

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

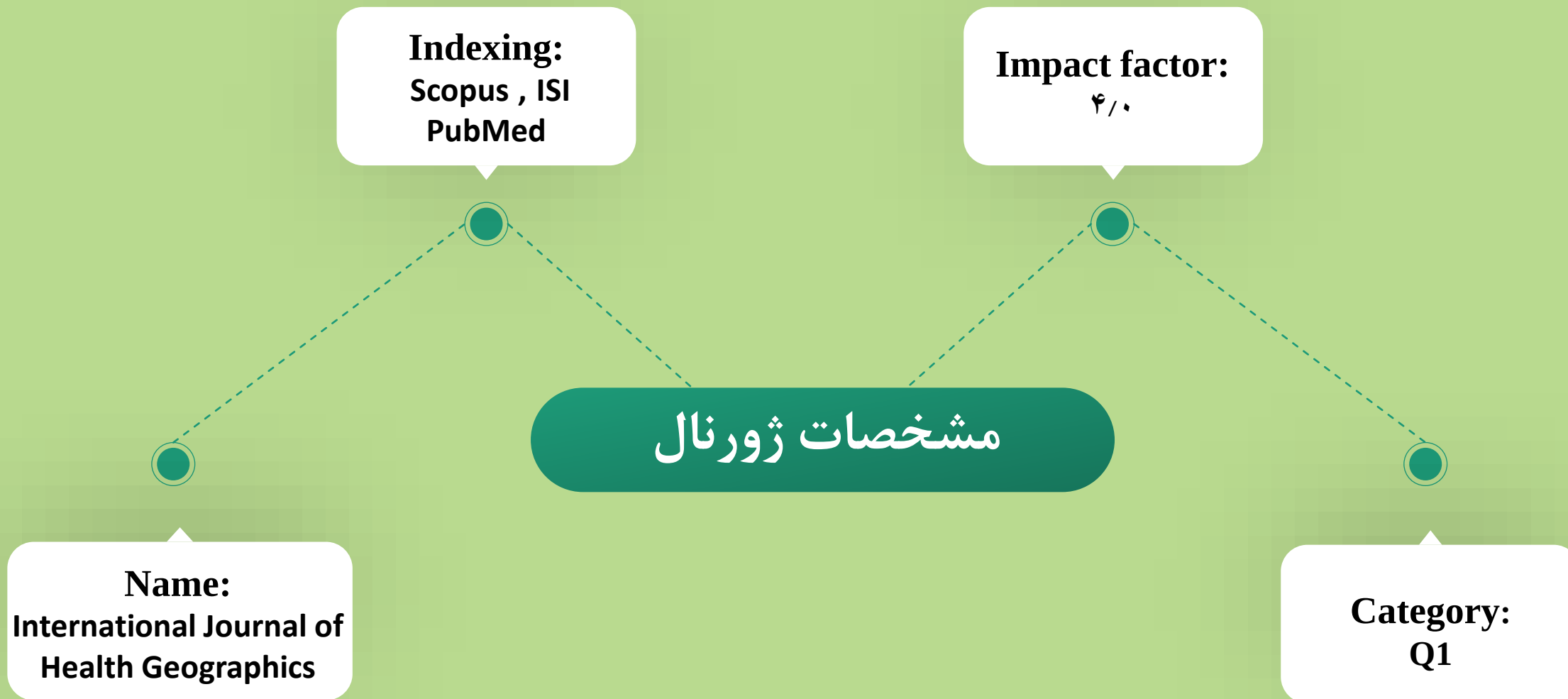
Socioeconomic and environmental determinants of asthma prevalence: a cross-sectional study at the U.S. County level using geographically weighted random forests

عوامل اجتماعی-اقتصادی و محیطی
تعیین کننده شیوع آسم:

یک مطالعه مقطعی در سطح شهرستان‌های ایالات
متحده با استفاده از جنگل‌های تصادفی وزن دار
جغرافیایی



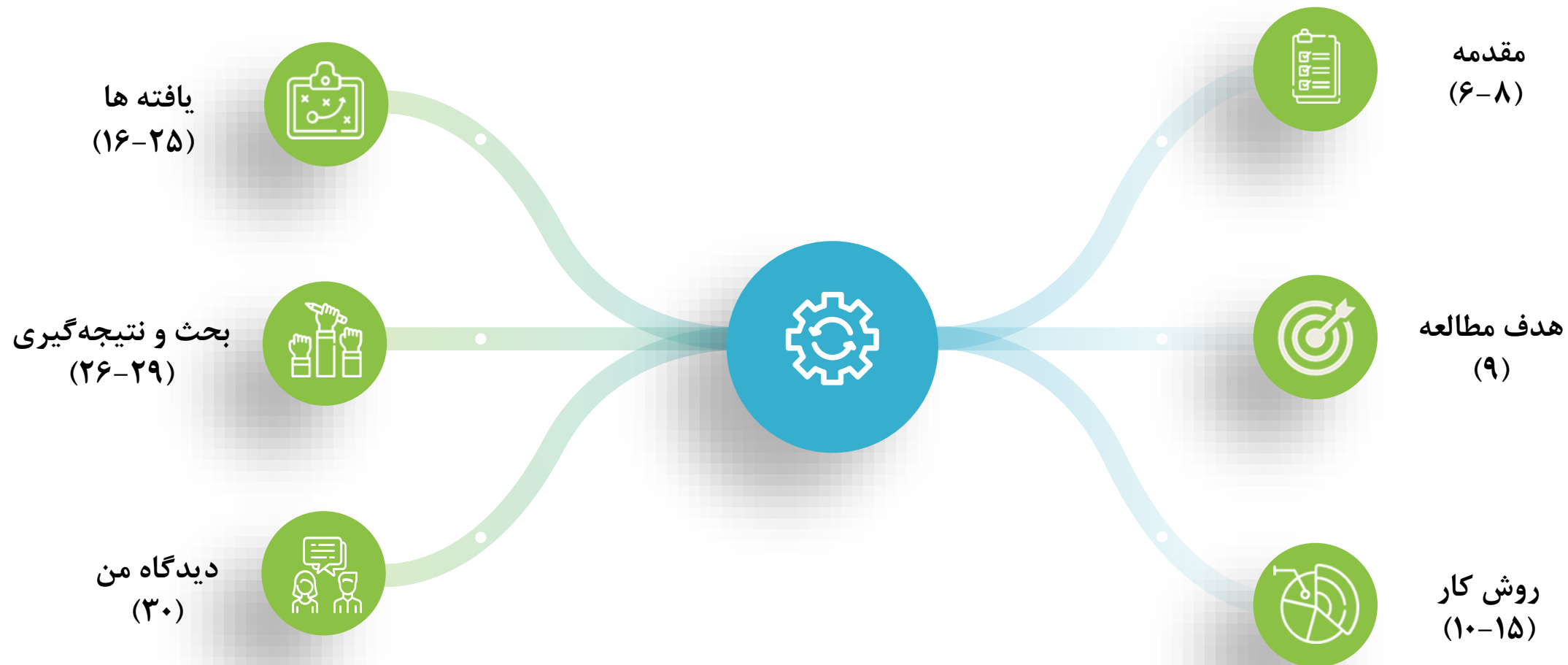
ارائه دهنده : ابوالفضل ندافی
استاد مشاور : جناب آقای دکتر مظاهری
تاریخ : ۱۴۰۴/۱۰/۱



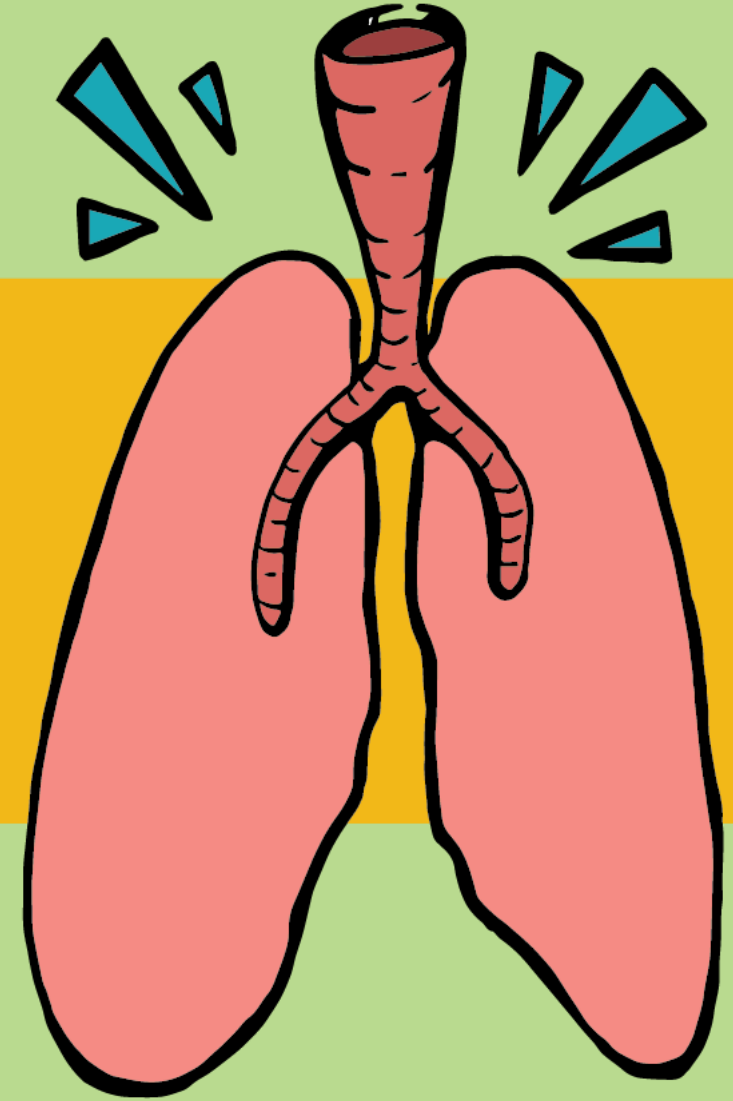
Abbreviations:

Name	Abbreviations	Meaning
GWRF	Geographically Weighted Random Forests	جنگل تصادفی وزن دار جغرافیایی
BRFSS	Behavioral Risk Factor Surveillance System	سیستم نظارت بر عوامل خطر رفتاری
RF	Random Forests	جنگل تصادفی
NDVI	Normalized Difference Vegetation Index	شاخص پوشش گیاهی نرمال شده

فهرست مطالب



مقدمه





➤ اهمیت بیماری آسم

آیا آسم بیماری مهمی است؟

در سال چند نفر در جهان به آسم مبتلا هستند؟

در سال چند نفر در ایالات متحده به آسم مبتلا هستند؟

چه عواملی در شیوع آسم نقش مهمی ایفا میکنند؟

اهداف مطالعه :

- هدف اصلی مطالعه:

✓ تحلیل رابطه بین میزان شیوع آسم و چندین عامل تعیین کننده اجتماعی و اقتصادی و محیطی در سطح شهرستان های ایالات متحده

- اهداف فرعی مطالعه :

✓ سنجش این موضوع که اهمیت نسبی هر یک از این عوامل خطر در مناطق مختلف جغرافیایی چگونه تغییر می کند.

✓ استفاده از مدل های یادگیری ماشین مانند GWRF برای در نظر گرفتن روابط غیرخطی و تعاملات بین متغیرها که در مطالعات قبلی نادیده گرفته شده بود.

روش کار



ملاحظات اخلاقی :



- این مطالعه از داده‌های BRFSS و اداره آمار ایالات متحده استفاده کرده است که همگی به صورت عمومی در دسترس هستند.

- در این نوع مطالعات که از داده‌های ثانویه و بدون نام استفاده می‌شود، معمولاً نیازی به دریافت تأییدیه اخلاقی جدید نیست؛ زیرا داده‌ها به گونه‌ای تجمعی (در سطح شهرستان) هستند که هویت هیچ فردی در آن‌ها قابل شناسایی نیست.

معیار های ورود و خروج:

افراد ۱۸ ساله و بالاتر

ساکنان خانوارهای غیرنهادی

از افراد پرسیده می شد که آیا تاکنون توسط یک متخصص سلامت به آسم مبتلا تشخیص داده شده اند و آیا همچنان آسم دارند یا خیر. پاسخ هایی که شامل «نمی دانم» یا «امتناع از پاسخ» بودند، از تحلیل حذف شدند و در برآوردهای ملی لحاظ نشدند.

متغیر ها و جمع آوری داده ها

فقر منطقه‌ای (poverty):

داده‌های فقر از *American Community Survey* استخراج شد. نرخ فقر بر اساس نسبت درآمد به خط فقر محاسبه می‌شود.

محدودیت در زبان انگلیسی (limited language):

این شاخص نسبت افرادی را نشان می‌دهد که توانایی محدودی در صحبت به زبان انگلیسی دارند. این داده نیز از *American Community Survey* به دست آمد.

جمعیت اقلیت (minority):

نسبت جمعیت غیرسفیدپوست و غیرهیسپانیک به کل جمعیت بزرگسال هر شهرستان.

فاقد بیمه درمانی (uninsured):

درصد افرادی که هیچ نوع پوشش بیمه سلامت ندارند.

شیوع سیگار کشیدن (Smoking):

شامل افرادی که حداقل ۱۰۰ نخ سیگار در طول عمر خود مصرف کرده‌اند و در زمان پیمایش به صورت روزانه یا گهگاهی سیگار می‌کشیدند.



متغیر ها و جمع آوری داده ها

• شیوع افسردگی (depression) :

افسردگی با استفاده از پرسشنامه استاندارد PHQ-9 سنجیده شد. نمره ۱۰ یا بالاتر به عنوان افسردگی در نظر گرفته شد.

• شیوع چاقی (obesity) :

درصد افراد چاق به ازای هر ۱۰۰,۰۰۰ نفر که از داده‌های BRFSS استخراج شد. شاخص توده بدنی بر اساس قد و وزن خوداظهاری محاسبه گردید.

• فضای سبز: (NDVI)

شاخص تفاضل نرمال شده پوشش گیاهی (NDVI) با استفاده از تصاویر ماهواره‌ای Landsat 8 و از طریق Google Earth Engine محاسبه شد. مقادیر مثبت تر نشان دهنده پوشش گیاهی بیشتر است.

• ذرات معلق ریز: (PM2.5)

داده‌ها از ایستگاه‌های پایش آژانس حفاظت محیط زیست آمریکا (EPA) استخراج شد. در مناطقی که اندازه‌گیری مستقیم وجود نداشت، از مدل‌های رگرسیون کاربری زمین و داده‌های ماهواره‌ای استفاده شد.

• غلظت ازن: (O3)

میانگین سالانه غلظت ازن بر حسب ppb برآورد گردید.

• دمای میانگین سالانه هوا (mean temperature) :

داده‌های دمایی که از دانشگاه ایالتی اورگن به دست آمد.



توسعه و اعتبار سنجی مدل :

- مدل Random Forest (RF) یک روش یادگیری تجمعی مبتنی بر درخت تصمیم است که روابط غیرخطی و تعامل متغیرها را بدون فروض آماری سخت‌گیرانه مدل‌سازی می‌کند. با این حال، RF تغییرات مکانی روابط را در نظر نمی‌گیرد. برای رفع این محدودیت، از Geographically Weighted Random Forest (GWRF) استفاده شد
- پارامترهای مدل شامل تعداد متغیرهای انتخاب‌شده در هر گره و تعداد درخت‌ها با استفاده از Random Grid Search تنظیم شدند.
- معیارهای ارزیابی شامل MSE، MAE، RMSE و R^2 بودند.





یافته ها

Table 1 Cross-validated prediction accuracy

	RF	GWRF
MSE	17.62	10.61
MAE	3.02	2.80
RMSE	4.20	3.20
R^2	0.82	0.89
Moran's I (residuals)	0.28 ($p < 0.05$)	0.0008 ($p > 0.05$)

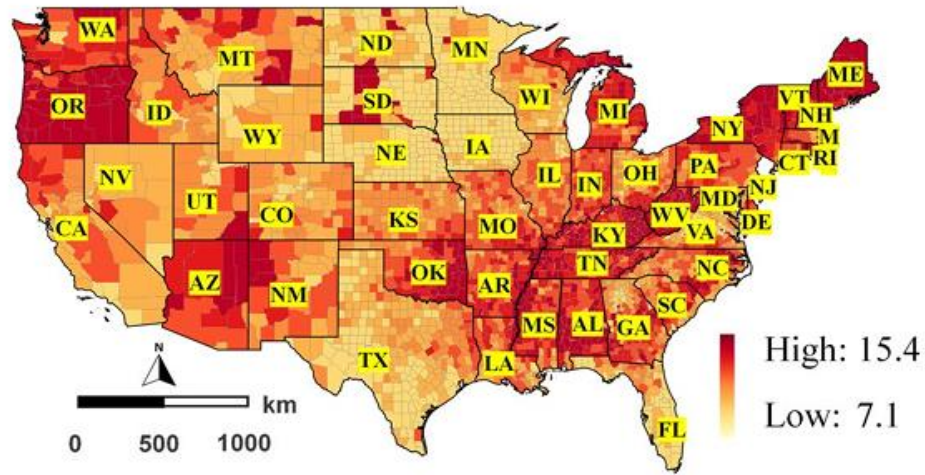
Table 1 Cross-validated prediction accuracy

	RF	GWRF
MSE	17.62	10.61
MAE	3.02	2.80
RMSE	4.20	3.20
R^2	0.82	0.89
Moran's I (residuals)	0.28 ($p < 0.05$)	0.0008 ($p > 0.05$)

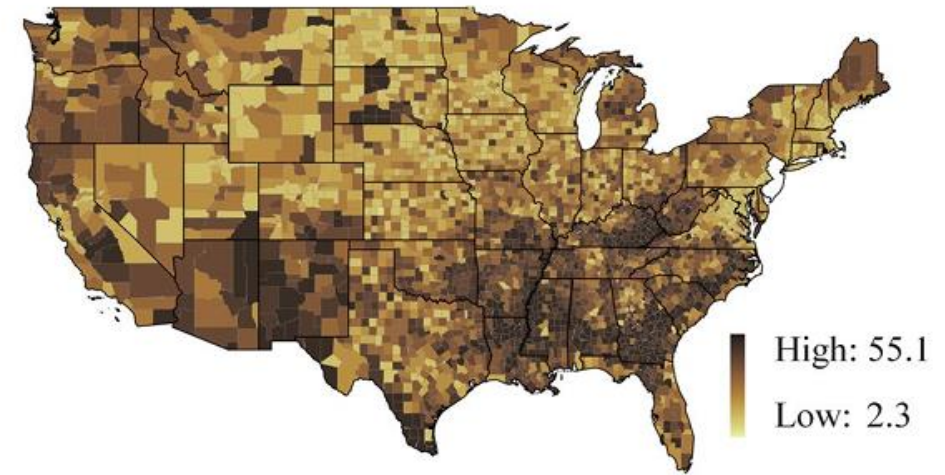
Table 1 Cross-validated prediction accuracy

	RF	GWRF
MSE	17.62	10.61
MAE	3.02	2.80
RMSE	4.20	3.20
R^2	0.82	0.89
Moran's I (residuals)	0.28 ($p < 0.05$)	0.0008 ($p > 0.05$)

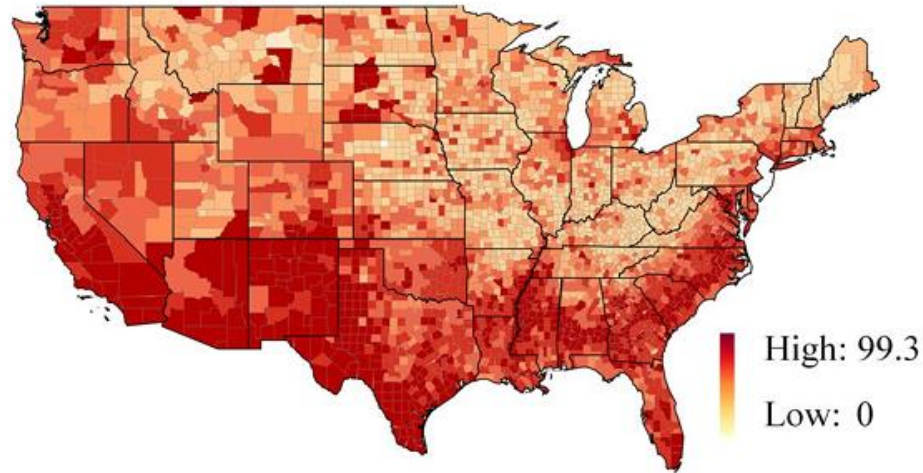
a) Asthma prevalence (%)



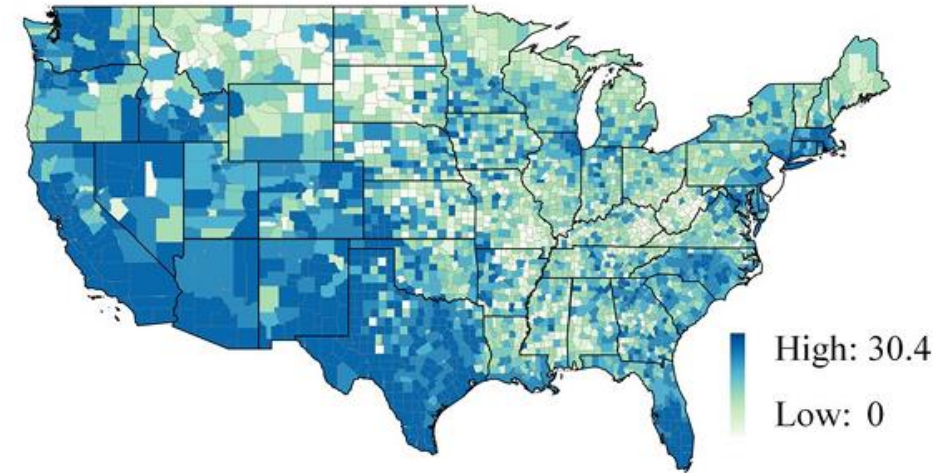
b) Poverty (%)



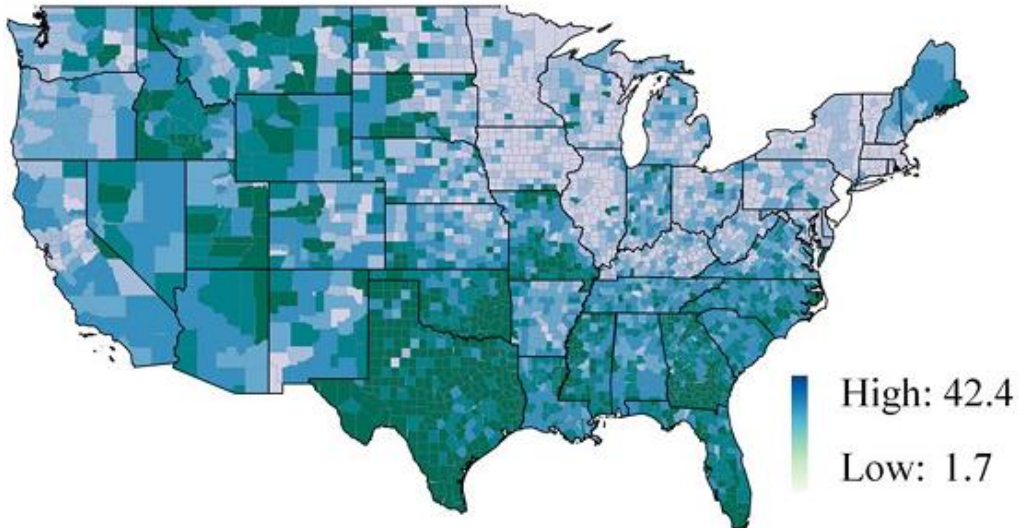
c) Minority (%)



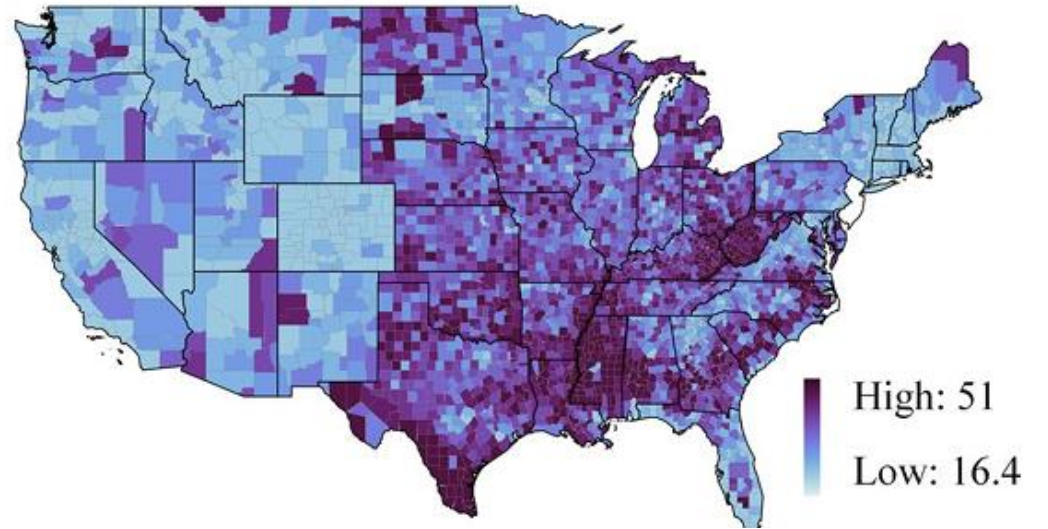
d) Limited language (%)



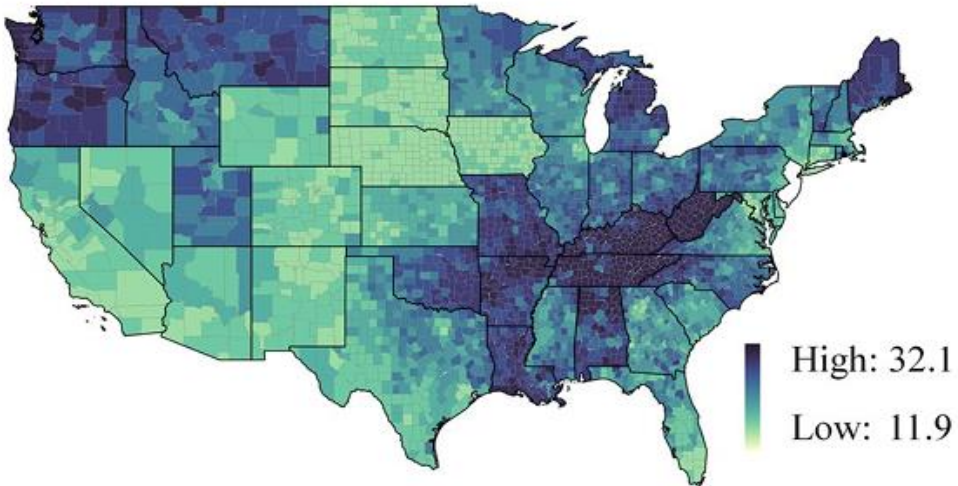
e) Uninsured (%)



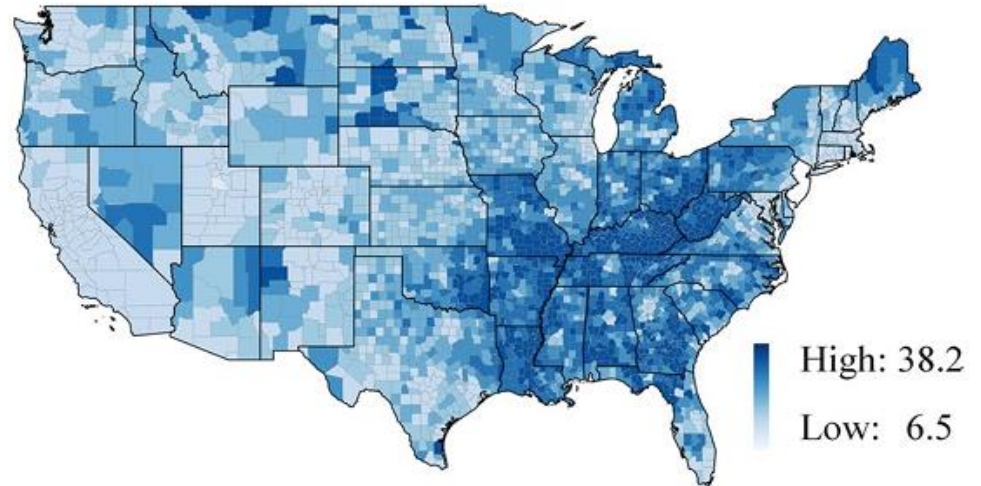
f) Obesity prevalence (%)



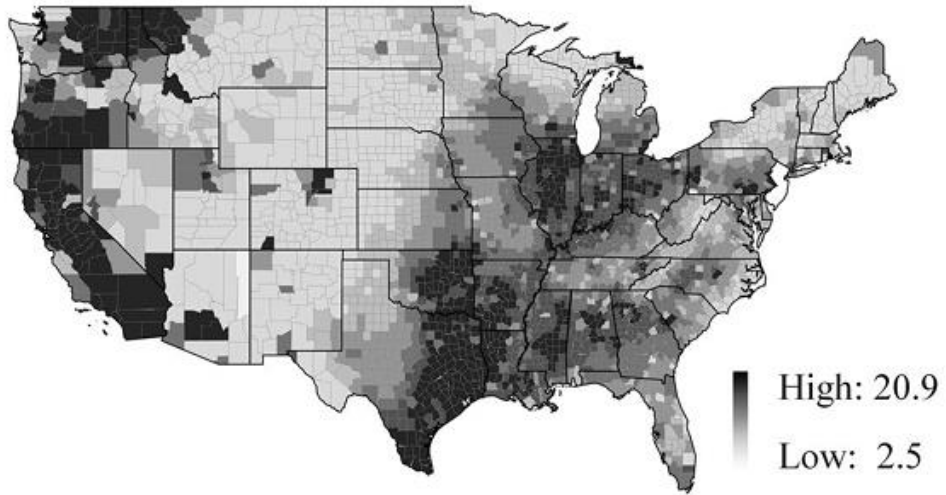
g) Depression prevalence (%)



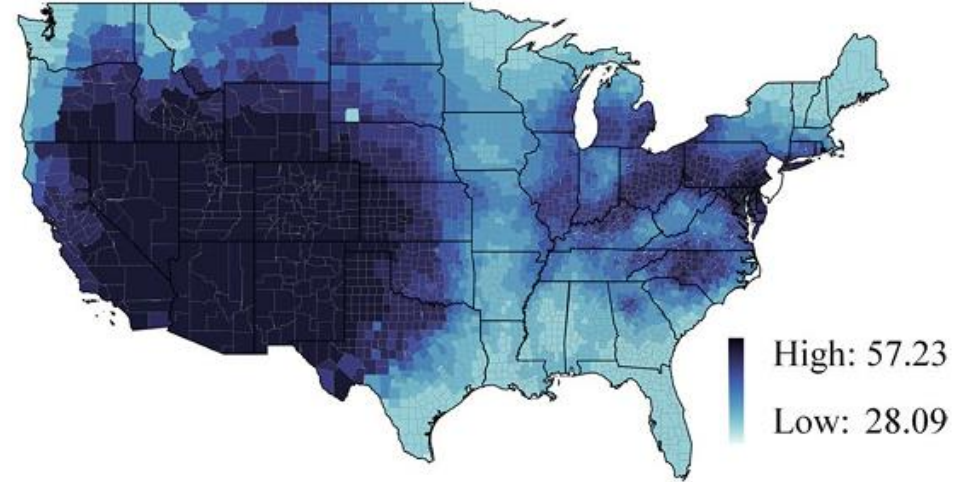
h) Smoking prevalence (%)



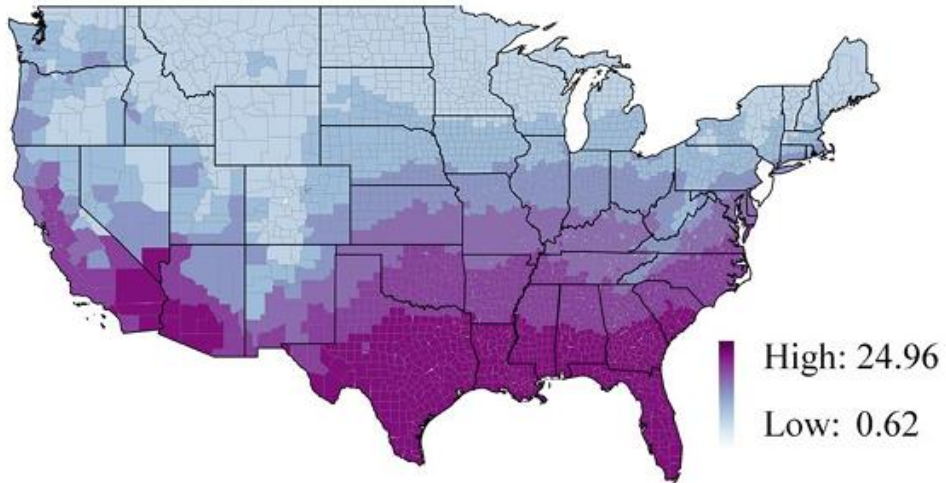
i) $PM_{2.5}$ concentration ($\mu g/m^3$)



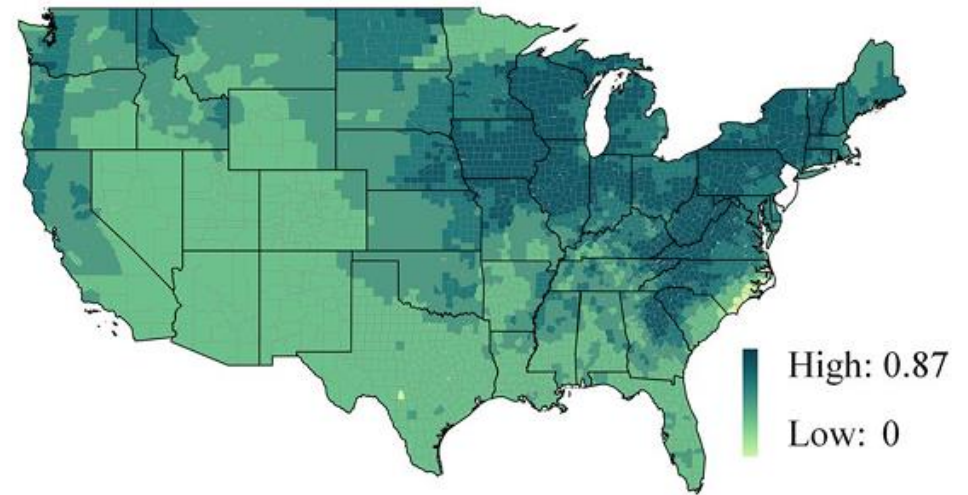
j) O_3 concentration (ppb)

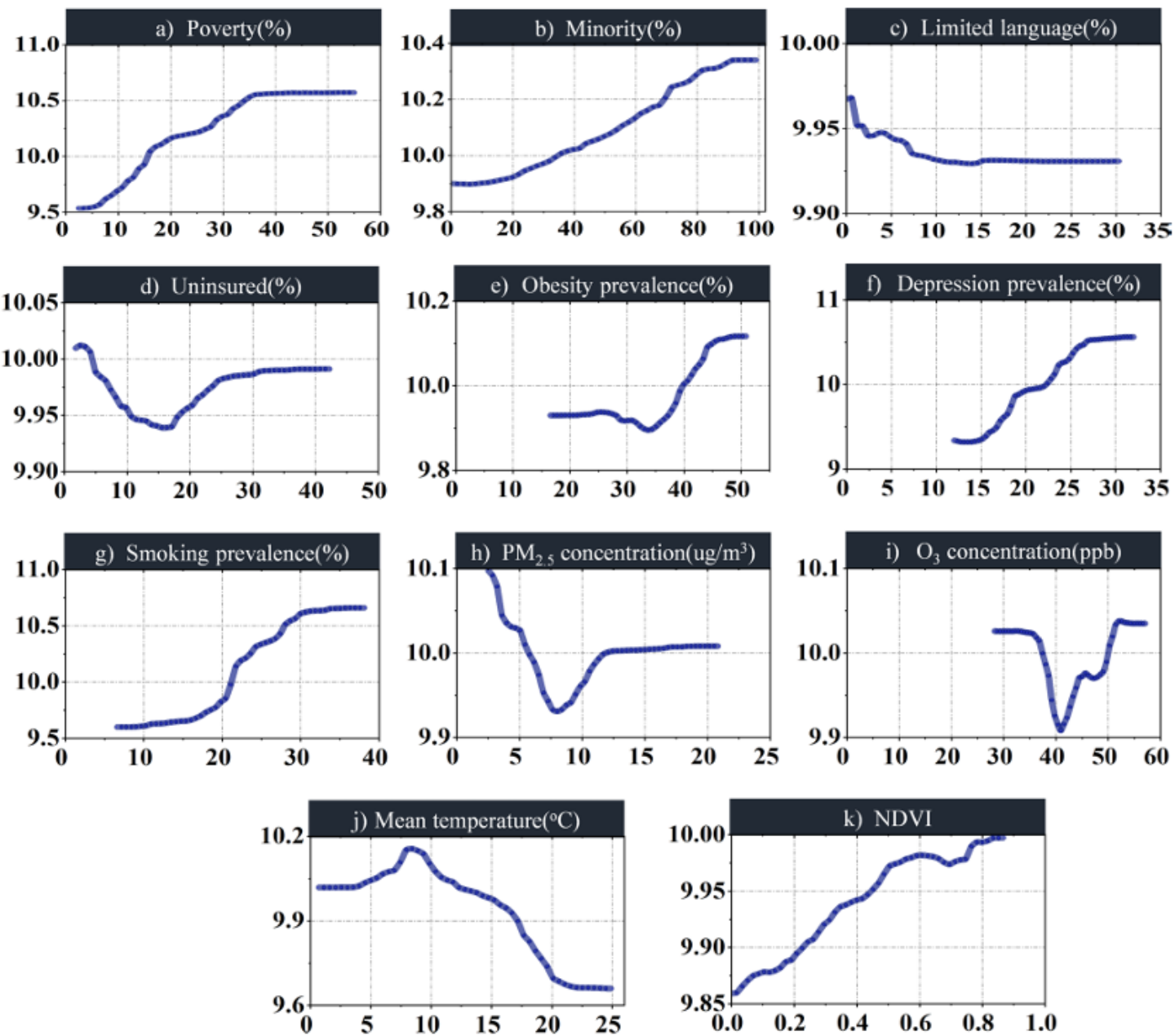


k) Mean temperature ($^{\circ}C$)



l) NDVI





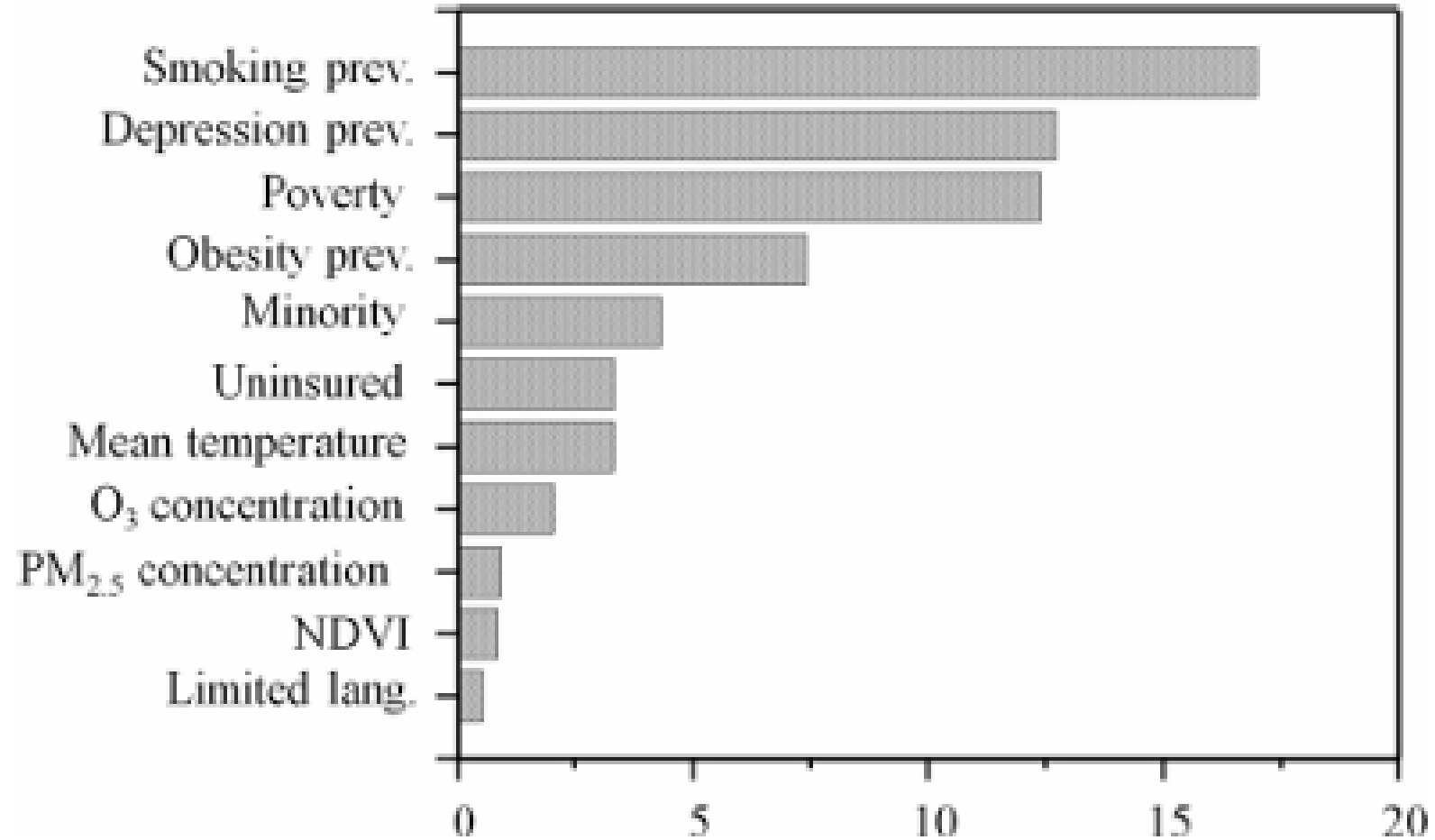
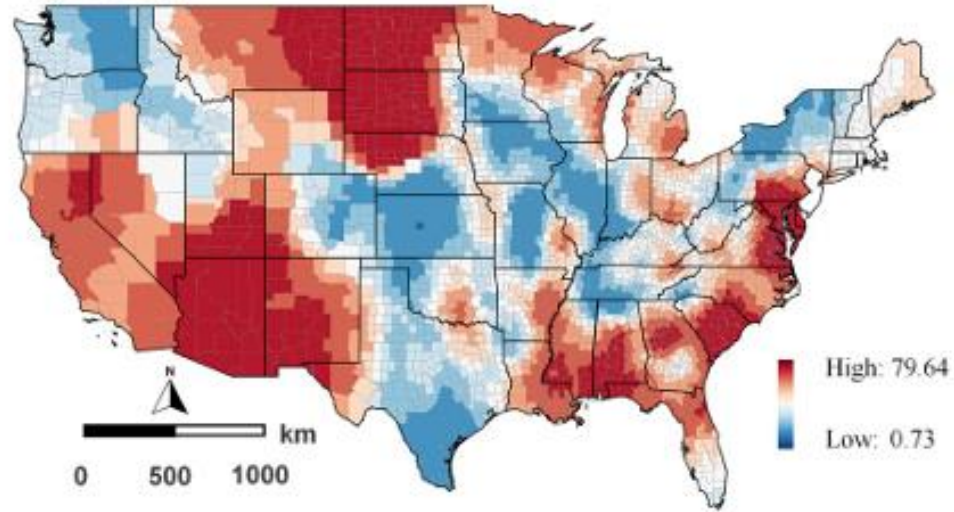
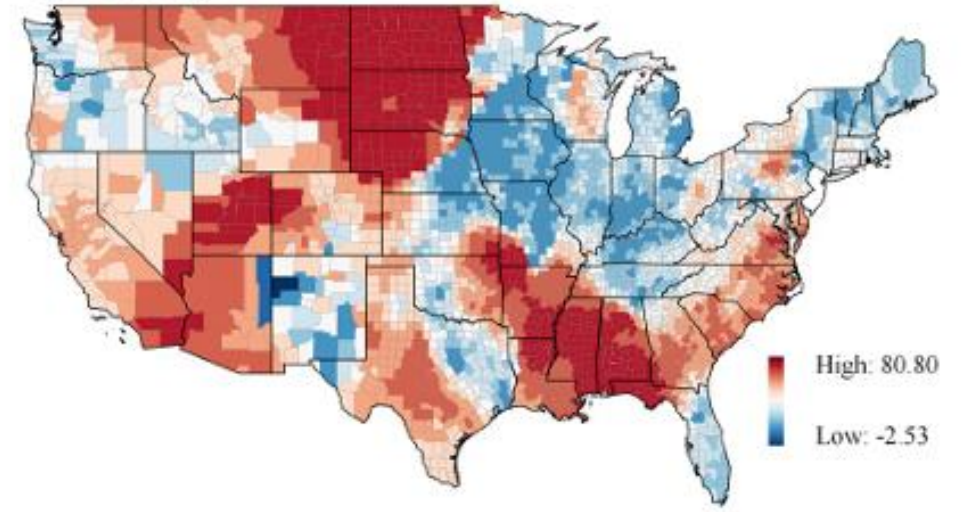


Fig. 3 Mean Variable importance based on the GWRF

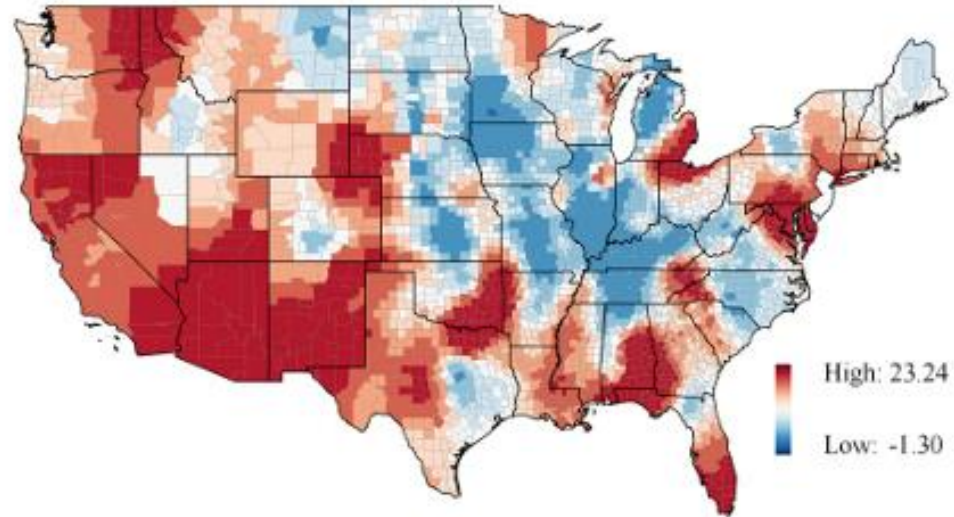
a) Poverty



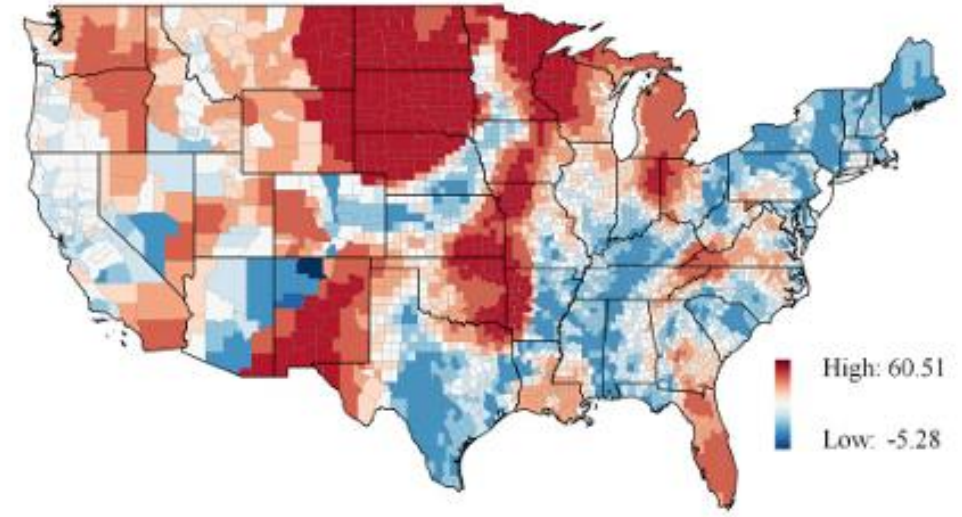
b) Minority



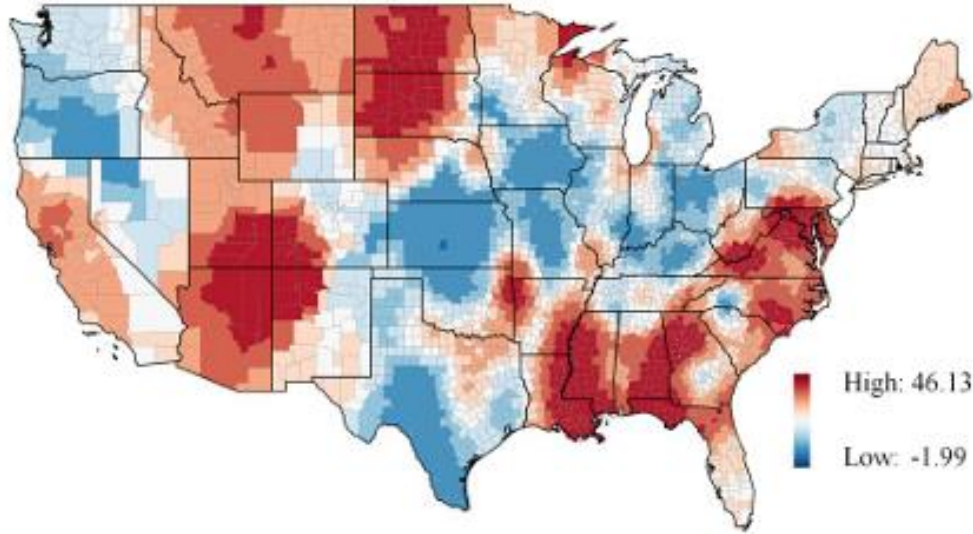
c) Limited language



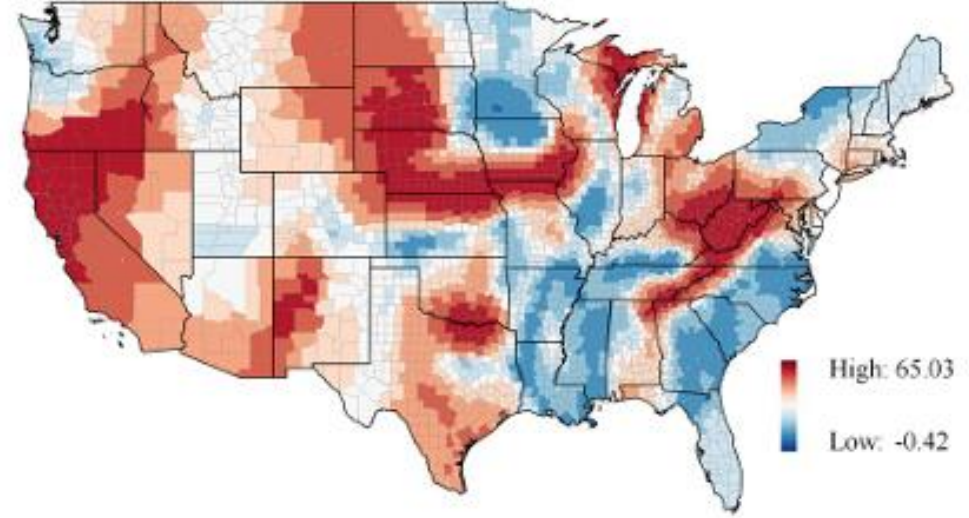
d) Uninsured



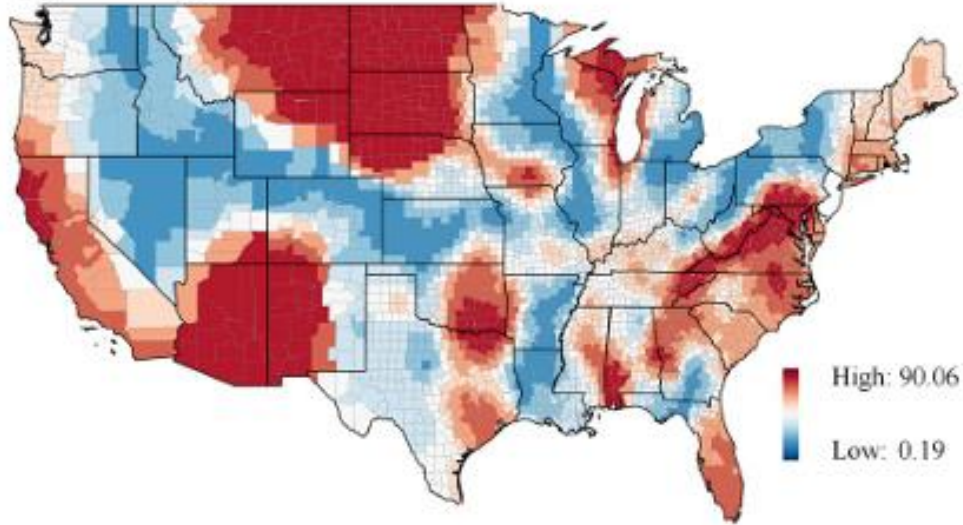
e) Obesity prevalence



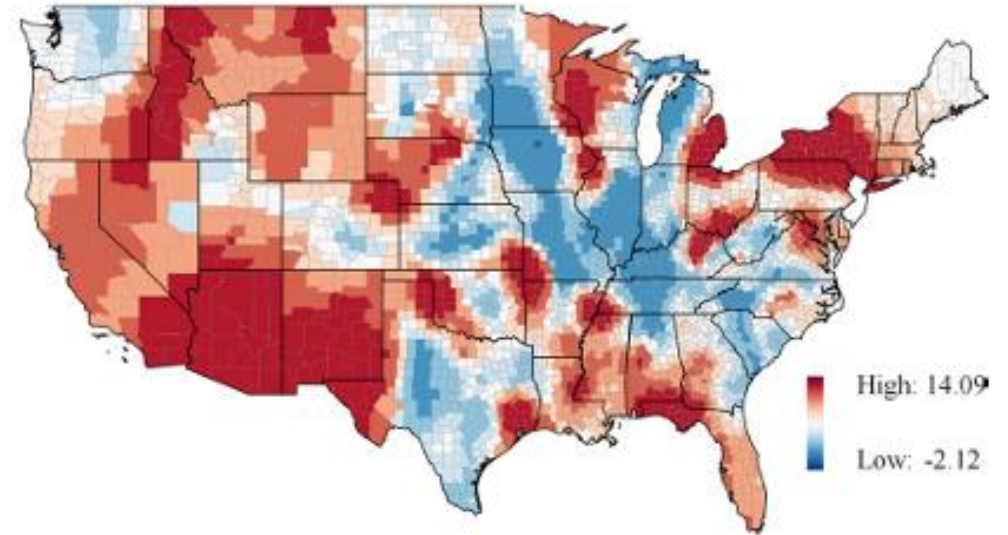
f) Depression prevalence



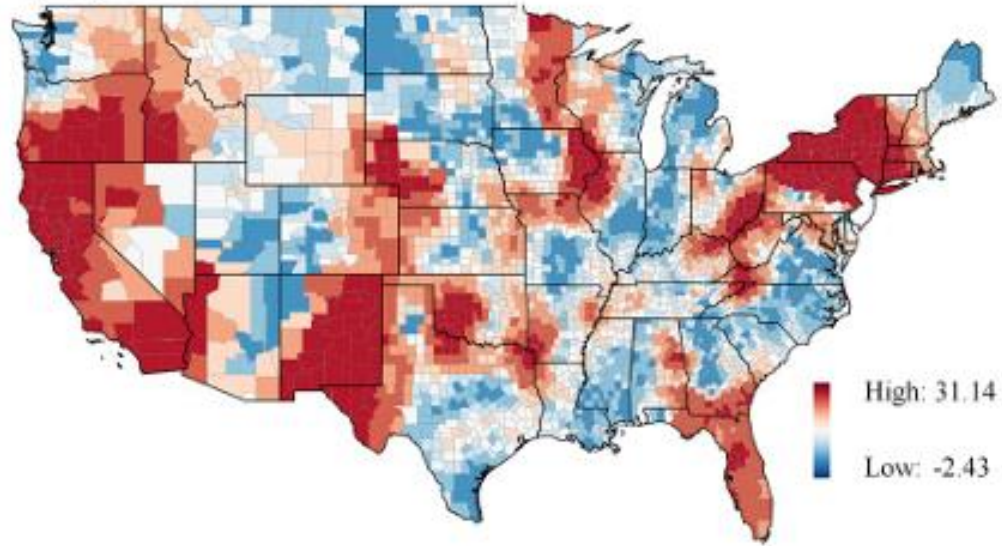
g) Smoking prevalence



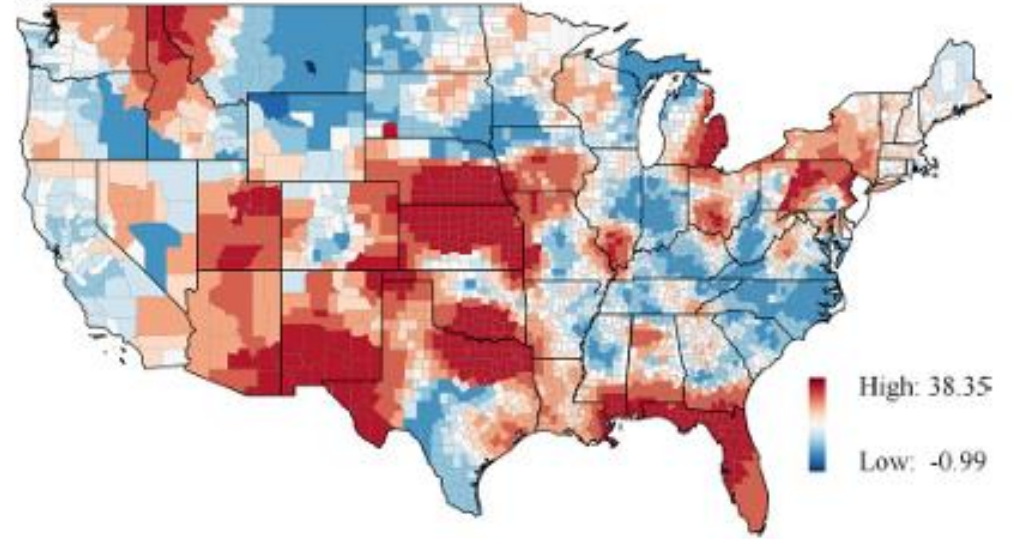
h) PM_{2.5} concentration



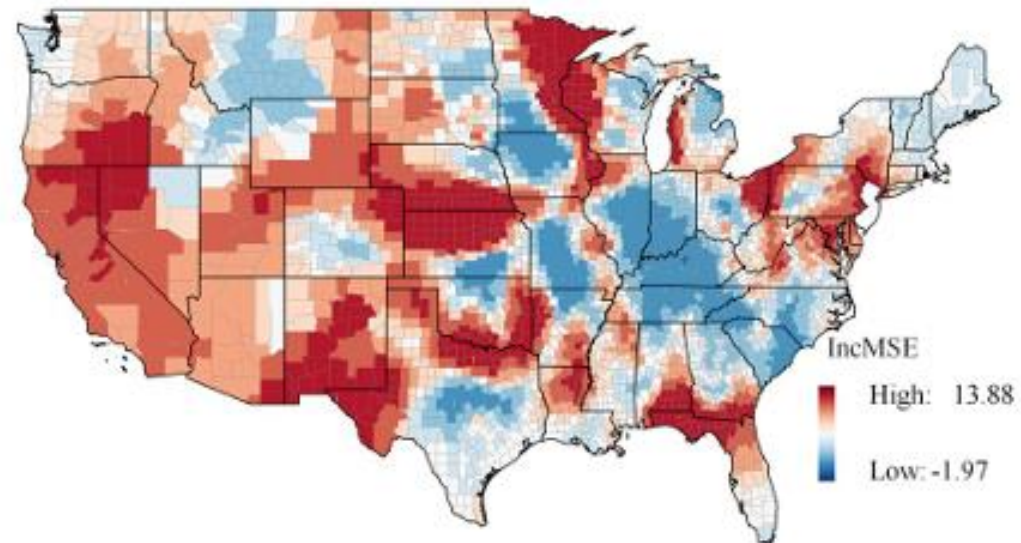
i) O₃ concentration



j) Mean temperature



k) NDVI



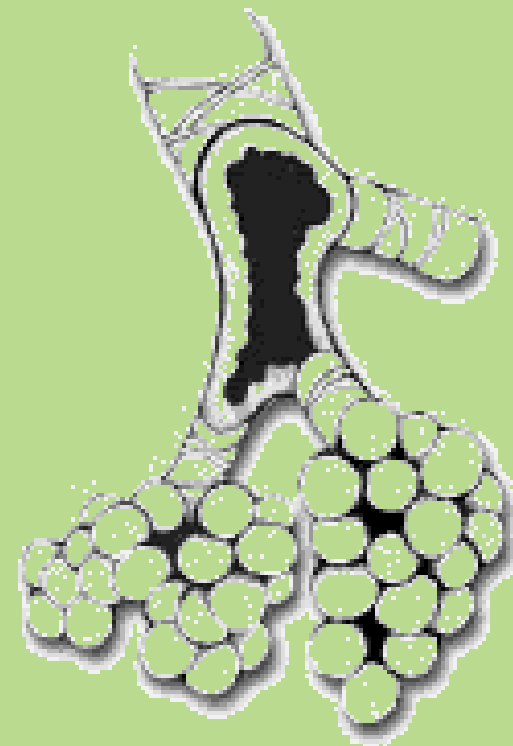
بحث و نتیجه گیری



نتیجه گیری:

این پژوهش با استفاده از مدل GWR² ثابت کرد که عوامل رفتاری و اجتماعی همچون سیگار، افسردگی و فقر، محرک‌های اصلی آسم در آمریکا هستند. نتایج نشان داد که اهمیت این ریسک فاکتورها در هر منطقه جغرافیایی متفاوت است و روابط آنها با بیماری لزوماً به صورت خطی و ساده نیست. همچنین مشخص شد شاخص‌های سلامت روان و وضعیت اقتصادی تأثیر بسیار عمیق‌تری نسبت به متغیرهای محیطی خالص مانند فضای سبز بر نرخ این بیماری دارند. در نهایت، محققان تأکید می‌کنند که سیاست‌های درمانی باید به جای یک نسخه واحد برای کل کشور، به صورت اختصاصی و متناسب با نیازهای هر منطقه طراحی گردند.

محدودیت ها:



خود اظهاری

به دلیل مقطعی و
اکولوژیک بودن داده‌ها،
امکان استنتاج دقیق
روابط علت و معلولی
وجود ندارد.

نتایج فقط در سطح
شهرستان قابل تفسیر
هستند و نمی‌توان آن‌ها
را به سطح تک‌تک
افراد تعمیم داد



نقاط قوت:



توجه به تفاوت
های
جغرافیایی

• برخلاف مطالعات قبلی، این مدل به طور صریح ناهمگنی فضایی را مورد ارزیابی قرار داده است

انعطاف
پذیری مدل

مدل GWRF به محققان اجازه داد تا روابط غیرخطی و تعاملات پیچیده بین متغیرها را به خوبی بررسی کنند.

نوآوری در روش شناسی

از اولین پژوهش‌هایی است که از یادگیری ماشین مکانی برای ارزیابی ارتباط بین بیماری، تعیین‌کننده‌های محیطی و آسیب‌پذیری اجتماعی در اپیدمیولوژی آسم استفاده کرده است.



دیدگاه من :

✓ استفاده از داده‌های ثبت‌شده در بیمارستان‌ها، پرونده‌های الکترونیک سلامت (EHR) یا نسخه‌های دارویی تجویز شده به جای خوداظهاری فردی.

✓ طراحی مطالعاتی که تغییرات محیطی (مثل کاهش آلودگی هوا یا تغییرات دما) را در طول ۱۰ سال رصد کرده و اثر آن را بر نرخ آسم در همان مناطق بسنجند.

✓ استفاده از مدل‌سازی چندسطحی که در آن داده‌های فردی (سن، ژنتیک، شغل) در سطح اول و داده‌های محیطی شهرستان (فقر، آلودگی هوا) در سطح دوم قرار می‌گیرند.



THANK YOU
FOR YOUR ATTENTION