

לאמידה מדויקת של ריכוזי זיהום אויר ישנה חשיבות רבה במחקרים סביבתיים ואפידמיולוגים היות ותוצאות מחקרים אלו מושפעות במידה רבה מהטעות ברמת החשיפה לזיהום. השימוש במודלים המבוססים על נתוני לווין לאמידת כמות החומר החלקיקי (PM) עלה משמעותית בשנים האחרונות. מודלים אלו מיישמים אלגוריתמי למידה סטטיסטיים המנתחים נתונים עתיים ומרחביים על מנת לספק הערכות לרמות זיהום האוויר באזורים בהם לא בוצעו מדידות. הערכות אלו משמשות לאחר מכן כמשתנים מסבירים במחקרי המשך ועל כן לאמינותן חשיבות רבה.

עם זאת, ניתוח נתונים המאופיינים בדפוס עתי או מרחבי טומן בחובו כמה אתגרים מורכבים אשר לעיתים אינם מוגדרים היטב ולחלקם אף לא נמצא פתרון. מתאם עתי-מרחבי מהווה אתגר לשיטות הסטטיסטיות המסורתיות, ובמיוחד לאלו המיועדות לחיזוי והערכת הביצועים. על כן, טכניקות סטטיסטיות מתאימות הממזערות את שגיאת החיזוי נחוצות על מנת להימנע מתוצאות אפידמיולוגיות מוטות העשויות להוביל למסקנות שגויות. אתגר נוסף קשור במורכבות העולה כאשר מודלים עתיים-מרחביים מחושבים מתוך בסיס נתונים בנפח גדול. לכן, על האלגוריתמים הסטטיסטיים לקחת בחשבון גם שיקולי מורכבות חישובית.

במחקר זה אנו מציעים ללמוד את הנושאים והאתגרים הקשורים באמידה ובהערכת הביצועים של מודלים המשמשים לחיזוי ריכוזי זיהום אויר בנתונים בעלי מתאם עתי-מרחבי. בפרט, מטרות המחקר הן: (א) להציע הרחבה לשיטות הסטטיסטיות המסורתיות לאמידת טעות החיזוי (כגון Cross-validation) במודלים לחיזוי רמת החשיפה ל-PM בנתונים עם דפוס עתי-מרחבי; (ב) לשפר את ביצועי החיזוי של המודלים המיושמים כיום וזאת על ידי מידול מבנה התלות, תוך הקטנת מורכבותם באמצעות שימוש במודלים ליניאריים; (ג) להתגבר על המורכבות החישובית הכרוכה בשימוש במודלים אלו ולאפשר את יישומם בבסיסי נתונים גדולים מאוד, זאת בין היתר בעזרת כלים אשר פותחו בתחומי המחקר העוסקים בייצוג מידע יעיל ובחישוב מקבילי.