Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company’s public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Image processing unravels the evolutionary pattern of SARS-CoV-2 against SARS and MERS through position-based pattern recognition

Reza Ahsan a, Mohammad Reza Tahsili b, Faezeh Ebrahimi d, Esmaeil Ebrahimie d, Mansour Ebrahimie d,e

a Department of Computer Engineering, Qom Branch, Islamic Azad University, Qom, Iran
b Department of Biology, School of Basic Sciences, University of Qom, Qom, Iran
c Faculty of Life Sciences and Biotechnology, Department of Microbiology and Microbial Biotechnology, Shahid Beheshti University, Tehran, Iran
d Genomics Research Platform, School of Life Sciences, College of Science, Health and Engineering, La Trobe University, Melbourne, Victoria, 3086, Australia
e School of Animal and Veterinary Sciences, The University of Adelaide, South Australia, 5371, Australia

1. Introduction

The coronavirus 2019 (COVID-19) was first identified and reported in patients with severe respiratory illnesses in Wuhan, China. The virus was a novel member of the coronavirus family, scientifically named Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [1–3]. Since its discovery, more than sixty million infection cases have been reported, including nearly 1.5 million fatalities [4]. The most worrisome features of SARS-CoV-2 are its apparent ability to spread readily, causing severe health problems in high-risk patients and the elderly, and its tendency to genetically mutate and recombine [5–7].

Coronavirus members belong to the subfamily Coronovirinae, the family Coronaviridae, and the Nidovirales order [8]. They are zoonotic pathogens transmitted to humans by direct contact with infected animals; many scientific reports claim SARS-CoV-2 originated in bats and transmitted to humans via intermediate host animals in a seafood market [9–11]. Coronaviruses’ genome is a single-stranded positive-sense RNA (+ssRNA) molecule with a genomic size range between 27–32 kbp, which contains at least six open reading frames (ORFs) [12–14]. The first ORFs (ORF1a/b) encodes a polyprotein1a,b (pp1a, pp1b), while the other ORFs located at the 3’ end encode at least four structural proteins: envelop glycoprotein spike (S), responsible for recognizing host cell receptors, Membrane (M) proteins, responsible for shaping the virions, the Envelope (E) proteins, responsible for virions assembly and release, and the Nucleocapsid (N) proteins, involved in packaging the RNA genome in the virions with roles in pathogenicity as an interferon (IFN) inhibitor [15,16]. In addition to the four main structural proteins, structural and accessory proteins are
species-specific, such as HE protein, 3a/b protein, and 4a/b protein. Upon entrance of the viral genome into the cytoplasm of the target cell, the positive-sense RNA genome translates into two polyproteins 1a,b (pp1a, pp1b), and are then processed into 16 Non-Structural Proteins (NSPs) to form a replication-transcription complex (RTC) that is involved in genome transcription and replication [17,18].

SARS-CoV-2 is the third novel coronavirus to cause a large-scale epidemic or pandemic in the current century after the SARS-CoV (SARS) in 2003 [19] and the Middle East Respiratory Syndrome Coronavirus (MERS) in 2012 [20] defined by WHO. As a large group of viruses with large crown-like peplomers, they are common among many animals and can cause gastrointestinal illnesses and respiratory diseases in humans. Before the SARS-CoV epidemic in 2003, this virus family was not considered deadly in humans and was thought to only cause mild symptoms in immunocompetent individuals, with a chance of lower respiratory illnesses such as pneumonia and bronchitis. However, in 2003, the first pandemics of the 21st century was caused by SARS in China, and MERS was reported in Saudi Arabi [21–23]. With a high fatality rate of 34.4%, MERS was confirmed to have infected approximately 2500 people, including 861 fatalities [24].

Deep learning and neural networks can learn from data that has gained immense attention in the recent SARS-COV-2 outbreak and pandemic. Artificial intelligence and deep learning methods have been used to diagnose and treat SARS-COV-2 by many researchers [25–27]. Recently, we developed a novel, robust, and reliable deep learning method by combining image processing techniques and Convolutional Neural Network (CNN) to analyze raw protein sequences. The proposed method could predict different influenza A virus subtypes by their surface protein sequences with 100% accuracy in our previously published work [28].

In the present study, to search for the cause of different pathogenicity and fatality rates in the three coronaviridae family members, we
converted the genomic and protein sequences of the SARS-COV-2, SARS, and MERS into a polynomial and binary image datasets (with each nucleotide or amino acid position as a variable/feature). Polynomial datasets were analyzed by attribute weighting (or feature selection) models to find the key differences in nucleotide positions in the three virus classes and any possible relations between the virus’s pathogenicity and its nucleotide sequences. Convolution deep neural networks analyzed binary genomic and proteomic datasets to develop a machine-learning algorithm to automatically distinguish between the three virus classes.

2. Materials and methods

This study’s flowchart is presented in Fig. 1. 1261 SARS-COV-2, 4699 SARS, and 136 MERS sequences in FASTA format were downloaded from the NCBI nucleotide website (https://www.ncbi.nlm.nih.gov/nuccore). The sequences were then converted to polynomial datasets (single nucleotide treated as an attribute; dataset contained ~ 30,000 attributes or variable) with the classes of viruses (two or three classes depending on the comparison techniques) set as the target variable.

As the spike (S) protein plays the most important role in receptor recognition and cell membrane fusion, the amino acids’ sequences of three viruses (SARS-COV2 with 2442 sequences, SARS with 2731 sequences and MERS with 349 sequences) were downloaded from NCBI; converted to polynomial datasets and compared with image processing algorithms (as follows).

2.1. Attribute weighting

The polynomial datasets were imported into RapidMiner software (RapidMiner GmbH, Westfalendamm 8744141 Dortmund, Germany). To identify the most important features or attributes (or nucleotide
Fig. 3. The results of seven attribute weighting models applied on the SARS-COV-2 vs. MERS, SARS-COV-2 vs. SARS, and SARS-COV-2 vs. MERS vs. SARS polynomial dataset. Weights closer to 1.0 show the most important variables (nucleotide positions).
position) that differ between SARS-COV-2, SARS, and MERS viruses, the following attribute weighting algorithms were applied to the dataset; as fully described in our previous papers [29,30]:

1. Weight by Information Gain

Attributes’ weights are computed using the information gain algorithm; the higher the feature’s weight, the more relevant or important it is considered.

2. Weight by Information Gain Ratio

This model calculates the weight of attributes by using the information gain ratio; to solve the drawback of information gain.

3. Weight by Rule

The weights are calculated by constructing a single rule for each attribute and calculating the errors. The higher the weight of an attribute, the more relevant it is considered.

4. Weight by Chi-squared statistic

The chi-squared statistic of the nonparametric statistical technique is used to calculate the weights of attributes concerning the class attribute, to determine if a distribution of observed frequencies differs from the theoretically expected frequencies.

5. Weight by Gini Index

The attributes’ weights are calculated for the target or label attribute by computing the Gini index of the class distribution when the dataset is split according to the attribute.

6. Weight by Uncertainty

By measuring the symmetrical uncertainty for target or label attribute, the weight of attributes is calculated. The higher the weight of an attribute, the more relevant it is considered.

7. Weight by Relief

Relief, a simple and effective algorithm, is considered one of the most prominent models for assessing the quality of features. The quality of features is computed based on their values distinguishing between the same and different classes near each other.

2.2. Image processing algorithms and virus classification

Each amino acid or nucleotide position was considered as a variable/feature. As the longest virus sequence comprises 29,999 nucleotides, the same number of features or columns, plus one more for the target variable, were created. Polynomial datasets were converted to binary images as described in our previous paper [28]. Each image’s dimension was set as $29,999 \times 5$ for the genomic sequences (ATCG and N – for not determined nucleotide). For each binary column, only one white pixel was considered. If the DNA sequence matched A, the first column was set as white, and the next four columns were black. For the C sequence, the second pixel, G the third, T the fourth and, N the fifth column was set as a white pixel. Using this method, the length of the binary image equals to the length of the sequence. Polynomial protein sequence datasets were converted to binary images as described in our previous paper [28]. For the amino acid datasets, 20 columns or features (plus one more for the target variable) were created and converted into binary images, based on our abovementioned published paper (Fig. 2A and B).

Binary (image) databases were analyzed by the Convolution deep Neural Networks (CNN), with three convolutions, pooling, and fully connected major layers. Full details of each layer and its functions have been previously described [28]. Here, the convolution layer made of 30,000 neurons (15,000 × 2) was set as the first input layer with a 2 × 12 filter and one stride. Then a Rectified Linear Unit (ReLU) layer was added to zero the resulting matrix’s negative values. To prevent over-fitting, a Max Pooling layer (7500 × 2 window) was placed after the convolutional layer to reduce the data’s size. The Softmax layer takes the Max Pooling layer’s output and converts it into the probability distribution of the classes (a number between 0 and 1, where the sum of the numbers equals one). The fourth layer of Convolution converted each image into a vector of $3750 \times 1$ (with ten filters and step one) and fed it again to another ReLU layer to zero the negative numbers. A fully connected layer with 30 neurons, and another ReLU layer followed by another fully connected layer with 2 or 3 neurons (based on the target feature classes, two or three classes of viruses), a Softmax layer, and finally a classification layer to determine the class of viruses applied. A GABOR filter with a 90-degree angle on horizontal images was applied to the binary image datasets at the preprocessing stage. The designed CNN was trained with 80% of the data, tested with 10%, and evaluated with the remaining 10% of the data with 20 epochs (see Fig. 1).

3. Results and discussion

Seven attribute weighting (or feature selection) models were utilized on 1262 SARS-COV-2, 4699 SARS, and 136 MERS sequence datasets. Each attribute weighting algorithm computes a weight for each variable (about 30,000 variables of viruses’ sequences generated) regarding the target or label variable (a column containing the virus class). The computed weights are normalized into intervals between 0 and 1; weights closer to 1.0 indicate that the target variable (the class of virus) can be realized using said variable. The variables (or nucleotide positions) that cannot differentiate between the target variable (the class of virus) gained weights close to 0.

3.1. SARS-COV-2 vs. MERS nucleotide attribute weighting

Five models out of seven gave weights higher than 0.70 to the 29,617th nucleotide position, while the Gini index, Chi-Squared, and Info gain models gave the weight of 1.0 to this feature. It means that by using this feature, Gini index, Chi-Squared, and Info gain models can distinguish between SARS-COV-2 and MERS viruses. As seen in Fig. 3-A, the other most important nucleotide positions identified by attribute weighting models were all located after the 29,417th nucleotide position.

- **SARS-COV-2 vs. SARS nucleotide attribute weighting**

In addition to the 22,297th and 22,356th nucleotide positions, nine other variables, between the 29,841th and 29,862nd positions, were given the highest possible weight of 1.0 by Rule and Info Gain attribute weighting models (Fig. 3-B).

- **SARS-COV-2 vs. SARS vs. MERS nucleotide attribute weighting**

In addition to the 22,297th and 22,417th nucleotide positions, nine other variables, between the 29,992nd and 29,999th positions, were given the highest possible weight of 1.0 by Rule and Info Gain attribute weighting models (Fig. 3-C).

- **Comparison of the 3’UTR of SARS-COV-2 and SARS**

The results of the attribute weighting models showed that nucleotides after the 29,400th position (which are almost entirely located at the 3’UTR) were the most important features in classifying the
Fig. 4. The results of seven attribute weighting models applied on the 3'UTR sequences of SARS-COV-2 vs. MERS, SARS-COV-2 vs. SARS, and SARS-COV-2 vs. MERS vs. SARS polynomial dataset. Weights closer to 1.0 show the most important variables (nucleotide positions).
above mentioned viruses. Genomic dataset of the 3’UTR of SARS-COV-2 and SARS viruses were extracted from the original Fasta file. All seven attribute weighting models selected the 29,683rd nucleotide as the most important feature for differentiating between the SARS-COV-2 and SARS viruses. Four (out of seven) models gave a weight of 1.0 to this nucleotide position, and three others allocated weights of 0.9, 0.8, and 0.6. As seen in Fig. 4, all important nucleotide positions were located after the 29,830th position (Fig. 4-A).

- **Comparison of the 3’UTR of SARS-COV-2 and MERS**

  The 29,617th and 29,621th positions were the two most important features selected by at least five attribute weighting models with weights higher than 0.7 (Fig. 4-B).

- **Comparison of the 3’UTR of SARS-COV-2, MERS, and SARS**

  When seven attribute weighting models compared the 3’UTR sequences of the three viruses, the only important nucleotide selected as the most important feature by four out of seven models, with weights higher than 0.95, was the 29,683rd nucleotide position. Also, six models gave this variable a weight higher than 0.5 (Fig. 4-C).

  The results of all attribute weighting models on viral polynomial datasets are presented in supplementary tables.

3.2. **Binary image processing of the Spike (S) proteins of SARS-COV2, SARS and MERS**

  The CNN model ran on image datasets of Spike (S) proteins of three classes of viruses (with 2442 SARS-COV-2, 2731 SARS, and 349 MERS sequences). The model accuracy in classifying the three classes of viruses based on their S proteins was just 48%. Other performance criteria such as precision and sensitivity were also less than 0.50 as shown in Fig. 5.

3.3. **Binary image processing of 3’UTR and Nucleoproteins**

  Based on the attribute weighting results, the most prominent differences were found at the end of the genomic positions, corresponding to the viruses’ nucleoprotein sequence. For further investigation, the genomic sequences of 3’UTR and proteomic sequences of nucleoproteins were extracted and subjected to the following image processing technique:

- **Binary image processing of SARS-COV-2 and SARS nucleoproteins**

  Convolution neural network was applied on SARS-COV-2 and SARS viruses’ binary image dataset. The dataset was divided into three parts, training set (80% of the data), test set (10% of the data), and validation set (10% of the data). As seen in the confusion matrix of Fig. 5, the CNN model was able to clearly differentiate between class 1 and class 2 (Fig. 6-A).

- **Binary image processing of SARS-COV-2 and MERS**

  Binary images of SARS-COV-2 and MERS protein sequences were trained, tested, and validated by the CNN model. The model correctly predicted their class from 59 samples in class 1 (SARS-COV-2; in 10% test dataset); the same was true for the 1132 samples of MERS in class 2.

  As seen in the confusion matrix (Fig. 6), the CNN model’s prediction accuracy was 100% in the two virus classes. Even when the trained CNN model was applied on just the validation set, the model accuracy was again 100% (Fig. 6-B).

- **Binary image processing of nucleoprotein of SARS-COV-2, SARS, and MERS**

  The CNN model’s confusion matrix ran on image datasets of three viruses (59 SARS-COV-2, 1132 SARS, and 20 MERS viruses) shown in Fig. 6 with 100% accuracy. The model’s performance for classifying four classes (class 1 = MERS, class 2 = SARS-COV-2, and class 3 = SARS) of viruses were at the highest possible rate (Fig. 6-C).

- **Binary image processing of SARS-COV-2 and SARS 3’UTR**

  SARS-COV-2 and SARS datasets were converted to binary images and analyzed by CNN. Eighty percent of the data were selected as a training set, and the two remaining 10% portions of the data were regarded as test and validation sets. As seen in Fig. 7-A, the CNN model’s accuracy in predicting the right classes of SARS or SARS-COV-2 was 0.99, while its precision and specificity were 98% with FPR of less than 1%.

- **Binary image processing of SARS-COV-2 and MERS 3’UTR**

  The precision, sensitivity, specificity, accuracy, and FPR of the developed CNN model on 80% of binary image datasets of SARS-COV-2 and SARS and tested on 10% test dataset were 100%, 0.93, 100%, 99%, and 0%, respectively (Fig. 7-B).

- **Binary image processing of SARS-COV-2, SARS, and MERS 3’UTR**

  CNN analyzed binary image datasets of SARS-COV-2 and SARS, and the result was shown in Fig. 7-C. The dataset was divided into three parts: training set (80% of the data), test set (10% of the data), and validation set (10% of the data). The model accuracy was 100% in the two virus classes. Even when the trained CNN model was applied on just the validation set, the model accuracy was again 100% (Fig. 7-B).
validation set (10% of the data). The developed model clearly classified all three classes of viruses into the right category with the highest possible accuracy of 100%.

4. Discussion

Convolution neural network combined with image processing techniques has been applied in numerous research areas, including vehicle license plate detection, face recognition, and medical image pattern recognition [31]. Recently, image processing techniques have been used in diagnosing coronavirus patients via chest x-ray images, PET, and SPECT imaging [32,33]. In this study, for the first time ever, we developed and deployed image processing, CNN algorithms, and machine-based prediction tools to differentiate between the three
members of the coronaviridae virus family. Herein, we again confirmed that the combination of deep neural networks and image processing techniques could be successfully employed to accurately classify SARS-COV-2, SARS, and MERS viruses by using their raw genomic and proteomic sequences. Firstly, the models were trained with 80% of the data, tested with 10%, and validated using the remaining 10%. Highly accurate models were used to convert the raw sequences into binary images in merely a few seconds, and the CNN + image processing techniques were subsequently applied to calculate the relation between image features and the virus classes. The raw sequences were directly converted to binary images (not polynomial datasets) and were then processed by machine-based learning algorithms to determine the class of viruses with 100% accuracy.

The models used here have been developed and validated in our

| Actual | Predicted SARS | Predicted COVID-19 | Predicted MERS | Predicted COVID-19 | Predicted SARS |
|--------|----------------|-------------------|----------------|-------------------|----------------|
| SARS   | 46             | 1                 | 13             | 0                 | 13             |
| COVID-19| 1              | 113               | 1              | 127               | 1              |
| FNR    | 0.01           |                   |                |                   |                |
| FNP    | 0.99           |                   |                |                   |                |
| Accuracy| 0.99           |                   |                |                   |                |

**Fig. 7.** Confusion matrix of MERS, SARS-COV-2, and SARS viruses’ 3’UTR sequences analyzed by CNN model, showing the model accuracy, sensitivity, specificity, FNR, FNP and precision in predicting the right class.
In this study, machine-based learning deep CNN models were also capable of classifying the three coronaviridae viruses based on either genomic or proteomic raw sequences with 100% accuracy. Thus, confirming our previous finding that this novel image recognition approach can be used as a simple, fast, and highly accurate tool for the classification of microorganisms. Even when the three coronaviruses were compared with the influenza A virus (unpublished results), the proposed models discriminated the four classes of viruses with the same accuracy (100%), proving this approach’s suitability among different virus families. When a predictive model showed the uppermost achievable accuracy of 100%, the model overfitting should always be checked. We have already excluded any potential of the developed model overfitting, carefully following the scientific approaches to rule-out this possibility (for more details and discussion, please see our previous paper [28]).

The importance of attribute weighting in selecting the most varying and important discriminatory features between classes has been previously reported. The models have been successfully used to find the different features among healthy and non-healthy patients [34,35], thermostable and un-thermostable enzymes [29], and even for subtyping the influenza viruses [35]. Combining this approach with other data mining tools, such as machine learning prediction and meta-analyses, has opened new vistas in bioinformatics classifications [30,36–38]. Herein, we employed the same procedures to find the prominent nucleotide positions that differ among the three classes of the coronaviridae virus family. As each model uses specific statistical methods to define the relation between each feature of the virus classes, it is expected that the outcome weights may be different. Therefore, the average or the sum of weights has proven a suitable classifier among three classes of viruses, which could be due to the evolutionary pattern and conserved nonstructural architecture.

In conclusion, novel image recognition convolution deep learning network algorithms and attribute weighting models were used to trace the structural differences at raw genomic and proteomic sequences for the three coronaviridae family members. The findings confirmed that this approach can easily be used to compare the genome and proteome of microorganisms. The simple, highly accurate, and expedient approach presented here can also be applied to many other biological issues. The structural differences at the end regions of the viruses highlighted by the attribute weighting and image processing models may play significant roles in viral pathogenesis and warrant further investigation.

Declaration of competing interest
The authors declare that they have no conflict of interest.

Acknowledgement
We would like to greatly thank the National Institute of Genetic Engineering and Biotechnology, IRAN for supporting this research and also thank Dr. Jean Winter, Adelaide Medical School, The University of Adelaide for editing of manuscript.

References
[1] G. Antonelli, M.R. Capobianchi, E. Riva, The SARS-CoV-2 epidemic: how the Italian public is being informed, Epipub 2020/04/05, Clin. Microbiol. Infect. 26 (6) (2020) 791–792, https://doi.org/10.1016/j.cmi.2020.03.007. PubMed PMID: 32246996; PubMed Central PMCID: PMCPMC7195308.
[2] Y.M. Bar-On, A. Flamholz, R. Phillips, R. Milo, SARS-CoV-2 (SARS-COV-2) by the numbers, Epipub 2020/04/02, Elife 9 (2020), https://doi.org/10.7554/elife.57309. PubMed PMID: 32288860; PubMed Central PMCID: PMCPMC7224649.
[3] G. Biondi Zoccai, G. Landoni, R. Canevale, E. Cavarretta, S. Sciarrati, G. Frati, SARS-CoV-2 SARS-COV-2: facing the pandemic together as citizens and cardiovascular practitioners, Epipub 2020/03/10, Minerva Cardioangiol. 68 (2) (2020) 61–64, https://doi.org/10.2376/0006-4725.2020.02520-0. PubMed PMID: 32150358.
[4] L.-L. Liang, C.-H. Tseng, H.J. Ho, C.-Y. Wu, SARS-CoV-2 mortality is negatively associated with test numbers and government effectiveness, Sci. Rep. 10 (1) (2020) 1–7.
[5] M. Ferioli, C. Cisternino, V. Leo, L. Pisani, P. Palange, S. Nava, Protecting healthcare workers from SARS-CoV-2 infection: practical indications, Epipub 2020/04/06, Eur. Respir. Rev. 29 (155) (2020), https://doi.org/10.1183/ 16000617.0609.2020. PubMed PMID: 32248146; PubMed Central PMCID: PMCPMC7176410.
[6] K. Hajifathalian, S. Mahadev, R.E. Schwartz, S. Shah, K. Sampath, F. Schnoll-Sussman, et al., SARS-CoV-2 infection (coronavirus disease 2019) for the gastrointestinal consultant, Epipub 2020/04/25, World J. Gastroenterol. 26 (14) (2020) 1546–1553, https://doi.org/10.3748/wg.v26.i14.1546. PubMed PMID: 32237904; PubMed Central PMCID: PMCPMC7116410.
[7] A.Y. Yu, R. Yu, T. Xiao, A. Fan, K. Zhou, J. Huang, A comprehensive Chinese experience against SARS-CoV-2 in ophthalmology, Epipub 2020/04/15, Eye Vis (Lond). 7 (2020) 19, https://doi.org/10.1186/s40662-020-00187-2. PubMed PMID: 32289038; PubMed Central PMCID: PMCPMC7136699.
[8] J. Carneiro Leao, T. Paula de Lima Gusmao, A. Machado Zarrar, J.C. Leao Filho, A. Barbocas Santos de Freitas, M. Moraes Silva, et al., Coronavirus - Old Enemies, New Friends! Oral Dis, Epipub 2020/06/01, 2020, https://doi.org/10.1111/odi.13447. PubMed PMID: 32475006.
[9] S.P. Adhikari, S. Mengers, Y.J. Wu, T.P. Mao, R.X. Ye, Q.Z. Wang, et al., Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (SARS-CoV-2) during the early outbreak period: a scoping review, Epipub 2020/03/19, Infect Dis Poverty 9 (1) (2020) 29, https://doi.org/10.1186/s40249-020-00646-0. PubMed PMID: 32183901; PubMed Central PMCID: PMCPMC7975921.
[10] A. Al-Mandhari, D. Samhouri, A. Abuhakim, R. Brennan, Coronavirus disease 2019 outbreak: preparedness and readiness of countries in the eastern mediterranean region, Epipub 2020/03/07, East. Mediterr. Health J. 26 (2) (2020) 136–137, https://doi.org/10.26719/2020.26.2.136. PubMed PMID: 32141588.
[11] M. Barry, M. Al Amiri, Z.A. Memish, SARS-CoV-2 in the shadows of MERS-CoV in the kingdom of Saudi arabia, Epipub 2020/03/17, J. Epidemiol Glob Health 10 (1) (2020) 1–3, https://doi.org/10.2991/jegh.k200128.003. PubMed PMID: 32175703.
[12] R. Cagliari, D. Forni, M. Clerici, M. Sironi, Coding potential and sequence analysis of coronavirus disease (SARS-COV-2) during the early outbreak period: a scoping review, Epipub 2020/03/04, J. Viral. 94 (12) (2020), https://doi.org/10.1128/JVI.00411-20. PubMed PMID: 32238584.
[13] R. Cagliari, D. Forni, M. Clerici, M. Sironi, Computational inference of selection underlying the evolution of the novel coronavirus, severe Acute respiratory syndrome coronavirus 2, Epipub 2020/04/03, J. Virol. 94 (12) (2020), https://doi.org/10.1128/JVI.00411-20. PubMed PMID: 32238584.
[14] R. Devendra, M. Kumar, S. Chakraborty, Genome analysis of SARS-CoV-2 isolates occurring in India: present scenario, Epipub 2020/06/05, Indian J. Public Health 64 (Supplement) (2020) S147–S155, https://doi.org/10.4103/ijph.IJPH_506_20. PubMed PMID: 32496247.
[15] Y.W. Chen, C.B. Yu, K.Y. Wong, Prediction of the SARS-CoV-2 (2019-nCoV) 3Clike protease (3CL(pro)) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates, Epipub 2020/04/22, F1000Res 9 (2020) 129, https://doi.org/10.26888/f1000research.22457.2. PubMed PMID: 32194494; PubMed Central PMCID: PMCPMC7062264.
[16] S. Kannan, P. Saikl Syed Ali, A. Sheeraz, K. Hemalatha, SARS-COV-2 (Novel Coronavirus 2019) - recent trends, Epipub 2020/03/07, Eur. Rev. Med. Pharmacol.
