This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106064, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Heuristic Model to Compute Indices for Classification of Incidence and Non-incidence of Thunderstorms over Ranchi with Atmospheric Parameter

Kanchan Bala[1], Sanchita Paul[1], Mili Ghosh[2]

[1] Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi 835215, India
[2] Department of Remote Sensing, Birla Institute of Technology, Mesra, Ranchi 835215, India

Corresponding author: Kanchan Bala, Email: (kbala2736@gmail.com)

.

ABSTRACT Prediction of incidences of thunderstorms using different techniques is quite well establsihed. In this paper, a heuristic equation is proposed which relates the correlation coefficient of atmospheric parameters with the number of hourly incidences of thunderstorms. There are four ways to compute the indices values from the proposed heuristic equation. These indices values are used in the classification of the hourly incidences of thunderstorms. The proposed equation and indices work well, as tested on 2018 pre-monsoon hourly atmospheric data and validated with April 2019 and 2020 hourly data. From four indices, the first index value is computed with normalized average values of parameters of only hourly incidence of thunderstorms data of 2016-2017 in the month-wise method. The other three indices use optimization techniques namely, the Teaching Learning Based Optimization (TLBO) technique, Differential Evaluation (DE), and Simulated Annealing (SA). TLBO shows the better classification of hourly incidence of thunderstorms for 2018 atmospheric data. TLBO is also precisely validating the hourly incidences of thunderstorms for April 2019 and April 2020 hourly atmospheric data. It performed better with 2020 data by 88%. The variations of atmospheric parameters before, after or, during the incidences of thunderstorms are also depicted.

INDEX TERMS Correlation coefficient, Thunderstorm, Teaching Learning Based Optimization, Differential Evolution, Simulated Annealing.

I. INTRODUCTION

The thunderstorm is a violent and short-lived weather phenomenon, which associates with heavy rain alongwith strong gusty wind, dense clouds, lightning, and thunder. Moist and warm air rises and shifts the updraft to a cooler area of the atmosphere, moisture in the updraft condenses to form cumulonimbus clouds and eventually precipitates. Columns of cooled air sink towards the earth and strike the ground. The electric charge accumulates on water droplets and ice in the cloud. Once these accumulated electric charges become large enough, lightning discharge occurs. When lightning passes through the air quickly and intensely, shock waves produce a sharp rumbling sound. These rumbling sounds are known as thunders [25]. The thunderstorm is a space scale varying from a few kilometers to 100 kilometers and a time sale varying from 30 minutes to several hours. A severe thunderstorm occurs seasonally that causes extensive damages to crops, properties, animals, and life [12].

Thunderstorms are generally classified on basis of their physical characteristics. These thunderstorms are continuous, but still, there are four-way to classify thunderstorms namely single-cell storms, multicell cluster storms, multicell line storms, and supercell storms. Single-cell storms are weak and produce brief weather events. They usually have lifetimes of 20-30 minutes and companioned with brief heavy rain and sometimes occasional weak tornadoes. Single-cell storms poorly organize and occur at random locations and times, it is difficult to predict where and when these will occur. Multicell cluster storms are another common type of thunderstorms. They consist of a group of cells that moving as one unit. As multicell cluster storm moves, each cell in the cluster takes the dominant cell. Multicell cluster storms have a life of several hours, but individual cells die within 20 minutes. Multicell cluster storms are more intense than single-cell storms. Multicell cluster storms produce heavy rains. Multicell line storms are squall lines. It consists of a long line of the storm that is a continuous well developed gust. These thunderstorms produce hail, heavy rain, and weak tornadoes. The supercell thunderstorms are highly organized and pose a high threat to life and property [55].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106064, IEEE Access

The thunderstorm is the most spectacular weather phenomenon that can occur at any time anywhere in the world. These storms are accompanied by torrential rainfall, hail, tornadoes, strong wind gust, lightning, and development due to intense convention [31]. In general, lift, moisture, and instability are three main ingredients, which are essentially present in the environment for the development of severe thunderstorms. Thunderstorms have a life span of less than one hour to several hours and a spatial extension of a few kilometers [45].

The purpose of the present study is to propose a heuristic equation with the indices values to classify the incidences of thunderstorms over Ranchi City during the pre-monsoon season. The heuristic equation is formed with atmospheric parameters. The proposed equation is based on the correlation of these atmospheric parameters with the incidences of thunderstorms. It has been tried to maximize the correlation so that an equation is obtained that appropriately classifies atmospheric meteorological data.

The main contributions of this manuscript are as follows:

1. This manuscript has proposed a new way to find the classification of the incidences of hourly thunderstorms. A heuristic equation with four indices classifies the incidences of thunderstorms.
2. This manuscript uses five years of hourly atmospheric pre-monsoon data from 2016 to 2020. A heuristic equation is formed with two years of hourly atmospheric pre-monsoon data from 2016 to 2017. The atmospheric pre-monsoon data of the year 2018 hourly has been used as test data to classify into TD or NTD class. April 2019 and 2020 hourly atmospheric data are used for validation purposes.
3. The proposed heuristic equation uses the correlation of hourly atmospheric data of parameters with the number of hourly incidences of thunderstorms. The equation is further optimized using multiplication and division with some numeric constant to the parameters in the equation.
4. This manuscript proposes to compute the four indices values for the classification of TD and NTD in the data set. All four indices' values use the same proposed heuristic equation. The first index uses the normalized average of parameters in the equation and the other three indices take the equation as a cost function.
5. Hourly variation of the atmospheric parameters during the incidences of thunderstorms can help in the nowcasting of thunderstorms.

In this regard, the paper has been structured as follows: Literature survey has been presented in section 2 and Section 3 describes the problem statement. The proposed methodology can be easily understood with an algorithm, Figure 1, and description specified in section 4. The result of classification using the proposed methodology has been presented in Section 5 which also shows the variation of parameters and frequency of occurrences of thunderstorms. Conclusion and future direction are precisely described in Section 6.

II. LITERATURE SURVEY

The northeastern and eastern part of India encloses Jharkhand, Bihar, Gangetic West Bengal, Orissa, Assam and other states of the northeast. These areas have experienced a high frequency of lightning and thunderstorms during the pre-monsoon months, which is locally named as Kal-baishaki or Nor'westers. These thunderstorms are in North West (NW) direction so; these are called Nor'westers [16], [28]. Nor'westers are not due to local heat storms. Two different air streams, west to northwesterly direction winds from the land origin and moist winds of Bay co-exist over West Bengal cause Nor'westers. Thus, there exists low pressure over West Bengal, Orissa, Assam, Bangladesh, Chota Nagpur Plateau, and the adjoining regions. Ranchi is situated at one such location (Chota Nagpur Plateau) which experiences high-frequency occurrence of thunderstorms during afternoon [54].

Forecasting thunderstorms is the most difficult task in weather prediction due to its small temporal-spatial scale [36] and due to lack of vertical and horizontal resolution of the numerical model [30].

Many researchers have utilized the Artificial Intelligence (AI) technique and soft computing in the study of weather phenomenon [18]-[20], [34]-[35], [42], [52]. Numerous scientists have applied AI in the prediction of thunderstorms [6], [8]-[13], [33]. Deep-learning neural network (DLNN) model has been introduced to predict thunderstorm occurrence up to 15 hrs in advance within 400 km2 South Texas domains [57]. Soft computing and data mining techniques for the prediction of thunderstorms have been used for many decades which has been summarized [2]-[3].

Many researchers have proposed the expert system approach to forecast thunderstorms. An expert system was used to predict the thunderstorm and severe weather events using a decision tree. This system assists in categorizing the thunderstorms such as severe, non-severe, dry microbursts, or wet microbursts [15]. A rule-based expert system for the prediction of the thunderstorm have been made for use of weather forecasters of armed forces. This system utilizes the sounding parameters to predict the critical value of vertical total, cross totals, total totals, Showalter index, and lift index in thunderstorm occurrences. Thus, the Thunderstorm Intelligence Prediction System (TIPS) assists in the forecasting of thunderstorms [29]. Stability indices were introduced for many decades in the predictability of the thunderstorm of a particular region. Several authors have studied stability indices for the prediction of thunderstorms [32], [43], [46], [48].

The evolutionary computing method has also been used in the optimization problem. Genetic Programming (GP) is one of them, which uses the same properties of natural selection in biological evolution. GP has been used in the prediction of thunderstorms [27]. Many researchers utilized wavelet techniques in the prediction of thunderstorms [4], [41]. The screening regression technique has also been applied to the lightning data with meteorological predictors. These meteorological predictors were obtained from a numerical forecast model that forecasts thunderstorms using derived equations [39]-[40]. Two forecasting techniques were applied to forecast the thunderstorm over Delhi during the pre-monsoon season. The first technique was a graphical method with fifteen stability indices in the combination of different pairs. The second technique used nine significant predictors to formulate a multiple regression equation, which gives the forecast of the thunderstorm in terms of a probabilistic manner. Multiple regression gives better results than graphical techniques [38]. Different types of thunderstorms (severe, ordinary, and total) associated with rainfall over urban metropolises have been separately analyzed during pre-monsoon and monsoon. Analysis of the different types of thunderstorms has been done using the correlation coefficient technique with sounding data. The frequency of thunderstorm analysis was yearly based, which indicated positive co-relation to all types of thunderstorms with pre-monsoon rainfall amount while on the other hand there is anti-correlation observed in monsoon [45]. In this paper a heuristic equation has been proposed that uses the correlation coefficient of atmospheric meteorological parameters with the incidences of thunderstorms in pre-monsoon.

III. PROBLEM STATEMENT

In the last two decades, the prediction of thunderstorms and lightning is an active area of research. Yet it is still a challenging work for researchers and forecasters due to its spatial and temporal extension.

The study site (Ranchi) has its importance due to the thunderstorms generated over the Chota Nagpur during pre-monsoon. These thunderstorms move in the direction of the North West [50].

In this paper, the focus is to classify the incidences of thunderstorms over Ranchi, which occur on hourly basis during the pre-monsoon. Some data have been used for validation of the proposed method.

IV. MATERIAL AND METHODS

A. RECORD AND DATA

The station Ranchi (85^0.26E, 23^0.25N) is in the south part of the Chota Nagpur plateau and is surrounded by forest, hilly topography, and humid subtropical climate [21], [47]. The soil has a sandy loam texture that has sand (60%), clay (31.3%), and silt (8.7%) [53]. Ranchi lies in a humid subtropical monsoon area of India. It has general hot wet summers and cold winters. Maximum rainfall takes place from July to September that accounts for more than 90% of the total rainfall [54].

This study includes the hourly atmospheric data and records of the incidences of thunderstorms over Ranchi in pre-monsoon from the year 2016 to 2020. These hourly atmospheric data and records of incidences of thunderstorms are collected from [56]. The data has been corroborated with data of IMD Ranchi. The meteorological data set and records during the rain hour have been rejected due to the possible effect of water droplets on data quality [51], [53]. Thus rainy hourly data are not included in the dataset.

Hourly meteorological data have a total of five parameters namely, humidity, temperature, sea level pressure, wind speed, and wind direction. Two-year hourly data from 2016 to 2017 are used to formulate the proposed heuristic equation and four indices. Testing of the proposed indices is made with 2018 hourly meteorological data. Data from April 2019 and 2020 is used to validate the heuristic equation and four indices.

B. METHODOLOGY

This manuscript uses four indices values from the same heuristic equation. First of all, the heuristic equation is developed on basis of the correlation between parameters and the incidences of thunderstorms during pre-monsoon. The final equation is obtained by multiplying and dividing some numeric constant. Four indices values were obtained that classify meteorological data into TD and NTD classes appropriately. All these four indices values have been obtained through the application of the final heuristic equation. The first index value has been computed as an average of the obtained values, which were obtained by putting the normalized average of parameters of only TD pre-monsoon months of two years hourly data from 2016 to 2017 to heuristic equation. Normalized average hourly incidence of TD data values of parameters gets through the normalization of average values. These average values were obtained by taking the average of the values of each parameter of only hourly incidence of TD data of particular months such as April. Similarly, we compute the average values of each parameter for May and June months. These average values of parameters are normalized using equation 1. In the second method, an optimization technique (TLBO) is used to obtain the second index value. TLBO uses the same heuristic equation as a cost function .The normalized average values of each parameter of only the hourly incidences of thunderstorms data of pre-monsoon months of two years hourly data from 2016 to 2017 are used to obtain the second index. The third and four index values have been obtained using DE and SA respectively with the same procedure as TLBO. The generation of the Heuristic equation and four indices are further described briefly.

1) HEURISTIC EQUATION AND INDICES GENERATION

The indices values are used to classify the hourly incidence of thunderstorms over Ranchi. These indices values were obtained through a heuristic equation. Preprocessing and normalization of the atmospheric parameters are required for the formulation of the heuristic equation. In preprocessing, wind direction and wind speed with calm values have been removed. This is because these two do not have numeric values, but in the formation of a heuristic equation, only numeric values are required. Other unavailable parameter values have also not been taken in the data set. Data has been normalized using the following formula [22].

$$Parameter_{normalization} = \frac{x_{value-}x_{min}}{x_{max-}x_{min}} \quad (1)$$

x_{value} is the value of the parameter, x_{min} and x_{max} are the minimum and maximum value of parameters respectively. An equation has been proposed that uses these correlations of parameters. This proposed equation is applied in the training dataset of two years of hourly data from 2016 to 2017 and then computes the first index value for pre-monsoon. The index values and proposed equations have been further applied in the testing dataset to classify the incidences of thunderstorms. A brief description of the computation of the four indices values and the proposed heuristic equation is described in the algorithm and block diagram.

2) ALGORITHM FOR PROPOSED HEURISTIC EQUATION AND FOUR INDICES

The algorithm and block diagram explain precisely the proposed heuristic equation and indices values generation. Heuristic equation generation explains the systematic formulation of the heuristic equation. In the mentioned algorithm, $x_{humidity}$, x_{temp}, $x_{Sea\ Level\ Pressure}$, and $x_{wind\ speed}$ are values of atmospheric parameters namely, humidity, temperature, sea level pressure, and wind speed respectively. In brief, Algorithm describes the formulation of proposed heuristic equations with the four indices values. First, the heuristic equation has been formulated on basis of the R^2 values (correlation coefficient) of the atmospheric parameters with the number of incidences of thunderstorms in pre-monsoon. Heuristic equation formulation is similar to the regression equation. The heuristic equation constructs with a positive correlation and negative correlation of parameter value with the number of TD days. These correlations of parameters have been taken as the positive value and a negative value in the equation. The obtained heuristic equation further transforms with multiplication and division of some numeric constant in the parameters of the equation to maximize the R^2 value of the equation. The numeric values have not been added or subtracted in the heuristic equation due to no improvement in the R^2 value of the equation. Thus, the final equation is used for further computation of all four indices' values. The algorithm for the proposed heuristic equation and indices generation is given below.

Input: $(x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}) \rightarrow$ {Hourly thunderstorm data (TD), hourly non-thunderstorm data (NTD)}

Output: $(x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}) \rightarrow$ TD or $(x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}) \rightarrow$ NTD

Step 1. Five years of hourly data have been downloaded from [56] which have $x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}$ with either TD or NTD class.

Step 2. All data have been preprocessed such as the speed with calm values and wind direction have been removed.

Step 3. Only hourly incidences of TD from pre-monsoon months of 2016 and 2017 have been selected to form the heuristic equation and first index value.

Step 4. Arrange these selected hourly incidences of TD data according to month-wise.

Step 5. Compute the total number of incidence of TD and the individual average of each parameter value ($x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}$) of only TD days for particular months (Ex. : April).

Step 6. Repeat step 5 for other months (May, June).

Step 7. Normalize the individual average of parameters (x_humidity, x_temp, x_(Sea Level Pressure), x_(wind speed)) ($x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}$) data of months (April, May, and June) of pre-monsoon months of 2016 to 2017 years. Table 2 shows these normalized values.

Step 8. Compute the correlation between the normalized average values of parameters. ($x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}$) and number of TD in pre-monsoon months.

Step 9. Formulate the heuristic equation using a correlation of normalized average values of parameters and the number of incidences of TD as follows.

a) Form the heuristic equation, a positive correlation of parameter value with the number of incidence of TD days as the positive value in the equation and negative correlation of parameters value with the number of incidence of TD days as a negative value in the equation.
b) Compute correlation value (R^2) of the obtained equation in step 9.a) with the number of incidence of TD days.
c) Transform the obtained equation in step 9.a) to maximize the R^2 value of the heuristic equation using some numeric values are multiply or divide into parameters in the obtained equation in step 9.a).
d) Use obtained equation in step 9.c) as a final heuristic equation, which is the proposed heuristic equation.

Step 10. Develop four prediction indices from the final heuristic equation.

Step 11. Compute the first index using forecasting the developed heuristic equation in normalizing the individual average of parameters ($x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}$) data for months (April, May, and June) of pre-monsoon months of 2016 to 2017 years. Table 2 shows the forecasted values in column $Classification_{index}$.

Step 12. Thus, the first index value is computed using averaging the obtained values from the application of the heuristic equation to the normalized average value of parameters ($x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}$) of pre-monsoon months of 2016 to 2017 years as in Table 2.

Step 13. Obtain other three indices using taking heuristic equation as a cost function in optimization techniques namely, TLBO, DE, and SA.

Step 14. Heuristic equation and the four indices values are used to test one year hourly normalized data of 2018 of Ranchi city (($x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}$) $\rightarrow$ TD or ($x_{humidity}, x_{temp}, x_{Sea\ Level\ Pressure}, x_{wind\ speed}$) $\rightarrow$ NTD).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106064, IEEE Access

Step 15. Validate the proposed heuristic equation and four indices with April 2019 and 2020 hourly data.
Step 16. Stop.

Figure 1 briefly describes a block diagram of an algorithm for generating the heuristic equation and four indices values The heuristic equation and four indices values are based on two-year hourly data. Only the data of incidences of TD have been separated from the data set. All TD data is arranged in a month-wise way. Finally, the normalization process applies to the average value of the parameters of the only incidence of TD. The number of incidence of TD and normalized average values of the parameters month-wise are arranged as shown in Table 2. Correlation (R^2 value) was computed with normalized average values of parameters and the number of incidences of TD month-wise. Obtained correlation of normalized average values of parameters and number of incidence of TD month-wise is shown in Table 1. A heuristic equation and the first index value form with the help of Table 1 and Table 2 respectively.

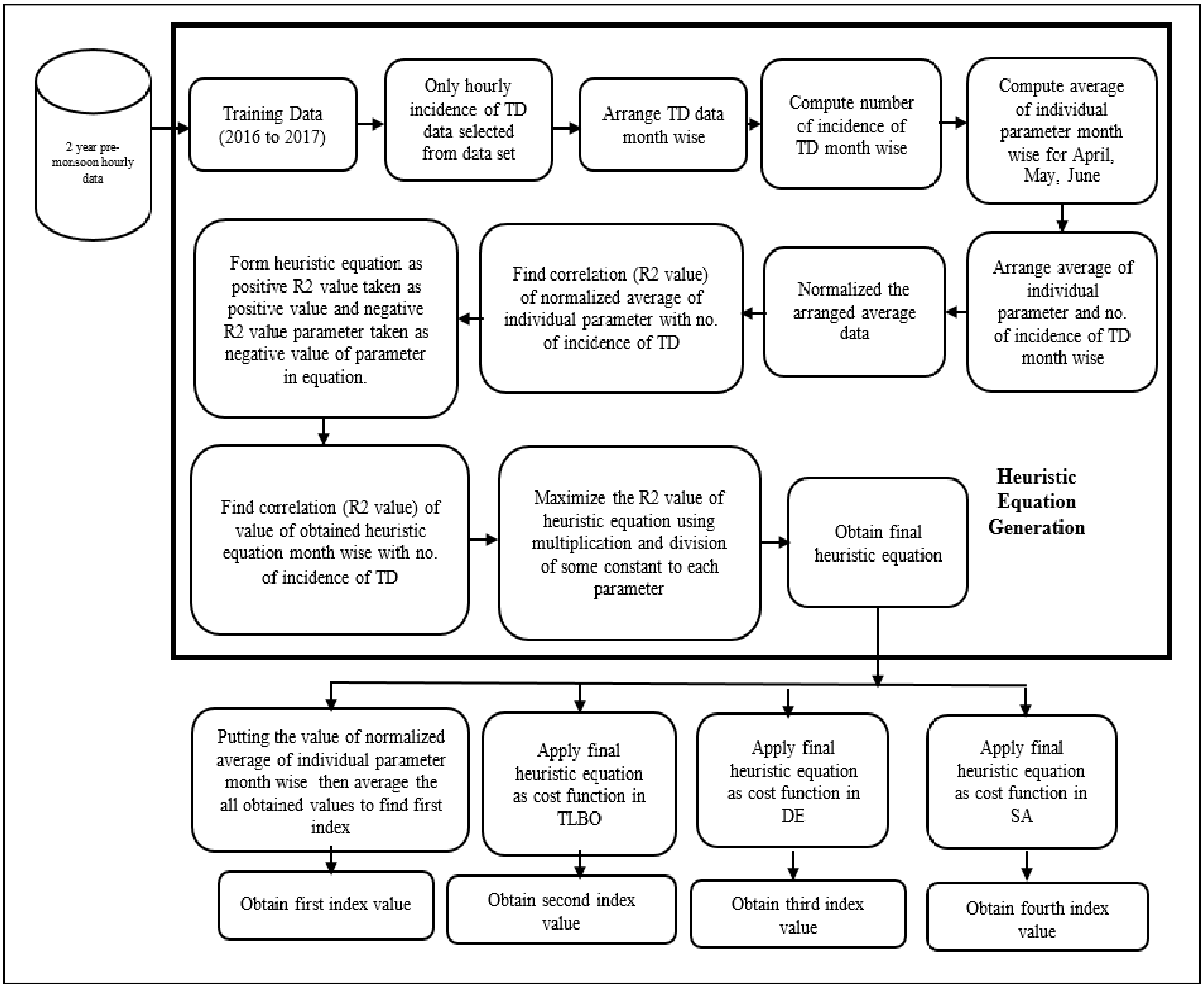

FIGURE 1. **Block Diagram of Generation of the Heuristic Equation and Four indices computation**

3) HEURISTIC EQUATION GENERATION

Several authors have proposed the correlation of predictors with the occurrence of thunderstorms. An equation was proposed on Basis of these correlations, such a proposed equation was indicative of the occurrence of thunderstorms [17], [38]. A regression equation was developed that was based on the correlation of predictors with the occurrence of thunderstorms. In this developed equation, positively correlated predictors were taken as positive, and negatively correlated predictors were taken as negative in the regression equation [17]. The proposed methodology of this manuscript has used a similar concept with little modification to formulate the heuristic equation for the classification of thunderstorm days. R^2 values of atmospheric parameters are given in Table 1. The following points should be noted in Table 1.

1. Humidity ($x_{humidity}$)→ +ve (Positive) (0.758)
2. Temperature (x_{temp}) → -ve (Negative) (0.94)
3. Wind Speed ($x_{wind\ speed}$) → +ve (Negative) (0.01)
4. Pressure (x_{SLP}) → -ve (Negative) (0.2142)

From these four R^2 value of parameters, the authors conclude the following

(a) Temperature and pressure are negatively correlated with the incidences of thunderstorms. This means that a lower value of temperature or pressure favors the incidence of thunderstorms in pre-monsoon.

(b) Wind speed and humidity are positively correlated with the incidence of thunderstorms. The higher value of wind speed/ humidity indicates favorable for occurrences of thunderstorms in pre-monsoon.

Thus, all parameters are physically significant and have indicative occurrences of thunderstorms or non-thunderstorms. How the proposed methodology of the heuristic equation ($Heuristic_{equ}$) formulated for incidence of thunderstorms is computed as follows. Since the value of R2 of wind speed and humidity have a positive that is a positive correlation of these two parameters with the number of incidences of thunderstorms. their values have been taken as the positive in the heuristic equation as in equation number 2. Pressure and temperature have a negative value of R2 that is a negative correlation with the number of incidences of thunderstorms. So, we take the negative value of temperature and pressure in the heuristic equation. Thus, the heuristic ($Heuristic_{equ}$) equation is proposed as.

$$Heuristic_{equ}=(x_{humidity} + x_{wind\ speed})-(x_{temp} + x_{SLP}) \quad (2)$$

The proposed heuristic equation is positively correlated with the number of incidences of thunderstorms with the R^2 value of 0.9241. Thus $Heuristic_{equ}$ equation has more correlated than any atmospheric surface layer parameters with the number of incidence of thunderstorm occurrences. The higher value of R^2 of the heuristic equation indicates the favorable for incidence of thunderstorms in pre-monsoon. It is needed to maximize the R^2 value of $Heuristic_{equ}$ equation. Some numeric constants are multiplied or divided in the above $Heuristic_{equ}$ equation, so that the increased correlation of the $Heuristic_{equ}$ equation with the number of incidence of thunderstorms over Ranchi. Addition and subtraction of some numeric constants does not increase the correlation of the $Heuristic_{equ}$ equation with the number of incidence of thunderstorms. Thus, increase the correlation of the $Heuristic_{equ}$ equation with the number of incidence of thunderstorms is done as follows.

First, consider the wind parameter of $Heuristic_{equ}$ equation to maximize R^2 value from 0.9241.

- R^2 value increases up to 0.9872 as multiple by 2 to the value of wind parameter of $Heuristic_{equ}$ equation.
- R^2 value decreases up to 0.5934 as multiple by 4 to the value of wind parameter of $Heuristic_{equ}$ equation.
- R^2 value decreases up to 0.8005 as divide by 2 to the value of wind parameter of $Heuristic_{equ}$ equation.

So, Wind parameter multiple with 2. The same procedure is done for humidity, temperature, and pressure parameters of $Heuristic_{equ}$ equation. Thus the final $Heuristic_{equ}$ equation is obtained as follows.

$$Heuristic_{equ} = (2 * x_{humidity} + 2 * x_{wind\ speed}) - (0.1 * x_{temp} + x_{SLP}) \quad (3)$$

The R^2 value of the above final $Heuristic_{equ}$ equation is 0.999. Thus, $Heuristic_{equ}$ equation is the function of humidity, temperature, wind speed, and sea level pressure.

4) INDICES GENERATION

Final $Heuristic_{equ}$ equation is forecasted to training data (2016 to 2017) to find the first index value. Thus, the first index value is computed using forecasting $Heuristic_{equ}$ equation 3 in normalized average values of individual parameters (humidity, temperature, wind speed, and sea level pressure) and then average the generated values month-wise as in Table 2 for pre-monsoon. Thus the analysis of hourly atmospheric data and records of thunderstorm over Ranchi from 2016 to 2017 led to the first index value of $Heuristic_{equ}$> 1.278 for pre-monsoon which indicates that if forecasted values are within this mention range, then the possibility of incidence of thunderstorms, otherwise non-thunderstorms.

The other three indices are generated using optimization techniques. These optimization techniques use the normalized average value of parameters of hourly incidence of only TD month-wise of two years hourly atmospheric data from 2016 to 2017 as lower and upper bound of parameters. The proposed heuristic equation is used as a cost function in optimization techniques. The heuristic equation is optimized to get the improved index. There are three optimization techniques namely, TLBO, SA, and DE used to generate the improved indices.

C. TLBO

Teaching Learning Based Optimization is a nature-based algorithm. It was a fascinating researcher and has lots of recognition due to high problem solving for the optimization problem. TLBO is used in many applications of different fields of science, technology, engineering. The algorithm adopts ideas from the natural teaching-learning method of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106064, IEEE Access

classroom and takes the best solution from various possible solutions [1]. TLBO algorithm is based on two phases 'Teacher Phase' and 'Learner Phase'.

1) TEACHER PHASE

It is the first phase of an algorithm where learners learn from a teacher. The teacher makes an effort to enhance the mean result of the class for the subject taught by her or him on their capability. For any iteration i, suppose 'p' number of subjects (design variable) 'q' number of learner (population size 1, 2… q) and $M_{j,i}$ be the average result of the learner for a particular subject 'j' (j=1, 2, ….P). $X_{total\text{-}lbest,i}$ is the best overall result containing all subjects together from the entire population of the learner which is the best learner *lbest*. However, usually consider that teacher is a highly learned person who teaches learners for better results. The best learner is identified as a teacher. Now find out the difference between the mean results of each subject with the result of teacher for every subject [44].

$$Diff\_Mean_{j,l,i} = r_i\ (X_{j,lbest}\text{-}\ S_F M_{j,i}) \quad (4)$$

where $X_{j,lbest,i}$ is best learner of subject j. S_F is a teaching factor that influences the value of mean to vary change. r_i is the random numbers in between [0, 1]. The value of S_F can be 1 or 2 and it is decided randomly as

$$S_F = round[1+rand(0,1)] \quad (5)$$

Where S_F is randomly generated by an algorithm and it is not a parameter of algorithm which have a value between 0 to 1 for better performance of algorithm either 1 or 2. Based on $Diff_Mean_{j,l,i}$updates the existing solution in the teaching phase as

$$X'_{j,l,i}=X_{j,l,i}+Diff\_Mean_{j,l,i} \quad (6)$$

Where $X'_{j,l,i}$ is acceptable only when it gives a better result of the function value. All the accepted function values are maintained and used as input to the learner phase. Thus, the learner phase depends on the teacher phase.

2) LEARNER PHASE

It is the second phase of the TLBO algorithm. In this phase, learners enhance their Knowledge using interaction with themselves. A learner randomly interacts with other learners to increase her or his knowledge. A learner learns new knowledge from others who have more knowledge than her or him. Consider 'n' is the population size and two learner A and B are randomly selected in such a way that $X'_{total-A,i} \neq X'_{total-B,i}$ Where $X'_{total-A,i}$ and $X'_{total-B,i}$ updated value function value of A and B respectively at the end of TLBO's teacher phase [44].

$$X''_{j,A,i}=X'_{j,A,i} + r_i\left(X'_{j,A,,i} - X'_{j,B,i}\right) \text{ If } X'_{total-A,i} < X'_{total-B,i} \quad (7)$$

$$X''_{j,A,i}=X'_{j,A,i} + r_i\left(X'_{j,B,,i} - X'_{j,A,i}\right) \text{ If } X'_{total-B,i} < X'_{total-P,i} \quad (8)$$

When $X''_{j,A,i}$ gives better function value, then only its value is acceptable. For the minimization problem equations (6) and (7) are used and equations (8) and (9) for maximization.

$$X''_{j,A,i}=X'_{j,A,i} + r_i(X'_{j,A,,i} - X'_{j,B,i})\text{If}X'_{total-B,i} < X'_{total-A,i} \quad (9)$$

$$X''_{j,A,i}= X'_{j,A,i} + r_i(X'_{j,B,,i} - X'_{j,A,i})\text{If}X'_{total-A,i} < X'_{total-B,i} \quad (10)$$

This TLBO algorithm was applied to the same heuristic equation and It uses the heuristic equation as a cost function. The cost function is optimized for finding the second index value. Thus, we have obtained the second index value as 1.1 using TLBO. While in the case of the above-proposed first index value, it was 1.278. These indices values are forecasted for hourly 2018 data and used to compare the result.

Basic TLBO algorithm

Step 1: Initializations

Population size= q
Number of Iteration= N
Number of design variable = p
Upper limit of design variable = Up
Lower limit of design variable =Lp

Step 2: Generate random population according to number of design variable and population size

Step 3: Evaluation of fitness function

Evaluate the fitness of all feasible solutions in the population and arrange them according to their fitness.

Step 4: Teaching phase

Modify the solution according to equations 4 to 6.

Step 5: Student phase

Modify the solution according to equations 7 to 10.

Step 6: Repeat the step 3 to 5 up to the maximum number of iteration

Step 7: stop

D. DE

DE is an evolutionary algorithm for solving real-valued parameters optimization problems [24]. This technique optimizes iteratively the given problem. This method is a metaheuristic that makes no or few assumptions for the problem [23]. In 1995, It was introduced by R. Storn and K.V. Price [49]. A detailed description and flow chat of DE are explained in [26].

Basic DE algorithm

Step 1. Take Parameter CR ∈ [01], F ∈ [1, 2], NP≥4

CR = crossover probability, LR=0.9
F = differential weight, F=0.8
NP = population size
Lb = lower bound
Ub = upper bound
X =agent which initialized with random position in population

Step 2. Do until a termination is reached,

Step 3. For each x in population do.

Step 4. Take three agents a, b & c randomly from population
Step 5. Take R ∈ {1,…….., n} Where R is random index and n is dimension of problem being optimize
Step 6. Compute new position y=[y1 ,………,yn] of agent
Step 7. For each i ∈ {1,…………..,n}, take r_i ∈ [0,1] Where r is random number
Step 8. If r_i < CR or i=R then set yi = xi
Step 9. If f(y) ≤ f(x)
Step 10. Then replace the agent x in the population with improved
Step 11. Pick the agent from the population that has the best fitness.
Step 12. End for
Step 13. End for
Step 14. End do

E. SA

SA is a nature inspire technique to solve the optimization problem. It is a general and efficient optimization technique that was proposed by Kirkpatrickt, Gelett, and Vecchi in 1983 and independently by Cerny in 1985. This technique is based on physical behavior system which emulates like when a solid is cooled slowly and then its structure is frozen at minimum energy [5].

SA Algorithm

Step 1. Take proposed equation e_0 to be optimized.
Step 2. Initialize different parameters: Temperature T, Wind Speed W, Humidity H, Sea level pressure P in equation e_0, reduction factor C, and Boltzann's constant K
Step 3. while termination is not satisfied do
Step 4. for number of different new solution
Step 5. Take a new solution $e_0+\Delta e$
Step 6. If $f(e_0+\Delta e) > f(e_0)$ then
Step 7. $f_{new} = f(e_0+\Delta e)$; $e_0 = e_0+\Delta e$
Step 8. Else
Step 9. $\Delta f = f(e_0+\Delta e)- f(e_0)$
Step 10. Random r (0, 1)
Step 11. If r>exp(-Δf/KT) then
Step 12. $f_{new} = f(e_0+\Delta e)$, $e_0 = e_0+\Delta e$
Step 13. else
Step 14. $f_{new} = f(e_0)$
Step 15. end if
Step 16. end if
Step 17. if = f_{new}
Step18. Decrease the temperature(T), humidity(H), sea level pressure(P), and wind speed(W) periodically T/H/W/P = C* T/H/W/P
Step 19. end for
Step 20. end while

V. RESULT AND DISCUSSION

Several atmospheric surface parameters are available for the prediction of the beginning of convection and pre-convective condition.

There are four types of thunderstorms; A, B, C, and D type occurred during pre-monsoon in the northeastern part of India (IMD T.N. 10, 1944). Type A develops mainly in the afternoon over West Bengal, Chota Nagpur Plateau (India), and Bangladesh and subsequently moves in the southeastern direction [54]. Ranchi is one of the places of the Chota Nagpur Plateau. Type A thunderstorms occur in the afternoon of the pre-monsoon month over Ranchi that depicts in Figure 2(a) – (c).

Figures 2 (a) – (f) shows the record of the incidence of thunderstorms during pre-monsoon and monsoon (April to September) of the period from 2016 to 2018. Thus Figure 2 shows the hourly frequency distribution of the incidence of thunderstorms from April to September. Figure 2, indicates that the maximum frequency of incidence of thunderstorms was found from 12:00 to 22:00 local time in pre-monsoon months. Even some months of monsoon have a maximum frequency of incidence of thunderstorms during these hours.

Figures 3 (a) – (d) depict that how the atmospheric meteorological parameters vary before, after, and during the occurrences of the thunderstorm. In all figures, Time in the hours, and values of atmospheric parameters were hourly data of the month of 'May 2018'. These figures take two consecutive days for diurnal variation of atmospheric meteorological parameters, one is NTD i.e 24 May 2018 and another day is TD, 25 May 2018 with thunderstorm hours from 19:00 to 21:30. Wind speed becomes high approximately one hour before the incidence of the thunderstorm. At the time of incidence of the thunderstorm and after or during the incidence of thunderstorms remain high and vary as in Figure 3 (a). On 24 May, air temperature variation depicted usual diurnal variation while on 25 May TD; the temperature gets down a few hours before the incidence of the thunderstorm. Temperature becomes minimum during or after thunderstorms as shown in Figure 3 (b). Thus, on TD; there is a sudden fall of temperature observed during the incidence of thunderstorm hour drop from 34^0 C to 25^0 C.

The humidity gets high at the time of incidence of the thunderstorm and remains high during the thunderstorm as in Figure 3 (c). Humidity rises by 56 % with two to three hours at the time of incidence of the thunderstorm.

Sea level pressure goes down before the occurrences of the thunderstorm and becomes minimum at the time of the incidence of thunderstorms as shown in Figure 3 (d). All these variations of parameters of atmospheric data during occurrences for Ranchi city observed in [47], [54].

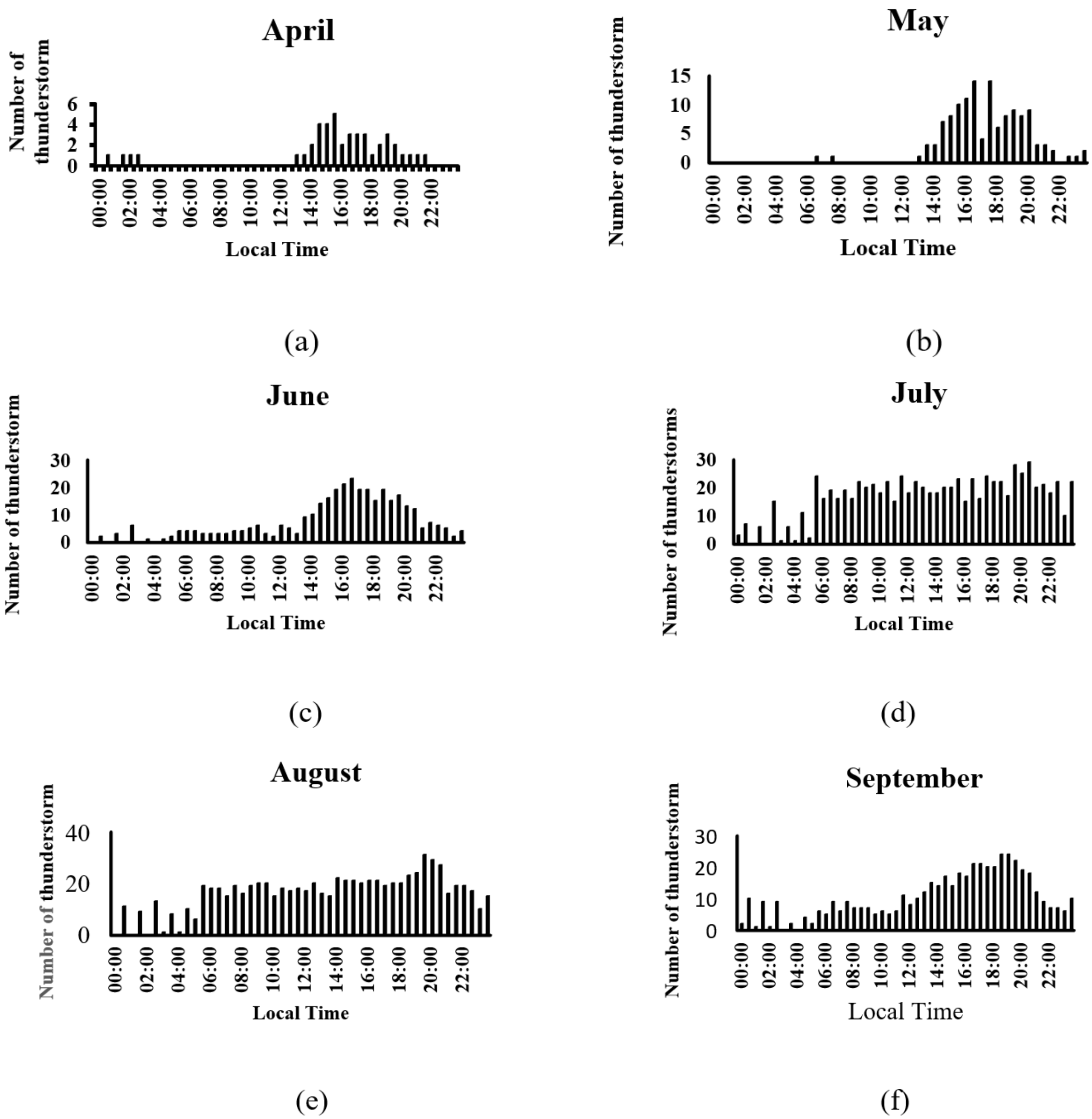

Figure 2: **Occurrence of the number of thunderstorms during monsoon and pre-monsoon.**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106064, IEEE Access

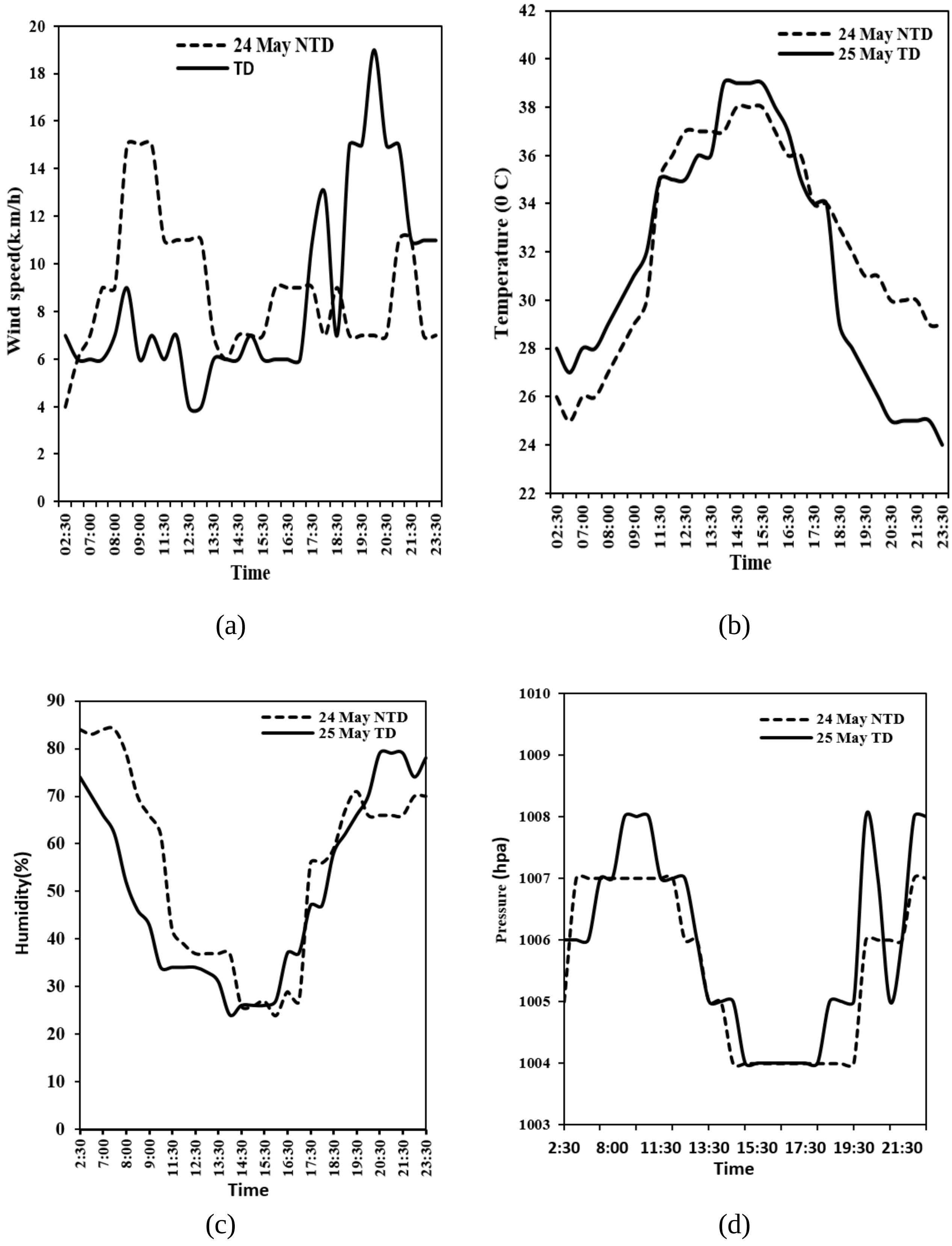

FIGURE 3. Diurnal variation of (a) Wind Speed, (b) Air Temperature, (c) Humidity, (d) atmospheric surface Pressure on 24 May 2018 (NTD), 25 May 2018 (TD) using hourly data.

TABLE 1. R^2 value of atmospheric parameters with numbers of incidence of thunderstorms

Parameters	Pre-Monsoon R^2 value	Monsoon R^2 value
Temperature	-0.942	+0.96
Wind Speed	+0.01	-0.16
Humidity	+0.758	-0.105
Pressure	-0.214	0.877

TABLE 2. First index value computation for pre-monsoon from 2016 to 2017 hourly data

Parameters	No. of TD	Temperature	Wind speed	Humidity	Pressure	$Classification_{index}$
April	2	1	0.37	0	0.90	-0.25
May	79	0.299	1	0.55	1	2.08
June	83	0	0	1	0	2.00
First index						**1.278**

Table 2 shows the computation of the first index value. First, we have collected only the hourly incidences of thunderstorms (TD) data from pre-monsoon months of two-year hourly data from 2016 to 2017. These collected TD data have been separated according to month-wise (April, May, June), also we computed month-wise the total number of hourly incidences of thunderstorms as in No. of TD column of Table 2. Average values of temperature parameter of only the hourly incidence of TD data month-wise for two year hourly data from 2016 to 2017 were computed. Similarly, average values of other parameters namely, wind speed, humidity, and pressure of the only hourly incidence of TD days month-wise were computed too. All these average values of parameters of TD days month-wise are normalized that forms the columns of temperature, wind speed, humidity, and pressure of Table 2. These normalized average month-wise TD days values of parameters are put in the heuristic equation which is the $Heuristic_{equ}$ the column of Table 2. The values of $Heuristic_{equ}$ the column has been averaged and gets the value of 1.278. This value is the first index value.

The first index value (1.278) and heuristic equation were applied to hourly data of the pre-monsoon month of 2018 and obtained results are shown in Table 3. The heuristic equation has been used as a cost function in optimization techniques to get the other three indices (1.1, 1.08, and 0.9). All three indices are applied along with the heuristic equation to hourly data of the pre-monsoon month of 2018 and obtained results shown in Tables 4, 5, and 6.

TABLE 3. Result for pre-monsoon months of 2018 hourly data using first index value (1.278)

Months	TD	Classified TD (Accuracy)	NTD	Classified NTD (Accuracy)	Accuracy
April	46	31 (67.39)	927	767 (82.74)	82.01
May	46	29 (63.04)	974	887 (91.06)	89.80
June	57	40 (70.17)	835	631 (75.56)	75.22
Total	149	100(67.11)	2736	2285 (83.51)	82.66

TABLE 4. Result for pre-monsoon months of 2018 hourly data using TLBO (1.1)

Months	TD	Classified TD (Accuracy)	NTD	Classified NTD (Accuracy)	Accuracy
April	46	35 (76.08)	927	692 (74.64)	74.71
May	46	33 (71.73)	974	804 (82.54)	82.05
June	57	45 (78.94)	835	653 (78.20)	78.25
Total	149	113(75.83)	2736	2149 (78.53)	78.40

TABLE 5. Result for pre-monsoon months of 2018 hourly data using DE (0.9)

Months	TD	Classified TD (Accuracy)	NTD	Classified NTD (Accuracy)	Accuracy
April	46	41 (89.13)	927	573 (61.81)	63.10
May	46	39 (84.78)	974	646(66.32)	67.15
June	57	48 (84.21)	835	436(52.21)	54.26
Total	149	128(85.90)	2736	1655 (60.48)	61.80

TABLE 6. Result for pre-monsoon months of 2018 hourly data using SA (1.08)

Months	TD	Classified TD (Accuracy)	NTD	Classified NTD (Accuracy)	Accuracy
April	46	37 (80.43)	927	672 (72.49)	72.86
May	46	34 (73.91)	974	793 (81.41)	81.07
June	57	45 (78.94)	835	535 (64.07)	65.02
Total	149	116(77.85)	2736	2000 (73.09)	73.34

Table 3 depicts the result of the first index value (1.278) that was applied to the one-year hourly data of 2018. In Table 3, both incidences of thunderstorms data (TD) and non-incidence of thunderstorms data (NTD) are classified well. The total number of TD data in pre-monsoon months is 149 days. The first index value depicts 100 incidences out of 149 TD data and 2285 non-incidence out of 2736 NTD data. Thus, classification accuracy for TD data is 67.11% and for NTD data is 83.11%. Thus total classification accuracy is 82.66% for pre-monsoon. May month has the highest overall accuracy of 89%, NTD accuracy of 91%, and TD classification accuracy of 63%. Although the classification accuracy using the first index value has good combined accuracy of classification for TD and NTD of pre-monsoon months, individual TD day's classification accuracy is not very satisfactory. So, the proposed heuristic equation has been used as input (Cost function) in the optimization model to improve the classification accuracy of incidence of TD. Three optimization models namely, TLBO, DE, and SA have been used for the improvement of accuracy of classification.

In the TLBO model, $Heuristic_{equ}$ an equation has been used as a cost function and the second index value of 1.1 is obtained using the minimization of the cost function. This cost function and the second index have been applied to 2018 hourly data. If the obtained values of the cost function in 2018 hourly data are greater than the second index value of 1.1 it belongs to the hourly incidence of TD otherwise it is NTD. Table 4 shows the result of the TLBO model. Only 36 incidences of hourly TD data out of 149 hourly TD days were not classified. The accuracy of hourly classification for TD and NTD are 75.83% and 78.53% respectively. The overall (TD+NTD) classification accuracy using TLBO with its index value of 1.1 is 78.40%. NTD and overall accuracy are low by approximately 4% as compared to the result for the first index value (1.278), but TD day classification accuracy is high by 8.72 %. The proposed equation and indices are validated with the hourly 2019 and 2020 atmospheric meteorological data. For validation of indices and equation, TLBO shows better results by 25%. than the first index method due to high TD accuracy. Thus TLBO performs better than the first index method. TLBO is also a better performer than other methods DE and SA. Although some papers used different data and techniques to predict the incidence of thunderstorms, the proposed method also gives higher accuracy than what is mentioned in [7], [14], [37].

In the DE algorithm, $Heuristic_{equ}$ equation has been taken as a cost function and gets the third index value of 0.9. This index value and heuristic equation as cost function have been applied in 2018 hourly atmospheric meteorological data to classify the incidence of TD day and NTD. Table 5 depicts the result of DE. It classifies a very low number of NTD. Overall (TD+NTD) and NTD classification accuracies are 60.53% and 61.84% respectively. Although TD days classification is appropriate and has the highest classification accuracy. thus DE technique does not work well and has the lowest performance as depicted in Table 5. DE algorithm does not classify approximately 40 % of total hourly testing data.

The same heuristic equation has been used as a cost function in SA and obtains a minimum cost function value of 1.08. The obtained minimum value acts as a fourth index value. Cost function and its index values are applied in 2018 hourly atmospheric meteorological data to classify the incidence of TD and NTD. Table 6 shows the result of the SA algorithm. Hourly classification accuracy of TD has good accuracy with 77.85% and for NTD accuracy is 73.09%. Hourly classification of NTD days has also been well classified for all pre-monsoon months except June. The overall (TD+NTD) classification accuracy using SA with its index value of 1.08 is 73.34%. NTD and overall accuracy are low by approximately 9% as compare to the result of the first index value (1.278), and 5% less as compare to TLBO but TD day classification accuracy is high with 10 % and 2% as compared to the first index and TLBO respectively.

Performance of the first index, TLBO, SA, and DE are validated with the hourly 2020 atmospheric meteorological data. In the validation of these models, TLBO shows better results than SA and DE due to high overall accuracy and low accuracy than the first index. In the case of TD days classification, TLBO shows better results than the first index by 25%. Thus TLBO shows the best among other models SA, DE, and first index. It shows SA is better than DE in terms of performance which is clear from Figure 4 to 6.

Figure 4 shows the classification accuracy of hourly incidence of TD data for month-wise and total pre-monsoon months. DE has the highest classification accuracy and the first index value achieves the lowest accuracy in April, May, June, and total pre-monsoon. TLBO and SA have approximately the same classification accuracy of incidence of TD, although SA has higher accuracy than TLBO except for June. Figure 5 and Figure 6 shows TLBO has better performance than SA, DE in terms of hourly NTD classification and total classification of pre-monsoon atmospheric data. Table 7 and Table 8 validate that TLBO performs better than SA, DE, and the first index method.

Figure 5 depicts the classification accuracy of hourly NTD data month-wise and total pre-monsoon months. The result shown in Figure 5 is just opposite to the result of the hourly TD classification shown in Figure 4. Like DE gets highest in NTD classification whereas in case of TD classification it has the lowest accuracy. The first index has the highest and DE has the lowest accuracy in April, May, June, and total pre-monsoon months for hourly NTD days. Month-wise overall (TD+NTD) classification is also demonstrated in Figure 6.

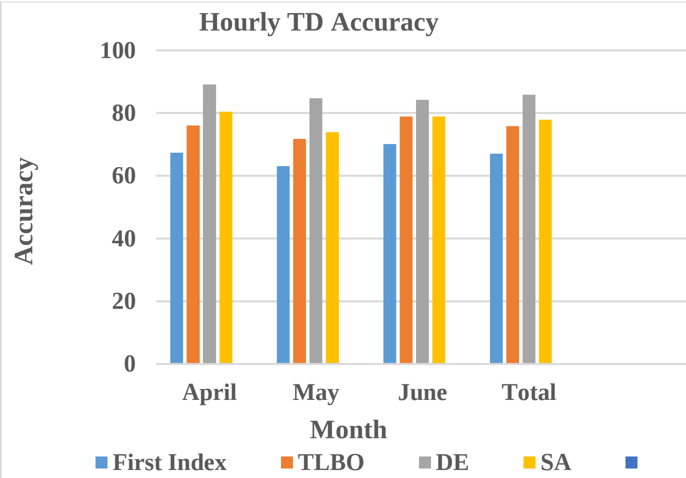

FIGURE 4. Month-wise TD accuracy and total Pre-monsoon accuracy

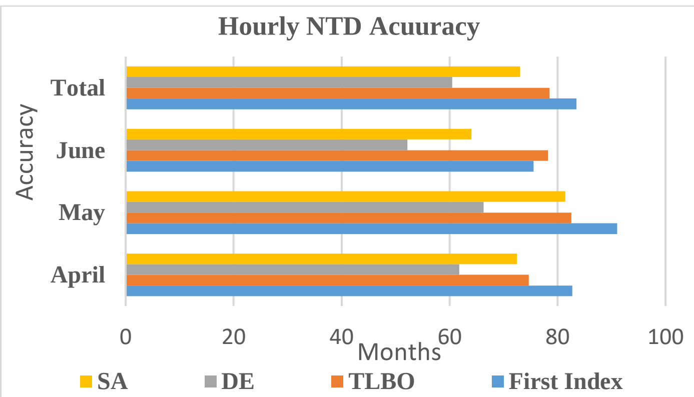

FIGURE 5. Month wise NTD accuracy and total Pre-monsoon accuracy

Figure 6 depicts classification accuracy month-wise and total accuracy in pre-monsoon months for TD+NTD days. DE algorithm did not perform well and has the lowest accuracy for April, May, June, and total accuracy in pre-monsoon. SA algorithm is better than DE algorithm but having low accuracy of classification as compared to TLBO model and first index value method in April, May, and June month and total pre-monsoon months. TLBO algorithm has better accuracy than the other two algorithms namely, DE and SA. TLBO has a lower accuracy than the first index method except for June month. Figure 6 depicts that the first index value method has better performance than the other three methods except for June month for hourly TD+NTD classification.

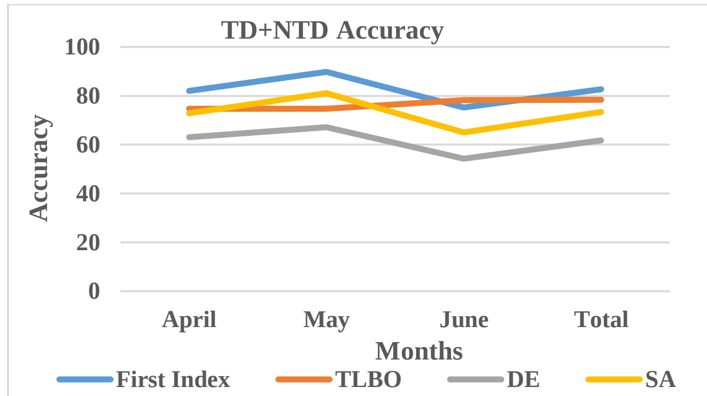

FIGURE 6. Month-wise overall (TD+NTD) accuracy and total pre-monsoon month accuracy.

It is clear from Figure 4 that the first method has the lowest accuracy of classification of incidence of TD for April, May, June, and total pre-monsoon month as compared to the other three algorithms. Thus TLBO shows better performance than the other three methods in classification of hourly incidence and non-incidence of thunderstorms of 2018 atmospheric data which is also validated with April 2019 and 2020 data. The validation TLBO again shows the best performance which is depicted in the Tables 7 and 8.

TABLE 7. Result for April month of 2019 hourly data using a different method

Months	TD	Classified TD (Accuracy)	NTD	Classified NTD (Accuracy)	Total Accuracy
First index	41	28 (68.29)	806	723(89.70)	88.66
TLBO	**41**	**30 (73.17)**	**806**	**677(83.99)**	**83.47**
SA	41	30 (73.17)	806	667(82.75)	82.29
DE	41	35 (85.36)	806	598(74.19)	74.73

TABLE 8. Result for April month of 2020 hourly data using a different method

Months	TD	Classified TD (Accuracy)	NTD	Classified NTD (Accuracy)	Total Accuracy
First index	32	15 (46.87)	470	435 (92.55)	89.64
TLBO	**32**	**23 (71.87)**	**470**	**420 (89.36)**	**88.24**
SA	32	23 (71.87)	470	417 (88.72)	87.64
DE	32	27(84.37)	470	399 (84.89)	84.86

Table 7 shows the result of the validation of the heuristic equation with April 2019 hourly data. In April month, the total number of TD and NTD are 847 days, out of which 41 days are TD while 806 days are NTD. The proposed method and other three optimization techniques namely TLBO, SA, and DE are applied on April 2019 hourly data. All these methods achieve good accuracy in the overall classification. DE has the highest accuracy for TD but the lowest accuracy in NTD and overall performance among the four techniques. DE is not a good choice as the best classifier among all classifiers. Thus, the first index and DE are not suitable for validating hourly atmospheric data. The proposed first method has the lowest classification accuracy for TD days. It is also not suitable for classification. The accuracy for TD for TLBO and SA techniques is the same. TLBO gives better performance than SA in terms of NTD and overall accuracy. This result is that TLBO is more suitable for the classification of April 2019 hourly data.

Table 8 depicts the result of the validation of the heuristic equation with April 2020 hourly data. In April month, the total number of TD and NTD days were 502, out of which only 32 TD and 470 NTD. The first index method achieves the highest accuracy for NTD and TD+NTD (Total) classification, but for TD it is not very appropriate. So, the first index cannot be declared as the best classifier. DE has the lowest accuracy among all techniques. Although DE seems to be the best classifier because for TD, NTD, and TD+NTD it gives approximately accuracy 84%. Not one classifier achieves as such the highest accuracy in TD classification as in shown in Table 8. As it is also clear from Figures 5 and 6, DE classification for NTD and TD+NTD is low as compared to other classifiers. So, DE cannot be the best choice for a classifier. TLBO and SA have the same TD classification accuracy, but NTD and TD+NTD classification accuracy are different and higher as depicted in Table 8. Thus TLBO is the best classifier among all proposed techniques.

VI. CONCLUSION AND FUTURE DIRECTION

Results and discussion revealed that figures (Figure 3 a-d) lead to a better understanding of atmospheric parameter variation before, during, and after the incidence of thunderstorms. These parameters contribute to the generation of the heuristic equation that is used in the classification of hourly incidence of thunderstorms for two years data from 2016 to 2017. Four indices have been derived from the proposed heuristic equation with only the hourly incidence of thunderstorms of two years data from 2016 to 2017. Table 3 to Table 6 shows the result of the application of heuristic and indices values to the hourly year 2018 atmospheric data. Classification of hourly incidence and non-incidence of thunderstorm of 2018 pre-monsoon data using the first index value has the highest overall (TD +NTD) accuracy than any other three indices values. TD classification accuracy using the first index value is not very appropriate. Although TLBO has lower NTD and overall (TD+NTD) accuracy than the first index, TD classification accuracy is higher than the first index value method. Thus, TLBO is better suited for classification using the first index value model. DE generates a third index value. This index value and cost function are forecasted in 2018 hourly atmospheric meteorological data to classify the incidence of TD day and NTD. It is depicted from Table 5 that DE has overall (TD+NTD) and NTD classification accuracies of 60.53% and 61.84% respectively, although TD accuracy is high. In June month, DE has 54% and 52% accuracy of overall (TD+NTD) and NTD respectively. Therefore, DE is not acceptable. SA has lower overall and NTD classification accuracy as compare to TLBO, and first index value. Thus from the above discussion, it can be concluded that TLBO is the most suitable index for the classification of hourly atmospheric pre-monsoon 2018 testing data. This result is validated with hourly 2019 and 2020 atmospheric meteorological data.

The proposed equation and indices are validated with the hourly 2019 and 2020 atmospheric meteorological data. For the validation of indices and equation, TLBO shows better results than the first index method by 25% due to high TD accuracy. Thus TLBO performs better than the first index method. TLBO is also a better performer than the other methods DE and SA which is shown in Table 7 and Table 8. TLBO is also better than DE and SA in terms of overall performance as shown in Figure 6. DE has the worst performance in overall accuracy. Thus, TLBO has better performance than the other three indices.

Based on the heuristic and its indices along with the daily atmospheric parameter variation pattern, it will able to improve for nowcasting purposes. Although this paper is not finding the nowcast of thunderstorms, it helps in nowcasting of thunderstorms which can be possible, if all atmospheric surface parameter variations and indices values are used. This proposed method can be applied to other study areas too if they can form their own heuristic equation based on the correlation of atmospheric parameters and check the variation of atmospheric parameters. In the future, an expert system can be developed using an index of each parameter such as temperature, pressure, humidity, and wind speed along with the index value of the proposed heuristic equation and pattern of variation of atmospheric parameters.

Conflict of Interest: Author Kanchan Bala declares that she has no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

[1] Aruna, S. Kalra, "Review of the teaching-learning based optimization Algorithm" *Indian Journal of computer science and Engineering (IJCSE).* 2017, pp. 319-323.
[2] K. Bala, D.K. Choubey, S. Paul. "Soft computing and data mining techniques for thunderstorms and lightning prediction: A survey", in *International Conference on Electronics and Aerospace Technology (ICECA), IEEE*, 2017, pp. 42-46.
[3] K. Bala, D.K. Choubey, S. Paul, M. Ghosh, "Classification techniques for thunderstorm and lightning prediction: A survey", *Soft-Computing Based Nonlinear Control System Design,* 2018, DOI: 10.4018/978-1-5225-3531-7.ch001.
[4] K. Bala, S. Paul, M. Ghosh, "Thunderstorm prediction using soft computing and wavelet", in *Proceeding of the Second International Conference on Microelectronics, Computing & Communication Systems (MCCS 2017)* 2017, pp. 109-118.
[5] D. Bertsimas, J. Tsitsklkis, "Simulated Annealing", *Statistical Science* vol. 8 no. 1, 1993, pp.10-15.
[6] S. Chaudhuri, A. Middey, "Applicability of bipartite graph model for thunderstorms forecast over Kolkata," *Advances in Meteorology,* 2009, pp. 1–12.
[7] D. Chatterjee, H. Chakrabarty, "Application of machine learning technique to predict severe thunderstorm using upper air data", *International Journal of scientific and engineering Research.*
[8] S. Chaudhuri, A. Middey, "Adaptive neuro-fuzzy inference system to forecast peak gust speed during thunderstorms", *Meteorology and Atmospheric Physics,* pp. 139–149, 2011.
[9] S. Chaudhuri, A. Middey, "Nowcasting thunderstorms with graph spectral distance and entropy estimation", *Meteorological Application,* vol. 18 no. 2, pp. 238–249, 2011.
[10] S. Chaudhuri, "Predictability of severe thunderstorms with fractal dimension approach", *Asian Journal of Water, air &. Environmental Pollution,* vol. 7 no. 4, pp. 81–87, 2010.

[11] S. Chaudhuri, "Convective energies in forecasting severe thunderstorms with one hidden layer neural net and variable learning rate backpropagation algorithm", *Asia-Pacific Journal of Atmospheric Sciences,* vol. 46 no. 2, pp. 173–183, 2010.
[12] S. Chaudhuri, "A probe for consistency in CAPE and CINE during the prevalence of severe thunderstorms: statistical-fuzzy coupled approach", *Atmospheric, and Climate Sciences*, vol. 4 no. 1, pp. 197–205, 2011.
[13] S. Chaudhuri, "Implementation of intuitionistic fuzzy logic to assess the predictability of severe thunderstorms". *Asia-Pacific Journal of Atmospheric Sciences,* vol. 47, no. 5, pp. 477–483, 2011.
[14] J. L. Cintineo, M. J. Pavolonis, J. M. Sieglaff, D.T. Lindsey, "Probabilistic nowcasting of severe convection", *National Weather Association Annual Meeting, Madison W1,* 2012.
[15] Colquhoun, J. R., "A decision tree method of forecasting thunderstorms and tornadoes", *Weather forecasting*, vol. 2, pp. 337-345, 1987.
[16] N. Desai, "Mechanism of Nor'weters of Bengal", *Indian Journal of Meteorology Geophysics*, pp. 74-76, 1950.
[17] V. B. Dhawan, A. Tyagi, M. C. Bansal, "Forecasting of thunderstorm in pre-monsoon over north" *India, Mausam*, vol. 59, no. 4, 433-444, 2008.
[18] A. Elshorbagy, S. P. Simonovic, U. S. Panu, "Performance evaluation of artificial neural networks for runoff prediction", *Journal of Hydrologic Engineering ASCE,* vol. 5, no. 4, pp. 424–427, 2000.
[19] A. K. Fernando, A. W. Jayawardena, "Runoff forecasting using RBF networks with OLS algorithm", *Journal of Hydrologic Engineering*, vol. no.3, pp. 203–209, 1998.
[20] W. W. Hsieh, T Tang, "Applying neural network models to prediction and data analysis in meteorology and oceanography", *Journal, Bulletin of the American Meteorological Society,* vol. 79, pp. 1855–1869, 1998.
[21] https://en.wikipedia.org/wiki/Ranchi.
[22] https://www.statisticshowto.datasciencecentral.com/normalized/
[23] https://en.wikipedia.org/wiki/Differential_evolution
[24] https://www.intechopen.com/books/artificial-intelligence-emerging-trends-andapplications/ differential-evolution-algorithm-in-the-construction-of-interpretable-classification-models
[25] https://www.britannica.com/science/thunderstorm
[26] Z. Huang, Y. Chen, "An Improved Differential Evolution Algorithm Based on Adaptive parameter", *Journal of Control Science and Engineering,* 2013.
[27] R. A. Jundi, M. Yasen, N. A. Madi, "Thunderstorms Prediction using Genetic Programming" in *2nd International Computer Sciences and Informatics Conference (ICSIC)*, 2017.
[28] Kessler, Edwin, Thunderstorm morphology and dynamic, U.S Department of commerce. National Oceanic and Atmospheric Administration, Environmental Research Laboratories, vol 2, 1982.
[29] R. R. Lee, J. E. Passner, "The development and verification of TIPS: an expert system to forecast thunderstorm occurrence", *Weather Forecast*, vol. 8, pp. 271–280, 1993.
[30] K. Lilly, "Numerical prediction of thunderstorms has its time come?", *J R Meteorology Society*, vol. 116, pp. 779–798, 1990.
[31] J. Litta, A.J., S. M. Indicula, and U. C. Mohanty, "Artificial neural network model in prediction of meteorological parameters during pre-monsoon thunderstorms". *International Journal of Atmospheric Science*, 2013, http://dx.doi.org/10.1155/2013/525383.
[32] J. Neumann, "The thunderstorm forecasting system at the Kennedy, Space Center". *Journal Applied Meteorology,* vol. 10, pp. 921–936, 1971.
[33] W. McCann, "A neural network short-term forecast of significant thunderstorms", *Weather Forecast* vol. 7, pp. 525–534, 1992.
[34] Marzban, G. Stumpf, "A neural network for damaging wind prediction", *Weather Forecast*, vol. 13, pp. 151–163, 1998.
[35] C. Marzban, A. Witt, "A Bayesian neural network for severe hail size prediction, *Weather Forecast*, vol. 16, pp. 600–610, 2001.
[36] I, Orlanski, 1975. "A rational subdivision of scales for atmospheric processes", *Bulletin of the* American *Meteorological Society*, vol. 56, no. 5, pp. 527–530, 1975.
[37] L. Ping, Q. Tao-rong, L. Yu-yuan, "The study on the model of thunderstorm forecast based on RS-SVM", *Journal of Convergence Information Technology*, vol. 8, no. 10, pp. 66-74, 2013.
[38] N. Ravi, U.C Mohanty, O.P. Madan, R.K. Paliwal. "Forecasting of thunderstorm in the pre-monsoon season at Delhi", *Meteorology Applied*, Vol. 6, pp. 29-38, 1999.
[39] R. M. Reap, "New-6-H thunderstorm probability forecasts for the west", *NOAA, National weather service Technical Procedures*, Bulletin No. 362, 1986.
[40] R. M. Reap, "Six-hour thunderstorm probability forecasts for Kansas-Okalahoma and the northeast United States", *NOAA, National weather service Technical Procedures*, Bulletin No. 385, 1990.
[41] K. K. C. Reddy, P. P. Anisha, P. L. V. Narasimha, 2014. "Detection of thunderstorms using data mining and image processing", in *Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT),* pp. 226-231, 2014.
[42] P. P. Rajurkar, U.C Kothyari , U. C. Chaube, "Modeling of the daily rainfall-runoff relationship with the artificial neural network", *Journal of Hydrologic Engineering*, pp. 96–113, 2004.
[43] N. Rasmussen, D. O. Blanchard, "A baseline climatology of sounding-derived supercell and tornado forecast parameters", *Weather Forecast*, vol. 13, pp. 1148–1164, 1998.
[44] R. V. Rao, "Review of application of TLBO algorithm and a tutorial for beginners to solve the unconstrained and constrained optimization problems", *Decision Science Letters* pp. 1–30, 2016.
[45] U. Saha, A. Maitra, S. K. Midya, G. K. Das, "Association of thunderstorm frequency with rainfall occurrences over an Indian urban metropolis", *Atmospheric Research*, vol. 138, pp. 240-252, 2014.
[46] P. Schultz, "Relationship of several stability indices to convective weather events in Northeast Colorado", *Weather Forecast*, vol. 4, pp. 73–80, 1989.
[47] A. Srivastava, M. Kumar, A. K. Dwivedi, "Studies of Atmospheric parameters during monsoon period associated with severe cyclone Aila and lightning occurrence over Ranchi", *Internal Scholarly Research Network,* 2012, doi:10.5402/2012/975790.
[48] H. M. Stone, "A comparison among various thermodynamic parameters for the prediction of convective activity", Part I, NOAA Technical Memorandum, NWS-ER 68, Washington DC, 1985.
[49] R. Storn, K. V. Price, "Differential Evolution - A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces", Technical Report TR-95-012, International Computer Science Institute, March 1995.
[50] STORM Science Plan: "Severe Thunderstorm Observations and regional Modeling (STORM) Programme Science Plan", Dept. of Science and Technology, Government of India, 118, 2005.
[51] C. Thomas, and T Foken, "Flux contribution of coherent structure and its implications for the exchange of energy and matter in a tall spruce canopy", *Bound-Layer Meteorology,* vol. 123, no. 2, pp. 317-337, 2007.
[52] S. Tokar, M. Y. Markus, "Precipitation–runoff modeling using artificial neural networks and conceptual models", *Journal of Hydrologic Engineering ASCE,* vol. 5 no. 2, pp. 156–161, 2000.
[53] B. Tyagi, A. N. V. Satyanarayana, "Assessment of turbulent kinetic energy budget and boundary layer characteristics during pre-monsoon thunderstorm season over Ranchi", *Asia-pacific journal of Atmospheric Sciences,* vol. 49, no. 5 ,pp. 587-601, 2013.
[54] B. Tyagi, A. N. V. Satyanarayana, M. Kumar, N. C. Mahanti, "Surface Energy and radiation budget over a tropical station: an observational study", *Asia-pacific Journal of Atmospheric Sciences,* vol. 48, no. 4, pp. 411-421, 2012.
[55] http://www.tsgc.utexas.edu/stars/tstypes.html
[56] http://en.tutiempo.net/records/verc.
[57] H. Kamangir, W. Collins, P. Tissot, and S. A. King, "A deep learning model to predict thunderstorms within 400 km2 South Texas domains" Meteorological Application, 9 March 2020. DOI: 10.1002/met.1905.

KANCHAN BALA received B.E (CSE) from Rajasthan University and M.Tech (CS) from Banasthali University, Jaipur, Rajasthan. She was a Project Assistant in the IC design department of CEERI-CSIR, Rajasthan. She is currently an Assistant Professor with Govt. college of DST, Bihar. Her research interest includes Image Processing, Soft Computing, Data Mining, Machine Learning, and Microwave Tube.

SANCHITA PAUL received the B. E degree in Computer Science and Engineering from Burdwan University, West Bengal, India, in 2004 and an M. E degree in Software Engineering from Birla Institute of Technology, Mesra, Ranchi, India, in 2006. She received her Ph. D degree in Computer Science and Engineering from Birla Institute of Technology, Mesra, Ranchi, India, in 2012.

She is currently an Assistant Professor with the Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India. She has published 8 SCI papers, 20 SCOPUS papers, and 30 other papers. Her research interest includes Cloud Computing, Machine Learning, Internet of Thing, Big Data analytics, Soft Computing, and Brain Signal Processing.

Mili GHOSH received M.S.C (Physics) from Bhagalpur University, M.Tech (Remote sensing) from B.I.T Mesra, Ranchi, and Ph.D. from IIT Roorkee. She is currently an Assistant Professor with B.I.T, Mesra in the Remote Sensing Department. She has 8 years of teaching and 14 years of research experience. She has published several SCI papers, SCOPUS papers, and other papers. Her research interest includes environmental modeling, Planetary Remote Sensing. Her awards and honors include the MHRD fellowship in M.Tech and Ph.D.