Journal of Telecommunication, Electronic and Computer Engineering
ISSN: 2180 – 1843 e-ISSN: 2289-8131 Vol. 17 No. 1 (2025) 1-9
jtec.utem.edu.my
DOI: https://doi.org/10.54554/jtec.2025.17.01.001

Evaluation Metrics for Air Quality Optimization Utilizing Machine Learning: PRISMA Review

Amir Hamzah Mohd Shaberi[1], Sumayyah Dzulkifly[1,2]*, Wang Shir Li[1] and Che Zalina Zulkifli[1,2]
[1]*Department of Software Engineering and Smart Technology, Faculty of Computing and Meta-Technology, Sultan Idris Education University, 35900 Tanjong Malim, Perak, Malaysia*
[2]*Centre of Education Green Technology (EduGreen), Faculty of Computing and Meta-Technology, Sultan Idris Education University, 35900 Tanjong Malim, Perak, Malaysia*

Article Info

Article history:
Received Jul 9th, 2024
Revised Aug 21st, 2024
Accepted Dec 2nd, 2024
Published Mar 19th, 2025

Index Terms:
Air Quality
Machine Learning
Evaluation Metrics
PRISMA

Abstract

Air quality, both indoors and outdoors, is crucial for public health as it affects respiratory conditions and overall well-being. Machine learning (ML) techniques offer innovative solutions for monitoring and predicting air quality. However, choosing the right ML algorithms and evaluation metrics is essential for creating accurate air quality prediction models, as these choices directly impact the accuracy and reliability of the results. This study aims to explore and summarize the literature on the use of ML techniques in predicting and optimizing air quality, addressing the urgent issue of air pollution. The review analyzed papers from two electronic databases, namely Scopus and Science Direct. Information on ML techniques for predicting air quality and identifying the main sources of pollution was extracted from 26 studies. The study focuses on common ML techniques employed in air quality prediction, including classification, deep learning, regression, ensemble learning, and combinations of regression and deep learning. It also identifies the evaluation metrics used to assess the performance of these models, such as recall, root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). By synthesizing existing knowledge, this study provides a comprehensive understanding of the metrics researchers use to measure the overall effectiveness of ML algorithms. It serves as a benchmark for future research and guides the selection of appropriate evaluation metrics in the field of air quality prediction.

*Corresponding Author: sumayyah.dzul@meta.upsi.edu.my

I. INTRODUCTION

Air quality refers to the condition of the air in our surroundings, specifically in terms of the presence of pollutants and their impact on human health and the environment. It is a crucial aspect of environmental health, directly affecting the well-being of both humans and ecosystems. Such harmful substances can cause discomfort, illness, and even result in millions of deaths each year [1]. Therefore, air pollution has become a major global concern, prompting experts to focus on monitoring, maintaining, and improving air quality [2].

In recent years, the integration of Machine Learning (ML) techniques has emerged as a promising approach for predicting and optimizing air quality. As societies deal with the consequences of air pollution, it becomes imperative to understand the effectiveness of ML models in this context [3]. The rapidly developing field of ML has shown significant potential in addressing environmental challenges, with a particular focus on predicting and optimizing air quality. With the ongoing growth of industries and urbanization, the harmful effects of air pollution have become increasingly apparent [4].

In this context, the selection of ML algorithms and evaluation metrics plays a crucial role in developing effective models for predicting air quality [5]. Various ML algorithms, including supervised, unsupervised, and ensemble methods, possess different strengths for handling the complexities of air quality data, ranging from predicting pollutant levels to identifying patterns and anomalies. The choice of the most suitable algorithm depends on factors such as the characteristics of the data, the specific objectives of the prediction, and the need for interpretability [6].

Equally important is the choice of evaluation metrics, which assess the performance and reliability of these models. Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared (R^2), and precision-recall scores provide insights into the models' performance and suitability for real-world applications. A thorough understanding of both the algorithms and metrics is essential for optimizing air quality models and ensuring they provide accurate, actionable insights.

A few Systematic Literature Reviews (SLR) have been conducted in the field of air quality. According to [7], mosses have been used as natural sensors to fill gaps in air quality records, particularly in developing economies that face air quality challenges. However, their study focused solely on moss-based passive biomonitoring. In addition, [8] reviewed urban air pollution by integrating ML in smart cities, but the review only examined supervised learning algorithms and their ability to accurately predict key pollutants, disregarding other potential approaches and applications of ML in air quality management. Furthermore, a study by [9] reviewed 30 papers to compare different methods for measuring and forecasting air pollution levels in cities.

However, the existing systematic literature reviews (SLRs) did not comprehensively discuss evaluation metrics for multiple algorithms in ML. Therefore, to address this gap in the literature, this study aims to comprehensively and systematically review multiple algorithms in ML and evaluation metrics, with a specific focus on predicting air quality for various purposes. By doing so, it seeks to provide a clearer understanding of how different models perform under different conditions. Consequently, this study is guided by two research questions as follows:

1. How do different ML algorithms perform in predicting multiple air quality parameters?

2. How effective are various evaluation metrics in assessing the performance of ML models for multipurpose air quality prediction?

II. MATERIALS AND METHODS

This systematic review follows the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) methodology, a standardized approach for evidence-based reporting in meta-analyses and systematic reviews. PRISMA is a global initiative created to address issues such as the lack of transparent and well-documented review methods in published review articles [9]. It provides a minimum set of items to ensure comprehensive reporting and transparency in research synthesis [8]. The PRISMA framework protocol consists of four stages: (1) identification, (2) screening, (3) eligibility, and (4) study selection. Figure 1 summarizes the stages in the PRISMA framework, following the guidelines of [10].

To address the objectives in this study, the process was divided into distinct stages. First, keywords and search terms, including synonyms and related terms, were identified by consulting thesauruses, dictionaries, encyclopedias, and existing research. Next, screening and eligibility criteria were established to select the most relevant articles from the database. After that, data extraction was conducted based on predefined research questions. Furthermore, the findings related to these questions were presented, highlighting comparisons and limitations in the field. These procedures are explained in more detail in the following sections. Finally, the review synthesizes the results to draw comprehensive conclusions and provide recommendations for future research directions in the application of ML to air quality prediction.

A. Identification

Two different databases, Scopus and Science Direct, were used to find relevant articles related to the research questions outlined in the introduction section. Search strings for the

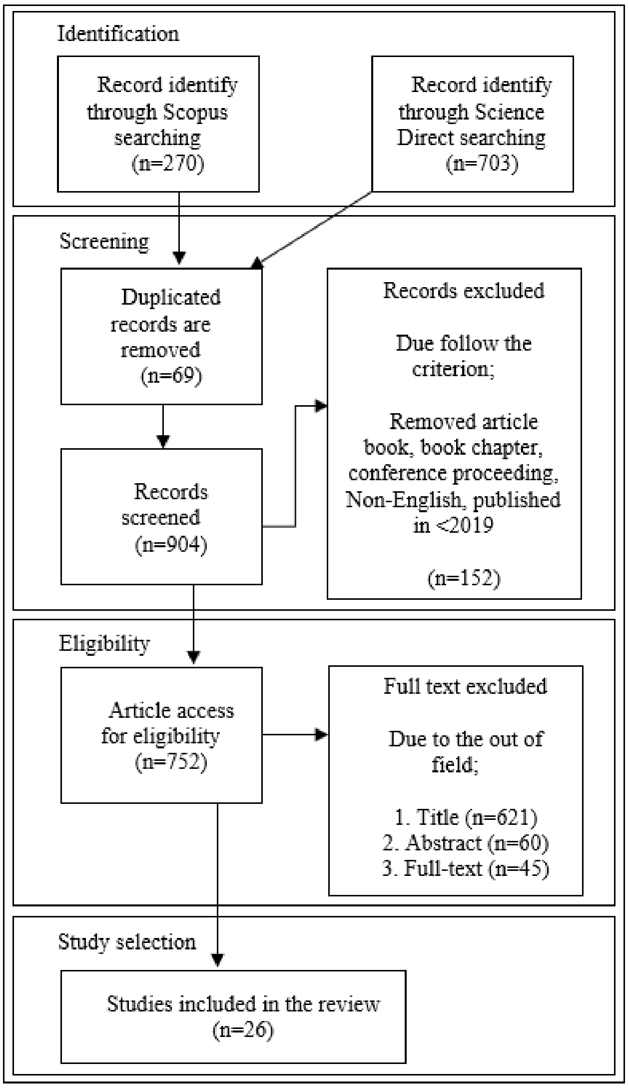

Figure 1. Flow diagram for the proposed search "Evaluation Metrics for Air Quality Optimization Utilizing Machine Learning."

Scopus and Science Direct databases, as displayed in Table 1, were created after identifying all relevant terms. In the initial stage of the systematic review process, a total of 973 articles were retrieved successfully from both databases. These articles were then evaluated based on the established inclusion and exclusion criteria. After this thorough screening process, the most relevant articles were chosen for in-depth analysis to ensure a comprehensive coverage of the research questions.

Table 1
Query search keywords for the two databases

Database	Query
Scopus	(("Indoor air quality" OR "Air quality") AND ("Machine learn*" OR "Artificial intelligence") AND ("Evaluation metrics"))
Science Direct	("Indoor air quality" OR "Air quality") AND ("Machine learning" OR "Artificial intelligence") AND ("Evaluation metrics")

B. Screening

In the initial screening phase, we made deliberate efforts to exclude duplicated articles, ensuring that only distinct and unique articles were considered for further analysis. As a result, a total of 69 duplicated articles were excluded in this phase. In the second phase, 904 articles were meticulously screened using carefully crafted inclusion and exclusion criteria, as shown in Table 2. The primary criterion prioritized research articles as the primary source of practical information. Consequently, publications such as systematic reviews, reviews, book series, books, chapters, and conference proceedings were excluded. Additionally, we focused exclusively on English-language articles, which is widely accepted in the academic community, to maintain consistency in the analytical process.

The study's timeframe spanned six years, from 2019 to 2024, allowing for a comprehensive examination of relevant literature within a specific period. To align with the research question, the review focuses specifically on research articles involving air quality prediction methods related to ML subdomains. Evaluation metrics were also considered, ensuring a focused and contextually relevant investigation. As a result, a total of 152 publications were excluded during the screening phase based on these strict parameters, demonstrating a commitment to selecting only high-quality and pertinent articles for subsequent research stages.

Table 2
Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Language	English	Non-English
Time line	2019 - 2024	< 2019
Literature type	Journal (Article)	Systematic reviews, reviews, book series, books, chapters and conference proceedings
Publication stage	Final	In Press

C. Eligibility

In the third phase, designated as the eligibility assessment, a total of 752 articles were gathered. During this stage, a comprehensive review of the titles and key content of all articles was conducted to verify their compliance with the inclusion criteria and their relevance to the current research objectives. Subsequently, 726 articles were excluded due to their lack of relevance to the research field, insignificant titles, or absence of abstracts aligned with the study's objectives supported by empirical evidence. Consequently, 26 articles were retained based on the criteria set for further examination.

D. Study Selection

Data relevant to the systematic review were extracted from the selected publications for further analysis. The extraction process included collecting the following information:

- Authors titles, abstracts and full-text.
- Year of publication and associated database.
- Focus on ML algorithms for air quality prediction and the use of evaluation metrics.

After extracting the data, the authors thoroughly synthesized and analyzed the included articles to address the research question. Additionally, they organized the extracted data based on the ML algorithms and evaluation metrics utilized in each study. This categorization allowed for a comprehensive analysis of the various approaches and methodologies employed in predicting air quality using ML techniques.

III. RESULTS AND FINDINGS

The systematic review identified 26 relevant articles that meet the criteria for computationally efficient ML algorithms designed to address the specific challenges of air quality. Table 3 provides a summary of these publications, organized based on criteria established by the researchers. These criteria include the common theme of the publications, articles and methods used, indoor parameters, outdoor parameters, and performance metrics.

A. Comprehensive Overview of Machine Learning Models and Evaluation Metrics in Air Quality Prediction

Table 3 shows the systematic literature review focused on using ML for air quality prediction. The review includes various studies that explore different models and evaluation metrics, providing a comprehensive examination of both indoor and outdoor parameters. The classification models primarily utilize random forest [11]-[14] and are evaluated using metrics such as specificity, precision, recall, accuracy, F1 score, AUC, and sensitivity. One study stands out as it incorporates multiple algorithms (decision tree, random forest, k-nearest neighbor, and support vector machine) to predict air quality and evaluate R^2 [16]. This approach demonstrates a strong assessment strategy to gauge the performance of these models in air quality prediction.

A substantial portion of the reviewed literature focuses on deep learning approaches. 12 studies [17]-[27], [35] utilized different architectures like LSTM, CNN, and unique combinations such as SA–EMD–LSTM and EEMD-SSA-LSTM. The evaluation metrics used in these studies include MAE, RMSE, SMAPE, MAPE, IA, TIC, R^2, and Absolute Average Deviation. This diversity of metrics highlights the complexity and multifaceted nature of evaluating deep learning models for air quality prediction, considering both spatial and temporal aspects.

The review also examined six regression models [13], [15], [29]-[32]. These models utilize various techniques like ANN, Extra Trees Regressor, Wavelet Artificial Neural Network, and Support Vector Regression. The evaluation metrics used for these models include MAE, RMSE, SMAPE, MAPE, and R. The inclusion of regression models in the literature is valuable as they offer continuous air quality prediction insights, which complements the categorical predictions made by classification models.

Ensemble learning is a collective intelligence approach to air quality prediction, demonstrated in four articles that utilize Gradient Boosting, AdaBoost and KPCA, and Stacking Ensemble [15], [33]-[35]. In this category, evaluation metrics include precision, recall, accuracy, error rate, F1 score, MAE, RMSE, IA and Pearson's correlation coefficient. The use of ensemble methods showcases the advantages of combining multiple models to enhance predictive performance. These articles emphasize the effectiveness of ensemble methods in improving the accuracy and reliability of air quality prediction models, providing valuable insights for air quality optimization.

Table 3
Summary of the research articles yielded from the proposed search criteria

Common Theme	Articles & Methods	Indoor parameters	Outdoor parameters	Performance metrics
Classification Models	2021 [11] Random Forest	Not Applicable	PM_{10}	Specificity, precision, recall, and overall accuracy
	2023 [12] Random Forest	Not Applicable	Horizontal visibility, Air temperature, Atmospheric pressure, Relative humidity, Mean wind speed, Dewpoint temperature, PM_{10}, NO_2, SO_2, O_3	Accuracy, recall, F1 score and precision
	2022 [13] Random forest	Not Applicable	$PM_{2.5}$, PM_{10}, and O3	Accuracy, area under the ROC Curve (AUC), recall, precision and F1 score
	2022 [14] Random forest	Not Applicable	$PM_{2.5}$	Accuracy, sensitivity, specificity, precision and recall
	2019 [15] Decision Trees	Not Applicable	Sulphur dioxide, nitrogen dioxide, PM_{10}, and $PM_{2.5}$	Precision, recall, accuracy, error rate and F1 score
	2024 [16] Decision Tree, Random Forest, K-Nearest Neighbor, and Support Vector Machine	VOC	Not Applicable	Accuracy and Coefficient of the determination (R^2)
Deep Learning	2022 [17] LSTM	Not Applicable	$PM_{2.5}$ and PM_{10}	Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Symmetric Mean Absolute Percentage Error (SMAPE)
	2023 [18] Convolutional Neural Network (CNN)	Not Applicable	$PM_{2.5}$ and PM_{10}	RMSE, MAE, Mean Absolute Percentage Error (MAPE), index of agreement (IA) and Theil inequality coefficient (TIC)
	2023 [19] LSTM	CO_2, PM and VOC	Not Applicable	MAE, RMSE and correlation coefficient (R)
	2021 [20] Spatiotemporal Association Analysis LSTM (STAA-LSTM)	Not Applicable	$PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3, NH3, and Pb	Absolute Average Deviation (AAD), RMSE, MAE, and R
	2022 [21] Extended Stationary Wavelet Transform (ESWT) and the Nested Long Short-term Memory (NLSTM) Neural Network	Not Applicable	$PM_{2.5}$	Coefficient of the termination (R^2), MAE, RMSE, and MAPE
	2023 [22] SA–EMD–LSTM: Combines Self-Attention (SA) mechanism, Empirical Mode Decomposition (EMD) algorithm, and Long-Short Term Memory (LSTM) network	$PM_{2.5}$	Not Applicable	R^2, RMSE, MAE, and MAPE
	2024 [23] Convolutional Long Short-Term Memory	Not Applicable	$PM_{2.5}$, PM_{10}, O_3, NO, NO_2, NOx, SO_2, CO, Toluene (TOL), Benzene (BEN) and Ethylbenzene (EBE)	RMSE, normalized RMSE (nRMSE), and MAE
	2022 [24] Combination of Long-Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)	Not Applicable	$PM_{2.5}$	R^2, MAE, and RMSE
	2023 [25] LSTM	Not Applicable	$PM_{2.5}$	MAE and R^2
	2023 [26] (VMD-TCN) : Variational mode decomposition (VMD) and Temporal convolutional network (TCN)	Not Applicable	Not state but discuss about Index	RMSE, MAPE MAE and R^2
	2023 [27] (EEMD-SSA-LSTM) : Ensemble Empirical Mode Decomposition- Sparrow Search Algorithm	Not Applicable	PM_{10}, SO_2, NO_2, O_3 and CO	RMSE and MAE
	2024 [28] Convolutional Neural Network (CNN) and Long-Short-Term Memory (LSTM)	Not Applicable	PM2.5 and PM10	R^2 and RMSE
Regression Models	2023 [29] ANN	Pressure and temperature	Not Applicable	R^2 and RMSE
	2023 [30] Extra Trees (ET) Regressor	Not Applicable	$PM_{2.5}$	RMSE, MAE, MAPE, and SMAPE
	2023 [31] Wavelet Artificial Neural Network (WANN)	Not Applicable	$PM_{2.5}$	Pearson correlation coefficient (R), MAPE, RMSE, and MAE

	2020 [32] Artificial neural network (ANN)	Not Applicable	Not state but discuss about API reading	MAPE, RMSE and computational time
	2022 [13] Support vector regression	Not Applicable	$PM_{2.5}$, PM_{10}, and O_3	RMSE and MAE
	2019 [15] Support Vector Regression	Not Applicable	Sulphur dioxide, nitrogen dioxide, PM_{10}, and $PM_{2.5}$	R, R^2, min max accuracy and mean absolute percentage error
Ensemble Learning	2019 [15] Stacking Ensemble	Not Applicable	Sulphur dioxide, nitrogen dioxide, PM_{10} and $PM_{2.5}$	Precision, Recall, accuracy, error rate and F1 score
	2022 [33] Gradient boosting	Not Applicable	$PM_{2.5}$	MAE, RMSE, IA and Pearson's correlation coefficient (R)
	2023 [34] XGBoost	Not Applicable	Not state but discuss about Index	MAE, RMSE, and R^2
	2024 [35] AdaBoost and Kernel Principal Component Analysis (KPCA)	$PM_{2.5}$	Not Applicable	R^2
Regression Models and Deep Learning	2023 [36] Combines support vector machine (SVM) + long short-term memory (LSTM)	Not Applicable	$PM_{2.5}$	MAE, MSE, RMSE, and (R^2)

A unique contribution to the literature is a study that combines Regression Models and Deep Learning [36]. In particular, this study integrates SVM and LSTM, and evaluates the hybrid model using metrics such as MAE, RMSE, MSE, and R^2. These metrics provide a comprehensive assessment that combines the strengths of both regression and deep learning paradigms. By demonstrating the potential of combining different ML approaches, this study shows how more accurate and reliable air quality predictions can be achieved. The integration of regression models and deep learning in this hybrid approach offers a novel and effective method for addressing complex challenges in air quality prediction.

Table 3 above presents a wide range of ML applications for predicting air quality, including different models and evaluation metrics. By considering both indoor and outdoor parameters, along with various models and metrics, the table provides a comprehensive view of the challenges and progress in this important field. Additionally, the table emphasizes the significance of using diverse evaluation metrics to accurately assess the performance of ML models. This comprehensive evaluation approach enables researchers and practitioners to choose the most appropriate models and metrics for predicting air quality in different situations, ultimately enhancing the effectiveness of air quality management strategies.

IV. DISCUSSION

This study delves into the realm of air quality prediction using ML models, with a specific focus on both indoor and outdoor parameters. The research commences with a comprehensive review of existing literature, which underscores the predominance of deep learning models in this field as opposed to other models. Additionally, this study investigates the evaluation metrics utilized across various model categories, uncovering a range of metrics designed to assess the predictive abilities of each model type. These metrics provide a thorough comprehension of the performance of ML models in air quality prediction, enabling a detailed analysis of their strengths and limitations in relation to both indoor and outdoor parameters.

A. Dominance of Deep Learning in Air Quality Prediction

The systematic literature review focused on indoor and outdoor parameters examines the use of ML models for air quality prediction. The findings reveal a clear preference for deep learning models over traditional approaches. Specifically, the study shows that deep learning techniques, such as LSTM, CNN, and their hybrid variations, are the dominant choices, as evidenced by 12 analyzed articles. This aligns with the results of [37], which demonstrate that deep learning models outperform shallow ML models and conventional data analysis techniques across various applications. This correlation emphasizes the growing consensus within the scientific community regarding the superior performance of deep learning methodologies. The widespread adoption of these models can be attributed to the complex and dynamic nature of air quality data, as well as the adaptability of deep learning models.

Firstly, air quality data, especially in the context of indoor and outdoor parameters, is complex and dynamic. To capture the intricate patterns and dependencies, models with high capacity are often required. Deep learning models, such as LSTMs, are well-suited for this task. They excel in modelling long-range dependencies and temporal relationships within sequential data, making them particularly effective for time-series prediction - a crucial aspect of air quality forecasting [38]. Furthermore, a study by [39] introduced an LSTM-Autoencoder-based hybrid deep learning model that achieved an impressive accuracy rate of 99.50% in detecting anomalies in Indoor Air Quality (IAQ) time-series data. LSTM models offer notable advantages in air quality prediction; they not only reduce execution time but also alleviate computational complexity during the training phase, making them more efficient for handling large and intricate datasets. Moreover, another study by [40] highlighted that deep learning models have the ability to solve complex data problems. Leveraging their deep neural network architectures, these models can automatically learn hierarchical features and patterns from raw data. This enables them to make accurate predictions and uncover hidden relationships that traditional models may overlook.

Secondly, the reviewed articles demonstrate the utilization of a wide range of deep learning architectures, such as STAA-

LSTM, Convolutional Long Short-Term Memory, and SA-EMD-LSTM, highlighting the versatility of deep learning in addressing different aspects of air quality prediction. The ability of deep learning models to adapt to various data modalities and structures contributes to their extensive application in the literature. Furthermore, the combination of deep learning with advanced signal processing techniques, as observed in models like EEMD-SSA-LSTM and VMD-TCN, further improves the model's capacity to extract relevant features and patterns from complex air quality data, resulting in comprehensive and accurate predictions. Additionally, the study by [39] supports the notion that employing stateful LSTM and 3D CNN effectively captures crucial spatiotemporal correlation characteristics, thereby enhancing the accuracy and stability of air quality predictions across different spatiotemporal scales.

While references to traditional models like Random Forests, Decision Trees, and Regression Models are still present in the literature, there is clearly a more limited focus on these approaches compared to the emerging techniques. In fact, only six articles discuss classification models and regression models, highlighting this narrower perspective. It is recognized that these traditional methods, though well-established, do not fully capture the complexities of modern air quality data, which often requires more sophisticated modelling techniques [41]. The limited representation in recent studies suggests a shift towards more advanced methods, such as deep learning, which have enhanced capabilities for handling large and complex datasets. This trend emphasizes the importance of exploring both traditional models and newer approaches to gain a comprehensive understanding of their effectiveness in air quality prediction.

The effectiveness of ensemble learning techniques, such as AdaBoost, Gradient Boosting, and Stacking Ensemble, is also acknowledged, although these approaches are used less frequently compared to deep learning techniques. Moreover, the existing study highlights a growing interest in combining ML models with domain knowledge, such as the physical and chemical principles that govern air quality. This integration aims to improve model interpretability and generalizability, indicating a shift towards more holistic approaches that leverage both data-driven techniques and domain expertise to enhance air quality prediction models [42]. By integrating empirical knowledge with advanced ML techniques, this holistic approach can lead to better performance in accurately predicting air quality, as well as providing insights into the underlying mechanisms driving air pollution.

Deep learning models are prevalent in systematic literature reviews on air quality prediction due to their proficiency in handling intricate data relationships, adaptability to different data modalities, and integration of advanced signal processing techniques for feature extraction. As a result, they improve prediction accuracy in the challenging domain of air quality forecasting. The use of deep learning models also underscores the importance of leveraging sophisticated algorithms to extract meaningful patterns from complex air quality datasets. This, in turn, enables more informed decision-making for environmental management. The prominence of these models demonstrates the ongoing evolution of ML techniques in addressing critical environmental issues like air pollution. It also sets a precedent for future research in the field.

B. Evaluation Metrics Across Model Categories

This study focused on predicting air quality using ML models. Various evaluation metrics were used to assess classification models like random forest and decision tree. These metrics include specificity, precision, recall, accuracy, F1 score, AUC, and sensitivity, providing a comprehensive understanding of the model’s ability to classify air quality conditions [43]. However, multi-class classification has a limitation in terms of its lower limits, as mentioned in [44]. The minimum value ranges between -1 and 0 and varies depending on the number and distribution of classes in the dataset. To overcome this limitation, deep learning models offer a key advantage by adapting to variations in air quality data, enhancing the accuracy and robustness of predictions across different environmental contexts.

For the deep learning models, which were a significant part of the reviewed articles, a variety of evaluation metrics are used. Deep learning is highly effective in automatically learning hierarchical representations from data, allowing for extraction of complex patterns and features without explicit programming [45]. As a data-driven technology that has gained attention in recent years, deep learning has made significant advancements in the field of optical metrology [46]. Metrics of these models include MAE, RMSE, SMAPE, MAPE, Index of Agreement (IA), Theil Inequality Coefficient (TIC), R^2, and Absolute Average Deviation. This diverse range of metrics reflects the complexity of evaluating the predictive capabilities of deep learning models, considering both spatial and temporal aspects in air quality scenarios. Consequently, many studies now focus on combining deep learning with regression techniques to capitalize on the strengths of both approaches, resulting in more robust and accurate air quality predictions.

Regression models, including ANN, Extra Trees Regressor, WANN, and Support Vector Regression, were assessed using metrics such as MAE, RMSE, SMAPE, MAPE, and R. These metrics aim to evaluate the accuracy and precision of predicted numerical values in regression scenarios. While regression models provide a straightforward and interpretable approach to analyze variable relationships, they assume linear relationships, which may oversimplify complex, non-linear patterns in the data [47]. Although regression models offer simplicity and interpretability, their effectiveness may be limited when dealing with non-linear relationships inherent in air quality data [48]. Therefore, there is a growing interest in exploring more advanced techniques, such as deep learning and ensemble learning, to capture these intricate relationships and enhance the accuracy of air quality predictions.

Ensemble learning, which includes gradient boosting and stacking ensemble models, combines classification and regression metrics. To assess the overall performance of the ensemble models, metrics such as precision, recall, accuracy, error rate, F1 score, MAE, RMSE, Index of Agreement (IA), and Pearson's correlation coefficient were used. A practical tutorial on bagging and boosting based ensembles for ML showed that these methods are effective in various scenarios, such as standard multi-class problems, categorical features, and large datasets [49]. The use of ensemble learning techniques is an effective strategy for improving air quality prediction models by leveraging the strengths of multiple models [42]. This approach not only improves predictive performance but also provides a comprehensive evaluation of model performance using diverse metrics.

One unique contribution in the literature review involved integrating regression models and deep learning, specifically by combining SVM and LSTM. The evaluation metrics for this hybrid model included MAE, RMSE, MSE, and R^2, showcasing a fusion of traditional regression and deep learning assessment criteria. This unique contribution to the literature highlights the potential benefits of combining SVM and LSTM networks, two powerful techniques in ML and time series analysis. While conventional time series models like SARIMA have been widely employed for air quality forecasting, they may struggle to capture complex nonlinear patterns in the data [50]. This integration demonstrates a novel approach to air quality prediction, leveraging the strengths of both regression models and deep learning to improve the accuracy and robustness of predictions in this challenging domain.

This systematic literature review found various evaluation metrics designed for each model category's unique characteristics. This approach ensures a thorough understanding of the strengths and limitations of ML models in predicting air quality, both indoors and outdoors. Additionally, the review emphasized the growing use of deep learning models in air quality prediction due to their ability to handle complex data relationships. The increasing prominence of ensemble learning techniques also indicates a trend towards using multiple models to enhance prediction accuracy. In summary, this review emphasizes the significance of advancing ML methodologies to tackle the evolving challenges in air quality prediction and management.

C. Addressing Gaps in Air Quality Prediction Studies

Many previous studies on predicting air quality have failed to recognize the importance of including continuous, dynamic occupancy data. While a few studies touch on this topic, very few delve deeply into the complex relationship between occupancy patterns and air quality factors. Occupancy is a key factor in managing indoor air quality [51], as the number of occupants directly affects the levels of CO_2, humidity, and other pollutants in enclosed spaces. By not analyzing continuous occupancy data, researchers miss an opportunity to develop more precise and responsive models for indoor air quality that can adapt to changes in occupancy in real-time. Additionally, incorporating continuous occupancy data into these models can lead to more effective strategies for controlling ventilation rates [52], optimizing energy consumption [53], and ultimately enhancing both indoor air quality and occupant comfort.

Another gap in this study is the limited discussion of key indoor variables. While some studies focus on a few variables such as CO_2 and PM levels, many critical variables are often overlooked. Factors like temperature, humidity, and VOCs can significantly affect IAQ and must be considered alongside other variables to comprehensively understand the dynamics of IAQ. Neglecting these variables can result in incomplete models that fail to capture the complexity of IAQ management [54]. Furthermore, the interaction between these indoor variables is often intricate, necessitating sophisticated modelling techniques to accurately predict IAQ levels. Future research should strive to incorporate a broader range of indoor variables into IAQ models to enhance their predictive capabilities and facilitate more effective strategies for managing indoor air quality.

In addition, many studies fail to take into account the impact of outdoor factors on indoor air quality. The quality of outdoor air can significantly affect IAQ, particularly in environments with poor ventilation or high levels of outdoor pollutants. Variables such as outdoor temperature, humidity, and pollution levels can directly impact the indoor air quality and must be considered when creating IAQ prediction models. If researchers overlook the inclusion of outdoor variables [55], they run the risk of developing models that do not accurately reflect real-world IAQ conditions. Moreover, the interaction between indoor and outdoor air can create complex dynamics that affect IAQ. Therefore, future studies should incorporate outdoor factors into IAQ models to achieve a more comprehensive understanding of the factors that affect indoor air quality.

Addressing the gaps in air quality prediction studies is crucial for advancing the understanding of indoor and outdoor air quality dynamics, as well as improving air quality management strategies. To achieve this, it is important to incorporate continuous occupancy data, consider key indoor variables, and account for outdoor variables. These measures will lead to more accurate and effective air quality prediction models. Taking a holistic approach to air quality research will enable researchers to gain insights that can contribute to the creation of healthier indoor and outdoor environments for occupants. Additionally, integrating these factors into air quality models can also support the development of smarter building management systems that optimize ventilation and air purification processes based on real-time IAQ data. Ultimately, these advancements will significantly improve occupant health, comfort, and overall well-being in both indoor and outdoor environments.

V. CONCLUSION AND FUTURE RESEARCH DIRECTION

In conclusion, this systematic literature review highlights the various approaches to air quality prediction using ML. Each approach is characterized by specific models and associated evaluation metrics. This study examined and analyzed 26 articles published in Scopus and Science Direct databases. The integration of classification models, deep learning techniques, regression models, and ensemble learning strategies offers a wide range of options for researchers and practitioners to choose from, depending on their specific air quality prediction needs. Among these approaches, deep learning techniques often stand out for their ability to handle complex datasets and capture intricate patterns. The careful consideration of evaluation metrics in each category ensures a thorough understanding of model performance and contributes to the ongoing development of effective air quality prediction methods. This rigorous evaluation not only helps identify the most suitable models for different prediction tasks but also drives advancements in methodologies to improve accuracy and reliability.

However, this literature review has certain limitations. Firstly, it only considered articles from two databases, which may not provide a comprehensive overview of the extensive literature on ML, evaluation metrics, and air quality. To overcome this limitation, future researchers should consider examining more than two databases to capture all the crucial advancements in these areas. Another limitation is the scarcity of studies that comprehensively address both indoor and outdoor air quality in a single paper. This restricts our understanding of how these environments interact and influence each other. To address this limitation, future research should aim to integrate both indoor and outdoor air quality data within a single framework, allowing for a holistic analysis of how environmental factors across different settings impact air quality. This study has the potential to make a significant impact in the field of air quality. It provides valuable insights and can guide the development of more effective strategies for managing air quality in diverse environments. Further exploration in this area could lead to enhanced policies and technologies for improving air quality on a broader scale.

ACKNOWLEDGMENT

This research is funded by the Malaysian Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS), grant number 2022-0092-104-02 (FRGS/1/2022/TK08/UPSI/02/1).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest related to the research, authorship or publication of this paper. All authors have contributed to the work without any financial, personal or professional interests that could be perceived as influencing the results or interpretation of the manuscript.

AUTHOR CONTRIBUTION

We, the undersigned authors, have made the following contributions to the systematic literature review presented in this manuscript:

Amir Hamzah and Sumayyah Dzulkifly conceived the presented idea and wrote the manuscript, with support from Wang Shir Li and Che Zalina Zulkifli. Sumayyah Dzulkifly, Wang Shir Li and Che Zalina Zulkifli also helped supervise the manuscript. All authors discussed the results and contributed to the final version of the manuscript. Amir Hamzah and Sumayyah Dzulkifly also played a key role in designing the research methodology and analyzing the data. The collaborative effort among all authors ensured a comprehensive and robust study, leading to meaningful contributions to the field of air quality research.

REFERENCES

[1] A. A. Almetwally, M. Bin-Jumah, and A. A. Allam, "Ambient air pollution and its influence on human health and welfare: an overview," Environmental Science and Pollution Research, vol. 27, no. 20, pp. 24815-24830, 2020.

[2] I. Tanasa, M. Cazacu, and B. Sluser, "Air quality integrated assessment: environmental impacts, risks and human health hazards," Applied Sciences (Switzerland), vol. 13, no. 2, pp. 1222, 2023.

[3] S. Karimi, M. Asghari, R. Rabie, and M. E. Niri, "Machine learning-based white-box prediction and correlation analysis of air pollutants in proximity to industrial zones," Process Safety and Environmental Protection, vol. 178, pp. 1009–1025, 2023.

[4] S. A. Sarkodie, P. A. Owusu, and T. Leirvik, "Global effect of urban sprawl, industrialization, trade and economic development on carbon dioxide emissions," Environmental Research Letters, vol. 15, no. 3, 2020.

[5] Y. Hashmy, Z. U. Khan, F. Ilyas, R. Hafiz, U. Younis, and T. Tauqeer, "Modular air quality calibration and forecasting method for low-cost sensor nodes," IEEE Sens J, vol. 23, no. 4, pp. 4193–4203, 2023.

[6] X. Dai, W. Shang, J. Liu, M. Xue, and C. Wang, "Achieving better indoor air quality with IoT systems for future buildings: opportunities and challenges," Science of the Total Environment, vol. 895, 2023.

[7] S. Chaudhuri and M. Roy, "Global ambient air quality monitoring: can mosses help? a systematic meta-analysis of literature about passive moss biomonitoring," Environment, Development and Sustainability, vol. 26, no. 3, pp. 5735-5773, 2024.

[8] I. Essamlali, H. Nhaila, and M. E. Khaili, "Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: a systematic review," Sustainability (Switzerland), vol. 16, no. 3, 2024.

[9] E. M. Jovanovska, V. Batz, P. Lameski, E. Zdravevski, M. A. Herzog, and V. Trajkovik, "Methods for urban air pollution measurement and forecasting: challenges, opportunities, and solutions," Atmosphere, vol. 14, no. 9, 2023.

[10] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," BMJ (Online), vol. 339, no. 7716, pp. 332-336, 2009.

[11] A. Tella, A. L. Balogun, N. Adebisi, and S. Abdullah, "Spatial assessment of PM10 hotspots using random forest, k-nearest neighbour and naïve bayes," Atmos. Pollut. Res., vol. 12, no. 10, 2021.

[12] S. H. Alhathloul, A. K. Mishra, and A. A. Khan, "Low visibility event prediction using random forest and K-nearest neighbor methods," Theor. Appl. Climatol., 155, pp. 1289-1300, 2024.

[13] C. H. Cheng and M. C. Tsai, "An intelligent time series model based on hybrid methodology for forecasting concentrations of significant air pollutants," Atmosphere (Basel), vol. 13, no. 7, 2022.

[14] S. Mahmud, T. B. I. Ridi, M. S. Miah, F. Sarower, and S. Elahee, "Implementing machine learning algorithms to predict particulate matter (PM2.5): a case study in the Paso del Norte region," Atmosphere (Basel), vol. 13, no. 12, 2022.

[15] J. K. Sethi and M. Mittal, "Ambient air quality estimation using supervised learning techniques," EAI Endorsed Transactions on Scalable Information Systems, vol. 6, no. 22, pp. 1–10, 2019.

[16] S. Nayak, S. B. Majumder, K. P. Thiruppathi, and S. Misra, "LiFe5O8-based gas sensor: a machine learning approach for selective VOCs detection," Mater. Chem. Phys., vol. 312, 2024.

[17] A. G. M. Mengara, E. Park, J. Jang, and Y. Yoo, "Attention-based distributed deep learning model for air quality forecasting," Sustainability (Switzerland), vol. 14, no. 6, 2022.

[18] K. Zhang, X. Yang, H. Cao, J. Thé, Z. Tan, and H. Yu, "Multi-step forecast of PM2.5 and PM10 concentrations using convolutional neural network integrated with spatial–temporal attention and residual learning," Environ. Int., vol. 171, 2023.

[19] M. Gabriel and T. Auer, "LSTM deep learning models for virtual sensing of indoor air pollutants: a feasible alternative to physical sensors," Buildings, vol. 13, no. 7, 2023.

[20] S. Abirami and P. Chitra, "Regional air quality forecasting using spatiotemporal deep learning," J. Clean. Prod., vol. 283, 2021.

[21] Y. Zeng, J. Chen, N. Jin, X. Jin, and Y. Du, "Air quality forecasting with hybrid LSTM and extended stationary wavelet transform," Build. Environ., vol. 213, 2022.

[22] E. Yuan and G. Yang, "SA–EMD–LSTM: A novel hybrid method for long-term prediction of classroom PM2.5 concentration," Expert. Syst. Appl., vol. 230, 2023.

[23] I. I. Prado-Rujas, A. García-Dopico, E. Serrano, M. L. Córdoba, and M. S. Pérez, "A multivariable sensor-agnostic framework for spatio-temporal air quality forecasting based on deep learning," Eng. Appl. Artif. Intell., vol. 127, 2024.

[24] N. Sarkar, R. Gupta, P. K. Keserwani, and M. C. Govil, "Air quality index prediction using an effective hybrid deep learning model," Environmental Pollution, vol. 315, 2022.

[25] P. R. Gokul, A. Mathew, A. Bhosale, and A. T. Nair, "Spatio-temporal air quality analysis and PM2.5 prediction over Hyderabad City, India using artificial intelligence techniques," Ecol. Inform., vol. 76, 2023.

[26] X. Sun, Z. Tian, and Z. Zhang, "A new decomposition-integrated air quality index prediction model," Earth Sci. Inform., vol. 16, no. 3, pp. 2307–2321, 2023.

[27] N. Zaini, A. N. Ahmed, L. W. Ean, M. F. Chow, and M. A. Malek, "Forecasting of fine particulate matter based on LSTM and optimization algorithm," J. Clean. Prod., vol. 427, 2023.

[28] X. Wang, M. Wang, X. Liu, Y. Mao, Y. Chen, and S. Dai, "Surveillance-image-based outdoor air quality monitoring," Environmental Science and Ecotechnology, vol. 18, 2024.

[29] M. H. Kim and H. J. Park, "Application of artificial neural networks using sequential prediction approach in indoor airflow prediction," Journal of Building Engineering, vol. 69, 2023.

[30] C. A. Pozo-Luyo, J. M. Cruz-Duarte, I. Amaya, and J. C. Ortiz-Bayliss, "Forecasting PM2.5 concentration levels using shallow machine

learning models on the Monterrey Metropolitan Area in Mexico," Atmos. Pollut. Res., vol. 14, no. 11, 2023.
[31] Q. Guo, Z. He, and Z. Wang, "Simulating daily PM2.5 concentrations using wavelet analysis and artificial neural network with remote sensing and surface observation data," Chemosphere, vol. 340, 2023.
[32] J. W. Koo, S. W. Wong, G. Selvachandran, H. V. Long, and L. H. Son, "Prediction of air pollution index in Kuala Lumpur using fuzzy time series and statistical models," Air Qual. Atmos. Health, vol. 13, no. 1, pp. 77–88, 2020.
[33] X. Ma, T. Chen, R. Ge, C. Cui, F. Xu, and Q. Lv, "Time series-based PM2.5 concentration prediction in Jing-Jin-Ji area using machine learning algorithm models," Heliyon, vol. 8, no. 9, 2022.
[34] N. H. Van, P. Van Thanh, D. N. Tran, and D. T. Tran, "A new model of air quality prediction using lightweight machine learning," International Journal of Environmental Science and Technology, vol. 20, no. 3, pp. 2983–2994, 2023.
[35] J. Wang, D. Wang, F. Zhang, C. K. Yoo, and H. Liu, "Soft sensor for predicting indoor PM2.5 concentration in subway with adaptive boosting deep learning model," J. Hazard. Mater., vol. 465, 2024.
[36] R. Ameri, C. C. Hsu, S. S. Band, M. Zamani, C. M. Shu, and S. Khorsandroo, "Forecasting PM 2.5 concentration based on integrating of CEEMDAN decomposition method with SVM and LSTM," Ecotoxicol. Environ. Saf., vol. 266, 2023.
[37] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electronic Markets, vol. 31, no. 31, 2021.
[38] Z. Zhang, S. Zhang, C. Chen, and J. Yuan, "A systematic survey of air quality prediction based on deep learning," Alexandria Engineering Journal, vol. 93, pp. 128-141, 2024.
[39] Y. Wei, J. Jang-Jaccard, W. Xu, F. Sabrina, S. Camtepe, and M. Boulic, "LSTM-autoencoder-based anomaly detection for indoor air quality time-series data," IEEE Sens. J., vol. 23, no. 4, pp. 3787–3800, 2023.
[40] X. Yao, D. Wang, T. Yu, C. Luan, and J. Fu, "A machining feature recognition approach based on hierarchical neural network for multi-feature point cloud models," J. Intell. Manuf., vol. 34, no. 6, pp. 2599–2610, 2023.
[41] J. K. Mohammedamin and Y. A. Shekha, "Indoor sulfur dioxide prediction through air quality modeling and assessment of sulfur dioxide and nitrogen dioxide levels in industrial and non-industrial areas," Environ. Monit. Assess., vol. 196, no. 5, 2024.
[42] S. I. Abba et al., "Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration," IEEE Access, vol. 8, pp. 157218–157237, 2020.
[43] L. Huang et al., "Recommendations on benchmarks for numerical air quality model applications in China - part 1: PM2.5 and chemical species," Atmos. Chem. Phys., vol. 21, no. 4, pp. 2725–2743, 2021.
[44] M. Grandini, E. Bagli, and G. Visani, "Metrics for multi-class classification: an overview," Accessed: 10 Aug 2024, [Online]. Available: http://arxiv.org/abs/2008.05756.
[45] M. M. Taye, "Understanding of machine learning with deep learning: architectures, workflow, applications and future directions," Computers, vol. 12, no. 5, 2023.
[46] C. Zuo et al., "Deep learning in optical metrology: a review," Light: Science and Applications, vol. 11, no. 1, 2022.
[47] A. Mousaei and Y. Naderi, "Machine learning-based regression models for state of charge estimation in hybrid electric vehicles: a review," Preprints, 2023. https://doi.org/10.20944/preprints202312.1938.v1.
[48] Y. Wu, S. Lin, K. Shi, Z. Ye, and Y. Fang, "Seasonal prediction of daily PM2.5 concentrations with interpretable machine learning: a case study of Beijing, China," Environmental Science and Pollution Research, vol. 29, no. 30, pp. 45821–45836, 2022.
[49] S. González, S. García, J. Del Ser, L. Rokach, and F. Herrera, "A practical tutorial on bagging and boosting based ensembles for machine learning: algorithms, software tools, performance study, practical perspectives and opportunities," Information Fusion, vol. 64, pp. 205–237, 2020.
[50] A. Gangwar, S. Singh, R. Mishra, and S. Prakash, "The state-of-the-art in air pollution monitoring and forecasting systems using IoT, big data, and machine learning," Wireless Personal Communications, vol. 130, no. 3, pp. 1699-1729, 2023.
[51] Y. K. Kim, Y. Abdou, A. Abdou, and H. Altan, "Indoor environmental quality assessment and occupant satisfaction: a post-occupancy evaluation of a UAE university office building," Buildings, vol. 12, no. 7, 2022.
[52] J. Kim, J. Il Bang, A. Choi, H. J. Moon, and M. Sung, "Estimation of occupancy using IoT sensors and a carbon dioxide-based machine learning model with ventilation system and differential pressure data," Sensors, vol. 23, no. 2, 2023.
[53] Y. Pan et al., "Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies," Advances in Applied Energy, vol. 10, 2023.
[54] Z. Wu, C. Ma, X. Shi, L. Wu, Y. Dong, and M. Stojmenovic, "Imputing missing indoor air quality data with inverse mapping generative adversarial network," Build. Environ., vol. 215, 2022.
[55] T. Ahmed, P. Kumar, and L. Mottet, "Natural ventilation in warm climates: the challenges of thermal comfort, heatwave resilience and indoor air quality," Renewable and Sustainable Energy Reviews, vol. 138, 2021.