Chapter 3
Classification of COVID-19 CT Scan Images Using Novel Tolerance Rough Set Approach

S. Nivetha and H. Hannah Inbarani

3.1 Introduction

As of date, Corona Virus Disease-19 (COVID-19) confirmed cases are 180,372,985 all over the world, with a mortality rate of 3,907,656 and a recovered rate of 165,098,641. The pandemic of 2019–2020 is a global public health emergency. The coronavirus affects not only the respiratory system but also other vital organs including the kidneys and liver [1]. Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV2) causes COVID-19, a coronavirus disease [2]. The Corona Virus Disease 2019 (COVID-19) is extremely infectious and can cause serious respiratory distress, pneumonia, multiple organ failure, and death. Symptoms can range from a common cold to fever, cough, shortness of breath, and acute respiratory problems, depending on the type of coronavirus. Handwashing periodically, wearing a mask, social distancing, and avoiding close contact with infected people are all general guidelines for preventing the spread of this coronavirus [3]. The COVID-19 virus has a 1-week incubation cycle, according to experts [4]. This is important because the infected patient serves as a virus carrier and unwittingly transmits the virus during this time. COVID-19 can be diagnosed using three different methods: blood tests, X-rays, and Computed Tomography (CT) scans [5, 6]. The use of CT scans and X-ray images of the chest has a very strong ability to diagnose the disease in the absence of common symptoms such as fever [7]. The first step is to identify the disease's symptoms and use distinct signs to correctly diagnose the coronavirus. To diagnose COVID-19, a medical professional will ask whether the patient has been

S. Nivetha (✉) · H. Hannah Inbarani
Department of Computer Science, Periyar University, Salem, Tamil Nadu - 636011, India
e-mail: nivethas@periyaruniversity.ac.in

F. Al-Turjman, A. Nayyar (eds.), *Machine Learning for Critical Internet of Medical Things*, https://doi.org/10.1007/978-3-030-80928-7_3

in touch if he/she wears the mask and whether he/she follows the distance and hygiene laws. The second step is to examine the symptoms of the patient. In the third step, the Reverse Transcription Polymerase Chain Reaction (RT-PCR) test is used. The fourth step is to use the radiological imaging method. The Nasopharyngeal Swab, which includes exposing a swab to paper strips that causes negative antibodies intended to bind to coronavirus antigens, is the most popular diagnostic method. Antigens bind to the strips, resulting in a visible result. The procedure is quite quick and is used at the point of care. The sensitivity of the nucleic acid test is modest, ranging from 60 to 71%. Laboratory tests use samples from the nasopharyngeal swab, throat swabs, sputum, and deep airway material. However, radiological procedures may have a higher sensitivity than lab tests. Chest Radiography (Chest X-Ray) is used in the early stages. CT and ultrasonography are used when chest radiography is inadequate. Finally, the results of the blood analysis are investigated. From these results, the medical professional determines if the patient is affected by COVID-19 or not. CT imaging is currently clinically adopted as the major approach to confirm positive and suspected positive cases of COVID-19. When the patients were admitted to the hospital, 86.2% of the CT scans revealed abnormal symptoms. Ground-Glass Opacity (56.4%) and bilateral patchy shadows are the most common symptoms of COVID-19 patients as seen on CT images (51.8%).

CT images were found to have a wide range of COVID-specific lung infection patterns, including the presence of ground-glass opacities, mixed ground-glass opacities, or consolidation; the presence of an air bronchogram, interlobular septal thickening, or cavitation; the presence of a different number of lobes afflicted by ground-glass or consolidative opacities; and the presence of pleural effusion; thoracic lymphadenopathy; underlying lung illness, such as tuberculosis, emphysema, or interstitial lung disease; and various opacity distribution patterns, such as peripheral, central, bilateral, focal, multi-lobar, and diffuse. CT is a more accurate imaging method for the chest, with greater sensitivity and reliability than Chest X-Rays (CXR) [8]. According to the National Health Commission of China, chest CT may be used to detect nCoV infection. CT is a non-invasive medical imaging technology that was preferred because it is recognized as a powerful method for advanced internal porosity detection and characterization [9]. A chest CT scan may provide an abundance of pathological facts. As a result of its high sensitivity, chest CT has been used as an alternative method to detect nCoV infection [10]. When compared to first RT-PCR from pharyngeal swab samples, chest CT has a better sensitivity for COVID-19 diagnosis.

The main goal of this study is to use CT scan images to identify COVID-19-infected patient. For data reduction based on attribute dependence, a rough set is a suitable strategy. Rough Set Theory has proven to be a useful method for solving a variety of problems, including representing unknown or imprecise knowledge, knowledge analysis, evaluating the quality and availability of information in terms of accuracy, identifying and evaluating data dependence, and reasoning based on uncertainty. The inability of Rough Set Theory (RST) to deal with real-valued data is its major flaw. The number of rough set applications used today is much broader than in the past, with applications mostly in medicine, database attribute analysis, and process control. The tolerance rough set model can work effectively with

real-valued (and crisp) data, resulting in minimal information loss [11]. The Novel Tolerance Rough Set Classification approach is used to incorporate a measure of feature value similarity and determine lower and upper approximations based on these measures of similarity. Tolerance rough sets are described by certain lower and upper approximations [12]. The efficacy of the proposed approach is compared with Decision Tree Classifier (DTC), Random Forest Classifier (RFC), Naive Bayes Classifier (NBC), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). In this paper, we present a Novel Tolerance Rough Set Classification (NTRSC) approach for handling COVID-19 medical diagnosis system.

3.1.1 Research Objectives and Contributions

Several papers are presented, with an emphasis on coronavirus detection using machine learning and deep learning techniques. The applications such as medical imaging diagnosis, disease tracking, protein structure prediction, drug discovery, and virus infectivity use machine learning algorithms and deep learning architectures. The main objective of this work is to classify COVID and NON-COVID images using Novel Tolerance Rough Set Classification approach. It includes four main tasks such as preprocessing, feature extraction, segmentation, and classification. Initially, the COVID images are denoised using median filter. Then, Gray-Level Co-occurrence Matrix (GLCM) features are generated for the dimensions 0°, 45°, 90°, and 135°. Segmentation of the COVID images is done by using Otsu thresholding. For COVID-19 and NON-COVID images, segmented feature values are extracted from the datasets which are continuous values. To use rough sets, a discretization step must be performed first, which sometimes results in data loss. Therefore, Novel Tolerance Rough Set Classification is implemented for COVID and NON-COVID CT scan image identification to overcome this disadvantage, which improves the diagnosis system's performance. The tolerance similarity measure is used to find lower and upper approximation-based similarity values. The proposed approach is evaluated by comparing it to existing algorithms such as Decision Tree Classifier, Random Forest Classifier, Naive Bayes Classifier, K-Nearest Neighbor, and Support Vector Machine. The overall classification accuracy of the proposed NTRSC approach is 95%, 88%, 96%, and 93% for the GLCM 0°, GLCM 45°, GLCM 90°, and GLCM 135° datasets, respectively, based on the experimental results.

3.1.2 Research Motivation

Nowadays, many kinds of research were done for the pandemic of the new coronavirus (COVID-19), which is dangerous and threatening to people's lives. This paper contributed a Novel Tolerance Rough Set Classification (NTRSC) approach for the classification of COVID-19 CT scan images. There are four phases to the process

1. In the first phase, a denoising operation is performed to remove artifacts in the images.
2. In the second phase, Otsu thresholding-based segmentation is applied to the images.
3. In the third phase, Gray-Level Co-occurrence Matrix (GLCM) is used to extract relevant features from the COVID-19 CT scan dataset.
4. In the fourth phase, Novel Tolerance Rough Set Classification is applied.

Then, the effectiveness of various classification algorithms is assessed using appropriate classification scales.

3.2 Related Work

G.D. Rubin et al. [13] discussed 14 main questions based on the expected value of the information that thoracic imaging will be expected to provide, which corresponded to 11 decision points within the three scenarios and three additional clinical situations. The findings were compiled into five major and three supplementary guidelines to assist medical professionals in the use of chest radiography and CT in the treatment of COVID-19. Zhao et al. [14] proposed COVID-mechanistic COVID-19 diagnosis models using a CT dataset with 349 positive COVID-19 CT images from 216 patients. The authors tested the dataset's utility for developing COVID-19 diagnosis models in experimental studies. In this work, diagnosis methods are built based on multitask learning and self-supervised learning with an F1 of 0.90, an area under the Receiver Operating Characteristic (ROC) curve (AUC) of 0.98, and a precision of 0.89 using a CT dataset. Shan et al. [15] proposed that VB.Net is a deep learning-based system that uses chest CT to automatically partition all lung and infection locations. CT. Zhao et al. [16] discussed the relationship between chest CT images and pneumonia. The results have shown that COVID-19 pneumonia patients have imaging features that can assist with the early detection of highly suspected cases as well as determining the seriousness and severity of the disease. Wang et al. [17] developed the algorithm called modified inception transfer learning model, which is followed by internal and external validation. The internal validation was 89.5% accurate overall, with a precision of 0.88 and a sensitivity of 0.87. The external research dataset revealed a total accuracy of 79.3%, with a precision of 0.83 and a sensitivity of 0.67. Gozes et al. [18] proposed and demonstrated that deep learning-based automated CT image analysis tools for COVID-19 identification, quantification, and monitoring can distinguish patients from healthy people. They achieved area under the ROC curve (AUC) was 0.996, the sensitivity was 98.2%, and the specificity was 92.2%. Hasan et al. [19] proposed a promising technique for using Convolutional Neural Network (CNN) to predict COVID-19 patients from a CT scan. The modified CNN architecture in the current state to detect COVID-19 is based on DenseNet. With a 95% recall rate, the findings outperformed 92% accuracy. To diagnose COVID-19 automatically, Maghdid et al. [20] used a

deep learning approach and transfer learning strategies. The structure is a hybrid of CNN and an enhanced AlexNet structure. On the X-rays and CT slice datasets, enhanced architecture accuracy hits 94.10%. Pathak et al. [21] presented Deep Transfer Learning (DTL) which is utilized to train the COVID-19 classification model. To avoid overfitting, tenfold cross-validation was used. The dataset's training and research ratios were set to 60% and 40%, respectively. A total of 10% of the training data was used for validation purposes, out of a total of 60%. The presented method obtains 96.2264% and 93.018% training and testing accuracy, respectively. Shaban et al. [22] proposed Enhanced K-Nearest Nneighbor (EKNN) which avoids the trapping issue of conventional KNN by using solid heuristics to choose the tested item's neighbors. EKNN selects only the qualified neighbors for classification after calculating the degree of both closeness and intensity of each neighbor of the tested object. EKNN will reliably detect infected patients with the least amount of time penalty. Dilbag Singh et al. [23] studied the deep learning model for classification of CT images based on Multi-Objective Differential Evolution (MODE) CNN. The proposed model is compared to models such as CNN, Adaptive Network-Based Fuzzy Inference System (ANFIS), and Artificial Neural Network (ANN). The proposed model outperforms competitive models in terms of accuracy, F-measure, sensitivity, specificity, and kappa statistics by 1.978%, 2.0928%, 1.8262%, 1.6827%, and 1.9276% respectively, in terms of ANN, ANFIS, and CNN models. Eduardo et al. [24] discussed artificially intelligent techniques that can determine if a person is infected with Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) by analyzing CT scans. They proposed eXplainable Deep Learning (xDNN) is a non-iterative method that relies solely on recursive calculations and prototypes. As a result, it is incredibly computationally efficient. xDNN produces 97.38%, 9.16%, 95.53%, 97.13%, and 97.36% in terms of accuracy, precision, recall, F1-Score, and AUC, respectively. xDNN outperforms the other deep learning models such as Residual neural Network (ResNet), GoogLeNet, Visual Geometry Group-16 (VGG-16), AlexNet, decision tree, and Adaptive Boosting (AdaBoost). Shaoping Hu et al. [25] proposed the weakly supervised deep learning framework for fully automatic identification and classification. This framework can help determine the specific location of COVID-19-induced lesions or inflammations. They can learn to detect and localize tumors on COVID-19 and Community-Acquired Pneumonia (CAP) and Non-Pneumonia (NP) CT images based just on image-level labels. The suggested framework provides good accuracy, precision, and AUC for classification, as well as promising qualitative visualization for lesion detections, according to experimental findings. Aayush Jaiswal et al. [26] proposed DenseNet201 to categorize patients as COVID-infected or not. The proposed model extracts feature from the ImageNet dataset using its learned weights and CNN. When compared to certain well-known deep transfer learning models such as Inception-ResNetV2, VGG16, and ResNet152V2, the DenseNet201-based CNN performs much better. The proposed system diagnoses chest CT scans as having 99.82%, 96.25%, and 97.4% training, testing, and validation accuracy, respectively. Siqi Liu et al. [27] discussed that COVID-19-related tomographic patterns on chest CTs from negative cases were generated using a Generative Adversarial

Network (GAN) model. The 2D network's Dice Similarity Coefficient (DSC) was increased from 0.623 to 0.645, while the 3D network's DSC increased from 0.657 to 0.706, which is equivalent to the inter-user variability DSC (0.7132 ± 0.1831). For the 2D network, Pearson's Correlation Coefficient (PCC) was improved from 0.908 to 0.939, and for the 3D network, it improved from 0.933 to 0.961, which is comparable to the inter-user variability range PCC = 0.957. Similarly, for the 2D network, the PCC for Percentage of High Opacity (PHO) improved from 0.906 to 0.927, and for the 3D network, it improved from 0.9099 to 0.9387. In addition to the research mentioned above, the following studies are described in detail. Table 3.1 presents the data of additional studies. When the table is reviewed, the methodology utilized in the studies, as well as the size and type of datasets, is shown.

Table 3.1 Summary of related work

References	Dataset	Technique
G.D. Rubin et al. [13]	Risk factors such as community conditions and resource constraints are represented in three different scenarios	To provide guidelines to physicians on the use of thoracic imaging in a range of healthcare environments
Zhao et al. [14]	https://github.com/UCSD-AI4H/COVID-CT	For binary classification of COVID-19 or NON-COVID-19, the DenseNet169 model was utilized
Shan et al. [15]	For validation, 300 CT images from 300 COVID-19 patients (from Shanghai) were gathered. For training, 249 CT images of 249 COVID-19 patients were gathered from other centers (outside Shanghai)	DL-based segmentation network: VB.Net
Zhao et al. [16]	In Hunan, China, data on 101 cases of COVID-19 pneumonia were collected retrospectively from four institutions	Early screening and tracking the diseases
Wang et al. [17]	Collected CT images from 259 patients, with 180 cases of standard viral pneumonia and the remaining 79 cases from three hospitals with reported SARS-CoV-2 nucleic acid testing	A deep learning-based prediction model
Gozes O et al. [18]	6150 CT images of the lungs with anomalies and lung masks	Deep learning-based automated CT image analysis tool
Hasan et al. [19]	2482 CT images in total while 1252 CT images were COVID-19 positive, and 1230 CT images were from noninfected COVID-19 images but who have other pulmonary diseases	Deep learning architecture DenseNet121 for image classification
Maghdid et al. [20]	5 different sources to form a dataset of 170 X-ray images and 361 CT images of COVID-19	An effective CNN model together with testing pre-trained AlexNet for the detection of COVID-19 images

(continued)

Table 3.1 (continued)

References	Dataset	Technique
Pathak et al. [21]	413 COVID-19 images, 439 – Normal or pneumonia images	COVID-19-infected patients are classified using a deep transfer learning algorithm
Shaban et al. [22]	216 COVID-19 images and 133 NON-COVID images	Enhanced KNN classification technique
Dilbag Singh et al. [23]	From January 21 to February 3, 2020, 73 individuals with COVID-19 were proactively collected in six hospitals	Multi-objective differential evolution-based convolutional neural networks
Eduardo Soares et al. [24]	2482 CT scans, of which 1252 corresponds to 60 patients identified with SARS-CoV-2 and 1230 CT scans corresponds to 60 patients not identified with SARS-CoV-2. These data have been collected from different hospitals in Sao Paulo, Brazil	eXplainable deep learning (xDNN) approach
Shaoping Hu et al. [25]	150 3D volumetric chest CT exams of COVID-19, CAP, and NP patients, respectively. In total, 450 patient scans acquired from two participating hospitals between September 2016 and March 2020 were included for further analysis	Weakly supervised deep learning method
Aayush Jaiswal et al. [26]	Dataset collected from Kaggle website. The dataset consists of a total of 2492 CT scans out of which 1262 are positive for SARS-CoV-2 infection, i.e., COVID-19 (+), and the rest of 1230 are negative for SARS-CoV-2 infection, i.e., COVID-19 (−)	DenseNet201-based Ddeep Transfer Learning (DTL)
Siqi Liu et al. [27]	Collected 2143 chest CTs, containing 327 COVID-19 positive cases, acquired from 12 sites across seven countries	Generative Adversarial Network (GAN)

3.3 Methods and Materials

Dataset is used in this work is available at https://github.com/UCSD-AI4H/COVID-CT [42] from the github repository: in this repository, 349 positive COVID CT images and 397 NON-COVID. The images in this COVID-19 CT Scan dataset have varying sizes : 153,491,and 1853 are the minimum, average and greatest heights respectively. These images were taken from 216 different patient instances. medRxiv and bioRxiv have been used to collect the positive images. To maintain uniform properties in our final dataset for studies, all images were transformed to Portable Network Graphics (.png) format. Both positive and negative class photographs were also downsized to 256*256.

3.3.1 Preprocessing

Preprocessing is a vital step in the automated disease detection process. Before computational processing, images are usually preprocessed by eliminating low-frequency background noise, normalizing the intensity of individual particle images, and removing or enhancing data images. Preprocessing includes conversion, image resizing, noise removal, and quality enhancement. Filtering is a technique for improving an image in which filters are primarily used to remove either the image's high frequencies, i.e., smoothing the image, or the image's low frequencies, i.e., enhancing or detecting edges [28]. The median filter resides on a rectangular region. During the filtering process, it alters the size of images based on the conditions. The median value in the three-by-three neighborhood around the corresponding pixel in the input images is stored in each output pixel. The filter's output is a single value that replaces the current pixel, value at (x, y) the time when S is oriented. The applied preprocessing technique not only saves time and also compares with various filtering techniques to determine the best pixel result using the median filter. Figures 3.1 and 3.2 describe the original COVID and NON-COVID images and different filtering techniques images.

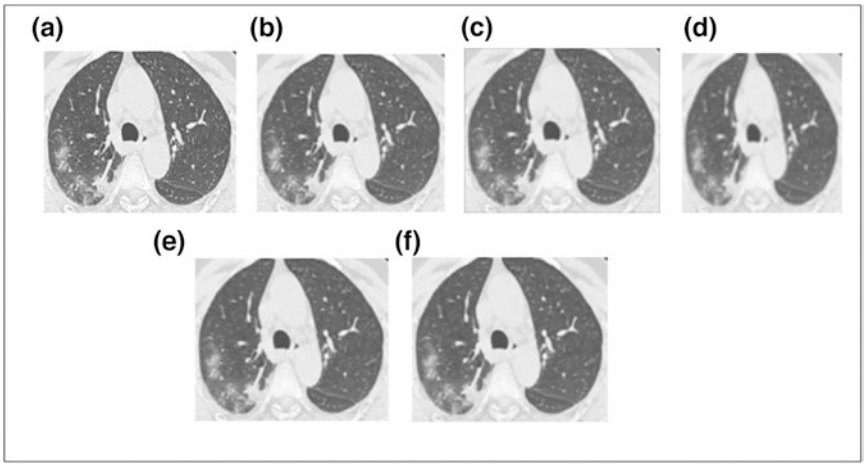

Fig. 3.1 Filtering using Different Techniques: (**a**) Input COVID Image, (**b**) Add Gaussian Noise to Image, (**c**) Gaussian Filter (**d**), Average Filter, (**e**) Median Filter, (**f**) Bilateral Filter

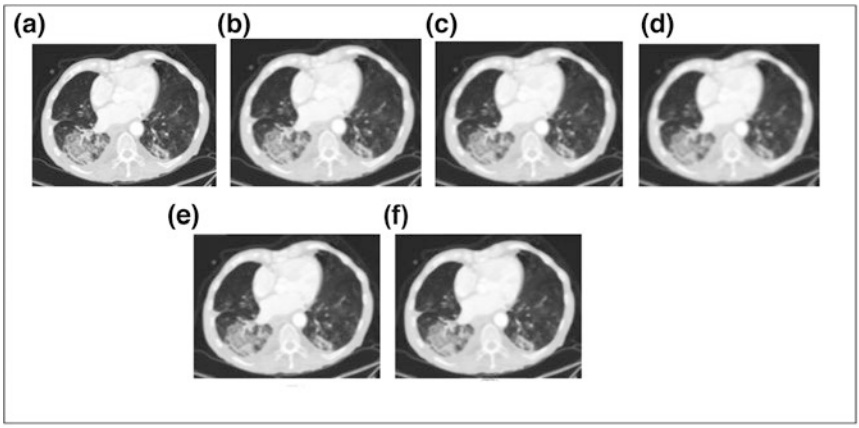

Fig. 3.2 Filtering using Different Techniques: (**a**) Input NON-COVID image, (**b**) Add Gaussian Noise to Image, (**c**) Gaussian Filter (**d**), Average Filter (**e**), Median Filter, (**f**) Bilateral Filter

3.3.2 Segmentation

The Otsu (Otsu) is a global adaptive binarization threshold image segmentation algorithm developed by Japanese researchers in 1979 [29]. As the threshold selection rule, this algorithm uses the maximum inter-class variance between the context and target image. The maximum between-class variance method is an alias of the Otsu method based on the same theory. According to the grayscale characteristics of the image, it divides it into foreground and background. The disparity between the two parts is the greatest when the better threshold is used. The maximum inter-class variance, which is a fairly common measure norm, is used by the Otsu algorithm. Otsu thresholding is used for the images in this study, and it is used to perform automatic thresholding in image processing. After segmenting the portion of an image, apply erosion operation to the images. The erosion process increases the number of pixels with a value of zero (background) while decreasing the number of pixels with a value of one (foreground). Figure 3.3 and Fig. 3.4 depict Otsu's thresholding-based segmentation applied on CT COVID and NON-COVID images.

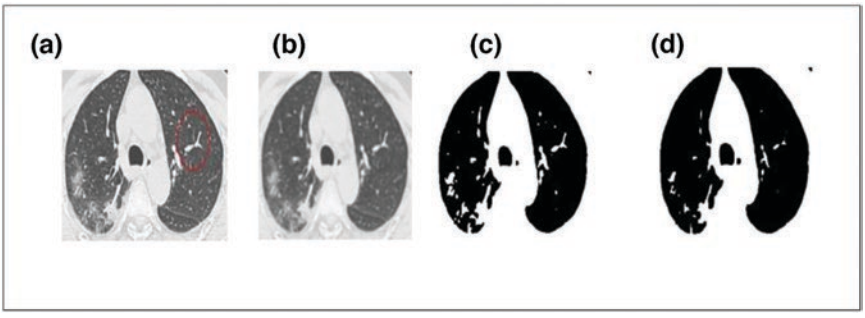

Fig. 3.3 Input and Output of CT COVID Images: (**a**) Original CT image of COVID Image, (**b**) Output of Median Filter for CT COVID image, (**c**) Apply Otsu's Thresholding for CT COVID Image, (**d**) The Output of Morphological Operation – Erosion of CT COVID

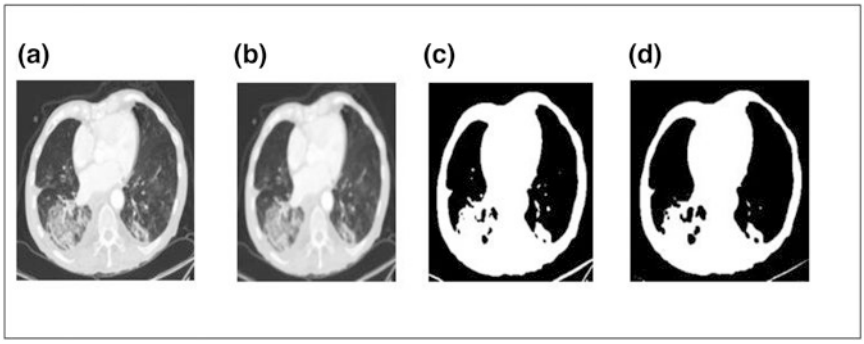

Fig. 3.4 Input and output of CT NON-COVID image: (**a**) original CT image of NON-COVID image, (**b**) output of median filter for CT NON-COVID image, (**c**) Otsu's thresholding applied to CT NON-COVID image, (**d**) output of morphological operation – erosion of CT NON-COVID image

3.3.3 Feature Extraction

A prominent texture-based feature extraction approach is the Gray-Level Co-occurrence Matrix (GLCM). The GLCM defines the textural relationship between pixels by executing an operation on the images second-order statistics. A matrix with the same number of rows and columns as the gray values is used to represent the GLCM properties of an image. For GLCM, two pixels are usually used [30]. The GLCM features used in this study are as follows: Contrast, Dissimilarity, Homogeneity, Energy, Correlation, and Angular Second Moment (ASM). GLCM is constructed in four spatial dimensions which are 0°, 45°, 90°, and 135°. For all segmented COVID and NON-COVID images, Gray-Level Co-occurrence Matrix is applied to extract the features in 0°, 45°, 90°, and 135° directions.

3.4 Background Study of Tolerance Rough Set

Pawlak introduced a new mathematical method called Rough Set Theory [31, 32] in the early 1980s to deal with vagueness and ambiguity in datasets. Many extensions of the rough set model have been suggested in terms of different criteria over the last 40 years, including the rough set model based on a soft rough set model, the rough soft set model, the fuzzy soft set model, tolerance relations, and rough fuzzy model [33]. The equivalence relation is used to identify Pawlak's rough set-based classification algorithms, which are only suitable for discrete datasets. If the attribute values of objects in typical rough sets are similar, they are grouped into equivalence groups. For continuous data, where values can only vary due to noise, this criterion

may be too stringent. When using the Pawlak model to handle continuous data, the computation cost keeps increasing. Several extensions of rough set theory have been developed to replace analogous relations, dimensionality reduction, and classification systems, such as the fuzzy rough set model [34], probabilistic rough sets [35], similarity rough set [36], tolerance relation rough set [37], decision-making rough set [38], and covering rough set [12].

3.4.1 Tolerance Rough Set

Tolerance Rough Set (TRS) analyzes the indeterminate data found in the limit region of tolerance rough sets using comparability steps. This technique is used to present an estimate of highlight value comparability as well as define the lower and upper approximations using these similitude metrics. Tolerance rough sets are characterized by certain lower and upper approximations [39].

3.4.1.1 Tolerance Information Systems

If A = (U, A) is an information system and B ⊆ A, then $INF(B) = \{Inf_B(x) : x \in U\}$ is the set of information vectors $Inf_B(x) = \{(a, a(x)) : a \in B\}$. If u ϵ INF(C) and B ⊆ C ⊆ A, then $u \mid B = \{(a, w) \in u : a \in B\}$. That is, $u \mid B$ is the restriction of u to B [12]. A tolerance information system is a pair (A, D) where

A = (U, *A*) is an information system,
D = $(D_B)_{B \subseteq A}$ and $D_B \subseteq$ INF(B) × INF(B) are a relation termed as the discernibility relation satisfying the conditions [12]:

(i) INF(B) × INF(B) − D_B is a tolerance (indiscernibility) relation.
(ii) $((u - v) \cup (v - u) \subseteq (u_0 - \upsilon_0) \cup (\upsilon_0 - u_0))$ & $uD_B\upsilon \rightarrow uD_B\upsilon_0$ for any u, v, u_0, υ_0 ϵ INF(B) that is D_B is monotonic with respect to discernibility property.
(iii) $non(uD_C\upsilon)$ implies $non(u| BD_B\upsilon|$ B) for any B ⊆ C and u, v ϵ INF(C).

$A(B, D_B)$ tolerance function, $I[B, D_B] : U \rightarrow P(U)$ is defined by $y \in I[B, D_B](x)$ if $non(Inf_B(x)D_BInf_B(y))$ for any x, y ϵ U. The set $I[B, D_B](x)$ is called the tolerance set of x [12].

3.4.1.2 Definition 1: Tolerance Similarity Measures

Let $x_iT_ax_j$ epitomize the similarity between x_i and x_j in terms of the tolerance threshold attribute. x_i and x_j are identical in terms of attribute 'a,' where T_a denotes the tolerance similarity threshold relation for attribute 'a,' whose value falls within the T_a [0,1] range. As a result, we can apply the standard similarity measure $S_a(x_i, x_j)$ to the T_a that can be determined by a simple distance,

$$S_a\left(x_i,x_j\right)=1-\frac{\left|a\left(x_i\right)-a\left(x_j\right)\right|}{\max_a-\min_a} \tag{3.1}$$

where $a(x_i)$ and $a(x_j)$ are attribute values concerning x_i and x_j, respectively, and max_a and min_a are the maximum and minimum values of attribute 'a' [7]. T_a's and S_a's relationship is depicted below,

$$x_i T_a x_j \leftrightarrow S_a\left(x_i,x_j\right)\geq\tau \tag{3.2}$$

where t_a is the attribute a-based similarity threshold. The similarity measure in classification is based on the normalized distance function as follows:

$$S_a\left(x_i,x_j\right)=1-\frac{\left|d(a\left(x_i\right),a\left(x_j\right)\right|}{\left|d_{\max}\right|} \tag{3.3}$$

The maximum distance between two attribute values $a(x_i)$ and $a(x_j)$ is denoted by d_{max},

$$d\left(a\left(x_i\right),a\left(x_j\right)\right)=a\left(x_i\right)-a\left(x_j\right) \tag{3.4}$$

which is the distance function between two objects in terms of attribute values [40]. The similarity measure S_A (x_i, x_j) between two objects x_i and x_j is defined as an arithmetic average of similarity measures of all attributes between two objects x_i and x_j:

$$S_A\left(x_i,x_j\right)\frac{1}{|A|}\sum_{\forall a\epsilon A}s_a\left(x_i,x_j\right) \tag{3.5}$$

The number of attributes in A is represented by |A|. When all attributes 'A' is considered at the same time, we can apply the tolerance related to the similarity measure as follows,

$$x_i t_A x_j \leftrightarrow S_A\left(x_i,x_j\right)\geq t\left(A\right) \tag{3.6}$$

where t(A) [0,1] is an image classification similarity threshold dependent on all attributes A.

3.4.1.3 Definition 2: Tolerance Rough Set

The degree of similarity is calculated for each function in the tolerance rough set method as follows [41]:

$$SIM_a(x,y) = 1 - \frac{|a(x) - a(y)|}{a_{\max} - a_{\min}} \tag{3.7}$$

The feature or attribute is 'a,' and the maximum and minimum values for the features taken are amax and amin. Similarity can be achieved for a subset of features P as follows:

$$(x,y) \in SIM_{P,\tau} \; iff \prod_{a \epsilon P} SIM_a(x,y) \geq \tau \tag{3.8}$$

$$(x,y) \in SIM_{P,\tau} iff \frac{\sum_{a \epsilon P} SIM_a(x,y)}{|P|} \geq \tau \tag{3.9}$$

Lower PτX and upper $\overline{P\tau X}$ approximations are demarcated as follows:

$$\underline{P_\tau X} = \left\{ x \mid SIM_{P,\tau}(\underline{x}) \subseteq X \right\} \tag{3.10}$$

$$\overline{P_\tau X} = \left\{ x \mid SIM_{P,\tau}(x) \cap X \neq \varnothing \right\} \tag{3.11}$$

Positive region and negative region are as follows:

$$POS_{P,\tau}(y) = \bigcup_{(X \varepsilon U \mid y \varepsilon U)} \underline{P_\tau X} \tag{3.12}$$

$$NEG_{P,\tau}(y) = \bigcup_{(X \varepsilon U \mid y \varepsilon U)} \overline{P_\tau X} \tag{3.13}$$

3.4.2 *Proposed Approach: Novel Tolerance Rough Set Classification (NTRSC)*

The proposed approach is shown in Fig. 3.5. This proposed NTRSC approach is used to provide a measure of feature value similarity and define lower approximations using the similarity measures. In the first phase, preprocessing and Otsu segmentation are applied to the COVID and NON-COVID images. In the second phase, GLCM features for four directions are extracted. In the third phase, the NTRSC approach was applied to the dataset. A classification dataset is separated into two parts: a training set and a testing set. The efficiency of a classifier is determined by presenting it with a testing set. In this NTRSC training approach, tolerance cosine similarity is constructed in step 1. In the second step, novel tolerance lower approximation of the dataset based on decision class C is constructed. With the help of tolerance approximation, we generate the rules which are certain rules. In the third

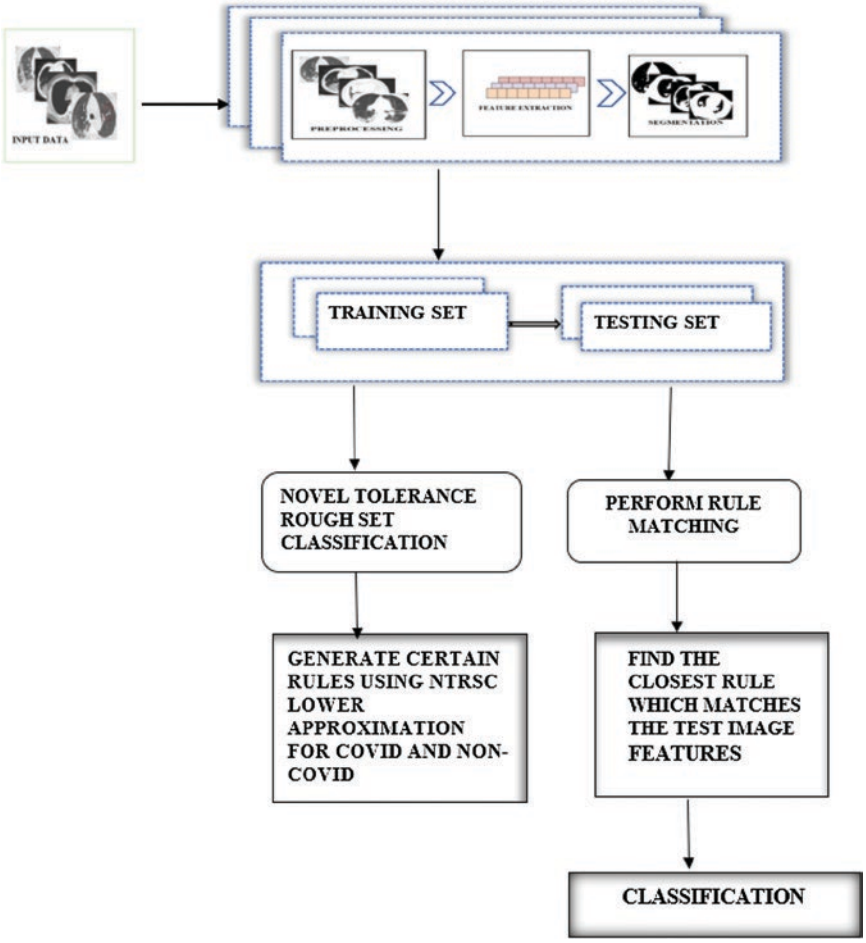

Fig. 3.5 The proposed NTRSC approach

step, certain rules are generated based on tolerance rough set lower approximation. In the testing algorithm matching, the closest decision rule is applied for the test data. Finally, classification measures are used to assess the effectiveness of different categorization procedures for COVID and NON-COVID diagnosis.

The experimental results for the proposed NTRSC and other benchmark algorithms are validated using classification validation accuracy measures. Table 3.2 shows the proposed Novel Tolerance Rough Set Classification (NTRSC) approach for the training dataset. Table 3.3 depicts the NTRSC testing algorithm. NTRSC algorithm is applied to generate certain rules for tolerance lower approximation space.

Algorithm 1: Proposed Approach – Novel Tolerance Rough Set Classification: Training Algorithm

Table 3.2 Proposed NTRSC approach for training dataset

Input*:* $<U, A \cup B>$,τ-threshold, $U = \{X_1, X_2, \ldots, X_n\}$, $A = \{a_1, a_2, \ldots, a_n\}$ is the set of conditional attributes; $D = \{d_1, d_2, \ldots, d_n\}$ is the decision attribute.
Output*:* Set of rules for each class.
Step 1: Extract the features from COVID and NON-COVID images using GLCM
Step 2: Define the tolerance relation for the conditional attributes using
$[x]_C \leftarrow C$
Step 3: Construct the equivalence relation for the decision attribute:
$[x]_D \leftarrow C$
Step 4: Compute tolerance cosine similarity for the tolerance relation:
$\mathbf{SIM}_a(x,y) = 1 - \frac{\lvert a(x) - a(y) \rvert}{a_{max} - a_{min}}$
Step 5: Construct the tolerance rough set lower approximation space for each class:
$\underline{P_\tau X} = \{x \mid SIM_{P,\tau}(x) \subseteq X\}$
Step 6: Generate certain rules for each class separately using a tolerance rough set based on lower approximation

Algorithm 2: Proposed Approach – Novel Tolerance Rough Set Classification: Testing Algorithm

Table 3.3 Proposed NTRSC approach for testing dataset

Input: Set of decision rules
Output*:* Decision values
Step 1: Extract the features for each image in the test set
Step 2: For each class, perform feature matching with decision rules
Step 3: Use the closest decision rule to classify the image
Step 4: Output decision

3.5 Experimental Results and Discussion

An experimental evaluation of the proposed Novel Tolerance Rough Set Classification for COVID and NON-COVID images is presented in this section. All the experiments were run on an Intel® Core™ i5-10210U CPU at 1.60–2.11 GHz machine with 8 GB RAM. The NTRSC approach was implemented in Python using Anaconda. Experiments are carried out using a CT scan dataset. The COVID-19 CT dataset used in this study is open to the public (https://github.com/UCSD-AI4H/COVID-CT) [42]. 349 CT images from COVID patients and 397 CT images from

Table 3.4 PSNR and SSIM values for various filters

Image	Metrics	Noisy image	Gaussian filter	Average filter	Median filter	Bilateral filter
CO	**PSNR**	31.9639	38.9170	33.1507	41.2520	39.41192
	SSIM	0.85376	0.9770	0.88605	0.98514	0.974788
N-CO	**PSNR**	34.0622	40.538	34.1514	45.7168	41.89566
	SSIM	0.93808	0.9919	0.94818	0.99558	0.988984

NON-COVID-19 patients are included in the dataset. The experiments are carried out in the following ways: GLCM is used to extract the features of the COVID and NON-COVID images. Features were extracted from both training and testing images. Then, using NTRSC, each test image is matched with a lower approximation generated rule and matches the closest decision rule for classifying the image. In comparison to other filtering methods, the median filter yields strong Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) values. The PSNR and SSIM values of COVID and NON-COVID images after noise removal using different filters are shown in Table 3.4.

A confusion matrix is a method of summarizing a classification algorithm's results. A confusion matrix helps to get the summary of prediction results on a classification problem. Count values are used to sum and break down the number of correct and wrong predictions by class. True positive refers to correctly predicted NON-COVID images that are classified correctly. True negative refers to correctly predicted COVID images that are labeled as COVID images. False positive refers to the incorrectly predicted COVID images, and false negative refers to incorrectly predicted NON-COVID images. Precision, recall, specificity, F-measure, and G-Mean are among the useful performance indicators computed in medical applications. The findings are reviewed and compared to those of other decision-making classifiers. Precision is the prediction of a positive observation, accuracy is the accurate prediction observation to the total observations, and F1-Score is the calculation of the weighted average of precision and sensitivity. G-Mean is the product of the prediction accuracies for both classes. In this paper, we show the performance of a classification algorithm using well-known metrics including Accuracy, Sensitivity, Specificity, Error Rate, Matthews Correlation Coefficient, Lift, Youden's Index, Balanced Classification Rate, and Balanced Error Rate. The complete interpretation for each metric is depicted in Table 3.5 [43–51].

Several machine learning algorithms such as Decision Tree Classifier, Random Forest Classifier, Naive Bayes Classifier, K-Nearest Neighbor, and Support Vector Machine were implemented for the classification of COVID and NON-COVID images. Performance values were evaluated via confusion matrix as shown in Table 3.6 for GLCM 0°, GLCM 45°, GLCM 90°, and GLCM 135°. According to the empirical results, the proposed NTRSC approach accurately classifies COVID and NON-COVID images. The proposed NTRSC approach produces the correct predictions as the same output results which prove that NTRSC provides the best result than other classifiers. From all the classifiers examined, GLCM 0° NTRSC has 0.95% accuracy. The proposed NTRSC achieved better classification accuracy

Table 3.5 Various evaluation metrics

Metrics	Formula
Precision	$\frac{TP}{TP+FP}$
Recall (sensitivity)	$\frac{TP}{TP+FP}$
Specificity	$\frac{TN}{TN+FP}$
Negative Predictive Value	$\frac{TN}{TN+FN}$
False Predictive Value	$\frac{FP}{FP+TN}$
False-Negative Value	$\frac{FP}{TP+FP}$
F1-score	$2.\frac{PRECISION*RECALL}{PRECISION+RECALL}$
G-Mean	$\sqrt{PRECISION*RECALL}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Error Rate	$\frac{FP+FN}{TP+TN+FP+FN}$
Matthews Correlation Coefficient	$(TP*TN)-(FP*FN)/SQRT((TP+FP*(TP+FN)*(TN+FP)*(TN+FN))$
Lift	$\frac{(TP/(TP+FP))}{((TP+FN)/(TP+TN+FP+FN)}$
Youden's Index	$SENSITIVITY+SPECIFICITY-1$
Balanced Classification Rate	$\frac{1}{2}(SENSITIVITY+SPECIFICITY)$
Balanced Error Rate	$1-BCR$

[a]*TN* True Positive, *TP* True Negative, *FP* False Positive, *FN* False Negative

than other methods, and the minimum error rate of 0.04 is represented in Table 3.7. It also proves the efficiency of the proposed NTRSC approach. The output of the decision-making algorithms for the GLCM 45° dataset in the proposed NTRSC algorithm has an 88% overall accuracy and a 0.12% error rate. The proposed decision-making algorithm earns the highest score on Youden index, i.e., 2.83. For GLCM 90° in the COVID dataset, the classification accuracy of the NTRSC is higher than that of Decision Tree, Random Forest Classifier, Naive Bayes Classifier, K-Nearest Neighbor, and Support Vector Machine. It also demonstrates that the

Table 3.6 Confusion matrix, precision, recall, F1-Score, support, and G-Mean for various classifiers

GLCM features	Classification algorithm	Desired output	Output result for confusion matrix CO	N-CO	Precision	Recall	F1-Score	Support	G-Mean
GLCM 0°	NTRSC	CO	97	7	0.96	0.93	0.95	104	0.94
		N-CO	4	116	0.94	0.97	0.95	120	0.95
	Decision Tree	CO	87	13	0.87	0.87	0.90	124	0.87
		N-CO	13	111	0.90	0.90	0.90	121	0.90
	Random Forest	CO	97	9	0.92	0.92	0.92	106	0.92
		N-CO	9	109	0.92	0.92	0.92	118	0.92
	Naive Bayes	CO	67	39	0.85	0.63	0.72	106	0.73
		N-CO	12	106	0.73	0.90	0.81	118	0.81
	KNN	CO	94	10	0.96	0.90	0.93	104	0.92
		N-CO	4	116	0.92	0.97	0.94	120	0.94
	SVM	CO	79	19	0.86	0.81	0.84	97	0.83
		N-CO	13	114	0.86	0.90	0.88	127	0.87
GLCM 45°	NTRSC	CO	98	8	0.97	0.70	0.81	97	0.82
		N-CO	10	108	0.81	0.98	0.89	127	0.89
	Decision Tree	CO	85	24	0.80	0.78	0.79	109	0.78
		N-CO	21	94	0.80	0.82	0.81	115	0.80
	Random Forest	CO	91	9	0.70	0.78	0.74	95	0.73
		N-CO	18	106	0.82	0.76	0.79	129	0.78
	Naive Bayes	CO	65	39	0.78	0.62	0.70	104	0.69
		N-CO	18	102	0.72	0.85	0.78	120	0.80
	KNN	CO	95	11	0.80	0.90	0.84	106	0.84
		N-CO	24	94	0.90	0.80	0.84	118	0.84
	SVM	CO	62	35	0.70	0.64	0.67	97	0.66
		N-CO	27	100	0.74	0.79	0.76	127	0.76
GLCM 90°	NTRSC	CO	110	9	0.99	0.94	0.96	119	0.96
		N-CO	2	103	0.94	0.99	0.96	105	0.96
	Decision Tree	CO	80	26	0.84	0.75	0.80	106	0.79
		N-CO	15	103	0.80	0.87	0.83	118	0.87
	Random Forest	CO	86	20	0.88	0.81	0.84	106	0.84
		N-CO	12	106	0.84	0.90	0.87	118	0.86
	Naive Bayes	CO	50	49	0.66	0.51	0.57	99	0.58
		N-CO	26	99	0.67	0.79	0.73	125	0.72
	KNN	CO	100	8	0.89	0.93	0.91	108	0.90
		N-CO	12	104	0.93	0.90	0.91	116	0.91
	SVM	CO	52	45	0.74	0.54	0.62	97	0.63
		N-CO	18	109	0.71	0.86	0.78	127	0.78

(continued)

Table 3.6 (continued)

GLCM features	Classification algorithm	Desired output	Output result for confusion matrix CO	N-CO	Precision	Recall	F1-Score	Support	G-Mean
GLCM 135°	NTRSC	CO	110	9	0.97	0.90	0.93	119	0.93
		N-CO	2	103	0.89	0.97	0.93	105	0.92
	Decision Tree	CO	90	14	0.82	0.87	0.84	104	0.84
		N-CO	20	100	0.88	0.83	0.85	120	0.84
	Random Forest	CO	98	8	0.89	0.92	0.91	106	0.90
		N-CO	12	106	0.93	0.90	0.91	118	0.91
	Naive Bayes	CO	72	38	0.82	0.65	0.73	110	0.73
		N-CO	16	98	0.72	0.86	0.78	114	0.78
	KNN	CO	97	4	0.88	0.96	0.92	101	0.91
		N-CO	13	110	0.96	0.89	0.93	123	0.92
	SVM	CO	68	29	0.77	0.70	0.74	97	0.73
		N-CO	20	107	0.79	0.84	0.81	127	0.81

Table 3.7 Performance metrics for various classifiers

GLCM features	Performance measures	NTRSC	DTC	RFC	NBC	KNN	SVM
GLCM 0°	Accuracy	0.95	0.88	0.92	0.77	0.94	0.86
	Sensitivity	0.95	0.87	0.91	0.84	0.95	0.85
	Specificity	0.94	0.89	0.92	0.73	0.92	0.86
	Error rate	0.04	0.11	0.08	0.22	0.06	0.13
	MCC	0.89	0.76	0.84	0.55	0.87	0.71
	NPV	0.96	0.91	0.92	0.88	0.91	0.89
	FPV	0.05	0.11	0.06	0.28	0.06	0.13
	FNV	0.04	0.09	0.08	0.13	0.09	0.14
	Lift	2.07	1.94	1.93	1.79	2.06	1.98
	Youden's index	2.90	2.76	2.83	2.57	2.87	2.72
	BCR	0.95	0.88	0.91	0.78	0.93	0.86
	BER	0.04	0.11	0.08	0.21	0.06	0.13
GLCM 45°	Accuracy	0.88	0.80	0.86	0.75	0.84	0.72
	Sensitivity	0.90	0.80	0.83	0.78	0.79	0.69
	Specificity	0.89	0.79	0.79	0.72	0.81	0.74
	Error rate	0.12	0.20	0.13	0.28	0.15	0.27
	MCC	0.73	0.62	0.70	0.44	0.69	0.43
	NPV	0.75	0.78	0.92	0.83	0.84	0.78
	FPV	0.17	0.19	0.06	0.29	0.21	0.25
	FNV	0.29	0.25	0.08	0.22	0.17	0.30
	Lift	1.91	1.64	1.87	1.64	1.68	1.60
	Youden's index	2.83	2.59	2.75	2.72	2.43	2.43
	BCR	0.91	0.79	0.87	0.86	0.71	0.71
	BER	0.08	0.20	0.12	0.13	0.15	0.28

(continued)

Table 3.7 (continued)

GLCM features	Performance measures	NTRSC	DTC	RFC	NBC	KNN	SVM
GLCM 90°	Accuracy	0.96	0.82	0.86	0.67	0.91	0.72
	Sensitivity	0.99	0.84	0.87	0.65	0.89	0.74
	Specificity	0.93	0.79	0.84	0.66	0.92	0.70
	Error rate	0.03	0.18	0.14	0.33	0.08	0.28
	MCC	0.92	0.74	0.73	0.73	0.85	0.42
	NPV	0.99	0.92	0.88	0.79	0.90	0.85
	FPV	0.06	0.20	0.13	0.30	0.05	0.29
	FNR	0.00	0.09	0.13	0.36	0.10	0.25
	Lift	1.87	1.77	1.85	1.48	1.85	1.71
	Youden's index	2.92	2.64	2.71	2.32	2.82	2.45
	BCR	0.96	0.82	0.85	0.66	0.91	0.72
	BER	0.03	0.17	0.14	0.33	0.08	0.27
GLCM 135°	Accuracy	0.93	0.88	0.91	0.76	0.92	0.78
	Sensitivity	0.97	0.81	0.89	0.81	0.88	0.77
	Specificity	0.89	0.87	0.92	0.72	0.96	0.78
	Error rate	0.06	0.15	0.08	0.24	0.07	0.21
	MCC	0.86	0.72	0.81	0.40	0.80	0.55
	NPV	0.97	0.84	0.96	0.79	0.89	0.84
	FPV	0.10	0.15	0.06	0.27	0.08	0.21
	FNR	0.02	0.18	0.10	0.25	0.11	0.22
	Lift	1.84	1.76	1.88	1.66	1.95	1.78
	Youden's index	2.86	2.69	2.82	2.53	2.84	2.55
	BCR	0.93	0.84	0.91	0.76	0.92	0.77
	BER	0.06	0.15	0.08	0.23	0.07	0.22

proposed NRSC approach is more efficient than the KNN classification algorithm. For GLCM 135°, the proposed NTRSC outperforms all the algorithms, and NTRSC produces the 0.93% accuracy with a 0.06 error rate.

The effectiveness of algorithms is computed using various validation measures. Figures 3.6 and 3.7 present the performance of various classifiers: GLCM 0°, GLCM 45°, GLCM 90°, and GLCM 135° datasets. From this figure, the proposed NTRSC approach produces the best accuracy for all datasets and minimum error rate, i.e., it outperforms the other algorithms.

A Receiver Operating Characteristic (ROC) curve is a graph that shows how well a classification model performs overall classification thresholds. Two parameters plotted on this curve are true-positive and false-positive rate. The proposed NTRSC approach outperformed the other current classification algorithms on all datasets, including GLCM 0°, GLCM 45°, GLCM 90°, and GLCM 135°. The NTRSC algorithm's curve appears in the ROC graph's top left border. This indicates that the

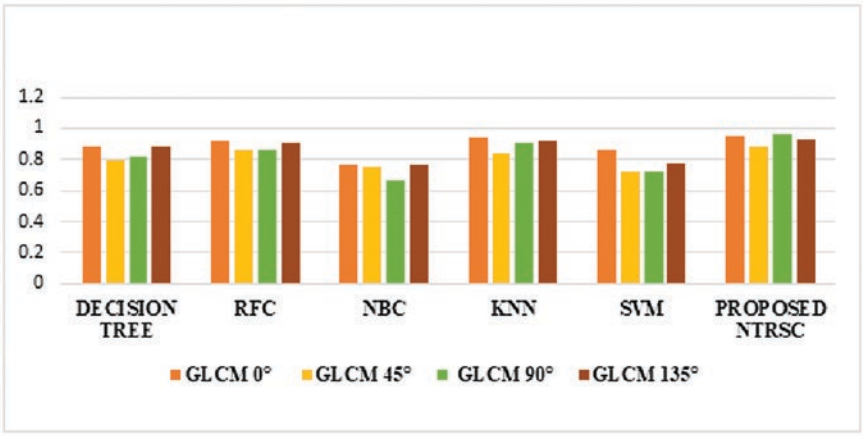

Fig. 3.6 Accuracy comparison for various classifiers

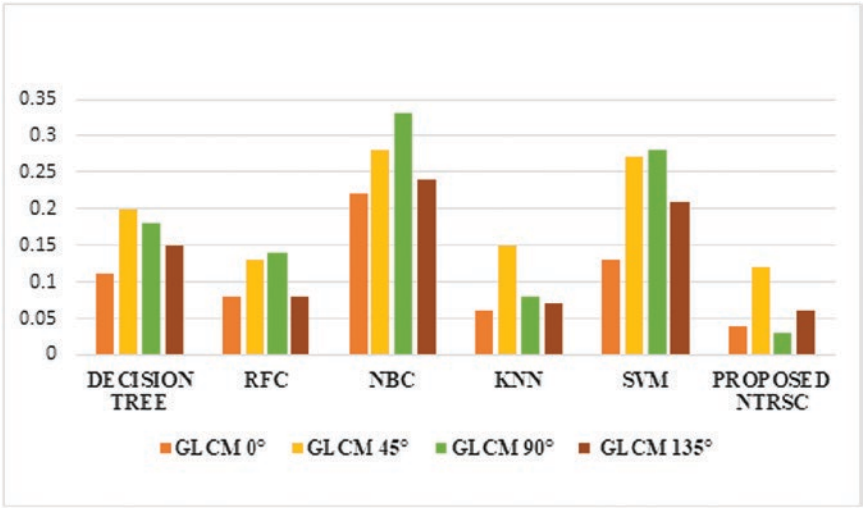

Fig. 3.7 Error rate comparison for various classifiers

proposed approach correctly differentiates between COVID and NON-COVID. The ROC curve comparison of the proposed NTRSC approach and current decision-making algorithms is shown in Figs. 3.8, 3.9, 3.10, and 3.11 for GLCM 0°, GLCM 45°, GLCM 90°, and GLCM 135° datasets.

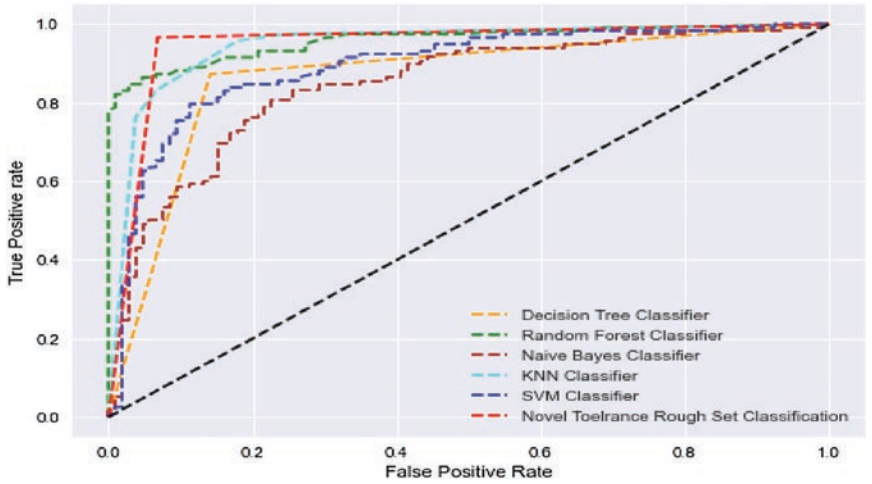

Fig. 3.8 ROC curve analysis NTRSC approach – GLCM 0°

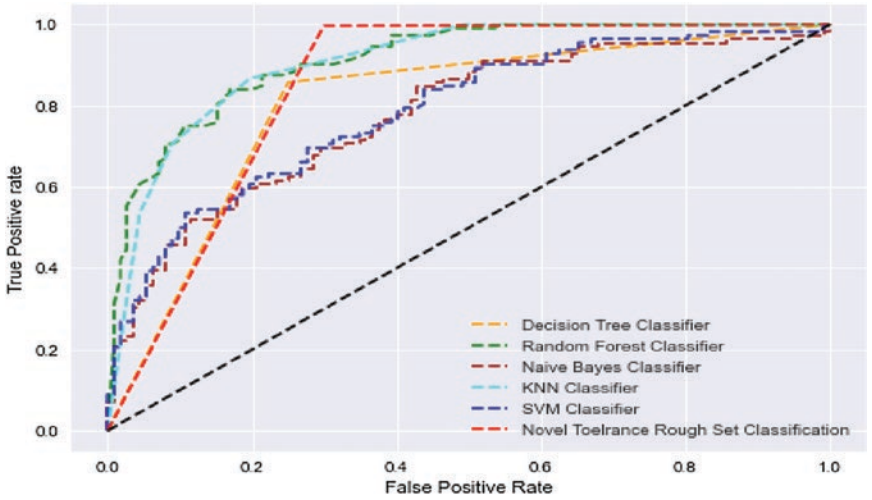

Fig. 3.9 ROC curve analysis NTRSC approach – GLCM 45°

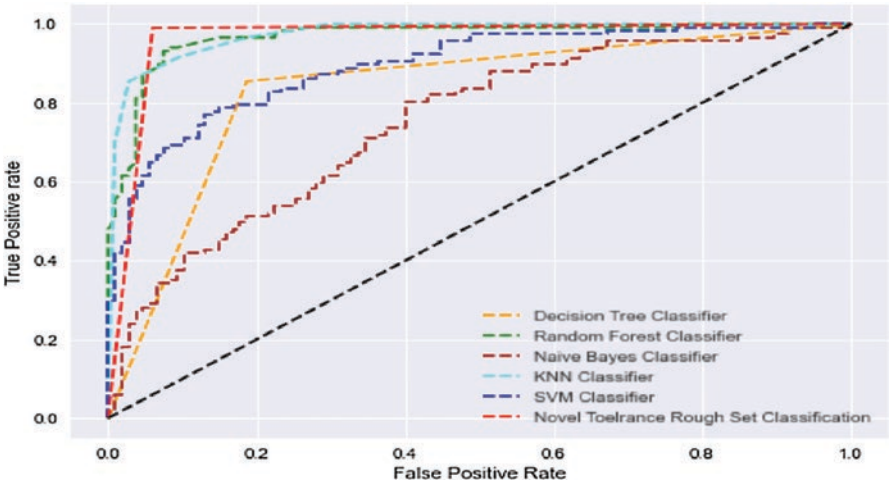

Fig. 3.10 ROC curve analysis NTRSC approach – GLCM 90°

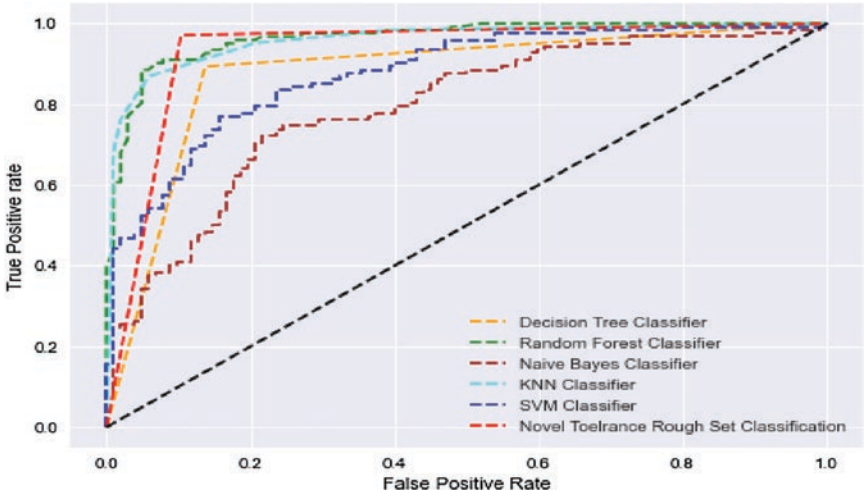

Fig. 3.11 ROC curve analysis NTRSC approach – GLCM 135°

3.6 Conclusion

Finally, in the current situation, where COVID-19 is still spreading, rapid and accurate diagnosis and disease progression analysis are critical. COVID-19 must be diagnosed early to treat and isolate individuals and prevent the virus from spreading. COVID-19 CT scan images for the GLCM 0°, GLCM 45°, GLCM 90°, and

GLCM 135° datasets were retrieved from the GitHub source, with accuracy values of 95%, 88%, 96%, and 93%, respectively. The chest CT dataset of COVID-19-infected patients and NON-COVID-19-infected patients is first decomposed into two sets: training and testing. Novel Tolerance Rough Set Classification (NTRSC) approach is applied in the training process, deciding on a collection of new information based on data acquired by lower rules and identifying the closest matches with the testing dataset. The experimental results were found to be highly compelling, and the approach was shown to be a helpful tool for COVID-19 screening on CT scan images of corona suspects. The presented approach is combined with current benchmark algorithms, and several classification measures are assessed. The obtained findings suggest that the proposed NTRSC approach has greater output accuracy with a low error rate than other algorithms. The obtained result demonstrates that the presented approach outperforms other comparative classification algorithms in terms of accuracy. Furthermore, the proposed algorithm achieves the greatest ROC. Finally, proposed Novel Tolerance Rough Set Classification approach for distinguishing COVID-19 and NON-COVID-19 images are classified accurately and helpful for diagnosis. Multiple illnesses such as pneumonia, bronchitis, and tuberculosis (TB), as well as COVID-19 of suspected persons with respiratory illness, can be detected in the future. As future work, the proposed approach might be extended to classify more lung illnesses. This can be accomplished by using multiclass classifier. For further studies, a larger dataset COVID-19 data will be collected, and deep learning architectures on the datasets will be tested.

Acknowledgments The first author immensely acknowledges the financial assistance under University Research Fellowship, Periyar University, Salem. The second author thankfully acknowledges the UGC-Special Assistance Program for the financial support for research under the UGC-SAP at the level of DRS-II (ref. no.: F.5-6/2018/DRS-II(SAP-II), 26 July 2018) in the Department of Computer Science, Periyar University, Salem, Tamil Nadu, India.

References

1. McIntosh, K., Hirsch, M. S., & Bloom, A. (2020). Coronavirus disease 2019 (COVID-19): Epidemiology, virology, and prevention. *The Lancet Infectious Diseases, 1*, 2019–2020.
2. Hannah Inbarani, H., & Nivetha, S. (2021). Prediction of COVID-19 fatality cases based on regression techniques. *European Journal of Molecular & Clinical Medicine, 7*(3), 696–719.
3. Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents, 55*(3), 105924.
4. Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., et al. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine, 172*(9), 577–582.
5. Fan, L., Li, D., Xue, H., Zhang, L., Liu, Z., Zhang, B., et al. (2020). Progress and prospect on imaging diagnosis of COVID-19. *Chinese Journal of Academic Radiology, 3*(1), 4–13.
6. Hu, Z., Song, C., Xu, C., Jin, G., Chen, Y., Xu, X., et al. (2020). Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Science China Life Sciences, 63*(5), 706–711.

7. An, P., Chen, H., Jiang, X., Su, J., Xiao, Y., Ding, Y., et al. (2020). Clinical features of 2019 novel coronavirus pneumonia presented gastrointestinal symptoms but without fever onset.
8. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet, 395*(10223), 497–506.
9. Hermanek, P., & Carmignato, S. (2017). Porosity measurements by X-ray computed tomography: Accuracy evaluation using a calibrated object. *Precision Engineering, 49*, 377–387.
10. Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., et al. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA, 323*(11), 1061–1069.
11. Mac Parthalain, N., & Shen, Q. (2009). Exploring the boundary region of tolerance rough sets for feature selection. *Pattern Recognition, 42*(5), 655–667.
12. Skowron, A., & Stepaniuk, J. (1996). Tolerance approximation spaces. *Fundamenta Informaticae, 27*(2, 3), 245–253.
13. Rubin, G. D., Ryerson, C. J., Haramati, L. B., Sverzellati, N., Kanne, J. P., Raoof, S., et al. (2020). The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the Fleischner society. *Chest, 158*(1), 106–116.
14. Zhao, J., Zhang, Y., He, X., & Xie, P. (2020). Covid-ct-dataset: A CT scan dataset about covid-19. *arXiv* preprint arXiv:2003.13865, 490.
15. Shan, F., Gao, Y., Wang, J., Shi, W., Shi, N., Han, M., et al. (2020). Lung infection quantification of COVID-19 in CT images with deep learning. *arXiv* preprint arXiv:2003.04655.
16. Zhao, W., Zhong, Z., Xie, X., Yu, Q., & Liu, J. (2020). Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: A multicenter study. *American Journal of Roentgenology, 214*(5), 1072–1077.
17. Appel, L. M., Franke, V., Bruno, M., Grishkovskaya, I., Kasiliauskaite, A., Schoeberl, U. E., et al. (2020). PHF3 regulates neuronal gene expression through the new Pol II CTD reader domain SPOC.
18. Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P. D., Zhang, H., Ji, W., et al. (2020). Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. *arXiv* preprint arXiv:2003.05037.
19. Hasan, N., Bao, Y., & Shawon, A. (2020). *DenseNet convolutional neural networks application for predicting COVID-19 using CT image*.
20. Maghdid, H. S., Asaad, A. T., Ghafoor, K. Z., Sadiq, A. S., Mirjalili, S., & Khan, M. K. (2021, April). Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms. In *Multimodal image exploitation and learning 2021* (Vol. 11734, p. 117340E). International Society for Optics and Photonics.
21. Pathak, Y., Shukla, P. K., Tiwari, A., Stalin, S., & Singh, S. (2020). Deep transfer learning-based classification model for COVID-19 disease. *Ing Rech Biomed*.
22. Shaban, W. M., Rabie, A. H., Saleh, A. I., & Abo-Elsoud, M. A. (2020). A new COVID-19 patients detection strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. *Knowledge-Based Systems, 205*, 106270.
23. Singh, D., Kumar, V., & Kaur, M. (2020). Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks. *European Journal of Clinical Microbiology & Infectious Diseases, 39*(7), 1379–1389.
24. Angelov, P., & Almeida Soares, E. (2020). SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. *medRxiv*.
25. Hu, S., Gao, Y., Niu, Z., Jiang, Y., Li, L., Xiao, X., et al. (2020). Weakly supervised deep learning for covid-19 infection detection and classification from CT images. *IEEE Access, 8*, 118869–118883.
26. Jaiswal, A., Gianchandani, N., Singh, D., Kumar, V., & Kaur, M. (2020). Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics*, 1–8.

27. Liu, S., Georgescu, B., Xu, Z., Yoo, Y., Chabin, G., Chaganti, S., et al. (2020). 3d tomographic pattern synthesis for enhancing the quantification of covid-19. *arXiv*. preprint arXiv:2005.01903.
28. Lysaker, M., Lundervold, A., & Tai, X. C. (2003). Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing, 12*(12), 1579–1590.
29. Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics, 9*(1), 62–66.
30. Albregtsen, F., Nielsen, B., & Danielsen, H. E. (2000, September). Adaptive gray level run length features from class distance matrices. In *Proceedings 15th international conference on pattern recognition. ICPR-2000* (Vol. 3, pp. 738–741). IEEE.
31. Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences, 11*(5), 341–356.
32. Pawlak, Z. (2002). Rough set theory and its applications. *Journal of Telecommunications and Information Technology*, 7–10.
33. Dubois, D., & Prade, H. (1990). Rough fuzzy sets and fuzzy rough sets. *International Journal of General System, 17*(2–3), 191–209.
34. Hu, Q., Yu, D., Xie, Z., & Liu, J. (2006). Fuzzy probabilistic approximation spaces and their information measures. *IEEE Transactions on Fuzzy Systems, 14*(2), 191–201.
35. Yao, Y. (2008). Probabilistic rough set approximations. *International Journal of Approximate Reasoning, 49*(2), 255–271.
36. Slowinski, R., & Vanderpooten, D. (2000). A generalized definition of rough approximations based on similarity. *IEEE Transactions on Knowledge and Data Engineering, 12*(2), 331–336.
37. Yao, Y., & Zhao, Y. (2008). Attribute reduction in decision-theoretic rough set models. *Information Sciences, 178*(17), 3356–3373.
38. Yao, Y., & Yao, B. (2012). Covering based rough set approximations. *Information Sciences, 200*, 91–107.
39. Stepaniuk, J., Kobayashi, S., Yokomori, T., & Tanaka, H. (1996, November). Similarity based rough sets and learning. In *Proceedings of the fourth international workshop on rough sets, fuzzy sets, and machine discovery* (pp. 18–22).
40. Li, X. L., Wang, T., & Du, Z. L. (2005, October). Audio retrieval based on tolerance rough sets. In *2005 international conference on neural networks and brain* (Vol. 3, pp. 1948–1951). IEEE.
41. Hu, Y. C. (2015). Flow-based tolerance rough sets for pattern classification. *Applied Soft Computing, 27*, 322–331.
42. https://github.com/UCSD-AI4H/COVID-CT
43. Bekkar, M., Djemaa, H. K., & Alitouche, T. A. (2013). Evaluation measures for model's assessment over imbalanced data sets. *Journal of Information Engineering and Applications, 3*(10).
44. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437.
45. Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research, 7*, 1–30.
46. Ganesan, J., Inbarani, H. H., Azar, A. T., & Polat, K. (2017). Tolerance rough set firefly-based quick reduct. *Neural Computing and Applications, 28*(10), 2995–3008.
47. Sayed, G. I., Hassanien, A. E., & Azar, A. T. (2019). Feature selection via a novel chaotic crow search algorithm. *Neural Computing and Applications, 31*(1), 171–188.
48. Inbarani, H. H., Kumar, S. U., Azar, A. T., & Hassanien, A. E. (2018). Hybrid rough-bijective soft set classification system. *Neural Computing and Applications, 29*(8), 67–78.
49. Kumar, S. S., Inbarani, H. H., Azar, A. T., & Polat, K. (2017). Covering-based rough set classification system. *Neural Computing and Applications, 28*(10), 2879–2888.
50. Inbarani, H. H., Banu, P. N., & Azar, A. T. (2014). Feature selection using swarm-based relative reduct technique for fetal heart rate. *Neural Computing and Applications, 25*(3), 793–806.
51. Inbarani, H. H., Bagyamathi, M., & Azar, A. T. (2015). A novel hybrid feature selection method based on rough set and improved harmony search. *Neural Computing and Applications, 26*(8), 1859–1880.