
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152

www.ijmece .com

Vol 12, Issue 2, 2024

A Machine Learning Approach for Rainfall Estimation Integrating Heterogeneous Data Sources

[1]SMT.S ARUNA, [2]LANKA GANESH SAI

[1](Assistant Professor), MCA, Swarnandhra College

[23]MCA, scholar, Swarnandhra College

ABSTRACT

Providing an accurate rainfall estimate at individual points is a challenging problem in order to mitigate risks derived from severe rainfall events, such as floods and landslides. Dense networks of sensors, named rain gauges (RGs), are typically used to obtain direct measurements of precipitation intensity in these points. These measurements are usually interpolated by using spatial interpolation methods for estimating the precipitation field over the entire area of interest. However, these methods are computationally expensive, and to improve the estimation of the variable of interest in unknown points, it is necessary to integrate further information. To overcome these issues, this work proposes a machine learning-based methodology that exploits a classifier based on ensemble methods for rainfall estimation and is able to integrate information from different remote sensing measurements. The proposed approach supplies an accurate estimate of the rainfall where RGs are not available, permits the integration of heterogeneous data sources exploiting both the high quantitative precision of RGs and the spatial pattern recognition ensured by radars and satellites, and is computationally less expensive than the interpolation methods. Experimental results, conducted on real data concerning an Italian region, Calabria, show a significant improvement in comparison with Kriging with external drift (KED), a well-recognized method in the field of rainfall estimation, both in terms of the probability of detection (0.58 versus 0.48) and mean-square error (0.11 versus 0.15).

1.INTRODUCTION

Protecting against flood dangers, managing river basins, and simulating erosion are all uses for hydrological impact modeling that rely on accurate rainfall estimates. In order to achieve this goal, rain gauges (RGs) are used to directly monitor the length and intensity of precipitations at specific locations.

We employ interpolation algorithms that are calculated using the data recorded by these RGs to estimate rainfall occurrences in regions that are not covered by them. There are a lot of different versions of these procedures that have been suggested in the literature. One of the most well-known and often utilized is the Kriging geo statistics method [1, 2].

When dealing with severe convective weather occurrences, it is crucial to have an accurate spatial reconstruction of the rainfall field. For example, sparse RGs may miss significant convective precipitation that forms in certain areas, and floods might occur even when no precipitation has fallen [3]. To get around this problem, there has been a recent uptick in the use of interpolation algorithms to combine different types of rainfall data in order to provide a more precise estimate [4].

Ordinary Kriging (OK), which is widely used, has the drawback of only being able to utilize data from one source as input. To get around this, Kriging with external drift (KED) became a popular method [5, 6]. In fact, KED permits the interpolation of a random field and, in contrast to OK, may account for secondary data. The most significant issue is the high computational cost and extensive resource requirements of these approaches.

Machine learning (ML) methods are the basis of an alternative strategy. Class imbalance, a significant number of missing characteristics, and the need to work gradually as new data become available are some of the difficult difficulties that must be overcome while employing these approaches. These problems are usually addressed by using ensemble techniques. As a classification approach, ensemble [7] combines several models trained with separate data sets or techniques to assign labels to previously unknown occurrences. The ensemble paradigm allows for the management of imbalanced classes and the reduction of error variance and bias, as compared to the situation of utilizing a single classification model. In particular, problems with rainfall estimates and strong weather event monitoring may be better handled using ensemble-based approaches. These techniques may also detect nonlinear relationships, such as those between sensor readings, cloud characteristics, and precipitation prediction. This paper introduces a machine learning (ML) approach to rainfall estimate using a hierarchical probabilistic ensemble classifier (HPEC) to tackle the primary challenges in this area. In cases when RGs are unavailable, the suggested method allows for precise rainfall estimate by combining data from many sources (e.g., RGs, radars, and satellites) and using an under sampling methodology to deal with the imbalanced classes issue that is prevalent in such cases.

For example, our method works well in practice when an official from the Department of Civil Protection (DCP) has to assess the likelihood of landslides and floods in a certain area due to heavy rainfall. Actual data pertaining to the southern Italian area of Calabria, supplied by the DCP, is used in the experimental assessment. Due to its complicated orography and striking climatic variations, Calabria serves as an excellent test site.

Here is a summary of what we have contributed.

1. Improved rainfall event estimations are produced by integrating three diverse data sources, namely RGs, radar, and Meteosat.
2. Several categorization approaches are evaluated using a real-world example involving the southern Italian area of Calabria, and a hierarchical probabilistic ensemble method is suggested.
3. Various ML-based approaches are evaluated using KED, a popular interpolation method in the hydrological sector, and pre-trained only on historical data.

Here is how the remainder of the article is structured. In Section II, we compare and contrast our method with others in the field and highlight the key points of difference. In Section III, we see the case study in action and learn about the primary data sources that went into developing the framework. The procedure for estimating the precipitation is detailed in Section IV. The findings and explanation of the experiments are presented in Section V.

2.LITERATURE SURVEY

Creating Predictive Models for the Indian Summer Monsoon Rainfall: How Topological Pattern Discovery and Support Vector Machines Fit In order to visualize the pattern of clustering behavior of yearly rainfall due to changes in monthly rainfall for each year, this study applies a growing hierarchical self-organizing map (GHSOM) to a dataset of Indian rainfall data that spans 142 years. The dataset is then used to group the yearly rainfall into smaller units. Additionally, it has been noted using support vector machine (SVM) that cluster creation has a beneficial effect on the rainfall forecast for the Indian summer monsoon. Statistical and graphical analysis have been used to display the results.

A model for predicting rainfall using artificial neural networks is titled as such.

Because it is so easy to train, the multilayer artificial neural network setup using the backpropagation algorithm is the most popular. Around eighty percent of all neural network development efforts use back-propagation. The learning cycle of a back-propagation algorithm consists of two stages: the first involves spreading the input patterns over the network, and the second involves adjusting the output by adjusting the network's weights. Character recognition, financial and weather prediction, face identification, and many more applications are all within the realm of possibility for the back-propagation-feedforward neural network.In order to put one of these applications into action, the study constructs testing and training data sets and determines the optimal number of hidden neurons for each layer. This study used artificial neural network models to examine the feasibility of forecasting the average rainfall across the Udupi district in Karnataka. Predictive models based on artificial neural networks have three layers of construction. In terms of hidden neurons, the models that are being compared vary.

Title: A Rainfall Prediction Model for the Near Future Employing Convolutional Neural Networks with Multiple Tasks

One of the most pressing issues in meteorological service is the forecast of precipitation, especially in the near future. Utilizing radar data or satellite photos to generate forecasts has been the primary emphasis of most current research. One other situation, however, involves the collection of a set of weather parameters from a number of different sensors installed at different observation locations. Even while site observations aren't always thorough, they can give useful information for weather prediction at surrounding sites, which hasn't been completely used in previous research. In order to address this issue, we provide a model of a multi-task convolutional neural network that can automatically acquire features from time series recorded at observation sites and use the correlation between these sites to their advantage in weather prediction. So far as we are aware, this is the first effort to forecast the quantity of rainfall in the near future using multi-site characteristics and deep learning approaches. More specifically, we model the correlations across sites and construct the learning challenge as an end-to-end multi-site neural network model. This enables us to apply the gained information from one site to other linked sites. Results from a battery of tests demonstrate that the suggested model beats a variety of baseline models, including ECMWF, and that the learnt site correlations are quite informative.

The Use of Deep Learning Models in Rainfall Prediction

One of the most important ways that all living things get their freshwater supply is via rainfall. Data on the impact of different climatic factors on precipitation totals is supplied by rainfall prediction models. A data-driven model for a time series dataset may now be created thanks to Deep Learning's self-learning data labels.

In addition to making predictions about future occurrences based on previous events, it can identify anomalies or changes in time series data. This paper discusses the use of Deep Learning Architectures (LSTM and ConvNet) to obtain models of rainfall precipitation. It determines which architecture is better, with LSTM having an RMSE of 2.55 and ConvNet having an RMSE of 2.44. The authors assert that Deep Learning models will be efficient and effective for modellers working with any time series dataset.

3. EXISTING SYSTEM

Despite the fact that our study aims to construct a run-off analysis, it is comparable to previous work that uses a probabilistic ensemble and blends two data sources (rain gauges and radar) and is based on the ensemble paradigm (e.g., [12]). A single runoff hydrograph is then determined by applying a blending approach to the outputs of the runoff hydrologic models. The experimental findings prove that the hydrologic models are reliable, which may lead to better flood warning decision-making. In order to get a probabilistic geographical analysis of the daily precipitation using rain gauges, Frei and Isotta [13] outline a method.

In the end, the model is just a Bayesian predictive distribution that measures the uncertainty in the data sampled from the station network; it stands for an ensemble of potential fields that depend on the observations. The approach's capacity to provide accurate forecasts for a hydrological partitioning of the area is shown by an examination of an actual case study situated in the European Alps. Using just high-resolution gauges, the authors of [14] suggest an intriguing investigation into the daily precipitations for Australia and a number of South and East Asian locations. By averaging the results of the investigations performed on each source, the chosen model may be determined. When comparing the global accuracy of the model's individual components, the authors stress that the ensemble method is superior. Extra data from other precipitation products may also be captured by the suggested model. These final two studies demonstrate that ensemble approaches may guarantee excellent outcomes in a rainfall estimate scenario by using an ensemble strategy to provide more precise forecasts. In contrast to our work, however, the combination tactics that have been used are quite basic, and the integration of diverse data sources is completely disregarded.

In their study, Chiaravalloti et al. [16] used RG-only data and the combined RG-radar product as benchmarks to examine the performance of three newly created satellite-based products: IMERG, SM2RASC, and an ingenious hybrid of the two. A better quality satellite rainfall product is obtained by combining IMERG with SM2RASC, and experiments show that IMERG performs well at temporal resolutions greater than 6 hours. The majority of alternative methods combine information from many sources, such as radars and satellite channels. Some of these rely on finding the right models to use the data to determine the parameters, which in turn take advantage of the relationship between clouds' optical and microphysical characteristics [17], [18]. Using statistical methods, further research [19–21] identify the models. Precipitation estimates derived from multispectral satellite data are given via Bayesian estimation in [22], whereas reference values are given by techniques that take radar data into consideration.

In their approach, which incorporates RG observations and satellite data and uses an interpolation methodology based on the

Kriging method, Verdin et al. [23] also use Bayesian estimation to estimate the model parameters. While each of these methods has the potential to provide intriguing outcomes, their flexibility and efficacy are often compromised due to the sensitive process of parameters estimate for each individual model. More adaptable methods grounded on ML techniques have lately been explored because to the very nonlinear nature of the relationships between sensor data, cloud features, and rainfall estimates. For example, in [24], the authors use ANNs in conjunction with support vector machines to solve the challenge of identifying convective occurrences and adjacent wet regions. The data sets are derived from the optical channels of the multispectral sensor on board the Meteosat Second Generation (MSG) satellites. In contrast to our work, RG measures are used primarily as a reference and not during the algorithm's training phase. In their proposal for an SVM-based method of rainfall estimate, Sehad et al. [25] combine input data from multispectral channels on MSG and create two models, one for the day and one for the night.

Only random forests (RFs) and RGs are used to verify the strategy, and the results are compared to comparable ANN-based methods. In another ANN-based method, detailed in [26], radar data are used as a reference for recognizing pixels that are wet, with the input being an image matrix. Using data from multispectral channels on MSG satellites, Kuhnlein et al. [27] estimate rainfall rates using RFs, and they also use the ensemble paradigm.

Disadvantages

- The hierarchical probabilistic ensemble classifier (HPEC) is not used in the system to forecast rainfall.
- The system uses ANNs, or artificial neural networks, to make predictions, however these predictions are inaccurate.

3.1 PROPOSED SYSTEM

For example, our method works well in practice when an official from the Department of Civil Protection (DCP) has to assess the likelihood of landslides and floods in a certain area due to heavy rainfall. The experimental assessment is carried out using actual data given by the DCP and pertaining to the area of Calabria in southern Italy. The complicated orography and considerable climatic fluctuation of Calabria make it an ideal testing site. The following is an

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

overview of our contributions.
1) To improve the accuracy of rainfall event estimations, three different data sources are combined: RGs, radar, and Meteosat.
2) A hierarchical probabilistic ensemble strategy is suggested after comparing several categorization algorithms on a real-life situation involving Calabria, a southern area in Italy.

3) Using a popular interpolation approach in the hydrological area (KED), we evaluate several ML-based algorithms that have only been trained on historical data.

The Benefits

In order to analyze the raw data in the suggested system, it is preprocessed. To solve the issue of class imbalance, an undersampling method is used. By combining RG, satellite, and radar data, the suggested system was able to train and test using efficient ML classifiers and provide an impact.

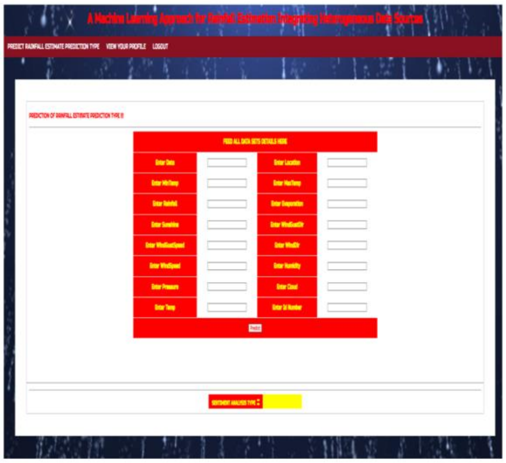
4. OUTPUT SCREENS

Registration

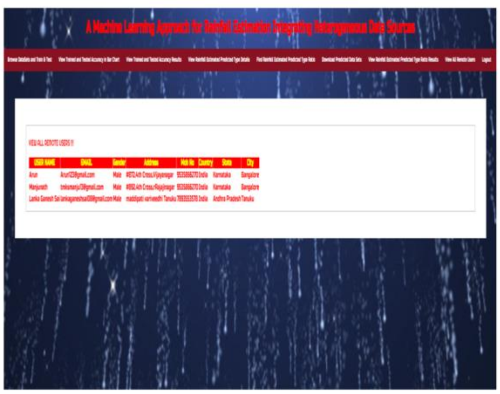
Login page

Upload Data

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

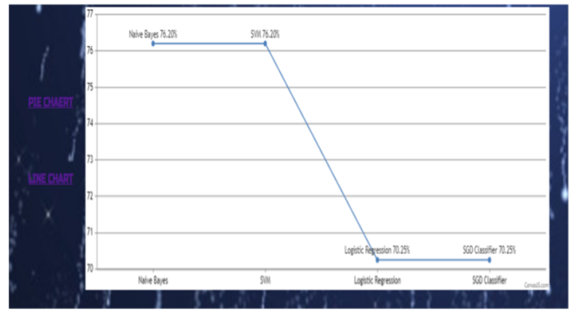


Service_provider

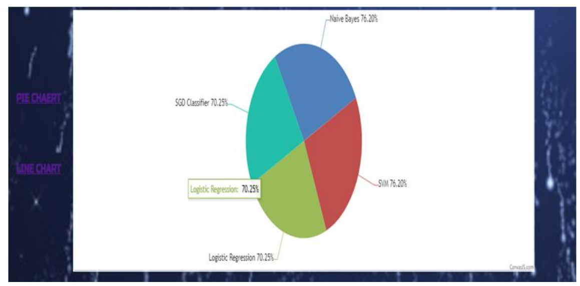


Accuracy

Line chart



Pie chart



5. CONCLUSION

The spatial rainfall field estimating method based on ML has been developed. In situations when RGs are not available, this approach may estimate rainfall by combining data from diverse sources like radars and satellites, which also take use of the spatial pattern recognition provided by radars and satellites. An HPEC enables the model used to predict the intensity of rainfall events after a preprocessing phase, after which a random uniform under sampling technique is applied. In the first stage, RF classifiers are trained as the basis of this ensemble. The second stage involves using a probabilistic metal earner to combine the estimated probabilities given by the base classifiers in accordance with a stacking schema. As compared to Kriging with external drift, a widely used and well-known approach in rainfall estimate, experimental findings performed on actual data given by the Department of Civil Protection demonstrate substantial improvements. The ensemble approach stands out for its superior ability to identify rainfall occurrences. The HPEC-obtained POD value of 0.58 and the MSE value of 0.11 are statistically superior than the KED-obtained values of 0.48 and0.15, respectively. When it comes to the latter two classes, which stand for strong rainfall events, HPEC is computationally more efficient than the Kriging approach, but the difference between the two is not statistically significant (based on the F-measure).

Because the Kriging method's complexity is cubic in the number of samples [51], it becomes computationally difficult to examine a large number of points, making the operation expensive from a computational standpoint. The ML algorithms (RF included) are quite sophisticated, however. One further advantage is that ensemble techniques may be easily parallelized and scaled up. We conclude that our technique offers significant benefits in this domain. Further, it is clear that all data sources contribute to the technique's strong performance when the impact of integration of the diverse data sources is analyzed. On all metrics, the algorithm's performance drops significantly once RG data is removed. In situations when one of the other two forms of data is removed, the deterioration is not as noticeable. However, the MSE is at its lowest (0.11) when all data are used, proving that all data sources must be used for optimal outcomes.

We want to evaluate the technique on a longer time period in future work to account for impacts caused by annual and seasonal variability, and we are also thinking about ways to gradually develop the flexible ensemble model using the additional data. Additionally, we want to assess the algorithm's performance in identifying locally concentrated heavy precipitation events, using time series analysis to dissect the distinct contributions of the radar and Meteosat characteristics

6.REFERENCES

[1] J. E. Ball and K. C. Luk, "Modeling spatial variability of rainfall over a

catchment," *J. Hydrologic Eng.*, vol. 3, no. 2, pp. 122–130, Apr. 1998.
[2] S. Ly, C. Charles, and A. Degré, "Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale. a review," *Biotechnologie, Agronomie, Société et Environnement*, vol. 17, no. 2, p. 392, 2013.
[3] H. S. Wheater *et al.*, "Spatial-temporal rainfall fields: Modelling and statistical aspects," *Hydrol. Earth Syst. Sci.*, vol. 4, no. 4, pp. 581–601, Dec. 2000.
[4] J. L. McKee and A. D. Binns, "A review of gauge–radar merging methods for quantitative precipitation estimation in hydrology," *Can. Water Resour. J./Revue Canadienne des Ressources Hydriques*, vol. 41, nos. 1–2, pp. 186–203, 2016.
[5] F. Cecinati, O. Wani, and M. A. Rico-Ramirez, "Comparing approaches to deal with non-gaussianity of rainfall data in Kriging-based radargauge rainfall merging," *Water Resour. Res.*, vol. 53, no. 11, pp. 8999–9018, Nov. 2017. [6] H. Wackernagel, *Multivariate Geostatistics: An Introduction With Applications*. Berlin, Germany: Springer, 2003.
[7] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
[8] B. J. E. Schroeter, *Artificial Neural Networks in Precipitation Now-Casting: An Australian Case Study*. Cham, Switzerland: Springer, 2016, pp. 325–339.
[9] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst.*, vol. 1, Dec. 2015, pp. 802–810.
[10] W.-C. Hong, "Rainfall forecasting by technological machine learning models," *Appl. Math. Comput.*, vol. 200, no. 1, pp. 41–57, Jun. 2008.
[11] A. Parmar, K. Mistree, and M. Sompura, "Machine learning techniques for rainfall prediction: A review," in *Proc. 4th Int. Conf. Innov. Inf. Embedded Commun. Syst. (ICIIECS)*, Mar. 2017, pp. 152–162.
[12] M. Lee, N. Kang, H. Joo, H. Kim, S. Kim, and J. Lee, "Hydrological modeling approach using radar-rainfall ensemble and multi-runoff-model

blending technique," *Water*, vol. 11, no. 4, pp. 1–18, 2019.

[13] C. Frei and F. A. Isotta, "Ensemble spatial precipitation analysis from rain gauge data: Methodology and application in the European alps," *J. Geophys. Res., Atmos.*, vol. 124, no. 11, pp. 5757–5778, Jun. 2019.

[14] J. L. Peña-Arancibia, A. I. J. M. van Dijk, L. J. Renzullo, and M. Mulligan, "Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method for regions in Australia and south and East Asia," *J. Hydrometeorology*, vol. 14, no. 4, pp. 1323–1333, Aug. 2013.

[15] V. Levizzani, P. Bauer, and F. J. Turk, *Measuring precipitation from Space: EURAINSAT and the Future*, vol. 28. Berlin, Germany: Springer, 2007.

[16] F. Chiaravalloti, L. Brocca, A. Procopio, C. Massari, and S. Gabriele, "Assessment of GPM and SM2RAIN-ASCAT rainfall products over complex terrain in southern italy," *Atmos. Res.*, vol. 206, pp. 64–74, Jul. 2018.

[17] B. Thies, T. Nauss, and J. Bendix, "Discriminating raining from nonraining cloud areas at mid-latitudes using Meteosat Second Generation SEVIRI night-time data," *Meteorolog. Appl.*, vol. 15, no. 2, pp. 219–230, 2008.

[18] M. Lazri, Z. Ameur, S. Ameur, Y. Mohia, J. M. Brucker, and J. Testud, "Rainfall estimation over a mediterranean region using a method based on various spectral parameters of SEVIRI-MSG," *Adv. Space Res.*, vol. 52, no. 8, pp. 1450–1466, Oct. 2013.

[19] H. Feidas and A. Giannakos, "Identifying precipitating clouds in greece using multispectral infrared Meteosat Second Generation satellite data," *Theor. Appl. Climatol.*, vol. 104, nos. 1–2, pp. 25–42, May 2011.

[20] H. Feidas, G. Kokolatos, A. Negri, M. Manyin, N. Chrysoulakis, and Y. Kamarianakis, "Validation of an infrared-based satellite algorithm to estimate accumulated rainfall over the mediterranean basin," *Theor. Appl. Climatol.*, vol. 95, nos. 1–2, pp. 91–109, Jan. 2009.

[21] A. Giannakos and H. Feidas, "Classification of convective and stratiform

rain based on the spectral and textural features of meteosat second
Generation infrared data," *Theor. Appl. Climatol.*, vol. 113, nos. 3–4,
pp. 495–510, Aug. 2013.

[22] M. Grecu and W. S. Olson, "Bayesian estimation of precipitation
from satellite passive microwave observations using combined radar–
radiometer retrievals," *J. Appl. Meteorol. Climatol.*, vol. 45, no. 3,
pp. 416–433, Mar. 2006.

[23] A. Verdin, B. Rajagopalan, W. Kleiber, and C. Funk, "A Bayesian
Kriging approach for blending satellite and ground precipitation observations,"
Water Resour. Res., vol. 51, no. 2, pp. 908–921, Feb. 2015.

[24] M. A. Tebbi and B. Haddad, "Artificial intelligence systems for rainy
areas detection and convective cells' delineation for the south shore
of Mediterranean Sea during day and nighttime using MSG satellite
images," *Atmos. Res.*, vols. 178–179, pp. 380–392, Sep. 2016.

[25] M. Sehad, M. Lazri, and S. Ameur, "Novel SVM-based technique to
improve rainfall estimation over the mediterranean region (North of
Algeria) using the multispectral MSG SEVIRI imagery," *Adv. Space
Res.*, vol. 59, no. 5, pp. 1381–1394, Mar. 2017.