Final Report

by Ornrakorn Mekchaiporn

Submission date: 05-Aug-2022 05:27PM (UTC+0700)
Submission ID: 1872467799
File name: Final_Report_without_ref.pdf (4.73M)
Word count: 6591
Character count: 35587

Spatiotemporal epidemiology and analysis of mental health conditions in Thailand

Abstract

Mental health disorders are one of the great health issues due to the high number of mental patients worldwide. In particular, 23% of the Thai population is mentally ill, but there are only a few who have access to prevention and treatments. The treatments' cost, lack of medical professionals, or discrimination from society can be a cause. Eventually, it will be a chronic and progressive course for those who do not have access to both prevention and therapy. The purposes of this study were : 1) to analyze the mental health disorders' spatial patterns, the results show that 8 out of 11 disorders have a spatial pattern. 2) to cluster the mental health disorders into a group using unsupervised machine learning. As a result, we divided 11 disorders into 3 groups; addiction (Alcoholism, Drug addiction(excluding Amphetamine addiction)) , mood disorder (Schizophrenia, Depression, Anxiety disorder), and brain disorder (Autistic disorder, Intellectual disabilities, and Learning disabilities). 3) to find the association between two disorders using a correlation matrix. In the discussion section, we discuss the results based on clinical evidence. The benefits of understanding both spatiotemporal patterns and the association between two disorders can be used to support mental health policies to reduce mortality, modify the public's attitudes towards mental patients, and allocate resources for mental health problems efficiently.

1. Introduction

In 2019, one in every eight individuals, or 970 million people worldwide, suffered from a mental condition, with anxiety and depression being the most common (1). Due to the COVID-19 pandemic, the number of individuals who suffer from anxiety and depression disorders greatly increased in 2020. Initial projections show that anxiety and major depressive disorders will rise by 26% and 28%, respectively, in just one year (2). In Thailand, any mental condition has a prevalence of 23.2%. However, there are far fewer people seeking treatment from mental health services (2). The Department of Mental Health, Ministry of Public Health (MOPH), reported that the average number of patients accessing mental health services in 2013–2021 was 2,500–4,000 people per day—or over a million people per year. That means one in every 70 people in Thailand has a mental condition and access to therapy (3). Although there are effective methods for both prevention and therapy, the majority of those who suffer from mental illnesses do not have access to them. Stigma, prejudice, and human rights violations are also commonplace (4).

A clinically significant impairment in a person's intellect, emotional control, or behavior is known as mental health. It is typically linked to distress or functional impairment in key areas. Mental diseases come in many different types. The term "mental health problems" can also be used to describe mental diseases. The latter is a more general phrase that encompasses mental disorders, psychosocial impairments, and (other) mental states connected to considerable suffering, functional disability, or danger of self-harm. This information sheet focuses on mental disorders as defined by the 11th revision of the International Classification of Diseases (ICD-11)(5).

Stigmatization of mental disorders has a severe, harmful effect on those who experience them, their family members, communities, and society as a whole. Public perceptions of mental disorders as being closely associated with danger and violence are frequently influenced by inaccurate and biased media representations, and they have a significant impact on the stigma associated with mental disorders (6). In a cross-sectional study from five distinct areas of Thailand, it was discovered that stigma and attitudes regarding mental illness had a major impact on whether or not people chose to use a mental health service (7). Thai society has a significant presence of animist and paranormal ideas. According to this belief system, nonliving entities can cause mental illness, which also contributes to the stigmatization of those who suffer from mental illness (8).

There are a few small spatiotemporal studies on mental disorders in other countries. For example, Chien-Yu Lin, using self-harm data, investigated the spatial distribution of self-harm incidence rates, their socioeconomic correlates, and sex/age differences (9). In particular, there were limited resources, and there were no spatiotemporal association studies between the disorders in mental health in Thailand before. Due to the impact of mental disorders, which should be a serious problem, we decided to study the spatiotemporal epidemiology of mental disorders in Thailand. The study includes the analysis in many ways—spatiotemporal analysis of each disorder using cluster detection technique (global and local autocorrelation: Moran's I), the spatiotemporal association between the disorders using Spearman correlation matrix, and unsupervised machine learning: K-prototypes clustering and agglomerative hierarchical clustering. We apply all these techniques to investigate the reported mental health service cases collected from the health data center (HDC), Department of Mental Health, Ministry of Public Health (MOPH), Thailand. In the discussion section, we discuss the results based on scientific evidence (9).

This study aims to advocate government policies that allocate resources to those regions where people seek mental health services, help Thai people realize the importance of mental disorders, reduce the stigma associated with mental disorders, modify attitudes and behaviors among the public interacting with those who have mental disorders under equitable and parity conditions, and increase access to mental health treatments (10).

2. Materials and Methods

2.1 Data source

The study is a retrospective analysis of reported mental health service cases collected from the health data center (HDC), Department of Mental Health, Ministry of Public Health (MOPH). All the cases were aggregated yearly mental health service cases notified during the years 2015 to 2021 at the provincial level, classified into eighteen categories in the health data center, and only eleven categories were used in the study (Dementia, Alcoholism, Drug addiction (excluding Amphetamine addiction), Schizophrenia, Depression, Anxiety disorder, Intellectual disabilities, Learning disabilities, Autistic disorder, Self-harm, and Epilepsy). ICD-10 mental disorders were assigned the codes F00.X - F99.X. Case report data includes the year and province of diagnosis. Thailand's mid-year population statistics were collected from official statistics registration systems. This study was conducted in Thailand, which collected data on a population of approximately 70 million people across 77 provinces with different geographies. The geographic coordinates and provincial boundary data are obtained from the GEO package file in the Global Administrative Region Database (GADM), a high-resolution territorial database containing provincial to subdistrict data for all countries across the world. Files for analysis are compiled using the Python programming language.

2.2. Spatial Pattern Analysis

The presence of systematic spatial variation in a mapped variable is referred to as spatial autocorrelation (11). It is used to test the spatial relations between near and far and to present the distribution of values. It is also known as statistics. The tendency for nearby areas or sites to have similar values means that geographically nearby values of a variable tend to be similar on a map, which is a positive spatial autocorrelation (12). Negative spatial autocorrelation, on the other hand, is characterized by dissimilar variant values next to each other (13). There are several statistical techniques for detecting its presence: Geary's C, Getis and Ord's G, and Moran's I. The Moran's I technique was chosen for this study to detect clusters.

2.2.1 Spatial Contiguity Matrices

A measure of contiguity is required in spatial autocorrelation analysis. Because spatial contingency, or weight matrix, W means that two spatial units share a common border of non-zero length (14), and they are used to define provincial connectivity relationships (neighborhood contingency). The weight matrix is an adjacency matrix that calculates the distance between province pairs (i,j) by conditions as spatial weights w_{ij} are 1 when i and j are neighbors, otherwise 0. which is a reason for calculating the weight to be computed in the equations of Global and Local Moran's I.

$$w \in \{w_{ij}\} \qquad (1)$$

$$w_{ij} = \begin{cases} 1, pair\ (i,j)\ is\ reachable \\ 0, otherwise \end{cases}$$

$$i, j \in \{n | 0 \leq n < N;\ n, N \in \aleph\}$$

Normally, the contiguity matrix is mainly split into 3 types: rook's case, bishop's case, or queen's (king's) case (13,15). The rook case is defined by a neighborhood between 4 locations adjacent to each cell (focus on horizontal and vertical joins only). Bishops only join diagonals of the relationship (joins with diagonal joins only). The queen is encompassed by neighbors as spatial units sharing a common edge or a common vertex (joins with both rook and bishop joins). As shown in Figure 1, from the nature of provinces in Thailand, there are borders of interconnected areas. This made the queen pattern appropriate for this study.

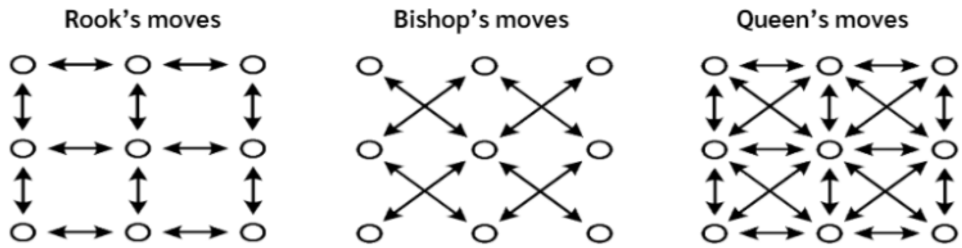


Figure 1: Type of neighborhood relations.

2.2.2 Global Spatial Detection (Global Moran's I)

Global Moran's I is a statistic that measures variable spatial autocorrelation. Correlation calculations based on location and magnitude features, which are cross-products for measuring attribute association, Finally, when attribute association is computed in Global Moran's I, the final result can be used to estimate the strength and determine whether the spatial correlation is positive or negative. It is shown as a clustered, fragmented, or random pattern (16). This statistic employs cross-products to assess attribute associations. The statistics equation is as follows:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \quad i \neq j \qquad (2)$$

is the number of provinces, x_i and x_j are the variable of interest at locations of i and j, $\bar{x}$ is the mean of x, and w_{ij} is the spatial weight value in the spatial matrix between locations of i and j(17).

Moran's I will have a value between -1 and 1. When the value is close to 1, it means that values in neighboring positions tend to cluster, whereas when the value is close to -1, it means that values in neighboring positions tend to cluster. The pattern has perfect dispersion, which means that high values are close to low values. Moran's I approaches zero, indicating that the data are distributed randomly (17). Furthermore, when p-value statistical significance indicates that the null hypothesis should be rejected, it means that values in adjacent positions tend to cluster similarly, with the value being closer to 1. The null hypothesis for this tool states that the values associated with features are randomly distributed (16). Only the spatial distribution patterns of Global Moran's I are known in spatial autocorrelation, but not the regions in which they occur. A deeper spatial interpretation is required for local Moran's I.

2.2.3 Local Spatial Detection (Local Moran's I)

The local Moran statistic was suggested by Anselin (1995). This is a spatial autocorrelation that can be used to evaluate the autocorrelation within local neighborhoods at the local level. Local measures produce values that represent the degree of spatial autocorrelation or identify spatial clusters and outliers of a variable (13). The statistical equation is as follows:

$$I_i = \frac{n\left(x_i - \bar{x}\right)\sum_{j=1}^{n} w_{ij}\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \tag{3}$$

where these notations are as described for Equation (2), but the values are from the local neighboring region.

To represent the result of local moran's I as a map that shows locations with significant statistics, is defined as a Local Indicators of Spatial Autocorrelation (LISA) cluster map. From the concept of LISA, it has two requirements. First, it provides an indication of the extent of spatial clustering of similar values with an assessment of significance. Second, it reveals a proportional spatial association between the sum of local statistics and global statistics (18,19). The categories describe the meaning of each cluster: Two categories are meant to suggest clusters (High-High, Low-Low), which use the Moran scatter plot. One can describe the value of this location as high, and nearby it is also high. On the other hand, if this location is low, then the surrounding location is also low for each of the two categories meant to be suggested outliers (High-Low, Low-High). That means the values of this location and the surrounding area are opposite to each other (17).

2.3 Spearman's rank correlation coefficient

The Spearman's correlation coefficient is the nonparametric measure of a non-linear and monotonic relationship between two ranked variables (20). In this study, the variables mean the disorder, so we use the Spearman's correlation coefficient to measure a relationship between two disorders. This method returns values between -1 and 1. For interpretation, if the value is closer to +1 or -1, the two disorders will have a more monotonic relationship to each other (20,21). A value of -1 means a perfect negative association between two disorders. A value of 1 means a perfect association of disorders. A value of 0 means there is no association between two disorders. The Spearman's correlation coefficient is used when variables are ordinal or variables aren't normally distributed (22). The statistical equation is as follows:

$$R_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)} \tag{4}$$

where R_s is spearman's rank correlation coefficient, Σd_i^2 is a sum of the squares between x-variable and y-variable ranks, and n is the number of data points of two disorders (23).

2.4 Cluster analysis

Cluster analysis is an unsupervised machine learning technique used to group similar characteristics of data (24,25). In this study, the clustering algorithms used were K-prototype Clustering and Agglomerative Hierarchical Clustering.

2.4.1 K-prototype Clustering

The k-prototypes algorithm is a cluster analysis method derived from the k-means and k-modes algorithms. Therefore, the k-prototypes algorithm can cluster with mixed numerical and categorical values (26–28). K-means handle the numerical values, and k-modes handle the categorical values. These clustering algorithms are unsupervised learning, which is used with unlabeled data to find groups in the dataset. In this study, the algorithm is used to find groups among the disorders. The algorithms assign each disorder to one of the K groups. The disorders are clustered based on the number of patients accessing mental health services per 100,000 population in every province (2015-2021) of the disorder. K-means uses mathematical measures (distance) to cluster numerical values (the number of patients accessing mental health services per 100,000 population). The lesser distance indicates more similarity. The centroids of each group are updated by this means. But for categorical values (the disorders' names), the distance cannot be mathematically measured. The k-modes algorithm, which uses modes instead of means, uses the dissimilarities (total mismatches) between the data points (disorders). The fewer dissimilarities indicate more similarity (29,30).

Tools for finding the optimal number (K) of clusters

1) Elbow method

The optimal number of clusters is a fundamental step for any unsupervised algorithm, and one of the most popular methods is elbow (31). The Elbow Method is a method used to determine the best number of clusters by comparing the number of clusters with the sum of squared errors (SSE). The statistical equation of SSE can be expressed as :

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad (5)$$

Where, y_i is a value of instance (the number of patients accessing mental health services per 100K population in every region (2015-2021) of the disorder) at i and $\bar{y}$ is a mean of instance. The sum of squared errors (SSE) is the sum of the average euclidean distance between each observed value and the centroid (32,33).

The optimal point selection is to select the least K that makes the lowest SSE value or the number of errors does not change significantly. It can also be observed by looking at the point where the graph looks like an elbow. The Elbow Method is said to measure the sum of the distances between the object (the disorder) and the centroid. Where the number of K increases, the SSE is significantly reduced. When the resulting SSE decreases, the slope of the curve will be smooth (smooth) and will form an angle that looks like an elbow. Which at this elbow point will be the point shown, the value of the optimal number of clusters (K).

2) Silhouette coefficient

The Silhouette coefficient is yet another of the most well-liked techniques for assessing the caliber of clustering. The difference between group separation and cohesiveness divided by the greater of the two is effectively the Silhouette coefficient. A model with more coherent groups has a greater silhouette coefficient, which runs from 1 to 1. In other words, samples are remote from nearby clusters if the silhouette coefficient is close to +1. It indicates that the disorders are clearly isolated from other groups. Negative values suggest that the instance may have accidentally been allocated to the incorrect cluster, while a value of 0 indicates that the instance is on or very near to the decision border between two closest neighbors (34).

Group separation (b) and group cohesion (a) must be determined in order to determine the silhouette coefficient. In contrast to group separation, which refers to the average distance between an item (the disorder) and all other items in the closest group, cluster cohesiveness measures the average distance between an item (the disorder) and all other items within the same group. The silhouette coefficient may be determined as shown below (34).

$$S = \frac{b-a}{max(a,b)} \quad (6)$$

Distance Measure ; Huang's Distance

Huang (1997) defined a distance measure for mixed variable data by combining the square Euclidean distance for numeric variables with a simple matching distance for categorical variables. In this study, the numeric variables are the number of patients accessing mental health services per 100,000 population in every province (2015-2021) for each disorder, and the categorical variables are the disorders' names.

$$d_{ij} = d_{ij}^N + \gamma d_{ij}^C \quad (7)$$

Where d_{ij}^N is the distance between numeric variables (the number of patients accessing mental health services), d_{ij}^C is the distance between categorical variables (the disorders' names),

$$d_{ij}^N = \sum_{k=1}^{P_n} (x_{ik} - x_{jk})^2 \quad (8)$$

$$d_{ij}^C = \sum_{k=1}^{P_c} \delta_c(x_{ik}; x_{jk}) \quad (9)$$

P_n are the number of numeric (the number of patients accessing mental health services) and P_c are the number of categorical variables (the disorders' names),

$\delta_c(x_{ik}; x_{jk})$ is the simple matching distance between object i and j in the categorical variable k (the disorders' name) is given as,

$$\delta_c(x_{ik}; x_{jk}) = \begin{cases} 0, when\ x_{ik} = x_{jk} \\ 1, when\ x_{ik} \neq x_{jk} \end{cases} \quad (10)$$

Then, the Huang (1997) distance between objects i and j is calculated by (35,36).

$$d_{ij} = \sum_{k=1}^{P_n} (x_{ik} - x_{jk})^2 + \sum_{k=1}^{P_c} \delta_c(x_{ik}; x_{jk}) \quad (11)$$

2.4.2 Agglomerative Hierarchical Clustering (AHC)

Hierarchical clustering or hierarchical cluster analysis (HCA) is a method of cluster analysis (37,38). There are two methods of hierarchical clustering, Agglomerative and divisive. In this study, we used the agglomerative hierarchical clustering method. In agglomerative hierarchical clustering, clusters are built from the bottom up (39). This clustering starts with each data point (or each disorder) as a cluster and then calculates the distance between the disorders by merging the two closest disorders to create a cluster and continues to do this until it eventually becomes a single large cluster (37,40). Hierarchical clustering can be represented by a dendrogram (37). A cut-point in the dendrogram is the level of similarity that is cut to obtain a cluster of the data (41). For the best cut-point, it is hard to claim, because there is no ground truth to which one could refer. Therefore, internal cluster validation is required (42). There are several methods to measure the distance between clusters for merging clusters, such as single linkage, complete linkage, average linkage, centroid method, and Ward's method (43). In this study, we used Ward's method.

1) Ward's method

Ward's method is based on a sum of squared errors (SSE) rationale that only works for the Euclidean distance between observations (44). In this study, it is the Euclidean distance between two disorders. The Euclidean distance is the measure of the distance between two points (45). If the two points are very similar, then it means that each piece of data will be very close to the other. This brings the euclidean value closer to zero. The Euclidean distance formula is shown in Eq. 13. Ward's method merges two clusters, which results in the smallest increase in the total sum of squared errors (43,44). Ward's Method is most suitable for quantitative variables (46). Ward's method formula according to Eq. 14.

$$d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]} \qquad (12)$$

Where d is the euclidean distance between two disorders, (x_1, y_1) is the coordinate of the first point (first disorder), and (x_2, y_2) is the coordinate of the second point (second disorder) (47).

$$\Delta(A, B) = \frac{n_A n_B}{n_A + n_B} \|m_A - m_B\|^2 \qquad (13)$$

Where m_j is the center of cluster j, n_j is the number of points in it, Δ is called the merging cost of combining the clusters A and B, and $\|m_A - m_B\|$ is the Euclidean distance between the two cluster centers (44,48,49).

3. Results

3.1 Descriptive analytics

There are 9,496,342 cases in the patient data for 11 mental health services collected from the Department of Mental Health's data system between 2015 and 2021. As shown in Figure 2, the total number of cases is calculated per 100,000 population and displayed in line chart format. The graph's data can be displayed in two periods: The number of mental health services provided prior to the COVID-19 outbreak until COVID-19 was detected (2015-2019). The number of patients accessing mental health services for Depression, Autism, Alcoholism, and Drug addiction (excluding Amphetamine addiction) increased steadily until 2018 and then decreased. The COVID-19 period's alcohol-control measures are expected to reduce the number of alcoholism cases. The second section displays the number of patients accessing mental health services following the detection of COVID-19 (2020–2021), found that the number of patients Depression, Alcoholism, and Drug addiction are expected to rise slightly, in 2019

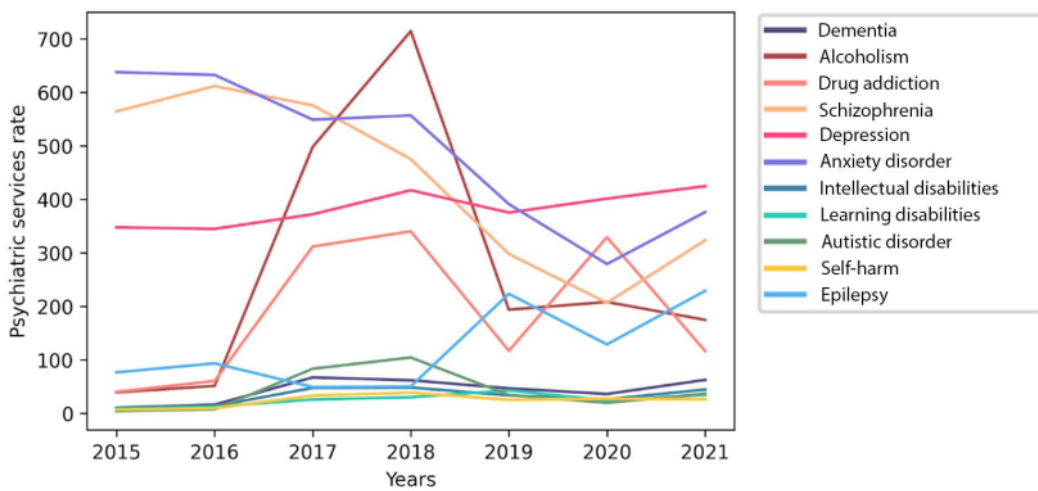


Figure 2: The patients accessing mental health services rate in Thailand since 2015-2021 (per 100k population).

The graph below depicts the percentage of the number of patients accessing mental health services in Thailand calculated per 100,000 population from 2015 to 2021. The total number of patients is 9,496,342 and is shown in bar chart format. Anxiety disorders, Schizophrenia, and Depression were the three disorders with the most mental health patients, with 2,259,434, 2,016,200, and 1,773,618 patients, respectively. Intellectual disabilities, Learning disabilities, and Self-harm were the three disorders with the fewest mental health patients, with 110,469, 118,088, and 147,807 patients, respectively.

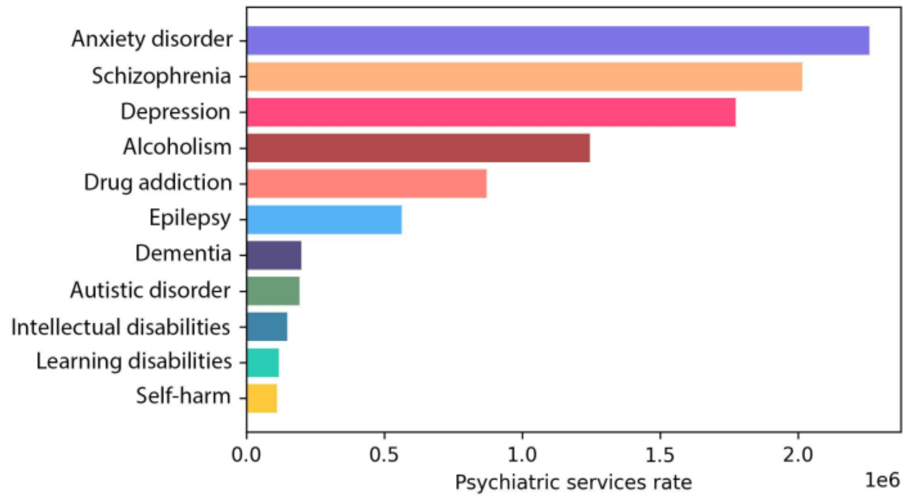


Figure 3: Total of the number of patients accessing services in Thailand since 2015-2021 (per 100k population).

Mapping

Figure 4 depicts the study region's geographical location, namely the provinces of Thailand, which has a total of 77 provinces. The number of patients accessing mental health services for all 11 mental health services from 2015 to 2021 was: Alcoholism, Drug addiction (excluding Amphetamine addiction), Schizophrenia, Depression, Anxiety disorder, Autistic disorder, Intellectual disabilities, Learning disabilities, Dementia, Self-harm, and Epilepsy. A geographical heat map was used to present the spatial distribution of the number of patients over a seven-year period, allowing the mapping to highlight the geographic distribution of disease-identifying mental disorders using different colors. The darker the color displayed in a province, the greater the amount of traffic in that province. The GEO package file in the Global Administrative Region Database (GADM) Version 3.6 was used to create all the maps.

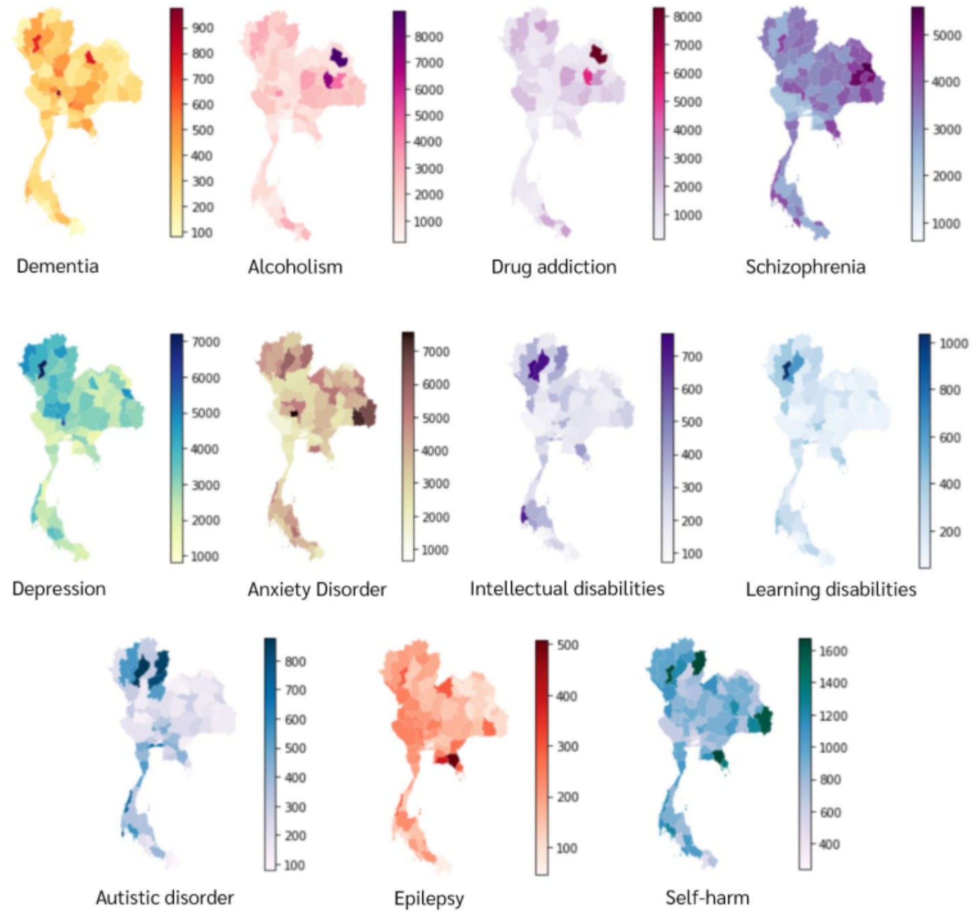


Figure 4: The map Shows the location of mental health services rate at region, namely the provinces of Thailand, which has a total of 77 provinces.

3.2 Unsupervised

3.2.1 Correlation matrix

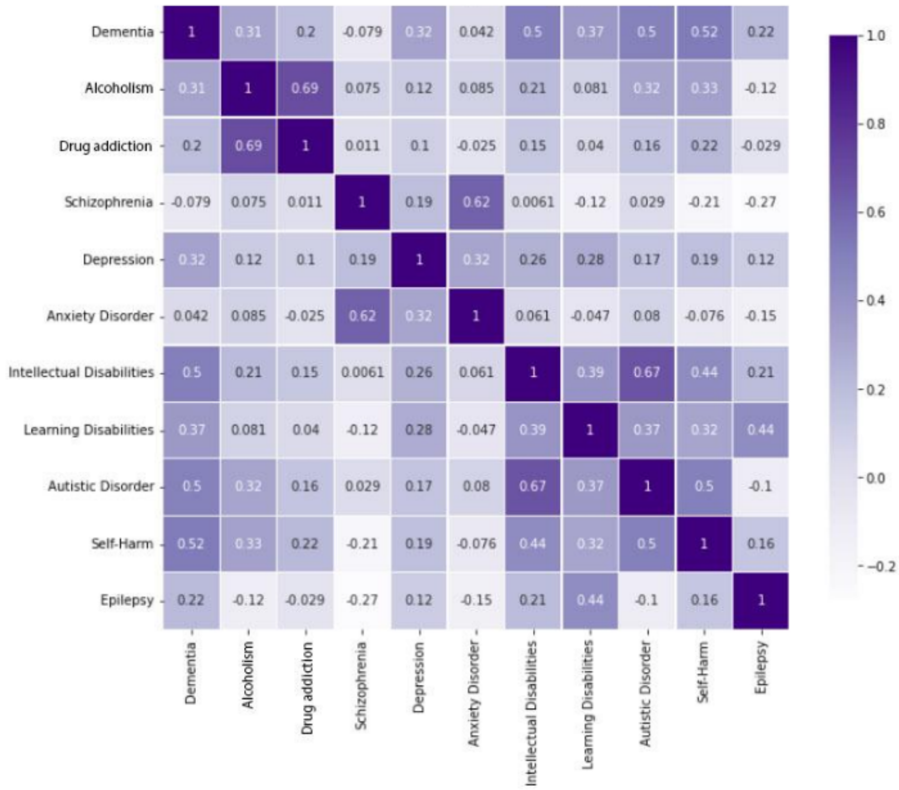
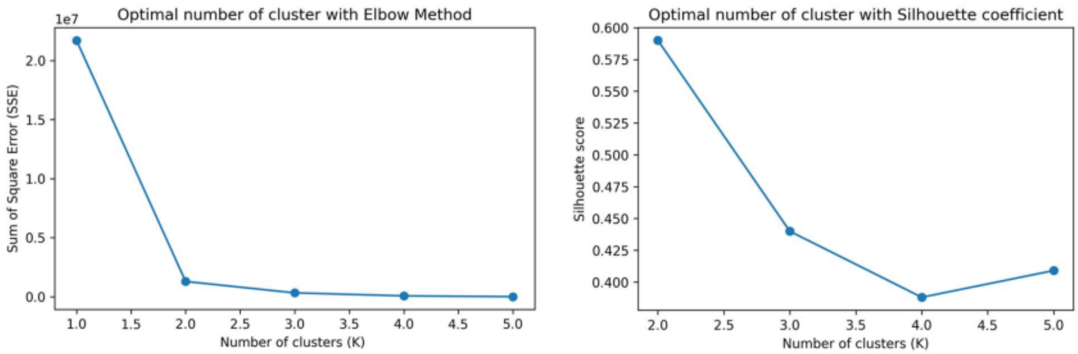


Figure 5: Total correlation of patients accessing mental health services in Thailand since 2015-2021 (per 100k population).

The correlation matrix shows 3 groups of disorders that have a positive linear relationship (correlate) with each other. The first group includes Alcoholism and Drug addiction (excluding Amphetamine addiction). The second group includes Schizophrenia, Depression and Anxiety disorder. The last group includes Intellectual Disabilities, Learning Disabilities, Autistic disorder and Self-harm. Dementia and Epilepsy are not highly correlated with the rest of the disorders.

3.2.2 K-Prototype

Below are the results of finding the optimal number of clusters using the Elbow method (Figure 6A) and the Silhouette coefficient (Figure 6B). In the elbow method, we can see a bend where the graph looks like an elbow at K = 2 and K = 3 in the left graph. In the right graph, it shows that the group of disorder is well-separated from other groups at K = 2 and K = 3 due to the higher silhouette coefficient than the other number of clusters (K). It indicates that 2 and 3 are the optimal number of clusters.



(A) (B)

Figure 6: The results of (A) finding the optimal number of clusters for K-prototype clustering using Elbow method and (B) Silhouette coefficient.

We applied K-Prototype to cluster the mental disorder. The algorithm clusters the disorders into groups based on the similarity of the number of patients accessing mental health services (per 100,000 population in every province 2015-2021) and the dissimilarity of the disorders' names. When using K = 2, it divides 11 disorders into 2 groups. The first group includes Schizophrenia, Depression and Anxiety disorder, and another group includes the rest. The result of the K-Prototype using K = 3 as represented in Figure 7 is quite similar to the correlation matrix, which also clusters disorders into 3 groups. In the correlation matrix, Dementia and Epilepsy are not highly correlated with the rest of the disorders. For the K-prototype clustering result, Dementia was assigned to the same group of Intellectual disabilities, Learning disabilities, Autistic disorder and Self-harm. Epilepsy was assigned to the same group of Alcoholism and Drug addiction (excluding Amphetamine addiction). As we can see, the first group using K = 2 and K = 3 has the exact same disorders. That means Schizophrenia, Depression and Anxiety disorder are well-separated from other disorders and have a comparable number of patients accessing mental health services to each other.

Schizophrenia Depression Anxiety Disorder	Dementia Intellectual Disabilities Learning Disabilities Autistic Disorder Self-Harm	Alcoholism Drug addiction Epilepsy

Figure 7: The result of K-Prototype.

3.2.3 Agglomerative hierarchical clustering

We applied agglomerative hierarchical clustering and Ward's method to cluster the mental disorders. As a result, the dendrogram in Figure 8 shows two major branches, A and B, by using a line cut-point to obtain clusters. We found three clusters, represented in Figure 9. The first group includes Schizophrenia, Depression and Anxiety disorders. Subsequently, the second group includes Dementia, Intellectual disabilities, Learning disabilities, Autistic disorder and Self-harm. The third group includes Alcoholism, Drug addiction and Epilepsy.

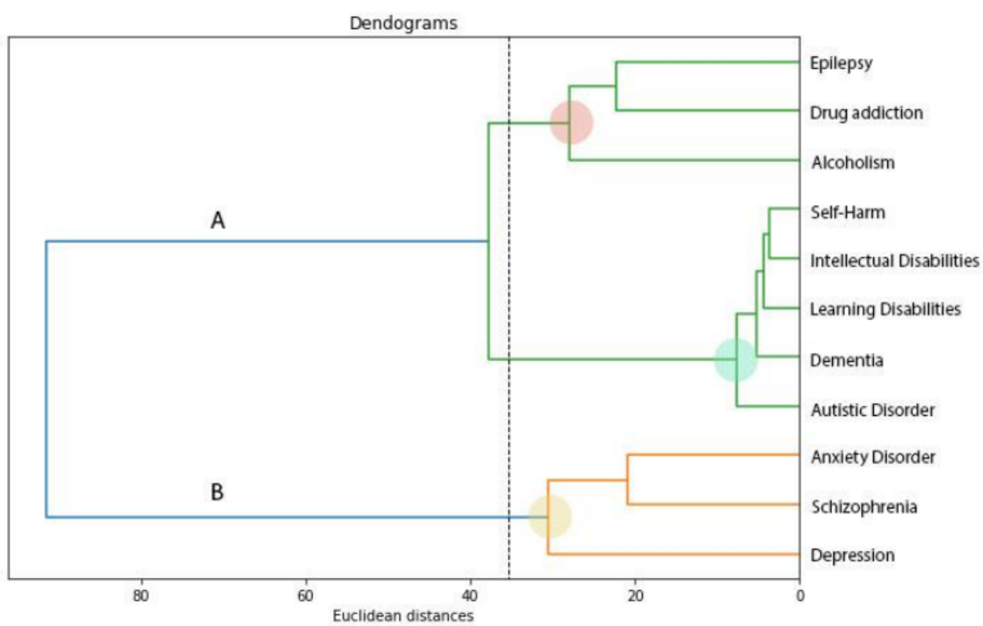


Figure 8: The Dendrogram of mental disorder shows two major branches: A and B by using line cut-point to obtain three clusters.

Schizophrenia Depression Anxiety Disorder	Dementia Intellectual Disabilities Learning Disabilities Autistic Disorder Self-Harm	Alcoholism Drug addiction Epilepsy

Figure 9: The result of Agglomerative hierarchical clustering.

3.3 Spatial Pattern Analysis

3.3.1 Global Spatial Detection

According to the description, Global Moran's I is a statistic used to measure variable spatial autocorrelation. The table below shows the value and p-value of Moran's I hypothesis test of mental health services accession in Thailand over a 7-year period from 2015 to 2021, calculated per 100,000 population. The GEO package file in the Global Administrative Region Database (GADM) Version 3.6 was used to create this analysis. According to Table 1's interpretation, there are 9 disorders with a p-value of 0.05 (significant), indicating that we can find the pattern of cluster or location in neighboring positions that tend to cluster but were unable to clearly identify the area.

However, there is a slight difference when looking at the yearly calculated p-value data. That is, the yearly p-value has an interesting cluster pattern and exceeds several years in only 8 disorders: Alcoholism, Drug addiction (excluding Amphetamine addiction), Schizophrenia, Depression, Anxiety disorder, Autistic disorder, Intellectual disabilities, and Learning disabilities. It is divided into three subgroups: addiction, mood disorder, and brain disorder. This is why the researchers decided to investigate the relationship by calculating the correlation coefficient between seven disorder pairs, as shown in Figure 11.

Diorders	I	P-value	Interpretation
Dementia	0.18973	0.008	clustered
Alcoholism	0.06467	0.121	random
Drug addiction	0.02064	0.26	random
Schizophrenia	0.48678	0.001	clustered
Depression	0.29424	0.001	clustered
Anxiety Disorder	0.24519	0.001	clustered
Intellectual Disabilities	0.28906	0.001	clustered
Learning Disabilities	0.22704	0.004	clustered
Autistic Disorder	0.33164	0.001	clustered
Self-Harm	0.28466	0.001	clustered
Epilepsy	0.23111	0.004	clustered

Table 1 The value and p-value of Moran's I hypothesis test of patients accessing mental health services in Thailand during 2015-2021 (per 100k population).

3.3.2 Local Spatial Detection

This map was created using the LISA cluster technique and analysis. It was used to assess the autocorrelation within local neighborhoods at the local level. Local measures can be used to identify spatial clusters and spatial outliers of each disorder. The interpretation of the defined color results is described in the method's section 2.2.3 Local Spatial Detection (Local Moran's I). The GEO package file in the Global Administrative Region Database (GADM) Version 3.6 was used to create all the maps.

Figure 10 shows a cluster pattern of detectable mental health services for each disorder in Thailand. For spatial interpretation, it was found that most clusters in the northern region had access to all 7 disorders, namely Depression, Anxiety disorder, Intellectual disabilities, Learning disabilities, Autistic disorder, Self-harm and Epilepsy. In the northeast region, there were clusters of the number of patients using mental health services for all 4 disorders, namely Alcoholism, Drug addiction, Depression and Schizophrenia. In the central region, there was a cluster of patients accessing mental health services for all 2 disorders, namely Dementia and Self harm. For the cluster, there were 2 clusters of patients accessing mental health services in the southern region: Anxiety disorder and Learning disabilities. In the eastern region, there is also a cluster of patients accessing mental health services for 2 disorders: Autistic disorder and Epilepsy. The last area, the western region, had a cluster of patients accessing mental health services only for learning disabilities. As shown in Figure 10,

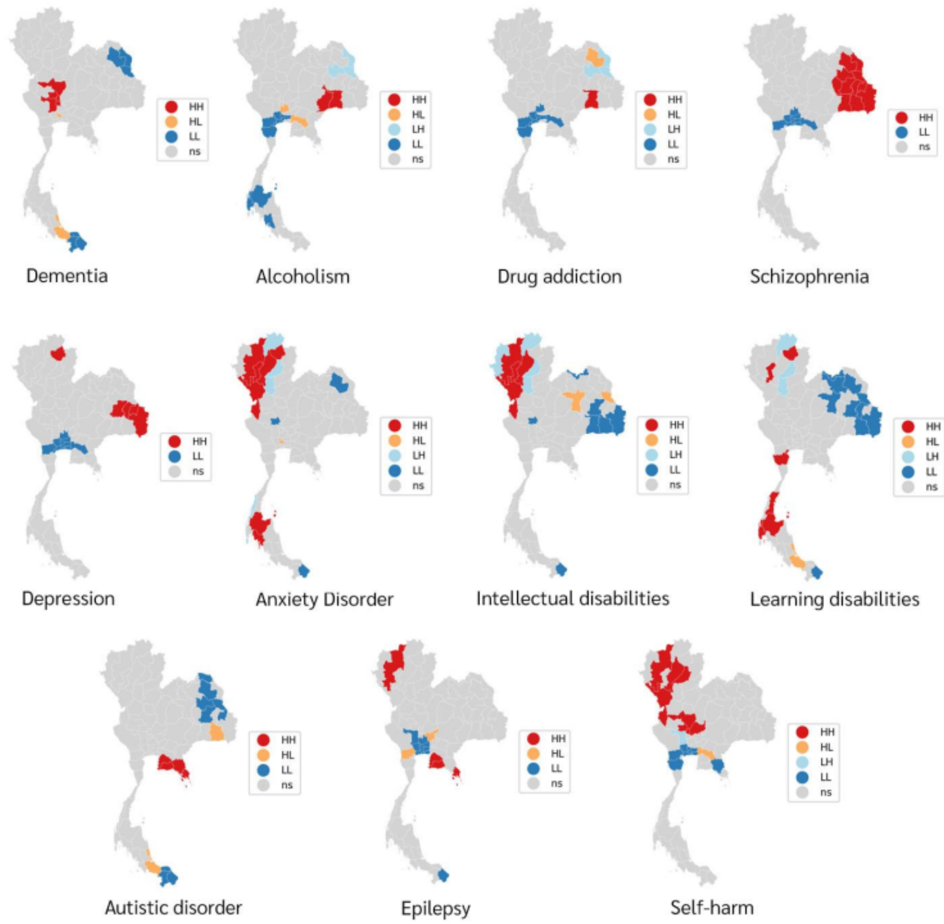


Figure 10: Maps showing 11 clusters of disorders of the number of patients accessing mental health services in Thailand at the provincial level using Local Moran's I statistic (LISA) used for a total of 7 years from 2015 to 2021.

Note: For interpreting clusters, High-High were labeled by red, High-Low were labeled by yellow, Low-High were labeled by sky blue, Low-Low were labeled by blue and Not significant were labeled by gray.

3.3.3 Disorder pairs

The correlation coefficient was calculated at a national level. The results are shown in Figure 5. In addition to arranging the disorders with similar correlation coefficients into the same cluster together. If determined using correlation data and pattern finding data calculated from global Moran's I analysis together, it is interesting to carry out a further study to find the correlation between pairs of disorders by calculating correlation detailed at the provincial level and using 7-year cumulative patients accessing mental health services data from 2015 to 2021. For interpretation, if a dark area is found in any province, it means these disorder pairs have a similar number of patients accessing mental health services that were continued for many years in such provinces. In other words, a relationship was found between these disorder pairs. Which found a total of 7 relationships with interest from 8 disorders as shown in Figure 11: Alcoholism and Drug addiction (excluding Amphetamine addiction) with a national correlation coefficient of 0.69. The second pair is Schizophrenia and Depression, with a correlation coefficient of 0.19. The third pair is Schizophrenia and Anxiety disorder, a correlation coefficient of 0.62. The fourth pair is Depression and Anxiety disorder with correlation coefficient of 0.32. The fifth pair is Autistic disorder and Intellectual disabilities with correlation coefficient of 0.67. The sixth pair is Autistic disorder and Learning disabilities with correlation coefficient of 0.37, and last disorder pairs is Intellectual disabilities and Learning disabilities, with a correlation coefficient of 0.39.

In terms of interpreting spatial outcomes for interrelated disorder pairs when calculating correlation at province level. It was found that most provinces in the northern region were related to Alcoholism and Drug addiction, Schizophrenia and Anxiety disorder. In addition, Autistic disorder, Intellectual disabilities and Learning disabilities that are the part of brain-related disorders were also found. In most provinces in the Northeast, there was a relationship between Alcoholism and Drug addiction, Schizophrenia and Anxiety disorder. The central and southern regions found the same pairs of disorders, namely, brain-related disorders (Autistic disorder, Intellectual disabilities and Learning disabilities). For the western ans eastern region, it was found a relationship of disorder pairs; Schizophrenia and Depression and the second pair is Depression and Anxiety disorder are shown in Figure 11.

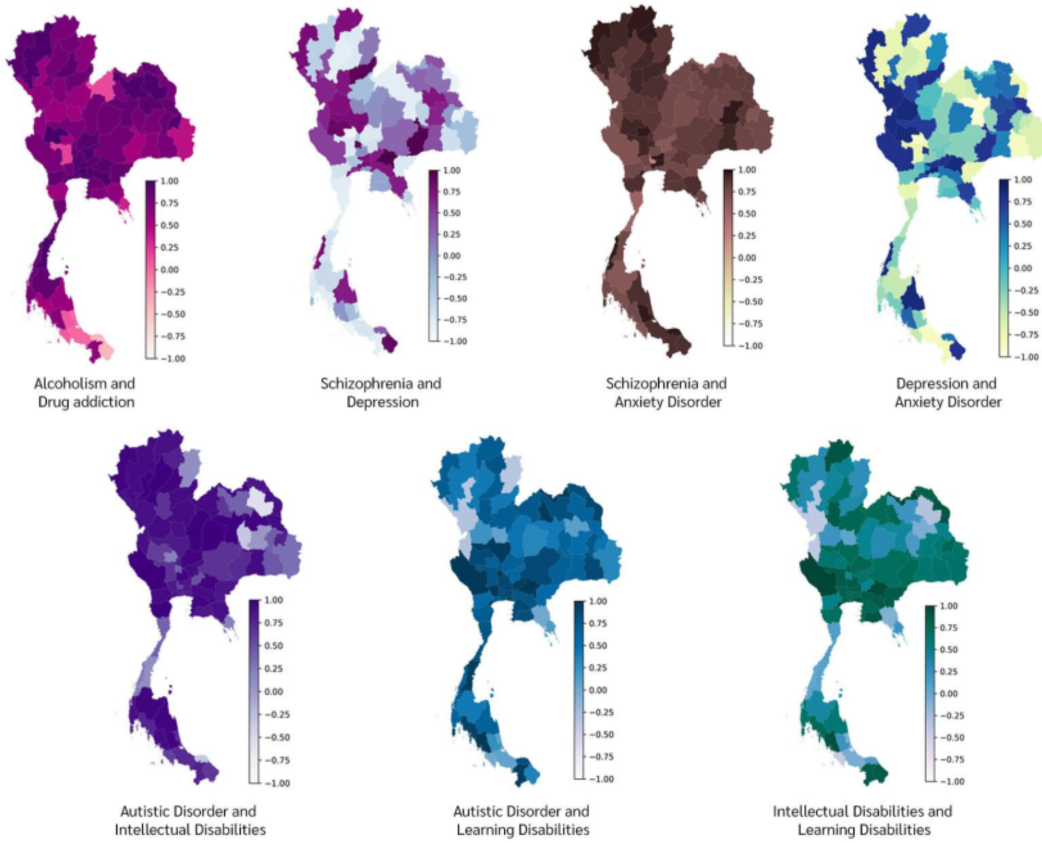


Figure 11: Map of Spearman's correlation of disorder pairs of the number of patients accessing mental health services in Thailand at the provincial level for a total of 7 years (2015-2021).

4. Discussion

In the discussion of spatial outcomes, it was found that the clusters of spatial autocorrelation patterns occurred in various regions. For example, the major clusters of spatial autocorrelation depression were discovered in several northern provinces where the Mind organization provided information on Seasonal Affective Disorder (SAD). SAD is a type of depression that occurs during specific seasons or times of year. Depression is characterized by a persistently low mood that interferes with daily activities. If you have SAD, you will experience depression during certain seasons or due to certain types of weather or temperature. It is available in both winter and summer (50). That is why the North has a large number of patient depressive services compared to other regions. The second disorder is drug addiction. According to an analysis of the current drug situation in 2017, most of the northern regions, especially the fringe areas, are conducive to the smuggling of drugs from neighboring countries that produce drugs, namely countries. Laos, Myanmar, and Cambodia (51). As a result, it may be the cause of the drug epidemic and a large number of patients seeking drug addiction services in the north. Then there was the cluster of spatial autocorrelation anxiety disorders, which were mostly found in the Northeast. It's a very hot climate. A description of cortisol: it's a hormone secreted by stress or anxiety. Cortisol levels are higher in the summer than in the winter (52). The northeastern region of Thailand has the highest average temperature. A rise in temperature can also cause agitation, palpitations, nausea, and fatigue, which is a common anxiety symptom. There are other disorders that were not discussed because there is limited information.

From all the methods we used (correlation matrix, K-prototype clustering, and agglomerative hierarchical clustering), there are three groups of disorders that cluster based on characteristics of the number of patients accessing mental health services in every province from 2015-2021.

The first group includes Alcoholism, Drug addiction (excluding Amphetamine addiction), and Epilepsy. These disorders are highly correlated due to the fact that abusing alcohol has a potential risk of other substance use, such as marijuana, cocaine, and heroin (at least one) (53). For the association between Alcoholic and Epilepsy, the available evidence suggests that the prevalence of epilepsy among alcoholics is at least triple that in the general population, and that alcoholism may be more prevalent among epileptic patients than in the general population (54).

The second group includes Schizophrenia, Depression and Anxiety disorder. From the result of the correlation, it was found that Schizophrenia and Anxiety disorder are more highly correlated than Schizophrenia and Depression. In this research, the prevalence of anxiety disorders was studied. The incidence of anxiety disorders was 45.16% in schizophrenia compared to controls. The research findings show that anxiety disorders are higher in schizophrenia than in the general population. It is also said that symptoms of anxiety tend to precede symptoms of schizophrenia (55). Likewise, the Conrad and Chapman study found that anxiety disorders are one of the early causes of schizophrenia (56). On the other hand, anxiety disorders may occur after treatment for schizophrenia due to concerns in daily life. The stress of being judged because of admission for schizophrenia is not widely accepted by society. This is reason enough to explain the relationship between schizophrenia and anxiety disorders. There was also a relationship between schizophrenia and depression. which had lower correlation values compared to schizophrenia and anxiety disorders. Therefore, when diagnosed with this type of disorder, it is less likely to be identified as schizophrenia (57). But when diagnosed with schizophrenia, depression may also be present. Therefore, the correlation calculation showed that the correlation was low.

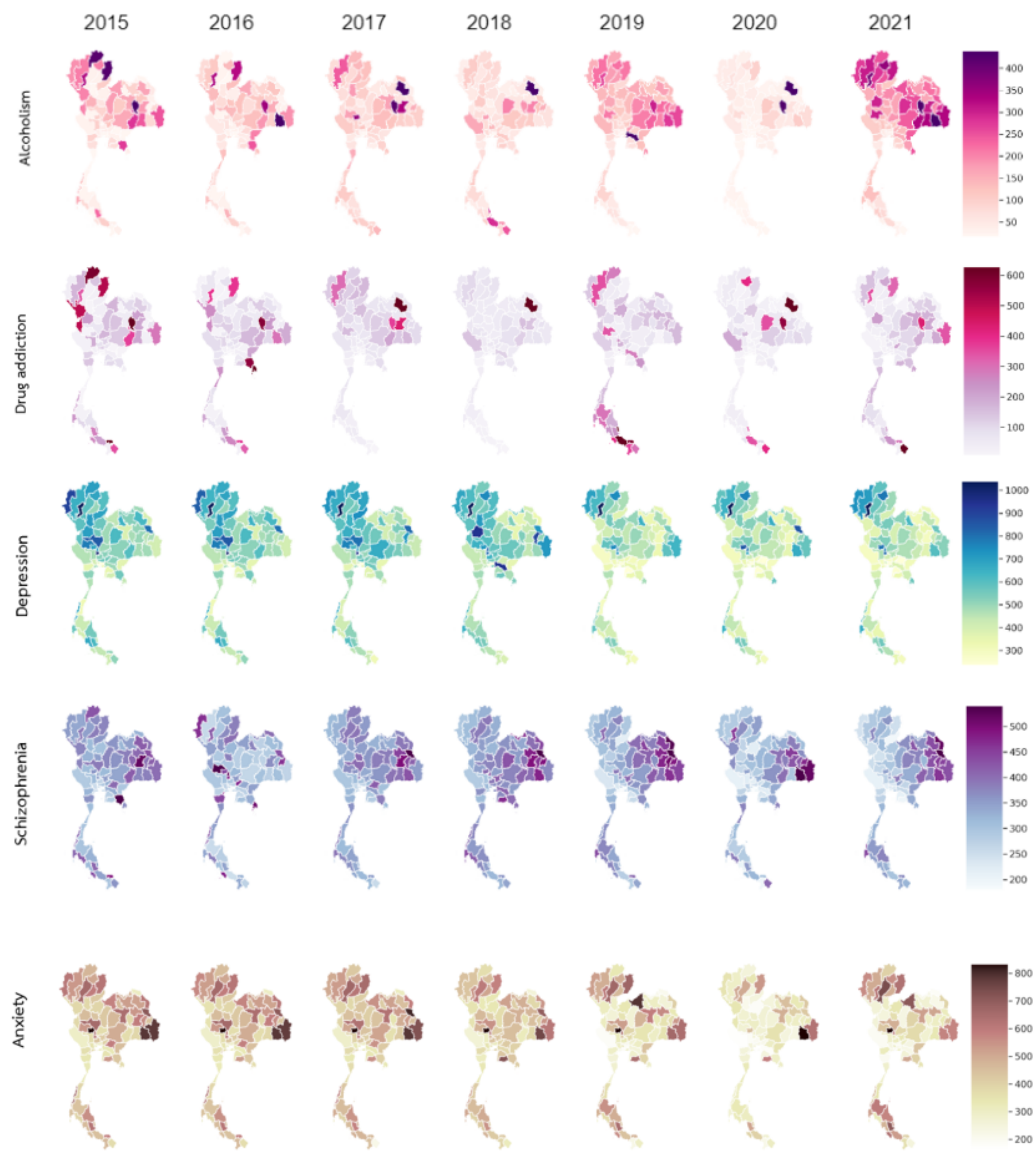
The last group includes Intellectual disabilities, Learning disabilities, Autistic disorder, Self-harm and Dementia. The correlation coefficient between Autistic disorder and Intellectual disabilities is higher than the correlation coefficient between Autistic disorder and Learning disabilities because Intellectual disabilities are the most common co-occurring disorder with Autistic disorder, and the greater severity of one of these two disorders appears to have effects on the other disorder (58). Dementia correlates with Autistic disorder and also correlates with Intellectual Disabilities. Vivanti et al. used US medical insurance records for 1.2 million individuals aged 30–64 years. The prevalence was highest among people with intellectual disability alone (7.10%), but was also considerably higher among people with ASD (4.04%) and people with ASD and intellectual disability (5.22%) than among the healthy population (0.97%) (59). For the association with Self-harm, the study of Blanchard et al. shows that Autistic disorder was associated with a substantially increased risk of self-injurious behaviors and suicidality, and people with Autistic disorder had 2.26-times higher odds of self-harm than those without ASD (60).

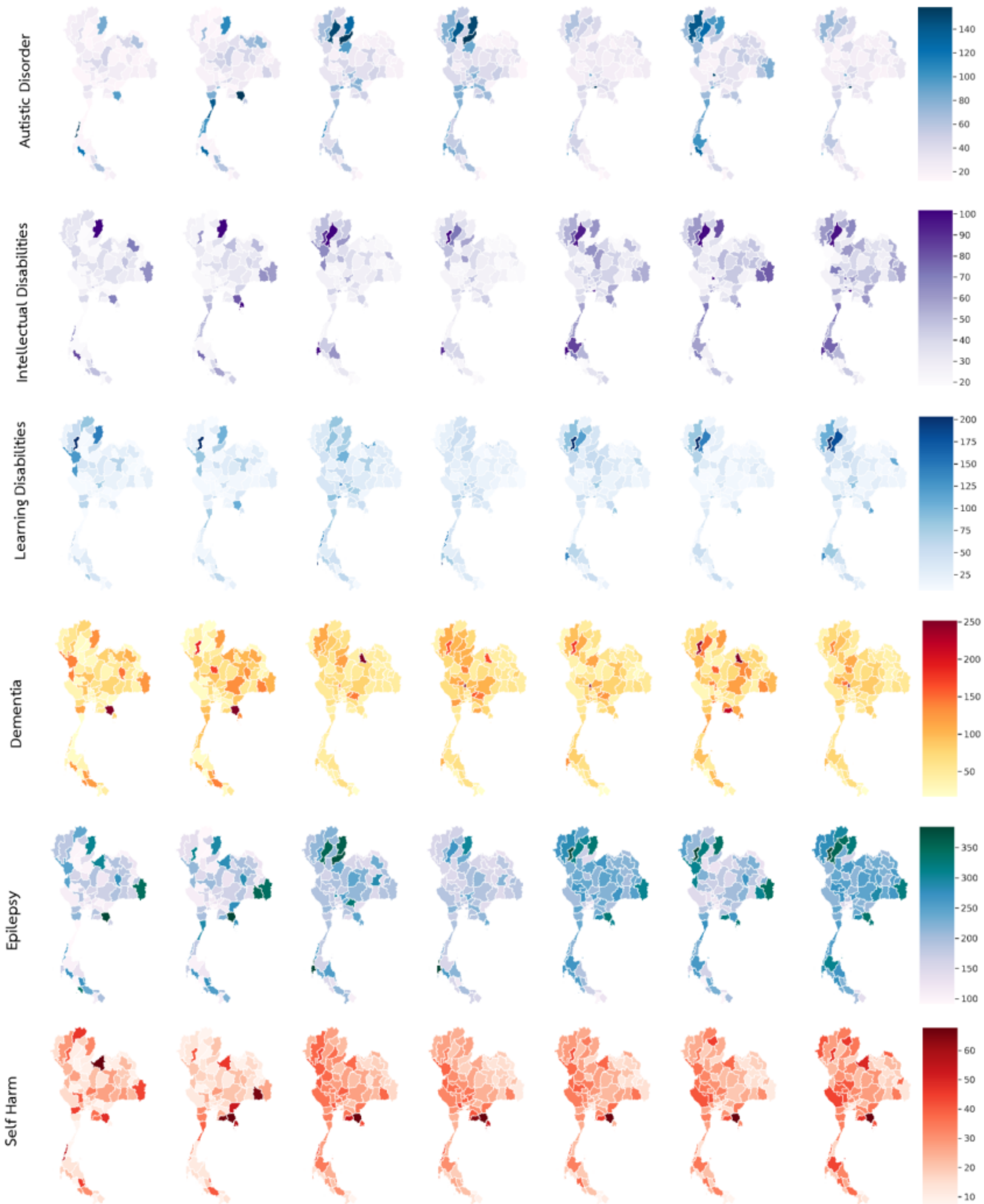
There are several limitations to this study. First, the data source was a mental health accessed services report and stigmatization of mental disorders may prevent patients from seeking treatment. Therefore, the number of patients accessing mental health services may not be the same as the actual number of patients that occur. Furthermore, the number and location of mental health services in each province may affect the difficulty of accessing the service and may explain why the number of patients accessing mental health services is low. Another limitation is that Bangkok Metropolis's mental health accessed services data was missing in 2017. Missing data can reduce the efficiency of statistical analysis (61). Moreover, many people may have more than one mental health diagnosis (62). All of the above limitations can cause bias in the data. As for the future work, we desired to reduce the limitations to the least.

5. Conclusions

The results of this study provided information about the spatiotemporal characteristics of mental disorders in Thailand and also revealed the limitations of the data, which have an impact on the overall analysis and further applications. For more accurate and interpretable results, such as geographic heat maps and associations between two disorders, the data used for analysis should represent the real characteristics of each disorder, which can lead to change, such as mental health policy by the government, attitudes and public interaction, and access to prevention and treatments.

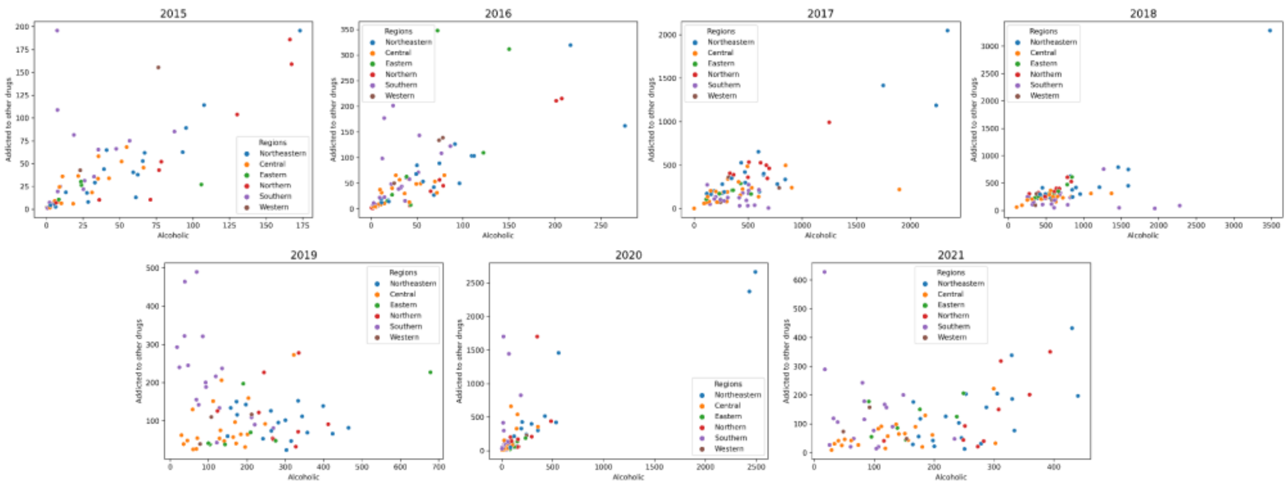
Appendix

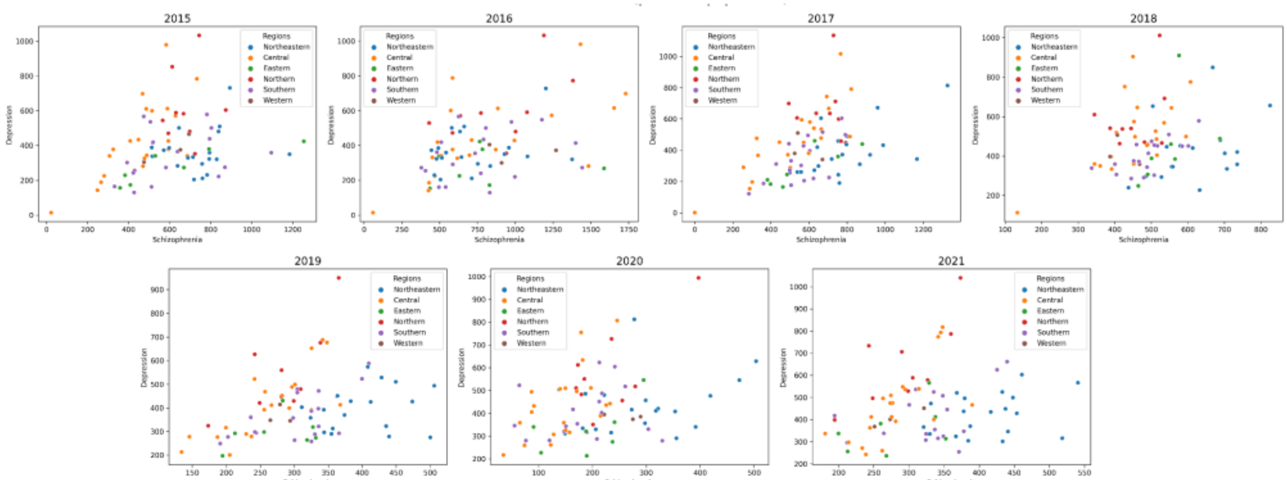

2015
2016
2017
2018
2019
2020
2021
Alcoholism
Drug addiction
Depression
Schizophrenia
Anxiety


Autistic Disorder
Intellectual Disabilities
Learning Disabilities
Dementia
Epilepsy
Self Harm

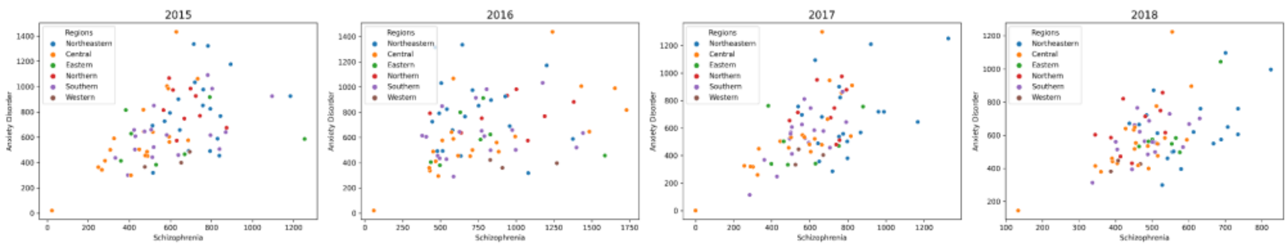
Scatter plot

Alcoholism and Drug addiction



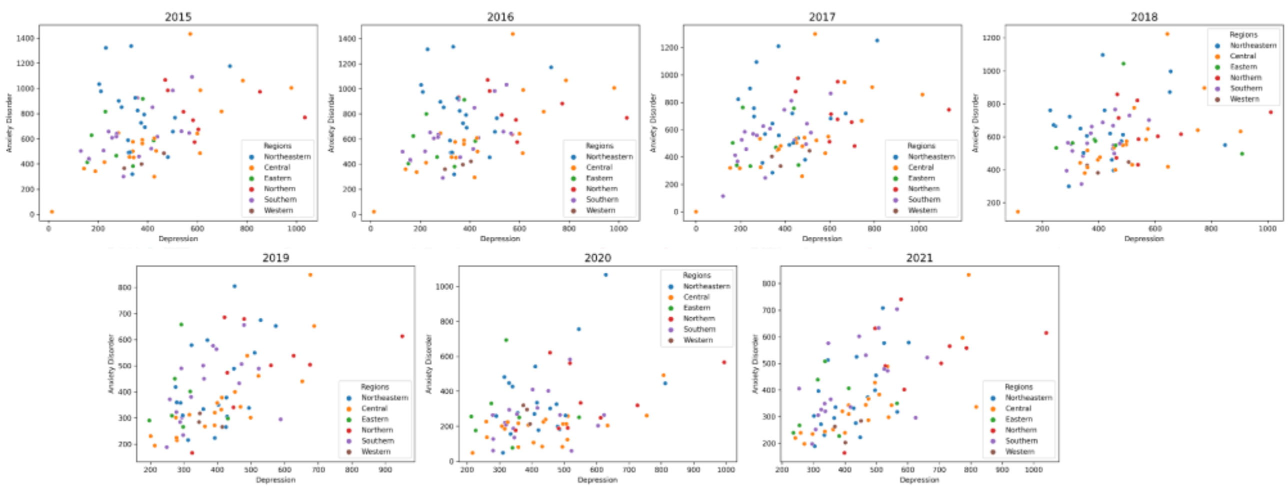


Schizophrenia and Depression

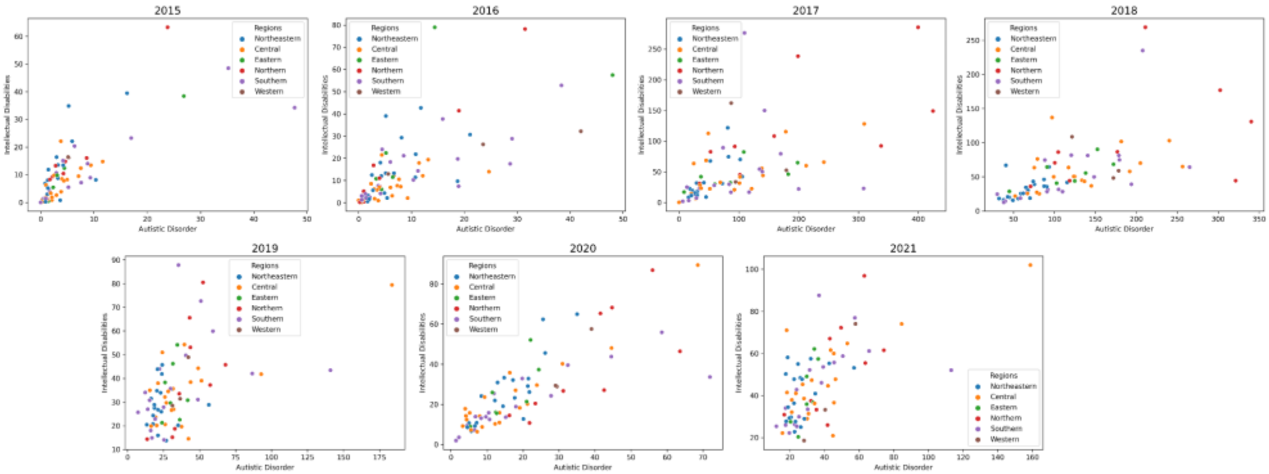


Schizophrenia and Anxiety Disorder

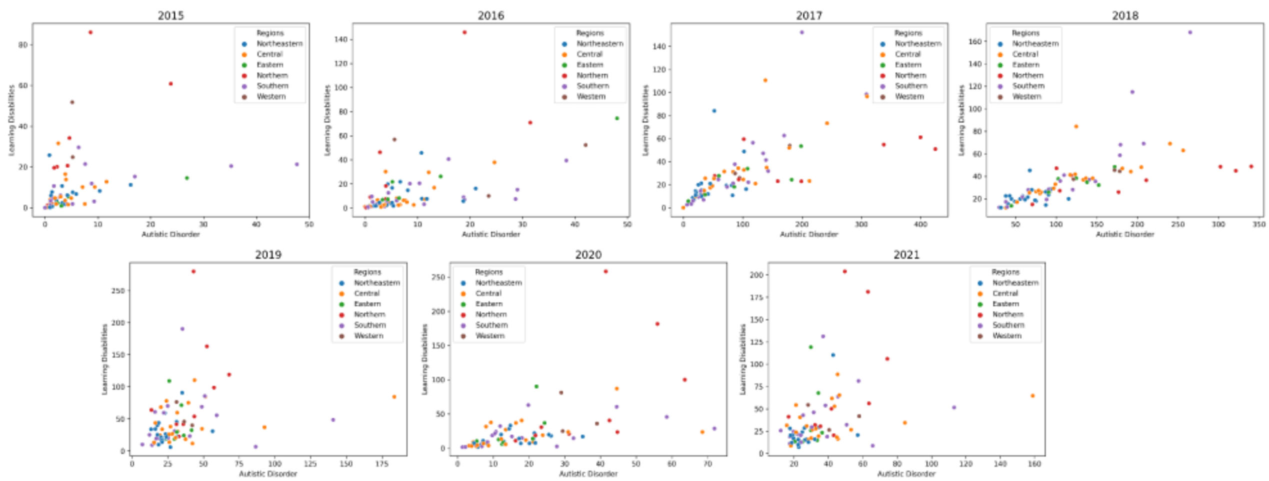
Depression and Anxiety Disorder



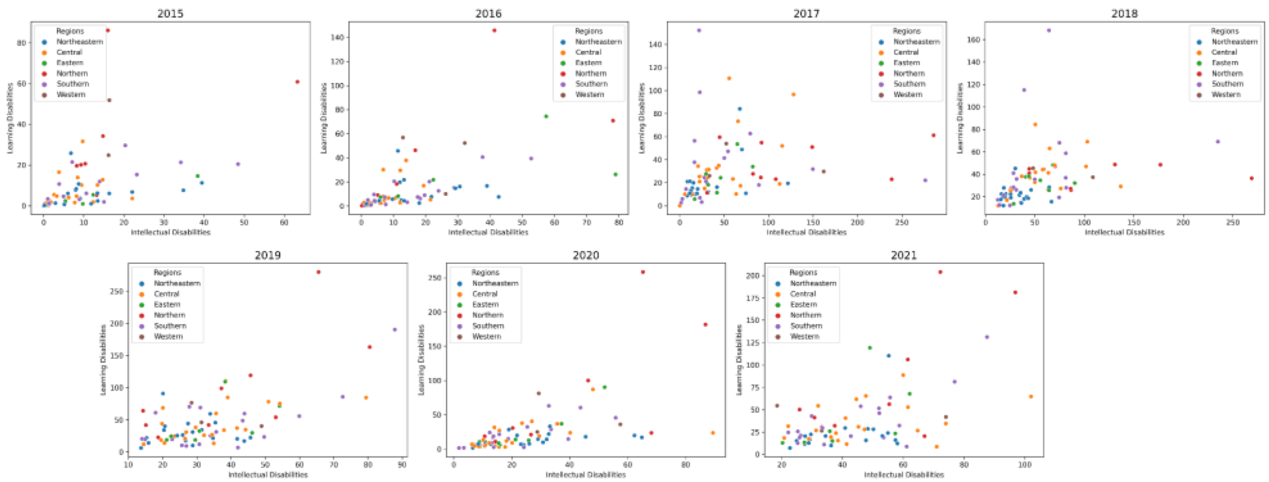
Autistic Disorder and Intellectual Disabilities



Autistic Disorder and Learning Disabilities



Intellectual Disabilities and Learning Disabilities



Global Moran's I

Years	Dementia	Alcoholism	Drug addiction	Schizophrenia	Depression	Anxiety	Intellectual Disabilities	Learning Disabilities	Autistic	Self-Harm	Epilepsy
2015	-0.0623	0.1887	0.0681	0.3147	0.2230	0.3007	-0.1273	0.1115	-0.1115	-0.0612	-0.0033
2016	0.0521	0.1611	0.1636	0.1073	0.2097	0.3050	0.0364	0.0263	0.0444	0.1699	0.1174
2017	0.1231	0.0438	0.1976	0.5679	0.2572	0.3047	0.2225	-0.0110	0.1550	0.3146	0.0883
2018	0.2354	0.0006	0.0127	0.4651	0.1699	0.2102	0.2423	0.2025	0.2930	0.4551	0.0943
2019	0.1745	0.3804	0.4067	0.6607	0.3066	0.2024	0.1670	0.2260	0.1173	0.3466	0.3464
2020	0.0713	0.0026	-0.0435	0.4614	0.2574	0.1802	0.0501	0.3414	0.1790	0.2769	0.1756
2021	0.2907	0.5893	0.1427	0.6204	0.3166	0.2748	0.0757	0.2524	0.1318	0.1830	0.4130

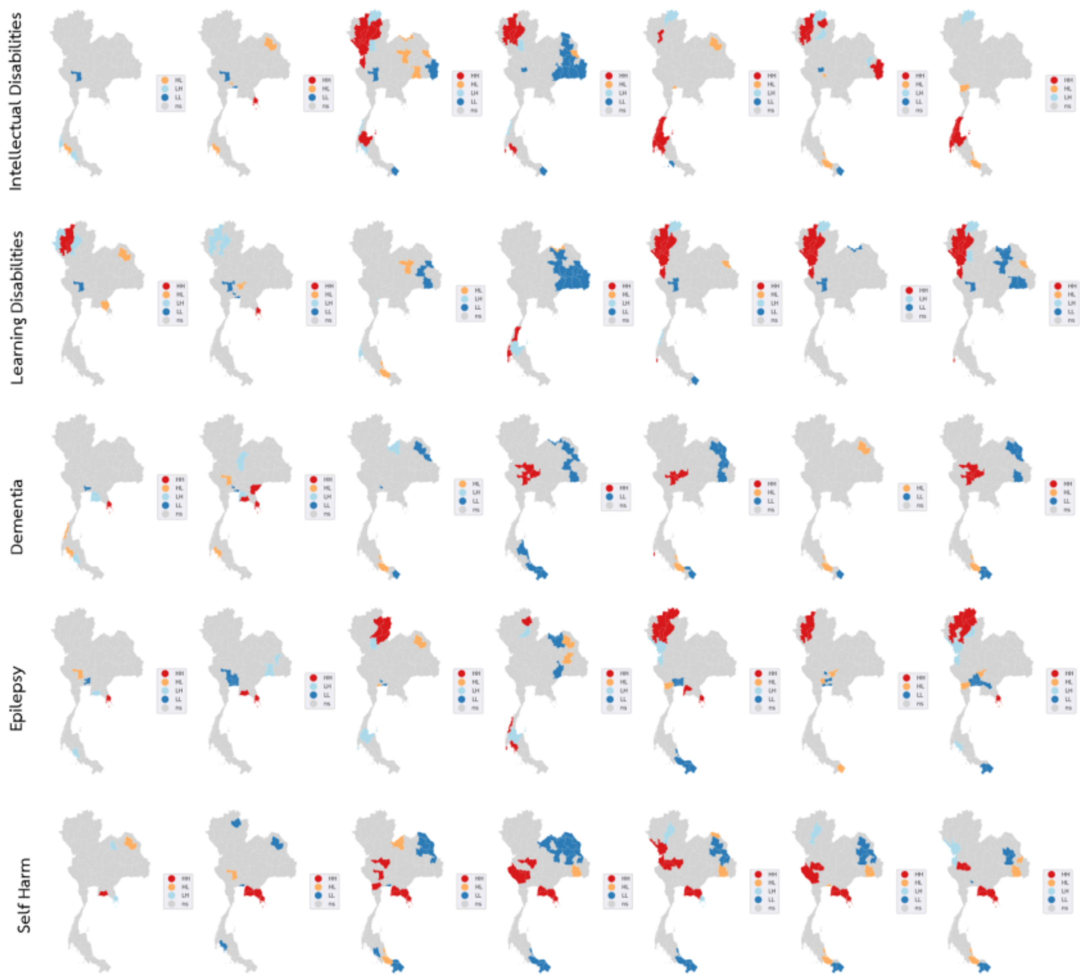
Table 2 The value of Moran's I hypothesis test of the number of patients accessing mental health services in Thailand in 2015-2021 (per 100k population).

Years	Dementia	Alcoholism	Drug addiction	Schizophrenia	Depression	Anxiety	Intellectual Disabilities	Learning Disabilities	Autistic	Self-Harm	Epilepsy
2015	0.244	0.006*	0.155	0.001*	0.002*	0.001*	0.041*	0.044*	0.037*	0.269	0.459
2016	0.201	0.011*	0.023*	0.057	0.004*	0.001*	0.232	0.238	0.185	0.011*	0.043*
2017	0.019*	0.213	0.009*	0.001*	0.001*	0.001*	0.006*	0.474	0.021*	0.001*	0.079
2018	0.001*	0.38	0.247	0.001*	0.01*	0.003*	0.001*	0.004*	0.001*	0.001*	0.06
2019	0.008*	0.001*	0.001*	0.001*	0.001*	0.005*	0.01*	0.003*	0.039*	0.001*	0.001*
2020	0.113	0.253	0.375	0.001*	0.002*	0.011*	0.168	0.001*	0.014*	0.001*	0.009*
2021	0.001*	0.001*	0.03*	0.001*	0.001*	0.001*	0.108	0.003*	0.027*	0.009*	0.001*

Table 3 The p-value of Moran's I hypothesis test of the number of patients accessing mental health services in Thailand in 2015-2021 (per 100k population).

Note: The significant value were marked by *.

LISA Moran's I


Intellectual Disabilities
Learning Disabilities
Dementia
Epilepsy
Self Harm

Final Report

ORIGINALITY REPORT

15%	10%	10%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.pubmedcentral.nih.gov Internet Source	1%
2	www.chemijournal.com Internet Source	1%
3	Manish Mathur. "Spatial autocorrelation analysis in plant population: An overview", Journal of Applied and Natural Science, 2015 Publication	1%
4	www.nature.com Internet Source	1%
5	Arthur W. K. Chan. "Alcoholism and Epilepsy", Epilepsia, 8/1985 Publication	1%
6	Submitted to Los Angeles City College Student Paper	1%
7	jamanetwork.com Internet Source	<1%
8	Submitted to Dallas Community College District Student Paper	<1%

9	raw.githubusercontent.com Internet Source	<1 %
10	www.stat.cmu.edu Internet Source	<1 %
11	Submitted to Kelvin Grove State College Student Paper	<1 %
12	Submitted to West Thames College London Student Paper	<1 %
13	docplayer.ru Internet Source	<1 %
14	Henwy Wibowo, Kristina Pestaria Sinaga. "Telecommunication Analytics Based on Customer Segmentation Using Unsupervised Algorithms", 2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS), 2021 Publication	<1 %
15	Submitted to University of Durham Student Paper	<1 %
16	Khan, Mahtaab, Deepika Dubey, Huma Gupta, and Amit Saxena. "Optimize Prototype Classification Based on GA ANT Hybrid Technique", 2013 5th International Conference on Computational Intelligence and Communication Networks, 2013. Publication	<1 %

17	Salvador Garcia-Ayllon. "Urban Transformations as an Indicator of Unsustainability in the P2P Mass Tourism Phenomenon: The Airbnb Case in Spain through Three Case Studies", Sustainability, 2018 Publication	<1 %
18	Submitted to Universiti Kebangsaan Malaysia Student Paper	<1 %
19	dokumen.pub Internet Source	<1 %
20	Ian Fyfe. "Deep learning distinguishes tauopathies", Nature Reviews Neurology, 2021 Publication	<1 %
21	ru.scribd.com Internet Source	<1 %
22	Submitted to Milton Keynes College Student Paper	<1 %
23	Reza Mohammadi, Mohammad Taleai, Somayeh Alizadeh, Omid Reza Abbasi. "STFTiS: Introducing a spatio‐temporal FTiS model to investigate the level of citizens' satisfaction of 311 non‐emergency services", Transactions in GIS, 2022 Publication	<1 %

24	repositorio.iscte-iul.pt Internet Source	<1 %
25	www.cambridge.org Internet Source	<1 %
26	Kim, D.W.. "Fuzzy cluster validation index based on inter-cluster proximity", Pattern Recognition Letters, 200311 Publication	<1 %
27	Li Peng, Xu-xi Wang. "What is the relationship between ecosystem services and urbanization? A case study of the mountainous areas in Southwest China", Journal of Mountain Science, 2019 Publication	<1 %
28	Ricardo A. Olea. "Geostatistics for Engineers and Earth Scientists", Springer Science and Business Media LLC, 1999 Publication	<1 %
29	Www.hindawi.com Internet Source	<1 %
30	journal.fapetunipa.ac.id Internet Source	<1 %
31	mdpi-res.com Internet Source	<1 %
32	media.neliti.com Internet Source	<1 %

33	pysal.org Internet Source	<1 %
34	zia207.github.io Internet Source	<1 %
35	"Geo-Informatics in Resource Management and Sustainable Ecosystem", Springer Nature, 2015 Publication	<1 %
36	Submitted to Erasmus University of Rotterdam Student Paper	<1 %
37	Michael Nnachebe Onah, Sally Field, Jason Bantjes, Simone Honikman. "Perinatal suicidal ideation and behaviour: psychiatry and adversity", Archives of Women's Mental Health, 2016 Publication	<1 %
38	Submitted to University of Greenwich Student Paper	<1 %
39	parasitesandvectors.biomedcentral.com Internet Source	<1 %
40	pjmhsonline.com Internet Source	<1 %
41	www.db-thueringen.de Internet Source	<1 %

42	www.repository.cam.ac.uk Internet Source	<1 %
43	E. Joseph Billo. "Linear Regression and Curve Fitting", Wiley, 2011 Publication	<1 %
44	Submitted to Victoria University of Wellington Student Paper	<1 %
45	Wen-Chong He, Ke Ju, Ya-Min Gao, Pei Zhang, Yin-Xia Zhang, Ye Jiang, Weibin Liao. "Spatial inequality, characteristics of internal migration, and pulmonary tuberculosis in China, 2011–2017: A spatial analysis", Research Square Platform LLC, 2020 Publication	<1 %
46	onlinelibrary.wiley.com Internet Source	<1 %
47	Michael Jerrett, Richard Burnett, Alette Willis, Daniel Krewski, Mark Goldberg, Patrick DeLuca, Norm Finkelstein. "Spatial Analysis of the Air Pollution--Mortality Relationship in the Context of Ecologic Confounders", Journal of Toxicology and Environmental Health, Part A, 2003 Publication	<1 %
48	Rosina Moreno, Raffaele Paci, Stefano Usai. "Innovation Clusters in the European Regions", European Planning Studies, 2006	<1 %

Publication

49	www.shipmarcltd.com Internet Source	<1%
50	A. D. van Ingen Schenau, C. G. Verschoor, C. Romers. " The crystal and molecular structure of hexakis(pyridine- -oxide)nickel(II) bis(tetrafluoroborate) ", Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry, 1974 Publication	<1%
51	Shuang Yin, Guojun Gan, Emiliano A. Valdez, Jeyaraj Vadiveloo. "Applications of Clustering with Mixed Type Data in Life Insurance", Risks, 2021 Publication	<1%
52	Submitted to University of Oxford Student Paper	<1%
53	docplayer.net Internet Source	<1%
54	www.mdpi.com Internet Source	<1%
55	Rencher. "Cluster Analysis", Wiley Series in Probability and Statistics, 02/22/2002 Publication	<1%
56	Xu. "Hierarchical Clustering", Clustering, 10/09/2008 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		