Abstract No: 98 ***ICT, Mathematics, and Statistics***

EMPIRICAL AND MULTIVARIATE EPIDEMIOLOGICAL ANALYSIS ON COVID-19 PANDEMIC

S.H.M.B.A. Samarasinghe and Y.P.R.D. Yapa*

[1]*Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka.*
**roshany@sci.pdn.ac.lk*

The spreading pattern of COVID-19 differs greatly across the countries based on different country-level factors, quarantine measures, and government policies. This study examines definite clusters of countries that exhibit similar patterns in the time series of COVID-19 daily confirmed cases per million worldwide. The country-level demographic, socioeconomic, and meteorological variables associated with the trend patterns were studied, and finally, the geographic and temporal distribution of the countries within each identified cluster were examined. The data set consists of 155 affected countries in the world as of August 7, 2020. Multivariate analysis techniques, principal component analysis, factor analysis, and cluster analysis in data mining were used to explore the hidden features of the COVID-19 merged dataset. Three distinct clusters of daily confirmed cases per million across the world were identified using time series clustering, and three univariate time series models were fitted for the average values of the series within each cluster by assuming countries within each cluster follow a similar distribution. An explanatory model was applied to identify the association of meteorological, demographic, and socioeconomic variables with each cluster pattern, and then the cluster solutions were validated. Assorted county-level meteorological, demographic, and socioeconomic variables appeared to have significant relationships with identified three clusters. The findings of this study can be used to determine the disease spread in countries with similar distributions and underlying factors.

Keywords: Cluster analysis, Data mining, Factor analysis, Principal component analysis, Time series