International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning Models to Analyze the Non-Linear-Lag Effect of Environmental Factors on the Occurrence of Schizophrenia

Jobin Thomas*[1], Murali P.[2]

Submitted: 11/01/2024 **Revised**: 17/02/2024 **Accepted**: 25/02/2024

Abstract: In recent years, the distributed lag non-linear Model(DLNM) has dominated over other techniques for measuring risk in environmental epidemiology. The impact of air pollutants or climate factors on schizophrenia is evident from the literature. This study aims to examine the influence of pollution and climate-related variables on the frequency of hospital admissions for individuals diagnosed with schizophrenia. We used DLNM and deep neural networks(DNNs) to explore the non-linear relationship between environmental variables and schizophrenia admissions in Bangalore City, India. The outcomes derived from the DLNM model reveal that the optimal forecast for hospital emergency visits is achieved with a lag of 3 days, resulting in a maximum RR value of 1.6 (95% confidence interval). Subsequently, DNN models, including the Convolutional Neural Network(CNN), hybrid CNN-LSTM, Long Short-Term Memory(LSTM), and the Gated Recurrent Unit(GRU), were employed, each with varying time steps, in the pursuit of refining predictive accuracy. These predictive models are evaluated by mean absolute error(MAE), the mean absolute percentage error(MAPE), mean square error(MSE), the root of mean square error(RMSE), and Symmetric Mean Absolute Percentage(SMAPE). We found results of deep learning models are consistent with the results of DLNM in predicting the number of admissions based on short-term environmental exposure. The short-term exposure-response relationship is evident in all models and it is proved through sensitivity analysis. CNN and GRU models have better performance than other models by using sigmoid activation functions. The CNN and GRU resulted with the lowest MAE(0.46, 0.49), MAPE(35.5%, 34.7%) and RMSE(0.73, 0.75).

Keywords: *Deep neural networks, time series analysis, lag effect, distributed lag nonlinear model, environmental factors*

1. Introduction

In recent years, scholars have examined the intricate relationship between environmental elements and human medical outcomes[1][2][3]. A subject of considerable scientific focus is the possible relationship connecting climate-related variables and the incidence of schizophrenia[4][5][6]. The appearance of psychosis in individuals diagnosed with schizophrenia is subject to the effect of several environmental variables, including particulate matter(PM), ambient temperature, nitrogen dioxide(NO2), relative humidity, and sulfur dioxide(SO2)[7][8]. This results in an increase in their admissions to mental health facilities. A commonly used surrogate measure for the occurrence of diseases is the utilization of medical facility admission data[9][10]. The dynamic character of the exposure-response association in schizophrenia prevalence is often seen due to a temporal delay of a few days[8][11]. The phenomenon of delay, sometimes referred to as "lag," is a subject of interest in our study. The potential existence of a lagged impact suggests that the connection between the independent variable and outcome variables is non-linear.

[1] *School of Computer Science and Engineering, Presidency University, Bangalore, India*
ORCID ID : 0000-0002-1708-5974
[2] *School of Computer Science and Engineering, Presidency University, Bangalore, India*
ORCID ID : 0000-001-7481-5307
** Corresponding Author Email: jobinthomas@presidencyuniversity.in*

Several studies have examined the delayed effects of pollutants on admissions for schizophrenia with a particular focus on metropolitan regions, which are usually non-linear[4][5]. The coexistence of many contaminants might result in a cumulative impact on health that surpasses the impact of each contaminant considered independently, which has not been investigated in these studies. Time-series regression techniques provide an adaptable structure for assessing this problem. Because of the ability of DLNM[12] to look at the non-linear and delayed link, researchers have extensively used this technique. The previous studies documented the environmental influence on schizophrenia.

The non-linear-lag impact of environmental variables on schizophrenia was studied across various parts of the world, such as Hefei,China[13][14] Queensland[37], TongLing,China[10], Sakai[8], Xi'an,China[15], Israel[3], Shandong[16] and Arizona[17]. These studies have shown that changes in air quality may have an immediate or delayed effect on schizophrenia patients, with a delay of three to six days. These studies used various pollutant variables, such as PM2.5[18][19], PM10, SO2, and NO2[7][10]. The primary climatic factors under investigation included atmospheric temperature (AT)[5][13][14] and sunshine[20]. Variables like relative humidity (RH)[20] and wind speed are used as cofounders, along with major contributor variables like temperature[14][15]. A statistically significant impact of short-term exposure to environmental variables on schizophrenia is shown in all of the aforementioned research, particularly during the three- to six-day lag period[11].

Following our lag selection criteria, we have determined that an ideal lag of 3 days is appropriate for analysing the data collected from Bangalore City. In this study of the regression type, we use the count of hospital admissions(NoA) as a surrogate measure for the occurrence of schizophrenia onset. Previous research has mostly focused on the examination of a single pollutant or, at most, three environmental elements in their investigations. All models were single-lag models with individual pollutant variables. The integration of pollutant absorption and climatic factors, such as heat and humidity levels, plays a significant role in the prognostication of schizophrenia hospitalizations, especially in densely populated urban areas with high levels of pollution. This study explores the intricacy of the combined effect of air quality variables on hospital admission via the use of diverse deep neural network(DNN) models. The DLNM technique is used to ascertain the combined exposure, whereby one variable is considered the primary contributor while the others are regarded as cofounders. In this research, we used deep neural networks(DNNs) to construct predictive models that investigate the cumulative impact of PM2.5, PM10, NO2, SO2, ambient temperature, and relative humidity on daily hospital admissions for schizophrenia.

Deep neural networks can acquire complex nonlinear associations of variables[21]. The effectiveness of conventional DLNM approaches may be limited when attempting to capture complex and nonlinear patterns that are complicated in nature. These models are specifically designed to effectively process and analyze data with a large number of dimensions, such as photos, text, and time-series data[22]. Deep learning models, such as recurrent neural networks(RNNs) have been specifically developed to effectively process sequential input that exhibits temporal relationships[23]. The mentioned models can accurately represent delayed effects and temporal patterns, both of which are crucial elements in the scope of DLNM. None of the literature on this particular problem is predictive; instead, it is a risk assessment type.

The power of neural networks in addressing the issue of environmental effects on schizophrenia has not been explored yet. The purpose of this study is to do a comparative analysis of deep learning predictive models, evaluating their performance in the context of multi-variate predictions. Specifically, the study examines the implications of three different activation functions (AFs) on the performance of these models. In this study, we propose deep neural networks such as CNN, Recurrent Neural Networks(RNNs) such as GRU and LSTM, and a combination of CNN&LSTM(CNN-LSTM) to assess the non-linear and lagged impact of environmental variables and to predict schizophrenia admissions in hospitals. In the rest of the paper, we elucidate the methodology used for data collection and processing, the models we considered, and their structure. The results of the comparison are presented, followed by a discussion and conclusion of this paper.

2. Data and Methods

To determine the occurrence of schizophrenia, we relied on the daily count of hospital admission data as a surrogate, following the methodology utilized in earlier investigations[24][5]. Due to the absence of automated records in these institutions, we collected patient admission statistics from entry/exit physical logs, which don't include any personal or confidential information. The data used includes only the date and the count of hospital visits from the metropolitan region of Bangalore. The dataset included 899 records collected over a period from 2018 to 2021. Day-wise average pollution statistics for the city of Bangalore were acquired from the Central Pollution Control Board of India. A set of air pollutants was obtained, namely SO2, NO2, and PM10, PM2.5, and data on the climate variables, namely relative humidity (RH) and atmospheric temperature (AT), were obtained from the India Meteorological Department (IMD) for the city of Bangalore. The average values of these parameters were used for our study. Table 1 provides a summary of the parameters of air quality. Both the air quality data and the number of hospital admissions are time series in nature. The forward fill method was used to substitute the missing data. Table 2 displays a concise example of the gathered data. Determining the best suitable lag value is crucial in the context of the time series nature of data. The test developed by Dickey-Fuller[25] was used to ascertain the optimal lag value. Among the many lag selection criteria used in the bound test, the lag is set at 3 days. In the context of time series, the knots have a logarithmic value system with a uniform distribution, and their greatest delay is 3 days. In our investigation, the autolag factor was assigned a value based on the Akaike Information Criterion (AIC).

2. 1 The Distributed Lag Non-Linear Model(DLNM)

The frequency of schizophrenic patient's admission follows a Poisson distribution. This study used a Poisson generalized linear regression technique along with a DLNM to investigate the association between the average concentration of environmental variables and hospitalizations for schizophrenia. Atmospheric temperature(AT) is considered a major contributor to schizophrenic onset, as per the literature. The other variables PM2.5, PM10, NO2, SO2, and relative humidity(RH) are considered cofounders. The typical model depiction proposed by Gasparrini et al.[26] is shown as $\gamma t \sim$

Poisson(μt).

$$\text{Log}(\mu t)=\beta AVGT_{t,l}+ \mathcal{S}\ (T,6)+(AVGP,3)+ \mathcal{S}\ (AVGH,3)+c \quad (1)$$

In this study, the variable "t" represents the day observed. The variable "vt" represents the daily count of hospitalizations on a given day. The variable "**c**" represents the model slope. The $\boldsymbol{AVGT_t}$', "**AVG$\boldsymbol{\mathcal{P}}$**" and " **AVG$\boldsymbol{\mathcal{H}}$**" are daily average temperature, pollutants, and humidity respectively. The cross-basis matrices '$\boldsymbol{AVGT_t}$', "**AVG$\boldsymbol{\mathcal{P}}$**" and " **AVG$\boldsymbol{\mathcal{H}}$**" are created by using the DLNM methodology. The variable "**β**" represents the vector of coefficients for "$\boldsymbol{AVGT_t}$", and "l" represents the lag in terms of days. Here we examined a lag of 0 to 3 days. $\boldsymbol{\mathcal{S}}$ is the natural cubic spline function, which was used to include the effects of a long-lasting pattern and seasonality, using three degrees of freedom. For pollutants and humidity, the degree of freedom used is 3. The coefficient β represents the main contributor to temperature. To consider the possibility of long-term health consequences, we included several lag durations in sensitivity tests.

2.2 Convolutional Neural Network (CNN)

CNNs are typically used for image data, but they can also be adapted for time series analysis, especially when spatial patterns are important[27[28]. We implemented a 1-dimensional(1D) CNN for time series prediction using a dataset with lagged features. The model is trained to predict the target variable which is the 'number of admissions'(NoA) based on the past four days' environmental values with six features as mentioned above. The architecture defines a sequential CNN model using Keras which includes various layers referred to as dense layers, which are a fundamental component of artificial neural networks. These layers consist of nodes, or neurons, that are linked to every neuron. The model is compiled using the Adam optimizer and the loss function used is MSE. The overall architecture is shown in figure1.

Table 1: Statistics of environmental data used

Env.Vari-ables→	PM2.5 μg/m^3	PM10 μg/m^3	NO2 μg/m^3	SO2 μg/m^3	A.T. °C	R.H. %
Total	899	899	899	899	899	899
Average	30.9	67	22	5.2	26	69.
Std.	12.8	36.1	9.7	2.1	4.2	17
Min.	13	19	9	4	17	25
Q1(25%	21	40	15	5	25	59
Q2(50%	27	57.5	22	6	27	71
Q3(75%	38	88	28	6	29	81
Max.	68	180	54	14	32	118

Table 2: Sample of the data set used

Month	Date	PM 2.5	PM 10	NO2	SO2	A.T.	R.H.	NoA
8	3	33.1	73.36	26	4	27	78	4
8	4	42.9	98.04	30	5	26	69	2
8	5	40.6	91.45	31	3	25	70	2
8	6	32.7	73.47	25	3	24	69	3
8	7	34.8	79.31	25	3	24	62	1

The layer of convolutional neural networks is defined as follows:

$$Z_i = \sigma(\mathrm{W}_i * X_{in} + b_i) \quad (2)$$

W_i is the convolutional filter weights and b_i is the bias for the i^{th} convolutional layer. X_{in} is the input time series data then flatten the output from the last pooling layer into a 1D vector.

$$Z_i^{'}=\text{pooling}(Z_i) \quad (3)$$

$$X_{flat}=\text{flatten}(Z_i^{'}) \quad (4)$$

The model used one fully connected layer to make predictions. ReLU(Rectified Linear Unit), Tanh(Hyperbolic Tangent), and Sigmoid are activation functions and activation functions used to compare performance.

$$Y_{out} = \sigma\left(\mathrm{W}_{fc} * X_{flat} + b_{fc}\right) \quad (5)$$

2.3 Long Short-Term Memory Network(LSTM)

Long Short-Term Memory is specifically developed to effectively collect and model long-term dependencies present in sequential data. We defined a simple Long Short-Term Memory (LSTM) model mathematically based on the given time series data[29]. Figure 2 represents the LSTM model we used for regression using a time series dataset with lagged features. To model the impact of environmental factors as a prediction factor of schizophrenia admission we designed an LSTM model as follows. The LSTM layer with 50 units is added to the model.

The LSTM equations for one-time step t are given by:

1. Input Gate (i_t):

$$i_t = \sigma(\mathrm{W}_{ii}.x_t + \mathrm{W}_{hi}.h_{t-1} + b_{hi}) \quad (6)$$

W terms represent weight matrices, x_t represents new input, b_h terms represent bias vectors and h_{t-1} is the previous hidden state. The sigmoid activation function, denoted by σ, is used in this context. Additionally, the symbol $\odot$ is used to represent element-wise multiplication.

2. Forget Gate (f_t):

$$f_t = \sigma\left(W_{if}.x_t + b_{if} + W_{hf}.h_{t-1} + b_{hf}\right) \tag{7}$$

3. Cell State Update (g_t)

$$g_t = tanh\left(W_{ig}.x_t + b_{ig} + W_{hg}.h_{t-1} + b_{hg}\right) \tag{8}$$

4. Cell State (c_t):

$$c_t = f_t \odot c_{t-1} \odot i_t \odot g_t \tag{9}$$

5. Output Gate (o_t)

$$o_t = \sigma(W_{io}.x_t + b_{io} + W_{ho}.h_{t-1} + b_{ho}) \tag{10}$$

6. Hidden State(h_t) :

$$h_t = o_t \odot \tanh\,(c_t) \tag{11}$$

7. Output Prediction(y_t):

y_t =Some output layer operation on h_t

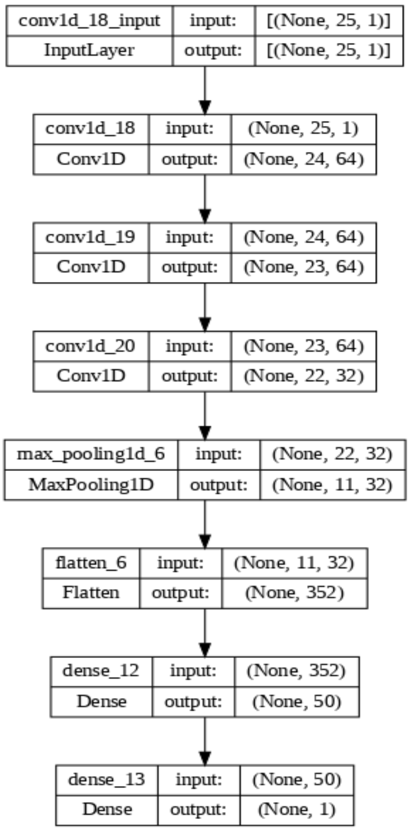

Fig 1: Input-output layered architecture of CNN model used with lagged features

2.4 Gated Recurrent Unit(GRU)

Similar to LSTMs but with a simpler structure, having only two gates (update and reset gates) compared to LSTM's three gates and used for sequential data processing[30]. To model the impact of environmental factors as a prediction factor of schizophrenia admission we designed the GRU model as follows. The first GRU layer is comprised of 64 units, whereas the subsequent GRU layer consists of 32 units. Figure 3 depicts the GRU model for regression using a time series dataset with lagged features. Here's a typical configuration for a GRU cell in terms of activation functions:

1. Update Gate (z): Sigmoid activation is commonly used here to determine the extent to which historical knowledge should be transmitted to subsequent periods.

$$z_t = \sigma(W_z.[h_{t-1}, x_t]) \tag{12}$$

2. Reset Gate (r): Sigmoid activation is also frequently used to determine how much of the past information to forget.

$$r_t = \sigma(W_r.[h_{t-1}, x_t]) \tag{13}$$

3. Candidate Hidden State (h_t'): Hyperbolic tangent (Tanh) is commonly used here to create a new candidate hidden state.

$$h_t' = \tanh\,(W_h.[r_t \odot h_{t-1}, x_t]) \tag{14}$$

4. Hidden State (h_t): The ultimate latent state results from the amalgamation of the preceding implicit state and the potential hidden state, with their respective contributions determined by the update and reset gates.

$$h_t = (1 - x_t) \odot h_{t-1} + z_t \odot h_t]) \tag{15}$$

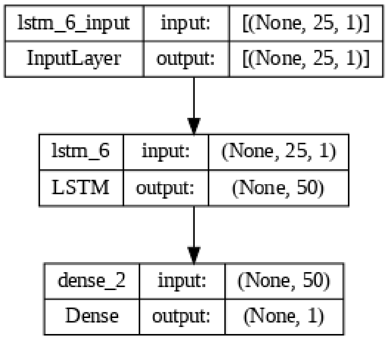

Fig 2: LSTM model for regression using a time series dataset with lagged features.

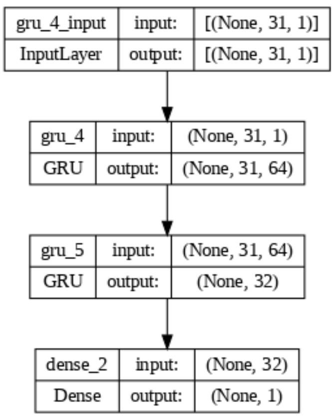

Fig 3: GRU model for regression using a time series dataset with lagged features

2.5 CNN-LSTM HYBRID MODEL

The formation of the CNN-LSTM hybrid model involves the merging of the CNN and LSTM architectures[31]. The CNN part is responsible for extracting spatial features from the input sequences, and the LSTM part captures temporal dependencies in the feature maps produced by the CNN. The input data consists of time series sequences with multiple features(e.g., PM2.5, PM10, NO2, SO2, AT, RH) and a target variable(NoA). Each sequence is represented as a matrix, where rows correspond to different time steps, and columns correspond to different features. The first part of the model is a CNN layer and the layer applies filters (kernels) to the input sequences, convolving over the time steps and features. The output of the CNN layer is a set of feature maps representing spatial patterns learned from the input sequences. The output of the CNN layer is reshaped to a 3D tensor to serve as input for the LSTM layer. The second part of the model is an LSTM layer, which captures temporal dependencies in the feature maps obtained from the CNN layer. The LSTM layer is responsible for processing the tensor that has been reshaped, namely along the time dimension. This procedure allows the layer to effectively learn and capture long-term relationships present in the data. The hidden states of the LSTM are updated by incorporating information from both the current input and the prior hidden state. This enables the model to effectively capture and represent sequential patterns. After the LSTM layer, one Dense layer for further processing and abstraction of features is added. This dense layer with a single neuron, represents the output for regression tasks. The activation function used is linear. The model is compiled with an appropriate optimizer and loss function (e.g., MSE for regression).

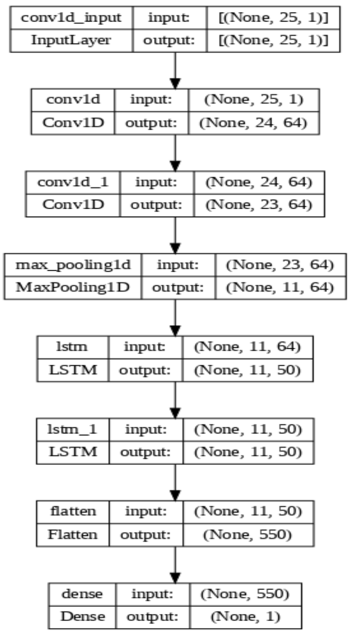

Fig 4: CNN-LSTM architecture for time series dataset with lagged features

2.6 Activation Functions(AF)

ReLU (Rectified Linear Unit), Tanh (Hyperbolic Tangent), and Sigmoid are activation functions commonly used in neural networks[32]. In recent architectures, ReLU is often the default choice for hidden layers due to its training efficiency[33]. In the case of GRUs or LSTMs are specifically designed to collect and model long-term dependencies within data that is sequenced, the hyperbolic tangent (Tanh) and sigmoid activations are commonly used within the gates and cell states due to their ability to control the flow of information. We have applied all three activation functions and compared the results to find the best-fitting deep learning model to address the multivariate time series problem of the exposure of environmental variables and its response to schizophrenia admission.

ReLU is Represented as

$$f(\mathrm{y})=\mathrm{MAX}(0,\mathrm{y}) \quad (16)$$

ReLU function is a mathematical operation that replaces all negative values with zero while leaving positive values unaltered. It is widely used in hidden layers due to its simplicity and the fact that it helps mitigate the vanishing gradient problem, allowing for faster training.

Tanh is Represented as

$$(\mathrm{x}) = \tanh(\mathrm{x}) = \frac{e^{2x}-1}{e^{2x}+1} \quad (17)$$

Tanh squashes the input values to the range [-1,1] mapping negative values strongly negative and positive values strongly positive. It is often used in the hidden layers of the network. Tanh is zero-centered, which can be beneficial for optimization.

Sigmoid is Represented as

$$f(\mathrm{y}) == \frac{1}{1+e^{-y}} \quad (18)$$

Sigmoid squashes provide input values within the range of [0,1]. It can also be used in hidden layers, but it is less common than ReLU and Tanh.

3. Results and Discussion

The research used the R and Python programming languages for its implementation. The assessment of the model's performance was carried out by using five error measures, namely MSE, RMSE, MAE, MAPE, and SMAPE. These metrics function as reference points for evaluating the efficacy of our regression-predicting models. A model that exhibits lower values of MAPE, MSE, MAE, and RMSE and the combination of these measures offers a full assessment of the predictive model's performance, providing valuable information into its level of accuracy and effectiveness in capturing the underlying patterns in the data. MSE and RMSE are sensitive to outliers because they involve squared differences. MAE is less sensitive. MAPE and SMAPE handle the case when the actual value is zero more gracefully than other metrics. In our investigation, we identified a strong association between external factors and the occurrence of hospital admissions of schizophrenia within the dataset specific to Bangalore, India. Adhering to the lag selection standards used in our dataset, we limited the lag values to a period of 3 days in our study. This restriction in lag values was implemented based on the specific considerations and parameters outlined in our analysis.

3.1. DLNM ANALYSIS RESULTS

The three-dimensional graphic shown in Figure 5 illustrates the projected impacts linked to the combined effect of environment variables at specific periods. The presented graphic depicts the exposure of environmental factors and their response to Schizophrenia in hospitals located in Bangalore, India. The results of our study demonstrate positive correlations that exhibit non-linear trends throughout a range of zero to four days of delay for, SO2, particulate matters, NO2, ambient temperature, and humidity. Figure 5 emphasizes the lag effects within the time frames of 0 to 3 days and 0 to 4 days. These lagged effects are quantified and expressed in a measure of relative risk(RR). The computation of relative risk entails the assessment of the ratio between the probability of occurrence of hospital admission that has been subjected to environmental factors, and the probability of occurrence of hospital admission that wasn't subjected to environmental factors. The presented graphs illustrate several patterns regarding the influence of the pollutants at minimum, modest, and excessive levels, taking into account the lag time effects. The corresponding 95% confidence intervals serve as a means to quantify the level of uncertainty related to these estimations. The comprehensive investigation described here enhances our understanding of the soon-after effect of poor air quality on the frequency of hospital admissions located in the city of Bangalore with a lag of zero to four days.

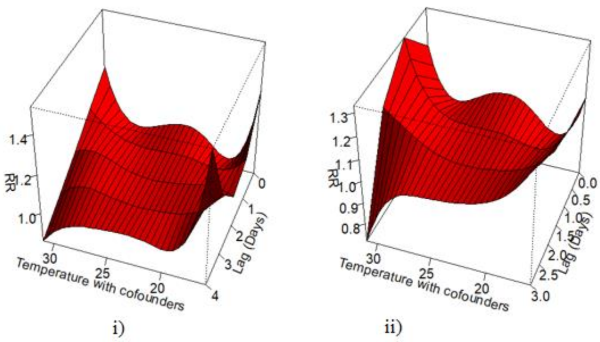

Fig 5: Relative Risk of temperature and its cofounders with a lag of 3(i) and lag of 4(ii) using DLNM

In our detailed examination, it has been repeatedly observed that the DLNM exhibits the lowest values for RMSE, MSE, MAE, and MAPE among all lag models ranging from lag 0 to lag 3. The steady performance shown in this study indicates that the DLNM is great when applied to the particular dataset about hospital admissions. In a broader framework, our examination discloses positive correlations, characterized by nonlinear temporal delays, among PM2.5, PM10, dioxides of nitrogen and sulfur, AT, RH, and the daily count of hospital admissions. Figure 6 depicts the results of the analysis, demonstrating that each pollutant and climatic factor displays distinct relationships with admissions about schizophrenia, distinctive to the lag period. Figure 6 shows single variable lag models in terms of relative risk using DLNM. The highest relative risk(RR) of schizophrenia admission is found with relative humidity(RH), 1.5 for high-level humidity at lag 0. The RR of 1.6 for low-level humidity at lag 3. The RR values at lag 0, which is the immediate effect are high compared to lag3. The RR of temperature at lag 3 is less, which means the effect of temperature on schizophrenia admission is comparatively less. However, as per the result, all environmental variables have a significant effect on the number of admissions of schizophrenia.

3.2 CNN, LSTM, AND GRU ANALYSIS RESULTS

Figure 7 shows the training loss and validation loss of the CNN model performed on our data to predict hospital admission based on a 4-day lag of environmental factors. The various possibilities of the activation function(AF) are explored to find best suitable model. The ReLU and Tanh are showing signs of overfitting the training data. The sigmoid AF fitted fine with our data on CNN. As seen in Figures 7(i) and 7(ii), the training loss continues to decrease because the model is becoming increasingly tailored to the training set. This overfitting arises when a model acquires an excessive level of knowledge of the training data, leading to certain adverse effects. A comparison of the predicted number of admissions(NoA) versus actual NoA with various activation functions is shown in Figure8. The CNN performed well with the sigmoid activation function compared to ReLU and Tanh as shown in Figure8(c). With the MSE error metric CNN produced the least value of 0.53 using sigmoidAF against 0.88 of ReLU. Table 4 and Table 5 indicate similar results with RMSE and MAE metrics.

MAPE calculates the percentage difference between predicted NoA and actual NoA, averages these percentage differences, and expresses the result as a percentage. In MAPE evaluation SigmoidAF outperformed ReLU and Tanh with 35.56% against 48.01% and 37.40%.

Figure 9 shows the training loss and validation loss of the LSTM model performed on our data to predict NoA. The data was prepared with additional features, which is the lag effect of environmental factors. As shown in this result both training and validation losses are decreasing and parallel movement suggests that the model is successfully acquiring knowledge from the training data and demonstrating strong generalization abilities when presented with new, unseen data. The parallel movement indicates that the model on the validation set is consistent with its effectiveness in the training data and capturing underlying patterns that generalize well. A comparison of the predicted number of admissions(NoA) versus actual NoA with activation functions(AF) ReLU, Tanh, and Sigmoid is shown in Figure 10. In the context of examining the lagged association between climate and admissions, it is observed that the performance of LSTM is still behind that of CNN. In this LSTM model, sigmoidAF has slightly better results than other AFs as shown in Figure 10(c). In error metrics, LSTM produced the least MSE value of 0.627 and 0.682 with sigmoid and ReLU respectively, as shown in Table 3.
Table 4 indicates similar results with the RMSE metric. In terms of MAE sigmoid looks better than other AFs, as shown in Table 5. In the evaluation of the percentage difference between predicted NoA and actual NoA, MAPE scored 49.4% with ReLU, 50.99%, and 50.38% with Tanh and Sigmoid AFs. Figure 11 depicts training loss vs validation loss in GRU prediction of NoA with various activation functions. As Figure 11 (i) and (ii) show both training and validation losses are decreasing indicating that the model is effectively gaining expertise from the provided training set and extrapolating to novel instances. With sigmoidAF the training loss is decreasing and parallel to the validation loss, implying that the model is exhibiting efficient learning capabilities and is not excessively fitting to the training data. GRU outperformed other deep learning techniques in consideration of all the error metrics. The MSE values of the GRU model are 0.666, 0.595, and 0.571 for ReLU, Tanh, and Sigmoid AFs respectively. The corresponding MAE values are 0.587, 0.515, and 0.492. For mean absolute percentage error also GRU outperformed CNN and LSTM. MAPE and SAMPE values with Sigmoid AF are 34.72% and 0.515% respectively as shown in Table 6 and Table 7. Figure 12 depicts a visual explanation of GRU prediction. The predicted number of admissions has excellent accuracy compared to other deep-learning models in our study.

3.3 CNN-LSTM MODEL ANALYSIS RESULTS

Finally, we explored a combination of CNN and LSTM on this particular problem of exposure-response relationship[33]. As mentioned above, six features of environmental variables and schizophrenia hospital admission variables along with their lag attributes were trained and tested. Results show that the training loss is decreasing and moving in parallel to the validation loss, as visualized in Figure 13. This observation implies that the model is exhibiting efficient learning behavior and is not excessively fitting to the training data. CNN-LSTM model has not performed as expected compared to traditional deep learning models. It produced high MSE values compared to other models which are 0.857, 1.079, and 1.079 for ReLU, Tanh, and Sigmoid respectively. As shown in Tables 3 to 5, the error metrics such as MSE, MAE, and RMSE showed poor performance with our dataset. The prediction of schizophrenia admission based on environmental variables is visualized in Figure 14. In effect, CNN-LSTM may not be a good choice in exposure-response kind of regression models using lag effect. The CNN-LSTM may not capture a non-linear relationship between environmental variables and schizophrenia admission.

To assess the robustness and reliability of the models in studying the exposure of pollution and climate factors and its response on schizophrenia admission, performed sensitivity analysis. In sensitivity analysis, we trained these deep learning models and the DLNM model with various lags of 1 to 4 as shown in Table 8 and Figure 15. Table 8 and Figure 15 indicate the performance of various models with a lag effect of 1 to 4 in terms of MSE. DLNM model has the lowest MSE on Lag1, Lag 2, and Lag 4(days), indicating it performed the best among the specified models. At lag 3, the CNN model produced the lowest MSE. Lower MSE values generally indicate better predictive performance, as they indicate smaller errors between the predicted NoA and the actual NoA. Lag-based sensitivity analysis measured with MAPE is shown in Table 9. The visual comparison is depicted in Figure 16, where GRU has the lowest MAPE on lag 2 and lag 3(day), indicating better performance in predicting values with a lag of 2 and 3. For lag 1, LSTM has the lowest MAPE, suggesting better performance at predicting the soon-after effect of environmental variables on schizophrenia admissions. On lag 4, CNN has the lowest MAPE, suggesting better performance in predicting the number of admissions with a lag of 4.

Table 3: MSE values of various models in comparison with DLNM and various activation functions

Activation function	CNN	LSTM	GRU	CNN-LSTM	DLNM
RELU	0.88	0.682	0.666	0.857	0.477
TANGENT	0.59	0.666	0.595	1.079	
SIGMOID	0.53	0.6269	0.571	1.079	

Table 4: RMSE values of various models in comparison with DLNM and various activation functions

Activation function	CNN	LSTM	GRU	CNN-LSTM	DLNM
RELU	0.94	0.826	0.816	0.925	0.690
TANGENT	0.77	0.816	0.771	1.038	
SIGMOID	0.73	0.791	0.755	1.038	

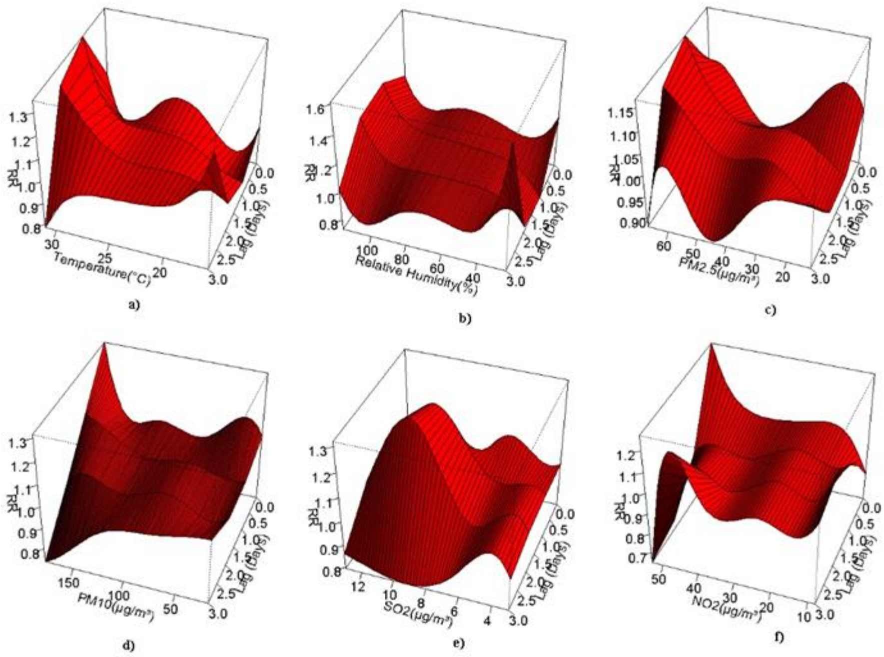

Fig 6: The use of DLNM for the assessment of risk associated with different environmental factors.

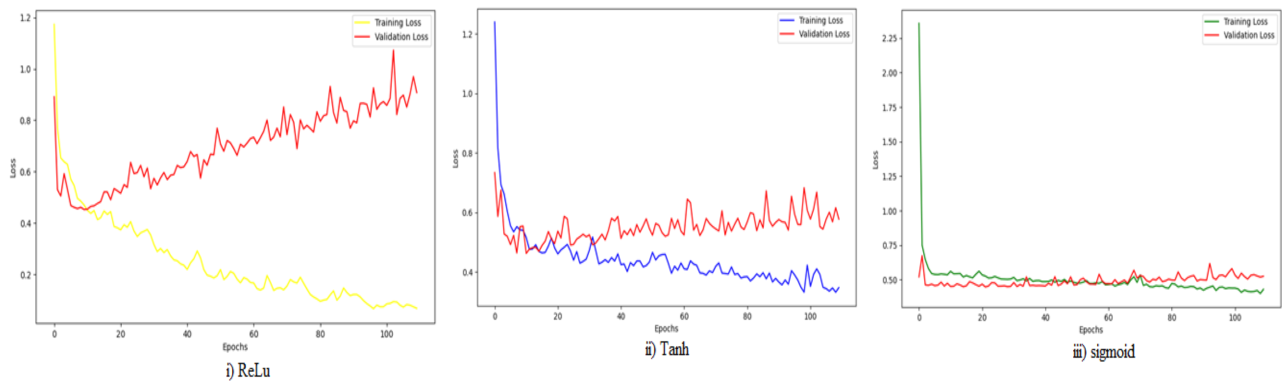

Fig 7: Training loss vs validation loss in CNN prediction of number of admissions with various activation functions

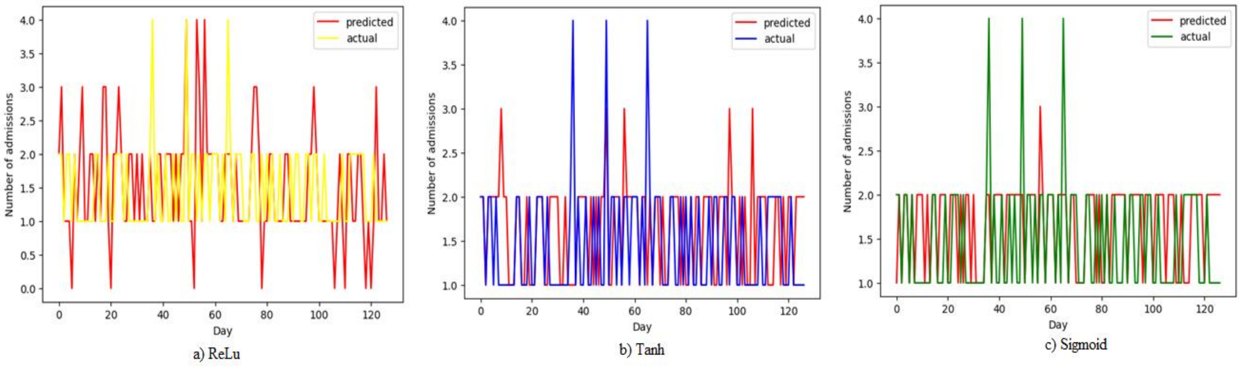

Fig 8: Prediction vs actual for number of admissions using CNN with various activation functions

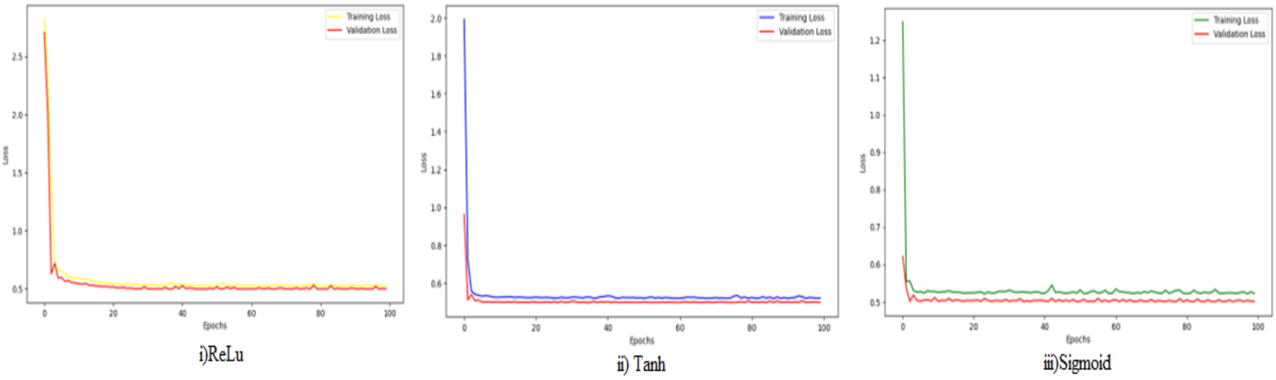

Fig 9: Training loss vs validation loss in LSTM prediction of number of admissions with various activation functions

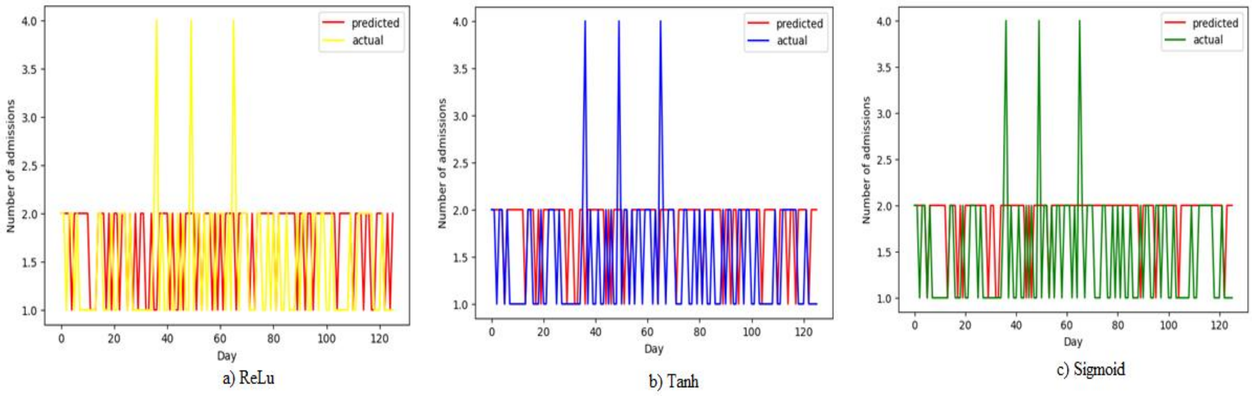

Fig 10: Prediction vs actual for the number of admissions using LSTM with various activation functions

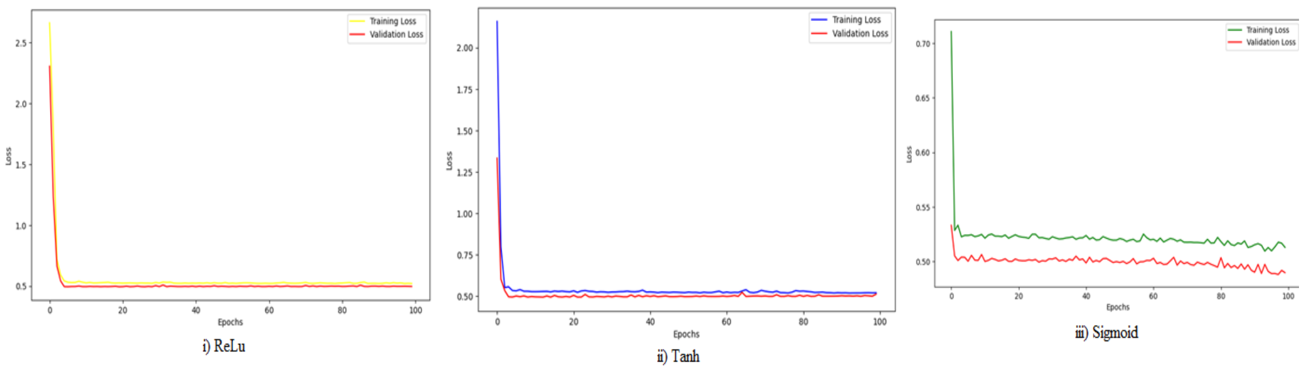

Fig 11: Training loss vs validation loss in GRU prediction of number of admissions with various activation functions

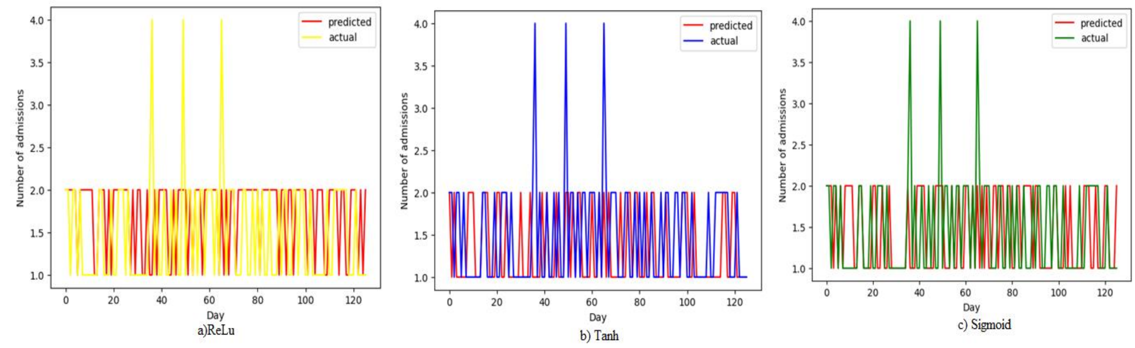

Fig 12: Prediction vs actual for number of admissions using GRU with various activation functions

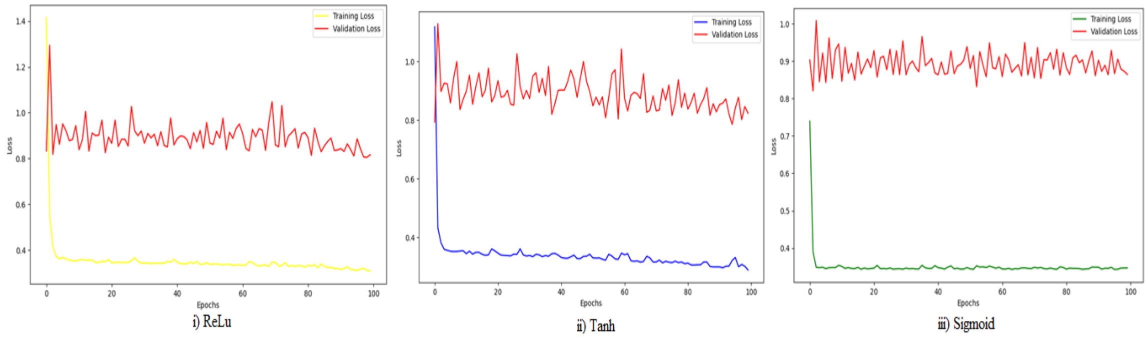

Fig 13: Training loss vs validation loss in CNN-LSTM hybrid model

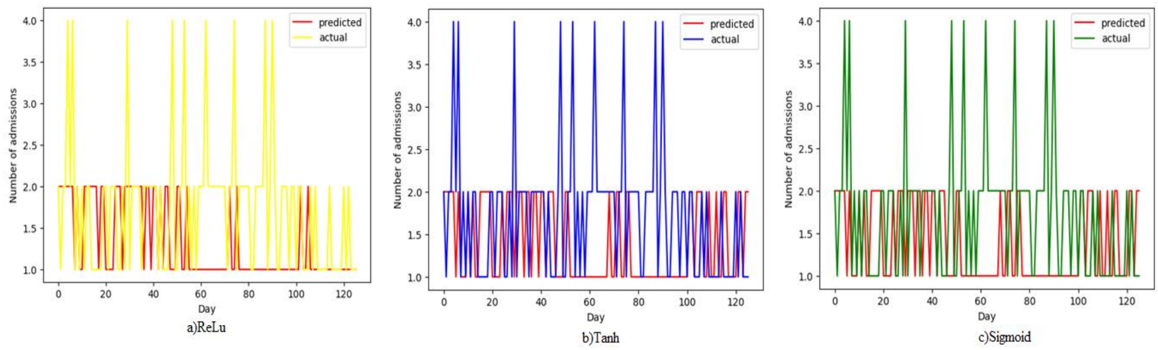

Figure 14: Prediction vs actual for number of admissions using CNN-LSTM hybrid model

Table 5: MAE values of various models in comparison with DLNM and various activation functions

Activation function	CNN	LSTM	GRU	CNN-LSTM	DLNM
RELU	0.64	0.603	0.587	0.587	
TANGENT	0.50	0.587	0.515	0.714	0.546
SIGMOID	0.46	0.579	0.492	0.714	

Table 6: MAPE values in percentage for various models with activation functions

Activation function	CNN	LSTM	GRU	CNN-LSTM	DLNM
RELU	48.01%	49.40%	48.61%	36.50%	
TANGENT	37.40%	50.99%	33.53%	39.88%	38.41%
SIGMOID	35.56%	50.38%	34.72%	39.88%	

Table 7: SMAPE values in percentage for various models with activation functions

Activation function	CNN	LSTM	GRU	CNN-LSTM	DLNM
RELU	0.65%	0.661%	0.654%	0.535%	0.745%
TANGENT	0.54%	0.675%	0.502%	0.570%	
SIGMOID	0.52%	0.677%	0.515%	0.570%	

Table 8: Lag effect of various models from day 1 to day 4 in terms of MSE

Lag(Day)	CNN (sigmoid)	LSTM (sigmoid)	GRU (sigmoid)	CNN-LSTM (Relu)	DLNM
Lag1	0.70	0.9603	0.6190	1.134	0.4768
Lag2	0.63	0.666	0.8888	0.9603	0.469
Lag3	0.53	0.6269	0.57142	0.857	0.4772
Lag4	0.66	0.66666	0.6587	0.97619	0.4779

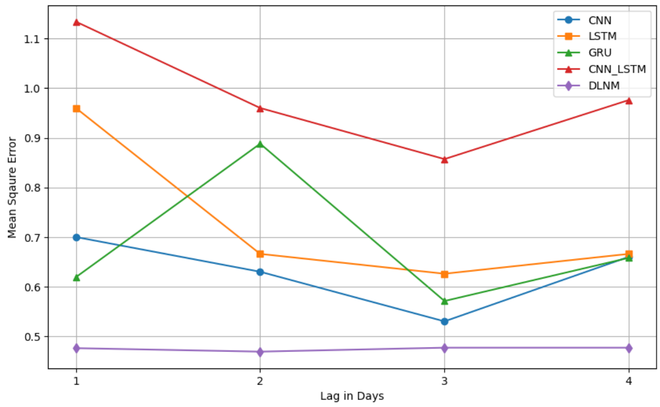

Fig 15: Lag effect of various models form day 1 to day 4 in terms of MSE

Table 9: Lag effect of various models from day 1 to day 4 in terms of MAPE(%)

Lag(Day)	CNN (sigmoid)	LSTM (sigmoid)	GRU (sigmoid)	CNN-LSTM (Relu)	DLNM
Lag1	49.41	25.0	43.25	36.50	37.961
Lag2	43.50	54.36	44.64	41.46	37.23
Lag3	35.56	50.38	34.72	36.50	38.41
Lag4	29.23	46.89	54.23	38.49	37.90

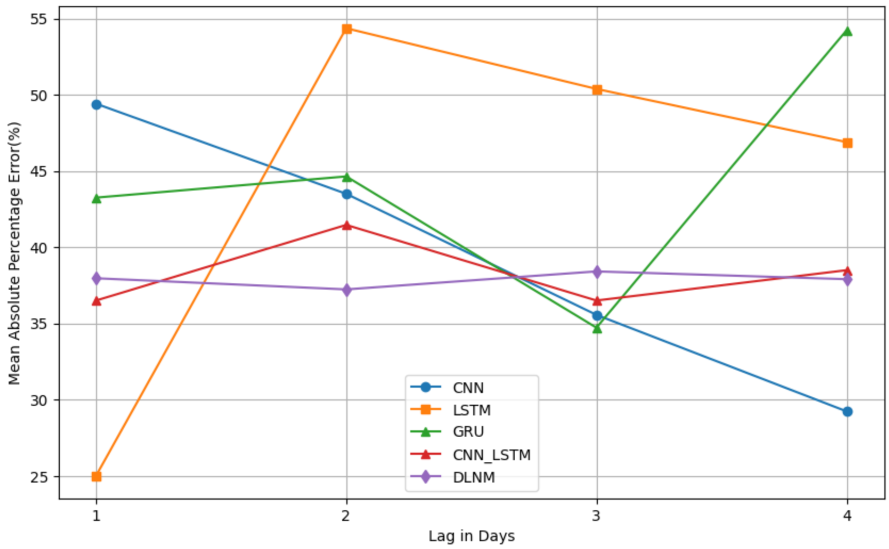

Fig 16: Lag effect of various models from day 1 to day 4 in terms of MAPE

4. Discussion

The results of deep learning models and DLNM using the data from Bangalore city, India is consistent with similar research happened in foreign cities, mainly in China. Our results using DLNM shows the highest correlation between PM2.5, PM10, NO2, SO2,AT,RH and hospital admission of schizophrenia during day 1 to 4 after exposure. This discovery is consistent with the results of several research that have successfully used Distributed Lag Non-linear Models to examine this issue and derive findings pertaining to lag-response associations. The exposure to PM2.5 and hospitalization response for schizophrenia occurs on the second and third day after exposure, consistent with findings described in other scholarly works[8]. In another investigation conducted by Liang et al. (2019), it was shown that there was a shorter lag correlation, which was attributed to the increased concentration of PM10. The negative impacts resulting from significant temperature changes were seen up to a lag of six days[14]. As per the optimal lag selection criteria of our data, we used a lag of up to 4 days. The effect of the highest temperature with a lag of 1 day and the lowest temperature with a lag of 3 days is evident in Bangalore city. The study conducted by Wang et al.[5] examined the influence of temperature on the beginning of schizophrenia. The researchers watched the participants for three consecutive days after their exposure to varying temperatures. Prior studies have shown that the risk effects of pollutants, namely NO2 and PM10, were seen to occur from lag 0 to lag 5. Similarly, the risk effects of SO2 were found to occur from lag 0 to lag 10 [10][15]. which strongly proves our results.

Most of the previous studies investigated the impact of a single environmental variable on hospital admission or a maximum of three pollutants. Multivariate exposure to pollutants and climate variables has been considered in our research. To get the exact sense of the combined effect of multiple variables on the dependent variable, which is NoA, and to get its non-linear effect we explored deep neural networks. The popular technique in literature was DLNM which may not measure the combined effect efficiently. Because DLNM considers one variable as a major contributor and other variables as cofounders. We consider our results using DNN to be more reliable for measuring the non-linear effect of the surrounding environment on sudden schizophrenia onset. There is no other research on this particular problem that investigated the possibility of deep learning techniques. So, we compare our results with DLNM and also with other studies that investigated non-linear modeling using neural networks. The study conducted by Lu et al.[33] provided evidence supporting the use of Long Short-Term Memory as a viable approach for identifying the delayed impact of PM2.5 on respiratory ailments. Hossain et al.[34] used non-linear methods of analysis to reconstruct lengthy periodic climate patterns and demonstrated the superiority of artificial neural networks over linear models. LSTM was used to train and model daily cases of influenza based on meteorological daily values and proved that there is a significant relationship between meteorological factors and influenza[35]. The study conducted by Mahmudimanesh et al.[36] presents a predictive model for estimating the death rate among individuals with cardiovascular conditions. Their model takes into account several air pollution indicators and utilizes a CNN-LSTM technique. They found that NO2 was also significant in lag 6. As shown in Figure 15, the models used in our study such as CNN, LSTM, and GRU are good in predicting the number of admissions based on environmental factors. The highest association is found at lag 3 even though lag association is evident from lag 1 to lag 4. Using the optimal lag of 3 the GRU has better performance in terms of error metrics as shown in table 3 to table 7.

5. Conclusion

The purpose of this analysis was to explore deep learning techniques to determine non-linear exposure-response relationships in time series analysis with lag effect. In this analysis using DLNM and deep learning algorithms, there exists a noteworthy link between exposure to environmental factors and the presentation of signs related to schizophrenia inside the urban region. of Bangalore is evident. The association between pollutants(PM2.5, PM10, SO2, NO2), climate factors(Temperature and humidity) and hospitalizations for schizophrenia is well established, and the findings remain consistent in sensitivity analysis. The predicted link between exposure and reaction was non-linear and models with combined effect of environment variables were performed under this assumption. Results indicate that the CNN and GRU models have better performance with sigmoid activation function in predicting hospital admissions. The efficiency of DNNs in finding underlying non-linear patterns is helpful in this problem. DLNM cannot represent the combined effect of all variables and so results are not reliable. This can be overcome using DNNs and it performed well with consistent results in all sorts of comparisons.

Ethical Approval

Human data or personally identifiable information is absent.

Funding

The allocation of financial resources does not apply to this research article.

References

[1] Yolton K, Khoury JC, Burkle J, LeMasters G et al. (2019) Lifetime exposure to traffic-related air pollution and symptoms of depression and anxiety at age 12 years. Environmental Research, 173, 199-206.

[2] Xu H, Jia Y, Sun Z, Su J, Liu Q S et.al (2022) Environmental pollution, a hidden culprit for health issues. Eco-Environment & Health, 1(1), 31-45.

[3] Yackerson N S, Zilberman A, Todder D, Kaplan Z (2014) The influence of air-suspended particulate concentration on the incidence of suicide attempts and exacerbation of schizophrenia. International journal of biometeorology, 58, 61-67.

[4] Qiu H, Tak-sun Yu I, Tse LA, Tian L, Wang X and Wong TW (2013) Is greater temperature change within a day associated with increased emergency hospital admissions for heart failure?. Circulation: Heart Failure, 6(5), pp.930-935.

[5] Wang S, Zhang X, Xie M, Zhao D, Zhang H et al. (2018) Effect of increasing temperature on daily hospital admissions for schizophrenia in Hefei, China: a time-series analysis. Public Health, 159, pp.70-77.

[6] Engemann K, Pedersen CB, Arge L, Tsirogiannis C et.al (2018) Childhood exposure to green space–a novel risk-decreasing mechanism for schizophrenia. Schizophrenia research, 199, 142-148.

[7] Ji Y, Liu B, Song J, Pan R, Cheng J, Wang H, & Su H (2022). Short-term effects and economic burden assessment of ambient air pollution on hospitalizations for schizophrenia. Environmental Science and Pollution Research, 29(30), 45449–45460.

[8] Eguchi R, Onozuka D, Ikeda K, Kuroda K et.al (2018) The relationship between fine particulate matter (PM 2.5) and schizophrenia severity. International archives of occupational and environmental health, 91, 613-622.

[9] Shiloh R, Shapira A, Potchter O, Hermesh H et.al (2005) Effects of climate on admission rates of schizophrenia patients to psychiatric hospitals. European Psychiatry, 20(1), 61-64.

[10] Duan J, Cheng Q, Luo X, Bai L, Zhang H, Wang S, et al. (2018). Is the serious ambient air pollution associated with increased admissions for schizophrenia?. Science of the total environment, 644, pp.14-19.

[11] Song R, Liu L, Wei N, Li X, Liu J et al. (2023) Short-term exposure to air pollution is an emerging but neglected risk factor for schizophrenia: A systematic review and meta-analysis. Science of The Total Environment, 854, p.158823.

[12] Gasparrini A (2016) Modelling lagged associations in environmental time series data: a simulation study. Epidemiology, 27, 835-842.

[13] Pan R, Zhang X, Gao J, Yi W, Wei Q et al. (2019). Impacts of heat and cold on hospitalizations for schizophrenia in Hefei, China: an assessment of disease burden. Science of the Total Environment, 694, 133582.

[14] Zhao D, Zhang X, Xie M, Cheng, Zhang H et al. (2016). Is greater temperature change within a day associated with increased emergency admissions for schizophrenia? Science of The Total Environment, 566, 1545-1551.

[15] Liang Z, Xu C, Cao Y, Kan HD, Chen RJ (2019) The association between short-term ambient air pollution and daily outpatient visits for schizophrenia: A hospital-based study. Environmental Pollution. 244, pp.102-108.

[16] Bai L, Yang J, Zhang Y, Zhao D & Su H (2020). Durational effect of particulate matter air pollution wave on hospital admissions for schizophrenia. Environmental Research, 187, 109571.

[17] Crank PJ, Hondula DM, & Sailor DJ (2023). Mental health and air temperature: Attributable risk analysis for schizophrenia hospital admissions in arid urban climates. Science of The Total Environment, 862, 160599.

[18] Qiu X, Danesh-Yazdi M, Wei Y, Di Q et al. (2022). Associations of short-term exposure to air pollution and increased ambient temperature with psychiatric hospital admissions in older adults in the USA: a case–crossover study. The Lancet Planetary Health, 6(4), e331–e341.

[19] Qiu X, Wei Y, Weisskopf M, Spiro A, Shi L, Castro E et al. (2023). Air pollution, climate conditions and risk of hospital admissions for psychotic disorders in U.S. residents. Environmental Research, 216, 114636.

[20] Gu S, Huang R, Yang J, Sun S, Xu Y, et al. (2019) Exposure-lag-response association between sunlight and schizophrenia in Ningbo, China. Environmental Pollution, 247, 285-292.

[21] Almeida JS (2002). Predictive non-linear modeling of complex data by artificial neural networks. Current opinion in biotechnology, 13(1), 72-76.

[22] Larochelle H, Bengio Y, Louradour J, & Lamblin P (2009). Exploring strategies for training deep neural networks. Journal of machine learning research, 10(1).

[23] Mishra S, Bordin C, Taharaguchi K, & Palu I (2020). Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature. Energy Reports, 6, 273-286.

[24] Sung TI, Chen MJ, Lin CY, Lung SC and Su HJ (2011) Relationship between mean daily ambient temperature range and hospital admissions for schizophrenia: results from a national cohort of psychiatric inpatients. Science of the total environment, 410, pp.41-46.

[25] Dolado JJ, Gonzalo J, & Mayoral L (2002). A fractional Dickey–Fuller test for unit roots. Econometrica, 70(5), 1963-2006.

[26] Gasparrini A, Armstrong B and Kenward MG (2010) Distributed lag non-linear models. Statistics in medicine, 29(21), pp.2224-2234.

[27] Mode GR, Hoque KA (2020). Adversarial examples in deep learning for multivariate time series regression. In 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (pp. 1-10). IEEE.

[28] Chung H, & Shin KS (2020). Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction. Neural Computing and Applications, 32, 7897-7914.

[29] Surakhi O, Zaidan MA, Fung PL, Hossein Motlagh N et al. (2021). Time-lag selection for time-series forecasting using neural network and heuristic algorithm. Electronics, 10(20), 2518.

[30] Wan R, Mei S, Wang J, Liu M, & Yang F (2019). Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting. Electronics, 8(8), 876.

[31] Dai H, Huang G, Wang J, Zeng H & Zhou F (2021). Prediction of air pollutant concentration based on one-dimensional multi-scale CNN-LSTM considering spatial-temporal characteristics: A case study of Xi'an, China. Atmosphere, 12(12), 1626.

[32] Nwankpa C, Ijomah W, Gachagan A & Marshall S (2018). Activation functions: Comparison of trends in practice and research for deep learning. arXiv preprint arXiv:1811.03378.

[33] Lu J, Bu P, Xia X, Yao L et al. (2020). A new deep learning algorithm for detecting the lag effect of fine particles on hospital emergency visits for respiratory diseases. IEEE Access, 8, 145593-145600.

[34] Hossain I, Rasel HM, Imteaz MA, & Mekanik F (2020). Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: a case study for Western Australia. Meteorology and Atmospheric Physics, 132, 131-141.

[35] Zhu H, Chen S, Lu W, Chen K, Feng Y et al. (2022). Study on the influence of meteorological factors on influenza in different regions and predictions based on an LSTM algorithm. BMC Public Health, 22(1), 1-17.

[36] Mahmudimanesh M, Mirzaee M, Dehghan A, & Bahrampour A (2022). Forecasts of cardiac and respiratory mortality in Tehran, Iran, using ARIMAX and CNN-LSTM models. Environmental Science and Pollution Research, 29(19), 28469-28479.

[37] Jahan S, Wraith D, Dunne MP, & Naish S (2021). Assessing evidence for seasonality of hospital admissions for schizophrenia in Queensland, Australia: a time series observational study. International journal of biometeorology, 65(12), 2025-2035.