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] and Mohamed Loey [3]

[1] Department of Information Technology, Faculty of Computers and Artificial Intelligence, Cairo University, Cairo 12613, Egypt; nourmahmoud@cu.edu.eg (N.E.M.K.)
[2] Department of Mathematics, University of New Mexico, Gallup Campus, NM 87301, USA; smarand@unm.edu (F.L.)
[3] Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha 13511, Egypt; mloey@fci.bu.edu.eg (M.L.)
*Correspondence Author: nourmahmoud@cu.edu.eg

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 over a limited COVID-19 chest x-ray dataset will be presented. The study relies on neutrosophic set theory, as it shows a huge potential for solving many computers problems related to the detection, and the classification domains. The neutrosophic set in this study is used for converting 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 as there is no benchmark dataset for COVID-19 chest X-ray until the writing of this research. The dataset consists of four classes and they are COVID-19, Normal, Pneumonia bacterial, and Pneumonia virus. The selected deep learning models for investigation 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, 36 experimental trails have been conducted and presented. A combination of training and testing strategies have been applied into dataset by splitting it to (90%-10%, 80%-20%, and 70%-30%) accordingly. According to the experimental results, the Indeterminacy (I) neutrosophic domain achieves the highest accuracy possible 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 higher testing accuracy, especially with limited datasets such as COVID-19 chest x-ray dataset which is investigated throughout this study.

Keywords: Coronavirus; Neutrosophic; deep transfer learning; COVID-19; SARS-CoV-2; CNN;

1. Introduction

Severe acute respiratory syndrome-related coronavirus (SARSr-CoV) is a kind of B-coronavirus that infects bats and some other mammals. In 2002–2004 SARSr-CoV flare-up was an epidemic covering severe acute respiratory syndrome (SARS). The Canton in South China was infected with B-coronavirus that lead to SARSr-CoV. SARSr-CoV was a kind of coronavirus as a family of the B-coronavirus (B- CoV) subgroup and was title as SARSr-CoV [1]. The subclasses of the coronaviruses family are A-coronavirus (A-CoV), B-coronavirus (B-CoV), C-coronavirus (C-CoV), and D-coronavirus (D-CoV) coronavirus [2]. Historically, SARSr-CoV, across 29 countries in the world,

infected over 8000 humans and at least 750 died. The 2019–2020 coronavirus epidemic is an ongoing scourge of coronavirus malady 2019 (COVID-19) created by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) [3]. The International Committee of Viruses and World health organization (WHO) official names the B-coronavirus of 2019 as SARS-CoV-2 and the malady as COVID-19 in the International Classification of Diseases (ICD) [4–6]. However, SARS-CoV-2 infected more than two million of humans with more than 200000 deaths, across 230 states, until the date of this lettering. It elucidates that the propagate rate of SARS-CoV-2 is greater than SRAS-CoV [7,8]. The SARS-CoV-2 transmission has been assured by the World Health Organization (WHO) with evidence of human-to-human conveyance from different cases outside China, namely in Vietnam [9], Italy [10], Germany [11], US [12]. On 26 April 2020, SARS-CoV-2 confirmed more than 2,920,660 cases, 829,075 recovered cases, and 203,622 death cases [13].

The theory of neutrosophic logic was proposed by Smarandache in 1995. Afterward, it has been unified and generalized by its founder in 1999 [14]. Since that date, neutrosophic logic has been used in many computer science fields including pattern recognition [15], image segmentation, and processing [16] and more. It contributes to solving many research and practical real-life problems in a lot of domains such as medicine [17], economics [18], space satellite[19], and agriculture…etc. Neutrosophy leads to a whole family of novel mathematical theories with an overview of not only classical but also fuzzy counterparts [20]. The term neutro-sophy means knowledge of neutral thought and this neutral represents the main difference between fuzzy and intuitionistic fuzzy logic and set [21,22]. Neutrosophic set has the required potentials of being a general framework for uncertainty analysis in data sets[21] and especially with images in the field of Artificial Intelligence and deep learning.

Deep Learning (DL) is a type of Artificial Intelligence (AI) concerned with methods inspired by the functions of people's brain [23]. For the time being, DL is quickly becoming an important method in image/video detection and diagnosis. DL used in medical x-ray/computed tomography diagnoses. DL upgrade medical diagnosis system (MDS) to realize great results, and implementing applicable real-time medical diagnosis system [24,25]. Convolutional Neural Network (ConvNet or CNN) is a mathematical type of DL architectures used originally to recognize and diagnose images. CNN's have masterful unusual success for medicinal image/video diagnoses and detection. In 2012, [26,27] introduced how ConvNets can boost many image/vision databases such as Modified National Institute of Standards and Technology database (MNIST) [28], Arabic digits recognition (ADR) [29], Arabic handwritten characters recognition (AHCR) [30], and big-scale ImageNet [31]. Deep Transfer Learning (DTL) is a CNN architecture that storing learning parameters gained while solving the DL problem and execute DTL to various DL problem. Many DTL models were introduced like VGG [32], Google CNN [33], residual neural network [34], Xception [35], Inception-V3 [36], and densely connected CNN [37].

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. In [38], proposed an early medical diagnosis system for Pneumonia chest X-ray images based on DTL models. In this academic research, an x-ray data [39] containing about 1600 healthy case, 4200 un-healthy pneumonia case. The trial score introduced that VGG DTL networks better than X-ception DTL network with an error rate of 19%. In [40], 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 [39]. 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. In [41], 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-rays data. They used an x-ray database called ChestX-ray14 proposed by Wang et al. [42]. In [43], 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, in [44] 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 [39] database reached a miss-classification error of 3.6%.

The modernity of this research is the outcome as follows: a) the proposed DTL models have end-to-end structure without hand-craft feature extraction and selection techniques. b) We show that neutrosophic is an effective method to generate x-ray data. c) Chest X-ray data are one of the best diagnostic methods for COVID19. d) The DTL architectures outcome high-performance measurement in a limited COVID19 chest x-ray. The rest of the paper is organized as follows. Segment 2 discusses the dataset used in our paper. Segment 3 presents the proposed DTL architectures, while Segment 4 identifies the carried-out results and our discussion. Finally, Segment 5 provides conclusions and directions for additional study.

2. Dataset

The COVID19 x-ray database applied in this paper [45] was introduced by Dr. Joseph Cohen. The Pneumonia [39] Chest X-Ray Images was used to create the introduced database with Cohen Covid-19 dataset. The Cohen dataset was collected from websites like Italian Society of Medical, Radiopaedia web and online publications. The created dataset [46] is organized into four categories normal, pneumonia bacterial, pneumonia virus, and COVID19. The dataset contains 306 x-ray images divided to 69 images for the COVID-19 class, 79 images for the normal class, 79 images for the pneumonia bacterial class, and 79 images for pneumonia virus class. Figure 1 illustrates samples of images used for this research. Figure 1 also illustrates that there is a lot of variation of x-ray image sizes and features that may reflect on the accuracy of the proposed model which will be presented in the next section.

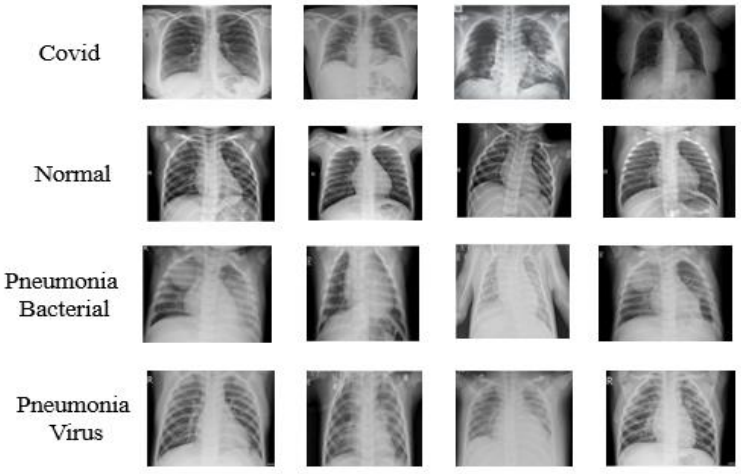

Figure 1. Samples of the used images in this research.

3. The Proposed Model

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 used as a preprocessing step while the DTL architectures used in the training, and the testing steps.

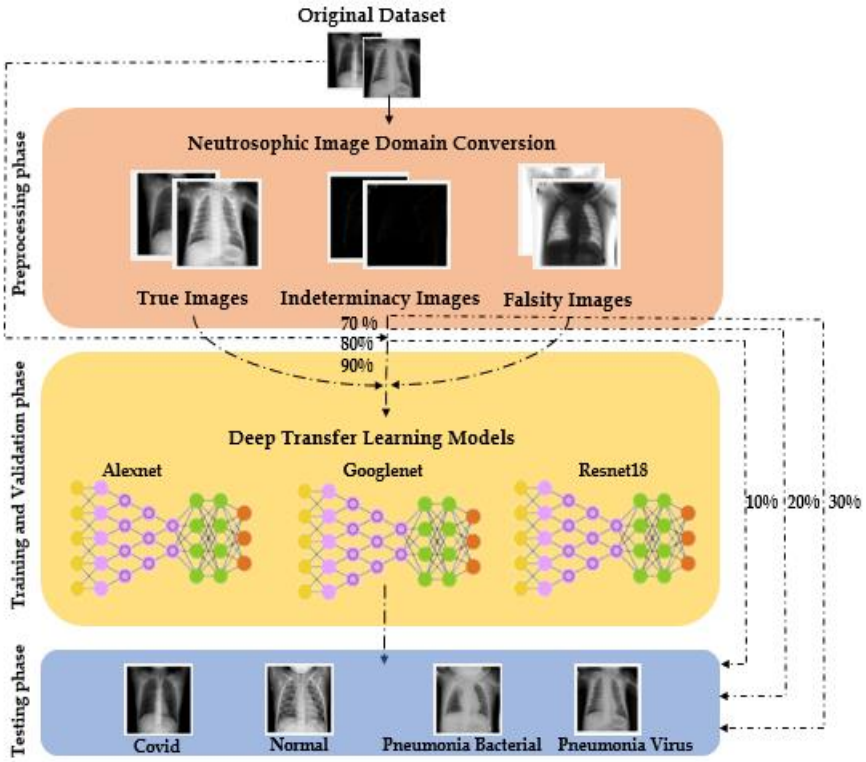

Figure 2. The introduced Neutrosophic /DTL model for the study.

3.1. Neutrosophic Image Domain Conversion

Neutrosophy (NS) is a theory sophisticated and created by Florentin Smarandache [47–49], NS is a useful and successful theory in analyzing uncertain situations. In NS theory, events are analyzed by subset them into three sets as true (T) significance the status is percentage of true, indeterminacy (I) significance the status is percentage of indefinite and 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 equations 1-5. $P(n,m)$ pixel in the image domain is converted to neutrosophic domain $P2NS(n,m)$ [50,51]:

$$P2NS_{NS}(n,m) = \{T_{n,m}, I_{n,m}, F_{n,m}\} \quad (1)$$

$$T_{n,m} = \frac{f(\bar{n},m) - \overline{f_{min}}}{\overline{f_{max}} - \overline{f_{min}}} \quad (2)$$

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

$$I_{n,m} = 1 - \frac{H(n,m) - \overline{H_{min}}}{\overline{H_{max}} - \overline{H_{min}}} \quad (3)$$

$$H(n,m) = abs(\, I(n,m) - I(\bar{n},m)\,) \quad (4)$$

Where $H(n,m)$ is the homogeneity value of T at (n,m), which is measured by the absolute value of the difference between intensity $f(n,m)$ and its local mean value $f(\bar{n},m)$. While $\overline{H_{max}}$ and $\overline{H_{min}}$ are the last and first peaks respectively, measured from the homogeneity image.

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

After the image is converted to the NS domain, the COVID19 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.

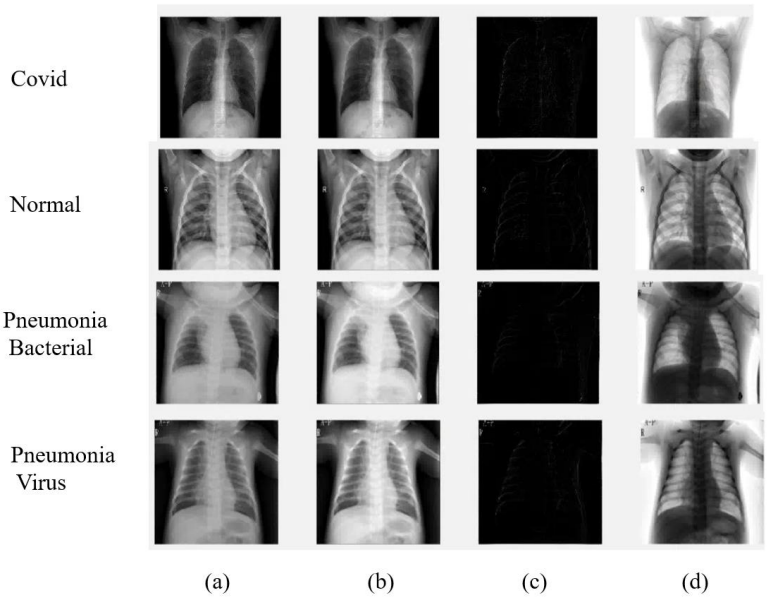

Figure 3. Different neutrosophic images domain for 4 classes in the dataset were (a) original images, (b) True significance, (c) Indeterminacy significance, and (d) Falsity significance images.

3.2. Deep Transfer Learning Model

Algorithm 1 introduces the proposed DTL model in detail. Each DTL model is trained with the COVID-19 x-ray Images database (I, Z); where I the set of N input data, each of size, 512 lengths × 512 widths, and Z have the corresponding label, $Z = \{z/z \in$ {normal; pneumonia bacterial; pneumonia virus; COVID19}}. Let V is a DTL architecture where $V =$ {Alexnet, Googlenet, Resnet18} be the set of CNN models. The x-ray database divided to train and test, training set $(\mathrm{I_{train}}; \mathrm{Z_{train}})$ for 80% for the training and then validation while 10% for the validating proved it is efficient in many types of academic papers [52–56]. The training data then divided into mini-batches, each of size $s = 32$, such that $(I_t;\ Z_t) \in (\mathrm{I_{train}}; \mathrm{Z_{train}});\ t = 1,2, \ldots, \frac{N}{n}$ and iteratively optimizes the CNN model $v \in V$ to reduce the functional loss as illustrated in Equation (6).

$$f_{loss}(w, I_u) = \frac{1}{s} \sum_{i \in I_u, z \in Z_u} f(v(I, w), Z) \tag{6}$$

where $v(I, w)$ is the ConvNet model that true label Z for input I given w is a weight and $f(.)$ is the multi-class entropy loss function.

Algorithm 1 Introduced algorithm.

1: **Input images:** COVID-19 Chest x-ray Images (I, Z); where $Z = \{z/z \in$ {normal; pneumonia bacterial; pneumonia virus; COVID-19}}
2: **Output model:** The DTL model that diagnosed COVID19 x-ray images $i \in I$
3: **Pre-processing steps:**
4: resize input images to dimension 512 height × 512 width
5: Generate neutrosophic x-ray images
6: normalize each image
7: install and reuse DTLmodels D= {Alexnet, Googlenet, Resnet18}

8: Update the last layer by 4 × 1 layer dimension.
9: **foreach** $\forall v \in v$ **do**
10: $\sigma = 0.01$
11: **for** epochs = 1 to 50 **do**
12: **foreach** mini-batch $(I_j;\ Z_j) \in (\mathrm{I}_{\mathrm{train}}; \mathrm{Z}_{\mathrm{train}})$ do
Update the coefficients of the DTL model $v(\cdot)$
if the miss-classification rate is increased for six epochs **then**
$\sigma = \sigma \times 0.01$
end
end
13: **end**
14: **end**
15: **foreach** $\forall i \in I_{test}$ **do**
16: Result of all DTL architectures, $v \in V$
17: **end**

4. Experimental Results

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

- 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 [52,53,55], but the testing accuracy wasn't 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 [56,57]. 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 [58]: 32
- Momentum: 0.9
- Epochs: 60
- Learning Rate [58]: 0.001
- Optimizer [59]: adaboost
- Early stopping [60]: 10 epochs

A large number of trials were performed to draw a full picture of the effectiveness and 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 [61] and they are presented from Equation (7) to Equation (10).

$$\text{Accuracy} = \frac{\text{TP+TN}}{(\text{TP+FP})+(TN+FN)} \quad (7)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP+FN})} \quad (9)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision*Recall}}{(\text{Precision+Recall})} \quad (10)$$

Where TP is the count of True Positive samples, TN is the count of True Negative samples, FP is the count of False Positive samples, and FN is the count 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 experimental results for the different neutrosophic domains. Finally, the third subsection will illustrate a comparative results analysis for the original, the neutrosophic domain according to the confusion matrix for the testing accuracy.

4.1. Original Dataset Experimental Results

As mentioned above different experimental environments were selected in this research for the research experiment. Table 1 presents the testing accuracy and 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 tight 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 1 illustrates interesting facts and they are **1)** The more data the deep learning models have, the higher testing accuracy they will achieve. **2)** The 80%-20%, and 70%-30% strategy achieved very

close results for the testing accuracy which means that those strategies are enough to reflect an accurate testing accuracy for the proposed evaluation model.

4.2. Neutrosophic Domains Experimental Results

The Neutrosophic domains included three types, and they are the True (T) neutrosophic domain, the Indeterminacy (I) neutrosophic domain, 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 section 3.1, 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 model achieves similar highest testing accuracy with 54.84% with an advantage for the Googlenet model in the achieved performance metrics while in 70%-30% strategy, Resnet18 model achieves the highest testing accuracy with 67.74%.

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 2 illustrates interesting facts and they are **1)** In the True (T) neutrosophic domain, more data doesn't mean higher 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 to 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. 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	0.7404	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 3 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 model achieves similar highest testing accuracy with 56.45% with an advantage for the Googlenet model in the achieved performance metrics while in 70%-30% strategy, Googlenet model achieves the highest testing accuracy with 65.59%.

Table 3 also shows interesting facts and they are **1)** In the Falsity (F) neutrosophic domain, more data doesn't mean higher 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 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 to 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 don't 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 70%-30% strategy, both Googlenet and Resnet18 models achieve similar highest testing accuracy with 73.12% with advantage for Googlenet model in the achieved performance metrics.

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**

Table 4 illustrates interesting facts and they are **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 have been revealed and helped the deep transfer models to achieve higher testing accuracy. **3)** More data doesn't mean achieving higher testing accuracy in Indeterminacy (I) neutrosophic domain, as in the 70%-30% strategy, the achieved testing accuracy was better than the achieved accuracy in 80%-20% strategy with 6.99% enhancement using Googlenet.

4.3 Comparative Result of Indeterminacy (I) neutrosophic domain with the original domain

Section 4.2 concluded that the Indeterminacy (I) neutrosophic domain achieved the highest possible testing accuracy in all experiment's trails. This section is dedicated for presenting 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 is a summary for the highest achieved testing accuracy selected from Table 1 and Table 4.

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**

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 also shows interesting facts and they are **1)** in the Indeterminacy (I) neutrosophic or the original domain, Googlenet model is the most dominant model 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 than the achieved testing accuracy in original domain. In the 70%-30% the Indeterminacy (I) neutrosophic domain achieves better accuracy with 10.75% than the original domain. **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 higher testing accuracy.

All the experimental outcomes show that converting to the Indeterminacy (I) neutrosophic domain from the original domain grantees achieving higher testing accuracy. Therefore, 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 Figures 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 testing accuracy 100% and for the normal class with 100%. In the 80%-20% strategy, The Googlenet model was able to detect COVID-19 with testing accuracy 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 testing accuracy 100% and for the normal class with 87.5%.

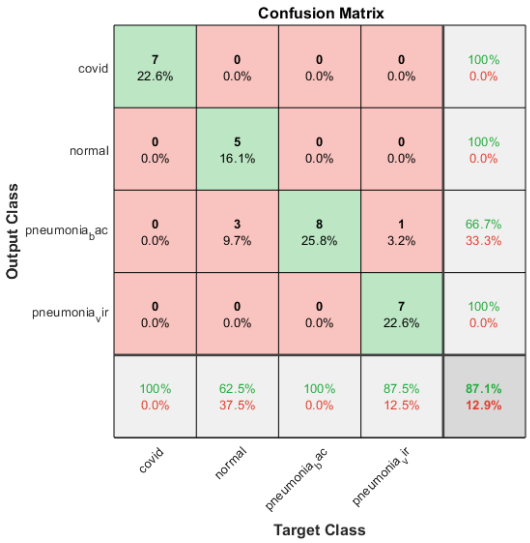

Figure 4. Confusion matrix for Alexnet in 90%-10% strategy for Indeterminacy (I) neutrosophic domain

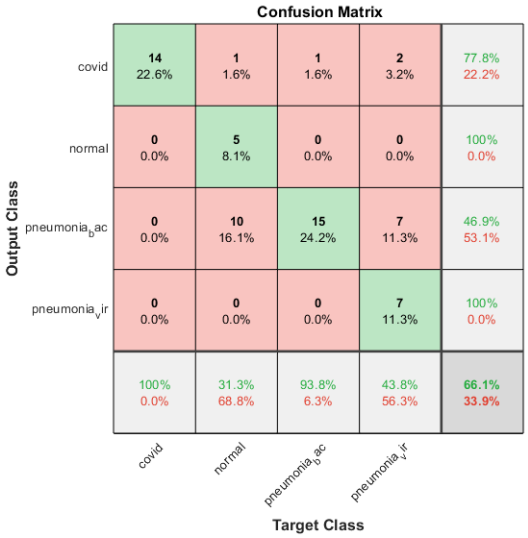

Figure 5. Confusion matrix for Googlenet in 80%-20% strategy for Indeterminacy (I) neutrosophic domain

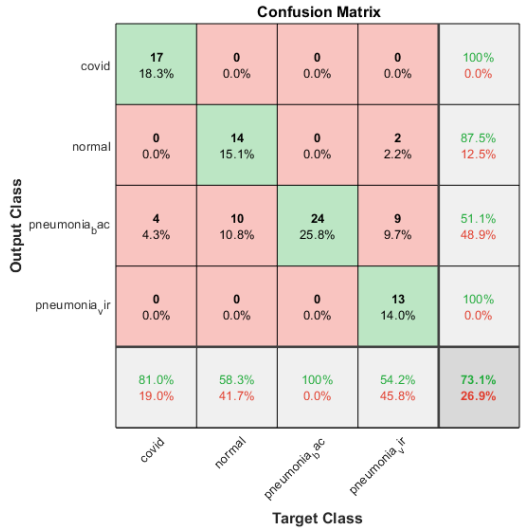

Figure 6. Confusion matrix for Googlenet in 70%-30% strategy for Indeterminacy (I) neutrosophic domain

This study concludes that using the neutrosophic set with deep learning models might be an encouraging transition to achieve higher testing accuracy, especially with limited datasets such as COVID-19 chest x-ray dataset which is investigated throughout this research.

5. 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 are 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, about 36 trails had been conducted and recorded. A combination of training and testing strategies by splitting the dataset into (90%-10%, 80%-20%, and70%-30%) were included in the experiments. Four domains of images are tested, and they were, 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 datasets such as COVID-19 dataset.

Author Contributions: All authors contributed equally to this work. All authors have read and agree to the published version of the manuscript.

Funding: This research received no external funding

Conflict of interest: The author declares no conflict of interest

References

1. Chang, L.; Yan, Y.; Wang, L. Coronavirus Disease 2019: Coronaviruses and Blood Safety. *Transfusion Medicine Reviews* **2020**, doi:https://doi.org/10.1016/j.tmrv.2020.02.003.
2. Wong, A.C.P.; Li, X.; Lau, S.K.P.; Woo, P.C.Y. Global Epidemiology of Bat Coronaviruses. *Viruses* **2019**, *11*, 174, doi:10.3390/v11020174.
3. Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *The Indian Journal of Pediatrics* **2020**, *87*, 281–286, doi:10.1007/s12098-020-03263-6.
4. Lai, C.-C.; Shih, T.-P.; Ko, W.-C.; Tang, H.-J.; Hsueh, P.-R. 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* **2020**, *55*, 105924, doi:https://doi.org/10.1016/j.ijantimicag.2020.105924.
5. Li, J.; Li, J. (Justin); Xie, X.; Cai, X.; Huang, J.; Tian, X.; Zhu, H. Game consumption and the 2019 novel coronavirus. *The Lancet Infectious Diseases* **2020**, *20*, 275–276, doi:10.1016/S1473-3099(20)30063-3.
6. Sharfstein, J.M.; Becker, S.J.; Mello, M.M. Diagnostic Testing for the Novel Coronavirus. *JAMA* **2020**, doi:10.1001/jama.2020.3864.
7. York, A. Novel coronavirus takes flight from bats? *Nature Reviews Microbiology* **2020**, *18*, 191, doi:10.1038/s41579-020-0336-9.
8. Rabi, F.A.; Al Zoubi, M.S.; Kasasbeh, G.A.; Salameh, D.M.; Al-Nasser, A.D. SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far. *Pathogens* **2020**, *9*, 231, doi:10.3390/pathogens9030231.
9. Phan, L.T.; Nguyen, T. V; Luong, Q.C.; Nguyen, T. V; Nguyen, H.T.; Le, H.Q.; Nguyen, T.T.; Cao, T.M.; Pham, Q.D. Importation and Human-to-Human Transmission of a Novel Coronavirus in Vietnam. *New England Journal of Medicine* **2020**, *382*, 872–874, doi:10.1056/NEJMc2001272.

10. Giovanetti, M.; Benvenuto, D.; Angeletti, S.; Ciccozzi, M. The first two cases of 2019-nCoV in Italy: Where they come from? *Journal of Medical Virology* **2020**, *92*, 518–521, doi:10.1002/jmv.25699.
11. Rothe, C.; Schunk, M.; Sothmann, P.; Bretzel, G.; Froeschl, G.; Wallrauch, C.; Zimmer, T.; Thiel, V.; Janke, C.; Guggemos, W.; et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *New England Journal of Medicine* **2020**, *382*, 970–971, doi:10.1056/NEJMc2001468.
12. Holshue, M.L.; DeBolt, C.; Lindquist, S.; Lofy, K.H.; Wiesman, J.; Bruce, H.; Spitters, C.; Ericson, K.; Wilkerson, S.; Tural, A.; et al. First Case of 2019 Novel Coronavirus in the United States. *New England Journal of Medicine* **2020**, *382*, 929–936, doi:10.1056/NEJMoa2001191.
13. Coronavirus (COVID-19) map Available online: https://www.google.com/covid19-map/ (accessed on Apr 26, 2020).
14. Smarandache, F. *A Unifying Field in Logics: Neutrosophic Logic.*; 1999; ISBN 978-1-59973-080-6.
15. Ali, M.; Deli, I.; Smarandache, F. The theory of neutrosophic cubic sets and their applications in pattern recognition. *IFS* **2016**, *30*, 1957–1963, doi:10.3233/IFS-151906.
16. 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.
17. *Neutrosophic Set in Medical Image Analysis*; Elsevier, 2019; ISBN 978-0-12-818148-5.
18. Christianto, V.; Smarandache, F. A Review of Seven Applications of Neutrosophic Logic: In Cultural Psychology, Economics Theorizing, Conflict Resolution, Philosophy of Science, etc. *J* **2019**, *2*, 128–137, doi:10.3390/j2020010.
19. 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.
20. Smarandache, F.; Broumi, S.; Singh, P.K.; Liu, C.; Venkateswara Rao, V.; Yang, H.-L.; 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.
21. 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.
22. 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**.
23. Rong, D.; Xie, L.; Ying, Y. Computer vision detection of foreign objects in walnuts using deep learning. *Computers and Electronics in Agriculture* **2019**, *162*, 1001–1010, doi:https://doi.org/10.1016/j.compag.2019.05.019.
24. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik* **2019**, *29*, 102–127, doi:https://doi.org/10.1016/j.zemedi.2018.11.002.

25. 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, doi:https://doi.org/10.1016/j.zemedi.2018.12.003.
26. Ciregan, D.; Meier, U.; Schmidhuber, J. Multi-column deep neural networks for image classification. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012; pp. 3642–3649.
27. Krizhevsky Ilya and Hinton Alex and Sutskever, G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **2012**, 1097–1105.
28. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **1998**, *86*, 2278–2324, doi:10.1109/5.726791.
29. 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, A.E., Shaalan, K., Gaber, T., Azar, A.T., Tolba, M.F., Eds.; Springer International Publishing: Cham, 2017; pp. 566–575.
30. El-Sawy, A.; Loey, M.; EL-Bakry, H. Arabic Handwritten Characters Recognition Using Convolutional Neural Network. *WSEAS Transactions on Computer Research* **2017**, *5*.
31. Deng, J.; Dong, W.; Socher, R.; Li, L.; Kai, L.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009; pp. 248–255.
32. 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.
33. 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.
34. 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.
35. 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.
36. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition; 2016; pp. 2818–2826.
37. 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.
38. Ayan, E.; Ünver, H.M. 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.

39. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.S.; Liang, H.; Baxter, S.L.; 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, doi:10.1016/j.cell.2018.02.010.
40. Stephen, O.; Sain, M.; Maduh, U.J.; Jeong, D.-U. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *Journal of Healthcare Engineering* **2019**, *2019*, 4180949, doi:10.1155/2019/4180949.
41. 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.
42. Wang, X.; Peng, Y.; Lu, L.; Lu, Z.; Bagheri, M.; Summers, R.M. 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.
43. Islam, S.R.; Maity, S.P.; Ray, A.K.; 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.
44. Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; de Albuquerque, V.H.C. A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images. *Applied Sciences* **2020**, *10*, 559, doi:10.3390/app10020559.
45. Cohen, J.P.; Morrison, P.; Dao, L. COVID-19 Image Data Collection. *arXiv:2003.11597 [cs, eess, q-bio]* **2020**.
46. Dataset Available online: https://drive.google.com/uc?id=1coM7x3378f-Ou2l6Pg2wldaOI7Dntu1a (accessed on Mar 31, 2020).
47. 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.
48. 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.
49. 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.
50. Anter, A.M.; Hassenian, A.E. CT liver tumor segmentation hybrid approach using neutrosophic sets, fast fuzzy c-means and adaptive watershed algorithm. *Artificial Intelligence in Medicine* **2019**, *97*, 105–117, doi:10.1016/j.artmed.2018.11.007.
51. 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, doi:10.1016/j.measurement.2019.07.058.

52. Khalifa, N.; Taha, M.; Hassanien, A.; Mohamed, H. Deep Iris: Deep Learning for Gender Classification Through Iris Patterns. *Acta Informatica Medica* **2019**, *27*, 96, doi:10.5455/aim.2019.27.96-102.
53. Khalifa, N.E.M.; Taha, M.H.N.; Hassanien, A.E.; Hemedan, A.A. Deep bacteria: robust deep learning data augmentation design for limited bacterial colony dataset. *International Journal of Reasoning-based Intelligent Systems* **2019**, doi:10.1504/ijris.2019.102610.
54. Lemley, J.; Bazrafkan, S.; Corcoran, P. Smart Augmentation Learning an Optimal Data Augmentation Strategy. *IEEE Access* **2017**, doi:10.1109/ACCESS.2017.2696121.
55. Khalifa, N.E.M.; Taha, M.H.N.; Ezzat Ali, D.; Slowik, A.; Hassanien, A.E. Artificial Intelligence Technique for Gene Expression by Tumor RNA-Seq Data: A Novel Optimized Deep Learning Approach. *IEEE Access* **2020**, doi:10.1109/access.2020.2970210.
56. Khalifa, N.; Loey, M.; Taha, M.; Mohamed, H. Deep Transfer Learning Models for Medical Diabetic Retinopathy Detection. *Acta Informatica Medica* **2019**, *27*, 327, doi:10.5455/aim.2019.27.327-332.
57. Khalifa, N.E.M.; Loey, M.; Taha, M.H.N. Insect pests recognition based on deep transfer learning models. *Journal of Theoretical and Applied Information Technology* **2020**, *98*, 60–68.
58. Smith, S.L.; Kindermans, P.-J.; Ying, C.; Le, Q.V. Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489* **2017**.
59. Žižka, J.; Dařena, F.; Svoboda, A.; Žižka, J.; Dařena, F.; Svoboda, A. Adaboost. In *Text Mining with Machine Learning*; 2019.
60. Prechelt, L. Automatic early stopping using cross validation: quantifying the criteria. *Neural Networks* **1998**, *11*, 761–767.
61. Goutte, C.; Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In; 2010.