Deep Learning-Driven Visual Analytics Framework for Real-Time Data Interpretation and Predictive Decision Making

[1]Mr Yograj Sharma*, [2]Ankur Tharwan, [3]Dr. Jagdish Makhijani and [4]Mr. S. Ayyappan

[1]Assistant Professor, Rustamji Institute of Technology, Gwalior, Madhya Pradesh-475001

Email: yograjgwl@gmail.com/ ORCID Id: 0009-0009-0829-238X

[2]Founder, EMS Academy, 82/121 Sector 8 Pratap Nagar Sanaganer Jaipur, Rajasthan-302033

Email: ankuremsacademy@gmail.com/ ORCID Id: 0009-0009-4006-5267

[3]Assistant Professor, Rustamji Institute of Technology, BSF Academy, Gwalior, Madhya Pradesh-475001

Email: j_makhijani@yahoo.com/ ORCID Id: 0000-0002-3601-3319,

[4]Assistant Professor, Department of Electronics and Communication Engineering, NPR College of Engineering and Technology, Natham, Dindigul, Tamil Nadu-624401

Email: ayps23@gmail.com/ ORCID: 0000-0001-9240-0244

*Corresponding Mail: yograjgwl@gmail.com

Abstract:

The rapid growth of real-time data generated from environmental sensors, social media platforms, and IoT-enabled systems has created significant challenges in extracting actionable insights for timely decision-making. This study introduces a novel approach called Deep Learning-Driven Visual Analytics Framework for Real-Time Data Interpretation and Predictive Decision Making that combines multiple data sources to enhance the precision of data interpretation and decision-making. The framework combines the Air Quality Dataset and Sentiment140 Dataset, which are used to merge environmental measurements with sentiment information from the public. A complete pre-processing pipeline is used, which includes missing value imputation, normalization, cleaning text, tokenisation, lemmatisation and creation of BERT embeddings. Relevant features are extracted from both sets of data, and combined into a single multimodal representation. Then a hybrid CNN-BiLSTM network is applied to capture spatial and temporal dependences for accurately prediction and classification. The experimental results show that the model has a better prediction performance for air quality, with an MAE of 0.041, RMSE of 0.072 and R^2 score of 0.97, and a more accurate results for sentiment classification with 98.83% accuracy and an F1 score of 98.17%. The proposed framework improves the real-time monitoring, predictive intelligence, and visual decision making for various environmental, healthcare, industrial and smart city applications.

Keywords: DL; Visual Analytics; Real-Time Data Interpretation; Predictive Decision Making; CNN-BiLSTM

1. INTRODUCTION

The exponential growth of digital data generated from diverse sources such as Internet of Things (IoT) devices, social media platforms, sensor networks, healthcare systems, financial transactions, and industrial automation has created significant challenges in extracting meaningful insights for timely decision-making [1]. In recent years, data-driven approaches have become more essential to organizations in various domains, enabling them to optimize operations, strengthen forecasting and aid in intelligent decision processes [2]. However, the massive volume, velocity, variety, and complexity of modern datasets often exceed the capabilities of traditional analytical methods, making real-time interpretation and actionable insight generation a critical challenge [3].

Visual analytics is a successful interdisciplinary area that integrates data visualization, human computer interaction and analytical reasoning for exploring and understanding complex datasets [4]. Visual Analytics transforms large volumes of data into visual interactive representations, allowing patterns, correlations, anomalies, and trends to be detected that might not be apparent from the raw numerical information [5-8]. However, traditional visual analytics systems have a problem when dealing with constantly changing, real-time data streams and high-dimensional data, and are not effective in dynamic settings where fast decisions must be made [9].

More recently, recent breakthrough advances in the fields of Artificial Intelligence (AI) and Deep Learning (DL) have achieved unprecedented successes in sophisticated data processing and analysis applications [10]. Various DL models such as CNNs, RNNs, LSTMs, and various forms of Transformer architectures can autonomously discover features at various levels of abstraction, discover intricate patterns, and produce an accurate output prediction even for complex structured and unstructured data [11-14]. Without extensive human-driven features engineering, the learned relationships for complex models provide significant improvements in various tasks related to large scale data and real time analytics.

Deep learning combined with visual analytics can provide a promising way to address the drawbacks of traditional data interpretation systems [15]. The deep learning concept can be used to build visual analytics frameworks that can process incoming data streams automatically, extract features intelligently, detect critical events and present the results via interactive visual dashboards [16]. The integration not only results in more accurate analysis, but also facilitates human understanding through the visualization of prediction results in an intuitive way [17]. This synergic environment allows users to discover data insights while enjoying the benefits of high-performance predictive capabilities, made possible by a mix of machine intelligence and human cognitive abilities.

The need for real-time data interpretation has been growing in critical application areas like healthcare monitoring, smart city management, cybersecurity, financial risk assessment, industrial process control and transportation systems as well as environmental monitoring [18-20]. There, missed or misinformed decisions can lead to substantial financial losses, security issues or even endanger life [21]. So, intelligent systems that are capable of continuously analyzing streaming data and predicting timely events is essential for proactive decision making [22]. This goal can be achieved with deep learning visual analytics framework by providing intuitive visual interfaces for ongoing monitoring, anomaly detection, trend forecasting, and scenario analysis.

Despite significant progress in both deep learning and visualization technologies, several challenges remain [23]. The challenges encompassed in these are the ability to process high dimensional, heterogeneous data, interpretability of the models, reduction of computational complexity, realtime processing needs, and user-friendly visual interfaces for decision makers [24]. In addition, most of the current systems emphasize predictive modeling or visualization tasks without providing a complete framework for integrating predictive modeling and visualization tasks into a single decision support system [25].

To address these challenges, this paper proposes a Deep Learning-Driven Visual Analytics Framework for Real-Time Data Interpretation and Predictive Decision Making by integrating environmental sensor data and social media sentiment information. The Air Quality dataset consists of historical air quality data that are retrieved and preprocessed, and then subjected to feature extraction using statistical analysis and BERT-based text embeddings. The Air Quality dataset includes historical air quality data, which are retrieved, preprocessed and used for feature extraction through statistical analysis and BERT-based text embeddings. The extracted features are brought together in a multimodal representation, which is then fed into a hybrid CNN-BiLSTM network that considers both temporal and spatial patterns. The framework provides real-time prediction, sentiment classification and trend forecasting features, and presents the results using visual analytics. Here are the research objective of the study follows as:

- To develop a Deep Learning-Driven Visual Analytics Framework that integrates environmental sensor data and social media sentiment data for real-time data interpretation and predictive decision-making.
- To collect and preprocess multimodal datasets, including the Air Quality Dataset and Sentiment140 Dataset, using data cleaning, normalization, feature engineering, and natural language processing techniques.
- To design and implement a hybrid CNN-BiLSTM model capable of learning spatial and temporal dependencies for accurate prediction, classification, and trend forecasting.
- To evaluate the performance of the proposed framework using regression and classification metrics such as MAE, RMSE, R^2 Score, Accuracy, Precision, Recall, and F1-Score.
- To compare the proposed CNN-BiLSTM framework with existing ML and DL models to demonstrate its effectiveness and predictive superiority.

2. RELATED WORK

In recent times, the role of artificial intelligence, machine learning and deep learning in environmental monitoring, air quality prediction and risk assessment has been proven. To predict the environmental health risks in real-time, Edeh et al. (2026) [26] created a machine learning-based approach that combines environmental, occupational, meteorological and demographic information, with high predictive power of the XGBoost model ($R^2 = 0.92$) and classification accuracy > 93%. Likewise, Ullah et al. (2026) [27] developed an AI-based expert system for real-time monitoring and forecasting of air quality based on Random Forest and Artificial Neural Networks with a deep reinforcement learning recommender system for adaptive environmental management, which achieved forecasting accuracies of 80–85%. Alam et al. (2026) [28] studied the machine learning methods for forecasting Air Quality Index (AQI) and found that Random Forest model has the highest R^2 of 0.88 when used with Air Quality Index (AQI) data from India. Moreover, Taştan et al. (2026) [29] proposed an optimized kNN model to improve the accuracy of calibration of low-cost air quality sensors using AQ-MultiCal, a no-code calibration platform. Overall, these studies underscore the potential of AI-powered predictive analytics to improve environmental monitoring precision and decision-making powers.

Table 1: Comparison of previous study

Authors & Year	Methodology / Techniques	Dataset / Application	Key Findings
Edeh et al. (2026)	Random Forest, Gradient Boosting, XGBoost, LSTM, SHAP-based XAI	Environmental, occupational, meteorological, and demographic data (43,200 records)	XGBoost achieved $R^2 = 0.92$; risk classification accuracy exceeded 93%; effective real-time environmental health risk prediction
Ullah et al. (2026)	Random Forest Regressor, ANN, Deep Reinforcement Learning (DRL)	Real-time air quality and meteorological data	Achieved 80–85% forecasting accuracy; generated adaptive alerts and nature-based recommendations
Alam et al. (2026)	Random Forest, DNN, ARIMA, SARIMA	Air Quality Index (AQI) prediction for India and Bangladesh	Random Forest achieved $R^2 = 0.88$ for India and outperformed all other models
Taştan et al. (2026)	kNN, ML-based sensor calibration platform (AQ-MultiCal)	Low-cost air quality sensor calibration	Optimized kNN achieved $R^2 = 0.990$; improved accessibility through no-code interface
Sammy et al. (2025)	CNNs (VGG16, GoogleNet, ResNet50), PCA, PSO Optimization	MODIS, Landsat-8, Sentinel-2 remote sensing datasets	ResNet50 achieved up to 95.2% accuracy with high F1-score and AUC
Ali et al. (2025)	Autoencoder, Random Forest, GBM, LSTM, Visual Analytics Dashboard	Customer behavior and marketing datasets	Autoencoder achieved ARI = 0.91; LSTM achieved $R^2 = 0.95$; improved marketing efficiency
Yang et al. (2025)	Hybrid CNN-LSTM, PCA	Financial data from CSMAR database (54,389 observations)	CNN-LSTM outperformed ARIMA, RF, XGBoost, and LSTM with MSE = 0.020
Adelusi et al. (2025)	Random Forest, SVM, Gradient Boosting, XAI	EHR, wearable sensor data, socioeconomic variables	Achieved AUC > 0.90 for cardiovascular disease prediction
Satish et al. (2025)	Random Forest, Logistic Regression	Tsunami prediction dataset (1995–2023)	Random Forest achieved 90% accuracy and 88% precision
Arun et al. (2025)	CNN, RNN, Reinforcement Learning	Smart building sensor networks	Improved energy efficiency, security, and sustainability metrics above 94%

In addition to its environmental applications, deep-learning-based predictive systems have found success in many other areas, showcasing the versatility of intelligent visual analytics systems. Few studies have applied deep learning for environmental monitoring using multispectral remote sensing data, and Sammy et al. (2025) [30] report classification results above 95% using ResNet50. Ali et al. (2025) [31] combined customer segmentation using an autoencoder with forecasting using LSTM and visualizing the dashboards, achieving an Adjusted Rand Index of 0.91 and predictive R^2 of 0.95. In financial forecasting, Yang et al. (2025) [32] proposed a hybrid CNN-LSTM model, which achieved superior performance compared with traditional machine learning models on financial forecasting and decision support systems. Adelusi et al. (2025) [33] proposed a machine learning-based cardiovascular disease prediction system in healthcare that combines EHRs and wearable sensor data with an explainable AI mechanism for transparency, yielding an AUC above 0.90. Likewise, Satish et al. (2025) [34] used Random Forest and Logistic Regression models for tsunami occurrence forecast which achieved 90% accuracy in prediction. A deep learning-based framework was suggested by Arun et al. (2025) [35] to enhance the energy efficiency, security, and sustainability indicators of smart buildings, achieving more than 94% improvement. While these studies have shown the effectiveness of multimodal data fusion, deep learning prediction and interactive visual analytics, there is still a lack of research that combines them in a single system for real-time data interpretation and predictive decision-making.The present study fills the gap in research that has not been explored extensively so far, namely by combining multimodal data fusion, deep learning prediction and interactive visual analytics within a single framework for real-time data interpretation and predictive decision-making.

3. RESEARCH METHODOLOGY

3.1 Dataset Collection

For the proposed Deep Learning-Driven Visual Analytics Framework for Real-Time Data Interpretation and Predictive Decision Making, researchers can use the following two publicly available datasets such as:

- **Air Quality Dataset**

The dataset contains 9357 instances of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device [36]. The device was located on the field in a significantly polluted area, at road level,within an Italian city. Data were recorded from March 2004 to February 2005 (one year) representing the longest freely available recordings of on field deployed air quality chemical sensor devices responses. The attributes in the data are data, month, year, $PM_{2.5}$, PM_{10}, NO, NO_2, NO_x, NH_3, CO, SO_2, O_3, Benzene, Toluene, AQI, and AQI_Bucket.

- **Sentiment140 dataset**

It contains 1,600,000 tweets extracted using the twitter api [37]. The tweets are categorized with annotations such as '0 = negative,' '2 = neutral,' and '4 = positive,' making them suitable for sentiment detection. Refer to Table 2 for specific details regarding the dataset.

Table 2: Description of Twitter Sentiment Dataset Attributes

S. No	Information	Description
1	Target	Tweet sentiment label (0 = negative, 2 = neutral, 4 = positive)
2	Ids	Tweet ID
3	Date	Tweet date
4	Flag	Query
5	User	User that tweeted
6	Text	Tweet text

The Data Acquisition Layer pulls and merges data from two sources which are heterogeneous: the Air Quality Dataset and the Sentiment140 Dataset. The Air Quality Dataset contains environmental data like CO concentration, benzene concentration, temperature and humidity gathered by sensor networks. The Sentiment140 Dataset consists of tweets labeled with sentiment, time of tweet and user-generated text. Both the sources are loaded into a centralized repository and synchronized using a temporal attribute. This integration allows the framework to collect real-world environmental data in addition to public sentiment trends, thereby enabling the comprehensive multimodal data to be analyzed, predicted, and displayed in real-time.

3.2 Data preprocessing

The Data Preprocessing Layer is responsible for improving data quality and transforming raw data into a structured format suitable for deep learning analysis. Since the Air Quality Dataset and Sentiment140 Dataset contain different types of data, separate preprocessing procedures are applied to each dataset.

- **Preprocessing of Air Quality Dataset**

The Air Quality Dataset includes measurements of environmental sensors, including CO, $PM_{2.5}$, PM_{10}, SO_2, and NO_2. There are several preprocessing steps in the dataset to increase data quality and reliability prior to analysis. The first step is to detect and fill the missing sensor readings with mean imputation methods to complete the data and ensure it is complete. Abnormal environmental conditions are detected by the Z-Score method, and sensor malfunction is detected and removed. Smoothing and filtering are used to reduce noise in the sensor signals. All the numerical features are then scaled using Min-Max scaling so that they are within the same range and do not penalize the model in training. Lastly, the data records are sorted by their timestamps to maintain the temporal relationship and prepare the data for time-series forecasting and predictive analytics.

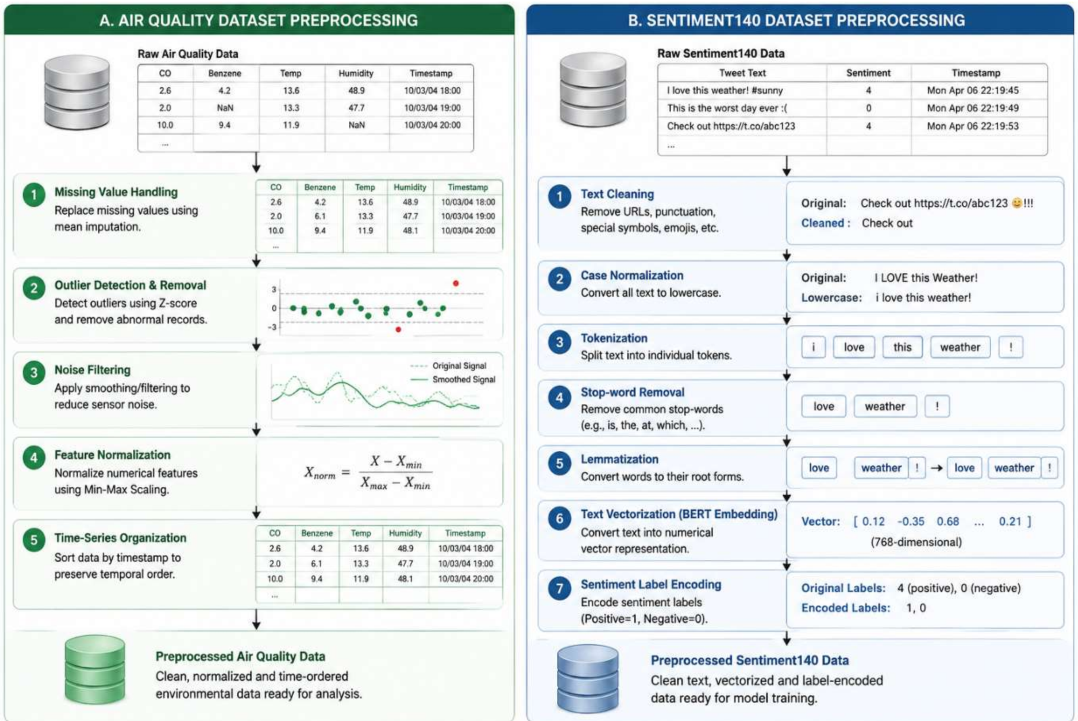


Figure 1: Data preprocessing steps on both datasets

- **Preprocessing of Sentiment140 Dataset**

The Sentiment140 Dataset is a collection of tweets featuring text and sentiment labels originated by users. Several Natural Language Processing (NLP) processing methods are used to preprocess the raw social media data, which are often noisy and unstructured. The tweets undergo removal of irrelevant characters, special symbols, emoji, punctuation and URL from the beginning. It is then transcribed to lowercase for uniformity throughout the data set. The tweets are tokenized to split them up into words and then remove frequently occurring words that have little semantic meaning through the stop word removal process. This is followed by lemmatization, which is used to convert words into their stem forms to eliminate word redundancy and consistency in the text. The cleaned text is then embedded using BERT to represent it in numbers, thus including the semantic and contextual information in the tweets. Last but not least, sentiment labels are translated into numerical values, whereafter the deep learning model for sentiment classification and predictive decision-making can be efficiently trained.

3.3 Feature Extraction Layer

The Feature Extraction Layer is designed to discover and extract useful features from the preprocessed Air Quality and Sentiment140 datasets. This stage aims to convert the numerical and textual data into feature vectors that are representative of the data and can be effectively used by the deep learning model. The goal is to simplify the data and avoid losing important patterns that are crucial to prediction and decision-making.

- **Feature Extraction from Air Quality Dataset**

After the preprocessing step, environmental sensor readings are evaluated in order to extract statistical, temporal, and domain-related features. Statistical features like mean, variance, standard deviation, minimum, and maximum values are extracted to characterize the behavior of the air quality variables. The temporal features like moving average, trend, and seasonality are extracted from timestamped data in order to evaluate the change in environmental conditions through time. Furthermore, important environmental features such as CO, $PM_{2.5}$, PM_{10}, SO_2, and NO_2 are selected as main predictor variables. These features provide valuable insights into air quality fluctuations and pollution trends.

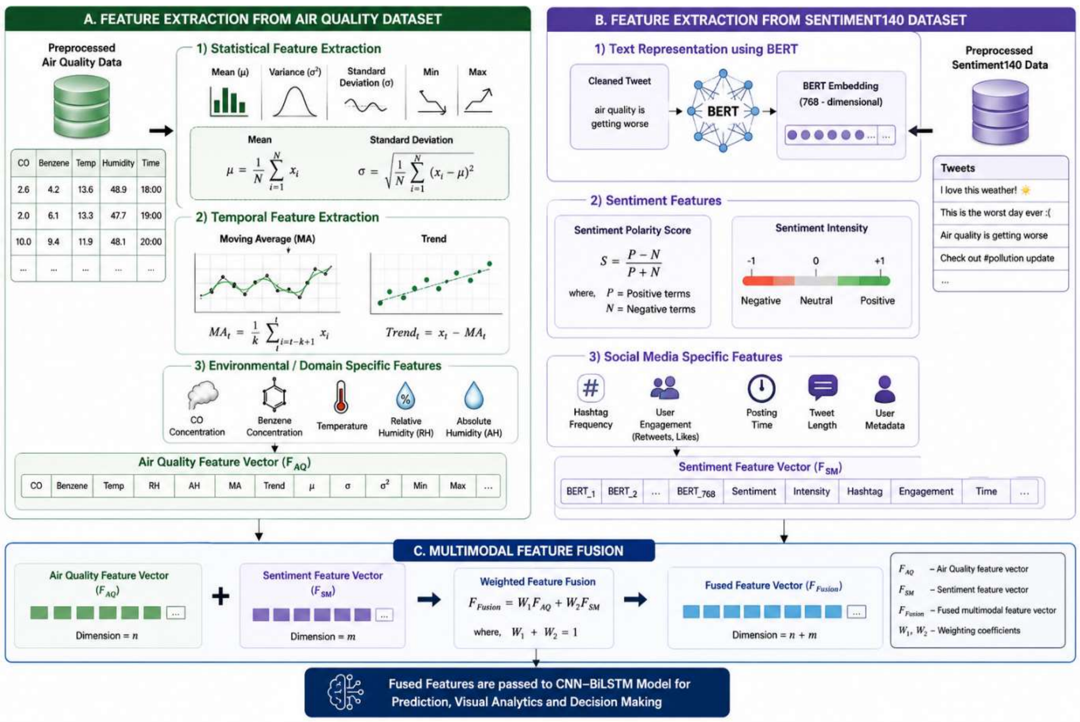


Figure 2: Feature extraction steps on both datasets

- **Feature Extraction from Sentiment140 Dataset**

The Sentiment140 dataset, on the other hand, is used for feature extraction to extract the semantic and contextual information contained in social media posts. Text cleaning and preprocessing is then followed by using BERT (Bidirectional Encoder Representations from Transformers) to create high dimensional contextual representations for each tweet. These embeddings represent the semantic meaning, sentiment and language relations within the text. Further attributes like sentiment polarity, sentiment intensity, hashtag frequency, posting time and user engagement information are captured to support the textual representation. These features help them interpret the sentiments of the public and identify trends in social media feeds.

- **Multimodal Feature Fusion**

The extracted environmental and social media features are combined to create a unified multimodal feature representation. Feature fusion integrates physical environmental observations with human-generated sentiment information, allowing the framework to capture complex relationships between environmental conditions and public reactions. The fusion equation can be written as:

$$F_{\text{Fusion}} = F_{AQ} + F_{SM}$$

where F_{AQ} denotes the Air Quality feature vector and F_{SM} represents the Sentiment140 feature vector. The resulting multimodal feature set is forwarded to the CNN-BiLSTM prediction model for real-time forecasting, visual analytics, and predictive decision-making.

3.4 Deep Learning Prediction Module

The Deep Learning Prediction Module is the analytical module of the proposed Deep Learning-Driven Visual Analytics Framework, which could accurately predict in real time by learning multimodal data [38]. This module uses a combination of Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) network to extract both spatial and temporal features from the fused Air Quality (AQ) and Sentiment140 (S140) datasets [39]. The CNN part is used to learn discriminative spatial representations and hidden feature patterns in the fused feature space, and the BiLSTM models the sequential dependencies and long-range temporal relationships both forward and backward [40]. The fusion of these two architectures allows for the full learning of environmental variations and sentiment dynamics in the public, which improves the accuracy of predictions and decisions-making capacity [41].

Let the fused multimodal input feature matrix be represented as:

$$X = [F_{AQ}, F_{SM}]$$

where F_{AQ} denotes the Air Quality feature vector and F_{SM} represents the Sentiment140 feature vector.

Initially, the CNN layer performs feature extraction using a convolution operation:

$$h_i = f\left(\sum_{j=1}^{k} W_j X_{i+j-1} + b\right)$$

where W_j denotes the convolution kernel weights, b represents the bias term, $f(\cdot)$ is the activation function, and h_i is the resulting feature map [42].

The Rectified Linear Unit (ReLU) activation function is employed to introduce nonlinearity:

$$ReLU(x) = \max(0, x)$$

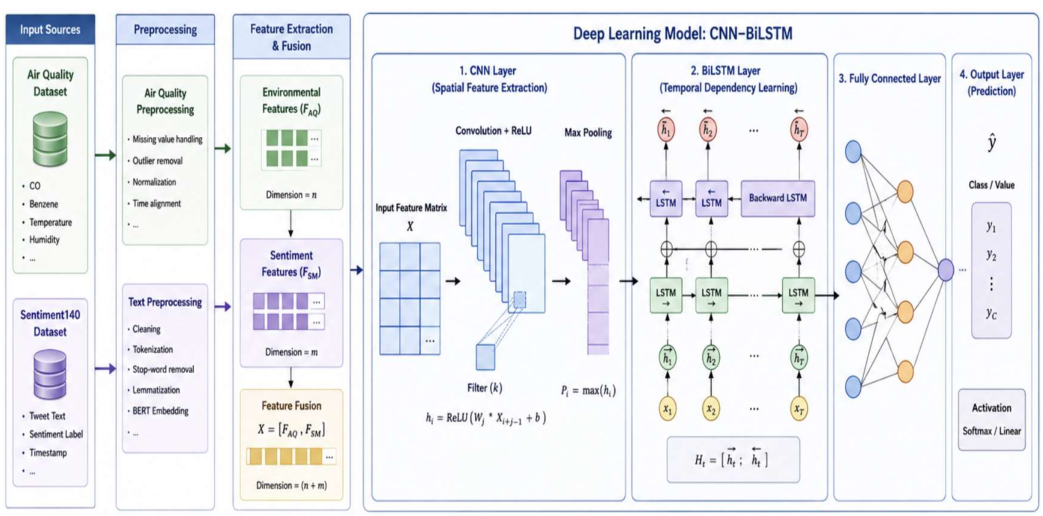


Figure 3: Architecture of proposed model

Subsequently, max-pooling is applied to reduce dimensionality and retain the most significant features:

$$P_i = \max(h_i)$$

The pooled feature representations are then supplied to the BiLSTM layer for temporal sequence modelling [43]. The hidden state of each LSTM cell is computed as:

$$h_t = o_t \odot \tanh(C_t)$$

where h_t is the hidden state, o_t is the output gate, C_t denotes the memory cell state, and $\odot$represents element-wise multiplication.

The forget gate is formulated as:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

where $\sigma(\cdot)$ denotes the sigmoid activation function [44].

The cell state is updated according to:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

where i_t is the input gate and $\tilde{C}_t$ denotes the candidate memory state.

To capture contextual information from both past and future observations, the forward and backward hidden states generated by the BiLSTM network are concatenated as:

$$H_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$$

The combined hidden representation is subsequently passed to a fully connected layer to generate prediction outputs:

$$Y = \sigma(WH_t + b)$$

where W represents trainable weights, b is the bias vector, and Y denotes the predicted output.

For multi-class classification tasks, the Softmax function converts the output scores into class probabilities:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

where $P(y_i)$ represents the probability associated with class i.

During model training, parameters are optimized using the Adam optimization algorithm by minimizing the cross-entropy loss function:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

where y_i denotes the true class label and $\hat{y}_i$ represents the predicted probability.

The hybrid CNN-BiLSTM model effectively learns complex nonlinear relationships between environmental and social sentiment indicators, generating highly discriminative feature representations [45]. These learned representations support real-time forecasting, anomaly detection, visual analytics, environmental risk assessment, and intelligent predictive decision-making within the proposed framework.

Table 3: Hyperparameter Configuration

Parameter	Value
Optimizer	Adam
Learning Rate	0.001
Epochs	50
Batch Size	64
Dropout	0.30
Activation Function	ReLU

3.5 Performance Evaluation

The effectiveness of the proposed Deep Learning-Driven Visual Analytics Framework is evaluated using a comprehensive set of classification and regression.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

where y_i represents the actual value and $\hat{y}_i$ denotes the predicted value.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

4. RESULT AND ANALYSIS

The proposed Deep Learning-Driven Visual Analytics Framework was evaluated using the Air Quality Dataset and Sentiment140 Dataset. A hybrid CNN-BiLSTM model was employed to learn spatial and temporal patterns from multimodal data sources. The experiments were conducted using a 75:25 train-test split and optimized using the Adam optimizer.

4.1 For Air Quality Dataset

- **Spearman correlation analysis Result**

To identify pollutant factors highly correlated with the AQI, this study employed Spearman's Rank Correlation Coefficient to analyze the relationships between individual pollutants and AQI. This method effectively measures the strength of monotonic associations between variables and is particularly suitable for datasets exhibiting non-normal distributions or nonlinear relationships. The analysis revealed that CO, $PM_{2.5}$, PM_{10}, SO_2, and NO_2 exhibited significant positive correlations with the AQI, indicating that these pollutants are the primary contributors to air quality deterioration. Consequently, these five variables were selected as input features for subsequent predictive modeling. The detailed results of the correlation analysis are presented in Figure 4.

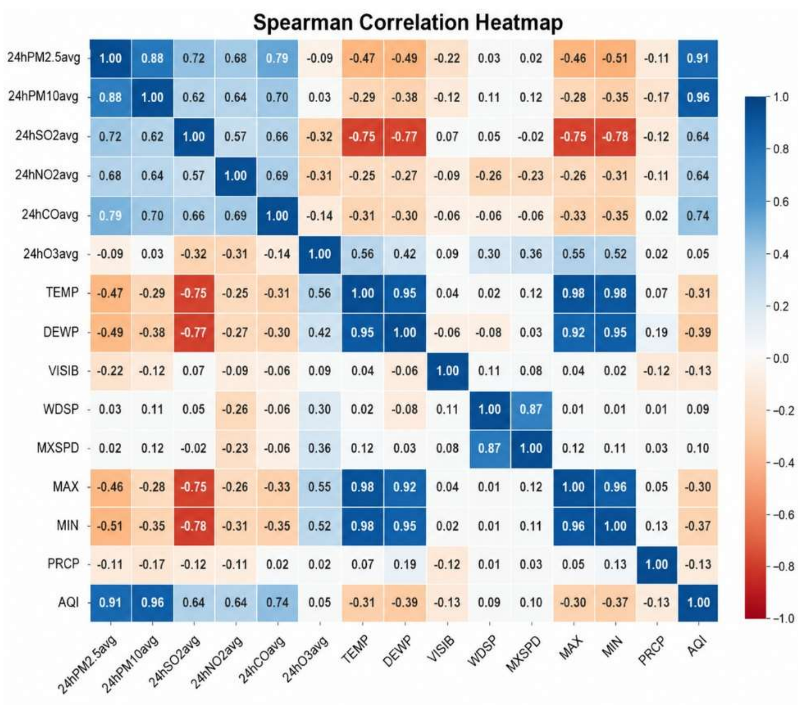


Figure 4: Spearman Correlation Heatmap

- **Data Distribution Result**

To comprehensively assess the effectiveness of data preprocessing, both visualization and statistical analysis were conducted for the five selected input features (CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2) and the target variable, AQI. First, boxplots were generated to intuitively illustrate the distribution patterns and potential outliers of each variable. These visualizations also facilitated the assessment of the concentration and dispersion of variables across different value ranges.

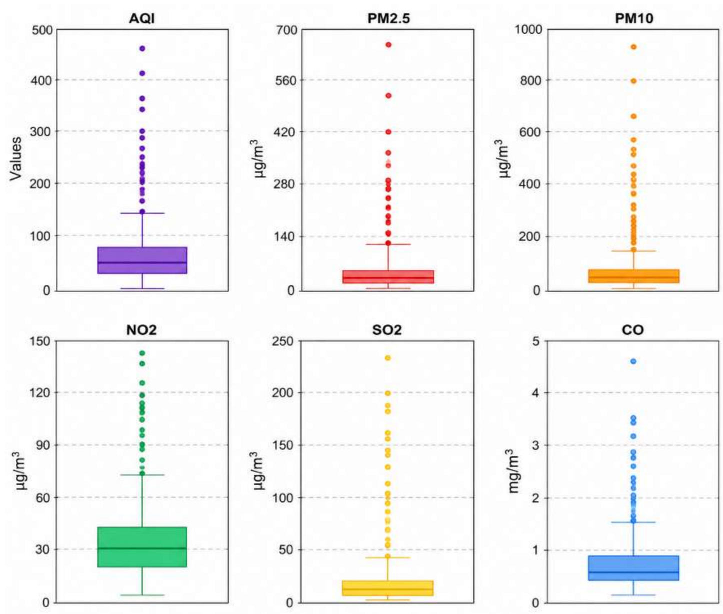


Figure 5: Box Plot Analysis

Furthermore, to ensure that missing values had been adequately imputed and to gain a deeper understanding of the overall data characteristics, a descriptive statistical summary was compiled. This summary included the total count, maximum, minimum, mean, median, standard deviation, first quartile (Q1), and third quartile (Q3) for each variable. The corresponding results are presented in Figure 3 and Table 4.

Table 4: Descriptive Statistics for Air Quality Data

Parameter	Count	Max	Min	Mean	Median	Std	25%	75%
AQI	3652	462	8	76.84	61	51.27	39	92
$PM_{2.5}$	3652	645	3	48.62	34	46.13	18	61
PM_{10}	3652	927	7	84.37	65	68.54	42	103
SO_2	3652	234	1	21.46	11	24.78	6	24
NO_2	3652	145	5	36.91	34	15.82	25	46
CO	3652	4.62	0.14	0.86	0.74	0.42	0.51	1.01

- **Evaluation Value of Proposed Model**

The predictive performance of the framework for environmental trend forecasting was tested with regression metrics. The regression performance of CNN, LSTM, Bi-LSTM and the proposed CNN-BiLSTM model are compared in terms of MAE, RMSE and R^2 score in Table X. The conventional CNN model resulted in an MAE of 0.118, RMSE of 0.164, and R^2 score of 0.82, which showed relatively weak prediction ability. The outcomes of the LSTM model were improved with the parameters of the MAE as 0.081, RMSE as 0.124, and R^2 score as 0.90. The Bi-LSTM model further improved the performance with an MAE of 0.064, an RMSE of 0.098, and an R^2 score of 0.94. CNN-BiLSTM was the best among all baseline models with the lowest MAE (0.041) and RMSE (0.072) and the highest R^2 (0.97), indicating its high performance in predicting the spatial and temporal dependencies in accurate prediction.

Table 5: Regression Analysis for Air Quality Prediction

Model	MAE	RMSE	R^2 Score
CNN	0.118	0.164	0.82
LSTM	0.081	0.124	0.90
Bi-LSTM	0.064	0.098	0.94
Proposed CNN-BiLSTM	0.041	0.072	0.97

- **Comparative Analysis**

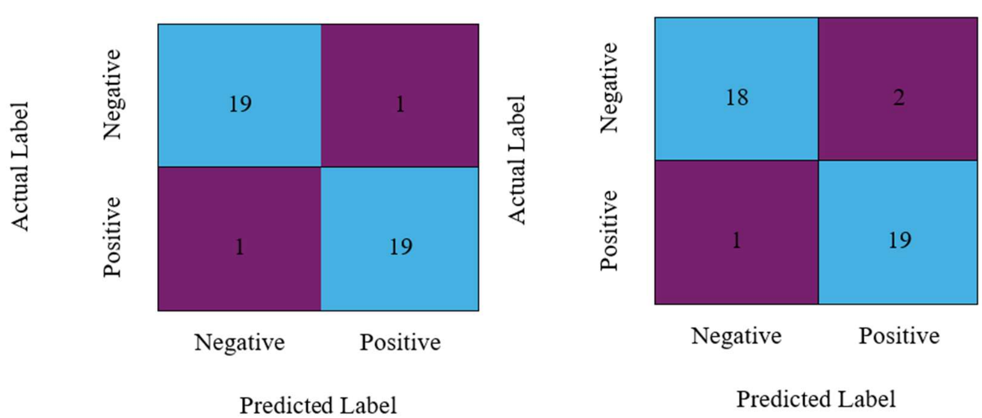
Table 6 shows the regression results of the proposed CNN-Bi-LSTM model with existing models. Artificial Neural Network (ANN) was used by Liang et al. (2020) with an MAE of 1.823, RMSE of 3.966, and R^2 score of 0.992. Natarajan et al. (2024) applied a Grey Wolf Optimization-based Decision Tree model with a MAE of 0.0735 and a RMSE of 0.1304 and an R^2 score of 0.98. Ravindiran et al. (2023) applied CatBoost, achieving an MAE of 0.60, RMSE of 0.76, and R^2 score of 0.99. Overall, the proposed CNN-Bi-LSTM model achieved the lowest prediction errors, with a MAE of 0.041 and a RMSE of 0.072, and a high value of R^2 (0.97), which showed that the model had a high level of predictive accuracy and robustness.

Table 6: Comparison analysis with previous model based on air quality dataset

Authors [Reference]	Methods	MAE	RMSE	R^2 Score
Liang et al., (2020) [46]	**ANN**	1.823	3.966	0.992
Natarajan et al., (2024) [47]	**Grey wolf optimization-DT**	0.0735	0.1304	0.98
Ravindiran et al., (2023) [48]	**Catboost**	0.60	0.76	0.99
Proposed Model	**CNN-Bi-LSTM**	**0.041**	**0.072**	**0.97**

4.2 For Sentiment140 dataset

The Sentiment140 dataset was used to evaluate sentiment prediction performance. Figure 6 shows the confusion metrics of proposed model.



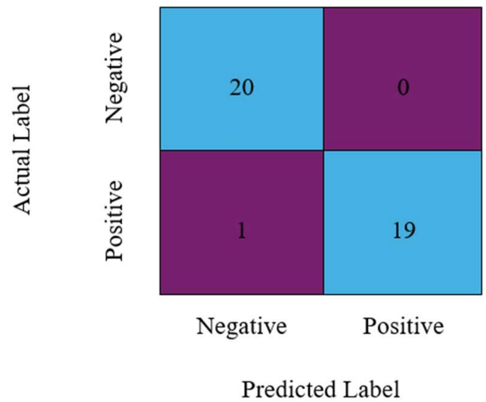


Figure 6: Confusion metrics of CNN, Bi-LSTM, and CNN-Bi-LSTM

- **Convolutional Neural network (CNN)**

The batch size was set to 32, the learning rate was set to 0.001, the optimizer was set to Ada-grade, and categorical cross-entropy was chosen as a loss function. They were able to achieve the best set of model training parameters. They trained and verified the CNN model for a total of 50 epochs, and the graphs that represent the model's accuracy are shown in Figure 7. There is a report on the categorization using CNN in Table 7.

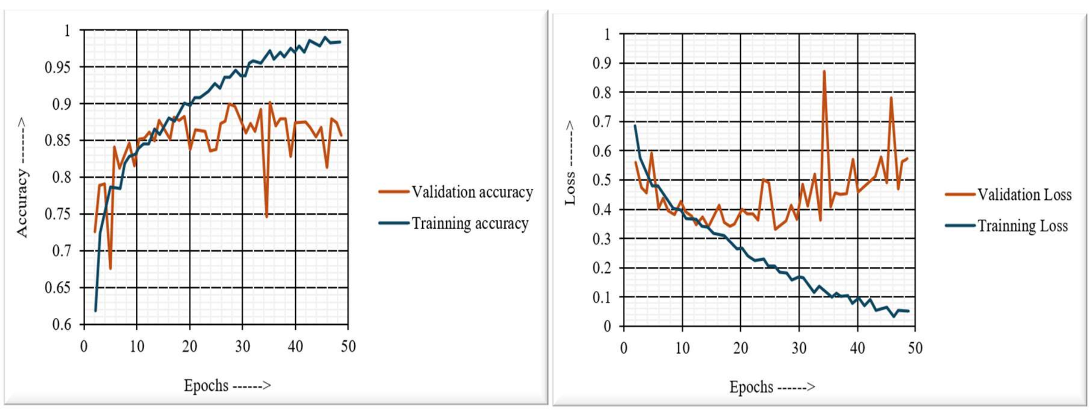


Figure 7: Accuracy and loss curve of the CNN model

Table 7: The result of the CNN model

S.No.	Performance Metrics	Result
1	**Accuracy**	97.50%
2	**Precision**	95.25%
3	**Recall**	100%
4	**F1-score**	97.56%

- **Bi-LSTM Model**

Figure 8 shows the results of Bi-LSTM validation and training/learning accuracies. The training accuracy, shown by the blue line, grows as the epoch count rises and reaches 100% after 50 epochs. The validation accuracy is shown by the brown curve, which starts at 96.56% and rises to 98.02% after 50 epochs. They stopped training after 50 epochs due to overfitting in the learning curve. Optimal training/validation accuracy was achieved by fine-tuning the number of training epochs.

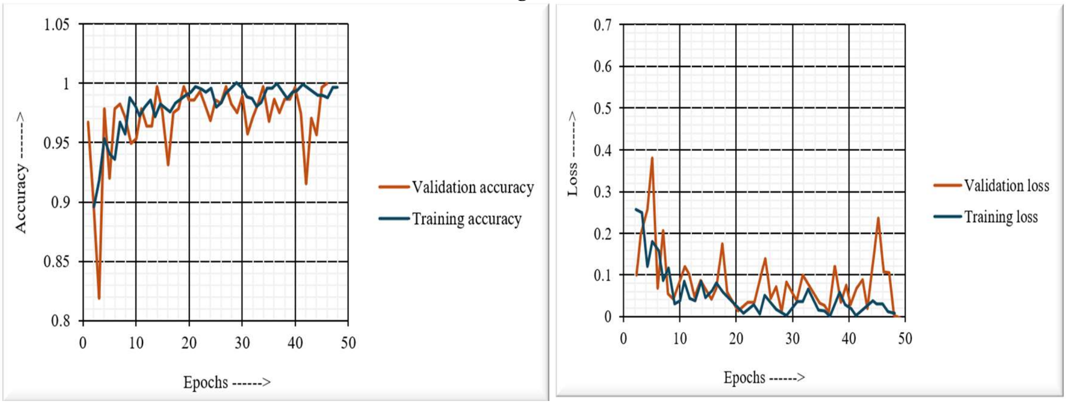


Figure 8: Accuracy and loss of the Bi-LSTM model

The training and validation loss for the Bi-LSTM model is shown in Figure 8. A score of 0.0 would imply that the whole learning process was outstanding and there were no errors observed. After 50 epochs, the training loss reached 0.0030 and the validation loss, which started at 0.110 and decreased to 0.01, both decreased steadily with an increasing epoch count. There is a report on the categorization using Bi-LSTM in Table 8.

Table 8: The result of the SVM model

S.No.	Performance Metrics	Result
1	**Accuracy**	97.29%
2	**Precision**	94.36%
3	**Recall**	97.29%
4	**F1-score**	95.80%

- **CNN-BiLSTM Model**

Figure 9 illustrate the performance of a CNN-BiLSTM model in terms of accuracy and loss during training and validation over 50 epochs. The accuracy curves show that training accuracy steadily increases, reaching nearly perfect values as the number of epochs progresses. This indicates that the model is effectively learning the patterns in the training data. On the other hand, the validation accuracy fluctuates between 0.9 and 0.95 after the initial few epochs. While this reflects good generalization to unseen data, the widening gap between the training and validation accuracies suggests possible overfitting.

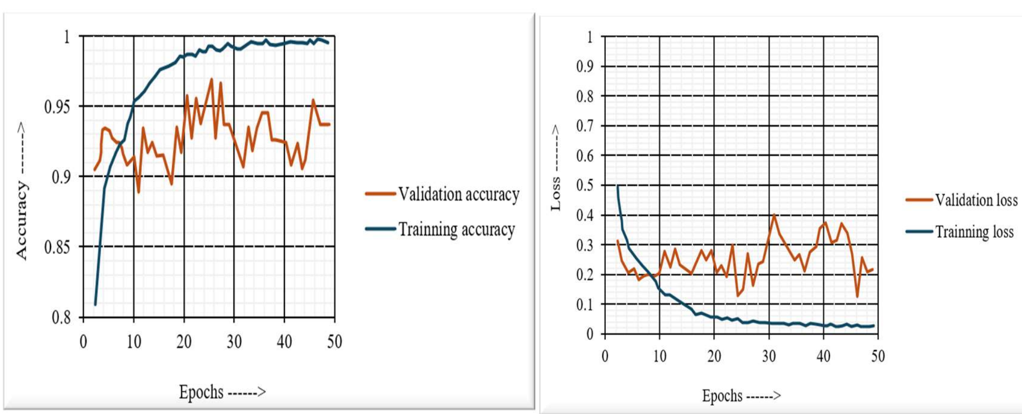


Figure 9: Accuracy curves of the CNN-BiLSTM model

The training loss shows a consistent decline, indicating improved fitting of the model to the training data. Meanwhile, the validation loss decreases initially but then stabilizes with some fluctuations, suggesting variability in the model's performance on the validation set. This behavior highlights the need for potential regularization or fine-tuning of hyperparameters to improve generalization further and reduce overfitting. Overall, the model demonstrates strong performance with a minor room for optimization. Additionally, the data that was used for this matrix matched the data that was tested. Table 9 displays the classification report that was generated by utilizing the suggested model.

Table 9: The result of the CNN-LSTM Model

S.No.	Performance Metrics	Result
1	**Accuracy**	98.83%
2	**Precision**	97.70%
3	**Recall**	98.64%
4	**F1-score**	98.17%

- **Comparison analysis**

They do comparative analysis in this section, which is divided into two parts. The first part is a comparison of the CNN-Bi-LSTM model with CNN and Bi-LSTM. In the second part, a comparative analysis of the suggested model with the approaches done in the past.

- **Comparison of the CNN-Bi-LSTM with CNN and Bi-LSTM**

The results of the CNN and Bi-LSTM models were compared with the findings of the CNN-Bi-LSTM model. The accuracy that is acquired with the use of the CNN model is 97.50%. The accuracy that is acquired with the use of the Bi-LSTM model is 97.29%. When applied to the same dataset, the RF model attained an accuracy of 98.83%. Figure 4.11 is a bar chart that illustrates the comparison of all the metrics between CNN, Bi-LSTM, and the CNN-Bi-LSTM model. Table 10 presents the results of this comparison.

Table 10: Comparison of the suggested model with CNN, Bi-LSTM, and the CNN-Bi-LSTM model

S.No.	Metrics	SVM	KNN	RF
1	**Accuracy (%)**	97.29	97.50	**98.83**
2	**Precision (%)**	94.36	95.25	**97.70**
3	**Recalling (%)**	97.29	100	**98.64**
4.	**F1-score (%)**	95.80	97.56	**98.17**

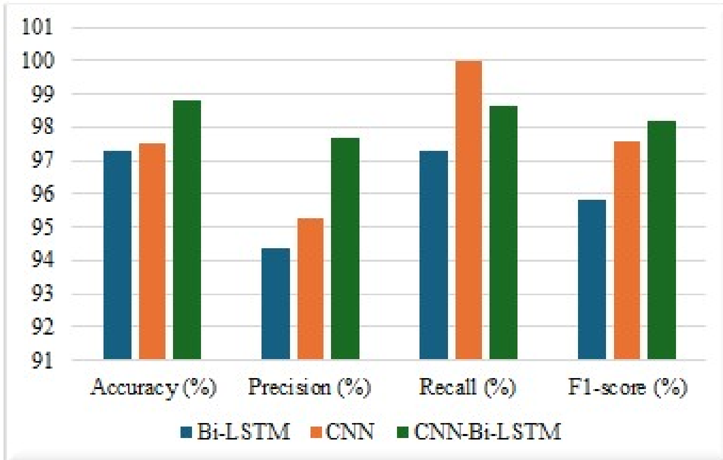


Figure 10: Graph of comparison of the Suggested Models

- **Comparison of the Suggested Model with Previous Work**

Figure 11 provides a comparative analysis of performance metrics, including $A_{accuracy}$, $P_{precision}$, R_{recall}, $F1_{score}$, across various previous studies: Loukili et al., Kusal et al., Tabany et al., Yu et al., and the proposed method. Loukili et al., (2023) achieved high and balanced performance, with all metrics exceeding 85%. Similarly, Kusal et al., (2023) showed slightly better results across all metrics, reflecting a consistent and reliable method. In contrast, Tabany et al., (2024) were comparatively weaker with 70% of accuracy and F1-score while other metrics slightly higher depicting low efficiency. Metrics in the experiment of Yu et al., (2024) improved greatly, the results being above 90% and proving that the method used is rather effective. Of all the methods examined, the proposed method was the best with nearly perfect results in terms of $A_{accuracy}$, $P_{precision}$, R_{recall}, $F1_{score}$. This proves the efficiency of the proposed approach when reaching the highest possible points for all the critical parameters that allowed giving it an advantage compared to the previous work (Table 11).

Table 11: Comparison of the suggested model with previous work

Methods Used	Accuracy (%)	Precision (%)	Recalling (%)	F1-score (%)
LR	90	85.4	78.6	81.4
BERT	78.91	79.08	78.75	78.86
SVM	70	63	70	62
LR	87.19	86.90	87.75	87.32
CNN-Bi-LSTM	**98.83**	**97.70**	**98.64**	**98.17**

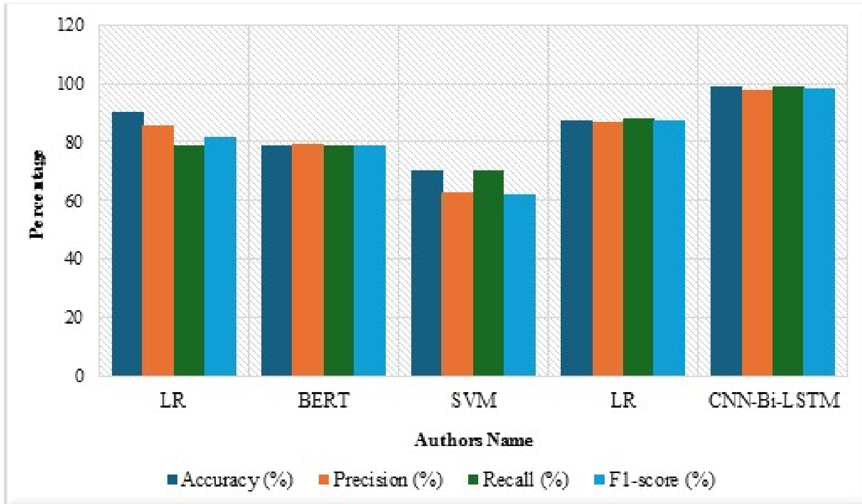


Figure 11: Bar graph of comparison models

5. CONCLUSION

This study proposed an integrated multimodal analytics platform for real-time interpretation of environmental sensor data and social media sentiment information, and introduced a Deep Learning-Driven Visual Analytics Framework for Real-Time Data Interpretation and Predictive Decision Making. The framework proposed was a combination of comprehensive data preprocessing, feature extraction, multimodal feature fusion, and a hybrid CNN-BiLSTM deep learning model, which optimized the predictive performance and facilitated intelligent decision-making. The evaluation results on the Air Quality Dataset and Sentiment140 Dataset showed the effectiveness of the approach proposed in capturing spatial and temporal pattern from various sources of heterogeneous data. The CNN-BiLSTM model outperformed the CNN, LSTM, and Bi-LSTM models in terms of performance, as evidenced by the regression analysis results with a lower MAE of 0.041, RMSE of 0.072, and R^2 score of 0.97. Likewise, the proposed framework demonstrated a high accuracy of 98.83%, precision of 97.70%, recall of 98.64%, and F1-score of 98.17%, thereby reinforcing the strength and robustness of the proposed framework. Predictive analytics combined with the interactive interpretation of visual data further increases situational awareness, helping to make effective decisions in a dynamic situation. The proposed framework provides an intelligent and scalable solution for real-time monitoring, forecasting and decision support in a wide variety of application domains, such as environmental management, smart cities, healthcare, and industrial analytics, among others.

REFERENCE

1. Patro, Pramoda, M. Sowmya, N. Srinivas, Subhranginee Das, Arul Kumar VP, and R. Dhanapal. "Big Data-Based Remote-Sensed Intelligent Visual Analytics for Environmental and Earth Monitoring Systems." *Journal of Applied Science and Technology Trends* (2025): 35-50.

2. Sammy, F., Thiyagarajan Chettier, Venkata Boyina, Harshala Shingne, Kamal Saluja, Manisha Mali, R. Sugumar, and A. Shobana. "Deep Learning-Driven Visual Analytics Framework for Next-Generation Environmental Monitoring." *Journal of Applied Science and Technology Trends* (2025): 114-122.

3. Kandregula, Narendra. "OPTIMIZING EDGE-TO-VISUALISATION PIPELINES: LEVERAGING MACHINE LEARNING FOR REAL TIME DECISION-MAKING AND PREDICTIVE ANALYTICS." (2025): 1532-42.

4. Lu, Shuaixin, Lingling Ba, Jie Wang, Min Zhou, Peiyao Huang, Xiaohua Zhang, Simo Pan, Xinmiao Zhou, Kai Wen, and Jing Sun. "Deep learning-driven approach for cataract management: towards precise identification and predictive analytics." *Frontiers in Cell and Developmental Biology* 13 (2025): 1611216.

5. Mamun, Md Nur Hasan. "Role of AI and Data Science in Data-Driven DecisionMaking for it Business Intelligence: A Systematic Literature Review." *Available at SSRN 5402976* (2025).

6. Adnan, Khan Muhammad, Taher M. Ghazal, Muhammad Saleem, Muhammad Sajid Farooq, Chan Yeob Yeun, Munir Ahmad, and Sang-Woong Lee. "Deep learning driven interpretable and informed decision making model for brain tumour prediction using explainable AI." *Scientific Reports* 15, no. 1 (2025): 19223.

7. Islam, Md Morshedul. "Artificial intelligence–driven predictive analytics framework for data-driven decision support in complex organizational systems." *American Journal of Data Science and Analytics* 7, no. 03 (2026): 163-207.

8. Raza, Hassan, A. Singh, Tsendayush Erdenetsogt, Muhammad Mohsin Kabeer, Muhammad Shahrukh Aslam, and Mazhar Farooq. "Machine Learning Driven Decision Making in the Modern Data Era." *PERFECT: Journal of Smart Algorithms* 3, no. 1 (2026): 11-22.

9. Nyoni, Rumbidzai. "Harnessing data analytics for predictive insights: advancing Decision-Making with big data innovations." *International Journal of Research Publication and Reviews* 6, no. 1 (2025): 2915-2936.

10. Strielkowski, Wadim, Andrey Vlasov, Kirill Selivanov, Konstantin Muraviev, and Vadim Shakhnov. "Prospects and challenges of the machine learning and data-driven methods for the predictive analysis of power systems: A review." *Energies* 16, no. 10 (2023): 4025.

11. Ramareddy, Sathish Kaniganahalli. "A CLOUD-BASED AI FRAMEWORK FOR REAL-TIME FINANCIAL DATA VISUALIZATION AND DECISION SUPPORT." *International Journal of Applied Mathematics* 38, no. 11s (2025): 1206-1225.

12. Agarwal, Navya Sonal, and Sanjay Kumar Sonbhadra. "A review on large language models for visual analytics." *arXiv preprint arXiv:2503.15176* (2025).

13. Yang, Aixiang. "Big data-driven corporate financial forecasting and decision support: a study of CNN-LSTM machine learning models." *Frontiers in Applied Mathematics and Statistics* 11 (2025): 1566078.

14. Zhang, Yunfan, Eakachat Joneurairatana, and Jirawat Vongphantuset. "An AI-Driven Decision Support Framework for Ergonomic Optimization in Fashion Manufacturing: Integrating Predictive Analytics and MCDM Techniques." *Decision Making: Applications in Management and Engineering* 7, no. 1 (2024): 786-802.

15. Yunfeng, Wang, and Othman Ibrahim. "Enhancing crowdfunding success with machine learning and visual analytics: Insights from chinese platforms." *IEEE Access* (2025).

16. Bhuiyan, Maniruzzaman, Mohammad Anisur Rahman, Afrin Binta Hoque, Md Ashrafuzzaman, and Anisur Rahman. "Advanced Analytics And Machine Learning For Revenue Optimization In The Hospitality Industry: A Comprehensive Review Of Frameworks." *American Journal of Scholarly Research and Innovation* 2, no. 02 (2023): 52-74.
17. El-Assady, Mennatallah, Wolfgang Jentner, Manuel Stein, Fabian Fischer, Tobias Schreck, and Daniel A. Keim. "Predictive visual analytics: Approaches for movie ratings and discussion of open research challenges." (2014).
18. Yang, Danlei, Chenhao Mi, Xiangcheng Lyu, Ying Xie, Zhenhua Luo, and Luofeng Huang. "Real-time machine learning-driven digital twin framework of a floating solar system in waves." *Energy Conversion and Management* 356 (2026): 121373.
19. Akintuyi, Olabimpe Banke. "Adaptive AI in precision agriculture: a review: investigating the use of self-learning algorithms in optimizing farm operations based on real-time data." *Research Journal of Multidisciplinary Studies* 7, no. 02 (2024): 016-030.
20. Durowoju, Emmanuel Segun, and Juwon Kehinde Olowonigba. "Machine learning-driven process optimization in semiconductor manufacturing: A new framework for yield enhancement and defect reduction." *Int J Adv Res Publ Rev* 1, no. 4 (2024): 110-130.
21. Ganesan, Thirusubramanian, and Mohanarangan Veerappermal Devarajan. "Integrating IoT, fog, and cloud computing for real-time ECG monitoring and scalable healthcare systems using machine learning-driven signal processing techniques." *International Journal of Information Technology and Computer Engineering* 9, no. 1 (2021): 202-217.
22. Yu, Yue, Yifang Wang, Yongjun Zhang, Huamin Qu, and Dongyu Liu. "InclusiViz: Visual Analytics of Human Mobility Data for Understanding and Mitigating Urban Segregation." *IEEE Transactions on Visualization and Computer Graphics* (2025).
23. Lyu, Guoliang. "Data-driven decision making in patient management: a systematic review." *BMC medical informatics and decision making* 25, no. 1 (2025): 239.
24. Cuzzocrea, Alfredo, Islam Belmerabet, Abderraouf Hafsaoui, and Carson K. Leung. "A Machine-Learning-Based Data Science Framework for Effectively and Efficiently Processing, Managing, and Visualizing Big Sequential Data." *Computers* 14, no. 7 (2025): 276.
25. Jamithireddy, Naren Swamy. "Visual Analytics of Internet Health: A Multidisciplinary Framework for Mapping Digital Infrastructure Performance." *Bridge: Journal of Multidisciplinary Explorations* 1, no. 2 (2025): 25-30.
26. Edeh, Chijioke George, Gloria Opoku Darkoh, Shobayo Ifeoluwanimi Praise, Julius Odemi Brown, Rufus Fidelis Ojuoluwa, Omotayo Christopher Afolabi, Waliu Temidayo Asamu, Ifeoluwa Odunayo Olofinsao, Nana Firdausi Hassan, and Akinsuyi Samson. "A Machine Learning-Based Framework for Real-Time Environmental Health Risk Prediction and Spatial Air Quality Intelligence in Urban-Industrial Ecosystems." Asian Journal of Advanced Research and Reports 20, no. 4 (2026): 1-20.
27. Ullah, Abaid, Muhammad Hamza Ejaz, Esha Seher Khalil, and Yasir Kareem. "Real time air quality monitoring and forecasting using AI and nature-based recommender system for climate-resilient air quality management." Fusion Journal of Engineering and Sciences (2026).

28. Alam, Nayema Sanzida, Marufa Akter Mitu, Md Foysal Siddik, Ahammad Hossain, Jayanta Das, Md Kamruzzaman, and A. H. M. Rahmatullah Imon. "Modeling and Predicting Air Quality Index of Bangladesh and India Using Machine Learning Approach." In The Nexus Between Geography and Sustainability: Exploring Emerging Tools and Techniques, pp. 455-482. Cham: Springer Nature Switzerland, 2026.

29. Taştan, Mehmet, Eren Cihan Karsu Asal, and Hayrettin Gökozan. "AQ-MultiCal: An Interactive No-Code Machine Learning Platform for Low-Cost Air Quality Sensor Calibration and Comparative Model Analysis." Sensors 26, no. 8 (2026): 2398.

30. Sammy, F., Thiyagarajan Chettier, Venkata Boyina, Harshala Shingne, Kamal Saluja, Manisha Mali, R. Sugumar, and A. Shobana. "Deep Learning-Driven Visual Analytics Framework for Next-Generation Environmental Monitoring." Journal of Applied Science and Technology Trends (2025): 114-122.

31. Ali, S. Md Shakir, P. Deivanai, Santanu Roy, Niharika Singh, Vinayak Vishwakarma, Prasanta Kumar Parida, and Kartik Dhiman. "Optimizing Marketing Strategies Through Customer Segmentation and Visual Analytics."

32. Yang, Aixiang. "Big data-driven corporate financial forecasting and decision support: a study of CNN-LSTM machine learning models." Frontiers in Applied Mathematics and Statistics 11 (2025): 1566078.

33. Adelusi, Bamidele Samuel, Damilola Osamika, M. C. Kelvin-Agwu, A. Y. Mustapha, A. Y. Forkuo, and Nura Ikhalea. "A machine learning-driven predictive framework for early detection and prevention of cardiovascular diseases in US healthcare systems." Engineering and Technology Journal 10, no. 5 (2025): 4727-4751.

34. Satish, Snehal, Hari Gonaygunta, Akhila Reddy Yadulla, Deepak Kumar, Mohan Harish Maturi, Karthik Meduri, Elyson De La Cruz, Geeta Sandeep Nadella, and Guna Sekhar Sajja. "Forecasting the Unseen: Enhancing Tsunami Occurrence Predictions with Machine-Learning-Driven Analytics." Computers 14, no. 5 (2025): 175.

35. Arun, M., Debabrata Barik, Nashwan Adnan Othman, Seepana Praveenkumar, and Kapura Tudu. "Investigating the performance of AI-driven smart building systems through advanced deep learning model analysis." Energy Reports 13 (2025): 5885-5899.

36. https://www.kaggle.com/datasets/fedesoriano/air-quality-data-set/code

37. https://www.kaggle.com/datasets/kazanova/sentiment140https://www.kaggle.com/datasets/kazanova/sentiment140

38. Rhanoui, Maryem, Mounia Mikram, Siham Yousfi, and Soukaina Barzali. "A CNN-BiLSTM model for document-level sentiment analysis." *Machine Learning and Knowledge Extraction* 1, no. 3 (2019): 832-847.

39. Tang, Chao, Yufeng Zhang, Fan Wu, and Zhuo Tang. "An improved cnn-bilstm model for power load prediction in uncertain power systems." *Energies* 17, no. 10 (2024): 2312.

40. Zhang, Jilin, Lishi Ye, and Yongzeng Lai. "Stock price prediction using CNN-BiLSTM-Attention model." *Mathematics* 11, no. 9 (2023): 1985.

41. Ayetiran, Eniafe Festus. "Attention-based aspect sentiment classification using enhanced learning through CNN-BiLSTM networks." *Knowledge-based systems* 252 (2022): 109409.

42. Chen, Yu, Ruixin Fang, Ting Liang, Zongyu Sha, Shicheng Li, Yugen Yi, Wei Zhou, and Huilin Song. "Stock Price Forecast Based on CNN-BiLSTM-ECA Model." *Scientific Programming* 2021, no. 1 (2021): 2446543.

43. Bohara, Bharat, Raymond I. Fernandez, Vysali Gollapudi, and Xingpeng Li. "Short-term aggregated residential load forecasting using BiLSTM and CNN-BiLSTM." In *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pp. 37-43. IEEE, 2022.

44. Thekkekara, Joel Philip, Sira Yongchareon, and Veronica Liesaputra. "An attention-based CNN-BiLSTM model for depression detection on social media text." *Expert systems with applications* 249 (2024): 123834.

45. Wang, Zhaohua, Longzhen Duan, Dongsheng Shuai, and Taorong Qiu. "Research on water environmental indicators prediction method based on EEMD decomposition with CNN-BiLSTM." *Scientific reports* 14, no. 1 (2024): 1676.

46. Liang, Yun-Chia, Yona Maimury, Angela Hsiang-Ling Chen, and Josue Rodolfo Cuevas Juarez. "Machine learning-based prediction of air quality." applied sciences 10, no. 24 (2020): 9151.

47. Natarajan, Suresh Kumar, Prakash Shanmurthy, Daniel Arockiam, Balamurugan Balusamy, and Shitharth Selvarajan. "Optimized machine learning model for air quality index prediction in major cities in India." Scientific reports 14, no. 1 (2024): 6795.

48. Ravindiran, Gokulan, Gasim Hayder, Karthick Kanagarathinam, Avinash Alagumalai, and Christian Sonne. "Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam." Chemosphere 338 (2023): 139518.