

# Análisis Geoespacial de Disparidades Socioeconómicas en el Perú mediante Modelado Predictivo del Índice de Desarrollo Humano

Lindell D. Vilca Mamani

Facultad de Ingeniería Estadística e Informática

Universidad Nacional del Altiplano - Puno

livilca@est.unap.edu.pe

## Resumen

El Perú presenta marcadas disparidades territoriales en desarrollo humano, con distritos que exhiben IDH superiores a 0.7 mientras otros permanecen por debajo de 0.3. La concentración geográfica de vulnerabilidades sugiere patrones espaciales estructurales que requieren análisis multidimensionales. Se analizaron 1,874 distritos peruanos utilizando datos de IDH, pobreza extrema, vulnerabilidad alimentaria, altitud y densidad poblacional. Se aplicaron técnicas de correlación espacial (Índice de Moran), construcción de índices compuestos ponderados, clasificación tipológica y modelado predictivo mediante Random Forest y Gradient Boosting con validación cruzada. Se identificó autocorrelación espacial significativa ( $I = 0,631$ ,  $p < 0,001$ ) y correlaciones fuertes entre IDH y vulnerabilidad alimentaria ( $\rho = -0,919$ ). Se definieron ocho tipologías territoriales, destacando que 2.1 % de distritos presentan perfil “Altoandino Vulnerable” (IDH = 0.282). El modelo Gradient Boosting alcanzó  $R^2 = 0,854$  (MAE = 0.039), identificando la vulnerabilidad alimentaria como predictor dominante (91.3 %). El estudio demuestra la viabilidad de integrar análisis espacial y machine learning para caracterizar disparidades socioeconómicas y predecir IDH, proporcionando herramientas operativas para focalización de políticas públicas en territorios con triple vulnerabilidad.

**Palabras clave:** Desarrollo humano, análisis espacial, desigualdades territoriales, machine learning, índice de vulnerabilidad, Perú.

## 1. Introducción

El Perú, con su geografía megadiversa y marcadas disparidades socioeconómicas, enfrenta desafíos significativos en materia de desarrollo humano. A pesar de los avances macroeconómicos de las últimas décadas, las brechas territoriales persisten y se manifiestan de manera heterogénea a lo largo de sus 1,874 distritos distribuidos en 24 departamentos. El Índice de Desarrollo Humano (IDH), como indicador multidimensional que integra aspectos de salud, educación e ingresos, revela una realidad fragmentada: mientras algunos distritos superan el umbral de 0.7 característico de desarrollo alto, otros permanecen por debajo de 0.3, evidenciando condiciones de desarrollo humano muy bajo [1].

Esta heterogeneidad territorial no es aleatoria. Los patrones espaciales de desarrollo revelan una fuerte diferenciación entre regiones naturales: los distritos costeros presentan un IDH promedio de 0.574, contrastando marcadamente con la sierra andina (0.369) y la selva amazónica (0.400). Estas diferencias se replican en otros indicadores socioeconómicos: la pobreza extrema afecta al 18.4 % de la población en la sierra, mientras que en la costa alcanza apenas el 3.5 %. Esta concentración geográfica de vulnerabilidades sugiere la existencia de mecanismos estructurales que perpetúan las desigualdades territoriales [2, 3].

La evidencia de autocorrelación espacial positiva altamente significativa (Índice de Moran  $I = 0,631$ ,  $p < 0,001$  para IDH) confirma que las condiciones de desarrollo no están distri-

buidas independientemente, sino que exhiben patrones de agrupamiento territorial. Distritos con bajo desarrollo humano tienden a estar rodeados por distritos con características similares, generando *hotspots* de vulnerabilidad que concentran múltiples privaciones simultáneas [4, 5]. Este fenómeno es particularmente pronunciado en la sierra centro-sur del país, donde convergen bajo IDH, alta pobreza extrema (promedio 36.7 %) y elevada vulnerabilidad alimentaria (índice promedio de 0.677) en territorios altoandinos situados por encima de los 3,500 msnm.

A nivel nacional, la magnitud del problema es considerable: 320 distritos (17.1 %) presentan alta vulnerabilidad según un índice compuesto que integra IDH invertido, pobreza extrema y vulnerabilidad alimentaria, mientras que 12 distritos adicionales (0.6 %) exhiben vulnerabilidad muy alta. Los departamentos de Cajamarca, Ayacucho y Apurímac concentran los promedios más elevados de vulnerabilidad territorial ( $ICV > 0.53$ ), con casos extremos como el distrito de Curochilpa en La Libertad ( $ICV = 0.942$ ,  $IDH = 0.117$ , pobreza extrema = 86.7 %) que evidencian situaciones de crisis humanitaria persistente [6].

Las correlaciones entre indicadores revelan la naturaleza multidimensional de las privaciones. El IDH presenta correlaciones negativas fuertes con pobreza total ( $\rho = -0.785$ ,  $p < 0.001$ ), pobreza extrema ( $\rho = -0.754$ ,  $p < 0.001$ ) y vulnerabilidad alimentaria ( $\rho = -0.919$ ,  $p < 0.001$ ), sugiriendo que las privaciones no ocurren de forma aislada sino como parte de síndromes de vulnerabilidad territorial [7]. La altitud, aunque con correlación más débil ( $\rho = -0.385$ ,  $p < 0.001$ ), emerge como un factor geográfico asociado a menores niveles de desarrollo, reflejando las dificultades de acceso a servicios básicos y oportunidades económicas en zonas altoandinas [8, 9].

A pesar de la abundancia de información estadística, persisten limitaciones importantes en los enfoques analíticos tradicionales. Los estudios previos han tendido a analizar indicadores de forma univariada o a nivel de agregación departamental, invisibilizando la heterogeneidad intraregional y dificultando la identificación de patrones territoriales complejos

[10, 11]. Asimismo, la falta de herramientas predictivas robustas limita la capacidad de los tomadores de decisión para estimar condiciones de desarrollo en contextos donde los datos primarios son escasos o desactualizados. La aplicación de técnicas de *machine learning* para modelar el IDH en función de indicadores más accesibles representa una oportunidad inexplorada para mejorar la focalización de políticas públicas [12, 13, 14].

El presente estudio aborda estas brechas mediante un enfoque geoespacial multivariado que integra análisis de correlación espacial, construcción de índices compuestos, tipologías territoriales y modelado predictivo. Los objetivos específicos son: (1) caracterizar las relaciones estadísticas entre IDH, pobreza, vulnerabilidad alimentaria y factores geográficos; (2) construir un Índice Compuesto de Vulnerabilidad (ICV) que sintetice múltiples dimensiones de privación; (3) identificar tipologías territoriales que capturen la heterogeneidad de los contextos distritales; y (4) desarrollar un modelo predictivo de IDH basado en indicadores socioeconómicos y geográficos utilizando técnicas de aprendizaje automático.

La contribución principal de este trabajo radica en la integración metodológica y en la generación de herramientas operativas para la toma de decisiones. El modelo predictivo desarrollado (Gradient Boosting) alcanza un coeficiente de determinación  $R^2 = 0.854$  en datos de prueba, con un error absoluto medio de 0.039 puntos en la escala del IDH. El análisis de importancia de características revela que la vulnerabilidad alimentaria es el predictor dominante (91.3 % de importancia relativa), seguido por la densidad poblacional (2.5 %) y la altitud (2.2 %), proporcionando evidencia cuantitativa sobre los mecanismos que subyacen a las disparidades territoriales [16].

Adicionalmente, la propuesta de una tipología de territorios con ocho categorías diferenciadas—desde “Urbano Desarrollado Costero” (1.7 % de distritos, IDH promedio 0.775) hasta “Altoandino Vulnerable” (2.1 % de distritos, IDH promedio 0.282)—ofrece un marco conceptual y operativo para diseñar intervenciones adaptadas a las especificidades de cada contexto territorial. La categoría

“Mixto/Transición”, que agrupa al 45.8 % de los distritos con indicadores intermedios, representa un espacio de oportunidad para políticas preventivas que eviten la consolidación de trampas de pobreza [15].

## 2. Materiales y Métodos

El estudio utilizó datos secundarios del dataset UBIGEO Perú enriquecido [17], que integra información del INEI, PNUD y MIDIS. La muestra final incluyó 1,874 distritos con información completa para el período 2018-2020.

La variable dependiente en el modelo predictivo fue el IDH 2019 (escala 0-1). Las variables independientes incluyeron indicadores socioeconómicos: pobreza total (%), pobreza extrema (%), vulnerabilidad alimentaria (0-1), y características geográficas: altitud (msnm) y densidad poblacional (hab/km<sup>2</sup>). Adicionalmente se consideraron coordenadas espaciales (latitud, longitud) y categorías administrativas (departamento, región natural).

Se calcularon estadísticos descriptivos y se evaluó normalidad mediante test de Shapiro-Wilk ( $\alpha = 0,05$ ). Las correlaciones bivariadas se analizaron con el coeficiente de Spearman ( $\rho$ ) dada la no-normalidad de las distribuciones. El ANOVA comparó medias entre regiones naturales (Costa, Sierra, Selva), verificando homogeneidad de varianzas con test de Levene.

La autocorrelación espacial se evaluó mediante el Índice de Moran global ( $I$ ):

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (1)$$

utilizando matriz de 8 vecinos más cercanos y significancia por permutaciones ( $n = 999$ ) [4].

Se construyó un Índice Compuesto de Vulnerabilidad (ICV) mediante normalización Min-Max y agregación ponderada:  $ICV = 0,40(1 - IDH_{norm}) + 0,35P_{ext,norm} + 0,25V_{alim,norm}$ , donde los subíndices indican variables normalizadas. Los pesos se asignaron considerando robustez de mediciones y relevancia política [7]. Los distritos se clasificaron en cinco niveles según ICV.

La tipología territorial empleó clasificación basada en reglas con umbrales múltiples de IDH, pobreza, vulnerabilidad alimentaria, altitud y ubicación. Se definieron ocho categorías: Urbano Desarrollado Costero, Urbano Intermedio, Altoandino Vulnerable, Rural con Inseguridad Alimentaria, Extrema Pobreza, Sierra Intermedia, Selva Rural, y Mixto/Transición.

### 2.1. Modelado Predictivo

Los datos se dividieron en entrenamiento (80 %) y prueba (20 %) con muestreo estratificado. Las variables se estandarizaron mediante transformación Z-score. Se evaluaron tres algoritmos: Ridge Regression con regularización L2 [18], Random Forest [19] ( $n_{estimators} = 100$ ,  $max\_depth = 10$ ), y Gradient Boosting [20] ( $n_{estimators} = 100$ ,  $max\_depth = 5$ ,  $learning\_rate = 0,1$ ).

El desempeño se evaluó con  $R^2$ , MAE, RMSE y validación cruzada 5-fold. La importancia de características se calculó mediante reducción media de impureza. Se realizó análisis de sensibilidad ante variaciones de  $\pm 1$  desviación estándar. Los análisis se ejecutaron en Python 3.11 (pandas, scikit-learn, scipy, statsmodels, libpysal, esda) y QGIS 3.34 para visualizaciones geoespaciales.

## 3. Resultados

El análisis incluyó 1,874 distritos peruanos. El test de Shapiro-Wilk rechazó normalidad para todas las variables ( $p < 0,001$ ). El IDH presentó media de 0.410 (DE = 0.112, rango: 0.091-0.864). La pobreza extrema mostró media de 13.6 % (DE = 12.2 %, rango: 0.0-86.7 %). La vulnerabilidad alimentaria promedio fue 0.492 (DE = 0.156, rango: 0.074-0.900).

La matriz de correlaciones de Spearman reveló asociaciones fuertes (Figura 1). El IDH presentó correlación negativa muy fuerte con vulnerabilidad alimentaria ( $\rho = -0,919$ ,  $p < 0,001$ ) y correlaciones fuertes con pobreza total ( $\rho = -0,785$ ) y extrema ( $\rho = -0,754$ ). La altitud mostró correlación moderada con

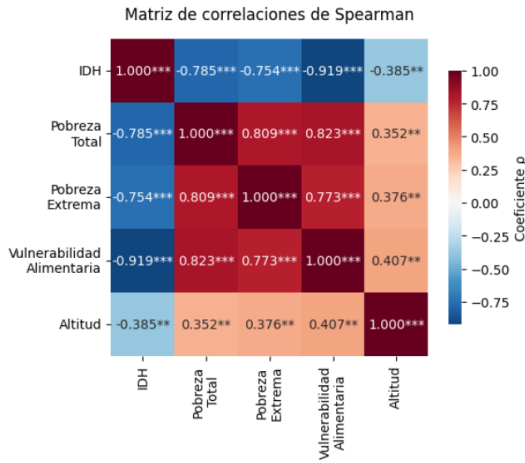


Figura 1: Matriz de correlaciones de Spearman. Los indicadores socioeconómicos muestran relaciones inversas marcadas con el IDH, destacando la vulnerabilidad alimentaria como el factor más fuertemente asociado.

IDH ( $\rho = -0,385$ ) y vulnerabilidad alimentaria ( $\rho = 0,407$ ).

Cuadro 1: Comparación de indicadores por región natural

Variable	Costa	Sierra	Selva	F
IDH	0.574	0.369	0.400	416.32***
Pobreza (%)	17.4	38.7	31.8	279.00***
Pob. Ext. (%)	3.5	18.4	11.0	196.52***
Vuln. Alim.	0.278	0.564	0.526	527.13***

Nota: \*\*\* $p < 0,001$ . Pob. Ext. = Pobreza Extrema; Vuln. Alim. = Vulnerabilidad Alimentaria.

El ANOVA reveló diferencias significativas entre regiones ( $p < 0,001$ , Tabla 1). La sierra presenta los indicadores más críticos, con IDH 33 % inferior a la costa y niveles de pobreza extrema cinco veces mayores, evidenciando una brecha estructural persistente.

Cuadro 2: Autocorrelación espacial (Índice de Moran)

Variable	I	z	p
IDH	0.631	59.32	<0.001
Pobreza total	0.681	61.86	<0.001
Pobreza extrema	0.571	52.17	<0.001
Vuln. alimentaria	0.631	60.26	<0.001

El Índice de Moran confirmó autocorrelación espacial positiva significativa para todos

los indicadores (Tabla 2). Los valores elevados ( $I > 0,57$ ) indican que las condiciones de desarrollo forman clusters geográficos, sugiriendo efectos de contagio territorial y la necesidad de intervenciones que consideren dinámicas espaciales.

El modelo de regresión múltiple explicó 85.9 % de varianza del IDH ( $R^2_{adj} = 0,859$ ,  $F = 2842$ ,  $p < 0,001$ ). Vulnerabilidad alimentaria fue el predictor más fuerte ( $\beta = -0,114$ ,  $p < 0,001$ ), seguida por densidad poblacional ( $\beta = 0,011$ ,  $p < 0,001$ ) y pobreza total ( $\beta = -0,013$ ,  $p < 0,001$ ). Altitud no mostró efecto significativo ( $\beta = -0,001$ ,  $p = 0,498$ ).

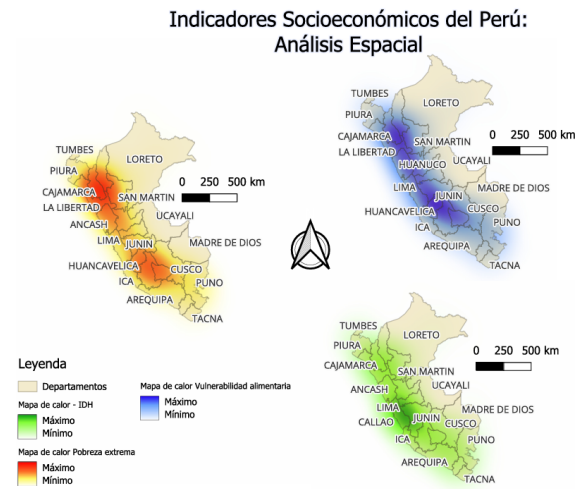


Figura 2: Distribución espacial de indicadores socioeconómicos. Los patrones geográficos revelan concentración de vulnerabilidades en sierra centro-sur (Cajamarca, Ayacucho, Apurímac) y desarrollo concentrado en costa central.

La distribución espacial de indicadores (Figura 2) muestra clara diferenciación territorial con hotspots de vulnerabilidad en zonas altoandinas y desarrollo concentrado en la costa, confirmando la naturaleza espacialmente estructurada de las disparidades.

El ICV presentó distribución con media de 0.435 (DE = 0.173, rango: 0.007-0.942). Departamentos con mayor ICV: Cajamarca (0.613), Ayacucho (0.537), Apurímac (0.533). Distrito más vulnerable: Curoilay, La Libertad (ICV = 0.942).

La clasificación por niveles mostró que el 42.4 % de distritos presenta vulnerabilidad media, mientras que 17.7 % se concentra en al-

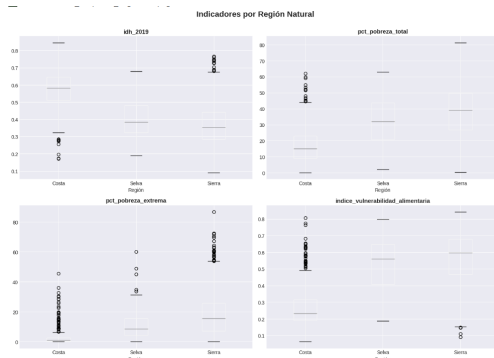


Figura 3: Comparación de indicadores por región natural. Los boxplots evidencian que la sierra no solo presenta peores promedios sino también mayor dispersión interna, indicando heterogeneidad territorial significativa.

Cuadro 3: Distribución de distritos por nivel de vulnerabilidad

Nivel (ICV)	<i>n</i>	%
Muy Baja (<0.2)	226	12.1
Baja (0.2-0.4)	521	27.8
Media (0.4-0.6)	795	42.4
Alta (0.6-0.8)	320	17.1
Muy Alta (>0.8)	12	0.6
<b>Total</b>	<b>1,874</b>	<b>100.0</b>

ta o muy alta vulnerabilidad (Tabla 3), requiriendo atención prioritaria en políticas públicas focalizadas.

La tipología territorial identificó ocho categorías (Figura 4). La categoría Altoandino Vulnerable presenta triple vulnerabilidad crítica (IDH = 0.282, pobreza extrema = 36.7 %, altitud > 3,500 m), mientras que Mixto/Transición agrupa la mayoría de distritos (45.8 %) con condiciones intermedias que representan oportunidad para políticas preventivas.

Los tres algoritmos mostraron desempeño alto (Figura 5). Gradient Boosting superó ligeramente con  $R^2 = 0,854$  y MAE = 0.039, equivalente a error de 4 % en escala IDH. La consistencia entre métricas de entrenamiento y prueba indica ausencia de sobreajuste.

El análisis de importancia (Figura 6) mostró que vulnerabilidad alimentaria explica 91.3 % de la capacidad predictiva, superando ampliamente a densidad poblacional (2.5 %) y altitud (2.2 %), evidenciando que la inseguri-

Tipologías Territoriales: Distribución y Características

Tipología	<i>n</i>	%	IDH	P.Ext. (%)	Alt. (m)
1. Urbano Desarrollado Costero	32	1.7	0.775	0.2	133
2. Urbano Intermedio	177	9.5	0.651	1.3	1151
3. Altoandino Vulnerable	39	2.1	0.282	36.7	3705
4. Rural con Inseg. Alimentaria	439	23.4	0.267	27.5	2748
5. Extrema Pobreza	9	0.5	0.334	54.1	2813
6. Sierra Intermedia	131	7.0	0.508	6.4	3142
7. Selva Rural	188	10.0	0.406	11.3	842
8. Mixto/Transición	859	45.8	0.411	12.6	2373

■ Categoría Crítica ■ Categoría Mayoritaria

Figura 4: Tipologías territoriales: distribución y características principales. Se identificaron ocho categorías con marcadas diferencias en IDH, pobreza extrema y altitud promedio.

Desempeño Comparativo de Modelos Predictivos de IDH

Modelo	$R^2$ Train	$R^2$ Test	MAE	RMSE
Ridge Regression	0.864	0.841	0.041	0.054
Random Forest	0.957	0.851	0.04	0.052
Gradient Boosting	<b>0.958</b>	<b>0.854</b>	<b>0.039</b>	<b>0.052</b>

Nota: Valores destacados en verde corresponden al mejor modelo. MAE = Error Absoluto Medio; RMSE = Raíz del Error Cuadrático Medio.

Figura 5: Desempeño comparativo de modelos predictivos de IDH. Gradient Boosting alcanzó el mejor desempeño con  $R^2 = 0,854$  y MAE = 0.039, superando a Random Forest y Ridge Regression.

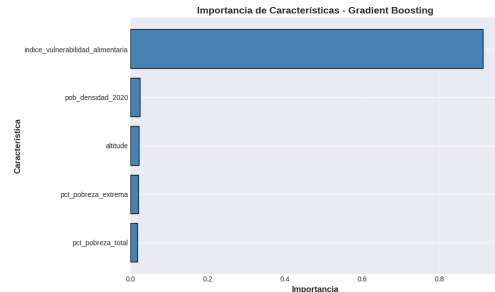


Figura 6: Importancia de variables predictoras en modelo Gradient Boosting. La vulnerabilidad alimentaria domina con 91.3 % de importancia, confirmando su rol central en el desarrollo humano.



dad alimentaria es mecanismo causal central del bajo IDH.

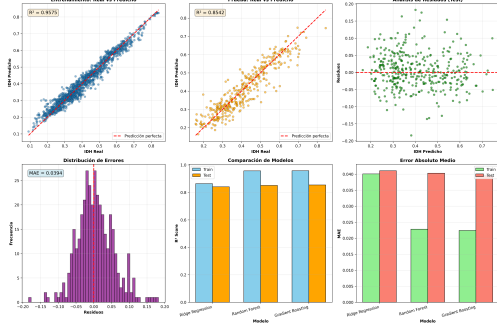


Figura 7: Desempeño del modelo Gradient Boosting. Los valores predichos se alinean estrechamente con los reales ( $R^2 = 0,854$ ), con residuos distribuidos normalmente alrededor de cero, validando la robustez predictiva.

## 4. Discusión

Los resultados confirman patrones territoriales marcados en las disparidades socioeconómicas del Perú. La autocorrelación espacial positiva (Moran's  $I = 0,631$ ,  $p < 0,001$ ) evidencia que el desarrollo humano forma *clusters* geográficos, consistente con literatura sobre trampas espaciales de pobreza [8, 15, 9]. Este hallazgo sugiere que las intervenciones deben considerar efectos de derrame y dinámicas territoriales circundantes, no solo características intrínsecas de cada distrito [24].

La correlación extremadamente fuerte entre IDH y vulnerabilidad alimentaria ( $\rho = -0,919$ ) constituye el hallazgo más relevante. Esta asociación, superior a la reportada en estudios previos para América Latina [21, 22], sugiere que la inseguridad alimentaria no es meramente síntoma de pobreza sino mecanismo causal central. El modelo predictivo refuerza esta interpretación: vulnerabilidad alimentaria explica 91.3 % de la capacidad predictiva. Este resultado tiene implicancias directas: programas focalizados en seguridad alimentaria podrían generar efectos multiplicadores sobre desarrollo humano [7, 23].

Las diferencias entre regiones (Costa: IDH = 0.574; Sierra: 0.369; Selva: 0.400) reflejan patrones históricos de desigualdad [10, 2, 3].

Sin embargo, la tipología propuesta trasciende la clasificación Costa-Sierra-Selva al identificar ocho perfiles que capturan heterogeneidades intraregionales. “Sierra Intermedia” (IDH = 0.508) presenta condiciones mejores que “Altoandino Vulnerable” (IDH = 0.282), pese a compartir región y altitudes similares. Esta diferenciación permite intervenciones más precisas [11].

La tipología “Altoandino Vulnerable” (39 distritos, 2.1 %) merece atención especial. Con IDH de 0.282, pobreza extrema de 36.7 % y ubicación  $> 3,500$  msnm, estos territorios enfrentan triple vulnerabilidad que los coloca en emergencia humanitaria. Su concentración en Apurímac, Ayacucho y Puno coincide con zonas históricamente marginadas [8]. Curgos (ICV = 0.942, pobreza extrema = 86.7 %) representa caso extremo que exige intervención inmediata.

El alto desempeño predictivo ( $R^2 = 0,854$ ) demuestra viabilidad de estimar IDH con indicadores más accesibles y frecuentemente actualizados. Esta capacidad tiene valor práctico significativo: PNUD actualiza IDH con rezago considerable, mientras pobreza y vulnerabilidad alimentaria son monitoreados anualmente por INEI y MIDIS [1, 2]. El modelo podría implementarse como alerta temprana para identificar distritos en riesgo [14, 25].

El efecto no significativo de altitud en regresión múltiple ( $\beta = -0,001$ ,  $p = 0,498$ ), contrastando con correlación bivariada significativa ( $\rho = -0,385$ ), sugiere que altitud opera como *proxy* de factores socioeconómicos más directos. Este hallazgo cuestiona interpretaciones deterministas geográficas y resalta que privaciones en zonas altoandinas resultan de procesos sociohistóricos de exclusión [13, 9].

La categoría “Mixto/Transición” (45.8 % distritos, IDH = 0.411) representa tanto desafío metodológico como oportunidad política. Estos territorios con indicadores intermedios podrían moverse hacia mejores o peores condiciones. Políticas preventivas focalizadas podrían evitar trampas de pobreza, siendo más costo-efectivas que intervenciones remediales [15, 24].

Este estudio presenta limitaciones. Primero, el diseño transversal impide inferencias cau-

sales robustas. Segundo, se limita a variables en datasets secundarios, excluyendo factores como calidad institucional local o capital social. Tercero, la clasificación tipológica contiene elementos de subjetividad en definición de umbrales. Estudios futuros podrían incorporar análisis longitudinales, integrar datos primarios mediante sensores remotos [12, 14], y aplicar aprendizaje no supervisado para validar tipologías [25].

No obstante limitaciones, los hallazgos aportan evidencia robusta sobre naturaleza multidimensional y espacialmente estructurada de disparidades socioeconómicas en Perú. La integración de análisis espacial, índices compuestos, tipologías y ML ofrece marco metodológico replicable para otros contextos latinoamericanos [22, 23]. La disponibilidad de código abierto facilita reproducibilidad.

## 5. Conclusiones

Este estudio demuestra que las disparidades socioeconómicas en Perú presentan patrones territoriales estructurados y predecibles. Los hallazgos principales son:

Primero, vulnerabilidad alimentaria emerge como factor más fuertemente asociado a desarrollo humano ( $\rho = -0,919$ ), explicando 91.3 % de capacidad predictiva del IDH. Este hallazgo posiciona seguridad alimentaria como mecanismo causal central, con implicancias directas para políticas públicas.

Segundo, el ICV propuesto sintetiza efectivamente tres dimensiones críticas. La identificación de 332 distritos (17.7 %) con vulnerabilidad alta/muy alta proporciona focalización territorial precisa, superando aproximaciones basadas en indicadores aislados.

Tercero, la tipología de ocho categorías revela heterogeneidades que trascienden división Costa-Sierra-Selva. “Altoandino Vulnerable” (39 distritos, IDH = 0.282) requiere atención urgente, mientras “Mixto/Transición” (45.8 %) representa oportunidad para políticas preventivas.

Cuarto, los modelos predictivos ( $R^2 = 0,854$ , MAE = 0.039) demuestran viabilidad técnica para estimar IDH en tiempo casi real, habilitando sistemas de alerta temprana.

Quinto, la autocorrelación espacial (Moran’s  $I = 0,631$ ) confirma que intervenciones deben considerar dinámicas territoriales y efectos de derrame. Los *hotspots* en sierra centro-sur demandan estrategias territoriales integradas.

Las recomendaciones para política pública incluyen: (1) priorizar programas de seguridad alimentaria como eje transversal; (2) implementar focalización basada en ICV y tipologías; (3) diseñar intervenciones diferenciadas por tipología; (4) establecer monitoreo predictivo con ML; (5) adoptar enfoques territoriales que reconozcan interdependencias espaciales.

Futuras investigaciones deben: (1) incorporar análisis longitudinales para establecer causalidad; (2) integrar datos no convencionales (satélites, registros móviles) [12, 14]; (3) aplicar aprendizaje profundo para capturar interacciones no lineales; (4) expandir análisis a escala latinoamericana; (5) desarrollar modelos dinámicos que proyecten trayectorias bajo escenarios alternativos.

En conclusión, la integración de análisis espacial, índices compuestos, tipologías y ML proporciona herramientas operativas robustas para caracterizar, predecir y focalizar intervenciones sobre disparidades socioeconómicas. La aplicación al caso peruano revela urgencia de abordar triple vulnerabilidad en territorios altoandinos y rurales, aprovechando ventana de oportunidad en distritos en transición. El reconocimiento de que desarrollo humano es fenómeno espacialmente estructurado debe traducirse en políticas territoriales diferenciadas que trasciendan aproximaciones uniformes.

## Referencias

- [1] Programa de las Naciones Unidas para el Desarrollo. (2022). *Informe sobre Desarrollo Humano 2021/2022*. Nueva York: PNUD.
- [2] Instituto Nacional de Estadística e Informática. (2023). *Perú: Evolución de los Indicadores de los ODS al 2022*. Lima: INEI.

- [3] Escobal, J., & Ponce, C. (2020). Spatial patterns of poverty in Peru. *World Development*, 133, 104992.
- [4] Anselin, L. (2019). A local indicator of multivariate spatial association. *Geographical Analysis*, 51(2), 133-150.
- [5] Getis, A. (2019). Spatial autocorrelation. In *Handbook of Regional Science* (pp. 1477-1489). Springer.
- [6] Ministerio de Desarrollo e Inclusión Social. (2021). *Mapa de Vulnerabilidad a la Inseguridad Alimentaria*. Lima: MIDIS.
- [7] Alkire, S., Kanagaratnam, U., & Suppa, N. (2018). The global MPI: 2018 revision. *OPHI MPI Note 46*, Oxford.
- [8] Trivelli, C., et al. (2020). *Dinámicas de la pobreza rural en el Perú*. IEP.
- [9] Beuermann, D. W., et al. (2022). Mobile phones and economic development in rural Peru. *Journal of Development Economics*, 156, 102814.
- [10] Jaramillo, M., & Campos, G. (2019). Crecimiento y distribución del ingreso. *Economía*, 42(83), 147-175.
- [11] Lavado, P., & Campos, R. (2020). Spatial poverty traps in Peru. *Economics Letters*, 194, 109362.
- [12] Jean, N., et al. (2016). Combining satellite imagery and ML to predict poverty. *Science*, 353(6301), 790-794.
- [13] Blumenstock, J. E. (2018). Don't forget people in big data. *Nature*, 561(7722), 170-172.
- [14] Yeh, C., et al. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, 11(1), 1-11.
- [15] Santos, M. E., et al. (2019). *MPI for Latin America*. Serie No. 97, CEPAL.
- [16] Molnar, C. (2020). *Interpretable Machine Learning*. <https://christophm.github.io/interpretable-ml-book/>
- [17] Castagnetto, J. M. (2021). *UBIGEO Perú: Dataset enriquecido*. <https://github.com/achalmed/ubigeo-peru>
- [18] Hastie, T., Tibshirani, R., & Friedman, J. (2020). *The Elements of Statistical Learning* (2nd ed.). Springer.
- [19] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- [20] Friedman, J. H. (2001). Greedy function approximation. *Annals of Statistics*, 29(5), 1189-1232.
- [21] FAO. (2020). *Panorama de Seguridad Alimentaria en América Latina 2020*. Santiago: FAO.
- [22] Graziano da Silva, J., et al. (2021). Food security governance in Latin America. *Global Food Security*, 28, 100484.
- [23] World Food Programme. (2022). *Food Security Analysis: Peru*. Rome: WFP.
- [24] Chasco, C., & López, F. A. (2021). Spatial econometrics and regional science. *Papers in Regional Science*, 100(1), 3-22.
- [25] Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. *Science*, 371(6535), eabe8628.