



BÜTÜNLEŞİK KALİTE YÖNETİM SİSTEMİ GERÇEKLEŞTİRME KANIT FORMU

Birim	Fen Edebiyat Fakültesi
Alt Birim	Matematik Bölümü
Dönem	
Gerçekleşme Tarihi	05.08.2021
Gerçekleştirilen Faaliyet Adı	COVID-19 PREDICTION METHODS USING HUMAN GENOMIC DATA
Gerçekleştirilemeyen Faaliyetler İçin Gerekçe	
Faaliyet Gerçekleştirme Açıklaması	Bölüm içi bilimsel etkinlik amaçlı Doç. Dr. Handan KÖSE tarafından yürütülen seminerler dahilinde Hilal ARSLAN'ın 05.08.2021 tarihinde saat 13.30 da “ COVID-19 PREDICTION METHODS USING HUMAN GENOMIC DATA ” adlı seminer http://toplanti.ahievran.edu.tr/b/han-qby-0dj adresinden online olarak gerçekleştirilmiştir.



BÜTÜNLEŞİK KALİTE YÖNETİM SİSTEMİ GERÇEKLEŞTİRME KANIT FORMU



SEMİNER

COVID-19 PREDICTION METHODS USING HUMAN GENOMIC DATA

Hilal ARSLAN

İzmir Bakırçay Üniversitesi Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği

**Faaliyet
Gerçekleştirme
Kanıt Görseli**

Özet:

COVID-19 caused by a novel SARS-CoV-2 virus rapidly spread worldwide and threatens the public health system as well as leads to severe outcomes. It is essential to quickly identify SARS-CoV-2 to prevent the spread of COVID-19 among people. In this study, we propose a COVID-19 detection method based on genome similarity of human SARS-CoV-2 and bat SARS-CoV-like coronavirus. By considering nucleotide similarities, we define three similarity features by comparing sequences of human SARS-CoV-2 and bat SARS-CoV-like coronavirus. The efficiency of the proposed features is evaluated by conducting various machine learning classifiers on 1000 genome sequences of SARS-CoV-2 and 612 genome sequences of the other types of human coronaviruses. Results are also shown on the datasets released by the current state of the art studies predicting SARS-CoV-2 from genome sequences. Experimental results show that using the integration of CpG based and similarity features achieves an accuracy of 99.8% on our dataset and 100% accuracy on the dataset released by the current state of the art studies. The proposed method remarkably achieves better results than the state-of-the-art COVID-19 prediction methods by using only five effective features. This underlines the efficiency of the proposed method.



BÜTÜNLEŞİK KALİTE YÖNETİM SİSTEMİ GERÇEKLEŞTİRME KANIT FORMU

The screenshot shows a Zoom meeting recording of a presentation. The main content is a slide titled "COVID-19 prediction methods using human genomic data" by Hilal Arslan. The slide features a red banner with the title and a 3D model of a coronavirus particle. Below the banner, the presenter's name "Hilal Arslan" is listed. The slide content includes a list of topics: "Covid-19 Pandemic", "Method", "Dataset", "Results", and "Conclusion". The text "A new COVID-19 detection method from human genome sequences using CpG island features and KNN classifier" is also visible. At the bottom of the slide, there are three small images: a globe with virus particles, a 3D model of a coronavirus, and a 3D model of a coronavirus with red spikes. The presentation is displayed in a browser window with a video feed of the presenter on the right. The browser address bar shows the URL: "toplanti.ahievran.edu.tr/playback/presentation/2.0/playback.html?meetingId=b1490d9c884a5b6b895e172ce3fff6b5c312a0f-1628158589161". The Zoom interface shows the presenter's name "Hilal Arslan Seminer" and a video feed of the presenter. The system tray at the bottom shows the time as 03:41 and the date as 11/8/2021.

Feature Extraction: CpG island

$$feature_1 = \text{ratio}(C) + \text{ratio}(G) \quad (1^*)$$

$$feature_2 = \frac{\text{ratio}(CG)}{\text{ratio}(C)\text{ratio}(G)} \quad (2^*)$$

AAAAAAGTT GCTGAACCTT TCCCCAACT CTGCCGTAGA GGCGGGAGTG GAGGCGGTG


#C = 13, #G = 20, #CG = 3,

Fig. 1. CpG based features. The numbers of C, G, and CG are 13, 20, and 3, respectively. Thus, $CGp = \text{ratio}(C) + \text{ratio}(G) = 0.55$, and $CpGo = \text{ratio}(CG)/(\text{ratio}(C)\text{ratio}(G)) = 0.69$

* Zeng et al., SCS: signal, context, and structure features for genome-wide human promoter recognition, 2010

Comparison of the state-of-the-art methods

Study	Method	Features	Sequence Dataset	Accuracy (%)
Naeem et al. (2020)	KNN	7 MI 1 DFT 1 DCT	76 SARS-CoV-2 76 SARS-CoV 76 MERS-CoV	100
Randhawa et al. (2020)	LD KNN SVM	DFT PCC	20 betaCoV 20 alphaCoV 20 deltaCoV	100
Lopez-Rincon et al. (2021)	CNN	3827 max pooling features	66 SARS-CoV-2 487 others	98.73
Arslan and Arslan (2021)	KNN with L1 metrics	2 CpG based features	1000 SARS-CoV-2 592 others	98.4
Proposed Method	MLP AdaBoost RF KNN DR SVM	2 CpG based and 3 similarity features	1000 SARS-CoV-2 92 AlphaCoV 523 BetaCoV 1 RaTG13	99.8



Değerlendirme:

(Form No: FR - 585; Revizyon Tarihi :...../...../.....; Revizyon No:.....)