



وزارت علوم ، تحقیقات و فناوری  
دانشگاه شهید بهشتی  
دانشکده علوم زمین  
گروه آموزشی سنجش از دور و GIS  
طرح تحقیق پایان نامه کارشناسی ارشد

عنوان : مدلسازی مکانی - زمانی خطرپذیری بروز بیماریهای حاد تنفسی در ایران

**Title: Spatial-temporal modeling of risk of acute respiratory diseases in Iran**

نوع تحقیق :



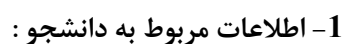
بنیادی



توسعه‌ای



کاربردی



نام : محسن      نام خانوادگی : دستاران      نام پدر : علیرضا

محل تولد : ساری شماره دانشجویی : 98421068

دوره : روزانه / شماره : سنحش از دور و سیستم اطلاعات جغرافیایی رشته :

آدرس و شماره تلفن تماس : 09368081727

**E-mail:** [Mohsen.Dastaran@gmail.com](mailto:Mohsen.Dastaran@gmail.com) & [m.dastaran@Mail.sbu.ac.ir](mailto:m.dastaran@Mail.sbu.ac.ir)

## 2- اطلاعات مربوط به استاد / استادان راهنما :

نام و نام خانوادگی : بابک میرباقری

آخرین مدرک تحصیلی: دکتری

تخصص اصلی : GIS

تعداد دانشجویان در حال راهنمایی :                      تعداد دانشجویان راهنمایی شده :

آدرس : تهران، اوین، دانشگاه شهید بهشتی، دانشکده علوم زمین، گروه سنجش از دور و GIS  
تلفن : 29903077

### 3- اطلاعات مربوط به استاد / استادان مشاور :

نام و نام خانوادگی : علیرضا شکبیا

آخرین مدرک تحصیلی: دکتری

تخصص اصلی : اقلیم‌شناسی

#### 4- جزئیات طرح پایان نامه :

##### الف ( بیان مسئله :



بیماریهای مزمن تنفسی (CRD) بیماریهای مجاری تنفسی و سایر ساختارهای ریه است. برخی از رایج ترین آنها بیماری انسدادی مزمن ریوی (COPD)، آسم، فشار خون ریوی و ... است (WHO, 2021). عوامل زیادی در به وجود آمدن این بیماریها دخیل هستند. آلودگی هوا، گرد و غبارها، مواد شیمیایی و ... در بلندمدت میتوانند باعث ابتلا به بیماریهای تنفسی شوند (Chiba & Abe, 2003).

در سالهای اخیر نیز در ایران مرگ و میر بر اثر بیماریهای تنفسی دارای افزایش قابل توجهی بوده است که بررسی عوامل اصلی بروز این نوع بیماریها مسئلهای مهمی است که نیازمند تحقیق بیشتری میباشد (Pishgar et al., 2020).

یکی از بخشهای مهم جغرافیای پزشکی، جغرافیای بیماری است که تحلیل مکانی عوامل پاتولوژیک و روابط آنها با مکان جغرافیایی را در بر میگیرد. سیستم اطلاعات جغرافیایی و فناوریهای مرتبط مانند سنجش از دور و آمار مکانی، تحلیل و مدلسازی بیماریها را ممکن میسازد (Cromley, 2003).

استفاده از هوش مصنوعی (AI) و یادگیری ماشین (ML) به عنوان یکی از زیر شاخه های هوش مصنوعی در پزشکی و به ویژه در بخش بیماریهای تنفسی، موضوعی است که به طور فزاینده ای در حال رشد است (Mekov et al., 2020).

یکی از روشهای مهم یادگیری ماشین که سابقه ای طولانی در مدلسازی بیماریها دارد رگرسیون خطی چندگانه<sup>1</sup> (MLR) می باشد. این روش یک سیستم بسیار انعطاف پذیر برای بررسی رابطه مجموعه ای از متغیرهای مستقل با یک متغیر وابسته ارائه می دهد. (Aiken et al., 2012)

همچنین یکی از مدل های معروف یادگیری ماشین رگرسیون بردار پشتیبان<sup>2</sup> (SVR) است که می تواند برای داده هایی که به صورت خطی قابل تفکیک نیستند، مورد استفاده قرار گیرد. در این روش امکان تعریف خطای قابل قبول وجود دارد تا بتوان مدل را بر اساس آن مقدار برازش داد (Sharp, 2020).

از دیگر مدل های مهم یادگیری ماشین، مدل جنگل تصادفی<sup>3</sup> (RF) است که یک الگوریتم قوی است که می تواند برای کارهای مختلفی از جمله رگرسیون و طبقه بندی استفاده شود. این روش بر پایه یادگیری ترکیبی<sup>4</sup> بنا شده است، به این معنی که یک الگوریتم جنگل تصادفی از تعداد زیادی درخت تصمیم کوچک تشکیل شده است که هر کدام پیش بینی های خود را تولید می کنند. مدل جنگل تصادفی پیش بینی های تک درخت های تصمیم را برای تولید پیش بینی دقیق تر ترکیب می کند تا بهترین نتیجه بدست آید (Thomas Wood, 2020).

در ایران برخی از مطالعاتی با موضوع تحقیق حاضر که از الگوریتم های یادگیری ماشین بهره مند بوده اند نیز مقایسه بین روش ها را نادیده گرفته و به یک روش برای مدل سازی بسنده کرده اند.

لذا در تحقیق حاضر ضمن استفاده از مدل های یادگیری ماشین به منظور مدلسازی مکانی بیماری های حاد

<sup>1</sup> Multiple Linear Regression

<sup>2</sup> Support Vector Regression

<sup>3</sup> Random Forest

<sup>4</sup> Ensemble method

تنفسی، مقایسه‌ای بین روش‌های مختلف انجام می‌گیرد تا بهترین مدل انتخاب شود.

#### ب) سؤالات تحقیق :

- از بین متغیرهای اقلیمی یا آلودگی هوا کدامیک در وقوع بیماریهای حاد تنفسی نقش موثر تری دارند؟
- کدام یک از مدل‌های یادگیری ماشین منتخب در مدلسازی بیماری‌های حاد تنفسی دقت بالاتری را نشان می‌دهند؟

#### ج) فرضیات تحقیق :

- آلودگی هوا نقش موثرتری را در وقوع بیماری‌های حاد تنفسی دارا می‌باشد.
- مدل جنگل تصادفی از دقت بالاتری در مدلسازی بیماریهای حاد تنفسی برخوردار است.

#### د) اهداف :

- مقایسه‌ی مدل‌های منتخب یادگیری ماشین به منظور تعیین بهترین مدل جهت پیش بینی بیماریهای حاد تنفسی
- تعیین متغیرهای موثر بر وقوع بیماری‌های حاد تنفسی

#### ه) بهره‌وران :

برخی از مهمترین ذینفعان از یافته‌ها و نتایج تحقیق فعلی عبارتند از:

- سازمان‌های بهداشت و درمان
- وزارت بهداشت
- محیط زیست
- دانشگاه‌های علوم پزشکی

#### \* و) جنبه جدید بودن تحقیق :

بر خلاف سایر مطالعات که اکثرا یک مدل را جهت پیش بینی وقوع بیماری مورد استفاده قرار داده اند، در این تحقیق چند مدل منتخب یادگیری ماشین با یکدیگر مقایسه خواهند شد.

### 5- روش پژوهش و مراحل انجام کار :

#### الف) روش تحقیق :

1. مطالعه منابع مربوط به موضوع مورد بررسی شامل روش کار، بیماری و متغیرها
2. جمع‌آوری و آماده سازی داده‌های بیماری و عوامل محیطی موثر بر آن
3. تحلیل داده‌ها و شناسایی الگوهای مکانی

4. اجرای مدل های یادگیری ماشین منتخب بر اساس پیشینه تحقیقاتی
5. ارزیابی دقت مدل های یادگیری ماشین و انتخاب بهترین مدل
6. تعیین میزان تاثیرگذاری متغیرهای مورد استفاده در تحقیق بر اساس مدل منتخب
7. تفسیر نتایج و ارائه گزارش نهایی

ب ( تکنیک کار / ابزار گردآوری اطلاعات :

1. استخراج داده های آلودگی هوا با استفاده از سامانه ی Google Earth Engine
2. استفاده از نرم افزار ArcGIS جهت تولید نقشه های خودهمبستگی مکانی و نقشه های خروجی حاصل از اجرای مدل ها
3. استفاده از زبان برنامه نویسی پایتون و نرم افزار R به منظور اجرای مدل های یادگیری ماشین
4. بکارگیری مدل های یادگیری ماشین نظیر ، SVR ، MLR و Random Forest به منظور تعیین بهترین مدل و ارزیابی میزان تاثیر گذاری عوامل مختلف

\* ج ( نمونه برداری یا نمونه گیری :

د ( روش تحلیل اطلاعات و داده ها :

در این تحقیق عملکرد مدل های یادگیری ماشین با استفاده از شاخص های آماری  $R^2$  ، MAE و RMSE ارزیابی خواهد شد.

ه ( جدول زمانی مراحل اجرا و پیشرفت انجام کار :

زمان لازم به ماه														فعالیت های هر مرحله
14	13	12	11	10	9	8	7	6	5	4	3	2	1	
											*	*	*	مرور منابع مرتبط با موضوع
										*	*	*		جمع آوری و آماده سازی داده ها
								*	*	*	*			تحلیل داده ها و شناسایی الگوهای مکانی
					*	*	*	*						اجرای مدل های یادگیری ماشین
			*	*	*									ارزیابی و مقایسه ی دقت مدل ها
		*	*											تفسیر نتایج
	*	*												تهیه نقشه
*	*	*	*											جمع بندی نتایج و نوشتن پایان نامه

6- پیشینه تحقیق :

پیشگام و همکاران (2020) با استفاده از داده های بیماری های تنفسی حاد در تهران به تحلیل آنها

پرداختند. در این تحقیق از روش‌های Global Moran و Anselin Local Moran's I برای خودهمبستگی مکانی استفاده شده است. نتایج این تحلیل روند افزایشی بیماران را نشان می‌دهد که الگوی توزیع مکانی دارند. Zhi-Yong و همکاران (2020) به مطالعه و ارزیابی تأثیر آلاینده‌های هوا بر بیماری انسدادی مزمن ریوی در چین پرداخته‌اند. نتایج این تحقیق نشان‌دهنده‌ی این است که پارامترهای  $PM_{2.5}$  و  $SO_2$  از مهمترین عوامل تأثیرگذار بر شیوع این بیماری در منطقه‌ی مورد مطالعه است. در این مطالعه از مدل MLR به منظور پیش‌بینی و کشف ارتباطات بین پارامترها استفاده شده است. به منظور ارزیابی مدل نیز از خطای MAPE استفاده شده است.

Fonseca و همکاران (2014) در مقاله‌ای به برآورد شیوع بیماری سیستم‌های ریویز در ایالت میناس گرایس برزیل پرداخته‌اند. آنها توانستند با دو مدل MLR و رگرسیون مکانی نقشه ریسک ایالت مورد مطالعه را تهیه کنند و نشان دهند که متغیرهای پوشش گیاهی، دما، بارش، توپوگرافی، بهداشت و شاخص‌های توسعه انسانی در نحوه شیوع بیماری مهم بودند.

Welsing و همکاران (2001) در مطالعه‌ای به بررسی رابطه بین رشد بیماری آرتریت روماتوئید و تخریب مفصل با میزان عملکرد بیماران بعد از بهبودی با استفاده از مدل MLR پرداختند. در نهایت مشخص شده است که تأثیر رشد بیماری و تخریب مفصل بر توانایی عملکرد در طول بیماری تغییر می‌کند. در ابتدا رشد و فعالیت بیماری تأثیر بیشتری دارد اما در انتها تخریب مفصل اثرگذاری بالاتری از خود نشان می‌دهد.

Nilashi و همکاران (2020) به منظور تشخیص به موقع بیماری پارکینسون به مشاهده‌ی پیشرفت این بیماری و پیش‌بینی آن پرداختند. به منظور مدلسازی و پیش‌بینی بیماری از نه روش مختلف استفاده شد که در نهایت ترکیبی از دو روش SVR و شبکه باور عمیق توانست پیش‌بینی بهتری از سایر تکنیک‌های یادگیری ماشین انجام دهد.

Zhou و همکاران (2019) در تحقیقی به بررسی ارتباط بین قرار گرفتن در معرض آلاینده‌ها و پذیرش روزانه کودکان مبتلا به بیماری‌های تنفسی پرداخته‌اند. پارامترهای استفاده شده شامل  $PM_{2.5}$ ،  $PM_{10}$ ،  $NO_2$  و  $SO_2$  می‌باشند. نتایج مدل SVR نشان می‌دهد که آلودگی هوا بخصوص پارامتر  $PM_{10}$  به طور قابل توجهی با تعداد پذیرش در بیمارستان، مدت بستری و هزینه اقتصادی بستری در کودکان مبتلا به بیماری‌های تنفسی ارتباط دارد.

Liu و همکاران (2019) به منظور هشدار به مردم در مواقع حساس و کنترل آلودگی هوا اقدام به پیش‌بینی کیفیت هوا کردند. آنها برای این کار از مدل‌های SVR و رگرسیون جنگل تصادفی (RFR) و به منظور ارزیابی دقت مدل‌ها از  $RMSE$  و  $R^2$  استفاده کردند. نتایج این تحقیق نشان داد که مدل مبتنی بر SVR در پیش‌بینی شاخص کیفیت هوا عملکرد بهتری داشته و مدل مبتنی بر RFR در پیش‌بینی غلظت NOX بهتر عمل کرده است.

Ong و همکاران (2018) با استفاده از مدل RF به تولید نقشه خطر توزیع بیماری تب‌دنگی در سنگاپور پرداخته‌اند. مدل RF با استفاده از داده‌های بیماری، محیط، جمعیت و حشره‌شناسی نقشه‌ی خطرپذیری بیماری را در چهار گروه مشخص کرده است و محققان را به این نتیجه رسانده است که مدل RF به دلیل پیش‌بینی قوی آن از پتانسیل بالایی جهت تولید نقشه خطرپذیری انتقال تب‌دنگی در سنگاپور برخوردار است.

رضوی و همکاران (2021) مدل سازی مناطق مستعد آسم را با استفاده از مدل جنگل تصادفی انجام داده‌اند. منطقه مورد مطالعه در این تحقیق شهر تهران می‌باشد. نتایج همبستگی مکانی و مدل RF نشان داده است که معیارهای فاصله از پارک‌ها و خیابان‌ها و همچنین آلاینده‌های  $PM_{2.5}$  و  $PM_{10}$  بیشترین تأثیر را در

7- منابع و مآخذ :

- Aiken, L. S., West, S. G., Pitts, S. C., Baraldi, A. N., & Wurpts, I. C. (2012). Multiple Linear Regression. In *Handbook of Psychology, Second Edition*. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118133880.hop202018>
- Chiba, H., & Abe, S. (2003). [The environmental risk factors for COPD--tobacco smoke, air pollution, chemicals]. *Nihon rinsho. Japanese journal of clinical medicine*, 61(12), 2101–2106. <http://europepmc.org/abstract/MED/14674317>
- Cromley, E. K. (2003). GIS and disease. *Annual Review of Public Health*, 24, 7–24. <https://doi.org/10.1146/annurev.publhealth.24.012902.141019>
- Great Learning Team. (2020). *Support Vector Regression in Machine Learning / What is SVM?* <https://www.mygreatlearning.com/blog/support-vector-regression/>
- Liu, H., Li, Q., Yu, D., & Gu, Y. (2019). Air quality index and air pollutant concentration prediction based on machine learning algorithms. *Applied Sciences (Switzerland)*, 9(19). <https://doi.org/10.3390/app9194069>
- Mekov, E., Miravittles, M., & Petkov, R. (2020). Artificial intelligence and machine learning in respiratory medicine. *Expert Review of Respiratory Medicine*, 14(6), 559–564. <https://doi.org/10.1080/17476348.2020.1743181>
- Nilashi, M., Ahmadi, H., Sheikhtaheri, A., Naemi, R., Alotaibi, R., Abdulsalam Alarood, A., Munshi, A., Rashid, T. A., & Zhao, J. (2020). Remote tracking of Parkinson's Disease progression using ensembles of Deep Belief Network and Self-Organizing Map. *Expert Systems with Applications*, 159, 113562. <https://doi.org/10.1016/j.eswa.2020.113562>
- Ong, J., Liu, X., Rajarethinam, J., Kok, S. Y., Liang, S., Tang, C. S., Cook, A. R., Ng, L. C., & Yap, G. (2018). Mapping dengue risk in Singapore using Random Forest. *PLoS Neglected Tropical Diseases*, 12(6), e0006587. <https://doi.org/10.1371/journal.pntd.0006587>
- Pishgar, E., Mohammadi, A., Bagheri, N., & Kiani, B. (2020). A spatio-temporal geodatabase of mortalities due to respiratory tract diseases in Tehran, Iran between 2008 and 2018: A data note. *BMC Research Notes*, 13(1), 1–12. <https://doi.org/10.1186/s13104-020-05319-4>
- Razavi-Termeh, S. V., Sadeghi-Niaraki, A., & Choi, S. M. (2021). Asthma-prone areas modeling using a machine learning model. *Scientific Reports*, 11(1), 1–16. <https://doi.org/10.1038/s41598-021-81147-1>
- Sharp, T. (2020). *An Introduction to Support Vector Regression (SVR)*. Towards Data Science. <https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2>
- Thomas Wood. (2020). *Random Forests Definition / DeepAI*. <https://deepai.org/machine-learning-glossary-and-terms/random-forest>
- Welsing, P. M. J., Van Gestel, A. M., Swinkels, H. L., Kiemeny, L. A. L. M., & Van Riel, P. L. C. M. (2001). The Relationship Between Disease Activity, Joint Destruction, and Functional Capacity Over the Course of Rheumatoid Arthritis. In *ARTHRITIS & RHEUMATISM* (Vol. 44, Issue 9). <https://doi.org/10.1002/1529-0131>
- WHO. (2021). *Chronic respiratory diseases*. [https://www.who.int/health-topics/chronic-respiratory-diseases#tab=tab\\_1](https://www.who.int/health-topics/chronic-respiratory-diseases#tab=tab_1)

Zhou, H., Wang, T., Zhou, F., Liu, Y., Zhao, W., Wang, X., Chen, H., & Cui, Y. (2019). Ambient Air Pollution and Daily Hospital Admissions for Respiratory Disease in Children in Guiyang, China. *Frontiers in Pediatrics*, 7, 400. <https://doi.org/10.3389/fped.2019.00400>

بله ☐ خیر ☒ - آیا برای طرح از سازمانهای دیگر تأمین اعتبار شده است ؟

در صورت مثبت بودن نام سازمان و میزان اعتبار را مشخص نمایید .

\* موارد ستاره دار بنا به ضرورت تکمیل گردد .

- دانشجو موظف است ابتدا راهنمای تکمیل فرم تحقیق پایان نامه کارشناسی ارشد را که در سایت دانشکده علوم زمین قرار دارد ، مطالعه و سپس اقدام به تکمیل آن نماید .

- تعهد می نماید محتویات پایان نامه نتیجه تحقیق ، مطالعات و استنتاجات اینجانب باشد .

نام و نام خانوادگی دانشجو : محسن دستاران تاریخ : / / 1400  
امضاء

- مراتب مورد تأیید اینجانبان استاد / استادان راهنما و استاد / استادان مشاور می باشد .

نام و نام خانوادگی استاد / استادان راهنما : تاریخ : / / 1400  
بابک میرباقری  
علیرضا شکیبا  
امضاء

نام و نام خانوادگی استاد / استادان مشاور : تاریخ : / / 1400  
امضاء