



Paper Type: Original-Application Paper



## Forecasting of COVID-19 Sixth Peak in Iran Based on Singular Spectrum Analysis

Morteza Abdolhosseini\* 

Department of Electrical Engineering, Technical and Vocational University, Tehran, Iran; [morteza\\_ab86@yahoo.com](mailto:morteza_ab86@yahoo.com).

**Citation:**



Abdolhosseini, M. (2023). Forecasting of COVID-19 sixth peak in Iran based on singular spectrum analysis. *Journal of decisions and operations research*, 8(1), 123-132.

Received: 05/11/2021

Reviewed: 07/12/2021

Revised: 18/02/2022

Accepted: 22/03/2022

### Abstract

**Purpose:** Coronavirus (COVID-19) is a pandemic that has affected all countries of the world. Forecasting the spread of corona disease will lead to the necessary measures to be taken by the authorities to control this disease. These include increasing vaccinations, quarantining cities and banning entry and exit, increasing the capacity of hospital beds, setting up round-the-clock vaccination centers, requiring the use of masks in public places, and observing social distances. Therefore, predicting such cases will reduce the number of corona cases and therefore reduce the mortality rate.

**Methodology:** In this paper, using the Singular Spectrum Analysis (SSA) algorithm, the sixth peak of coronavirus in Iran is predicted by considering the current situation. To improve the grouping process of the SSA algorithm, eigenvalues have been selected in the optimization process, so that the predicted time series of which has been significantly improved according to the error-index.

**Findings:** Comparing the proposed method with other forecasting methods include Autoregressive Integrated Moving Average (ARIMA), Fractional ARIMA (ARFIMA), TBATS, and Neural Network Autoregression (NNAR), it is observed that the forecasting error is acceptable and the SSA method can be used for forecasting.

**Originality/Value:** This article predicts a new case of COVID-19 using efficient method SSA and the presented results confirm the effectiveness of the proposed method.

**Keywords:** COVID-19, Corona sixth peak, Singular spectrum analysis, Time series forecasting.

Corresponding Author: [morteza\\_ab86@yahoo.com](mailto:morteza_ab86@yahoo.com)

 <http://dorl.net/dor/20.1001.1.25385097.1402.8.1.7.9>



Licensee. **Journal of Decisions and Operations Research**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).



نوع مقاله: پژوهشی-کاربردی

## پیش‌بینی پیک ششم بیماری کووید-۱۹ در ایران بر مبنای آنالیز طیفی منفرد (SSA)

مرتضی عبدالحسینی\*

گروه مهندسی برق، دانشگاه فنی و حرفه‌ای، تهران، ایران.

### چکیده

**هدف:** بیماری کرونا و ویروس (کووید-۱۹) یک بیماری همه‌گیر است که همه کشورهای جهان را درگیر کرده است. پیش‌بینی روند گسترش بیماری کرونا منجر خواهد شد که تدابیر لازم جهت کنترل این بیماری از سوی مسئولین انجام شود. این موارد شامل افزایش واکسیناسیون، قرنطینه کردن شهرها و ممنوعیت ورود و خروج، افزایش ظرفیت تخت‌های بیمارستانی، ایجاد مراکز واکسیناسیون شبانه‌روزی، الزام به استفاده از ماسک در اماکن عمومی و رعایت فواصل اجتماعی به صورت کلی آمادگی لازم جهت برخورد با پیک جدید بیماری کرونا است. لذا پیش‌بینی چنین مواردی سبب کاهش آمار مبتلایان به کرونا و لذا کاهش نرخ مرگ‌ومیر خواهد شد.

**روش‌شناسی پژوهش:** در این مقاله با استفاده از آنالیز طیفی منفرد (SSA)، پیش‌بینی میزان پیک ششم مبتلایان کرونا با لحاظ کردن وضع فعلی صورت گرفته است. به منظور بهبود فرآیند گروه‌بندی الگوریتم SSA، انتخاب مقادیر ویژه به صورت فرآیند بهینه‌سازی صورت گرفته است به طوری که سری زمانی پیش‌بینی شده با توجه به شاخص خطای مدنظر به طور قابل توجهی بهبود یافته است.

**یافته‌ها:** با مقایسه روش پیشنهادی با سایر روش‌های پیش‌بینی شامل میانگین متحرک خود همبسته یکپارچه (ARIMA)، کسری (ARFIMA)، و خود همبسته شبکه عصبی (NNAR)، مشاهده می‌شود که خطای پیش‌بینی به حد قابل قبولی بوده و می‌تواند روش SSA جهت پیش‌بینی مورد استناد قرار گیرد.

**اصالت/ارزش افزوده علمی:** در این مقاله با استفاده از روش کارآمد SSA، موارد مبتلا جدید کرونا و ویروس را پیش‌بینی می‌کند و نتایج ارایه شده اثربخشی روش پیشنهادی را تایید می‌کند.

کلیدواژه‌ها: آنالیز طیفی منفرد، کووید-۱۹، پیش‌بینی سری زمانی، پیک ششم کرونا، SSA.

### ۱- مقدمه

سازمان بهداشت جهانی<sup>۱</sup> در ۱۱ مارس ۲۰۲۰، کووید-۱۹ را یک بیماری همه‌گیر اعلام کرد. شیوع سریع این ویروس مردم بیش از ۲۰۰ کشور را مبتلا کرده است. با شیوع بیماری کرونا از اوایل سال ۲۰۲۰ در جهان، تاکنون نزدیک ۲۴۰ میلیون نفر به این بیماری مبتلا شده و حدود ۵ میلیون نفر جان خود را از دست داده‌اند. در ایران تاکنون شاهد ۵ پیک کرونا بوده‌ایم که حدود ۱۲۳ هزار نفر جان خود را از دست داده‌اند.

<sup>۱</sup> World Health Organization (WHO)

\* نویسنده مسئول

morteza\_ab86@yahoo.com

<http://dorl.net/dor/20.1001.1.25385097.1402.8.1.7.9>





در حال حاضر، تعداد مبتلایان و فوت‌شدگان همچنان در حال افزایش است، با نرخ سرایت بسیار بالا، تقریباً در همه موارد کشورهای آسیب‌دیده این بیماری سلامت انسان را به‌طور جدی تهدید می‌کند و تاثیر قابل‌توجهی در زمینه‌های مختلف مانند توسعه اقتصادی، گردشگری، روابط اجتماعی، سبک زندگی و سیاست بین‌المللی دارد.

اخیراً، مطالعات متعددی برای مدل‌سازی همه‌گیری کووید-۱۹ با استفاده از روش‌های مختلف انجام شده است. به‌عنوان مثال، یک مدل حافظه کوتاه‌مدت برای آموزش داده‌ها با استفاده از الگوریتم  $LSTM-SAE$  به‌عنوان یک مطالعه اولیه در مرجع [1] مورد استفاده قرار گرفت و به‌عنوان پایه‌ای برای آزمایش سایر انواع  $ANN$  عمل کرد. سپس، شبکه‌های رمزگذار خودکار اصلاح‌شده ( $MAE$ ) به‌عنوان مدل نهایی برای پیش‌بینی کووید-۱۹ در برزیل استفاده شده است. همچنین، به‌منظور پیش‌بینی تعداد موارد مثبت گزارش‌شده برای ۳۲ ایالت هند، از مدل‌های مبتنی بر یادگیری عمیق در مرجع [2] استفاده شده است. در مرجع [3]، یک روش تکرار ساده برای پیش‌بینی استفاده شده است که فقط به مقادیر روزانه موارد تاییدشده به‌عنوان ورودی نیاز دارد. در مرجع [4]، ابتدا، مدل‌های افزودنی عمومی ( $GAMs$ ) برای تخمین سه پارامتر وابسته به زمان شامل سرعت انتقال، میزان بازیابی و میزان مرگ‌ومیر ناشی از شیوع کووید-۱۹ در چین و سپس با استفاده از تعداد مبتلایان کووید-۱۹ در ایران، تعداد بیماران در ایران پیش‌بینی شد. یک مطالعه مقایسه‌ای از پنج روش یادگیری عمیق در مرجع [5] برای پیش‌بینی تعداد موارد جدید و موارد بهبودیافته پیشنهاد شده است. الگوریتم‌های شبکه عصبی مکرر  $LSTM$ ،  $LSTM$ ،  $RNN$ ،  $LSTM$  دوطرفه،  $GRU$  و  $VAE$  برای پیش‌بینی جهانی موارد کووید-۱۹ بر اساس داده‌های ایتالیا، اسپانیا، فرانسه، چین، ایالات متحده و استرالیا استفاده شده‌اند. در مرجع [6]، یک مدل ترکیبی شامل تبدیل منحنی دوبعدی، الگوریتم  $CSSA$  و تکنیک یادگیری عمیق برای تعیین بیمار مبتلابه و ویروس کرونا با استفاده از تصاویر اشعه ایکس پیشنهاد داده شده است.

اخیراً تلاش‌های زیادی باهدف پیش‌بینی شیوع بیماری کووید-۱۹ با استفاده از مدل‌های سری زمانی انجام شده است. به‌عنوان مثال، از تکنیک هموارسازی نمایی در مرجع [7] برای پیش‌بینی تجمعی روزانه موارد تاییدشده، مرگ‌ومیر و بهبودیافته از بیماری کووید-۱۹ استفاده شده است. مدل روند خطی و تکنیک‌های هموارسازی نمایی دوگانه در مرجع [8] به‌منظور پیش‌بینی شیوع کووید-۱۹ در مالزی، تایلند و سنگاپور آزمایش شده است. از مدل  $ARIMA$  در مرجع [9] برای پیش‌بینی کل موارد آلوده ایالات متحده، برزیل، هند، روسیه و اسپانیا از ۱۵ فوریه تا ۳۰ جون ۲۰۲۰ استفاده شده است. برای پیش‌بینی موارد جدید تاییدشده روزانه، مرگ‌ومیر و بهبودیافتگان در پاکستان به مدت ده روز، از یک مدل بردار اتورگرسیو در مرجع [10] استفاده شده است. تحلیل سری‌های بیزین در مرجع [11] با استفاده از داده‌های روزانه کووید-۱۹ در ژاپن تا ۳۱ مارس ۲۰۲۰ انجام شده است. یک مدل ترکیبی جدید از تجزیه موجک گسسته و مدل‌های  $ARIMA$  در مرجع [12] توسعه داده شده است تا یک ماه زودتر پیش‌بینی موارد مرگ در ایتالیا، اسپانیا، فرانسه، انگلستان و آمریکا انجام شود.

در سال ۲۰۲۱ نیز چندین مقاله به پیش‌بینی شیوع بیماری کووید-۱۹ پرداخته است. در مرجع [13] با استفاده از روش یادگیری عمیق به پیش‌بینی شیوع بیماری در ایران تا سپتامبر ۲۰۲۱ پرداخته است. پیش‌بینی مبتلایان بیماری کووید-۱۹ در کشورهای ایتالیا، اسپانیا و روسیه با استفاده از سری‌های زمانی و مدل‌های شبکه عصبی در مرجع [14] ارایه شده است. پیش‌بینی کووید-۱۹ در هند با استفاده از مدل  $ARIMA$  در مرجع [15] بیان شده است. در مرجع [16]، با استفاده از تکنیک یادگیری ماشین مدلی تولید می‌کند که به‌طور دقیق تعداد مرگ‌ومیرها و موارد کووید-۱۹ گزارش‌شده در ایالات متحده را تا ۱۰ هفته پیش‌بینی کرده است. در مرجع [17]، یک مدل شبکه عصبی فازی برای پیش‌بینی تعداد موارد مبتلا و مرگ‌ومیر ناشی از این بیماری در کشورهایی که بیشتر با این بیماری درگیر هستند پیشنهاد شده است. تجزیه سری‌های زمانی گزارش‌شده از موارد تایید شده و مرگ‌ومیر را به یک جزو روند، فصلی و نامنظم با استفاده از روش‌های یادگیری ماشینی در مرجع [18] مورد بررسی قرار گرفته است.

باوجود تلاش‌های زیاد برای مدل‌سازی همه‌گیری بیماری کووید-۱۹، در این مقاله از آنالیز طیفی منفرد جهت پیش‌بینی موارد مبتلا جدید به کرونا در ایران در ۵۰ روز آینده استفاده شده است. پیش‌بینی دقیق تعداد افراد مبتلا جدید بسیار مهم است و می‌تواند به مسئولین در مورد تاثیر جهانی و ویروس کرونا و برنامه‌ریزی صحیح در آینده، مانند برآورد تعداد موردنیاز تخت‌های بیمارستانی یا تغییر قوانین فاصله‌گذاری اجتماعی و قرنطینه کمک کند.

الگوریتم  $SSA$  یک روش سریع تحلیل سری‌های زمانی بوده است. این تکنیک غیر پارامتری به‌طور گسترده‌ای در زمینه‌های مختلف مانند پردازش سیگنال، امور مالی، اقتصاد، پردازش تصویر، هواشناسی، مهندسی، پزشکی، زیست‌شناسی و ژنتیک استفاده می‌شود. با توجه به

پتانسیل بالای SSA برای پیش‌بینی داده‌های آینده، این روش می‌تواند یک پیش‌بینی قابل اعتماد برای داده‌های سری زمانی بیماری کووید-۱۹ ارائه دهد و لذا در این مقاله از این روش استفاده شده است.

در این مقاله، از الگوریتم بهینه‌ساز شاهین هریس<sup>۱</sup> مبتنی بر جمعیت و با الهام گرفتن از طبیعت، به منظور بهبود فرآیند گروه‌بندی الگوریتم SSA استفاده شده است. ایده اصلی الگوریتم HHO، رفتار مشارکتی و سبک تعقیب شاهین هریس در طبیعت است که به نام یورش غافلگیرانه شناخته می‌شود. در این استراتژی هوشمند، چندین شاهین به صورت مشارکتی تلاش می‌کنند تا یک طعمه را از جهات مختلف غافلگیر کنند. شاهین هریس می‌تواند الگوهای تعقیب و گریز متنوعی را بر اساس ماهیت پویای سناریوها و الگوهای شکار از خود نشان دهد [14].

هدف از این مطالعه بررسی مزایای بالقوه تحلیل طیفی منفرد برای پیش‌بینی تعداد موارد تاییدشده روزانه، ناشی از کووید-۱۹ است که از متغیرهای اصلی مورد توجه است. نتایج دو رویکرد پیش‌بینی در SSA، یعنی *vector* و پیش‌بینی *recurrent*، با نتایج حاصل از سایر تکنیک‌های پیش‌بینی سری‌های زمانی متداول مقایسه می‌شود. این روش‌ها شامل میانگین متحرک اتورگرسیو (ARIMA)، کسری (ARFIMA)، هموارسازی نمایی (TBATS) و شبکه عصبی اتورگرسیو (NNAR) است. در نهایت، بهترین مدل پیش‌بینی بر اساس اندازه‌گیری دقت<sup>۲</sup> انتخاب می‌شود و برای پیش‌بینی ۵۰ روز آینده اعمال می‌شود. این پیش‌بینی‌ها می‌تواند در پیش‌بینی رفتارهای آینده این بیماری و تصمیم‌گیری بهتر کمک کند.

این مقاله از چند جهت به ادبیات مربوط به پیش‌بینی همه‌گیری کووید-۱۹ کمک می‌کند:

۱. نسخه بهینه روش‌های پیش‌بینی *vector SSA* و *recurrent SSA*، برای پیش‌بینی تعداد موارد تاییدشده روزانه ناشی از بیماری کووید-۱۹ استفاده شده است. به منظور بهبود فرآیند گروه‌بندی الگوریتم SSA، انتخاب مقادیر ویژه به صورت فرآیند بهینه‌سازی صورت گرفته است به طوری که سری زمانی پیش‌بینی شده با توجه به شاخص خطای مدنظر به طور قابل توجهی بهبود یافته است.
۲. به منظور ارزیابی پتانسیل SSA برای پیش‌بینی افراد مبتلابه جدید، عملکرد SSA را با سایر تکنیک‌های پیش‌بینی سری‌های زمانی متداول از جمله *ARFIMA*، *ARIMA*، هموارسازی نمایی، *TBATS* و *NNAR* مقایسه شده است.
۳. بهترین مدل پیش‌بینی بر اساس اندازه‌گیری دقت خطای میانگین مربع ریشه (RMSE) انتخاب می‌شود که یک معیار متداول در پیش‌بینی سری‌های زمانی است و برای پیش‌بینی ۵۰ روز آینده کاربرد دارد. این پیش‌بینی‌ها ممکن است به دولت و سایر سازمان‌ها کمک کند تا استراتژی‌های خود را تغییر داده و منابع موجود را مطابق وضعیت پیش‌بینی شده بهینه کنند.

ادامه این مقاله به شرح زیر تنظیم شده است. در بخش دوم روش SSA معرفی شده است و در بخش سوم شبیه‌سازی و نتایج ارائه شده است. در بخش چهارم و پنجم نیز به ترتیب نتیجه‌گیری و مراجع بیان شده است.

## ۲- روش SSA

سری زمانی غیر صفر با مقادیر مثبت  $Y_T = (y_1, \dots, y_T)$  را با طول کافی  $T$  در نظر بگیرید. هدف اصلی SSA، تجزیه سری زمانی اولیه به مجموع چندین سری زمانی است که آن سری‌های زمانی عبارتند از ترند + مولفه‌های پرودیکی و شبه پرودیکی + نویز. سپس با حذف نویز و با استفاده از مولفه ترند و مولفه‌های پرودیکی و شبه پرودیکی به بازسازی سری زمانی پرداخته می‌شود. تکنیک SSA شامل دو مرحله تجزیه و بازسازی است که هر دو آن‌ها شامل دو مرحله مجزا می‌باشند. در مرحله اول سری زمانی تجزیه شده و در مرحله دوم بازسازی سری زمانی انجام می‌شود و از سری بازسازی شده (بدون نویز) برای پیش‌بینی استفاده شده است. شکل ۱ روند الگوریتم SSA را نشان می‌دهد. در این شکل تمامی بخش‌های الگوریتم پیشنهادی نشان داده شده که در ادامه به تفصیل توضیح داده خواهد شد.

<sup>1</sup> Harris Hawks Optimization (HHO)

<sup>2</sup> Root Mean Squared Error (RMSE)





شکل ۱- روند الگوریتم SSA.  
Figure 1- Process of SSA algorithm.

### ۲-۱- قدم اول در تجزیه: تعبیه

سری زمانی با طول پنجره  $L$  را در نظر بگیرید. بردارهای پس فاز  $X_i = (y_{i-1}, \dots, y_{i+L-2})^T, (i = 1, \dots, K = T - L + 1)$  را بازسازی کرده و این بردارها را به صورت  $X = [X_1, \dots, X_K]$  که به عنوان ماتریس مسیر شناخته می‌شوند، ترکیب می‌شود. این ماتریس یک ماتریس هنگل است به این معنا که تمام عناصر در امتداد قطر  $i + j = \text{const}$  باهم برابر است.

### ۲-۲- قدم دوم در تجزیه: تجزیه مقدار منفرد (SVD)

از آنالیز بردار  $X^T X$ ، مقادیر ویژه  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$  به دست آورده می‌شود، به طوری که  $d = \text{rank}(X^T X)$  و  $U_i$  بردار منفرد سمت چپ و  $V_i$  بردار منفرد سمت راست است. با استفاده از تجزیه مقدار منفرد، سری زمانی رابطه (۱) به دست می‌آید.

$$X \approx \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T \quad (1)$$

به طوری که  $X_i$  ماتریس‌های با رنک واحد هستند [21].

### ۲-۳- قدم اول در بازسازی: گروه‌بندی کردن

در این مرحله، یک پارتیشن از مجموعه اندیس‌های  $d, \dots, 1, j = 1, \dots, M$  مجموعه مجزا  $I_1, \dots, I_M$ ، متناظر با شکستن ماتریس اولیه  $X_i$  به  $M$  گروه است. حال اگر  $I = \{i_1, \dots, i_p\}$  باشد ماتریس  $X_I$  به صورت رابطه (۲) تعریف می‌شود:

$$X_I = X_{i_1} + \dots + X_{i_p} \quad (2)$$

حال با محاسبه ماتریس  $X_I$  برای  $I = I_1, \dots, I_M$  و جایگزینی آن‌ها در رابطه (۱)، رابطه (۳) به دست می‌آید:

$$X = X_{I_1} + \dots + X_{I_M} \quad (3)$$

روش انتخاب مجموعه‌های  $I = I_1, \dots, I_M$  گروه‌بندی نامیده می‌شود [22]. در اینجا جهت گروه‌بندی کارآمد با استفاده از الگوریتم بهینه‌سازی  $HHO$ ، مقادیر ویژه به‌گونه‌ای انتخاب شده است که شاخص خطای سری زمانی پیش‌بینی شده حداقل شود.

ایده اصلی این مرحله بازسازی مولفه‌های معین غیر احتمالی سری زمانی است. اگر  $z_{ij}$  بیانگر عناصر ماتریس  $Z$  باشد، در این صورت جمله  $k^{th}$  از سری منتج با متوسط گیری  $z_{ij}$  روی همه  $i, j$  به طوری که  $i + j = k + 2$  باشد به دست می‌آید. به این روش، متوسط سازی قطری یا هنکل سازی ماتریس گفته می‌شود. متوسط سازی قطری یک عملگر خطی و نگاشتی از ماتریس مسیر سری زمانی اولیه به خود آن سری زمانی است. با اعمال روش هنکل سازی به همه مولفه‌های ماتریس رابطه (۳)، رابطه (۴) به دست می‌آید:

$$X = \tilde{X}_{I_1} + \dots + \tilde{X}_{I_m} \quad (4)$$

به طوری که  $\tilde{X}_{I_i} = HX$  و

$$y_t = \sum_{k=1}^m \tilde{y}_t^{(k)}, \quad (5)$$

که در آن  $(\tilde{y}_t^{(k)}, \dots, \tilde{y}_T^{(k)}) = \tilde{Y}_T^{(k)}$  متناظر با ماتریس  $X_{I_k}$  است [22].

### ۵-۲- انتخاب پارامتر SSA (طول پنجره)

طول پنجره  $L$  تنها پارامتر در مرحله تجزیه سری زمانی است. انتخاب مناسب طول پنجره افزایش دقت SSA را به دنبال خواهد داشت. اگر بدانید که سری زمانی ممکن یک مولفه پریودیکی با پریود عدد صحیح داشته باشد بهتر است طول پنجره متناسب با آن پریود گرفته شود. در این مقاله  $L = 301$  در نظر گرفته شده است.

جدول ۱ آثار طول پنجره را در SSA نشان می‌دهد. هرچه طول پنجره بیش تر افزایش یابد، کوچک ترین مقدار منفرد کاهش یافته و ترند هموارتر و منظم تر می‌شود. اگر طول پنجره کاهش یابد، کوچک ترین مقدار منفرد افزایش یافته و ترند تقریبی و نامنظم تر می‌شود زیرا آن شامل مولفه‌های دیگر هم می‌شود.

جدول ۱- اثر طول پنجره (L).

Table 1- Effect of window length (L).

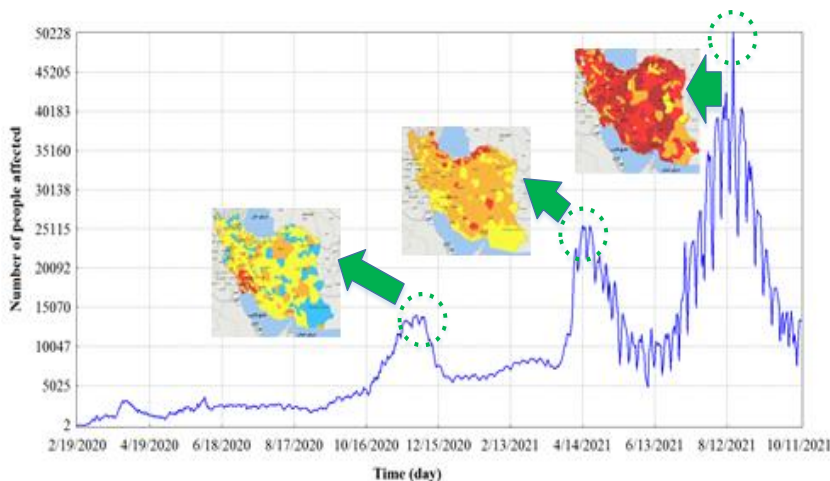
Window Length	Rank of $T_X$	$\sqrt{\lambda_1}$	Main Trend	
			smooth	regular
L↑	↑	↓	smooth	regular
L↓	↓	↑	rough	irregular

### ۳- شبیه سازی و ارزیابی نتایج

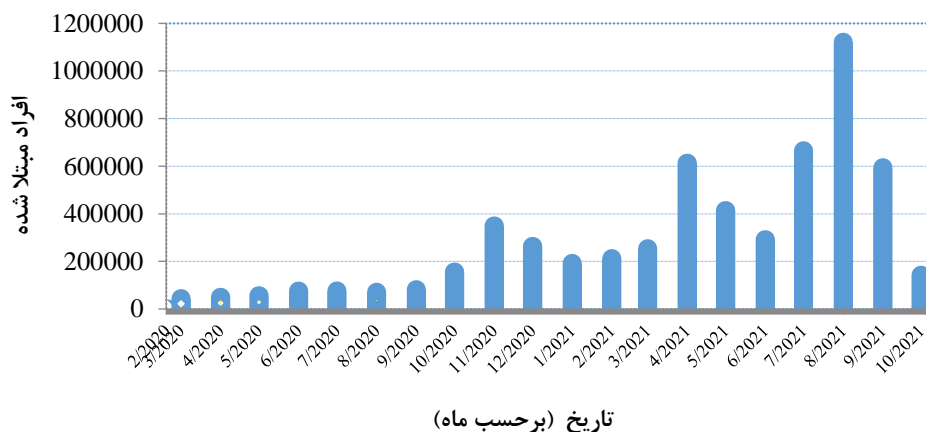
روند مبتلایان به کووید-۱۹ در ایران از تاریخ ۱۹ فوریه ۲۰۲۰ تا ۱۰ اکتبر ۲۰۲۱ به صورت روزانه با استفاده از مجموعه داده‌های مرکز علوم و مهندسی سیستم‌ها (CSSE) در دانشگاه جانز هاپکینز برای پیش بینی تعداد موارد تایید شده، در شکل ۲ نشان داده شده است [23]. همان طور که در این سری زمانی مشخص است تا ۱۰ اکتبر پنج پیک کرونا ایجاد شده است. بزرگ ترین پیک کرونا روز ۱۷ اوت ۲۰۲۱ رخ داده که حدود ۵۰۲۲۸ نفر مبتلای روزانه به کرونا توسط وزارت بهداشت در مان و آموزش پزشکی اعلام گردید. متأسفانه با افزایش آمار مبتلایان به کرونا، مرگومیر نیز افزایش یافته به طوری که در پیک پنجم کرونا در روز ۲۴ اوت ۷۰۹ نفر در یک روز جان خود را بر اثر کرونا از دست دادند.

دیتاهای مورد بررسی به صورت ماهیانه در چارت شکل ۳ نیز نشان داده است. همان طور که مشاهده می‌شود در این بازه زمانی در ماه اوت بیشترین میزان مبتلا به بیماری گزارش شده است.





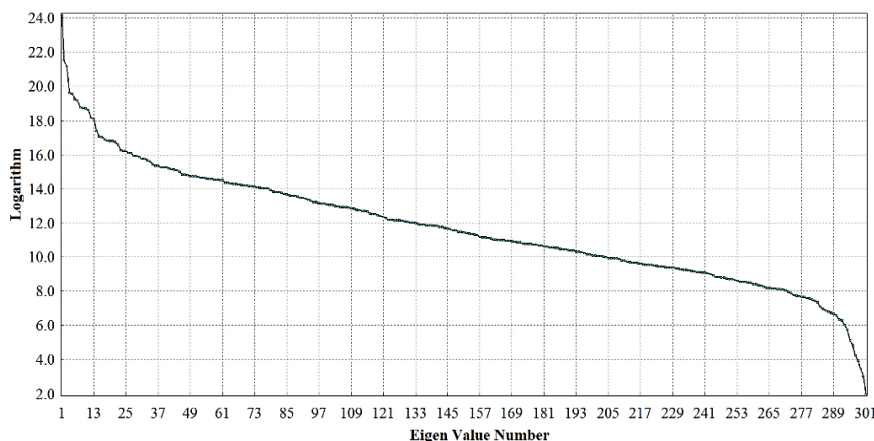
شکل ۲- سری زمانی روزانه مبتلایان بیماری کرونا.  
Figure 2- Daily time series of corona patients.



شکل ۳- تعداد افراد مبتلا شده به صورت ماهیانه.  
Figure 3- Number of infected people monthly.

### ۳-۱- نحوه گروه‌بندی کردن

در اینجا دو روش برای گروه‌بندی کردن ارائه شده است. روش اول، بررسی نقاط شکست در طیف مقادیر ویژه است. شکل ۴ نمودار لگاریتم ۳۰۱ مقدار ویژه اول برای مبتلایان بیماری کرونا رسم شده است. همان‌طور که از شکل دیده می‌شود مقادیر ویژه (۱)، (۲ و ۳)، (۴ و ۵)، (۶ و ۷)، (۸ و ۹ و ۱۰ و ۱۱)، (۱۲ و ۱۳)، (۱۴)، (۱۵، ۱۶، ۱۷، ۱۸، ۱۹) و (۲۰ و ۲۱) هارمونیک‌های اصلی سری زمانی مبتلایان بیماری کرونا را تشکیل می‌دهد.

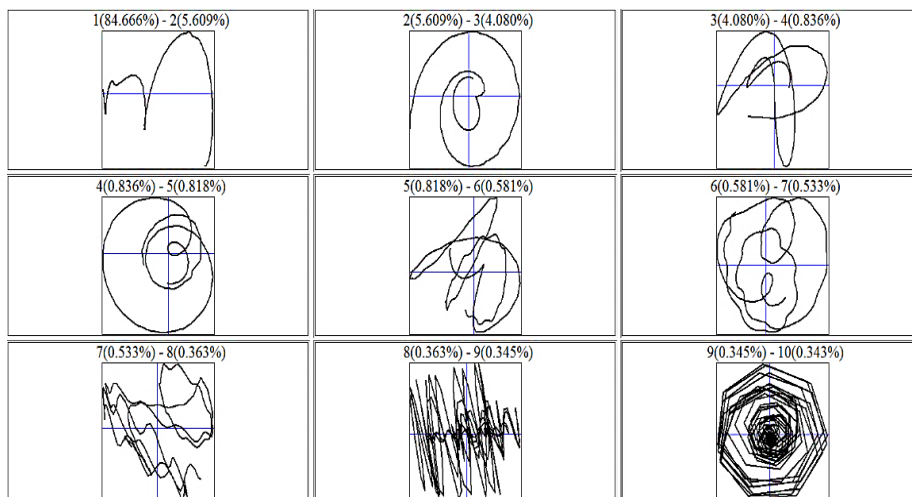


شکل ۴- لگاریتم ۵۰ مقدار ویژه اول.  
Figure 4- The logarithm of 50 first eigenvalues.





روش دیگر برای انجام گروه‌بندی مناسب، استفاده از نمودار scatter plot است. یک هارمونیک خالص با فرکانس  $\omega$ ، فاز، دامنه خاص و موقعیت مناسب، به طوری که  $p = \frac{1}{\omega}$  یک عدد صحیح باشد، آن وقت نقاط نمودار scatter plot روی یک  $p$  ضلعی منتظم قرار می‌گیرد. از آنجایی که  $p$  یک عدد صحیح است، آن یک پریود از هارمونیک است. شکل ۵ نمودار scatter plot جفت بردارهای هارمونیکی را تا هارمونیک دهم نشان می‌دهد.

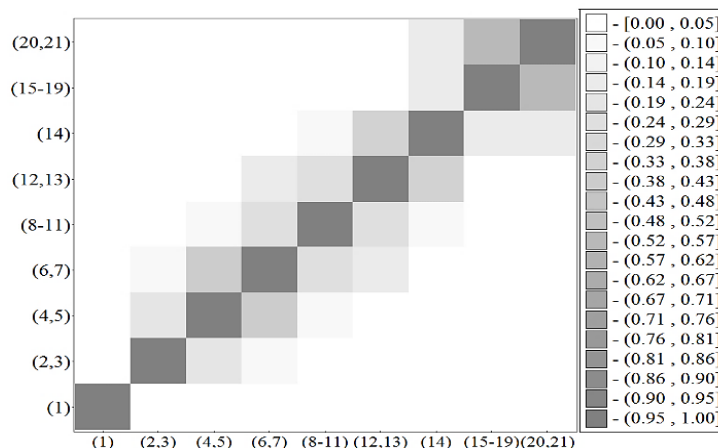


شکل ۵- scatter plot جفت بردار ویژه هارمونیکی.  
Figure 5- Scatter plot of harmonic eigenvector pairs.

### ۳-۲- بازسازی سری زمانی

اطلاعات بسیار مفیدی برای تشخیص تفکیک‌پذیری و شناسایی گروه در ماتریس به اصطلاح همبستگی  $W$  وجود دارد. این ماتریس شامل کسینوس‌های وزنی زاویه‌ای بین اجزای سری زمانی بازسازی شده است. وزن‌ها تعداد ورودی‌های عبارت‌های سری زمانی را در ماتریس مسیر آن منعکس می‌کند.

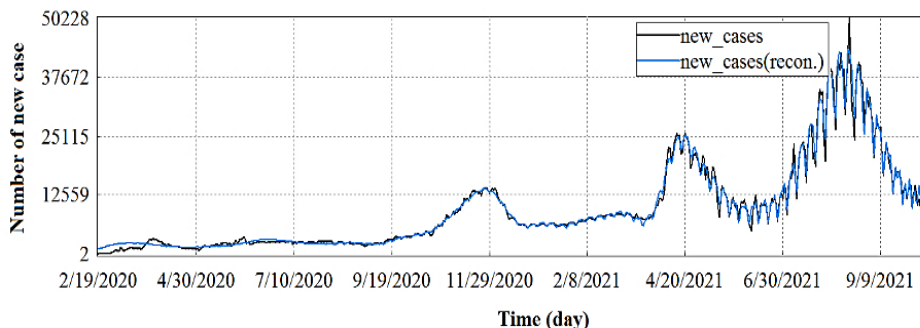
بنابراین، با ماتریس همبستگی  $W$  بین سری‌های بازسازی اولیه می‌توان گروه‌هایی از اجزای سری مرتبط را پیدا کرد و از این اطلاعات برای گروه‌بندی بعدی استفاده کرد. یکی از قوانین این است که اجزای همبسته را در گروه‌های مختلف قرار ندهید. همچنین، برای بررسی تجزیه گروه‌بندی شده می‌توان از همبستگی  $W$  استفاده کرد. به راحتی می‌توان ماتریس همبستگی  $W$  بین اجزای سری را به صورت گرافیکی در مقیاس سفید-سیاه به تصویر کشید، جایی که همبستگی‌های کوچک به رنگ سفید نشان داده می‌شود، در حالی که همبستگی با مدوله‌ای نزدیک به ۱ به رنگ سیاه نشان داده می‌شود. شکل ۶ ماتریس همبستگی  $W$  را با توجه به گروه‌بندی انجام‌شده نشان می‌دهد.



شکل ۶- ماتریس همبستگی  $W$ .  
Figure 6- W correlation matrix.

همان‌طور که قبلاً ذکر شد، در این بخش تلاش می‌شود که با استفاده از زوج مقادیر ویژه انتخاب‌شده، به بازسازی سری زمانی پرداخته شود. شکل ۷ بازسازی سری زمانی در مقیاسه با سری زمانی اولیه را نشان می‌دهد.





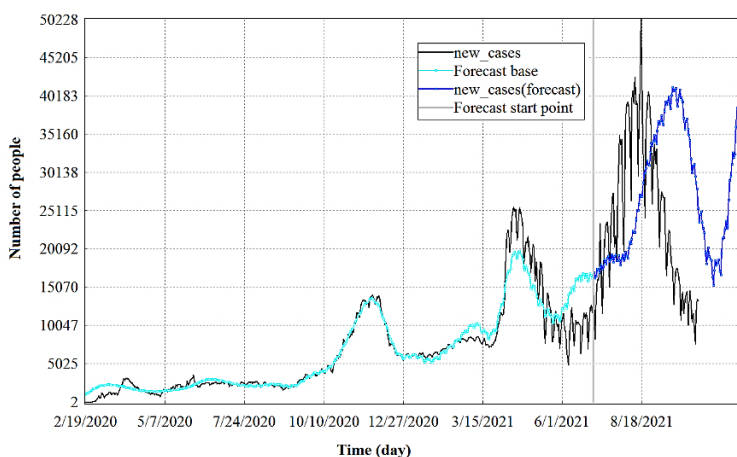
شکل ۷- سری زمانی بازسازی شده (خط) در مقایسه با دیتای واقعی (نقطه).  
**Figure 7- Reconstructed time series (line) compared to real data (dots).**

### ۳-۳- پیش‌بینی با SSA

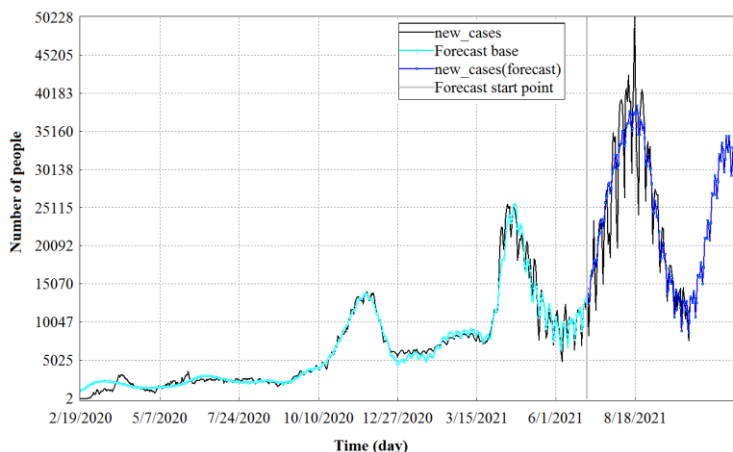
پیش‌بینی سری زمانی با الگوریتم SSA می‌تواند به سری زمانی که فرمول بازگشتی خطی (LRF) را برآورده می‌سازد، اعمال شود. سری زمانی  $y_T$  یک LRF مرتبه  $d$  را تامین می‌کند اگر اعداد  $a_1, \dots, a_d$  چنان وجود داشته باشد تا رابطه (۶) برقرار باشد:

$$\tilde{y}_{i+d} = \sum_{k=1}^d a_k y_{i+d-k}, \quad 1 \leq i \leq T-d. \quad (6)$$

و این ضرایب با استفاده از تجزیه SVD به دست می‌آید. شکل‌های ۸ و ۹ پیش‌بینی با SSA با دوروش *Recurrent* و *Vector* را نشان می‌دهد. همان‌طور که در شکل ۸ مشاهده می‌شود، در ۵۰ روز آینده احتمال وقوع یک ششم کرونا در ایران البته با تعداد افراد مبتلا کمتر نسبت به پیک پنجم مشاهده می‌شود.



شکل ۸- مقایسه مقدار پیش‌بینی شده با مقدار واقعی آن روش Vector-SSA.  
**Figure 8- Comparison of the predicted value with the actual value of the Vector-SSA method.**

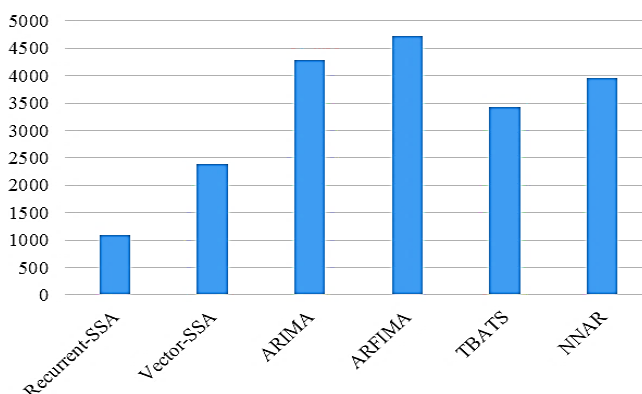


شکل ۹- مقایسه مقدار پیش‌بینی شده با مقدار واقعی آن با استفاده از روش Recurrent-SSA.  
**Figure 9- Comparing the predicted value with its actual value using the recurrent-SSA method.**

به منظور ارزیابی پتانسیل SSA برای پیش‌بینی افراد مبتلابه جدید، عملکرد SSA را با سایر تکنیک‌های پیش‌بینی سری‌های زمانی متداول از جمله *ARFIMA*، *ARIMA*، هموارسازی نمایی، *TBATS* و *NNAR* مقایسه می‌کند. برای محاسبه دقت روش‌ها، از شاخص خطای مربع میانگین ریشه (*RMSE*) استفاده شده است. *RMSE* برای پیش‌بینی کاربردی فراوان دارد و رابطه آن به صورت زیر است:

$$RMSE = \left[ \frac{1}{N} \sum_{i=1}^N (L_{ACT} - L_{FOR})^2 \right]^{1/2}. \quad (7)$$

شکل ۱۰ مقایسه شاخص خطای *RMSE* روش پیشنهادی را با سایر روش‌ها نشان می‌دهد.



شکل ۱۰ - مقایسه شاخص خطای *RMSE* روش‌های پیش‌بینی.

Figure 10- Comparison of the *RMSE* error index of forecasting methods.

بر اساس میزان خطایی که تکنیک‌های پیش‌بینی *Recurrent-SSA* و *Vector-SSA* نسبت به سایر مدل‌ها در افق ۵۰ روز دارند، این روش‌ها گزینه‌های مناسبی برای پیش‌بینی تعداد موارد تاییدشده روزانه ناشی از کووید-۱۹ هستند.

#### ۴- نتیجه‌گیری

شیوع اخیرا جهانی کووید-۱۹ بسیاری از کشورهای جهان را تحت تاثیر قرار داده است. با توجه به افزایش تعداد افراد تازه آلوده و تنگناهای سیستم مراقبت‌های بهداشتی، پیش‌بینی تعداد بیماران آتی مفید خواهد بود. این مطالعه باهدف پیش‌بینی کارآمد برای تخمین موارد جدید در ایران به مدت ۵۰ روز با استفاده از مجموعه داده‌های رسمی وزارت بهداشت، درمان و آموزش پزشکی و تاثیر اقدامات کنترلی بر روی گسترش کووید-۱۹ است.

بدون شک پیش‌بینی سری‌های زمانی مورد استفاده می‌تواند محدودیت‌های قابل توجهی شامل تصمیمات مقامات بهداشتی، در دسترس بودن واکسن، تغییر آلودگی‌ها و مرگ‌ومیرها به صورت روزانه و هم‌چنین ماهیت پویای همه‌گیری دارد؛ یعنی در شرایط واقعی، پیش‌بینی سری‌های زمانی عموماً فقط در کوتاه‌مدت می‌تواند دقیق باشد. با این وجود، اگر شرایط فعلی (محدودیت‌های قرنطینه، در دسترس بودن واکسن، سرعت واکسیناسیون و ظرفیت بیمارستان‌ها) را اصلاح کنیم، می‌توانیم سناریوی وضعیت توسعه را با توجه به شرایط داده‌شده در میان‌مدت یا بلندمدت مشاهده کنیم. برای به دست آوردن نتایج دقیق، توصیه می‌شود که داده‌ها حداقل به صورت هفتگی به روز شوند، زیرا برخی از عوامل موثر بر روند انتشار ویروس وجود دارد که می‌تواند به طور قابل توجهی بر انتخاب مدل و دقت پیش‌بینی‌های به دست آمده تاثیر بگذارد.

در این مقاله، از نسخه بهینه دو تکنیک پیش‌بینی *SSA*، یعنی *Recurrent-SSA* و *Vector-SSA*، برای پیش‌بینی تعداد موارد تاییدشده روزانه ناشی از کووید-۱۹ استفاده شده است. هم‌چنین گروه‌بندی مقادیر ویژه جهت بازسازی سری زمانی با استفاده از الگوریتم بهینه‌سازی *HHO* صورت گرفته است. به منظور ارزیابی عملکرد این رویکردها بر اساس معیار *RMSE*، نتایج پیش‌بینی با نتایج سایر روش‌های پیش‌بینی سری زمانی از جمله *ARFIMA*، *ARIMA*، *TBATS* و *NNAR* مقایسه شده است. نتایج شبیه‌سازی نشان می‌دهد که تکنیک بهینه *Recurrent-SSA* می‌تواند ابزاری قدرتمند برای پیش‌بینی تعداد موارد تاییدشده روزانه بیماری کووید-۱۹ باشد و نسبت به سایر روش‌های پیش‌بینی خطای کمتری ارائه دهد. همان‌طور که در نمودار شکل ۱۰ مشاهده می‌شود، میزان شاخص خطا در الگوریتم پیشنهادی *Recurrent-SSA* نسبت به الگوریتم *ARIMA*، ۷۴/۲۵٪ کاهش، نسبت به الگوریتم *ARFIMA*، ۷۶/۷۱٪ کاهش



و نسبت به الگوریتم‌های *TBATS* و *NNAR* به ترتیب ۶۷/۹٪ و ۷۲/۲٪ کاهش پیدا کرده است. هم‌چنین الگوریتم *Vector-SSA* نسبت به الگوریتم‌های *TBATS*، *ARFIMA*، *ARIMA* و *NNAR* به ترتیب ۴۴/۳۳٪، ۴۹/۶۴٪، ۳۰/۵۹٪ و ۳۹/۹٪ کاهش یافته است.

## منابع

- [1] Pereira, I. G., Guerin, J. M., Silva Júnior, A. G., Garcia, G. S., Piscitelli, P., Miani, A., Distante, C., & Gonçalves, L. M. G. (2020). Forecasting COVID-19 dynamics in Brazil: a data driven approach. *International journal of environmental research and public health*, 17(14), 5115.
- [2] Arora, P., Kumar, H., & Panigrahi, B. K. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: a descriptive case study of India. *Chaos, solitons & fractals*, 139, 110017.
- [3] Perc, M., Gorišek Miksić, N., Slavinec, M., & Stožer, A. (2020). Forecasting COVID-19. *Frontiers in physics*, 8, 1-5. <https://doi.org/10.3389/fphy.2020.00127>
- [4] Zareie, B., Roshani, A., Mansournia, M. A., Rasouli, M. A., & Moradi, G. (2020). A model for COVID-19 prediction in Iran based on China parameters. *MedRxiv*, 23(4), 244-248.
- [5] Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-series data: a comparative study. *Chaos, solitons & fractals*, 140, 110121. <https://doi.org/10.1016/j.chaos.2020.110121>
- [6] Altan, A., & Karasu, S. (2020). Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. *Chaos, solitons & fractals*, 140, 110071. <https://doi.org/10.1016/j.chaos.2020.110071>
- [7] Petropoulos, F., & Makridakis, S. (2020). Forecasting the novel coronavirus COVID-19. *PLoS one*, 15(3), e0231236. <https://doi.org/10.1371/journal.pone.0231236>
- [8] Konarasinghe, K. M. U. B. (2020). Forecasting COVID-19 spread in Malaysia, Thailand, and Singapore. *Journal of new frontiers in healthcare and biological sciences*, 1(2), 1-13.
- [9] Sahai, A. K., Rath, N., Sood, V., & Singh, M. P. (2020). ARIMA modelling & forecasting of COVID-19 in top five affected countries. *Diabetes & metabolic syndrome: clinical research & reviews*, 14(5), 1419-1427.
- [10] Khan, F., Saeed, A., & Ali, S. (2020). Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using vector autoregressive model in Pakistan. *Chaos, solitons & fractals*, 140, 110189. <https://doi.org/10.1016/j.chaos.2020.110189>
- [11] Iwata, K., Doi, A., & Miyakoshi, C. (2020). Was school closure effective in mitigating coronavirus disease 2019 (COVID-19)? time series analysis using bayesian inference. *International journal of infectious diseases*, 99, 57-61.
- [12] Singh, S., Parmar, K. S., Kumar, J., & Makkhan, S. J. S. (2020). Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. *Chaos, solitons & fractals*, 135, 109866. <https://doi.org/10.1016/j.chaos.2020.109866>
- [13] Abdollahi, J., Irani, A. J., & Nouri-Moghaddam, B. (2021). Modeling and forecasting spread of COVID-19 epidemic in Iran until Sep 22, 2021, based on deep learning. <https://doi.org/10.48550/arXiv.2103.08178>
- [14] Abotaleb, M., & Makarovskikh, T. (2021). System for forecasting COVID-19 cases using time-series and neural networks models. *Engineering proceedings*, 5(1), 46. <https://doi.org/10.3390/engproc2021005046>
- [15] Darapaneni, N., Reddy, D., Paduri, A. R., Acharya, P., & Nithin, H. S. (2020). Forecasting of COVID-19 in India using ARIMA model. *2020 11th IEEE annual ubiquitous computing, electronics & mobile communication conference (UEMCON)* (pp. 0894-0899). New York, USA: IEEE.
- [16] Ramazi, P., Haratian, A., Meghdadi, M., Mari Oriyad, A., Lewis, M. A., Maleki, Z., Vega, R., Wang, H., Wishart, D. S., & Greiner, R. (2021). Accurate long-range forecasting of COVID-19 mortality in the USA. *Scientific reports*, 11(1), 13822.
- [17] Niksirat, M., & Nasser, S. H. (2021). Forecasting of the number of cases and deaths due to corona disease using neuro-fuzzy networks. *Journal of decisions and operations research*, 5(4), 414-425.
- [18] Doornik, J. A., Castle, J. L., & Hendry, D. F. (2021). Modeling and forecasting the COVID-19 pandemic time-series data. *Social science quarterly*, 102(5), 2070-2087.
- [19] Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris hawks optimization: algorithm and applications. *Future generation computer systems*, 97, 849-872.
- [20] Zhigljavsky, A. (2010). Singular spectrum analysis for time series: introduction to this special issue. *Statistics and its interface*, 3(3), 255-258. <https://dx.doi.org/10.4310/SII.2010.v3.n3.a1>
- [21] De Carvalho, M., Rodrigues, P. C., & Rua, A. (2012). Tracking the US business cycle with a singular spectrum analysis. *Economics letters*, 114(1), 32-35.
- [22] Hassani, H. (2007). Singular spectrum analysis: methodology and comparison. *Journal of data science*, 5(1), 239-257. <https://mpra.ub.uni-muenchen.de/id/eprint/4991>
- [23] CSSE, & Johns Hopkins University. (2020). *COVID-19 data repository by the center for systems science and engineering (CSSE) at Johns Hopkins university*. <https://github.com/CSSEGISandData/COVID-19>

## ضمایم

داده‌های تکمیلی: داده‌های بیماری کووید-۱۹ در ایران از طریق وبسایت <https://github.com/CSSEGISandData/COVID-19> قابل استفاده است.

