TPM Vol. 32, No. S2, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

GRAPH CONVOLUTIONAL NETWORKS FOR PREDICTING DISEASE OUTBREAKS IN PUBLIC HEALTH SURVEILLANCE

[1]K. SWARUPA RANI
ASSISTANT PROFESSOR, DEPARTMENT OF INFORMATION TECHNOLOGY, P.V.P SIDDHARTHA INSTITUTE OF TECHNOLOGY, VIJAYAWADA, ANDHRA PRADESH, INDIA. EMAIL: swarupapvpsit@gmail.com

[2]B ARUNA DEVI
PROFESSOR, DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING, DR NGP INSTITUTE OF TECHNOLOGY, COIMBATORE, TAMILNADU, INDIA. EMAIL: arunadevi@drngpit.ac.in

[3]R JAYADURGA
DEPARTMENT OF COMPUTER SCIENCE, CHRIST UNIVERSITY, YESHWANTPUR(YPR) CAMPUS, BENGALURU, KARNATAKA, INDIA. EMAIL: jayadurgar@gmail.com

[4]MEENAKSHI K
PROFESSOR, DEPARTMENT OF MATHEMATICS, VTU(RC), CMR INSTITUTE OF TECHNOLOGY, BENGALURU, INDIA. EMAIL: meenakshik531@gmail.com

[5]K. PRAVEENA
ASSISTANT PROFESSOR, DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING, SCHOOL OF ENGINEERING, MOHAN BABU UNIVERSITY, TIRUPATI, ANDHRA PRADESH, INDIA. EMAIL: kakarlapraveenasvec@gmail.com

[6]VENKATA RAMANA K
ASSOCIATE PROFESSOR, DEPARTMENT OF CSE, QIS COLLEGE OF ENGINEERING AND TECHNOLOGY, QIS COLLEGE OF ENGINEERING AND TECHNOLOGY, ONGOLE, ANDRAPRADESH, INDIA.
EMAIL: ramanakaveripakam@gmail.com

Abstract

Aim: With the help of CT scans, we construct a detection module by adhering to a procedure for COVID-19

Background: The current public health crisis, which is known as SARS-CoV-2, has resulted in a number of fatalities and has caused extensive economic disruption worldwide.

Methodology: By adhering to a pre-processing, feature-extraction, and detection strategy, we are able to construct a detection module that is capable of identifying COVID-19 patients via CT images. Following the extraction of features using a Grey Level Co-occurrence Matrix (GLCM), the next step in the image pre-processing process is classification using Graph Convolutional Networks (GCN).

Contribution: The objective of the simulation is to assess the performance of the model by making use of a number of different CT imaging datasets that contain images depicting a significant number of patients individually.

Findings: With a detection rate of 98% and a mean average percentage error (MAPE) that is lower than 0.2, the outcomes of the simulation reveal that the recommended method beats the traditional procedures that are currently in use.

Keywords: Viral Detection, Covid outbreak, GCN, Deep Learning

1. INTRODUCTION

As the world continues to expand at a dizzying rate, the importance of technological growth is growing even in nations that have already achieved high levels of development. Numerous facets of day-to-day life, including as education, business, communication, military operations, and health care are becoming increasingly dependent on the uses of new technology (Anno et al., 2022). When identifying symptoms and performing digital patient triage making accurate diagnoses, the healthcare institution is an essential area that demands modern technology as soon as possible (Wang et al., 2022).

A medication has not been able to cure the COVID-19 virus as of yet; however, the World Health Organisation has allowed the use of numerous combination medications that are based on alcohol, hydrogen peroxide and isopropyl alcohols due to the high reactivity level that these drugs have with the virus (Saravanan et al., 2023).

Some examples of diagnostic instruments that could potentially benefit from the implementation. One of the approaches that are now to identify infections is a technology that is capable of functioning on its own, such as thermal screening (Yadav et al., 2024).

For the purpose of determining whether the study addresses the existing knowledge gaps, the researchers utilised a machine learning technique to investigate the COVID-19 literature (Saravanan et al., 2023). The advantages of these AI-based techniques is the possibility of reducing the amount of time required to identify and cure COVID-19 infections, as indicated in (Shesayar et al., 2023).

Data mining with AI help enhance the accuracy of case discovery, which is especially important in light of the rapid spread of the pandemic difficulties that it has brought about. The treatment of the health problem will then be more successful as a result of this information. It is possible that artificial intelligence could be helpful in finding a solution to the problem in a variety of different ways (Yu et al., 2022). COVID-19 was a pandemic that spread over the world and posed a threat to human existence. According to systematic reviews (Yu et al., 2023), computers are capable of doing a wide variety of tasks on their own by utilising statistical models and machine learning (ML) training methodologies.

The accuracy of predictions made with the help of machine learning algorithms has made them all but routine in recent years. The availability of a new database of low quality online is one of the challenges that machine learning methods confront, however in comparison to more conventional approaches, these methods face less challenges. In order to make predictions, the researchers made use of the most effective machine learning model that was suitable for the dataset that they were working with (Fritz et al., 2022).

The use of machine learning in conjunction with data analytics makes it possible to identify patterns that were not previously visible (Du et al., 2021). There has been a significant increase in the utilisation of machine learning algorithms for the goal of finding and classifying risk indicators as well as the interactions that occur between them (Sun et al., 2024).

Preprocessing, feature extraction, and COVID-19 patient detection are all utilised by the detection module of the study, which makes use of previously published methods. The properties that are collected via the use of Graph Convolutional Networks (GCN) and Grey Level Cooccurrence Matrix (GLCM) are the basis for which image categorization is performed.

The novelty of the paper is given below:

- With the help of CT scans, the authors construct a detection module by adhering to a procedure that involves pre-processing, feature extraction, and patient detection for COVID-19.
- The process of extracting and classifying image features begins with the preprocessing of the images, which is the initial stage in the process that ultimately results in GCN classification.

2. RELATED WORKS

The use of machine learning as a categorization model has a great deal of promise (La Gatta et al., 2020). To put it another way, machine learning models are instruments that permit the establishment of the functional, dependent, or structural relationships that exist between the variables that are input and those that are output. Explicit algorithms or automated learning approaches will not be able to construct these relationships in the majority of circumstances (Li et al., 2023).

Through the utilisation of ML, it is possible to forecast the confirmed cases and fatality rates in the future (Tariq et al., 2023). It is divided into two categories by the components that are utilised to classify machine learning. Through the utilisation of a fault-identification neural network, the implementation of a genetic algorithm enables the

calculation of the suitable weight for merging data from the perception results of a large number of nodes, the elimination of nodes that are not necessary, and the identification of problem nodes (Oliveira et al., 2022).
Artificial intelligence (AI), which encompasses a wide variety of learning strategies through its many different methodologies. The following are some examples of machine learning models that make use of ML techniques: classification, clustering, regression, dimensionality reduction, anomaly detection, and reward maximisation (Yang et al., 2022).
In this context, SVM are used to undergo proper training on labelled sets. These training sets generally demonstrate that, provided with sufficient data, machine learning models are produces a ground-truth. (Wang et al., 2022) contends that the primary focus of UL algorithms is not on giving ground-truth results but rather on locating data patterns.
However, rather than using a convolutional neural network methodology, (Ma et al., 2022) analysed 184,319 reported cases. This was in contrast to the supervised learning approach that they used in their earlier study. In order to uncover strategies to prevent the SARS-CoV-2 epidemic from occurring again, researchers examine the medical records of individuals who died from the virus. Developing a machine-learning that assist primary care clinicians in determining which patients were at a high risk of getting COVID-19 and which patients would require intubation or mechanical breathing was the goal that we set out to accomplish.
In the paper (Burkholz et al., 2021) developed a model that applies a number of different algorithms, one of which is the K-nearest neighbours classifier, in order to accurately identify COVID-19 cases in patients by assessing medical information in the appropriate context.
Supervised learning in (Wang et al., 2023) using neural network do data analysis. According to the findings of their research, an embedded software platform that allowed to sped up the process of development, adoption, and rollout in the healthcare industry. Therefore, even in the most complex situations, the system is capable of simulating automated lung segmentation and opacity assessment to a human-level quality in as little as ten days, which makes it appropriate for use in medical applications (Xie et al., 2022).
Supervisory learning and regression analysis were the methodologies that (Li et al., 2022) decided to use as their approach of choice. The polynomial regression emerged as the clear victor among the four models that were assessed; this model will serve the government well in both the short term and the long term as it assists in the planning of its operations for the upcoming month. In addition, the programme included a well-known log, which assisted in the maturation of the regression model and provided a forecast regarding the peak and end of the epidemic in the year 2020. In addition, each and every piece of information on the total number of COVID-19 instances is right at your fingertips.
As a supplemental method for the investigation, they also made use of supervised learning and an artificial neural network. When compared to other successful tactics, the findings demonstrated that machine learning models achieved the highest level of accuracy in predicting significant COVID-19 patterns. Some research indicates that artificial intelligence might be effective in properly estimating the risk of COVID-19 patients, which might improve the management of outcomes and the efforts to triage patients (Jiang et al., 2022).

3. PROPOSED METHOD

Figure 1 illustrates the processes that were utilised to develop these models. Both the training and testing sets include preprocessing, feature extraction, and classification as components of their respective sets. This would allow for optimising the classification process.

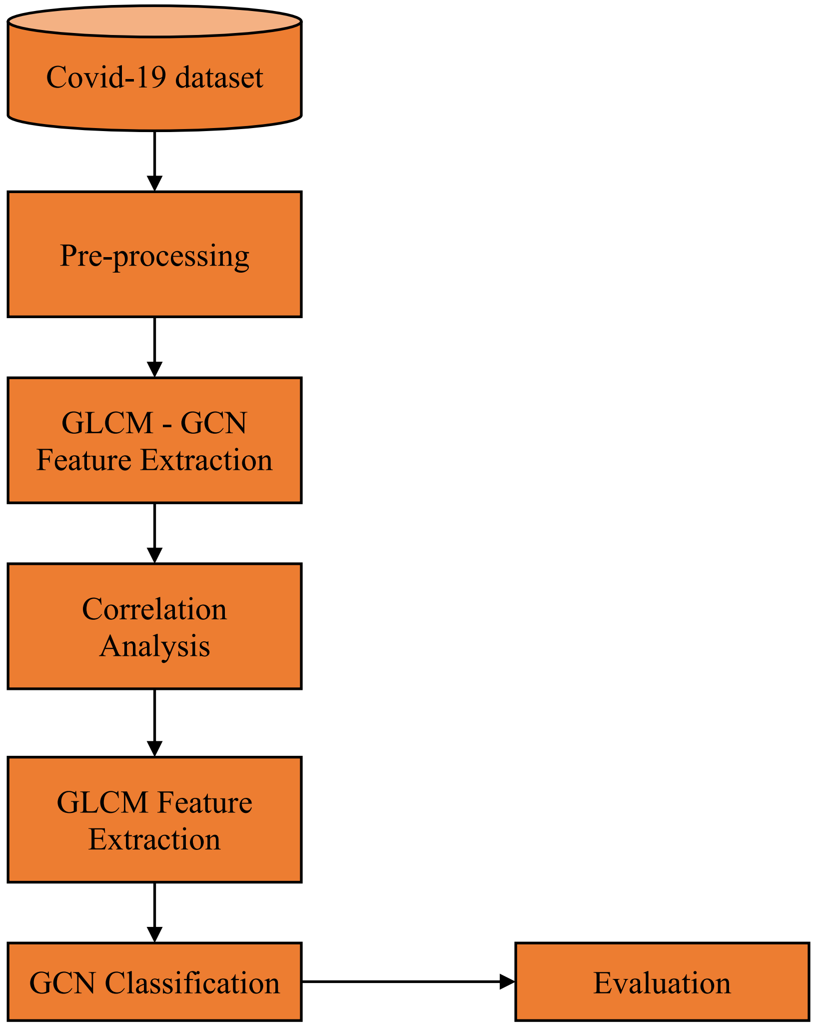

Figure 2: Proposed Disease Outbreak Prediction

GLCM Feature Extraction

A great number of image processing methods make use of a technique known as Grey Level Co-Occurrence Matrix (GLCM) based feature extraction. For the purpose of determining the texture relationship between each individual image pixel, the GLCM makes use of second-order statistics. In most cases, even two pixels will be sufficient for this function. For the purposes of this discussion, the GLCM is responsible for determining the coupling frequency of these pixel brightness values. It is a depiction of the frequency with which pairs of pixels appear, to put it another way. An image must have a mathematical representation of its GLCM features in order to be considered. It is important that the number of rows and columns in the matrix that you construct corresponds to the grayscale levels that are present in the image. Within this matrix, the grey values that are present in each row and column represent the statistical probability of the second order. The transient matrix, on the other hand, will be exceptionally large because of the distributed nature of the intensity values. As a consequence of this, the existing workflow becomes significantly more difficult. This analysis took into consideration the following qualities of the GLCM:

By making use of these characteristics, a GLCM feature matrix has the potential to describe an image with a reduced number of parameters.

The research takes use of a technique that involves the extraction of features. In order to classify the pixels, this method makes use of the following:

$$LBP(P,R)=\sum_{P=0}^{P-1} s\left(g_p-g_c\right)2^P$$

$$s(x)=\begin{cases}1 & x\geq 0\\ 0 & x<0\end{cases}$$

Where

g_c - gray value of central pixels,

g_p - neighborhood value for the center pixel.

P - total number of neighborhoods and

R - radius of the neighborhood.

We use LBP to generate feature matrices, each of which contains 10 features. Utilising these property for classifier training is the next step.

GCN

GCN nodes allow for the replication of brain processes. Binary digits or a combination of the two are used by brain neuronal connections to communicate and exchange information. Every neuron's precise weight determines its function and part in the system. The levels comprising a GCN may be referred to as input, extraction, or classification layers, depending on the describing party. Various people use various languages to define these stages. Every layer must perform a certain function and convert input into meaningful data to add to the ultimate outcome. An important component of neuronal activity is the activation and transfer function. Below are the steps needed to convert the weighted input into the finished product:

$z=\sum_{nn=1} w_i x_i + w_b b$

where,

w - weight,

b – bias and

x – input

4. RESULTS AND DISCUSSIONS

We conducted the supervised machine learning algorithms in Python, which consists of 708 images of both benign and malignant objects, using a laptop with an i5 CPU, 8 GB of random access memory (RAM), and a processing speed of 2.8 gigahertz. With all the needed libraries installed, we could build models and examine correlations in a Python notebook.

Use the following metric to determine whether or not something is accurate:

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

Using the sensitivity metric, the following is an example of how to determine how accurate the model was in estimating the percentage of patients who tested positive for COVID-19:

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

Specificity shows the proportion of patients whose negative results are in fact positive diagnoses, is as follows:

$$\text{Specificity} = \frac{TP}{TN+FP}.$$

where

TP - true positive,

TN - true negative,

FP - false positive and

FN - false negative.

With the help of an epidemiology dataset that had both positive and negative cases, the machine learning models for COVID-19 infection were training. Some machine learning approaches, including GCN, SVM, and ANN, were utilised in the construction of these models. Before beginning the process of creating the model, it was essential to investigate the correlation coefficients that existed between the independent and dependent characteristics. All of the dependent features in the dataset generate a strong positive correlation with the independent variables, which creates a robust positive association. Only a modest correlation exists between the independent variable and any of the associated dependent characteristics.

These categories are positive and negative. By utilising this approach, we are able to train the model with eighty percent of the data while simultaneously verifying the classifier with twenty percent of the data. Using sensitivity, specificity, and accuracy measures, we are able to evaluate the efficiency and precision with which the models are able to make predictions.

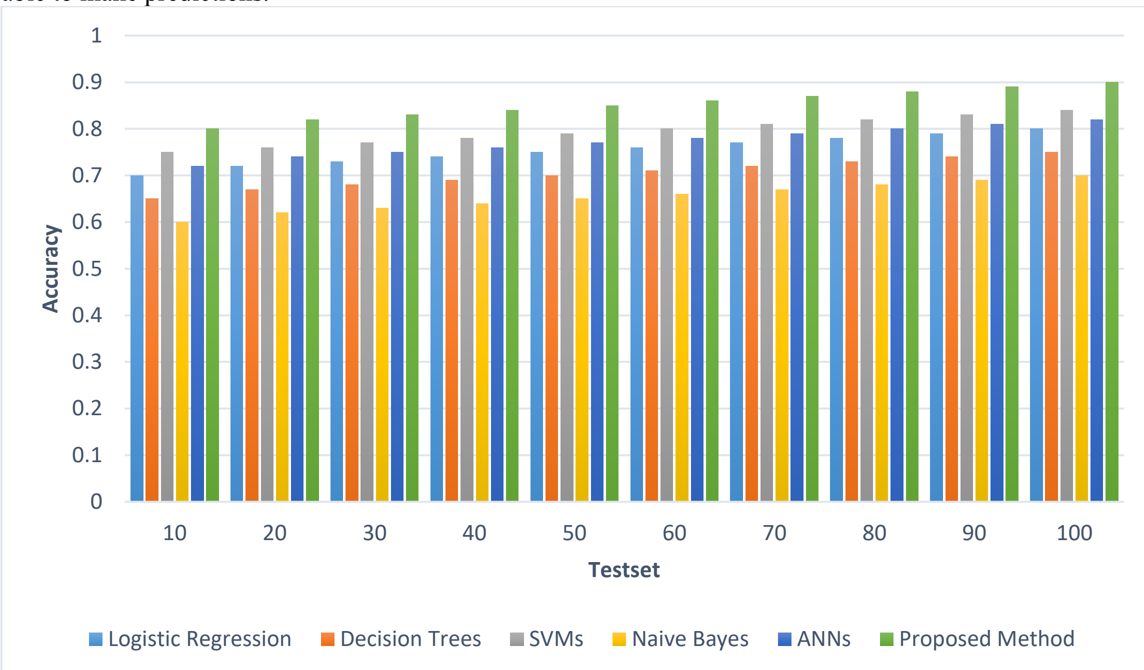

Figure 3: Accuracy

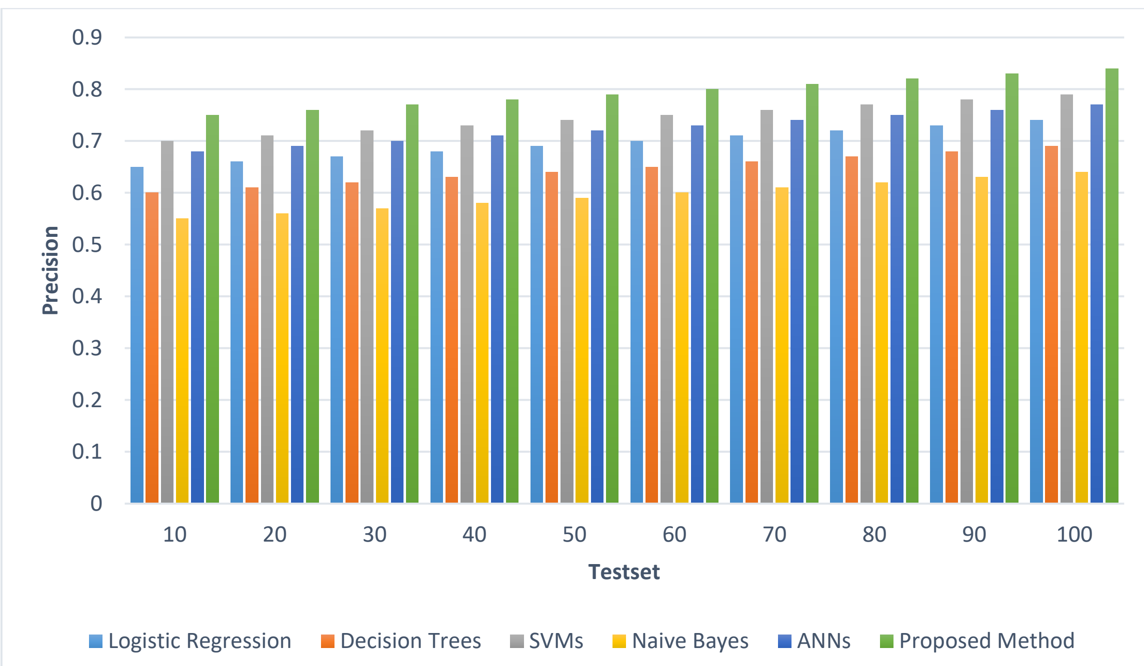

Figure 4: Precision

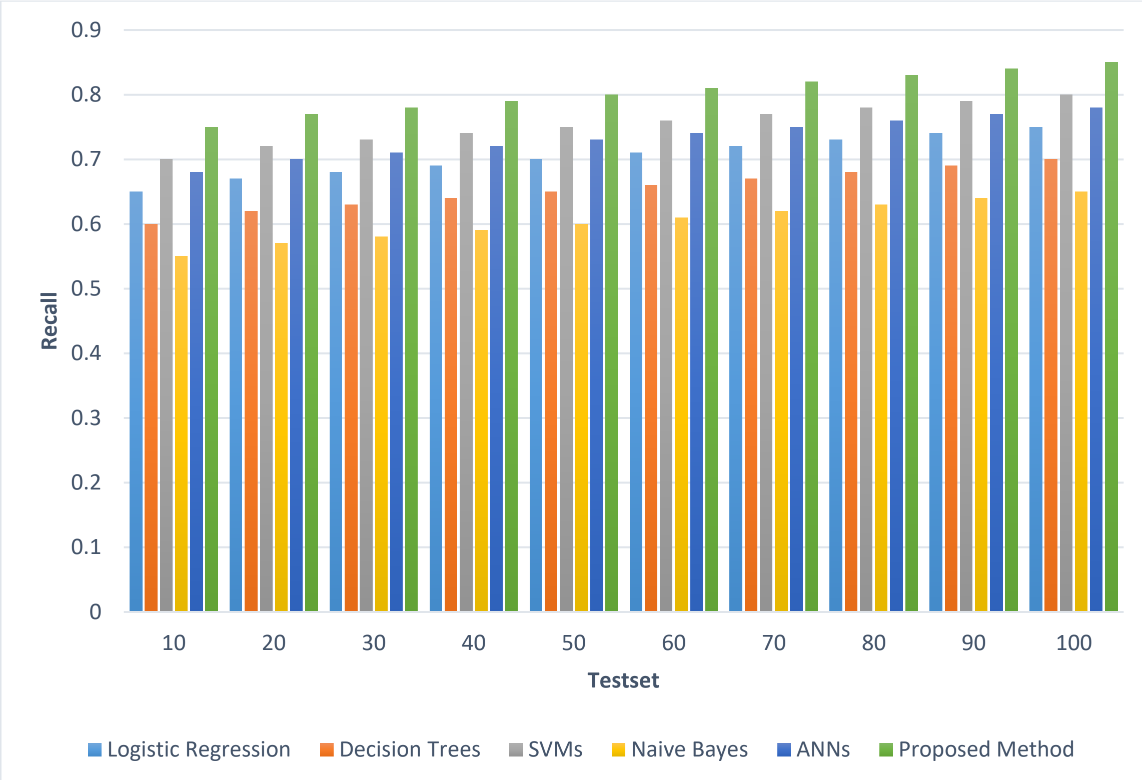

Figure 5: Recall

Figure 6: F-measure

Figure 7: MAPE

DISCUSSION AND LIMITATIONS

Currently, there is a lack of a warning system for COVID-19 patients, which could potentially lead to a quicker diagnosis. For the purpose of developing classification models for the virus, this research made use of a GCN supervised learning dataset that contained both positive and negative COVID-19 cases. Through the use of their respective accuracy measures, the results of analysing the overall performance of each model are displayed in Figures 3–7.
The research demonstrates that supervised machine learning models can effectively classify COVID-19 cases. Despite the strong positive correlation between independent and dependent variables, the models as depicted in Figures 3-Figure 6. Also, the further metrics like Sensitivity and specificity metrics further helps to evaluate the models' performance, highlighting their potential in accurately predicting infections. However, limitations include the absence of an early warning system for prompt diagnosis, underscoring the need for further refinement and real-world application of these models.

5. CONCLUSIONS

This work includes a detection module that comprises preprocessing, feature extraction, and the identification of COVID-19 patients through CT images. This module is present in this study. It is necessary to perform GLCM-based fundamental image processing and feature extraction on the images before moving on to algorithmic classification. In the course of the simulation, the researchers will have the opportunity to evaluate the model's capacity to predict CT images from a wide variety of patients. Training the model uses up 80% of the data, while post-development testing and evaluation use the remaining 20% of the data. The decision tree model emerged victorious with a success rate of 94%, despite the fact that every model was capable of producing accurate results. Additionally, when compared to the average of the solutions that are currently available, the technique that has been suggested is 2% better than the average. As a consequence of this, SVM models achieved a combined sensitivity and specificity of 93%, which was superior to the performance of any other models. There is a possibility that in the future, additional enhancements to feature extraction will be attainable with the assistance of advanced models that permit the possibility of a reduction in the number of features.
In future enhancement this work focuses on integrating the machine or deep learning models into real time diagnostic system which improves the accuracy and faster detection of Covid 19 disease. Also, it should focus on expanding the

diverse dataset and comprehensive model of patient data set which improves the robustness and generalizability. Efforts should also be directed towards developing an early warning system to enhance proactive healthcare responses.

REFERENCES

[1] Anno, S., Hirakawa, T., Sugita, S., & Yasumoto, S. (2022). A graph convolutional network for predicting COVID-19 dynamics in 190 regions/countries. *Frontiers in Public Health, 10*, Article 911336. https://doi.org/10.3389/fpubh.2022.911336

[2] Wang, Y., Zeng, D. D., Zhang, Q., Zhao, P., Wang, X., Wang, Q., & Cao, Z. (2022). Adaptively temporal graph convolution model for epidemic prediction of multiple age groups. *Fundamental Research, 2*(2), 311-320. https://doi.org/10.1016/j.fmre.2022.01.009

[3] Saravanan, V., Parameshachari, B. D., Hussein, A. H. A., Shilpa, N., & Adnan, M. M. (2023, November). Deep learning techniques based secured biometric authentication and classification using ECG signal. In *2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICIICS58830.2023.10204233

[4] Yadav, R. S., Singh, A., & Yadav, S. P. (2024, March). The smart fuzzy-based image recognition model for modern healthcare. In *2024 2nd International Conference on Disruptive Technologies (ICDT)* (pp. 250-254). IEEE. https://doi.org/10.1109/ICDT55535.2024.00047

[5] Saravanan, V., Sankaradass, V., Shanmathi, M., Bhimavarapu, J. P., Deivakani, M., & Ramasamy, S. (2023, May). An early detection of ovarian cancer and the accurate spreading range in the human body by using a deep medical learning model. In *2023 International Conference on Disruptive Technologies (ICDT)* (pp. 68-72). IEEE. https://doi.org/10.1109/ICDT55535.2023.00107

[6] Shesayar, R., Agarwal, A., Taqui, S. N., Natarajan, Y., Rustagi, S., Bharti, S., & Sivakumar, S. (2023). Nanoscale molecular reactions in microbiological medicines in modern medical applications. *Green Processing and Synthesis, 12*(1), Article 20230055. https://doi.org/10.1515/gps-2023-0055

[7] Yu, S., Xia, F., Wang, Y., Li, S., Febrinanto, F. G., & Chetty, M. (2022). PANDORA: Deep graph learning-based COVID-19 infection risk level forecasting. *IEEE Transactions on Computational Social Systems*. ttps://doi.org/10.1109/TCSS.2022.3147526

[8] Yu, S., Xia, F., Li, S., Hou, M., & Sheng, Q. Z. (2023). Spatio-temporal graph learning for epidemic prediction. *ACM Transactions on Intelligent Systems and Technology, 14*(2), 1-25. https://doi.org/10.1145/3579196

[9] Fritz, C., Dorigatti, E., & Rügamer, D. (2022). Combining graph neural networks and spatio-temporal disease models to improve the prediction of weekly COVID-19 cases in Germany. *Scientific Reports, 12*(1), Article 3930. https://doi.org/10.1038/s41598-022-07887-4

[10] Du, Y., Wang, H., Cui, W., Zhu, H., Guo, Y., Dharejo, F. A., & Zhou, Y. (2021). Foodborne disease risk prediction using multigraph structural long short-term memory networks: Algorithm design and validation study. *JMIR Medical Informatics, 9*(8), Article e29433. https://doi.org/10.2196/29433

[11] Sun, C., Fang, R., Salemi, M., Prosperi, M., & Rife Magalis, B. (2024). DeepDynaForecast: Phylogenetic-informed graph deep learning for epidemic transmission dynamic prediction. *PLOS Computational Biology, 20*(4), Article e1011351. https://doi.org/10.1371/journal.pcbi.1011351

[12] La Gatta, V., Moscato, V., Postiglione, M., & Sperli, G. (2020). An epidemiological neural network exploiting dynamic graph structured data applied to the COVID-19 outbreak. *IEEE Transactions on Big Data, 7*(1), 45-55. https://doi.org/10.1109/TBDATA.2019.2952497

[13] Li, D., Ren, X., & Su, Y. (2023). Predicting COVID-19 using lioness optimization algorithm and graph convolution network. *Soft Computing, 27*(9), 5437-5501. https://doi.org/10.1007/s00500-022-06181-0

[14] Tariq, S., Tariq, S., Kim, S., Woo, S. S., & Yoo, C. (2023). Distance adaptive graph convolutional gated network-based smart air quality monitoring and health risk prediction in sensor-devoid urban areas. *Sustainable Cities and Society, 91*, Article 104445. https://doi.org/10.1016/j.scs.2022.104445

[15] Oliveira, L. C., Oliva, J. T., Ribeiro, M. H., Teixeira, M., & Casanova, D. (2022). Forecasting the COVID-19 space-time dynamics in Brazil with convolutional graph neural networks and transport modals. *IEEE Access, 10*, 85064-85079. https://doi.org/10.1109/ACCESS.2022.3193269

[16] Yang, C., Song, H., Tang, M., Danon, L., & Vigfusson, Y. (2022, August). Dynamic network anomaly modeling of cell-phone call detail records for infectious disease surveillance. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 4733-4742). https://doi.org/10.1145/3539597.3561763

[17] Wang, L., Adiga, A., Chen, J., Sadilek, A., Venkatramanan, S., & Marathe, M. (2022, June). Causalgnn: Causal-based graph neural networks for spatio-temporal epidemic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 11, pp. 12191-12199). https://doi.org/10.1609/aaai.v36i11.21157

[18] Ma, Y., Gerard, P., Tian, Y., Guo, Z., & Chawla, N. V. (2022, October). Hierarchical spatio-temporal graph neural networks for pandemic forecasting. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management* (pp. 1481-1490). https://doi.org/10.1145/3511808.3511882

[19] Burholz, R., Quackenbush, J., & Bojar, D. (2021). Using graph convolutional neural networks to learn a representation for glycans. *Cell Reports, 35*(11), Article 109250. https://doi.org/10.1016/j.celrep.2021.109250

[20] Wang, M., Yang, B., Liu, Y., Yang, Y., Ji, H., & Yang, C. (2023). Emerging infectious disease surveillance using a hierarchical diagnosis model and the Knox algorithm. *Scientific Reports, 13*(1), Article 19836. https://doi.org/10.1038/s41598-023-46631-0

[21] Xie, F., Zhang, Z., Li, L., Zhou, B., & Tan, Y. (2022, September). EpiGNN: Exploring spatial transmission with graph neural network for regional epidemic forecasting. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 469-485). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-08734-9_27

[22] Li, Q., Pan, Q., & Xie, L. (2022). Prediction of spread trend of epidemic based on spatial-temporal sequence. *Symmetry, 14*(5), Article 1064. https://doi.org/10.3390/sym14051064

[23] Jiang, J. Y., Zhou, Y., Chen, X., Jhou, Y. R., Zhao, L., Liu, S., & Wang, W. (2022). COVID-19 surveiller: Toward a robust and effective pandemic surveillance system based on social media mining. *Philosophical Transactions of the Royal Society A, 380*(2214), Article 20210125. https://doi.org/10.1098/rsta.2021.0125