GoogleNet have been widely used by researchers for the classification of weather images. Big Data frameworks such as Spark and Hadoop have also been explored to process big datasets.

III. CHALLENGED AND LIMITATIONS OF EXISTING APPROACHES

Many existing weather classification techniques make use of expensive sensors and huge manpower for classifying weather images. These methods rely on human observations, thus are more prone to errors and are very time-consuming. Numerical Weather Prediction (NWP) models are also being widely used for forecasting the weather conditions, but the techniques are quite expensive and rely on the power of supercomputers for data processing.

Recent work on weather classification includes classification of weather conditions from images. CNN has an advantage of automatically detecting the important features from images without any human intervention. For getting the optimized results, CNN models require huge amount of processing power and large datasets for training. Further, training weather images using CNN models also require tuning number of hyper-parameters such as number of convolutional and max-pooling layers, kernel size, regularization techniques, etc.

To overcome the above mentioned issues, transfer learning techniques are applied to transfer the knowledge learned from pre-trained deep CNN models to our weather image dataset. As transfer learning utilizes the features extracted from pre-trained CNN models, training time is considerably reduced. It too eliminates the need for tuning of number of hyper-parameters. Transfer learning has proved its potential in number of applications such as natural language processing, sentiment classification, text classification, spam email detection, video classification, drug efficacy classification, etc. [31].

Further, the size of weather images generated is also huge. Existing weather classification techniques exploiting the transfer learning techniques do not make use of big data technologies as well as lacks scalability. Thus, there is a need of weather classification framework which can scale on number of machines and can support big weather image datasets.

IV. PROPOSED METHODOLOGY

To overcome the limitations discussed in previous section, a weather image classification framework has been proposed and implemented using spark platform. Fig. 1 describes the proposed framework which comprises of three steps: image loading, feature learning and multi-class classification. Firstly, weather images are loaded into the spark via the readSchema() method of the ImageSchema library and a dataframe is created. Afterward, image data is given as input to the pre-trained deep CNN model to automatically learn hierarchical feature representation. At this stage, either we can freeze all the layers of the pre-trained models or some of its layers are unfreeze and fine-tuned to learn the desired features [32]. The final step is to use the learned features from the deep CNN models as input to the classifier model which classifies the outdoor weather image into its appropriate category.

In step 2, 5 deep CNN networks i.e. InceptionV3, Xception, ResNet50, VGG16, and VGG19 have been explored for feature learning and experiments have been conducted to determine the model which gave the highest accuracy. VGG16 and

VGG19 is 16 and 19 layered deep neural network whereas ResNet50 is 50 layered deep Convolutional neural networks. Inception V3 is also a very powerful neural network which is 42 layers deep and has an advantage of performing multiple different transformations. Xception is a 71 layers deep neural network and incorporates depth-wise separable convolutions. All the layers from the pre-trained CNN models have been frozen in this case study.

In the last stage of proposed framework, one of the two classifiers i.e. Logistic Regression and Random Forests has been implemented for classification of weather images. Logistic Regression is a supervised ML technique which uses a sigmoid activation function for solving classification problems. To reduce the over-fitting and generalization error, the Logistic Regression classifier is trained with elastic net regularization. Two hyper-parameters for Logistic Regression need to be optimized; alpha (α) which is used to assign the weight to both L1 and L2 penalty and lambda (λ), a regularization parameter that defines the balance between minimizing the training error and model complexity. If α is 0, the model follows L2 regularization while if α is 1, the model follows L1 regularization.

Random Forests is another popular supervised machine learning technique based on ensemble learning which combines the results of a large number of weak decision trees to form a robust classifier. As shown in Fig. 2, 10 different combinations of the networks are possible and the same has been implemented using the Sparkdl library. Extensive experiments have been performed to determine the CNN model and classifier which provided the maximum accuracy for weather image classification.
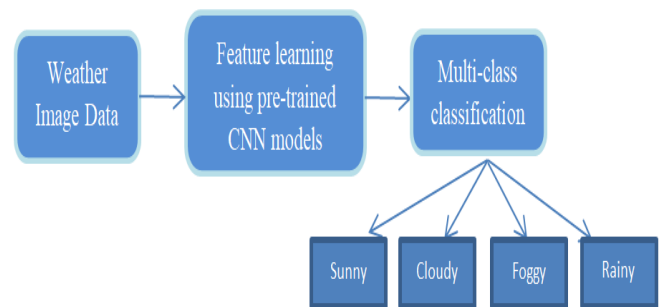

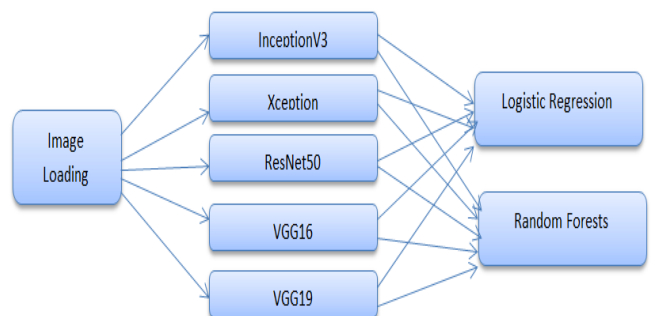Fig. 1. Proposed framework for weather image classification.

Fig. 2. Proposed architecture for weather classification.

V. EXPERIMENTAL SETUP

To evaluate the efficiency of the proposed framework, Multi class weather image classification datasets freely available on Kaggle.com has been used [33]. Dataset comprises 1125 images of four different categories i.e. cloudy, rainy, shiny, and sunrise. Sample images from each target class have been shown in Fig. 3. Categorization of target classes for the dataset has been presented in Fig. 4.
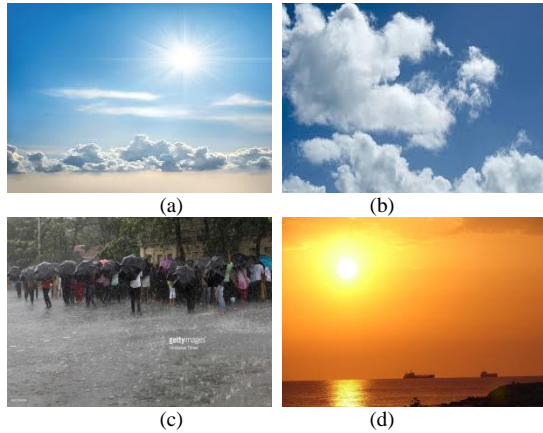

Fig. 3. Outdoor weather images (a) shine (b) cloudy (c) rainy (d) sunrise.

For the experimental setup, spark version 2.3.0 has been set up on a local PC (i.e. driver and the executor are set up on the same local machine) with 16GB RAM (12 GB RAM is allocated to spark) and 1 TB hard disk. Python version 3.7 and Java version 1.8 has been installed as shown in Table I. Further, there are more than 150 parameters in Spark which needs to be tuned for any spark application to achieve optimized results. There is no default value of a parameter that fits every application; values of these parameters vary with workload and the type of application.
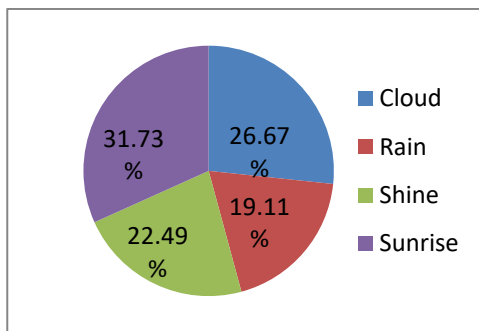

Fig. 4. Categorization of target class.

TABLE I. LOCAL PC CONFIGURATION

Parameter	Value
RAM	16GB
Hard Disk	1TB
OS	64-bit Windows 10
Spark version	2.3.0
Java	1.8
Python	3.7

Optimal tuning of spark parameters i.e. driver memory, driver cores, executor cores, executor memory, number of partitions etc. improves the performance of spark application to the greater extent [34,35,36,37]. As size of data is not very large, 4g memory is sufficient to store the data as well as to run the CNN models with Logistic Regression as classifier. While for Random Forest classifier, minimum 8g driver memory is required. Number of cores allocated to driver/executor too impacts the application's performance. If a cluster has n number of cores, number of partitions for input data should not be less than n as it might lead to under-utilization of resources. Few cores will be left idle and resources would not be utilized efficiently. Thus, number of input partition should either be equal or greater than number of cores. In this research work, number of cores and partitions has been set to 2 and 4, respectively.

To evaluate the efficiency of proposed framework, following metrics have been used:

1) Accuracy: It can be classified as the ratio of truly predicted samples to the total number of predictions.

2) Precision: It can be defined as the ratio of truly predicted positive values to the total number of positive predictions.

3) Recall: It can be defined as the ratio of true positives and total positive samples.

4) F1 score: Calculated as the harmonic mean of precision and recall.

5) Training Time: Time required for training the model.

VI. RESULTS AND DISCUSSION

For Logistic Regression classifier, impact of regularization parameter, alpha and lambda on model's accuracy has been studied. Accuracy values for different value of alpha have been observed and results are shown in Table II. It can be interpreted that accuracy of the Logistic Regression model trained using ResNet50 model is maximum when the value of alpha lies between 0 and 0.5.

TABLE II. IMPACT OF ALPHA ON MODEL'S ACCURACY

Alpha	Accuracy
0	97.28
0.1	97.28
0.2	97.28
0.3	96.8
0.4	96.8
0.5	97.28
0.6	96.8
0.7	96.38
0.8	95.92
0.9	95.92
1	96.28

Further, 5 fold Cross-Validation method has been used to determine the optimal lambda value of Logistic Regression.

Numerous values of λ i.e. (0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 1.0) have been examined and the results concluded that the model achieves higher accuracy for λ value of 0.01. Thus, for implementing the Logistic Regression classifier, α and λ has been set to 0.5 and 0.01, respectively.

For Random Forest classifier, the number of trees in forests impact performance of classifier to a greater extent. Thus, for different values of trees, its impact on the accuracy and run time of an algorithm has been studied and results are shown in Table III. It can be interpreted that till a certain value i.e. 150, the accuracy value of the model increases with an increase in the number of trees in the forests. Afterwards, the accuracy value decreases. Random Forest classifier provided the best accuracy value of 97.73% with 150 trees in the forests. It can also be observed that the increase in the number of trees has no significant impact on the training time of an algorithm.

Train and test data have been divided in the ratio 80:20. Table IV describes the accuracy, F1 score, precision, recall, and training time for 10 different possible architectures specified in Fig. 2. It can be interpreted that with the InceptionV3 pre-trained model and Logistic Regression classifier, the accuracy value is maximum i.e. 97.77%. Factorized convolutions, grid size reduction, asymmetric convolutions and use of auxiliary classifiers are some the significant features of InceptionV3 model resulting in its higher accuracy value. Instead of getting deeper, InceptionV3 extracts features in more depth.

TABLE III. IMPACT OF NUMBER OF TREES ON THE PERFORMANCE OF PROPOSED NETWORK

No. of trees	Accuracy	F1 score	Precision	Recall	Time
10	94.57	94.55	94.68	94.57	14 min 24 sec
50	96.39	96.38	96.4	96.38	16 min 3 sec
100	97.28	97.29	97.41	97.28	14 min 35 sec
125	97.29	97.29	97.32	97.28	13 min 58 sec
150	97.73	97.73	97.82	97.73	13 min 1 sec
200	96.38	96.4	96.49	96.38	14 min 5 sec

TABLE IV. OUTPUT OF PROPOSED FRAMEWORK FOR NUMEROUS PRE-TRAINED DEEP CNN MODELS

	Accuracy	F1 score	Precision	Recall	Time
ResNet50+LR	97.28	97.29	97.33	97.28	7 min 35 sec
VGG16+LR	94.57	94.61	94.83	95.57	10 min 12 sec
VGG19+LR	94.12	94.13	94.31	94.12	6 min 35 sec
InceptionV3+LR	97.77	97.74	97.75	97.74	6 min 42 sec
Xception+LR	95.02	95	95.11	95.02	8 min 2 sec
ResNet50+RF	97.73	97.73	97.82	97.73	13 min 1 sec
VGG16+RF	96.38	96.35	96.46	96.38	28 min 1 sec
VGG19+RF	93.66	93.64	93.62	93.66	26 min 45 sec
InceptionV3+RF	91.4	91.5	91.7	91.4	24 min 14 sec
Xception+RF	90.04	89.98	90.32	90.04	50 min 50 sec

Further, the second highest accuracy value i.e. 97.73% has been achieved with the ResNet50 pre-trained model and Random Forests classifier. ResNet50 pre-trained model with the Logistic Regression classifier too gave promising results. The Xception model trained using the Random Forests classifier provided the least accurate value of 90.04%. It can also be concluded that the time taken by the framework for training the dataset with the Logistic Regression classifier is quite low compared to the Random Forests.

Table V demonstrates the confusion matrix for weather image classification for dataset 1 with Inception V3 pre-trained CNN model and Logistic Regression classifier, where the row represents the actual labels and the column represents the predicted labels. As shown in Table III, the total number of test images is 221. The number of test images for classes cloudy, foggy, rain, shine, and sunshine is 45, 64, 47, and 65, respectively. Out of 221, 216 images are predicted true by the classifier.

TABLE V. CONFUSION MATRIX FOR INCEPTION V3 LOGISTIC REGRESSION MODEL

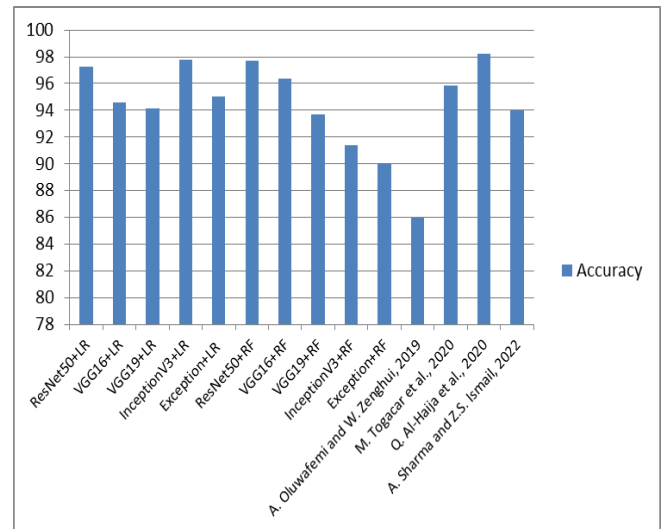|	Cloudy	Rainy	Shine	Sunrise	Total
Cloudy	44	0	1	0	45
Rainy	1	63	0	0	64
Shine	0	0	45	2	47
Sunrise	0	0	1	64	65
Total	45	63	47	66	221

Fig. 5. Comparison of the proposed framework with the existing techniques.

The performance of proposed framework has been compared with the existing techniques utilizing the same weather image dataset for classification. Fig. 5 shows the comparison of accuracy value of the proposed framework with the four existing approaches. Our proposed framework yields the maximum accuracy value of 97.77% when Inception V3 model and Logistic Regression classifier (InceptionV3+LR) has been used. Models proposed by Oluwafemi and Zenghui (2019), Togacar et al. (2020), and Sharma and Ismail (2022) had lower accuracy values of 86%, 95.85% and 94%

respectively [14,38,39]. Model proposed by Al-Haija et al. (2020) had the accuracy value of 98.22%, thus performed slightly better than the proposed InceptionV3+LR model in terms of accuracy [40]. Further, Model proposed by Al-Haija et al. had precision and sensitivity values of 96.5% and 96.4%, respectively while the proposed InceptionV3+LR model has better precision and sensitivity values i.e. 97.75% and 97.74% respectively. Hence, InceptionV3+LR model has better ability to capture false negatives and true positives than existing techniques. Time taken by the proposed framework for training is also very less i.e. 6 min 42 sec.

VII. CONCLUSION

Weather Prediction is a very vital activity performed before numerous day-to-day activities. Deep CNN is a deep learning technique widely used to classify image datasets but the techniques require very huge datasets and high computing devices for achieving optimal results. Transfer learning is proven to be effective in scenarios with smaller datasets and the unavailability of high computing devices. It helped in minimizing the learning time by utilizing the features learned from the previously trained deep CNN models.

In this paper, a framework based on transfer learning has been proposed for classifying the weather conditions from outdoor images and the performance of various CNN models has been compared for accuracy, precision, recall, f1 score, and training time. From the results, it can be observed that with the InceptionV3 pre-trained model and Logistic Regression classifier, the accuracy value is maximum i.e. 97.77%. Further, it can be concluded that the time taken by the framework for training the dataset with the Logistic Regression classifier is quite low compared to the Random Forests. The comparison has also been done with the existing techniques and the results proved the efficiency of the proposed framework. Further, the proposed framework is scalable and has support for big datasets. In the future, some layers from pre-trained CNN models can be un-freeze and fine-tuned to learn better hierarchical representation.

REFERENCES

[1] B. Zhao, X. Li, X. Lu, and Z. Wang, "A CNN-RNN architecture for multi-label weather recognition," Computer Science, Computer Vision and Pattern Recognition, 2019.

[2] M. Elhoseiny, S. Huang, and A. Elgammal, "Weather classification with deep convolutional neural networks," International Conference on Image Processing (ICIP), IEEE, Canada, 2015.

[3] J. C. Villarreal Guerra, Z. Khanam, S. Ehsan, R. Stolkin, and K. McDonald-Maier, "Weather classification: A new multi-class dataset, data augmentation approach and comprehensive evaluations of convolutional neural networks," NASA/ESA Conference on Adaptive Hardware and Systems (AHS), pp. 305-310, 2018.

[4] M. Alam and M. Ajmad, "Weather forecasting using parallel and distributed analytics approaches on big data clouds," Journal of Statistics and Management Systems, vol. 12 no. 4, pp. 791-799, 2020.

[5] N. Yang, L. Westfall, and P. Dalvi, "A weather prediction model with big data," Proceedings of Student-Faculty Research Day, CSIS, 2018.

[6] K. A. Ismail, M.A. Majid, J. Mohamed Zain, and N.A. Abu Bakar, "Big data prediction framework for weather temperature based on MapReduce algorithm," IEEE Conference on Open Systems, 2016.

[7] R. Dhoot, S. Agrawal, S. Kumar, and M. Sivagami, "Implementation and analysis of Arima model and kalman filter for weather forecasting in spark computing environment," 3rd International Conference on Computing and Communications Technologies (ICCCT), IEEE, 2019.

[8] D. Jayanthi and G. Sumathi, "Weather data analysis using spark – An In-memory computing framework," International Conference on Innovations in Power and Advanced Computing Technologies [i-PACT2017], IEEE, 2017.

[9] A. Sudmant, V. Viguie, Q. Lepetit, L. Oates, A. Datey, A. Gouldson, and D. Watling, "Fair weather forecasting? The shortcomings of big data for sustainable development, a case study from Hubballi-Dharwad, India," Sustainable Development, vol. 24, no. 6, pp. 1237-1248, 2021.

[10] P.C. Reddy and A.S. Babu, "Survey on weather prediction using big data analytics," Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, 2017.

[11] X. Ren, X. Li, K. Ren, J. Song, Z. Xu, K. Deng, and X. Wang, "Deep learning-based weather prediction: A survey," Big Data Research, Elsevier, vol. 13, 2020.

[12] T.T.K. Tran, S.M. Bateni, , S.J. Ki, and H. Vosoughifar, "A review of neural networks for air temperature forecasting," Water, vol. 13, no. 9, 2021.

[13] I. Gad and D. Hosahalli, "A comparative study of prediction and classification models on NCDC weather data," International Journal of Computers and Applications, vol. 44 no. 5, pp. 414-425, 2022.

[14] A. Sharma and Z.S. Ismail, "Weather classification model performance: Using CNN, Keras-TensorFlow," 1st International Conference on Applied Computing & Smart Cities (ICACS21), vol. 52, 2022.

[15] Y. Wang and Y. Li, "Research on multi-class weather classification algorithm based on multi-model fusion," IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 2251-2255, 2020.

[16] H. Xiao, F. Zhang, Z. Shen, K., Wu, and J. Zhang, " Classification of weather phenomenon from images by using deep convolutional neural network", Earth and Space Science," vol. 8, 2021.

[17] M.R. Ibrahim, J. Haworth, and T. Cheng, "WeatherNet: Recognizing weather and visual conditions from street-level images using deep residual learning," Computer Science, Computer Vision and Pattern Recognition, 2019.

[18] J. Xia, D. Xuan, L. Tan, and L. Xing, "ResNet15: Weather recognition on traffic road with deep Convolutional neural network," Advances in Meteorology, Hindwai, 2020.

[19] M.M. Dhananjaya, V.R. Kumar, and S. Yogamani, "Weather and light level classification for autonomous driving: Dataset, baseline and active learning," Computer Science, Computer Vision and Pattern Recognition, 2019.

[20] K.J. Matteussi, B.F Zanchetta, G. Bertoncello, J. Santos, J. Anjos, and C. Geyer, "Analysis and performance evaluation of deep learning on big data," IEEE Symposium on Computers and Communications, Spain, 2020.

[21] X. Zhao, S. Qi, B. Zhang, H. Ma, W. Qian, Y. Yao, and J. Sun, "Deep CNN models for pulmonary nodule classification: model modification, model integration, and transfer learning," Journal of X-Ray Science and Technology, vol. 27, no. 4, pp. 615-629, 2019.

[22] Z. Li, Y. Li, J. Zhong, and Y. Chen, "Multi-class weather classification based on multi-feature weighted fusion method," 2nd International Conference on Oil & Gas Engineering and Geological Sciences, IOP Conf. Series: Earth and Environmental Science, 2020.

[23] Z. Zhang and H. Ma, "Multi-class weather classification on single images," International Conference on Image Processing (ICIP), IEEE, pp. 4396-4400, 2015.

[24] L. Kang, K. Chou, and R. Fu, "Deep learning based weather image recognition," International Symposium -on Computer, Consumer and Control (IS3C), pp. 384-387, 2018.

[25] Z. Chen, F. Yang, A. Lindner, G. Barrenetxea, and M. Vetterli, "How is the weather: automatic inference from images," 19th IEEE International Conference on Image Processing, pp. 1853-1856, 2012.

[26] C. Lu, D. Lin, J. Jia, and C.K. Tang, " Two-class weather classification," IEEE Conference on Computer Vision and Pattern Recognition, pp. 3718-3725, 2014.

[27] W.T. Chu, X.Y. Zheng, and D.S. Ding, "Image2Weather: A large-scale image dataset for weather property estimation," Second International Conference on Multimedia Big Data (BigMM), IEEE, pp. 137-144, 2016.

[28] Q. Li, Y. Kong, and S. Xia, "A method of weather recognition based on outdoor images," International Conference on Computer Vision Theory and Applications (VISAPP), pp. 510-516, 2014.

[29] S. Wang, Y. Li, and W. Liu, "Multi-class weather classification fusing weather dataset and image features," In: Z. Xu, X. Gao, Q. Miao, Y. Zhang, J. Bu (eds) Big Data, Big Data, Communications in Computer and Information Science, vol. 945, Springer, Singapore, 2018.

[30] S.V.D.M. Challa and H.K. Vaishnav, "Weather categorization using foreground subtraction and deep transfer learning," In: H. Saini, R. Sayal, R. Buyya, G. Aliseri (eds), Innovations in Computer Science and Engineering, Lecture Notes in Networks and Systems, vol. 103, Springer, Singapore, 2020.

[31] K. Weiss, T.M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," Journal of Big Data, vol. 3, 2016.

[32] R. Ribani and M. Marengoni, "A survey of transfer learning for Convolutional Neural Networks." 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), Brazil, pp. 47-57, 2019.

[33] G. Ajayi, "Multi-class weather dataset for image classification," Mendeley Data, V1, 2018.

[34] M. Rahman, J. Hossen, and C. Venkataseshaiah, "SMBSP: A Self-Tuning Approach using Machine Learning to Improve Performance of Spark in Big Data Processing," 7th International Conference on Computer and Communication Engineering (ICCCE), IEEE, Malaysia, Sep. 2018.

[35] P.N. Saipraveen and G.S. Nagaraja, "Parameter Tuning of Apache Spark based Applications for Performance Enhancement," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 5, May 2020.

[36] Q. Chen, K. Wang, Z. Bian, I. Cremer, G. Xu, and Y. Guo, "Simulating Spark Cluster for Deployment Planning, Evaluation and Optimization," 6th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH), IEEE, Portugal, Jul. 2016.

[37] H. Zhang, Z Liu, and L.Wang, "Tuning Performance of Spark Programs," International Conference on Cloud Engineering (IC2E), IEEE, USA, Apr. 2018.

[38] A.G. Oluwafemi and W. Zenghui, "Multi-Class Weather Classification from Still Image Using Said Ensemble Method," Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), IEEE, South Africa, 2019.

[39] M. Togaçar, B. Ergen, and Z. Comert, "Detection of weather images by using spiking neural networks of deep learning models," Neural Computing and Applications, vol. 33, pp. 6147–6159, 2020.

[40] Q.A. Al-Haija, A. Smadi, and S. Zein-Sabatto, "Multi-Class weather classification using ResNet-18 CNN for autonomous IoT and CPS applications," International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, USA, 2020.