



המחלקה להנדסת מערכות

שיטת ויזואליזציה חישובית לגילוי תרופות

תקציר מנהלים

מאת: שי אלון

ת.ז.: 038505665

מנחה: ד"ר אברהם יוסיפוף

אפקה - המכללה האקדמית להנדסה בת"א

תקציר מנהלים

המחלקה להנדסת מערכות

תואר שני

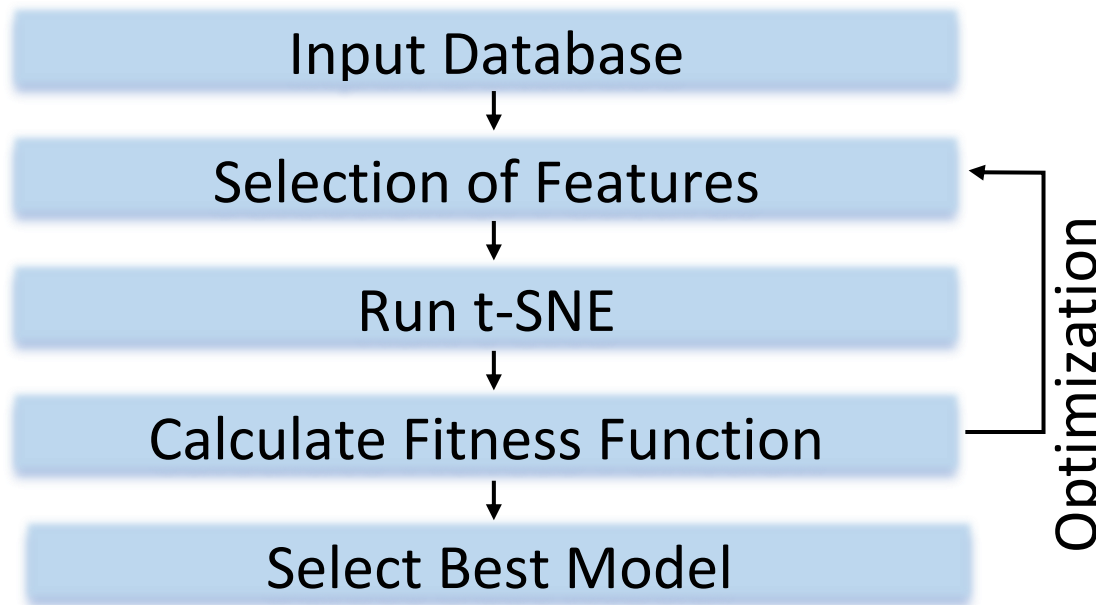
שיטת ויזואליזציה חישובית לגילוי תרופות

מאת שי אלון

העלות של תהליך המחקר ופיתוח של תרופות חדשות כיום מוערכת על ידי חברות התרופות בכחמש וחצי מיליארד דולר והוא כל הזמן במגמת עלייה. השלב הראשון של התהליך האמור - השלב הקדם-קליני - נחשב בדרך כלל לכל כך מסוכן ולא רווחי עד כי תעשיית התרופות נטשה אותו לחלוטין ולריק נכנסו ארגונים לא ממשלתיים ומוסדות אקדמיים העוסקים בגילוי תרופות חדשות ממניעים שאינם רווח.

ההשערה המועלית במאמר זה היא כי שיטת חישוב יעילה יכולה להיות מנוצלת באופן משמעותי כדי להפחית את העלויות של השלב הקדם קליני עד כדי ששלב זה יהיה כדאי כלכלית. השערה זו מודגמת על בסיס נתונים של קולטני חלבון G מצומדים (GPCRs) אשר, בעוד שהינם מהווים רק 3% לערך של מטרות מולקולריות ידועות, הם מהווים 45% מתוך כלל מטרות התרופות הידועות ולכן טמון בהם פוטנציאל מצוין לגילוי תרופות.

המאמר מדגים שימוש במאגר של למעלה משבעת אלפים תרכובות בעלות למעלה ממאה ושלושים מאפיינים. השיטה החישובית המתוארת במאמר זה משתמשת באלגוריתם גנטי כדי לזהות את קבוצת התכונות האופטימליות בהן ישתמש בתהליך צמצום ממדי T-SNE. תהליך זה נבחר כי הוא מעניק עדיפות למרחקים קצרים על פני מרחקים גדולים ולכן משמש טוב יותר כבסיס לשיטת סיווג השכן הקרוב ביותר. הקריטריונים להצלחת האלגוריתם הם הקריטריוני קרבה לשכן - כלומר, ההנחה הסבירה היא כי השכן הקרוב ביותר של תרכובת המתוייגת חיובית יש את אותו תג.



תוצאות האלגוריתם (כפי שניתן לראות בטבלה שלהלן) הוכיחו כי ההשערה נכונה וכי תהליך חישובי יכול לזהות מועמד פוטנציאלי בדיוק גבוה ולצמצם את מספר בדיקות המעבדה ביותר מ-90%.

בטבלה הבאה Q-1 פירושו ההסתברות לכך שתרכובת לא ידועה הקרובה ביותר לשכן שהינו תרכובת המפעילה סוכן מסוים (למשל דופמין) תפעיל את אותו סוכן. היות וההסתברות הממוצעת היא 97.7% נובע כי באמצעות זה אלגוריתם על מסד נתונים דומה נקבל מידה רבה של ודאות לגבי פעילות התרכובת לפני בדיקת המעבדה.

Table 1 Result Summary

	<i>Q-1</i>	<i>Q-3</i>	<i>Q-5</i>	<i>Trust</i>
<i>Dopamine</i>	0.9650	0.9650	0.9610	0.0870
<i>Adrenoceptors</i>	0.9830	0.9850	0.9840	0.0770
<i>Histamine</i>	0.9670	0.9630	0.9640	0.0780
<i>Muscarinic</i>	0.9930	0.9930	0.9930	0.0760
<i>Serotonin</i>	0.9380	0.9340	0.9210	0.0900
<i>Average</i>	0.9692	0.9680	0.9646	0.0816

למרות שמורכבות האלגוריתם אינה מתאימה למסדי נתונים בקנה מידה גדול עדיין הוכח כי הוא יעיל מאוד כאשר נעשה בו שימוש בקנה מידה קטן יותר (כמו מסד הנתונים 7000 על ידי 130 בשימוש במאמר זה).