

COX Regressive Winsorized Correlated Convolutional Deep Belief Boltzmann Network for Covid-19 Prediction with Big Data

Dr. K.Sankar¹, Dr. Divya Rohatgi² and S.Balakrishna Reddy³

^{1,3}CVR College of Engineering, Dept. of ET, Hyderabad, India

Email: sankarkrish@cvr.ac.in, sama.balakrishnaireddy@gmail.com

²Amity University, Dept. of CSE, Mumbai, India

Email: divi.rohatgi@gmail.com

Abstract—Big data analytics in the health care industry is a very promising method of integrating, discovering, and analyzing huge volume of complex heterogeneous data. Some of the techniques were not well suited for medical data classification with minimum time consumption. To improve disease prediction accuracy, a novel technique Cox Regressive Winsorized Correlated Convolutional Deep Belief Boltzmann Network (CRWCCDBBN) is introduced. The proposed CRWCCDBBN technique uses multiple layers for performing the different processes like feature selection and classification. Initially, feature selection is performed using Cox Regression to minimize the dimensionality of the data by finding the Cox partial log-likelihood between the two features. After the feature selection, the classification is performed using Winsorized Correlation coefficient by analyzing the training and testing disease data. Based on the classification results, Covid 19 disease is correctly diagnosed with a minimum error by updating the weight value. Finally, the gradient descent first-order iterative function is used to find the local minimum error. Experimental evaluation of CRWCCDBBN technique is carried out using a Covid 19 Dataset with different performance metrics such as accuracy, precision, recall, f-measure, and prediction time with respect to a number of patient data. The observed result proves that the proposed CRWCCDBBN technique achieves higher prediction accuracy with a minimum time consumption than the state-of-the-art methods.

Index Terms— Big Data, Convolutional Deep Belief Boltzmann Network, Cox Regression, feature selection, Winsorized Correlation coefficient, gradient descent first-order iterative function.

I. INTRODUCTION

A novel virus disease called COVID-19 caused in the world with various problems. But there are still limitations in the real-time detection of COVID-19 such as a large volume of datasets, imbalance classes, and a misclassification rate of models. Therefore, a novel deep learning-based model aims to investigate for effective detection of COVID-19. A hybrid deep learning prediction model CNN-LSTM was introduced in [1] to correctly forecast the COVID-19. However, it failed to test the algorithm on the large time-series datasets to determine suitability and correctness. Deep-LSTM ensemble model was introduced in [2] to predict the Covid-19. But it was not efficient to provide better results with minimum time. Supervised machine learning models were

developed in [3] for accurate Prediction of COVID-19 Infection. But the deep feature learning was not performed to improve the prediction performance. Three algorithm selection models were developed in [4] to make a COVID-19 predictive risk score. However, the accurate prediction was not performed. Three different approaches based on integrating the multi-head attention, CNN, and LSTM with the Bayesian optimization algorithm were introduced in [5] to predict the COVID-19 time-series data. But it failed to extract the informative features from time-series data. Machine learning models were introduced in [6] for predicting the COVID-19. However, the deep learning models were not applied for accurate COVID-19 disease prediction. A hybrid model was developed in [8] based on artificial intelligence and an expert system for COVID-19 Evolution. But, it failed to minimize the error rate. A dynamic hybrid model was developed in [9] to automatically predict COVID-19 pandemic. But it failed to use an effective tool for prediction and analysis of the dynamics of the COVID-19 pandemic. A federated machine learning model was introduced in [10] for COVID-19 prediction. However, the designed federated models were not accurate.

A. Contributions of the article

In order to overcome the existing issues, a novel CRWCCDBBN technique is developed with the following contribution:

- A novel CRWCCDBBN technique is developed to enhance the COVID-19 prediction with help of feature selection and classification.
- A Cox Regression is applied to Convolutional Deep Belief Restricted Boltzmann Network to minimize the dimensionality of the data by selecting the relevant feature from the big dataset based on Cox partial log-likelihood measure. This helps to minimize the disease prediction time.
- To improve the accuracy and minimize the error rate, Winsorized Correlation coefficient is applied to a Convolutional Deep Belief Restricted Boltzmann Network by analyzing the training and testing disease data. Finally, the Heaviside step activation function is applied to analyze the correlation value and predict the disease with higher accuracy.
- At last, experiments were conducted to estimate the performance of the proposed CRWCCDBBN technique along with conventional deep learning methods. Moreover, the CRWCCDBBN technique is tested on novel COVID 19 datasets based on different performance metrics.

II. LITERATURE REVIEW

Depth wise separable convolution neural network (DWS-CNN) with a Deep support vector machine (DSVM) was introduced in [11] for COVID-19 classification. However, it failed to attain maximum classification results. A binary classifier based on stacking ensemble model with deep neural networks was developed in [12] for prediction of post-COVID-19 infection. However, the various feature selection methods were not implemented to handle the high-dimensional data. Long Short-Term Memory (LSTM) based approach was introduced in [13] for COVID-19 prediction. But the performance accuracy was not focused on. A robust IoT-based deep learning model was introduced in [14] for accurate prediction of COVID-19. But time complexity of accurate prediction was not focused. The Deep Learning model was introduced in [15] to predict the COVID-19. But the complexity involved COVID-19 prediction was not performed. A novel deep learning framework with CNN and stacked Bi-GRU was developed in [16] for predicting and analysing COVID-19 cases. However, the designed approach failed to predict the COVID-19. Social Network Services Big Data Analysis Framework was developed in [17] for COVID-19 Outbreak Prediction. But accurate prediction was not focused. Ensemble learning, autoregressive, and moving regressive (EAMA) model was introduced in [18] to predict the impact of Covid-19. However, the designed model failed to integrate different deep learning models for accurate Covid-19 prediction. An application of a deep learning model was introduced in [19] to predict the novel coronavirus (COVID-19). However, it was difficult to handle the big data for achieving higher accuracy. Adaptive neuro-fuzzy inference system and enhanced beetle antennae search swarm optimization were developed in [20] to improve the overall performance of the prediction model. However, the deep feature learning model was not implemented for classification, regression.

III. METHODOLOGY

Nowadays, a severe disease named COVID-19 is caused by a virus SARS-CoV-2 infects humans and causes fever, cough, shortness in breathing, and many other serious issues. The epidemic of severe COVID-19 disease is spreading all over the world. Medical health care systems are in urgent need to detect this epidemic disease at an earlier stage with the support of new emerging technologies like deep learning and Big Data systems. In this

situation, the accurate COVID-19 prediction is a main difficult task. Based on the motivation, a deep learning technique called CRWCCDBBN is introduced in this section by combining the feature selection and classification to improve the accurate prediction.

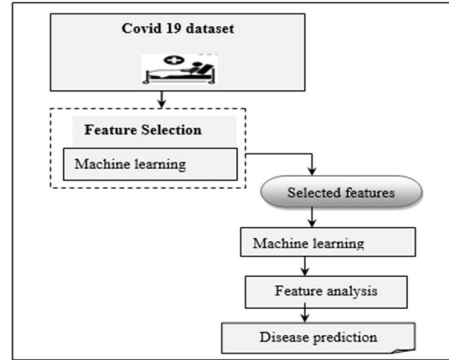


Figure 1. Block Diagram of proposed CRWCCDBBN technique

Fig. 1 given above illustrates a block diagram of the proposed CRWCCDBBN technique with the feature selection and classification. First, the multiple data input is acquired from the corresponding big data dataset. The overall proposed CRWCCDBBN technique with the feature selection and classification. First, the multiple data input is acquired from the corresponding big data dataset. The overall proposed CRWCCDBBN technique is partitioned into two parts, namely, feature selection and disease prediction. The proposed CRWCCDBBN technique first performs the feature selection as a way of choosing the subset of the more relevant features from the original features set by eliminating the irrelevant features.

Input: Big dataset, Number of features $s_1, s_2, s_3, \dots, s_m$, patient data $pd_1, pd_2, pd_3, \dots, pd_n$ Output: Increase the disease prediction accuracy Begin 1: Collect number of features $s_1, s_2, s_3, \dots, s_m$ as input at the input layer 2: For each features, hidden layer 3: For each features s_i 4: Measure the Cox partial log-likelihood between the two features ' $\phi_L(s_i c_k)$ ' 5: if $(R_w > 0)$ then 6: feature is selected as relevant 7: else 8: feature is selected as irrelevant 9: End if 10: End for 11: End for 12: For each patient data -----hidden layer 13: Measure the correlation between the testing and training data ' R_w ' 14: If $(R_w > 0)$ then 15: Activation function ' ϕ ' returns '1' 16: Data is classified as Covid_19 confirmed cases 17: else 18: Activation function ' ϕ ' returns '0' 19: Data is classified as Covid_19 normal cases 20: end if 21: Return (classification results) at the output layer 22: End for End
--

Algorithm 1: Cox Regressive Winsorized Correlated Convolutional Deep Belief Boltzmann Network

Algorithm 1 explains the step-by-step process of Cox Regressive Winsorized Correlated Convolutional Deep Belief Boltzmann Network. The number of features and data are collected from big dataset. Cox Regression is applied in the first hidden layer to select the relevant features based on the Cox partial log-likelihood measure to minimize the time complexity of classification. In the second hidden layer, the Winsorized Correlation Coefficient is applied to analyze the training and testing disease data. The Correlation between training and testing disease data is employed to predict the disease with higher accuracy at output layer.

IV. EXPERIMENTAL SET UP

Experimental evaluation of proposed CRWCCDBBN technique and existing CNN-LSTM [1] Deep-LSTM ensemble model [2] are carried out in Java language using Novel Corona Virus 2019 dataset is taken from the Kaggle. The dataset consists of 33 features and more than 10, 00,000 instances. The numbers of features are

patient ID, age, sex, city, country, province, chronic_disease_ binary, symptoms, travel history, and so on. Among these features, relevant features are selected and perform the classification for disease prediction. The patient's age is higher, and the temperature is higher than 39°C and the travel history is lesser than 14 days, and the chronic disease is present, then the patient is classified as the heavy risk of COVID19. In this way, all the patient data are correctly classified with higher accuracy and minimum error rate.

V. RESULT

The experimental results of the CRWCCDBBN technique and existing CNN-LSTM [1] Deep-LSTM ensemble model [2] are discussed with respect to various performance metrics such as Accuracy, Precision, Recall, F-measure, and prediction time.

TABLE I. PERFORMANCE OF THE ACCURACY USING THREE METHODS

Number of patient data	Accuracy (%)		
	CRWCCDBBN	CNN-LSTM	Deep-LSTM ensemble model
10000	95	92	90
20000	94.5	92.5	89
30000	95.66	93.33	92
40000	96	93	91.25
50000	95.6	92.4	90.2
60000	96	93.5	91.83
70000	95.28	93.14	92.28
80000	95.5	93.25	91.87
90000	95.33	93.55	92.22
100000	95.2	93.8	92.6

Table 1 illustrates performance analysis of accuracy versus a number of patient data. The observed results of CRWCCDBBN technique are compared to the results of existing methods. With these analytical results, CRWCCDBBN technique is increased by 3% when compared to [1] and 4% when compared to [2]. Fig. 2 illustrate the performance analysis of precision versus a number of patient data. Overall performance results indicate that CRWCCDBBN technique is increased by 2% and 3% when compared to existing [1] [2] respectively. Table 2 illustrates the performance of recall with numbers of patient data. The observed results reveal that recall is improved by CRWCCDBBN technique than two existing methods. Performance of recall is improved by 1% and 2% when compared to existing [1] [2] respectively.

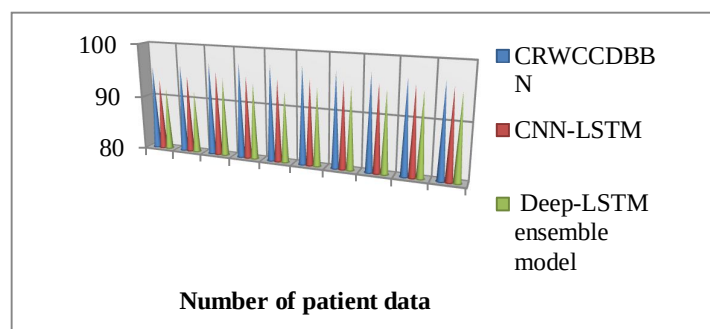


Figure 2. Performance analysis of precision

The performance analysis of F-measure using three methods is plotted in Fig. 3. Among three methods, the CRWCCDBBN technique achieves a higher performance of F-measure. The Results demonstrates that F-measure is significantly improved by 1%, and 3% as compared to the existing methods [1] [2] respectively. Table 3 shows that COVID 19 prediction time with number of patient data. The overall comparison results display that prediction time is decreased by 8% and 16% when compared to the CNN-LSTM [1] and Deep-LSTM ensemble model [2] respectively.

TABLE II. PERFORMANCE OF THE RECALL USING THREE METHODS

Number of patient data	Recall (%)		
	CRWCCDBBN	CNN-LSTM	Deep-LSTM ensemble model
10000	98.86	97.61	96.25
20000	97.81	97.12	95.80
30000	98.21	97.06	96.01
40000	98.4	97.25	96.36
50000	97.87	96.71	96.13
60000	98.06	97.64	97.04
70000	97.88	97.20	96.85
80000	98.01	97.02	96.57
90000	97.98	97.23	96.59
100000	98.08	97.41	96.73

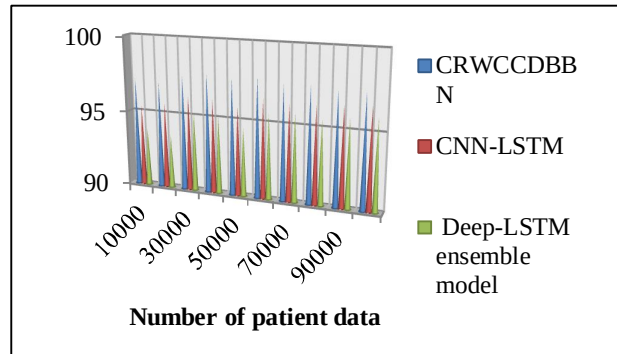


Figure 3. Performance analysis of F-measure

TABLE III. PERFORMANCE OF THE PREDICTION TIME USING THREE METHODS

Number of patient data	Prediction time		
	CRWCCDBBN	CNN-LSTM	Deep-LSTM ensemble model
10000	53	60	68
20000	62	68	72
30000	69	75	81
40000	76	84	92
50000	85	90	100
60000	96	108	120
70000	105	112	126
80000	112	120	128
90000	121.5	126	135
100000	122	130	140

VI. CONCLUSION

Early prediction of COVID-19 is used for minimizing the large burden on the healthcare community. A new deep learning-based technique called CRWCCDBBN for performing the COVID-19 prediction with higher accuracy. CRWCCDBBN technique is used to accurately detect, prevent and diagnose the Covid-19 disease. The CRWCCDBBN technique consists of two different methods namely feature selection and classification. As a result, Covid 19 disease is accurately detected with minimum error. Experimental evaluation is carried out with a big Covid 19 disease dataset. The overall comparison results designate that our CRWCCDBBN technique provides better performance.

REFERENCES

- [1] Sourabh Shastri, Kuljeet Singh, Sachin Kumar, Paramjit Kour, Vibhakar Mansotra, “Deep-LSTM ensemble framework to forecast Covid-19: an insight to the global pandemic”, *International Journal of Information Technology*, Springer, 2021, Pages 1-11.
- [2] Shwet Ketu & Pramod Kumar Mishra, “India perspective: CNN-LSTM hybrid deep learning model-based COVID-19 prediction and current status of medical resource availability”, *Soft Computing*, Springer, Volume 26, 2022, Pages 645–664.
- [3] L. J. Muhammad, Ebrahim A. Algehyne, Sani Sharif Usman, Abdulkadir Ahmad, Chinmay Chakraborty & I. A. Mohammed, “Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset”, *SN Computer Science*, Springer, Volume 2, 2021, Pages 1-13.
- [4] Walter Ageno, Chiara Cogliati, Martina Perego, Domenico Girelli, Ernesto Crisafulli, Francesca Pizzolo, Oliviero Olivieri, Marco Cattaneo, Alberto Benetti, Elena Corradini, Lorenza Bertù, Antonello Pietrangelo & List of contributors, “Clinical risk scores for the early prediction of severe outcomes in patients hospitalized for COVID-19”, *Internal and Emergency Medicine*, Springer, 2021, Pages 1-8.
- [5] Hossein Abbasimehr and Reza Paki, “Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization”, *Chaos, Solitons and Fractals*, Elsevier, Volume 142, 2021, Pages 1-14.
- [6] Yazeed Zoabi, Shira Deri-Rozov & Noam Shomron, “Machine learning-based prediction of COVID-19 diagnosis based on symptoms”, *npj Digital Medicine*, Volume 4, 2021, Pages 1-5.
- [7] Celestine Iwendi, Kainaat Mahboob, Zarnab Khalid, Abdul Rehman Javed, Muhammad Rizwan & Uttam Ghosh, “Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system”, *Multimedia Systems*, Springer, 2021, Pages 1-15.
- [8] Luis Fernando Castillo Ossa, Pablo Chamoso, Jeferson Arango-López, Francisco Pinto-Santos, Gustavo Adolfo Isaza, Cristina Santa-Cruz-González, Alejandro Ceballos-Marquez, Guillermo Hernández and Juan M. Corchado, “A Hybrid Model for COVID-19 Monitoring and Prediction”, *Electronics*, Volume 10, 2021, Pages 1-13.
- [9] Maher Ala’raj, Munir Majdalawieh, Nishara Nizamuddin, “Modeling and forecasting of COVID-19 using a hybrid dynamic model based on SEIRD with ARIMA corrections”, *Infectious Disease Modelling*, Elsevier, Volume 6, 2021, Pages 98-111.
- [10] Ameni Kallel, Molka Rekik & Mahdi Khemakhem, “Hybrid-based framework for COVID-19 prediction via federated machine learning models”, *The Journal of Supercomputing*, Springer, 2021, Pages 1-28.
- [11] Dac-Nhuong Le, Velmurugan Subbiah Parvathy, Deepak Gupta, Ashish Khanna, Joel J. P. C. Rodrigues & K. Shankar, “IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and classification”, *International Journal of Machine Learning and Cybernetics*, Springer, Volume 12, 2021, Pages 3235–3248.
- [12] Aditya Gupta, Vibha Jain, Amritpal Singh, “Stacking Ensemble-Based Intelligent Machine Learning Model for Predicting Post-COVID-19 Complications”, *New Generation Computing*, Springer, 2021, Pages 1-21.
- [13] Muhammad Awais, Mohsin Raza, Nishant Singh, Kiran Bashir, Umar Manzoor, Saif ul Islam, Joel J. P. C. Rodrigues, “LSTM-Based Emotion Detection Using Physiological Signals: IoT Framework for Healthcare and Distance Learning in COVID-19”, *IEEE Internet of Things Journal*, Volume 8, Issue 23, 2021, Pages 16863 – 16871.
- [14] Jayachitra, V.P.Nivetha, SNivetha, RHariniR, “ “A cognitive IoT-based framework for effective diagnosis of COVID-19 using multimodal data” *Biomedical Signal Processing and Control*, Elsevier, Volume 70, 2021, Pages 1-16.
- [15] Connor Shorten, Taghi M. Khoshgoftaar & Borko Furht, “Deep Learning applications for COVID-19”, *Journal of Big Data*, Springer, Volume 8, 2021, Pages 1-54.
- [16] Sahil Ahuja, Nitin Arvind Shelke & Pawan Kumar Singh, “A deep learning framework using CNN and stacked Bi-GRU for COVID-19 predictions in India”, *Signal, Image and Video Processing*, Springer, 2021, Pages 1-8.
- [17] Abir EL Azzaoui, Sushil Kumar Singh, Jong Hyuk Park, “SNS Big Data Analysis Framework for COVID-19 Outbreak Prediction in Smart Healthy City”, *Sustainable Cities and Society*, Volume 71, 2021, Pages 1-14.
- [18] Senthilkumar Mohan, John A, Ahed Abugabah, Adimoolam M, Shubham Kumar Singh, Ali kashif Bashir, Louis Sanzogni, “An approach to forecast impact of Covid-19 using supervised machine learning model”, *Software: Practice and Experience*, 2021, Pages 1–17.
- [19] Amanullah Asraf, Md. Zabirul Islam, Md. Rezwanul Haque & Md. Milon Islam, “Deep Learning Applications to Combat Novel Coronavirus (COVID-19) Pandemic”, *SN Computer Science*, Springer, Volume 1, 2020, Pages 1-7.
- [20] Miodrag Zivkovic, Nebojsa Bacanin, K. Venkatachalam, Anand Nayyar, Aleksandar Djordjevic, Ivana Strumberger, Fadi Al-Turjman, “COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach”, *Sustainable Cities and Society*, Elsevier, Volume 66, 2021, Pages 1-22.