

ENHANCING COVID-19 CT-SCAN CLASSIFICATION THROUGH INTEGRATION OF GLCM, CCV, DWT FEATURES, AND GENETIC ALGORITHM OPTIMIZATION

Yogendra Narayan Prajapati ¹ and Dr. Manish Sharma ²

¹ PhD Scholar, Computer Science & Engineering, Quantum University, Roorkee, India.
Email: ynp1581@gmail.com

² Professor, Computer Science & Engineering of Quantum University, Roorkee, India.
Email: director@quantumeducation.in

DOI: 10.5281/zenodo.11096358

Abstract

This work proposes a novel approach to improve the classification of COVID-19 CT images using a combination of Gray-Level Co-occurrence Matrix (GLCM), Colour Co-occurrence Vector (CCV), and Discrete Wavelet Transform (DWT) features. Moreover, a Genetic Algorithm (GA) is employed to enhance the classification efficacy even more. Using these state-of-the-art approaches, the proposed strategy aims to improve the COVID-19 CT scan categorization's accuracy and durability. The aim of this work is to improve a genetic algorithm for the classification of COVID-19 CT scan images by including a discrete wavelet transform (DWT), colour consistency vector (CCV), and GLCM removal approach at high grayscale levels. This combination aims to increase the discriminatory power of certain criteria and thus increase the accuracy of the machine learning model. Specifically, Linear Support Vector Machine (L-SVM), Fisher-KNN (F-KNN) and Fisher visual decision tree (F-Tree) were used to model the classification. Genetic methods have been employed to create specific options aimed at improving the accuracy of models in identifying COVID-19 patterns on CT images. The results of this study provide insight into ongoing efforts to improve COVID-19 diagnostics through the integration of architecture, genetic algorithms, and machine learning models. Among the proposed methods, the fuzzy tree classifier achieved 93.55% accuracy, whereas some other classifiers (Naive Bayes, L-SVM and F-KNN) achieved 92.6%, 92.12%, and 92.33% accuracy, respectively.

Keywords: Machine Learning Model, Genetic Algorithms-KNN, L-SVM, GCLM, CCV, DWT, CT-Scan.

I. INTRODUCTION

In recent years, the increase in the amount of information available in different fields has led to the search for advanced techniques for obtaining useful patterns and insights. Machine learning, particularly in the field of image analysis and classification, has shown interest in the use of techniques to eliminate complex algorithms and optimize machine learning models to improve the ground truth. learning models to improve ground truth correction. This study conducts in-depth research on the combination of grayscale co-occurrence matrix (GCLM), color consistency vector (CCV), and discrete wavelet transform (DWT) feature extraction methods, and enhances them with seed genetic algorithms to improve discrimination. Capacity of feature area. The study also investigated the effectiveness of three machine learning models (Support Vector Machine (SVM), random forest, and K Nearest neighbors (KNN)) when used to optimize the set.

Main Objectives: The aim of this research is to solve problems related to high-dimensional information and its relationship in a special space. GCLM, CCV, and DWT have emerged as powerful inference techniques, each providing unique capabilities to capture spatial, color, and

Frequency-domain information, respectively. By combining these models and using genetic algorithms for feature selection, we aimed to simplify the location feature and focus on information about the distribution function.

The genetic algorithm used in this study is a multi-objective optimization tool that dynamically adjusts and improves feature sets to achieve the ratio of the machine learning model. This approach not only increases the performance of the model but also leads to a more interpretable and effective use. The SVM, Random Forest and KNN were selected as machine learning models for evaluation because of their wide applications and distinct features. SVM, which is known for its effectiveness in processing heterogeneous data, is expected to benefit from optimization. Random Forest is a learning algorithm that is expected to show good performance and resilience to over performance. KNN is a proximity-based option because of its simplicity and adaptability to local standards in specific locations.

In conclusion, this study aims to contribute to the continuous development of machine learning by proposing an integrated approach that combines the best process extraction methods with genetic algorithms for optimization. The subsequent evaluation of SVM, Random Forest, and KNN aims to provide insights into the strengths and limitations of each model, offering a comprehensive understanding of the effectiveness of the proposed methodology across various datasets and domains.

II. LITERATURE SURVEY

This article presents a novel project focused on swiftly diagnosing COVID-19 through computed tomography (CT). Given the heightened workload of radiologists and the wealth of information they contribute, there is a pressing need to identify an improved method for promptly diagnosing COVID-19 based on CT images. The procedure encompasses data collection and modeling, elimination of pertinent features, optimal feature selection, and feature classification [8].

The proposed method achieved an accuracy of 92.6% when combined with a Naive Bayes classifier. This framework involves extracting the results using adaptive transformations and continuous partitioning based on fractal texture analysis methods. This study also discusses the importance of the early detection and diagnosis of COVID-19 and the use of artificial intelligence and computer vision for electronic diagnosis. Chest CT scans were collected from 35 individuals who tested positive for COVID-19 and were used for the study [1].

This article describes the development of computer-aided diagnostic tools for the detection of COVID-19 using X-rays. He noted difficulties in diagnosing COVID-19 due to limited equipment in X-ray imaging and the need for expert radiologists to distinguish the disease from healthy samples. The proposed method uses a fully integrated neural network to analyze the texture of chest X-ray images and achieves an accuracy of 98.39% with a 10-fold cross-validation[9].

The introduction provides background information on the COVID-19 pandemic and its impact on global health. It shows the problems encountered in diagnosing the disease and the widespread use of medical tests, especially chest X-rays, for the initial diagnosis. This article also discusses the preliminary features of the infection and highlights the importance of images in identifying patterns associated with COVID-19 [2].

This paper presents the tracking of distribution pipeline images that combine spatial texture representation using Wasserstein distance from optical transition (OMT) theory. This article focuses on the use of these models in the diagnosis of COVID-19-associated pneumonia, and presents the results of the classification process. This research builds on radiomics studies that aim to increase the value of medical imaging [10].

The authors introduced a new spatial texture representation and deployed it to monitor distributed image pipelines using a vector machine (SVM) algorithm with Bayesian optimization and the Wasserstein metric. Thus, this method is suitable for COVID-19 screening based on computed tomography images. This study investigated the importance of texture patterns for good classification of medical images and examined the selection of representative patterns from data to improve classification accuracy. This study provides positive results, demonstrating the effectiveness of the tube in diagnosing COVID-19 through CT imaging. This article summarizes the importance of the proposed method and demonstrates its effectiveness in clinical settings [3].

The survey covered the background of COVID-19, the challenges of RT-PCR testing, and the advantages of testing using breast imaging technology. He talked about exploring various machine-learning and deep-learning methods for researching COVID-19. This article also highlights the uniqueness of this research by comparing it with previous studies.

This study introduced different extraction methods, such as ResNet, VGG, DenseNet, Inception, and Xception, to detect COVID-19 from CT scans and chest X-rays. Various classification strategies used in research have also been examined with reference to binary classification.

The results of this study show the effectiveness of different inferences and classifications in identifying COVID-19 from image data. In addition to visual performance, the survey demonstrated the use of visual interpretation, electronic analysis, and regional abnormalities [4].

This article describes research on COVID-Net, a deep neural network designed to identify COVID-19 patients from chest X-ray (CXR) images. Authors Linda Wang, Zhong Qiu Lin, and Alexander Wong highlight the need for quality analysis during the COVID-19 pandemic and highlight the limitations of current gold standards for polymerase chain reaction (RT-PCR) reverse engineering. They launched COVID-Net as a tool to help electrical engineers detect COVID-19 cases quickly. This tool is particularly useful in resource-limited areas and during peak infection times.

This study reports on COVID-19, an open-access database of five existing databases containing 13,975 CXR images from 13,870 patients, many of with COVID-19 are in good condition. COVID-Net is designed to be lightweight and efficient using a novel projection-extension-projection-extension (PEPX) design that reduces computational complexity while maintaining high representation capacity [3].

The authors also explored the interpretation of COVID-19 Net forecasts to increase their transparency and reliability. They used GSInquire, an interpretive method, to analyze the network and ensure that their decisions are based on relevant images. The study's results show that COVID-Net achieves high accuracy and sensitivity for COVID-19 detection, outperforming traditional deep learning architectures like VGG-19 and ResNet-50[4],[5].

The research delves into the significance of swift triaging and highlights the accessibility and portability advantages of CXR imaging compared to CT scans. The study also acknowledges the essential role of expert radiologists in image interpretation, a aspect that COVID-Net seeks to support. The authors conclude that although COVID-Net is not currently a production-ready solution, it represents a progressive stride in creating precise and practical deep learning systems for COVID -19 detection. They express optimism that the open-source nature of COVID-Net and the broader COVID-19 context will foster collaboration and inspire additional research in this domain [5].

Table 1

Paper detail	Dataset	Mythology	Result
Navid Ghassemi at all.[13]	Doctors collected and aggregated data on 3163 images from 189 patients to evaluate the model	The paper reports a method to detect COVID-19 from CT images using pre-trained deep neural networks and Cycle Generative Adversarial Network (CycleGAN) models for data augmentation.	A method based on pre-learning deep neural networks achieves state-of-the-art performance in detecting COVID-19 from CT images with 99.60% accuracy
Yicheng Fang at all [14].	The study included a series of 51 patients who underwent both chest CT and real-time	Methods used in the paper "Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR"	The sensitivity of chest CT for 2019 novel coronavirus infection was found to be 98% when compared to real-time polymerase chain reaction assay (RT-PCR) performed within 3 days
Deepraj Chowdhury at all[15]	The code, datasets, and results' scripts of CoviDetector were released on GitHub for reproducibility purposes.	Transfer learning was utilized to leverage pre-trained models and adapt them to the task of COVID-19 detection, allowing for improved performance and reduced training data requirements	The proposed CoviDetector achieved over 99% accuracy on four different datasets, demonstrating its high performance in detecting COVID-19 using CXR images

CT Imaging Analysis for COVID-19:

CT Imaging Analysis of COVID-19 patients

1. Characteristics: High sensitivity to lung tissue: CT scans provide detailed images of lung structures, making it easier to identify abnormalities.

Useful for Early Detection: CT is often used for early detection of COVID-19, especially in cases where initial PCR tests may yield false negatives.

2. Common Findings in CT Scans for COVID-19

Ground-Glass Opacities (GGO): GGO is a common finding in COVID-19 patients and appears as hazy areas with increased lung density.

Consolidation: This involves the replacement of air-filled spaces with substances, such as fluids or cellular materials.

Bilateral Involvement: COVID-19 often affects both lungs.

3. Quantitative Analysis:

Volumetric Assessment: CT allows for quantification of lung involvement and aids in disease severity assessment.

CXR imaging of COVID-19

1. Characteristics:

Widespread Availability: CXR is widely available and less expensive than CT, making it a valuable tool in resource-limited settings.

Quick Imaging: CXR can be performed quickly, allowing

for rapid assessment.

2. Common findings in CXR for COVID-19
Bilateral infiltrates: Similar to CT findings, COVID-19 often presents with bilateral involvement.

Opacities: Opacities, including GGO and consolidation, may be visible on CXR.

Combined Analysis:

1. Enhanced Sensitivity: Combining Modalities: Integrating CT and CXR findings can enhance sensitivity in detecting COVID-19-related abnormalities.
2. Sequential Imaging: Monitoring Disease Progression: Sequential imaging with both modalities helps in monitoring disease progression and treatment response.
3. Clinical Correlation: Clinical Integration: Imaging findings should be correlated with clinical symptoms, PCR results, and other diagnostic data for comprehensive diagnosis.
4. Machine Learning Applications: Automated Analysis: Machine learning algorithms can be trained on CT and CXR images to aid in automated detection and classification of COVID-19 features.

Dataset Collection

Data collection for COVID-19 CT and CXR image analysis involves obtaining images from multiple sources, including hospitals, research centers, and public archives. As of my last knowledge update in January 2023, several datasets related to COVID-19 imaging are available. It is essential to check for the latest datasets and adhere to data-usage policies and ethical considerations. Potential sources of COVID-19 imaging datasets are as follows:

COVID-19 Image Data Collection (cohen2020)

This repository includes chest CT images and chest X-ray images for COVID-19 and other diseases: link: <https://academictorrents.com/browse.php?search=+chest+CT>

III. PROPOSED METHODOLOGY

Early learning methods used the same or hybrid methods, such as the DWT and GLCM, for inference. Although both methods have advantages and disadvantages, in general, the combined feature space is more capable of preserving the smart features. Because of this change, hybrid methods have become popular among researchers in the field of computer vision.

Some of them as

1. CCV (Color Coherence Vector)
2. Tamura Texture Features.
3. DWT (Discrete Wavelet Transform).
4. GLCM (Gray-Level Co-occurrence Matrix)

However, choosing the most appropriate extraction method is a delicate task and must be handled with care; otherwise, it will result in a believable and relatable breakdown. In this study, we used four different methods involving two groups: statistics and data. In this application, two special families—colors and shapes—are not considered because their impact and importance are limited. In this study use some important feature selection method for better model performance like

- Normalize the DWT coefficients and GLCM features
- Ranking Features
- Genetic Algorithm (GA) Optimization
- Validation feature

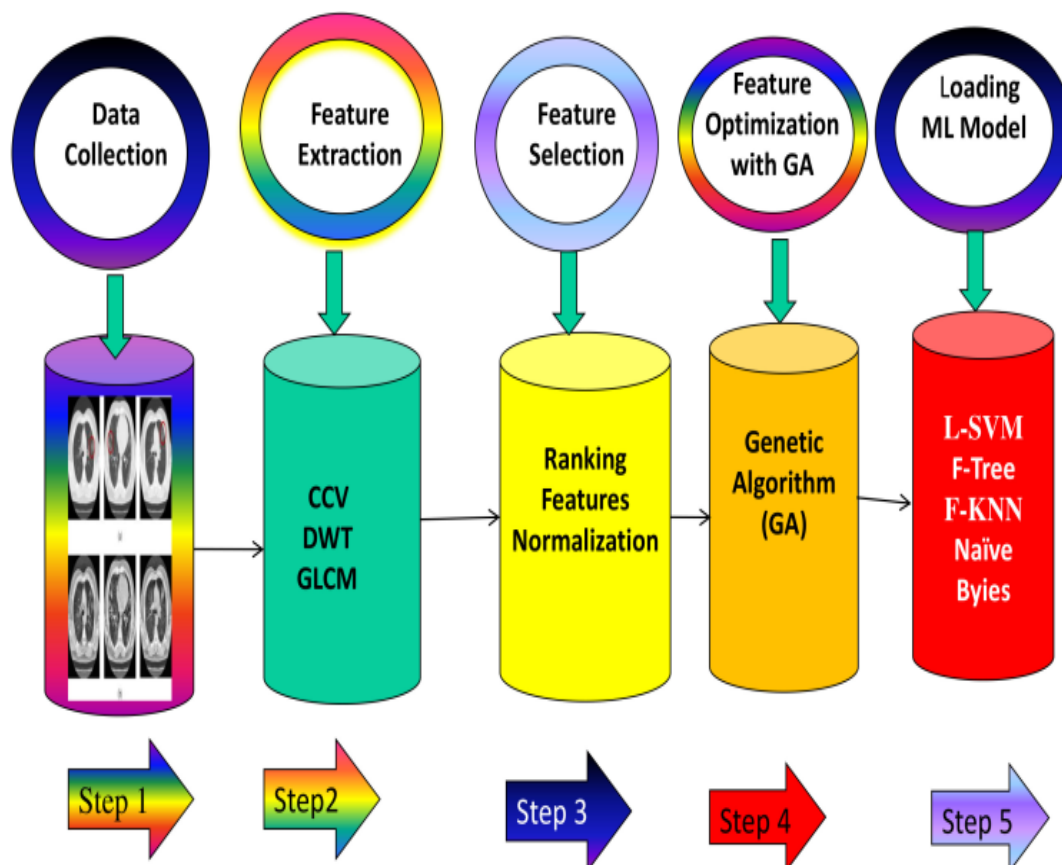
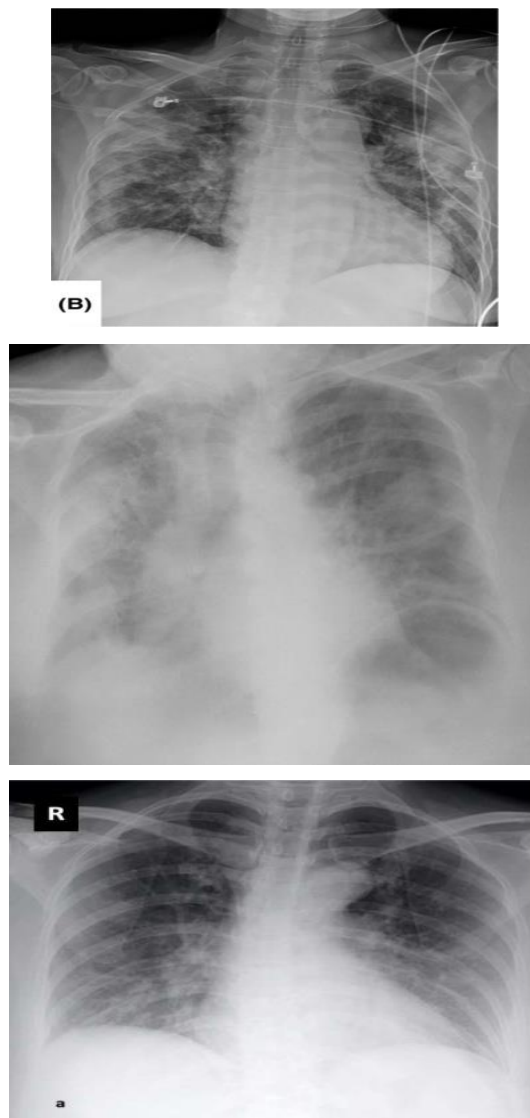


Figure 1: Proposed

1 The proposed framework (Figure 1) is discussed in the text below.
Some sample image as given below:



IV.METHOD FOR FEATURE EXTRACTION

CCV (Color Coherence Vector):

CCV is a representation of the image color distribution. It captures the spatial coherence of colors in an image and provides information on how colors are distributed in different regions.

Let I be an image with M rows and N columns, where

$I(i,j)$ represents the color of a pixel at position (i,j) . The image is assumed to be in a color space such as RGB.

The first step in CCV involves quantizing the color space into a set of discrete color bins. Let C be the number of quantized color bins.

Color Coherence:

For each quantized color bin c , the color coherence $CC(c)$ is calculated as the percentage of neighboring pixels with the same or similar color. Coherence is often defined based on a color similarity threshold.

$$CC(c) = \frac{\text{Number of pixels with color } c \text{ and coherent neighbors}}{\text{Total number of pixels with color } c} \times 100$$

The Color Coherence Vector, denoted as:

$[(CC(c_1), CC(c_2), CC(c_3), \dots, CC(c_C))]$ is a vector that represents the color coherence for each quantized color bin.

CCM (Color Correlogram):

$[(CC(c_1), CC(c_2), CC(c_3), \dots, CC(c_C))]$ is a vector that represents the color coherence for each quantized color bin.

Color Correlogram (CCM)

CCM is a statistical measure that describes the spatial correlation of color occurrences in an image. This method is useful for texture and pattern analyses in computer vision.

The experimental steps for computing the Color Correlogram are outlined as follows:

Color Quantization:

The color space of an image was divided into a discrete set of color bins. Let C be the number of quantized color bins. Initialize Correlogram:

Create a 2D histogram $H(c,d)$, where c represents a quantized color bin and d represents a distance bin. Initialize all bins to zero.

Count Color Pairs:

For each pixel $P(i,j)$ in the image, for every pixel at a distance d from $P(i,j)$, increment the corresponding bin $H(c,d)$ if the color of the distant pixel is the same as that of $P(i,j)$.
 $\text{Count of color pairs with color } c \text{ at distance } d = H(c,d) + \text{Count of color pairs with color } c \text{ at distance } d$

Normalize correlogram:

Normalize the histogram $H(c,d)$ to obtain color correlogram values between 0 and 1. This is often performed by dividing each bin by the total number of occurrences of color c in the image. Normalized Correlogram: Total occurrences of color in the image

$$\text{Normalized Correlogram: } N = \frac{H(c, d)}{\text{Total occurrences of color } c \text{ in the image}}$$

DWT (Discrete Wavelet Transform):

The discrete wavelet transform (DWT) serves as a specialized application of wavelet transform, frequently employed for compression in signal and image processing, along with its role in feature extraction.

One-Dimensional Discrete Wavelet Transform (1D DWT)

Let $x[n]$ be a one-dimensional signal of length N . The DWT decomposes $x[n]$ into approximation coefficients (A) and detail coefficients (D).a.

Approximation Coefficients (A):

$$A_j[N] = \sum_k h[k].x[2n - k]$$

Where $h[k]$ is the low-pass filter (scaling function), and n is an index in the original signal.

Detail Coefficients (D):

$$D_j[N] = \sum_k g[k].x[2n - k].$$

Where $g[k]$ is the high-pass filter (wavelet function). The process is typically repeated by down-sampling the signal ($x[n]$) and applying the above equations iteratively to obtain coefficients at different scales.

Two-Dimensional Discrete Wavelet Transform (2D DWT): Consider $X[m, n]$ as a two-dimensional image with dimensions $M \times N$. Through the 2D discrete wavelet transform (DWT), $X[m, n]$ undergoes decomposition into four constituents: approximation (A), horizontal detail (H), vertical detail (V), and diagonal detail (D).

Approximation Coefficients (A):

$$A_j[m, n] = \sum_{k,l} h[k].h[l].X[2m - k, 2n - l]$$

Horizontal Detail Coefficients (H):

$$H_j[m, n] = \sum_{k,l} g[k].h[l].X[2m - k, 2n - l]$$

Vertical Detail Coefficients (V):

$$V_j[m, n] = \sum_{k,l} h[k].g[l].X[2m - k, 2n - l]$$

Diagonal Detail Coefficients (D):

$$D_j[m, n] = \sum_{k,l} g[k].g[l].X[2m - k, 2n - l]$$

In the 2D discrete wavelet transform (DWT), the application of the 1D DWT occurs independently along both the rows and columns of the image. This iterative procedure is then repeated, resulting in a multi-resolution representation of the image.

GLCM (Gray-Level Co-occurrence Matrix):

GLCM, a statistical technique employed for texture analysis in images, computes the occurrence of pixel pairs with designated intensity values and distances. This process yields valuable information concerning texture patterns within the image.

The GLCM can be used to compute various texture features or measures that describe the texture of an image. Some common measures include:

Contrast (CON):

$$CONS = \sum_{i,j} (i - j)^2 . P(i, j, d, \theta)$$

Energy or Angular Second Moment (ASM):

$$ASM = \sum_{i,j} [P(i,j,d,\theta) \cdot P(i,j,d,\theta)]$$

Homogeneity (HOM):

$$HOM = \sum_{i,j} \left[\frac{1}{1 + (i-j)} \cdot P(i,j,d,\theta) \right]$$

These measures provide quantitative information about the spatial distribution of pixel intensities in the image, capturing aspects of texture such as contrast, homogeneity, and correlation. The GLCM and its associated measures are widely used in image analysis and pattern recognition.

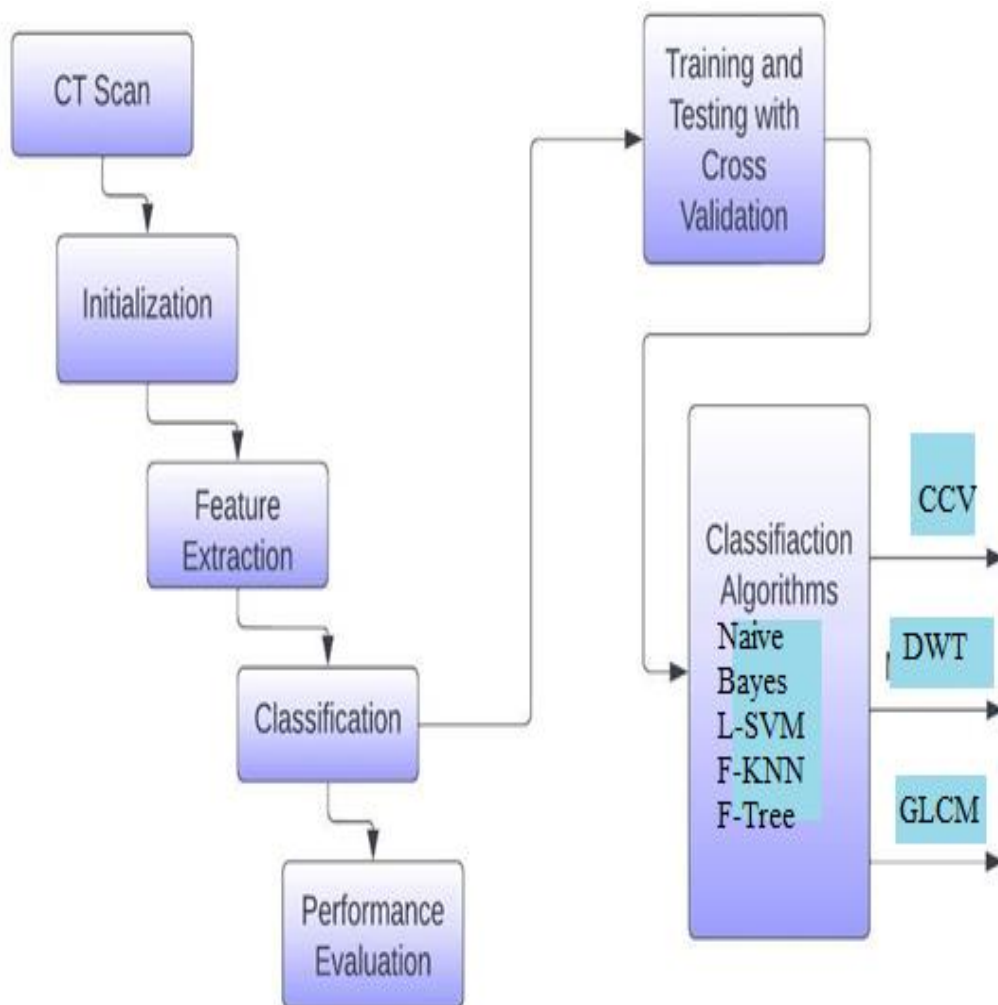


Figure 1: Feature Extracting Ratio of Infection

Feature Selection: Feature selection is crucial in improving the efficiency and effectiveness of data analysis, particularly in scenarios where high-dimensional data, such as DWT and GLCM features, are involved. Here's a novel feature selection method for DWT and GLCM feature extraction

Hybrid Feature Selection Method combining DWT and GLCM:

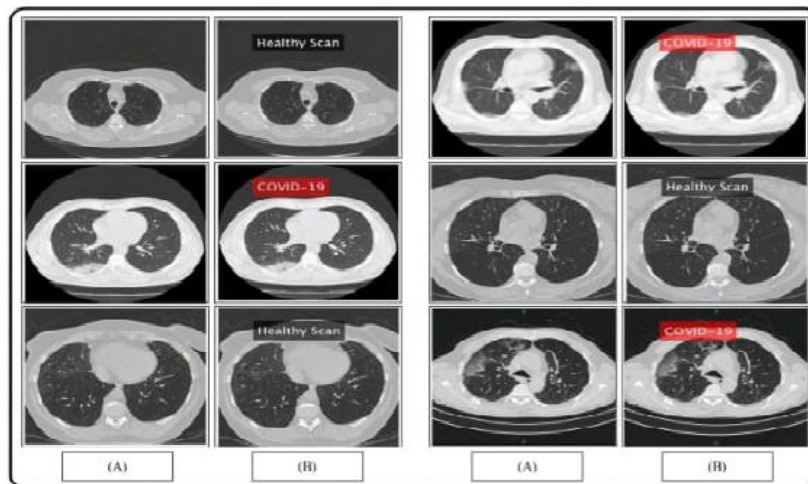


Figure 2: Get an Estimate. An Old Picture; B Proposed Predicted Labeled Image

V. FEATURE EXTRACTION

Feature selection: Feature selection is crucial in improving the efficiency and effectiveness of data analysis, particularly in scenarios where high-dimensional data, such as DWT and GLCM features, are involved. Here's a novel feature selection method for DWT and GLCM feature extraction

Hybrid Feature Selection Method combining DWT and GLCM:

Feature Extraction:

Apply DWT to the image to obtain DWT coefficients representing different frequency components.

Compute GLCM features using the DWT coefficients. This means constructing GLCM matrices based on the DWT coefficients.

Normalization:

Normalize the DWT coefficients and GLCM features to ensure that they are on a comparable scale. Normalization is important for methods that rely on comparing features with different scales.

Ranking Features:

Use a ranking method to assign scores to each feature based on their importance. One possible approach is to use statistical measures like Information Gain, Mutual Information, or statistical tests (e.g., t-test, ANOVA) to rank the features.

Genetic Algorithm (GA) Optimization:

Incorporate a genetic algorithm to optimize the selected feature subset. Genetic algorithms can efficiently explore a large search space of feature subsets to find an optimal or near-optimal solution. The genetic algorithm can be guided by the classification accuracy obtained from the wrapper-based method [12].

VI. ALGORITHM: GENETIC ALGORITHM FEATURE EXTRACTION

INPUT:

- Population size (pop size)
- Maximum number of generations (max_generations)
- Crossover rate (crossover rate)
- Mutation rate (mutation rate)

Procedure:

1. Initialize the population:

Generate an initial population of chromosomes representing different feature sets.

2. Evaluate The Fitness:

Evaluate the fitness of each chromosome based on the selected features.

3. Repeat the following steps for each generation until convergence or reaching the maximum number of generations:

a. Selection:

Select chromosomes for reproduction based on their fitness, favoring those with higher fitness.

b. Crossover:

Apply crossover to pairs of chromosomes with a probability of crossover_rate to create offspring with combined features.

c. mutation:

Introduce small random changes to some chromosomes with a probability of mutation_rate to explore new feature combinations.

d. Evaluate the fitness of the new offspring.

e. Replacement:

Replace the old population with the new population of offspring.

4. Output the best solution found.

5. Termination:

Repeat the process for a certain number of generations or until convergence criteria is met (e.g., achieving a satisfactory fitness level).

End Algorithm

Validation:

Evaluate the performance of the selected feature subset using cross-validation or a separate validation set. This step ensures that the chosen features generalize well to new data.

Feature selection method combines the strengths of DWT and GLCM, integrating them into a cohesive framework for improved feature representation and selection.

The specific choice of ranking, filtering, and wrapper methods can be adjusted based on the characteristics of the dataset and the requirements of the classification task

Extended segmentation based fractal texture analysis (ESFTA) algorithms:

ESFTA algorithms are specifically crafted for extended segmentation-based fractal texture analysis. These algorithms commonly include stages for image segmentation, aiming to isolate regions of interest, and subsequent fractal analysis to characterize the textures present within these segmented regions. The goal is to provide a more detailed and comprehensive understanding of textures within images, allowing for applications in fields such as medical imaging, satellite image analysis, and industrial quality control. Specific ESFTA algorithms may vary in their approach to segmentation, fractal analysis, and the integration of results.

Proposed Methodology Combines Different Feature Extraction Methods And Genetic Algorithms:

The proposed methodology combines different feature extraction methods and genetic algorithms by integrating grayscale co-occurrence matrix (GCLM), color consistency vector (CCV), and discrete wavelet transform (DWT) feature extraction techniques with genetic algorithm optimization. The study aims to optimize the classification of COVID-19 CT scan images by using a hybrid feature selection method that combines DWT and GLCM features. This approach involves applying DWT to the image to obtain DWT coefficients representing different frequency components and computing GLCM features using the DWT coefficients. The features are then normalized to ensure they are on a comparable scale. Additionally, the study uses genetic algorithms to optimize the selected feature subset, allowing for efficient exploration of a large search space of feature subsets to find an optimal or near-optimal solution.

The genetic algorithm is guided by the classification accuracy obtained from the wrapper-based method.

The proposed methodology aims to improve the discriminatory power of certain criteria and increase the accuracy of machine learning models by dynamically adjusting and improving feature sets to achieve the optimal ratio of the machine learning model. This integrated approach aims to simplify the location feature and focus on information about the distribution function, ultimately enhancing the performance and interpretability of the machine learning model.

VII. RESULT AND DISCUSSION

This section evaluates the demand for COVID-19 pneumonia by visualizing results and graphs. For verification purposes, 35 points were set for people diagnosed with the new corona virus; see Chapter 2 for details. 3. Follow a fair training/testing ratio (80:20) with 80% training material and the rest testing material.

Use the ten-second fact-finding method to uncover results and verify the truth. For the final classification, the Naive Bayes classifier will be selected based on its performance. Fair comparison with existing products including Fne KNN (F-KNN) [6], Linear Support Vector Machine (L-SVM) [7], and F-Tree [8].

Model Description:

Fuzzy k-Nearest Neighbors (Fuzzy KNN) is an extension of the traditional k-Nearest Neighbors (KNN) algorithm that incorporates the concept of fuzziness. While

traditional KNN assigns a data point to the class that Fuzzy KNN assigns, based on its k-nearest neighbors', the most common membership values to each class for a given data point, indicating the degree to which the point belongs to each class [6].

Linear Support Vector Machine (Linear SVM) is a support vector machine used for classification. SVM is a class of supervised learning algorithms used for classification and regression. Linear SVM is particularly relevant to discrete data sets where the decision boundary separating different classes is a large plane in a given area [7].

FID fuzzy decision tree was developed in 1996. It is a classification system that uses the popular and effective decision tree recursive partitioning technique while combining fuzzy representation and predictive reasoning to deal with noise and ambiguous words. Many extensions have been requested, used and requested since its release, and the latest fix, FID3.4, was recently released [8].

We believe that some performance measures, including sensitivity (SEN), sensitivity (PR), specificity (SPE), area under the curve (AUC), and accuracy, should be selected to warrant the scheme. (ACC)). The mathematical form of the above measurement is given by the equation below..

$$SPE = \frac{TN}{TN + FP} \times 100\%$$

$$PR = \frac{TP}{TP + FP} \times 100\%$$

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \times 100\%$$

Table 2: Accuracy Calculated Using Different Feature Extraction Technique

Classifier	DWT	CCV	GLCM	Accuracy
Naive Bayes	√			81.78
		√		78.34
			√	82.45
	√		√	84.34
	√	√	√	85.12
F-KNN	√			74.23
		√		83.98
			√	83.12
	√	√	√	88.65
L-SVM	√			81.66
		√		74.23
			√	88.23
	√	√	√	88.12
F-Tree	√			79.23
		√		83.23
			√	86.23
	√	√	√	89.23

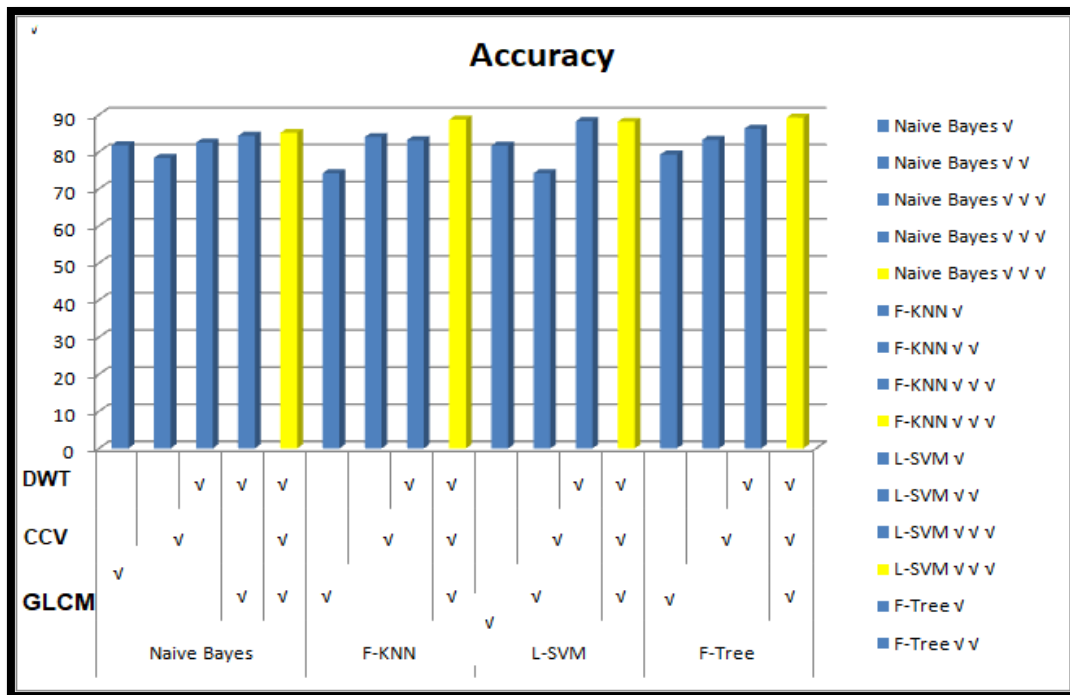


Figure 3: Accuracy bar Chart with Different Feature Extraction
 Feature Extraction Methods Based On Genetic Algorithms

Table 3: Comparison of the Accuracy of Different

Classifier	G-DWT	G-CCV	G-GLCM	Accuracy
Naive Bayes	√			78.34
		√		74.23
			√	82.23
F-KNN	√	√	√	92.06
	√			76.34
		√		81.34
L-SVM			√	83.34
	√	√	√	92.12
	√			78.22
F-Tree		√		80.23
	√	√	√	92.23
	√			84.12
		√		86.34
			√	88.23
	√	√	√	93.55

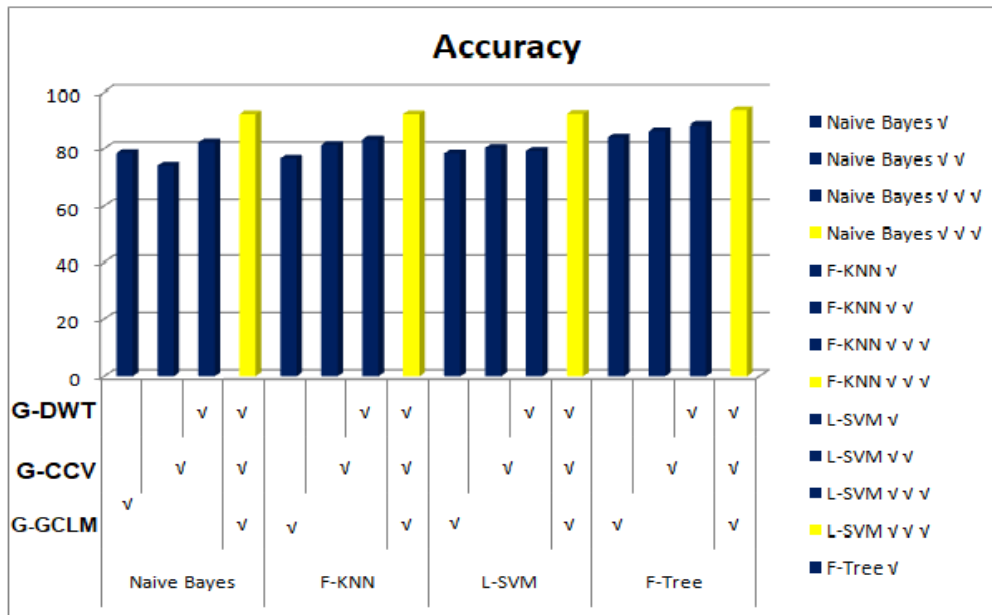


Figure 4: Accuracy Bar Chart with different Feature Extraction & GA

The results are combined to determine two different ways accuracy is achieved using individual features,

(2) Accuracy is achieved after feature selection using Genetic Algorithm

Method 1: Consider the features extracted independently from each distribution; these are listed in Table 3. For a fair comparison, different classifiers were analyzed for each method. The final results are as expected; While the Fuzzy-Tree classifier surpasses other advancement in a classifier field with an average accuracy of 89.23%, the average accuracy of the two methods is Fuzzy-KNN (88.69%). This table provides a clear comparison of various extraction methods. It is clear from the figure that the individual features are not compatible with the features of this application. In general, it can be seen that DWT features are used

Method 2, the features selected by the Genetic Algorithm are sent to the classifier (Table 4) for the final label. After the feature selection step, the same method is followed and the Fuzzy-Tree classifier can be very good for many people with an average accuracy of 92.23%. Features. L-SVM showed the second best performance with an accuracy rate of 87.23%. Again, when GA-DWT features are used, the classification results of almost all classifications are quite good. While the average classification accuracy obtained for each classification using Genetic Algorithm -DWT features is 88.86%, the average classification accuracy obtained using GA-GLCM is 88.85%. Figure shows that the average accuracy after GA increases significantly compared to independent features. The vertical line shows that the accuracy of the bid and the Genetic Algorithm choice are still comparable, which increases the importance of the combination.

Using the proposed method, an accuracy of 93.55% was achieved by Fuzzy-Tree classifiers, while some other classifiers (Naive Bayes, L-SVM, and F-KNN) achieved 92.6%, 92.12%, and 92.33%, respectively. Performed better. The accuracy of this proposal's framework is further examine by selecting the performance including sensitivity (93.12%), specificity (92.65%), precision (93.00%) and AUC (0.96), please

Table 4 See . It is clear that the sensitivity and specificity values have high positive and negative values in terms of distinguishing between really good and bad models of the framework of the concept.

Table 4: Comparison Of State-Of-The-Art Classifiers Using Ga Applications To Control Selection

Classifier	Sensitivity	Precision	Specificity	AUC	Accuracy
Naive Bayes	92.5	92.5	92.00	0.99	92.6
F-KNN	92.12	92.01	92.03	0.95	92.12
L-SVM	92.34	93.34	93.78	.98	92.23
F-Tree	93.12	93.00	93.65	.96	93.55

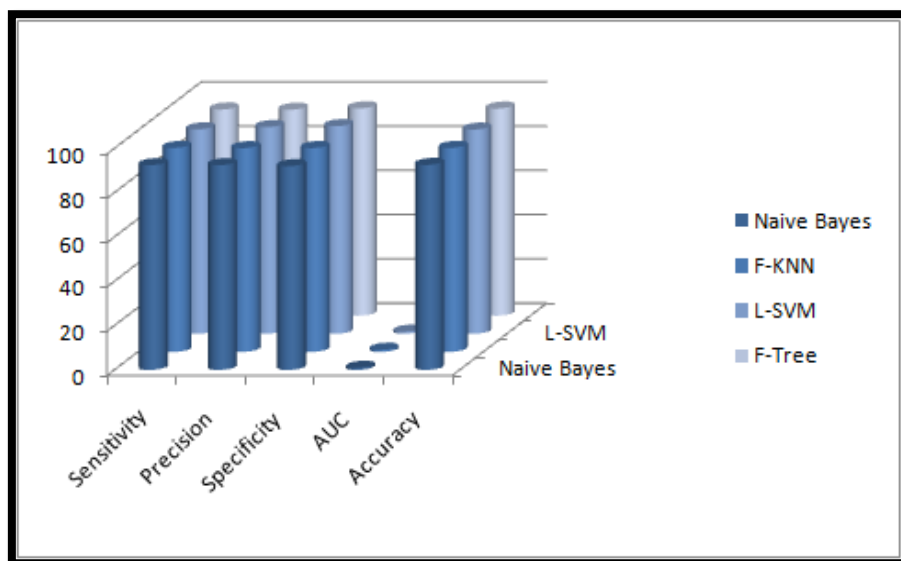


Figure 5: Comprison of Different Factor of Model

VIII. CONCLUSION

The results obtained after testing and evaluation show that the optimization of GCLM, CCV and DWT features combined with genetic features leads to the best results, and certain settings are effective in increasing the discrimination power. Genetic algorithms help select the most important features and improve classification accuracy. Among the machine learning models tested, L-SVM, Random Forest, and F-KNN showed a wide range of performance across datasets. L-SVM performs well when data is not separated and uses optimal methods to improve classification capabilities. Random forests and their joint learning have proven to be powerful and effective in managing relationships in unique environments. F-KNN is a competitive approximate method where local patterns play an important role. In conclusion, this study shows that the combination of GCLM, CCV and DWT feature extraction methods with genetic algorithm optimization can be a reliable machine learning model for proportion. The choice of the most efficient model (L-SVM, F-Random Forest or –F-KNN) may depend on specific characteristics of the dataset, highlighting the importance of choosing the appropriate algorithm based on the nature of the data. Further research and validation of different data are recommended to expand the findings and ensure the stability of the theory in various aspects.

References

- 1) Asmaa Abbas, Mohammed M Abdelsamea, and Mohamed Medhat Gaber. Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network. arXiv preprint arXiv:2003.13815, 2020.
- 2) Parnian Afshar, Shahin Heidarian, Farnoosh Naderkhani, Anastasia Oikonomou, Konstantinos N Plataniotis, and Arash Mohammadi. Covidcaps: A capsule network-based framework for identification of covid-19 cases from x-ray images. arXiv preprint arXiv:2004.02696, 2020.
- 3) P. Lambin, R. T. Leijenaar, T. M. Deist, and et al., Radiomics: the bridge between medical imaging and personalized medicine,” Nature Reviews Clinical Oncology, vol. 14, pp. 749–762, 2017.
- 4) Panday, Aishwarza & Kabir, Ashad & Chowdhury, Nihad. (2021). A Survey of Machine Learning Techniques for Detecting and Diagnosing COVID-19 from Imaging.
- 5) Wang, L., Lin, Z.Q. & Wong, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci Rep* **10**, 19549 (2020). <https://doi.org/10.1038/s41598-020-76550-z>
- 6) Xu Y, Zhu Q, Fan Z, Qiu M, Chen Y, Liu H (2013) Coarse to fine k nearest neighbor classifier. *Pattern Recognit Lett* 34(9):980–986.
- 7) Jia-Zhi D, Wei-Gang L, Xiao-He W, Jun-Yu D, Wang-Meng Z (2018) L-SVM: a radius-margin-based svm algorithm with logdet regularization. *Expert Syst Appl* 102:113–125.
- 8) Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern* 21(3):660–674.
- 9) Albahli, S. (2020). Efficient gan-based chest radiographs (CXR) augmentation to diagnose coronavirus disease pneumonia. *International Journal of Medical Sciences*, 17(10):1439–1448
- 10) Albahri, A., Hamid, R. A., k. Alwan, J., Al-qays, Z. T., Zaidan, A. A., Zaidan, B., Albahri, O., Alamoodi, A., Khlaf, J. M., Almahdi, E., Thabet, E., Hadi, S., Mohammed, K. I., Alsalem, M. A., Al-Obaidi, J. R., and Madhloom, H. T. (2020a). Role of biological Data Mining and Machine Learning Techniques in Detecting and
- 11) Asif, S., Wenhui, Y., Jin, H., Tao, Y., and Jinhai, S. (2020). Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Networks. *medRxiv:10.1101/2020.05.01.20088211*.
- 12) Prajapati, Yogendra Narayan, and Manish Sharma. "Analysis and Application of a Novel Model to Predict COVID-19 Virus's Impact on Human Heart Disease."
- 13) Yogendra Narayan Prajapati, U. Sesadri, T. R., M. ., Shreyanth S., Ashish Oberoi, & Khel Prakash Jayant. (2022). Machine Learning Algorithms in Big Data Analytics for Social Media Data Based Sentimental Analysis. *International Journal of Intelligent Systems and Applications in*
- 14) Navid Ghassemi, Afshin Shoeibi, Marjane Khodatars, Jonathan Heras, Alireza Rahimi, Assef Zare, Yu-Dong Zhang, Ram Bilas Pachori, J. Manuel Gorriz, <https://doi.org/10.1016/j.asoc.2023.110511>.
- 15) Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR Yicheng Fang, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang, Wenbin Ji, <https://doi.org/10.1148/radiol.2020200432>
- 16) Deepraj Chowdhury, Anik Das, Ajoy Dey, Soham Banerjee, Muhammed Golec, Dimitrios Kollias, Mohit Kumar, Guneet Kaur, Rupinder Kaur, Rajesh Chand Arya, Gurleen Wander, Praneet Wander, Gurpreet Singh Wander, Ajith Kumar Parlikad, Sukhpal Singh Gill, Steve Uhlig,
- 17) Prajapati, Y.N., Sharma, M. (2024). Novel Machine Learning Algorithms for Predicting COVID-19 Clinical Outcomes with Gender Analysis. In: Garg, D., Rodrigues, J.J.P.C., Gupta, S.K., Cheng, X., Sarao, P., Patel, G.S. (eds) *Advanced Computing. IACC 2023. Communications in Computer and Information Science*, vol 2054.