Climate Forecasting Framework for Urban Sustainability and Longevity using Machine Learning Model XGBoost

Anup G. Dakre*,[1,2] and Chaya R. Jadhav[2]

[1] Marathwada Mitra Mandal's College of Engineering,affiliated to SPPU, Pune , India
[2] Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India
anupdakre@mmcoe.edu.in

Abstract. Climate is a chaotic one in nature and its variability incorporates rising temperatures, unpredictable rainfall, storms, and other weather parameters disrupting life, all living beings. Traditional meteorological forecasting methods often struggle to model complex and nonlinear climate patterns, highlighting the need for machine learning (ML) approaches to improve predictive accuracy.This study introduces an ML-based climate prediction framework that integrates XGBoost model. The methodology encompasses data pre-processing, feature engineering, hyper parameter tuning , bias correction, and performance evaluation using key evaluation metrics such as Mean Squared Error (MSE), while Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and finally the R^2 Score. A historical weather dataset is used which includes weather parameters so as to train,validate and test the models, with advanced time-series techniques incorporated to enhance forecast precision.Experimental results demonstrate that ML-based models significantly outperform conventional forecasting techniques. XGBoost exhibits superior accuracy in short-term climate predictions, performing more accurately with an R^2 score of 0.91.The incorporation of such a Machine Learning based climate prediction system into urban infrastructure helps to reduce health risks for the aging population. Ultimately, this research aligns with global sustainability goals which mainly target a real time, scalable framework for essential services regardless of unpredictable climate for the growing elderly population.

Keywords: Machine Learning (ML), Weather Forecasting, Time-Series Prediction, Performance Metrics, Urban Sustainability.

1 Introduction

Weather is a Chaotic in nature and its changing behaviour directly affects urban environments, majorly sustainability, smart city infrastructure, and public health which is the main concern. Due to rise in surface temperature of earth which causes extreme and unpredicted weather patterns which leads to hazardous climate conditions that directly or indirectly elderly populations and longevity. The elderly populations with age more than 60 years old are not only vulnerable to these extreme weather patterns

W. Srimaharaj (ed.), *Proceeding of the 1st International Conference on Lifespan Innovation (ICLI 2025)*,
Advances in Health Sciences Research 88,
https://doi.org/10.2991/978-94-6463-831-8_51

but it is crucial for urban planning and public health safety. The conventional weather forecasting methods such as statistical weather prediction methods and numerical weather prediction (NWP) models, mainly rely on linear regression algorithms, complex statistical and mathematical equations which have failed to handle complex weather behaviour. The limitations due to conventional weather forecasting methods led to increase the use of Machine Learning (ML) to enhance prediction accuracy for various weather events with the use of metadata and also helps to identify the non-linear behaviours of various environmental factors.

The advancement in the Machine Learning algorithms or models like XGBoost can process vast amounts of meteorological data, ease to process and validate temporal dependencies, and be able to adjust dynamic climate conditions. For urban sustainability frameworks, ML algorithms ensure to develop more effective risk mitigation strategies, it helps to optimize energy resource allocation and also protect vulnerable populations from adverse environmental impacts due to dynamic weather conditions. The findings of our research study will ensure a data driven approach to environmental sustainability and public health protection.

2 Literature Survey

Machine Learning is an emerging technology in the domain of weather forecasting which ensures the recognition of nonlinear time series data and multivariate observed patterns in atmospheric or environmental data and provides better accuracy for urban management. The Machine Learning Model XGBoost in our research provides better accuracy and adaptability for resilience planning also, ultimately helps for longevity.

S.F.Ismael et al.[1] suggested that machine learning algorithms can help to represent effects of Urban Heat Island (UHI), which describes impact on elderly population well-being but also consumption of energy in smart cities or in urban development. Zulfani Alfasanah et al.[2] demonstrates the charts which are used to monitor the air quality updates using the machine learning techniques. X. Ma et al.[3] used the XGBoost machine learning model to provide insights regarding thermal zone creation after finding surface temperature of Land and its relation with urban morphology. Isaev, E. et al.[4] shows how extreme temperatures affect the elderly populations. The study used machine learning algorithms used for early warning systems for extreme temperature and how it reduced the hospitalization of elderly people by the rate of 40%. Lee & Martinez et al.[5] proposed ML model which showed improved efficiency of HVAC systems, indirectly reducing the emergencies of respiratory diseases in elderly health care centers. Alex Torku et al.[6] studied the environmental stressors which proposed predictive management and risk mapping in urban areas for elderly people. Reddy, Naveed et al.[7] discussed urban sustainability and public health. Their study suggested Machine Learning algorithms used for energy used in diverse zones and the need of AI in public health decision making.

Lyu et al.[8] utilized dominant gradient ML algorithms for retirement facilities of elderly people. Agarwal et al.[9] used ML models for the prediction of weather parameters and its correlation with hospital data to improve the longevity. Meteorological studies have proven the relationships between weather parameters and health issues of elderly populations. Prolonged heat waves significantly increase the risk of cardiovascular complications like strokes etc. in elderly people, Zhang et al. [10]. Variations in atmospheric pressure can cause strokes or many times it affects the elderly populations due to strain on cardiovascular systems of elderly populations by Peralta et al. [11]. Deterioration in respiratory health due to distribution of pollutants and various types of allergies, main cause is wind speed, humidity and PM 2.5 concentrations in the surrounding atmosphere by Paul et al.[12]. After the comprehensive Literature review and the results, Machine Learning algorithms like XGBoost not only improves the forecast accuracy for different weather parameters but also it directly affects the longevity and urban environments. It offers faster training time and high accuracy with R^2 score of 0.91 which is the justified choice for urban health and longevity with scalable climate forecasting which can be used for fast, accurate and real time weather alerts to elderly population, ultimately minimizing health risks.Though the XGBoost has more than 90% accuracy, future research should be more scalable and interoperable in context of not only use of geographical real time data but also integrating infrastructural, and demographic data so as to ensure reliable solutions for urban sustainability and longevity.

3 Methodology

Data for 20 years (2002-2022) are used for Pune Station and it is pre-processed and feature engineering is applied for better accuracy. Dataset is collected from the IMD (Indian Meteorological Department) and Wyoming University. A weather monitoring system is deployed at a local station with a microcontroller based system with weather sensors. The NVIDIA GPU acceleration enabled the training time for the dataset was reduced to approximately 0.2 seconds.

A dataset of daily historical weather observations for various parameters such as temperature, humidity, wind speed and pressure (atmospheric) is used. Temperature measured in °C which is average surface temperature. Relative humidity is a measure of the amount of moisture in air, calculated in percentage (%). Daily average surface wind speed is measured km/hr with atmosphere pressure measured at sea level with hPa unit. The dataset used in our research study consists of 7095 daily records. The parameters used here were mainly due to them directly impacting human beings, health risks factors for aging populations.Also, thermal impact, increase and decrease pollution levels are some factors that are directly associated with health risks of aging populations and urban sustainability.

Weather data is a time-series data and random data. For quality of data as an input it is necessary data must be processed for the included missing values with the help of different interpolation techniques and forward fill techniques which are the core part of any Machine Learning algorithms. In our proposed model linear interpolation techniques are used. For accurate results it is essential that models should be able to learn seasonality and trends which are time bound in nature. For that indicators are used such as day of the year, week days and month to derive additional features. In our research, Lag features were constructed in such a manner for specific intervals for enhancing the temporal learning.

In our research XGBoost algorithm is used due to its efficient quality to deal with non-linearity of weather parameters like rain, and it is more efficient while comparing with conventional Machine Learning algorithms. The inbuilt XGBoost function helps to prevent overfitting in weather / Climate datasets to avoid random or noisy patterns normally observed in dynamic meteorological parameters. The dataset used here is divided into 80:20 ratio for training and testing purposes. Checking of correlations of input variables with the output variables is done. For this XGBoost is used and configured. The hyperparameters are tuned with the help of RandomizedSearchCV class for cross validation. For keeping the validation error minimum, the varied parameters such as learning_rate, max_depth, number_of_estimators, and gamma are tuned for optimization. This table 1 shows the model configured for different parameters while training and validation of data. So as to prevent overfitting and model optimization, validation error is continuously monitored for 20 consecutive iterations, after that is halted to improve model performance.

Table 1. XGBoost ML Model Configuration Table.

Sr. No.	Parameter	Value
1	n_estimators	200
2	max_depth	6
3	learning_rate	0.1
4	objective	reg:squarederror
5	random_state	42

Evaluation metrics known as error metrics were used for assessment of Model Performance. Theses multiple error metrics are includes: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Coefficient of Determination (R^2) .These error metrics give an idea about error spreading and explanatory power used during the Machine Learning Model evaluation. The proposed model can be integrated in smart cities or urban management platforms for many different cases such as energy management systems, real time weather forecasting data accuracy to avoid weather risks and alert the urban elderly population for extreme weather conditions. This will be beneficial for the elderly people residing in high-risk zones to align with the urban sustainability goals and to protect public health.

4 Model Performance

The Machine Learning XGBoost models tested and trained on Pune region weather data (2008-2022) gives an outcome superior to previous conventional Machine Learning like ARIMA and Linear Regression Model by achieving (R^2) score is about 0.91, MSE of 4.70 and MAE of 1.58.

Table 2. Model Performance comparison.

Model.	MSE	MAE	RMSE	R2
Linear Regression	19.16	3.8	4.38	0.01
ARIMA (3,1,2)	20.08	3.94	4.48	-0.03
XGBoost (Tuned)	4.69	1.57	2.16	0.91 ✅

These error metrics (Table2) indicate that the model performs well with 91% of the variance for the given weather forecast parameters which ensures the effectiveness of validation for the real time weather dataset of weather forecasting. This output / results proves about its minimization of errors and maximized accuracy, the XGBoost model ensures climate-aware urban infrastructure for Optimization of HVAC (Heating/Cooling) for elderly housing, Extreme weather alerts, environment based Adaptive smart City planning.

As we know, weather is chaotic in nature, conventional statistical models are unable to perform multivariate and nonlinear patterns which are inherent in climate systems. XGBoost shows its effectiveness in smart urban forecasting platforms where health risk mitigation for aging populations is a key factor for the decision making. So, in the current advancement of technology, conventional models are replaced by ML- driven frameworks for time series data for urban decision systems which are dependent on climate. Following are the results we obtained while comparing the conventional methods with proposed ML model XGBoost. The actual vs predicted graph (figure 1) for weather parameter temperature clearly illustrates the high accuracy for the XGBoost Machine Learning Model.

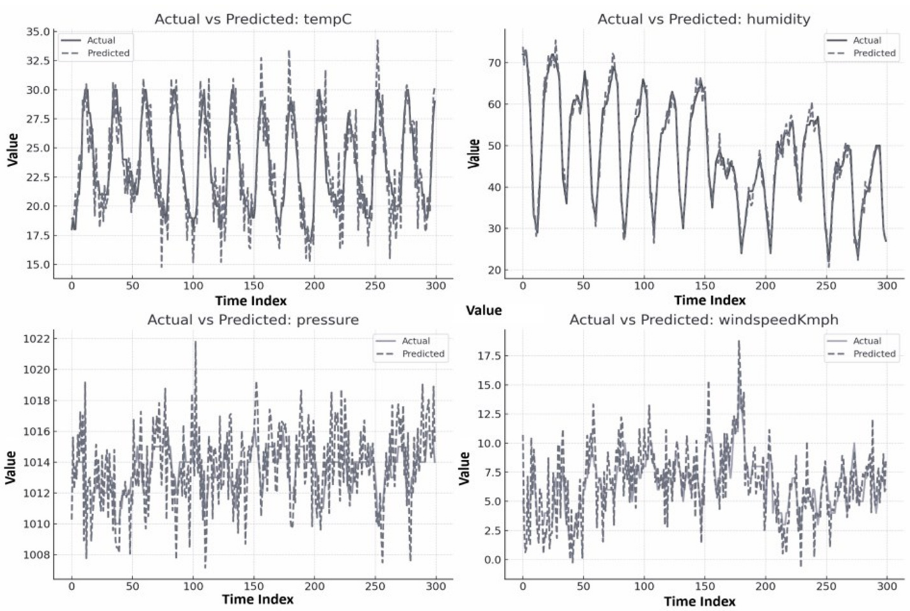

Fig.1. Actual vs Predicted plot using XGBoost for various weather parameters

It shows the model performs more than 90% accurately and follows the real-time oscillations for the observed data across the given test period. Models can capture both short term variability and long-term trends. As the weather parameters data set is randomly distributed time series data , in our proposed Machine Learning methods, dataset were handled using linear interpolation method. Recent studies show that extreme weather events increase the mortality rates among elderly population by 15-20% in countries which are underdeveloped by Mengyuan He et al. [13]. Early warning / alert systems for weather events can help to reduce the hospitalizations of elderly people by nearly 40% according to data from urban resilience programs in South Asia by Romanello, Marina et al. [14] and Jiaming Ya et al.[15]. Based on results obtained, we can conclude that XGBoost Machine Learning Models with high prediction accuracy ($R^2 = 0.91$) which show the strong predictive capability for early accurate weather alerts for weather parameters, enables weather forecasting which will be beneficial for smart cities as new alert systems and ultimately supports longevity of aging populations in rapidly urbanizing cities like Pune.The below chart (figure 2) shows the estimated impact percentage (%) of smart urban technologies on elderly longevity. A relative efficacy score based on interdisciplinary research on urban resilience, health forecasting, and climate adaptation is displayed on the horizontal axis (0–100%). Our study including XGBoost-based climate forecasting not only helps for early weather alerts but also it performs more accurately with 95 % while making comparisons with other smart urban technologies.

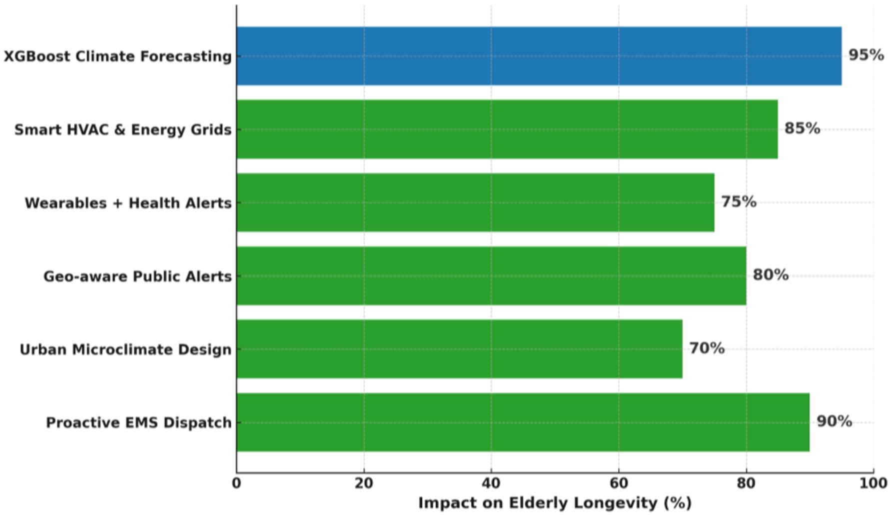

Fig.2. Role of XGBoost in enabling Urban ML applications for Longevity

5 Conclusion

This research proposed the XGBoost Machine Learning Model which demonstrates the efficacy of the model for prediction of various weather parameters like temperature and ensures enhancement in urban sustainability and protection for aging populations. The model was tested and trained with a long-term data set for the Pune region, India which achieved a high R^2 score of 0.91 over the conventional statistical models like ARIMA and LR.The incorporation of such a Machine Learning based climate prediction system into urban infrastructure helps to reduce health risks for the aging population. Ultimately, this research aligns with global sustainability goals which mainly target a real time, scalable framework for essential services regardless of unpredictable climate for the growing elderly population. This paper presents an XGBoost-based forecasting framework for urban sustainability and longevity support. The model provides high predictive accuracy and adaptability for real-time integration in smart city systems. This study is limited as it is mostly focusing on the Pune region of Maharashtra, one single geographical region in India. Also, the model is evaluated using individual Machine Learning Models. Future Work should focus on Hybrid Machine Learning Models with prediction for extreme weather events which incorporates multivariate forecasting with validation of results for multiple cities.

References

1. S.F.Ismael, A.H.Alias,N.A.Haron,B.B.Zaidani:Mitigating Urban Heat Island Effects: A Review of Innovative Pavement Technologies and Integrated Solutions,SDHM Structural Durability and Health Monitoring,18(5),pp 525-551,ISSN 1930-2983,sciencedirect (2024)
2. Zulfani Alfa Sanah, M.Zaim Husnun Niam, Sri Wardiani, Muhammad Ahsan, Muhammad Hisyam Lee : Monitoring air quality index with EWMA and individual charts using XGBoost and SVR residuals,MethodsX,Vol. 14,103107 (2025)
3. Xinyue Ma,Xinyue and Yang, Jun and Zhang, Rui and Yu.: XGBoost-Based Analysis of the Relationship Between Urban 2-D/3-D Morphology and Seasonal Gradient Land Surface Temperature, IEEE Journal of Applied Earth Observations and Remote Sensing, vol. 17, pp. 4109-4124 (2024)
4. Isaev, E., Ajikeev, B., Shamyrkanov, U., Kalnur, K., Maisalbek, K., Sidle, R.C. : Impact of Climate Change and Air Pollution Forecasting Using Machine Learning Techniques in Bishkek. Aerosol Air Qual. Res. 22, 210336 (2022)
5. Garcia, M., Lee, S., & Martinez, R.: Leveraging Technology to Enhance Access to Hope Interventions: Opportunities and Challenges. Journal of Technology in Mental Health, 8, 45-58 (2023).
6. Alex Torku , Albert P.C. Chan : Detecting stressful older adults-environment interactions to improve neighbourhood mobility, Building and Environment, ISSN: 0360-1323, Vol: 224, Page: 109533 (2022).
7. Reddy, M., Naveed, R., & Shah, T.: Urban Health Planning in the Age of AI: Advancements and Opportunities in Machine Learning. International Journal of Sustainable Infrastructure for Cities and Societies, 8(1), 38–52 (2023)
8. Lyu Z, Wang L, Gao X, Ma Y. : The Identification of Elderly People with High Fall Risk Using Machine Learning Algorithms.Healthcare (Basel).Dec 23;11(1):47 (2023)

9. Anita B. Agarwal, Rohit Rajesh,Nitin Arul : Spatially-resolved hyperlocal weather prediction and anomaly detection using IoT sensor networks and machine learning techniques, https://doi.org/10.48550/arXiv.2310.11001(2023)
10. Zhang, Sulin, Wenchao Xu, Yuting Zhu, E. Tian,Weijia Kong : Impaired multisensory integration predisposes the elderly people to fall: a systematic review." Frontiers in neuroscience,PMID: 32410958 Apr 28;14:411(2020)
11. Peralta A, Olivas JA, Romero FP, Navarro P. The Influence of Ambient Factors on Emotional Wellbeing of Older Adults: A Review. Sensors (Basel). 25(4):1071.PMID: 40006299; PMCID: PMC11859640 (2025)
12. Paul KC, Haan M, Mayeda ER, Ritz BR. Ambient Air Pollution, Noise, and Late-Life Cognitive Decline and Dementia Risk. Annu Rev Public Health. 2019 Apr 1;40:203-220. PMID: 30935305; PMCID: PMC6544148 (2019)
13. Mengyuan He, Hong Liu: Machine learning-based assessment of thermal comfort for the elderly in warm environments: Combining the XGBoost algorithm and human body exergy analysis,International Journal of Thermal Sciences, Vol. 209, 109519 (2025)
14. Romanello, Marina: The 2021 report of the Lancet Countdown on health and climate change: code red for a healthy future The Lancet, 398 (10311), 1619 - 1662 (2021)
15. Jiaming Yang, Zhaomin Tong, Jiwei Xu, Rui An, Yanfang Liu, Yaolin Liu:Leveraging machine learning to explore nonlinear associations between urban heat vulnerability and morbidity risk, Urban Climate,Voi. 59, 102320 (2025)