Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Novel Customized Sequential Deep Learning Model (CSDLM) For Covid-19 Risk Prediction

Mrs. Ramya. V[1]*, Dr. K. S. Thirunavukkarasu[2]

[1]Ph.D Scholar, Department of Computer Science, VISTAS, Pallavaram, Chennai, India,

Mail id: ramyavofficial89@gmail.com, Orchid: https://orcid.org/0000-0002-8435-6199

[2]Assistant Professor, Department of Computer Science, VISTAS, Pallavaram, Chennai, India

MailId:thirukst@gmail.com, Orchid:https://orcid.org/0000-0002-3791-5824

ARTICLE INFO

Received: 30 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

One of the most devastating pandemics in human history was caused by the coronavirus 2019 (COVID-19). In order to initiate early intervention, we sought to forecast a bad prognosis among severe patients. It is more likely that the virus may mutate, leading to the formation of highly replicating pathogenic forms, as the number of probable COVID-19 cases rises everyday. A Deep Learning model predicting high risk for COVID-19 from a patient's present symptoms, status, and medical history is the primary goal of this study. We used a Customized Sequential Deep Learning represent (CSDLM) to represent the highly risky COVID-19 situations. A thick layer in this model indicates that every neuron is related to every neuron in surrounding levels, and each layer is closely connected to the one before it. A wide range of measures, including accuracy, precision, recall, and F1_score, were used to thoroughly assess the top performance of the suggested model. This comprehensive assessment approach shed light on how well and how resilient the model predicted COVID-19 high-risk scenarios.

Keywords: covid-19, CSDLM, High risk prediction, Deep learning model.

I. INTRODUCTION

A family of big RNA viruses known as coronaviruses has been around since the mid-1960s. The common cold and other mild to severe infections of the upper respiratory tract are caused by them. Coronaviruses such as SARS-CoV and MERS-CoV are well-known in the human population. When SARS-CoV first emerged in the south part of china, Guangdong in 2003, it was subsequently named. There was a 2012 outbreak of MERS-CoV in Saudi Arabia. Wuhan, a city in Hubei Province, China, revealed new cases of coronavirus infection in December 2019. Covid19 was the name given to the new virus on January 7, 2020. It can lead to symptoms such as anosmia, fever, dry cough, myalgia, and gastrointestinal issues. [5]. In March 2020, the World Health Organization (WHO) declared a global pandemic due to the widespread transmission of COVID-19. As of January 22, 2021, the WHO reported over 96 million cases of COVID-19 and more than two million deaths caused by the virus. [1].

Governments and healthcare agencies may greatly benefit from early epidemic prediction in order to respond promptly to outbreaks. It will make sure that resources are used in a planned way and reduce the effect. By restricting social gatherings and enforcing travel restrictions, locationspecific trend prediction may help lessen the danger of infectious illness transmission at the community level. In a similar vein, infectious illness projections help policymakers get medical and financial resources ready.

Strategic decision-makers may benefit from social impact forecasting as well [2]. Lab findings may be somewhat off-base at times, and patients don't always provide precise enough descriptions of their health issues. Because they could lack expertise in certain areas, specialists struggle to make conclusions about the diseases. To solve this problem and have a positive impact on society, There is a need for a disease prediction system that combines medical information with other systems to enhance its accuracy and effectiveness. [3].

As healthcare resources are limited, medical professionals might use prediction models that include several characteristics or traits to assess the likelihood of infection or adverse infection outcomes. This would help with patient triage. Advances in big data analytics have shifted the focus from traditional disease prediction methods to more modern approaches. In an effort to improve the accuracy of risk classification, several studies have used automated feature selection from massive datasets. The need to quickly and publicly disclose pertinent COVID-19 research results has prompted the development and publication of models spanning the gamut from rulebased markingapplications to sophisticated MLsystems (deep learning) [4,5].

Machine learning is one of several approaches to analyzing complicated data sets; nonetheless, deep learning within machine learning has recently emerged as the most prominent subfield in the discipline. Compared to earlier technological approaches, deep learning excels in processing complicated data, efficiently responding to unstructured data, extracting the primary characteristics of multi-dimensional data, and implementing a classification strategy with more accuracy. More individuals can learn about deep learning and become involved because of this. In recent years, significant advancements have been made in the development of deep learning technologies. It has had a profound impact on various fields, including image recognition, voice recognition, and natural language processing. At the same time, deep learning's application in disease prediction has gained momentum, yielding remarkable results. Recently, deep learning models have demonstrated the ability to directly extract relevant features from raw electronic health data in several areas, including computational phenotyping, risk prediction, and diagnostic prediction.

Heart failure prediction is one application of attention-based recurrent neural networks (RNNs), which are well-suited for risk prediction tasks. To improve efficiency, convolutional neural networks (CNNs) are employed to capture the specific temporal features of patient visits and predict the likelihood of diseases. Long short-term memory (LSTM), a type of recurrent neural network (RNN), is particularly effective at managing and forecasting critical events with long intervals and delays in time series data.Nevertheless, when it comes to medical issues, the time interval between patients' various hospitalizations varies, and traditional LSTM isn't good at learning the important features of a patient's medical condition, so it can't be used practically [8]. One of the most well-known models in deep learning is the Convolutional Neural Network (CNN). A convolutional neural network (CNN) has several layers that apply convolution filters on local features; each compute unit reacts to a tiny section of input data, and the network may leverage the data's underlying structure (such as the 2D structure of picture data) to its advantage. Although convolutional neural network (CNN) models were first developed for computer vision, they have now shown to be useful in several fields, including word embedding learning and search query retrieval. Text mining has seen the widespread use of convolutional neural networks (CNNs) for a variety of purposes, including product feature mining, document categorization, sentence modeling, and many more, ever since Collobert's work on token-level applications [9].

II. RELATED WORKS

According to Zhenyu Dai et al. [10], there is currently little hope for the worldwide control of COVID-19. Patients with severe and critical illness have a reasonably significant death rate despite the fact that COVID-19 has not a high overall mortality rate on an international scale. Death rates among individuals with severe illnesses exceeded 50%, as reported by the World Health Organization. Handling these critical issues in a timely and suitable way is obviously of the utmost importance. Actually, fast case diagnosis has proven achievable in most locations and nations. Thus, the key to treating COVID-19 is figuring out how to limit the development of moderate to severe illness in these people.o predict the likelihood of severe COVID-19, we developed a nomogram, which demonstrated well-fitting calibration curves for both 7-day and 14-day predictions. The nomogram achieved a notable concordance index of 0.86 (95% CI 0.83-0.89). We then created a scoring model (COVID-19-AACC) based on this nomogram, which attained a high concordance index of 0.85 (95% CI 0.81-0.90). To further enhance the model, incorporating additional markers, such as genetic data and imaging, is necessary.

In their study, Madhumita Pal et al. [11] compared K-NN and MLP, two ML methods used for cardiovascular disease prediction. In comparison to K-NN, which achieves an accuracy of 73.77%, MLP achieves an accuracy of 82.47%. With the MLP algorithm, the diagnosis rate was 86.41%, whereas with the K-NN method, it was 86.21%. The medical industry has a lengthy and expensive process for diagnosing CVD. In the event of an error, the proposed method can be highly beneficial for doctors, suggesting that machine learning could serve as a clinical tool for CVD detection. Compared to other methods, the developed MLP model consistently outperforms them in illness prediction.

According to S. Akila et al. [12], DT and MLP were used in a hybrid approach to predict the risk of coronary heart disease (CHD) in drivers who work in the industry based on clinical, chemical, and habitation factors. An essential component of ANNs, MLPs are multilayer neural feedforward networks trained using the back propagation approach to detect mistakes. In order to fine-tune the ANN, we choose sets of variable parameters based on pair-wise correlations between database parameters and CHD diagnosis.This initial step of DT prediction obtained an accuracy of 98.66%. Level 2 of MLP prediction analyses CHD risk instances and misclassified examples from level 1 and achieves an accuracy of 96.66%. Health data's predictive potential is more influenced by its specificity and sensitivity than its accuracy.

A novel RCNN-based model for illness risk assessment is proposed by MOHD USAMA et al. [13]. By utilizing both structured and unstructured data, the model extracts detailed aspects of chronic infarction disease, enhancing the accuracy of previous models. Experimental results show that the proposed model achieves an accuracy of 96.02%, outperforming previously published results in chronic illness risk prediction. Future studies will use dilated convolution to better forecast the likelihood of deadly illnesses by identifying their early warning signs.

A new method for collaborative filtering-enhanced deep learning was suggested in a study by Xin Li et al. [14]. Collaborative topic regression (CTR) was used first in the suggested method to extract value from both structured and unstructured data in order to fill in the gaps caused by missing information. After that, the prediction was made using the trained deep neural network. A real-world dataset of readmissions from diabetes patients was used to test the suggested method on a large and complicated medical issue. Our system was shown to be able to efficiently use both structured and unstructured data to uncover hidden relationships, reliably estimate dataset missing values, and combine feature learning for complicated data prediction in their experiments. They also compared it to other intelligent algorithms like Decision Tree, Naive Bayes, and Support Vector Machine (SVM). The outcome of the extensive comparison trials showed that the suggested strategy was superior to these other alternatives.

MD Samiul Islam et al. [15] detected CVD early on, allowing physicians to intervene quickly and accurately treat patients. The whole country might be inspired to understand comprehensive and individualised healthcare by this kind of proactive therapy. Consequently, health informatics may help in the early detection and prevention of CVD by identifying the major risk factors for a particular location. Also, even the younger generation, which is already at risk for CVD due to its lack of discipline, might benefit from data-driven healthcare therapy by receiving the emotive lifestyle that CVD patients need. Traditional lifestyle changes, such as exercising regularly, going on walks regularly, eating healthily, and getting enough sleep at the right times, may also be aided by this. The proposed attention-based LSTM for the Intelligent Healthcare Platform could shift the focus towards prognosis and risk factors of cardiovascular diseases (CVD) rather than just detection.

According to Mohamed Djerioui et al. [16], LSTM networks, with their memory capabilities, can optimize the number of hidden layer nodes in a network for predicting cardiac diseases and addressing time series issues with strong correlations. In comparison to the multi-layer perceptron model, the proposed model performs exceptionally well in predicting cardiac events. The results highlighted that LSTM significantly enhanced the efficiency of the predictive system. Healthcare system design and deployment of cardiac disease prediction systems, as well as cost reductions, may be greatly impacted by using this approach.

Exclusion and inclusion criteria for the study's subjects and data were strictly followed by Chuanyu Hu et al. [17]. Finally, when compared to other models that have been published, our prediction model is competitive, and our findings are similar to those of others. The model is linear, with regression coefficients showing the effects of the predictors, making it both simple—comprising just four variables—and highly interpretable. Additionally, the final prediction model was validated. So that doctors could use our prediction model more easily, they also created a companion online tool.

A multi-label classification issue, multi-disease risk prediction was investigated by Tingyan Wang et al. [18] in their modeling work. An actual hospital dataset verified the suggested technique. Our research shown that many stakeholders, including patients, healthcare providers, policymakers, and medical experts, may benefit from various degrees of aggregated patient diagnoses. By aggregating 3-digit ICD codes, LSTM networks achieved an exceptional exact-match score of 98.90% on the MIMIC dataset and 95.12% on the GenCare dataset. With 4-digit ICD code aggregation, the scores improved to 96.60% and 96.83%, respectively. These results highlight the ability of LSTM networks to effectively predict future disease risks for patients.A hospital information system might also include the study's findings as a function support module to help doctors and nurses make better decisions when they need them.

According to Ali A. El-Solh et al. [19], it is important to validate predictions made by risk models before using them on cohorts with a varied distribution of patient characteristics. Predicting the trajectory of a pandemic in its early phases is difficult, as seen by the diversity in projected results in this research. Improving prediction models requires laying the groundwork for platforms that can deploy proven models and assess their efficacy in the future.

An significant public health medical issue, osteoporosis risk factor estimate has been studied, and Dimitrios H. Mantzaris et al. [20] brought up the prospect of using artificial neural networks in medical decision-making. The development of precise and timely methods that do not expose patients to radiation is very important due to its frequency and the harmful repercussions it may have on patients. Both feed-forward neural networks (PNNs) and multi-layer perceptrons (MLPs) with back-propagation algorithms formed the basis of advanced artificial neural approaches. Many have called MLPs "black boxes" because to the fact that their internal connections are not linear and do not follow standard statistical patterns.

Summary:

- Coronavirus comes in many forms and is found all over the globe. Coronavirus infection was the agent responsible for the 2003 SARS pandemic and the 2013 Middle East respiratory disease (MERS).
- Deep learning (DL), a subset of machine learning, utilizes a set of algorithms designed to model high-level abstractions of data. It is also referred to as deep structured learning, deep machine learning, and hierarchical learning.

- Cardiovascular disease, coronary atherosclerosis, and pulmonary infections are among the many additional illnesses that may have their fine-grain characteristics extracted using deep learning methods.

III. PROPOSED METHOD

Kaggle was used to get the dataset for the COVID-19 forecast. The input features were prepared for predictive modeling by handling null values, removing unnecessary features, and handling categorical information during data preparation. Separate training and test sets were created from the dataset after preprocessing.

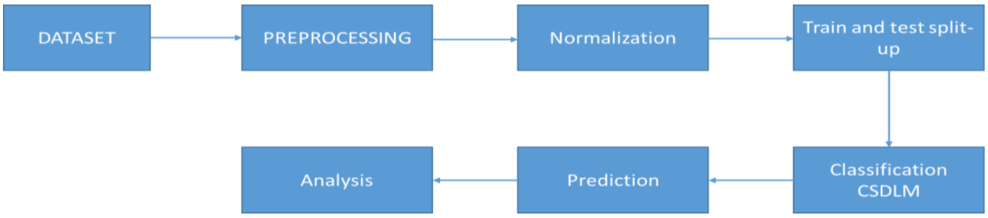


Figure 1. Proposed Block Diagram

The associations between the input characteristics were then assessed using the Pearson correlation coefficient approach. Both the training and test sets of data were then stripped of associated characteristics. An Adaptive Sequential Deep Learning Model (CSDLM) was used for the purpose of forecasting COVID-19 danger situations. The layers in this model are densely interconnected, with a thicker layer representing full connectivity between all neurons in that layer and those in the layer beneath it.

The model was trained using the specified training data, which allowed it to recognize and internalize dataset patterns. Then, the trained model was used to forecast COVID-19 cases with a high risk level in the test data. The proposed model was evaluated using various performance metrics, including recall, accuracy, precision, and F1-score.This comprehensive assessment approach shed light on how well and how resilient the model predicted COVID-19 high-risk scenarios.

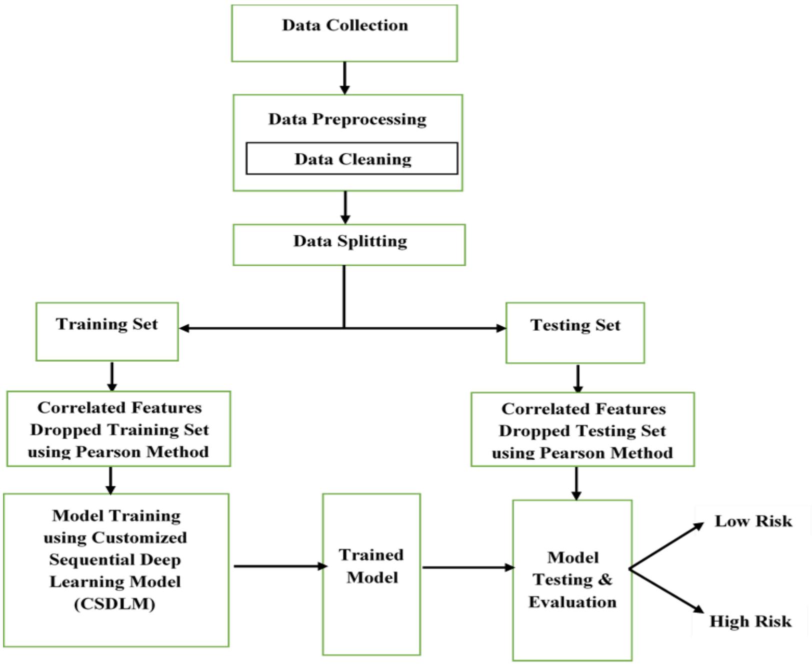


Figure 2. Proposed Architecture

i. Sequential Deep Learning Model:

When optimizing an ANN, the optimum approach is to use the supervised learning method to find the values for all of the parameters, including the "input weights" and the "output weights." Discovering the values of an ANN's parameters is the first step in establishing an ANN model. At this stage, the ANNs are trained using the observed values, or training data, and optimization technique. The correlation coefficient (R) is used to improve performance, whereas the root-meansquared error (RMSE) function is now employed as a fitness function for evaluating the ANNs' performance [21].

One unique architecture in deep learning is the sequential deep learning model, which stacks layers one after another as shown in figure 3. A linear flow of data is created in the network when the output of one layer is used as the input of the following layer in this sort of model. Frameworks such as Keras rely on the sequential model, a basic structure that facilitates the simple development of neural networks via the sequential addition of layers. Many deep learning tasks are well-suited to sequential models because of their simplicity and ease of construction. Therefore, a tailored model for classification with high features is suggested for risk prediction owing to COVID-19.

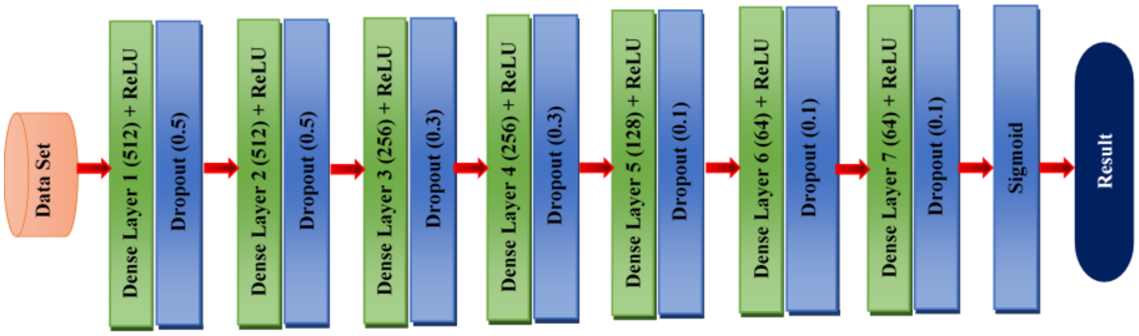


Figure3. Architecture of the proposed CSDLM model

ii. Adam optimizer:

As a deep learning optimization method, it is among the most popular choices. Adam leverages the estimated first and second moments of gradients to compute individual adaptive learning rates for different parameters, deriving its name from adaptive moment estimation.Adam combines the pros of two popular methods: AdaGrad, which excels at sparse gradients, and RMSProp, which is useful in online and non-stationary contexts. It includes both momentum and rmsprop. Along with a bias adjustment method, the update procedure only takes the smooth gradient into account.

Adam has many significant traits. Momentum serves as a direct estimate of the gradient's first-order moment. Adam incorporates bias-corrected estimates of both first-order and second-order moments to address their initialization at the origin. The hyperparameters ρ1\rho_1ρ1 and ρ2\rho_2ρ2 control the exponential decay rates of the moving averages, which are iteratively updated by the algorithm. These moving averages include the squared gradient ut and the exponential moving average of the gradient. Moving averages are estimations of the mean and uncentered variance of the gradient, the first and second raw moments, respectively.

The initial and last moment variables, m and u, are required by the Adam method. Time step t updates the biased first moment estimate and the gradient, after calculating the latter.

$$m_t \leftarrow \rho_1 m_{t-1} + (1-\rho_1) g_t \tag{1}$$

$$u_t \leftarrow \rho_2 u_{t-1} + (1-\rho_2) g \odot g \tag{2}$$

Adam has several benefits. A little adjustment to the learning rate is necessary initially. It is also a simple strategy that does not change when the gradients are scaled diagonally. It requires very minimal memory and is efficient computationally. In addition, Adam works well for nonstationary goals and issues with sparse gradients and high levels of noise [22].

- A stack of seven thick layers forms the basis of the suggested model's design. Layers one and two each have 512 neurons, whereas layers three and four each have 256. The fifth layer has 128 neurons, the sixth and seventh have 64 neurons apiece, and the bottom two levels have no neurons at all. To minimize the risk of overfitting, steps are taken to ensure the model generalizes well to unseen data., we apply dropout regularization to each thick layer at a rate of 0.5, 0.3, and 0.1, respectively.

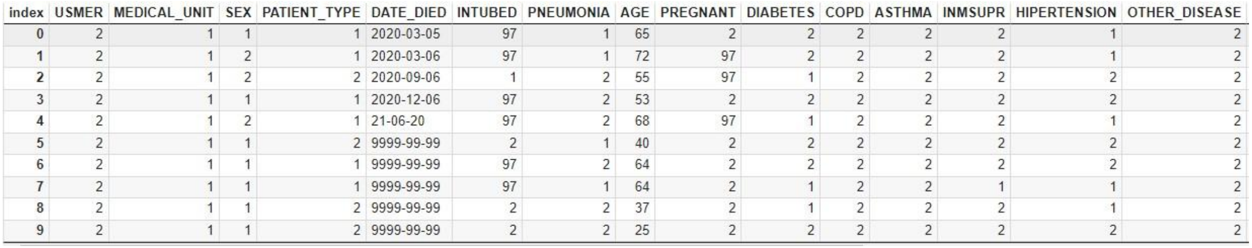
- A sigmoid activation function is utilized in the model's last output layer. Because of its compatibility with the prediction job and its ability to facilitate the intended result, the sigmoid activation function was purposefully chosen.

- Using the ADAM optimizer, the model is optimized during training. By dynamically modifying learning rates to match the individual needs of each model parameter, the ADAM optimizer is renowned for its efficiency and versatility.

- Using a binary cross-entropy loss function, we can make predictions. For COVID-19 high-risk prediction, this loss function is ideal for binary classification tasks as it quantifies the disparity between actual and predicted outcomes, aiding the model in making accurate predictions.

IV. RESULT AND DISCUSSION

The Govt. of Mexico supplied the data source. An large amount of de-identified patient data, inlcudes medical history and current conditions, is included in this collection. Eleven distinct characteristics make up the raw dataset. Out of 22,500 patient records, 897 were found to be positive and 3603 were found to be negative.



index	USMER	MEDICAL_UNIT	SEX	PATIENT_TYPE	DATE_DIED	INTUBED	PNEUMONIA	AGE	PREGNANT	DIABETES	COPD	ASTHMA	INMSUPR	HIPERTENSION	OTHER_DISEASE
0	2	1	1	1	2020-03-05	97	1	65	2	2	2	2	2	1	2
1	2	1	2	1	2020-03-06	97	1	72	97	2	2	2	2	1	2
2	2	1	2	2	2020-09-06	1	2	55	97	1	2	2	2	2	2
3	2	1	1	1	2020-12-06	97	2	53	2	2	2	2	2	2	2
4	2	1	2	1	21-06-20	97	2	68	97	1	2	2	2	1	2
5	2	1	1	2	9999-99-99	2	1	40	2	2	2	2	2	2	2
6	2	1	1	1	9999-99-99	97	2	64	2	2	2	2	2	2	2
7	2	1	1	1	9999-99-99	97	1	64	2	1	2	2	1	1	2
8	2	1	1	2	9999-99-99	2	2	37	2	1	2	2	2	1	2
9	2	1	1	2	9999-99-99	2	2	25	2	2	2	2	2	2	2

Figure 4. Data_set used for the preented model

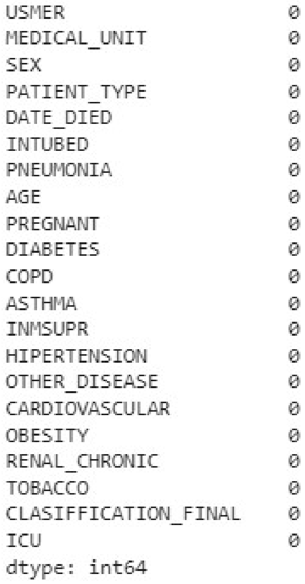
A sample of the dataset used in the presentedwork is displayed in fig. 4.

The upcoming features are included in the dataset that was used for the proposed model:

- sex: 1 (1_female) and 2 (2_male).
- age
- • Classification: results of the COVID-19 test. Values 1 to 3 indicate varying degrees of COVID-19 diagnosis; values 4 or above indicate no COVID-19 carrier status or inconclusive results.

- patient type: 1(returned home) and 2 (hospitalization).
- pneumonia: have air sacs inflammation or not.

- pregnancy
- diabetes
- copd: Has Chronic obstructive pulmonary disease or not.
- asthma
- inmsupr: Immunosuppressed or not.
- hypertension
- cardiovascular: Heart or blood vessels related disease.
- renal chronic: chronic renal disease or not.
- other disease
- obesity
- tobacco
- usmr: treated patients in first, second, or third level medical units.
- medical unit: kind of National Health System facility that gave the treatment.
- intubed: if the ventilator was attached to the patient.
- icu: provides information on the patient's admission status to an intensive care unit.
- date died: provides information on the patient's admission status to an intensive care unit.



```
USMER                   0
MEDICAL_UNIT            0
SEX                     0
PATIENT_TYPE            0
DATE_DIED               0
INTUBED                 0
PNEUMONIA               0
AGE                     0
PREGNANT                0
DIABETES                0
COPD                    0
ASTHMA                  0
INMSUPR                 0
HIPERTENSION            0
OTHER_DISEASE           0
CARDIOVASCULAR          0
OBESITY                 0
RENAL_CHRONIC           0
TOBACCO                 0
CLASIFFICATION_FINAL    0
ICU                     0
dtype: int64
```

Figure 5. Preprocessed data

The suggested methodology's preprocessed data is shown in Figure 5. The transformation of raw data into a format appropriate for training models is known as data preprocessing and is an essential part of machine learning workflows. The data is cleaned, transformed, and prepared for improved model performance using a range of approaches and procedures included in it. In this case, the preprocessing method includes managing missing values, normalizing the data, and dividing the data into train and test sets. Methods for dealing with missing values include eliminating features or occurrences that have missing data or imputation, which involves filling missing values with a statistic such as the mean, median, or mode. Bringing all numerical characteristics to the same scale, as in Min-Max scaling or Standardization (zero mean and one standard deviation). Creating three separate sets from the dataset: one for training models, one for validating their parameters, and a third for testing their performance.

In this case, the data_set is partitioned into 2 sub divisions.: one for testing and the another for training.

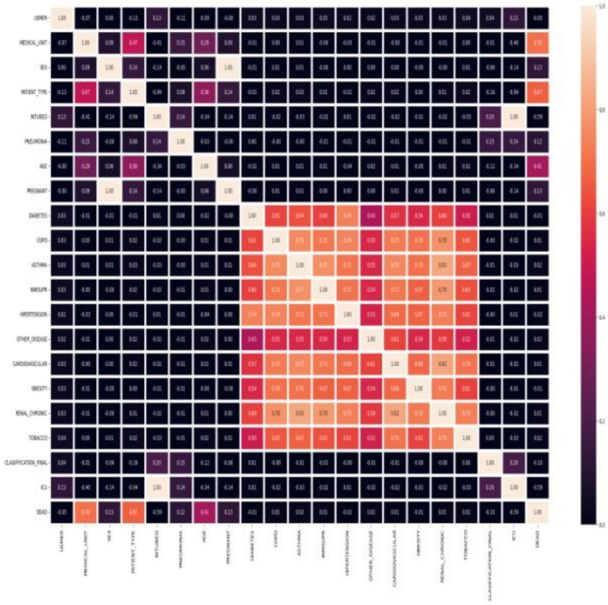


Figure 6. Correlated Features Extraction and Dropping using the Pearson Correlation Coefficient Technique

The correlation values among the input variables are shown in Figure 6. Greater correlations, as shown by higher absolute values, impact the interrelationships of these variables. For better feature selection and model interpretation, this representation helps to grasp inter-variable relationships.

Table I. Proposed Model Structure (CSDLM)

Layer	**Output shape**	**Param#**
1.Dense (Dense)	512	9216
2.Dropout (Dropout)	512	0
3.Dense_1 (Dense)	512	262656
4.Dropout_1 (Dropout)	512	0
5.Dense_2 (Dense)	256	131328
6.Dropout_2 (Dropout)	256	0
7.Dense_3 (Dense)	256	65792
8.Dropout_3 (Dropout)	256	0
9.Dense_4 (Dense)	128	32896
10.Dropout_4 (Dropout)	128	0
11.Dense_5 (Dense)	64	8256
12.Dropout_5 (Dropout)	64	0
13.Dense_6 (Dense)	64	4160
14.Dropout_6 (Dropout)	64	0
15.Dense_7 (Dense)	1	65

Metrics like accuracy of training process and loss are often utilized to assess a model's performance in the training.

1. Training Accuracy: For our training dataset, this statistic shows what percentage of instances were properly labeled. Divide the sum of all training-set instances by the number of accurate predictions to get the accuracy rate. If the model is learning to match the training data effectively, then the training accuracy will be greater.

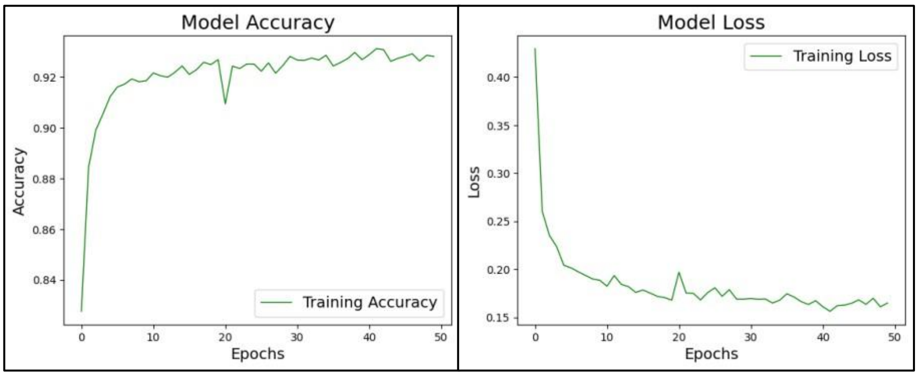


Figure 7. Model Accuracy and model loss

2. Training Loss: This statistic compares the training data's real values (ground truth) with its expected values (prediction error). It is a measure of the accuracy with which the model's predictions correspond to the genuine goals. Binary cross-entropy is a common loss function used for classification problems. Minimizing this loss during training entails getting the model's predictions as near to the real values in fig. 6 as feasible.

Our model's learning and improvement performance may be better understood by comparing the training accuracy and loss numbers produced at various epochs (training iterations).

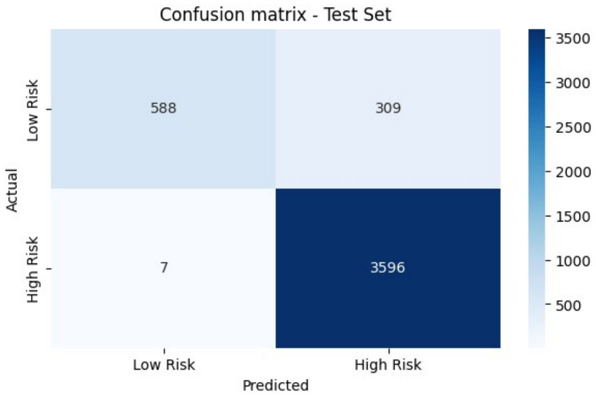


Figure 8. Confusion matrix of the proposed model

grouping tasks make use of confusion matrices, which display the number of properly and erroneously categorized cases of each class, to provide insight into the model's effectiveness. Here is a confusion matrix with certain numbers that need to be explained:

True Positive (TP): The count of cases that were appropriately categorized as low risk. In this case, TP = 588.

False Positive (FP): The total count of cases that were incorrectly classified as low risk despite being high risk is represented by FP, which in this scenario is 7.

True Negative (TN): The total number of cases accurately classified as high risk is denoted by TN, which in this instance is 3596.

.False Negative (FN): The tot.count of cases when low risk categories were mistakenly assigned high risk categories. In this case, FN = 309.

Table II. Validation comparison of the proposed model.

methodology	**Precision**	**Recall**	**F1 Score**	**Accuracy**
DT	0.86	0.88	0.86	0.8124
MLP	0.88	0.89	0.87	0.8764
LSTM	0.91	0.89	0.89	0.8812
CSDLM	0.93	0.93	0.92	0.9297

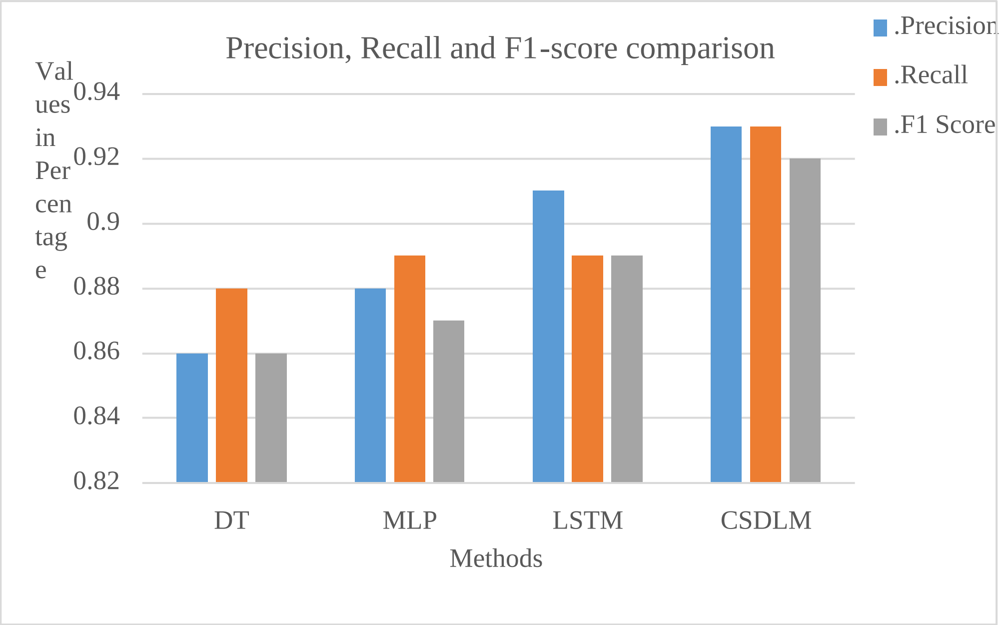


Figure 9. Validating parameters comparison on proposed and existing methodologies.

Table II and figure 9 provide the validation parameters, such as recall, precision, and f-score, for both the new and current systems. Based on the results of the comparison, it is evident that our suggested CSDLM is the most efficient system. The decision tree model performs the worst of all the options.

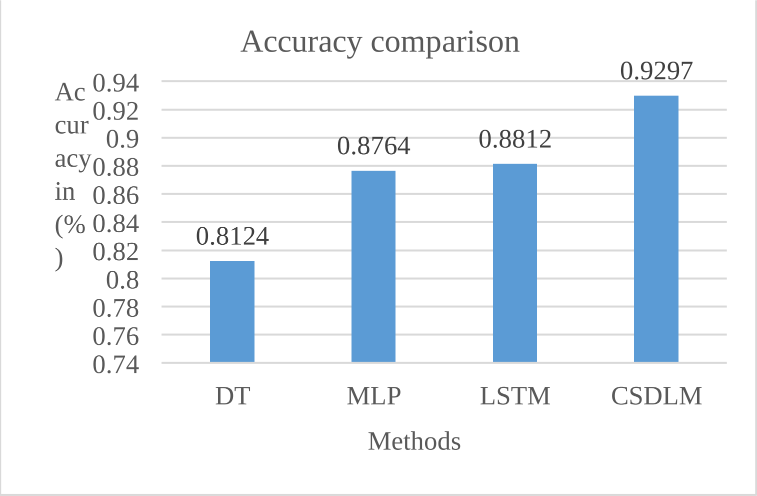


Figure 10. Accuracy comparison of existing and proposed method

In Fig. 10, we can see a justification of the accuracy of Decision tree, MLP, RNN-LSTM, and CSDLM on the dataset that was utilized for the proposed technique. The suggested approach is obviously superior than all previous methods, as shown in the chart. The suggested approach outperforms LSTM by 4.9%, MLP by 5.3%, and decision tree by 11.7%.

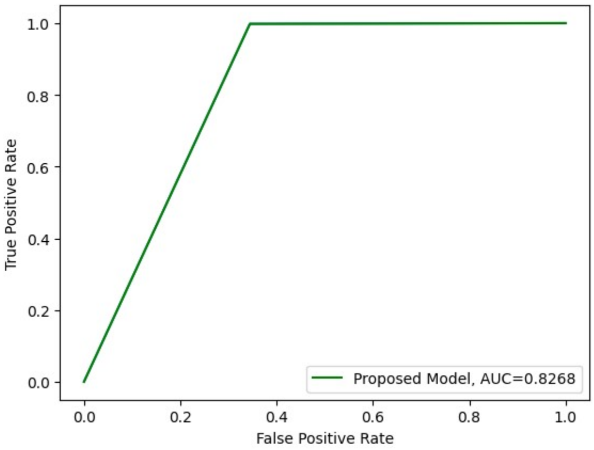


Figure 11. ROC curve of the proposed method

An area under the curve (AUC) that is greater than zero implies superior performance in ROC curve analysis. With an AUC of 1, we can perfectly distinguish between positive and negative examples, but an AUC of 0.5 is equivalent to random guessing. The model demonstrates a strong ability to differentiate between the two groups based on predicted probabilities, as evidenced by an AUC score of 0.8268, significantly outperforming random chance.

CONCLUSION

The coronavirus has spread throughout the whole world. As of August 31, 2021, it is responsible for about 4 million fatalities.Scientists made enormous attempts to mitigate the pandemic impacts after the horrifying spread of COVID-19. A great deal of time and energy was devoted to finding a way to counter this virus by scientists and researchers. One of the early strategies to fight the infection was using DL techniques.Here, we present a new deep learning method for COVID-19 illness prediction. The CSDLM is a custom-built sequential deep learning model. Predicting and comparing the intensity of viral symptoms is a common application of DL algorithms, particularly for those with chronic diseases. The Mexican government's COVID-19 patient dataset is processed using DT, MLP, and LSTM methods. The collection included a variety of disease-related data, including 21 distinct characteristics. The proposed system surpasses current state-of-the-art methodologies, achieving a remarkable performance with an accuracy of 92.97%, a recall of 93%, a precision of 92%, and an F1-score of 92% in predicting risk levels associated with COVID-19.

REFERENCES

1. Alballa, N., & Al-Turaiki, I. (2021). Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review. *Informatics in Medicine Unlocked*, *24*, 100564. https://doi.org/10.1016/j.imu.2021.100564
2. Pal, R., Sekh, A. A., Kar, S., & Prasad, D. K. (2020). Neural network based country wise risk prediction of COVID-19. *Applied Sciences (Switzerland)*, *10*(18). https://doi.org/10.3390/APP10186448
3. Saranya, G., &Pravin, A. (2020). A comprehensive study on disease risk predictions in machine learning. *International Journal of Electrical and Computer Engineering*, *10*(4), 4217–4225. https://doi.org/10.11591/ijece.v10i4.pp4217-4225
4. Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., Bonten, M. M. J., Damen, J. A. A., Debray, T. P. A., De Vos, M., Dhiman, P., Haller, M. C., Harhay, M. O., Henckaerts, L., Kreuzberger, N., Lohmann, A., Luijken, K., Ma, J., Andaur Navarro, C. L., ... Van Smeden, M. (2020). Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *The BMJ*, *369*. https://doi.org/10.1136/bmj.m1328
5. Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease Prediction by Machine Learning over Big Data from Healthcare Communities. *IEEE Access*, *5*, 8869–8879. https://doi.org/10.1109/ACCESS.2017.2694446
6. Xie, S., Yu, Z., &Lv, Z. (2021). Multi-disease prediction based on deep learning: A survey. *CMES - Computer Modeling in Engineering and Sciences*, *127*(3). https://doi.org/10.32604/CMES.2021.016728
7. Ma, F., You, Q., Gao, J., Zhou, J., Suo, Q., & Zhang, A. (2018). Risk prediction on electronic health records with prior medical knowledge. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1910–1919. https://doi.org/10.1145/3219819.3220020
8. Junwei, K., Yang, H., Junjiang, L., &Zhijun, Y. (2019). Dynamic prediction of cardiovascular disease using improved LSTM. *International Journal of Crowd Science*, *3*(1), 14–25. https://doi.org/10.1108/IJCS-01-2019-0002
9. Cheng, Y., Wang, F., Zhang, P., & Hu, J. (2016). Risk prediction with electronic health records: A deep learning approach. *16th SIAM International Conference on Data Mining 2016, SDM 2016*, 432–440. https://doi.org/10.1137/1.9781611974348.49
10. Dai, Z., Zeng, D., Cui, D., Wang, D., Feng, Y., Shi, Y., Zhao, L., Xu, J., Guo, W., Yang, Y., Zhao, X., Li, D., Zheng, Y., Wang, A., Wu, M., Song, S., & Lu, H. (2020). Prediction of COVID-19 Patients at High Risk of Progression to Severe Disease. *Frontiers in Public Health*, *8*(November), 1–9. https://doi.org/10.3389/fpubh.2020.574915
11. Pal, M., Parija, S., Panda, G., Dhama, K., &Mohapatra, R. K. (2022). Risk prediction of cardiovascular disease using machine learning classifiers. *Open Medicine (Poland)*, *17*(1), 1100–1113. https://doi.org/10.1515/med-2022-0508
12. Akila, S., &Chandramathi, S. (2015). A hybrid method for Coronary Heart Disease risk prediction using decision tree and multi layer perceptron. *Indian Journal of Science and Technology*, *8*(34). https://doi.org/10.17485/ijst/2015/v8i34/85947

13. Usama, M., Ahmad, B., Wan, J., Hossain, M. S., Alhamid, M. F., & Hossain, M. A. (2018). Deep feature learning for disease risk assessment based on convolutional neural network with intra-layer recurrent connection by using hospital big data. *IEEE Access*, *6*, 67927–67939. https://doi.org/10.1109/ACCESS.2018.2879158
14. Li, X., & Li, J. (2018, September). Health risk prediction using big medical data-a collaborative filtering-enhanced deep learning approach. In *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)* (pp. 1-7). IEEE.
15. Islam, M. S., MuhamedUmran, H., Umran, S. M., & Karim, M. (2019). Intelligent Healthcare Platform: Cardiovascular Disease Risk Factors Prediction Using Attention Module Based LSTM. *2019 2nd International Conference on Artificial Intelligence and Big Data, ICAIBD 2019*, *May*, 167–175. https://doi.org/10.1109/ICAIBD.2019.8836998
16. Djerioui, M., Brik, Y., Ladjal, M., &Attallah, B. (2020). Heart Disease prediction using MLP and LSTM models. *2020 International Conference on Electrical Engineering, ICEE 2020*. https://doi.org/10.1109/ICEE49691.2020.9249935
17. Hu, C., Liu, Z., Jiang, Y., Shi, O., Zhang, X., Xu, K., Suo, C., Wang, Q., Song, Y., Yu, K., Mao, X., Wu, X., Wu, M., Shi, T., Jiang, W., Mu, L., Tully, D. C., Xu, L., Jin, L., … Chen, X. (2020). Early prediction of mortality risk among patients with severe COVID-19, using machine learning. *International Journal of Epidemiology*, *49*(6), 1918–1929. https://doi.org/10.1093/ije/dyaa171
18. Wang, T., Tian, Y., &Qiu, R. G. (2020). Long short-term memory recurrent neural networks for multiple diseases risk prediction by leveraging longitudinal medical records. *IEEE Journal of Biomedical and Health Informatics*, *24*(8), 2337–2346. https://doi.org/10.1109/JBHI.2019.2962366
19. El-Solh, A. A., Lawson, Y., Carter, M., El-Solh, D. A., &Mergenhagen, K. A. (2020). Comparison of in-hospital mortality risk prediction models from COVID-19. *PLoS ONE*, *15*(12 December), 1–13. https://doi.org/10.1371/journal.pone.0244629
20. Mantzaris, D. H., Anastassopoulos, G. C., &Lymberopoulos, D. K. (2008). Medical disease prediction using artificial neural networks. *8th IEEE International Conference on BioInformatics and BioEngineering, BIBE 2008*, *February 2015*. https://doi.org/10.1109/BIBE.2008.4696782
21. Hamadneh, N. N., Khan, W. A., Ashraf, W., Atawneh, S. H., Khan, I., &Hamadneh, B. N. (2021). Artificial neural networks for prediction of covid-19 in Saudi Arabia. *Comput. Mater. Contin*, *66*, 2787-2796.
22. Soydaner, D. (2020). A comparison of optimization algorithms for deep learning. *International Journal of Pattern Recognition and Artificial Intelligence*, *34*(13), 2052013.