Imaging and Radiation Research (2023) Volume 6 Issue 1
doi: 10.24294/irr.v6i1.5451

ORIGINAL RESEARCH ARTICLE

Classification of some epidemics through microscopic images by using deep learning. Comparison

Laura Brito, Roberto Rodríguez*

Institute of Cybernetics, Mathematics & Physics (ICIMAF), Havana 10 400, Cuba

*** Corresponding author:** Roberto Rodríguez, rrm@icimaf.cu

ABSTRACT

In this study, we utilized a convolutional neural network (CNN) trained on microscopic images encompassing the SARS-CoV-2 virus, the protozoan parasite "*plasmodium falciparum*" (causing of malaria in humans), the bacterium "*vibrio cholerae*" (which produces the cholera disease) and non-infected samples (healthy persons) to effectively classify and predict epidemics. The findings showed promising results in both classification and prediction tasks. We quantitatively compared the obtained results by using CNN with those attained employing the support vector machine. Notably, the accuracy in prediction reached 97.5% when using convolutional neural network algorithms.

Keywords: deep learning; supervised learning; convolutional neural networks; support vector machines; training; neural network architectures

ARTICLE INFO

Received: 26 March 2023
Accepted: 16 April 2023
Available online: 22 May 2023

1. Introduction

Deep learning (DL) has been widely used in many fields of modern life[1,2]. DL is a subfield within machine learning (ML), and it does not require any human-designed rule to work. DL, rather needs and uses large amounts of data to establish and map a given input to specific relationships or labels. Segmentation and classification tasks using DL unlike ML (where the latter requires performing a set of sequential steps guided by the specialist), with DL, one can automatically learn a set of features from an input database, and to carry out these tasks without human intervention.

In the last decade, DL have had successes very outstanding, which have improved the quality of human life in a remarkable way with an additional accuracy in diagnosis of diseases, in study of epidemics, in research of microscopic images, in discovery of new drugs, as well as in many other areas. Literature has pointed out that the average accuracy of disease diagnosis of a DL network has been superior to that of many medical specialists[3].

Computer vision researchers increasingly use algorithms from DL to help build robust, intelligent and reusable imaging interpretation systems. For this reason, artificial systems that learn and adapt represent an important challenge in computer vision research.

Today, automated analysis of microscopic biomedical images has become very important, especially as traditional medicine starts to shift to a more preventive and predictive paradigm. Automated

analysis of microscopic images remains a considerably challenging task, mainly because microscopic images are complex and variant[4,5]. Moreover, the difference between disease and nondisease cases many times is subtle. Therefore, accurate automated analysis of microscopic images requires the development of innovative and adaptive computational models. DL approaches to microscopic image analysis, often coupled to other intelligent algorithms, have shown great promise[6,7].

Very likely that due to unhealthy and unhygienic lifestyles, poor eating habits and climate change coupled with stress; many types of viruses and bacteria are transmitted to humans producing diseases that become true epidemics, accentuated in low-income and underdeveloped countries. Such are the cases of the epidemics of malaria, cholera and covid-19, where millions of people around the world have been affected and have died due to these pandemics. For that reason, the development of new algorithms by using techniques of artificial intelligence and, basically, DL in order to face new pathologies will be always welcome in the field of medicine[8].

In this study, we utilized a convolutional neural network (CNN) trained on microscopic images encompassing the SARS-CoV-2 virus, the protozoan parasite "*plasmodium falciparum*" (causing of malaria in humans), the bacterium "*vibrio cholerae*" (which produces the cholera disease) and non-infected samples (healthy persons) to effectively classify and predict epidemics. The findings showed promising results in both classification and prediction tasks. We quantitatively compared the obtained results by using CNN with those attained employing the support vector machine. Notably, the accuracy in prediction reached 97.5% when using convolutional neural network algorithms.

The rest of the paper is organized as follows: In section 2, the materials and methods are given, and we slightly outlines some theoretical and algorithmic aspects. Here, we will specify on the database used. Section 3 contains the obtained results and discussion. We will describe our conclusions in section 4.

2. Materials and methods

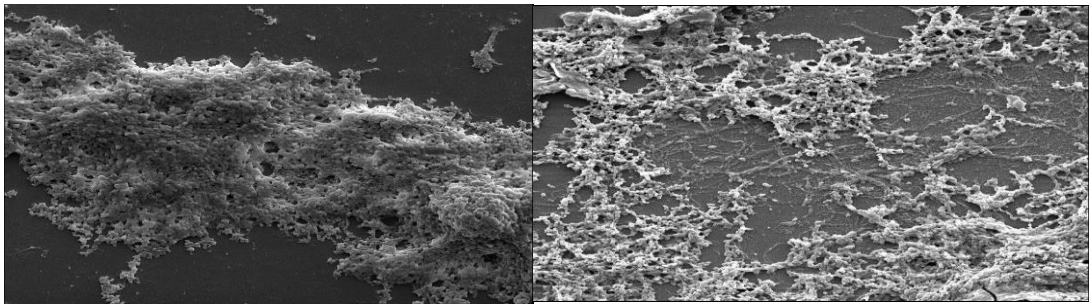
2.1. Medical methodology

In the fields of medicine and biology, detection of certain microorganisms that affect the human health is often rigorous and costly. Sometimes the results take an undetermined amount of time, time that is essential for the preservation of life.

For those reasons, predicting infections through the study of microorganisms using intelligent computational techniques is very convenient in order to reduce the clinical process of patient, reducing the waiting time and the cost involved.

We obtained the microscopic image samples of SARS-CoV-2 virus in the same way as in the studies of Rodríguez et al.[4,5]. In **Figure 1**, we show two microscopic images of a patient infected with SARS-Cov-2 virus.

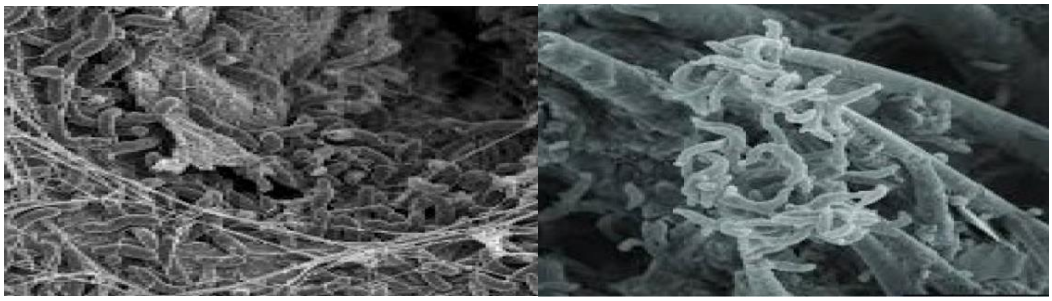
The "*vibrio cholerae*" is the causative agent of cholera, an acute diarrheal disease that occurs in form of epidemic outbreaks. The *vibrio cholerae* species are classified according to their somatic lipopolysaccharide antigens into different serogroups. The genes responsible for the O antigen synthesis are present in an area of genome called wbf. Of the two hundred serogroups identified, only serogroup O1 and O139 are recognized as the only ones responsible for cholera epidemic.



(a) **(b)**

Figure 1. Two examples of SARS-CoV-2 virus, **(a)** and **(b)** original microscopic images.

In **Figure 2**, we show microscopic images of blood samples from positive patients of cholera.

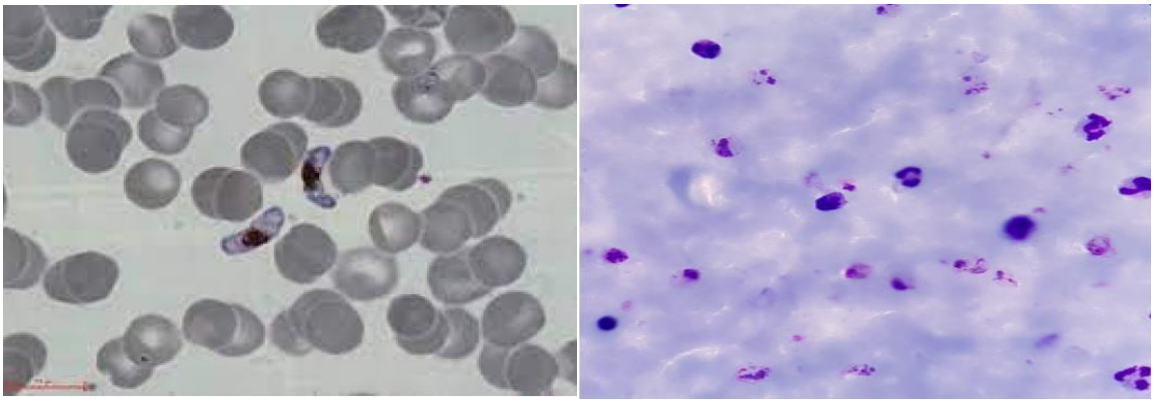


(a) **(b)**

Figure 2. Two examples of microscopic images of blood samples from positive patients infected with cholera.

The *plasmodium falciparum* is a protozoan parasite that causes the most virulent form of human malaria and kills, every year at least, thousands of children. In the case of malaria, the infection is caused by the entry of the parasite into the erythrocytes, which is responsible for acute and severe malaria. All this is a complex and dynamic process[9,10]. Therefore, the identification and classification of this blood-stage infection through microscopic images using automated artificial intelligence techniques is of vital importance for a good diagnosis and more effective treatment.

In **Figure 3**, we show microscopic images of blood samples from positive patients of malaria.



(a) **(b)**

Figure 3. Two examples of microscopic images of blood samples from positive patient infected with malaria.

The identification and classification of structures of all these microorganisms through the study of high-resolution microscopy are essential to know the etiological agents of many epidemics. For this reason, the development of new artificial intelligence techniques based on deep learning to the early study of new epidemics will always be welcome in the field of medicine.

2.2. Deep learning

We use the deep learning (DL), employing high-resolution microscopy, for the study of the structures of microorganisms, as it is known that the DL does not need any human-designed rules to work. DL uses large volumes of data in order to establish and map the given input to specific relations, and has the ability of learning a feature set from input data. Unquestionably, the successes of DL have been outstanding, improving the quality of human life with an additional accuracy in diagnosis of many pathologies, and in the discovery of new drugs, among many other good results[3]. However, despite the undisputed merits of DL, its main disadvantage is that need of a large database that serve as training set. Otherwise, training can result in poor quality and an undesired result can occur, especially in problems of visual pattern recognition and classification[11].

Undoubtedly, to learn and classify tens of thousands of objects and patterns from a huge number of images, one needs a model that has a large learning capacity; and on the other hand, such a model must have prior knowledge to be able to compensate for data that one does not have. One such class of models is deep convolutional neural networks (CNNs)[12]. In this paper, we quantitatively compared the obtained results by using CNNs with those attained employing the support vector machine (SVM).

2.2.1. Database

For an effective comparison, we built a standard database (microscopic images were resized to a size of 100 × 100 pixels) and then performed the training task. We used a specific metric (to be detailed in the experimental results section) because the database was unbalanced in the number of microscopic images per class, which was evident in the learning models, especially in the cholera class.

Thus, the database of microscopic images (7270 samples) encompassed the SARS-CoV-2 virus (1055 micro-images), the protozoan parasite "*plasmodium falciparum*" (3180 micro-images), the bacterium "*vibrio cholerae*" (905 micro-images) and non-infected samples (2130 micro-images).

We separated the database composed of four classes into two groups: the set of microscopic images for the training and another for the validation process.

2.2.2. Proposed architectures

In practice, one can quantitatively measure the performance of a deep learning model previously trained. Thus, we implemented different metrics to evaluate the performance of the learning process in order to improve the predictive power of the models. In this work, we use the following metric: Accuracy, recall, F1-score, confusion matrix and precision[13].

Table 1. Description of features of the proposed CNN models.

Considerations	Model I	Model II	Model III	Model IV
Database	Total number of microscopic images: 7270 Micro-images of healthy samples (non-infection): 2130 Micro-images of *plasmodium falciparum* samples: 3180 Micro-images of Sars-CoV-2 samples: 1055 Micro-images of *vibrio cholerae*: 905			
Normalization	Data Augmentation and Drop Out[14–16]			
Layers	6 ReLu layers, 6 max pooling layers, drop out 50% of neurons in the hidden layer	6 ReLu layers, 6 max pooling layers, drop out 25% of neurons in the hidden layer	3 ReLu layers, 3 max pooling layers, drop out 50% of neurons in the hidden layer	3 ReLu layers, 3 max pooling layers, drop out 25% of neurons in the hidden layer
Optimization	Root Mean Square Propagation[17]			
Loss function	Weighted Cross-Entropy Loss Function[17]			
Validation	We choose from initial dataset the 20% of them			

We proposed four models of CNN architectures, and selected the best one according to the evaluation metrics. In **Table 1**, we will describe the characteristics of the architectures of the different CNN models. In addition, in the experimental results, we will present the obtained scores by metrics for each proposed models. In that section, one will appreciate why the Model IV was the chosen one and whose architecture is the following:

a) Data augmentation via ImageDataGenerator[17].
b) Convolutional layer with ReLu activation function: with 3 × 3 kernel size for convolution (contains 32 neurons).
c) Max pooling layer: 2 × 2 pooling size.
d) Convolutional layer with ReLu activation function: with 2 × 2 kernel size for convolution (contains 64 neurons).
e) Max pooling layer: 2 × 2 pooling size.
f) Convolutional layer with ReLu activation function: with 2 × 2 kernel size for convolution (contains 64 neurons).
g) Max pooling layer: 2 × 2 pooling size.
h) Flatten layer.
i) ReLu dense layer (contains 256 neurons)[18].
j) Drop out 25% of neurons in the hidden layer.
k) SoftMax classification layer.

In **Figure 4**, we show the architecture of Model IV.

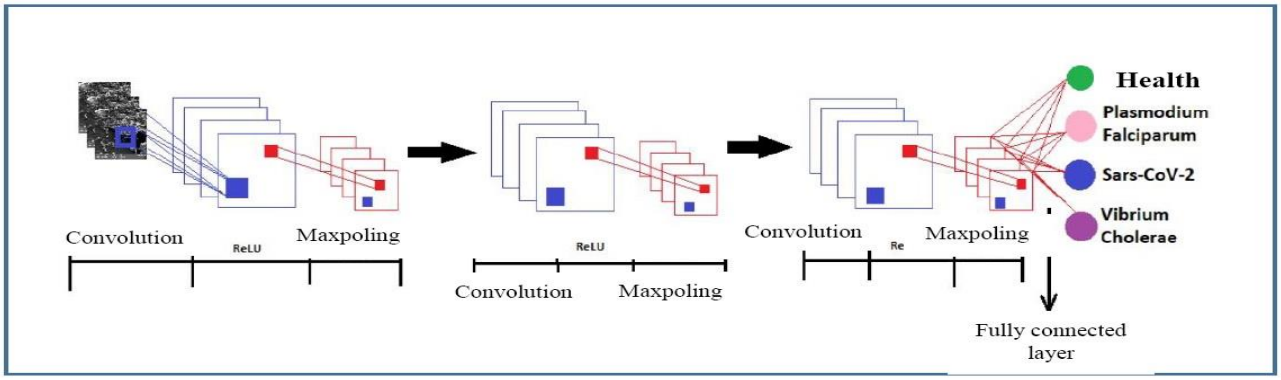


Figure 4. Architecture of Model IV.

3. Experimental results: Analysis and discussion

Due to the variety of components and hyper-parameters possible to employ in a deep neural network model, one can establish a large number of combinations in the creation of such model. For this reason, it was necessary to carry out a study of the state of art of different architectures used in classifications and predictions[15,16,19,20] (see **Table 1**).

We will show in **Tables 2–5** the results of the evaluation metrics for the different models, and will carry out a quantitative comparison among them in order to select the best model in the prediction of these epidemics.

In order to measure the performance of trained algorithms in multi-class databases (to classification and prediction tasks), it is possible to use three types of metrics: micro-average method, macro-average method or weights-average method. For example, one can use the macro-average method when one wants to know how the system performs overall across the database, but should not come up with any specific decision with this average. In other words, the macro-average method considers all classes as basic elements of the calculation: each class has the same weight in the average, so that there is no distinction between highly and poorly populated classes. Therefore, it is not convenient to select this macro-average approach when the datasets are

very unbalanced. However, micro-average method can be a useful measure when the datasets varies in size[20].

In this research, the datasets are very unbalanced (see the difference in size between the *plasmodium falciparum* class and *vibrio cholerae* class). For this reason, we chose the micro-average method for carrying out the quantitative comparison, and to determine, from the proposed CNN models, which one had the best performance. Now, we will show all the metrics in Tables for a better visualization and comparison of them.

Table 2. Results of evaluation metrics for the Model I.

	Classes			
Metrics	**Healthy**	***Plasmodium falciparum***	**Sars-CoV-2**	***Vibrio cholerae***
Precision	0.9942	0.9574	0.8736	0.9565
Recall	0.9157	1.0	0.9880	0.5238
F1-score	0.9534	0.9782	0.9273	0.6769
Precision (micro)	0.9556			
Precision (macro)	0.9454			
Precision (weights)	0.9573			
Recall (micro)	0.9556			
Recall (macro)	0.8569			
Recall (weights)	0.9556			
F1-score (micro)	0.9556			
F1-score (macro)	0.8840			
F1-score (weights)	0.9523			
Confusion Matrix	$\begin{bmatrix} 174 & 16 & 0 & 0 \\ 0 & 518 & 0 & 0 \\ 0 & 0 & 83 & 1 \\ 1 & 7 & 12 & 22 \end{bmatrix}$			
Accuracy	0.9556			

F1-score assesses the performance of classification model starting from the confusion matrix, where F1-score can be interpreted as a weighted average between precision and recall, and this reaches its best value at one (1) and worst score at zero (0). In the multi-class cases, F1-score should involve all the classes, and so this metric can have two different specifications: Micro F1-Score and Macro F1-Score, where the macro model considers all the classes as basic elements of the calculation[21,22]. Thus, one can see that the Model IV obtained the best score when using the micro average method (remember that micro averaging considers all the units together, without taking into consideration possible differences between them).

To further deep why the Model IV was selected. For example, if one analyzes the Accuracy metric, it is evident that the highest was that of Model IV (0.9846), and the same is true for the micro precision metric (0.9844) and the micro F1-score (0.9845). The same is true when analyzing the confusion matrices, which the best was that of Model IV.

Table 3. Results of evaluation metrics for the Model II.

	Classes			
Metrics	**Healthy**	***Plasmodium falciparum***	**Sars-CoV-2**	***Vibrio cholerae***
Precision	1.0	0.9333	0.9230	1.0
Recall	0.8368	1.0	1.0	0.6904
F1-score	0.9111	0.9651	0.9600	0.8169
Precision (micro)	0.9472			
Precision (macro)	0.9641			
Precision (weights)	0.9508			
Recall (micro)	0.9472			
Recall (macro)	0.8818			
Recall (weights)	0.9472			

Table 3. (*Continued*).

	Classes			
Metrics	**Healthy**	***Plasmodium falciparum***	**Sars-CoV-2**	***Vibrio cholerae***
F1-score (micro)	0.9472			
F1-score (macro)	0.9133			
F1-score (weights)	0.9450			
Confusion Matrix	$\begin{bmatrix} 159 & 31 & 0 & 0 \\ 0 & 518 & 0 & 0 \\ 0 & 0 & 84 & 1 \\ 0 & 6 & 7 & 29 \end{bmatrix}$			
Accuracy	0.9472			

Table 4. Results of evaluation metrics for the Model III.

	Classes			
Metrics	**Healthy**	***Plasmodium falciparum***	**Sars-CoV-2**	***Vibrio cholerae***
Precision	1.0	0.9829	0.9438	1.0
Recall	0.9842	1.0	1.0	0.7380
F1-score	0.9920	0.9913	0.9710	0.8493
Precision (micro)	0.9832			
Precision (macro)	0.9816			
Precision (weights)	0.9837			
Recall (micro)	0.9832			
Recall (macro)	0.9305			
Recall (weights)	0.9832			
F1-score (micro)	0.9832			
F1-score (macro)	0.9509			
F1-score (weights)	0.9823			
Confusion Matrix	$\begin{bmatrix} 187 & 3 & 0 & 0 \\ 0 & 518 & 0 & 0 \\ 0 & 0 & 84 & 0 \\ 0 & 6 & 5 & 31 \end{bmatrix}$			
Accuracy	0.9832			

Therefore, to be precise, in the selection of a better deep learning model for automatic multi-class classification, working with the Micro approach is the most correct way, on all when the datasets is very unbalanced. In this case, if one obtain a poor performance on small classes is not important, since the number of units belonging to those classes is small compared to the database size[21].

Table 5. Results of evaluation metrics for the Model IV.

	Classes			
Metrics	**Healthy**	***Plasmodium falciparum***	**Sars-CoV-2**	***Vibrio Cholerae***
Precision	0.9947	0.9904	0.9325	0.9687
Recall	0.9947	1.0	0.9880	0.7380
F1-score	0.9947	0.9951	0.9595	0.8378
Precision (micro)	0.9844			
Precision (macro)	0.9716			
Precision (weights)	0.9844			
Recall (micro)	0.9844			

Table 5. (*Continued*).

	Classes			
Metrics	**Healthy**	***Plasmodium falciparum***	**Sars-CoV-2**	***Vibrio Cholerae***
Recall (macro)	0.9302			
Recall (weights)	0.9844			
F1-score (micro)	0.9845			
F1-score (macro)	0.9468			
F1-score (weights)	0.9835			
Confusion Matrix	$\begin{bmatrix} 189 & 1 & 0 & 0 \\ 0 & 518 & 0 & 0 \\ 0 & 0 & 83 & 1 \\ 1 & 4 & 6 & 31 \end{bmatrix}$			
Accuracy	0.9846			

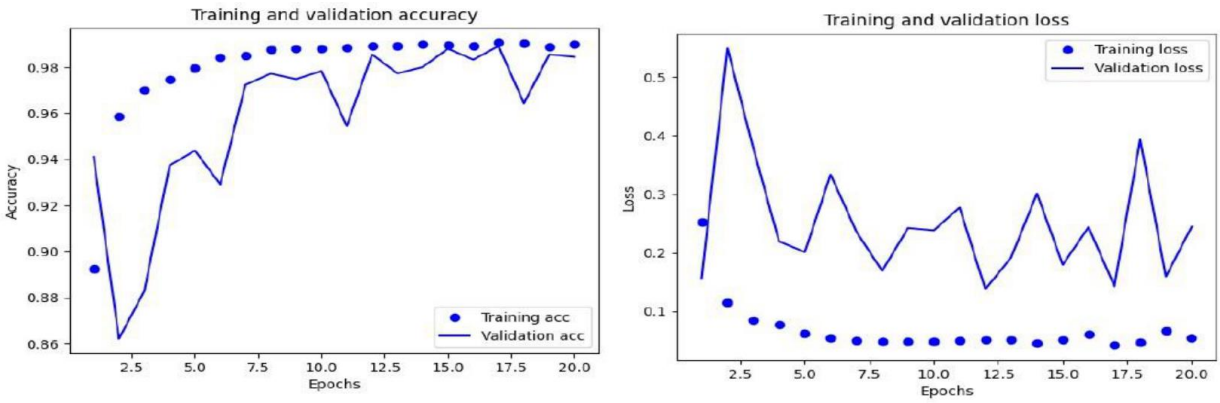
3.1. Learning mode

Once we selected the model, the following step was to carry out the training, which was performed in 20 epochs, and for each epoch 500 validation steps were realized.

Figure 5a shows the adjustment of the weights, which was produced in an ascending way by epochs. For example, when selecting accuracy as metric and despite presenting a little overfitting, one can said that the adjustment had a desired behavior, that is, a correct learning process. In this case, overfitting is denoted in the peaks of the continuous line that appeared by epochs (validation set curve), which means that the training was not completely uniform with respect to the accuracy metric, although there were epochs that were somewhat close to the training set curve (dotted curve).

However, despite the non-uniformity in the learning due to the appearance of some peaks in certain epochs that move away from a monotonous upward behavior, these peaks decreased in frequency as the epochs progressed, which indicated again a good learning process.

Figure 5b shows the graphic of the loss function. One can see in the curve that a bit of overfitting is also evident, when observing the random peaks that emerged with the advancement of epochs. This behavior indicated that the neuron weights were not uniformly adjusted in the validation process at each epoch, as happened in the training.



(a) **(b)**

Figure 5. Learning curves of Model IV, **(a)** accuracy; **(b)** loss function.

In many cases, one can evaluate the models through the analysis of graphics (see **Figure 5**), which sometimes this do not provide very accurate results when one works with a single metric and wants comparing different models. For this reason, one must use other evaluation metrics (precision, recall, F1-score, etc.) to carry out a more analytical and in-depth study of models.

3.2. Comparison of the obtained results with CNNs and support vector machine (SVM)

We will carry out a quantitative comparison between the obtained results in the prediction of epidemics using CNNs with those achieved using SVM. For this comparison, we also worked with four SVM models, which are shown in **Table 6**. Here, our goal is not to give an exhaustive explanation of SVM, which is a well-known machine learning technique. This method was chosen for comparison because it has proven to be effective, and it was convenient to compare the results obtained with CNNs with a classical method such as SVM.

Table 6. Description of features of the SVM models.

Aspects	**Model V**	**Model VI**	**Model VII**	**Model VIII**
Dataset	Total number of microscopic images: 7270 Micro-images of healthy samples (non-infection): 2130 Micro-images of *plasmodium falciparum* samples: 3180 Micro-images of Sars-CoV-2 samples: 1055 Micro-images of *vibrio cholerae*: 905			
Kernel	Linear	Sigmoid	Radial Basis	Polynomial
Validation	We choose the 25% of initial dataset through the K-Fold algorithm[22]			

Based on the metrics mentioned above, we carried out tests with each of models shown in **Table 6**. Due to space constraints, we will only present the obtained results with Model VIII, which was the model chosen. For example, the Model VIII, relative to the other SVM models, had the highest score of metrics. Thus, micro accuracy metric, micro precision metric and micro F1-score were higher in this model, and the Confusion matrix had lower false positives and negatives. These are shown in **Table 7**.

In the quantitative comparison between the Model IV (CNN) and the Model VIII (SVM), we obtained an Accuracy of 98%, which is a measure of a good performance of both algorithms.

However, when the quantitative analysis is carried out by using a single metric, one should not be absolute and certify that it is in presence of fully reliable models without a detailed study of the database. In effect, the accuracy considers (in the numerator) the sum of true positive (TP) and true negative (TN) elements, and at the denominator, the sum of all the entries of the confusion matrix (elements incorrectly classified by the

model). Therefore, the Accuracy returns an overall measure of how much the model correctly predicted on the entire dataset, and each unit contributed with the same weight to the accuracy value.

Table 7. Results of evaluation metrics for the SVM model VIII.

	Classes			
Metrics	**Healthy**	***Plasmodium falciparum***	**Sars-CoV-2**	***Vibrio cholerae***
Precision	1.0	0.99	0.99	0.77
Recall	1.0	0.99	1.0	0.77
F1-score	1.0	0.99	0.99	0.77
Precision (micro)	0.9841			
Precision (macro)	0.9364			
Precision (weights)	0.9841			
Recall (micro)	0.9841			
Recall (macro)	0.9390			
Recall (weights)	0.9844			
F1-score (micro)	0.9841			
F1-score (macro)	0.9377			
F1-score (weights)	0.9841			

Table 7. (*Continued*).

	Classes			
Metrics	**Healthy**	***Plasmodium falciparum***	**Sars-CoV-2**	***Vibrio cholerae***
Confusion Matrix	$\begin{bmatrix} 303 & 0 & 0 & 0 \\ 0 & 761 & 1 & 9 \\ 0 & 0 & 84 & 0 \\ 0 & 9 & 0 & 30 \end{bmatrix}$			
Accuracy	0.9841			

But, in the multiclass classification, a class can have a more important weight than another, since there will be classes with a high number of units and others with few ones. In this situation, as it happened in our database, the highly populated classes will have higher weight compared to the smallest ones. When the datasets are imbalanced, the *Accuracy* tends to hide classification errors for classes with less elements, since those classes have few weight compared to the biggest ones. Therefore, one should address the analysis towards other metrics, in order to carry out a global study in the validation of the performance of a model.

For example, in the case of the selected CNN model, an analysis by class using the micro F1-score metric showed a score we presented in **Table 8**. Remember that F1-score is the harmonic mean between precision and recall.

Table 8. Results of the F1-score for CNNs.

Datasets	**% F1-score**
Plasmodium falciparum	99.51
Sars-CoV-2	95.95
Vibrio cholerae	83.78
Healthy	99.87

It is clear that the results shown in **Table 8** are very promising. These results highlighted that a deep convolutional network is capable of achieving reliable classifications on a challenging dataset by using supervised learning. Here, we verified that the network degraded significantly as the number of deep layers and the degree of neuron drop out increased, resulting in a loss of approximately 4% in network performance.

In **Table 9**, we showed the obtained results using SVM. One can see that these results via SVM are also remarkable. The SVM model showed higher reliability in the prediction of SARS-CoV-2. However, the CNN model had better prediction in *Vibrio Cholerae*, which is the class with the smallest size.

Table 9. Results of the F1-score for SVM.

Datasets	**% F1-score**
Plasmodium falciparum	99.00
Sars-CoV-2	99.00
Vibrio cholerae	77.00
Healthy	99.78

We will carry out an analysis of these results. It is known that despite the undisputed merits of deep learning (DL), its main disadvantage is that it needs of a large database that serve as training set. The larger the database, the more the network learns, but more time for training too. However, in this case and due to the somewhat black-box behavior of DL algorithms (CNNs), it is not completely clear why it learned SARS-CoV-2 with a score more lower than SVM, and at the same time, the *vibrio cholerae* with a higher percent than SVM. Then, one should expect that the results of DL in the classification and prediction of epidemics should improve with the increase of datasets.

On the other hand, the obtained results with SVM evidenced that one should not completely discard some machine learning algorithms. Sometimes, in many practical problems when one does not have large datasets, it is best to use hybrid techniques that might mitigate this situation, and at the same time, to provide satisfactory results[4,5]. Nevertheless, despite the imbalance of datasets, the obtained results with DL can be considered promising, and further development in this direction will be welcome.

4. Conclusions

In this work, we utilized a convolutional neural network (CNN) trained on microscopic images encompassing the SARS-CoV-2 virus, the protozoan parasite "*plasmodium falciparum*", the bacterium "*vibrio cholerae*" and non-infected samples to effectively classify and predict epidemics. The findings showed promising results in both classification and prediction tasks. We quantitatively compared the obtained results by using CNN with those attained employing the support vector machine. Notably, the accuracy in prediction reached 98% when using convolutional neural network algorithms.

However, when analyzing other metrics, the SVM model performed better for the *SARS-CoV-2* virus than the DL model, while the opposite was true for *vibrio cholerae*. We concluded that due to the somewhat black-box behavior of DL algorithms (CNNs), it is not completely clear why it learned *SARS-CoV-2* with a score lower than SVM. Then, we should expect that the results of DL in the classification and prediction of epidemics should improve with the increase of datasets.

In future work, we will carry out further research and analysis based on these results, increasing the datasets and trying to achieve a better balance among all classes.

Author contributions

Conceptualization, RR and LB; methodology, RR and LB; software, LB; validation, LB and RR; formal analysis, RR; investigation, RR and LB; resources, RR; data curation, LB and RR; writing—original draft preparation, RR; writing—review and editing, RR; visualization, RR; supervision, RR; project administration, RR; funding acquisition, RR. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. Moorthy J, Gandhi UD. A Survey on Medical Image Segmentation Based on Deep Learning Techniques. Big Data and Cognitive Computing. 2022; 6(4): 117. doi: 10.3390/bdcc6040117
2. Kaur A, Singh Y, Neeru N, et al. A Survey on Deep Learning Approaches to Medical Images and a Systematic Look up into Real-Time Object Detection. Archives of Computational Methods in Engineering. 2021; 29(4): 2071-2111. doi: 10.1007/s11831-021-09649-9
3. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data. 2021; 8(1). doi: 10.1186/s40537-021-00444-8
4. Rodríguez R, Mondeja BA, Valdés O, et al. SARS-CoV-2: enhancement and segmentation of high-resolution microscopy images—Part I. Signal, Image and Video Processing. 2021; 15(8): 1713-1721. doi: 10.1007/s11760-021-01912-7
5. Rodríguez R, Mondeja BA, Valdes O, et al. SARS-CoV-2: theoretical analysis of the proposed algorithms to the enhancement and segmentation of high-resolution microscopy images—Part II. Signal, Image and Video Processing. 2022; 16(3): 595-604. doi: 10.1007/s11760-021-02045-7
6. Basu A, Senapati P, Deb M, et al. A survey on recent trends in deep learning for nucleus segmentation from histopathology images. Evolving Systems. 2023; 15(1): 203-248. doi: 10.1007/s12530-023-09491-3
7. Yang T, Luo Y, Ji W, et al. Advancing biological super-resolution microscopy through deep learning: a brief review. Biophysics Reports. 2021; 7(4): 253. doi: 10.52601/bpr.2021.210019
8. Rodríguez R, Sossa JH. Mathematical Techniques for Biomedical Image Segmentation. Encyclopedia of Biomedical Engineering. Published online 2019: 64-78. doi: 10.1016/b978-0-12-801238-3.99989-6
9. Ledón T, et al. *Vibrio cholerae* O139: Emergence, evolution, and genetic structure of CTXΦ (Spanish). Revista CENIC Ciencias Biológicas. 2007; 38(1): 062-067.
10. Haldar K, Mohandas N. Malaria, erythrocytic infection, and anemia. Hematology. 2009; 2009(1): 87-93. doi: 10.1182/asheducation-2009.1.87
11. Yang W, Zhang X, Tian Y, et al. Deep Learning for Single Image Super-Resolution: A Brief Review. arXiv. 2019; arXiv:1808.03344v3.
12. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM. 2017; 60(6): 84-90. doi: 10.1145/3065386
13. Borja-Robalino R, Monleón-Getino A, Rodellar J. Standardization of performance metrics for classifiers (Spanish). Revista Ibérica de Sistemas e Tecnologias de Informação. 2020; E30: 184-196.
14. Traore BB, Kamsu-Foguem B, Tangara F. Deep convolution neural network for image recognition. Ecological Informatics. 2018; 48: 257-268. doi: 10.1016/j.ecoinf.2018.10.002
15. Karim A, Singh J, Mishra A, et al. Toxicity prediction by multimodal deep learning. Pacific Rim Knowledge Acquisition Workshop. 2019; 2: 142-152.
16. Available online: http://www.evanlray.com/stat344ne_s2020/materials/20200226_generators_data_augmentation/20200225_stuff/20200225_stuff.pdf 2020 (accessed on 6 March 2023).
17. Ho Y, Wookey S. The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. IEEE Access. 2020; 8: 4806-4813. doi: 10.1109/access.2019.2962617
18. Javid AM, Das S, Skoglund M, et al. A ReLu dense layer to improve the performance of neural networks. arXiv. 2020; arXiv2010.13572v1.
19. Karim A, Mishra A, Newton MAH, et al. Efficient Toxicity Prediction via Simple Features Using Shallow Neural Networks and Decision Trees. ACS Omega. 2019; 4(1): 1874-1888. doi: 10.1021/acsomega.8b03173
20. Margherita G, Enrico B, Giorgio V. Metrics for Multi-class Classification: An Overview. arXiv. 2020; arXiv:2008.05756v1.
21. Opitz J, Burst S. Macro F1 and Macro F. arXiv. 2021; arXiv:1911.03347v3.
22. Murphy KP. Machine learning: A probabilistic perspective. MIT Press; 2012.