Predicting the Impact of Socio-Demographic Risk Factors on COVID-19 Based on Hybrid ANN-CNN Model

Yahaya Mohammed Sani
Department of Computer Science
Makerere University
Kampala, Uganda
saniy5502@gmail.com

Benjamin Davou Pam
Department of Computer Science
University of Jos
Jos, Nigeria
bengypam@gmail.com

Mamman Adamu
Department of Computer Science
Federal University of Technology
Minna, Nigeria
rakiya85@gmail.com

2023 First International Conference on the Advancements of Artificial Intelligence in African Context (AAIAC) | 979-8-3503-3013-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/AAIAC60008.2023.10465311

***Abstract*—Global health has been greatly influenced by the COVID-19 pandemic, especially in low- and middle-income nations like Nigeria. Despite the catastrophic effects of the pandemic, little is known about how sociodemographic risk factors that affects the number of COVID-19 infections and deaths in Nigeria. Using Spearman heat map correlation analysis, this study examined these parameters and developed a hybrid ANN-CNN model to forecast the influence of sociodemographic characteristics against COVID-19 confirmed cases and mortality cases in Nigeria. The Nigerian COVID-19 confirmed and death cases data from May 1, 2020, to April 30, 2021, as well as sociodemographic risk factor statistics, were the datasets used in this study. The experiment was completed by training and testing the models, and based on MAEs and RMSEs models performance evaluation metrics, the developed Hybrid ANN-CNN model outperformed the other five state-of-the-art machine learning models involving Multiple Linear Regression (MLR), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Long Short Term Memory (LSTM), and Least Absolute Shrinkage and Selection Operator (LASSO). With mean absolute errors of (0.0157) and (0.0181) for confirmed and death cases, respectively, the developed Hybrid ANN-CNN model outperformed the others. Similarly, with RMSEs of (0.0842) and (0.0813) for confirmed and death cases, respectively, the developed Hybrid ANN-CNN model fared better than other models. The developed Hybrid ANN-CNN model can be helpful in tracking and containing pandemic outbreaks, both in the present and the future.**

Keywords*—*Socio-demographic, Multiple Linear Regression, Artificial Neural Network, Convolutional Neural Network, Least Absolute Shrinkage and Selection Operator, Hybrid ANN-CNN, Long Short Term Memory and COVID-19.

I. INTRODUCTION

The coronavirus disease (COVID-19) outbreak in 2019 has resulted in a serious worldwide health catastrophe that continues to dominate agendas for socioeconomic, global health, and intervention programs [13]. The Global South bears a greater percentage of the epidemic's effects, especially in low- and middle-income nations where the rates of the infections and mortality were higher. This was true even though the COVID-19 pandemic had a major and catastrophic effect on developed countries. The higher number of COVID-19 infected cases reported in Nigeria between 2020 and 2021 [13]serves as proof for this. The absence of context-specific evidence about the socio-demographic risk factors for COVID-19 infection and death cases in Nigeria persisted despite the growing COVID-19 burden [6]. It may be possible to manage COVID-19 cases by identifying the sociodemographic risk factors linked to them [17]. Additionally, examining the relationship between COVID-19 outbreaks and sociodemographic characteristics in Nigeria will open the door to the development of a model that will offer practical indicators that have not, as far as we are aware, been examined in the literature [8]. This emphasizes the need for an improved model that makes use of machine learning techniques to monitor and evaluate outbreaks and produce more accurate forecasts. This is due to the fact that modeling the effects of underlying variables and disease mortality is crucial for developing effective control methods for COVID-19 case transmission and hazards. We are motivated in this study, therefore, to provide solution to the aforementioned problem by analyzing the socio-demographic risk factors associated with COVID-19 infection and death cases in Nigeria using Spearman heat map correlation analysis, and developed a hybrid ANN-CNN model that does feature extraction and performs better even with limited dataset since medical records data are limited in supply, particularly, COVID-19 cases in Nigeria. In addition, a formal representation of the model was provided. After completing the experiment by training and testing the model, the performance of the developed Hybrid ANN-CNN model was compared with five other state-of-the art machine learning models, including: Multiple Linear Regression (MLR), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Long Short Term Memory (LSTM) and Least Absolute Shrinkage and Selection Operator (LASSO) respectively. The developed ANN-CNN model outperformed the five other models based on MAEs and RMSEs performance evaluation metrics.

The contributions in this study is as follows:

- A hybrid ANN-CNN model was developed to predict the impact of socio-demographic factors in the spread of COVID-19 cases with feature extraction and better performance under limited dataset, since medical records are scarce, most especially COVID-19.
- The formal representation of the hybrid ANN-CNN model using mathematical expressions was developed.
- An architecture of the Hybrid ANN-CNN model for predicting the impact of socio-demographic factors in the spread of COVID-19 cases was designed.
- The developed hybrid ANN-CNN model outperformed other state-of-the-art models evaluated on MAEs and RMSEs performance comparison metrics.

Thus the reminder of the sections includes; literature review, methodology, results and discussion and conclusion.

II. RELATED WORK

Using real-time PCR testing, Elimian [6] evaluated participants who were screened for COVID-19 positivity and subsequently died between February 27, 2020, and June 8, 2020. In order to find variables that were independently linked with both outcome variables, multivariable logistic regression analyses were carried out. The results were shown as adjusted OR (aOR) and 95\% CI. 10,517 of the 36,496 patients who underwent COVID-19 screening had confirmed cases. 295 deaths were reported among the 3215 confirmed cases with sufficient clinical data. Male sex and older age were factors that were independently linked to testing positive for COVID-19 (aOR 1.11, 95\%CI 1.04-1.18). Individuals who were 51 years of age or older had a higher risk of mortality following COVID-19.

Using LASSO analysis, Lopez [10] conducted a study based on 10,550 confirmed COVID-19 cases reported during the first wave in the municipality of Barcelona. According to the study, older population structures and larger population densities were associated with higher rates of COVID-19 cases. Conversely, the rate of COVID-19 infection was inversely correlated with the proportion of residents with a secondary education and the population that was born in nations with high Human Development Index (HDI).

Using a logistic regression model, Peres [16] examined the sociodemographic variables linked to COVID-19 in-hospital mortality in Brazil. 228,196 adult hospitalized COVID-19 positive patients with a predetermined outcome made up the study population. After controlling for education level, region of residence, and comorbidities, there was an independent correlation between increased in-hospital mortality and black/brown race (odds ratio ¼ 1.15; 95\% confidence interval ¼ 1.09e1.22). Black and brown patients showed reduced hospital resource utilization and increased in-hospital mortality among hospitalized Brazilian adults with COVID-19.

From March 2020 to December 2020, Sohrabi [17] investigated the sociodemographic factors associated with COVID-19-positive patients in the Tehran province. The sociodemographic information of 205,654 patients was examined. A multiple logistic regression model and the chi-squared test were employed to evaluate the relationship between the study variables and the disease severity. Male sex and older age were linked to worse outcomes. These results should alert medical professionals to the risk variables linked to a bad prognosis for COVID-19 patients and assist policymakers in identifying susceptible populations for public health campaigns aimed at halting the disease's spread and fatalities.

From late February to early June 2020, Birenbaum [2] looked at the macro-level trends of the first wave of COVID-19 occurrence in Israel. Three socio-demographic variables—socioeconomic status, age, and population density were the focus of the investigation. Morbidity rates were positively associated with population density in both descriptive statistics and regressions. Both the socio-economic status and the size of the elderly population were found to be significantly related to the morbidity rate.

The previous studies examined to determine the effect of socio-demographic factors on COVID-19 transmissions, but unfortunately, did not take into account the use of machine learning models that specialize in feature extraction based on feature importance to handle the non-linear relationship between socio-demographic factors and COVID-19 cases. Additionally, none of the earlier research employed a hybrid ANN-CNN to predict the effect of socio-demographic on COVID-19 cases with enhanced model performance, while resolving the issue of a limited dataset connected with medical record data, particularly COVID-19. This gives our research a strong basis to address the gaps found, which is what this study set out to achieve.

III. METHODOLOGY

A. Dataset Collection and Data Description

COVID-19 daily confirmed cases, daily death cases, and socio-demographic variables in Nigeria were all collected from multiple sources in this study. The datasets but not limited to: Nigeria COVID-19 confirmed cases with a total of 90,786 and death cases datasets with a total of 990 deaths \cite{covid19microsite}, socio-demographic variables which consists of: unemployment data [11], population density and area in square kilometre data [5], literacy level, poverty head count [18]. All datasets were collected for six Nigerian states, with one representing each of Nigeria's six geopolitical zones, namely Adamawa-North East, FCT-North-Central, Kano-North West, Enugu-South East, Rivers-South South, and Lagos-South West respectively.

B. Normalization and Feature Scaling Techniques

In order to facilitate the training process, the test data were inversely transformed and the training data were rescaled using a proposed two-step feature scaling method, which is a combination of max absolute values and minmax feature scaling techniques. To put it simply, the maxabs scaler divides each value by the maximum value after taking the absolute maximum value of each column. The data is scaled using this approach between [-1, 1]. As an example, minmax scales values between -1 and 1, where -1 is the lowest value and 1 is the maximum value [15].

C. Training and Testing Data Split

Eighty percent of the datasets utilized in this study were allocated for testing, while the remaining twenty percent were used for training.

D. Existing Machine Learning Models

Developing a model that performs well in a new and unexplored data set is the main objective of any regression endeavor. To effectively generalize to test data, a machine learning model needs to learn the training set without overfitting [7].

1) Multiple Multiple Linear Regression

Severally, a supervised learning technique used in machine learning and general data science is the multiple linear regression model. Although multiple linear regression models have been around for a while, a number of recent developments have significantly expanded their application [7]. One drawback of this technique is that linear regression may perform appallingly when there are outliers in the data. Poor performance may also result from the relationship between predictor and target variables, which is based on the assumption that they always have the same relationship.

The equation for multiple linear regression looks like this:

$$y = \beta \circ + \beta 1x1 + \beta 2x2 + ,........, + \beta pxp + \varepsilon \quad (1)$$

The unknown parameter β0 can be referred to as the regression constant, and the regression coefficients are β1……where β0, β1 are p +1. The dependent variable is Y, and p is x1, x2,...,xp. Independent variables are general variables that are simultaneously measurable and controllable \cite{xiao2021}.

2) Least Absolute Shrinkage and Selection Operator (LASSO)

By performing both variable selection and regularization, regression analysis with LASSO enhances the statistical regression model's predictability and interpretability. When there are many independent variables, we usually wish to isolate the smaller subset with the largest effects. The main advantage of LASSO sparsity is that it makes interpretation easier to understand. One major downside of LASSO is that it often selects only one variable from a set of tightly connected variables since it lacks a grouping attribute.

$$\sum_{i=1}^{n}\left(yi-\sum_{j}xi_{j}\beta_{j}\right)^{2}+\lambda\sum_{j=1}^{p}|\beta_{j}| \qquad (2)$$

The L1 penalty's strength is controlled by λ. Essentially, λ represents the amount of shrinkage: There are no dropped parameters if λ = 0. The estimate and the result of linear regression are equivalent. The number of coefficients that are set to zero and removed increases as λ rises; in theory, all coefficients are destroyed when λ = ∞. The bias rises as λ increases. The variance rises as λ falls. In this work, the LASSO employed λ = 0.0001 [19].

3) Artificial Neural Network (ANN)

There are numerous benefits when ANN computational methods are contrasted with traditional machine learning models. In addition to learning an input-output mapping from a learner, an ANN's nonlinear components allow it to adjust its synaptic weights to better fit its surroundings. It can also manage partial information and react when there is uncertainty. When using ANN models, it can be difficult to decide on the suitable training data attributes, the network architecture (number of layers and nodes), and the overfitting prevention technique.

In mathematical terms, an ANN neuron is explained by the following equations:

$$\sum_{j=0}^{n}w_{kj}x_{j} \qquad (3)$$

In this case, x0=1 and w0 are regarded as the bias [14], [1]. In this study, artificial neural network (ANN) had three layers: 64, 16, and 1. Its activation function was ReLU, loss was equal to MAE, batch size was 32, and it was stopped after 100 epochs.

4) Convolutional Neural Network (CNN)

CNNs are a subset of deep learning models that have been popular in many machine learning applications and are drawing interest from a variety of domains, including nonlinear regression. The three "building blocks" of conventional CNNs are convolution, pooling, and fully linked layers. The convolution and pooling layers in layers one and two extract features, while the fully connected layer in layer three translates the extracted features to the output, such as linear and nonlinear regression. The CNN, which is made up of several mathematical operations including convolution, a specific kind of linear operation, requires a convolution layer in order to function. Compared to generic artificial neural networks, CNN has several advantages, such as: local connections efficiently reduce parameters and speed up convergence by restricting each neuron's connections to a small subset of neurons from the preceding layer. Additional reductions in parameters and dimensionality are achieved via weight sharing, which permits connections to share the same weights.

$$Y_{j}=f\left(\sum_{i=1}^{n}\left(xj*wij\right)+bj\right) \qquad (4)$$

The input feature xi is represented by the letter wij, which is the weight value of the link between the input feature xi and the neuron j. n features are delivered to the neuron j concurrently. The bias value, bj, represents the intrinsic state of neuron j. Yj is neuron j's output. The activation function, also called the tanh (x) Rectified Linear Unit Function, is represented by the notation f(cdot). Even though CNN is extensively used and has many benefits, convolution is challenging to handle because of several issues, such as poor generalization, a lack of equivariance, and poor performance in cluttered environments [9], [20].

This study's CNN, ConV1D, employed a Dense of three layers (64, 16, and 1), flattened, with ReLU as the activation function, loss = MAE, batch batch size=32, and ended the experiment after 100 epochs.

5) Long Short Term Memory(LSTM)

Specialized deep neural network design served as the foundation for long short-term memory (LSTM) models, which have become an important model for both time-series and non-time series forecasting. Additionally, non-linearity can be learned via LSTM. Since the learning parameter cells of LSTMs are computationally expensive, a significant disadvantage of LSTMs is that they need more training data to learn more efficiently [12], [3].

The functions of the LSTM units are represented by the following equations:

$$f(t)=sigm\left(WfX(t)+Ufh(t-1)+bf\right) \qquad (5)$$

$$I(t)=sigm\left(WIX(t)+UIh(t-1)+bI\right) \qquad (6)$$

$$\bar{C}(t)=\tanh\left(WCX(t)+UCh(t-1)+bC\right) \qquad (7)$$

$$C(t)=f(t)\odot C(t-1)+l(t)\odot\bar{C}(t) \qquad (8)$$

$$\circ(t)=sigm\left(W\circ X(t)+U\circ h(t-1)+b\circ\right) \qquad (9)$$

$$h(t)=\circ(t)\odot\tau\left(C(t)\right) \qquad (10)$$

W, U, and b stand for the weight matrices for the inputs, outputs, hidden layer, and bias vector in equations (5)–(10). The element-wise multiplication operation in the same equations is denoted by (_). The two activation functions are the sigmoid (sigm) and the hyperbolic tangent (tanh) [3]. This study's LSTM had a batch_size of 32, a loss of MAE, a Dense of 64, 16, and 1, and an activation function of ReLU. The experiment terminated after 100 epochs.

E. *The Proposed Sequential Hybrid ANN-CNN Model for Predicting Socio-Demographic Risk factors Against COVID-19 Cases*

The proposed sequential Hybrid ANN-CNN model is an integration of ANN and CNN models to form a consistent whole model. Figure 1 below depicts the architecture of the developed Hybrid ANN-CNN model. Firstly, ANN which is a critical component the proposed Sequential Hybrid ANN-CNN model pride itself for carrying out robust non-linear machine learning operation and even with small amount of data. On the other hand, CNN carries out feature extraction, and the fully connected layer, for both linear and nonlinear regression, maps the extracted features to the final output. Both ANN and CNN perform well with both spatial and time series data, and each neuron in CNN is no longer connected to every other neuron in the preceding layer, but only to a select few neurons. This reduces parameters, speeds up convergence, and facilitates weight sharing. These properties inherent in the two models overcome the drawbacks of the individual models discussed above with good performance as shown in Figure 1.

The mathematical equation for the proposed Sequential Hybrid ANN-CNN is as follows:

$$X(S) \rightarrow Y \tag{11}$$

$$X(S) = (x1s, x2s, \ldots.., xns) \tag{12}$$

$$Y = (y1, y2, \ldots.., yn) \tag{13}$$

Where X(s) and Y are the proposed Sequential Hybrid ANN-CNN inputs and outputs variables respectively.

Suppose:

X(S), represents the socio-demographic variables involving five parameter which are: unemployment, population density, area in square kilometre and poverty head count as independent variables while Y represents the corresponding dependent variables, which are the confirmed and death cases.

The proposed Sequential Hybrid ANN-CNN network was trained using a COVID-19 confirmed and death cases and socio-demographic datasets to predict the impact of socio-demographic variables consisting of five unemployment, population density, area in square kilometre and poverty head count as independent variables against COVID-19 confirmed and death cases. The network architecture was implemented using neurons and ConV1D layer. The network consists of three linear layers of 64, 16 and 1 units respectively. The training was performed with the ADAM optimiser and the activation function was ReLU for non-linear operation, kernel_regularizer = L1 to overcome the problem of overfitting, loss = MAE, batch_size=32 and the experiment was terminated after 100 epochs. Eighty percent of the datasets samples were utilized for training, while the remaining twenty percent were used for testing for each dataset. To estimate a set of future case predictions, the testing process was repeated in the manner described below:

$$\hat{Y} = Estimate(x1si, xs2si, \ldots., xnsi), 1 \leq i \leq N \tag{14}$$

Where N denote the test data.

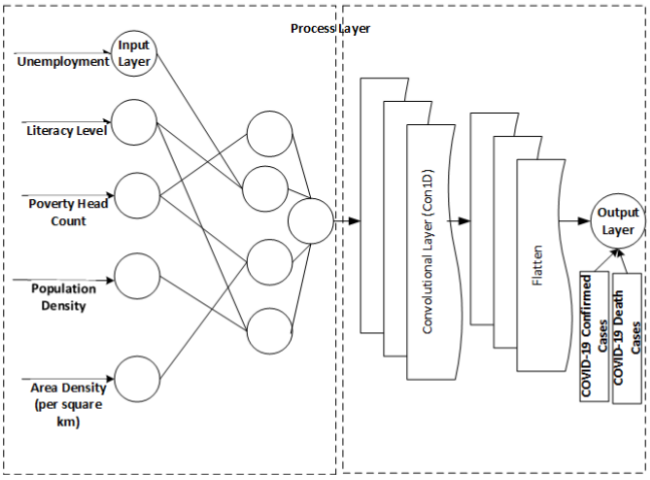


Fig. 1. Hybrid ANN-CNN Model

F. *Performance Evaluation*

As for the loss functions used to measure models performance, Mean Absolute Error (MAE) was used and the mathematical equation is as follows:

$$\frac{1}{N}\sum_{i=1}^{D} | xi - x | \tag{14}$$

where, N = the number of errors, $\sum$ = summation symbol and |xi - x| = the absolute errors.

And Root Mean Square Error (RMSE) which is the square root of Mean Square Error (MSE) was also used to further the depth of error analysis and to justify the good performance of the developed sequential Hybrid ANN-CNN model over MLR, ANN, CNN, LSTM and LASSO respectively. The mathematical equation for RMSE is given below:

$$\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}i)^2}{N}} \tag{15}$$

IV. Results and Discussion

A. *Spearman Heat Map Correlation Analysis*

Adopting Spearman's heat map correlation analysis in this study, the correlation analysis results were revealed as follows: in Enugu states, confirmed cases was found to be weakly positively correlated with unemployment with p-value of (0.0011), literacy (0.0010), area (km2) (0.001), population density (in h%km2) (0.001). Similarly, deaths was also weakly correlated with unemployment (0.0006), literacy level, area (km2) (0.0004), population density (in h\%km2) (0.0006). In the FCT, confirmed cases were positively correlated with unemployment with a p-value of (0.27), literacy (0.24), negatively correlated with area (km2) (-0.26), positively correlated with population density (in h\%km2) (0.27) and negatively correlated with poverty headcount (-0.2). Similarly, deaths were weakly correlated with unemployment with a p-value of (0.038), literacy (0.003), negatively correlated with area (km2) (-0.031), weakly positively correlated with population density (in h\%km2) (0.0038) and weakly negatively correlated with poverty headcount. Also for Kano State, confirmed cases were weakly positively correlated with unemployment with a p-value of (0.12), negatively correlated with literacy (-0.1), weakly positively correlated with area (km2) (0.1), weakly positively correlated with population density (in h%km2) (0.11). Similarly, deaths were weakly correlated with unemployment with a p-value of (0.022), negatively correlated with literacy level (-0.021), weakly correlated with area (km2) (0.02), positively correlated with population density (in h%km2) (0.02) and weakly negatively correlated with poverty headcount.

In Lagos State, confirmed cases were positively correlated with unemployment with p-value of (0.25), literacy (0.2), negatively correlated with area (km2) (-0.24) and positively correlated with population density (in h\%km2) (0.23). while deaths cases were weakly correlated with unemployment with a p-value of (0.004), literacy (0.004), negatively correlated with area (km2) (-0.003), weakly positively correlated with population density (in h%km2) (0.004), and weakly negatively correlated with poverty headcount (-0.003). Finally, for Rivers State, confirmed cases were positively correlated with unemployment with a p-value of (0.11), area (km2) (0.1), negatively correlated with population density (in h^km2) (-0.12) and weakly correlated with poverty headcount (0.01). Deaths were very weakly positively correlated with unemployment with a p-value of (0.07), area in (km2) (0.077), negatively correlated with population density (in h%km2) (-0.072) and weakly correlated with poverty headcount (0.073).

TABLE I. SOCIO-DEMOGRAPHIC FACTORS VS CONFIRMED CASES (MAE)

State	Linear	ANN	CNN	LSTM	LASSO	Hybrid
Adamawa	0.0278	0.0267	0.0282	0.0280	0.0278	**0.0157**
Enugu	0.1165	0.1198	0.1205	0.1151	0.1164	**0.1125**
FCT	0.0976	0.1081	0.1313	0.1060	0.0962	**0.0949**
Kano	0.1015	0.1079	0.1269	0.1010	0.0977	**0.0971**
Lagos	0.0988	0.1227	0.1323	0.1083	0.0981	**0.0944**
Rivers	0.0962	0.1030	0.1146	0.1095	0.0957	**0.0956**

Table I shows the results of predicting of socio-demographic risk factors against COVID-19 confirmed cases for the six states which includes but not limited to, Adamawa, Enugu, FCT, Kano, Lagos and Rivers states with six machine learning models: the developed Hybrid ANN-CNN, MLR, ANN, CNN, LSTM and LASSO and their performance compared based on MAEs, with Hybrid ANN-CNN performing better in all the states. The developed Hybrid ANN-ANN performed better than the other other models for all the six states with the MAEs of (0.0157) for Adamawa, Enugu (0.1125), FCT (0.0949), Kano (0.0971), Lagos (0.0944) and Rivers state (0.0956).

TABLE II. SOCIO-DEMOGRAPHIC FACTORS VS DEATH CASES (MAE)

State	Linear	ANN	CNN	LSTM	LASSO	Hybrid
Adamawa	0.0438	0.0418	0.0438	0.0387	0.0438	**0.0181**
Enugu	0.0340	0.0217	0.0484	0.0250	0.0226	**0.0216**
FCT	0.0941	0.0710	0.0720	0.0696	0.0763	**0.0690**
Kano	0.0724	0.0742	0.0707	0.0718	0.0724	**0.0571**
Lagos	0.0135	0.0206	0.0149	0.0138	0.0135	**0.0134**
Rivers	0.1076	0.1111	0.1295	0.1062	0.1044	**0.1027**

Secondly, Table II showed the results of predicting socio-demographic risk factors against COVID-19 deaths for the six states including Adamawa, Enugu, FCT, Kano, Lagos and Rivers states with six machine learning models. The developed Hybrid ANN-CNN, MLR, ANN, CNN, LSTM and LASSO based on MAEs, with the developed Hybrid ANN-CNN outperforming the other five models in all the states. With an MAEs of (0.0181) for death cases in Adamawa, Enugu (0.0216), FCT (0.0690), Kano (0.0571), Lagos (0.0134), and Rivers State (0.1027). The developed hybrid ANN-ANN model outperformed the other models in all the six states.

TABLE III. SOCIO-DEMOGRAPHIC FACTORS VS CONFIRMED CASES (RMSE)

State	Linear	ANN	CNN	LSTM	LASSO	Hybrid
Adamawa	0.0495	0.0496	0.0515	0.0495	0.0488	**0.0482**
Enugu	0.1753	0.1648	0.1694	0.1742	0.1752	**0.1570**
FCT	0.1331	0.1425	0.1797	0.1423	0.1331	**0.1111**
Kano	0.1428	0.1544	0.1401	0.1413	0.1390	**0.1362**
Lagos	0.1610	0.1653	0.1630	0.1690	0.1610	**0.1343**
Rivers	0.1220	0.1463	0.1371	0.1302	0.1285	**0.1217**

Thirdly, Table III showed the outcomes of predicting socio-demographic risk factors against COVID-19 confirmed cases for the six states of Adamawa, Enugu, FCT, Kano, Lagos, and Rivers using six machine learning models, the developed Hybrid ANN-CNN, MLR, ANN, CNN, LSTM, and LASSO based on RMSEs, with the proposed Hybrid ANN-CNN performing better in all states. For all the six states, the developed hybrid ANN-ANN outperformed the other five models, with RMSEs of (0.0842) for Adamawa, Enugu (0.1570), FCT (0.1111), Kano (0.1362), Lagos (0.1343), and Rivers State (0.1217).

TABLE IV. SOCIO-DEMOGRAPHIC FACTORS VS DEATH CASES (RMSE)

State	Linear	ANN	CNN	LSTM	LASSO	Hybrid
Adamawa	0.0907	0.0910	0.0909	0.0853	0.0907	**0.0813**
Enugu	0.1736	0.0383	0.0650	0.0388	0.0387	**0.0365**
FCT	0.1331	0.0843	0.1150	0.0845	0.0842	**0.0842**
Kano	0.1416	0.1408	0.1450	0.1460	0.1416	**0.1321**
Lagos	0.0233	0.0216	0.0237	0.0231	0.0232	**0.0210**
Rivers	0.1305	0.1559	0.1313	0.1385	0.1376	**0.1291**

Finally, Table IV shows he results of six machine learning models that predicted socio-demographic risk factors on COVID-19 death cases for the six states of Adamawa, Enugu, FCT, Kano, Lagos, and Rivers. The developed Hybrid ANN-CNN, MLR, ANN, CNN, LSTM, and LASSO models were used and their performance compared based on RMSE, with the Hybrid ANN-CNN outperforming the others in every state. Furthermore, for all the six states, the developed hybrid ANN-CNN outperformed the other five models, with RMSEs of (0.0813) for Adamawa, Enugu (0.0365), FCT (0.0842), Kano (0.1321), Lagos (0.0210), and Rivers State (0.1291) respectively.

B. Discussion

The coronavirus disease (COVID-19) outbreak in 2019 has resulted in a serious global health issue that continues to rule agendas for global health, socioeconomic policy, and intervention initiatives, according to [13]. The Global South carries a larger share of the epidemic's expenses, particularly in terms of morbidity and mortality rates in low-income and middle-income countries, notwithstanding the COVID-19 outbreak's enormous and devastating consequences on developed nations. The notable rise in COVID-19 infection cases that Nigeria reported between 2020 and 2021 [13] provides proof to this. Despite the COVID-19 burden, there is insufficient evidence to determine the sociodemographic risk factors for COVID-19 infections and subsequent mortality in Nigeria. [6]. In order to manage COVID-19 pandemic, it is important to identify socio-demographic factors that are related to them [17].

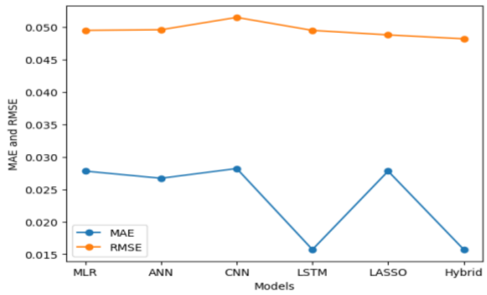
In this study we adopted Spearman's heat map correlation analysis and the developed Sequential Hybrid ANN-CNN model. We were motivated to look into the socio-demographic risk factors connected to mortality and COVID-19 positive cases in the Nigerian population. The developed Hybrid ANN-CNN model performance was compared with five other machine learning models, which includes, ANN, CNN, LSTM, and LASSO, after the experiment was completed by training and testing the models based on the MAE and RMSE performance criteria, the developed Hybrid ANN-CNN model outperformed the five other state-of-the-art models.

Similarly, in this study, in Enugu states, confirmed cases was found to be weakly positively correlated with unemployment, literacy level, area (km2), population density (in h\%km2). Similarly, for deaths, was also weakly correlated with unemployment, literacy level, area (km2), population density (in h\%km2). In the FCT, confirmed cases were positively correlated with unemployment, literacy, negatively correlated with area (km2), positively correlated with population density (in h\%km2) and negatively correlated with poverty headcount.

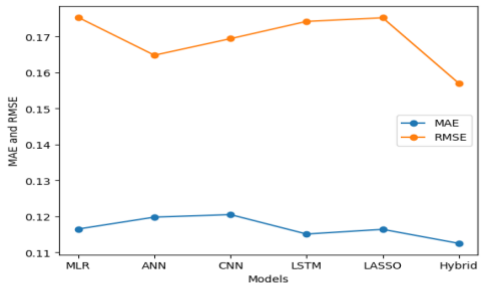
Deaths were weakly correlated with unemployment, literacy, negatively correlated with area (km2), weakly positively correlated with population density (in h\%km2) and weakly negatively correlated with poverty headcount. Also for Kano State, confirmed cases were weakly positively correlated with unemployment, negatively correlated with literacy level, weakly positively correlated with area (km2), weakly positively correlated with population density (in h\%km2). Similarly, deaths were weakly correlated with unemployment, negatively correlated with literacy, weakly correlated with area (km2), positively correlated with population density (in h\%km2) and weakly negatively correlated with poverty headcount.

In Lagos State, confirmed cases were positively correlated with unemployment, literacy level, negatively correlated with area (km2), positively correlated with population density (in h\%km2). Deaths were weakly correlated with unemployment, literacy, negatively correlated with area (km2), positively correlated with population density (in h\%km2), and weakly negatively correlated with poverty headcount. Finally, for Rivers State, confirmed cases was positively correlated with unemployment, area (km2) (0.1), negatively correlated with population density (in h\^km2) and weakly correlated with poverty headcount. Deaths were weakly correlated with unemployment, area in (km2), negatively correlated with population density (in h\%km2) and weakly correlated with poverty headcount. Finally, on models performance in the prediction of socio-demographic risk factors against COVID-19 confirmed and death cases, Hybrid ANN-CNN outperformed MLR, ANN, CNN, LSTM and LASSO in all the six states studied based on MAE and RMSE performance evaluation metrics.

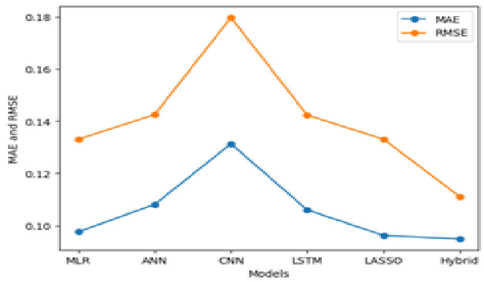
Finally, the developed Hybrid ANN-CNN Model contributions centres around: predicting the impact of socio-demographic factors on the spread of COVID-19 cases with feature extraction and better performance with limited dataset, since medical records are scarce. The formal representation of the hybrid ANN-CNN model using mathematical expressions was developed. An architecture of the developed Hybrid ANN-CNN model for predicting the impact of socio-demographic factors on the spread of COVID-19 cases was designed. The developed hybrid ANN-CNN model outperformed the five other state-of-the-art models evaluated based on MAE and RMSE performance comparison metrics.



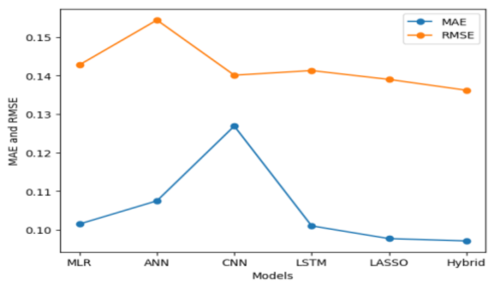
(a) Adamawa



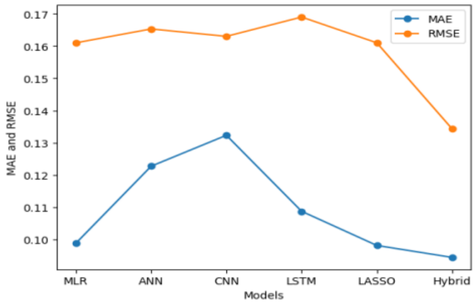
(b) Enugu



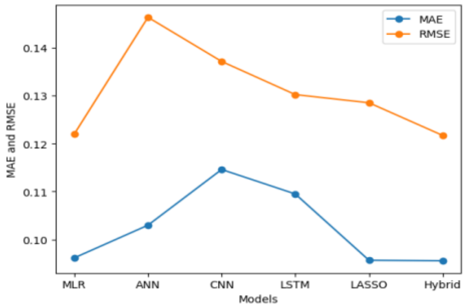
(c) FCT



(d) Kano

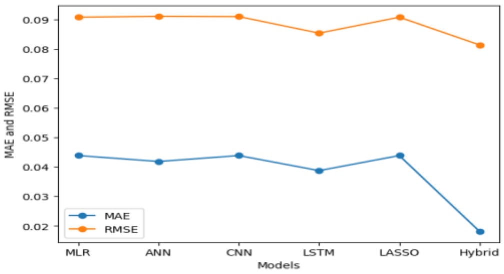


(e) Lagos

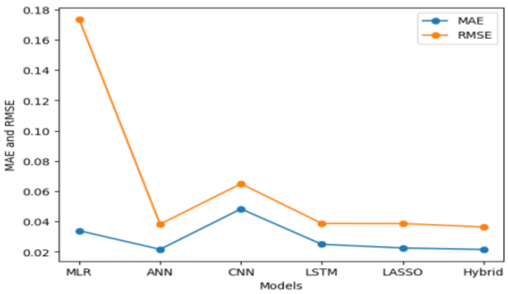


(f) Rivers

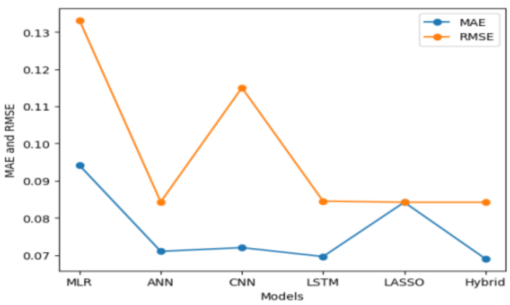
Fig. 2. Showing Models Evaluation Performance Comparison for Confirmed Cases for Adamawa, Enugu, FCT, Kano, Lagos and Rivers for Based on MAEs and RMSEs



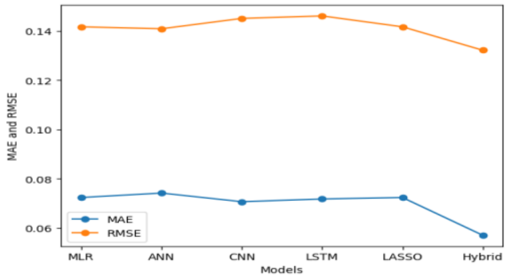
(a) Adamawa



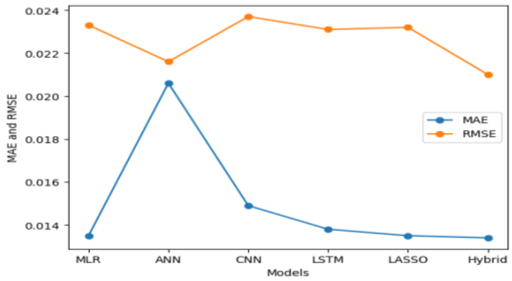
(b) Enugu



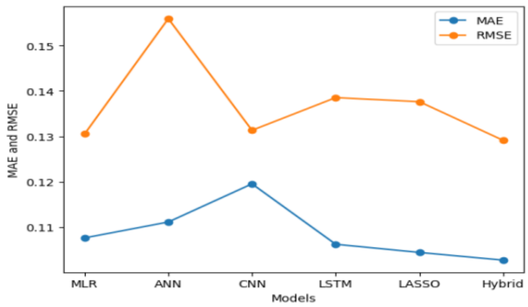
(c) FCT



(d) Kano



(e) Lagos



(f) Rivers

Fig. 3. Showing Models Evaluation Performance Comparison for Death Cases for Adamawa, Enugu, FCT, Kano, Lagos and Rivers for Based on MAEs and RMSEs.

V. Conclusion

We developed a Hybrid ANN-CNN model in this study to predict the effect of sociodemographic risk factors against COVID-19 confirmed and fatality cases. We then evaluated its performance with five other machine learning models, including MLR, ANN, CNN, LSTM, and LASSO. The models' performance was evaluated by comparing them using the MAE and RMSE evaluation criteria. In order to achieve this, the developed Hybrid ANN-CNN beat the other five models at the conclusion of the experiment based on MAE and RMSE for each of the six states, which are Adamawa, Enugu, FCT, Kano, Lagos, and Rivers. This was done in light of the results displayed in Tables I through IV above. The developed Hybrid ANN-CNN Model has helped predict the impact of socio-demographic factors on the spread of COVID-19 cases with feature extraction and better performance under limited dataset, since medical records data are scarce. These are the main contributions made by the study, which can be seen in the results obtained. A formal mathematical expression representation of the hybrid ANN-CNN model was developed. The produced Hybrid ANN-CNN model had an architecture that was designed. To the best of our knowledge, no research has, however, used the Hybrid ANN-CNN model from the literature that is currently accessible and achieved a higher performance in predicting the influence of sociodemographic risk factors against COVID-19 confirmed and mortality cases in Nigeria. Policy makers and epidemiologists can benefit from the study's findings by using the prediction model and sociodemographic risk indicators linked to COVID-19 cases to identify potential alert systems for managing pandemic outbreaks in the present and the future.

REFERENCES

[1] Bikku, T.: Multi-layered deep learning perceptron approach for health risk prediction. Journal of Big Data 7(1), 1-14 (2020)

[2] Birenbaum-Carmeli, D., Chassida, J.: Covid-19 in israel: sociodemographic characteristics of first wave morbidity in jewish and arab communities. International Journal for Equity in Health 19(1), 1-13 (2020)

[3] Bolboac˘a, R., Haller, P.: Performance analysis of long short-term memory predictive neural networks on time series data. Mathematics 11(6), 1432 (2023)

[4] COVID, N.C.: Microsite. covid-19 nigeria. 2020,[cited 4 september,2020] (19)

[5] EGIDI, S.A., USHIE, C.A.: Population distribution, density and developmentindicators in nigeria: The cross river state example.

[6] Elimian, K.O., Ochu, C.L., Ebhodaghe, B., Myles, P., Crawford, E.E., Igumbor, E., Ukponu, W., Olayinka, A., Aruna, O., Dan-Nwafor, C.,et al.: Patient characteristics associated with covid-19 positivity and fatality in nigeria: retrospective cohort study. BMJ open 10(12), e044079 (2020)

[7] Emmert-Streib, F., Dehmer, M.: High-dimensional lasso-based computational regression models: regularization, shrinkage, and selection.Machine Learning and Knowledge Extraction 1(1), 359-383 (2019)

[8] Hasan, S.M., Anisha, A.M., Adnin, R., Eliza, I.J., Tarin, I., Afroz, S.,Islam, A.A.A.: Revealing influences of socioeconomic factors over disease outbreaks. In: ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS). pp. 490 - 512 (2022)

[9] Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J.: A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems (2021)

[10] L´opez-Gay, A., Spijker, J., Cole, H.V., Marques, A.G., Triguero-Mas, M., Anguelovski, I., Mar´ı-Dell' Olmo, M., Mo´denes, J.A., A´ lamo-Junquera, D., L´opez-Gallego, F., et al.: Sociodemographic determinants of intraurban variations in covid-19 incidence: the case of barcelona. J Epidemiol Community Health 76(1), 1-7 (2022)

[11] MAKINDE, L.O., ADEGBAMI, A.: Unemployment in nigeria. Ilorin Journal of Administration and Development 5(2), 71-77 (2019)

[12] Malakar, S., Goswami, S., Ganguli, B., Chakrabarti, A., Roy, S.S.,Boopathi, K., Rangaraj, A.: Designing a long short-term network forshort-term forecasting of global horizontal irradiance. SN Applied Sciences 3, 1-15 (2021)

[13] Mansour, S., Abulibdeh, A., Alahmadi, M., Ramadan, E.: Spatial assessment of covid-19 first-wave mortality risk in the global south. The Professional Geographer 74(3), 440-458 (2022)

[14] Mas, J.F., Flores, J.J.: The application of artificial neural networks to the analysis of remotely sensed data. International Journal of Remote Sensing 29(3), 617-663 (2008).

[15] Pandey, A., Jain, A.: Comparative analysis of knn algorithm using various normalization techniques. International Journal of ComputerNetwork and Information Security 11(11), 36 (2017)

[16] Peres, I.T., Bastos, L., Gelli, J.M., Marchesi, J., Dantas, L., Antunes, B.,Mac , aira, P., Bai ˜ ao, F., Hamacher, S., Bozza, F.A.: Sociodemographic factors associated with covid-19 in-hospital mortality in brazil. Public Health 192, 15-20 (2021)

[17] Sohrabi, M.R., Amin, R., Maher, A., Bahadorimonfared, A., Janbazi,S., Hannani, K., Kolahi, A.A., Zali, A.R.: Sociodemographic determinants and clinical risk factors associated with covid-19 severity: a cross-sectional analysis of over 200,000 patients in tehran, iran. BMC infectious diseases 21(1), 474 (2021)

[18] National Bureau of Statistics (NBS), N.B.: Poverty and inequality in nigeria (2020)

[19] Xiao, Y., Jin, Z.: The forecast research of linear regression forecast model in national economy. Open Access Library Journal 8(8), 1-17 (2021)

[20] Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K.: Convolutional neural networks: an overview and application in radiology. Insights into imaging 9, 611-629 (2018).