Research Article

Spatio-Temporal Analysis of Infectious Diseases for Hotspot Identification and Evolution: A Special Case of COVID-19 in Punjab Pakistan

Fida Muhammad[1], Resham Hafeez[2], Fatima Khalique[3*], Mehroz Sadiq[1], Saba Mahmood[1], Farah Akif[1] and Moazzam Ali[1]

[1]Department of Computer Science, Bahria University, E8, Islamabad, Pakistan; [2]Islamabad Medical and Dental College, Shaheed Zulfiqar Ali Bhutto Medical University, Islamabad, Pakistan; [3]Center of Excellence in Artificial Intelligence (CoE-AI), Department of Computer Science, Bahria University, Islamabad, Pakistan.

Abstract: Pandemics have far-reaching and complex effects on human society, and can have a profound impact on the way we live, work, and interact with one another. Although, the substantial advancement in healthcare industries is an undeniable fact, still, there is a crucial need to resolve the inadequacies to overcome the glitches. The widespread of pandemics in general and corona virus in particular has motivated researchers to study the patterns of spread and evolution of diseases in population. The increase rate of the COVID-19 has captured the attention of researchers to analyze the COVID-19 cases on the basis of spatio-temporal patterns occurring on several hotspots. This study provides the foundations for analyzing the communicable diseases in context of both spatio, and temporal features. The aim of the hotspot spatial analysis is to determine the most impacted regions by the COVID-19. This research involves investigation of disease hotspots in space and time on the Covid dataset for 36 districts of Punjab Pakistan. The data is investigated using ideal hot spot analysis, which scrutinizes the location of hot spot sites over a certain span of time. Spatio-temporal hotspot identification of coronavirus can help in understanding the spread and transmission dynamics of the virus over time and space. It can aid in identifying areas with high infection rates and allow for targeted interventions to control the spread of the virus. This information can be useful in informing public health policy and resource allocation for managing the pandemic.

Received: December 25, 2024; **Accepted:** September 05, 2025; **Published:** November 15, 2025
***Correspondence:** Fatima Khalique, Center of Excellence in Artificial Intelligence (CoE-AI), Department of Computer Science, Bahria University, Islamabad, Pakistan; **Email:** fkhalique.buic@bahria.edu.pk
Citation: Muhammad, F., R. Hafeez, F. Khalique, M. Sadiq, S. Mahmood, F. Akif and M. Ali. 2025. Spatio-temporal analysis of infectious diseases for hotspot identification and evolution: A special case of COVID-19 in Punjab Pakistan. *Journal of Engineering and Applied Sciences,* 44: 36-36.
DOI: https://dx.doi.org/10.17582/journal.jeas/44.36.46
Keywords: Spatio-temporal analysis, Spatial hotspots, Getis-Ord Gi, Communicable diseases, disease trends, COVID-19

Introduction

Pandemics pose significant challenges to public health systems, particularly in low- and middle income countries where infrastructure and real-time data analysis capabilities are limited. COVID-19 has emphasized the need for accurate and timely spatial surveillance system (Zumla *et al*., 2024; Li *et al*., 2024). While substantial advancements in healthcare technologies are notable, gaps remain in leveraging spatio-temporal analytics for proactive disease management. Recent studies from 2022 to 2024 underscore the importance of real-time, localized hotspot identification to guide interventions (Wang *et al*., 2024; Oh *et al*., 2024). Our approach builds on this emerging body of work by implementing spatio-temporal hotspot analysis using the Getis-Ord Gi* statistic (pronounced G-i-star) and Mann-Kendall trend tests specifically for Punjab, Pakistan. Unlike some regional studies which utilize static heatmaps or only temporal forecasting, we focus on dynamic interaction between space and time to track the evolution of disease spread at the district level (Ansari *et al*., 2024; Chohan, 2024).

Spatio-temporal disease patterns investigation date back to first epidemic spreads of common diseases and involves, for example, identification of seasonal patterns of influenza transmission, the clustering of tuberculosis cases in urban areas, and the emergence and spread of COVID-19 in different regions of the world. As of now, the reported cases of COVID-19 have surpassed 34.2 million globally (Liu *et al*., 2020; Mahmood and Tariq, 2023). Unfortunately, the individuals with preexisting issues regarding health, such as heart syndromes and diabetics problems are more vulnerable to the virus. The increase rate of the COVID-19 has captured the attention of researchers to analyze the COVID cases on the basis of Spatio-temporal patterns occurring on several hotspots.

Spatio-temporal investigation of disease spread is important because it can provide valuable insights into the dynamics of the spread and transmission of a disease over time and space. This information can help in identifying the areas and populations that are most affected by the disease and guide targeted interventions to control its spread. By analyzing spatio-temporal data, researchers can also identify patterns and trends in the spread of the disease, which can aid in developing more accurate models to predict future outbreaks and in-form public health policy and resource allocation. Moreover, spatio-temporal investigation can help to understand the impact of environmental, social, and behavioral factors on disease transmission, which can provide opportunities to implement preventive measures and mitigate the spread of the disease.

Hotspot identification is crucial step in spatio-temporal investigations of a time series data. Hotspot analysis is identified either by the study of point distribution or by the spatial configuration of the hitting points in a space (Lai *et al*., 2020). The high concentration of points within a precise area is contrasted to a complete spatial randomness model, which depicts that the points occurrences transpire randomly, i.e. homogeneous spatial poison process. While the researchers are busy in investigative study of the patterns of the points, a comparison is made between the density of all points within specified hotspots and a complete spatial randomness model. The spatial randomness model is being responsible to depict the progression in which all the point events transpire on random phenomenon. In addition, there are several hot spot approaches which determine the spatial patterns of events occurring in a specified region. The study reveals the significant growth of hot spot analysis in public health and epidemiological research, as well as in other disciplines such as crime mapping (Lai *et al*., 2020; Kim *et al*., 2024).

Hot spots can be characterized as not being static. They vary according to the spatio-temporal changes. Spatial analysis on hotspots has long-drawn-out its importance to allocate the particular resources according to the disease and hit points on a particular region. The reseach needs to expand on temporal basis as well. This problem has high potential to solve by targetting some particular hotspots to analyze the temporal change according to number of patients. The reason of this study is necessary, so, particular resources can be provided to the regions as per their requirement. Thus, the aim of the hotspot analysis is to determine the most impacted regions by the COVID-19.

This research involves the implementation of Getis Ord Gi* on the Covid dataset for 36 districts (hotspots) of Punjab Pakistan. This study provides the foundations for analyzing the communicable diseases in context of both spatio, and temporal features. The analysis is able to identify statistically significant hot regions and cold regions. An adjacent environment is examined for each characteristic in a hot spot study. It is possible that a high-value feature might not be a hot spot. A hot spot feature will have a high value of events and must be surrounding by feature with high value of event. While a cold spot feature will have a high value of events and must be surrounding by feature with high value of event. Hot spot analyses are widely performed for demography retail analysis, voting patterns, industrial geographic location, disease, traffic accidents, kitchen fires, and crimes. Frequently, frequency of disease cases is used as hot spot indication (Pitalua Rodriguez *et al*., 2019; Sreeram and Shanmugam, 2018; Khan *et al*., 2021).

Related work

The literature review on disease spread reveals that hotspots of any kind are not static. Also, it is not in routine to discuss their temporal features. Bejon *et al*. (2010) conducted their research in certain rural areas of Kenya by identifying several hotspots. They work for the investigation of the parasitemia by taking the blood samples from sick children, known as febrile children. The purpose of this task is to diagnose for parasitemia without the symptoms of ailment and to check the antibodies against malaria parasites. Furthermore, they work for the detection of transmission heterogeneity and hotspots by applying "spatial scan statistic". The researchers have classified hotspots in two main classes; febrile malaria hotspots, and stable hotspots. Although, febrile malaria hotspots remain in location for the time period of one or two years, while the stable hotspots are symptoms-free parasitemia which remained for years after the first appearance. The drawbacks due to the lack of understanding of the hotspots involve high risks after another while omitting the regions with real high perils (Xu, 2022; Ali *et al*., 2020; Kertész *et al*., 2019; Ertunç *et al*., 2013).

Jana and Sar (2016) contributed to hotspot detection through the application of cluster-outlier analysis and the Getis-Ord Gi* statistic. Their study focused on evaluating the performance of the second cycle of primary education across various Indian states. To identify geographic clusters of educational performance, they employed a cluster-outlier analysis method based on Inverse Distance Weighting (IDW), enabling the detection of statistically significant hotspot regions in the education sector. In this research, the authors use Gi* statistic to achieve their goals. Both the analysis, (cold spots and hot spots analysis) perform their important role in determining the geographical clusters of the second cycle of primary education in India. The findings of their study indicate that effective identification of high and low-risk regions is able to help to improve the developmental plans and policies at the second cycle of primary education. Manepalli *et al*. (2011), has identified the hotspots for the accident occurs on Arkansas highway. This study shows that the authors use the Geo-graphical information system for the identification of hot spots. They perform contrast kernel density estimate (K) and Gi* statistics analysis to achieve their goals in better way. However, some of the factors have high capacity to be considered while performing these analyses. These techniques need a solid statistical base and precise crash placement.

Gi* can be used to locate hot spots in some circumstances. Ruswandi *et al*. (2021), have worked for the hashtag #2019GantiPresiden by giving the indication that it has reached half of the Indonesia's provinces during the time period of three months effective from May 2018. The authors of this research have used the Gi* statistics for optimal hotspots analysis to detect the victim's regions by utilizing Indonesia political opposition hashtag data collected from Twitter. The results of this article show that the capital of Indonesia, Daerah Khusus Ibukota Jakarta (DKI Jakarta), has the largest concentrations of hotspots. Bejon *et al*. (2010), contribute their struggle in determining the cases of people with parasitemia and without parasitemia. To locate hotspots of febrile illness and clusters of hotspots of asymptomatic parasitemia, they use a Bernoulli model with a discrete Poisson model. Furthermore, the authors of this study use the SaTScan software, where the scanning windows used by the spatial scan statistic move across the space. In this article, an equal distribution of cases throughout the population is, also, considered. The statistical importance of the hot spot is subsequently assessed, taking into consideration the results of several tests for the various possible cluster locations and sizes. In addition, 5600 incidences of febrile illness were predicted throughout 32,452 person-years of monitoring in 256 homesteads across three research cohort with the highest prevalence. The results indicate that there were 0.49 and 0.82 malaria cases per child per year (cpy), on average.

Srikanth and Srikanth (2020), underwrite their research for identification of hotspot areas of traffic accident. The article exposes the use of Kernel Density Estimation (KDE) and Gi* statistics to detect hot spot areas. The KDE approach divides the entire research region into a specific number of cells. KDE is assessed by covering each crash location with a smooth, symmetrical surface, by summing the values of all the surfaces for the reference site inside the search radius, and using a mathematical function, the kernel function, to calculate density per unit area. Statistical significant hotspots are found using the KDE and Gi* spatial analysis methodologies for the city under observation and investigation. This study uses the city of Des Moines in Polk County, United States as research area, for illustration reasons. While the Gi* statistic displays the statistical importance of the hot spots discovered in KDE, the KDE gives a visual depiction of the locations of accident hot spots. However, the grid size and bandwidth must be chosen properly to have clear visuals in KDE, In this investigation, the ideal grid size and bandwidth were determined to be 100 m and 500 m, respectively. In other instances, the hot spots were either merged in higher bandwidths or could not be recognized (Wang *et al*., 2011; Kalinic and Krisp, 2018).

Materials and Methods

The methodology consists of multi-step pipeline culminating in hotspot evolution prediction. The pipeline includes gathering the COVID data for the region of investigation, exploratory data analysis to preprocess the data, hotspot analysis to create the time-space cube, and analyzing emergent hotspots. The pipeline of suggested study is depicted below in Figure 1.

The data set has been collected from the National Information Technology Boardy (NITB) to evaluate the results. This data covers the Corona Virus cases districts of Punjab, March 2020 to February 2021.

The two categories: i.e., geographical and temporal, are representing the dataset. The attributes of dataset are depicted in Table 1.

Table 1: *Description of COVID-19 dataset.*

Attribute	Description
Reporting name	Hospital or lab name that reported the case
Age	Patient age
Gender	Patient gender
Address	Patient location
Date reported	The date on which case was reported
Province	Name of province where case was reported
District	Name of district in which case was reported

We selected 36 districts in Punjab, Pakistan based on the availability and completeness of COVID-19 case data reported between March 2020 and February 2021. Weekly intervals were chosen instead of monthly intervals to ensure finer granularity and early detection of emerging hotspots. The dataset comprises 179,685 COVID-19 cases reported over a 12-month period. Exploratory data analysis revealed 1,018 records with missing age information and 107 with missing gender. Additionally, district information was absent in 8,573 records, while 1,356 records lacked both address and district details. For preprocessing, the dataset was filtered to include only cases from the Punjab province.

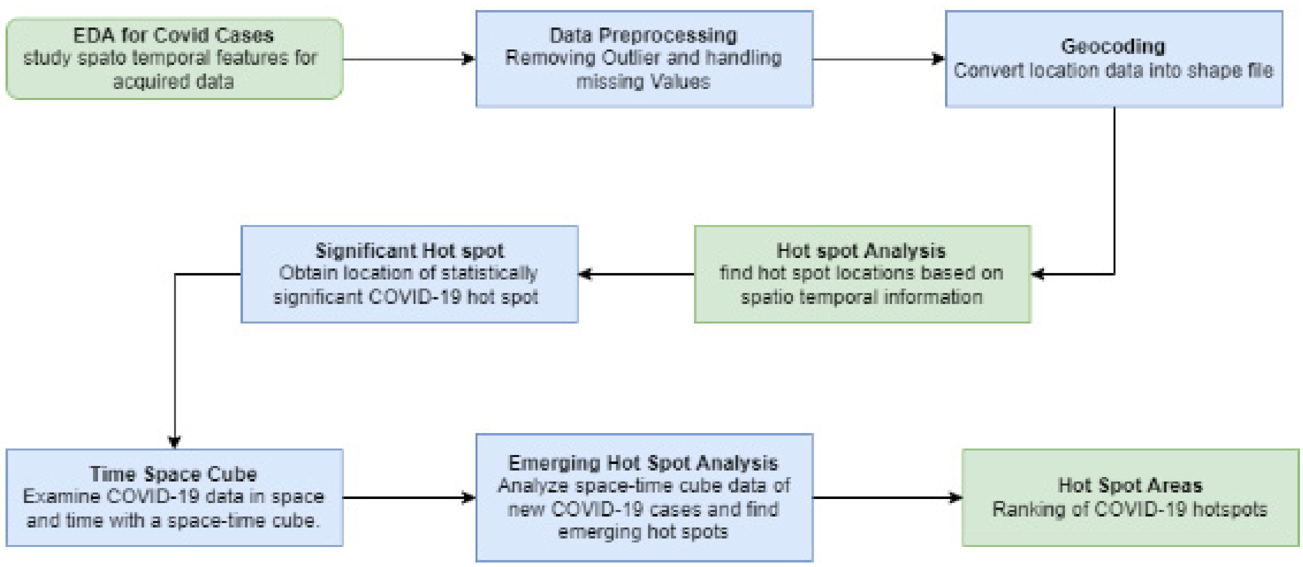

Figure 1: *Spatio-temporal analysis workflow for COVID-19 hotspot detection, outlining key steps from exploratory data analysis and preprocessing to geocoding, hotspot identification, time-space cube generation, and ranking of emerging hotspot areas.*

Links Researchers

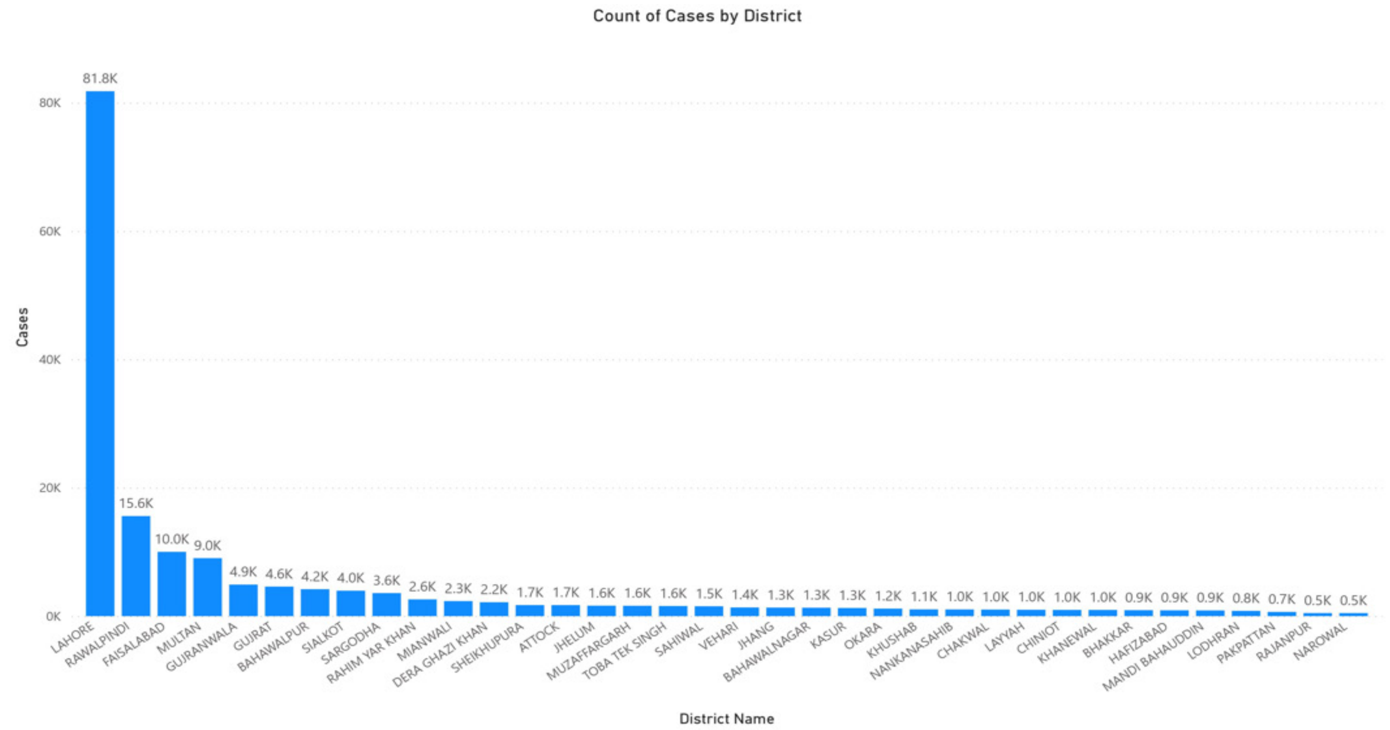

Figure 2: *COVID Cases by each district.*

Among the 8,573 records with missing district data, 7,217 had address fields available. Districts were inferred from these addresses using substring matching, as geocoding was not feasible due to inconsistent and improperly formatted address entries. For records where district information could not be recovered (7,252 cases), the entries were excluded from further analysis. Figure 2 presents the distribution of reported COVID-19 cases across districts in Punjab. Lahore emerged as the most severely affected district, with approximately 82,000 cases reported over the year, followed by Rawalpindi with 16,000 cases.

Hot spot analysis

Spatial data often exhibit spatial dependency, meaning that geographically proximate locations tend to share similar characteristics. However, many conventional statistical tests assume independence among observations (Svensson *et al.*, 2018). This study employs the Getis-Ord Gi* statistic to perform hotspot analysis across districts in the dataset. Importantly, a district with a high number of cases is not necessarily a statistically significant hotspot unless it is also surrounded by neighboring districts with similarly high case counts (Colak *et al.*, 2018).

We computed Moran's I statistic on weekly case aggregates to validate the spatial clustering assumption. The Moran's I values were significantly positive across most weeks ($p < 0.01$), indicating spatial dependence, thus justifying the application of Getis-Ord Gi*. To assess whether a district constitutes a statistically significant hotspot, we evaluate the z-scores and p-values associated with each district. A high positive z-score with a low p-value indicates a strong spatial clustering of high values (hotspot), while a low negative z-score with a low p-value signifies clustering of low values (cold spot). In other words, a district is classified as a hotspot if it shows both a high case count and is situated within a cluster of other high-case districts. The Getis-Ord Gi* statistic is calculated using the following Equation 1:

$$G_i^* = \frac{\sum_{j=1}^{t} W_{i,j} d_j - \overline{M} \sum_{j=1}^{t} W_{i,j}}{S \sqrt{\frac{\left[t \sum_{j=1}^{t} W_{i,j}^2 - \left(\sum_{j=1}^{t} W_{i,j} \right)^2 \right]}{t-1}}} \quad \ldots(1)$$

Where; dj represent the total cases in a district j, Wi, j represent spatial weight between districts i and j, t represents the total districts, and M is the mean value and S is the standard deviation for each district cases with j = 1 to n.

Each district in the dataset has a z-score as the Gi* statistic returned for it. The clustering of high values becomes denser for statistically significant positive z-scores as the z-score increases (hot spot). The clustering of low values is more pronounced for

Links Researchers

statistically significant negative z-scores the smaller the z-score (cold spot). We calculate the z score and p-value for the cases of corona and perform three experiments by changing the value of k (nearest number of neighbours) to observe the behaviour of z score.

Spatio-temporal trend analysis

For the analysis of spatiotemporal trend, we create space-time cube of the output (hot spot, cold spot, and not significant) received after applying Gi* on the corona weekly data. Space-time cubes illustrate the temporal evolution of corona cases within a specific district.

In a space-time cube, each cube represents the times in a week and each bin in the cube represents the specific district. The study reveals that the value of k = 3 is set to make the time-space cube of Gi* statistics output.

The selection of k = 3 was made after evaluating the outputs of the Getis-Ord Gi* analysis under different values of k, as discussed in the following section. Figure 3 illustrates the space–time cube constructed for the COVID-19 dataset, where each vertical column (bin) represents the aggregation of cases for a specific district during a given week. The height of each column corresponds to the temporal progression of the data, allowing both spatial and temporal patterns of disease spread to be visualized simultaneously. This enables the identification of emerging clusters, hotspots, and trends across different districts over time, providing valuable insight into how the pandemic evolved geographically and temporally.

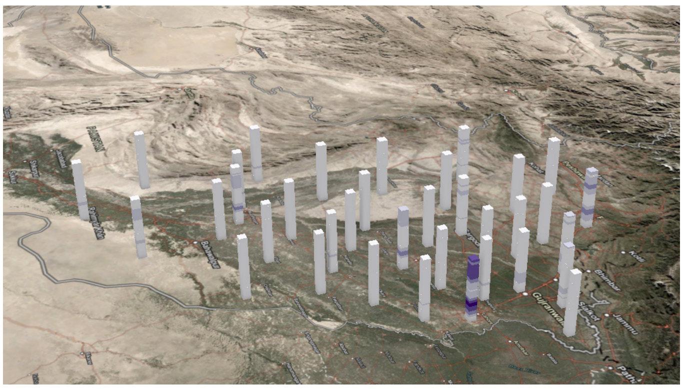

Figure 3: *Space-time cube representation of weekly COVID-19 case data across districts in Punjab, where each bin captures the spatio-temporal classification of a district as a hotspot, cold spot, or not significant based on Gi* z-scores and p-values.*

Spatio-temporal trends were analyzed using this space-time cube to identify emerging hotspot and cold spot districts. Each bin is associated with a z-score, p-value, and classification as a hotspot, cold spot, or not significant. The Mann-Kendall trend test was then applied to evaluate temporal patterns in these classifications across both space and time.

The resulting trend analysis enables the identification of evolving spatial clusters based on COVID-19 case counts and statistical significance. The Spatio-Temporal Trend Analysis algorithm categorizes districts into one of three cluster types, continuous, periodical, or irregular, based on their hotspot behavior, as summarized in Table 2.

Table 2: *Trend analysis clusters.*

Hot spot class	Description	Legend
Irregular pattern	A region that is neither a hot spot nor a cold spot	
Continuous hot spot	A region that is statistically significant hot spot for 90% of the weekly time window and there is no clear trend change in the cluster over time window.	
Periodical hot spot	A region that fluctuates between being a hot spot and an irregular pattern. It is a statistically significant hot spot for less than 90% of the weekly time window but there is no cold spot in this region over time.	

Results and Discussion

Three separate experiments were conducted using different values of k (number of nearest neighbors) to identify hotspot and cold spot districts, as outlined in the methodology section. The results of these experiments, along with the spatio-temporal trend analysis, are presented below. These results are grounded in the Gi* statistic defined in Equation 1, where each district receives a z-score indicating the degree of clustering of COVID-19 cases. The z-scores and corresponding p-values, computed for varying neighborhood sizes (k = 2, 3, 4), form the statistical basis for the hotspot and cold spot patterns illustrated in Figures 4–9.

Figure 4 displays the weekly Gi* analysis using k = 2. A choropleth map illustrates the spatial distribution of hotspots and cold spots, with subfigures (a), (b), (c), and (d) corresponding to weeks 1, 9, 23, and 2, respectively. The results clearly show that hotspot locations vary across weeks. Red indicates statistically significant hotspots at the 99% confidence level, orange at 95%, and light orange at 90%. Grey represents districts that are not statistically significant as either hotspots or cold spots, while blue shades indicate cold spots.

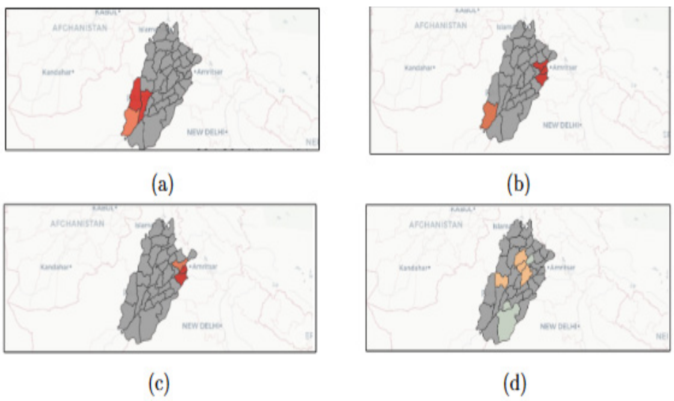

Figure 4: *Weekly Hotspot Distribution Using k=2. Red: Hotspot at 99% confidence, Orange: 95%, Light Orange: 90%, Grey: Not Significant, Blue Shades: Cold Spots.*

In week 1, Rawalpindi and its surrounding districts emerged as hotspot regions. By week 2, Sargodha and Faisalabad appeared as hotspots at the 90% confidence level, while Bahawalnagar and Lodhran were identified as cold spots. Notable shifts in hotspot intensity are observed in Lahore and its neighboring districts in weeks 9 and 23.

Figure 5 presents a 3D plot summarizing weekly hotspot patterns across all districts. The plot reveals that Lahore, Kasur, and Sheikhupura consistently remained hotspot areas for most of the year, with only brief periods of deviation spanning approximately 4 to 5 months.

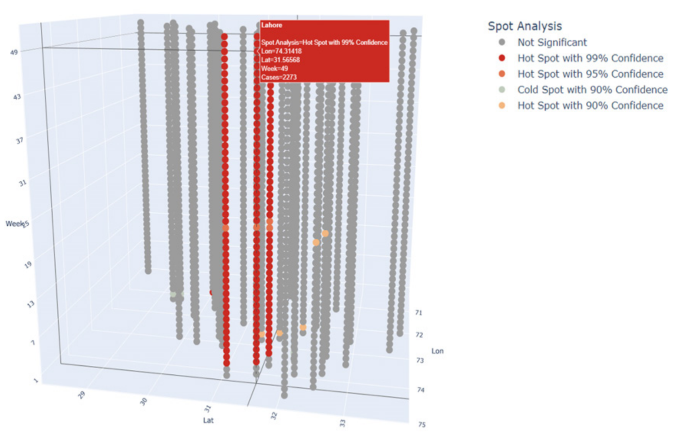

Figure 5: *3D interactive plot for visualizing weekly hotspots using 2 neighbours.*

Hence, hot spot areas are Rawalpindi and its neighbours during initial times, then, the hot spot areas moved to Lahore and its neighbours for the remaining year.

Furthermore, the second experiment is performed by using the k = 3 value to increase the experimental analysis. Now, hot spot and cold spot areas are calculated based on 3 nearest neighbors.

The results by using k = 3 are similar for k = 2 except for 1 week. Figure 6, (c) shows the Week 24, while, the results of other weeks are same. The experimental results reveal that the same pattern of hot spot districts has been followed.

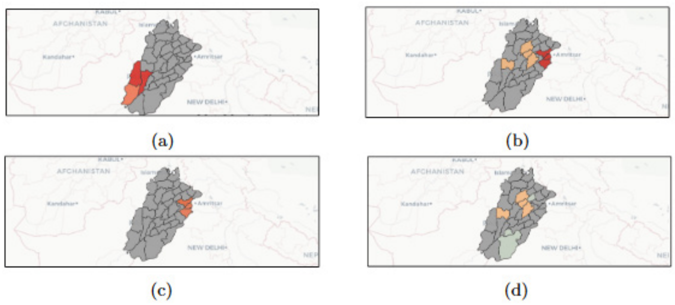

Figure 6: *Gi* weekly analysis using 3 neighbours.*

Initially Rawalpindi and its surroundings are hot spots and then it moves to Lahore and its surroundings. However, it intensifies the efficiency of the experimental results.

The 3D plot in Figure 7 for k = 3 is also similar to k = 2, i.e., Lahore and its adjacent districts are the hot spot areas throughout the whole year, approximately.

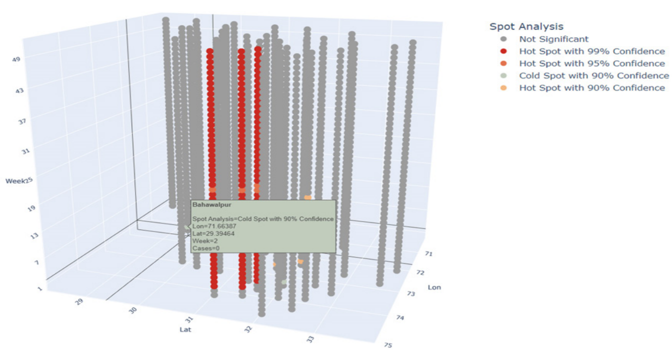

Figure 7: *3D interactive plot for visualizing weekly hotspots using 3 neighbours.*

For our third and last experiment to hit the hot spot and cold spot districts, we change the values of k as k = 4 shown in Figure 8. Here, z-score and p-values of each district depends on its four nearest neighbors.

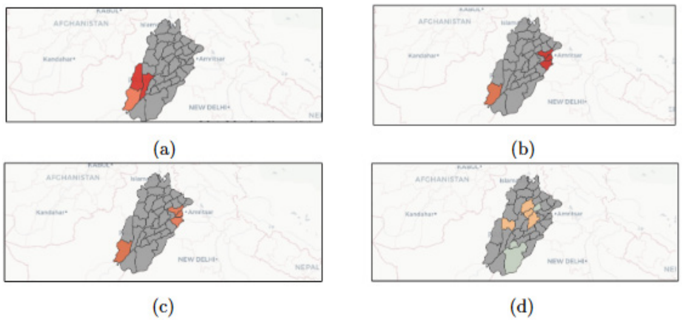

Figure 8: *Gi weekly analysis using 4 neighbours. The choropleth map for k = 4 is, also, similar as the maps by k = 2 and k = 3, approximately.*

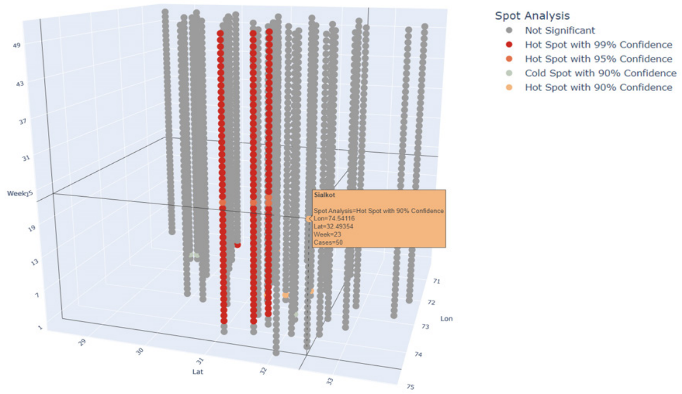

Figure 9: *3D interactive plot for visualizing weekly hotspots using 4 neighbours.*

The same pattern of hot spots and cold spots districts are identified by GI statistics with different values of k.

The 3D plots generated using k = 3 closely resemble those produced with other values of k, consistently identifying the same hotspot and cold spot districts. Across all three experiments, the results showed approximately 99% similarity, indicating minimal variation in hotspot detection regardless of the chosen neighborhood size. Based on this consistency, k = 3 was selected for the spatio-temporal trend analysis.

Figure 10 illustrates the results of the spatio-temporal trend analysis, categorizing districts into three types of hotspot patterns: irregular, continuous, and periodical. The analysis reveals that Narowal and Sheikhupura exhibited continuous hotspot behavior, remaining statistically significant hotspots with no change throughout the year. In contrast, Lahore and Kasur were identified as periodical hotspots, remaining hotspots for over 45 weeks, without being classified as cold spots during any time interval. This indicates that these districts experienced sustained high transmission rates, although not continuously significant across all weeks.

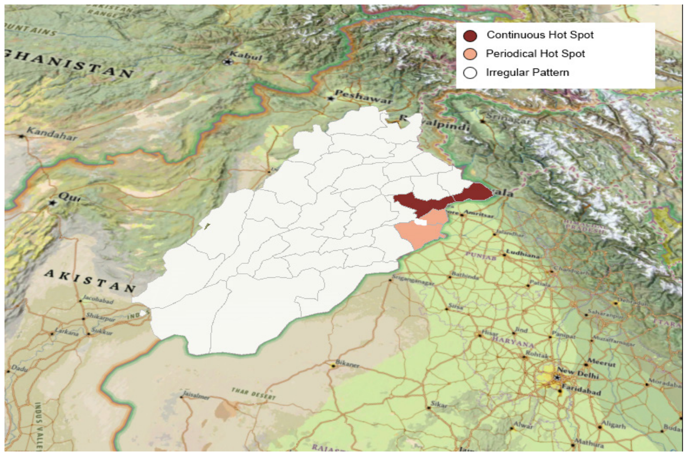

Figure 10: *COVID-19 Spatio temporal trend analysis.*

These findings underscore the value of spatio-temporal analysis in informing public health strategies. By identifying persistent and emerging hotspot regions, authorities can prioritize interventions, allocate resources more effectively, and implement targeted containment measures (Yin *et al*., 2025). The approach demonstrated in this study provides a robust framework for managing future outbreaks of similar infectious diseases.

The research highlights key districts, particularly Rawalpindi (in the initial weeks), followed by Lahore, Kasur, and Sheikhupura, as consistent high-risk areas. These hotspot shifts are reflected in the statistically significant Gi z-scores derived from Equation 1, where positive values indicated persistent clustering of high case counts. For example, Rawalpindi's high z-scores in early weeks shifted to Lahore and neighboring districts in later weeks, showing evolving clusters of COVID-19 transmission. Some districts, such as Kasur, alternated between significant hotspot and irregular patterns, suggesting fluctuating infection dynamics. While the granularity of district-level data limited full coverage, the categorization of hotspot types, grounded in Gi* significance testing, provides a detailed understanding of disease dynamics. This strengthens the case for using spatio-temporal hotspot analysis to guide targeted interventions, resource allocation, and future forecasting.

Conclusion

The presented study offers a comprehensive spatio-temporal analysis of COVID-19 case distribution at the district level across Punjab Province, Pakistan, over a one-year period. By leveraging spatial statistics, particularly the Getis-Ord Gi* algorithm, the study identifies statistically significant hotspot and cold spot districts and tracks their evolution on a weekly basis. To ensure robustness and validate spatial clustering, three separate experiments were conducted using different values for the number of nearest neighbors (k = 2, 3, and 4). The consistency of results across these experiments confirms the stability of identified patterns and strengthens the reliability of the hotspot detection process. Spatio-temporal trend analysis, performed using space-time cubes and Mann-Kendall trend tests categorized districts based on the persistence and frequency of hotspot occurrences. This added layer of analysis provides actionable insights into the trajectory of the outbreak, highlighting areas that require long-term mitigation measures versus those with episodic spikes. This study contributes a replicable analytical framework that can be applied to other regions or diseases where spatial clustering is a key feature of transmission. The results hold significant value for public health authorities by enabling evidence-based, geographically targeted resource allocation, and supporting strategic planning for outbreak control and prevention.

Acknowledgements

The authors thank the National Information Technology Board (NITB) of Pakistan for providing access to the COVID-19 dataset. We acknowledge Bahria University and the Center of Excellence in Artificial Intelligence (CoE-AI) for providing research infrastructure and computational resources. Special appreciation to the healthcare workers across Punjab Province who collected and reported the case data during the pandemic period.

Novelty Statement

This study introduces a dynamic spatio-temporal framework combining Getis-Ord Gi* statistics with Mann-Kendall trend tests to categorize COVID-19 hotspots into continuous, periodical, and irregular patterns across Punjab's districts. The approach advances beyond static analysis by using space-time cubes to simultaneously track spatial clustering and temporal evolution of disease transmission for targeted public health interventions.

Author's Contribution

Fida Muhammad: Conceptualized the research framework, designed the spatio-temporal analysis methodology, implemented the Getis-Ord Gi* statistical analysis, and contributed to manuscript writing and revision.

Resham Hafeez: Provided medical expertise on COVID-19 epidemiology, assisted in data acquisition from NITB, performed geocoding and address matching for missing district information, contributed to data quality assessment, and provided data interpretation from public health perspective.

Fatima Khalique: Led the research design, supervised the overall study, coordinated the spatio-temporal trend analysis using Mann-Kendall tests, and was responsible for manuscript preparation and final review.

Mehroz Sadiq: Performed data preprocessing, conducted exploratory data analysis, implemented the space-time cube generation, and contributed to results visualization.

Saba Mahmood: Reviewed and edited the manuscript, contributed to critical revision of the document, and assisted in final manuscript preparation.

Farah Akif: Reviewed and edited the manuscript, contributed to manuscript formatting, and assisted in document preparation.

Moazzam Ali: Assisted in literature review, reviewed and edited the manuscript, and helped in manuscript formatting and reference management.

Generative AI and AI-assisted technology statement

The authors declare that no generative artificial intelligence (AI) or AI-assisted technologies were used in the conception, design, data collection, analysis, interpretation, or writing of this manuscript. All research methodologies, statistical analyses, data processing, and manuscript preparation were conducted using conventional research methods and standard software tools without the assistance of AI-powered writing, analysis, or content generation platforms.

Conflict of interest

The authors have declared no conflict of interest.

References

Ali, H., Delair, R., Pham, D.T., Alam, S. and Schultz, M., 2020. Dynamic hot spot prediction by learning spatial-temporal utilization of taxiway intersections. 2020 International Conference on Artificial Intelligence and Data

Analytics for Air Transportation (AIDA-AT), pp. 1-10. https://doi.org/10.1109/AIDA-AT48540.2020.9049186

Ansari, N., Kabir, F., Khan, W., Khalid, F., Malik, A.A., Warren, J.L. and Nisar, M.I., 2024. Environmental surveillance for COVID-19 using SARS-CoV-2 RNA concentration in wastewater. A study in District East, Karachi, Pakistan. Lancet Reg. Health Southeast Asia, pp. 20. https://doi.org/10.1016/j.lansea.2023.100299

Bejon, P., Williams, T.N., Liljander, A., Noor, A.M., Wambua, J., Ogada, E., Olotu, A., Osier, F.H., Hay, S.I., Färnert, A. and Marsh, K., 2010. Stable and unstable malaria hot spots in longitudinal cohort studies in Kenya. PLoS Med., 7: e1000304. https://doi.org/10.1371/journal.pmed.1000304

Chohan, U.W., 2024. Coronavirus v. Pakistan: Early successes in the epidemiological battle. In: COVID-19 in South Asia. Routledge India. pp. 79-91. https://doi.org/10.4324/9781003477563-8

Colak, H.E., Memisoglu, T., Erbas, Y.S. and Bediroglu, S., 2018. Hot spot analysis based on network spatial weights to determine spatial statistics of traffic accidents in Rize Turkey. Arabian J. Geosci., 11: 1-11. https://doi.org/10.1007/s12517-018-3492-8

Ertunç, E., Çay, T. and Mutluoğlu, O., 2013. Intersection road accident analysis using geographical information systems: Antalya (Turkey) example. 2013 7th International Conference on Application of Information and Communication Technologies, pp. 1-5. https://doi.org/10.1109/ICAICT.2013.6722713

Jana, M. and Sar, N., 2016. Modeling of hot spot detection using cluster outlier analysis and Getis-Ord Gi statistic of educational development in upperprimary level, India. Model. Earth Syst. Environ., 2: 1-11. https://doi.org/10.1007/s40808-016-0122-x

Kalinic, M. and Krisp, J.M., 2018. Kernel Density Estimation (KDE) vs. HotSpot Analysis - Detecting Criminal Hot Spots in the City of San Francisco. Agile 2018 - 21st Conference on Geo-information Science.

Kertész, G., Felde, I., Szénási, S. and Nádai, L., 2019. Road accident black spot analysis to support autonomous vehicle control. 2019 International IEEE Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE), pp. 225-228. https://doi.org/10.1109/CANDO-EPE47959.2019.9110973

Khan, M.Z., Khan, M.U.G., Saba, T., Razzak, I., Rehman, A. and Bahaj, S.A., 2021. Hot-spot zone detection to tackle Covid19 Spread by fusing the traditional machine learning and deep learning approaches of computer vision. IEEE Access, 9: 100040-100049. https://doi.org/10.1109/ACCESS.2021.3094720

Kim, D., Cánovas-Segura, B., Campos, M. and Juarez, J.M., 2024. Visualization of spatial–temporal epidemiological data: A scoping review. Technologies, 12: 31. https://doi.org/10.3390/technologies12030031

Lai, C.C., Liu, Y.H., Wang, C.Y., Wang, Y.H., Hsueh, S.C. and Yen, M.Y., 2020. Asymptomatic carrier state acute respiratory disease and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARSCoV-2): Facts and myths. J. Microbiol. Immunol. Infect., 53: 404-412. https://doi.org/10.1016/j.jmii.2020.02.012

Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J. and Hsueh, P.R., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): The epidemic and the challenges. Int. J. Antimicrob. Agents, 55: 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924

Li, Y., Li, J., Dang, Y., Chen, Y. and Tao, C., 2024. Adverse events of COVID-19 vaccines in the United States: Temporal and spatial analysis. JMIR Publ, Hlth, Surv., 10: e51007. https://doi.org/10.2196/51007

Liu, Y., Gayle, A.A., Wilder-Smith, A. and Rocklöv, J., 2020. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J. Travel Med., 27: 1-4. https://doi.org/10.1093/jtm/taaa021

Mahmood, S. and Tariq, Z., 2023. Mapping Covid-19 incidence hotspots in Pakistan using spatial-statistical approach. Int. Geoinf. Days, 7: 286-289.

Manepalli, U.R., Bham, G.H. and Kandada, S., 2011. Evaluation of hot-spots identification using kernel density estimation and getis-ord on I-630. 3rd International Conference on Road Safety and Simulation.

Oh, J., Lee, M., Kim, M., Kim, H.J., Lee, S.W., Rhee, S.Y., Yon, D.K., 2024. Incident allergic diseases in post-COVID-19 condition: multinational cohort studies from South Korea, Japan and the UK. Nat. Commun., 15: 2830. https://doi.org/10.1038/s41467-024-47176-w

Pitalua Rodriguez, M.A., Mengel, S., Gittner, L.S. and Khan, H.M.K., 2019. Automated hot-spot identification for spatial investigation of disease indicators. 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), pp. 9-18. https://doi.org/10.1109/BigDataService.2019.00009

Ruswandi, B., Zulkifli, Z., Raspiani, R. and Sutanto, T.E., 2021. Detecting hot spots using GIS analysis: Case study of Indonesia's political opposition hashtag. 2021 9th International Conference on Cyber and IT Service Management (CITSM), pp. 1-7. https://doi.org/10.1109/CITSM52892.2021.9588911

Sreeram, S. and Shanmugam, L., 2018. Autonomous robotic system based environmental assessment and dengue hot-spot identification. 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC and CPS Europe), pp. 1-6. https://doi.org/10.1109/EEEIC.2018.8493849

Srikanth, L. and Srikanth, I., 2020. A case study on kernel density estimation and hot spot analysis methods in traffic safety management. 2020 International Conference on Communication Systems and Networks (COMSNETS), pp. 99-104. https://doi.org/10.1109/COMSNETS48256.2020.9027448

Svensson, V., Teichmann, S.A., Stegle, O., 2018. SpatialDE: Identification of spatially variable genes. Nat. Methods, 15: 343-346. https://doi.org/10.1038/nmeth.4636

Wang, Y., Chen, X. and Xue, F., 2024. A review of Bayesian spatiotemporal models in spatial epidemiology. ISPRS Int. J. Geo-Inf., 13: 97. https://doi.org/10.3390/ijgi13030097

Wang, Z., Wu, J. and Yu, B., 2011. Analyzing spatio-temporal distribution of crime hot-spots and their related factors in Shanghai, China. 2011 19th International Conference on Geoinformatics, pp. 1-6. https://doi.org/10.1109/GeoInformatics.2011.5981000

Xu, T., 2022. Demand analysis of taxi passenger-carrying hot spot areas based on XGBoost algorithm. 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 682-686. https://doi.org/10.1109/ICAIBD55127.2022.9820001

Yin, S., Liu, Z., Yu, S., Li, Y., An, J., Wang, D. and Luan, X., 2025. Geographic variations, temporal trends, and equity in healthcare resource allocation in China, 2010–21. J. Glob. Health, 15: 04008. https://doi.org/10.7189/jogh.15.04008

Zumla, A., Peiris, M., Memish, Z.A. and Perlman, S., 2024. Anticipating a MERS-like coronavirus as a potential pandemic threat. Lancet, 403: 1729-1731. https://doi.org/10.1016/S0140-6736(24)00641-X