CHAPTER 2

LITERATURE REVIEW

1.1 Introduction to Case Study

Climate change has emerged as a significant global challenge in recent years, impacting various sectors, including the preservation of cultural heritage sites. The increasing frequency and intensity of extreme weather events, along with gradual shifts in temperature, humidity, and wind patterns, have highlighted the need for adaptive solutions to safeguard these invaluable assets.

One promising approach to address these challenges involves the application of advanced algorithms, such as Random Forest and XGBoost, for microclimate monitoring and prediction at cultural heritage sites. These techniques can help preserve and protect these valuable assets by analyzing temperature, humidity, and wind speed data, which are crucial factors in the conservation of these sites.

This case study focuses on the implementation of Random Forest and XGBoost algorithms for microclimate monitoring and prediction at two cultural heritage sites in Johor Bahru, Malaysia: the Johor Bahru High Court and the Sultan Ibrahim Building. By leveraging these advanced techniques and utilizing Power BI dashboards for data visualization and analysis, this research aims to enhance the understanding of site-specific microclimates and inform effective conservation strategies for these historic landmarks.

Through continuous assessment and refinement of these methods, researchers, conservators, and heritage site managers can work together to develop improved strategies for preserving cultural heritage sites like the Johor Bahru High Court and the Sultan Ibrahim Building under changing environmental conditions. By adopting a collaborative approach, we can ensure the protection and preservation of these invaluable assets for future generations to appreciate and learn from, even in the face of challenges posed by climate change.

1.2 Importance of Preserving Cultural Heritage Sites

The preservation of cultural heritage sites holds immense significance for society, history, and identity, as these sites serve as tangible reminders of our shared past, providing valuable insights into the cultural, social, and economic development of human civilizations (Lowenthal, 1985). By protecting and maintaining these sites, we ensure the continuity of our cultural memory and allow future generations to appreciate and learn from the rich tapestry of human history (UNESCO, 1972). Moreover, cultural heritage sites contribute to a sense of belonging and pride within communities, fostering social cohesion and promoting intercultural dialogue (Smith, 2006). Furthermore, preserving these sites can offer economic benefits, as they often attract tourism and stimulate local economies (Timothy & Boyd, 2003). Given these multifaceted advantages, it is crucial to develop and implement strategies to safeguard cultural heritage sites against various threats, including the impact of microclimate factors, to ensure their longevity and continued cultural relevance.

1.3 Impact of Microclimate Factors on Cultural Heritage Sites

Microclimate factors, such as humidity, temperature, wind speed, play a significant role in the deterioration of cultural heritage sites. Existing studies have established the adverse effects of these factors on various materials and structures, leading to both physical and chemical degradation (Cassar, 2005; Camuffo, 2014).

Temperature fluctuations, especially in the presence of moisture, can lead to the expansion and contraction of materials like stone, brick, and mortar, resulting in cracks, delamination, and structural damage (Camuffo, 2014). Moreover, extreme temperatures can accelerate the decay of organic materials, such as wood and textiles, commonly found in cultural heritage sites (Cassar, 2005).

Humidity is another critical factor in the deterioration process. High humidity levels can cause moisture to accumulate in porous materials, leading to the growth of mold and bacteria, which can weaken and damage the structure (Lankester & Brimblecombe, 2012). Additionally, the presence of moisture can facilitate the dissolution of soluble salts in porous materials, causing efflorescence and subflorescence, further compromising structural integrity (Cassar, 2005).

Wind speed, particularly in combination with rain, can exacerbate the erosion of building materials and increase the rate of material loss from structures (Cassar, 2005). Moreover, high wind speeds can cause physical damage to fragile elements, such as decorative features and stained-glass windows (Camuffo, 2014).

In summary, understanding the impact of microclimate factors on cultural heritage sites is crucial for developing effective preservation strategies. By identifying and mitigating the risks associated with these factors, we can better protect these invaluable resources and ensure their continued existence for future generations.

1.4 Traditional Methods for Cultural Heritage Sites Preservation

Traditional methods for cultural heritage site preservation often rely on reactive maintenance approaches. These approaches involve responding to issues and damage after they have already occurred, rather than anticipating and preventing them. Reactive maintenance has several limitations, making it necessary to explore proactive and preventive measures for the preservation of cultural heritage sites (Staniforth, 2013).

Delayed intervention is one of the limitations of reactive maintenance, as it occurs after the damage has been detected, leading to further deterioration or irreversible loss of cultural elements (Muñoz Viñas, 2002). Additionally, reactive maintenance can be expensive, especially if the damage requires extensive interventions and specialized expertise (Stovel, 2005). Incomplete recovery can also be an issue, as advanced damage can result in the loss of original features or materials, compromising the site's authenticity and historical value (Muñoz Viñas, 2002).

Moreover, interventions during reactive maintenance can be invasive or destructive, leading to further damage or exposing other areas to new risks (Matero, 1999).

To address these limitations, there is a need to shift towards proactive and preventive measures, such as regular monitoring, preventive conservation, maintenance planning, and capacity building for local stakeholders. Regular inspections and monitoring can help identify early signs of deterioration or potential threats (Caple, 2008), while preventive conservation can reduce or eliminate risk factors contributing to the site's deterioration, such as controlling humidity and temperature (Muñoz Viñas, 2002). Maintenance planning, including preventive measures and timely interventions, can also help address potential issues (Caple, 2008). Capacity building through training and education for local stakeholders can further enhance the site's preservation efforts (Ashley-Smith, 2016).

In conclusion, the preservation of cultural heritage sites requires a shift towards proactive and preventive measures, which can minimize the risk of irreversible damage, maintain the site's authenticity and historical value, and reduce the overall cost of preservation efforts. By adopting regular monitoring, preventive conservation, maintenance planning, and capacity building for local stakeholders, cultural heritage sites can be better preserved for future generations (Muñoz Viñas, 2002; Caple, 2008; Ashley-Smith, 2016).

1.5 Machine Learning Algorithms

This study develops machine learning-based methods for microclimate monitoring and prediction at cultural heritage sites, using the supervised learning concept. This involves training a classifier to assign labels to specific data points or regions in the dataset, enabling it to identify hidden patterns and signatures of various labelled factors and make accurate predictions. To ensure effective monitoring and prediction using a variety of data sources, it is crucial to use classifiers that can handle large-scale data and achieve high accuracy quickly. The study focuses on two classifiers, XGBoost and Random Forest, which are both capable of achieving these requirements.

1.5.1 Random Forest

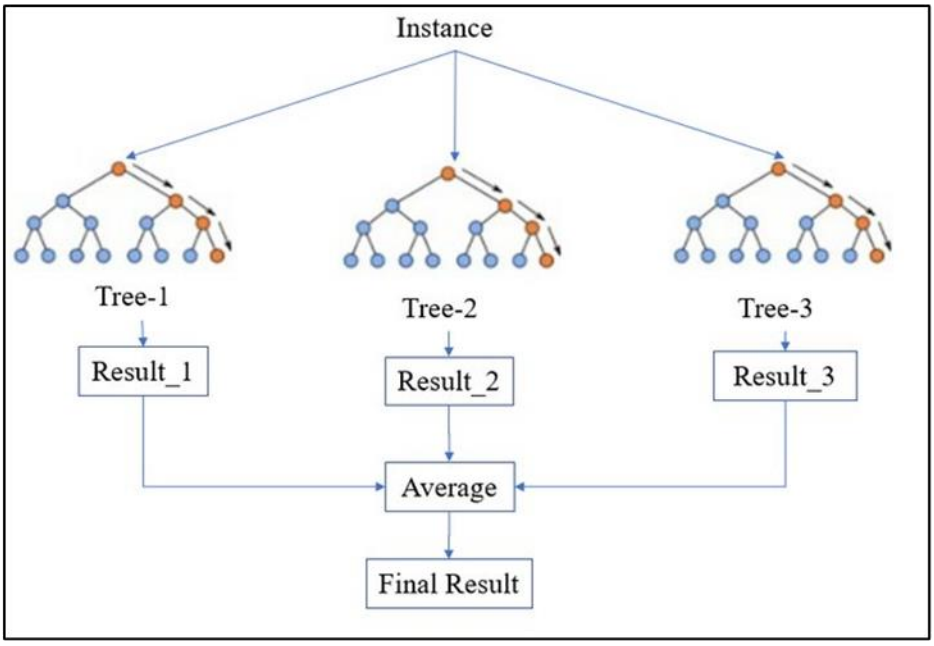

Figure 1: Random Forest Model Architecture

Breiman's Random Forest algorithm, introduced in 2001, is a widely used ensemble learning model that is known for its versatility in performing various tasks such as classification, regression, clustering, interaction detection, and variable selection (Rahmati et al., 2017; Belgiu and Drăguţ, 2016). This learning method leverages the aggregation of decision trees, which divide input data based on specific parameters in a tree-like structure (Ma and Cheng, 2016; Breiman, 2001) (see Fig. 1). Unlike other learning methods, Random Forest is designed to handle complex datasets with high dimensionality, noisy, and missing data, making it particularly useful for microclimate monitoring and prediction at cultural heritage sites.

Each decision tree in a Random Forest model is built using a bootstrapped sample of the data, with nodes split according to the optimal subset and randomly selected predictors at each stage (Araki et al., 2018; Rahmati et al., 2017). The final classification is based on the majority vote of the decision trees, and output is generated accordingly (Micheletti et al., 2014; Rahmati et al., 2017). This approach helps prevent overfitting, where a model learns the training data too well and fails to generalize well to new data. Random Forest's robustness and high-performance capabilities have made it a popular choice in various fields, including image analysis, remote sensing, and ecology.

Furthermore, Random Forest is a highly flexible model that can handle a wide range of input variables, including categorical and continuous variables, and can deal with missing data points by imputing values. The algorithm can also determine the importance of each input variable in predicting the output, enabling researchers to identify the most influential parameters for microclimate monitoring and prediction at cultural heritage sites. Additionally, researchers have developed various extensions and modifications to improve the algorithm's efficiency, such as parallel computing, pruning techniques, and feature importance measures (Balogun et al., 2021; Tella et al., 2021).

One notable feature of Random Forest is its ability to handle interactions between variables, which is important in predicting microclimate parameters at cultural heritage sites. The algorithm can identify and model complex interactions between multiple variables, allowing researchers to better understand the relationships between environmental factors and microclimate patterns. This feature is particularly valuable in cultural heritage sites, where environmental conditions can vary significantly and interactions between environmental factors can be complex.

In conclusion, Random Forest is a powerful machine learning algorithm that has proven to be a valuable tool for microclimate monitoring and prediction at cultural heritage sites. Its robustness, flexibility, and high-performance capabilities make it an attractive choice for handling complex datasets with high dimensionality, noisy, and missing data. Moreover, the algorithm's ability to identify and model interactions between variables provides researchers with valuable insights into the complex relationships between environmental factors and microclimate patterns.

1.5.2 XGBoost

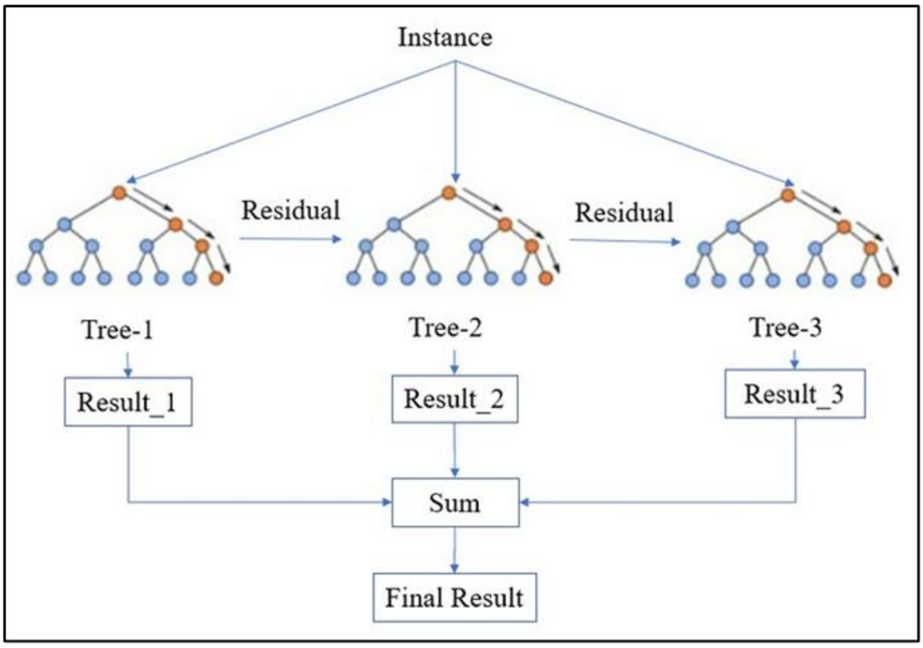

Figure 2: XGBoost Model Architecture

XGBoost is a popular machine learning algorithm that is commonly used for classification tasks. It belongs to the family of boosting algorithms, where multiple weak learners are combined to create a strong model. The algorithm works by iteratively adding decision trees to the model and adjusting their weights based on the error rate of the previous trees. The result is a highly accurate classifier that can handle large and complex datasets.

One of the key advantages of XGBoost is its ability to handle missing data effectively. The algorithm can use surrogate splits to compensate for missing data points, resulting in improved accuracy and robustness in the presence of missing data. XGBoost is also highly optimized for parallel computing, enabling it to process large volumes of data quickly and efficiently.

XGBoost has demonstrated high performance and accuracy when dealing with large-scale, multi-class data in various fields, including remote sensing, medical diagnosis, and natural language processing. Studies have shown that XGBoost can outperform other popular classification algorithms, such as Random Forest and

Support Vector Machines (SVM), in terms of accuracy and efficiency (Bhagwat & Shankar, 2019; Zamani Joharestani et al., 2019; Rumora et al., 2020). This makes it an attractive choice for microclimate monitoring and prediction at cultural heritage sites, where large datasets and high-dimensional feature spaces are common.

XGBoost is highly scalable, which makes it an ideal choice for handling large volumes of satellite data. This enables researchers to perform microclimate monitoring and prediction in real-time, providing valuable insights into the environmental conditions at cultural heritage sites. Additionally, XGBoost is highly optimized for feature selection, allowing researchers to identify the most influential variables for microclimate monitoring and prediction.

In conclusion, XGBoost is a powerful machine learning algorithm that offers several unique advantages for microclimate monitoring and prediction at cultural heritage sites. Its ability to handle missing data, parallel computing, scalability, and feature selection capabilities make it an attractive choice for researchers and practitioners in this field. By leveraging XGBoost's powerful capabilities, researchers can gain valuable insights into the environmental conditions at cultural heritage sites, enabling them to develop more effective strategies for managing and preserving these invaluable assets for future generations.

1.6 Comparative Analysis of Previous Case Studies and the Uses of Machine Learning in Cultural Heritage Preservation

Several studies in the cultural heritage field apply machine learning (ML) techniques for tasks such as automatic text recognition, image annotation, and user preference recommendations. However, the use of ML in conservation science and heritage preservation studies is limited. These studies primarily focus on identifying and classifying materials or structures or using ML to monitor cultural heritage collections or sites for abnormalities. For instance, Zou et al. employed deep learning on image data to locate missing or damaged heritage components in historical buildings, while Kejser et al. used ML to classify the acidity of historic paper samples. Pei et al. utilized machine learning to predict household mite infestation based on indoor climate conditions and found that the extreme gradient boosting (XGBoost) model was the most suitable approach.

Case Study	Method Used	Target Site/Subject	Main Outcomes
Yu et al. (2022)	Convolutional Neural Network Deep Learning	Dunhuang Mogao Grottoes, China	Detected wall painting deterioration; informed preventive measures
(Kumar et al. (2019)	Logistic Regression, Support Vector Machine	Damaged Heritage Sites from 2015 Nepal Earthquake	Classify heritage and not-heritage sites; damage or no damage
Prieto et al., (2017)	Multiple Linear Regression, Fuzzy Logic Models	100 parish churches, located in Seville, Spain	Identifies relevant variables for the functional degradation of the churches.
Gonthier et al., (2019)	Support Vector Machines	Child Jesus, the crucifixion of Jesus, Saint Sebastian	Recognition of iconographic elements in artworks.
Valero et al., (2019)	Logistic Regression, Multi Class Classification	Chapel Royal in Stirling Castle, Scotland	Identifies loss of material defects and discoloration on the walls.

1.7 Implementation of Random Forest and XGboost in Microclimate Monitoring and Prediction

Researchers have been exploring the performance of XGBoost and Random Forest classifier algorithms for microclimate monitoring and prediction in various studies. These algorithms have proven to be effective in providing valuable insights for monitoring and managing microclimate factors in different environments.

In a study by J. Angelin Jebamalar & A. Sasi Kumar (2019), a hybrid light tree and light gradient boosting model were used for predicting PM2.5 levels. The proposed method captured PM2.5 data using a sensor with Raspberry Pi and stored it in the cloud, where the hybrid model was used for prediction. The hybrid model outperformed other algorithms, including Linear Regression, Lasso Regression, Support Vector Regression, Neural Network, Random Forest, Decision Tree, and XGBoost. Despite its advantages in handling large amounts of data and requiring less space, the hybrid model's limitation was its time-consuming nature.

In a study by Maryam Aljanabi (2020), the authors compared Multilayer Perceptron, XGBoost, Support Vector Regression, and Decision Tree Regressor to predict ozone levels based on temperature, humidity, wind speed, and wind direction. After preprocessing the data and performing feature selection, XGBoost emerged as the superior model for predicting ozone levels on a day-to-day basis.

Soubhik et al. (2018) compared various algorithms, including Linear Regression, Neural Network Regression, Lasso Regression, ElasticNet Regression, Decision Forest, Extra Trees, Boosted Decision Tree, XGBoost, K-Nearest Neighbor, and Ridge Regression, to predict air pollutant levels. They found that XGBoost provided better accuracy due to the arrangement of features in decreasing order of importance for predicting upcoming values. Haotian Jing & Yingchun Wang (2020) used XGBoost to predict the air quality index. By employing weak classifiers and using the shortcomings of previous weak classifiers to form a strong classifier, XGBoost reduced the error between predicted and actual values. However, it was susceptible to outliers and unwanted air pollutants, as it took the previous value into account.

Mejía et al. (2018) determined PM10 levels best with Random Forest but found that it did not accurately predict the levels of dangerous pollutants. However, Random Forest had the advantage of working with incomplete datasets. Pasupuleti et al. (2020) compared Decision Tree, Linear Regression, and Random Forest for predicting air pollutant levels using meteorological conditions and data from the Arduino platform. Random Forest provided more accurate results due to reduced overfitting and error. However, it required more memory and incurred higher costs.

In summary, XGBoost and Random Forest have been applied in various case studies for microclimate monitoring and prediction, with both algorithms demonstrating their effectiveness in predicting air pollutant levels. While they have their respective limitations, these advanced techniques offer valuable tools for researchers and practitioners seeking to understand and manage the air quality in different environments.

1.7.1 Comparison Between Random Forest and XGBoost Algorithms

Criteria	**XGBoost**	**Random Forest**
Model Type	Gradient boosting decision tree ensemble (Chen & Guestrin, 2016)	Decision tree ensemble (Breiman, 2001)
Learning Approach	Gradient boosting, optimizing loss function (Friedman, 2001)	Bagging, independent decision trees combined through majority voting or averaging (Liaw & Wiener, 2002)
Handling Missing Data	Imputation or treating missing values as separate categories (Chen & Guestrin, 2016)	Imputation or treating missing values as separate categories (Breiman, 2001)

Overfitting Prevention	Shrinkage and regularization (Chen & Guestrin, 2016)	Averaging results of multiple decision trees (Breiman, 2001)
Interpretability	Can provide feature importance information	Easier to interpret due to simpler decision tree structure (Breiman, 2001)
Speed and Scalability	Slower in training due to sequential nature (Chen & Guestrin, 2016)	Faster and more parallelizable due to independent tree construction (Breiman, 2001)
Performance	Compare using MAE, RMSE, R-squared (Caruana & Niculescu-Mizil, 2006)	Compare using MAE, RMSE, R-squared (Caruana & Niculescu-Mizil, 2006)
Feature Importance	Can rank input variables by importance (Chen & Guestrin, 2016)	Can rank input variables by importance (Breiman, 2001)
Hyperparameter Tuning	Requires tuning, may be more sensitive to hyperparameter settings (Probst, Wright, & Boulesteix, 2019)	Requires tuning, may be less sensitive to hyperparameter settings (Probst, Wright, & Boulesteix, 2019)
Memory Usage	Less memory usage due to sequential nature (Chen & Guestrin, 2016)	More memory usage due to storage of multiple decision trees (Breiman, 2001)

1.8 Chapter Summary

Through this chapter, the study focuses on the preservation of cultural heritage sites through machine learning-based microclimate monitoring, with a specific focus on the application of Random Forest and XGBoost algorithms at two heritage sites in Johor Bahru, Malaysia. The review begins with an overview of the impact of climate change on cultural heritage sites and the importance of their preservation. It then explores the impact of microclimate factors on cultural heritage sites, including temperature, humidity, and wind speed, and the traditional reactive methods used for preservation. The limitations of reactive maintenance and the need for a shift towards proactive and preventive measures are discussed, such as regular monitoring and preventive conservation. Finally, the review explains the use of machine learning algorithms in microclimate monitoring and prediction, specifically Random Forest and XGBoost, and their application in this study. The review highlights the significance of using machine learning-based approaches for preserving cultural heritage sites and the potential benefits of incorporating them into preservation strategies.

REFERENCES

Lowenthal, D. (2015). The Past is a Foreign Country. Cambridge: Cambridge University Press.

UNESCO. (1972). Convention Concerning the Protection of the World Cultural and Natural Heritage. Paris: UNESCO.

Smith, L. (2006). Uses of Heritage. London: Routledge.

Timothy, D. J., & Boyd, S. W. (2003). Heritage Tourism. Harlow: Prentice Hall.

Cassar, M. (2005). Climate Change and the Historic Environment. London: English Heritage.

Camuffo, D. (2014). Microclimate for Cultural Heritage: Conservation, Restoration, and Maintenance of Indoor and Outdoor Monuments. Amsterdam: Elsevier.

Lankester, P., & Brimblecombe, P. (2012). The impact of future climate change on historic interiors. Science of The Total Environment, 417-418, 248-254.

Yu, T., Lin, C., Zhang, S., Wang, C., Ding, X., An, H., Liu, X., Qu, T., Wan, L., You, S., Wu, J., & Zhang, J. (2022). Artificial Intelligence for Dunhuang Cultural Heritage Protection: The project and the dataset. International Journal of Computer Vision, 130(11), 2646–2673. https://doi.org/10.1007/s11263-022-01665-x

Kumar, P., Ofli, F., Imran, M., & Castillo, C. (2020). Detection of disaster-affected cultural heritage sites from social media images using Deep Learning Techniques. Journal on Computing and Cultural Heritage, 13(3), 1–31. https://doi.org/10.1145/3383314

Staniforth, S. (2013). Historical Perspectives on Preventive Conservation. Getty Conservation Institute.

Stovel, H., Stanley-Price, N., & Killick, R. G. (2005). Conservation of living religious heritage: Papers from the ICCROM 2003 forum on living religious history: Conserving the sacred. International Centre for the Study of the Preservation and Restoration of Cultural Property.

Muñoz Viñas (2002) Contemporary theory of conservation, Studies in Conservation, 47:sup1, 25-34, DOI: 10.1179/sic.2002.47.Supplement-1.25

Ashley-Smith, J. (2016).Risk assessment for object conservation. Routledge.

Caple, C. (2012). Preventive conservation in museums. Routledge.

Matero, F. (1999). Lessons from the Great House: Condition and treatment history as prologue to site conservation and management at Casa Grande Ruins National Monument. Conservation and Management of Archaeological Sites, 3(4), 203–224. https://doi.org/10.1179/135050399793138482

Prieto, A. J., Silva, A., de Brito, J., Macías-Bernal, J. M., & Alejandre, F. J. (2017). Multiple linear regression and fuzzy logic models applied to the functional service life prediction of Cultural Heritage. Journal of Cultural Heritage, 27, 20–35. https://doi.org/10.1016/j.culher.2017.03.004

Gonthier, N., Gousseau, Y., Ladjal, S., & Bonfait, O. (2019). Weakly supervised object detection in artworks. Lecture Notes in Computer Science, 692–709. https://doi.org/10.1007/978-3-030-11012-3_53

Valero, E., Forster, A., Bosché, F., Hyslop, E., Wilson, L., & Turmel, A. (2019). Automated defect detection and classification in ashlar masonry walls using machine learning. Automation in Construction, 106, 102846. https://doi.org/10.1016/j.autcon.2019.102846