A Study of the Prediction Algorithm for Identifying Reports of Skin Diseases in Hyderabad

R. Sukruta[1] Udaya Sri Gidigam[2]
[1,2] Assistant Professor, Department of CSE, Geethanjali College of Engineering and Technology, Cheeryal (vi), Hyderabad, Telangana, India.

Abstract:
Hyderabad is highly affected by climate change and is reported to be a highly prone skin disease-endemic area. This study investigates the association between skin diseases and climate factors. For selecting the best-fit climate prediction method for skin diseases occurrence in Hyderabad city, we have considered 3 different machine learning regression models namely: Poisson Distributed Lag Model [PDLM], Seasonal Autoregressive Integrated Moving Average Model [SARIMA] and other model Standard Multiple Regression [SMR] directed toward investigate the relationship between skin diseases and climate attributes incident during the time period 2000 to 2018. We verify the models lag predicting skin diseases for the time duration of January month to December month 2019 using the MAPE. ROC curves were considered to examine the prediction rate of a skin disease outburst. The results show that relative humidity and temperature are significant parameters which promote skin diseases where there is no rainfall effect. The PDLM model presents the finest fitting prediction of skin disease occurrence and identification of an outbreak when analyzed for a 6, 9, and 12 month time. Nevertheless, the SARIMA algorithm enacts a better prediction of skin disease occurrence for a short 3 month time period. The standard multiple regressions present a high loss prediction of skin disease incidence. From our results we are encouraged to carry out an extensive analysis to validate and examine the modelwith more data in Hyderabad city and contribute in prevention and control of skin diseases at an early stage.

Keywords: Poisson Distributed Lag Model, Standard Multiple Regression, SARIMA algorithm, Standard Multiple Regression

1. Introduction:

Skin Cancer related Diseases is the most widespread disease during summer and winter around the world. Skin cancer occurrence has increasedsince 3 decades,the skin is the organ most exposed toenvironmental UVR and it is roughly calculated that 40 lakhs skin diseases take place each year. Nearly2 crore people reside in skin cancer effected countries[1].Tropical areas America, Australia, India, central Asia andPacific locations are mostly bundled with skin diseases, and in which southeast india and the central Asia area bear nearly 30% of the presentworldwide diseases bundle due to skin[2]–[4]. Some type of Skin are vulnerable toor higherTemperature that result from worldwide global warming which cover an increasingoccurrence of skin diseases [5]–[7]. The annual investigation by has confirm the increase of skin diseases at higher temperature. In-room temperature UV radiations getcarcinogenic. In human beings, theoccurrence of squamous cell sabal and cell carcinoma rises by 45% and 2.4%, respectively, per 1C rise in atmospheric temperature [8]. These estimates correlate to a rise in the virtual UV dose of around 1.5% per 1C rise. Skin cancer can give rise to a change of skin color with more than 6 typesof skin diseases. At present there are specific medicines available for treating skin cancer and vaccine to prevent skin disease. the most constructive way to control this type of disease is through control of population and proving prevention medicine by forecasting occurrence of diseases using advanced machine learning techniques at particular area [9]–[12].
A prior warning method, utilize prediction models asan effective method for skin diseases outbreak control and preparedness. Climatic attributes likehumidity, rainfall, and temperature have been examine broadlyin studies for their prospective prediction task as an initial warning method to prevent skin cancer as a climate-sensitive disease[13][8], [14]–[17].Previous studies reveal strong association or Interconnection between skin diseases and climate factors like humidity, rainfall and temperature[1], [9]–[12], [18]–[22]. Nevertheless, the skin disease prediction method has diverse due totheir very complexity, research location and methodology[7].Therefore, recognition of the best fit prediction algorithms may be suitability to develop and validated for a particular geographical location.[23][24], [5], [25]

Hyderabad capital of telangana state is prone to be the most vulnerable for climate change and is reported to be with highest inflatedoccurrence of skin diseases in state. Nevertheless, there is a model by using climate attribute to assist evolving and initial warning method at effected location. The main objective of this study is. (1) To inspect the relationship or among climate attributes and skin diseases.(2) To examine the finest algorithm for predicting skin disease occurrence using climate attributes in and around the Hyderabad

2. Data Collection

In this study, data of the quarterly andmonthly number of reported skin cancer cases in Hyderabad city from January month 2000 to December month 2018 was collected from the disease controldepartment of Hyderabad medical center (HMC).HMC is the chief agency in charge of analysis and obtaining reported data for prevention of contagious in the city. Daily weather data from 18 Jan 2000 to 24 December 2018 was collected from IMD IndianMetrologicalDepartment Pune India. The data was collected throughsatellite sensors at Hyderabad city including daily Maximum, Average and Minimum Temperature, dailyrainfall, humidity. The obtaincomplete data has to be converted to average, meantemperature and maximum and increasing rainfall and monthly humidity for investigation.

2.1. Data Analysis

In theinitialstage, the relation amongmetrological attributes was inspect by utilize different machine learning regression methodsStandard Multiple Regression, Poisson distributed lag model and Seasonal autoregressive integrated moving average .Above 3 modelsare most commonly applied in a literaturethatto assesses the association between climatic changes and skin diseases, in the next step, the algorithms were validated to examine the machine learning models for prediction of skin cancer incidence and skin diseases outbreak.

2.2. Standard MultipleRegression

The initialphase of the investigation applied crosscorrelation technique to analysis the relationship among the skin diseases outbreak and various lags of the exploratory attributes like rainfall, humidity, temperate to regulate the best time interval for the initialmodel. which is expressed in Eq. 1.

$$In(S) = Y_o + Y_1T_{min} + Y_{max} + Y_3H_{min} + Y_4H_{max} + Y_5R \quad .....Eq\ (1)$$

Where S is the monthly count of skin diseases, Tmin and $T_{max}$are monthly minimum and maximum of temperature; H_{min} and H_{max}asmonthly min and maxhumidity and rainfall is denoted as R.

To avertthe remarkable co-linearityof associated lags involves in the machine learning methods, we design aadvanceanalytical attribute using the abrade value of two lasheswhich is possess the veryhighest correlation among the predictedattributes, Prior to utilizing the attribute in the last approach. By using a few statistical tests like Kolmogorov Smirnow,augmentedDickyfalter test, and Shapiro-wire are examined normality andstability to make it very clear that there was no single violated beliefs for SMR. Furthermore validation of skin diseases,wehavebuild a multiple regression algorithm using collected data during the study 2000 to 2018 and then we have train the model to predict outbreak case from 2018.

2.3. SARIMA Model

TheSARIMA algorithm is first proposed by author Jenkins and we have try to hybrid the model and expanded using skin disease occurrence data from 2000 to 2018.Than the train algorithm was used to forecast skin disease occurrencein 2018.

ARIMA algorithmwith S as seasonal SARIMAtry to examination seasonal period, which is express by SARIMA[x, y, z] [X, Y, Y] $_s$,is given through model Eq-(2)

$$Pt = \frac{\Theta z(B)\Theta y(Bs) + at\phi x(Bs)}{\Theta x(Bs)\Theta(S)(1-B)d(1-Bs)D} + x. \quad \text{.....Eq (2)}$$

Where $\Theta_x(B^s)$ is the seasonal auto regression(AR) $\Theta_x(B)$ is the AR operator, $\Theta_y(B)$ moving average(MA) Θ_y is seasonal and ordinary differentcomponents, X external variable Pt is dependent variable and a_t is white-nose.

SARIMA model involves many proceeding stepsthrough the study. Initial, the variancevalue of skin disease count was stimulatedby applyinga natural lag variation of dataset.By testing with Dicky fuller metrics. By using time series data mean is stabilize the selection of a sequence of different [D and d] attributes was found on time series plot and comparing with their standard deviations the time series with more stabilize mean is consider. Next phase, the examination of the order of non-seasonal and seasonal [AR(X,x),MA(Y,y)] attributes were managed using theACF and PACF.With the help of a few feasible methods was screened using AIC for the goodness of best-fit comparison among distinct methods.Depends On the AIC lower score algorithm is selected for prediction. At leastthe best model, fine-tune for climate attribute,which was remarkably associated with skin disease occurrence,was develop for the time period from 2000 to 2018.Then we can predict skin disease cases in 2019.

2.4. PDLM Model

A Distributed lag algorithm [DLM] merge with the passion time series model was adapted to predict the consequence of climate attribute on skin diseases and to build a prediction model. Construable of a number are involved to build the model. Initially,forchecking for very long duration seasonality and trend of collected data, we create a pair of basis variable it's a part of main time variable,make use of 3 degree of freedom annually, In creating the spline basis, 4 knots annually was used 8 year multiply by 4 and minus by 1 to generate mathematical spline function. Even though is no agreement to choose a degree of freedom. From the previous studies, 3 years are consider and pound as a balance between if adequate manage trends and seasonality in monthly datafrom which gives more precise informationbywhich we can estimate the exposure effects. All they have variable involve in the PDLM model for both multivariateregression and bivariant analysis subsequent. Next phase auto regression

of skin disease cases recommended very strong autocorrelation between current and past cases; in PACF [fig 2B] a cut off for 1 month of a lag time. Few studies carried out research on climate and infectious diseases strongassociation among them.Consequently, in this study,we inspect lag time extend from 2 to 3 monthly and to choose optimal lag time for the prediction model. In the third phases due to delay, it affects the independent variable on skin diseases that were determined by systematic literature review and cross-correlation function.Therefore we formulate a Poisson modelby interconnecting to identify autocorrelation order. A backward phase-wise analysis was considered for selecting the final model in the level which is approximant vale 0.04.with the help of AIC value as lowest help to build the best predicting model for skin diseases in upcoming years.

$$In(\mu_t) = \beta_o + \sum^{2}_{x=1} \beta_{t-x}\, AR + \sum^{p}_{L=1} \beta_{tl}(Temp_{l,p}) + \sum^{p}_{L=1} \beta_{hl}(Humid_{l,p}) + \sum^{p}_{L=1} \beta_{rl}(Rain_{l,p}) + S(t) + \varepsilon_t + In(POP).$$

Where $In(\mu_t)$ is average predict skin diseases at model; β_r is constant number of skin diseases cases; AR_{t-x} is skin diseases cases at lagperiod k, $Temp_{l,p}$, $Humid_{l,p}$, $Rain_{l,p}$ are matrix for DLM to rainfall, humidity, temp respective; β_{t-x} is parameter of auto regression at lagperiod k; L is lag month; β_{tl}, β_{hl}, β_{rl} are coefficient for $Rain_{l,p}$,$Humid_{l,p}$, $Temp_{l,p}$ in lag period; p consider to be maximum lag; S(t) is the Spline function. Which help to control seasonally trends.

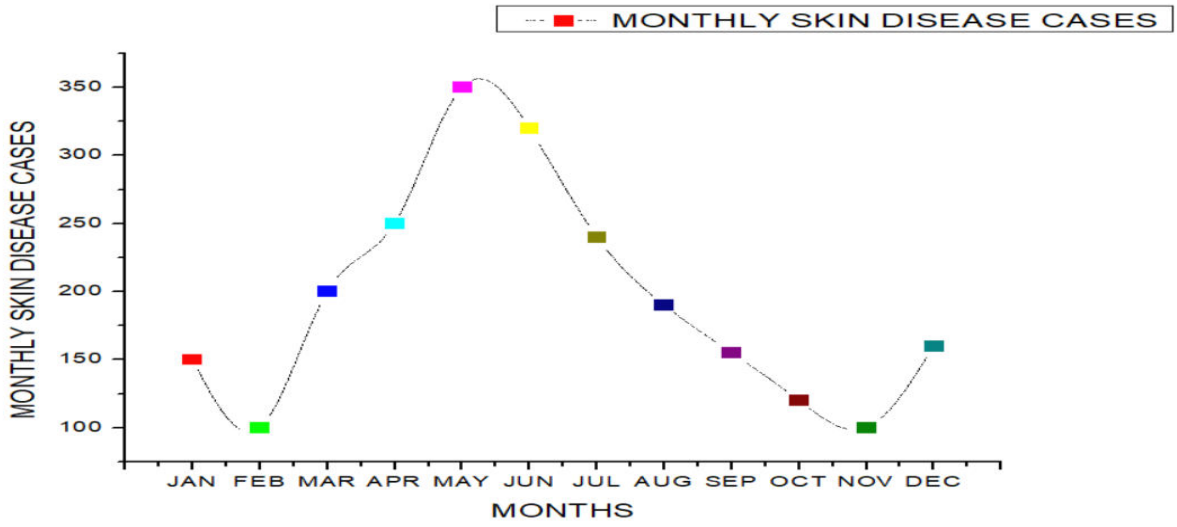


Fig.1 Monthly counts of skin disease cases (2000-2018) in Hyderabad city.

2.5. Sensitivity Test and Model Validation.

The prediction accuracy of each and every model they evolve was abscess by comparing precision outbreakagainst real worldrecognized skin diseases cases, With the help ofsome metrics like mean absolute percentage error[MAPE] was applied to verify and validate predicting model for the different time period 3 lags,6 lags,9 lags

and 12 lags and models with lower MAPC value consider as best fit model for further analysis.

$$MAPE = \frac{1}{n} \sum^{n}_{t=1} | x_t - \hat{x} / x_t |$$

Where n is a number of the month, x_t and $\hat{x}$ are observed cases and predicted cases no. of skin diseases in a month.

Our Study indicate, we have examine outbreak of skin disease cases increased 80th% of skin disease cases recorded thought out the time interval from 2000 to 2018. Wear additionallythe optimal algorithm used as ROC to calculate and analyze the sensitivity ofthetrain model in identifying their skin diseases outbreak during study, The ROC metric is used to analyze True positive Rate(TP) or sensitivity oftrain modelto predict outbreak against the False Positive Rate(FP) [1-specificity]. ROC curve gives a detailed outbreak and non-outbreak of cases.

3. Result:

3.1 Descriptive analysis

There were 6000 skin disease cases in the city of Hyderabad during study. On the mean monthly indecentnumberof skin diseases,instances were 65 withoccurrence rate as 5.4per 99999 person in a month. The occurrenceestimate the rising from summerperiod [Feb – June] to wets period [July – Dec]the excessive monthly occurrence rate 80 per 100000 person month was in April, and lowest occurrence rate is 2.5 per 99999 people was in the ofJanuary. This is show in [Fig 1].

Variables	Minimum	Maximum	Average	Median	Std.Deviation
Skin diseases cases	10	629	127	105	96
Min Temp	23.2	27.4	25	25.1	0.93
Max Temp	29.3	36	32	32..3	1.3
Min humidity	50.9	71	62	62	2.7
Max humidity	90.5	100	95.5	95	2.6
Rainfall	0	445	127	121	109

Table 1 Summary of skin diseases and climate attributes from 2000 to 2018 in Hyderabad.

In regards of climate variable data the range of rainfall varies from 5MM in the monthJanuary to 160 MM in the month of May the mean peak temp varies from 37 °C in the month of Feb& March to 43 °Cin the month of April&May andthe average min Temp varies from 20C to 26C in November, Avg min & max humidity varies from 88%

march to 92% may and from 49% in Nov & Dec to 60% in January. Respectively in table1 show different between dependent &independent variable using statically method.

3.2 Multiple linear regressions

The outcome of autocorrelation occurrenceshows that the heightscorrelationamong skin diseases rate and min temp were initiate at lag1, lag2(r,0.3), maximum temp at lag3,lag4,and lag5(r,0.34 and 0.31), min humidity at lag12 and lag1 (r,0.53,0.51), max humidity at lag3,lag4 (r,0.35,0.37) and relative rainfall at lag2,lag3 (0.53,0.45).based on this a new independent variable build added;$Tmax_{3,4}$,$Tmin_{1,2}$,$Hmax_{12,1}$ and $R_{2,3}$.In table 2 mention the effects of climate attributes ($Tmax_{3,4}$,$Tmin_{1,2}$,$Hmax_{12,1}$ and $R_{2,3}$) on skin disease and shows that gradually decreases in skin diseases instance value β= -0.1 and p value = 0.04 for correlated with 1C° rise in Maximum temp, because 1C° increase maximum humidity value β = 0.04,1MM of relative rainfall(40.002) monthly skin diseases significantly increase with a levelsignificant value of 0.06.

Variable	Root MSE	R-Squared	p-value	95% CI	Coefficients
Model 1					
$T_{min2,3}$ [a]	0.59	0.39	0.07	-0.5-0.009	-0.3
$T_{max4,5}$ [b]			0.7	-0.2-0.18	0.04
$H_{mino,1}$ [c]			0.09	-0.0008-0.0097	0.06
$H_{maxo,1}$ [d]			0.9	-0.060-0.0078	0.010
$R_{1,2}$ [e]			0.0002	0.002-0.007	0.004
Constant			0.3	-2.9-11.9	4.45
Model 2					
$T_{min2,3}$	0.57	0.43	0.06	-0.39-0.03	-0.19
$H_{mino,1}$			0.06	-0.005-0.94	0.06
$R_{1,2}$			<0.02	0.003-0.007	0.005
Constant			<0.02	2.0-9.9	5.89

Table 2 SMR coefficient of the skin disease vs. Min and Max temperature, rainfall, and humidity for 2002 -2011 for Hyderabad city.

The model exhibit that min humidity and min temperature were no statistically significantly related to change in no. of skin diseases. After removing insignificant variables, the remaining independent variable are $Hmax_{2,1}$, $Tmax_{3,4}$, and $R_{1,2}$ were significantly related to skin disease with much greater prediction strength R value = 0.38 compared with previous value = 0.35; RMSE0.49 compare with 0.51 ad second model used to predict skin diseases cases from Jan to Dec 2018.

3.3 SARIMA Analysis

In time series analysis, the lag of skin diseases cases confers the normal distribution i.e. ShapiroWalk Test with P-value 0.51 comparing with original skin diseases case with P-value <0.01 which has the lowest distribution. The plot betweenACF and PACF using collected data sets from 2000-2018 mention in figure.2 [A and B]. The ACF metrics exhibits the very strong seasonal hidden information of skin diseases that has confirmed to add S with seasonal to build SARIMA model (X,Y,Z) with periodic length S is 12 and other is non seasonal (x,y,z).Weanticipating or imagine as skin diseases occurrence has a different seasonal period.

PACFmetics recommended that the value of X between 1 and 2 with the period interval of 12 month lags, as PAC was almost Zero at all lags excluding lag3 and lag5. The ACF recommends a moving average value Y from 3 to 5; mention that autocorrelation is all zero expect 4 with the period time of 12 months (Fig 2). Next differencing value d = 1 on plot ACF shoes a significant cut off 1 month lag among 12 months. (Fig.2 C), a basic test indicates that there is significantstability in data using Dickey- fuller Test, P<0.0.1 which is compared with original data i.e. P>0.4.Nevertheless, from the ACF still shows the seasonal patterns at

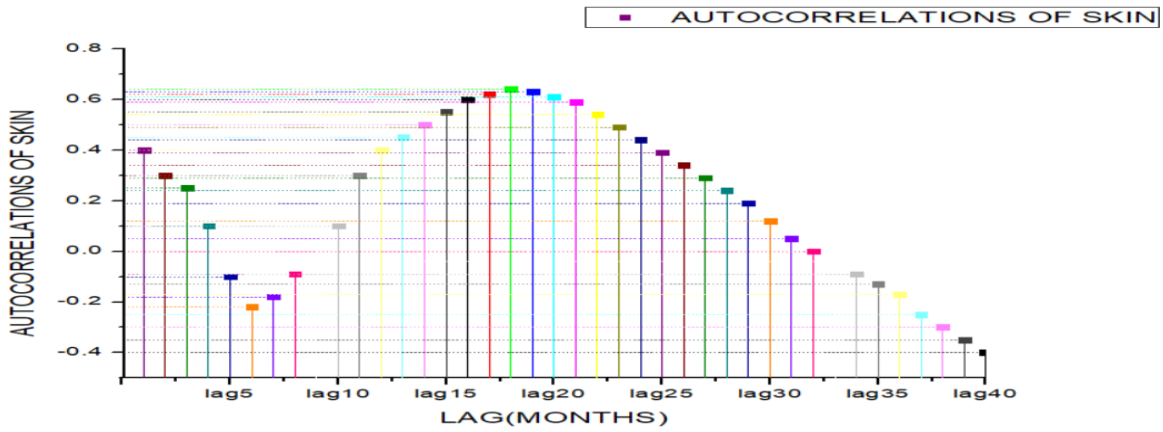


Fig.2. Autocorrelation functions calculate using log transformed skin disease cases from 2000 to 2018 in Hyderabad.

In Fig.2 with 12 lags, clearly shows better to add seasonal differenceD = 1 month. Throughout 12 months autocorrelation shows positive significance, were SAR considers the value of x = 1.Table 3 exhibits the value of VIC and AIC for the SARIMA model. Selecting different x and y the model which has lowest BIC, AIC consider being best model as SARIMA (1, 1, and 1) (1, 1 and 0)$_{12}$ alter as average temp,rainfall and humidity (BIC 38,AIC 29) respectively.The investigation of residuals does not have any significant in ACF,shown in Fig.3A. The inverse graph shows a reasonable probability of residual in Fig.3B by Ljung- Box test establish residual value statistically not depended on P value greater than 0.04(Fig 3C).On the hand The SARIMA establishes the strong statistically significant result onmonthly temperature as β = 0.19 and p =

0.01, humidity value β = 0.09 and p is <0.01, but rainfall is not significant with value as β = -0.007 and p = 0.2.Then SARIMA model is train to predict skin disease cases at Hyderabad.

3.4PoissonDistribution Analysis

The correlation between skin disease cases and climate factors confer a uniform sine wave oscillation at 6 monthsfor one rotation. 2-5 months are significant for min temp, humidity 2-3 months, 3-6 months interval for a max temp, 0-3 month for max rainfall and humidity. by removing the no significant attribute from the tested model, the final model shows all variable was; 2 auto regression value of skin 2-4 min humidity, lag 2-3of rainfall by adjusting Spline function of the population and time(table4).

Time series Model	BIC	AIC
Simple SARIMA		
(1,1 and 1)(1,1 and 0)$_{12}$	150	138
(1,1 and 2)(1,1 and 0)$_{12}$	153	139
(1,1 and 3)(1,1 and 0)$_{12}$	154	137
Complex SARIMA		
(1,1 and 1)(1,1 and 0)$_{12}$	44	33
(1,1 and 2)(1,1 and 0)$_{12}$	49	36
(1,1 and 3)(1,1 and 0)$_{12}$	54	39

Table 3 Ridings of BIC metrics and AIC metrics by different SARIMA model combination value (x, 1 and y) (1, 1 and 0)$_{12}$.

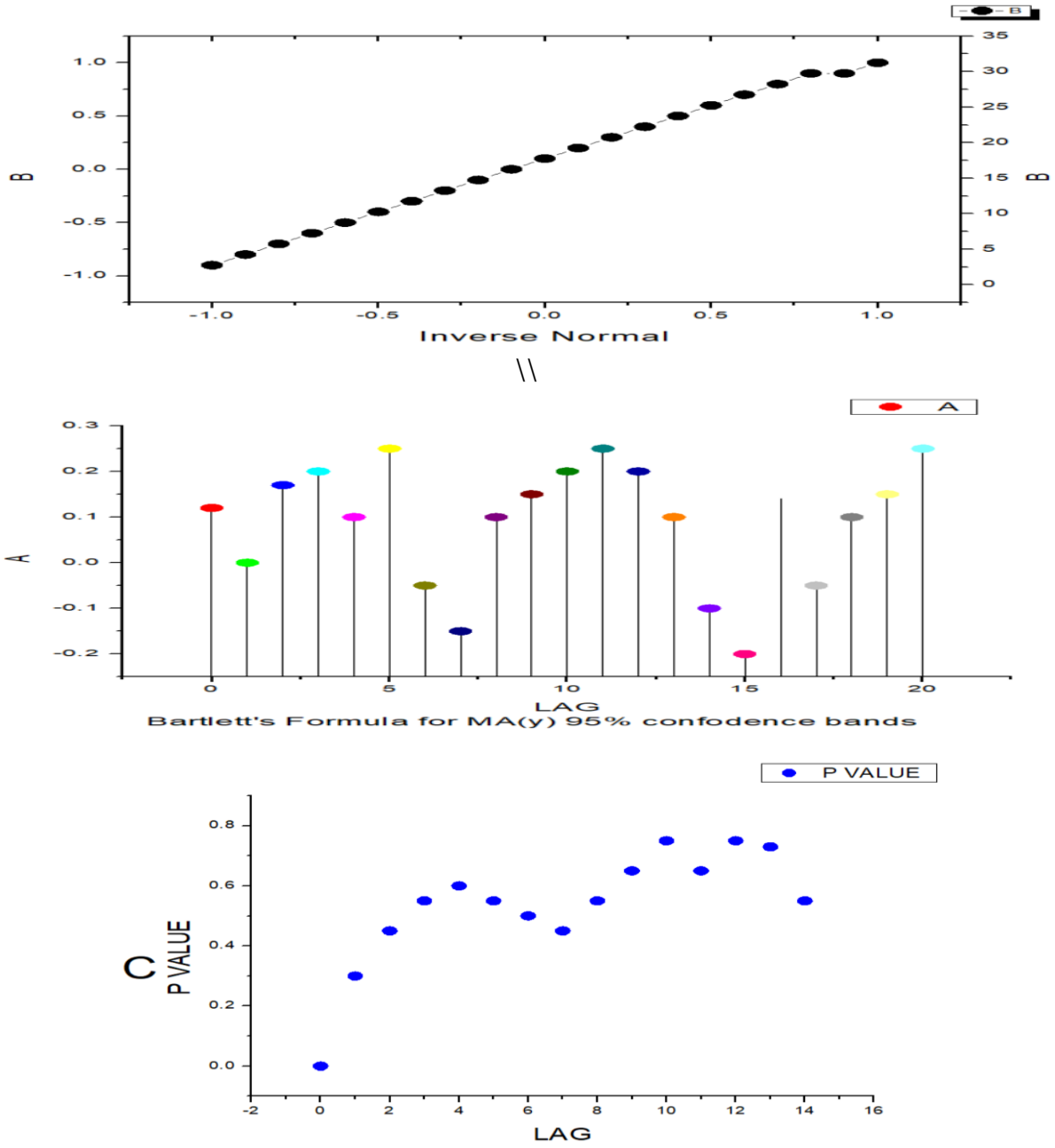


Fig.3.Graphical representation of SARIMA model: [A] and [B] plot of standardized residuals of Bartlett's formula and inverse, and [C] p value.

Variable	P-value	Coefficients	95% CI
Lag 1 Skin cases	<0.0.2	-0.001	-0.00027-(-0.0015)
Lag 2 Skin cases	<0.0.2	-0.002	-0.0036-(-0.0027)
Lag 1 Max Temp	<0.0.2	0.273	0.209-0.318
Lag 2 Max Temp	<0.0.2	0.220	0.168-0.29
Lag 1 Max Humidity	<0.0.2	0.046	0.034-0.058
Lag 2 Max	<0.0.2	0.061	0.047-0.074

Humidity			
Lag 3 Max Humidity	<0.0.2	0.025	0.009-0.0019
Lag 1 Rainfall	<0.0.2	0.002	0.00090-0.0019
Lag 2 Rainfall	<0.0.2	0.002	0.00097-0.0018
Lag 3 Rainfall	<0.0.2	0.008	0.0004-0.0013
Constant	<0.0.2	15.1	10.3-9.9

Table 4 Poisson distributed lag model regression coefficients of skin diseases vs. maximum and minimum temperature, rainfall, and humidity

Maximum humidity and maximum temperature were statistically significant in the ML model. So that variable was removed from the building a final model. The R^2 (0.74) of out Ml model intimated in the previous skin disease cases; min humidity, min temp, rainfall trends and seasonality explained 74 of the variances of the monthly skin disease distribution.

3.5 Validation of 3 models

The model predictsinstancesby using PDLM, SMR and SARIMA algorithmmany observedinstancesare mention in Figure 4. By combining passion with distribute lag model give the best-fit prediction of skin disease instancesfor a period of 12 lag months with MAPE value = 0.07.Next SARIMA model with MAPE = 0.24 and lowest MAPE value = 0.64 found in standard multiple models. For prediction different time interval is consider for better prediction SARIMA model than PDLM for time period of 3 months with MAPE value = 0.02, but the PDLM model shows much better prediction for time period of 6 months with MAPE value = 0.31,9 month with MAPE value 0.31 vs. 0.09, and for 12 lag months with MAPE value = 0.24 vs. 0.07.

ROC metrics in Figure.5 shows that the last model PDLM is consider to be best fit model for diagnosticthe skin disease outbreak (>90thpercentile of skin disease instances) range of sensitivity and CI from 96 to 98% through outbreaksfrom 2000 to 2004, 2006 and 2009. It is estimated that ROC area for the time interval 2000 – 2009 shows that PDLM performed at 98%(95% CI,97-99%) sensitivity between non-outbreak and outbreak period (Fig.5A) other hand SARIMA algorithmrepresent at 90%(95% CI,75-97%) shown in Figure and SMR at 85 (95% CI,55-85%) sensitivity among non-outbreak and outbreak[Fig.5C].

Machine learning Model	Lag Months	MAPE
SMR model	0 to 3 lag	.17
	0 to 6 lag	.31
	0 to 9 lag	.52
	0 to 12 lag	.70
SARIMA model	0 to 3 lag	.04
	0 to 6 lag	.35
	0 to 9 lag	.36
	0 to 12 lag	.29
PDLM model	0 to 3 lag	.23
	0 to 6 lag	.17
	0 to 9 lag	.14
	0 to 12 lag	.11

Table 5 Mean absolute percentage error (MAPE) at different time intervals of predictions.

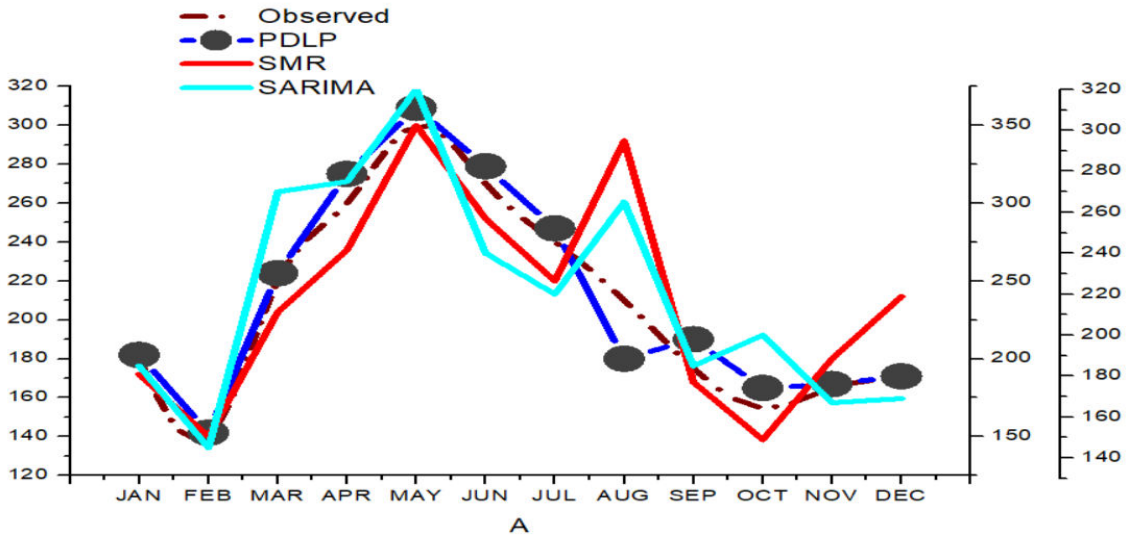
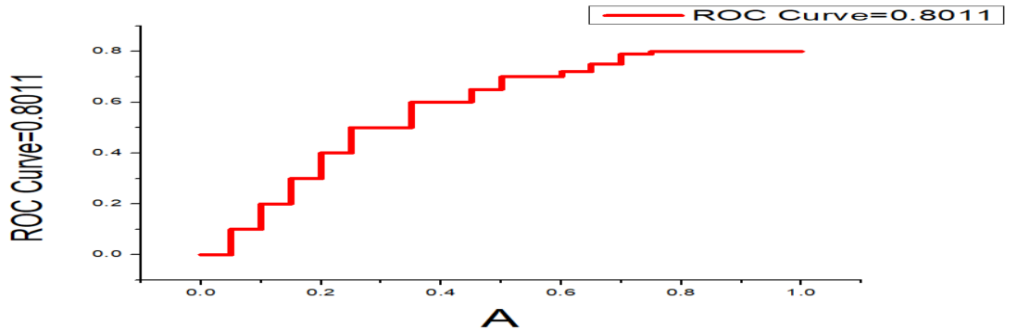


Fig.4. Skin diseases cases in Hyderabadregion for 2019.brown color line:actual values, Red color line SMR; cream color line: SARIMA model, blue color line: PDLM model.



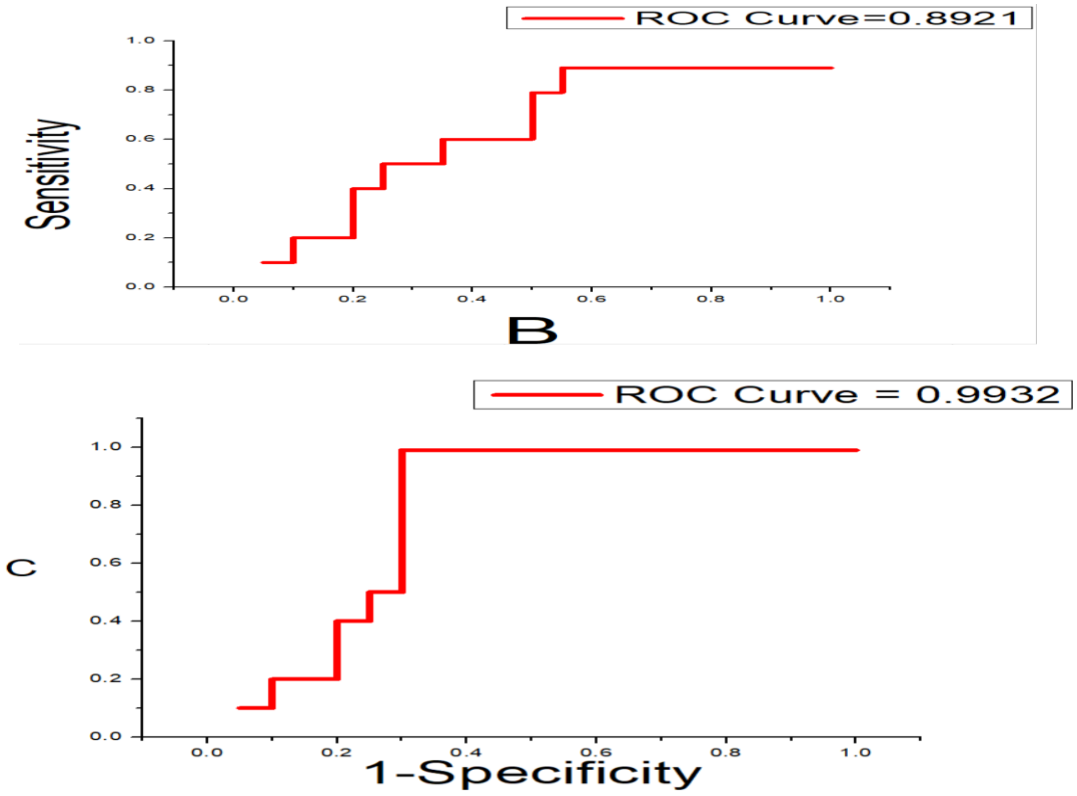


Fig.5. Model accuracy tested by Roc for prediction of skin diseases cases: [A] SMR model, [B] SARIMA model, [C] PDLM model.

3.6. Discussion:

Epidemiological records show that, the relationship among skin disease cases and climate factors was tested by different ML regression modes, and using the best fit model for skin disease prediction was validated by few validation techniques. The final result exhibits that relative humidity and temperature are significantly related to change in skin diseases occurrence confidently throughout the model, while rainfall shows association with skin diseases two of 3 models (PDLM and MLR). The Poisson model exhibits better prediction of skin cases indicate for 6, 9 and 12 months time period and detection of the outbreak[1], [18], [19]. Nevertheless, SARIMA shows thebetter prediction of skin diseases cases for 3 months time interval. The SMR model exhibits a lower prediction rate among the other 2 models in a short time period[26], [27].

The association between skin disease cases and climate factors. From the literature from the previous studies, we have to consider the climate factors Temp, Humidity, Rainfall that impact skin disease transmission. Ina few study rainfall not consider significant factors [17].Some many studies give strong evidence about the association between climate factors and other infectious diseases[28]. The plausible cases of skin disease temp association state as a higher temp increase rate of a skin-related system that leads to 6 types of skin diseases[1][22], [23], [29]. The same relative humidity also

givesa significant association between that which skin diseases expose to outbreak to outside temperature and rainfall was no significant but few studies consider rainfall also.

Prediction model for infectious diseases like Dengue, Zika, Chikungunya etc has been developed and applied by many research groups by using mathematically and statistically which diseases with the problem of small to complex [30].Nevertheless, none of the retrospectivestudies has not attempted to compare the dissimilarity machine learning modelsby experimenting with different datasets while the retrospective studies did it. Through our literature survey, few machine learning algorithms like logistic regression and multiple algorithms notperformed better prediction and autocorrelation metrics fail to find the seasonal trend changes that lead to week and wrong prediction rates [13], [31], [32]. When comparing with the SARIMA model outperform better among all machine learning algorithms. The present study exhibits that PDLM performs much good prediction of skin disease cases when compared to the SARIMA model for 5-12 months. This is because ofthe usage of autocorrelation 2 in the PDLM model vs. 1 in the SARIMA model. By using spline function among PDLM and SARIMA model, we have tried to control seasonal and trends and different lag time periods 3, 6, 9 and 12 months have used for validation. Future works should befocus on a better understanding of the dataset and applyingfew advance machine learning withfeaturing techniques[28], [30], [33], [34].

Some of the limitation of the studies is availability of medical record data for long time that make a time constraint and out study not tried to involve other factors like environmental,demographic and socioeconomic, etiological, education and humanbehavior have been associated with skin diseases incidence Global temperatures continue to rise[8], [35]–[37].Nevertheless, all the above factors can be considered for the long term not in the short term. Some other factors like population growth, air travel and unplanned urbanization should be considered in further studies. Previous studies found that only 70% of any infectious cases are registered near 30% of cases are not register and unreported[8].At some location infectious diseases like dengue,zika,chikungunyaetc are treated by the local clinic which is not reported in thesestudies. Osmania hospital management establishesa prevention and control center that helps to trace the infectious diseases with 24 hours. Reported is consider in our study that helps to analyze the health issue in and around that monthly skin disease cases consistent throughout the study period. Irrespective of limitation,this study gives enough research gaps for applying machine leaning algorithms for prevention and control of any diseases as an early warning sign.In the Hyderabad city, this machine learning devolved method can apply in the preventive local center which can be considered as humidity, rainfall and temperature to predict skin disease cases outbreak and risk level of anyinfectious disease incidence [35], [36]. Using such a model can reduce susceptible individuals. Early preparedness for general hospital

medical stores and needs, spreading prevention measures to the public in advance that control skin diseases can be efficient,adequate,controlled and prevented[8], [37], [38].

4. Conclusion:

Evidence from this study can inform public health interventions that high-level vulnerability of skin diseasescases to a variance of climatic attributes, like humidity, rainfall, temperature. Our studyfurther proposes that climate attributesdo have major connectionswith public health in Hyderabad city. The validation of algorithm methods used has suggested that either PDLM or SARIMA is fit for predicting skin diseases incidence in Hyderabad, SARIMA model shows better prediction rate for early 3 months and other PDLM model presents the better prediction in next time interval i.e.6 to 12 months. High precision in predicting skin disease outbreak of the Distributed Lag Model will be efficientfor prevention or controlskin disease epidemics in Hyderabadcity. We suggest a further examination to validate the model and analyze the likehoodof incorporating weather based skin disease early warning to skin disease case surveillance method.Further, research can transfer the prediction method as a friendly operative tool for use inthe local area. This would enhance early warning methods for prevention of skin diseases in Hyderabad city future alternative socio-economic pathways.

Reference:

1. A. K. Verma, S. Pal, and S. Kumar, “Comparison of skin disease prediction by feature selection using ensemble data mining techniques,” *Informatics Med. Unlocked*, p. 100202, 2019.
2. K. L. Ebi and J. Nealon, “Dengue in a changing climate,” *Environ. Res.*, vol. 151, pp. 115–123, 2016.
3. T. Chakraborty, S. Chattopadhyay, and I. Ghosh, “Forecasting dengue epidemics using a hybrid methodology,” *Physica A*, vol. 527, p. 121266, 2019.
4. R. Conde and G. Sa, “Prediction of dengue outbreaks in Mexico based on entomological , meteorological and demographic data,” pp. 1–14, 2018.
5. V. J. Jayaraj, R. Avoi, N. Gopalakrishnan, and D. B. Raja, “Acta Tropica Developing a dengue prediction model based on climate in Tawau , Malaysia,” *Acta Trop.*, vol. 197, no. February, p. 105055, 2019.
6. Y. Chen, C. Wenhan, M. I. C. Chen, and A. R. Cook, “The utility of LASSO-based models for real time forecasts of endemic infectious diseases : A cross country comparison,” *J. Biomed. Inform.*, vol. 81, no. September 2017, pp. 16–30, 2018.
7. M. Cabrera and G. Taylor, “Spatial and Spatio-temporal Epidemiology Modelling spatio-temporal data of dengue fever using generalized additive mixed models,” *Spat. Spatiotemporal. Epidemiol.*, vol. 28, no. 2019, pp. 1–13, 2020.

8. R. Lowe *et al.*, "Articles Climate services for health : predicting the evolution of the 2016 dengue season in Machala , Ecuador," *Lancet Planet. Heal.*, vol. 1, no. 4, pp. e142–e151, 2016.
9. Y. Peng, Z. Wu, and J. Jiang, "A novel feature selection approach for biomedical data classification," *J. Biomed. Inform.*, vol. 43, no. 1, pp. 15–23, 2010.
10. H. Kaur and V. Kumari, "Applied Computing and Informatics Predictive modelling and analytics for diabetes using a machine learning approach," *Appl. Comput. Informatics*, no. xxxx, pp. 0–5, 2018.
11. J. N. Taroni *et al.*, "MultiPLIER : A Transfer Learning Framework for Transcriptomics Reveals Systemic Features of Rare Article MultiPLIER : A Transfer Learning Framework for Transcriptomics Reveals Systemic Features of Rare Disease," *Cell Syst.*, vol. 8, no. 5, p. 380–394.e4, 2019.
12. C. Zhu, C. Uwa, and W. Feng, "Informatics in Medicine Unlocked Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques," *Informatics Med. Unlocked*, no. January, p. 100179, 2019.
13. M. Chen, P. Zhou, D. Wu, L. Hu, M. Mehedi, and A. Alamri, "AI-Skin : Skin disease recognition based on self-learning and wide data collection through a closed-loop framework," vol. 54, no. May 2019, pp. 1–9, 2020.
14. O. Titus, P. Stephenson, M. Zahirul, and A. W. Taylor-robinson, "Long-term predictors of dengue outbreaks in Bangladesh : A data mining approach," *Infect. Dis. Model.*, vol. 3, pp. 322–330, 2018.
15. S. Wongkoon, M. Jaroensutasinee, and K. Jaroensutasinee, "Development of temporal modeling for prediction of dengue infection in Northeastern Thailand," *Asian Pac. J. Trop. Med.*, vol. 5, no. 3, pp. 249–252, 2012.
16. H. M. Aburas, B. G. Cetiner, and M. Sari, "Expert Systems with Applications Dengue confirmed-cases prediction : A neural network model," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4256–4260, 2010.
17. R. A. Vinarti and L. M. Hederman, "A personalized infectious disease risk prediction system," *Expert Syst. Appl.*, vol. 131, pp. 266–274, 2019.
18. D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction : A review," *Egypt. Informatics J.*, vol. 19, no. 3, pp. 179–189, 2018.
19. S. Khan, R. Ullah, S. Shahzad, N. Anbreen, M. Bilal, and A. Khan, "Analysis of Tuberculosis Disease through Raman Spectroscopy and Machine Learning," *Photodiagnosis Photodyn. Ther.*, 2018.
20. P. Mohapatra, S. Chakravarty, and P. K. Dash, "An improved cuckoo search based extreme learning machine for medical data classi fi cation," *Swarm Evol. Comput.*, pp. 1–25, 2015.
21. V. K. Shrivastava, N. D. Londhe, R. S. Sonawane, and J. S. Suri, "Biomedical Signal Processing and Control A novel approach to multiclass psoriasis disease risk stratification : Machine learning paradigm," *Biomed. Signal Process. Control*,

vol. 28, pp. 27–40, 2016.
22. S. U. P, H. Hebbar, and A. Govindakarnavar, "Informatics in Medicine Unlocked A technology framework for remote patient care in dermatology for early diagnosis," vol. 15, no. March, 2019.
23. T. Bikku, S. Rao, and A. Rao, "A contemporary feature selection and classification framework for imbalanced biomedical datasets," *Egypt. Informatics J.*, vol. 19, no. 3, pp. 191–198, 2018.
24. P. Guo *et al.*, "Science of the Total Environment An ensemble forecast model of dengue in Guangzhou , China using climate and social media surveillance data," *Sci. Total Environ.*, vol. 647, pp. 752–762, 2019.
25. S. Rao, R. Mopuri, and S. Naish, "Spatial distribution and cluster analysis of dengue using self organizing maps in Andhra Pradesh , India , 2011 – 2013," *Parasite Epidemiol. Control*, vol. 3, no. 1, pp. 52–61, 2018.
26. S. Atique, S. S. Abdul, C. Hsu, and T. Chuang, "Asian Paci fi c Journal of Tropical Medicine Meteorological influences on dengue transmission in Pakistan," *Asian Pac. J. Trop. Med.*, vol. 9, no. 10, pp. 954–961, 2016.
27. A. Awad, M. Bader-el-den, J. Mcnicholas, and J. Briggs, "International Journal of Medical Informatics Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach," *Int. J. Med. Inform.*, vol. 108, no. September, pp. 185–195, 2017.
28. J. De Torre, A. Valls, and D. Puig, "A deep learning interpretable classifier for diabetic retinopathy disease grading," no. xxxx, 2019.
29. [31] M. Shafenoor, Y. Kia, and K. Dewi, "Telematics and Informatics Identification of significant features and data mining techniques in predicting heart disease," *Telemat. Informatics*, vol. 36, no. August 2018, pp. 82–93, 2019.
30. [32] H. Jin and H. Young, "Learning representations for the early detection of sepsis with deep neural networks," *Comput. Biol. Med.*, vol. 89, no. August, pp. 248–255, 2017.
31. S. Chatterjee, D. Dey, S. Munshi, and S. Gorai, "Biomedical Signal Processing and Control Extraction of features from cross correlation in space and frequency domains for classification of skin lesions," *Biomed. Signal Process. Control*, vol. 53, p. 101581, 2019.
32. G. Jácome, P. Vilela, and C. Yoo, "Ecological Informatics Social-ecological modelling of the spatial distribution of dengue fever and its temporal dynamics in Guayaquil , Ecuador for climate change adaption," *Ecol. Inform.*, vol. 49, no. July 2018, pp. 1–12, 2019.
33. Z. Xu *et al.*, "Using dengue epidemics and local weather in Bali , Indonesia to predict imported dengue in Australia," *Environ. Res.*, vol. 175, no. May, pp. 213–220, 2019.
34. G. Jácome, P. Vilela, and C. Yoo, "Present and future incidence of dengue fever in Ecuador nationwide and coast region scale using species distribution

modeling for climate variability ' s e ff ect," *Ecol. Modell.*, vol. 400, no. November 2017, pp. 60–72, 2019.

35. M. Fakhruddin, P. Setia, and K. Putra, "Assessing the interplay between dengue incidence and weather in Jakarta via a clustering integrated multiple regression model," *Ecol. Complex.*, vol. 39, no. May, p. 100768, 2019.
36. D. Petrova, R. Lowe, A. Stewart-ibarra, J. Ballester, S. Jan, and X. Rodó, "Sensitivity of large dengue epidemics in Ecuador to long-lead predictions of El Niño," *Clim. Serv.*, no. xxxx, p. 100096, 2019.