Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 3, https://carcinogenesis.com/

Assessing the Impact of Climate-Change & Developing A Predictive Model for Heatstroke Incidents in Pune, India

Harsha Bhute[1], Avinash Bhute[2], Bharati P. Vasgi[3], Chetan Chauhan[4], Manohar K. Kodmelwar[5], Seema Vanjire[6]

[1]Department of Information Technology, Pimpri Chinchwad College of Engineering, Nigdi, Pune, Maharashtra, India, harsha.bhute@pccoepune.org, ORCID:0000-0002-1522-4053

[2]Department of Computer Engineering, Shah & Anchor Kutcchi Engineering College, Chembur, Mumbai, Maharashtra, India, avinash.bhute@pccoepune.org, ORCID: 0000-0003-3236-686X

[3]Department of Information Technology, Marathwada Mitra Mandal's College of Engineering, Pune, Maharashtra, India, bharativasgi@gmail.com, ORCID: 0000-0003-2722-5480

[4]Department of Computer Engineering, Vishwakarma University, Pune, Maharashtra, India, er.chouhan@gmail.com, ORCID: 0000-0001-7973-0644

[5]Department of Information Technology, Vishwakarma Institute of Technology, Pune, Maharashtra, India. manohar.kodmelwar@vit.edu, ORCID: 0000-0001-5248-528X

[6]Department of Computer Engineering, Vishwakarma University, Pune, Maharashtra, India, seema.vanjire@vupune.ac.in, ORCID: 0000-0001-6463-1539

ABSTRACT

The increase in occurrences of heatwaves in Pune, India, creates the urgency of developing predictive models to safeguard public health. This research endeavours to construct a robust predictive model to estimate the impact of heatwaves on Pune's population, leveraging machine learning algorithms and historical weather data spanning five years (2016-2020). By analysing meteorological attributes like temperature, humidity, dew point, and wind speed, we aim to discern patterns indicative of heatwave occurrences, enabling proactive mitigation strategies. Through an exhaustive literature review, we explored the potential of machine learning in forecasting heat-related illnesses and identified suitable algorithms for our predictive model. Random Forest Regression, CatBoost Regressor, and XGBoost Algorithm emerged as promising candidates due to their effectiveness in similar contexts. The dataset used comprises comprehensive weather attributes, supplemented by synthetic data representing heat stroke cases under Indian weather conditions. Our study's findings hold promise for policymakers and healthcare authorities, offering actionable insights to mitigate the adverse effects of heatwaves on public health in Pune and beyond.

Keywords: *Predictive Analytics, Heat Stroke, Climate Change, Heat wave, Machine Learning, Random Forest Regression, CatBoost Regressor, XGBoost Algorithm, Weather Data, Temperature, Humidity*.

How to Cite: Harsha Bhute, Avinash Bhute, Bharati P. Vasgi, Chetan Chauhan, Manohar K. Kodmelwar, Seema Vanjire, ., (2025) Assessing the Impact of Climate-Change & Developing A Predictive Model for Heatstroke Incidents in Pune, India, *Journal of Carcinogenesis, Vol.24, No.3*, 688-697.

1. INTRODUCTION

Climate change is a global challenge impacting billions of lives. Pune has seen an increase in heatwave cases in recent years impacting the health of citizens[1].As these heatwaves are increasing the risk of heat related illnesses and deaths are also Increasing. Understanding the heatwave impacts developing mitigation strategies is crucial for protecting the lives of the citizens. There is a gap in accurately predicting and reducing the impact of heatwaves on individuals. Current methods fail to do this as the population is increasing and the risk of heatwave is also increasing. The main goal of this research paper is to create a predictive model for estimating the number of people affected by heatwaves in Pune. Analyze past data on temperature, humidity, wind speed to identify patterns and trends in heat wave impacts. Temperature, Dew point, humidity and wind speed. These features were selected for predicting the heatstroke cases in Pune. The historic data (daily) for weather was taken from Global Historical Weather and Climate Data Website for five years (2016-2020). After doing literature survey we found out that random forest regressor , XGBoost , CatBoost these algorithms will be suitable for our problem statement. During our research we implemented all of these algorithms and compared their accuracies. The research goal was to make a predictive model that can help in making better decisions.

In the face of escalating heatwaves exacerbated by climate change, Pune, India, grapples with the imperative to safeguard public health. This research aims to develop a predictive model to estimate the impact of heatwaves on Pune's population, leveraging machine learning and historical weather data. By analysing meteorological attributes, including temperature, humidity, dew point, and wind speed, the study seeks to discern patterns indicative of heatwave occurrences, facilitating proactive mitigation strategies. Aligned with Sustainable Development Goals (SDGs) 3, 11, 13, and 15, this research contributes to the agenda of promoting good health, sustainable cities, climate action, and land resilience, offering actionable insights to policymakers and healthcare authorities to mitigate heat wave-related health risks effectively.

2. LITERATURE SURVEY

A study in Korea constructed a Random Forest (RF) model to forecast weekly heat-related damages by analysing four years of statistical, meteorological, and floating population data.Key variables considered in the model include temperature, humidity, wind speed, vulnerable occupational groups, insurance premiums, personal income, floating population, and registered population[1]. A model employing the random forest technique to forecast heat stroke occurrence for heatwaves in China-The study aimed to predict heat stroke occurrences during heatwaves in China using a random forest model. Three years of data from typical hot cities in China were utilised, incorporating meteorological and socioeconomic status (SES) factors [2]. Machine learning-based mortality prediction model for heat-related illness-The study focused on establishing and validating a machine learning-driven mortality prediction model for hospitalised patients with heat-related illness. Extracting data from 2393 hospitalised patients. Twenty-four predictor variables, encompassing patient characteristics, vital signs, and laboratory test results upon hospital admission, were employed for machine learning analysis[3].

Heat Wave Prediction Using Recurrent Neural Networks Based on Deep Learning-The study employs a deep learning approach using recurrent neural networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, to predict heat waves invulnerable regions. Historical meteorological data from 2016 to 2020 for five cities in Telangana, India[4].Monthly heatwave prediction in Sweden based on Machine Learning techniques with remote sensing data-This study focuses on predicting heatwave occurrences in Sweden using machine learning (ML) techniques and remote sensing data. The research extracts 21 features from remote sensing data and employs feature selection techniques to determine the most relevant variables. Five ML classifiers - Logistic Regression, Gaussian Naive Bayes, K-Nearest Neighbors, Random Forest, and XGBoost - are utilized, with hyperparameter tuning[5].Regional Heat Wave Prediction Using Graph Neural Network and Weather Station Data-This study proposes a novel approach for predicting regional heat waves using a Graph Neural Network (GNN) model trained on weather station data. The research leverages historical weather records from the Global Surface Summary of the Day (GSOD) dataset, spanning 15 years from 2006 to 2020, collected from 91 weather stations across the CONUS region[6].

Utilizing machine learning approaches for forecasting sea surface temperature and marine heatwave occurrence: a case study of the Mediterranean Sea-This study focuses on predicting sea surface temperature (SST) and marine heatwaves (MHWs) in the Mediterranean Sea using ML methodologies, including random forest (RForest), long short-term memory (LSTM), and convolutional neural network (CNN), are deployed to predict both sea surface temperature (SST) trends and occurrences of marine heatwaves (MHWs) up to a week in advance[7]. Deep Learning-Based Extreme Heatwave Forecast - This study uses PlaSim climate model data to predict heat waves in the Northern Hemisphere mid-latitudes above 30°N. It focuses on surface temperature (Ts) and geopotential height at 500 hPa (Zg), employing a CNN architecture for high dimensional data[8]. Climate Change Forecasting using Machine Learning Algorithms - This study uses Kaggle and Earth System Research Laboratory datasets to analyse temperature trends and CO2 concentrations since 1750 and 1958. Employing machine learning techniques like Random Forest and XGBoost Regressors, it aims to predict temperature changes based on CO2 concentrations[9]. Heatstroke predictions using machine learning algorithms, weather information - This paper develops prediction models for heat strokes using various methods, including generalized linear model , generalized additive model , random forest, and extreme gradient boosting decision tree . GAM emerges as the best model with the least root-mean squared error[10]. Artificial intelligence and machine learning to do climate change research and preparedness - The abstract mentions the potential of machine learning (ML) algorithms in climate analysis, citing previous studies, and proposes their use to analyze complex climate interactions, improve weather forecasting, and address challenges in Earth System Models (ESMs)[11].

Medium-Term Prediction for Ambulances Desire of Heat Stroke - This study employs Partial Least Squares Regression (PLS) to forecast the number of heat stroke cases in Kobe City. Weekly weather forecast data and heuristic adjustments for rainy days are used. The model considers regional characteristics and threshold temperatures, utilising data from the

Japan Meteorological Agency (JMA) to aid emergency service planning[12]. Impact of Climate Change and Air Pollution Forecasting Using Machine Learning Techniques in Bishkek - This study utilised air quality and meteorological data from Kyrgyzhydromet (2016-2018) to develop a forecasting model for Bishkek, Kyrgyz Republic. Various machine learning algorithms including ANN, HPT, XgbR, RFR, KNN, DTR, LaR, and LR were employed to predict air quality and assess climate change impact on pollution[13]. An Analysis of Climate Change Based on ML and an Endoreversible Model - This research introduces a Finite-Time Thermodynamics (FTT) approach to predict surface temperature changes, considering factors like albedo, greenhouse gases, luminosity changes, and dissipation. Machine learning models are employed to evaluate FTT's effectiveness and compare predictive accuracy with experimental data[14]. AI-enabled strategies for climate change adaptation: protecting communities, infrastructure, and businesses from the impacts of climate change - The abstract highlights climate change as a global challenge and presents artificial intelligence (AI) as a promising tool for adaptation. It discusses AI's role in identifying vulnerable areas, simulating climate scenarios, and assessing risks for businesses and infrastructure. Ethical considerations, transparency, fairness, and equity in AI applications are emphasised[15].

Impact Analysis of Climate Change on Floods in an Indian Region Using Machine Learning - The study utilised historical weather data from the India Meteorological Department (IMD) to define 20 parameters and forecast future flood occurrences in Maharashtra state. Machine learning models like ANN, LightGBM, and LSSVM were employed. Simulation experiments predicted flood events under different socioeconomic pathways (SSPs) scenarios[16]. The research paper examines the relationship between annual crop production and climate variables using regression models, with the log-polynomial regression model identified as the best fit. While traditional methods are emphasised, integrating machine learning algorithms like decision trees or random forests could enhance predictive accuracy and provide deeper insights into complex data relationships, potentially improving crop production projections and understanding the impact of climate change on agriculture[17]. The research employs the Distribution-based Scaling (DBS) Method to bias-correct General Circulation Model (GCM) data, involving adjustments to wet fraction and transforming rainfall distributions. Evaluation compares climate statistics among observed, raw GCM, and bias-corrected GCM data, including rainfall characteristics. Climate change signals are analysed across GCM projections using trend analysis methods like Mann-Kendall and Student's t tests[18]. For modelling climate change impacts, techniques like dynamical downscaling with regional climate models (RCMs) or statistical downscaling are used to generate detailed local forecasts from global models. Algorithms such as decision trees, random forests, support vector machines (SVM), and neural networks are commonly applied to analyse climate data and forecast changes in temperature, precipitation, and extreme weather events. Selection of the algorithm depends on factors such as desired accuracy, data quality, and computational resources available[19].

The study outlines an evaluation of machine learning algorithms—Linear Regression, Support Vector Regression (SVR), Lasso, and ElasticNet—in predicting annual global warming trends. It aims to establish a robust statistical model capturing the relationship between average annual temperature and greenhouse gas concentrations like carbon dioxide, methane, nitrous oxide, and sulphur hexafluoride. Linear Regression is identified as the most accurate method for temperature prediction based on extensive analysis[20]. Machine learning aids climate monitoring via remote sensing for tasks like deforestation detection and disaster damage assessment. It accelerates scientific discoveries by optimising systems and expediting computations in simulations. ML addresses data challenges, utilising techniques like transfer learning, and its roadmap involves learning, collaborating, engaging stakeholders, and deploying impactful solutions to combat climate change[21]. The study investigates the impact of climate change on forests in India, highlighting the necessity for implementing strategies aimed at mitigating vulnerability and promoting sustainable management practices. It suggests using global models for temperature projections but notes uncertainties in rainfall predictions. The text calls for dynamic vegetation modelling, adaptation strategy development, and exploring links between mitigation and adaptation. It also emphasises estimating market mitigation potential and assessing climate change impacts on carbon sinks[22].

Gather diverse datasets encompassing meteorological, demographic, land use, and health data. Preprocess the data by addressing missing values, outliers, and normalisation. Extract features relevant to urban climate changes and health outcomes, then select appropriate machine learning models, splitting the data for training, validation, and testing. Train and optimise models, evaluate performance using metrics like MAE or AUC-ROC, interpret results, and validate on unseen data, iterating for refinement as necessary [23].

Machine Learning is a powerful tool which can be used to make predictions, analyse data in various fields.ML algorithms process large amounts of data and find hidden patterns. In climate change study ML allows to create predictive models that consider various factors to predict the impact. We studied various research papers to understand the potential risks caused by climate change, especially those associated with heat waves.Our aim was to gain insights from existing research on climate change impacts and how machine learning approaches can help in risk assessment. We studied twenty one research papers, each providing valuable insights on how to adapt and mitigate the impacts of climate change. Studying these papers gave insights on the ability of machine learning to predict Heat Stroke cases. One insight is the Random The forest algorithm, employed for precise forecasting of heatwave risk in Korea.In this study they developed a random forest (RF) model for weekly prediction of heat-related damages in South Korea using four years of statistical, meteorological, and floating population data.The model underwent training and assessment utilising diverse metrics, including mean absolute error, root mean squared error, and root mean squared logarithmic error.Key variables considered in the model include temperature, humidity, wind speed, vulnerable occupational groups, insurance premiums, personal income, floating population, and registered population. We studied the importance of parameters such as temperature, humidity and wind speed when predicting heat wave impacts. Another study was done in China for predicting heat stroke cases where three years of data from typical hot cities in China were utilised. By studying we decided to select features such as Temperature , Dew point , humidity and wind speed for predicting the heatstroke cases.The algorithms which we found common in these research papers are Random forest regressor , XGBoost and CatBoost.In our study we used these three algorithms for predicting heat stroke cases.

3. METHODOLOGY

This study adopts a data-driven approach to analyze historical weather patterns, assess the impact of climate change, and develop a predictive model for estimating heatstroke incidents in Pune, India. The methodology comprises five key stages: data collection, data preprocessing, feature selection, model development, and model evaluation.

The architecture of the proposed predictive framework is designed to collect, preprocess, analyze, and model meteorological and health-related data to estimate heatstroke incidents in Pune, India. The complete workflow is illustrated in Figure 1.

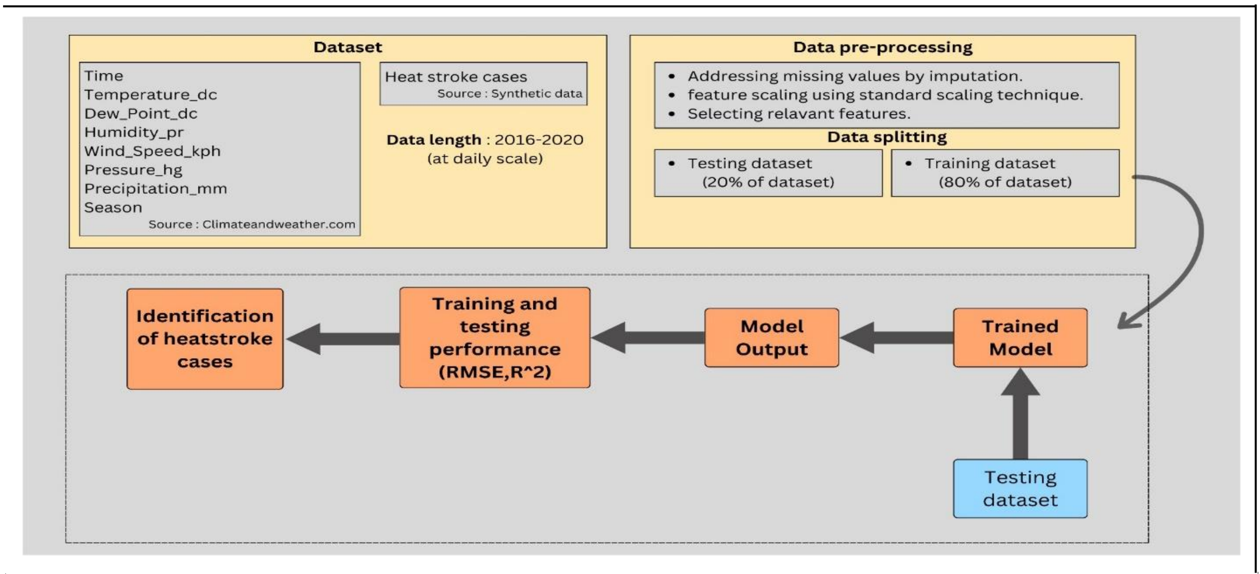

Figure 1: Proposed Model Workflow and Architecture

3.1 Data Collection

We collected the dataset from https://weatherandclimate.com/india, covering the period from 2016 to 2020. The dataset includes attributes such as Time, Temperature (°C), Dew Point (°C), Humidity (%), Wind Speed (kph), Pressure (hg), Precipitation (mm), and Season. For heat stroke cases, synthetic data generated to match Indian weather conditions was used, aiming to ensure accuracy. The finalised dataset comprises average temperature, wind speed, and humidity, with heat stroke cases being the target variable. For this model we used synthetic data for heatstroke cases. This data was generated according to the Indian weather conditions by considering summer months and October heatwaves.

Dataset Used Time:
Used to allow the model to capture patterns and trends over time. It helps the model understand how weather conditions change throughout the day, week, month, or year.
Temperature:
The temperature measured in °C , it is modified and reduced as per need. In the raw data, the temperature of each day . Those data are converted into average temperature for better and ease usage.
Humidity:
The humidity is measured in %, It is standardised according to use. In the raw dataset 85 % of data is given ,Remaining data is collected from the other sources.
Dew Point:
Dew point represents the amount of moisture in the air. Including dew point as an attribute helps the model account for

humidity levels and understand the likelihood of precipitation or fog formation,
Wind Speed:
The wind speed is measured in kph.Wind speed affects how hot or cold we feel, moves clouds around, and influences the development of storms. Including wind speed in weather predictions helps capture its impact on temperature, cloud movement, and storm patterns.
Pressure:
The Pressure is measured in hg ,. By including pressure in weather predictions, the model can understand how air pressure changes affect the weather.
Precipitation:
The Precipitation is measured in mm.Precipitation data tells us how much moisture is coming down from the sky, like rain, snow, or hail. By using this information in weather predictions, the model can forecast when and how much it's going to rain, snow, or hail, helping people prepare for the weather ahead.
Season:
The season attribute sorts information based on the time of year, such as winter, spring, summer, or fall. This helps the model recognize how weather shifts with each season, like getting colder in winter and warmer in summer, so it can make better predictions about things like temperature and rainfall.
heatstroke_cases:
The heatstroke_cases data is synthetic data which was generated according to Indian weather conditions.

3.2 Data Preprocessing

The research methodology began with a comprehensive data exploration phase aimed at understanding the structure and characteristics of the dataset. Initial scrutiny involved examining the layout and content of the dataset, particularly focusing on the first few rows to grasp its organisation. Furthermore, an assessment for missing values was conducted to ensure data completeness. Missing data was addressed through imputation techniques to maintain the integrity of the dataset. Subsequently, data preparation for modelling commenced by segregating features and target variables. Key features such as temperature, wind speed, dew point, pressure, and humidity were identified as inputs, while the 'heatstroke_cases' variable was designated as the target for prediction. To gain insights and visualise data distributions, histograms and correlation metrics were employed. Additionally, standard scaling techniques were applied to ensure uniformity in feature scales. As part of feature selection procedures, features such as precipitation and season were omitted from the dataset, deemed irrelevant for predicting heat stroke cases in Pune. This thorough preparation laid the groundwork for robust modelling and analysis.

3.3 Machine Learning Algorithms

In our study about heat stroke cases, we used different algorithms to understand the weather and predict how likely it is for people to get heatstroke. We used Random Forest Regression , CatBoost Regressor and XGBoost Algorithms. All the algorithms were implemented to find the best one to predict the heatstroke cases.

Random Forest Regression

Random Forest is like a team of decision trees working together to make predictions. Each decision tree considers different aspects of the data and votes on the outcome. It's great for studying heat waves because it's good at handling lots of information without getting confused. It's also smart enough to figure out which weather factors are most important for predicting heat stroke cases. In our model, we trained the Random Forest using weather data like humidity, temperature, and wind speed. This helped us understand which weather conditions are most likely to lead to heatstroke, which is really important for planning how to keep people safe during heat waves.

Step 1) Dataset is represented as follows

$$D = \{(X_1, y_1), (X_2, y_2) \dots (X_n, y_n)\}$$

Where:
$X_i = [X_{i1}, X_{i2}, \dots, X_{in}]$→ Feature vector for the ith day (temperature, humidity, etc.)
y_i → Heatstroke cases on the ith day
n → Total number of observations
m → Number of meteorological features
Step 2) Bootstrap Sampling
Random Forest builds B independent regression trees using bootstrap sampling. For each tree Tb where b=1,2,…,B, a random sample Db is drawn from D with replacement.

$$D_b \approx Bootstrap(D)$$

Step 3 – Decision Tree Prediction
Each decision tree Tb predicts an output $\widehat{y_b}$. For given input X
For a single regression tree, the prediction at a leaf node is the mean of target values in that region

$$\hat{y}_b(X) = \frac{1}{N_b}\sum_{i=R_b} y_i$$

Where:
Rb → Region (leaf) of the tree where X falls
Nb → Number of training samples in region Rb

Step 4 – Final Random Forest Prediction
The final prediction of the Random Forest is the average of all tree predictions

$$\hat{y}(X) = \frac{1}{B}\sum_{b=1}^{B} \left(\hat{y}_b(X)\right)$$

Where:
$\hat{y}(X)$→ Final predicted heatstroke cases
B → Total number of trees in the Random Forest
$\left(\hat{y}_b(X)\right)$ → Prediction from the bth tree

Step 5 – Feature Importance (Optional for Your Paper)
Feature importance in Random Forest is computed using the Mean Decrease in Impurity (MDI)

$$FI(f) = \frac{1}{B}\sum_{b=1}^{B} \sum_{t\in T_b} \Delta I_t(f)$$

Where:
$\boldsymbol{FI(f)}$→ Importance of feature f
$\Delta\boldsymbol{I_t(f)}$→ Reduction in impurity (variance) at node t caused by feature f
B → Total number of trees

CatBoost Regressor
CatBoost Regressor is like a super-smart algorithm that's really good at dealing with different kinds of information, even if it's not all numbers. It's like having a genius friend who can quickly figure out the patterns in data and make predictions. In our research, we used CatBoost to study heat waves because it's great at understanding how different weather factors, like humidity and temperature, are connected to heat stroke cases. It can also handle missing information without any problems, which is handy when dealing with real-world data. With CatBoost, we were able to predict heat stroke cases more accurately, helping us plan better ways to protect people during heat waves.

$$\hat{y}(X) = F_T(X) = F_0 + \beta\sum_{t=1}^{T} f_t(X)$$

Where:
$F_T(X)$→ Final predicted heatstroke cases
F_0 → Initial model prediction (mean of yi)
$f_t(X)$→ Contribution of the tth tree
T → Total number of boosting iterations

XGBoost Algorithm
XGBoost Algorithm is a powerful tool for analysing data and making predictions. It's like having a supercomputer that's really good at understanding complex relationships between different factors. In our study of heatwaves, we used XGBoost to analyse weather data and predict how likely it is for people to get heatstroke. XGBoost is smart enough to handle missing information in the data and figure out which weather factors are most important for predicting heat stroke cases. By using XGBoost, we were able to make more accurate predictions, which is really important for planning ahead and keeping people safe during heatwaves.

4. EVALUATION METRICS

After model training, we evaluated the performance of our predictions using two key metrics: Root Mean Squared Error (RMSE) and R-squared (R2) score. These metrics provide insights into the accuracy and explanatory power of our models. Lower RMSE values indicate better predictive accuracy, while higher R2 scores signify better model fit to the data.R2 score assesses how much of the variance in the target variable is explained by the model.

Lower RMSE and higher R2 indicate better model performance. Scatter plots were generated to compare actual vs predicted values. Additionally, a time-series plot was generated to show trends in actual vs predicted data.These evaluation metrics and visualisations offered valuable insights into the effectiveness of the predictive models and their ability to capture the underlying patterns in the data.

5. RESULTS AND DISCUSSION

In this study, we evaluated the performance of three popular machine learning algorithms, namely Random Forest Regression, CatBoost Regressor, and XGBoost Algorithm, on a given dataset. The evaluation metrics used to assess the models' performance include Root Mean Squared Error (RMSE), R-squared (R^2) score, and accuracy.

The results indicate that Random Forest Regression achieved a remarkable RMSE of 0.95 and an R^2 score of 0.8259, translating to an accuracy of 82.60%. Following closely, the CatBoost Regressor exhibited competitive performance with an RMSE of and an impressive R^2 score of 0.8373, resulting in an accuracy of 83.73%. Conversely, the XGBoost Algorithm demonstrated the highest R^2 score among the models, reaching 0.99, albeit with a slightly higher RMSE of 0.8087, corresponding to an accuracy of 80.88%.

Overall, our findings suggest that all three algorithms perform admirably in predicting the target variable.The selection of an algorithm may vary depending on the specific requirements of the task, necessitating consideration of the most suitable option for optimal performance. Random Forest Regression and CatBoost Regressor showing superior performance in terms of accuracy and predictive power.

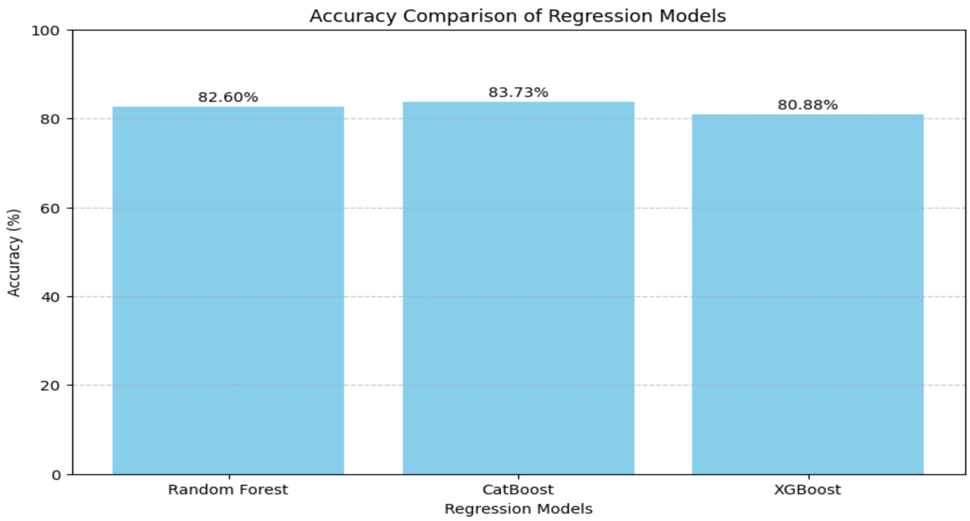

Figure 2: Comparison of Accuracy

Table 1: Machine Learning algorithms Evaluation Results

ML Algorithms	Root Mean Squared Error(RMSE)	R - squared (R^2 Score)	Accuracy
Random Forest Regression	0.95	0.8259	82.60%
CatBoost Regressor	0.91	0.8373	83.73%
XGBoost Algorithm	0.99	0.8087	80.88%

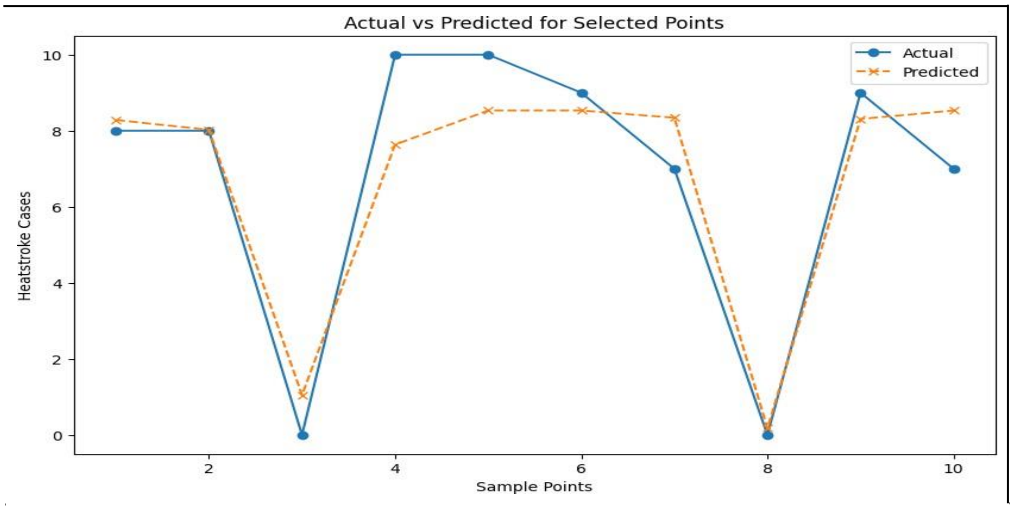

Figure 3 : Actual vs Predicted-RF CatBoost Regressor

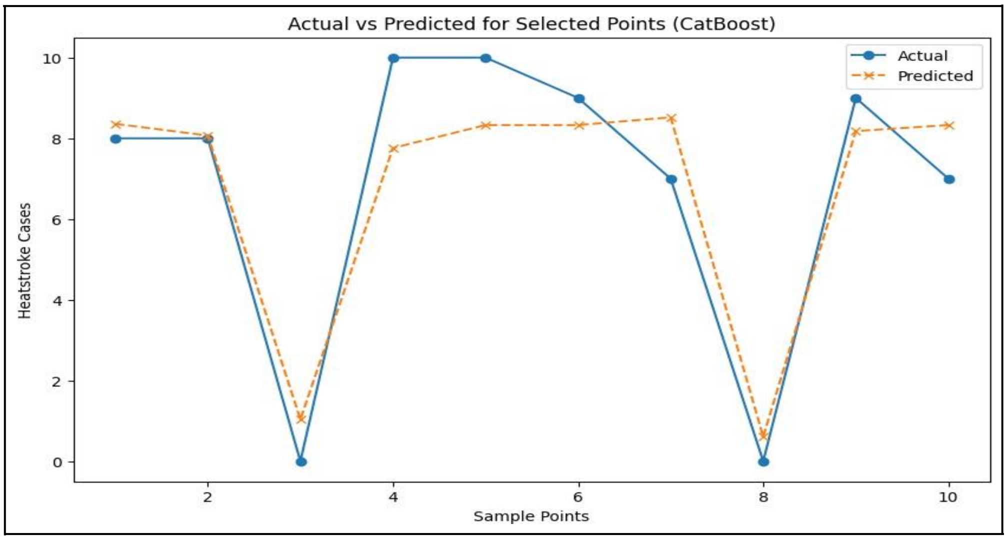

Figure 4 : Actual vs Predicted- for selected points (CatBoost Regressor)

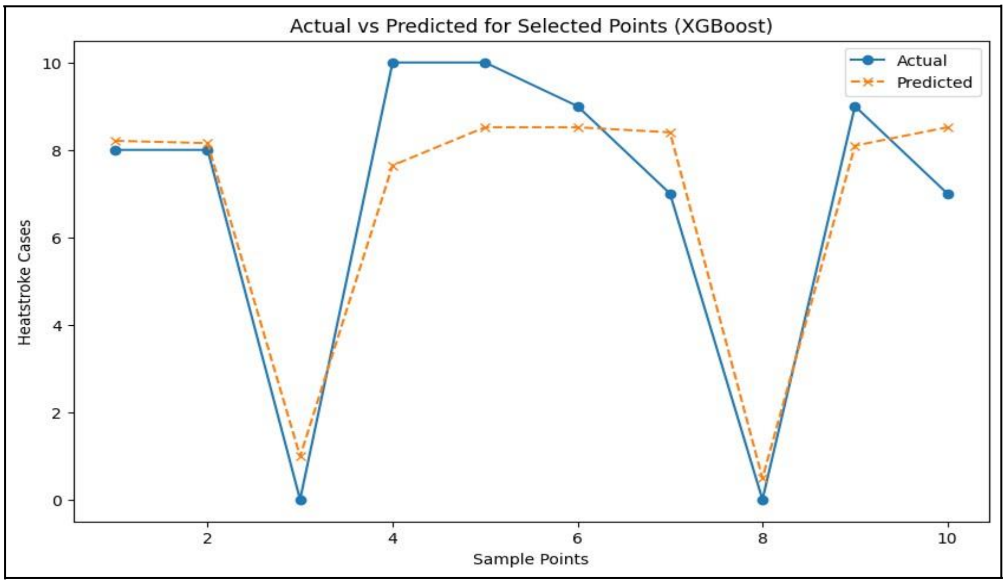

Figure 5: Actual vs Predicted-XGBoost

6. CONCLUSION

Random Forest Regression, CatBoost Regressor,XGBoost algorithms were used in study of approaches to predict heat stroke cases in Pune. The objective was to develop a predictive model that could effectively learn from historical data of heatstroke occurrences. The result of the study demonstrated how successfully these algorithms were used to make predictive models. The Random Forest Regression and CatBoost Regressor perform with accuracy of 82.60% and 83.73% respectively. Whereas the XGBoost showed 80.88% accuracy. These results demonstrate that these algorithms can be used to capture the relationship between weather data and heat stroke cases. This study concentrated on how machine learning algorithms may clarify prediction processes and improve precision of weather forecasts.

REFERENCES

[1] https://trends.google.com/trends/explore?cat=45&date=all&geo=IN-MH&q=He atstroke&hl=en-GB
[2] Park, Minsoo, Daekyo Jung, Seungsoo Lee, and Seunghee Park. "Heatwave damage prediction using a random forest model in Korea." Applied sciences 10, no. 22 (2020): 8237.
[3] Wang, Yanwen, Qingqing Song, Yanjun Du, Jiaonan Wang, Jinhui Zhou, Zonghao Du, and Tiantian Li. "A random forest model to predict heatstroke occurrence for heatwave in China." Science of the total environment 650 (2019): 3048-3053.
[4] Kim, Yesuel, and Youngchul Kim. "Explainable heat-related mortality with random forest and SHapley Additiv exPlanations (SHAP) models." Sustainable Cities and Society 79 (2022): 103677.
[5] Byagar, Sarita, Araddhana Arvind Deshmukh, Kirti Wanjale, Vinod S. Wadne, Nidhi Ranjan, and Rupali Gangarde. "Heat Wave Prediction Using Recurrent Neural Networks Based on Deep Learning." International Journal of Intelligent Systems and Applications in Engineering 12, no. 1s (2024): 612-619.
[6] Li, Zhuoran. "Monthly heatwave prediction in Sweden based on Machine Learning techniques with remote sensing data." (2023).
[7] Li, Peiyuan, Yin Yu, Daning Huang, Zhi-Hua Wang, and Ashish Sharma. "Regional heatwave prediction using graph neural network and weather station data." Geophysical Research Letters 50, no. 7 (2023): e2023GL103405.
[8] Bonino, Giulia, Giuliano Galimberti, Simona Masina, Ronan McAdam, and Emanuela Clementi. "Machine learning methods to predict Sea Surface Temperature and Marine Heatwave occurrence: a case study of the Mediterranean Sea." EGUsphere 2023 (2023): 1-22.
[9] Jacques-Dumas, Valérian, Francesco Ragone, Pierre Borgnat, Patrice Abry, and Freddy Bouchet. "Deep learning-based extreme heatwave forecast." Frontiers in Climate4 (2022).
[10] Rajest, S. Suman, Bhopendra Singh, Ahmed J Obaid, R. Regin, and Karthikeyan Chinnusamy, eds. "Recent developments in machine and human intelligence." (2023).
[11] Ogata, Soshiro, Misa Takegami, Taira Ozaki, Takahiro Nakashima, Daisuke Onozuka, Shunsuke Murata, Yuriko Nakaoku et al. "Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts." Nature communications 12, no. 1 (2021): 4575.
[12] Huntingford, Chris, Elizabeth S. Jeffers, Michael B. Bonsall, Hannah M. Christensen, Thomas Lees, and Hui Yang. "Machine learning and artificial intelligence to aid climate change research and preparedness." Environmental Research Letters 14, no. 12 (2019): 124007.
[13] Nakai, Tetsuya, Sachio Saiki, and Masahide Nakamura. "Medium-Term Prediction for Ambulance Demand of Heat Stroke using Weekly Weather Forecast." In 2021 8th International Conference on Internet of Things: Systems, Management and Security (IOTSMS), pp. 1-8. IEEE, 2021.
[14] Isaev, Erkin, Boobek Ajikeev, Urmatbek Shamyrkanov, Kenjebek-uulu Kalnur, Karimov Maisalbek, and Roy C. Sidle."Impact of climate change and air pollution forecasting using machine learning techniques in Bishkek." Aerosol and Air Quality Research 22, no. 3 (2022): 210336.
[15] Vázquez-Ramírez, Sebastián, Miguel Torres-Ruiz, Rolando Quintero, Kwok Tai Chui, and Carlos Guzmán Sánchez-Mejorada. "An Analysis of Climate Change Based on Machine
[16] Learning and an Endoreversible Model." Mathematics 11, no. 14 (2023): 3060.
[17] Jain, Harshita, Renu Dhupper, Anamika Shrivastava, Deepak Kumar, and Maya Kumari. "AI-enabled strategies for climate change adaptation: protecting communities, infrastructure, and businesses from the impacts of climate change." Computational Urban Science 3, no. 1 (2023): 25.
[18] Vage, Sarthak, Tanu Gupta, and Sudip Roy. "Impact Analysis of Climate Change on Floods in an Indian Region Using Machine Learning." In International Conference on Artificial Neural Networks, pp. 370-383. Cham: Springer Nature Switzerland, 2023.
[19] Kelkar, Saurabh M., Ashwini Kulkarni, and K. Koteswara Rao. "Impact of climate variability and change on crop production in Maharashtra, India." Current Science 118, no. 8 (2020): 1235-1245.India"
[20] Rana, Arun, Kean Foster, Thomas Bosshard, Jonas Olsson, and Lars Bengtsson. "Impact of climate change on rainfall over Mumbai using Distribution-based Scaling of Global Climate Model projections." Journal of Hydrology: Regional Studies 1 (2014): 107-128.

[21] Singh, Hukum, and Manoj Kumar. "Climate change and its impact on Indian Himalayan forests: current status and research needs." In Climate Change: Impacts, Responses and Sustainability in the Indian Himalaya, pp. 223-242. Cham: Springer International Publishing, 2022.
[22] Bisht, Akhilesh, and K. N. Meera. "Predicting the future tendencies in global warming using orthogonal projection of vectors." In AIP Conference Proceedings, vol. 2755, no. 1 AIP Publishing, 2023.
[23] Rolnick, David, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross et al. "Tackling climate change with machine learning." ACM Computing Surveys (CSUR) 55, no. 2 (2022): 1-96.
[24] Chaturvedi, Rajiv K., R. Tiwari, and N. H. Ravindranath. "Climate change and forests in India." International Forestry Review 10, no. 2 (2008): 256-268.
[25] Kim, Yesuel, and Youngchul Kim. "Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models." Sustainable Cities and Society 79 (2022): 103677.
[26] Choudhary, Manisha, Monika Lokhande, Rushikesh Borse, and Avinash Bhute. "A machine learning approach to aid paralysis patients using EMG signals." In Advanced Data Mining Tools and Methods for Social Computing, pp. 107-125. Academic Press, 2022.
[27] Tawade, S., Bhute, A., Pagar, R., Bhute, H., Deoghare, R. and Patankar, T., 2024, August. Trust-Based Routing Protocols for MANETs: Ensuring Security and Reliability. In 2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA) (pp. 1-7). IEEE.
[28] Bhute, Avinash, and Harsha Bhute. "A Comprehensive Review on Digital Innovations for Renewable Energy and Conservation." Digital Innovations for Renewable Energy and Conservation (2025): 1-20.
[29] Bhute, Avinash, Rakesh S. Patil, Shreyas Sanghai, Prajakta Yadav, and Sanketkumar Sathe. "Functional Electrical Stimulation for Hand Movement of Person." In Proceedings of International Conference on Machine Intelligence and Data Science Applications: MIDAS 2020, pp. 465-481. Singapore: Springer Singapore, 2021.
[30] Agrawal, Pratham, Anchalaa Jha, and Avinash Bhute. "Comprehensive Exploration of Deepfake Detection Using Deep Learning." In International Symposium on Intelligent Informatics, pp. 79-95. Singapore: Springer Nature Singapore, 2023.