

GeneDetect

Поиск генов, ассоциированных с риском развития
ишемического инсульта с использованием графа знаний



Kseniya Shafalovich

 <https://t.me/KseniyaSh2036>

 shafalovich.ks@phystech.edu

Федеральное государственное автономное образовательное учреждение высшего образования
«Московский физико-технический институт (национальный исследовательский университет)»

1 курс Магистратура

Центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
healthnet

ЦЕЛЬ

создание алгоритма для поиска и исследования генов, ассоциированных с риском развития ишемического инсульта с использованием графа знаний.

ПРЕДМЕТ

анализ генетических данных, сосредоточенных на идентификации генов, которые могут играть ключевую роль в патогенезе ишемического инсульта.

ОЖИДАНИЯ

выявление генов, которые имеют сильную ассоциацию с риском развития ишемического инсульта, а также определение их биологической роли и влияние на патогенез заболевания.

Исследователи
и
медицинские
специалисты

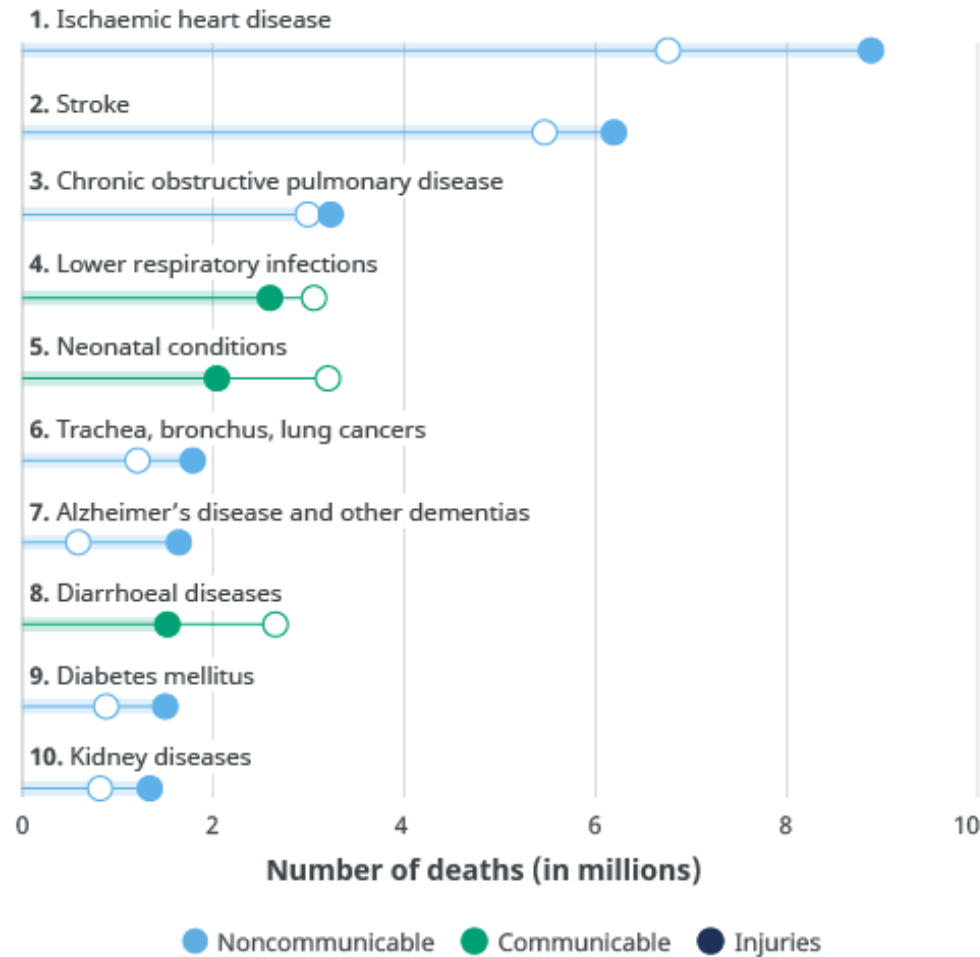
Вторая
по
значимости
причина
смертности

Неизвестные
генетические
факторы

Сложность
анализа
огромного
объема
научных
данных

Leading causes of death globally


○ 2000 ● 2019



Source: WHO Global Health Estimates.

Решение проблемы

Модель графа знаний, основанная на международных стандартных медицинских терминах и внешних данных, будет способствовать более точному представлению и взаимосвязи медицинской информации о ишемическом инсульте в лабораторных исследованиях.



Краткое описание технологии

Получение данных
Загрузка списка генов, связанных с ишемическим инсультом, из базы данных DisGeNET и сети белок-белковых взаимодействий (PPI).

Предварительная обработка и очистка данных
Преобразование данных PPI в формат, пригодный для работы с программами расчета встраивания (embeddings) графов, удаление дубликатов, обработка отсутствующих значений.
Встраивание графа сети PPI. Выбор метода встраивания графов - Node2vec, DeepWalk или LINE.
Расчет встраивания графов на примере сети PPI с помощью выбранного метода.
Оптимизация параметров, например, p и q для Node2vec.

Применение метода машинного обучения SVD
Обучение моделей с использованием выбранных алгоритмов на подготовленных данных.
Оценка производительности моделей с использованием кросс-валидации и метрик, таких как площадь под ROC-кривой (AUROC) и F1-мера.

Интерпретация результатов
Интерпретация значимости предсказанных генов и их связь с ишемическим инсультом.

Прогнозирование и оценка результатов
Прогнозирование генов, связанных с ишемическим инсультом, на основе обученных моделей.
Оценка качества прогнозов с использованием метрик, таких как точность, полнота, F1-мера.
Проверка значимости прогнозов с использованием дополнительных исследований.

Оптимизация гиперпараметров
Проведение поиска по сетке для определения оптимальных гиперпараметров выбранных алгоритмов.

Анализ рынка

С ростом количества данных медицинского генетического тестирования оборот сегмента интерпретации генетических тестов может расти на 14,5–20% в год, отмечают аналитики НТИ.

Ключевые факторы роста рынка:

Внедрение технологий ИИ и машинного обучения в медицинские исследования способствует росту рынка анализа генов.

Переход к открытому доступу к данным и совместным исследованиям может увеличить доступность данных и способствовать росту рынка.

TAM

*цифрового генома
ишемического
инсульта 5,5
миллиарда долларов.*

SAM

7 миллиардов
руб.

SOM

0,9 миллиарда
руб.

Ключевые партнеры	Ключевые виды деятельности	Ценностное предложение	Взаимоотношения с клиентом	Потребительские сегменты
<p>Медицинские учреждения и лаборатории</p>	<p>Анализ генных сетей, разработка алгоритмов анализа, медицинские исследования</p>	<p>Инновационная технология анализа генных сетей для выявления генетических маркеров инсульта</p>	<p>Предоставление инструментов для анализа Обеспечение поддержки и консультаций</p>	<p>Исследовательские институты Медицинские центры Фармацевтические компании</p>
	<p>Ключевые ресурсы</p>		<p>Каналы сбыта</p>	
	<p>Генетические данные Алгоритмы анализа Экспертиза</p>		<p>Выставки Публикации в профильных изданиях Партнерство с мед. учрежд. и лабораториями</p>	
<p>Структура затрат</p>		<p>Потоки поступления дохода</p>		
<p>Научные исследования Техническое обеспечение Административные расходы Маркетинг и реклама</p>		<p>Модель подписки Партнерство с клиниками и исследовательскими центрами Продажа аналитических услуг Привлечение инвестиций</p>		

Сравнение с аналогами

критерии	GeneDetect	Genotek	Университетские лаборатории	Исследовательские центры	Фарм. компании
Снижение себестоимости исследования	+	-	-	-	-
Повышение точности анализа	+	+	-	-	-
Увеличение скорости исследования	+	-	-	-	-
Экспертиза и опыт	-	+	+	+	+
Ресурсы и финансирование	-	+	+	+	+
Коммерциализация	+	+	-	-	+

ПЛАН РАЗВИТИЯ
ПРОЕКТА

ЭТАП 0
(5 мес.)

Разработка прототипа продукта

100 000 руб.

ЭТАП 1
(3 мес.)

Испытание разработанного продукта GeneDetect по функциональным свойствам и технико-экономическим характеристикам

700 000 руб.

ЭТАП 2
(4 мес.)

Сертификация продукта...
Производство: контрактное.
Выход продукта на рынок.

300 000 руб.

ЭТАП 3
(5 мес.)

Увеличение продаж продукта.
Производство: контрактное/собственное.

450 000 руб.

Итого бюджет проекта :

1 550 000 руб.

Команда

Управление проектом
Разработка

Шафалович Ксения Сергеевна

Опыт руководящей работы,
опыт продвижения нового
продукта, навыки
программирования python
Магистратура Менеджмент
1 курс Магистратура
Анализ данных в медицине
1 курс ДПО Биоинформатика
курсы бухгалтерский учет

Научный руководитель
Консультирование
Атаева Ольга Муратовна
КТН
ФИЦ ИУ РАН старший
научный сотрудник



Привлекаемые
специалисты:

- frontend разработчик
- Бухгалтер
- юрист

Сотрудничество:
Геномный центр

Проект **GeneDetect** находится на стадии идеи. В течение этого года проект будет развиваться и совершенствоваться, проводиться исследования, анализ данных, разработка алгоритмов и моделей. Задачей будет создание прототипа или первой версии продукта, который сможет быть презентован потенциальным пользователям, инвесторам или партнерам.

Участие в акселераторе ФИЗТЕХ Идея

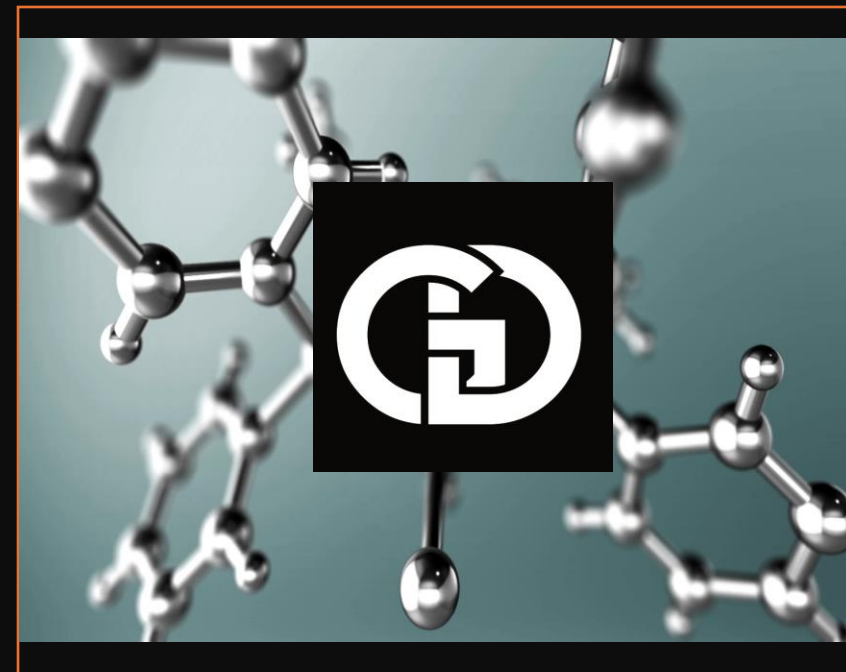
Подготовка заявки к участию в конкурсе студстартап в 2024г.

Через год после выхода на стадию MVP проект сможет представить конкретные результаты и достижения, опираясь на данные и обратную связь от пользователей и рынка. Это позволит более четко определить дальнейшие шаги развития проекта, привлечь дополнительные ресурсы и инвестиции, а также начать масштабирование и коммерциализацию продукта.

СПАСИБО

Шафалович
Ксения Сергеевна

+79646222054
shafalovich.ks@phystech.edu



ИСТОЧНИКИ
Дохода

Модель
подписки

Сотрудничество с
клиниками и
исследовательскими
центрами

Продажа
аналитических
услуг

Привлечение
инвестиций

Маркетинг
и
продвижение

Дополнительные источники данных

- **Базы данных генов и биологических процессов.**
GeneCards, HGNC, NCBI.
- **Базы данных о взаимодействиях белков.**
- **Клинические данные.**
Информация о пациентах с диагнозом ишемического инсульта, включая результаты обследований и тестов.

Модель графа знаний, основанная на международных стандартных медицинских терминах и внешних данных, будет способствовать более точному представлению и взаимосвязи медицинской информации о ишемическом инсульте.



Ожидаемые результаты

Составить базу данных генетических вариантов, связанных с ишемическим инсультом, с использованием доступных источников информации.

Создать графовую структуру, отражающую связи между генами, биомедицинскими публикациями и ассоциациями с ишемическим инсультом.

Разработать алгоритмы и методики анализа графов знаний для выявления генов, имеющих высокую степень ассоциации с ишемическим инсультом.

Провести анализ полученных данных и результатов, оценить точность и надежность выявленных ассоциаций.

Написать научную статью или отчет о проделанной работе и полученных результатах для публикации в научном журнале или презентации на конференциях.

Риски и стратегии снижения рисков

Технические риски: Недостаточная производительность алгоритмов или оборудования.

Стратегия: Проведение тщательного тестирования и оптимизации перед полным масштабированием и использованием.

Недостаток данных: Ограниченный доступ к качественным генетическим данным.

Стратегия: Сотрудничество с другими исследовательскими группами для обмена данными.

Финансовые риски: Недостаточное финансирование для завершения проекта.

Стратегия: Разработка бизнес-плана с детальным бюджетом и поиск дополнительного финансирования через гранты, инвестиции или партнерские отношения.

Этические риски: Нарушение приватности данных и недобросовестное использование результатов исследования.

Стратегия: Соблюдение всех норм и правил конфиденциальности данных и проведение этической экспертизы перед началом исследования.

Дополнительные источники данных

Модель графа знаний, основанная на международных стандартных медицинских терминах и внешних данных, будет способствовать более точному представлению и взаимосвязи медицинской информации о ишемическом инсульте.

