

Evaluation and Analysis of the Evolution and Completeness of Building Data in Volunteered Geographic Information (Case Study: OpenStreetMap Building Information in Mashhad City)

Hadiyeh, Hassan nia Badrabad^۱, Babak, Mirbagheri^۱, Masoud Minaei^{۲*}

۱. Center for Remote Sensing and GIS Research, Faculty of Earth Sciences, Shahid Beheshti University, Tehran, Iran

۲. Dep of Geography, Ferdowsi University, Mashhad, Iran. GISSRS: Lab, Mashhad, Iran

* Corresponding Author: m.minaei@um.ac.ir

Abstract

Introduction and Objectives: The advent of digital sensors, smartphones, crowdsourced applications, and vibrant social media platforms has dramatically transformed the landscape of geographic data accessibility. This transformation is even more pronounced with the emergence of user-generated content, a hallmark of Web ۲.۰ technologies. This participatory phenomenon is termed Volunteered Geographic Information (VGI), of which OpenStreetMap (OSM) serves as a prominent illustration. OSM offers an extensive repository of detailed and frequently updated geospatial data concerning physical infrastructures, including buildings, roads, and other vital landmarks. Though OSM provides this wealth of data free of charge and continues to expand its database, the quality of the information necessitates meticulous evaluation to determine its suitability for various practical applications. Research across the globe has underscored the efficacy of VGI; however, there remains a significant gap in studies investigating the quality of spatial information pertaining specifically to buildings within OSM, particularly in regions like Iran. This study aims to address this gap by focusing on three critical aspects: first, it will evaluate the quality of building data captured in OSM; second, it will analyze the enhancements and overall completeness of this data from the years ۲۰۱۸ to ۲۰۲۳; and third, it will explore the intricate relationship between land use patterns and population density, utilizing the building data derived from OSM. Through this comprehensive examination, the study hopes to shed light on the integrity of VGI in representing urban landscapes.

Materials and Methods: The study area, encompassing districts ۹ and ۱۱ of Mashhad, was divided into a grid of ۱ x ۱ km cells, serving as the foundational framework for our analysis. Within each cell, we assessed three key metrics from ۲۰۱۸ to ۲۰۲۳: building density, completeness, and the quantity of buildings documented in OpenStreetMap (OSM) data. Notably, the ratio of officially registered charges to the overall count of charges within the voluntary dataset indicates the extent to which official data is represented in the voluntary dataset. This metric reflects the completeness of the voluntary data relative to official sources. Subsequently, we analyzed the correlation between building density and data completeness for each grid

cell. Additionally, demographic parameters, including population and land use, were incorporated into the evaluation process.

Results and Discussion: The construction data sourced from OpenStreetMap (OSM) across the 9th and 11th districts of Mashhad has witnessed an impressive twentyfold increase from ۲۰۱۸ to ۲۰۲۲. This extraordinary expansion is particularly evident in the eastern and northeastern sectors of the city, where development has surged. Among the various land use categories identified, residential areas, followed by residential-commercial zones, commercial spaces, and educational institutions, display the highest frequency of construction activities in that specific order. Moreover, there exists a notable and robust correlation between the OSM building data and the population residing within each building block. However, this connection has exhibited significant fluctuations throughout the analyzed period, indicating varying patterns of growth and development. Furthermore, an intriguing relationship emerges between the density index and the completeness of the data; in regions where the density of OSM building data is greater, users tend to exhibit a heightened willingness to create and delineate new buildings, reflecting an active engagement in urban development.

Conclusion: An in-depth examination of completeness index values reveals a noteworthy discrepancy in the OpenStreetMap (OSM) dataset, which significantly underrepresents the actual number of buildings when compared to official records. This research compellingly illustrates that, although OSM data can serve as a valuable resource for geographical analyses in urban environments—especially within bustling metropolises such as Mashhad—there exists a pressing necessity for enhancements and validation of the data to attain a higher level of accuracy. Moreover, a meticulous evaluation of key performance indicators can provide profound insights that not only enhance the reliability of the OSM dataset but also bolster its applications in urban planning and other related fields, thereby paving the way for more informed decision-making processes.

Keywords: VGI, OpenStreetMap, Building, Completeness, Mashhad

ارزیابی و تحلیلی بر ارتقاء و کامل بودن داده‌های ساختمانی در اطلاعات جغرافیایی داوطلبانه (مطالعه موردی):

اطلاعات ساختمانی OpenStreetMap در شهر مشهد

هدیه حسن نیا بدرآباد^۱، بابک میرباقری^۱، مسعود مینائی^{۲*}

۱. مرکز مطالعات سنجش از دور و GIS، دانشکده علوم زمین، دانشگاه شهید بهشتی، تهران، ایران

۲. گروه جغرافیا، دانشگاه فردوسی مشهد، مشهد، ایران. آزمایشگاه علم / سیستم اطلاعات جغرافیایی و سنجش از دور، مشهد، ایران

*نویسنده عهده دار مکاتبات: m.minaei@um.ac.ir

چکیده

مقدمه و هدف: افزایش فزاینده حسگرهای دیجیتالی، تلفن‌های هوشمند، برنامه‌های کاربردی جمع‌سپاری و رسانه‌های اجتماعی موجب افزایش سریع حجم داده‌های جغرافیایی شده‌است. در این راستا، منبعی نوین و قابل توجه از اطلاعات جغرافیایی به شکل محتوای وب توسط کاربران در دسترس قرار گرفته‌است. این منبع عمدتاً توسط فناوری‌هایی تحت عنوان Web ۲.۰ پشتیبانی می‌شود. در چهارچوب جغرافیای نوین، افراد داده‌های مربوط به محیط خود را جمع‌آوری و آپلود می‌کنند و از این طریق به بهبود داده‌های موجود ارائه شده توسط دیگران کمک می‌کنند. این روش نوین با عناوینی همچون اطلاعات مکانی مردم‌گستر (VGI) Volunteered Geographic Information شناخته می‌شود. یکی از منابع VGI، داده‌های موجود در وب‌سایت OpenStreetMap (OSM) است که انواع گوناگونی از داده‌های مکانی را شامل می‌شود. استفاده از داده‌های OSM مزیت‌های بسیاری دارد زیرا این داده‌ها به صورت رایگان در دسترس هستند و به سرعت به روز می‌شوند. پایگاه داده OSM هم از نظر تعداد کاربران و هم از نظر حجم داده‌های مکانی به سرعت در حال توسعه است. از این رو، به یک مجموعه داده جهانی، پرکاربرد و با کیفیت بالا تبدیل شده‌است. علیرغم اینکه دسترسی به چنین داده‌هایی جذاب به نظر می‌رسد اما کیفیت آنها نیاز به ارزیابی دارد، بنابراین تناسب آنها برای اهداف خاص قابل ارزیابی است. علیرغم مطالعات انجام شده در سطح جهان و مزایای بهره‌برداری از اطلاعات مکانی مردم‌گستر (VGI) در ایران، تحقیقات بسیار کمی در ارتباط با اطلاعات مکانی مرتبط با ساختمان‌ها در OSM انجام شده‌است. از این رو در پژوهش حاضر (۱) به ارزیابی کیفیت داده‌های ساختمانی OSM پرداخته می‌شود؛ (۲) روند ارتقاء و کامل بودن ساختمان‌ها در داده‌های OSM در یک دوره زمانی ۶ ساله (۲۰۱۸-۲۰۲۳) مورد بررسی قرار می‌گیرد و (۳) ارتباط بین کاربری اراضی و جمعیت با داده‌های ساختمانی OSM ارزیابی می‌شود.

مواد و روش‌ها: منطقه مورد مطالعه (مناطق ۹ و ۱۱ مشهد) به یک شبکه‌ی سلولی به اندازه 1×1 کیلومتر تقسیم شد و این مساحت به عنوان واحد جغرافیایی مبنای محاسبات قرار گرفت. برای هر واحد جغرافیایی، سه معیار (تراکم ساختمانی، کامل بودن و تعداد ساختمان در داده‌های OSM) در منطقه مورد مطالعه برای سال‌های ۲۰۱۸ تا ۲۰۲۳ بررسی شد. لازم به ذکر است، تعداد عوارض رسمی متناظر در مجموعه داده داوطلبانه به تعداد کل عوارض مجموعه داده رسمی، میزان حضور داده‌های رسمی را در مجموعه داده داوطلبانه نشان می‌دهد. این مقدار همان میزان کامل بودن داده‌های داوطلبانه نسبت به داده‌های رسمی است. در گام بعدی، رابطه بین تراکم و کامل بودن داده‌های ساختمانی برای هر سلول شبکه مورد محاسبه قرار گرفت. همچنین دو پارامتر جمعیت و کاربری اراضی در بررسی‌ها مورد استفاده قرار گرفتند.

نتایج و بحث: داده ساختمانی OSM در مناطق ۹ و ۱۱ مشهد از ۲۰۱۸ تا ۲۰۲۳، ۲۰ برابر افزایش پیدا کرده است. به ویژه برای مناطق شرق و شمال شرقی این افزایش مشهودتر است. همچنین کاربری اراضی مسکونی، مسکونی-تجاری، تجاری و آموزشی به ترتیب، بیشترین فراوانی را به خود اختصاص داده اند. علیرغم همبستگی مثبت و معنادار بین داده های ساختمانی OSM و جمعیت هر بلوک ساختمانی، این همبستگی به صورت نوسانی در بازه زمانی مورد بررسی متغیر بوده است. مقدار همبستگی ها میان شاخص تراکم و کامل بودن نشان می دهد در مکان هایی که تراکم داده ساختمانی در داده OSM بالاتر بوده است، کاربران تمایل بیشتری برای ایجاد و ترسیم ساختمان جدید داشته اند.

نتیجه گیری: با توجه به مقادیر شاخص کامل بودن می توان نتیجه گرفت که در مقایسه با داده های رسمی، ساختمان های به ثبت رسیده در OSM به طور قابل توجهی از تعداد واقعی ساختمان ها در داده رسمی کمتر هستند. با توجه به این تحقیق، می توان گفت که داده های OSM برای تحلیل جغرافیایی ساختمان ها در شهرهای بزرگ مانند مشهد مناسب هستند، اما برای استفاده های بیشتر و دقیق تر، نیاز به تکمیل و بهبود این داده ها وجود دارد. همچنین بررسی شاخص های تاثیرگذار می تواند اطلاعات سودبخشی در جهت استفاده مطمئن تر از این داده ها در اختیار کاربران قرار دهد.

کلیدواژه ها: VGI، OpenStreetMap، ساختمان، کامل بودن، مشهد

۱- مقدمه

افزایش فزاینده حسگرهای دیجیتال، تلفن های هوشمند، برنامه های کاربردی جمع سپاری^۱ و رسانه های اجتماعی موجب افزایش سریع حجم داده های جغرافیایی شده است (Mobasheri ۲۰۱۷). در این راستا، منبعی نوین و قابل توجه از اطلاعات جغرافیایی به شکل محتوای وب توسط کاربران در دسترس قرار گرفته است. این منبع عمدتاً توسط فناوری هایی تحت عنوان Web ۲.۰ پشتیبانی می شود (Elwood et al. ۲۰۱۲). در چهارچوب جغرافیای نوین^۲، افراد داده های مربوط به محیط خود را جمع آوری و آپلود می کنند و از این طریق به بهبود داده های موجود ارائه شده توسط دیگران کمک می کنند. این روش نوین با عناوینی همچون اطلاعات جغرافیایی داوطلبانه (VGI)^۳ و داده های جغرافیایی جمع سپاری (CSGD)^۴ شناخته می شود (Minaei, ۲۰۲۰). بنابراین داده های جغرافیایی داوطلبانه آنلاین و رایگان با رویکردی مبتنی بر جمع سپاری پدید آمده اند. پدیده جمع سپاری به طور کلی و اطلاعات جغرافیایی

^۱ Crowdsourcing

^۲ Neogeography

^۳ Volunteered Geographic Information (VGI)

^۴ Crowdsourcing Geographic Data (CSGD)

داوطلبانه (VGI) به طور خاص، یک الگوواره جدید است که می‌تواند به غنی‌سازی چارچوب‌های موجود در علم اطلاعات جغرافیایی^۱ کمک کند (Sun and Du ۲۰۱۷, Bakillah and Liang ۲۰۱۶). یکی از منابع VGI، داده‌های موجود در وب‌سایت OpenStreetMap (OSM) است که انواع گوناگونی از داده‌های مکان مبنا از جمله جاده، ساختمان، کاربری اراضی و عوارض طبیعی مثل رودخانه‌ها را شامل می‌شود (Minaei ۲۰۲۰). پایگاه داده OSM هم از نظر تعداد کاربران و هم از نظر حجم داده‌های مکانی به سرعت در حال توسعه است و حجم عظیمی از کاربران و سازمان‌ها در به‌روزرسانی داده‌ها مشارکت دارند. از این‌رو، به یک مجموعه داده جهانی، پرکاربرد و با کیفیت بالا تبدیل شده‌است. ضمن اینکه در حال حاضر علاوه بر معابر، سایر داده‌ها نظیر پارسل‌های شهری و بسیاری دیگر از نقاط مهم شهری را نیز پشتیبانی می‌کند و لذا به‌عنوان مرجعی معتبر در پژوهش‌های جغرافیایی و برنامه‌ریزی شهری به شمار می‌رود. به همین دلیل در این مقاله این بستر برای بررسی انتخاب شده‌است. (Minaei ۲۰۲۰ and Zhou et al. ۲۰۲۲). علیرغم اینکه دسترسی به چنین داده‌هایی جذاب به نظر می‌رسد اما کیفیت آنها نیاز به ارزیابی دارد، بنابراین تناسب آنها برای اهداف خاص قابل ارزیابی است (Goodchild ۲۰۰۸). طبق تعریف سازمان جهانی استاندارد (ISO)، کیفیت داده‌های جغرافیایی شامل شش عنصر اصلی: کامل‌بودن^۲، سازگاری منطقی^۳، دقت موقعیتی، دقت معنایی، کیفیت زمانی و قابلیت استفاده می‌باشد (Kresse and Fadaie ۲۰۰۴). یکی از داده‌های OSM، ساختمان‌ها هستند که با توجه به اهمیت‌شان مطالعات متعددی برای ارزیابی کیفیت آنها بر اساس معیارهای کیفی و همچنین اهداف کاربردی مختلف انجام شده‌است. Hecht و همکاران در سال ۲۰۱۳، کیفیت داده‌های ساختمانی OSM را برای ایالت‌هایی در آلمان در مقایسه با داده رسمی ارزیابی کردند. بر طبق این پژوهش، کامل‌بودن ساختمان‌های OSM در نواحی شهری در مقایسه با مناطق روستایی بیشتر است. Fan و همکاران (۲۰۱۴)، کیفیت داده‌های ساختمانی OSM را در شهر مونیخ کشور آلمان ارزیابی کردند. آنها داده‌ها را برای پنج شاخص کیفیت (کامل‌بودن، دقت معنایی، دقت موقعیتی و دقت شکلی) با استفاده از داده‌های رسمی بررسی کردند و به این نتیجه رسیدند که دقت معنایی و کامل‌بودن در داده‌های مونیخ بالا است، درحالی‌که از نظر دقت موقعیتی به طور متوسط چهار متر انحراف وجود دارد. Törnros و همکاران (۲۰۱۵)، به دنبال ارزیابی کامل‌بودن داده‌های ساختمانی در جنوب غربی آلمان دو روش سنجش کامل بودن بر اساس تعداد ساختمان و سنجش کامل بودن بر اساس مساحت جهت مقایسه داده‌های ساختمانی OSM و داده‌های رسمی را مقایسه کردند. نتایج این تحقیق نشان داد، در روش سنجش بر اساس تعداد، معمولاً نرخ کامل‌بودن ساختمان‌های OSM کمتر از مقدار واقعی برآورد می‌شود. همچنین در روش سنجش بر اساس مساحت معمولاً نرخ کامل‌بودن داده‌های OSM بیشتر از مقدار واقعی برآورد می‌شود. بنابراین این مقاله

^۱ Geographic Information Science (GIScience)

^۲ Completeness

^۳ Logical consistency

استفاده از روش‌های پیش‌پردازش داده‌ها را برای رسیدن به نرخ دقیق‌تر توصیه می‌کند. Zhou (۲۰۱۸)، در مقاله‌ای به شناسایی رابطه بین تراکم و کامل‌بودن داده ساختمان OSM در نواحی شهری برای تخمین کیفیت این داده‌ها پرداخت. نتایج این مقاله نشان داد که بین کامل‌بودن و تراکم ساختمانی داده‌های ساختمانی OSM یک رابطه خطی وجود دارد. همچنین، تراکم ساختمانی یک معیار تاثیرگذار برای تخمین کامل‌بودن داده‌های ساختمانی OSM است. Tian و همکاران در سال ۲۰۱۹، در پژوهشی، به تحلیل ارتقاء، کامل‌بودن و الگوی جغرافیایی ساختمان‌های OSM در چین پرداختند. بر طبق نتایج این تحقیق، تعداد ساختمان‌های OSM از سال ۲۰۱۲ تا ۲۰۱۷ تقریباً ۲۰ برابر افزایش داشته‌اند. طول جاده و شاخص اقتصادی، دو پارامتر تاثیرگذار بر توسعه ساختمان‌های OSM در چین هستند. Zhang و همکاران (۲۰۲۲)، به ارزیابی کامل‌بودن ساختمان‌ها در OSM با استفاده از اطلاعات جمعیتی پرداختند. این مطالعه با استفاده از داده‌های جمعیتی، کامل‌بودن ساختمان‌های OSM را ارزیابی کرد و نشان داد که استفاده از داده‌های جمعیتی برای ایجاد داده‌های قابل اعتماد و معتبر ساختمانی روشی موثر برای ارزیابی کامل‌بودن ساختمان‌ها در OSM است. Zhou و همکاران (۲۰۲۲)، به ارزیابی وضعیت کامل‌بودن ساختمان‌ها برای ۱۲۹۷۵ شهر در سطح جهانی پرداختند. آنها از داده‌های جمعیت در این ارزیابی استفاده کردند. نتایج این تحقیق نشان داد برای ۷۵٪ از این شهرها میزان کامل‌بودن کمتر از ۲۰٪ است و فقط برای ۹٪ از شهرهای مورد مطالعه شاخص کامل‌بودن بالای ۸۰٪ تخمین زده شد. ادبی و همکاران (۱۴۰۰)، پژوهشی با هدف ارزیابی روند تکمیل بلوک‌های ساختمانی کلانشهر تهران در داده‌های OSM در سال‌های ۲۰۱۱ تا ۲۰۲۰ میلادی انجام داده‌اند. نتایج ایشان نشان می‌دهد در دو سال آخر بازه زمانی منتخب، داده‌های بلوک ساختمانی OSM از نظر تعداد عوارض و کامل‌بودن اطلاعات هندسی افزایش چشمگیری داشته‌است و کامل‌بودن این داده‌ها از ۰/۱۸ در سال ۲۰۱۱ به ۲/۷ در سال ۲۰۲۰ رسیده‌است. سدیدی و همکاران (۱۴۰۳)، در مقاله‌ای با هدف ارزیابی استفاده از هوش مصنوعی در تکمیل داده‌های داوطلبانه کاربری اراضی OSM در شهر کرج از روش شبکه عصبی کانولوشنی و سپس روش واحد مبنا برای ارزیابی میزان کامل‌بودن داده‌های OSM استفاده کردند. نتایج این تحقیق نشان می‌دهد میزان کامل‌بودن بلوک‌های ساختمانی OSM در کل منطقه مطالعاتی برابر با ۳/۶ درصد است. همچنین این پژوهش نشان دهنده نرخ پایین کامل‌بودن بلوک‌های ساختمانی و اراضی جنگلی و نرخ بالای کاربری کشاورزی و زمین‌های میوه است.

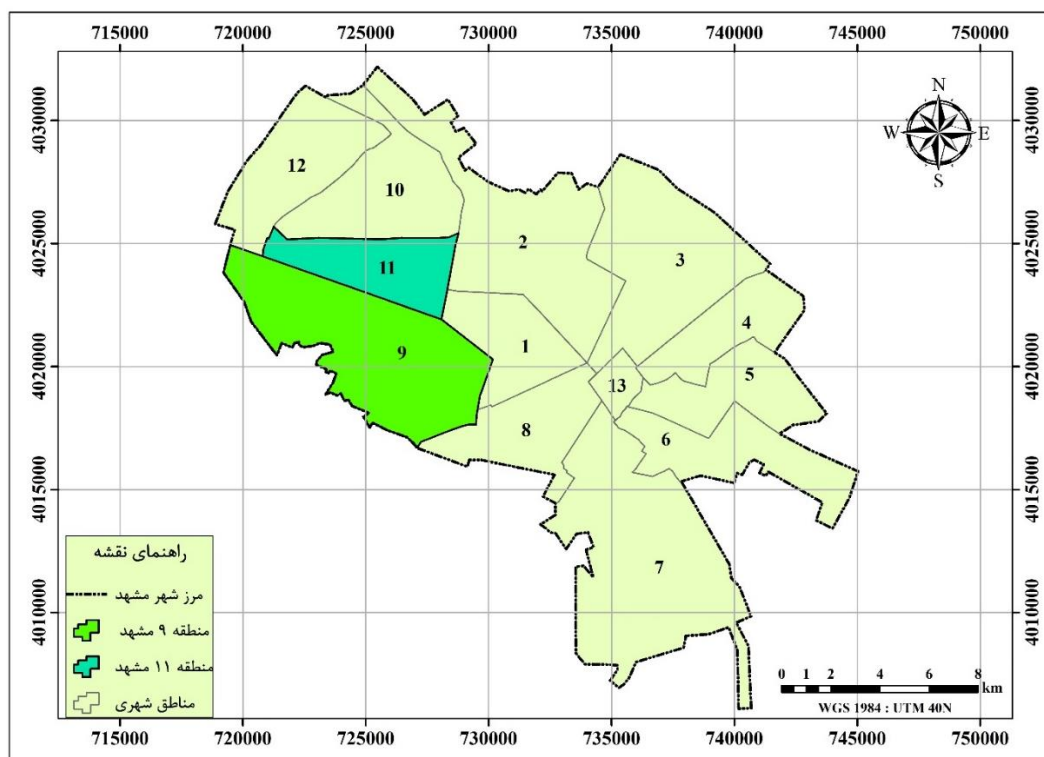
علیرغم مطالعات انجام شده در سطح جهان و مزایای بهره‌برداری از اطلاعات جغرافیایی داوطلبانه (VGI) در ایران، تحقیقات بسیار کمی در ارتباط با اطلاعات مکانی مرتبط با ساختمان‌ها در OSM انجام شده‌است. از این رو در پژوهش حاضر روند ارتقاء و کامل‌بودن ساختمان‌ها در داده‌های OSM در یک دوره زمانی ۶ ساله (۲۰۱۸-۲۰۲۳) مورد بررسی قرار می‌گیرد. همچنین ارتباط بین کاربری اراضی و جمعیت با داده‌های ساختمانی OSM ارزیابی می‌شود. در ادامه مقاله بخش‌های، منطقه مورد مطالعه و داده‌های

مورد استفاده تشریح خواهد شد. پس از آن بخش روش‌شناسی و در نهایت بخش های یافته‌ها و بحث و نتیجه‌گیری در مقاله ارائه خواهد شد.

۲- مواد و روش‌ها

۲-۱- منطقه مورد مطالعه

محدوده مورد مطالعه این پژوهش، مناطق ۹ و ۱۱ شهر مشهد در استان خراسان رضوی است. این مناطق دارای داده‌های OSM ثبت‌شده بیشتری هستند و لذا در فرآیند تحقیق امکان ارزیابی و مقایسه دقیق‌تری را فراهم می‌آورند. منطقه ۹ با ۴۱.۳۵۷ و منطقه ۱۱ با ۱۵.۸۰۸ کیلومتر مربع مساحت به ترتیب ۱۳/۲ و ۵/۰۵ درصد از مساحت کل شهر را تشکیل می‌دهند؛ شکل (۱). مجموع جمعیت این دو منطقه ۵۲۷.۲۰۳ نفر بر طبق گزارش رسمی است. کاربری اراضی غالب در این مناطق نیز به ترتیب مسکونی، مسکونی-تجاری، تجاری و آموزشی است (به استثناء مناطق ثبت نشده). داده‌های مورد نیاز در این مطالعه به صورت شیپ فایل‌های جامع از وبسایت OSM به نشانی <https://download.geofabrik.de> برای سال‌های ۲۰۱۸ تا ۲۰۲۳ بارگیری شد. پس از دانلود داده‌ها ابتدا لایه‌های اطلاعاتی مازاد حذف شده و تنها اطلاعات مرتبط با ساختمان‌ها استخراج شد. پس از آن پیش‌پردازش‌های لازم از جمله استخراج ساختمان‌های منطقه مورد مطالعه، یکسان‌سازی فرمت داده‌های رسمی و داده‌های OSM، یکسان‌سازی سیستم مختصات و حذف خطاهای توپولوژیکی مطالعه انجام گردید.



شکل (۱): موقعیت مناطق ۹ و ۱۱ در شهر مشهد

۲-۲- روش‌شناسی

۲-۲-۱- بررسی تعداد ساختمان‌ها در داده OSM

با توجه به افزایش تعداد ساختمان در طول زمان، استفاده از معیار تعداد برای ارزیابی کامل بودن ساختمان‌ها در OSM پیشنهاد شده‌است. به لحاظ تئوری، اگرچه تعداد نمی‌تواند میزان کامل بودن را به‌طور مشخص نشان دهد، اما تعداد ساختمان‌های OSM ارتباط مستقیمی با کامل بودن داده ساختمانی OSM در یک منطقه دارد (Tian et al. ۲۰۱۹).

۲-۲-۲- بررسی تراکم ساختمان‌ها در داده OSM

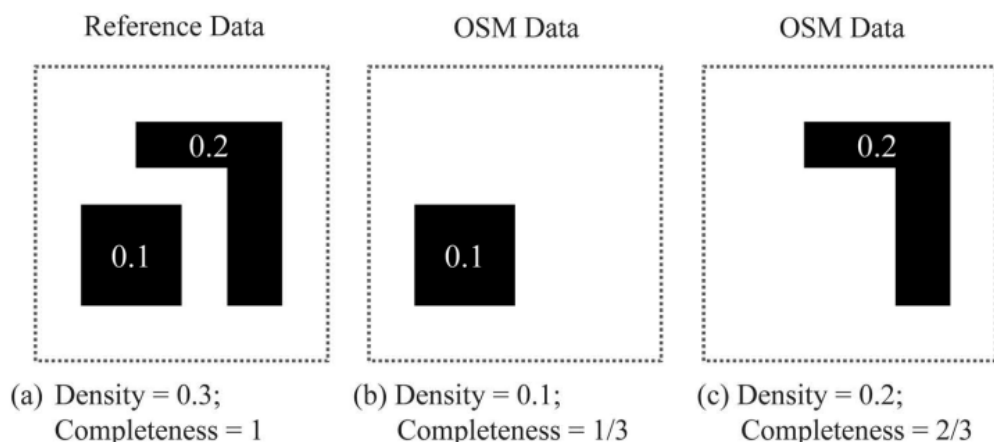
تراکم ساختمان^۱ (نسبت پوشش ساختمان) به درصد کل مساحت ساختمانی در یک واحد جغرافیایی اشاره دارد و غالباً به عنوان یک معیار در زمینه طراحی و برنامه‌ریزی شهری استفاده می‌شود (Tian et al. ۲۰۱۹).

^۱ OSM Building Density

این شاخص مطابق رابطه ذیل محاسبه می‌شود:

$$\text{تراکم (ساختمان)} = \frac{\text{مجموع مساحت ساختمان}}{\text{مساحت واحد جغرافیایی}} \times 100 \quad \text{رابطه (۱):}$$

تراکم ساختمان، ممکن است بین ۰٪ (به این معنی که هیچ ساختمانی در آن واحد جغرافیایی وجود ندارد) تا ۱۰۰٪ (به این معنی که این واحد جغرافیایی کاملاً پوشیده از ساختمان است) متفاوت باشد. به عنوان مثال، شکل (a) دو ساختمان را در یک واحد جغرافیایی نشان می‌دهد. با فرض اینکه مساحت واحد جغرافیایی ۱ باشد، بنابراین تراکم ساختمان به این شکل محاسبه می‌شود:

$$((0.1 + 0.2) \div 1) \times 100\% = 30\%$$


شکل (۲): محاسبه کامل بودن و تراکم ساختمان

۲-۲-۳- بررسی کامل بودن ساختمان‌ها در داده OSM

یکی از روش‌های پرکاربرد ارزیابی معیار کامل بودن، مقایسه ساختمان‌ها در داده‌های OSM با داده‌های تولید شده در منابع رسمی است (Adabi et al., ۲۰۲۲). به‌طور کلی عوارض در دو مجموعه داده (رسمی و داوطلبانه) به دو دسته دارای متناظر و بدون متناظر تقسیم می‌شوند و از این داده‌ها برای تعیین پارامتر کامل بودن استفاده می‌گردد. در واقع، تعداد عوارض رسمی متناظر در مجموعه داده داوطلبانه به تعداد کل عوارض مجموعه داده رسمی، میزان حضور داده‌های رسمی را در مجموعه داده داوطلبانه نشان می‌دهد که این مقدار همان میزان کامل بودن داده‌های داوطلبانه نسبت به داده‌های رسمی است. شاخص کامل بودن ساختمان‌ها در داده OSM با استفاده از رابطه ۲ بررسی می‌گردد:

$$\text{رابطه (۲):} \quad \text{کامل بودن (ساختمان)} = \frac{\text{مجموع مساحت ساختمان در داده OSM}}{\text{مجموع مساحت ساختمان در داده رسمی}} \times 100$$

کامل بودن ساختمان‌ها در OSM ممکن است بین ۰٪ (به این معنی که هیچ ساختمانی در این واحد جغرافیایی وجود ندارد) تا ۱۰۰٪ (به این معنی که کل مساحت لایه OSM و ساختمان‌های لایه رسمی در یک واحد جغرافیایی برابر است) متفاوت باشد. به طور مثال، کامل بودن ساختمان‌ها در OSM در شکل (۲ b) به صورت $(0.1 + 0.2) \div 0.1 = 33\%$ و شکل (۲ c) به صورت ۰.۲ (۶۶.۷٪) $(0.1 + 0.2) \div$ محاسبه می‌شود. هرچند، در صورتیکه در یک واحد جغرافیایی چه در لایه رسمی و چه در لایه OSM هیچ ساختمانی وجود نداشته باشد، کامل بودن برابر ۱۰۰ درصد است.

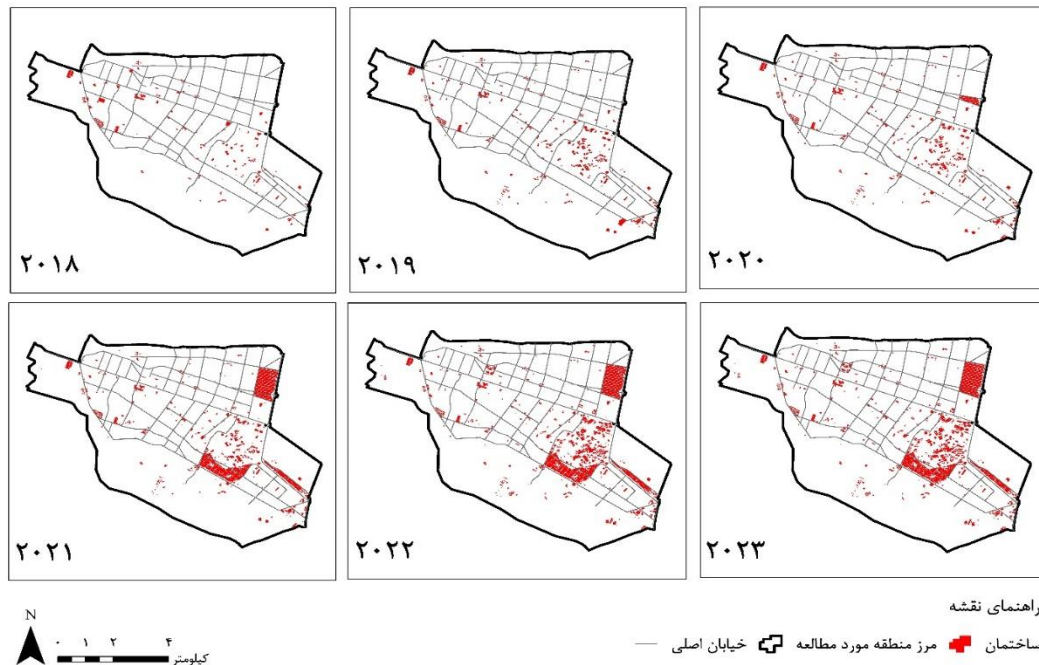
۲-۲-۴- رابطه بین شاخص تراکم و کامل بودن ساختمان‌ها در داده OSM

به منظور بررسی رابطه بین کامل بودن و تراکم ساختمان‌های OSM یک لایه شبکه^۱ (۱×۱ کیلومتر) برای هم‌پوشانی مورد استفاده قرار می‌گیرد. پژوهش Zhou در سال ۲۰۱۸ نشان می‌دهد که این اندازه در محدوده شهری می‌تواند آستانه مناسبی برای اندازه واحد جغرافیایی باشد. از این رو ساختمان‌های داده OSM در سطح شبکه‌ای به این اندازه در نظر گرفته شده‌است. سپس هر دو معیار بررسی (کامل بودن و تراکم ساختمانی) برای هر سلول شبکه مطابق با فرمول ۱ و ۲ محاسبه می‌شود و رابطه بین این دو شاخص تحلیل می‌گردد.

۳- نتایج و بحث

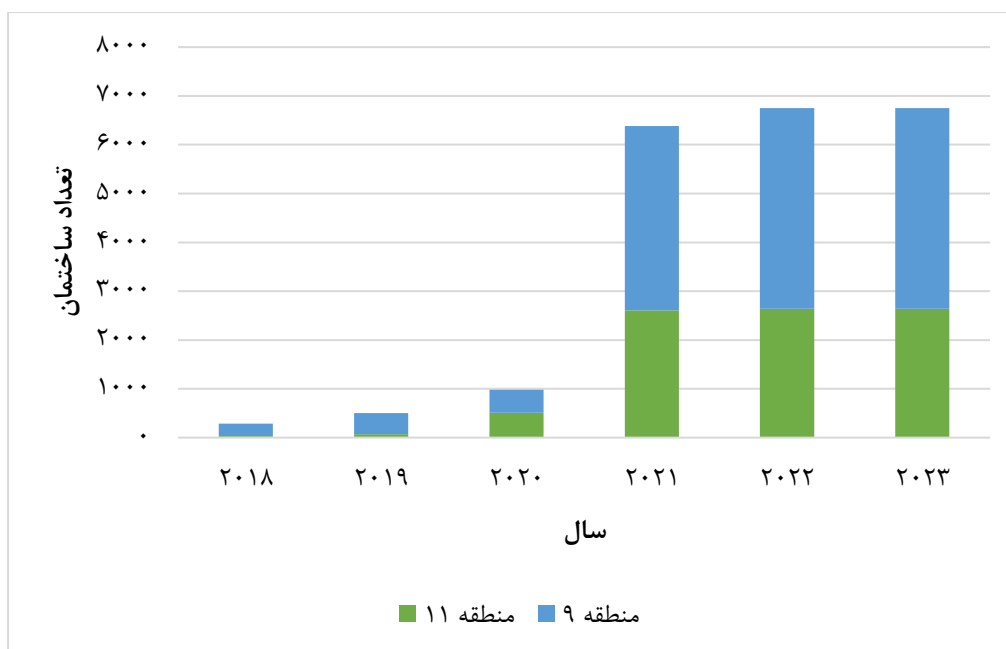
۳-۱- تحلیل بر پایه تعداد ساختمان‌های داده OSM

^۱ Grid



شکل (۳): ساختمان‌های OSM از سال ۲۰۱۸ تا ۲۰۲۳ در منطقه مورد مطالعه

شکل (۳) روند افزایش تعداد ساختمان‌های OSM را از سال ۲۰۱۸ تا ۲۰۲۳ نشان می‌دهد. در سال ۲۰۱۸ تعداد ۲۸۷ ساختمان وجود دارد. با کمی افزایش در سال‌های ۲۰۱۹ و ۲۰۲۰ تعداد ساختمان‌ها به ترتیب به ۵۰۲ و ۹۷۹ مورد می‌رسد. این در حالی است که تعداد ساختمان‌ها از سال ۲۰۲۰ تا ۲۰۲۱ با تقریباً ۵۵۱ درصد افزایش، به ۶۳۸۱ ساختمان رسیده‌است. روند افزایشی مجدداً در بین سال‌های ۲۰۲۱ تا ۲۰۲۳ کند می‌شود. تعداد ساختمان‌ها از ۲۰۲۱ تا ۲۰۲۳ حدوداً ۶ درصد افزایش پیدا می‌کند و به تعداد ۶۷۵۲ می‌رسد (جدول ۱)، (شکل ۴). این روند در تحقیق Tian و همکاران (۲۰۱۹) نیز به چشم می‌خورد. در پژوهش ایشان، تعداد ساختمان‌های OSM میان سال‌های ۲۰۱۲ تا ۲۰۱۷ در چین حدود ۲۰ برابر افزایش داشته‌است. همچنین پژوهش Herfort و همکاران در سال ۲۰۲۳ نشان می‌دهد که بین سال‌های ۲۰۱۰ تا ۲۰۱۳، تقریباً ۸۰ درصد به تعداد ساختمان‌ها در داده OSM شهر پاریس اضافه شده‌است.



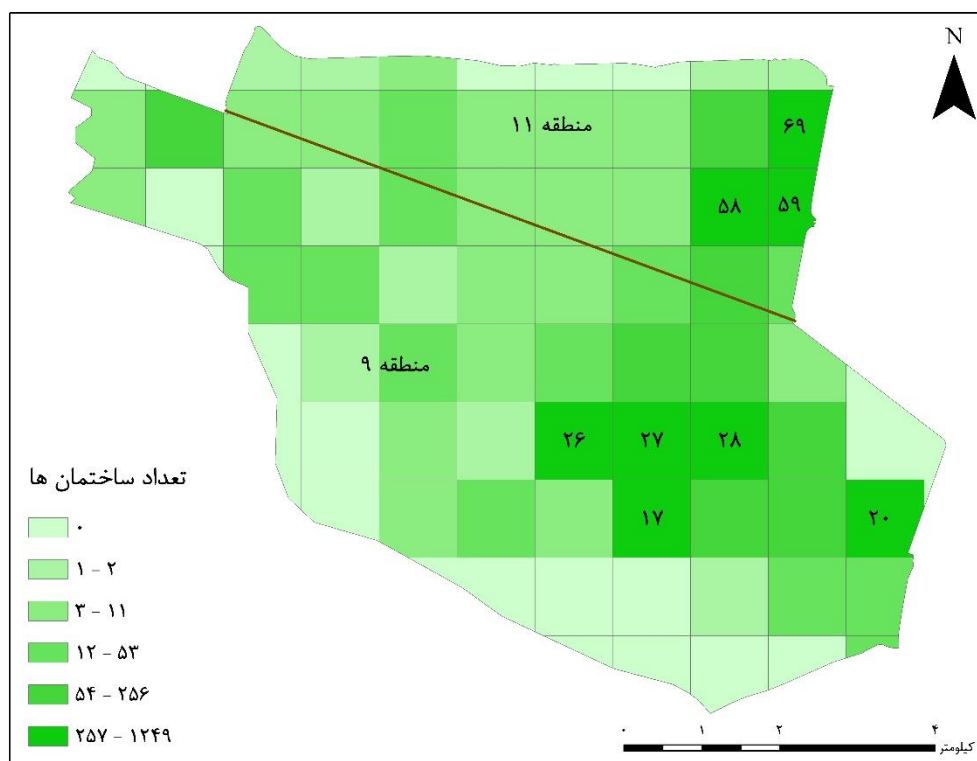
شکل (۴): تعداد ساختمان‌های OSM در مناطق ۹ و ۱۱ شهر مشهد از سال ۲۰۱۸ تا ۲۰۲۳

جدول (۱): تعداد و مساحت ساختمان‌های OSM در مناطق ۹ و ۱۱ شهر مشهد از سال ۲۰۱۸ تا ۲۰۲۳

سال	۲۰۱۸	۲۰۱۹	۲۰۲۰	۲۰۲۱	۲۰۲۲	۲۰۲۳
تعداد	۲۸۷	۵۰۲	۹۷۹	۶۳۸۱	۶۷۵۲	۶۷۵۲
مساحت (m ^۲)	۳۷۷۴۳۳.۴۲	۵۷۳۸۲۸.۶۷	۶۰۹۷۳۲.۶۹	۱۳۲۷۱۸۹.۴۵	۱۵۵۸۲۱۷.۸۹	۱۵۶۲۷۰۴.۱۸

۳-۲- پراکندگی، جمعیت و کاربری اراضی ساختمان‌ها در داده OSM

با توجه به شکل (۵)، که در آن شماره سلول‌های منتخب مشخص است، در بیشتر سلول‌ها، کمتر از ۱۰۰ ساختمان در داده OSM وجود دارد. در صورتیکه، ۸ سلول از ۵۱ سلولی که داده در آنها ثبت شده است، بیش از ۱۰۰۰ داده ساختمانی را شامل می‌شوند. این سلول‌ها اکثراً در شرق و شمال شرق منطقه مورد مطالعه واقع شده‌اند. تعداد ساختمان به ثبت رسیده و جمعیت این ۸ سلول منتخب در جدول شماره (۲) نشان داده شده‌است. همچنین کاربری اراضی آنها در جدول شماره (۳) مشخص است. بطور کلی توزیع کاربری‌های غالب در این ۸ سلول به ترتیب، مسکونی، مسکونی-تجاری، تجاری و آموزشی است. به بیان دیگر بیشترین ترسیم بیشتر در پیرامون و داخل دانشگاه فردوسی مشهد و پیرامون پارک ملت و چهارراه میلاد هستند که از نظر آموزشی و تجاری بسیار حائز اهمیت می‌باشند.



شکل (۵): توزیع تعداد داده‌های ساختمانی OSM در مناطق ۹ و ۱۱ مشهد در سال ۲۰۲۳ در هر سلول شبکه

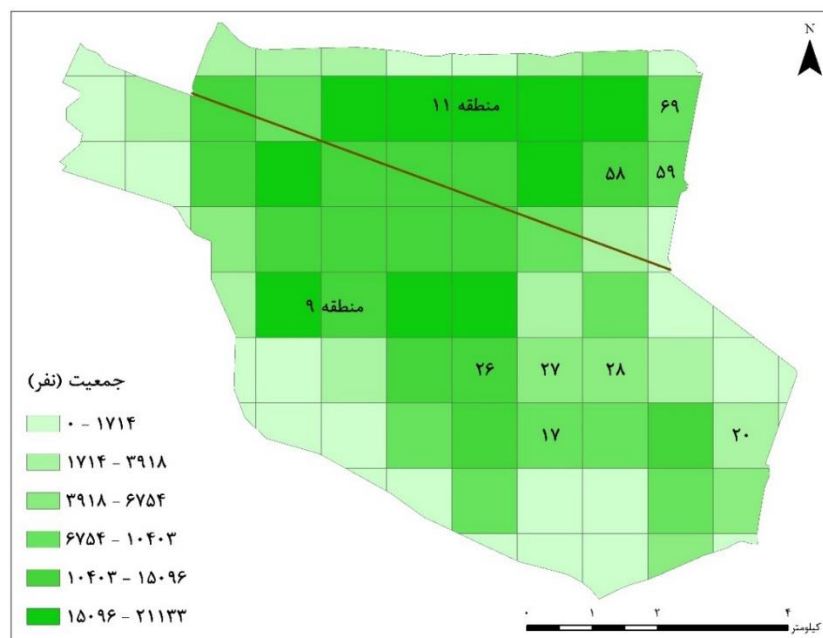
جدول (۲): جمعیت و تعداد ساختمان ثبت شده داده OSM سال ۲۰۲۳ در سلول‌های منتخب

شماره سلول	۶۹	۲۷	۵۹	۲۶	۲۸	۵۸	۱۷	۲۰
تعداد	۱۵۵۳	۱۲۴۹	۱۱۱۴	۶۲۴	۵۳۶	۵۲۰	۳۱۱	۲۹۲
ساختمان								
رتبه	۱	۲	۳	۴	۵	۶	۷	۸
جمعیت (نفر)	۱۰۲۸۵	۵۶۰۸	۸۱۴۹	۱۳۸۸۲	۵۲۱۰	۱۴۲۵۱	۹۹۸۵	۳۸۴۸

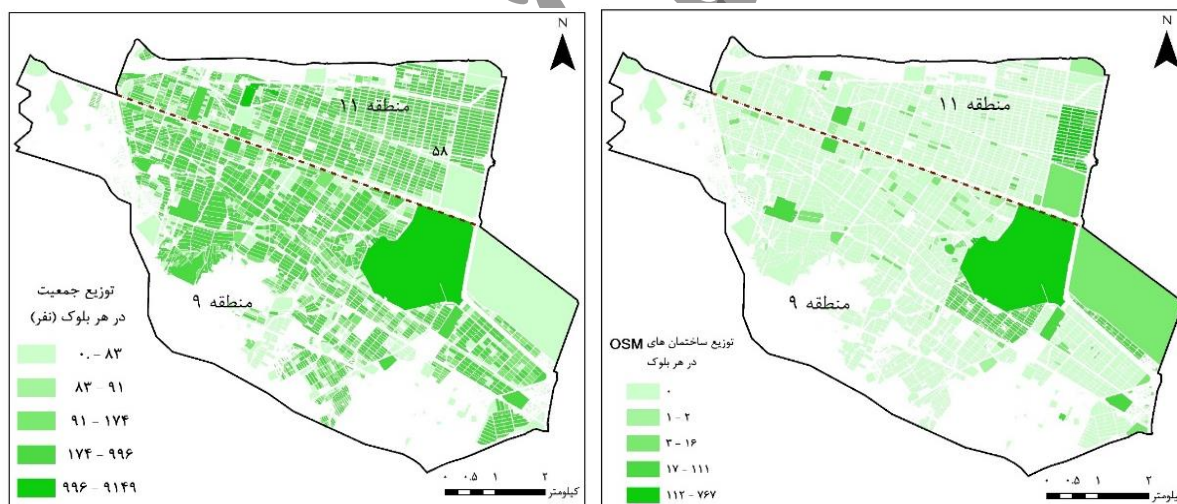
جدول (۳): کاربری اراضی ثبت شده در سلول‌های منتخب

شماره سلول	۱۷	۲۰	۲۶	۲۷	۲۸	۵۸	۵۹	۶۹
آموزشی	۲	۱	۵	۱	۰	۰	۰	۰
تجاری	۱۴	۱۳	۹	۲۱	۱۳	۲۸	۴	۶
مسکونی	۱۰۸۶	۵۳۹	۱۴۳۳	۸۲۵	۵۶۸	۱۴۷۷	۸۱۱	۸۹۸
مسکونی - تجاری	۱۳۵	۵۱	۲۴۷	۲۲۸	۸۸	۴۰۸	۹۱	۷۹

در بررسی رابطه بین داده‌های ساختمانی OSM و جمعیت، با توجه به شکل (۶) و شکل (۷ چپ)، جمعیت قابل توجهی از منطقه مورد مطالعه (بیش از ۴۰ درصد) در ۸ سلول منتخب ساکن هستند. این مسئله می‌تواند نشان‌دهنده تاثیر جمعیت بر ترسیم ساختمان در داده‌های OSM باشد.



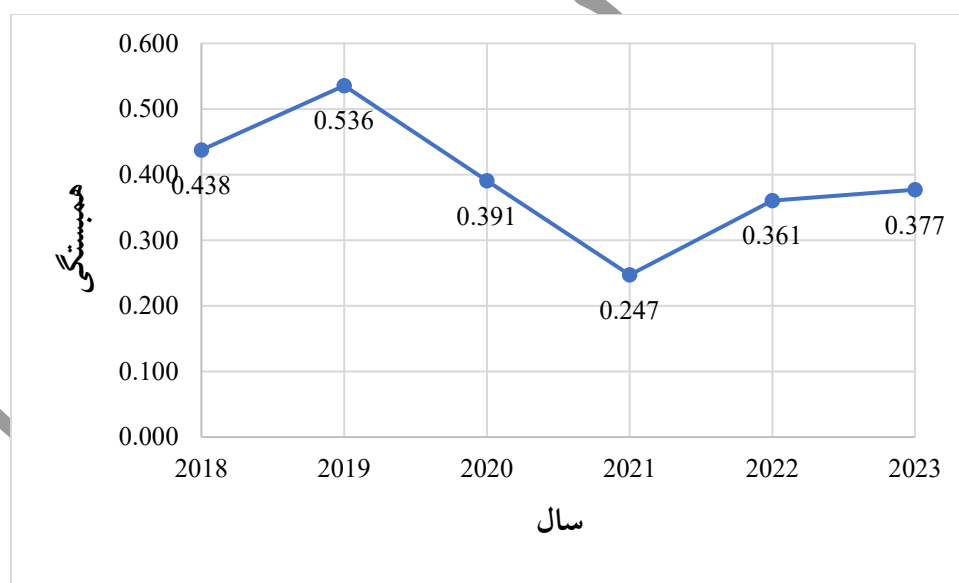
شکل (۶): توزیع جمعیت در مناطق ۹ و ۱۱ مشهد در سال ۲۰۲۳ در هر سلول شبکه



شکل (۷) راست، توزیع تعداد داده های ساختمانی OSM در مناطق ۹ و ۱۱ مشهد در سال ۲۰۲۳ در هر بلوک و سمت

چپ، توزیع جمعیت در مناطق ۹ و ۱۱ شهر مشهد در هر بلوک

با توجه به شکل ۸، در سال ۲۰۱۸ میان تعداد ساختمان‌های ذخیره‌شده در OSM و جمعیت هر بلوک، همبستگی مثبت وجود دارد. بنابراین به طور کلی، با افزایش تراکم جمعیتی در هر بلوک شهری، تعداد ترسیم‌ها افزایش یافته‌است. در سال ۲۰۱۹، ضریب همبستگی نشان‌دهنده رابطه مثبت قوی‌تری در مقایسه با سال قبل است. به این معنی که تعداد ساختمان‌های ذخیره‌شده در OSM بیشتر با جمعیت هر بلوک مرتبط است. ضریب همبستگی برای سال‌های ۲۰۲۰ تا ۲۰۲۳ کاهش پیدا می‌کند اما همچنان نشان‌دهنده یک همبستگی مثبت بین تعداد ساختمان‌های ذخیره‌شده در OSM و جمعیت در هر بلوک شهری است. با این حال، ضریب همبستگی سال ۲۰۲۳ کمی بیشتر از دو سال قبل خود است. می‌توان نتیجه گرفت که همبستگی مثبتی بین تعداد ساختمان‌های ذخیره‌شده در OSM و جمعیت در هر بلوک شهری در سال‌های (۲۰۱۸-۲۰۱۹) وجود دارد. با این حال، این همبستگی در سال‌های بعدی (۲۰۲۰-۲۰۲۳) ضعیف شده و رابطه میان آنها کمتر مشخص است. بنابراین، میان جمعیت هر بلوک و افزایش میزان ترسیم ساختمان‌های OSM رابطه معناداری وجود دارد. اما این همبستگی در سال‌های مختلف، به صورت نوسانی، متغیر است.

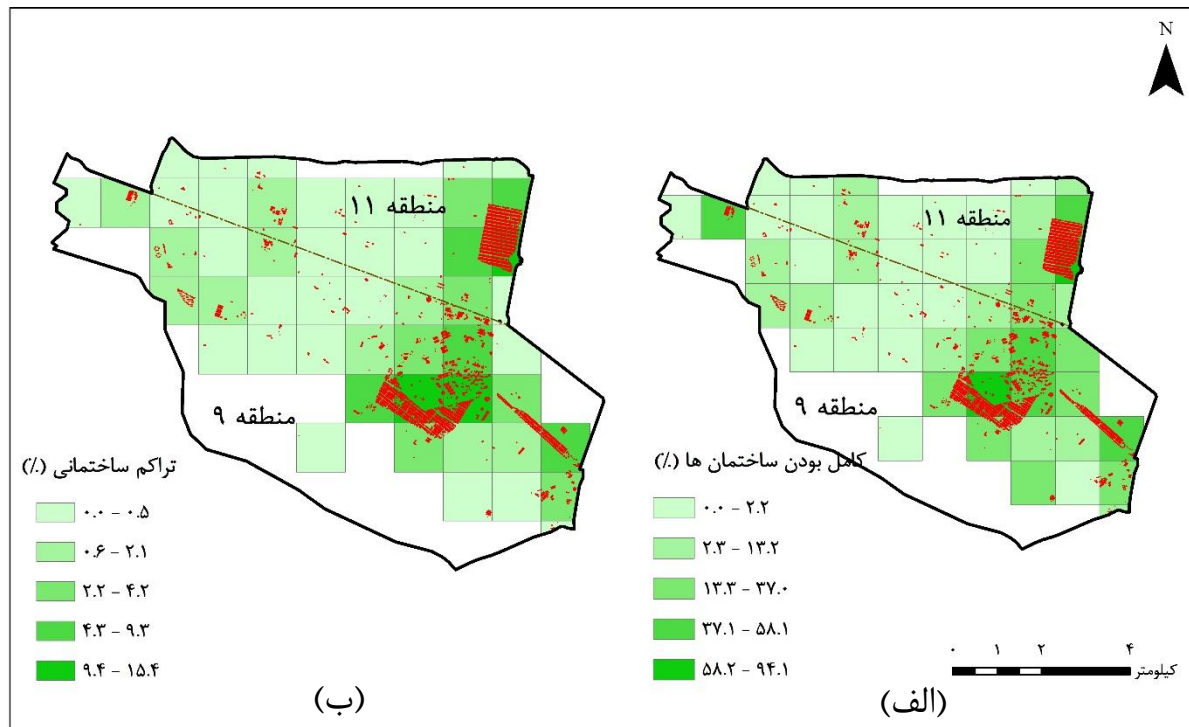


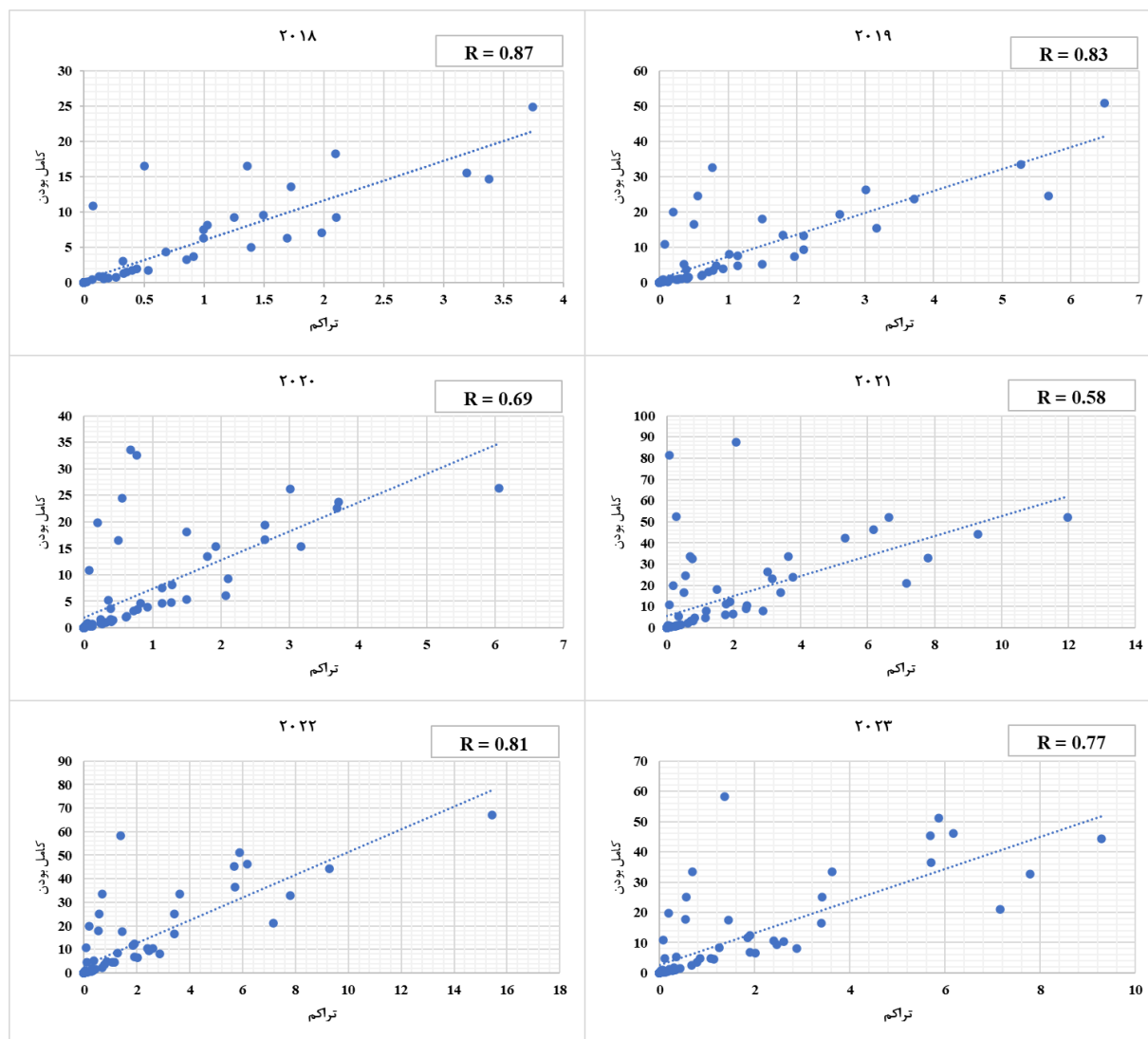
شکل (۸): همبستگی میان تعداد ساختمان‌ها در داده OSM و جمعیت هر بلوک از سال ۲۰۱۸ تا ۲۰۲۳

۳-۳- تحلیل برپایه شاخص تراکم و کامل بودن

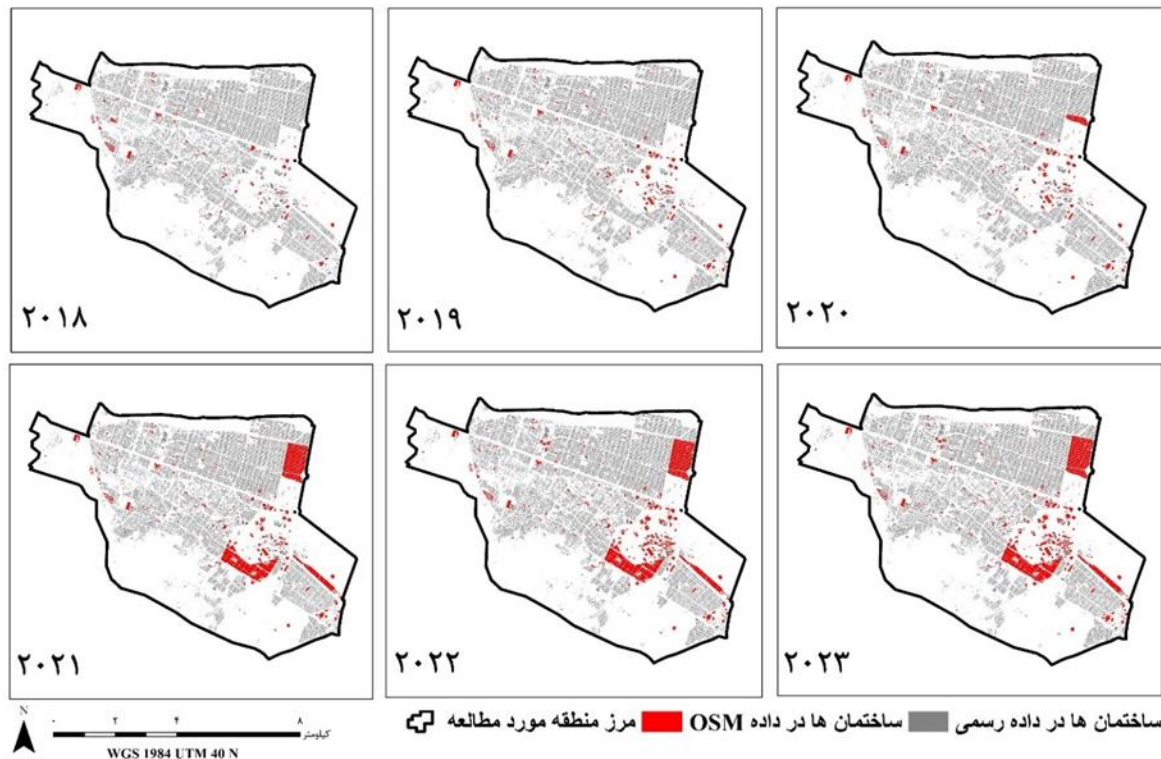
با توجه به شکل (۱۲) که تعداد ساختمان‌های OSM را در مقایسه با داده‌های رسمی موجود نشان می‌دهد و همچنین شکل (۹) الف، که بر اساس شاخص کامل بودن منطقه مورد مطالعه را نمایش می‌دهد، به خوبی بیانگر آن است که ساختمان‌های داده OSM نسبت به داده رسمی بسیار ناقص هستند. تحقیقات انجام شده توسط Hecht و همکاران در سال ۲۰۱۳ و Herfort و همکاران در سال ۲۰۲۳ نیز به همین نتیجه رسیده‌اند.

شکل (۹) نشان می‌دهد در قسمت‌های شرق و شمال شرق مناطق ۹ و ۱۱ مشهد، شاخص تراکم و شاخص کامل بودن داده‌های ساختمانی در OSM بالا است. برای بررسی رابطه بین تراکم ساختمانی و کامل بودن داده‌های ساختمانی OSM از ۲۰۱۸ تا ۲۰۲۳ ضریب همبستگی بین این دو شاخص در هر سلول شبکه محاسبه شد (شکل ۱۰). مقدار همبستگی‌ها که حداکثر ۰.۸۷ در سال ۲۰۱۸ و حداقل ۰.۵۸ در سال ۲۰۲۱ هستند نشان می‌دهند در مکان‌هایی که تراکم داده ساختمانی در داده OSM بالاتر بوده‌است، کاربران تمایل بیشتری برای ایجاد و ترسیم ساختمان جدید داشته‌اند. این پدیده نشان می‌دهد زمانیکه در یک مکان مشخص از نقشه، عملیات ترسیم و ایجاد ساختمان آغاز می‌شود، این فرآیند در سال‌های بعدی نیز ادامه می‌یابد. این مسئله بیانگر این است که کاربران OSM انگیزه بیشتری برای مشارکت در ایجاد و به روزرسانی مناطق شناخته شده دارند و به مرور زمان جزئیات بیشتری را به این مناطق اضافه می‌کنند. این یافته‌ها مطابق با تحقیقات انجام شده توسط Zhou و همکاران (۲۰۱۸) است. در تحقیق ایشان نیز رابطه بین تراکم و کامل بودن داده‌های ساختمان شهری را در OpenStreetMap بررسی شده‌است. آنها دریافتند مناطقی که تراکم ساختمانی بالاتری دارند، داده‌های کامل‌تری نیز وجود دارد. که نشان می‌دهد کاربران به احتمال زیاد در مناطقی که تراکم داده‌های ساختمانی بالا است مشارکت بیشتری می‌کنند. همچنین بر اهمیت تراکم به عنوان یک عامل مؤثر بر کامل بودن داده‌های ساختمان OSM تأکید می‌کند.





شکل (۱۰): همبستگی میان شاخص کامل بودن و تراکم در مناطق ۹ و ۱۱ شهر مشهد از سال ۲۰۱۸ تا ۲۰۲۳



شکل (۱۱): ارتقاء داده‌های ساختمانی OSM در مناطق ۹ و ۱۱ شهر مشهد از ۲۰۱۸ تا ۲۰۲۳

۴- نتیجه‌گیری

از آنجاییکه مجموعه داده‌های رسمی از ساختمان‌ها برای بسیاری از شهرها وجود ندارد و یا به طور رایگان در دسترس نیست، مجموعه‌ای از داده‌های ساختمانی رایگان می‌تواند برای پژوهشگران بسیار سودمند باشد. در این پژوهش به تحلیل ارتقاء و کامل بودن داده‌های ساختمانی در OSM پرداخته شد. به این منظور، ابتدا داده‌های ساختمان‌ها در OSM و روند تهیه و پیش‌پردازش آنها مورد ارزیابی قرار گرفت. در ادامه، منطقه مورد مطالعه (مناطق ۹ و ۱۱ مشهد) به یک شبکه‌ی سلولی به اندازه 1×1 کیلومتر تقسیم شد و این مساحت به عنوان واحد جغرافیایی مبنای محاسبات قرار گرفت. برای هر واحد جغرافیایی، سه معیار (تراکم ساختمانی، کامل بودن و تعداد ساختمان در داده‌های OSM) در منطقه مورد مطالعه برای سال‌های ۲۰۱۸ تا ۲۰۲۳ بررسی شد. سپس رابطه بین تراکم و کامل بودن داده‌های ساختمانی برای هر سلول شبکه مورد محاسبه قرار گرفت. همچنین دو پارامتر جمعیت و کاربری اراضی در بررسی‌ها مورد استفاده قرار گرفتند. طبق نتایج:

داده ساختمانی OSM در مناطق ۹ و ۱۱ مشهد از ۲۰۱۸ تا ۲۰۲۳، ۲۰ برابر افزایش پیدا کرده‌است. به ویژه برای مناطق شرق و شمال شرقی این افزایش مشهودتر است. همچنین کاربری اراضی مسکونی، مسکونی-تجاری، تجاری و آموزشی به ترتیب، بیشترین

فراوانی را به خود اختصاص داده‌اند. علیرغم همبستگی مثبت و معنادار بین داده‌های ساختمانی OSM و جمعیت هر بلوک ساختمانی، این همبستگی به صورت نوسانی در بازه زمانی مورد بررسی متغیر بوده‌است. مقدار همبستگی‌ها میان شاخص تراکم و کامل بودن نشان می‌دهد در مکان‌هایی که تراکم داده ساختمانی در داده OSM بالاتر بوده‌است، کاربران تمایل بیشتری برای ایجاد و ترسیم ساختمان جدید داشته‌اند. همچنین، با توجه به مقادیر شاخص کامل بودن می‌توان نتیجه گرفت که در مقایسه با داده‌های رسمی، ساختمان‌های به ثبت رسیده در OSM به طور قابل توجهی از تعداد واقعی ساختمان‌ها در داده رسمی کمتر هستند.

با توجه به این تحقیق، می‌توان گفت که داده‌های OSM برای تحلیل جغرافیایی ساختمان‌ها در شهرهای بزرگ مانند مشهد مناسب هستند، اما برای استفاده‌های بیشتر و دقیق‌تر، نیاز به تکمیل و بهبود این داده‌ها وجود دارد. همچنین بررسی شاخص‌های تاثیرگذار می‌تواند اطلاعات سودبخشی در جهت استفاده مطمئن‌تر از این داده‌ها در اختیار کاربران قرار دهد.

در این پژوهش سعی شده شاخص‌هایی مورد استفاده قرار گیرند که اعتبار آنها در تحقیقات معتبر پیشین مورد تایید قرار گرفته است و بدیهی است که این شاخص‌ها کامل نبوده و دارای نقاط ضعف نیز خواهند بود، از جمله عدم توانایی تشخیص ساختمان‌هایی ترسیم‌شده‌ای که در محیط خارجی وجود ندارند. بنابراین پیشنهاد می‌شود این مسئله در تحقیقات آتی مورد توجه قرار گیرد.

همچنین، شاخص جمعیت نشان داد، در بلوک‌هایی که جمعیت بیشتری در آنها ساکن بوده‌اند، میزان ترسیم ساختمان‌های جدید OSM نیز بیشتر بوده‌است. علیرغم وجود همبستگی مثبت بین تعداد ترسیم‌ها و جمعیت در سال‌های مختلف، این همبستگی نوسانی بوده‌است و این نشان می‌دهد علاوه بر جمعیت، باید شاخص‌های دیگر نیز بررسی شود. از این رو پیشنهاد می‌گردد در پژوهش‌های آتی رابطه ترسیم ساختمان‌ها با سایر شاخص‌ها مد نظر قرار گیرد.

- Adabi, R., Ali Abbaspour, R., & Chehreghani, A. (۲۰۲۲). An assessment on building blocks completion process in the Volunteered Geographic Information (VGI) for using in smart cities. *Scientific- Research Quarterly of Geographical Data (SEPEHR)*, ۳۰(۱۲۰), ۲۷-۴۲. <https://doi.org/10.22131/sepehr.2022.201.02>.
- Bakillah M, Liang S., (۲۰۱۶). Open Geospatial Data, Software and Standards, ۱(۱), ۱-۲. <https://doi.org/10.1187/s40960-016-004-1>.
- Elwood, S., Goodchild, M. F., & Sui, D. Z. (۲۰۱۲). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the association of American geographers*, 102(۳), ۵۷۱-۵۹۰. <https://doi.org/10.1080/00045608.2011.590607>.
- Fan H, Zipt A, Fu Q, Neis P., ۲۰۱۴. Quality Assessment for Building Footprints Data on OpenStreetMap. *International Journal of Geographical Information Science*, ۲۸(۴): ۷۰۰-۷۱۹. <https://doi.org/10.1080/13658816.2013.867490>.
- Goodchild F. M., ۲۰۰۷. Citizens as Sensors: The World of Volunteered Geography. *GeoJournal*, ۶۹:۲۱۱-۲۲۱. <https://doi.org/10.1007/s10708-007-9111-y>.
- Hecht R, Kunze C, Hahmann S., ۲۰۱۳. Measuring Completeness of Building Footprints in OpenStreetMap over Space and Time. *International Journal of Geo-Information*, ۲(۴): ۱۰۶۶-۱۰۹۱. <https://doi.org/10.3390/ijgi2041066>.
- Herfort B, Lautenbach S, Porto de Albuquerque J, Anderson J, Zipf A., ۲۰۲۳. A Spatio-Temporal Analysis Investigating Completeness and Inequalities of Global Urban Building Data in OpenStreetMap. *Nature Communications*, ۱۴(۱), ۳۹۸۵. <https://doi.org/10.1038/s41467-023-39698-7>.
- Kresse W, Fadaie K., ۲۰۰۴. ISO Standards for Geographic Information. Springer: Berlin. <https://doi.org/10.1007/978-3-707-08039-9>.
- Mobasheri Amin., ۲۰۱۷. A Rule-Based Spatial Reasoning Approach for OpenStreetMap Data Quality Enrichment; Case Study of Routing and Navigation. *Sensors*, ۱۷(۱۱): ۲۴۹۸. <https://doi.org/10.3390/s17112498>.
- Minaei M., ۲۰۲۰. Evolution, Density, and Completeness of OpenStreetMap Road Networks in Developing Countries: The Case of Iran. *Applied Geography*, ۱۱۹, ۱۰۲۲۴۶. <https://doi.org/10.1016/j.apgeog.2020.102246>.
- Sadidi J, tamnia F, rezaian H. Assessment of using artificial intelligence in completeness of Volunteer Geographic Information. A case study for Open Street Map (OSM) landuse data. *Journal of Spatial Analysis Environmental Hazards* ۲۰۲۴; ۱۱ (۱): ۱. <https://doi.org/10.1111/jsaeh.11112>.

Sun Y, Du Y., ۲۰۱۷. Big Data and Sustainable Cities: Applications of New and Emerging Forms of Geospatial Data in Urban Studies. *Open Geospatial Data Software and Standards*, ۲(۱): ۱-۴. <https://doi.org/10.1186/s40960-017-0037-0>.

Tian Y, Zhou Q, F, X., ۲۰۱۹. An Analysis of the Evolution, Completeness, and Spatial Patterns of OpenStreetMap Building Data in China. *ISPRS International Journal of Geo-Information*, ۸(۱), ۳۵. <https://doi.org/10.3390/ijgi8010035>.

Törnros, T., Dorn, H., Hahmann, S., & Zipf, A. (۲۰۱۵). Uncertainties of completeness measures in OpenStreetMap—A case study for buildings in a medium-sized German city. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, ۲, ۳۵۳-۳۵۷. <https://doi.org/10.5194/isprsannals-II-3-W0-353-2015>.

Zhou Q., ۲۰۱۸. Exploring the Relationship Between Density and Completeness of Urban Building Data in OpenStreetMap for Quality Estimation. *International Journal of Geographical Information Science*, ۳۲(۲), ۲۵۷-۲۸۱. <https://doi.org/10.1080/13658816.2017.1390883>.

Zhou Q, Zhang Y, Chang K, Brovelli M. A., ۲۰۲۲. Assessing OSM Building Completeness for Almost ۱۳,۰۰۰ Cities Globally. *International Journal of Digital Earth*, ۱۵(۱), ۲۴۰۰-۲۴۲۱. <https://doi.org/10.1080/17513758.2022.2109000>.

Zhang Yuheng, Zhou Qi, Brovelli M. Antonia, Li. Wanjing., ۲۰۲۲. Assessing OSM Building Completeness Using Population Data. *International Journal of Geographical Information Science*, ۳۶(۷): ۱۴۴۳-۱۴۶۶. <https://doi.org/10.1080/13658816.2021.2023108>.