

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.852

Шеров  
Эрадж Бахтиёрович

Поиск аналогов и оценка совместимости лекарственных средств путем анализа  
их состава

### **АВТОРЕФЕРАТ**

на соискание степени магистра информатики и вычислительной техники  
по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель  
Сиротко Сергей Иванович  
кандидат физ.-мат. наук, доцент

Минск 2017

## КРАТКОЕ ВВЕДЕНИЕ

Медицина располагает многими видами воздействия на организм больного человека. Однако наиболее значительной и распространенной является фармакотерапия – воздействие лекарственными веществами.

В настоящее время различные производители выпускают большое количество лекарственных средств, в том числе имеющих схожий состав и аналогичные действия.

Задача комплексной характеристики набора лекарственных средств требует систематизации анализа сведений из многих источников. При этом исходные сведения могут быть представлены в различной форме и быть неформализованными, неструктурированными.

Обработка такой информации – сложная задача, требующая привлечения как математических методов, так и современных компьютерных технологий.

Согласно статистике, только в одних Соединённых Штатах Америки по бюджету, дженерики, так называемые аналоги лекарств-оригиналов, сохраняют покупателям приблизительно от 8 до 10 миллиардов долларов в год в розничных аптеках. Ещё больше миллиардов сохраняется, когда больницы используют дженерики.

Взаимодействия лекарства с лекарством (ВЛЛ) происходят во время их совместного введения. Они являются одной из главных причин неблагоприятных реакций на лекарства и приводят к увеличению расходов на здравоохранение. Многие ВЛЛ не идентифицируются на стадии клинических испытаний и сообщаются после одобрения лекарств для клинического использования. Такие ВЛЛ часто приводят к заболеваемости и смертности пациентов, составляя 3–5% всех ошибок в лечении стационарных больных. Клинические ВЛЛ могут вызывать серьезные социальные и экономические проблемы. Таким образом, существует острая необходимость в обнаружении или определении ВЛЛ до того, как одобрены или введены лекарства.

В последнее время сетевые подходы фармакологии, такие как сетевая стратегия разработки лекарств, создали новую парадигму для открытия лекарственных препаратов. Поэтому разработка машинно-обучающей модели с использованием многомерных свойств лекарственного средства может быть перспективной стратегией для прогнозирования неизвестных ВЛЛ.

# **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

## **Цели и задачи исследования**

Целью диссертационного исследования является разработка алгоритмических и программных средств, ориентированных на автоматизации подбора аналогов и анализе совместимости лекарственных средств.

Достижение поставленной цели связано с решением следующих задач:

- Изучение предметной области.
- Ознакомление с подходами к технологиям обработки больших объемов неструктурированной информации.
- Построение алгоритмов подбора и анализа.
- Проектирование набора программных средств для реализации подбора и алгоритмов анализа.

Объект исследования – источники первичной информации о лекарственных средствах, технологии и алгоритмы машинного обучения.

Предмет исследования – алгоритмы подбора и анализа больших объемов данных.

## **Личный вклад соискателя**

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя С. И. Сиротко заключается в формулировке целей и задач исследования.

## **Апробация результатов диссертации**

Основные положения и результаты диссертационного исследования докладывались и обсуждались на 53-й научной конференции аспирантов, магистрантов и студентов БГИУР, секция «Программное обеспечение информационных технологий» и VI Республиканской научной конференции студентов, магистрантов и аспирантов ГГУ им. Ф.Скорины, секция «Новые материалы и технологии».

## **Опубликованность результатов диссертации**

По теме диссертации опубликованы 2 печатные работы, из них 1 работа в сборнике трудов и материалов научной конференции БГУИР, 1 работа в сборнике трудов и материалов научной конференции ГГУ имени Ф.Скорины.

## **Структура и объем диссертации**

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, списка использованных источников, списка публикаций автора и приложений.

В первой главе представлен анализ предметной области, предложена общая структура вывода гетерогенного сетевого подхода для прогнозирования взаимодействий лекарственных препаратов.

Вторая глава посвящена разработке алгоритмов, математических моделей и методов, применяемых для оценки взаимодействия лекарственных веществ.

В третьей главе представлена практическая реализация ПО для поиска аналогов, а также разработка прогнозирующих моделей с использованием ранее разработанных алгоритмов.

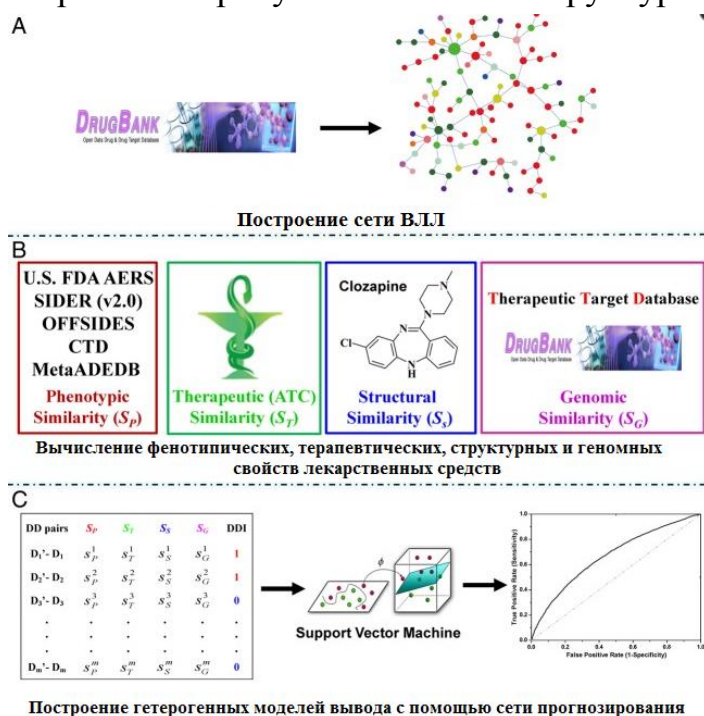
В четвертой главе рассмотрены промежуточные и итоговые результаты работы программных средств, а также проведён их статистический анализ.

Общий объем работы составляет 71 страницу, из которых основного текста – 59 страниц, 26 рисунков на 22 страницах, список использованных источников из 53 наименований на 3 страницах.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В **первой главе** рассмотрены предметная область, источники сбора первичной информации, технологии обработки больших объемов неструктурированных данных, а также кратко описана предложенная общая структура вывода гетерогенного сетевого подхода (ГСП) для прогнозирования взаимодействий лекарственных препаратов. На рисунке 1 показана структура вывода ГСП.



**Рисунок 1 – структура вывода ГСП**

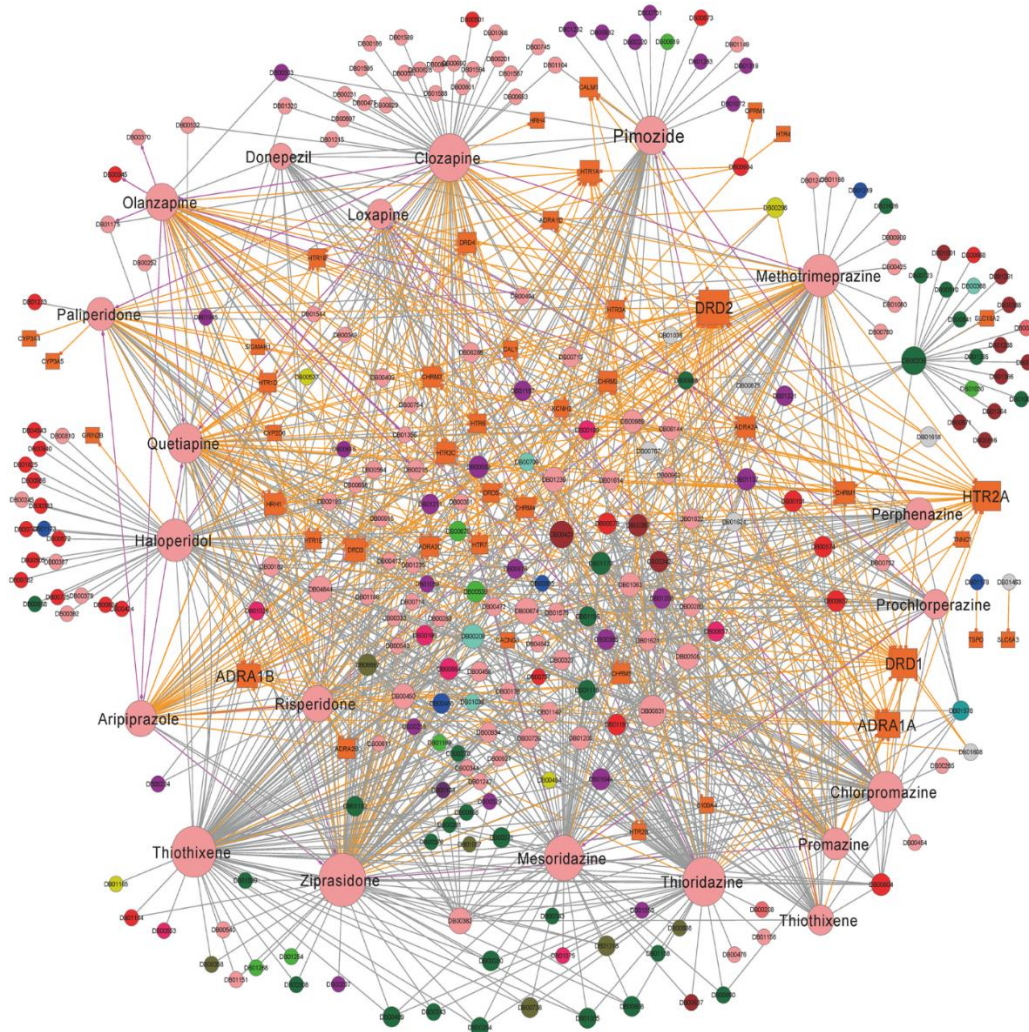
Во **второй главе** рассмотрены методы измерения четырёх типов сходства лекарственных средств, а также алгоритмы машинного обучения, применяемые для реализации программного комплекса и построения сети межлекарственных взаимодействий.

В **третьей главе** рассмотрены основные этапы разработки программных средств по поиску аналогов и оценки взаимодействий между лекарственными препаратами.

В **четвертой главе** рассмотрены полученные на каждом из этапов результаты, визуализирована сеть межлекарственных взаимодействий, а также проведен анализ полученных результатов.

В результате было разработано программный комплекс, позволяющий искать аналоги лекарственных средств по разным характеристикам, показана структура ГСП с использованием пяти моделей интеллектуального машинного обуче-

ния для прогнозирования межлекарственных взаимодействий, а также построена сеть взаимодействия препарат-препарат и препарат-мишень для антипсихотических лекарственных средств. На Рисунке 2 изображена данная сеть.



**Рисунок 2 - Сеть взаимодействия препарат-препарат и препарат-мишень для антипсихотических лекарственных средств**

## ЗАКЛЮЧЕНИЕ

В ходе выполнения диссертационного исследования были изучены различные модели и алгоритмы вычислительного прогнозирования межлекарственного взаимодействия. Основной упор был сделан на исследование возможностей технологий машинного обучения. Понимание взаимодействия лекарства с лекарством (ВЛЛ) является важным шагом в развитии лекарств и совместном введении лекарств. В настоящее время FDA США и крупнейшие мировые фармацевтические компании заинтересованы в разработке и применении вычислительного прогнозирования и оценки ВЛЛ.

В данной работе показана структура гетерогенного сетевого подхода (ГСП) и использованы пять моделей интеллектуального машинного обучения для прогнозирования ВЛЛ.

В течение последнего десятилетия сообщалось о нескольких возможных методах для вычислительного прогнозирования ВЛЛ. По сравнению с предыдущими методами одним из преимуществ ГСП является то, что его модели построены только с использованием четырех сходств, как особенностей каждой пары «лекарство–лекарство».

Хотя несколько известных моделей имеют более высокую производительность, чем данный метод, эти модели построены с использованием высокомерных функций, тем самым увеличивая сложность и «черный ящик» своих моделей.

В данной работе была построена структура ГСП с удовлетворительной производительностью для прогнозирования ВЛЛ посредством интеграции на основе машинного обучения по четырем типам сходства лекарственных средств.

Основываясь на предсказании ВЛМ, в работе был расширен подход (например, основанная на машинном обучении интеграция четырех типов сходств с лекарством: фенотипические, терапевтические, структурные и геномные сходства) для прогнозирования ВЛЛ.

Получена гипотеза ГСП, утверждающая, что если у двух препаратов есть высокое лекарственное фенотипическое, терапевтическое, структурное или геномное сходства, у них высокая вероятность ВЛЛ.

Было обнаружено, что структурное сходство положительных пар ВЛЛ значительно выше, чем у случайных пар. Это наблюдение согласуется с предыдущими работами, показывающими, что информация о сходстве лекарственного средства может быть использована для прогнозирования ВЛЛ.

Кроме того, было обнаружено, что систематическая интеграция различных источников данных, в том числе лекарственные ADR, кодовые АТС–коды,

химическая структура лекарственного средства и геномная информация, могут повысить эффективность прогнозирования ВЛЛ по сравнению с использованием только отдельных источников данных.

В рамках работы представлена структура ГСП для прогнозирования взаимодействия лекарств путём использования фенотипического, терапевтического, структурного и геномного сходств. Применено пять предсказательных моделей на основе машинного обучения на большом наборе данных лекарственных взаимодействий из базы данных DrugBank. В ходе исследования нескольких новых предсказанных лекарственных взаимодействий была продемонстрирована потенциальная полезность структуры ГСП для идентификации антипсихотических взаимодействий лекарств.

Таким образом было показано, что интеграция фенотипического, терапевтического, структурного и геномного сходств с машинным обучением представляют собой простую, но эффективную стратегию прогнозирования неизвестных лекарственных взаимодействий.



## СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1] Шеров Э.Б., Информационное обеспечение для анализа совместимости и поиска аналогов лекарственных средств /Шеров, Э. Б., Шелоник И.А., Холматов М. Дж // Компьютерные системы и сети: материалы 53-й научной конференции аспирантов, магистрантов и студентов (Минск, 2 – 6 мая 2017 г.). – Минск: БГУИР, 2017. – С. 197 – 199.

[2] Шеров Э. Б., Применение технологий BIG DATA для оценки совместимости лекарственных средств на рецепторном уровне взаимодействия друг с другом / Э. Б. Шеров // Новые материалы и технологии: материалы VI-ой научной конференции студентов, магистрантов и аспирантов (Минск, 26 – 28 апреля 2017 г.) – Гомель: ГГУ, 2017. – [В печати].

[3] Шеров, Э. Б. Новый «взгляд» на атом, как положение началу принципа соответствия / Э. Б. Шеров // Великие преобразователи естествознания: Нильс Бор : материалы юбилейных XXV Междунар. чтений (Минск, 16–17 марта 2017 года). – Минск : БГУИР, 2017. – С. 256 - 257

[4] Шелоник, И. А. Поиск профиля в социальной сети по фотографии человека / И. А. Шелоник, Э. Б. Шеров, М. Д. Холматов // Компьютерные системы и сети: материалы 53-й научной конференции аспирантов, магистрантов и студентов (Минск, 2 – 6 мая 2017 г.). – Минск: БГУИР, 2017. – С. 202 – 203.

[5] Холматов, М. Дж. Средства адаптивной настройки и подбора контента для музыкального портала / М. Дж. Холматов, И. А. Шелоник, Э. Б. Шеров // Компьютерные системы и сети: материалы 53-й научной конференции аспирантов, магистрантов и студентов (Минск, 2 – 6 мая 2017 г.). – Минск: БГУИР, 2017. – С. 199 – 200.