A Machine Learning Approach for Early Identification and Prevention of Covid-19 like Global Pandemics

Mr. Sandeep Kumar Mathariya*[1], Dr. Hemang Shrivastava [2]

Abstract: In addition to causing enormous global financial and social losses, the COVID-19 pandemic was a once-in-a-century occurrence that claimed a great deal of human lives. Reducing the number of victims can be achieved in part by precisely and early detection of the disease. The unanticipated surge in cases has also resulted in severe limits on the number of scans conducted and the amount of time radiologists may spend analyzing the data to assess the severity of the diseases and potential advancements in the future. Therefore, automated procedures that might lessen the burden on the healthcare system to offer prompt and accurate diagnoses are being researched. Furthermore, determining future possible hotspots for these diseases might aid in the selection of micro-containment zones, which can be a proactive measure to impede the disease's rapid spread. Numerous methods based on deep learning and machine learning have been researched for picture categorization, finding possible hotspots, or both. This work offers a data-driven approach to categories Covid-19 images and pinpoint possible pandemic hotspots, which will help with treatment and prevent pandemics from spreading in the future. Moreover, a deep neural network based regression model has been developed which predicts the number of new active cases in the days ahead. This approach is allows for early assessment of the disease spread and allows for timely management and early prevention of widespread pandemics. To assess the effectiveness of the suggested method, a comparison with the most recent iterations of deep learning and machine learning algorithms has been provided. Findings show that, when compared to baseline methods, the suggested data-driven strategy with probabilistic categorization performed better. Moreover, the proposed regression model outperforms baseline state of the art models in terms of accuracy of prediction..

Keywords: Coronavirus, COVID-19 Pandemic, Automated Detection, Probabilistic Classification, Deep Neural Networks.

1. Introduction

In contemporary times, human encounters with pandemics have been sporadic, with the most recent significant outbreak being the Spanish flu in the previous century [1]. The emergence of the Covid-19 pandemic marked a catastrophic event of unprecedented scale, resulting in significant loss of life [2]. This crisis led to a substantial financial downturn and economic slowdown, the ramifications of which continue to be felt today [3]. As the pandemic gradually waned due to widespread containment efforts, vaccine development, and the establishment of herd immunity, it became evident that a more proactive approach is imperative to effectively tackle similar future crises. There are valuable lessons to be gleaned from this current catastrophe to enhance anticipation and preparedness for future pandemics [4]. The importance of early detection and swift action is underscored as a crucial lesson. The delayed recognition of the severity of COVID-19 allowed it to spread globally before adequate countermeasures could be implemented [5]. To expedite the identification and mitigation of future threats, it is essential for pandemic preparedness efforts to prioritize the enhancement of surveillance systems, international collaboration, and information exchange [6].

The strain that the COVID-19 epidemic placed on healthcare

services exposed several weaknesses. To get ready for future pandemics, investments in healthcare infrastructure are required. These include having enough medical supplies, qualified medical workers, and adaptable systems that can manage spikes in cases [7]. Synchronization and international collaboration are essential. Today's interconnected world means that combating pandemics need collaboration from all stakeholders. Improving international collaboration in areas like immunization distribution, data interchange, and resource allocation will be crucial to managing and preventing future health catastrophes. [8].

Advancements in vaccine research, antiviral therapy, and diagnostic technologies are crucial for pandemic preparedness [9]. Sustaining dedication to research and innovation will enhance global readiness to address future pandemics [10]. A significant discovery from the current COVID-19 pandemic is the utilization of data-driven machine learning models for both diagnosis and strategic planning of containment zones, effectively curbing the widespread and rapid transmission of the disease. Figure 1 depicts the digital technology based framework for surveillance. and detection of pandemics like Covid-19.

This paper aims at developing a machine learning model which can identify potential hotspots and predict future rate of increase of Covid cases for early detection and prevention of the spread of pandemic likecases. Also a data driven Covid classification technique has been proposed [11].

COVID-19 detection utilizing Bayesian networks involves a sophisticated data-driven approach that integrates various factors to assess the likelihood of infection.

1 Department of Computer Science and Engineering

1 Research Scholar, SAGE University, Indore, (M.P) India

ORCID ID : 0000-0002-3795-5455

2 Research Supervisor, SAGE University, Indore, (M.P.), India

2drhemang.shrivastava@sageuniversity.in

* Corresponding Author Email: mathariya@gmaill.com

Fig.1 Digital Technology infrastructure for pandemic detection and surveillance**.**

. By incorporating information such as symptoms, exposure history, and test results, Bayesian networks provide a probabilistic framework for analyzing and predicting COVID-19 cases. These networks consider the conditional dependencies between different variables, allowing for more accurate and nuanced assessments of the disease's presence. Through Bayesian inference, the model can update its predictions as new data becomes available, enabling healthcare professionals to make informed decisions regarding testing, treatment, and containment measures. This approach not only enhances the efficiency of COVID-19 detection but also contributes to a more comprehensive understanding of the pandemic's dynamics and spread.

Detecting COVID-19 hotspots through machine learning involves leveraging large datasets and sophisticated algorithms to identify regions with heightened transmission rates. Machine learning models analyze various factors such as population density, mobility patterns, case counts, and demographic information to predict areas at higher risk of outbreaks. These models can uncover hidden patterns and correlations that may not be immediately apparent to human observers, enabling proactive measures to be implemented in vulnerable areas. By continuously updating and refining their predictions based on real-time data, machine learning algorithms empower authorities to allocate resources more effectively, implement targeted interventions, and mitigate the spread of the virus. This data-driven approach to hotspot detection not only aids in controlling the current pandemic but also lays the groundwork for better preparedness and response to future outbreak.

2. Background

2.1 Image Based Classification

COVID-19 detection through the analysis of CT and X-ray images using machine learning (ML) and deep learning (DL) models has emerged as a promising approach to aid in diagnosis. These imaging techniques offer valuable insights into the presence and severity of lung abnormalities associated with the virus. ML and DL models are trained on large datasets of annotated CT and X-ray images, learning to distinguish between normal and COVID-19-infected lungs based on visual patterns and features.
Convolutional neural networks (CNNs), a type of DL model,

have shown particular efficacy in automatically extracting relevant features from medical images. By analyzing subtle textures, patterns, and structural changes indicative of COVID-19 infection, these models can provide rapid and accurate assessments, assisting radiologists and clinicians in making informed decisions. Additionally, the use of ML and DL for COVID-19 detection in imaging data offers the potential for scalability and efficiency, enabling healthcare systems to handle large volumes of cases more effectively.

However, continued research and validation are necessary to ensure the reliability and generalizability of these models across diverse populations and imaging protocols. Overall, leveraging ML and DL techniques for COVID-19 detection in CT and X-ray images holds promise for improving diagnostic accuracy and aiding in the management of the pandemic.
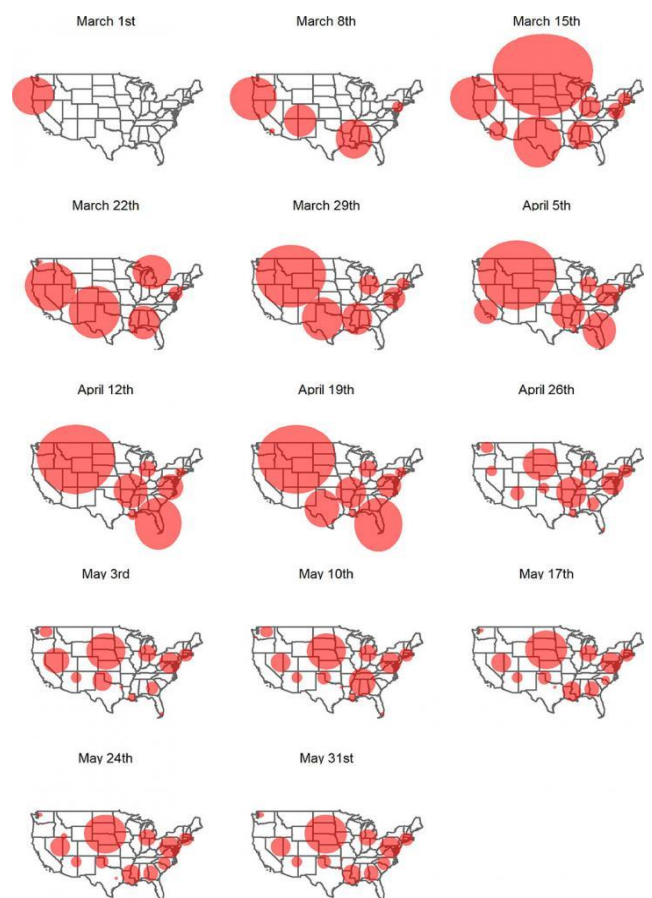

Fig.2 A Weekly Surveillance of Covid Hotspot Clusters in USA (hohl et al., spatial and spatio-temporal epidemiology (2020)

Figure 2 depicts the case distribution (weekly) for global pandemic distribution during Covid-19.

2.2 Prediction of Covid Hotspots

Machine learning models for detecting hotspots play a crucial role in identifying regions or areas with heightened activity or risk related to various phenomena, including disease outbreaks, natural disasters, and social unrest. These models leverage large datasets containing information such as historical records, demographic data, environmental factors, and real-time observations to predict the likelihood of hotspot occurrences. One commonly used approach involves clustering algorithms, which group spatial data points based on similarity, allowing for the

identification of clusters or concentrations of events.

Another approach involves spatiotemporal modeling, where machine learning algorithms analyze both spatial and temporal patterns to predict the emergence and evolution of hotspots over time. These models can incorporate diverse data sources, including satellite imagery, social media feeds, and sensor data, to provide comprehensive insights into hotspot dynamics.Additionally, ensemble methods, which combine predictions from multiple models, can enhance the robustness and accuracy of hotspot detection. By identifying hotspots early, decision-makers can allocate resources, implement targeted interventions, and mitigate potential risks, ultimately enhancing public safety and resilience. However, ongoing research and validation are essential to ensure the reliability and effectiveness of machine learning models for hotspot detection across different contexts and applications.

2.3 Prediction of future Covid Cases using Machine learning.

Predicting future COVID-19 cases using machine learning involves leveraging historical data, epidemiological factors, and other relevant features to forecast the spread of the virus over time. One common approach is time series forecasting, where machine learning algorithms analyze trends and patterns in past case counts to make predictions about future numbers of infections.

These models can incorporate various input variables such as population density, mobility data, government interventions, and healthcare capacity to improve the accuracy of predictions. Additionally, ensemble methods and deep learning techniques such as recurrent neural networks (RNNs) can capture complex temporal dependencies and nonlinear relationships in the data, enhancing the predictive capabilities of the models.

Continuous updating of the models with new data allows for adaptive forecasting, enabling authorities to adjust strategies and allocate resources in response to changing dynamics. While machine learning models for predicting COVID-19 cases have shown promise, it's essential to consider uncertainties and limitations inherent in the data and model assumptions. Moreover, ongoing validation and refinement of these models are crucial for ensuring their reliability and effectiveness in supporting decision-making and public health interventions.

Machine learning (ML) has been instrumental in developing various solutions to address the challenges posed by COVID-19.ML techniques, particularly time-series analysis and deep learning, have been used to model the spread of COVID-19, predict case trajectories, and forecast healthcare resource needs. These models take into account factors such as population demographics, mobility patterns, interventions, and environmental factors to provide insights for policymakers and healthcare providers. ML models can predict individual risk factors for COVID-19 complications based on patient demographics, comorbidities, and other clinical variables. These models enable personalized treatment plans and interventions for high-risk individuals, leading to better outcomes.

Overall, machine learning plays a critical role in advancing our understanding of COVID-19, developing effective interventions, and guiding public health responses to mitigate the impact of the pandemic, as well as early detection and prevention of such pandemics in future.

3. Methodology:

The methodology has a addresses 3 major challenges to Covid-19 like pandemics, which are:

1. Data driven models to classify the diseases.
2. Identifying potential hotspots in future.
3. Predicting new active cases in future.

These three approaches are expected to result in early detection and prevention of Covid like pandemics.

The proposed approach entails image-processing followed by feature extraction and classification to classify positive and negative cases [24]. While several deep learning approaches such as variants of CNNs are available at our disposal, yet the copious amounts of data and processing power needed to train the algorithms is extremely large to attain significantly high classification accuracy [25]. Alternatively an image processing based approach coupled with feature extraction is presented in this paper so as to overcome the challenge of extremely large datasets [26].

3.1 Data Pre-Processing

The first step though is the removal of noise from raw captured images, whose sources can be:

1) Addition of electronic noise in the image due to the use of amplifiers in the sensing device which is also termed as white or Gaussian noise [27].
2) The abrupt change or spikes in the analog to digital converters sued in the circuity of the fundus image causing salt and pepper noise patterns [28].
3) The multiplicative noise effect due to the inconsistent gain of the adaptive gain control (AGC) circuity used for capturing or retrieving the fundus image29 [].
4) The lack of pixels while capturing the image resulting in frequency mean valued interpolations in the reconstructed image causing Poisson image [30]-[31].

The removal of noise effects is fundamentally important as noisy images would result in erroneous feature extraction leading to inaccurate classification of the CT/MRI images [32]. Contrary to conventional Fourier based methods, the wavelet transform is made us of non-smooth Kernel functions such as Mayer, Haar, Coif etc [33]-[34]. The essence of the transform lies in the fact that the wavelet transform separates the low frequency and high frequency components as the approximate co-efficient (C_A) and detailed co-efficient (C_D) of the transform. Generally, detailed co-efficient (C_D) and a low frequency resolution component termed as the detailed co-efficient (C_A) [35]..

Retaining the low frequency component (C_A) while discarding the higher frequency component (C_D) for a number of iterations helps in removal of the baseline noise of the system [36]. One of the most effective hyperspectral image restoration techniques is based on the sub-band decomposition of images into low pass and high pass signal values using the wavelet transform [37]. The wavelet transform, unlike the conventional Fourier methods uses non-linear and abruptly changing kernel functions which show efficacy in analysing abruptly fluctuating signals such as images. The continuous and the discrete wavelet transforms are computed as:

$$CWT(x,s,\delta) = s^{\frac{1}{2}} \int_{-\infty}^{\infty} x(t) \emptyset^* \left(\frac{t-s}{\delta}\right) dt \quad \text{(1)}$$

Where,

$s, \delta \in R$ represent the scaling (dilation) and shifting (translation) constants constrained to the condition $\delta \neq 0$.

$\emptyset^*$ is the Wavelet Family or Mother Wavelet

t is the time variable

$x(t)$ is the time domain data.

For implementing the wavelet transform on the image dataset, the sampled version of the continuous wavelet transform yields the discrete wavelet transform given by:

$$DWT(x,m,n) = \delta_0^{m\frac{-1}{2}} \sum_i x(i) \emptyset^* \left[\frac{n-is_0^m}{s_0^m}\right] \quad \text{(2)}$$

Where,

$x(i)$ is the discrete $k \times 1$ vector.

s_0^m is the discrete scaling constant.

is_0^m is the discrete shifting constant.

The discrete wavelet transform yields two distinct low and high pass values based on the number of levels of decomposition and wavelet family given by the approximate co-efficient (CA) and detailed co-efficient (CD) [38]. The approximate co-efficient values are typically the low pass values containing the maximum information content of the image while the detailed co-efficient values account for the noisy spectral part. Retaining the low pass co-efficients and recursively discarding the high pass co-efficients allows to de-noise the image [39]. The choice of the wavelet family impacts the estimation of the noise gradient vector given by [40]:

$$G_N = k \frac{\nabla I}{\nabla I_F} \quad \text{(3)}$$

The value of the second order normalizing gradient as a function of spatial co-ordinates is given by:

$$q(x,y) = \sqrt{\frac{c_1(\nabla I/I_F)^2 + c_2(\nabla^2 I/I_F)^2}{(1+c_3(\nabla^2 I/I_F)^2}} \quad \text{(4)}$$

Here,

I denotes the original image.

I_F denotes the fused image after normalization.

G_N denotes the normalizing gradient.

∇ represents the gradient.

∇^2 represents the Laplacian.

3.2 Feature Extraction
The next step is the statistical feature extraction of image features expressed as [41]:

a) Mean or average value:

$$Mean \text{ or } \mu = \frac{1}{N}\sum_i^N f_i X_i \quad \text{(5)}$$

b) Standard Deviation:

$$sd = \sqrt{\frac{1}{N}\sum_i^N (X_i - \mu)^2} \quad \text{(6)}$$

c) Energy which is also considered as the secondary moment:

$$Energy = \sum_{i,j}^n |A_{i,j}|^2 \quad \text{(7)}$$

d) Variance is the squared value of s.d. given by:

$$variance = sd^2 \quad \text{(8)}$$

e) Contrast which is the deviation among the mean and differential change in illuminance:

$$Contrast = \sqrt{\frac{1}{mn}\sum_{i,j}^{m,n}[X(i,j) - \mu(i,j)]^2} \quad \text{(9)}$$

a) Entropy which is the statistical average information content defined as:

$$E = -P(I_{x,y})log_2 I_{x,y} \quad \text{(10)}$$

b) Homogeneity which is the similarity among the pixel value distribution:

$$H = \sum_{i,j}^{m,n} \frac{P_{IJ}}{1-[i-J]^2} \quad \text{(11)}$$

c) Correlation which is the similarity overlap among pixel values:

$$Correlation_{i,j} = \sum_{i,j}^{m,n} \frac{(i-u_x)(J-\mu_j)P_{j_{x,y}}}{sd_x sd_y} \quad \text{(12)}$$

d) Root Mean Square Value which is defined as the squared root of the squared mean of values in the random distribution defined as:

$$rms = \sqrt{\frac{\sum_{i=1}^n X_i}{n}} \quad \text{(13)}$$

The normalizing factor for the gray covariance matrix (GLCM) is defined as:

$$N = \frac{X_{i,j}}{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1} * X_{i,j}} \quad \text{(14)}$$

Here,

X_i denotes random variable X

f denotes the frequency of occurrence

$I_{x,y}$ denotes the image

m, n denotes pixels

$mean$ denotes avg. illuminance
A denotes the amplitude

N denotes levels of normalized GLCM matrix

$p_{i,j}$ denotes the normalized GLCM matrix

P denotes probability

The design of the automated classifier is critically important as the accuracy of classification critically depends on the design of the classifier. Generally, positive and negative CT/MRI images show overlapping feature values. Hence a probabilistic approach is often effective. An image processing is practically carried out in hospitals to medical facilities, hence deep learning algorithms which need large computational resources may render infeasibility to even a novel and accurate approach [42]. This leads to a natural inclination towards the Bayesian Regularization algorithm [43].

the labelled data vector $Tr = [f_1 \ldots \ldots . f_1$ is fed to

the bayesian regularized ann. the brann is chosen as it is an effective classifier. it works on the principle of Baye's theorem of conditional probability. After the BRANN is trained, in the testing phase, the BRANN calculates the probability of an element to belong to a particular category [44]. For a multi-class decision, the higher probability of a particular class decides the category of the data. In case of the BRANN tries to find out the probability of an image to be actually positive based on the probability before passing the judgement. For this, the important assumption which the BRANN makes is that of the accuracy of the classifier [45]. This is dependent on the training accuracy which is available to the classier (on completion of training) and the number of positive images in the dataset (already available to the classifier as the dataset provided by the user, which the classifier assumes to be true). The same logic applies to the negative images [46].

The training rule for the approach is based on the Bayes theorem of conditional probability which is effective for classifying overlapping feature vectors, based on a penalty $\rho = \frac{\mu}{v}$. The weights are updated based on the modified regularized cost function [47]:

$$F(w) = \mu w^T w + v[\frac{1}{n}\sum_{i=1}^{n}(p)] \qquad (15)$$

If   Network error are generally low.

else if   Network errors tend to increase, in which case

the weight magnitude should be reduced so as to limit errors (Penalty).

This is done be maximizing the weight Posteriori Probability using the Bayes theorem of Conditional Probability as:

$$\qquad (16)$$

The proposed training algorithm is presented next.

3.3 Proposed Algorithm:

Step.1: Initialize weights w and learning rate μ randomly, set maximum iterations as $Maxitr = 1000$, $e_{tolerance} = 10^{-6}$

Step.2: $for\ i = 1:Maxitr$, do

Step.3: $for\ (k = 1:n)$,

Step.4: Retain (C_A) while discarding (C_D)

Step.5: $Minimize\ (J = [\![x - WW^T x]\!])$ and compute $M = B_{k\times l}D$

Step.6: Compute $X_{LV} = mean(g(\sum X_{t-kh}), \epsilon_t)$

Step.7: $if\ i \le Maxitr\ \&\&\ if\ J \le e_{tolerance}$

Step.8: Minimize: $Argmin\left(\frac{\epsilon_{inst} - \epsilon_{opt}}{\epsilon_{opt}}\right) \forall \alpha, \mu$

Step.9: Compute $w_{x+1} = w_k - \mu\beta(k)$

Step.10: Compute $\beta(k) = \alpha\beta(k-1) + \nabla C(w_k)$

Step.11: $else$

Step.12: Truncate training

Step.13: $end\ if$

Step.14: Compute MSE, MAPE, R^2

Step.15: $end\ for$

Step.16: $end\ for$

The performance metrics to be computed are [47]:

Accuracy (Ac): It is mathematically defined as:

$$Ac = \frac{TP+TN}{TP+TN+FP+FN} \qquad (17)$$

Recall: It is mathematically defined as:

$$Recall = \frac{TP}{TP+FN} \qquad (18)$$

Precision: It is mathematically defined as:

$$Precisiosn = \frac{TP}{TP+FP} \qquad (19)$$

F-Measure: It is mathematically defined as:

$$F - Measure = \frac{2.Precision.Recall}{Precision+Recall} \qquad (20)$$

Here,
TP, TN, FP and FN denote the true positive, true negative, false positive and false negative rates respectively.

Bayesian networks have been explored for COVID-19 detection as they offer a probabilistic framework to model the relationships between various factors associated with the disease. In this context, Bayesian networks can integrate information from different sources such as symptoms, exposure history, demographic data, and test results to assess the likelihood of COVID-19 infection in individuals [48].The structure of a Bayesian network represents the dependencies between variables, and the network is augmented with probabilities that quantify the likelihood of certain events given the values of other variables. For COVID-19 detection, the network might include nodes representing symptoms (e.g., fever, cough, shortness of breath), clinical findings (e.g., lung abnormalities on imaging), test results (e.g., PCR or antigen test outcomes), and risk factors (e.g., age, comorbidities, recent travel or exposure).By updating the probabilities in the network based on observed data (e.g., new symptoms reported by the patient, test results), Bayesian inference can be used to calculate the posterior probability of COVID-19 infection. This approach allows for a more nuanced assessment that takes into account the interplay between different factors and their uncertainties.

One advantage of Bayesian networks is their ability to handle incomplete or noisy data, making them suitable for real-world medical applications where data may be sparse or uncertain. Additionally, the transparency of Bayesian networks enables clinicians to interpret the reasoning behind the model's predictions, which can enhance trust and facilitate clinical decision-making [49].However, developing a robust Bayesian network for COVID-19 detection requires careful consideration of the variables to include, the structure of the network, and the estimation of accurate probabilities from data. Moreover, ongoing validation and refinement are essential to ensure the reliability

and effectiveness of the model in diverse populations and clinical settings.

3.4 Regression Learning

Managing pandemic like situations need estimate of the future number of active cases so as to deviceapproaches for management and effective prevention [50]. The following parameters have been considered while predicting future cases:

1. Date
2. Confirmed Cases
3. Deaths
4. Recovered
5. Active New Cases

The target variable happens to be the active new cases. While several other parameters may affect the cause of increase or decrease in cases, such as demographics, spatial coordinates etc., the empirical attributes have been used for analysis. A regression learning model has been developed for the purpose employing deep neural networks.
1) Mean Absolute Percentage Error
2) Number of Iterations

The approximated 2nd order derivative based learning algorithmcomputing the Hessian Matrix is employed to update the weights of the network I n each iterations, given by [51]:

$$w_{k+1} = w_k - \alpha \left[\frac{\partial^2 e}{\partial w^2}\right]^{-1} \frac{\partial E}{\partial w} \tag{21}$$

Here,

w_k & w_{k+1} denote the weights of the present and subsequent iterations.

α denotes the learning rate.

e denotes the error in the present iterations.

4. Experimental Results

The experiments are carried out of MATLAB 2022a with the Deep Learning Toolbox. The experiment is performed on a Windows Machine with i5-9300H Processor enabled with NVIDIA GTX GPU and RAM of 8GB. The software package used in Matrix Laboratory (Matlab), 2022a.

The classification of images into Covid positive and negative cases has been done using an annotated dataset with 1000 images comprising of both positive and negative cases of Covid have been used. The images are .jpg images in this experimental setup, but the system is compatible with all other common image data types such as .png, .tiff etc. The images acquired are .jpg images which three colour channels viz. R, G and B. All the images are first converted to common dimensions of (256 x 256). The features are then used to train a Deep Bayes Net. The performance metrics chosen are the accuracy of classification and prediction error in terms of the TP, TN, FP and FN values respectively. The images are presented in the sequence of occurrence in the experimental setup followed by a detailed explanation and significance of each image obtained at each step.

The hotspot identification results have been evaluated in terms of the classification accuracy computed through the confusion matrix. Geographical location and number of cases have been used a independent variables to identify hotspots.

Finally, a six month span ranging from January 2020 to July

2023 has been used to predict future active cases, through regression analysis. The MAPE and R^2 values have been used for performance evaluation.
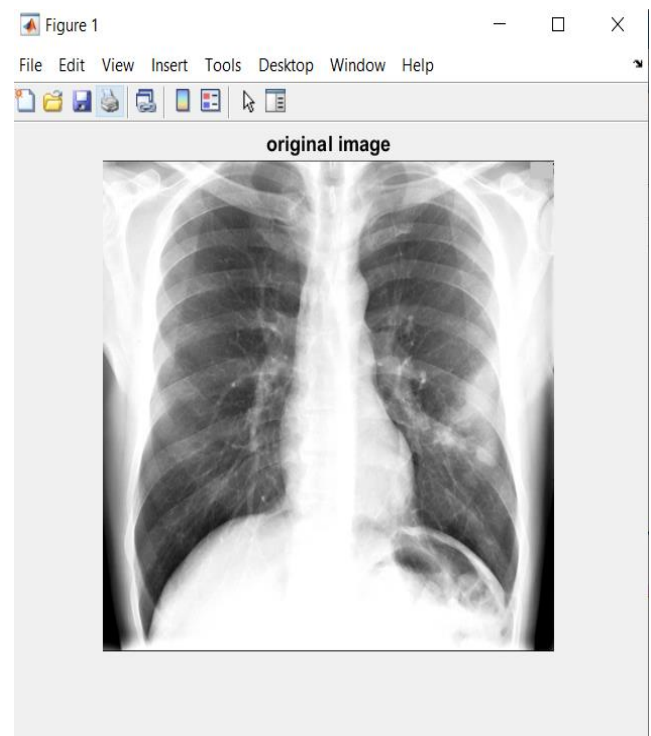

Fig.3 Original Image

Fig.4 Segmentation

Figure 3 depicts the original image while figure 4 depicts the segmented image, where the affected region in localized.
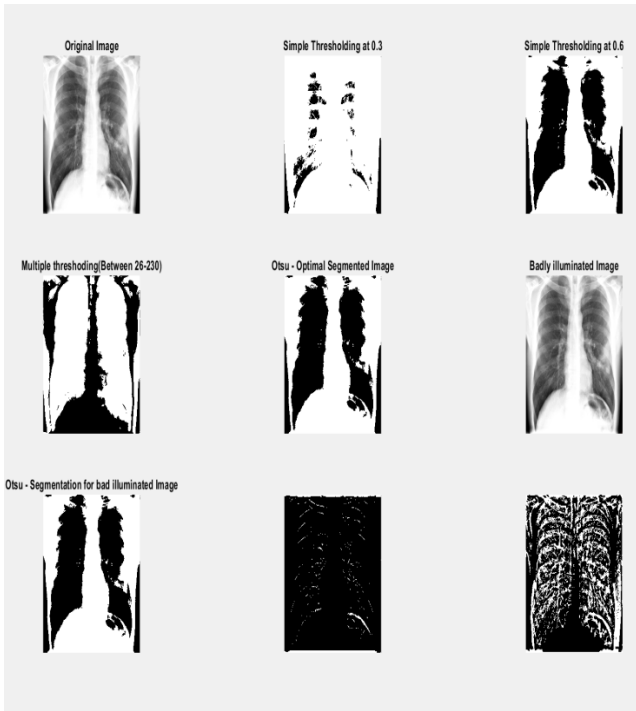
Fig.5 Variations in Segmentation based of threshold

Figure 5 depicts the various segmentation results on the image under interest.

The statistical features of the wavelet decomposition tabulated in table 2 followed by the image features computed subsequent to the DWT decomposition, in table 1.

Table.1 Statistical Analysis

S.No.	Parameter	Values	Class
1.	Minimum	0	
2.	Maximum	0.9295	
3.	Mean	0.3415	Original Image
4.	Median	0.3703	
5.	Standard Deviation	0.1524	
6	Mean Absolute Deviation	0.05355	
7.	Minimum	0.002951	
8.	Maximum	0.9165	Approximate Co-efficient values
9.	Mean	0.3415	
10.	Median	0.3706	
11.	Standard Deviation	0.1511	
12.	Mean Absolute Deviation	0.05354	
13.	Minimum	-0.1592	
14.	Maximum	0.1592	Detailed Co-efficient values
15.	Mean	0	
16.	Median	0	
17.	Standard Deviation	0.01232	
18.	Mean Absolute Deviation	0.005539	

After the DWT decomposition, the feature values of the images are to be computed as defined in the feature extraction section. The feature values computed from the images need to be fed to the proposed machine learning model for pattern recognition for both positive and negative cases. As an illustration, two separate CT images have been analysed using the proposed algorithm and their features have been tabulated in table 2. It can be observed that the statistical feature values have identical values for both positive and negative cases of covid, which necessitates the use of an accurate classifier.

Table.2 Image Features

Features	Normal Fundus Image	Glaucoma Fundus Image
Contrast	0.362500000000000	0.328409090909091
Correlation	0.151836572326215	0.179081293145884
Energy	0.700218879132232	0.723267045454546
Homogeneity	0.915729166666667	0.921979166666667
Mean	0.00652645881744487	0.00227212218788613
Standard Deviation	0.106429183594269	0.106604988215597
Entropy	3.53356796596205	3.43675888527571
RMS	0.106600358177805	0.106600358177805
Variance	0.0111929089863435	0.0111892474804911
Smoothness	0.923435573787053	0.807650945366392
Kurtosis	6.63089615157184	6.70718347384146
Skewness	0.512491012022217	0.483428823467213

Now, one of the most effective ways to ensure the accuracy and coherence of the feature extraction process is the distribution analysis of feature values for a multitude of images. Although there can be variations in images of a particular class of image in any dataset, yet the feature values should depict a certain amount of coherence over a multitude of images. Thus the feature sets for multiple images in the dataset (taken 10 for the sake of brevity and ease of analysis) has been depicted in figure 6. Figure 12 clearly indicates a similarity in the values of the features extracted from the images. For instance, feature 1 (for all the 10 images) clearly show a much lesser magnitude of value compared to feature 11 as depicted in figure 6.
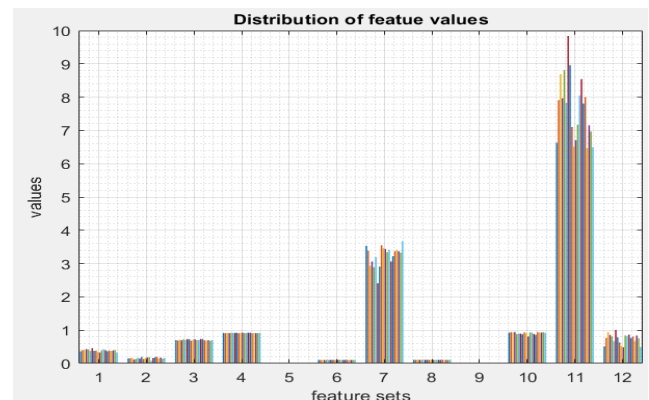

Fig.6 Distribution of feature values

A similar analysis can be done using the mesh plot depicted in figure 7. The mesh renders a three dimensional view of the feature value distribution. A coherent inference can be drawn from figures 6 and 7 depicting the fact that the feature distribution for similar class of features over a multitude of images show similarity in values. For instance, feature 1 (for multiple images) shows much lesser value compared to feature 11 (for multiple images). This is the exact same illustration taken in figure 6.
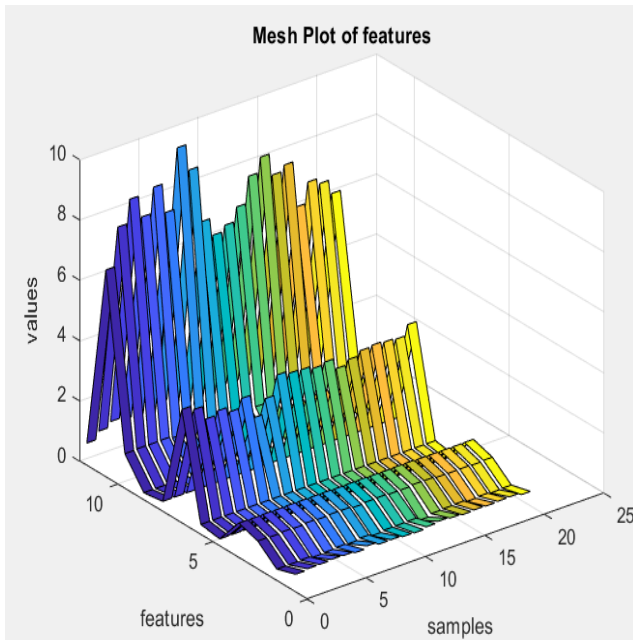
Fig.7 Mesh Plot of features

As machine learning algorithms often suffer from imbalanced instances, this phenomenon has been carefully considered while data set preparation. Imbalanced instances to imbalanced class distributions occur when the samples or observations of one of the classes is either much higher or much lower compared to the other class or classes. This may result in misleading results as machine learning algorithms tend to statistically ignore the class distributions. This caveat is eliminated in this experiment by following almost an equal share of positive and negative image classes. This can also been seen in the confusion matrix for the testing case. As the positive and negative cases have an identical class distribution, hence resampling (over sampling, under sampling or SMOTE (Synthetic Minority Oversampling Technique) has not been performed.

Three sub cases of the discriminant analysis are again considered in this experiment which happen to be:

1) Linear Discriminants.
2) Quadratic Discriminants.
3) Optimizable Discriminants.

It is observed that for a parallel pools for the training, the 3 case feature selection attains convergence in 30 iterations. The linear discriminant results in an accuracy of 98%, the quadratic discriminant results in an accuracy of 97.33% and the optimizable discriminant results in an accuracy of 98%.
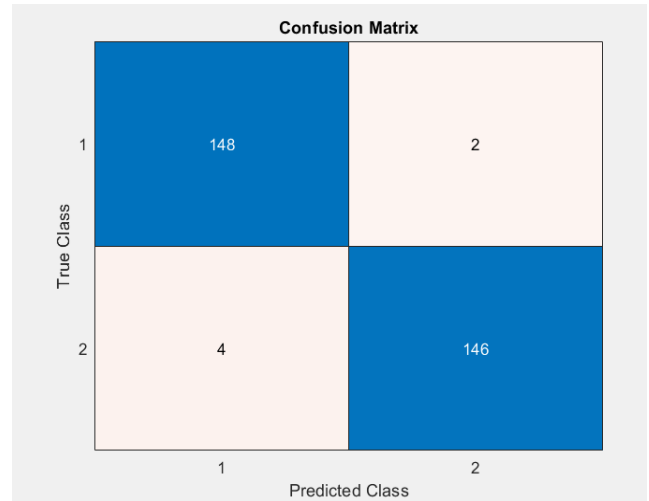

Fig.8 Confusion Matrix for Image Classification

The performance of the proposed approach is can be evaluated in terms of the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) rates of the confusion matrix. Out of the 1000 images of the dataset, 70% i.e. 700 images have been used for training and the rest of the 30% i.e. 300 images have been used for testing. The TP, TN, FP and FN values are depicted in the Confusion Matrix in figure 9. Based on the TP, TN, FP and FN values, the accuracy, sensitivity/recall, specificity, precision and F-Measure values have been computed and tabulated in table 5. The proposed approach attains an accuracy of 0.98, sensitivity or recall value of 0.9866, specificity of 0.9733, precision of 0.9736 and F-Measure of 0.9800. The mean training time for the proposed approach has been 12 minutes for the training dataset alone.

Table 3 Performance Metrics

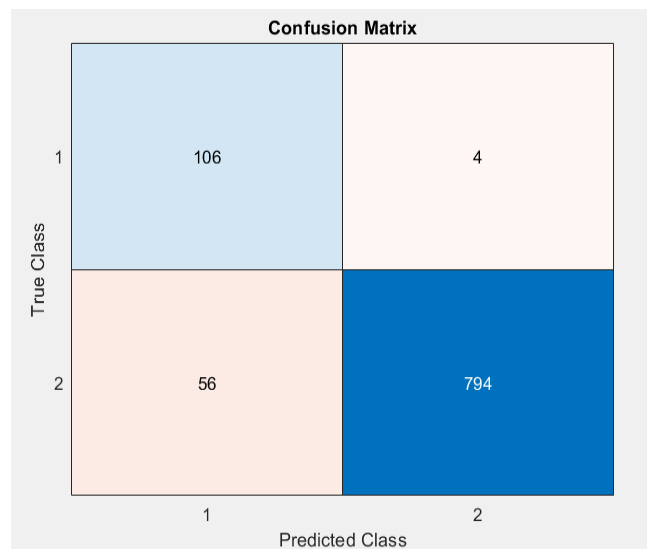| Accuracy%	Sensitivity Or Recall%	Specificity%	Precision%	F-Measure
0.980	0.9866	.9733	.9736	.9800

Fig.9. Confusion matrix for hotspot identification

While there can be many more parameters which may affect the area to be a hotspot, the above parameters are the most significant. The data parameters employed in this study are:

1. Age Bracket
2. Gender
3. Detected Area
4. Detected State
5. Current Status
6. Travel History
7. Contacted From
8. Nationality
9. Type of Transmission
10. Identity
11. No. of Cases
12. Containment Zone (target)

The approach attains a classification accuracy of 93.75% for the hotspot identification.To evaluate the performance of the proposed work in comparison to the contemporary approaches in the domain has been presented in table 4.

Table 4 Comparative Analysis w.r.t. existing work

Method	Accuracy
Mollaloa et al. (2021)	91%
Kahn et al. (2020)	79%
Alsan et al. (2022)	96.29
Akbarimajd et al. (2022)	72%
Momeny et al. (2021)	80.8%
Das et al. (2020)	97.4%
Zebari et al. (2022)	89.87%
Gannour et al.	97%
Proposed	**98%**

Table 5 presents a comparative study with respect to contemporary existing work in the domain. It can be observed from table 5 that the proposed work attains relative higher accuracy compared to existing methods in the domain. The improvement in the results can be attributed to the following reasons:

1) Image enhancement employing illumination correction and histogram normalization compensating inconsistencies in image capturing.

2) Iterative noise removal for denoising fundus images for accurate feature extraction.

3) Computing stochastic feature and subsequent feature optimization to enhance the training efficacy.

4) While deep learning models such as the CNN, RCNN, ResNet etc. may have the advantage of avoiding additional effort in handpicking features and feature combinations, they lose control over choosing the features to be used to train a model. The proposed approach with stochastic features to train a Bayesian Deep Neural Network attains relatively higher accuracy of classification compared to benchmark techniques. Moreover, the approach is also capable of identifying potential hotspots based on geographical and statistical survey features with an accuracy of 93.75%

The prediction of the Covid-19 cases has been doe based on a deep neural network architecture trained with five empirical parameters which allow us to predict the number of new active cases in future for Covid. The data statistics are analyzed followed by pattern recognition and regression analysis. The time range of January to July 2020 has been chosen for the purpose which exhibited the most amount of volatility in the number of new cases. The results are presented next.
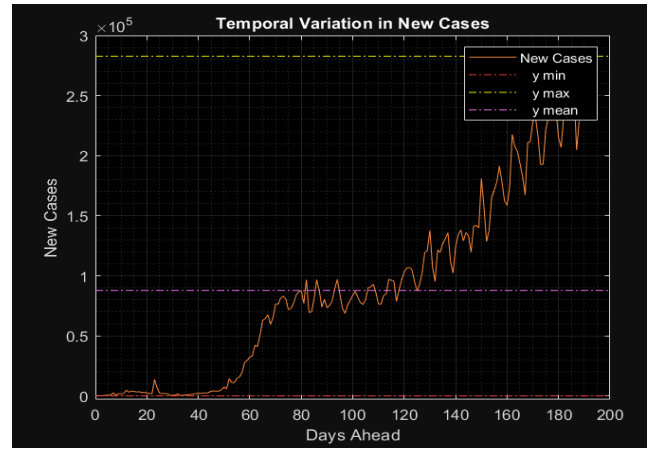

Fig.10 Raw Data

Figure 10 depicts the raw time series data for the active number of future cases over a 188 day period from January 2020 to July 2020. The minimum, maximum and mean have been marked.
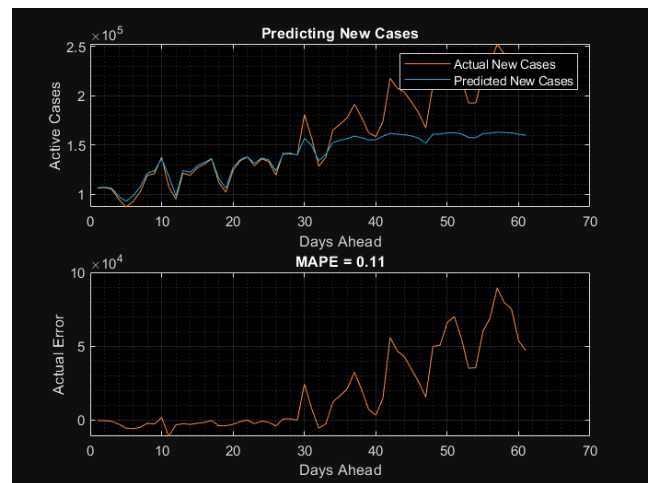

Fig.11 Prediction (Active Cases)

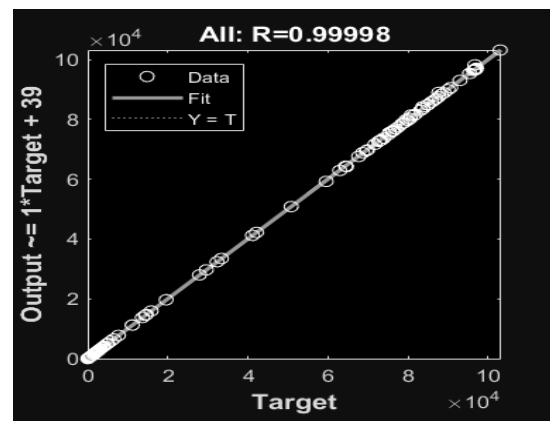Figure 11 depicts the forecasting MAPE for the proposed approach which happens to be only 0.11%/

Fig.12 Regression (Active Cases)

Figure 12 depicts the regression analysis (overall) for the data analyzed. The regression can be seen to be 0.9998.

A comparative analysis with existing baseline approaches has been presented in table 5.

Table.5 Comparative Predictive Analysis.

S.No.	Parameter	Value
1.	Sample Span	6 months
2.	ML Model	Back Propagation
3.	Training	Optimized 2nd Order Hessian
4.	Iterations to convergence	
5.	R^2	0.9998
6.	MAPE (proposed)	0.11
7.	MAPE, Balli. [52]	2.0762 (RF) 0.1853 (LR) 0.8179 (MLP) 0.1247 (SVM)

A comparative analysis with previous work in terms of baseline techniques such as Random Forecast (RF), Logistic Regression (LR), Multi Layer Perceptron (MLP) and Support Vector Machine (SVM) clearly indicates that the proposed deep neural network approach with 2nd Order Hessian Optimized back propagation beats the models in terms of MAPE of prediction. This makes the model more dependable and accurate in terms of estimating the severity of the disease in future.

5. Conclusion

This paper aims at addressing the pivotal challenges of Covid-19 like unprecedented pandemics, wherein it is mandatory to arrest the spread of the disease through early detection and prevention. It was observed from the Covid-19 scenario that the most effective way to reduce the severity and human casualties was to prevent the spread and increase in new cases in the first place. The methodology designed in this paper tries to address the three fundamental challenges pertaining to pandemics which are automated detection, identification of hotspots and predicting future trends in active new cases. The probabilistic Deep BayesNet has been proposed in this paper for both automated classification of cases as well as for identification of hotspots. Further a regression learning based approach employing the Optimized back propagation deep neural network model has been employed to predict future active cases. The results indicate that the proposed approach not only beats existing baseline techniques in terms of classification accuracy, but also attains lower prediction MAPE compared to state of art regression models. Thus, the proposed methodology developed addresses the fundamental challenges associated with pandemics and aids early detection and prevention in such cases.

Author contributions

Mr. Sandeep Kumar Maharaja: Data collection, Implementation of proposed model, compilation of results and drafting manuscript.

Dr. Hemang Shrivastava: Conceptualization of Research Problem and Methodology, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] J. Miah, R. H. Khan, S. Ahmed and M. I. Mahmud, "A comparative study of Detecting Covid 19 by Using Chest X-ray Images– A Deep Learning Approach," 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023, pp. 0311-0316

[2] A. Hussain, S. U. Amin, H. Lee, A. Khan, N. F. Khan and S. Seo, "An Automated Chest X-Ray Image Analysis for Covid-19 and Pneumonia Diagnosis Using Deep Ensemble Strategy," in IEEE Access, 2023, vol. 11, pp. 97207-97220,.

[3] S. Mishra and A. Mantri, "Design of COVID19 Detection based on Relative Eccentric Feature Selection using Deep Vectorized Regressive Neural Network for Corona Virus," 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), Raichur, India, 2023, pp. 1-6.

[4] J Somasekar, PP Kumar, A Sharma, "Machine learning and image analysis applications in the fight against COVID-19 pandemic: Datasets, research directions, challenges and opportunities", Proceedings in Matrials Today, Elsevier 2021.

[5] J. De Moura et al., "Deep Convolutional Approaches for the Analysis of COVID-19 Using Chest X-Ray Images From Portable Devices," in IEEE Access, 2020, vol. 8, pp. 195594-195607

[6] S Kumar, RD Raut, BE Narkhede," A proposed collaborative framework by using artificial intelligence-internet of things (AI-IoT) in COVID-19 pandemic situation for healthcare workers", International Journal of Healthcare Management, Taylor and Francis, 2020, vol.13, no.4, pp. 337-345.

[7] JB Awotunde, SO Folorunso, RG Jimoh, "Application of artificial intelligence for COVID-19 epidemic: an exploratory study, opportunities, challenges, and future prospects", Artificial Intelligence for COVID-19. Studies in Systems, Decision and Control, Springer, 2021, vol 358.

[8] S. Hu et al., "Weakly Supervised Deep Learning for COVID-19 Infection Detection and Classification From CT Images," in IEEE Access, 2020, vol. 8, pp. 118869-118883.

[9] J. C. Clement, V. Ponnusamy, K. C. Sriharipriya and R. Nandakumar, "A Survey on Mathematical, Machine Learning and Deep Learning Models for COVID-19 Transmission and Diagnosis," in IEEE Reviews in Biomedical Engineering, 2022, vol. 15, pp. 325-340.

[10] H. S. Alghamdi, G. Amoudi, S. Elhag, K. Saeedi and J. Nasser, "Deep Learning Approaches for Detecting COVID-19 From Chest X-Ray Images: A Survey," in IEEE Access, 2021, vol. 9, pp. 20235-20254.

[11] T Ozturk, M Talo, EA Yildirim, UB Baloglu, "Automated detection of COVID-19 cases using deep neural networks with X-ray images", Computers and Biology and Medicine, Elsevier 2020, col.121, 103792.

[12] J. C. Clement, V. Ponnusamy, K. C. Sriharipriya and R. Nandakumar, "A Survey on Mathematical, Machine Learning and Deep Learning Models for COVID-19 Transmission and Diagnosis," in IEEE Reviews in Biomedical Engineering, 2022, vol. 15, pp. 325-340

[13] M. M. Islam, F. Karray, R. Alhajj and J. Zeng, "A Review on Deep Learning Techniques for the Diagnosis of Novel Coronavirus (COVID-19)," in IEEE Access, vol. 9, pp. 30551-30572.

[14] Y Zhang, Q Zhang, Y Zhao, Y Deng, H Zheng, "Urban spatial risk prediction and optimization analysis of POI based on deep learning from the perspective of an epidemic", International Journal of Applied Earth Observation and Geoinformation, , Elsevier, 2022, vol. 112, 102942.

[15] L Xu, R Magar, AB Farimani, "Forecasting COVID-19 new cases using deep learning methods", Computers in biology and medicine,

Elsevier 2022, vol.144, 105342

[16] AS Kwekha-Rashid, HN Abduljabbar, B Alhayani, "Coronavirus disease (COVID-19) cases analysis using machine-learning applications", Applied Nanoscience, Springer, 2023, vol.31, pp. 2013–2025.

[17] A Mollalo, B Vahedi, S Bhattarai, LC Hopkins, "Predicting the hotspots of age-adjusted mortality rates of lower respiratory infection across the continental United States: Integration of GIS, spatial statistics and machine learning algorithms", International Journal of Medical Informatics, Elsevier, 2021, vol. 142, 104248.

[18] SD Khan, L Alarabi, S Basalamah -, "Toward Smart Lockdown: A Novel Approach for COVID-19 Hotspots Prediction Using a Deep Hybrid Neural Network", Journal of Computers, MDPI, 2020, vol.9, no.9, pp1-16.

[19] A Akbarimajd, N Hoertel, MA Hussain, "Learning-to-augment incorporated noise-robust deep CNN for detection of COVID-19 in noisy X-ray images", Journal of Computational Science, Elsevier 2022, vol.63, 101763.

[20] M Momeny, AA Neshat, MA Hussain, S Kia, "Learning-to-augment strategy using noisy and denoised data: Improving generalizability of deep CNN for the detection of COVID-19 in X-ray images", Computers in Biology and Medicine, Elsevier, 2022, vol.136, 104704.

[21] MF Aslan, MF Unlersen, K Sabanci, A Durdu, "CNN-based transfer learning–BiLSTM

[22] NN Das, N Kumar, M Kaur, V Kumar, D Singh, "Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays", IBRM, Elsevier, 2022, vol.43, no.2, pp.114-119.

[23] D. A. Zebari, A. M. Abdulazeez, D. Q. Zeebaree and M. S. Salih, "A Fusion Scheme of Texture Features for COVID-19 Detection of CT Scan Images," 2020 International Conference on Advanced Science and Engineering (ICOASE), 2020, pp. 1-6.

[24] O. El Gannour, S. Hamida, B. Cherradi, A. Raihani And H. Moujahid, "Performance Evaluation of Transfer Learning Technique for Automatic Detection of Patients with COVID-19 on X-Ray Images," 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), 2020, pp. 1-6.

[25] SH Kassania, PH Kassanib, MJ Wesolowskic, "Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: a machine learning based approach", Biometrics and Biomedical Engineering, Elsevier, 2021, vol.41, no.3, pp. 867-879.

[26] V Bolon-Canedo, B Remeseiro, "Feature selection in image analysis: a survey", Artificial Intelligence Review, Springer 2020, vol. 53, pp. 2905–2931.

[27] R. Miikkulainen, J Liang, E Meyerson, A Rawal, "Evolving deep neural networks", Artifical Intelligence in the Age of Neural Networks and Brain Computing, Academic Press, 2019, pp.293-312.

[28] A Zargari Khuzani, M Heidari, SA Shariati, "COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images", Scientific Repots, Nature, 2021, vol.11, no. 9887.

[29] M Shafiq-ul-Hassan, GG Zhang, K Latifi, "Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels", Scientific Reports, Nature, 2018, vol.8, no. 10545.

[30] M Toğaçar, B Ergen, Z Cömert, "COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches", Computers in Biology and Medicine, Elsevier 2020, vol.121, 103805.

[31] SH Kassania, PH Kassanib, MJ Wesolowskic, "Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: a machine learning based approach", Biometrics and Biomedical

Engineering, Elsevier 2021, vol.41, no.3, pp. 867-879.

[32] E. -S. M. El-Kenawy et al., "Advanced Meta-Heuristics, Convolutional Neural Networks, and Feature Selectors for Efficient COVID-19 X-Ray Chest Image Classification," in IEEE Access, 2021, vol. 9, pp. 36019-36037.

[33] K. K. Singh and A. Singh, "Diagnosis of COVID-19 from chest X-ray images using wavelets-based depthwise convolution network," in Big Data Mining and Analytics, 2021, vol. 4, no. 2, pp. 84-93.

[34] B. King, S. Barve, A. Ford and R. Jha, "Unsupervised Clustering of COVID-19 Chest X-Ray Images with a Self-Organizing Feature Map," 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), 2020, pp. 395-398.

[35] ALA Dalal, MAA Al-qaness, Z Cai, EA Alawamy, "IDA: Improving distribution analysis for reducing data complexity and dimensionality in hyperspectral images", Pattern Recogniton, Elsevier 2023, vol.134, no. 109096.

[36] R Mostafiz, MS Uddin, MM Reza, MM Rahman, "Covid-19 detection in chest X-ray through random forest classifier using a hybridization of deep CNN and DWT optimized features", Journal of King Saud University - Computer and Information Sciences, Elsevier 2022, vol.34, no.6, Part B, pp. 3226-3235.

[37] U Muhammad, MZ Hoque, M Oussalah, "SAM: Self-augmentation mechanism for COVID-19 detection using chest X-ray images", Lnowledge Based Systems, Elsevier 2022, vol.241, 108207.

[38] I. Ahmad, M. Basheri, M. J. Iqbal and A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," in IEEE Access 2018, vol. 6, pp. 33789-33795.

[39] P. Padilla, M. Lopez, J. M. Gorriz, J. Ramirez, D. Salas-Gonzalez and I. Alvarez, "NMF-SVM Based CAD Tool Applied to Functional Brain Images for the Diagnosis of Alzheimer's Disease," in IEEE Transactions on Medical Imaging, 2012 vol. 31, no. 2, pp. 207-216

[40] G. Rathee, S. Garg, G. Kaddoum, Y. Wu, D. N. K. Jayakody and A. Alamri, "ANN Assisted-IoT Enabled COVID-19 Patient Monitoring," in IEEE Access, vol. 9, pp. 42483-42492.

[41] C. Zhou, J. Song, S. Zhou, Z. Zhang and J. Xing, "COVID-19 Detection Based on Image Regrouping and Resnet-SVM Using Chest X-Ray Images," in IEEE Access, 2021, vol. 9, pp. 81902-8191.

[42] S. K. Zhou et al., "A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises," in Proceedings of the IEEE, 2021, vol. 109, no. 5, pp. 820-838.

[43] S. Tiwari, P. Chanak and S. K. Singh, "A Review of the Machine Learning Algorithms for Covid-19 Case Analysis," in IEEE Transactions on Artificial Intelligence.

[44] T. Sercu, C. Puhrsch, B. Kingsbury and Y. LeCun, "Very deep multilingual convolutional neural networks for LVCSR," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 4955-4959.

[45] MA Kızrak, Z Müftüoğlu, T Yıldırım, "Limitations and challenges on the diagnosis of COVID-19 using radiology images and deep learning", Data Science for COVID-19, Computational Perspectives , Science Direct 2021, pp.91-115.

[46] K. Sanjar, A. Rehman, A. Paul and K. JeongHong, "Weight Dropout for Preventing Neural Networks from Overfitting," 2020 8th International Conference on Orange Technology (ICOT), Daegu, Korea (South), 2020, pp. 1-4.

[47] L Borawar, R Kaur, "ResNet: Solving Vanishing Gradient in Deep Networks", Proceedings of International Conference on Recent Trends in Computing, Springer 2023, pp 235–247.

[48] M. A. Awal, M. Masud, M. S. Hossain, A. A. -M. Bulbul, S. M. H. Mahmud and A. K. Bairagi, "A Novel Bayesian Optimization-Based Machine Learning Framework for COVID-19 Detection From

Inpatient Facility Data," in IEEE Access, vol. 9, pp. 10263-10281, 2021.

[49] M Canayaz, S Şehribanoğlu, R Özdağ, "COVID-19 diagnosis on CT images with Bayes optimization-based deep neural networks and machine learning algorithms", Neural Computing, Springer, 2022, vol.34, pp.5349–5365.

[50] F. Rustam et al., "COVID-19 Future Forecasting Using Supervised Machine Learning Models," in IEEE Access, vol. 8, pp. 101489-101499, 2020.

[51] W Chen, P Wang, J Cheng, "Towards mixed-precision quantization of neural networks via constrained optimization", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 5350-5359.

[52] S Ballı, "Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods", Chaos, Solitons & Fractals, Elsevier, 2021, vol.142, 110512