

Конкретная задача проекта

- Конкретной задачей является разработка новых и совершенствование существующих способов совместной обработки **разнородных мультимодальных данных** на основе выделения их **существенных признаков** методами глубокого обучения.
- Полученные результаты будут **апробированы** на модельных и реальных мультимодальных данных из области гамма-астрономии, а именно данных, получаемых с помощью гибридной системы детекторов, регистрирующих широкие атмосферные ливни (ШАЛ) в **эксперименте TAIGA**.

План из заявки в РФФИ для 2023:

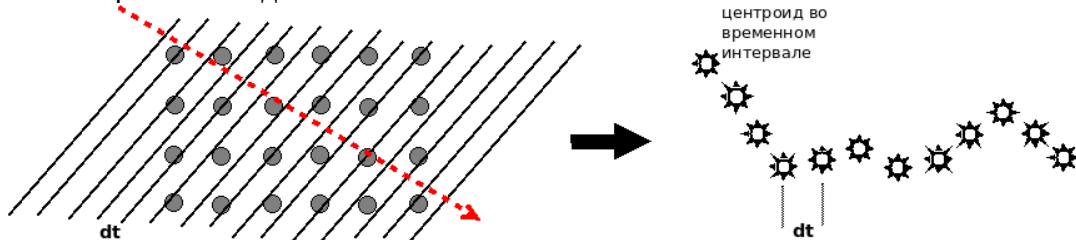
Планируется выполнить следующие работы:

1. выполнить **аналитический обзор**, включая:
 - a) выделение и отбор существенных признаков для наборов эксп. данных;
 - b) совместный анализ разнородных данных, поступающих из нескольких источников, включая, но не ограничиваясь различные типы АЕ и CNN;
 - c) обработки данных наземных экспериментов в области астрофизики частиц;
2. создать **выборки данных** на основе методов Монте-Карло для тестирования разрабатываемых методов, алгоритмов и их программных реализаций;
3. подготовить и обосновать список **критериев** эффективности совместной обработки данных;
4. **теоретическое** исследование/отбор **методов ML** для выделения и отбора **существенных признаков** для наборов данных;
5. адаптировать существующие и/или **разработать** новые **методы** и алгоритмы выделения и отбора **существенных признаков**;
6. создать **нейросетевые модели** для выделения и отбора существенных признаков для наборов данных;
7. разработать методики **интерпретации признаков** в терминах предметной области;
8. теоретическое исследование/отбор **методов ML для совместного анализа** разнородных данных
 - a) особое внимание должно быть уделено совместному анализу данных **изображений и пространственно-временных данных**;
9. доклады на международные конференции и **2 статьи** WoS/Scopus/RSCI.

План "light" А: существенные признаки

1. Выделение/отбор существенных признаков для данных TAIGA-HiSCORE
 - a) учет географического положения станций (стационарные данные, т. е. не input)
 - b) моменты? (аналог параметров Хилласа???)
2. Кластеризация по времени??
 - a) географические центры кластеров — динамические величины

Иллюстративная идея



- b) положения центроидов = входные данные для определения направления прихода
 - c) ближе к "изображению"
 - d) распределение интенсивностей без учета времени ↔ энергия (??)
3. данные TAIGA-HiSCORE ↔ AE ???
 4. замечание: в литературе "Event data", нет общей методологии. Пример: Crime Analysis.

План "light" Б: перебор вариантов нейросетей

1. Примитивный (на уровне score или усреднение энергий)

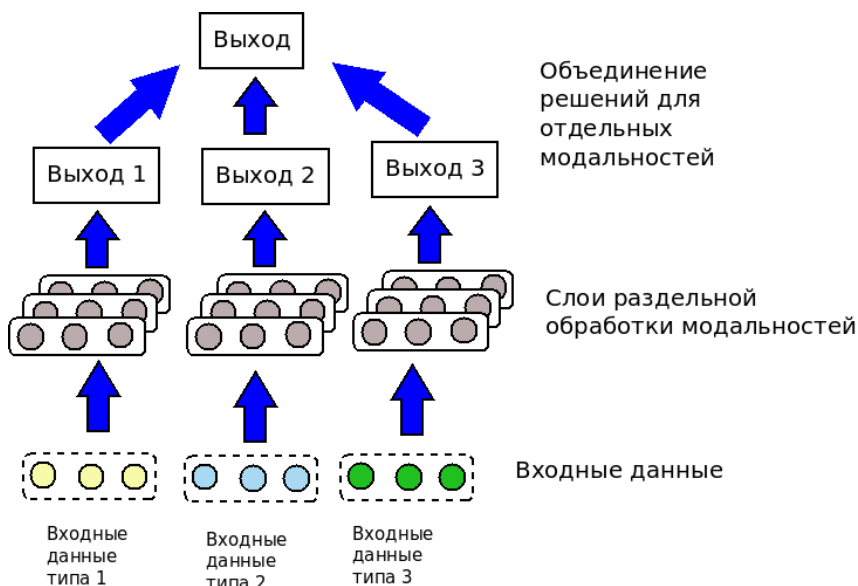


Рисунок 1: В подходе позднего слияния объединяются решения по подмоделям для каждой модальности.

2. На уровне классификатора

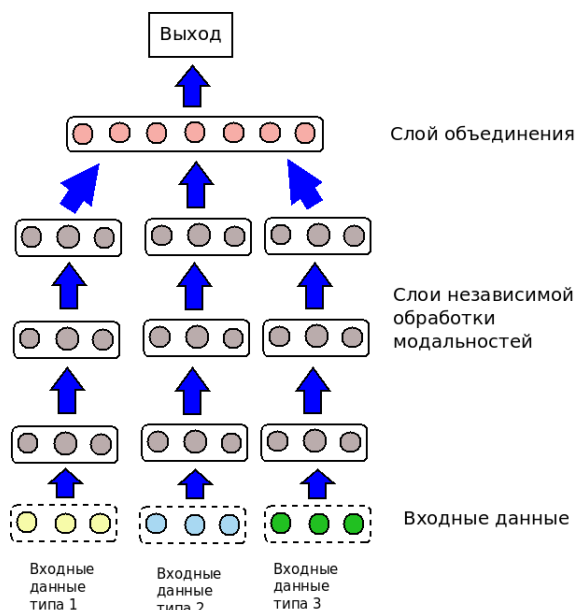


Рисунок 2: Возможный вариант промежуточного подхода к объединению: сначала модальности обрабатываются независимо, а затем выход совместного слоя непосредственно вводится в функцию принятия решения.

3. Промежуточный слой объединения

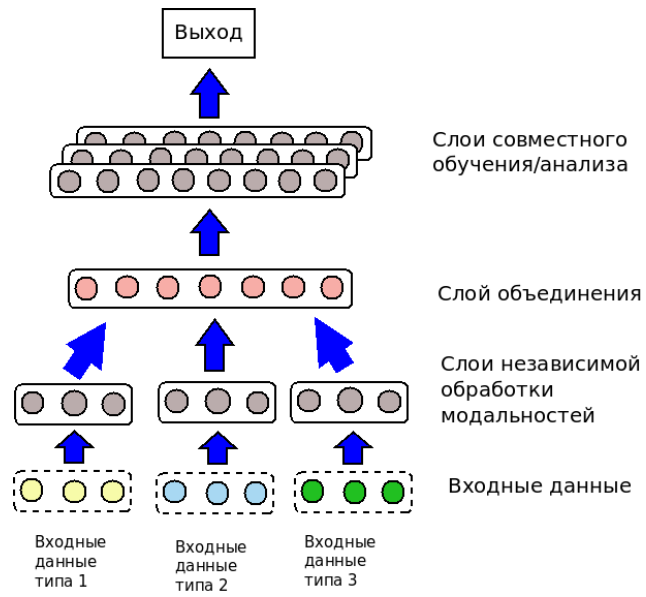


Рисунок 3: Возможный вариант промежуточного подхода к объединению: сначала модальности обрабатываются независимо, а затем - совместно