IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 6, December 2026, pp. 5623~5632
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i6.pp5623-5632

Integrating neural clustering and causal inference to explain global life expectancy variations

Ilyas Tammouch[1], Soumaya Nouna[2], Abdelamine Elouafi[3], Abdessamad Elmotawakkil[4], Aziz Soulhi[1]
[1]Department of Mining and Environment, National School of Mines of Rabat (ENSMR), Rabat, Morocco
[2]Laboratory for Analysis and Modeling of Systems and Decision Support (LAMSAD), Department of Mathematics and Informatics, ENSA Berrechid, Hassan First University of Settat, Berrechid, Morocco
[3]SIV Laboratory, National School of Business and Management of Dakhla (ENCG Dakhla), Faculty of Sciences, Ibn Zohr University, Agadir, Morocco
[4]Department of Computer Science, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco

Article Info

Article history:

Received Oct 26, 2025
Revised Aug 9, 2026
Accepted Aug 31, 2026

Keywords:

Causal discovery
Centroid neural network
Life expectancy
Linear non-Gaussian acyclic model
Neural clustering
Socioeconomic determinants

ABSTRACT

This study investigates global disparities in life expectancy through an integrated framework combining centroid-based neural clustering, spatial visualization, and causal inference. Using country-level data on life expectancy, suicide rate, and gross domestic product (GDP) per capita for the period 2015–2019, this study applies several clustering algorithms and introduce a centroid neural network (CENTNN) an unsupervised competitive learning model to capture complex, non-linear relationships, and reveal latent groupings of nations. Validation metrics confirm the CENTNN's superior ability to identify coherent socioeconomic–health profiles. The resulting clusters are projected onto a choropleth world map, revealing geographically structured disparities that mirror the global north–south divide in health outcomes. To uncover directional dependencies among indicators, we employ the linear non-Gaussian acyclic model (LiNGAM), which identifies a strong positive causal path from GDP per capita to life expectancy and a negative directional relationship from suicide rate to life expectancy. Together, these methods provide complementary perspectives: CENTNN clustering delineates structural heterogeneity, the choropleth contextualizes it spatially, and LiNGAM quantifies the underlying causal mechanisms. The findings highlight that both economic and psychosocial factors jointly shape global life expectancy and demonstrate the value of combining neural representation learning with causal discovery in public health research.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Ilyas Tammouch
Department of Mining and Environment, National Higher School of Mines Rabat (ENSMR)
Rabat, Morocco
Email: tammouch@enim.ac.ma

1. INTRODUCTION

Understanding why populations in some countries live significantly longer than others remain a fundamental question at the intersection of public health, economics, and artificial intelligence (AI). Despite improvements in healthcare and living standards, substantial disparities in life expectancy persist, reflecting the complex interactions among socioeconomic development, mental health, and access to healthcare [1]. Although economic prosperity is a major determinant of longevity, it cannot fully explain these disparities, as psychosocial and behavioral factors, including suicide prevalence, also contribute substantially to population health outcomes [2], [3]. These challenges require analytical approaches capable of uncovering both latent

structural patterns and the causal mechanisms underlying global longevity disparities [4]. Recent advances in machine learning have enabled the discovery of hidden patterns in complex socioeconomic and health data through neural learning techniques [5]–[7].

In this study employ the centroid neural network (CENTNN), a competitive learning algorithm that adaptively learns cluster prototypes from data [8]–[11]. While recent deep clustering frameworks, such as deep embedded clustering (DEC), improved deep embedded clustering (IDEC), and variational deep clustering, have demonstrated excellent performance for high-dimensional representation learning, their advantages are most evident for large-scale datasets with complex feature spaces. Given the relatively low-dimensional country-level dataset analyzed in this study, CENTNN provides a computationally efficient and interpretable alternative while retaining the ability to capture complex socioeconomic structures. Although clustering effectively identifies structural similarities among countries, it cannot distinguish association from causation, limiting its usefulness for policy-oriented decision making. To address this limitation, the proposed framework integrates the linear non-Gaussian acyclic model (LiNGAM), a causal discovery approach that infers directed relationships among continuous variables under non-Gaussian assumptions [12]. Applying LiNGAM enables the analysis to move beyond correlational patterns by inferring directional relationships among gross domestic product (GDP) per capita, suicide rate, and life expectancy under its model assumptions. Integrating CENTNN with LiNGAM establishes an explainable artificial intelligence (XAI) framework in which competitive neural learning uncovers latent population structures while causal discovery identifies directed relationships within them. This integration extends the analysis beyond descriptive clustering toward interpretable explanations of global health disparities [12].

The proposed framework is evaluated using a harmonized global dataset, enabling the analysis of worldwide socioeconomic and health disparities. The objective of this study is to investigate global life expectancy disparities through an integrated framework combining neural clustering, spatial visualization, and causal discovery. Specifically, it identifies country profiles, visualizes their spatial distribution, and infers causal relationships among GDP per capita, suicide rate, and life expectancy. This study hypothesize that life expectancy is jointly shaped by economic and psychosocial factors, with GDP per capita exerting both direct and indirect effects through mental health outcomes. By combining unsupervised representation learning with causal inference, the proposed framework provides complementary structural and causal insights into global longevity disparities, demonstrating the value of XAI for population health analysis.

2. DATA AND PREPROCESSING

2.1. Data sources

This study uses a harmonized global dataset accessed through Kaggle [13], covering 2015–2019 and comprising 193 countries. The underlying data originate from international sources, including the World Health Organization for suicide rates, the World Bank for GDP per capita, and United Nations demographic statistics for life expectancy, with Kaggle serving as the data access layer. The dataset includes country-level, sex-disaggregated life expectancy at birth (years), age-standardized suicide rates per 100,000 inhabitants, and GDP per capita. Monetary values were expressed in constant 2017 U.S. dollars to ensure cross-country comparability. Data harmonization involved aligning country identifiers, standardizing variable formats, and matching temporal coverage across indicators. Missing values, representing less than 5% of the data, were handled through mean imputation for continuous variables, while observations with invalid country identifiers were excluded. Supplementary demographic variables, including education and fertility rates, were retained for descriptive analysis only because of incomplete country coverage and were not included in the core clustering or causal inference models. After cleaning and harmonization, the final dataset comprised 2,938 country-level observations. Table 1 summarizes the variables, units, descriptions, and data sources used in the analysis.

Table 1. Description of variables and data sources

Variable	Unit	Description
Life expectancy (men)	Years	Male life expectancy at birth by country
Life expectancy (women)	Years	Female life expectancy at birth by country
Suicide rate	Per 100,000 inhabitants	Age-standardized suicide mortality rate
GDP per capita	USD (constant)	GDP per capita as an economic indicator
Education rate	Percentage (%)	Proportion of population with secondary or higher education
Fertility rate	Children per woman	Average number of births per woman

2.2. Data cleaning

Data integration began by harmonizing country names and filtering out entries with excessive missing values. The GDP per capita variable, originally a string containing commas and currency symbols, was converted to numeric using str. replace and pd.to_numeric [14]. Missing values (<5%) were handled via mean imputation for continuous variables, while entire rows with missing country identifiers were excluded. Mean imputation was selected due to the very small proportion of missing data, minimizing distortion of global distributions. Given the macro-scale nature of the analysis, this approach preserves cross-country trends. Sensitivity checks using alternative substitutions such as median imputation produced qualitatively similar clustering behavior, indicating robustness of the preprocessing strategy.

2.3. Transformation and normalization

Before applying clustering algorithms, all continuous features were standardized using StandardScaler to ensure equal influence on distance-based methods [15]. Correlation analysis was conducted to assess the relationships between variables and confirm their theoretical relevance for clustering [16]. A key preprocessing step was featuring selection. Only real-valued, continuous, and non-categorical indicators were retained for causal modeling (LiNGAM) [12].

3. METHOD

3.1. Comparative approach

To identify global socioeconomic-health profiles, four clustering algorithms were evaluated: k-means, agglomerative hierarchical clustering, Gaussian mixture models (GMM) [10], and CENTNN [9]. This comparison assessed conventional distance-based methods against competitive neural learning for modeling complex data structures. Each algorithm was trained on GDP per capita, suicide rate, and gender-disaggregated life expectancy. The optimal number of clusters (k) was determined using the Elbow method based on the within-cluster sum of squares (WCSS) [17], expressed as (1).

$$WCSS = \sum_{i=1}^{k} \sum_{x_j \in C_i} \| x_j - \mu_i \|^2 \quad (1)$$

Where x_j represents a data point belonging to cluster C_i, and μ_i denotes centroid of that cluster. By plotting WCSS as a function of k, the "elbow point" a sharp inflection where marginal gains in cluster compactness diminish was used as the optimal cluster count. Model selection relied on internal validation diagnostics including the Elbow method (WCSS), using silhouette score and Davies-Bouldin index [18], [19] to assess cluster quality and separation. The best-performing model was selected as the principal model for subsequent analyses.

3.1.1. Centroid neural network

To capture the complex relationships among economic capacity, mental health, and life expectancy, this study employs the CENTNN, an unsupervised competitive learning model [9]. CENTNN adaptively refines cluster prototypes according to the underlying data distribution. Unlike conventional centroid-based methods, it provides an interpretable and computationally efficient representation of latent socioeconomic-health structures [8].

Formally, given an input feature vector $x_i \in \mathbb{R}^n$ and a set of centroids $\{w_j\}_{j=1}^{k}$, the CENTNN determines the winning neuron j^*according to the minimum squared distance criterion, as in (2).

$$j^* = arg\ min \left\| x_i - w_j \right\|^2 \quad (2)$$

Once the winning neuron is identified, its corresponding centroid weight vector is updated via gradient descent according to the rule, as in (3).

$$w_j = w_j + \eta\left(x_i - w_j\right), if\ j = j^* \quad (3)$$

Where η represents the learning rate. This dynamic update process, inspired by competitive learning, enables the centroids to progressively adapt to the underlying distribution of the input data. Over successive iterations, the centroids converge toward equilibrium positions that define the stable prototypes of the dataset each representing a distinct socioeconomic–health profile.

3.2. Spatial visualization (choropleth mapping)

To enhance the interpretability of CENTNN-derived clusters, country memberships were visualized using a choropleth map based on the Robinson projection [20]. Each country was color-coded according to

its assigned cluster, representing distinct combinations of GDP per capita, suicide rate, and life expectancy. This visualization translated learned cluster structures into interpretable geographic patterns [21].

3.3. Linear non-Gaussian acyclic model causal inference framework

After defining coherent global clusters, the study applied the LiNGAM to infer causal relationships among life expectancy, suicide rate, and GDP per capita [12]. Rooted in independent component analysis (ICA), LiNGAM enables causal discovery from observational data under the assumption that variables are linearly related and affected by statistically independent, non-Gaussian disturbances [22]. Unlike correlation- or regression-based approaches, which quantify statistical associations without distinguishing cause from effect, LiNGAM infers directed causal relationships from observational data under non-Gaussian assumptions. This enables the identification of plausible causal pathways, providing a more interpretable understanding of how socioeconomic and behavioral factors jointly influence life expectancy [4].

The model assumes a linear structural relationship represented as (4).

$$X = BX + \varepsilon \tag{4}$$

Where $X = (x_1, x_2, \dots, x_n)^T$ denotes the observed variable vector, B is the coefficient matrix encoding causal dependencies (with zeros on the diagonal to maintain acyclicity), and $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ represents independent, non-Gaussian error terms. The key identification principle of LiNGAM lies in exploiting the non-Gaussianity of ε to uniquely estimate B without external causal priors.

In this study, the DirectLiNGAM algorithm [23] was implemented using Python's LiNGAM library. Bootstrapping was performed to assess the stability of the inferred causal edges. To account for developmental heterogeneity, LiNGAM was applied separately within each CENTNN-defined cluster, establishing a two-stage framework that combines neural clustering to identify structural similarities among countries with causal discovery to infer directional relationships. This integration provides complementary structural and causal insights into the determinants of life expectancy.

4. RESULTS AND DISCUSSION

4.1. Comparative approach

To identify meaningful groupings of nations according to shared health and socioeconomic profiles, several clustering algorithms were evaluated, including k-means, hierarchical agglomerative clustering, GMM, and the CENTNN. Each method was trained on standardized country-level indicators encompassing GDP per capita, suicide rate, and life expectancy (for men and women). Figure 1 shows a clear Elbow at k = 4, indicating the optimal balance between cluster compactness and interpretability. Beyond this point, reductions in WCSS become marginal, supporting the selection of four clusters for subsequent CENTNN clustering and LiNGAM analysis. Clustering performance was evaluated using the silhouette score, Davies-Bouldin index, and Calinski-Harabasz index [18], [19], [24], which assess cluster separation and compactness.

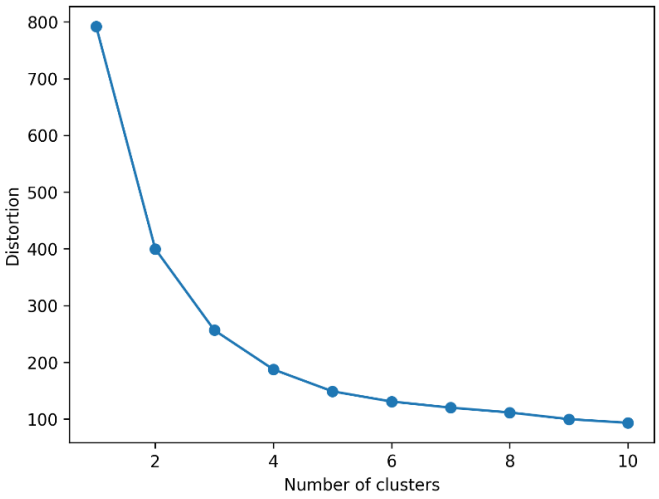

Figure 1. Elbow curve showing optimal cluster count

Comparative evaluation, as shown in Table 2 demonstrates that CENTNN achieved the strongest overall performance according to the selected internal validation metrics. Its higher silhouette score (0.481) indicates stronger within-cluster cohesion, while the lower Davies–Bouldin index (0.667) reflects improved inter-cluster separation. The highest Calinski-Harabasz index further confirms the formation of compact and well-defined country profiles. Beyond these quantitative improvements, the superior performance of CENTNN stems from its competitive neural learning mechanism, which iteratively adapts cluster prototypes to the underlying data distribution rather than relying on fixed distance-based partitions. This adaptive representation enables the model to capture latent socioeconomic–health heterogeneity that conventional methods may overlook, particularly among transitional economies exhibiting similar economic characteristics but divergent mental health and longevity profiles. Consequently, CENTNN generates clusters that are statistically well separated and suitable for subsequent causal analysis.

Table 2. Comparison of clustering algorithms

Algorithm	Silhouette score	Davies-Bouldin index	Calinski-Harabasz index
K-means	0.302	1.04	90.97
Hierarchical	0.466	0.681	142.89
GMM	0.397	0.892	110.225
CENTNN	0.481	0.667	154.234

4.2. Centroid neural network cluster analysis and interpretation

The CENTNN-based clustering produced four distinct and interpretable country profiles, summarized in Table 3. Table 3 summarizes the mean socioeconomic and health characteristics across CENTNN-defined clusters. Cluster 0, comprising high-income nations, shows high life expectancy and low suicide rates, indicating a favorable socioeconomic-health profile. Cluster 1 includes upper-middle-income countries in rapid economic transition, where GDP is rising but suicide rates remain relatively elevated [3]. Cluster 2 represents emerging economies with moderate GDP and improving life expectancy but variable mental health outcomes. Cluster 3 includes low-income nations facing systemic socioeconomic challenges, characterized by shorter life spans and lower socioeconomic indicators.

Table 3. Mean indicator values by CENTNN cluster

Cluster	Life expectancy (years)–men	Life expectancy (years)–women	GDP per capita (USD)	Suicide rate (per 100,000)
0	80.93	84.67	52 268.7	13.32
1	72.98	78.78	9 224.7	29.21
2	74.12	78.95	9 321.2	15.31
3	63.62	69.06	2 446.7	13.82

To enhance interpretability, a sensitivity view of feature influence was adopted. Centroid separation indicates that GDP per capita mainly drives inter-cluster structure, distinguishing development levels, while suicide rate refines boundaries among economically similar countries and gender-specific life expectancy strengthens intra-cluster cohesion. Overall, CENTNN captures both macroeconomic and behavioral dimensions in its learned representation.

4.3. Choropleth interpretation and global patterns

To contextualize the CENTNN's latent structure, a choropleth world map, as shown in Figure 2 visualized the geographic distribution of the clusters. The map displayed a clear global north–south divide, with high-life-expectancy nations concentrated in North America, Western Europe, and East Asia, while lower-longevity clusters dominated Sub-Saharan Africa and parts of South Asia. Transitional economies in Latin America and Eastern Europe occupied intermediary zones.

Figure 2 highlights the socioeconomic gradient underlying the clusters: high-longevity countries concentrate in Western Europe, North America, and parts of East Asia, whereas low-longevity clusters dominate Sub-Saharan Africa, South Asia, and Eastern Europe, with transitional economies in between. This spatial structure shows that longevity disparities are geographically patterned and linked to historical, economic, and infrastructural inequalities [2], [25]. By capturing these multidimensional dependencies more effectively than linear models, the CENTNN provides a powerful tool for interpreting global inequality in health outcomes.

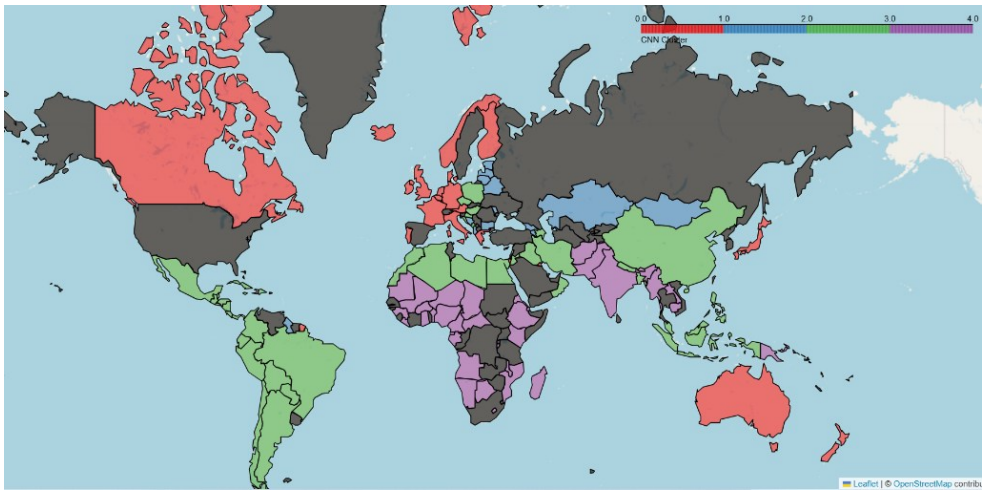

Figure 2. Choropleth map of CENTNN-defined clusters

4.4. Linear non-Gaussian acyclic model causal inference results

The LiNGAM causal inference analysis provided quantitative support for the directional relationships hypothesized through CENTNN-derived insights [12]. Across clusters, the model identified a consistent directional relationship from GDP per capita to life expectancy, suggesting that economic prosperity is positively associated with longevity [16], [26]. Interestingly, the directionality between suicide rate and life expectancy varied by cluster, consistent with prior findings that mental health can significantly moderate longevity outcomes [3]. The model's capacity to capture these asymmetrical relationships underscores the necessity of contextualizing causality within socioeconomic strata [4]. To assess statistical reliability, a bootstrap resampling procedure was applied within the LiNGAM framework. Repeated subsampling indicated the stability of coefficient magnitudes and edge directions across iterations. The GDP per capita → life expectancy paths consistently exhibited low variability, while suicide rate → life expectancy links remained negative and stable in most samples, indicating that the inferred causal structure is robust and not driven by random noise.

Figure 3 presents the global causal structure inferred by the LiNGAM model. The results confirm a robust, positive causal pathway from GDP per capita → life expectancy (men and women), establishing economic prosperity as a fundamental driver of longevity. Additionally, a weaker yet consistent suicide rate → life expectancy link was detected, with a negative direction, indicating that higher suicide prevalence modestly reduces overall life expectancy. These results empirically validate the theoretical model that economic and psychological dimensions jointly determine survival outcomes.

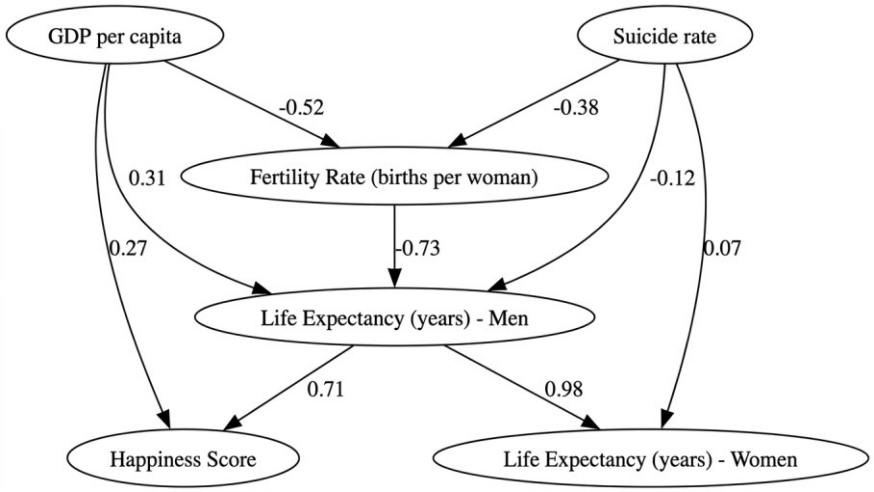

Figure 3. Directed causal graph estimated using LiNGAM

Figure 4 decomposes these causal relationships across CENTNN clusters, revealing substantial heterogeneity by socioeconomic context. In high-income clusters, the GDP → life expectancy path remains dominant but is moderated by suicide rate, which exerts a small but statistically significant negative influence. Conversely, in lower-income clusters, GDP shows directional relationships with life expectancy and suicide rate, suggesting that socioeconomic context influences the structure of these relationships. Transitional economies display more heterogeneous directional relationships, reflecting shifting health dynamics during economic modernization. This cluster-wise analysis suggests that relationships associated with longevity are context-dependent; patterns observed in affluent nations may not generalize to developing regions. Consequently, causal models of health must account for these stratified dynamics to inform equitable and region-specific policy interventions. Although the proposed framework captures major economic and psychosocial drivers of longevity, several relevant confounding factors, including healthcare expenditure, income inequality, education quality, fertility dynamics, and social protection, are not explicitly modeled. These variables may act as mediators, moderators, or latent confounders, influencing both cluster formation and the estimated causal pathways. Consequently, the reported relationships should be interpreted within the scope of the selected indicators. Future work should incorporate these factors to improve causal interpretability and further mitigate omitted-variable bias.

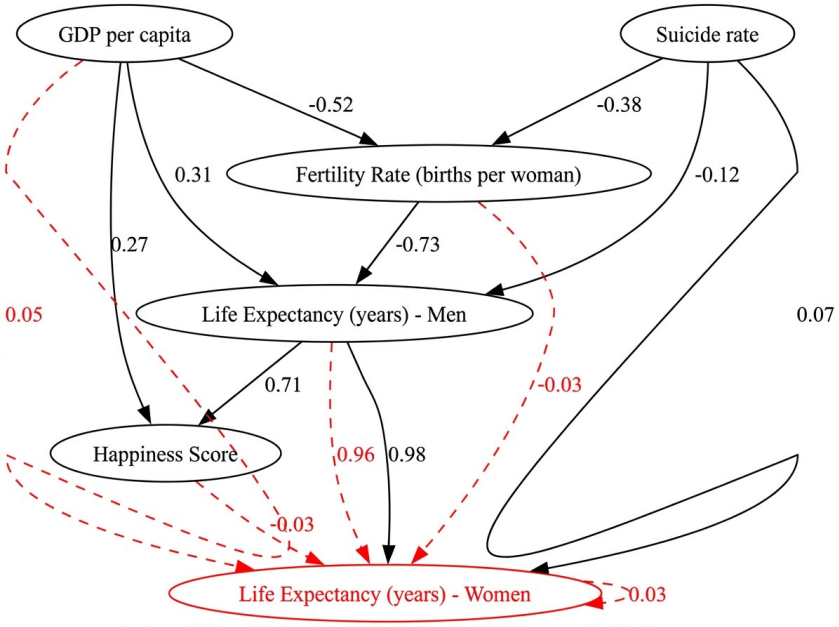

Figure 4. Cluster-wise LiNGAM causal networks

The robustness of the causal relationships was assessed through bootstrapping, yielding the edge-frequency plot in Figure 5. The GDP → life expectancy pathway appeared in over 90% of bootstrap iterations, indicating high structural stability in the inferred relationship. Other edges, particularly those linking suicide rate with life expectancy, showed moderate variability, suggesting sensitivity to local social and cultural contexts. This stability analysis reinforces the reliability of the LiNGAM results and strengthens the validity of the causal inferences. Countries in high-longevity clusters benefit from stable causal mechanisms: economic prosperity fosters healthcare access and psychosocial well-being, reinforcing a positive feedback loop. In contrast, countries in low-longevity clusters exhibit lower economic capacity and life expectancy alongside less favorable suicide-rate profiles. These patterns highlight the interconnected nature of socioeconomic and health conditions across countries, underscoring the need for integrated health and economic policies. Beyond empirical validation, these relationships have important implications for public health and socioeconomic planning: understanding how economic capacity and mental health interact to shape life expectancy can inform cross-sector strategies that align economic growth with mental health interventions, strengthen preventive care, and prioritize resources for populations most vulnerable to psychosocial and financial stress.

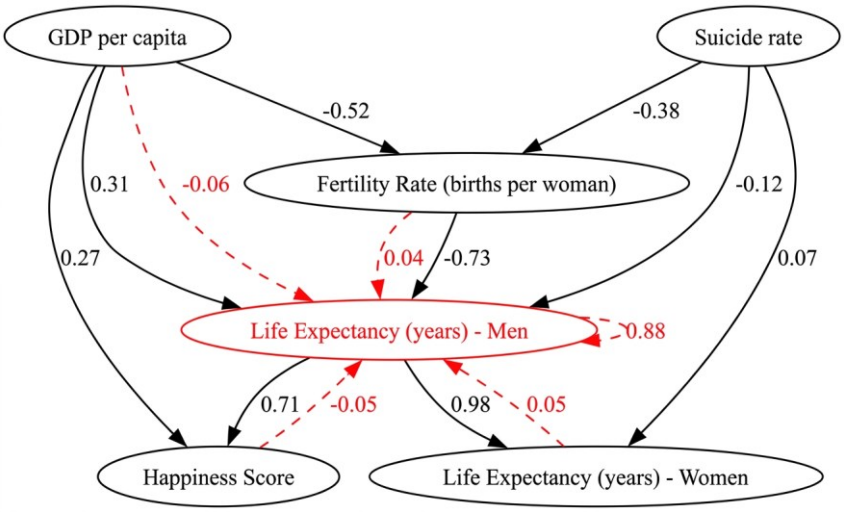

Figure 5. Bootstrapped stability of LiNGAM causal edges

5. CONCLUSION

This study proposed an XAI framework that integrates CENTNN clustering, choropleth spatial visualization, and LiNGAM causal discovery to investigate global disparities in life expectancy. By combining competitive neural learning with causal inference, the proposed methodology not only identifies latent socioeconomic–health structures but also explains the directional relationships underlying these patterns. The findings indicate that GDP per capita is the dominant upstream variable in the inferred structure, while suicide rate shows a weaker negative directional relationship with life expectancy across socioeconomic contexts. Beyond the empirical findings, the proposed framework demonstrates how neural representation learning and causal discovery can be integrated to provide interpretable decision-support for complex public health systems. These results highlight that improving life expectancy requires coordinated economic and mental health strategies tailored to heterogeneous national contexts. Future work should benchmark the proposed framework against recent deep clustering architectures while incorporating additional socioeconomic indicators and longitudinal data to improve causal robustness and external validity.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ilyas Tammouch	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Soumaya Nouna		✓		✓		✓		✓	✓	✓	✓	✓		
Abdelamine Elouafi	✓		✓	✓			✓			✓	✓		✓	
Abdessamad Elmotawakkil	✓		✓	✓			✓			✓	✓		✓	
Aziz Soulhi	✓		✓	✓			✓			✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest related to this work.

INFORMED CONSENT

This section is not applicable, as this study did not involve human participants or personally identifiable information.

ETHICAL APPROVAL

This section is not applicable, as this study did not involve human or animal subjects and used only non-sensitive computational or public data.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the Kaggle at https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who/data, reference [13].

REFERENCES

[1] World Health Organization, "The global health observatory: explore a world of health data," *who.int*, 2023. Accessed: Accessed: Aug. 4, 2025. [Online]. Available: https://www.who.int/data/gho

[2] M. Marmot, "The health gap: the challenge of an unequal world," *The Lancet*, vol. 386, no. 10011, pp. 2442–2444, Dec. 2015, doi: 10.1016/S0140-6736(15)00150-6.

[3] M. Naghavi, "Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the global burden of disease study 2016," *BMJ*, vol. 364, no. l94, Feb. 2019, doi: 10.1136/bmj.l94.

[4] J. Peters, D. Janzing, and B. Schölkopf, "Elements of causal inference: foundations and learning algorithms," in *Adaptive Computation and Machine Learning series*, Cambridge, United Kingdom: The MIT Press, 2017.

[5] A. L. Beam and I. S. Kohane, "Big data and machine learning in health care," *JAMA*, vol. 319, no. 13, Apr. 2018, doi: 10.1001/jama.2017.18391.

[6] T. Kohonen, "Essentials of the self-organizing map," *Neural Networks*, vol. 37, pp. 52–65, Jan. 2013, doi: 10.1016/j.neunet.2012.09.018.

[7] A. Nouna, S. Nouna, M. Mansouri, and A. Boujamaa, "Categorizing hyperspectral imagery using convolutional neural networks for land cover analysis," *International Journal of Informatics and Communication Technology*, vol. 14, no. 2, pp. 393–404, Aug. 2025, doi: 10.11591/ijict.v14i2.pp393-404.

[8] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long, "A survey of clustering with deep learning: from the perspective of network architecture," *IEEE Access*, vol. 6, pp. 39501–39514, 2018, doi: 10.1109/ACCESS.2018.2855437.

[9] D.-C. Park, "Centroid neural network with spatial constraints," in *Advances in Knowledge Discovery and Data Mining-(PAKDD 2009)*, Berlin, Germany: Springer, 2009, pp. 915–922, doi: 10.1007/978-3-642-01307-2_96.

[10] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annual Review of Statistics and Its Application*, vol. 6, no. 1, pp. 355–378, Mar. 2019, doi: 10.1146/annurev-statistics-031017-100325.

[11] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 19–41, Jan. 2000, doi: 10.1006/dspr.1999.0361.

[12] S. Shimizu, "Lingam: non-Gaussian methods for estimating causal structures," *Behaviormetrika*, vol. 41, no. 1, pp. 65–98, Jan. 2014, doi: 10.2333/bhmk.41.65.

[13] K. Rajarshi, "Life expectancy and suicide rates dataset," *Kaggle.com*, 2017. Accessed: Aug. 4, 2025. [Online]. Available: https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who/data

[14] W. McKinney, "Data structures for statistical computing in Python," in *9th Python in Science Conference (SciPy 2010)*, 2010, pp. 56–61, doi: 10.25080/Majora-92bf1922-00a.

[15] L. Buitinck *et al.*, "API design for machine learning software: experiences from the scikit-learn project," Sep. 2013, *arXiv: 1309.0238.*

[16] V. Varbanova, N. Hens, and P. Beutels, "Determinants of life-expectancy and disability-adjusted life years (DALYs) in European and Organisation for Economic Co-operation and Development (OECD) countries: a longitudinal analysis (1990–2019)," *SSM - Population Health*, vol. 24, Dec. 2023, doi: 10.1016/j.ssmph.2023.101484.

[17] A. E. Ezugwu *et al.*, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, Apr. 2022, doi: 10.1016/j.engappai.2022.104743.

[18] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.

[19] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.

[20] E. Novikova, *Best map projections*. Cham, Switzerland: Springer Nature, 2025, doi: 10.1007/978-3-031-78334-0.

[21] M. Monmonier, "Lying with maps," *Statistical Science*, vol. 20, no. 3, pp. 215–222, 2005, doi: 10.1214/088342305000000241.

[22] A. Hyvärinen, "Independent component analysis: recent advances," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, Feb. 2013, doi: 10.1098/rsta.2011.0534.

[23] S. Shimizu *et al.*, "DirectLiNGAM: a direct method for learning a linear non-Gaussian structural equation model," *Journal of Machine Learning Research*, vol. 12, no. 33, pp. 1225−1248, 2011.

[24] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, Jan. 1974, doi: 10.1080/03610927408827101.

[25] United Nations Development Programme, "Human development report 2020-the next Frontier: human development and the Anthropocene," *hdr.undp.org*, 2020. Accessed: Aug. 4, 2025. [Online]. Available: https://hdr.undp.org/content/human-development-report-2020

[26] A. A. V. Raalte, "What have we learned about mortality patterns over the past 25 years?," *Population Studies*, vol. 75, pp. 105–132, Dec. 2021, doi: 10.1080/00324728.2021.1967430.

BIOGRAPHIES OF AUTHORS



Dr. Ilyas Tammouch is a professor at the National Higher School of Mines of Rabat (ENSMR), Rabat, Morocco. His research interests include machine learning, deep learning, data analysis, evaluation systems, and AI applications in socioeconomic and health domains. He has authored and co-authored several scientific publications and actively contributes to research in computational intelligence and decision-support systems. He can be contacted at email: tammouch@enim.ac.ma.



Soumaya Nouna is a researcher at the Laboratory for Analysis and Modeling of Systems and Decision Support (LAMSAD), Department of Mathematics and Informatics, ENSA Berrechid, Hassan First University of Settat, Berrechid, Morocco. She holds a Ph.D. in Mathematics and AI and specializes in mathematics, machine learning, and deep learning. Her research interests include differential equations, machine learning algorithms, and AI. She has authored numerous research articles and continues to advance her expertise through research in mathematics and AI. She can be contacted at email: s.nouna@uhp.ac.ma.



Abdelamine Elouafi is a professor at the National School of Business and Management of Dakhla (ENCG Dakhla), SIV Laboratory, Faculty of Sciences, Ibn Zohr University, Morocco. His research interests include machine learning, deep learning, data analysis, and system evaluation for decision-making applications. He holds a master's degree in Decision Informatics from Sultan Moulay Slimane University and has been actively involved in teaching and research in AI and data science. He can be contacted at email: a.elouafi@uiz.ac.ma.



Abdessamad Elmotawakkil is a researcher and faculty member in the Department of Computer Science, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco. His research interests focus on computer science, AI, data analysis, and information systems. He contributes to research and teaching activities in computational and applied intelligence. He can be contacted at email: Abdessamad.Elmotawakkil@uit.ac.ma.



Aziz Soulhi is a full professor at the National Higher School of Mines Rabat (ENSMR) in Morocco, where he leads the Management and Engineering of Systems (MIS) research team. He completed his Ph.D. in Production, Automation, and Industrial Computing from École Centrale de Lille, France, in 2000. His extensive academic research centers on industrial engineering, supply chain 4.0, total productive maintenance (TPM), fuzzy logic decision-making models, and the integration of AI in logistics and production systems. He can be contacted at email: soulhi@enim.ac.ma.