Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Detection of Covid-19 using PDE based filter, GLSZM Features and Classifiers

S.Sanjayprabu[1, a)], R.Sathish Kumar[1], N.Bhuvaneswari[1], R.Karthikamani[2]

[1]Department of Mathematics, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, Tamil Nadu, India.

[2]Department of ECE, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India.

a) Corresponding author: sanjayprabu.phd@gmail.com

Article Info
Page Number: 1141-1152
Publication Issue:
Vol. 72 No. 1 (2023)

Article History
Article Received: 15 October 2022
Revised: 24 November 2022
Accepted: 18 December 2022

Abstract: Radiologists can diagnose Covid-19 using information from a chest X-ray scan. However, X-ray visual analysis takes time. Consequently, it's essential to create algorithms for the automatic detection of COVID-19 from CT scans. To identify Covid-19 cases, we provide a logistic regression, KNN, and boosted tree classifiers in this study. The proposed method uses the fourth-order PDE-based filter as a preprocessing before extracting GLSZM features and making the classification decision. In comparison to results from the classifiers, these results are more trustworthy and explicable. According to experimental findings, the technique can produce good performance with an accuracy of 98.42%.

Keywords: Fourth order PDE, GLSZM, Logistic regression, KNN, boosted tree.

I.Introduction

Globally, the Covid-19 virus continues to have a major effect on people's personal life quality and health. Within a few months, it had almost reached every corner of the globe [1]. According to data, more than 0.3 million individuals have died worldwide as a result of COVID-19 just 2 months after even the World Health Organization identified it as a worldwide epidemic [2]. Just around 267 million people were infected with it worldwide and resulted in more than 5.2 million fatalities by 8 December 2021. Government restrictions put in place to stop the spread of this deadly disease have a severe effect on every living thing in every country. [3]. COVID-19-related disorders can range in intensity from modest ailments to significant, even fatal, ailments. (RT-PCR) testing is recognized as effective for the diagnosis of COVID-19, despite its limitations. Inside the respiratory tract, SARS-CoV-2 genetic material is found using an RTPCR technique [4]. For RTPCR, equipment is required, and it takes at least 24 hours to reach a decision. The diagnosis may require confirmation by another RTPCR or other testing because it is not entirely accurate. Using chest X-ray imaging, doctors may monitor patients having COVID-19 symptomatology while they are waiting for RTPCR findings or in circumstances where RTPCR tests are negative. Images from chest X-rays are a crucial diagnostic tool for assessing a variety of respiratory disorders, including fever. However, they can be difficult to comprehend and take some time.

The development of a method that will help doctors identify COVID-19 is essential. Early identification of COVID-19 may be able to save a life of a patient by providing the necessary

medical assistance as soon as possible. One approach that has recently been used to overcome issues with image processing is deep learning. It has been found that it has significant impacts on a number of areas, including medicine, agriculture, remote sensing, and gesture recognition [5, 6, 7]. It is used in medicine to diagnose and classify a wide range of illnesses, including cancer [10], various ulcer types [8, 9], and skin problems [8, 9]. Such algorithms for deep learning allow for accurate disease diagnosis by medical professionals.

Figure 1 shows the workflow of the proposed method. The following is the format for this paper. In Section 2, the Covid-19 imaging methods and the collection and preprocessing of X-ray image data are described and the anisotropic diffusion filtering and GLSZM feature extraction technique in X-ray images are discussed in Section 3. The observations of three classifiers are shown in section 4. The outcomes are detailed in Section 5. In Section 6, the analysis reaches its conclusion.

Figure 1. Workflow

II.Material and Methods

The first step in any image processing system is image acquisition. Image acquisition means collecting image data from the database. The open-source Kaggle database was utilized for this study's classification of Covid 19 chest X-rays. This database, which is open source and offers an analysis of each image's flaws using efficient algorithms, has 392 chest X-rays images and the related ground truth images. The 392 images consist of 196 normal images and 196 aberrant images. The raw chest X-ray images for this paper's second phase are RGB images, and noise must be removed for the accuracy of the images to order to convert them to grayscale. These modifications are made in accordance with the norms and regulations of MATLAB. Additionally, this procedure is necessary because it enhances some of the source images that are important for subsequent processing while also working to improve the image data. The following steps are used to prepare images for grayscale conversion from RGB images and filtering. Figure 2 shows normal and Covid-19 chest X-ray images.

Figure 2. Chest X-ray images of Normal and Covid-19

Pre-processing is done to make an image better so that we can analyze it more successfully. Using preprocessing, we can get rid of undesired distortions and enhance certain elements that are crucial to the method we're working on. The Mean, Median, Laplacian, Wiener, Anisotropic diffusion filter, and Fourth order PDE filters are used in this study to pre-process images from the dataset. A Fourth order PDE filter is applied to images from the dataset based on some parameter values. Table 1 depicts the result of some parameter values of the various filters.

FILTERS	PSNR	SNR	MSE
Mean	4.42377662	2.54E+02	2.35E+04
Median	34.3947103	150	23.637924
Laplacian	4.72734988	2.19E+04	255
Wiener	32.145816	47	39.67334
Anisotropic diffusion	46.61619	1.417297	4
Fourth order PDE	40.5131315	5.777863	7

Table 1. Comparison of Filters

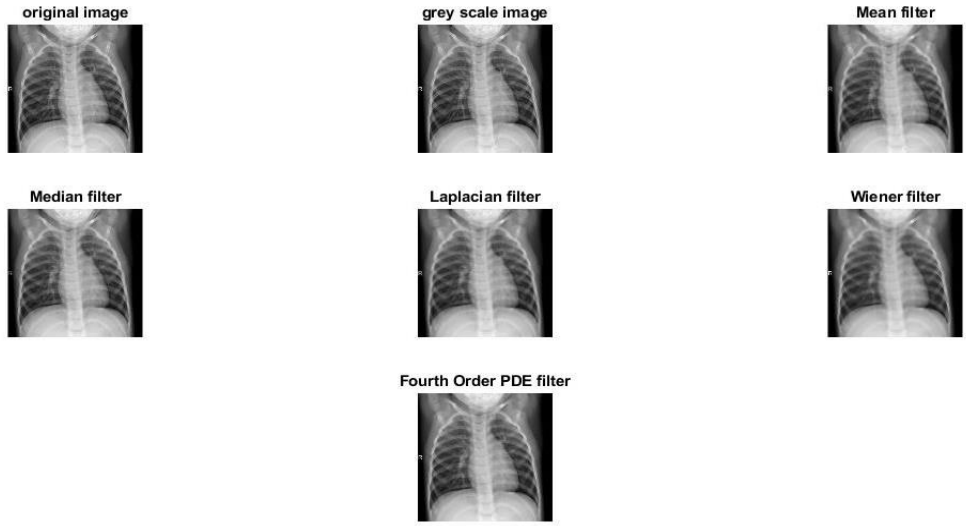

Figure 3. Filtered Chest X-ray images

III.FILTERING AND FEATURE EXTRACTION

Filtering is an important step in the image processing system. Because image filters remove the noises efficiently. In this anisotropic diffusion (sometimes referred to as Perona Malik diffusion) is used to remove unwanted distortion from an image without losing information that is essential to subsequent steps in the process, such as edges, corners, or other defining features of the input data that are required for feature extraction. The pixel intensity functions

(u), the timescale (t), and the diffusion coefficient (c) are all variables to consider. Anisotropic diffusion (1) is defined as:

$$\frac{\partial u}{\partial t} = \operatorname{div}(c(\Delta u)\Delta u)$$

Applying anisotropic diffusion to images allows noise to be reduced without distorting the edges. When the diffusion rate is held constant, anisotropic diffusion can be calculated using the heat equation, which is equivalent to Gaussian blurring. Even while it seems to help with noise reduction, this method also randomly and completely blurs edges. When the diffusion coefficient is specified as a threshold function, as in Perona Malik, the governing equation favors diffusion (and consequently filtering) over smooth pixel intensity zones while suppressing it everywhere else in the input image. Therefore, the image's components are preserved while the noise is filtered out. The GLSZM feature extraction method is used in the next step to extract features from the output filter image.

The volume of associated, homogenous zones of a particular size and intensity within the volume is described by a grey-level size-zone matrix. The $(m,n)^{th}$ item of the GLSZM p(m,n) indicates the quantity of interconnected gray level m and size n regions. In order to describe heterogeneity at a local level, GLSZM characteristics thus describe affected homogenous regions within the affected volume [13]. Let:

- $p(m,n)$ is the $(i,j)^{th}$ element in the existing GLSZM p,
- N_g the count of the image's distinct pixel intensities,
- N_z the extent of the most substantial, homogenous area within the amount of interest,
- N_s the overall number of homogeneous areas (zones), where $N_s = \sum_{m=1}^{N_g}\sum_{n=1}^{N_z} p(m,n)$
- p_z the distribution of all the zones with sizes j, added together, where $p_z(n) = \sum_{m=1}^{N_g} p(m,n)$
- p_g the distribution of grey levels m over the total number of zones, where $p_g(m) = \sum_{n=1}^{N_z} p(m,n)$
- N_p the quantity of image voxels, where $N_p = \sum_{n=1}^{N_z} np_r$
- μ_r average zone size, where $\mu_r = \sum_{m=1}^{N_g}\sum_{n=1}^{N_z} jp(m,n|\theta)$
- μ_g the average gray level, where $\mu_g = \sum_{m=1}^{N_g}\sum_{n=1}^{N_z} ip(m,n|\theta)$

- **Small Zone Emphasis (SZE)**

$$SZE = \sum_{n=1}^{N_z} \frac{p_z}{n^2}$$

- **Large Zone Emphasis (LZE)**

$$LZE = \sum_{n=1}^{N_z} n^2 p_z$$

- **Gray Level Non-uniformity (GLN)**

$$GLN = \sum_{m=1}^{N_g} p_g^2$$

- **Zone-Size Non-uniformity (ZSN)**

$$ZSN = \sum_{m=1}^{N_g} p_Z^2$$

- **Zone Percentage (ZP)**

$$ZP = \frac{N_s}{N_p}$$

- **Low Gray-Level Zone Emphasis (LGZE)**

$$LGZE = \sum_{m=1}^{N_g} \frac{p_g}{m^2}$$

- **High Gray-LevelZone Emphasis (HGZE)**

$$HGZE = \sum_{m=1}^{N_g} m^2 p_g$$

- **Small Zone High Gray-Level Emphasis (SZLGE)**

$$SZLGE = \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} \frac{p(i,j)}{m^2 n^2}$$

- **Large Zone Low Gray-Level Emphasis (LZLGE)**

$$LZLGE = \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} \frac{p(m,n)n^2}{m^2}$$

- **Large Zone High Gray-Level Emphasis (LZHGE)**

$$LZHGE = \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} p(m,n)n^2 m^2$$

- **Gray-Level Variance (GLV)**

$$GLV = \frac{1}{N_g \, X \, N_Z} \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} \left(ip(m,n) - \mu_g\right)^2$$

- **Zone Size Variance (ZSV)**

$$ZSV = \frac{1}{N_g \, X \, N_Z} \sum_{m=1}^{N_g} \sum_{n=1}^{N_z} (np(m,n) - \mu_z)^2$$

Features	Normal	Abnormal
SZE	0.80021	0.67983

LZE	590.21	625.64
GLN	0.00522	0.0062
ZSN	0.59599	0.42852
ZP	0.60025	0.35265
LGZE	0.00132	0.00028
HGZE	17376.5	27944.5
SZLGE	0.00067	0.00016
SZHGE	13685.7	17669.1
LZLGE	585.81	0.02942
LZHGE	84608.2	3.80E+04
GLV	6.30E-05	0.00014
ZSV	2.15E-08	6.16E-08

Table 2. GLSZM feature values of normal and abnormal images.

IV. Classification

In this section we are going to discuss three types of classifiers. They are logistic regression, KNN, booster tree classifiers. The above classifiers identify and classify normal and abnormal images from the extracted feature values.

Machine learning has adapted the classifications method known as logistic regression from statistics. Logistic regression is a statistical method for examining a sample where one or much more distinct factors can affect a result. A classifier is that any method with a given criteria that discretizes the output as classes. Because it is a regression of class probabilities, logistic regression is a fitting name for the technique.

An algorithm for classifying data is logistic regression. It is employed to predict a binary outcome using a number of variables (1/0, Yes/No, True/False). We used dummy variables to express binary or categorical outcomes. Logistic regression is a special example of linear regression when the main result is classified and the log of the probability is used as the regression model. It determines the likelihood of an event occurring an event will occur by fitting the experimental data to a binary logistic regression function[14].

The generalised linear model's core equation is:

$$g\big(E(y)\big) = \alpha + \beta x_1 + \gamma x_2$$

Here g() is the link equation, E(y) is the is the targeted variable's predicted value and $\alpha + \beta x_1 + \gamma x_2$ is the linear predictor (α, β, γ to be predicted).

$$\log\left(\frac{p}{1-p}\right) = y$$

Here $\log\left(\frac{p}{1-p}\right)$ is the link function. If p is the likelihood of success, $1-p$ will be the likelihood of failure. A non-linear association can be modelled linearly by applying a logarithmic transformation to the outcome variable.

The next type of classification method is K-Nearest Neighbor. KNN, or K-nearest Neighbor, is one of the supervised learning methods that is frequently used to classify data based on how its neighbors are classified. KNN keeps track of all of the available instances and classifies new cases based on a similarity score. The supervised object classification algorithm K-Nearest Neighbors (k-NN) gains knowledge from a labelled training set by swallowing the training and testing X and its labels Y and learning to map the input X to the desired output Y. Undoubtedly the simplest machine learning algorithm is the k-NN. The model only employs training data; that is, it learns the entire training set and outputs the category with the huge number of the closest 'k' neighbors identified by some distance measure [15].

The test image to be predicted is given to the model after it has stored the training set for prediction. After calculating the distance between each image in the training images and the test image, the model finds the 'k' training set that is most similar to the test dataset. The class is then produced via a voting process, typically a majority vote, based on the labels of these ‘k’ neighbours.

It is possible to utilise different distance metrics to determine distances, such as the L1 distance function, which adds up the differences between pixels in two images..

$$d_1(I_1, I_2) = \sum p|Ip_1 - Ip_2|$$

The final classifiacation techniques is boosted tree algorithm. Each tree depends on earlier trees when there is boosting. By adjusting the residual of the trees that came before it, the algorithm learns. As a result, boosting tends to increase accuracy while carrying a slight risk of reduced coverage in a decision tree ensemble. Based on the LightGBM algorithm, this component [16].

The Boosted Trees Model is an example of an additive model that combines judgements from a series of base models to provide predictions. Formally, we can express this class of models as follows:

$$g(x) = f_0(x) + f_1(x) + f_2(x) + \cdots$$

where the total of the simple base classifiers f_i makes up the final classifier, denoted as g. Each base classifier in the boosted trees model is a straightforward decision tree. Model ensembling is a wide term for the use of numerous models to improve predictive performance.

V. Experimental Results

When evaluating the classifier's performance for classifying chest X-ray images, sensitivity, specificity, accuracy, error rate, F1 Score, and MSE of the classifier are taken into

consideration. Due In fact, 10% of the data was used for training and 90% was used for testing. Below is a list of the formula for standard parameters to find the classifiers' performance.

$$PI = \frac{PC - MC - FA}{PC} X100$$

Perfect Classification (PC), False Alarm (FA), and Missed Classification (MC) are the used in this. The mathematical formulas for the specificity, sensitivity, and accuracy is

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

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

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

The objective is to understand the training process for the Mean Square Error Classification Model.

$$\text{MSE} = \frac{1}{\text{N}} \sum_{\text{I}=1}^{\text{N}} (\text{P}_\text{i} - \text{Q}_\text{i})^2$$

Where P_i indicates the value of the observed over a specific time, N is the number of samples per patient, and Q_i denotes the target's value over the representative j.

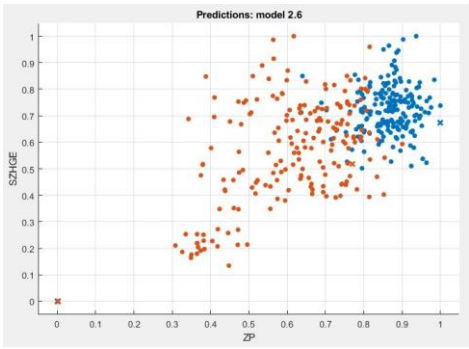

Figure 4. Scatter plot of Logistic Regression

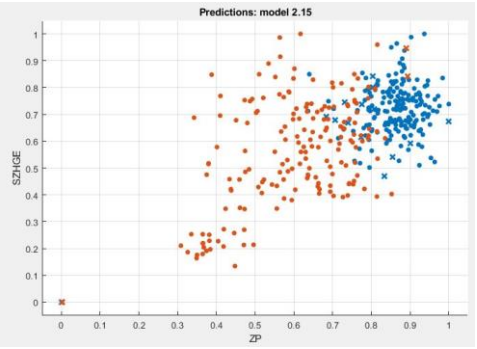

Figure 5. Scatter plot of KNN

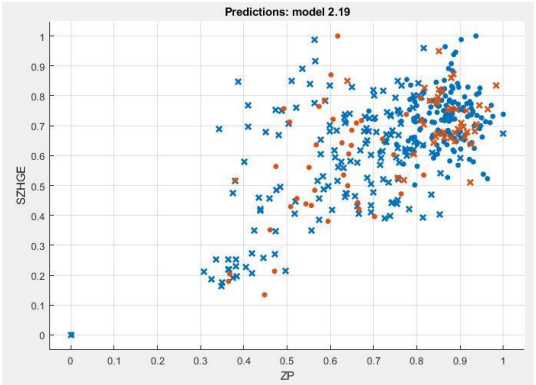

Figure 6. Scatter plot of boosted tree

Figure 4, 5, and 6 dipicts the Scatter plot of the three classifiers Logistic Regression, K-Nearest Neighbor, and Boosted Tree and the below Figure 7, 8, and 9 shows the confusion matrix of the three classifiers.

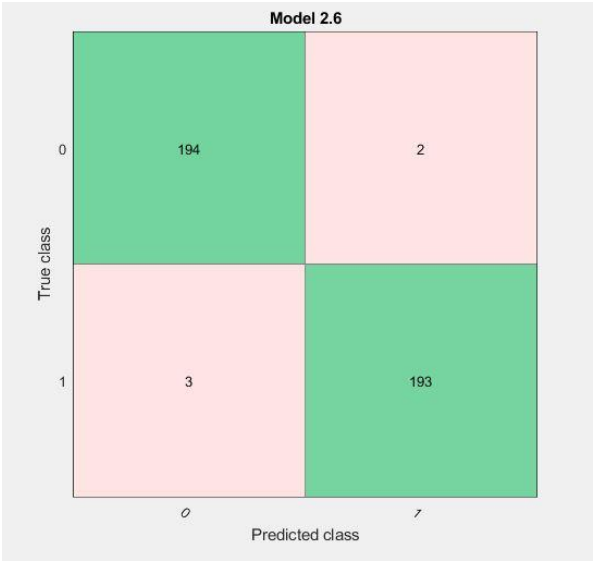

Figure 7. Confusion matrix of Logistic Regression

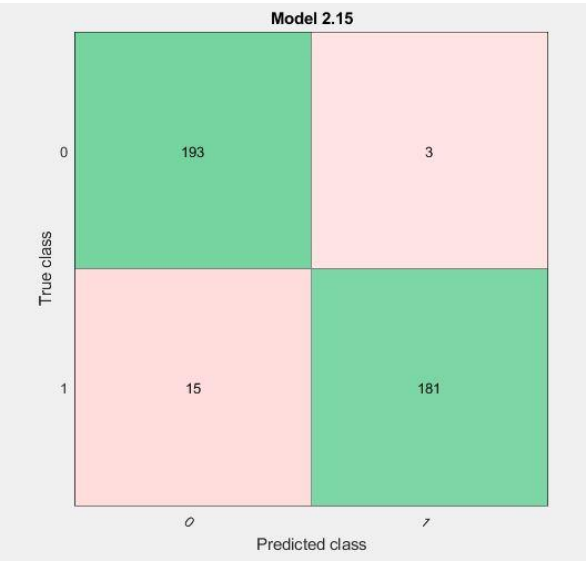

Figure 8. Confusion matrix of KNN

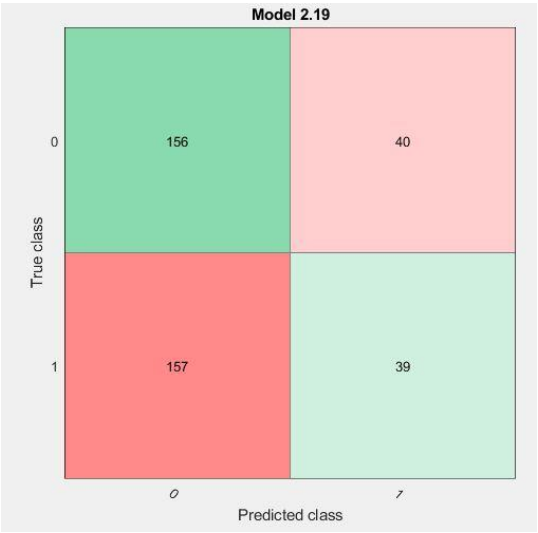

Figure 9. Confusion matrix of Boosted tree

	Classifiers		
Parameters (%)	**Logistic Regression**	**KNN**	**Boosted Tree**
Sensitivity	98.97959	98.46939	79.59184
Specificity	98.46939	92.34694	19.89796
Accuracy	**98.72449**	95.40816	49.7449

Errorrate	1.27551	4.591837	50.2551
Precision	98.47716	92.78846	49.84026
F1 Score	98.72774	95.54455	61.29666
Jacard Metric	97.48744	91.46919	44.19263
Balanced Classifier Rate	98.72449	95.40816	49.7449
MCC	0.974502	0.90987	-0.00636

Table 3. Classifiers' Performance

VI.Conclusion

The protracted procedure of processing medical images involves preprocessing, feature extraction, and classification. This paper offers a way for deciding on the best classification method for classifying medical images. Throughout the preprocessing phase, a fourth-order PDE filter was applied. From the filtered image, the features must be extracted, GLSZM is computed individually in the Feature Extraction stage. Logistic Regression, KNN, and Boosted Tree classifiers are used in the classification step to compare the effectiveness. The results of the classifier are compared based on accuracy. It has been demonstrated through testing the proposed methodology that Logistic Regression outperforms the other two classification methods when employed with GLSZM features on images from a Covid-19 chest X-ray with 98.42%.

References

1. Wang, W.; Xu, Y.; Gao, R.; Lu, R.; Han, K.; Wu, G.; Tan, W. Detection of SARS-CoV-2 in different types of clinical specimens. JAMA 2020, 323, 1843–1844.
2. Hafiz M. N. Iqbal, Kenya D. Romero-Castillo, Muhammad Bilal and Roberto Parra-Saldivar, The Emergence of Novel-Coronavirus and its Replication Cycle – An Overview, J. Pure Appl. Microbiol., 2020; **14**(1): 13-16. https://doi.org/10.22207/JPAM.14.1.03.
3. Ji T, Liu Z, Wang G, Guo X, Akbar Khan S, Lai C, Chen H, Huang S, Xia S, Chen B, Jia H, Chen Y, Zhou Q. Detection of COVID-19: A review of the current literature and future perspectives. Biosens Bioelectron. 2020 Oct 15;166:112455. doi: 10.1016/j.bios.2020.112455. Epub 2020 Jul 21. PMID: 32739797; PMCID: PMC7371595.
4. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar;579(7798):270-273. doi: 10.1038/s41586-020-2012-7. Epub 2020 Feb 3. Erratum in: Nature. 2020 Dec;588(7836):E6. PMID: 32015507; PMCID: PMC7095418.

5. Sahu, Kalicharan & Minz, Sonajharia. (2022). Self-adaptive-deer hunting optimization-based optimal weighted features and hybrid classifier for automated disease detection in plant leaves. Expert Systems. 39. 10.1111/exsy.12982.
6. Mehshan Ahmed Khan, Muhammad Attique Khan, Fawad Ahmed, Mamta Mittal, Lalit Mohan Goyal, D. Jude Hemanth, Suresh Chandra Satapathy, Gastrointestinal diseases segmentation and classification based on duo-deep architectures, Pattern Recognition Letters, Volume 131, 2020, Pages 193-204, ISSN 0167-8655,https://doi.org/10.1016/j.patrec.2019.12.024.
7. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big Data Analytics in Healthcare. Biomed Res Int. 2015;2015:370194. doi: 10.1155/2015/370194. Epub 2015 Jul 2. PMID: 26229957; PMCID: PMC4503556.
8. M. Ur Rehman, F. Ahmed, M. Attique Khan, U. Tariq, F. Abdulaziz Alfouzan et al., "Dynamic hand gesture recognition using 3d-cnn and lstm networks," Computers, Materials & Continua, vol. 70, no.3, pp. 4675–4690, 2022.
9. Gulzar, Y.; Khan, S.A. Skin Lesion Segmentation Based on Vision Transformers and Convolutional Neural Networks—A Comparative Study. Appl. Sci. 2022, 12, 5990. https://doi.org/10.3390/app12125990.
10. Sai Pandraju, T. K., Samal, S., Saravanakumar, R., Yaseen, S. M., Nandal, R., & Dhabliya, D. (2022). Advanced metering infrastructure for low voltage distribution system in smart grid based monitoring applications. Sustainable Computing: Informatics and Systems, 35 doi:10.1016/j.suscom.2022.100691
11. Sharma, R., & Dhabliya, D. (2019). A review of automatic irrigation system through IoT. International Journal of Control and Automation, 12(6 Special Issue), 24-29. Retrieved from www.scopus.com
12. Sharma, R., & Dhabliya, D. (2019). Attacks on transport layer and multi-layer attacks on manet. International Journal of Control and Automation, 12(6 Special Issue), 5-11. Retrieved from www.scopus.com
13. Chen Zhao, Renjun Shuai, Li Ma, Wenjia Liu, and Menglin Wu. 2022. Segmentation of skin lesions image based on U-Net + +. Multimedia Tools Appl. 81, 6 (Mar 2022), 8691–8717. https://doi.org/10.1007/s11042-022-12067-z.
14. Ünlü, E.I.; Çinar, A. Segmentation of Benign and Malign lesions on skin images using U-Net. In Proceedings of the 2021 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), Zallaq, Bahrain, 29–30 September 2021; IEEE: New York, NY, USA, 2021; pp. 165-169.
15. Ullah, F.; Ansari, S.U.; Hanif, M.; Ayari, M.A.; Chowdhury, M.E.H.; Khandakar, A.A.; Khan, M.S. Brain MR Image Enhancement for Tumor Segmentation Using 3D U-Net. Sensors 2021, 21, 7528. https://doi.org/10.3390/s21227528
16. Chu, A., Sehgal, C. M., & Greenleaf, J. F. (1990). Use of gray value distribution of run lengths for texture analysis. Pattern Recognition Letters, 11(6), 415-419. https://doi.org/10.1016/0167-8655(90)90112-F.
17. Brzezinski J.R. and Knafl G.J, Logistic regression modeling for context-based classification. Database and Expert Systems Applications, 1999. Proceedings. Tenth International Workshop on, 755-759, 1999doi: 10.1109/DEXA.1999.795279,(1999).

18. H.M Zelelew, A.T Papagiannakis,"A volumetric thresholding algorithm for processing asphalt concrete x-ray CT images". International journal of pavement engineering, Sept. 2007.
19. H. Prosper, Multivariate discriminants, “Ensemble learning”, EPJ Web of Conferences 4 02001, 2010, SOS’08 - School of Statistics.