Classification of corona virus(COVID -19) using images by Machine Learning Methods

Dr. B. Shadaksharappa
Principal & Professor (CSE)
Sri Sairam College of Engineering,
Anekal, Bengaluru 562106. Karnataka, India.
Email: bichagal@sairamce.edu.in
principal@sairamce.edu.in

P. Ramkumar
Assistant Professor (CSE)
Sri Sairam College of Engineering,
Anekal, Bengaluru 562106. Karnataka, India.
Email: ramkumarkohila@gmail.com
ramkumarp.cse@sairamce.edu.in

Abstract:

Novel Corona virus Disease (NCD) is an extremely dangerous infectious disease with long span of incubation and various clinical demonstrations, which comprises a major impact on life and public health. Despite, the world has reached a very high level of technological applications in medicine, this NCD disease creates a panic among the people in worldwide. In the year 2020 it's been originated in china as an epidemic disease. Afterwards due to migration of people in all over the world it becomes a pandemic disease. Due to this, the mortality ratio is always being increased gradually all over the world everyday in the year 2020. Even though there is a highly curable medicine is available, none of the medicine has supported to cure this dangerous disease. So, this paper proposed a method for classification based on the extracted features. This feature will be extracted by using the Support Vector Machine (SVM). Precision, accuracy and F score metrics have been used to evaluate the performance of the classification.

Keywords
Corona virus, Classification, Feature Extraction, SVM

1. Introduction

Disease can be classified into endemic, epidemic and pandemic. Endemic disease is a disease which prevails throughout a year such as malaria. Epidemic disease is a disease which originates at one area and spreads worldwide either through air or water. This is also called as communicable disease. One of such a dangerous disease caused by the virus in the year 2020 called as corona virus. Corona viruses are the type of viruses which cause severe diseases for human beings, birds and mammals. When any human affected by this type of virus may suffer by fever with high temperature, headache, cough and sneezing issues. Normally these type of diseases caused by the other types of viruses which can be cured. But these symptoms caused by the coronavirus are a critical one and none of the exact medicine has been identified. Even though some of the medical experts suggested to administer hydroxychloroquine and azythromycin tablets for this type of disease, these two tablets are not overhaul the disease. This type of viruses belongs to the sub family of orthocronoviranae and which is an enveloped viruses of nucleocapsid of helical symmetry and positive sense single stranded RNA genome. The size of the genome varies from 27 to 34 kilo bases. The novel coronavirus (COVID-19) is identified in the year 2019 in china. This is a new one which has not been identified previously in human.

This type of viruses can be estimated 5-6 days for the incubation period of 2.2 to 3.6 for the reproduction rate. This rate is higher than the seasonal and pandemic influenza. There is certain prevention methods have been taken to control the spread of this viruses which are not effective to control the virus spread in all over the world.

2. Related Work:

The disease caused by the viruses may increase mortality and morbidity in this world. To combat with this type of disease medical experts have used anti- viral drugs and virus specific vaccine. Anyhow, Broad Spectrum Antiviral Agents (BSAA) can give an enhanced protection of common population from materializing and re- materializing viral diseases to expedite the arsenal of present antiviral options.

In the year 2019 December novel corona virus flu epidemic arises in china, and it has been spread in the entire country rapidly. In the beginning period of the epidemic the country accepted to control the strategy and applied a sequence of core parameters around this considered point which includes common mobilization, increasing case isolation and avoid close contact monitoring management, avoiding epidemic areas and traffic control to diminish personal association and increase social distance with a perspective to control the epidemic as soon as possible in restricted area such as Wuhan.

Kapiriri. Et.al proposed the politics of disease epidemics a Comparative Analysis of the SARS, Zika, and Ebola. In this paper the author addressed about this type of viruses have become increased in all over the world. Various methods of the politics of epidemics were famous in the literature. While marginalized populations were identified to have borne the brunt of all disease outbreaks in the various contexts, the prevalence of their narratives within the reviewed literature was limited. Regardless of income setting, there is a need to provide voice to the most marginalized communities during an epidemic.

Lai.et.al proposed severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. Emergence of severe acute respiratory syndrome corona virus 2 (SARS-CoV-2, previously provisionally named 2019 novel corona virus or 2019-nCoV) diseases (COVID-19) in China at the end of 2019, has caused a large global outbreak and a major public health issue. As of February 11, 2020, data from the WHO has shown that more than 43,000 confirmed cases have been identified in 28 countries/regions, with more than 99% of the cases being detected in China. On January 30, 2020, World Health Organization (WHO) has declared COVID-19 as the sixth public health emergency of international concern.

Luo. Et.al Proposed Can Chinese Medicine Be Used for Prevention of Corona Virus Disease 2019 (COVID-19) A Review of Historical Classics, Research Evidence and Current Prevention Programs. Since December 2019, an outbreak of corona virus disease 2019 (COVID-19) occurred in Wuhan, and rapidly spread to almost all parts of China. This was followed by prevention programs recommending Chinese medicine (CM) for the prevention. In order to provide evidence for CM recommendations, we reviewed ancient classics and human studies. METHODS: Historical records on prevention and treatment of infections in CM classics, clinical evidence of CM on the prevention of severe acute respiratory syndrome (SARS) and H1N1 influenza, and CM prevention programs issued by health authorities in China since the COVID-19 outbreak were retrieved from different databases and websites till 12 February, 2020. Research evidence included data from clinical trials, cohort or other population studies using CM for preventing contagious respiratory virus diseases.

Novel Corona virus Pneumonia Emergency Response Epidemiology, Team Proposed The epidemiological characteristics of an outbreak of 2019 novel corona virus diseases (COVID-19) in China An outbreak of 2019 novel corona virus diseases (COVID-19) in Wuhan, China has spread quickly nationwide. Here, we report results of a descriptive, exploratory analysis of all cases diagnosed as of February 11, 2020. Methods: All COVID-19 cases reported through February 11, 2020 were extracted from China's Infectious Disease Information System. Analyses included: 1) summary of patient characteristics; 2) examination of age distributions and sex ratios; 3) calculation of case fatality and mortality rates; 4) geo-temporal analysis of viral spread; 5) epidemiological curve construction; and 6) subgroup analysis.

3. Statistical features of dataset

Table1 Four Various subsets obtained from patch regions

Subset	Patch Dimension	Number of Corona virus Patches	Number of Non- Corona virus Patches
Subset 1	16x16	6490	5219
Subset 2	32x32	2211	924
Subset 3	48x48	307	552
Subset 4	64x64	106	67

The above table depicts the collected abdominal image from 53 affected cases. The regions of the patches have been cropped with 150 CT images. From the selected regions these patched were extracted.

3.1 Visual features of Dataset

From different CT tools the images in the datasets have been acquired. It would be very difficult for classification. Since, the infected areas of corona bvirus in CT images might be represented by grey level values and the same may represent in non affected areas. The below image represent the non affected areas acquired from various CT tools.
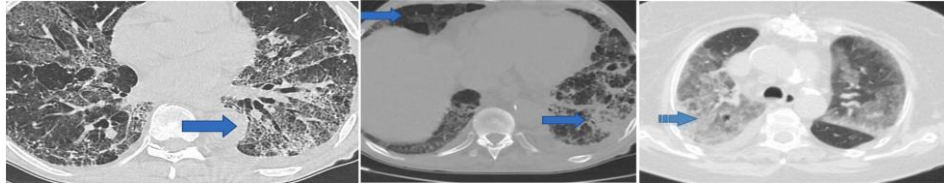

Figure1. The labeled affected areas on CT images.

The above figure reveals that there are certain grey levels in various CT tools. So this would be a drawback for classification. Figure2 demonstrates about patch sample and patch region from four different subsets.
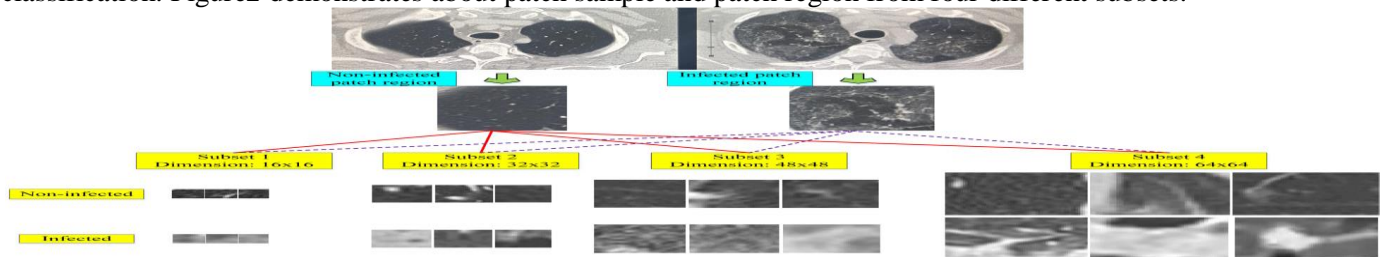

Figure 2. Sample images for affected and non-affected situations for all subsets

4. Proposed Method

The proposed method of classification consists of two stages. First, without feature extraction process, the classification methods have been implemented on four different subsets. The subsets were converted into vector and classified by SVM. Secondly, five different feature extraction methods such as Grey Level Co occurrence Matrix (GLCM), Local Directional Pattern(LDP),Grey Level Run Length Matrix(GLRLM), Grey Level Size Zone Matrix(GLSZM) and Discrete Wavelet Transform(DWT) .

Stage 1
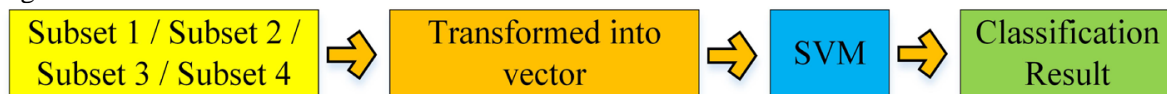
Stage 2
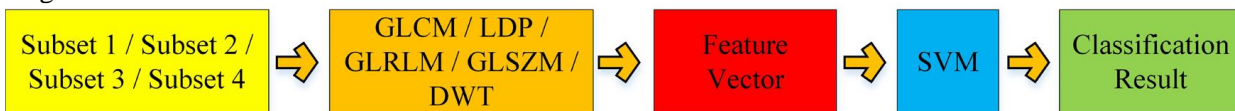
Figure 1. Classification process for Stage 1 and Stage 2

4.1. Feature Extraction Techniques

The feature set have been framed by using DWT, GLSZM, GLRLM, LDP, GLCM are used for corona virus classification. To classify the extracted features SVM classifier method is used. Since, it is a powerful binary classifier. The following feature extraction methods are used in this proposed method.

- ➢ Discrete Wavelet Transform
- ➢ Grey Level Size Zone Matrix
- ➢ Grey Level Run Length Matrix
- ➢ Local Directional Pattern

1826

➢ Grey Level Co-Occurrence Matrix

4.1.1. Discrete Wavelet Transform

This DWT classify the image into frequency sub bands with the help of h low pass filter and g high pass filter. Diagonal details, vertical details, horizontal details, approximation co-efficient represent the high frequency, vertical frequency horizontal high frequency, lowest frequencies in both the directions respectively.

4.1.2. Grey Level Size Zone Matrix

It is a feature extraction method developed by GLRLM algorithm. This method extracts the small zone emphasis, grey-level non-uniformity, long zone emphasis, low grey-level zone emphasis, zone percentage, size zone non-uniformity, small zone low grey-level emphasis, small zone high grey-level emphasis, large zone low grey-level emphasis, grey-level variance as well as size zone variance features from all subsets.

4.1.3. Grey Level Run Length Matrix

This method extracts high level texture features. Let L be the longest Run, R be the number of grey levels, and P be the number of pixels in the image. A matrix in GLRLM is represented as R x L and every p(i,j | θ)elements provides the number of presence in the θ direction the run length of j and I grey level. This method extracts the low grey-level run emphasis, run percentage, run-length non-uniformity, grey-level non-uniformity, long-run emphasis, short-run emphasis and high grey level run emphasis.

4.1.4. Local Directional Pattern

This method is used to combine the directional elements. Let I be an image on (xc, yc) and i_c *be the* density.

$YYYY$Ş $(xxcc, yycc) = \sum_{n=0}^{7} s(in - ic) \cdot 2^n$

This matrix will be converted into a vector for classifier input.

4.1.5. Grey Level Co-Occurrence Matrix

This method is used to derive the second degree statistical features on images. This comprises various angles between the pixels of an image. Let a co-occurrence matrix is calculated from an image and it is represented as *P=[p(i, j | d, Θ)]*. This techniques used to extract *inverse difference* features, maximum probability, cluster prominence, cluster shade, dissimilarity, autocorrelation, information measures of correlation 2, information measures of correlation 1, difference variance, entropy, difference entropy, sum entropy, sum variance, sum average, inverse difference moment, sum of squares: variance, correlation, contrast, *angular secondary moment* from all subsets.

5. Experiment Results

The corona virus has been classified into two stages. Without feature extraction stage 1 sub sets are classified. Feature extraction processes were implemented in stage 2 for all subsets and all the extracted features were classified. This proposed method assessed by using 3 metrics. The following metrics are used Precision (PRE), Accuracy(ACC) and F Score.

$$Precision = TP / (FP+TP)$$
$$Accuracy = (TN +TP) / (TN + TP + FP+FN)$$
$$F\text{-}score = (2*TP)/(2*TP+FN+FP)$$

FN, FP, TN, TP values are represents False Negative, False Positive, True Negative and True Positive respectively.

5.1. Classification Results of Subset 1

Subset 1 contains 6490 affected patches and 5912 non- affected patches. These both patches have been classified with the help of stage 1 and stage 2

The acquired classification results are shown below.

Table 2. Classification Results for Subset 1

Stage	Feature Extraction	Number of Features	Cross Validation	Evaluation Metrics (mean (%) ± std)		
				PRE	ACC	F- Score
Stage1	x	256	2- fold	70.77±0.1	66.67±0.3	68.99

		256	5- fold	73.62±3.5	70.2±3.1	71.35±5
	x	256	10- fold	73.99±7.1	70.62±8.9	72.19±7.8
Stage2	DWT	64	2 fold	98.33±0.3	97.5±0.1	97.26±0.1
	DWT	64	5 fold	98.35±0.2	97.7±0.3	97.47±0.4
	DWT	64	10 fold	98.4±0.5	97.81±0.3	97.6±0.4
	GLSZM	13	2 fold	99.48±0.1	98.54±0.2	98.4±0.3
	GLSZM	13	5 fold	99.62±0.1	98.71±0.3	98.58±0.4
	GLSZM	13	10 fold	99.6±0.2	98.77±0.2	98.65±0.2
	GLRLM	7	2 fold	95.93±0.1	92.99±0.2	95.71±0.1
	GLRLM	7	5 fold	96.34±0.1	93.65±0.5	96.1±0.1
	GLRLM	7	10 fold	96.41±0.6	93.75±1.1	96.2±0.6
	LDP	256	2 fold	48.66±0.9	43.16±0.8	41.94±0.5
	LDP	256	5 fold	49.55±5.4	43.86±6.9	42.57±8.1
	LDP	256	10 fold	48.8±4.8	43.37±5.9	41.14±5.9
	GLCM	19	2 fold	98.04±0.8	99.6±0.6	97.8±0.9
	GLCM	19	5 fold	98.22±1.7	99.61±0.9	97.97±1.9
	GLCM	19	10 fold	98.22±2.2	99.63±1.2	97.98±2.6

The above table expressed that, the best classification results were obtained as 99.6 in 10 fold cross validation of stage 2 using GLSZM feature extraction method.

5.2. Classification Results of Subset 2

Table2. Result for the classification of subset 3

Stage	Feature Extraction	Number of Features	Cross Validation	Evaluation Metrics (mean (%) ± std)		
				PRE	ACC	F- Score
Stage1	x	256	2- fold	75.9±0.4	80.09±0.1	79.1±0.1
	x	256	5- fold	78.07±2.8	81.64±2.2	80.54±2.3
	x	256	10- fold	81.64±1.9	77.87±2.67	80.62±1.9
Stage2	DWT	64	2 fold	99.47±0.9	99.37±0.6	99.31±0.7
	DWT	64	5 fold	99.57±0.4	99.32±0.3	99.25±0.3
	DWT	64	10 fold	99.47±0.2	99.18±0.2	99.09±0.7
	GLSZM	13	2 fold	93.2±0.2	92.54±0.3	91.7±0.4
	GLSZM	13	5 fold	93.29±1.9	92.78±1.1	91.98±1.3
	GLSZM	13	10 fold	93.41±1.1	93.93±1.8	92.69±1.2
	GLRLM	7	2 fold	92.99±0.2	95.93±0.1	95.71±0.1
	GLRLM	7	5 fold	93.65±0.5	96.34±0.1	96.1±0.1
	GLRLM	7	10 fold	93.75±1.1	96.41±0.6	96.2±0.6
	LDP	256	2 fold	46.47±0.4	50.94±0.3	45.05±0.1

LDP	256	5 fold	46.03±1.1	50.71±0.9	43.66±1.8
LDP	256	10 fold	46.09±1.8	50.7±1.5	44.19±2.1
GLCM	19	2 fold	98.48±0.3	98.19±0.2	98.02±0.2
GLCM	19	5 fold	98.96±0.4	98.79±0.3	98.69±0.4
GLCM	19	10 fold	99.1±0.4	98.91±0.2	98.81±0.2

The above table exhibited that 99.37% have been obtained as a best classification result in stage 2 in DWT feature extraction method with 10 fold cross validation.

5.3. Classification Results of Subset3

Table 4. Results for the classification of subset 3

Stage	Feature Extraction	Number of Features	Cross Validation	Evaluation Metrics (mean (%) ± std)		
				PRE	ACC	F- Score
Stage1	x	1024	2- fold	70.77±0.1	66.67±0.3	68.99
	x	1024	5- fold	73.62±3.5	70.2±3.1	71.35±5
	x	1024	10- fold	73.99±7.1	70.62±8.9	72.19±7.8
Stage2	DWT	64	2 fold	99.11±0.3	99.61±0.6	99.01±0.3
	DWT	64	5 fold	99.47±0.5	100	99.41±0.5
	DWT	64	10 fold	99.64±1.1	100	99.6±1.3
	GLSZM	13	2 fold	96.49±2.6	95.9±0.3	95.43±0.4
	GLSZM	13	5 fold	98.08±1.9	96.97±1.5	96.57±1.8
	GLSZM	13	10 fold	98.52±2.6	97.15±2.3	96.76±2.7
	GLRLM	7	2 fold	68.57±1.6	67.73±2.9	59.85±6.1
	GLRLM	7	5 fold	70.54±5.6	69.34±3.2	62.48±4.8
	GLRLM	7	10 fold	70.72±7.6	69.89±5.5	63.22±8
	LDP	256	2 fold	48.66±0.9	43.16±0.8	41.94±0.5
	LDP	256	5 fold	49.55±5.4	43.86±6.9	42.57±8.1
	LDP	256	10 fold	48.8±4.8	43.37±5.9	41.14±5.9
	GLCM	19	2 fold	98.04±0.8	99.6±0.6	97.8±0.9
	GLCM	19	5 fold	98.22±1.7	99.61±0.9	97.97±1.9
	GLCM	19	10 fold	98.22±2.2	99.63±1.2	97.98±2.6

The above table expressed about 99.64% classifications with 10 folds cross validation in stage 2 of DWT feature extraction method.

5.4. Classification results of subset 4

Subset 3 has 107 affected patches and 76 non affected patches . In Stage 1 and Stage 2 methods are used for classifying these patches. Table 5 shows the obtained classification results.

Stage	Feature Extraction	Number of Features	Cross Validation	Evaluation Metrics (mean (%) ± std)		
				PRE	ACC	F- Score

Stage1	x	2304	2- fold	70.77±0.1	66.67±0.3	68.99
	x	2304	5- fold	73.62±3.5	70.2±3.1	71.35±5
	x	2304	10- fold	73.99±7.1	70.62±8.9	72.19±7.8
Stage2	DWT	576	2 fold	99.11±0.3	99.61±0.6	99.01±0.3
	DWT	576	5 fold	99.47±0.5	100	99.41±0.5
	DWT	576	10 fold	99.64±1.1	100	99.6±1.3
	GLSZM	13	2 fold	96.49±2.6	95.9±0.3	95.43±0.4
	GLSZM	13	5 fold	98.08±1.9	96.97±1.5	96.57±1.8
	GLSZM	13	10 fold	98.52±2.6	97.15±2.3	96.76±2.7
	GLRLM	7	2 fold	68.57±1.6	67.73±2.9	59.85±6.1
	GLRLM	7	5 fold	70.54±5.6	69.34±3.2	62.48±4.8
	GLRLM	7	10 fold	70.72±7.6	69.89±5.5	63.22±8
	LDP	2304	2 fold	48.66±0.9	43.16±0.8	41.94±0.5
	LDP	2304	5 fold	49.55±5.4	43.86±6.9	42.57±8.1
	LDP	2304	10 fold	48.8±4.8	43.37±5.9	41.14±5.9
	GLCM	19	2 fold	98.04±0.8	99.6±0.6	97.8±0.9
	GLCM	19	5 fold	98.22±1.7	99.61±0.9	97.97±1.9
	GLCM	19	10 fold	98.22±2.2	99.63±1.2	97.98±2.6

Table 2, Table 3, Tabl;e4, Table5 demonstrated that the best performance have been achieved by extracting features from patches. 90% of classification accuracy has been accomplished in 10 fold cross validation methods for DWT, GLSZM and GLCM. Among all of these methods GLSZM methods by using 10 fold cross validation method have given best classification results.

The following figure expressed about optimum classifier structure

Figure 4. The optimum classifier structure for detection of the affected patches

GLSZM techniques extracts feature vector from the features of the patches. The vector is differentiated by using various structures of SVM and this would be acquired inn training phase. The classification performance of mean is obtained using SVM method of classification

6. Conclusion

The COVID 19 was initially identified in china and then it spread in and around the world. This also has behavior of other pneumonia. This paper presented an various types of images related to corona virus. So Five different feature extraction methods have been used to identify the feature set which classifies the affected patches with high accuracy. The proposed method has given 99.68% of classification accuracy.

References

1. The epidemiological characteristics of an outbreak of 2019 novel corona virus diseases (COVID-19) in China The Novel Corona virus pneumonia Emergency Response Epidemiology Team
2. Biotech start-ups hit by corona virus work stoppages "C&EN Global Enterprise "DOI: 10.1021/cen-09807-buscon3
3. Andersen, Petter I.; Ianevski, Aleksandr; Lysvand "Discovery and development of safe-in-man broad-spectrum antiviral agents" International Journal of Infectious Diseases DOI https://doi.org/10.1016/j.ijid.2020.02.018
4. Bai, S. L.; Wang, J. Y.; Zhou "Analysis of the first cluster of cases in a family of novel corona virus pneumonia in Gansu Province " Zhonghua Yu Fang Yi Xue Za Zhi DOI: 10.3760/cma.j.issn.0253-

9624.2020.0005

5. Burki, Talha "Outbreak of corona virus disease 2019"The Lancet Infectious Diseases DOI: 10.1016/S1473-3099(20)30076-1

6. Chen, W.; Wang, Q.; Li " Early containment strategies and core measures for prevention and control of novel corona virus pneumonia in China" Zhonghua Yu Fang Yi Xue Za Zhi DOI: 10.3760/cma.j.issn.0253-9624.2020.03.003

7. Chen, Yun; Guo, Yao "Structure analysis of the receptor binding of 2019-nCoV" Biochemical and Biophysical Research Communications DOI: https://doi.org/10.1016/j.bbrc.2020.02.071

8. Chen, Y.; Jin, Y. L.; Zhu "The network investigation on knowledge, attitude and practice about Novel corona virus pneumonia of the residents in Anhui Province" Zhonghua Yu Fang Yi Xue Za Zhi DOI: 10.3760/cma.j.issn.0253-9624.2020.0004

9. Chen, Yusha; Pradhan "What are we doing in the dermatology outpatient department amidst the raging of 2019-nCoV? " Journal of the American Academy of Dermatology DOI: https://doi.org/10.1016/j.jaad.2020.02.030

10. Cyranoski, David "When will the corona virus outbreak peak? "Nature DOI: doi:10.1038/d41586-020-00361-5

11. Du, B.; Qiu, H. B.; Zhan" Pharmacotherapeutics for the New Corona virus Pneumonia" Zhonghua Jie He He Hu Xi Za Zhi DOI: 10.3760/cma.j.issn.1001-0939.2020.0012

12. Ena, J.; Wenzel, R. P." A Novel Corona virus Emerges" Rev Clin Esp DOI: 10.1016/j.rce.2020.01.001

13. Kapiriri, Lydia; Ross, Alison" The Politics of Disease Epidemics: a Comparative Analysis of the SARS, Zika, and Ebola diseases" Global Social Welfare DOI: 10.1007/s40609-018-0123-y

14. Lai, Chih-Cheng; Shih "Severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges" International Journal of Antimicrobial Agents DOI: https://doi.org/10.1016/j.ijantimicag.2020.105924

15. Luo, Hui; Tang "Can Chinese Medicine Be Used for Prevention of Corona Virus Disease 2019 (COVID-19)? A Review of Historical Classics, Research Evidence and Current Prevention Programs Chin J Integr Med DOI: 10.1007/s11655-020-3192-6