

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра динамического моделирования и биомедицинской инженерии

Авторегрессионный анализ ЭЭГ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 461 группы
направления 12.03.04 Биотехнические системы и технологии
факультета nano- и биомедицинских технологий

Горбачева Дмитрия Андреевича

Научный руководитель

доцент, к.ф.-м.н

должность, уч. степень, уч. звание

И. В. Сысоев

подпись, дата

инициалы, фамилия

Зав. кафедрой, д.ф.-м.н., доцент

должность, уч. степень, уч. звание

Е.П. Селезнев

подпись, дата

инициалы, фамилия

Саратов 2019

Введение

Актуальность выпускной квалификационной работы. Метод энцефалографии это один из наиболее часто применяемых методов для изучения мозга человека. Электроэнцефалография дает возможность качественно и количественно анализировать функциональное состояние головного мозга и его реакцию при действии раздражителей и при выполнении различных видов деятельности. В клинической практике ЭЭГ используется для диагностики ряда психических и неврологических заболеваний. Существует ряд характерных измерений ЭЭГ, специфичных для определенного заболевания, зная которые, можно своевременно выявлять признаки болезней и лечить пациентов.

Данная работа направлена на изучение возможности создания авторегрессионных моделей и их тренировки по данным энцефалограммы для определения активности головного мозга объекта в текущий момент. В нашем случае модель тренировалась и тестировалась по данным энцефалограммы мышей во время медленного сна. В работе описан процесс создания моделей, при помощи которых можно будет установить активность головного мозга. В дальнейшем, развивая эту тему, можно изучить возможность применения авторегрессионной модели для определения состояние объекта в недалеком будущем. Например, натренировав математическую модель на объектах перед эпилептическим припадком, исследовать точность определения вероятности возникновения припадка у других объектов по данным энцефалограммы в реальном времени.

Цели и задач выполняемой выпускной квалификационной работы. Основной задачей в этой работе было построение математической модели по данным энцефалограммы мыши во время медленного сна и проверка этой модели на других мышах в том же состоянии, чтобы выявить возможность применения авторегрессионных моделей для определения функционального состояния мышей. Для этого были поставлены следующие цели:

- Построить авторегрессионную модель.
- Написать программу для считывания данных ЭЭГ в заданном промежутке времени.

- Написать программу с реализацией поиска коэффициентов модели по данным ЭЭГ мышей.
- Оптимизировать поиск коэффициентов, убрать лишние, которые не вносят существенной пользы.
- Построить прогноз энцефалограммы по полученной сокращенной модели для нескольких мышей по данным во время медленного сна и в отличное от сна время.
- Рассчитать ошибку для полученных энцефалограмм, сравнивая с эталонными данными, на основе которых строилась ЭЭГ.

Характеристика материалов исследования. Для построения и обучения авторегрессионной модели были использованы записи ЭЭГ 3 взрослых мышей во время медленного сна: V16-5926, V16-5927 и Trip8b-215. Данные поступали с частотой дискретизации 2083.(3) Гц от трех датчиков:

- VPM - датчика подключенный к таламусу.
- SoSeCx - датчик подключенный к отделу коры головного мозга, принимающему данные с таламуса.
- MoCx - датчик подключенный к отделу коры головного мозга, отвечающему за принятие решений.

Описание структуры выпускной квалификационной работы. Выпускная квалификационная работа состоит из введения, основной части, содержащую 2 главы, заключения и списка используемой литературы, состоящей из 10 ссылок. Работа изложена на 32 страницах, содержит 15 рисунков.

Основное содержание работы

В главе 1 сделан литературный обзор регрессионных моделей, объяснена возможность применения таких моделей для описания случайных процессов, выявлены основные формулы авторегрессионных моделей и показаны основные принципы подбора формул для анализа временных рядов экспериментальных данных. Имея представления о свойствах случайного процесса и его характеристиках, можно подобрать необходимую авторегрессионную модель с нужным порядком для описания этого процесса и его дальнейшего анализа.

Авторегрессионная модель — математическая формула, в которой рассчитываемый элемент числового ряда зависит от суммы предыдущих элементов. Количество предыдущих элементов d является порядком модели. Эта модель является одной из наиболее популярных конструкций при моделировании по наблюдаемым временным рядам.

Описаны экспериментальные данные энцефалограммы. Запись ЭЭГ — суммарная электрическая активность миллионов нейронов, представленной преимущественно потенциалами дендритов и тел нервных клеток: возбуждательными и тормозными постсинаптическими потенциалами и частично - потенциалами дендритов и тел нервных клеток. То есть ЭЭГ позволяет получить визуализированный результат функциональной активности головного мозга. Объяснены значения сигналов, которые поступают с датчиков:

- Сигналы с таламуса, получающего импульсы всех видов чувствительности, кроме обонятельных, и передающего их в кору больших полушарий.
- Сигналы с отдела коры головного мозга, принимающего данные с таламуса.
- Сигналы с отдела головы головного мозга, формирующего ответный импульс на внешние раздражители по его обработки нейронами.

Выявлено, что из-за невозможности на данный момент однозначно рассчитать значение данных от энцефалограммы в конкретный момент времени, экспериментальный материал можно считать случайным процессом. Обусловлена возможность его анализа при помощи регрессионных моделей.

Показаны основные этапы построения и улучшения математических моделей:

1. Оценка экспериментальных данных и результатов измерений, постановка цели моделирования.
2. Формирование содержательной модели.
3. Выбор наиболее подходящего математического аппарата, вид и числ уравнений, вид функции.
4. Подгоняются коэффициенты уравнений.
5. Проверка качества получившейся модели.

Установлена основная формула авторегрессионной модели

$$x_{n+1} = \sum_{i=0}^{d-1} c_i x_{n-i}, \quad (1)$$

которая использована для описания данных ЭЭГ. В этой формуле x_{n+1} — прогнозируемое значение, x_{n-i} — предыдущие значения, c_i — коэффициенты модели, d — порядок модели.

Выявлено, что для нахождения коэффициентов c_i из формулы (1) необходимо решить систему линейных уравнений методом Гаусса, который обычно записывается следующим образом:

$$Ax = b, \quad (2)$$

где A - матрица параметров, b - вектор известных значений, x - вектор искомых значений.

Метод Гаусса реализуется по следующему принципу:

- 1 На первом этапе осуществляется прямой ход. Путем элементарных преобразований над строками систему приводят к ступенчатой или треугольной форме. Среди элементов первого столбца матрицы выбирают ненулевой, перемещают его на крайнее верхнее положение перестановкой строк и вычитают получившуюся после перестановки первую строку из остальных строк, домножив её на величину, равную отношению первого элемента каждой из этих строк к первому элементу первой строки, обнуляя тем самым столбец под ним. После того, как указанные преобразования были совершены, первую строку и первый столбец мысленно вычеркивают и продолжают пока не останется матрица нулевого размера. Если на какой-то из итераций среди элементов первого столбца не нашёлся ненулевой, то переходят к следующему столбцу и проделывают аналогичную операцию.
- 2 На втором этапе осуществляется так называемый обратный ход, суть которого заключается в том, чтобы выразить все получившиеся базисные переменные через небазисные и построить фундаментальную систему решений, либо, если все переменные являются базисными, то выразить в численном виде единственное решение системы линейных уравнений. Эта процедура начинается с последнего уравнения, из которого выражают соответствующую базисную переменную и подставляют в предыдущие

уравнения, и так далее до первого уравнения. Каждой строчке соответствует ровно одна базисная переменная, поэтому на каждом шаге, кроме последнего, ситуация в точности повторяет случай последней строки.

В конце главы была изложена реализация построения авторегрессионной модели для энцефалограммы. Была написана программа для расчета параметров c_i уравнения (1) методом Гаусса:

- Составлялось M линейных уравнений (1) с возможностью указания порядка d модели по данным энцефалограммы x_n с неизвестными коэффициентами c_i . При чем M много больше d .
- При решении системы этих линейных уравнений методом Гаусса находились коэффициенты модели порядка d .
- Для коэффициентов указанного порядка рассчитывалась среднеквадратичная ошибка по формуле:

$$S^2 = \frac{1}{M-d} \frac{\sum_i (x_i - \hat{x}_i)^2}{G^2}, \quad (3)$$

где \hat{x}_i — рассчитанное значение по параметрам модели, G^2 — дисперсия находящаяся по формуле:

$$G^2 = \frac{1}{M-1} \sum_{i=1}^d (x_i - \bar{x})^2, \quad (4)$$

где \bar{x} — среднее арифметическое по выборке. Значение \hat{x}_i находится по формуле (1).

Этот алгоритм использовался при расчетах для 3 мышей с 3 каналов данных энцефалограммы во время медленного сна. Все расчеты коэффициентов порядка d и их ошибки сохранялись в файл для дальнейшего анализа с указанием порядка, названия наблюдаемого объекта и канала, с которого считывались данные.

Для дальнейшего анализа параметров и тестирования верности расчетов нахождения коэффициентов c_i был реализован модуль, который вычислял x_{n+1} значения по материалам ЭЭГ с использованием этих коэффициентов. Из данных, на основе которых строится энцефалограмма, берется количество d (по порядку модели) предыдущих значений относительно текущего и по формуле (1) находится x_{n+1} значение. Таким образом на основе этого алгоритма можно

построить рассчитанный по коэффициентам авторегрессионной модели прогноз энцефалограммы.

Так же этот функционал использован для тестирования достоверности расчетов параметров. Для этого указываются начальные данные x_0 , x_1 , константы c_0 , c_1 , длина генерируемого сигнала N , порядок модели D и дисперсия шума, при формировании сигнала. По формуле (1) рассчитываются значения x_i на основе начальных данных. Далее по сгенерированному сигналу находятся параметры c_i . После параметры c_i , найденные по алгоритму для расчета, сравниваются с указанными в начальных условиях. Для показателей находится квадратичная ошибка

Ниже представлены результаты тестов.

#	x_0	x_1	c_0	c_1	D	N	Дисперсия шума
1	2	6	0.8	-0.15	2	5	0
2	2	6	0.8	-0.15	2	5	0.01
3	2	6	0.8	-0.15	2	5	0.1
4	0.05	0.1	0.02	0.001	2	5	0.00001

Таблица 1.1 Начальные данные для тестирования расчетов параметров авторегрессионной модели

#	c_0	c_1	Квадратичная ошибка
1	0.8	-0.15	$2.33 * 10^{-32}$
2	0.79980587	-0.14860556	$9.57079477 * 10^{-6}$
3	0.79950399	-0.14430346	0.0002121
4	0.01992553	0.0010275	$3.57383308 * 10^{-8}$

Таблица 1.2 Результаты тестирования расчета параметров авторегрессионной модели

Как видно по результатам тестов, разработанный алгоритм нахождения параметров модели позволяет достаточно точно находить их. Таким образом, была разработана и оттестирована программа для расчета параметров авторегрессионной модели (1). Реализован расчет прогноза по этой формуле.

В главе 2 изложены способ построение и способы оптимизации авторегрессионной модели. При формировании математической модели всегда необходимо определится количество свободных членов K уравнения для описания процесса. Очень малый порядок многочлена зачастую не годится, Так как не дает возможности с хорошей точностью аппроксимировать сложные зависимости.

Такие модели называются недообученными. Другим крайним случаем является переобучение модели. Такая ситуация может возникнуть если число коэффициентов K равно числу точек N рассматриваемого ряда. Такая модель воспроизводит только наблюдаемую реализацию со всеми ее шумами. Из-за этого она будет очень плохо предсказывать в среднем новые наблюдения, так как при новых наблюдениях конкретные значения помехи будут другие.

Для исключения коэффициентов использован критерий наличия одинакового знака у коэффициентов. Для поиска коэффициентов с одинаковыми знаками использованы данные мыши В16-5926. Для каждого канала записи медленного сна рассчитаны коэффициенты математической модели (1) для ее порядков d от 0 до 300 по 20 записям. Найдены среди 20 записей сна коэффициенты с одинаковыми знаками. Далее были сохранены индексы массива этих коэффициентов.

Таким образом модель оптимизирована и подстроена под описание процесса медленного сна для одной мыши. Для дальнейшего обучения модели необходимо для второй мыши В16-5927 произвести те же расчеты что и для В16-5926. И по установленному критерию найти коэффициенты с одинаковыми знаками уже среди усредненных параметров модели у первой и второй мышей. Необходимо сохранить порядковые номера найденных коэффициентов с одинаковыми знаками.

Имея сокращенную математическую модель, можно построить по исходным данным ЭЭГ прогнозируемую энцефалограмму. Для этого используем уже рассчитанные коэффициенты модели для записи сна наблюдаемой мыши и ее исходный материал. По формуле (1) для расчета x_{n+1} элемента используем коэффициенты модели мыши, индексы которых совпадают с индексами нашей сокращенной модели. То есть, мы будем использовать не d коэффициентов из модели для этой мыши, а k коэффициентов, чьи индексы совпадают с индексами сокращенной модели. Коэффициенты, которые не попадают под это условие, примем за 0.

Исходные данные ЭЭГ наблюдаемых объектов для расчетов будут браться по следующему принципу. Берется любая из 20 записей энцефалограммы за 10 секунд до начала сна, которое устанавливается по маркерам к экспериментальным данным, 10 секунд во время медленного сна и 10 секунд после этого промежутка.

В итоге имеется запись ЭЭГ длительностью 30 секунд и частотой дискретизации 2083.(3) Гц, в которую входит эпизод со сном. По этой записи с применением сокращенной модели будут рассчитаны все x_{n+1} элементы энцефалограммы. Длительность рассчитанной энцефалограммы тоже составит 30 секунд. Ключевые точки $n = 20833$ — начало записи сна и $n = 41666$ — окончание записи сна.

Далее необходимо сравнить исходные данные ЭЭГ и рассчитанный прогноз, который будет получен из общей сокращенной модели. Для всех элементов x_n прогнозируемой энцефалограммы рассчитывается квадрат ошибки:

$$E^2 = (x - \bar{x})^2, \quad (5)$$

где x - элемент исходной выборки, \bar{x} - элемент прогнозируемой ЭЭГ.

Были построены графики зависимости квадратичной ошибки от индекса элемента в данных ЭЭГ для мышей B16-5926 и B16-5927, у которых материалы записей сна использовались для оптимизации модели, и для мыши Trip8b-215, чьи данные для оптимизации не использовались.

Зависимость квадрата ошибки прогнозируемых значений от индекса ряда энцефалограммы по данным ЭЭГ мыши B16-5926-sleep

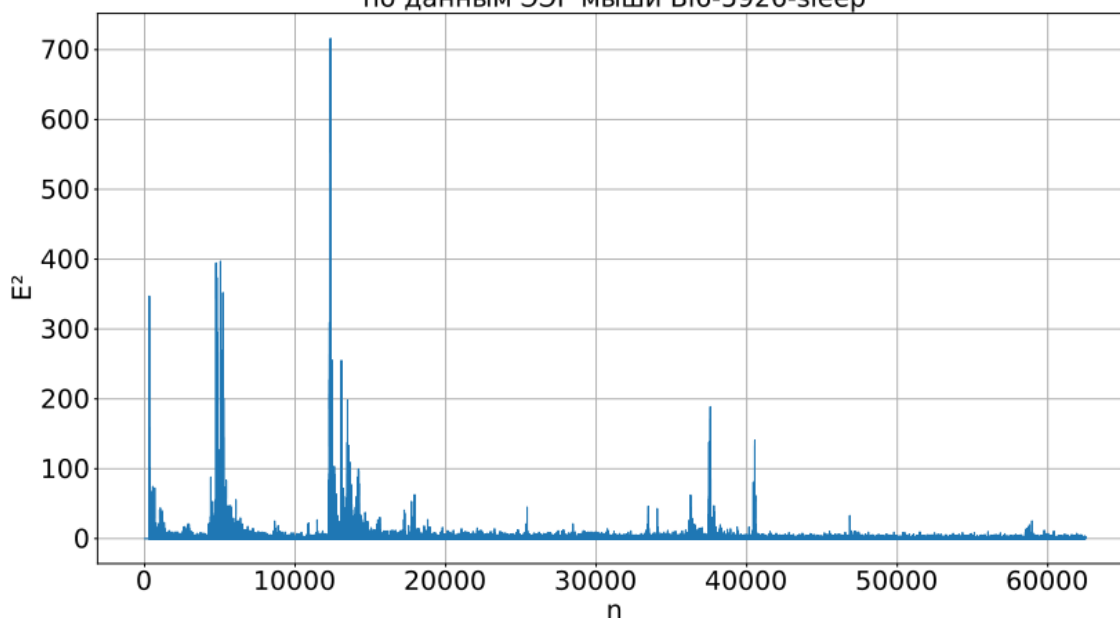


Рисунок 1 График зависимости квадрата ошибки прогнозируемых значений ЭЭГ E^2 по формуле (5) для всех прогнозируемых значений x_n по индексам n . Прогноз рассчитывался по формуле (1) с использованием коэффициентов C_i , индексы которых совпадали с индексами сокращенной модели. Прогноз производился по данным записи ЭЭГ мыши B16-5926 с канала SoSeCx. До момента $n = 20833$ мышь была в отличном от сна состоянии, в промежутке от 20833 до 41666 был медленный сон, после 41666 — неизвестное состояние

По рисунке 1 видно наличие больших ошибок перед началом записи сна и конца. Присутствуют небольшие ошибки во время записи сна. В целом момент сна установлен верно. По данному графику можно отличить момент записи сна.

Зависимость квадрата ошибки прогнозируемых значений от индекса ряда энцефалограммы по данным ЭЭГ мыши B16-5927-sleep

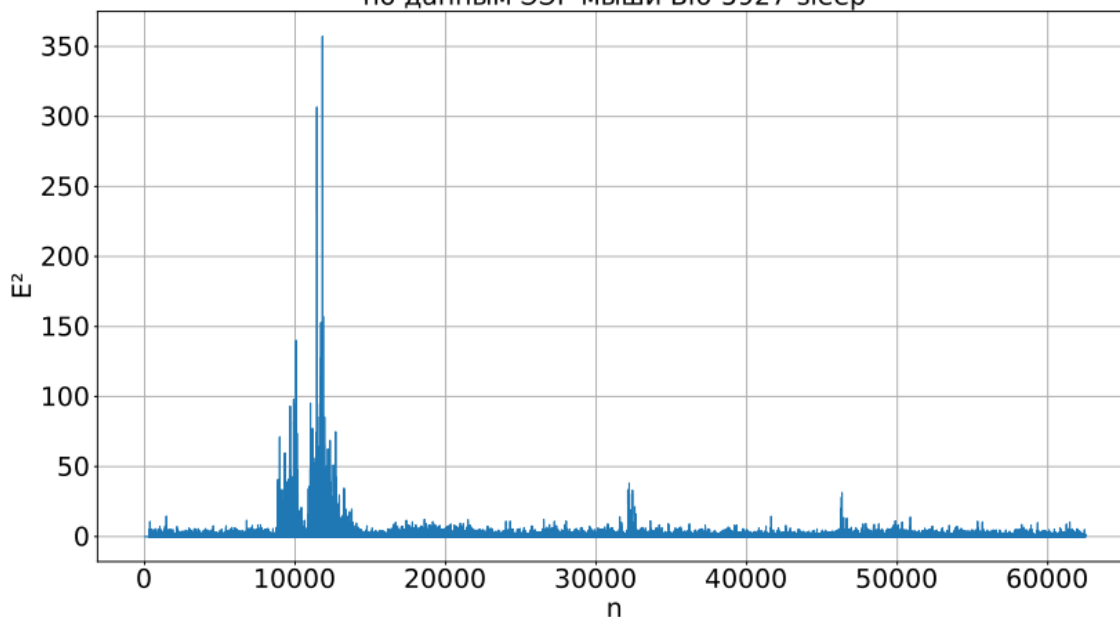


Рисунок 2 График зависимости квадрата ошибки прогнозируемых значений ЭЭГ E^2 по формуле (5) для всех прогнозируемых значений x_n по индексам n . Прогноз рассчитывался по формуле (1) с использованием коэффициентов c_i , индексы которых совпадали с индексами сокращенной модели. Прогноз производился по данным записи ЭЭГ мыши B16-5927 с канала SoSeCx. До момента $n = 20833$ мышь была в отличном от сна состоянии, в промежутке от 20833 до 41666 был медленный сон, после 41666 — неизвестное состояние

На рисунке 2 Можно заметить наличие больших ошибок перед началом записи сна. Присутствуют небольшие ошибки во время записи сна. В целом момент сна установлен верно. По данному графику можно отличить момент момент записи сна.

Зависимость квадрата ошибки прогнозируемых значений от индекса ряда энцефалограммы по данным ЭЭГ мыши Trip8b-215-sleep

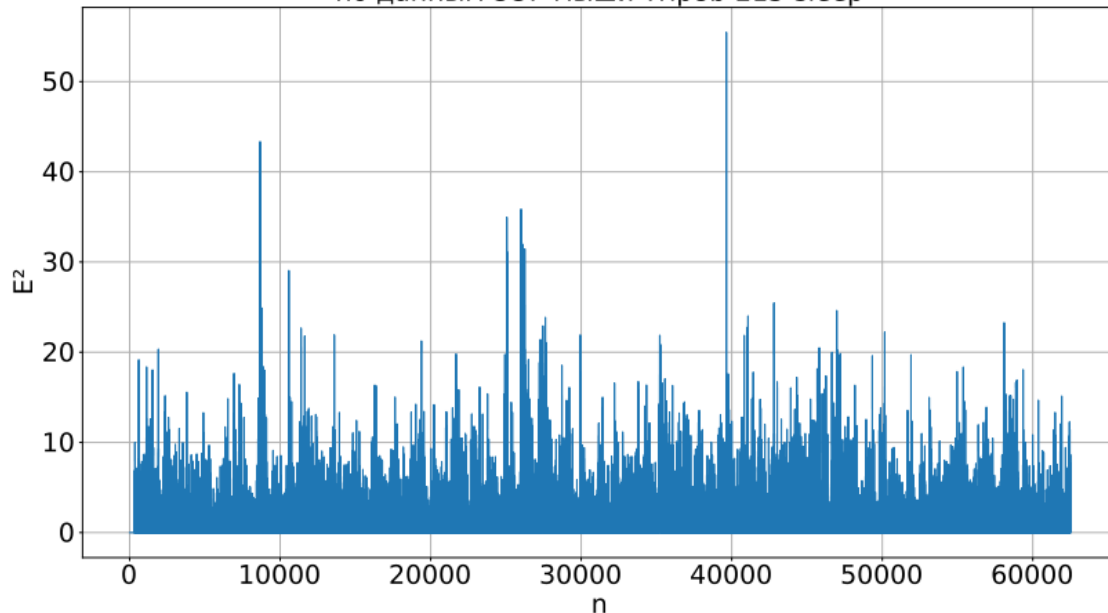


Рисунок 3 График зависимости квадрата ошибки прогнозируемых значений ЭЭГ E^2 по формуле (5) для всех прогнозируемых значений x_n по индексам n . Прогноз рассчитывался по формуле (1) с использованием коэффициентов c_i , индексы которых совпадали с индексами сокращенной модели. Прогноз производился по данным записи ЭЭГ мыши Trip8b-215 с канала SoSeCx. До момента $n = 20833$ мышь была в отличном от сна состоянии, в промежутке от 20833 до 41666 был медленный сон, после 41666 — неизвестное состояние

На рисунке 3 видно наличие больших ошибок во время всей записи ЭЭГ мыши. Определить в какой момент времени мышь находилась в состоянии медленного сна невозможно.

Заключение

В данной работе были рассмотрены основные принципы и этапы построения авторегрессионных математических моделей для случайного процесса, способы и необходимость оптимизации моделей. Для анализа были использованы записи электроэнцефалограммы сна взрослых мышей, для которых подбирались, строились и анализировались авторегрессионная модель.

Была написана программа которая позволяет считывать данные энцефалограммы в задаваемые моменты времени, реализован расчет параметров модели с заданным порядком, расчёт ошибки прогноза и разброса оценок коэффициентов модели. Подобран и реализован критерий для оптимизации модели. Реализовано формирование данных энцефалограммы по исходным

материалами и сокращенной модели. Написана оценка точности формирования ЭЭГ относительно оригинальных данных.

Удалось построить оптимизированную модель по данным 2 мышей. Было показано, что при использовании этой модели для определения активности мозга на мышах, по которым эта модель обучалась, можно по данным ЭЭГ определить в каком физиологическом состоянии находился объект.

При анализе мыши, материалы которой не были задействованы при обучении математической модели, не удалось определить ее состояние по записи энцефалограммы. Из чего следует, что модель нуждается в дальнейшей анализе, обучении. Следует проверить возможности обучения модели по нескольким мышам и определить оптимальное количество этих мышей. После по полученным данным проверить достоверность применяемого критерия оптимизации, модернизировать его.