A Novel Approach of CT Images Feature Analysis and Prediction to Screen for Corona Virus Disease (COVID-19)

3 authors, including:

Ahmed Abdullah Farid
Arab Academy for Science, Technology & Maritime Transport
7 PUBLICATIONS **2** CITATIONS

Hatem Khater
Horus university
22 PUBLICATIONS **46** CITATIONS

Some of the authors of this publication are also working on these related projects:

Optimization of Genetic Algorithm For Breast Cancer Detection View project

Comparison case study between clinical and neuroimaging for Alzaheimers disease prediction View project

A Novel Approach of CT Images Feature Analysis and Prediction to Screen for Corona Virus Disease (COVID-19)

Ahmed Abdullah Farid [1, *], Gamal Ibrahim Selim [1,2], and Hatem Awad A. Khater[3]

Abstract: The paper demonstrates the analysis of Corona Virus Disease based on a probabilistic model. It involves a technique for classification and prediction by recognizing typical and diagnostically most important CT images features relating to Corona Virus. The main contributions of the research include predicting the probability of recurrences in no recurrence (first time detection) cases at applying our proposed approach for feature extraction. The combination of the conventional statistical and machine learning tools is applied for feature extraction from CT images through four images filters in combination with proposed composite hybrid feature extraction (CHFS). The selected features were classified by the stack hybrid classification system(SHC). Experimental study with real data demonstrates the feasibility and potential of the proposed approach for the said cause.

Keywords: Deep Learning, Composite Hybrid Feature Selection, Machine learning, Stack Hybrid Classification, CT-Image, MPEG7 edge histogram feature extraction, CNN.

———————————— ◆ ————————————

1. Introduction

Data mining skills involved in biomedical sciences and investigate for providing prediction for help to identify the disease and classify it correctly[1]. Screening large numbers of reported cases for successful isolation and treatment is a priority to control the spread of Corona Virus Disease (COVID-19). Pathogenic laboratory testing is the scientific gold standard but, given significant false-negative results, it is time-consuming. There is an urgent need for quick and accurate diagnosis methods to combat the disease. Based on COVID-19 radiographic improvements in CT scans, we tried to create a deep learning algorithm that could extract the graphical characteristics of COVID-19 to provide a pre-pathogenic clinical diagnosis and thus save critical time for disease control[2-3]. Even before clinical symptoms emerge, computed tomography diagnose irregularities in patients with laboratory-confirmed coronavirus, according to a new case report[51-53]. It is yet another critical piece of evidence showing the central role of the modality in stopping the lethal epidemic. The case, reported in Clinical Imaging on February 22, reports that of a 61-year-old asymptomatic man admitted to a Chinese hospital 1,000 miles outside Wuhan after claiming close contact with an infected person[60]. In addition to identifying early abnormalities, CT showed a result that was not seen in any other COVID-19 instances. As well as those previously diagnosed with standard viral pneumonia and SARS, we obtained 250 CT photographs of pathogen-confirmed COVID-19 events from the Kaggle database web. Our proposed hybrid feature extraction of four filters (MPEG-7 edge histogram filter with Gabor filter- pyramid of rotation-invariant local binary pattern histograms - fuzzy 64-bin histogram), which analyzes a low-level feature of an image can extract the features and provide a statistical hypothesis,[74],[75],[76]. Our proposed model using composite hybrid attribute selection (CHFS) to achieve high accuracy in prediction and improve the feature extraction methods [2-15] with hybrid classification techniques for combine multi-classifiers to improving an in-depth investigation. Testing is employed to see which feature vectors/elements are most informative to differentiate different image classes. Also, using

(CNN) for relatively little pre-processing compared to other image classification algorithms and traditional classifiers. The article planned as follows. The next section discusses the literature review of other authors who have used data mining and its relative of machine learning algorithm to analyze coronavirus. Section 3 describes the proposed technique used for feature extraction from CT images datasets with the CHFS model and four image filters. Section 4 describes the method used for the stack hybrid classification process and convolution neural network (CNN) in comparison with traditional classifiers, whereas section 5 describes the experiments and evaluation. Section 6 discusses the results. Finally, section 7 presents the paper summary and conclusions.

2. Literature Review

 (Fei Shan) developed DL-based segmentation that uses the "VB-Net" neural network to segment CT scans of COVID-19 infection regions. The device is equipped using 249 patients with COVID19 and tested using 300 new patients with COVID-19. A human-in-the-loop (HITL) methodology is introduced to assist radiologists in developing automatic recording of each case to speed up the manual delineation of CT images for analysis. To assess DL-based system performance, the Dice similarity coefficient, volume differences, and infection percentage (POI) are calculated on the validation collection between automatic and manual segmentation checks. EOLBREAK (Xiaowei Xu) observed that in the early stage of identification of viral RNA from sputum or nasopharyngeal swab, the real-time reverse transcription-polymerase chain reaction (RT-PCR) had a relatively low positive rate regarding evaluating COVID-19 (named World Health Organisation). COVID-19 computed tomography (CT) imagery manifestations had their characteristics, which vary from other forms of viral pneumonia, such as viral influenza-A pneumonia. Hospital doctors often lobby for another clear diagnostic criteria for this new type of pneumonia at the earliest possible opportunity.

The accuracy of the chest CT (Yicheng Fang) was higher than that of RT-PCR (98% vs. 71% respectively, p<.001). The explanations for viral nucleic acid detection's low efficiency may include: 1) premature production of nucleic acid detection technology; 2)

$$G_c[i,j] = Be^{-\frac{(i^2+j^2)}{2\sigma^2}} \cos(2\pi f(i\cos\theta + j\sin\theta))$$

$$G_s[i,j] = Ce^{-\frac{(i^2+j^2)}{2\sigma^2}} \sin(2\pi f(i\cos\theta + j\sin\theta))$$

variability in the detection rate from different manufacturers; 3) weak viral load of patients; or 4) inadequate clinical sampling. (Xingzhi Xie, Zheng Zhong) observed that most instances have similar features on CT pictures, such as GGO or combined and merged GGOs. The peripheral spread of 2019-nCoV pneumonia is likely to occur with a longitudinal, multifocal lower lung involvement[6-8]. Given negative RT-PCR samples, CT characteristics of viral pneumonia may be highly suspect for 2019-nCoV infection in the case of regular clinical presentation, and exposure to other persons with 2019-nCoV. In these cases, it is essential to consider repeat swab testing and patient isolation.

Modern image edge detection algorithms include both first-order differential operators (i.e., operators Roberts, Prewitt, Sobel, and Canny) and second-order differential operators (i.e., operators Laplacian which LoG) and can extend to a wide variety of applications. Through integrating this with mathematical mechanics, Wang and Liu used the Roberts operator to identify vehicle image edges and distinguish vehicle license plate positions. While the detection mentioned above algorithms have the benefits of being simple and easy to implement and delivering excellent performance in real-time, they also have obvious shortcomings. The Roberts operator extracted image edge function is relatively rough and offers imprecise edge positions. The edge features which the Prewitt operator extracts have wide margins and many discontinuities. Likewise, the Sobel director does not provide precise picture edge coordinates. The Laplacian operator highly noised sensitive, and the LoG operator can not remove salt and pepper noise in an image.

3. Proposed Work

In the proposed work, the CT images dataset collected from the online access Kaggle benchmark dataset. The medical dataset contains several CT images for Sars and Covid-19 older adults— the proposed layout in Fig 1,2.
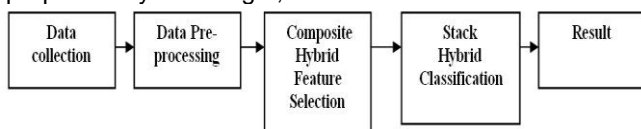

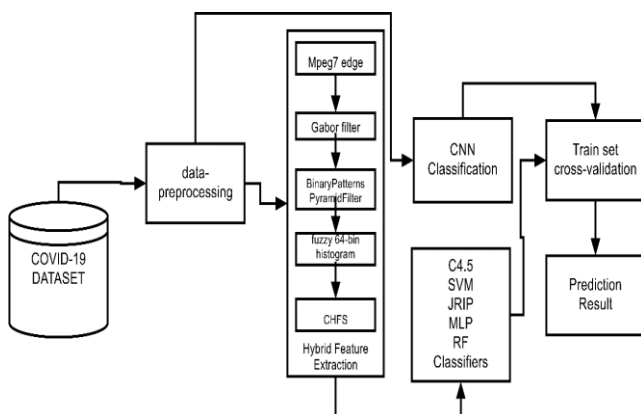Fig. 1 The proposed layout for the clinical dataset

Fig. 2 The proposed feature extraction and classification for covid-19

The following steps explain the mechanism of the proposed work on CT images dataset :

3.1. Data Collection

The CT images dataset has two classes of images both in training as well as the testing set containing a total of around ~51 images each segregated into the severity of Sars and coronavirus (online access Kaggle benchmark dataset,2020),[74],[75],[76]:
i.Covid-19
ii.Sars

3.2. Data Pre-processing

In the real world, data collected tend to be not wholly complete, noisy and conflicting, detection missing of data, data irregularity, prevent the errors, and decrease the data to be analyzed would lead to massive payouts for decision making (H. Witten & Eibe Frank ,2008),[74],[75],[76].

For machine learning and analytics, the selection of features is the process of selecting a subset of specific features (variables, predictors) for use in model building. Component selection methods used for many reasons(H. Witten & Eibe Frank ,2008):
i.simplification of the models to make them easier for researchers/users to understand
ii.shorter training periods
iii.preventing the curse of dimensionality
iv.improved generalization by minimizing overfitting.

3.3.proposed Hybrid Feature Selection Approach on CT-Images

There is a variability of approaches used to obtain images. Some of these are the mean and difference dependent on the bi-orthogonal wavelet filter, image retrieval based on shape. Edges are an essential image feature that is present between an objective and a background and between two targets, two zones, or two primitives. Much of the details for a picture borne in the margins. A picture edge is usually a group of pixels for which the gray-level values demonstrate a shift in phase(Jianfang Cao,2018),"[12],[22]",[74],[75],[76].

3.3.1.MPEG7 Histogram Filter

The histogram is the form most widely used to describe an image's composition of any global function. Translation and rotation of the picture invariant, a normalization of the histogram corresponds to the invariance of the distance. According to the edge histogram description in MPEG-7, an extra histogram bin may easily be created from the local 5-bin edge histogram of each 4 *4 sub-image. A statistical hypothesis testing is employed to see which feature vectors/elements are most informative to differentiate different image classes. The histogram is very useful for indexing and extracting pictures using the above property.

3.3.2 Gabor Image Filter

Gabor filter used for texture analysis, which implies that it mainly analyzes if there is some different frequency information in the picture in specific directions across the point or area of analysis in a regional region. It can define by a sinusoidal wave (a plane wave for 2D Gabor filters) multiplied by a Gaussian function. A set of Gabor filters with different frequencies and orientations may assist in extracting useful features from an image. In the discrete domain, the two-dimensional Gabor filters are given by(2)-(3),

(2)
(3)

3.3.3.Pyramid of Rotation-Invariant Local Binary Pattern Histograms Image Filter

The local binary sequence (LBP) commonly used in the classification of textures. The modern LBP methods define only micro-texture picture structures, such as edges, corners, points, although many of them show excellent texture classification efficiency. This situation could still not be changed, although the technique of multi-resolution research used in local binary pattern methods(Ojala et al., 2002). The texture operator LBP has become a simplified approach in different applications. These can describe as a unifying solution to historically different computational and structural texture research frameworks.

To the LBP operator, the following terminology used: LBPP, Ru2. The subscript represents a neighborhood using the operator (P, R). Superscript u2 indicates that only standard patterns used and the other patterns marked with a single label. After obtaining the LBP-labeled picture fl(x, y), the LBP histogram can identify as

$$Hi = \sum x, yI\{fl(x,y) = i\}, i = 0, \dots, n-1, \quad (4)$$

N is the number of different LBP operators labels, and I{A } is 1 if A is valid and 0 if A is false. When the picture patches with different sizes applied to the histograms, the histograms must standardize to achieve a coherent description:

$$Ni = Hi \sum n - 1j = 0Hj \quad\quad (5)$$

3.3.4 Fuzzy 64-bin Histogram Image Filter

The Fuzzy 64-bin focused on color vision is usually not typically shown in RGB. The better model of HVS is the so-called opposing color type. The competing color space has three component fuzzy 64-bin bases on color expectations typically not better represented in RGB. (C. A. Bouman,2007):

• O1 is a luminance component

• O2 is the red-green channel

• O2 = G − R

• O3 is the blue-yellow channel

• O3 = B − Y = B − (R + G)

Consequences of Opponent Channel CSF Luminance channel is Bandpass function

• Wide bandwidth ⇒ high spatial resolution.

• Low-frequency cut-off is ⇒ insensitive to the average luminance level. Chrominance channels are

• Lowpass function

• Lower bandwidth ⇒ low spatial resolution.

• Low pass ⇒ sensitive to absolute chromaticity (hue and saturation). (C. A. Bouman,2007)

3.3.5. Applying Hybrid feature selection architecture model on CT- images

The author combines four filters to feature extraction process considered for the optimal selection feature set from CT- images, as shown in fig 3.
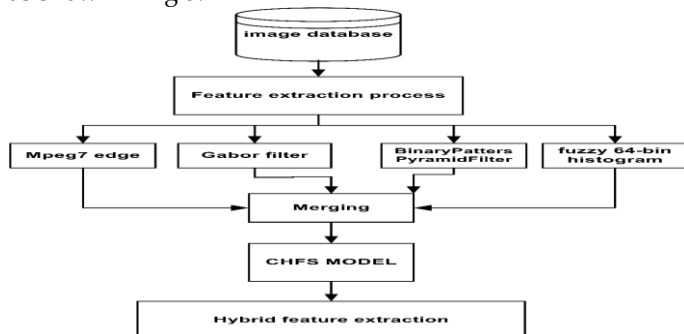

Fig. 3 The proposed hybrid feature selection of CT- images architecture

3.3.6. Proposed Composite Hybrid Feature Selection Model (CHFS)

Feature selection is the approach of taking a subset of relevant features for use in model construction(Chen & ...& F. Li,2010), and combines the advantages of three feature selection approaches (Filter (IG,GR)–Wrapper(improved (Genetic Algorithm)) with Embedded(C4.5)).

3.3.6.1 Composed Hybrid feature selection architecture

The author combine of three feature extraction technique considered for the optimal selection feature set, and this method is information gain (IG) - gain ratio (GR) and Optimized Genetic Algorithm(Huang & C.,2012) (Yanan Mao & Dingyuan Fan,2016)[18-26]. As shown in fig 4.
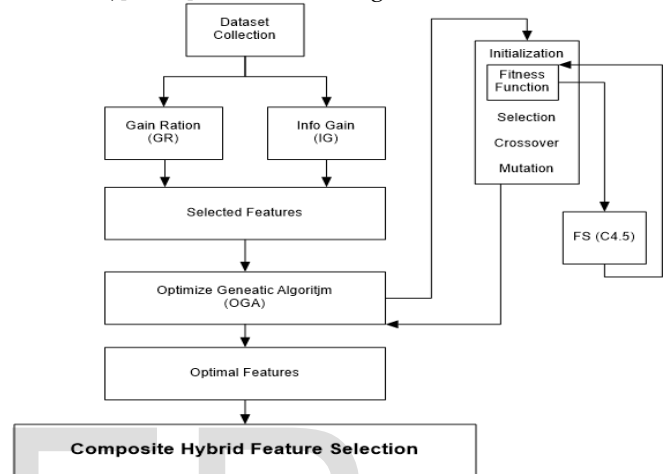

Fig.4 The proposed composite hybrid feature selection architecture

In Figure 4, The superior ranked features taken as the optimized subset, the threshold established the size of the resumption feature subset and amount of attributes in the dataset that appear in the range(44~50) (Vemuri & et al.Jack & C. R.,2008) (Devi &Cuckoo ,2013)[15-22].

3.3.6.2 Information Gain feature selection

The calculation of the information gain for only one attribute according to the algorithm below (Aouatif Amine & ...& Rziza Driss ,2011)[28-32]: This gain measure gives the effect of the features, and the following algorithm selects features that are larger than the threshold,[74],[75],[76].

3.3.6.3 Gain Ratio Feature Selection

A decision tree can be a simple form when non-terminal nodes perform tests on many attributes to the effect of decision outcomes (J.R. Quinlan,1986)[38]. The gain ratio is

$$GainRatio(A) = Gain \frac{A}{} QplitInfoA\,(Q) \quad (6)$$

3.3.6.4 Optimized Genetic Algorithm (OGA)

The authors propose a method to modify a general genetic algorithm to evaluates specified attributes on training data or a separate testing set and uses a decision tree (J.R. Quinlan,1986) [32-38]to estimate the 'merit' of a set of attributes to produce an optimized feature subset with genetic search elevation strategy to recognize the features. All feature selection techniques should use

an evaluation function together with a search strategy to achieve the optimal feature set by(Huang & C.,2012) [74],[75],[76].

It is unable to be realized to search all subsets to find out an optimal subset and need much effort to indicate whether a particular feature is present or not in the chromosome, one, and zero used. One in a gene position refers to feature and zero to absent(Yanan Mao & Dingyuan Fan,2016).The number of features and what are the features that are to be present in a chromosome are guided by information gain (IG) and gain ratio (GR).

The initial population created using input values of IG and GR of the values present in the chromosome. After Generated the population, the individuals evaluated using a fitness function. There is no general approach to find the fitness function for a genetic algorithm. It is a heuristic approach and depends on the used application. So the authors nominate a C4.5 classifier to be used as a fitness function because C4.5 has some utility of handling both continuous and discrete attributes and training data with missing attribute values, pruning trees after creation - C4.5 goes back through the tree once it has been created and try to eject branches that do not help by replacing them with leaf nodes (Dash & H. Liu, 1997) . (J.R. Quinlan,1986) . The following algorithm selects a feature from the set of features that are gained by OGA, gain ratio, and Information gain, as shown in fig 5.
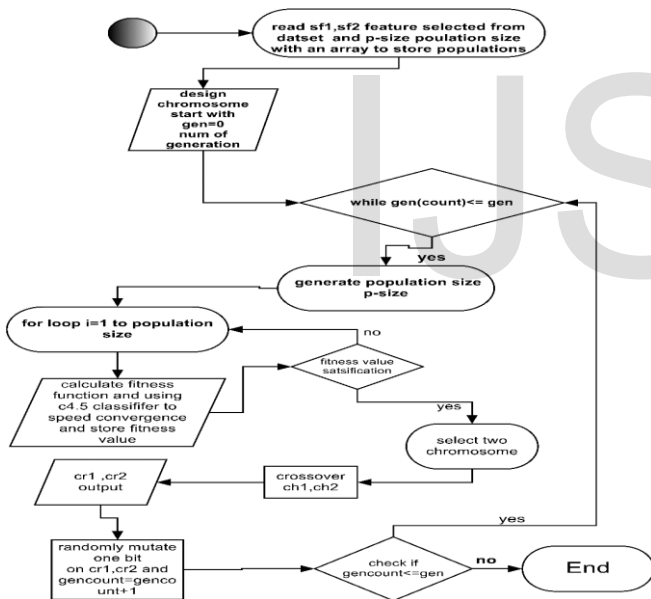

Fig. 5 Proposed optimized genetic algorithm with the c4.5 decision tree

3.4. Proposed Stack Hybrid Classification Model

A weka software tool (Weka online open-source accessed,2018)[16] shows the list of black-box classifiers. These algorithms, in general, are used to classify the medical dataset.

### 3.4.1. Two learning evaluators can be used to evaluate the dataset

- Training set: the classifier separates a dataset to test and training data.
- Cross-validation: in case of 10 fold cross-validation (Divya Jain, Vijendra Singh,2018).

3.4.2. Stacking technique

Ensemble methods are learning methods that contain a set of classifiers for classifying data by taking a weighted point of their predictions (Leo Breiman,1996),"[25],[74],[75],[76]". The authors

combine multiple classifiers to get the maximum efficiency of classification accuracy and overcome the weakness of individual classifiers in the classification process on potential patients. Classifiers, as shown in fig 6.

Fig. 6the proposed framework of stack hybrid classification based on the CHFS model
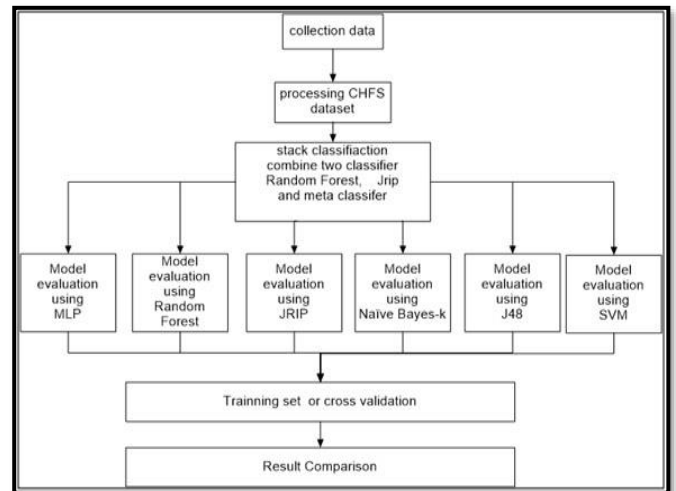
The author chooses permanent (Jrip, RF) based on a result from the below table2, which achieves a good indicator of a fitness function problem.

Weka explorer is for exploring the data(weka,2019)[16]. Using explorer preprocessing can be performed for data. The result-oriented attribute can be selected, and the results can visualize. The experimenter used to understand the learning curve of stack hybrid classification based on (CHFS) and compare results with individual traditional classifiers without (CHFS),[74],[75],[76]. The knowledge flow interface lets the user represent learning algorithms and data sources into the required configuration. It enables the user to specify the data flow by connecting components representing data sources, pre-processing tools learning algorithms evaluation method, and visualization modules,"[55],[75]".

3.4.3. Metrics used in health check systems for evaluation

The different performance metrics generally used to explore the performance of the different models like sensitivity, accuracy, precision, and f-measure (Divya Jain & Vijendra ,2017) . Accuracy: can be calculated by divide number of accurate predictions by the total number of all predications (Klöppel & ...&Frackowiak & R. S,2008)

$$Accuracy = \frac{TN + TP}{fP + tP + fN + tN} \qquad (6)$$

Inside equation
(FN = False Negative, FP = False Positive, TN = True Negative, TP = True Positive)

3.4.5.A Stack, Hybrid Classification Model, Based On Hybrid Feature Selection OF(MPEG7-Gabour filter-binary patterns pyramid – fuzzy 64 bin histogram) on CT-images Dataset

In below figure 7 shows the architecture of all processes from feature extraction output will coming through the four image filter to extract sensitive unseen data to prepare it into the

classification process"[35],[74],[75],[76]". Also, the classification process takes two ways of the test. The first way to consider the CNN approach in analyzing the MRI images and its major for extract and correctly classify the images with less error percentage depends on its algorithms,[74],[75],[76]. The other one we involved the data mining techniques and machine learning classifiers to in-depth into extracted data from four image filter in the feature extraction step. Besides, the comparison between stack hybrid classification with CNN classification[59].
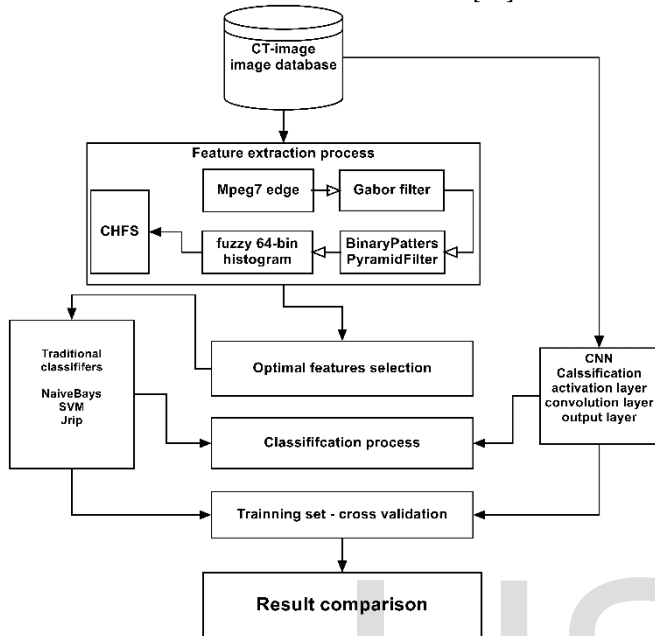
Optimized GA Fitness Fun(C4.5)(OGA)		Embedded	15

Besides using four image filters to extract the features from MRI images, which extract 82 features from MPEG-7 filter and 142 features after applying Gabor filter with 898 features from binary patterns pyramid with final applying of the fuzzy 64-bin histogram to produce total 1473 features from CT-images dataset. Besides, applying our proposed CHFS with optimization of genetic algorithm to get the last 15 sensitive features.

If 10- fold cross-validation applied. For summarized results in the table (2) from the weka software of the analytics dataset, we used the (ROC) curve to evaluate the output result from our proposed architecture model, as shown in fig 8.

Table 2: Experimental result of Accuracy pre-post CHFS model with various classifiers on CT-images images

Classifier	Pre- (CHFS) Feature Selection Accuracy By using four image filter with 1473 feature	Post – (CHFS) Feature Selection Accuracy
SVM	54.93%	84.31%
Naïve Bayes	48.66%	92.15%
JRIP	52.78%	76.47%
Random Forest	61.97%	92.95%

The overall improvement in our experiment on traditional classifiers produce 36.87% after using our proposed feature extraction from CT-images. The figure below shows it summarize for all 3for all results in table 6.
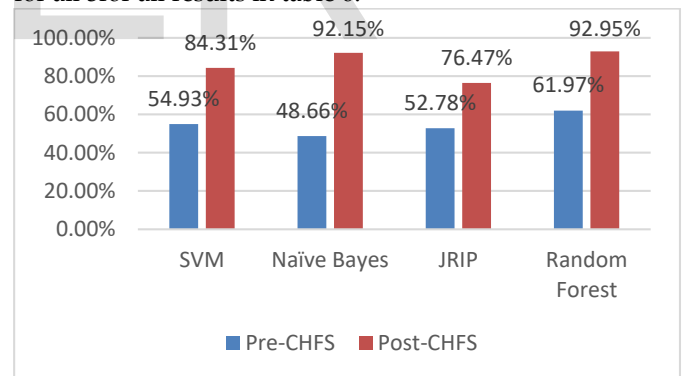

Fig.7 the proposed framework of stack hybrid classification on CT- images

4.Experimental Results and Evaluation

The CSV file of the medical Dataset (clinical test) to patients of Alzheimer's dementia from oasis.org(online access Oasis Medical Dataset,2019) loaded to the weka tool. All experiments evaluated by the receiver operating characteristic (ROC) curve, accuracy, F-measure (Dina A. Ragab & Omneya Attallah,2019).

4.1.Result from proposed (CHFS) feature selection model

The CT-images dataset collected from Kaggle database website with a total of 51 images each segregated into the severity of COVID-19, first we apply four image filter mention in the proposed architecture above in fig 6 on MRI images to extract the mathematical data analysis relation to applying the CHFS model and the result of feature extracted shown below in table1.

Table 1:Number of feature extraction from hybrid-image-filter-CHFS feature selection model

	Method	No.of Features
MPEG-7	Image filter	82
Gabor	Image filter	142
Binary patterns pyramid	Image filter	898
fuzzy 64 bin histogram	Image filter	1473
IG, GR	Filter	2104
GA	Wrapper	289

Fig.8 Summarized accuracy results of pre-post (CHFS) feature selection

4.2.Applying Convolution neural network (CNN) on CT-images of covid-19

The author's aims to make a comparison between the proposed model and convolution neural network (CNN), We tested a set of CT-images of COVID-19 from the kaggle.com benchmark web of dataset science, as shown in below figure18, to perform the accuracy of early-screen diagnosis by the convolution neural network.
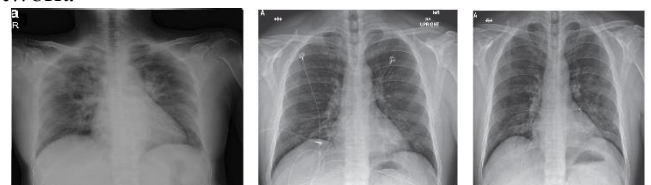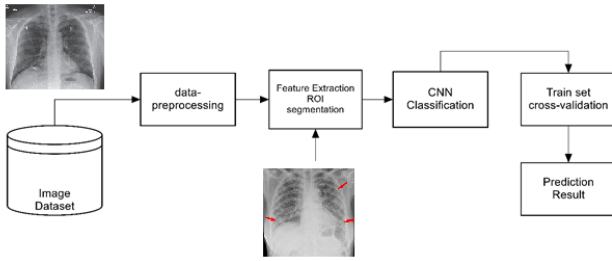
Fig. 9 the types of covid19-ct-images

Fig. 10 shows the prediction model for CNN CT-images

In the below figure 20, we propose the result of ct-images-CNN classification on Kaggle 51 CT-images dataset by using the convolution neural network in weka software data mining, and the result was 94.11% of accuracy, the result evaluated by ROC curve – f-measure


```
=== Confusion Matrix ===

 a  b   <-- classified as
38  1 |  a = COVID-19
 2 10 |  b = SARS
```
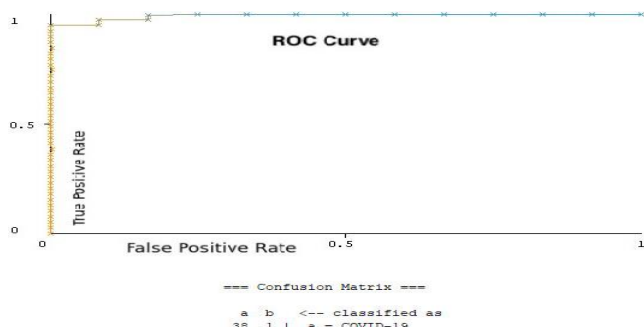
Fig.11 ROC curve of CT-images category and confusion matrix of CNN classification

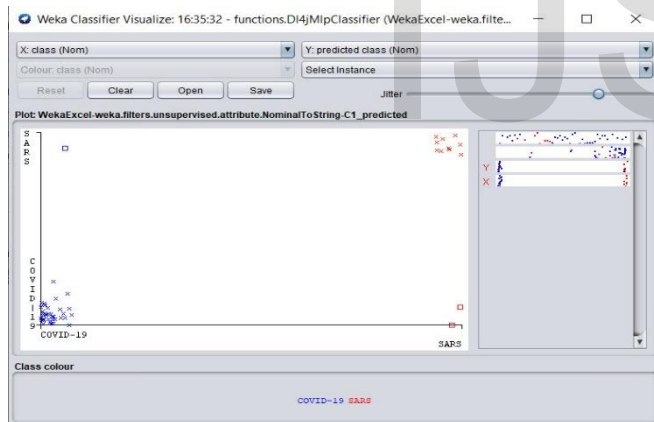And in the below figure 12 shown the error curve of CNN classification.

Fig.12 shown the error curve of CNN classification process

Table3: The details of result from CNN-Model

F-measure	TPR	ROC	PRC
94%	94.1%	99.4%	99.4%

The classification accuracy of the convolution neural network (CNN) was 94.11% and f-measure 94%, which considered least than the classification accuracy of our proposed model.

4.3.Applying a Novel Hybrid classification model on CT-images

The author applying a novel stack hybrid classification as described above in section 3.4.5 on the CT-images dataset to improve the classification accuracy of CT-images by using a traditional algorithm of classifiers but differently as a

combination process of(Jrip, Random Forest)classifiers with stacked meta classifiers of (Naive Bayes-SVM-Jrip-Random Forest). In table 4, the result of the comparison of accuracy between pre-post applying our proposed classification model based on the four image filter with CHFS feature extraction model and in figure 13,14, the ROC curve of all classifiers process with the result of accuracy and f-measure precision-recall-true positive rate-false positive rate for our proposed model as shown in table 4.

Table 4: Experimental result of Accuracy pre-post Stack Hybrid model with various classifiers on 10-fold-cross validation

Classifier	Accuracy Pre- Stack Hybrid (POST-CHFS)	Accuracy Post – Stack Hybrid- (POST-CHFS)
SVM	84.31%	86.27%
Naïve Bayes	92.15%	96.07%
JRIP	76.47%	94.13%
Random Forest	92.95%	86.35%
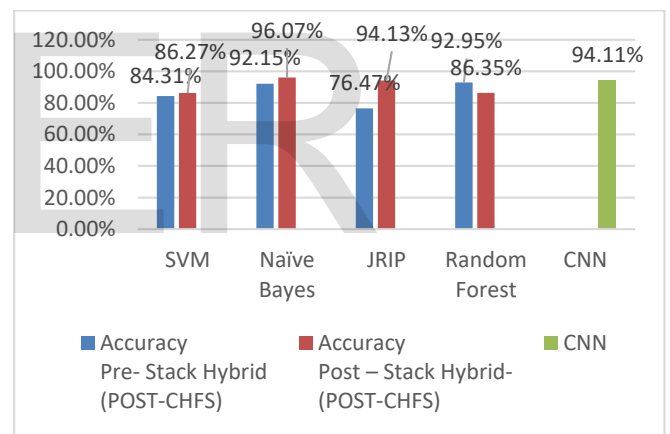
Fig.13 The summerize result of accuracy pre-post proposed Stack Hybrid Classification in comparison with CNN classification
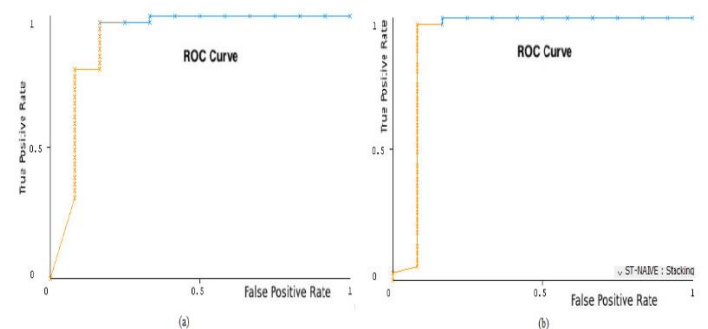

Fig.14 ROC curve for pre-post stack hybrid classifications based on image filter-CHFS(a) ROC curve pre- Hybrid classification model but(after proposed extraction model)(b) ROC curve post- Hybrid classification model

Table 9. The avg. detailed accuracy for CT-images-dataset class of naïve Bayes Stacked hybrid classification

5. Discussion

The authors compare the results of the proposed feature extraction model and stack hybrid classification on covid19-Sars CT-images data in different cases pre-post feature extraction model and pre-post proposed stack hybrid classification and the result compared with CNN model on CT-images dataset. These comparisons, according to our proposed model presented in this study, were reduced a false negative rate and showed a relatively high overall accuracy with more accurate results. The best result extracted from our proposed model was CHFS-Stacked (jrip, RF) with Naïve Bayes classifier accuracy of 96.07% compared to the output result of CNN accuracy of 94.11%. After that, the author applying a novel feature extraction on Ct-images with four image filters later with the CHFS model of optimizing the GA algorithm to best select sensitive features from CT-images. The reduction of

class	TP Rate	FP Rate	Preci-	Recall	F-Meas.	ROC Area	PRC Area
Covid19	0.974	0.083	0.974	0.974	0.974	0.918	0.929
Sars	0.917	0.026	0.917	0.917	0.917	0.918	0.930
Avg.	0.961	0.070	0.961	0.961	0.961	0.918	0.929

selected features from 1473 to 15 features. In the next stage, we applying our proposed stack hybrid classification on 15 features among four different classifiers to obtain the improves in the classification process.

6.Conclusion

In this work, the author aims to early diagnosis of covid19 by using a benchmark CT-Images dataset on our proposed combination of four image filters with composed hybrid feature selection (CHFS) model. That combines the advantages of three filter feature selection approaches and optimizes the Genetic Algorithm (OGA) by improving the initial population generating and genetic operators. Using the results of filter approaches as some prior information with using the J48 decision tree classifier as a fitness function instead of probability and random selection to speed up convergence and select the best features, After that using the selected feature in stack hybrid classification and combine three classifiers with improving the prediction and accuracy. The proposed model shows better than the traditional classification approaches for optimum feature selection and improvement of the classification process and effectively reduced the false-negative rate with high accuracy when using a Naïve Bayes as a meta-classifier in a hybrid classification method with 96.07% compared to 48.66% of usage individually which is considerably better than the previous state-of-the-art result. The results of the proposed model show an accurate classify Ct-Images COVID19. Such findings show the proof-of-principle for using artificial intelligence to derive radiological characteristics for timely and precise diagnosis of COVID-19.

References

[1]Kalló, Gergő, Miklós Emri, et al. "Changes in the Chemical Barrier Composition of Tears in Alzheimer's Disease Reveal Potential Tear Diagnostic Biomarkers." PLoS One, vol. 11, no. 6, Public Library of Science, June 2016, p. e0158000.

[2] J. Escudero, et al. Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease. IEEE transactions on biomedical engineering 2013; 60(1): 164-168.

[3] Jyoti Soni et al. "Predictive Data Mining for medical diagnosis: An Overview of Heart Disease Prediction"International Journal of Computer Applications (0975 –8887), Volume 17, No. 8, March 2011."

[4] Williams, Jennifer A., Alyssa Weakley, Diane J. Cook, and Maureen Schmitter-Edgecombe. (2013) "ML techniques for diagnostic differentiation of mild cognitive impairment." In Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence: 71-76...

[5] Chi, C. L., Oh, W., &Borson, S. (2015) "Feasibility Study of a Machine Learning Approach to Predict Dementia Progression," In Healthcare Informatics (ICHI), International Conference: 450-450

[6] Chyzhyk, D., &Savio, A. (2010) "Feature extraction from structural MRI:

---

[1]Department of Computer Engineering, Faculty of Engineering & Technology, Arab Academy for Science Technology and Maritime Transport(AASTMT), Cairo, Egypt.

[2]Professor Dr. Gamal Selim, (email: dgamal55@yahoo.com or dgamalselim@aast.edu)

[3] Associate Professor Dr. Hatem Awad Aly Khater, Faculty of Engineering, Horus University in Egypt (email: hkhater@horus.edu.eg).

*,[1] Eng.Ahmed Abdullah Farid -Corresponding author (email:

ahmed.abdullah.farid@gmail.com or portal.ahmed@student.aast.edu).

data from OASIS database," University of The Basque Country, Internal research publication.

[7] S. R Bhagya Shree and Dr. H. S. Sheshadri "An initial investigation in the diagnosis of Alzheimer's disease using various classification techniques" IEEE ICCIC, 2014.

[8] Tina R. Patil and Mrs. S. S. Sherekar "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification" International Journal of Computer Science and Applications Vol. 6, No.2, Apr 2013

[9] Jihad Ali et al. "Random Forests and Decision Trees" IJCSI International Journal of Computer IS, Vo. 9, Issue 5, No 3, September 2012

[10] Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., ... & Jack, C. R. (2008) "Alzheimer's disease diagnosis individual subjects using structural MR images: validation studies," Neuroimage, 39(3): 1186-1197.

[11] Datta, P., and Shankle, W. R., &Pazzani, M. (1996) "Applying ML to an Alzheimer's database," In Conference proceedings of the AAAI symposium: 25-27.

[12] Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., ... &Frackowiak, R. S (2008)"Automatic classification of MR scans in Alzheimer's disease." Brain, 131(3): 681-689.

[13] So, A., Hooshyar, D., Park, K. W., & Lim, H. S. (2008) "Early Diagnosis of Dementia from Clinical Data by Machine Learning Techniques," Applied Sciences, 7(7): 651.

[14] Ramirez, J., Górriz, J. M., Salas-Gonzalez, D., Romero, A., López, M., Álvarez, I., & Gómez-Río, M. (2013) "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features." Information Sciences, 237: 59-72.

[15] Sheshadri, H. S., Shree, S. B., & Krishna, M. (2015) "Diagnosis of Alzheimer's Disease Employing Neuropsychological and Classification Techniques," In IT Convergence and Security (ICITCS), 2015 5th International Conference: 1-6.

[16]Alzheimer's clinical medical dataset available online at https://www.oasis-brains.org/#Data (accessed on 2 Jan 2019) OASIS-3: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P50AG00561, P30NS09857781, P01AG026276, P01AG003991, R01AG043434, UL1TR000448, R01EB009352. AV-45 doses provided by Avid Radiopharmaceuticals, a wholly-owned subsidiary of Eli Lilly OASIS-3: https://doi.org/10.1101/2019.12.13.19014902

[17] Data mining: Practical machine learning tools and techniques by Ian H Witten and Eibe Frank, published by Elsevier, second edition 2008.

[18]Aouatif Amine Ali Elakad Mohammed Rziza Driss "GA-SVM and Mutual Information based Frequency Feature Selection for Face Recognition" GSCM-LRIT, Faculty of Sciences, Mohammed V University, B.P. 1014 Rabat, Morocco(2011).

[19] J.R. Quinlan, Induction of Decision Trees, Machine Learning 1: pp.81-106, Kluwer Academic Publishers, Boston, (1986).

[20] Leo Breiman. "Stacked Regressions" Machine Learning, 24, 49-64 (1996)

[21] Duarte Ferreira et al. "Applying data mining techniques to improve diagnosis in neonatal jaundice" BMC Medical Informatics and Decision Making, 0-5, 2012

[22] Jiawei Han, Micheline Kamber, Jian Pei "Data Mining: Concepts and Techniques" published by Elsevier, Third edition, 2012

[23] Anil Rajput, Ramesh Prasad Agarwal, Meghna Dubey S.P. Saxena, and Manmohan Raghuvanshi" J48 and JRIP Rules for EGovernance Data" International Journal of Computer Science and Security (IJCSS), Volume (5): Issue (2), Pp 201, 2011.

[24]Guyon I, Elisseeff A An introduction to variable and feature selection J Mach Learn Res 2003; 3(Mar):1157–82.

[25] Feature selection and classification systems for chronic disease prediction: A review Divya Jain, Vijendra SinghThe NorthCap University, Gurugram 12,2017, India, [online] http://creativecommons.org/licenses/by-nc-nd/4.0

[26] Huang, C., A hybrid stock selection model using genetic algorithms and support vector regression. Applied Soft Computing, 2012. 12(2): p.807-818

[27] K. Tejeswinee, S.G. Jacob. Binary classification of Cognitive Disorders: Investigation on the Effects of Protein Sequence Properties in Alzheimer's and Parkinson's disease. IAENG-IMECS 2017: 166-170

[28] Joliffe IT. Principal component analysis, series: Springer series in statistics. 2. Springer; NY: 2002

[29]O. Elsayed, K. Mahar, M. Kholief and H. A. Khater, "Automatic detection of the pulmonary nodules from CT images," 2015 SAI Intelligent Systems Conference (IntelliSys), IEEE, London, pp.742-746, 2015, DOI: 10.1109/IntelliSys.2015.7361223.

[30]Yasser Fouad, Hatem Khater, Mostafa Setta, Ashraf Alsaid, "proposed Approach for Automatic Underwater Object Classification," International Journal of Research and Surveys, ICIC -EL, Volume 12, No. 12, pp. 1202-1212, 2018, DOI:10.24507/icicel.12.12.1205.

[31] The national center for health statistics report for 2019 (https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm).

[32] Weka open source software for data mining .availble online (https://www.cs.wakito.ac.nz/) accessed 26 Dec 2019.

[33] Chen, F. and F. Li, A combination of feature selection approach with SVM in credit scoring. Expert Systems with Applications, 2010. 37(7): p. 4902-4909.

[34] Chen, Y. -W., and Lin, C. -J., Combining SVMs with various feature selection Strategies, Available from http://www.csie.ntu.edu.tw/~cjlin/papers/features.pdf, 2005.

[35] Dash, M., and H. Liu, Feature Selection for Classification. Intelligent Data Analysis, 1997: p. 131C156.

[36] Devi, K.N., V.M. Bhaskaran, and G.P. Kumar, Cuck

o ,Optimized SVM for Stock Market Prediction, IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded, and Communication Systems (ICJJECS)2015, 2015.

[37] Dai, W., Y.E. Shao, and C. Lu, Incorporating feature selection method into support vector regression for stock index forecasting. Neural Computing and Applications, 2013. 23(6): p. 1551-1561.

[38] Yanan Mao, Zuoquan Zhang, Dingyuan Fan, "Hybrid feature selection based on improved genetic algorithm for stock prediction," 2016 6th International Conference on Digital Home,2016.

[39] Dina A. Ragab, Maha Sharks and Omneya Attallah," Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers," Diagnostics 2019, 9, 165; DOI:10.3390/diagnostics9040165.

[40] David P Salmon and Mark W. Bondi"Neuropsychological Assessment of Dementia" Access NIH public, PubMed central, US national library of medicine Institutes of Health, May 2010.

[41] Divya Jain, Vijendra Singh," Feature selection and classification systems for chronic disease prediction: A review," Egyptian Informatics Journal 19 (2018) 179–189,2018

[42] Ramon Casanova, Fang-Chi Hsu, Kaycee M Sink, Stephen R Rapp, Je_ D Williamson, Susan M Resnick, Mark A Espeland, Alzheimer's Disease Neuroimaging Initiative, et al. Alzheimer's disease risk assessment using large-scale machine learning methods. PloS one, 8(11):e77949, 2013.

[43] Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Cli_ord Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clinics of America, 15(4):869{877, 2005.

[44] Elizabeth G Kehoe, Jonathan P McNulty, Paul G Mullins, and Arun LW Bode.Advances in MRI Biomarkers for the Diagnose of AD. Biomarkers in Medicine, 8(9):1151{1169, 2014.

[45] Arthur W Toga and Karen L Crawford. The AD Neuroimaging informatics core: A decade in review. Alzheimer's and Dementia: The Journal of the Alzheimer's Association, 11(7):832{839, 2015.

[46] Michael W Weiner, Dallas P Veitch, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Jesse Cedarbaum, Robert C Green, Danielle Harvey, Cli_ord R Jack, William Jagust, et al. 2014 Update of the AD Neuroimaging: A review of papers published since its inception. Alzheimer's and Dementia: The Journal of the Alzheimer's Association, 11(6):e1{e120, 2015.

[47]A benchmark Alzheimer's MRI segmentation Dataset from online https://www.kaggle.com/tourist55/alzheimers-dataset-4-class-of-images (accessed online 12/05/2019).

[48]N. Woods, O. Longe, and A. Roberts, "A Sobel edge detection algorithm based system for analyzing and classifying image-based spam," Journal of Emerging Trends in Computing and Information Sciences, vol. 3, no. 4, 2012.View at: Google Scholar.

[49] G.Saianilkumar1, K.Ravindrakbabu2, B.V. Nagendra Prasad Reddy3, "Medical Image Retrieval Based On Edge Histogram Descriptor," International Journal of Engineering Science Invention ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 www.ijesi.org ||Volume 4 Issue 4 || April 2015 || PP.22-27.

[50] Jianfang Cao," Implementing a Parallel Image Edge Detection Algorithm Based on the Otsu-Canny Operator on the Hadoop Platform, "Computational Intelligence and Neuroscience, Volume 2018 |Article ID 3598284 | 12 pages | https://doi.org/10.1155/2018/3598284.

[51] Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019[J]. N Engl J Med. 2020 Jan 24. DOI:10.1056/NEJMoa2001017.

[52]Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia[J]. N Engl J Med. 2020 Jan 29. DOI:10.1056/NEJMoa2001316.

[53] Cohen J, Normile D. New SARS-like virus in China triggers alarm[J]. Science.2020 Jan 17; 367(6475):234-235. DOI: 10.1126/science.367.6475.234.

[54]Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus(2019-nCoV) by real-time RT-PCR[J]. Euro Surveill. 2020 Jan; 25(3). DOI: 10.2807/1560-7917.ES.2020.25.3.2000045.[55]Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019novel coronavirus in Wuhan, China[J]. Lancet. 2020 Jan 24. PII: S0140-6736(20)30183-5. DOI:10.1016/S0140-6736(20)30183-5.

[56]Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster[J]. Lancet. 2020 Jan 24. PII: S0140 6736(20)30154-9. DOI:10.1016/S0140-6736(20)30154-9.

[57]Liu X, Guo S, Yang B, et al. Automatic Organ Segmentation for CT Scans Based on Super-Pixel and Convolutional Neural Networks[J]. Journal of Digital Imaging,2018, 31(6).26 / 29

[58]Gharbi, Michaël, Chen J, Barron J T, et al. Deep Bilateral Learning for Real-Time Image Enhancement[J]. Acm Transactions on Graphics, 2017, 36(4):118.

[59]Hesamian M H, Jia W, He X, et al. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges[J]. Journal of Digital Imaging, 2019,32(8).

[59]Akagi M, Nakamura Y, Higaki T, et al. Correction to Deep learning reconstruction improves the image quality of abdominal ultra-high-resolution CT[J].European Radiology, 2019, 29(8).

[60] Nardelli P, Jimenez-Carretero D, Bermejo-Pelaez D, et al. Pulmonary Artery-Vein Classification in CT Images Using Deep Learning[J]. IEEE Transactions on Medical Imaging, 2018, PP(99):1-1.

[61]Zhu W, Huang Y, Zeng L, et al. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy[J]. medical physics, 2019, 46(2).

[62]Huang P, Park S, Yan R, et al. Added Value of Computer-aided CT Image Features for Early Lung Cancer Diagnosis with Small Pulmonary Nodules: A Matched Case-Control Study. Radiology. 2018 Jan; 286(1):286-295. DOI:1148/radiol.2017162725.

[63] Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with deep three-dimensional learning on low-dose chest computed tomography. Nat Med. 2019 Jun;25(6):954-961. DOI: 10.1038/s41591-019-0447-x.

[65]Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin 27 / 29cancer with deep neural networks. Nature. 2017 Feb 2;542(7639):115-118. DOI:101038/nature21056. Epub 2017 Jan 25.

[66] Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks[J]. Radiology, 2017:162326.

[67]Wu W, Li X, Du P, et al. A Deep Learning System That Generates Quantitative CT Reports for Diagnosing Pulmonary Tuberculosis. arXiv preprint arXiv:1910.02285, 2019.

[68]Li L, Huang H, Jin X. AE-CNN Classification of Pulmonary Tuberculosis Based on CT Images[C]// 2018 9th International Conference on Information Technology in Medicine and Education (ITME). IEEE Computer Society, 2018.

[69]Onisko A, Druzdzel M, Wasyluk H. Learning Bayesian network parameters from small data sets: application of Noisy-OR gates[J]. International Journal of Approximate Reasoning, 2001, 27(2):165-182.

[70]Milletari F, Navab N, Ahmadi S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. arXiv preprint arXiv:1606.04797,2016.

[71]Jeffrey P Kanne. Chest CT Findings in 2019 Novel Coronavirus (2019-nCoV) Infections from Wuhan, China: Key Points for the Radiologist[J].Radiology, 2020.

[72]Chung M, Bernheim A, Mei X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV) [J]. Radiology, 2020.

[73] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J].28 / 292015.

[74] Abdullah Farid, A.; Selim, G.; Khater, H. Applying Artificial Intelligence Techniques for Prediction of Neurodegenerative Disorders: A Comparative Case-Study on Clinical Tests and Neuroimaging Tests with Alzheimer’s Disease. Preprints 2020, 2020030299 (DOI: 10.20944/preprints202003.0299.v1).

[75] Abdullah Farid, A.; Selim, G.; Khater, H. A Composite Hybrid Feature Selection Learning-Based Optimization of Genetic Algorithm For Breast Cancer Detection. Preprints 2020, 2020030298 (DOI: 10.20944/preprints202003.0298.v1).

[76] Abdullah Farid, A.; Selim, G.; Khater, H. Applying Artificial Intelligence Techniques to Improve Clinical Diagnosis of Alzheimer’s Disease. Preprints 2020, 2020030297 (DOI: 10.20944/preprints202003.0297.v1).