A Study of the Neutrosophic Set Significance on Deep Transfer Learning Models: an Experimental Case on a Limited COVID-19 Chest X-ray Dataset

Nour Eldeen M. Khalifa[1] · Florentin Smarandache[2] · Gunasekaran Manogaran[3,4] · Mohamed Loey[5]

Abstract

Coronavirus, also known as COVID-19, has spread to several countries around the world. It was announced as a pandemic disease by The World Health Organization (WHO) in 2020 for its devastating impact on humans. With the advancements in computer science algorithms, the detection of this type of virus in the early stages is urgently needed for the fast recovery of patients. In this paper, a study of neutrosophic set significance on deep transfer learning models will be presented. The study will be conducted over a limited COVID-19 x-ray. The study relies on neutrosophic set and theory to convert the medical images from the grayscale spatial domain to the neutrosophic domain. The neutrosophic domain consists of three types of images, and they are the True (T) images, the Indeterminacy (I) images, and the Falsity (F) images. The dataset used in this research has been collected from different sources. The dataset is classified into four classes {COVID-19, normal, pneumonia bacterial, and pneumonia virus}. This study aims to review the effect of neutrosophic sets on deep transfer learning models. The selected deep learning models in this study are Alexnet, Googlenet, and Restnet18. Those models are selected as they have a small number of layers on their architectures. To test the performance of the conversion to the neutrosophic domain, more than 36 trials have been conducted and recorded. A combination of training and testing strategies by splitting the dataset into (90–10%, 80–20%, 70–30) is included in the experiments. Four domains of images are tested, and they are, the original domain, the True (T) domain, the Indeterminacy (I) domain, and the Falsity (F) domain. The four domains with the different training and testing strategies were tested using the selected deep transfer models. According to the experimental results, the Indeterminacy (I) neutrosophic domain achieves the highest accuracy possible with 87.1% in the testing accuracy and performance metrics such as Precision, Recall, and F1 Score. The study concludes that using the neutrosophic set with deep learning models may be an encouraging transition to achieve better testing accuracy, especially with limited COVID-19 datasets.

✉ Mohamed Loey
  mloey@fci.bu.edu.eg

  Nour Eldeen M. Khalifa
  nourmahmoud@cu.edu.eg

  Florentin Smarandache
  smarand@unm.edu

  Gunasekaran Manogaran
  gmanogaran@ieee.org

[1] Department of Information Technology, Faculty of Computers and Artificial Intelligence, Cairo University, Cairo 12613, Egypt

[2] Department of Mathematics, University of New Mexico, Gallup Campus, NM 87301, USA

[3] University of California, Davis, USA

[4] College of Information and Electrical Engineering, Asia University, Taichung City, Taiwan

[5] Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha 13518, Egypt

Introduction

Severe acute respiratory syndrome–related coronavirus (SARS-CoV-1) is a kind of B-coronavirus that infects bats and some other mammals. SARS-CoV-1 was a kind of coronavirus as a family of the B-coronavirus (B- CoV) subgroup and was title as SARSr-CoV. Historically, SARS-CoV-1, across 29 countries in the world, infected over 8000 humans and at least 750 died [1, 2]. In 2019, a coronavirus epidemic is an ongoing scourge of coronavirus malady 2019 (COVID-19) created by SARS-CoV-2 [3]. However, SARS-CoV-2 infected more than four million humans with more than 300,000 deaths, 1.6 million recovered cases, and 300,000 death cases [4]. It elucidates that the propagate rate of SARS-CoV-2 is greater than SRAS-CoV [5, 6].

The theory of neutrosophic logic was proposed by Smarandache in 1995. Afterward, it has been unified and generalized by its founder in 1999 [7]. Since that date, neutrosophic logic has been used in many computer science fields including pattern recognition [8], image segmentation, and processing [9], and more. It contributes to solving many research and practical real-life problems in a lot of domains such as medicine [10], economics [11], space satellite[12], and agriculture. Neutrosophy leads to a whole family of novel mathematical theories with an overview of not only classical but also fuzzy counterparts [13]. The term neutrosophy means knowledge of neutral thought, and this neutral represents the main difference between fuzzy and intuitionistic fuzzy logic and set [14, 15]. Neutrosophic set has the required potentials of being a general framework for uncertainty analysis in data sets [14] and especially with images in the field of Artificial Intelligence and deep learning. Deep Transfer Learning (DTL) is a type of Artificial Intelligence (AI) concerned with methods inspired by the functions of people's brain [16]. For the time being, DTLs like VGG, ResNet, and DenseNets [17–21] are quickly becoming an important method in image/video detection and diagnosis based such as. DL is used in medical x-ray/computed tomography diagnoses. DL upgrade the medical diagnosis system (MDS) to realize great results, and implementing an applicable real-time medical diagnosis system [22, 23].

This part is dedicated to works on the recent x-ray academic researches for applying DL in the field of MDS in chest x-ray diagnosis. Ayan and Ünver [24] proposed an early medical diagnosis system for Pneumonia chest x-ray images based on DTL models. In this academic research, x-ray data [25] containing about 1600 healthy cases, 4200 un-healthy pneumonia cases. The trial score introduced that VGG DTL networks better than the X-ception DTL network with an error rate of 19%. In Stephen et al. [26], it introduced a new method of diagnosing the existence of pneumonia from chest x-ray database samples based on a CNN architecture with augmentation algorithms trained based on an x-ray database [25]. The results the model improves medical x-ray diagnosis with a miss-classification rate of 12.88% in training miss-classification rate is 18.35% in the validation. Varshni et al. [27] introduced DTL architectures as feature extractors followed by various classifiers (k-nearest neighbors, naïve Bayes, support vector machine, and random forest algorithm) for the diagnose of healthy/unhealthy chest x-ray data. They used an x-ray database called ChestX-ray14 proposed by Wang et al. [28]. Islam et al. [29] introduced a Compressed Sensing (CS) with DTL architectures for automatic identification of pneumonia on the x-ray database to assist the medical physicians. The chest x-ray database used for this research contained about 5800 x-ray images of (healthy / unhealthy). The suggested simulation results have shown that the proposed DTL architectures diagnose pneumonia from a chest x-ray with an error rate of 2.66%. Finally, Chouhan et al. [30] proposed an ensemble DTL architecture that combines results from all DTL architectures for the identification of chest pneumonia x-ray based on the concept of DL. The suggested model based on Kermany et al. [25] database reached a miss-classification error of 3.6%. In this study, we introduced a neutrosophic study based on DTL to classify COVID-19 infection versus non-COVID-19 diseases. We hypothesized that the DTL with neutrosophic would help doctors in detecting COVID-19 x-ray scan images.

Data and Methods

The selected COVID-19 x-ray dataset used in this research was acquired from [25, 31]. It was collected from different websites such as the Italian Society of Medical, Radiopaedia web, and online publications. The created dataset is organized into four categories normal, pneumonia bacterial, pneumonia virus, and COVID19. The dataset contains 306 x-ray images divided into 79 of normal, 69 of COVID-19, 79 of pneumonia bacterial, and 79 of pneumonia virus. Fig. 1 illustrates samples of images used for this research.

The methodology adopted in this research is to propose a model that can correctly classify the x-rays images with the different 4 classes. The proposed model includes two main components: the first component is neutrosophic domain conversion, while the second component is the transfer learning architectures. Figure 2 illustrates the proposed neutrosophic/DTL model for the study. The neutrosophic image domain conversion is used as a preprocessing step while the DTL architectures are used in the training and the testing steps.
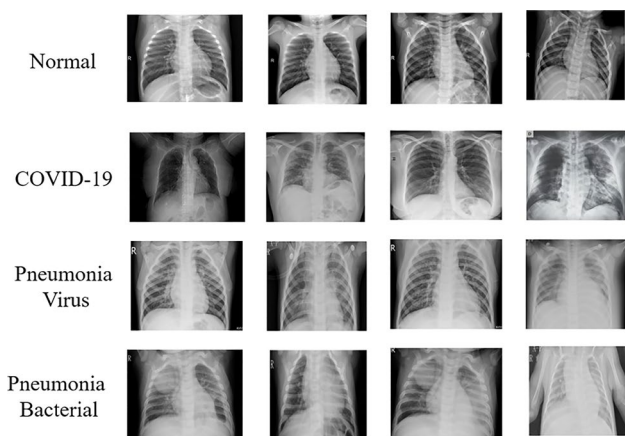
Fig. 1. Samples of chest x-ray dataset

Neutrosophic Image Domain Conversion

Neutrosophy (NS) is a theory sophisticated and created by Florentin Smarandache [32–34]. NS is a useful and helpful theory in computing fuzzy situations. In NS theory, events are computed by subset them into three sets as true (*T*) significance, the status is percentage of true; as indeterminacy (*I*) significance, the status is percentage of indefinite; and as falsity (*F*) significance, the status is percentage of false, where *t* varies in T subsets. In image processing such as object and edge detection, all pixels

of the image are subdivided into T, I, and F subsets. Then, the edge detection/object process of the image is performed through necessary operations on these subsets. The input image converts to the neutrosophic domain as shown in Eqs. 1–5. *P*(*n*, *m*) pixel in the image domain is converted to neutrosophic domain *P2NS*(*n*, *m*) [35, 36]:

$$\text{P2NS}_{\text{NS}}(n,m) = \{T_{n,m}, I_{n,m}, F_{n,m}\} \tag{1}$$

$$T_{n,m} = \frac{f\left(\bar{n},m\right) - \bar{f}_{\min}}{\bar{f}_{\max} - \bar{f}_{\min}} \tag{2}$$

where $I\left(\bar{n},m\right)$ is the local average value of related pixels $f_{\min}$ and $f_{\max}$ variables correspond to the last and first peaks measured from those pixels with a value higher than the maximum local average of the histogram.

$$I_{n,m} = 1 - \frac{H(n,m) - \bar{H}_{\min}}{\bar{H}_{\max} - \bar{H}_{\min}} \tag{3}$$

$$H(n,m) = \text{abs}(I(n,m) - I\left(\bar{n},m\right)) \tag{4}$$

where $H(n,m)$ is the homogeneity value of *T* at (n,m), which is measured by the absolute value of the difference between

Fig. 2 The introduced Neutrosophic /DTL model for the study
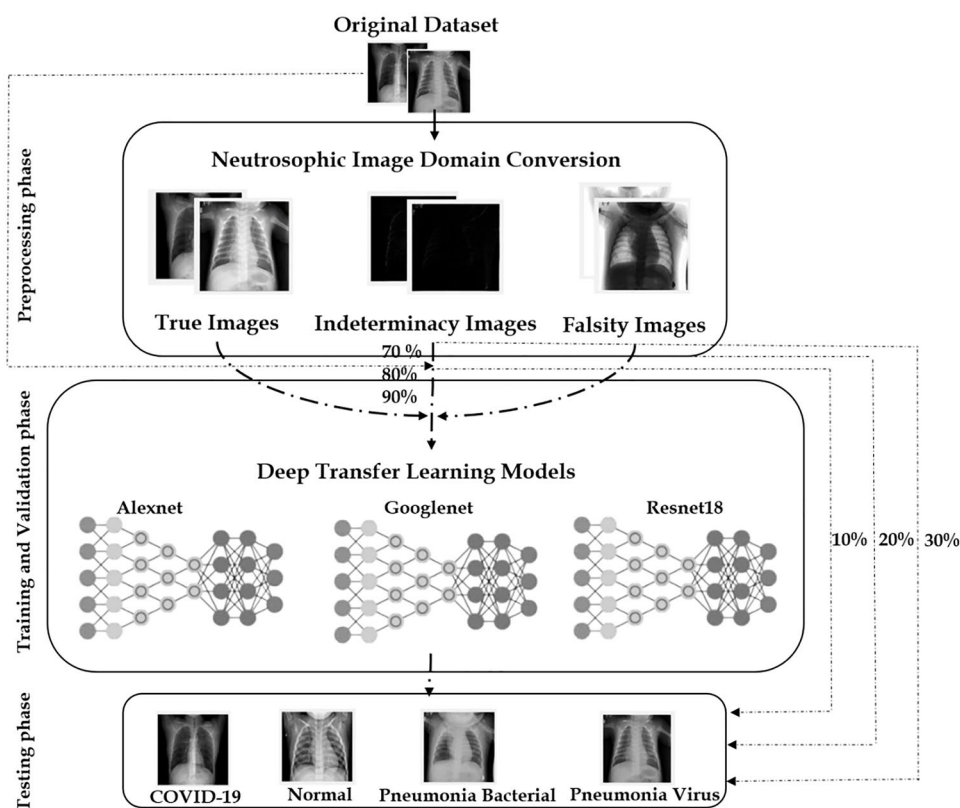
Algorithm 1: DTL algorithm.

1: **Input:** Images
2: **Output:** Performance measurements
3: **for** *DTL model* **do**
4:    *learning rate* = 0.001
5:    **for** *iteration* = 1 to 50 **do**
6:       **for** *image in training dataset* do
7:          **if** the miss-classification rate is increased for ten *iteration* **then**
8:             *update model learning rate*
9:          **end**
10:          *Update model coefficients*
11:       **end**
12:    **end**
13: **end**
14: **for** *image in test dataset* **do**
15:    *Calculate performance measurements for all DTL architectures*
16: **end**

---

intensity $f(n, m)$ and its local mean value $\bar{f}\left(\bar{n}, m\right)$. While $\bar{H}_{\max}$ and $H_{\min}$ are the last and first peaks respectively, measured from the homogeneity image.

$$F_{n,m} = 1 - T_{n,m} \qquad (5)$$

After the conversion of the image to the NS domain, the COVID-19 chest x-ray (object) is kept in the $T_{n,m}$ domain, the edges are in the $I_{n,m}$ domain, and the background is kept in the $F_{n,m}$ domain. Figure 3 presents samples of images after the conversion neutrosophic image domain in the different domains for every class in the dataset.
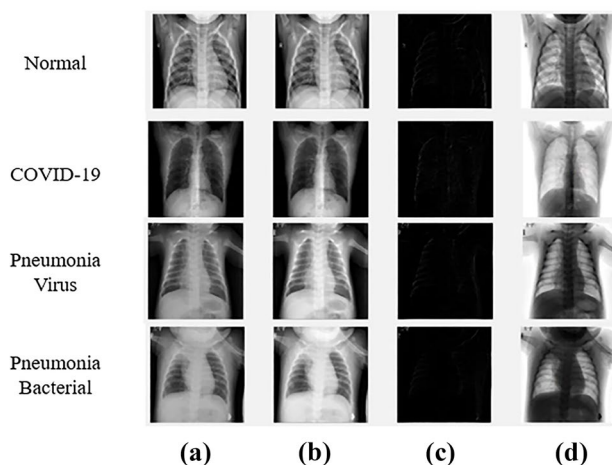

Fig. 3 Different neutrosophic images domain for 4 classes in the dataset were (**a**) original images, (**b**) True significance, (**c**) Indeterminacy significance, and (**d**) Falsity significance images

Deep Transfer Learning Model

In this work, the procedure of the DTL follows algorithm 1. The procedure starts with the input image. The input image will be fed into the DTL model with a learning rate of 0.001, and the number of epochs equal to 50. The training procedure will start. If there is no enhancement for the validation accuracy for 10 epochs, the update rate will be changed else the weights of the network will be updated. After the completion of the training phase, the performance measurements will be calculated to evaluate the performance of the DTL model.

Results

The introduced model for the evaluating of neutrosophic sets with deep transfer models was implemented using a software package (MATLAB). The experiments were conducted on a computer server with 96 GB of RAM, and an Intel Xeon processor (2 GHz). The development was CPU specific. About 36 recorded experiments were conducted in this study. The experiments included the following setups:

- Different Image domains

  - The Original dataset domain (grayscale).
  - The True (T) neutrosophic domain.
  - The Indeterminacy (I) neutrosophic domain
  - The Falsity (F) neutrosophic domain.

- Different training and testing strategies

  - 70% for the training – 30% for the testing.

- 80% for the training – 20% for the testing.
- 90% for the training – 10% for the testing.

●Different deep transfer models

- Alexnet.
- Googlenet.
- Restnet18.

The authors of this research tried first to build their deep neural networks based on the works presented [37–39], but the testing accuracy was not acceptable. So, the alternative way is to use deep transfer learning models. Using deep transfer models proved its efficiency in many types of research such as work presented in previous studies [40–44]. The Alexnet, Googlenet, and Restnet18 models are selected in this study as they have a small number of layers on their architectures which will reflect on decreasing the training time, consumed memory, and processing time. All the experimental results have been tested according to the following hyperparameters for the training, and the testing phases:

- Batch size [45]: 32
- Momentum: 0.9
- Epochs: 50
- Learning Rate [45]: 0.001
- Optimizer [46]: Adaboost
- Early stopping [47]: 10 epochs

The early stopping [47] plays an essential role in reducing training time and computational complexity. It triggers when the no improvement of the validation accuracy for a certain number of epochs. In the proposed model, the training stops if there no improvement for 10 epochs.

A large number of trials were performed to draw a full picture of the effectiveness and the significance of using neutrosophic sets in different experimental environments with different deep learning models. To evaluate the performance of the neutrosophic set in deep transfer learning models, performance matrices are needed to be investigated through this study. The most common performance measures in the field

of deep learning are Accuracy, Precision, Recall, and F1 Score [48], and they are presented from Eqs. 6–9.

$$Accuracy = \frac{True_{Positive} + True_{Negative}}{(True_{Positive} + False_{Positive}) + (True_{Negative} + False_{Negative})} \quad (6)$$

$$Precision = \frac{True_{Positive}}{(True_{Positive} + False_{Positive})} \quad (7)$$

$$Recall = \frac{True_{Positive}}{(True_{Positive} + False_{Negative})} \quad (8)$$

$$F1\ Score = 2 * \frac{Precision * Recall}{(Precision + Recall)} \quad (9)$$

where $True_{Positive}$ is the number of True Positive samples, $True_{Negative}$ is the number of True Negative samples, $False_{Positive}$ is the number of False Positive samples, and $False_{Negative}$ is the number of False Negative samples from a confusion matrix. The experimental results will be presented in three subsections: the first subsection will discuss the experimental results for the original dataset. The second subsection will introduce the different neutrosophic domain experimental results. Finally, the third subsection will illustrate a comparative results analysis for the original, the neutrosophic domain according to the confusion matrix for the highest achieved testing accuracy.

Original Dataset Experimental Results

Table 1 presents the testing accuracy and the performance metrics for the original dataset. The table clearly shows that in the 90–10% strategy, the Resnet18 model achieves the highest testing accuracy with 74.19% with close scores in performance metrics. In the 80–20% strategy, Googlenet achieves the highest testing accuracy with 64.52%, while in a 70–30% strategy, the Googlenet model achieves the highest testing accuracy with 62.47% and close to Resnet18 which achieved 61.29%.

Table 1 Testing accuracy and performance metrics for the original dataset

Training/Testing	Deep transfer model	Recall	Precision	*F* score	Testing accuracy
90–10%	Alexnet	0.6875	0.7582	0.7211	0.6774
	Googlenet	0.7188	0.8167	0.7646	0.7097
	Resnet18	0.7500	0.7639	0.7569	0.7419
80–20%	Alexnet	0.5781	0.6222	0.5994	0.5645
	Googlenet	0.6563	0.7477	0.6990	0.6452
	Resnet18	0.6250	0.681	0.6518	0.6129
70–30%	Alexnet	0.5506	0.5706	0.5604	0.5376
	Googlenet	0.6354	0.6814	0.6576	0.6237
	Resnet18	0.6250	0.6929	0.6572	0.6129

Table 2 Testing accuracy and performance metrics for the True (T) neutrosophic domain

Training/Testing	Deep transfer model	Recall	Precision	*F* score	Testing accuracy
90–10%	Alexnet	0.6563	0.6979	0.6764	0.6452
	Googlenet	0.6563	0.7500	0.7000	0.6452
	Resnet18	0.5938	0.6556	0.6231	0.5806
80–20%	Alexnet	0.5625	0.5868	0.5744	0.5484
	Googlenet	0.5625	0.6139	0.5871	0.5484
	Resnet18	0.5156	0.5893	0.5500	0.5000
70–30%	Alexnet	0.6310	0.7433	0.6825	0.6237
	Googlenet	0.6860	0.7462	0.7149	0.6774
	Resnet18	0.6979	0.7565	0.7260	0.6882

Table 1 illustrates interesting facts, and they are the following: (1) The more data the deep learning models have, the more testing accuracy they will achieve [49]. (2) The 80–20% and 70–30% strategy achieved close results for testing accuracy which means that those strategies are enough and reflect the accurate testing accuracy for the model.

Neutrosophic Domains Experimental Results

The neutrosophic domains included three types, and they are the True (T) neutrosophic domain, the Indeterminacy (I) neutrosophic domain, and the Falsity (F) neutrosophic domain. Those neutrosophic domains will be experimented on in this section to measure their performance under different experimental conditions. Table 2 presents the testing accuracy and performance metrics for the True (T) neutrosophic domain. As illustrated in "Neutrosophic Image Domain Conversion," the True (T) image is the averaging of the original image as every pixel is averaged by it is neighbors with a window of choice. The choice of the window in the study is 5 pixels.

Table 2 illustrates that in the 90–10% strategy, both Alexnet and Googlenet model achieves similar highest testing accuracy with 64.52% with an advantage for the Googlenet model in the achieved performance metrics. In the 80–20% strategy, also both Alexnet and Googlenet models achieve similar highest testing accuracy with 54.84% with an

advantage for the Googlenet model in the achieved performance metrics, while in the 70–30% strategy, the Resnet18 model achieves the highest testing accuracy with 67.74%.

Table 2 illustrates interesting facts, and they are the following: (1) In the True (T) neutrosophic domain, more data does not mean better accuracy in those deep learning architectures' as in the 70–30% strategy; the highest testing accuracy is achieved by 68.82% all over the other strategies. (2) The images on the True (T) neutrosophic domain are averaged images, which means that some of the important features of images are concealed which negatively affect the achieved testing accuracy if it is compared with the original experimental results presented in Table 1 for the 90–10%, and the 80–20% strategy.

The second neutrosophic domain to be experimented on is the Falsity (F) neutrosophic domain. This domain is the opposite of the True (T) neutrosophic domain. In the Falsity (F) domain, all pixel's values are inverted, it is expected that some features will be concealed, and other features will be revealed in images. Table 3 presents the testing accuracy and performance metrics for the (F) Falsity domain.

Table 3 illustrates that in the 90–10% strategy, both Alexnet and Googlenet models achieve similar highest testing accuracy with 64.52% with an advantage for the Googlenet model in the achieved performance metrics. In the 80–20% strategy, also both Alexnet and Googlenet models achieve similar highest testing accuracy with 56.45% with

Table 3 Testing accuracy and performance metrics for the Falsity (F) domain

Training/Testing	Deep transfer model	Recall	Precision	*F* score	Testing accuracy
90–10%	Alexnet	0.6563	0.6714	0.6638	0.6452
	Googlenet	0.6563	0.7404	0.6958	0.6452
	Resnet18	0.5938	0.6408	0.6164	0.5806
80–20%	Alexnet	0.5781	0.7153	0.6394	0.5645
	Googlenet	0.5781	0.6249	0.6006	0.5645
	Resnet18	0.5469	0.5759	0.5610	0.5323
70–30%	Alexnet	0.6131	0.7250	0.6644	0.6022
	Googlenet	0.6667	0.7083	0.6869	0.6559
	Resnet18	0.6548	0.7036	0.6783	0.6452

Table 4 Testing accuracy and performance metrics for the Indeterminacy (I) domain

Training/testing	Deep transfer model	Recall	Precision	F score	Testing accuracy
90–10%	Alexnet	0.8750	0.9167	0.8953	0.8710
	Googlenet	0.8125	0.8458	0.8288	0.8065
	Resnet18	0.7813	0.8534	0.8157	0.7742
80–20%	Alexnet	0.6406	0.8386	0.7264	0.6290
	Googlenet	0.6719	0.8116	0.7352	0.6613
	Resnet18	0.6406	0.7688	0.6989	0.6290
70–30%	Alexnet	0.7158	0.7440	0.7296	0.7097
	Googlenet	0.7336	0.8464	0.7860	0.7312
	Resnet18	0.7396	0.8294	0.7819	0.7312

an advantage for the Googlenet model in the achieved performance metrics, while in 70–30% strategy, the Googlenet model achieves the highest testing accuracy with 65.59%.

Table 3 also shows interesting facts, and they are the following: (1) In the Falsity (F) neutrosophic domain, more data does not mean better accuracy in those deep learning architectures' as in the 70–30% strategy, the highest testing accuracy is achieved by 65.59% all over the other strategies. (2) The images on the Falsity (F) neutrosophic domain are the inversion of the True (T) domain, which means that some of the important features of images are concealed which negatively affect the achieved testing accuracy if it is compared with the original experimental results are presented in Table 1 for the 90–10%, and the 80–20% strategy. (3) The results presented in Table 3 are very close to results presented in Table 2, which means the Falsity (F) neutrosophic domains do not add extra value for the grayscale images and can be discarded in some applications depending on their nature.

The third neutrosophic domain to be experimented on is the Indeterminacy (I) neutrosophic domain. This domain contains the absolute edges in the image. In the Indeterminacy (I) domain, all pixel values are resulted from subtracting the original pixel value from the average pixel value in the True (T) neutrosophic domain. Table 5 presents the testing accuracy and performance metrics for the Indeterminacy (I) domain.

Table 5 illustrates that in the 90–10% strategy, Alexnet achieves the highest testing accuracy with 87.10% with the highest achieved performance metrics scores. In the 80–20% strategy, Googlenet achieved the highest testing accuracy with 66.13%, while in the 70–30% strategy, both Googlenet and Resnet18 models achieve similar highest testing accuracy with 73.12% with an advantage for the Googlenet model in the achieved performance metrics.

Table 4 illustrates interesting facts, and they are the following: (1) In the Indeterminacy (I) neutrosophic domain, all the achieved testing accuracies are better than all the achieved testing accuracies in the Falsity (F), the True (T), and the original domain. (2) The images on Indeterminacy

(I) neutrosophic domain are the absolute difference between the original and the averaged image in the True (T) domain. Those are very important features that are revealed in an image and helped deep transfer models to achieve higher testing accuracy. (3) More data does not mean achieving higher testing accuracy in Indeterminacy (I) neutrosophic domain; in the 70–30% strategy, the achieved testing accuracy was better than the achieved accuracy in 80–20% strategy with 6.99% better using Googlenet.

Comparative Result of Indeterminacy (I) Neutrosophic Domain with the Original Domain

"Neutrosophic Domains Experimental Results"concluded that the Indeterminacy (I) neutrosophic domain achieved the highest possible testing accuracy in all experiment's trails. This section is dedicated to present a comparison result between the Indeterminacy (I) neutrosophic domain with the original domain with deeper performance metrics to evaluate the performance of the Indeterminacy (I) domain. Table 5 presents a comparative result of the achieved testing accuracy between the Indeterminacy (I) and the original domain.

Table 5 shows that the Indeterminacy (I) neutrosophic domain achieved the highest testing accuracy in all training and testing strategies with 87.10% (Alexnet), 66.13% (Googlenet), and 73.12% (Googlenet) in the 90–10%, 80–20%, and 70–30% accordingly.

Table 5 Testing accuracy for the Indeterminacy (I) and original domain

Training/Testing	Domain	Deep transfer learning model	Highest testing accuracy
90–10%	Original	Resnet18	0.7419
	Indeterminacy (I)	Alexnet	0.8710
80–20%	Original	Googlenet	0.6452
	Indeterminacy (I)	Googlenet	0.6613
70–30%	Original	Googlenet	0.6237
	Indeterminacy (I)	Googlenet	0.7312

Fig. 4 Confusion matrix for Alexnet in 90–10% percent strategy for Indeterminacy (I) neutrosophic domain

Fig. 5 Confusion matrix for Googlenet in 80–20% percent strategy for Indeterminacy (I) neutrosophic domain

Table 5 also shows interesting facts, and they are the following: (1) In the Indeterminacy (I) neutrosophic or the original domain, the Googlenet model is the most dominant in achieving the highest accuracy possible as it contains 20 layers in its architecture if it is compared with Alexnet and Resnet18 which contains 8, and 18 layers. (2) The Indeterminacy (I) neutrosophic greatly affects the testing accuracy; in the 90–10% the Indeterminacy (I) neutrosophic domain achieves better accuracy with 12.91% more. In the 70–30% the Indeterminacy (I) neutrosophic domain achieves better accuracy with 10.75% more. (3) In the Indeterminacy (I) neutrosophic domain, deep transfer models can learn with fewer data as illustrated in the 70–30% strategy, the Googlenet achieves better accuracy than the 80–20% strategy. That means the model can generalize whatever the amount of the data existed. While in the original domain, more data means better testing accuracy.

All the experimental outcomes show that converting to the Indeterminacy (I) neutrosophic domain from the original domain grantees achieving better testing accuracy. The Indeterminacy (I) neutrosophic need further investigations to prove it is efficient for the detection of COVID-19 among the other classes. The confusion matrices for the Indeterminacy (I) neutrosophic domain for the different deep transfer models are presented in Figs. 4, 5, and 6.

The figures show that the testing accuracy for the COVID-19 class in the different training and testing strategies are acceptable. For the 90–10% strategy, The Alexnet model was able to detect COVID-19 with a testing

accuracy of 100% and for the normal class with 100%. In the 80–20% strategy, The Googlenet model was able to detect COVID-19 with a testing accuracy of 77.8% and for the normal class with 100%. While in the 70–30% strategy, The Googlenet model was able to detect COVID-19 with a testing accuracy of 100% and for the normal class with 87.5%.

Fig. 6 Confusion matrix for Googlenet in 70–30% percent strategy for Indeterminacy (I) neutrosophic domain

Conclusion and Future Works

According to the World Health Organization (WHO), coronaviruses are a family of viruses that lead to sicknesses ranging from the common cold to more severe diseases. With the advancements in computer science, detection of this type of virus is urgently needed. In this paper, a study of neutrosophic significance on the deep transfer learning model is presented. The neutrosophic domain consisted of three types of images and they are, the True (T) images, the Indeterminacy (I) images, and the Falsity (F) images. The dataset used in this research had been collected from different sources as there is no benchmark dataset for COVID-19 chest x-ray until the writing of this research. The dataset consisted of four classes, and they were COVID-19, Normal, Pneumonia bacterial, and Pneumonia virus. This study aimed to review the effect of neutrosophic sets on deep transfer learning models. The selected deep learning models in this study were Alexnet, Googlenet, and Restnet18. Those models were selected as they had a small number of layers on their architectures that will reflect on reducing the consumed memory and training time. To test the performance of the conversion to the neutrosophic domain, more than 36 trials had been conducted and recorded. A combination of training and testing strategies by splitting the dataset into (90–10%, 80–20%, 70–30) were included in the experiments. Four domains of images are tested, and they are the original images, the True (T) neutrosophic images, the Indeterminacy (I) neutrosophic images, and the Falsity (F) neutrosophic images. The four domains with the different training and testing strategies were tested using Alexnet, Googlenet, and Restnet18 deep transfer models. According to the experimental results, the Indeterminacy (I) neutrosophic domain achieved the highest accuracy possible in the testing accuracy and performance metrics such as Precision, Recall, and F1 Score. The study concluded that using the neutrosophic set with deep learning models might be an encouraging transition to achieve better testing accuracy, especially with limited COVID-19 x-ray datasets. One of the potential future works is trying the proposed model on bigger datasets. Also, include deeper transfer learning for experimental investigation such as Resnet50, and X-ception model with the neutrosophic theory.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Chang L, Yan Y, Wang L. Coronavirus disease 2019: coronaviruses and blood safety transfusion medicine reviews. 2020. https://doi.org/10.1016/j.tmrv.2020.02.003

2. Loey M, Smarandache F, Khalifa M.N.E. Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. Symmetry 2020;12:651. https://doi.org/10.3390/sym12040651.

3. Singhal T. A review of coronavirus disease-2019 (COVID-19). The Indian Journal of Pediatrics. 2020;87:281–6. https://doi.org/10.1007/s12098-020-03263-6.

4. Coronavirus (COVID-19) map Available online: https://www.google.com/covid19-map/ (accessed on Apr 26, 2020).

5. York A. Novel coronavirus takes flight from bats? Nature Reviews Microbiology. 2020;18:191. https://doi.org/10.1038/s41579-020-0336-9.

6. Rabi FA, Al Zoubi MS, Kasasbeh GA, Salameh DM, Al-Nasser AD. SARS-CoV-2 and coronavirus disease what we know so far. Pathogens. 2019;2020(9):231. https://doi.org/10.3390/pathogens9030231.

7. Smarandache F. A unifying field in logics: neutrosophic logic. 1999; ISBN 978-1-59973-080-6.

8. Ali M, Deli I, Smarandache F. The theory of neutrosophic cubic sets and their applications in pattern recognition. IFS. 2016;30:1957–63. https://doi.org/10.3233/IFS-151906.

9. Salama A. Basic structure of some classes of neutrosophic crisp nearly open sets and possible application to GIS topology. Neutrosophic Sets and Systems. 2015;7:18–22.

10. Neutrosophic Set in Medical Image Analysis; Elsevier, 2019; ISBN 978-0-12-818148-5.

11. Christianto V, Smarandache F. A review of seven applications of neutrosophic logic. Cultural Psychology, Economics Theorizing, Conflict Resolution, Philosophy of Science, etc. 2019;2:128–37.

12. Bausys R, Kazakeviciute Januskeviciene G, Cavallaro F, Usovaite A. Algorithm selection for edge detection in satellite images by neutrosophic WASPAS method. Sustainability. 2020;12:548.

13. Smarandache F, Broumi S, Singh PK, Liu C, Venkateswara Rao V, Yang HL, Patrascu, I Elhassouny. A. Introduction to neutrosophy and neutrosophic environment. In Neutrosophic Set in Medical Image Analysis; Elsevier, 2019; pp. 3–29 ISBN 978-0-12-818148-5.

14. Majumdar, P. Neutrosophic sets and its applications to decision making. In Computational intelligence for big data analysis; Acharjya, D.P., Dehuri, S., Sanyal, S., Eds.; Adaptation, Learning, and Optimization; Springer International Publishing: Cham, 2015; Vol. 19, pp. 97–115 ISBN 978-3-319-16597-4.

15. Smarandache, F. Neutrosophic set is a generalization of intuitionistic fuzzy set, inconsistent intuitionistic fuzzy set (picture fuzzy set, ternary fuzzy set), pythagorean fuzzy set, q-rung orthopair fuzzy set, spherical fuzzy set, etc. arXiv:1911.07333 [math] 2019.

16. Rong D, Xie L, Ying Y. Computer vision detection of foreign objects in walnuts using deep learning. Computers and Electronics in Agriculture. 2019;162:1001–10. https://doi.org/10.1016/j.compag.2019.05.019.

17. Liu S, Deng W. Very deep convolutional neural network based image classification using small training sample size. In Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). 2015; pp. 730–734.

18. Szegedy C, Wei L, Yangqing J, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015; pp. 1–9.

19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on

Computer Vision and Pattern Recognition (CVPR). 2016; pp. 770–778.

20. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017; pp. 1800–1807.

21. Huang G, Liu Z, Maaten L. v. d, Weinberger K.Q. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017; pp. 2261–2269.

22. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik. 2019;29:102–27. https://doi.org/10.1016/j.zemedi.2018.11.002.

23. Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. Zeitschrift für Medizinische Physik. 2019;29:86–101. https://doi.org/10.1016/j.zemedi.2018.12.003.

24. Ayan E, Ünver HM. Diagnosis of pneumonia from chest x-ray images using deep learning. In Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). 2019; pp. 1–5.

25. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018;172:1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010.

26. Stephen O, Sain M, Maduh UJ, Jeong D-U. An efficient deep learning approach to pneumonia classification in healthcare. Journal of Healthcare Engineering. 2019;2019:4180949. https://doi.org/10.1155/2019/4180949.

27. Varshni D, Thakral K, Agarwal L, Nijhawan R, Mittal A. Pneumonia detection using CNN based feature extraction. In Proceedings of the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). 2019; pp. 1–7.

28. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017; pp. 3462–3471.

29. Islam SR, Maity SP, Ray AK, Mandal M. Automatic detection of pneumonia on compressed sensing images using deep learning. In Proceedings of the 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). 2019; pp. 1–4.

30. Chouhan V, Singh SK, Khamparia A, Gupta D, Tiwari P, Moreira C, Damaševičius R, de Albuquerque VHC. A novel transfer learning based approach for pneumonia detection in chest x-ray images. Applied Sciences. 2020;10:559. https://doi.org/10.3390/app10020559.

31. Cohen JP, Morrison P, Dao L. COVID-19 image data collection. arXiv:2003.11597 [cs, eess, q-bio] 2020.

32. Smarandache F. Neutrosophic masses indeterminate models. Applications to information fusion. In Proceedings of the The 2012 International Conference on Advanced Mechatronic Systems. 2012; pp. 674–679.

33. Smarandache F, Vlădăreanu L. Applications of neutrosophic logic to robotics: an introduction. In Proceedings of the 2011 IEEE International Conference on Granular Computing. 2011; pp. 607–612.

34. Deli I, Ali M, Smarandache F. Bipolar neutrosophic sets and their application based on multi-criteria decision making problems. In Proceedings of the 2015 International Conference on Advanced Mechatronic Systems (ICAMechS). 2015; pp. 249–254.

35. Anter AM, Hassenian AE. CT liver tumor segmentation hybrid approach using neutrosophic sets, fast fuzzy c-means and adaptive watershed algorithm. Artificial Intelligence in Medicine. 2019;97:105–17. https://doi.org/10.1016/j.artmed.2018.11.007.

36. Özyurt F, Sert E, Avci E, Dogantekin E. Brain tumor detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy. Measurement. 2019;147:106830. https://doi.org/10.1016/j.measurement.2019.07.058.

37. Khalifa N, Taha M, Hassanien A, Mohamed H. Deep iris: deep learning for gender classification through iris patterns. Acta Informatica Medica. 2019;27:96. https://doi.org/10.5455/aim.2019.27.96-102.

38. Khalifa NEM, Taha MHN, Hassanien AE, Hemedan AA. Deep bacteria: robust deep learning data augmentation design for limited bacterial colony dataset. International Journal of Reasoning-based Intelligent Systems. 2019. https://doi.org/10.1504/ijris.2019.102610.

39. Khalifa NEM, Taha MHN, Ezzat Ali D, Slowik A, Hassanien AE. Artificial intelligence technique for gene expression by tumor RNA-Seq Data: a novel optimized deep learning approach. IEEE Access 2020, https://doi.org/10.1109/access.2020.2970210.

40. Khalifa N, Loey M, Taha M, Mohamed H. Deep transfer learning models for medical diabetic retinopathy detection. Acta Informatica Medica. 2019;27:327. https://doi.org/10.5455/aim.2019.27.327-332.

41. Khalifa NEM, Loey M, Taha MHN. Insect pests recognition based on deep transfer learning models. J Theor Appl Inf Technol. 2020;98:60–8.

42. Loey M, ElSawy A, Afify M. Deep learning in plant diseases detection for agricultural crops: a survey Available online: www.igi-global.com/article/deep-learning-in-plant-diseases-detection-for-agricultural-crops/248499 (accessed on Apr 11, 2020).

43. El-Sawy A, EL-Bakry H, Loey M. CNN for Handwritten Arabic digits recognition based on LeNet-5 BT - Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016; Hassanien AE, Shaalan K, Gaber T, Azar AT, Tolba MF, Eds.; Springer International Publishing: Cham. 2017; pp. 566–575.

44. El-Sawy A, Loey M, EL-Bakry H. Arabic handwritten characters recognition using convolutional neural network. WSEAS Transactions on Computer Research 2017, 5.

45. Smith SL, Kindermans PJ, Ying C, Le QV. Don't decay the learning rate, increase the batch size. arXiv preprint arXiv:1711.00489 2017.

46. Žižka J, Dařena F, Svoboda A, Žižka J, Dařena F, Svoboda A. Adaboost. In Text Mining with Machine Learning; 2019.

47. Prechelt L. Automatic early stopping using cross validation: quantifying the criteria. Neural Networks. 1998;11:761–7.

48. Goutte C, Gaussier E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In; 2010.

49. Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621 2017.