International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 12 ǁ December 2022 ǁ PP. 110-115

Crowdsourcing Approaches for Weather Nowcasting: A Review

Riya Gori, Priyanka Hotchandani, Bipasha Paul Shukla
[1] Trainee, SAC-SMART ISRO, Ahmedabad, Department of Computer Science, K.J. Somaiya College of Engineering
[2] Trainee, SAC-SMART ISRO, Ahmedabad, Department of Computer Science, K.J. Somaiya College of Engineering
[3] Space Applications Centre ISRO, Ahmedabad, India

Abstract

Smart gadgets, social media/web 2.0, amateur weather stations and sensors, and citizen scientists may all provide information on the condition of the atmosphere. People are both data producers and consumers, hence crowdsourcing is essential in highly populated areas, data-deficient regions, or nations where established meteorological networks are dwindling. It may be able to solve problems with observations' temporal and geographical representativeness. However, there are still questions about how and where to crowdsource data from, how to evaluate the quality of this data, and how precisely the scientific and social applications of this data will be implemented. Weather Nowcasting or very short-range forecasting is a strong potential candidate for crowdsourcing applications. This paper performs a review of the various approaches for crowdsourcing and weather nowcasting, and explores them.

Keywords: *crowdsourcing, nowcasting, citizen science, prediction, machine learning, radar, satellite data*

Date of Submission: 01-12-2022 Date of acceptance: 10-12-2022

I. INTRODUCTION

The proper use of nowcasting tools and heeding of warnings might greatly aid in the improvement of preventive measures and the lowering of disaster-related losses. One of the new approaches which can enhance the capabilities is crowdsourcing the weather data. The needs of the general public require a wider range of nowcasting goods than the needs of specialists. A significant and growing quantity of data is currently being gathered from a wide range of non- traditional sources, including smartphone sensors, home weather stations, and public surveys.

In the past, crowdsourcing has relied on a dispersed network of independent contributors to address a specific issue. Various crisis response systems used a variety of data sources, some of which are crowdsourced or obtained through social media. Crowdsourcing now includes dispersed networks of portable sensors that may be activated and maintained via an open call for participation [1].

The World Meteorological Organization (WMO) keeps track of extreme weather events that annually claim thousands of lives and monitors the state of the world's climate. Extreme weather, climate, and water events in 2021 included record-breaking heat and rain, deadly fires, and crippling drought, with effects on people, the economy, and the environment that would last far longer than the year itself [2].

Over the past 50 years, there have been five times as many disasters, primarily as a result of climate change, more extreme weather, and better reporting. There is a clear pattern of extreme weather events increasing in intensity and frequency due to both anthropogenic climate change and global population growth. A number of methods are used to upload data that has been automatically acquired by sensors, such as semi-automated methods (in which data is collected by a sensor but manually uploaded), or methods that use human-generated data (data that is manually collected, entered and uploaded).

An increasing number of meteorological agencies are looking to harness the power of crowdsourcing data from weather watchers and spotters. It can be used to improve forecasting and warning serviceson a very short time horizon (nowcasting) or serve as prospective on-the-ground data for forecasts and warnings[3].

II. CONVENTIONAL DATA SOURCES

An experimental analysis was done on utility of various sources of data for machine-learning-based nowcasting of hazards related to thunderstorms where three types were chosen: ground- based radar data, satellite-based imagery and lightning observations. With the use of machine learning, these techniques aimed to give a systematic evaluation of the utility of various data sources for nowcasting thunderstorm related dangers.

The current generation of geostationary satellites, which, in comparison to the previous generation, provide higher resolution images, extra picture channels, and lightning data, are of special interest to this study since they want to understand how they will affect thunderstorm nowcasting. Multiple data sources like NEXRAD radar data, Satellite imagery: GOES ABI, Lightning data: GOES GLM, Digital elevation model: ASTER were analysed. Radar variables are good predictors for all predictands, especially radar-defined targets [9].

ABI satellite imagery improves all predictands' performance, but less than radar data. GLM lightning data is very valuable for lightning prediction; for other targets, they give relatively moderate gains, but may nevertheless enhance nowcasting performance, especially when radar data is unavailable. Satellite data may be utilised to give ML- based nowcasts in areas without radar coverage, such as across seas and in less-developed countries without ground- based radar networks. ECMWF prediction data, although being judged important by the ML algorithm, does not assist the nowcast according to the data exclusion analysis, since the relevant information content is present in the other observations.

The Indian Space Research Organization (ISRO) launched INSAT-3D (Indian National Satellite System) on July 26, 2013, with a multi-spectral imager covering visible to longwave infrared. This launch enabled the retrieval of temperature and water vapour profiles over India with higher temporal and vertical resolutions and altitude coverage, among other parameters. The INSAT-3D data was also matched to radiosonde measurements from 34 India Meteorological Department sites. Over India, comparisons were also done with data from other satellites including AIRS, MLS, and SAPHIR, as well as ERA-Interim and NCEP reanalysis data sets. With high spatial and temporal resolutions, INSAT-3D provides improved coverage over India. Except in the high tropospheric and lower stratospheric areas, there is a clear correlation in temperature between INSAT-3D and in situ data. INSAT-3D temperature data are of excellent quality and may be used directly to improve predictions for India [17].

Atmospheric data can also be used to forecast cities' weather. Using WMO as reference points, a statistical validation of citizen weather stations (CWS) observations is performed in this method. For each city, principal components analysis (PCA) time series from all WMO stations are constructed, and then statistical scores between WMO-PCA time series and each CWS time series are determined. Pearson correlation, the Kolmogorov-Smirnov two-sample statistical test, and the Kolmogorov-Smirnov normal test of remaining samples were the applied statistical scores (difference between two sources). In addition, the cross-correlation between each pair of CWS is determined. CWS that are not strongly linked with other CWS or PCA WMO (corr 0.8) are eliminated. Moreover, a technique is presented for predicting or even real-time forecasting of meteorological conditions via social media. It is believed that users would report on the weather and provide precise geographical coordinates with each post. The terms "sunny," "warm," and "cold" are considered as a human depiction of the weather. However, the user messages include far more terms that have nothing to do with weather. However, only relevant information is recovered using the named entity recognition (NER) method of Natural Language Processing. Here, a term frequency-inverse document frequency (TF- IDF) is presented in order to generate a quantitative representation of postings. Additionally, the postings are combined by day and turned to a tf-idf matrix. This is a regression problem where the input data is a vector of daily postings and the predictor is the current daily temperature collected from a WMO station [6].

III. CROWDSOURCING APPROACHES

Data is both produced and consumed by people. In highly inhabited places, data-deficient regions, or nations where established meteorological networks are deteriorating, crowdsourcing plays a critical role. It offers a great deal of promise to solve problems with observations' temporal and geographical representativeness. It is used for precise scientific and social applications using crowdsourced weather and climate data, including how and where to get data from, as well as how to rate the accuracy of this data. This paper presents a systematic literature review and delves into some of the approaches which are in place for the same [1].

A. Citizen science for Disaster Risk Reduction

Citizen science involves the participation of individuals or groups in developing new scientific information. Citizens can play a variety of roles in a project, from passive data gatherers to interpreters to collaborators and engagers. Similar to citizens, scientists can be involved in all aspects of a project - from leading it to working alongside them. Greater impact projects, especially those with broader social and societal benefits, demand more work and resources. There is still a lot of untapped potential and scientific difficulties revolving around citizen science such as ensuring proper data quality and ethical issues related to exploiting public data. Due to developing and reasonably priced technical advancements, there has been an increase in the number of citizen scientific initiatives.

The amount of participation from scientists and the public volunteers in citizen science programmes varies greatly. The improvement of the science-society discourse can all be facilitated by projects along this spectrum. Project typologies are classifications based on categories used to identify project responsibilities. Scientists' influence over the project as instructors, collaborators, or co-creators ranging from being a scientist lead to having shared ownership; and citizens' influence in the project as sensors, interpreters, engagers, or collaborators playing a passive role or being fully collaborative are two commonly used typologies. When creating citizen science initiatives, typologies may be utilized to assist these projects define and accomplish their goals [2].

B. Crowdsourcing Online

Crowdsourcing may be divided into "animate" and "inanimate" categories. While animate crowdsourcing incorporates some type of human interaction, inanimate crowdsourcing uses data from a variety of sensors and sensor networks [1].

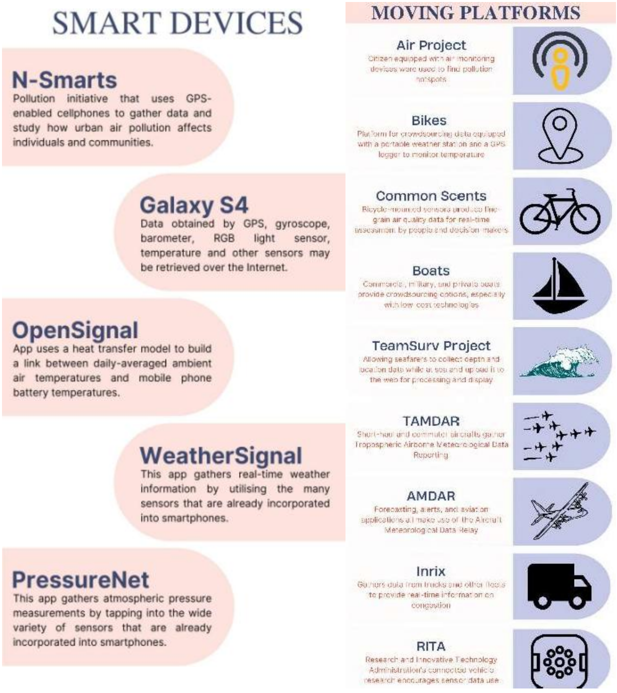
a) Social Media

Social networking platforms increasingly offer geo- located, time-stamped data. Social media can be used for crisis management since it can monitor situations and disseminate messages to key demographics quickly and effectively. A social weather application called "Metwit" enables users to contribute and receive data about the current weather using a variety of weather symbols. Similar to this, "Weddar" is a "people driven" application that requests users to use coloured symbols to express how they feel. In order to offer real-time signals for use by police and other members of society, services like "Twitcident" monitor, filter, and analyse tweets linked to events, risks, and crises. Storms have also been mapped using this method. The fact that social media feeds produce a lot of "noise" and false information is a drawback since it can lead to biased information being magnified.

b) In-Situ Sensors

Inexpensive, internet-enabled sensors and devices are available for remote personal, research, and operational use. Wi-Fi, Bluetooth, and machine-to-machine SIM cards can carry data, contributing to the Internet of Things and making a lot of data available. Some allow you to provide "metadata" while other freely available software can feed web pages with real-time sensor data. Instead of relying on presented information, accuracy and sensitivity of sensors must be analysed, calibrated, and compared.

c) Smart devices and Moving Platforms



N-Smarts
Pollution initiative that uses GPS-enabled cellphones to gather data and study how urban air pollution affects individuals and communities.

Galaxy S4
Data obtained by GPS, gyroscope, barometer, RGB light sensor, temperature and other sensors may be retrieved over the Internet.

OpenSignal
App uses a heat transfer model to build a link between daily-averaged ambient air temperatures and mobile phone battery temperatures.

WeatherSignal
This app gathers real-time weather information by utilising the many sensors that are already incorporated into smartphones.

PressureNet
This app gathers atmospheric pressure measurements by tapping into the wide variety of sensors that are already incorporated into smartphones.

Air Project
Citizen equipped with air monitoring devices were used to find pollution hotspots

Bikes
Platform for crowdsourcing data equipped with a portable weather station and a GPS logger to monitor temperature

Common Scents
Bicycle-mounted sensors produce fine-grain air quality data for real-time assessment by people and decision makers

Boats
Commercial, military, and private boats provide crowdsourcing options, especially with low cost technologies

TeamSurv Project
Allowing seafarers to collect depth and location data while at sea and upload it to the web for processing and display

TAMDAR
Short-haul and commuter aircrafts gather Tropospheric Airborne Meteorological Data Reporting

AMDAR
Forecasting, alerts, and aviation applications all make use of the Aircraft Meteorological Data Relay

Inrix
Gathers data from trucks and other fleets to provide real-time information on congestion

RITA
Research and Innovative Technology Administration's connected vehicle research encourages sensor data use

Fig. 1. Examples of Smart Devices and Moving Platforms

C. Numerical Weather Prediction (NWP): Assimilating Crowdsourced Data

Smartphones can gather and bias-correct air pressure data. Inaccurate smartphone location and sensor filtering causes poor data quality. In this approach, smartphone pressures were bias-corrected using machine learning to adjust for elevation issues. Random forests were trained to predict and resolve smartphone pressure errors. Pressure mistakes dropped 80% after bias correction. Under 20% of smartphone pressure was discarded. Bias-corrected smartphone pressures improved altimeter setting, 2-minute temperature, and 2-minute dew point forecasts. Data Assimilation for the REsilient City project (DARE) combines satellite, CCTV, and vehicle data to anticipate urban natural hazards including floods and high impact weather. High numbers of affordable observations are easily accessible, which is a benefit. However, data ownership, intermittency, heterogeneity, data provenance, and high data volumes are all difficulties [5].

D. Hail Reports

A crowdsourcing app Meteoswiss, reports hail in parts of Switzerland through popular crowdsourcing animated radar precipitation website. After a simple plausibility check, hail reports may be animated throughout the previous 24 h. Timing, place, and size of hail may be entered. The current time and phone location are default report input values. Smartphone monitoring shows location. Manually modifying a name or ZIP code diminishes spatial accuracy by hundreds of meters (depending on the size of the ZIP code area). The minute may be changed. Location and/or time must be manually given to filter reports. User chooses a hailstone classification system. "No hail," "coffee bean," In September 2017, golf ball and tennis ball replaced ">5 CHF." Table 1 Graupel (5 mm) is "smaller than a coffee bean" (5 mm). Other categories replace ">5 CHF" and identify dubious reports. The app stores the location, event time [in CEST (UTC + 2 h) in summer and CET (UTC +1 h) in winter], hailstone size, and submission time (when the user clicks "send"; Fig. 1b). Report "no hail." "No hail" reports distinguish hail from no-hail zones near storms. "No hail" reports aren't false alarms in this research [7].

E. Weather Applications or Apps

Apps like WarnWetter were studied using stakeholder mapping to find crucial parameters. These pertinent criteria were compared to prior crowdsourcing participants' experience and possible reporting needs. To ensure future data transfer compatibility, a selection of categories and values was selected as a compromise between the expectations of various stakeholders and existing crowdsourcing approaches. The app had a plausibility check to prevent misleading observations. Overall, user photographs were helpful, especially in high-impact situations like slippery conditions. Even though reporting was anonymous, there were few fraudulent reports and few negative complaints. 0.01% of photos were labelled as offensive, and most were false positives. The post-processing principle is that the temperature field must be closer to observed values at a specific observation time while gradually reverting to NWP raw forecast fields after a few hours. MET Norway uses crowdsourced data where Netatmo's personal weather stations provide the most data. Since March 2018, these data have been used to post-process forecasts on Yr (a weather app created in partnership by MET Norway and The Norwegian Broadcasting Corporation). An NWP model's two-meter temperature forecast is updated based on current readings over Scandinavia. Post-processing runs hourly, whereas data assimilation runs every six hours [8].

F. Weather Observations Website

Met Office and RMS launched it in 2011. WOW users can share observations with the world by submitting them in one of three formats: a quick manual observation with a condensed present weather code, a full manual observation, or raw data automatically uploaded from the Automatic Weather Station. The WOW site map may indicate air temperature, gale/hail/thunder/snow, mean sea level pressure (MSLP), soil temperatures, snowfall, sunshine duration, and ground condition. Users can display their observations in a dashboard, table, or graph. Europe has the most WOW users as compared to Australia, New Zealand, and the US. The official UK climate network's Voluntary Climate Network observers send daily reports using WOW. Some National Meteorological Services use it by hosting their websites at the Met Office or using the WOW API feed. It combines crowdsourcing data with observations from the Met Office's network of automated surface stations, the Meteorological Monitoring System. Nowcasting helps meteorologists predict high-impact weather occurrences like severe thunderstorms. By comparing WOW data to neighbouring MMS stations, bias is removed. Interpolating data onto a 4-km grid after bias reduction creates parameter maps. It helps meteorologists predict where thunderstorms may occur. WOW has 1.5 billion observations and 300–400 new users each month [10].

IV. COMPARISON BETWEEN CONVENTIONAL AND CROWDSOURCED APPROACHES

When it comes to conventional data sources like NEXRAD, GOES GLM, GOES ABI, increasingly weather forecasting and warning applications depend on integrated data from a range of asynchronous geographically distributed systems. The weather radar is a vital tool that delivers timely updates and comprehensive coverage. On regional scales, it has been demonstrated that the combination of the primary radar with subsidiary radars (either fixed or mobile) with satellite data, with automated meteorological measurements from satellites, and with a network of ground-based meteorological instruments reporting in real time provides enhanced nowcasting and short-term forecasting capabilities. Such capabilities improve severe local storm warnings (including predictions of storm commencement, development, and decay) whereas coming to crowdsourced approaches, its potential use for NWP is undeniable, as is their utility for verification and forecasting. As compared to NWP data assimilation, the use of crowdsourced observations in nowcasting or post-processing is seen as simpler and less difficult [5].

V. CONCLUSION

Globally, extreme weather events have a large detrimental social and economic impact, and it is expected that these effects will get worse as a result of both population growth and weather extremes brought on by climate change [2]. Due to technological developments over the past few decades, such as the proliferation of cellphones, it is now possible to acquire vast amounts of information about the effects of hazard occurrences, such as earthquakes and severe weather. Locals' observations of the consequences of severe weather on the ground provide researchers with information that may be used to enhance predictions and alerts as well as better understand these dangers, such as how earthquake shaking is felt and how people react.

The accuracy and utility of crowdsourced data must be high in order to use it for extreme weather events that harm property, life, and infrastructure [11].

Participation of citizens, amalgamation of multitudes of data sources and models may result in better accuracy and longer lead time of prediction.

Acknowledgements:

We would like to extend our gratitude to the Director of Space Applications Centre (SAC), ISRO

REFERENCES

[1]. C.L. Muller, L. Chapman, S. Johnston, C. Kidd, S. Illingworth, G. Foody, A. Overeem, and R.R. Leigh, “Crowdsourcing for climate and atmospheric sciences: current status and future potential,” in Int. J. Climatol. 35: 3185–3203 (2015).

[2]. Lauren J. Vinnell, Julia S. Becker, Anna Scolobig, David M. Johnston, Marion Tan, Lisa McLaren, “Citizen science initiatives in high-impact weather and disaster risk reduction,” Australasian Journal of Disaster and Trauma Studies Volume 25, Number 3, December 2021.

[3]. Thomas Kox, Henning W. Rust, Bianca Wentzel, Martin Göber, Christopher Böttcher, Jonas Lehmke, Elisabeth Freundl Matthias Garschagen, “Build and measure: Students report weather impacts and collect weather data using self-built weather stations,” Australasian Journal of Disaster and Trauma Studies Volume 25, Number 3, December 2021.

[4]. Tatiana Goded, Marion L. Tan, Julia S. Becker, Nick Horspool, Silvia Canessa Rand Huso, Jonathan Hanson, David M. Johnston, “Using citizen data to understand earthquake impacts: Aotearoa New Zealand’s earthquake Felt Reports,” Australasian Journal of Disaster and Trauma Studies Volume 25, Number 3, December 2021.

[5]. Kasper S. Hintz, Katharine O'Boyle, Sarah L. Dance, Saja Al-Ali, Ivar Ansper, Dick Blaauboer, Matthew Clark, Alexander Cress, Mohamed Dahoui, Rónán Darcy, JuhanaHyrkkanen, Lars Isaksen, Eigil Kaas, Ulrik S. Korsholm, Marion Lavanant, Gwenaelle Le Bloa, Emilie Mallet, Callie McNicholas, Jeanette Onvlee-Hooimeijer, Bent Sass, Valeria Siirand, Henrik Vedel, Joanne A. Waller and Xiaohua Yang, “Collecting and utilising crowdsourced data for numerical weather prediction: Propositions from the meeting held in Copenhagen, 4-5 December 2018,” in Atmospheric Science Letters on behalf of the Royal Meteorological Society, 2019

[6]. Amir Uteuova, Anna Kalyuzhnayaa, Alexander Boukhanovskya, “The cities weather forecasting by crowdsourced atmospheric data,”, 8th International Young Scientist Conference on Computational Science.

[7]. Hélène Barras, Alessandro Hering, Andrey Martynov, Pascal-Andreas Noti, UrsGermann, and Olivia Martius, “EXPERIENCES WITH >50,000 CROWDSOURCED HAIL REPORTS IN SWITZERLAND,” American Meteorological Society, August 2019.

[8]. Dr. Harald Kempf,” Experience from large-scale crowdsourcing via weather apps” Australasian Journal of Disaster and Trauma Studies Volume 25, Number 3, December 2021.

[9]. Jussi Leinonen, Ulrich Hamann, UrsGermann, and John R. Mecikalski, “Nowcasting thunderstorm hazards using machine learning: the impact of data sources on performance,” in Nat. Hazards Earth Syst. Sci., 22, 577–597, 2022.

[10]. Peter J. Kirk, Matthew R., Clark and Ellie Creed, “Weather Observations Website,” Weather - February 2021, Vol. 76, No. 2.

[11]. Thomas Nipen, Ivar Seiderstad, Cristian Lussana, Nina E. Larsgård and Mareile Wolff, “Crowdsourced data improves temperature forecasts on Yr.no,” in Norwegian Meteorological Institute.

[12]. Worden, C.B., Thompson, E.M., Hearne, M., and Wald, D.J., “ShakeMap Manual Online: technical manual, user’s guide, and software guide,” U. S. Geological Survey, 2020.

[13]. Goded, T., Horspool, N., Gerstenberger, M., Coomer, M.A., Becker, J.S., McBride, S., Canessa, S., and Lewis, A, “A comparison between GeoNet’s ‘Felt RAPID’ and ‘Felt Detailed’ online questionnaires,” in Proceedings of the New Zealand Society of Earthquake Engineering Technical Conference, Wellington, New Zealand, Paper 181, 8pp..

[14]. Heiss, W. H., McGrew, D. L., and Sirmans, D., “Nexrad: next generation weather radar (WSR-88D),” Microwave J., 33, 79+, 1990.

[15]. Hering, A., Germann, U., Boscacci, M., and Sénési, S., “Operational thunderstorm nowcasting in the Alpine region using 3D- radar severe weather parameters and lightning data,” in Proceedings of ERAD 2006.

[16]. Hering, A., Sénési, S., Ambrosetti, P., and Bernard-Bouissières, “Nowcasting thunderstorms in complex cases using radar data,” in WMO Symposium on Nowcasting and Very Short Range Forecasting.

[17]. Madineni Venkat Ratnam, Alladi Hemanth Kumar, and Achuthan Jayaraman, “Validation of INSAT-3D sounder data with in situ measurements and other similar satellite observations over India,” in Atmos. Meas. Tech. Discuss. (2016).